Fusion of heterogeneous data in non-destructive testing and structural health monitoring using Echo State Networks

Adam J. Wootton

A Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Keele University

June 2018

Abstract

Failure to monitor the condition of key infrastructure such as roads and bridges can result in costly closures, but the economic impact could be lessened by early intervention. Non-destructive testing (NDT) examines structures without causing damage, while structural health monitoring (SHM) monitors a structure throughout its life. This thesis presents a machine learning approach to fusing heterogeneous sensor modalities that can be systematically applied to improve sensor interpretation and reduce reliance on expertise. For the first time, echo state networks (ESNs) were used in two separate NDT and SHM data fusion case studies.

The NDT-based study looked at detecting defects in steel reinforcement, teaching ESNs to combine magnetic flux leakage (MFL) and cover depth data in order to compensate for variation in MFL amplitude with increasing cover depth. Using seven different cover depths between 42.5 mm and 289 mm, the fusion approach offered improved performance for $42.5\,mm < depth < 205\,mm$ and the most consistent calculated optimal output threshold, demonstrating the ease of systematic application.

In the SHM-based study, data from the National Physical Laboratory (NPL) footbridge monitoring project was processed by a suite of ESNs to detect, localise, classify and assess damage caused by deliberate interventions. A novel approach of combining physical and environmental sensors in order to model a different modality of physical sensor made it possible to use the residual to observe damage trends and locations, which also led to the isolation of a faulty strain gauge. There was additional success in distinguishing between different intervention types and producing a metric to express the damage level.

Across both studies, the ESN approach to heterogeneous data fusion improved upon non-fusion-based alternatives. This suggests that future work should consider structures that are in regular use, combining further sensor modalities and the development of bespoke data fusion software.

Acknowledgements

The analytical technique used as a comparison in Chapter 3 was developed by Dr James Austin of Keele University and Emmanuel Mahé of L'École Supérieure de Céramique Industrielle, Limoges. I would like to extend grateful thanks to The UK National Physical Laboratory (NPL) for funding and conducting the NPL footbridge project, particularly Dr Elena Barton and Dr John Blackburn, who provided the resultant data. The original leaf sample spectroradiometry dataset used in Appendix B was provided by Dr Sarah Taylor and Dr Austin Haffenden of Keele University.

Over the course of this project, I have been lucky enough to receive help, guidance and distraction from a great many people, to whom I am incredibly thankful. First of all, I owe a huge debt of gratitude to my two doctoral supervisors, Professor Peter Haycock and Dr Charles Day. I could not have asked for two more attentive and supportive supervisors, and I would particularly like to thank Peter for having given me the opportunity to undertake a PhD in the first place. Their help throughout my work has been invaluable, and the lessons that I have learnt from them both will stand me in good stead for the rest of my career. I would also like to thank John Butcher for his guidance on the RC toolbox.

I would like to thank Bob Rennison, formerly of SciCorr, for giving me the chance to work on the development of the EMAD technique and the experience of performing on-site surveys. Thanks are also due to those associated with SciCorr who gave me help and insights into the application of the EMAD technique in industry: Roger Dale, Tom Moore, Rob Painting, Danny Westhorpe and Mike Thomas.

I am also grateful to those fellow researchers at Keele who have had the dubious honour of sharing an office with me. Louis Major passed on some valuable words of

wisdom while showing me the ropes in my first few months as a PhD student, while my semi-regular lunches at the KPA with Chris Marshall were usually a welcome relief from ESN optimisation and data analysis. Thanks also to George Joseph, Mutiq Almutiq, Ryad Soobhany, James Rooney, Geethamala Francisco, Kaseem Olorisade, Rafael Valero-Fernandez and Amy Martin for making our various offices such enjoyable places to work.

I have been lucky enough to receive assistance from a number of different people at Keele University. Alan Dudley was good company whenever I had to work on the concrete test bed and gave me good advice in my early days as a PhD student, while Phil Callaghan and Steve Wye were always on hand to provide experimental equipment. The technical staff at the School of Computing and Mathematics, Ian Marr, Tom Cooper, Steve de Cruz and Ash Leake provided me with all of the computing support that I could ask for whenever I needed software or yet another new computer. I would like to thank all of my colleagues at the Foundation Year, but would especially like to mention Libby Smith for all of her help with contracts and paperwork.

I would never have been able to undertake this PhD without the support of my parents, who I can never thank enough. Lunches and gaming sessions with my younger brother, Peter, provided the chance to unwind and talk through any problems that I was having. I hope that you gain as much from your Mathematics PhD as I have gained from my PhD. Clare, my fiancée, has kept me sane over the last three years with her love and support. I can't wait to begin our future together.

Contents

List of Figures

List of Tables

Glossary

AFM Active Field Measurement.

ANN Artificial Neural Network.

AT Analytical Technique.

AUC Area Under Curve.

BSN Bridge-based Sensor Network.

CI Computational Intelligence.

ELM Extreme Learning Machine.

EMAD Electromagnetic Anomaly Detector.

ESN Echo State Network.

FPR False Positive Rate.

GPR Ground-Penetrating Radar.

JDL Joint Directors of Laboratories.

LSM Liquid State Machine.

MCC Matthews Correlation Coefficient.

MFL Magnetic Flux Leakage.

MLP Multilayer Perceptron.

NDT Non-Destructive Testing.

NPL National Physical Laboratory.

NPV Negative Predictive Value.

PCC Pearson Correlation Coefficient.

PPV Positive Predictive Value.

Rebar steel reinforcing bar.

RM Remanent Magnetism.

RNN Recurrent Neural Network.

ROC Receiver Operator Characteristic.

SHM Structural Health Monitoring.

SVM Support Vector Machine.

TDNN Time-Delay Neural Network.

WDBC Wisconsin Diagnostic Breast Cancer.

CHAPTER 1

Introduction

1.1 The Cost of Failing Infrastructure

In the World Economic Forum's 2016-17 Global Competitiveness Report, the quality of the United Kingdom's roads was ranked as 27th in the world, not only behind France, Germany and the United States, but also Namibia, Ecuador and Oman [1]. This follows on from a ranking of 29th in 2015-16 [2], 30th in 2014-15 [3], 28th in the 2013-14 [4] and 24th in 2012-13 [5]. In 2013, an HM Treasury policy paper stated that £10 billion would need to be spent on repairs to roads by 2021 [6], while the Highways Agency, which is responsible for only 2.4% of roads in England, spent £288,596,000 on road repairs in the period 2014-15 [7]. The situation is hardly better elsewhere. In the United States alone, the cost of repairing and upgrading deficient bridges is estimated to be $123 billion in 2017 [8], while the deteriorating infrastructure is expected to cost the US economy an average of 400,000 jobs per year up until 2040 [9]. Indeed, in 2014, the US Federal Highway Administration found that 24% of bridges in the US were structurally deficient or functionally obsolete [10]. A technique capable of detecting the early onset of damage, so as to prevent costly repairs at a later date, would clearly be of great benefit.

There have been several high-profile bridge collapses in recent years, each of which calls into question the efficacy of current approaches to monitoring the health of the built environment. On the 10th of January 2016, after just 42 days of use, the $100,000,000

Nipigon River bridge in Canada failed [11]. An independent investigation found that a progressive fracture of 40 bolts was responsible for the failure [12]. In the US, the collapse of the Tex Wash bridge in July 2015 led to 100 miles of Interstate 10 being closed for five days at an estimated cost to the economy of $1,500,000 per day [13]. The US Federal Highway Administration had given the bridge an 'A' grade and a sufficiency rating of 91.5 out of 100 just twelve months previously [14], which is particularly concerning given that 6,428 other bridges in California were deemed deficient in the same set of inspections [10]. Poor infrastructure maintenance has also led to tragic consequences, as 65 people were killed in Guinea when a bridge collapsed under the weight of a truck in 2007 [15]. The 2007 collapse of the I-35W bridge in Minneapolis (shown in Figure 1.1), caused by gusset plates in the bridge reaching their yield limit, resulted in 13 fatalities and 145 injuries [16]. Clearly, more work needs to be done to provide an accurate assessment of the condition of the infrastructure, so that failures such as these can be prevented.

Figure 1.1: The aftermath of the I-35W bridge collapse in 2007 (from Rofidal [17]).

In 2013, Vanniamparambil et al. identified one of the chief problems facing structural

engineers: while many different modalities of sensor are available for the assessment of many different aspects of structural integrity, there is no single monitoring technique that can provide a comprehensive picture of the state of a structure [18]. They concluded that a data fusion approach should be developed in order to detect deterioration signs, locate weak regions and assess damage severity. White's Data Fusion Lexicon defines data fusion as 'a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats as well as their significance' [19]. It aims to overcome the limitations of individual sensors in order to provide a more accurate and reliable estimate of the world-state [20]. This thesis demonstrates how Computational Intelligence (CI) can be used to work towards this goal, as heterogeneous data have been fused for damage detection in two separate case studies. The first of these case studies considered defect detection in reinforced concrete structures, while the second looked at online damage detection for a 'real-world' footbridge.

1.2 An Introduction to Echo State Networks

One flavour of Artificial Neural Network (ANN) is particularly suited to these two case studies: the Echo State Network (ESN). ESNs are a relatively recent addition to the Recurrent Neural Network (RNN) class of techniques, and were designed to be able to process time-series data. Furthermore, they are efficient to train and have been successfully applied in similar case studies in the past. They were developed with an engineering perspective in mind, and have proven to be more effective than traditional ANN methods at predicting chaotic time-series data, while also having a much shorter training time [21]. The two case studies added an extra layer of complexity to the data fusion task by using longitudinal data with a spatial or temporal extent. Any processing technique would need to not only model the relationship between the input and output data, but how changes in the input data over several consecutive datapoints should affect the output data. One of the most useful aspects of ESNs is their ability to recall

past inputs through presence of a short term memory, which can be controlled by a careful tuning of the network parameters [22]. These ESN architecture features made them a particularly good fit for the heterogeneous data fusion approaches presented in this thesis.

Alongside Liquid State Machines (LSMs) [23], ESNs belong to the field of Reservoir Computing [24]. However, an LSM's spiking neurons are significantly more difficult to tune to a desirable dynamic range, making them less well-suited to practical applications than ESNs, which are generally easier to configure [25]. The general ESN architecture has three principal features. The first of these is an input layer of neurons which, via a fully connected weight matrix, projects into the second feature, a typically much larger, high dimensional 'kernel' of sparsely interconnected reservoir neurons. Each of the reservoir neurons is in turn fully connected to a layer of output neurons. A schematic diagram of this architecture can be seen in Figure 1.2. In addition to these connections, it is also possible to connect the input units to the output units directly, or to have the output units feed back into the reservoir.

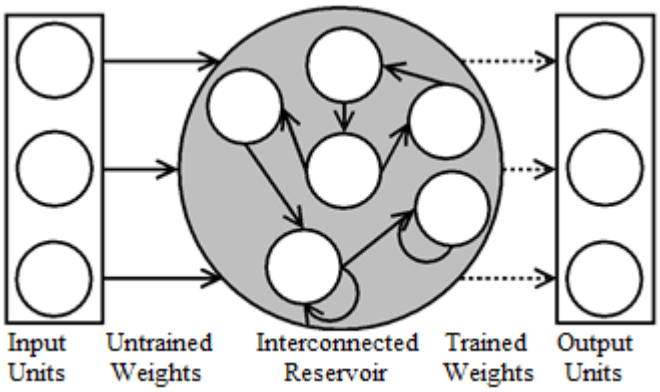


Figure 1.2: A typical ESN topology. Only the weights between the reservoir neurons and the output units are trained.

Since their inception in 2001, ESNs have become popular due to their modelling capacity, modelling accuracy, biologically plausible recurrence, extensibility and ability to overcome the vanishing gradient problems traditionally associated with gradient descent RNN training procedures [26]. This is achieved by keeping most of the weighted connections between neurons unchanged during training, whereas other ANN approaches require all of these connections to be trained. Indeed, only the weighted connections

between the reservoir neurons and the output neurons are trained; all other weighted connections are randomly generated at network initialisation and left unchanged throughout. Ridge regression [27] has been found to be a good approach to training the weighted connections between the reservoir and output neurons [28–30].

A detailed description of ESNs, including a review of recent advances and applications, can be found in Chapter 2.

1.3 Non-Destructive Testing of Reinforced Concrete

The most straightforward method for assessing the health of reinforced concrete structures, such as a bridge deck, is a visual inspection. In such an inspection, an expert would survey the concrete surface for visible indicators of corrosion or deterioration, such as cracking, small cavities ('pitting') and depressions ('spalling') [31]. While it is the most popular type of inspection (Moughty and Cases noted in 2017 that the majority of bridge data is obtained this way [32]), it is dependent on both the competence of the inspector and signs of corrosion on the surface [33]. This is especially problematic, since it is possible that there may be significant deterioration inside the concrete that cannot be observed on the surface [34]. Consequently, a visual inspection is likely to underestimate the full extent of the corrosion of the steel reinforcing bars (rebars) [35]. Both underestimation and overestimation can have negative consequences, since the former could lead to significant problems going unrepaired, while the latter could prompt potentially costly and unnecessary repairs to be performed. Visual inspection is, therefore, an insufficient method for the full evaluation of the level of corrosion in a reinforced concrete structure. One outcome of this shortcoming is the development of Non-Destructive Testing (NDT) techniques, which are capable of interrogating the rebar for information on the level of corrosion, without requiring the prior removal of the concrete. A successful, systematically applied computational approach to processing the data from these NDT techniques would overcome the variation in interpretation by individual inspectors.

This section details some of these NDT approaches, including the Electromagnetic

Anomaly Detector (EMAD) technique, which is the focus of the first case study, found in Chapter 3.

1.3.1 Corrosion of steel in reinforced concrete

The work performed in Chapter 3 looked at ways of fusing NDT techniques for the detection of corrosion in rebar. The mechanisms by which steel rebar can be corroded in concrete are well understood, and detailed guides to these have been published by Tuutti [36], Broomfield [37] and, more recently, Poursaee [38]. A brief outline of the corrosion process is given below.

1.3.1.1 Passivation

Concrete is very alkaline, usually with a pH of 12 - 13. When steel is immersed in concrete with a high pH and no chlorides are present, anodic and cathodic reactions begin. The anodic reaction leads to the formation of iron cations:

$$\mathrm{Fe} \longrightarrow + \mathrm{Fe}^{2+} + 2\,\mathrm{e}^{-}$$

while the cathodic reaction leads to the formation of hydroxyl anions:

$$\tfrac{1}{2}\mathrm{O}_2 + \mathrm{H}_2\mathrm{O} + 2\,\mathrm{e}^{-} \longrightarrow 2\,\mathrm{OH}^{-}$$

These two products combine and form a stable film around the steel [39]. This passivating layer is a dense, ultrathin film of metal oxides and hydroxides, which can protect it from corrosion by forming a physical barrier between the steel and corrosive elements [40], reducing the steel's dissolution rate to a negligible value [38]. However, once depassivation, the breakdown of the passive layer, has occurred, corrosion is initiated. This depassivation is caused by chloride attack and carbonation, both of which can affect the concrete simultaneously or separately.

1.3.1.2 Chloride attack

Precisely why, how, and when steel depassivates due to chlorine attack is not clearly understood [38]. However, Montemor et al. have given a detailed guide to chloride-induced corrosion [39].

There are several possible sources of the chloride ions that are required for chloride attack, such as the use of chloride-contaminated ingredients in the mixing process, or environmental factors such as seawater and deicing salts. There are three proposed models for how chloride attack can then cause depassivation. The first of these is the adsorption-displacement model, where the adsorption of Cl^- and the simultaneous displacement of O_2 from the passive layer destroy the passive layer. The chemico-mechanical model suggests that the chloride ions lower the surface tension, causing cracks and flaws due to the repulsive forces between large numbers of adsorbed ions. The final model is the migration-penetration model, where ion migration causes localised acidification, breaking down the passive film.

Depassivation and the penetration of Cl^- to the steel does not necessarily immediately lead to corrosion. Concrete is able to bind chlorine ions, forming calcium chloroaluminate and immobilising the ions. Consequently, a minimum number of ions ('the chloride threshold') must be reached before corrosion can begin.

1.3.1.3 Carbonation

Carbonation can occur when carbon dioxide is absorbed by the concrete through the environment, penetrating by diffusion through air-filled pores. Steffens et al. have detailed the carbonation process in depth [41]. It can be described by the reaction of the calcium hydroxide in the concrete and the penetrated carbon dioxide:

$$Ca(OH)_2 + CO_2 \longrightarrow CACO_3 + H_2O$$

The practical effect of this is to reduce the pH of the concrete to less than 8.3. When the carbonation penetrates to the depth of the rebar, this drop in pH leads to depassivation.

1.3.1.4 Corrosion of steel

Once corrosion is initiated, the anodic and cathodic reactions change, depending on the local conditions and the composition of the concrete. Poursaee reported four possible anodic reactions, and two possible cathodic reactions [38]. The anodic reactions were:

$$3\,Fe + 4\,H_2O \longrightarrow Fe_3O_4 + 8\,H^+ + 8\,e^-$$

$$2\,Fe + 3\,H_2O \longrightarrow Fe_2O_3 + 6\,H^+ + 6\,e^-$$

$$Fe + 2\,H_2O \longrightarrow HFeO_2^- + 3\,H^+ + 2\,e^-$$

$$Fe \longrightarrow Fe^{2+} + 2\,e^-$$

while the cathodic reactions were:

$$2\,H_2O + O_2 + 4\,e^- \longrightarrow 4\,OH^-$$

$$2\,H_2O + 2\,e^- \longrightarrow H_2 + 2\,OH^-$$

These corrosion products occupy a greater volume than the original steel, causing internal stress that can compromise the structural integrity of the concrete.

1.3.2 Detection of corrosion using the Electromagnetic Anomaly Detection technique

The EMAD technique is an NDT approach that uses Remanent Magnetism (RM) and Magnetic Flux Leakage (MFL) in order to detect breaks in rebars. It has been developed at Keele University over a number of years, and details of the technique were first published in 2006 [42]. It has since been the subject of further publications, which have detailed both the development of the equipment [43] and the application of CI approaches to the automated analysis of data captured by the technique [44–46]. While not widely used, the EMAD technique is a promising approach to the NDT of rebar, and has been applied commercially by SciCorr Ltd. in order to assess structures such as bridges, car parks and dual carriageways. This section describes the theory behind the EMAD technique and how it is practically applied. A detailed account of the physical and mathematical principles behind MFL in general was given by Sawade and Krause in 2010 [47].

The two pieces of EMAD equipment are the EMAD probe, shown in Figure 1.3, and the energiser, which can be seen in Figure 1.4. In a typical rebar scan, the energiser, which is a powerful electromagnet, is passed over the rebar in question. When a ferromagnetic material, such as the steel used in rebar, is placed in a magnetic field that is sufficiently large to magnetically saturate the material, that material will retain some

Figure 1.3: The EMAD probe, which was used throughout Chapter 3.

magnetization once the field has been removed. This is remanent magnetism, and is the basis of the EMAD technique. When the energiser is run along the length of a steel bar, it leaves the bar in a magnetized state, a process referred to as energisation. At this point, it is possible to consider the magnetic flux as flux lines flowing through the material from one pole, at the point where the energisation started, to the other, at the point where the energisation finished. Magnetic flux lines always take the path of least magnetic reluctance, which is analogous to resistance in electrical circuits. Magnetic reluctance is related to the magnetic permeability of the material that the flux travels through by Equation 1.1:

$$\mathcal{R} = \frac{l}{\mu A} \tag{1.1}$$

In Equation 1.1, $\mathcal{R}$ is the reluctance, l is the length of the circuit in m, μ is the magnetic permeability of the material in henries per metre and A is the cross-sectional area of the circuit in m^2. As non-magnetic materials, both air and concrete, for cases where the rebar is encapsulated, have a permeability very close to the that of free space,

Figure 1.4: The energiser, which was used throughout Chapter 3.

$4\pi \times 10^{-7}$ H m^{-1} [48]. This means that a typical rebar has a higher permeability than its immediate surroundings. As a result of this, the flux lines travel mainly through the steel. However, if a defect is introduced to the bar, the space available for the flux to travel through is reduced. This forces more of the flux through the path of lower permeability. Since air and concrete both have a lower permeability than steel, the flux is perturbed. This is shown in Figure 1.5, which depicts the magnetic field and recorded magnetic flux around two magnetised bars. This is the principle of MFL, and is how defects are detected using the EMAD technique. A typical scan using the EMAD technique will not be of one single bar, but several bars arranged in a mesh. In this case, each bar in the mesh must be energised in turn.

After energisation has been completed, the EMAD probe is then used to detect defects. The EMAD contains a triaxial magnetic probe which measures the three Cartesian components of the magnetic flux, as shown in Figure 1.5. The probe is passed over each rebar and records data every 4.71 mm. Figure 1.5 shows the data recorded by the EMAD probe when passed directly over a defect on an energised rebar. The Z axis readings for the flux can be treated as the magnitude of the flux leakage and the X

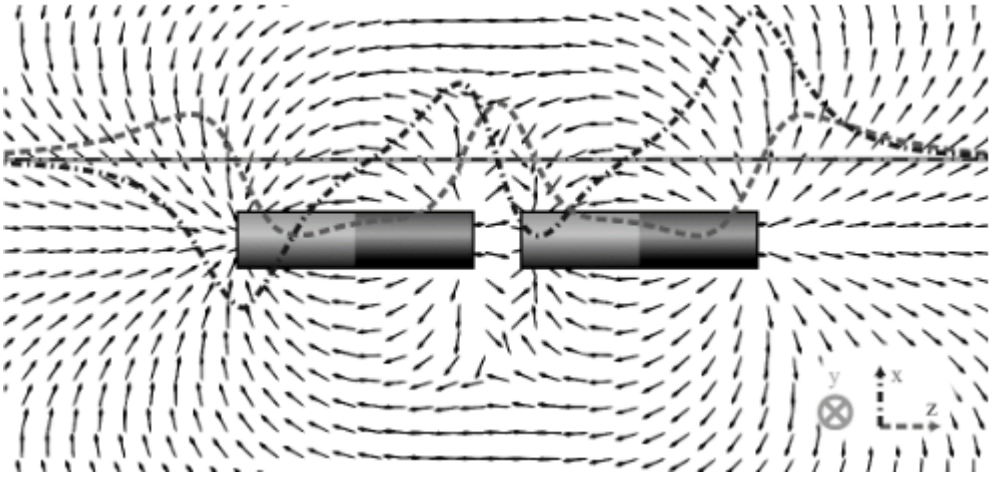


Figure 1.5: The principle of MFL (from Diederich and Vogel [49]).

axis readings can be treated as the radial component of the flux leakage. Consequently, the Z axis data rises up to a peak at the centre of the defect, before then returning to its previous level. The X axis data, meanwhile, initially rises and then decreases from a maximum down to a minimum, before returning to zero outside of the flux leakage region. Providing that the probe was perfectly aligned with the rebar during the scan, the Y axis data should give no signal at all. Although the Y axis data is recorded, it is rarely used in practice.

There are a few possible confounding signals that are an inherent part of using MFL to detect defects. The most common example of this is the 'end effect', a natural consequence of the energisation procedure. After energisation, a large magnetic pole is created at the end of the rebar or, in cases where the full rebar was not energised, at the point where the energisation finished. End effects are usually much larger than defect signals, but the two can still be easily confused. In the case of some rebar meshes, some flux leakage is observed at points where the longitudinal and transverse rebars cross over, leading to a 'ripple' effect. Although these signals resemble the characteristic defect signal, they do not correspond to any fault in the steel.

Since corrosion in rebar is usually local - typically, there may be a few centimetres of corrosion and a much longer section of clean steel [37] - one of the key challenges with any defect detection technique is the accurate and precise location of potential faults. Defects can be located using an Analytical Technique (AT), which is described in

more detail in Chapter 3. In Chapter 3, a RNN approach is applied to EMAD data for improved defect detection accuracy, before data fusion is performed using the EMAD probe and a covermeter.

1.3.3 Magnetic flux leakage approaches to NDT of reinforced concrete

The EMAD technique is not the first attempt at using MFL to find faults in rebars. The method was first proposed by Kusenberger and Barton in 1981, where the pair recommended the use of MFL based on their success in using it to detect faults in steel in other contexts [50]. This led to the granting of a US patent for a method of inspecting rebars in prestressed concrete beams by generating a magnetic field close to the beam and using a Hall effect sensor to find anomalies [51]. Note that this method is different to the EMAD technique, in that it does not exploit RM. Rather, the Hall effect sensor is placed at the centre point between the two poles of an electromagnet, where there is theoretically a zero field. Measurements are then taken during energisation in a process referred to here as Active Field Measurement (AFM). The principle of AFM can be seen in Figure 1.6. AFM has some advantages over RM, in that measured flux leakage signals have a greater amplitude and smaller width [49]. Past studies have found that in practice, however, RM is superior to AFM, since the latter was only able to detect clean breaks [52]. Since these early publications, several groups have attempted to implement methods for using MFL to detect faults in rebar.

The most active of these is the group of researchers from the Fraunhofer Institute. Scheel and Hillemeier first published their work in English in 1997 [54], although an earlier work was published in German by Hillemeier et al. in 1989 [55]. By way of comparison, the first significant paper on the EMAD technique came in 2006 [42], and details of a precursor to the EMAD probe were not published until 2009 [43]. In the 1997 paper, Scheel and Hillemeier used a remanent magnetism technique and investigated the relationship between the peak to peak amplitude of the X component of the magnetic flux signal and the width of the defect. This work has since developed into a portable

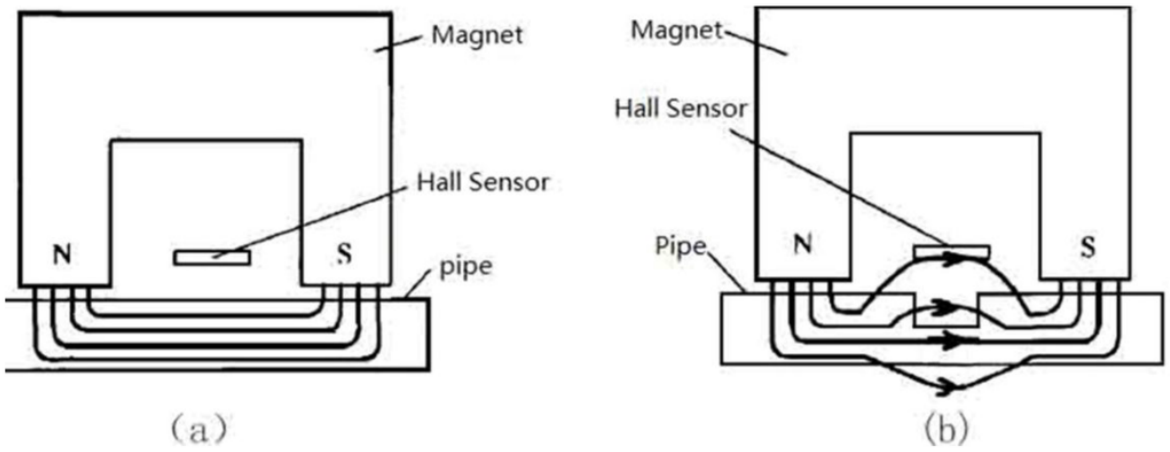


Figure 1.6: The principle of Active Field Measurement. In (a), the pipe has no metal loss, while in (b) the pipe has a defect. A pipe is comparable to a steel reinforcing bar (from Shi et al. [53]).

system capable of magnetizing and scanning a 3.5 m wide section of road in one run [56], and although RM methods have mostly been used, the group has also considered AFM [57]. The most recent paper fromt he group was published in 2015, and employed the RM method, which had been developed for commercial use [58]. The first paper on the EMAD technique in its present state was published in 2014 [46]. However, the data analysis for the method is limited to a manual inspection of the magnetic flux data by an expert, with the aim of detecting the characteristic defect signature. The group have also used MFL for other applications, such as the detection of corrosion damage in prestressed concrete poles [59].

Separate work on using MFL to detect faults in reinforced concrete has been done by a group at the University of Toledo, and this was first published in 1998 [60]. Although a paper from the group in 2013 claimed that there are no MFL methods that are currently ready for field work [61], a paper was published in 2014 that used AFM of both MFL and main magnetic flux (MMF) to locate damage [62]. The MMF is simply the flux that flows through the steel during the magnetization process. In order to measure the MFL and MMF, Hall effect sensors were placed both on the poles of the magnet used and at the centre point between the two poles. The level of corrosion was then assessed by comparing the data taken by each sensor with a computerised magnetic field simulation.

Some of the most recent work on using MFL to find faults in rebars was performed by Vogel, Wolf and Diederich at ETH Zurich. Initial tests used permanent neodymium magnets to magnetize 50 cm long rebar samples in a laboratory setting, and then measuring the RM using a tri-axial magnetometer [63, 64]. In these experiments, a simple algorithm was used in order to detect breaks, with limited success. The method has not yet moved from the laboratory to 'real-world' test sites, and in one of the most recent papers, the technique was adjusted to use AFM rather than RM [49]. Software for the evaluation and interpretation of results is currently in development [65].

There have been several other attempts to use MFL to inspect the condition of rebars. Göktepe first looked at crack detection in industrial machine parts using MFL in 2001 [66], but later went on to use AFM for the inspection of steel rebars [67]. This was ultimately limited to locating bars within a sample using peaks in the z component of the magnetic flux. Jensen et al. conducting a feasibility study in 2013 that involved the on-site scanning of a bridge barrier rail using an AFM method, but this work was still at an early stage [68]. Krause et al worked on the detection of tendon ruptures in prestressed members of bridges with MFL by using a four-channel superconducting quantum interference device in place of the typical Hall effect sensor and then manually analysing the recorded data [69]. Although the system showed promise for periodically monitoring rebars, no further work was conducted after 2002.

The use of MFL is not just limited to the NDT of rebars. Indeed, MFL has long been used for defect detection in pipelines, having been first used by the British Naval Architects Association to find defects on steel pipe in 1868 [70]. This inspection is generally carried out by self-contained units that travel through pipelines at 2 – 6 miles per hour and take AFM [71]. The popularity of MFL for defect detection in pipelines is such that there have been efforts to embed ferrite particles in the polymer matrix composites used in modern pipelines, so that they can be analysed using MFL [72]. Machine learning techniques such as neural networks have frequently been used to process the resultant datasets, most recently for defect depth estimation [73], but also for more straightforward tasks such as defect detection [74], which has so far only been

considered in the context of NDT of rebars by Butcher [45]. Other recent examples of the use of MFL include in the detection of defects in steam generator tubes [75] and the detection of cracks in large liquid atmosphere storage tanks [76]. The related magnetic rubber testing method, where magnetic particles are suspended in a liquid silicone that is then poured on a test site for the detection of cracks, has been used for aerospace applications since the 1970s [77].

While the use of MFL for the NDT of rebar has been subject to plenty of research, it is clear that there are some gaps in the literature that merit further investigation. Firstly, advanced techniques for processing MFL data are lacking. With the exception of the work done by Butcher [45], machine learning techniques have not been used for defect detection. Even then, the data was simply classified as 'defect' or 'no defect', and no data visualisation was provided. This thesis addresses this gap by not only detailing the application a machine learning technique, ESNs, to MFL data, but also by presenting easily interpretable contour maps, that allow defects to be spatially located. Secondly, MFL has yet to be fused with other NDT techniques for the more accurate detection of defects in rebar. One of the main contributions of this thesis is that for the first time, MFL was fused with rebar cover depth measurements, as might be provided by a covermeter.

1.3.4 Alternative techniques for the non-destructive testing of reinforced concrete

There are many alternative NDT approaches to detecting faults in rebar that could potentially be used to complement the EMAD technique, and a brief review of each these is given below. Since the EMAD technique looks principally at corrosion of the rebar, this review only considers those techniques which can be used to investigate the condition of the rebar. Guides to the NDT techniques that can be used for monitoring the health of concrete structures in general have been given by Verma et al. and Helal et al. [78, 79].

1.3.4.1 Half-cell potential

Half-cell potential is one of the most widely used NDT techniques for the inspection of rebar. First proposed in 1957, the method relies on the potential difference between the rebar and a reference electrode [80]. The corrosion process in the rebar causes it to act as a mixed electrode, and connection with the reference electrode on the surface causes it to act as an electrical cell [81]. This requires direct contact between the reference electrode and the rebar, and so some concrete has to be removed to create this connection. The reference electrode itself can then be moved along the surface of the concrete, with the potential difference periodically recorded. The potential difference between the reference electrode and rebar increases as the corrosion activity increases.

According to the American Society of the International Association for Testing and Materials standard C876, if the potentials for a given area are more positive than -0.2 V, there is a greater than 90% probability that no reinforcing steel corrosion is occurring in that area, while a value more negative than -0.35 V indicates a 90% probability of corrosion occurring, while the results should be plotted on a contour map [82]. However, it has been found that the method is sensitive to temperature, chloride concentration and fly ash content [83], and that as a consequence of this, results obtained using half-cell potential are unreliable without compensation [84]. Assouli et al. recommend that the local weather, in particular the humidity and temperature, on both the day of the survey and the days prior to the survey, should be taken into account when assessing the results [85]. The EMAD technique could potentially be fused with half-cell potential in order to increase confidence in the results, since an anomalous MFL signal in a region that half-cell potential suggests has a high corrosion potential would be more likely to be a defect than an anomalous MFL signal in a region with a low corrosion potential, providing that some rebar is still present. If the steel has been completely corroded, then the corrosion process will have stopped and the potential readings will be low, making it impossible to detect such a fault using half-cell potential. However, because of the sensitivity of the half-cell potential technique to environmental parameters, any accurate model would also need to take into account several other modalities of sensor

data. This would, in turn, require a prohibitive amount of training data in order to build a model, since there would need to be enough training data to account for all of the possible variations between the different modalities. On that basis, half-cell potential is unsuitable for fusion with the EMAD technique.

1.3.4.2 Galvanostatic pulse

The galvanostatic pulse method bears some similarity to the half-cell potential method. An external counter-electrode is placed on the surface of concrete and impresses a small amplitude, short interval anodic current pulse on the rebar. The change in potential over time is then used to assess the corrosion rate [86]. Typically, this current is in the range of 10 to 200 μA and is applied for ten seconds [87]. The corrosion rate is calculated by recording the time required for the potential difference to reduce to 37% of its initial value [88].

It has been found that the galvanostatic pulse method is particularly sensitive to both the sodium chloride content in the concrete and the cover depth [89]. It has also been shown that the response to the method is largely due to the resistance of the concrete, meaning that a map produced by the method is effectively a concrete resistivity map. Consequently, it was recommended that the technique should be performed alongside the half-cell technique, so as to avoid misinterpretation of the results [39]. The utility of galvanostatic pulse is therefore limited, and it cannot be usefully fused with EMAD data.

1.3.4.3 Linear polarization resistance

Linear polarization resistance is also closely related to the half-cell potential technique, and uses similar apparatus to determine the corrosion rate by recording the change in potential between the rebar and a half-cell when the current is changed. This is used to calculate the polarization resistance, which is the ratio between the voltage applied to the rebar and the step of the current change. The corrosion current is then calculated by dividing a constant, typically 26 mV for steel rebar, by the polarization resistance

[90, 91]. The greater the corrosion current, the greater the rate of corrosion.

Feliu et al. noted a number of shortcomings with linear polarisation resistance, in particular the fact that its sensitivity to different environmental conditions means that the result of a survey on any given day may not be representative of the average corrosion rate [92]. The authors claim that the practice of confirming the results with a half-cell potential scan is also flawed, due to its being similarly sensitive to environmental factors. Consequently, like half-cell potential and the galvanostatic pulse approaches, a prohibitive amount of training data containing known ground truth defects, gathered in a range of conditions, would be required in order to produce a working corrosion detection model. Furthermore, a measurement of corrosion rate may not necessarily adequately complement MFL data. While a high corrosion rate might indicate that it is more likely that a defect may be present in an area, the corrosion rate at a clean break in the rebar would be zero, since no steel would be present at all.

1.3.4.4 Time domain reflectometry

Time domain reflectometry has been used for a number of years as a method of detecting faults in transmission lines. When used for detecting faults in reinforced concrete, each rebar is treated as a transmission line. A pulse generator is used to send an electrical pulse down both the rebar and an adjacent, parallel sensor wire. An oscilloscope then records the echoes that return, with damage to the rebar observed as a characteristic spike in the impedance that can be calculated using the voltage and current in the wire [93]. Liu et al. were able to detect and characterise pitting corrosion, surface corrosion and even voids in the concrete using the method [94]. However, the technique is designed so that the sensor wire is embedded in the concrete, along with the rebar. It is still possible to assess the rebar for corrosion by placing the wire on the surface of the concrete, but this reduces the sensitivity of the technique to reduction in rebar cross-section. Its use is, therefore, limited unless it is embedded in the original structure during construction. This makes the method unsuitable for surveying existing structures alongside the EMAD technique.

1.3.4.5 Electrical resistivity

Unlike the other techniques considered to this point, the electrical resistivity method investigates a specific property of the concrete, electrical resistivity, in order to determine the condition of the rebar. The resistivity is measured by placing two or more electrodes on the surface of the concrete, applying a voltage between these electrodes and measuring the resultant current. It is common for a resistivity probe to have four electrodes, with two used for applying a voltage and two for measuring the current, since this has been found to reduce errors [95]. This current is carried by ions dissolved in the pore liquid, and so is greater for concrete with a higher water to cement ratio. The resistance is then calculated using Ohm's law, and a conversion factor is used to convert this value into the resistivity of the concrete [96]. The resistivity is closely linked to the degree of the concrete's pore saturation and, below a critical pore saturation value, the resistivity is sufficiently high to prevent corrosion [97]. Feliu et al. found in 1989 that the rebar corrosion rate is inversely proportional to concrete resistivity, since areas of low resistivity are areas that are relatively easily penetrated by chlorides [98]. Indeed, the rebar can be said to be in an active state of corrosion if the resistivity is lower than 10 kΩ cm, since it is likely that the passive layer of the steel will have been broken down by chloride attack. Conversely, rebar will exhibit passive behaviour if the resistivity is greater than 30 kΩ cm, since it is unlikely that the chloride penetration would be sufficient to cause depassivation [99]. Unlike the other methods reviewed here, resistivity is generally not recorded continuously, but is instead systematically recorded at several points on the surface to be scanned. This is because resistivity cannot be measured directly over the rebar, since the current may be carried by the rebar instead of the concrete. The exact position of the rebar in the concrete must therefore be known.

The instantaneous resistivity is sensitive to a number of environmental factors. The relative humidity at the time of the scan affects the resistivity [100], as does the temperature [96]. Consequently, as with linear polarization resistance, the resistivity at any given time may not be representative of the average resistivity, and hence the average corrosion rate. Much like half-cell potential, any data fusion model that sought

to use the electrical resistivity method would need to have a bank of training data gathered in a range of different environmental conditions. Since the EMAD technique is not sensitive to these parameters, it would not be helpful to fuse it with these methods, as it would serve to introduce additional potential sources of variability to the final result.

1.3.4.6 Ground penetrating radar

Ground-Penetrating Radar (GPR) is a versatile technique, and has also been used in NDT applications, such as rebar detection [101] and delamination detection [102]. The GPR method is relatively straightforward: an antenna generates short radio pulses at frequencies between 1 and 5 GHz, which reflect off objects in the concrete in different ways depending on their electrical conductivity and dielectric constant [103]. In the case of rebar, corrosion can reduce the rebar diameter and alter the dielectric constant for the adjacent concrete. The subsequent data processing, however, is known to be highly complex [104]. This data processing is made more complex by the changing behaviour of GPR signals in different media, including wet and dry concrete [105]. Although the use of GPR for corrosion detection is still an active area of research (see, for example, work by Hasan and Yazdini [106], and Martino et al. [107]), much of this work is still concerned with how best to interpret the data, and Tarussov et al. recommended that GPR is used as an imaging technique, rather than a measurement technique [103]. In their work, image analysis software was used for visual identification of defects, rather than the numerical analysis of the data that is performed when using the EMAD technique. This makes it unsuitable for data fusion in this particular context.

It should be noted that although GPR with a magnetic bias field has been used to detect corrosion, this work was still in a laboratory setting when last used, and so is not yet suitable for use in the field [108, 109].

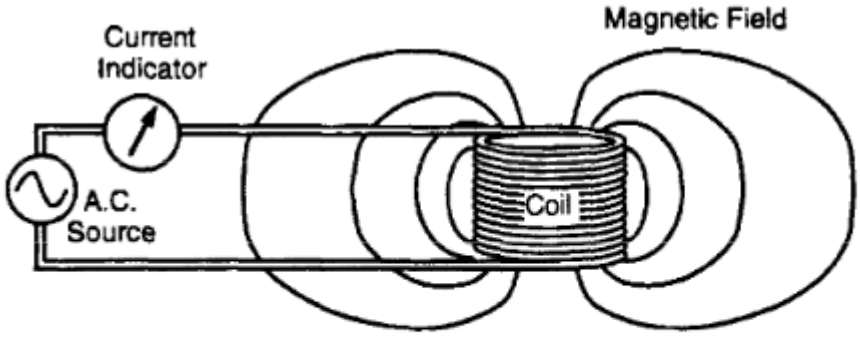
1.3.4.7 Ultrasonic guided waves

Ultrasonic guided waves have been a popular NDT in several applications, in particular the detection of defects in oil and gas pipelines, where methods have been developed for processing data arising from permanently embedded systems [110]. In reinforced concrete, the method relies on the propagation of waves through the rebar. A pulser/receiver is attached to two exposed ends of the rebar to be scanned so that ultrasonic waves of different frequencies can be propagated through the rebar from one end received at the other. The attenuation of the amplitude of the signal is then calculated for each different frequency and used to determine if there has been loss of section in the rebar [111]. Alternatively, changes to the time of flight of the waves in the sample can be used to detect reduction in the diameter of the rebar [112]. The method can also be used to discern between different types of corrosion [113]. However, it cannot be used to identify the precise location of defects. Furthermore, an advantage of the EMAD technique is that it is totally non-destructive and does not require direct contact with the rebar. Fusion with ultrasonic guided wave data would necessitate the breaking out of concrete at the end of each rebar to be scanned, which is undesirable. Beard et al. reported that the inspection range for a complete break would be limited to 0.8 m for wires with a diameter of 5 mm [114]. This is an additional problem, since the reinforcing meshes used in Chapter 3 had 5 mm diameters and were over 4 m in length. In light of all of this, it is not suitable to fuse the ultrasonic guided waves technique with the EMAD technique.

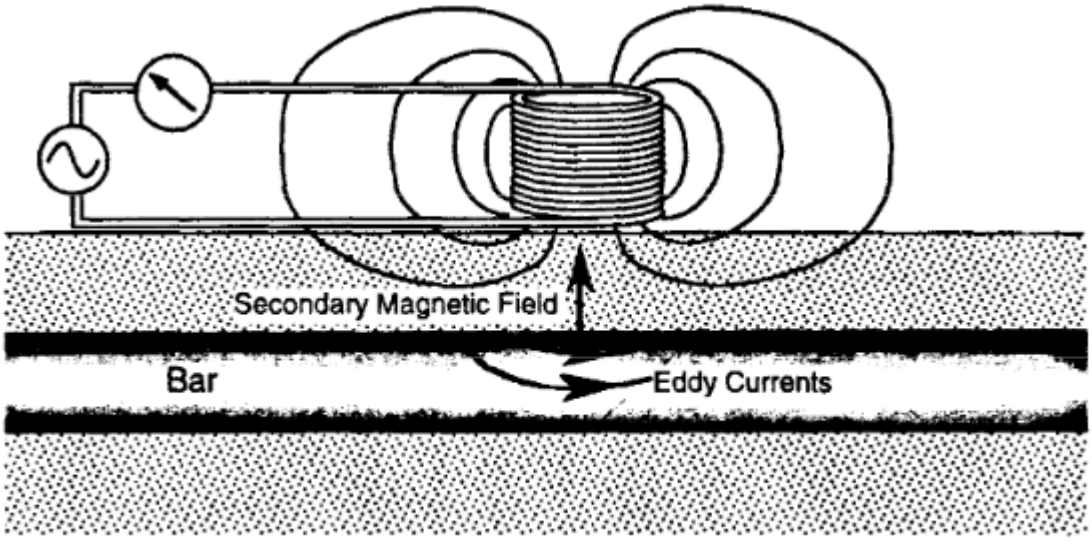
1.3.4.8 Covermeter

Covermeters do not assess damage to the rebar, but instead use electromagnetic pulse induction to map the rebar and, given the diameter of the rebar, determine the depth of the cover. A covermeter can be considered to be a simple coil connected to an alternating current electricity supply. When a coil carrying an AC current is brought near a conductor, the changing magnetic field induced in the coil by the AC current causes circulating eddy currents on the surface of the conductor via electromagnetic

induction. This is what happens when a covermeter is brought in close proximity to a rebar. The eddy currents on the surface of the rebar give rise to a secondary alternating magnetic field, which induces a secondary current in the covermeter. This secondary current opposes the original AC current, reducing the net current in coil and giving the impression of the impedance of the coil increasing. This impedance is compared with the impedance when no conductor is present in order to calculate the location and depth of rebars [115, 116]. This principle can be seen in Figure 1.7.



(a) Coil in air results in a characteristic current amplitude



(b) Interaction with reinforcing bar causes changes in coil impedance and current amplitude

Figure 1.7: The eddy current principle for covermeters (reproduced from Carino [115]).

The idea of combining the EMAD technique with covermeter data is appealing for two reasons. Firstly, the magnitude of the signals received by the EMAD technique depends directly on the vertical distance of the energiser and EMAD from the rebar. The greater the vertical distance from which the scan is conducted, the smaller the

effect of the energiser on the magnetic state of the rebar and the smaller the magnitude of the signals recorded by the EMAD probe. It is possible that a very small anomaly at a depth of 10 mm below the concrete would, in the raw EMAD data, appear to be a more significant problem than a complete loss of section at a depth of 280 mm. In a 'real-world' scenario, the rebar is not guaranteed to be close to the surface, or even at a constant distance from the surface. A CI technique that has been trained exclusively on data obtained at roughly the same distance from the rebar may learn that smaller amplitude signals are less serious than large amplitude signals. Context about the distance of the scan from the rebar is likely to be important for an accurate determination of rebar condition.

Secondly, Pailes observed that cover depth is a significant factor in the health of reinforcing bars, since rebars in areas of low cover deteriorate more quickly than those in areas with a great depth of concrete cover [34]. Indeed, Marsh described the failure to achieve the specified concrete cover as 'probably the greatest single factor influencing the premature deterioration of reinforced concrete'. For these two reasons, it was decided that the EMAD data would be most productively fused with covermeter data.

1.3.5 Fusion of data from the non-destructive testing of reinforced concrete

While data fusion has yet to be used with any of the MFL techniques for NDT, this is not true of the other NDT techniques. Indeed, since 2007, Huston et al. have repeatedly called for data fusion in the assessment of rebar condition [33, 117–121], with Kohl et al. also suggesting in 2005 that an automatic data fusion algorithm for the NDT of concrete would be useful [122]. Breyssea gave two principle reasons for the use of data fusion in this context. Firstly, if a measurement is sensitive to two parameters, a second measurement that is sensitive to one of these parameters can be used to invert the system and quantify each parameter. Secondly, if a technique is sensitive to one influential parameter and a bias, measuring a second parameter that is similarly sensitive to that bias can reduce the effect of the bias [123]. There are three

major systems of data fusion for the NDT of reinforced concrete: OSSCAR/BetoScan [124, 125], RABIT [126] and VOTERS [127]. These are summarised below, along with several other attempts at fusing a variety of different NDT techniques.

1.3.5.1 VOTERS

VOTERS, the product of a research group at Northeastern University, is perhaps the most well developed of these methods. First published in 2012 following a 2009 research grant [127], VOTERS employed accelerometers, laser height sensors and tyre pressure sensors, along with acoustic sensors for the detection of tyre-induced vibration and sound waves, so as to determine the surface texture [128], a GPR array for the mapping of reinforcement bars and pavement layers [129], radar for the measurement of near-surface defects [130] and video footage for crack detection [131, 132]. An overview of this system was published by Birken et al. in 2014 [133]. The sensors were mounted in a transit van, and data was collected using the Sirom^3 framework [134, 135]. Data recorded by VOTERS were fused by the PAVEMON system, which produced a pavement condition index for the road surveyed by VOTERS, classified surface features as 'cracks', 'manholes' or 'no cracks', produced a priority list for repairs, and predicted the effect of extreme weather on road condition [136]. In the first stage, certain features of the raw tyre pressure, accelerometer and laser height sensor data were extracted and transformed before being processed using stepwise regression, which gave a predicted pavement condition index in the range of 0 - 100. Features from the acoustic sensors, tyre pressure sensors and accelerometers were then fed into a support vector machine, which classified the condition of the road at any given point as 'crack', 'manhole' or 'no features'. The repair prioritisation was performed using a simple weighted equation, which was based on factors such as the pavement condition index, average daily traffic and benefit-to-cost ratio. Finally, stepwise regression was used for quantifying the effect of extreme weather on roads [137]. However, this approach does not fully utilise all of the sensors available, and does not take into consideration factors such as the actual condition of the steel reinforcing bars that would be mapped by the GPR.

1.3.5.2 OSSCAR/BetoScan

Rather than embedding sensors in a transit van, the OSSCAR approach used a multi-sensor scanner, while the related BetoScan method used a self-navigating robot. Both of these methods were the result of a collaboration between researchers from several different German research institutes and industrial partners. Betoscan was designed to autonomously drive over large surfaces, simultaneously recording data on its eight scanners [125]. These consisted of an air temperature and air humidity scanner, two microwave scanners for the humidity of the concrete surface and volume, an eddy current covermeter to record concrete coverage, ultrasound scanners for concrete thickness measurements, radar scanners to give general information on the rebars and half-cell potential [138]. In 2014, Cotic et al. developed an image fusion scheme that combined the data from half-cell potential, covermeter and microwaves using three different unsupervised clutering schemes, namely the fuzzy c-means, Gustafson-Kessel and possibilistic-fuzzy c-means algorithms. By doing this, it was possible to produce a map that represented each point as either an area of severe corrosion, an area with high salinity content or undamaged [139].

OSSCAR, meanwhile, was designed as a more versatile device that could also be used for scanning smaller areas [124]. OSSCAR featured three different sensors: radar for geometrical information on the rebar, ultrasonic-echo for further geometrical information and eddy-current covermeter for concrete coverage and rebar position [140]. The data fusion method for OSSCAR was a simple weighted addition of the values from the ultrasonic-echo scan and the radar scan, with the weights calibrated on tests performed on a polyamide block, which could then be combined with the covermeter data to produce a 3D map of the reinforcing bars [141]. Unlike the VOTERS system, both OSSCAR and BetoScan considered the condition of rebars, but the data fusion approaches used were straightforward, especially the weighted addition used by OSSCAR. Consequently, OSSCAR only allowed for the visualisation of the steel and while BetoScan could indicate regions that were affected by corrosion, it did not allow for the sort of precise location of defects of which the EMAD probe is capable.

1.3.5.3 RABIT

Similar to BetoScan, RABIT aimed to provide autonomous scanning of concrete bridge decks using a plurality of sensors, as part of research funded by the Federal Highways Administration in the USA [142]. RABIT used electrical resistivity, impact echo, ultrasonic surface waves and GPR in order to detect rebar corrosion, delamination and concrete degradation [143]. In the most recent work, each of the scanners was used to fulfil a separate role. Gucunski et al. used the impact echo sensor data to detect delamination, while the ultrasonic surface waves were used to assess the quality of the concrete [144]. La et al., meanwhile, used GPR to detect deterioration such as corrosion or delamination, electrical resistivity to detect corrosion, ultrasonic surface waves to assess the concrete elastic modulus and impact echo to detect delamination [145]. In all of these works, no actual data fusion was performed, and each dataset was analysed separately. The only exception to this has been in two other papers on RABIT, in which a 3D visualisation of the rebar was created, where delamination was presented as thin surfaces in the deck volume, low quality concrete was presented as a cloudy zone and corrosion through the use different coloured sections of rebar [143, 146]. There was also an attempt by Kim et al. to create an overall condition index based on the impact echo, electrical resistivity and GPR data [147]. In this case, although the techniques each looked at a different aspect of the condition of the bridge deck, data fusion could potentially be used to give an overall assessment of the bridge deck condition. Although corrosion and delamination were considered, RABIT was unable to precisely locate defects in the rebar.

In the group's most recent work, a prototype robot was developed and fitted with cameras and infrared sensors, with a crack detection algorithm applied to stitched 3D images [148]. Le et al., meanwhile, developed a robot that used GPR and electrical resistivity sensors [149]. The GPR data was processed with a Naive Bayes classifier and used to locate rebar, while the resistivity measurements were then mapped to these results in order to indicate corrosion levels. Again, these two papers did not actually use data fusion.

1.3.5.4 CI approaches to data fusion in the NDT of reinforced concrete

Although VOTERS, OSSCAR/BetoScan and RABIT are perhaps the most developed multi-sensor techniques for the assessment of reinforced concrete, they are by no means the only attempts to integrate heterogeneous sensors for this purpose. There have been a number of attempts to apply CI techniques for the fusion of different data types. Volker and Shokouhi were able to detect honeycombing in reinforced concrete by applying three different clustering algorithms - K-Means, Fuzzy C-Means and DBSCAN - to impact-echo, ultrasound and GPR data, finding that DBSCAN provided the best results [150]. The two also applied the Hadamard product - an algebraic operation for pixel-wise multiplication of same-size matrices - and Dempster-Shafer rule of combination - an approach that looks at the probabilities of each possible hypothesis - for the fusion of the same data [151]. Using recorded temperature, dissolved oxygen, salinity, pH, oxidation/reduction potential and corrosion rate data, Jue-Long et al. found that a three layer feedforward backpropagation ANN was suitable for data fusion for assessing corrosion in reinforced concrete sea walls [152], although in earlier work it was found that a support vector regression method outperformed a back-propagation neural network [153]. Sadowski, meanwhile, was able to assess the corrosion rate of rebar using a simple ANN to fuse air temperature, AC resistivity over the rebar, AC resistivity remote from the rebar and DC resistivity over the rebar [154]. This successful application of basic CI techniques to the data fusion problem in this context suggests that there is some scope for the application of more sophisticated machine learning methods, such as ESNs.

1.3.5.5 Other data fusion approaches to NDT of reinforced concrete

There have also been many attempts to use more straightforward data fusion approaches. As recently as 2014, Vanniamparambil et al. assessed the condition of concrete masonry walls by simply manually comparing digitally correlated digital photographs of a wall with acoustic emission data, as recorded by sensors that detected the high frequency waves emitted by the material due to deterioration [155]. A similar approach was used by Khan et al., who manually assessed active infrared monitoring and ultrasonic

tests [156]. Extracting useful conclusions by this method required some quite complex analysis of the data, and a machine learning approach could potentially have reduced the amount of manual analysis. Similarly, in work by Hola et al. [157] and Gorzelanczyk et al. [158], a scan of a concrete surface was performed with impulse response and possible delamination identified, before a second technique was used (impact-echo in the case of the former, ultrasonic tomography in the latter) to confirm these results. The approach was used by Yaghi for the fusion of infrared thermography and GPR in order to investigate a range of possible modes of concrete faults, including rebar corrosion [159]. Kohl et al. looked at different combinations of radar, ultrasonic data and impact echo data for grouting fault detection in reinforced concrete [160]. In their most recent work, radar data was fused with the ultrasonic and impact echo data separately using a simple 'max of amplitudes' method [161]. In that work, the authors suggested that an automated method might be preferable.

A more manual approach was used as part of the French SENSO project, which aimed to combine different NDT techniques for assessing the health of concrete structures [162]. Different aspects of ultrasonic, impact echo, resistivity, GPR and capacitance measurements were combined separately using principal component analysis and data inversion to find the porosity and water saturation of concrete [163]. This methodology required a lot of manual analysis to determine the correlation between different factors and then obtain the final results, with the result that predictive error for porosity decreased from 36.7% without fusion to 4.1% with fusion. A research group in Chile used multiple linear regression to combine ultrasonic pulse velocity, resistivity, resonance frequency and hammer test rebound in order to find the static modulus of elasticity for hydraulic concrete structures [164, 165]. The choice of a linear model, however, necessitated the generation of multiple models, so as to ensure that any of the techniques that did not correlate well with the static modulus of elasticity were identified and removed.

Many more of the attempts at data fusion for the assessment of the health of reinforced concrete have been based on pixel-level fusion. Huston et al. produced

maps of the results obtained from a bridge deck that was surveyed using chain drag, impact echo, half-cell potential and GPR, and then simply took the average value for the corresponding pixels in each map to produce an overall map of the condition of the bridge [166]. Moselhi et al. found bridge deterioration by performing pixel-level image fusion of GPR and infrared thermography data [167]. Cui et al. investigated the corrosion of rebar with half-cell potential and GPR by considering the level of agreement between pixels in the maps produced by each technique, producing a greyscale plot in the process [168]. Maser surveyed an area with GPR and infrared thermography, and then simply produced a map that overlayed the two sets of results [169]. These methods are not unlike the method used by the commercially available CoverMaster software, which allows a user to overlay recorded covermeter and half-cell data [170].

1.3.5.6 Industrial interest in data fusion

Finally, it is worth mentioning reports from the Netherlands Organisation for Applied Scientific Research and Oak Ridge National Laboratory in the USA. In 2012, Oak Ridge National Laboratory produced a report for the US Department of Energy on NDT for the evaluation of concrete in light water nuclear reactors [171]. In the proposed programme of work included in the report, a working group was set up to develop a data fusion algorithm that could combine the many methods of NDT that were available. Meanwhile, the Netherlands Organisation for Applied Scientific Research set out its 2015 - 18 action plan in an annual report and gave one of its planned activities as the production of a data fusion model that could combine heterogeneous NDT sensor data in order to detect rebar corrosion [172]. In addition to the academic interest in the topic demonstrated by body of literature detailed above, these two reports clearly show that the demand for the fusion of heterogeneous NDT data for the assessment of reinforced concrete also extends into industry.

1.3.5.7 Summary

What the current body of literature on the fusion of NDT techniques for the evaluation of reinforced concrete shows is that, while this is an active area of study, there is scope for new work and development, in the shape of both the EMAD technique and ESNs. Of the three well-developed techniques that have been used in a 'real-world' setting, not one out of VOTERS, OSSCAR/BetoScan and RABIT had the capability to locate accurately specific defects in rebar. MFL, and the EMAD technique in particular, should work well with other techniques in this regard, and could contribute to the production of an accurate model of the health of a given reinforced concrete structure. The literature also indicates that there is a current interest in heterogeneous data fusion in this context, since it is the subject of both continuing academic research and industrial interest. The combination of the EMAD technique and a covermeter, as presented in Chapter 3, represents a novel and relevant contribution to this active area of research.

1.4 Bridge-based Sensor Networks for Structural Health Monitoring of Bridges

Structural Health Monitoring (SHM) is defined by Farrar and Worden as 'the process of implementing a damage identification strategy for aerospace, civil and mechanical engineering infrastructure' [173]. Where SHM differs from NDT is that it aims to diagnose the state of a structure at every stage of its life and, over time, provide a prognosis of factors such as evolution of damage or remaining usable life [174]. This online monitoring approach is in contrast with NDT, which is generally used in an offline fashion for the diagnosis and characterisation of damage on the day of a survey.

A popular option for the SHM of bridges is the Bridge-based Sensor Network (BSN), where multiple sensors are embedded in a bridge and passively monitor its behaviour. Reviews of the general use of BSNs in SHM have been published by Lynch and Loh [175], Jain et al. [176] and Dragos and Smarsly [177], while a tutorial on the basics of

damage detection using BSNs was given by Viswanathan and Varshney [178].

The remainder of this section reviews past approaches to fusing data arising from different modalities of embedded sensors on bridges. Note that this does not include work in which multiple sensors of the same type were fused, such as the MODEM approach proposed by Bhuiyan et al. [179], nor machine learning approaches to damage detection that use only one type of sensor, a review of which has been published by Hughes and Correll [180]. This review relates to the work in Chapter 4, where temperature, tilt and strain sensors, embedded on a footbridge, were fused in order to detect, localise, classify and assess damage.

1.4.1 Data fusion for damage detection with bridge-based sensor networks

In 2007, Glaser et al. identified that work on data fusion for sensor networks installed in civil infrastructure was lacking [181]. Since then, several studies have examined the problem of damage detection in bridges using BSNs.

1.4.1.1 Removal of seasonal trends

One of the oldest and most popular approaches to processing BSN data is to remove seasonal trends. The idea behind the removal of seasonal trends is that the majority of the variation in sensor data is caused by seasonal effects, such as changes in temperature. By removing these effects, any remaining variation in the sensor data ('novelty') could potentially be indicative of damage. This was first attempted by Sohn et al. in 1999, when a linear model was used to remove temperature effects from accelerometer data taken from the Alamosa Canyon Bridge in New Mexico [182]. Ko et al., meanwhile, used ANNs to correlate temperature sensor and accelerometer readings obtained from the Ting Kau Bridge, so as to remove environmental effects [183]. Gu et al. used ANNs to filter temperature effects from vibration-based data [184]. Yu and Ou developed a sensor to monitor temperature, acceleration and strain, and used sensor-embedded arithmetic to correct acceleration readings [185]. Xiao et al. fitted a bridge with infrared strain

gauges, temperature sensors and tilt sensors and looked at the relationship between strain and temperature, but did not perform any data fusion [186].

A slight modification to this idea can be seen in the cointegration approach developed by Cross et al. [187], which was applied by Cross and Worden to purge the Z24 bridge accelerometer data of trends seen in the temperature sensor data by finding the cointegrating vector that related the two variables [188]. Once this vector had been found, new monitored data was projected onto that vector, filtering out temperature effects. The filtered residual data could then be used for damage detection, as any deviations in the residual were taken to be indicative of damage. A similar approach was used by Worden et al. on both the Z24 bridge and the Tamar suspension bridge [189].

1.4.1.2 Residual generation

The idea of a residual generation approach is also a popular one, although residuals are usually found by calculating the difference between the data produced by a sensor, and the output of a model of the sensor behaviour that is based on one or more alternative sensors. Peeters et al. used an autoregressive exogeneous model to predict the behaviour of the Z24 bridge accelerometers based on the temperature sensor data, and used outliers in the residual to indicate damage [190, 191]. Cross has used temperature, traffic loading, wind speed and route mean square deck acceleration data as inputs to a Gaussian Process regression model, which predicted the lowest natural frequency of the Tamar bridge, with residual outliers used to indicate damage[192], while a response surface model with the same inputs has also been used [193]. Laory et al. used several techniques, such as ANNs, random forest and support vector regression, to predict the natural frequency of the Tamar suspension bridge based on wind, temperature and traffic data, but did not go on to perform any damage identification [194]. Similarly, de Battista et al. created a finite element model of the Tamar suspension bridge and used this to explain the relationships between different variables, but did not go on to detect damage [195].

In other studies, the calculation of residuals is perhaps the most commonly used

method for damage detection. Kijewski-Correa et al. modelled the relationship between strain gauge data and acceleration for a simulated bridge using an autoregressive moving average technique, and then determined the residual [196]. If the residual was greater than a certain threshold value, it was assumed that damage had occurred. Kromanis and Kripakaran set up a laboratory-based model of a steel truss, used support vector regression to model strain gauge data based on temperature sensor data and then used a signal subtraction method to detect anomalies in the residuals [197].

1.4.1.3 The Z24 Bridge and Tamar Suspension Bridge

Much of the work on data fusion in SHM using BSNs has been done by researchers at the Universities of Sheffield and Exeter, who have used a range of approaches, most of which were applied to one of two major 'real-world' case studies. The first of these was the Z24 highway bridge in Switzerland, which was decommissioned, fitted with a plurality of environmental sensors and accelerometers, and monitored for twelve months between October 1997 and October 1998 [198]. The second of these major case studies was the Tamar suspension bridge in Plymouth, which was fitted with strain gauges, tower displacement sensors, level sensors, accelerometers and a number of different environmental sensors, all of which recorded data between 2006 and 2009 [199, 200]. Both bridges were built in the 1960s (1963 for Z24, 1961 for Tamar) and experienced decades of use before being fitted with sensors. Where the two differ is that the Z24 bridge had been taken out of use, but the Tamar suspension bridge continued to be used normally for the duration of the project.

While the Z24 and Tamar bridges provided widely used datasets, the possibility of fusing structural and environmental data has rarely been considered. When it has, this has sometimes taken the form of purging the structural data of temperature effects. Dervilis et al., identified outliers in the natural frequency of the Z24 bridge using robust multivariate outlier statistics and then used cointegration to remove environmental effects, before identifying which outliers were due to damage [201]. An affinity propagation algorithm - where exemplars are found in order to form clusters -

[202] and heteroscedastic Gaussian processes - Gaussian processes that do not assume constant noise across the input - [203] have also been used for filtering environmental effects on the Z24 bridge.

1.4.1.4 Alternative approaches for damage detection with data fusion

Flammini et al. took a different approach to those discussed so far, proposing the DETECT model for fault classification [204]. However, the model was ultimately only applied to the monitoring of railway infrastructure [205]. Zhang, Jo and Smyth and Wu have all used Kalman filters to detect damage by estimating sensor readings at different locations on a bridge. Smyth and Wu used Kalman filtering to estimate velocity and displacement based on simulated acceleration data [206], Zhang fused accelerometers, strain gauges and displacement transducers [207], and Jo used strain and acceleration sensors only [208]. Sim et al. used damage locating vectors to combine strain gauge and accelerometer data on a numerical model of a bridge truss and looked for damage indicators such as changes in stiffness [209]. In later work by the same group, strain gauge and accelerometer data from the Sorok bridge in South Korea were fused by a multimetric, data-based algorithm, in order to predict bridge displacement, a potential indicator of deterioration [210, 211]. Bruschetta et al. combined cable elongation data, recorded by strain gauges, with force measurements, recorded by elasto-magnetic sensors, and *a priori* knowledge of the bridge design and laboratory calibration to estimate the load on a bridge near to Trento in Italy [212]. Similarly, Zonta used strain gauge and thermometer data to estimate steel cable tension using Bayesian inference [213]. Sun et al. proposed a framework for data fusion in SHM that used an adaptive resonant theory neural network and adaptive fuzzy inference to combine global positioning satellite displacement data with wind speed data, in order to determine the structure health index for a cable bridge [214]. Cho et al. used a Kalman filter to fuse strain and acceleration data obtained from a concrete test bridge so that the bridge's displacement could be estimated [215]. Zhang and Xu used a finite element model, Kalman filters and a radial basis function ANN to combine strain gauges, accelerometers and displacement

sensors for damage detection, although this work was only applied to laboratory-based steel beam [216].

1.4.1.5 Summary

The present literature leaves a number of potential gaps for further work to be done. Indeed, it was noted by Kulkarni et al. in 2011 that ANNs are potentially very well suited to data fusion applications in BSNs, but have seldom been used [217]. Residual generation is an attractive option, since it allows for the deviation of bridge behaviour to be quickly identified, but only simple regression models have been used so far; the noise tolerance and short term memory of ESNs could allow for better models to be generated. Subtleties in the performance of bridges and the temporal relationship between factors affecting bridge behaviour justify the use of CI techniques that are well-suited to time-series data. Furthermore, once anomalous behaviour is identified, further interpretation is required to determine the cause. No approach to date has attempted to use both environmental and structural sensors in combination to model a different structural sensor.

1.4.2 Sensor Validation

Also of interest is the issue of sensor validation, or the performance monitoring of sensors. Sensor failure in SHM applications can potentially lead to the erroneous identification of structural damage, due to the presence of anomalous signals in the data. It is, therefore, important that any technique used for processing BSN data is sufficiently robust to account for sensor faults [218]. Yung identified eight types of sensor failure: hardover, bias, spike, stuck, erratic, cyclic, drift and nonlinear [219]. The characteristics and possible causes of these faults are recorded below [220–223].

1. **Hardover** A sudden and full deflection in sensor reading, usually caused by sudden changes in the operating environment, or an offset error in the sensor system.

2. **Bias** A sensor is offset from its true value by a constant amount. This is often due to ageing sensor components.

3. **Spike** An abrupt, but short, deviation from the regular signal, often caused by external disturbances or a fault in the operating environment.

4. **Stuck** The sensor value remains fixed at a constant value due to a blockage on the sensing line.

5. **Erratic** Random measurements are returned either continuously or intermittently, due to the ageing of the sensor.

6. **Cyclic** Cyclic variations in the signal due to factors such as temperature changes.

7. **Drift** A bias is systematically or gradually added to the sensor reading, caused by sensor contamination or ageing.

8. **Nonlinear** Similar to the cyclic fault, a nonlinear variation is added to the signal.

Despite the increasing use of BSNs in SHM, there is a relative paucity of research into sensor validation in an SHM context [224, 225]. A general review of sensor validation in SHM has recently been published by Yi et al. [226]. Much of the work that has been done has focused on piezoelectric sensors, which are typically used to monitor structural health using either modal analysis, electromechanical impedance or Lamb-wave propagation [227]. Several different groups have looked at using changes in one of the capacitance, impedance, admittance or susceptance of the sensors to find and isolate faulty piezoelectric sensors [225, 228–233]. In an alternative approach, Mulligan et al. were able to correct faulty sensors by analysing changes in the physical transducer modal damping at frequencies around piezoceramic resonance and then using a finite element model to produce a correction factor [234], while Liang et al. isolated sensors that produced outlying values and reworked sensor-paths as a means of reconstructing faulty sensor data [235].

For other modalities of sensor, one typical validation approach is to attempt to model the regular behaviour of the sensor and to then calculate the difference between the

model and the actual sensor behaviour, known as the residual. The methods that have been used for residual generation are given in Table 1.1, along with other approaches to sensor validation.

The problem with all of these sensor validation approaches is that they assume that the structure in question is in good condition, and that anomalous sensor readings are due to sensor faults, rather than structural faults. Hamid et al. used Dempster-Shafer theory of evidence to fuse sensors after calculating a simple reliability value based on the number of false readings given by a sensor over a set window of readings [264], but this approach did not actually identify specific sensor faults. There has only been one credible attempt to detect both structural damage and sensor failure in SHM. In work performed by Liu et al., natural frequency correlation was used to detect faulty sensors, which could then be isolated and removed. The remaining undamaged sensors were then used to examine the vibration pattern emitted when the structure was vibrating under one of its natural frequencies, which allowed for damage to be detected [265]. However, the applicability of this approach to the SHM of bridges was not considered, and the work was only trialled on a laboratory structure. The data fusion approach presented in Chapter 4 contains not only a means of detecting, locating, characterising and quantifying damage, but also of allowing for the observation of sensor faults on a 'real-world' bridge, and isolating those sensors as soon as any faults developed. More details on this can be found in Section 4.4.2.

1.5 Thesis Outline

This thesis will explore the application of a relatively new RNN technique, the ESN, to the problem of heterogeneous data fusion in a civil engineering context, in particular for the detection of damage in concrete structures, such as bridges. To this end, two separate, quasi-real-world case studies have been performed, with each examining a different scenario in which data fusion can be of benefit. In the first of these case studies, presented in Chapter 3, EMAD and covermeter data were combined in order to locate more accurately the position of defects in rebar. Although the presence and

Technique	**Purpose of Use**	**References**
Modal Filtering	Residual Generation	[224, 236]
Principal Component Analysis	Residual Generation	[237–241]
Auto-Regressive and Auto-Regressive Exogeneous Models	Residual Generation	[242, 243]
H_∞ filters	Residual Generation	[244]
Kalman Filters	Residual Generation	[245]
ANN	Residual Generation	[246, 247]
Naïve Bayesian Classifiers	Residual Generation	[248]
Auto-Regressive Exogeneous Models	Modelling the Relationship Between Two Sensors	[249]
Auto-Regressive Exogeneous Models	Correcting Faulty Sensor Values	[250]
Clustering Algorithm	Sensor Validation	[251]
Sliding Mode Observer	Estimate Sensor Measurement Coherence	[252]
Kullback-Leibler Divergence-based Analytical Model	Sensor Fault Amplitude Estimation	[253]
State-Space Model	Create an Error Function	[254]
Largest Empty Rectangle Problem	Sensor Validation	[255]
Null Space-based Algorithm	Detect Sensor Faults	[256]
Adaptive Differential Evolution Algorithm	Detect Sensor Faults	[257]
Damage Locating Vector	Sensor Validation	[258]
Bayesian Belief Network	Fault Detection in Strain Gauges	[259]
k-Medoid Clustering Algorithm	Detection and isolation of sensor fault	[260]
Minimum Mean Square Error Estimate, Multiple Hypothesis Test, Generalised Likelihood Ratio	Detect, Identify and Quantify Sensor Damage	[261]
Multivariable statistical model	Detect sensor faults	[262]
Bayesian inference	Detection and isolation of sensor faults	[263]

Table 1.1: Different techniques used for sensor validation

location of defects was controlled, the EMAD and covermeter data were both obtained from two real, physical reinforcing meshes that were surveyed in the same fashion that a typical reinforced structure would be surfaced. In the second, presented in Chapter 4, tilt, temperature and strain data, taken over the course of a three year footbridge monitoring project, were fused, allowing for the detection, location, classification and assessment of damage sustained by the bridge, along with the development of faults in the sensors. While potentially damaging events were manually enacted, the footbridge was a real structure that had been in regular use since the 1960s when the monitoring project began, and was still subject to environmental effects for the duration of the project. The fact that these case studies were rooted in quasi-real-world scenarios is significant, since Khaleghi noted in a 2013 survey of the data fusion literature that only 6% have a 'practical perspective' that does not use simulated data [266].

1.6 Research Questions

This work was carried out in order to answer the following research questions:

1. Can heterogeneous data fusion lead to improved sensor data interpretation using a relatively new RNN technique from the field of reservoir computing, ESNs?

2. Can any ESN heterogeneous data fusion improvements be demonstrated in quasi-real-world scenarios for:

   (a) NDT, and

   (b) SHM?

   rather than relying on simulated data streams?

3. Can systematically applying the properties of ESNs for heterogeneous data fusion affect the extent to which expert interpretation might otherwise be required?

Figure 1.8 gives a schematic guide the the themes and data modalities in each case study, and where the research questions are answered.

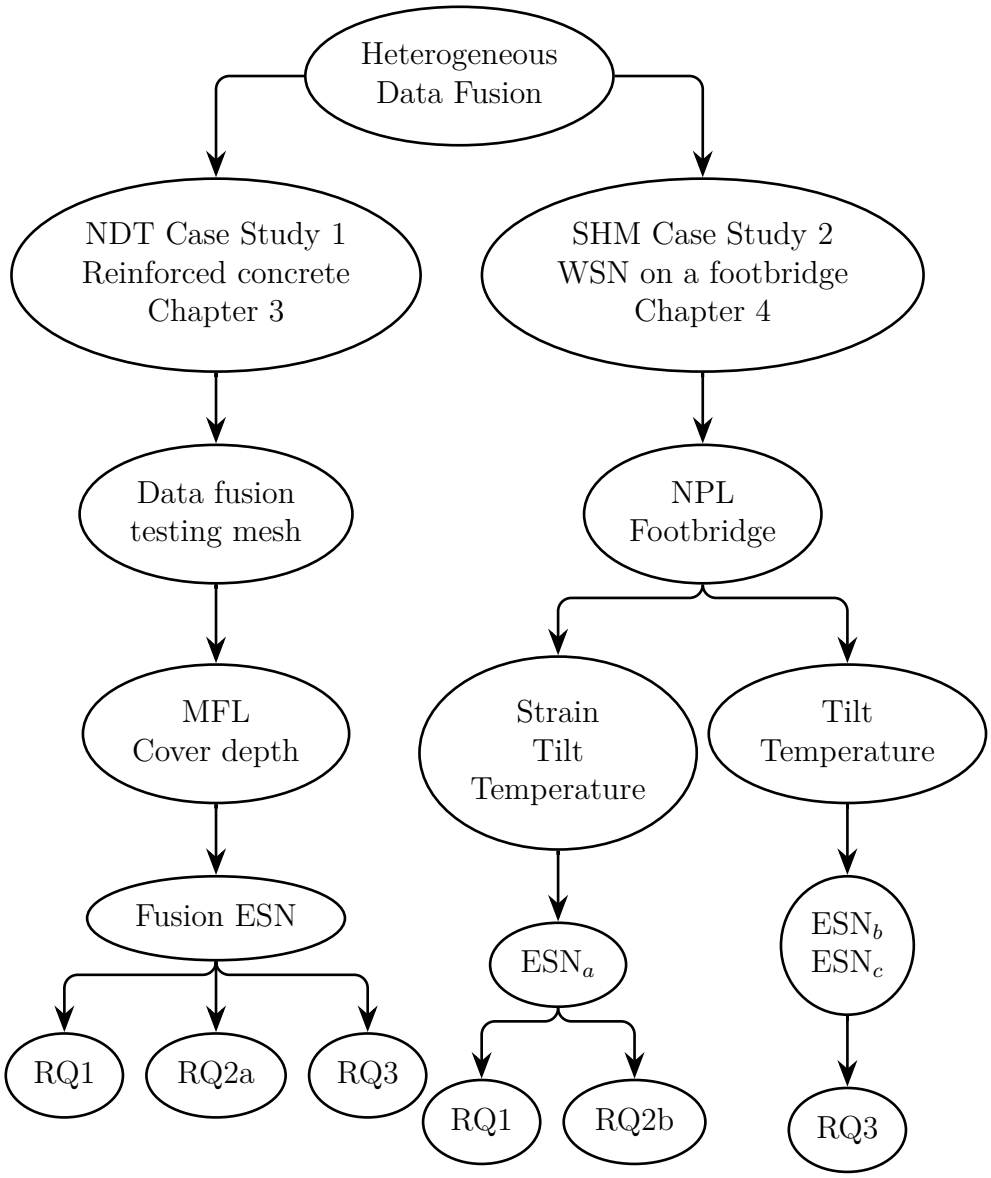


Figure 1.8: The themes present in the thesis, and how they answer the research questions. The NDT-based case study used magnetic flux leakage (MFL) data and cover depth data arising from a steel reinforcing mesh, while the SHM-based case study used strain, tilt and temperature data obtained from the National Physical Laboratory (NPL) footbridge by a bridge-based sensor network (BSN).

1.7 Contribution to Science and Novelty

In the above review of the existing literature, a number of gaps were identified as potential areas for research. This thesis aimed to fill these gaps, broaching new areas in both SHM and NDT alike.

1.7.1 Novelty of research

In seeking to answer the two research questions, the two case studies that are presented in Chapters 3 and 4 make novel contributions to NDT and SHM, respectively:

- As shown in Chapter 2, the use of ESNs for data fusion in SHM and NDT has not previously been attempted, and the work in this thesis enters new territory in this regard.
- Heterogeneous data has not previously been fused in order to detect the location of defects on rebar.
- No past work has attempted to combine any MFL technique with covermeters.
- Chapter 4 presents a 'suite' of ESNs for simultaneously detecting, localising, characterising and assessing damage to a monitored bridge.
- Chapter 4 also shows how environmental sensors and physical sensors can be combined together in the modelling of other physical sensors.

1.7.2 Industrial impact of research

As well as being of academic interest, the work contained in this thesis is also relevant to a number of directions presently being pursued in industry. The Netherlands Organisation for Applied Scientific Research stated in its 2015 - 18 action plan that one planned activity was the development of a data fusion model for the detection of corrosion in rebar using heterogeneous NDT sensor data [172], while Oak Ridge National Laboratory, working for the US Department of Energy, set up a working group to develop a data

fusion algorithm for NDT [171]. The American Society of Civil Engineers' 2017 report card mentioned that embedded sensor technologies allow for the earlier identification of structural problems in bridges, suggesting that any method that could better exploit these sensor technologies would be of great benefit.

Furthermore, the EMAD is not an experimental technique; it is a method that has been used commercially to inspect the condition of several 'real-world' structures. The improvement in the interpretation of EMAD data in Chapter 3 will deliver a genuine industrial impact, as the techniques can be used to improve the accuracy of EMAD surveys. This, in turn, will allow faults to be detected and repaired in a timely fashion, before a critical structural failure can occur.

Pailes identified that one of the problems with using NDT techniques is that the best threshold value to apply to results can vary depending on certain environmental factors, and does not always correspond to recommended thresholds given in the literature. This means that in order to obtain the best results, a solid understanding of the different factors that can influence measurements is required [34]. This is also true when using the EMAD, which is not sensitive to environmental factors but does have a dependency on the distance between the probe and the rebar, discussed in more detail in Chapter 3. This work has shown for the first time that using data fusion to account for variance in this distance led to an optimal threshold that was almost constant, especially when compared to the much more widely varying AT, which is the present approach to processing EMAD data.

While the work in Chapter 4 deals with a 'real-world' bridge that was taken out of regular use and manually damaged over a period of time, other published monitoring projects, such as the Tamar suspension bridge, looked at data collected from a large structure that was under constant use. The increasing use of such sensor networks on 'real-world' bridges means that the methods and principles developed in Chapter 4 can be applied to these structures in order to detect and monitor damage.

A further discussion of the novel contributions of this work and the practical applications of the thesis can be found in the concluding chapter, Chapter 6.

1.8 List of publications

The following papers have been published as part of the research carried out in this thesis:

A. J. Wootton, C. R. Day, and P. W. Haycock. Echo state network applications in structural health monitoring. in *Proceedings of the 53rd Annual Conference of the British Institute of Non-Destructive Testing (NDT 2014)*, pages 289 - 300, September 2014.

A. J. Wootton, C. R. Day, and P. W. Haycock. An echo state network approach to structural health monitoring. in *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1 - 7, July 2015.

A. J. Wootton, S. L. Taylor, C. R. Day, and P. W. Haycock. Optimizing echo state networks for static pattern recognition. *Cognitive Computation*, 9(3):391-399, 2017.

A. J. Wootton, J. B. Butcher, T. K. Kyriacou, C. R. Day, P. W. Haycock. Structural health monitoring of a footbridge using echo state networks and NARMAX. *Engineering Applications of Artificial Intelligence*, 64:152-163, 2017.

J. B. Butcher, A. V. Rutter, A. J. Wootton, C. R. Day, J. Sule-Suso. Artificial Neural Network Analysis of Volatile Organic Compounds for the detection of lung cancer. in *Advances in Computational Intelligence Systems. UKCI 2017. Advances in Intelligent Systems and Computing*, 650:152-163, 2018.

1.9 Thesis Layout

The remainder of the thesis is organised as follows. Chapter 2 looks at past CI approaches to the problem of data fusion and presents the CI approach used here. Chapter 3 presents the first case study, wherein EMAD and covermeter data were fused for the accurate location of defects in reinforced concrete. Chapter 4 presents the second case study, where three modalities of sensor data were fused for the detection, localisation, classification and assessment of damage to a 'real-world' footbridge. These case studies

are brought together and discussed in Chapter 5, while Chapter 6 concludes the thesis.

CHAPTER 2

Computational Intelligence Approaches

2.1 Introduction

This chapter explores developments in data fusion, considering past CI approaches. Following this, the CI technique that was used for the work presented in this thesis, the ESN, is presented in more detail, along with recent ESN applications.

2.2 Data Fusion

Perhaps the simplest, and most concise, explanation of why a data fusion approach can be valuable was given by Bellot et al. in 2002 [267], who described four possible ways in which data fusion may improve the performance of a system:

1. **Representation** The data at the end of the fusion process has a greater granularity or level of abstraction than each input dataset, providing a richer understanding than the original sensors.

2. **Certainty** The probability of the results being correct following fusion should be greater than the probability prior to fusion.

3. **Accuracy** Similar to the gain in certainty, the standard deviation of results should be smaller, with noise and errors eliminated.

4. **Completeness** Bringing together multiple sources of information on an environment should result in a more complete picture of the environment.

Consequently, an ideal system for fusing heterogeneous data would take in sources of data with limited individual use and provide insights that can not be derived from the individual sensors alone, suppressing noise and giving a fuller overall picture. The benefits of the data fusion approaches presented in Chapters 3 and 4 are demonstrated by an assessment against these four criteria.

Although multisensor data fusion has been used in fields such as NDT, as detailed in Chapter 1, it was initially developed for military applications. Indeed, it was a United States Department of Defence working group, the Joint Directors of Laboratories (JDL) Data Fusion Working Group, that was responsible for the most commonly used data fusion model: the JDL model [268, 269]. Since its publication in 1991, the creators proposed revisions in 1999 [270] and 2004 [271]. It is a conceptual model that decomposes the data fusion problem into five levels of fusion:

- **Level 0: Source preprocessing** The preparation of sensor data for fusion.
- **Level 1: Object refinement** For example, combining data relating to a target so as to classify that target.
- **Level 2: Situation refinement** The use of fusion to estimate of the relationship between different targets, such as temporal behaviour.
- **Level 3: Threat refinement** The use of fusion to estimate the effects of the target's behaviour.
- **Level 4: Process refinement** Optimisation of the data fusion process.

As an example of how this framework can be used, the work presented in Chapter 4 seeks to characterise different types of damage to a footbridge (Level 1), estimate the temporal relationship between different types of sensor (Level 2) and estimate an overall level of damage (Level 3). It should be noted, however, that the model has been

subject to criticism due to its data-centred and abstract nature. This has led to the formulation of a number of alternative models [272].

In 2013, Khalegi published a comprehensive review of the state of the art in the field of data fusion [266]. This review identified a number of challenges for data fusion approaches, some of which are particularly pertinent for the work in this thesis. For example, problems such as data imperfection, where sensor measurements feature uncertainty and imprecision, outliers in sensor data, differing data modalities, and the presence of dynamic time-variant phenomena are particularly relevant.

Khaleghi also identified a major drawback in the present body of literature: of over 50 papers that were surveyed, only 6% examined data fusion from a 'practical perspective'. The other papers were limited to using simulated data, or unrealistic tests. Of those papers that did take a more 'practical' approach, many suffered from a lack of knowledge of the ground truth, preventing any in-depth performance evaluation. An important feature of the work presented in Chapters 3 and 4 is that real, empirically recorded data from real sensors was used and that the ground truth was known.

While the scope of this thesis is limited to NDT and SHM, data fusion has been applied across several other domains. A further review published by Khaleghi in 2015 gave examples of how fusion has been used for military applications such as landmine detection and target tracking, but also non-military applications like air traffic control, healthcare and environment monitoring [273].

The many different methods applied to data fusion are not detailed exhaustively here, since the most common of these are covered in existing reviews of the state of the art [274–277]. Among these are some of the most popular approaches, such as Kalman filtering, wavelet transform and probabilistic grids. Of particular interest are CI techniques, especially the ANN, which can be trained to learn the relationship between different input and output variables. ANNs have long been used for data fusion, as demonstrated by Luo and Kay's 1989 review of data fusion techniques, which listed a number of ANN approaches [278]. ANNs have also been used in the domain of NDT (see, for example, work by Helifa et al. [279], Sambath et al. [280], and Sutcliffe

and Lewis [281]) and SHM (for example, the work done by Abdeljaber et al. [282], Chatterjee et al. [283] and Dworakowski et al. [284]).

One flavour of ANNs, the 'Deep Learning' or 'Convolutional' ANN, has been the subject of an increasing amount of research in recent years. For example, they have been used to combine image and radar data [285], image and text data [286], and LiDAR and aerial photography [287]. However, certain practical considerations have to be made relating to the implementation of any data fusion technique. There have even been attempts to integrate Deep Belief networks with ESNs [288–290], although this work is still at an early stage of development. One requirement for the approaches used here was that the computational requirements during both training and testing would not be prohibitive. For example, in Chapter 3, one desirable outcome of the work would be that a surface could be surveyed using the EMAD and covermeter, and the results viewed on-site. Deep Learning networks require considerable amounts of time and computational resources during training, and often have to be implemented on graphics processing units with greater parallel processing power [291–293]. In one example from 2012, two graphics processing units had to be used, since the 3 GB of memory available on a GTX 580 was insufficient for the Deep Learning network that was implemented, which also required 1.2 million samples of training data [294]. Sufficient levels of processing power and data would not be readily available for training in the applications considered here, and so Deep Learning networks were discounted as an option.

2.3 Echo State Networks

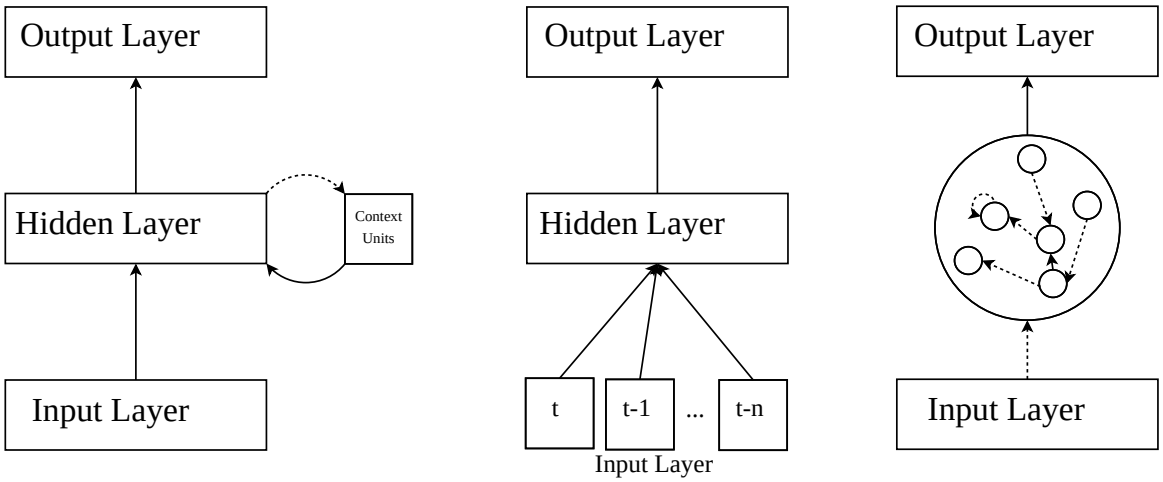
It was explained in Section 1.2 that ESNs are particularly suited to the complexities of the spatially and temporally longitudinal data used in Chapters 3 and 4. The remainder of this chapter is dedicated to fully detailing ESNs and exploring recent applications in a variety of fields.

2.3.1 ESNs in the field of RNNs

RNNs are a class of ANNs that fundamentally differ from more traditional feedforward ANNs, such as Multilayer Perceptrons (MLPs), in that they have at least one cyclic path of connections. They have been the subject of several reviews in the past, including those by Medsker and Jain [295] and Cardot and Bone [296].

While MLPs transfer data forwards through a number of different layers of neurons, an RNN will also include connections that feed data back into the network [297]. A simple Elman RNN model might have a similar architecture to an MLP, with an input layer, hidden layer and output layer, but with the distinction of having a 'contextual layer' of neurons that take the output of the hidden layer and feed it back into the hidden layer at the following timestep (see Figure 2.1) [298]. This contextual layer sets the simple Elman RNN apart from other varieties of ANN by allowing the network to 'remember' past network states, but at the cost of making the training procedure more complex. Although the gradient-based training procedures that are typically used by ANNs have been adapted for use with RNNs, they are susceptible to long training times, along with vanishing gradient (where the error gradient used for training weights quickly reduces to zero) and exploding gradient (where the error gradient grows large exponentially) problems [299].

One alternative machine learning approach to the sort of temporally or spatially varying data that can arise from NDT and SHM projects is the Time-Delay Neural Network (TDNN) [300]. The TDNN has previously been compared to ESNs in an NDT context by Butcher in 2012, and offers a way to include information about previous timesteps in a dataset by effectively expanding the size of the input layer through the use of delay lines [301]. While TDNNs fall are a type of feedforward network, rather than an RNN, the fact that the hidden layer has access to both the current timestep and multiple previous timesteps gives the networks the ability to learn temporal variations. A schematic diagram of the typical TDNN architecture is given in Figure 2.1, where it is compared to an Elman RNN and an ESN. Like an RNN, the number of weighted connections involved and the backpropagation training algorithm mean that TDNNs



(a) Jordan RNN (b) Time-Delay NN (c) Echo State Network

Figure 2.1: A side-by-side comparison of the schematic diagrams for (a) Elman RNNs, (b) TDNNs and (c) ESNs. Weighted connections that are trained are represented using a solid arrow, while weighted connections that are randomly generated at initialisation and subsequently fixed are represented with dashed arrows. Only the sparsely interconnected reservoir of the ESN allows for a more flexible short term memory than the fixed memory lengths of Elman RNNs and TDNNs.

are susceptible to long training times and vanishing gradient problems [302].

As stated in Section 1.2, ESNs are a form of RNN that use a sparsely interconnected reservoir of neurons in order to learn temporal patterns. In effect, they provide a new training rule that is designed to overcome the problems with RNN training procedures [26]. The presence of context units in a traditional RNN adds complexity and irregularity to the network weight search-space, which, in turn, increases both the time required to find optimal network weights and the chance of becoming trapped in local minima. In contrast, the ESN approach is to keep all of the weighted connections constant throughout training, with the exception of those between the reservoir and output layer. Instead of using more complex training methods such as backpropagation through time, a linear approach such as ridge regression will typically be used.

There is an additional drawback to using a standard RNN or TDNN for processing timeseries data: the strictly controlled temporal extent of the network's memory. The context units in an Elman RNN only allow for one additional timestep of information to be fed into the network, while the number of timesteps used by a TDNN is determined prior to network initialisation. Finding the best number of timesteps to include requires

a degree of optimisation, and the greater the number of timesteps, the greater the computational resources required for training. A fixed window of past inputs may also not necessarily be suitable for all applications. For example, in the NDT case study given in Chapter 3, the range of datapoints over which magnetic anomalies can be seen varies depending on a number of factors.

By way of contrast, carefully controlling the degree of recurrence in the reservoir gives ESNs their key characteristic: the echo state property [22]. This is simply defined as the state of the ESN, $\mathbf{S}$ at any time t depending not only on the input at t, but also on the input at all previous time steps. The state of the network at t is governed by an echo function, $\mathbf{E}$, which maps the history of the matrix of network inputs, $\mathbf{u}$, to the current network state, as shown in Equation 2.1 [303, 304]:

$$\mathbf{S}(t) = \mathbf{E}(..., \mathbf{u}(t-1), \mathbf{u}(t)) \tag{2.1}$$

Rather than having a fixed memory length, the sparsity of the reservoir allows for a more flexible short-term memory, making ESNs more suitable for the two case studies presented here than the standard RNNs or TDNNs.

2.3.2 ESN parameters

One advantage of ESNs is the ability to tune a range of network parameters, adjusting the behaviour of the ESN and potentially allowing for improved performance in practical applications. The remainder of this section deals with these tunable parameters, a general guide to which is given by Lukosevicius [29].

2.3.2.1 Leak rate

Although ESN reservoir neurons were initially simple additive units with sigmoid activation functions, most ESN reservoirs now use leaky integrator neurons, first proposed by Jaeger in 2007 [305]. The activation of a leaky integrator neuron in an ESN reservoir is given by Equation 2.2 [306]:

$$\mathbf{x}(t) = f((1-\delta)\mathbf{x}(t-1) + \delta(\mathbf{W}_{inp}^{res}\mathbf{u}(t) + \mathbf{W}_{res}^{res}\mathbf{x}(t-1))) \tag{2.2}$$

In Equation 2.2, f is the activation function (typically tanh) $\mathbf{W}_{inp}^{res}$ is the input layer to reservoir weight matrix, $\mathbf{x}(t)$ is the vector of the activations of the reservoir neurons at t, $\mathbf{W}_{res}^{res}$ is the reservoir weight matrix (drawn randomly from a Z distribution) and δ is the leak rate, which determines the extent to which ESN reservoir neurons' activations decrease over a period of time. In this way, the leak rate can be used to modify the behaviour of the reservoir, since higher leak rate values have been shown to lead to faster reservoir dynamics [307, 308]. It can also be seen from Equation 2.2 that adjusting the leak rate allows for the network's dependence on both past and present inputs to be fine tuned. A leak rate value very close to unity will increase the ability of the ESN to recall past inputs, but will also decrease the ESN's ability to recall the most recent inputs [306].

2.3.2.2 Spectral radius

The short term memory of an ESN is also affected by the spectral radius. The spectral radius is used to uniformly scale the reservoir weights, and the dependence of the final reservoir weights on the spectral radius can be seen in Equation 2.3:

$$\mathbf{W}_{res}^{res} = \frac{\alpha \times \mathbf{W}_{res}^{\prime res}}{|\lambda_{max}|} \tag{2.3}$$

In Equation 2.3, $\mathbf{W}_{res}^{res}$ is the final reservoir weight matrix, λ_{max} is the maximum eigenvalue of $\mathbf{W}_{res}^{\prime res}$, which represents the initial reservoir weight matrix, and α is a scaling factor. At this point, it is important to note that there is some confusion over these terms in the literature. Initially, the spectral radius was defined as λ_{max}, with α termed the spectral radius scaling factor [303, 306]. However, it has since become common to refer to α as the spectral radius (see, for example, work by Tong et al. [309] and Venayagamoorthy and Shishir [310]), since α is a tunable parameter and λ_{max} is simply a property of the initial reservoir weight matrix that was generated. In this

thesis, the latter convention is used, and the term 'spectral radius' refers to α. An ESN with a value of α smaller than one is guaranteed to have the echo state property, although it is possible to obtain the echo state property with a value of α greater than one, if zero is not an admissible input [22, 303]. After the work of Buehner and Young in 2006 [311] there was, for a period, a commonly held assumption that having a spectral radius less than unity was not just sufficient for the echo state property, but necessary. However, this was corrected by Yildiz et al., who showed that the echo state property can be obtained for spectral radii much greater than one, so long as the driving input is sufficiently strong [312]. Small values of α generally yield a network with a very short memory of past inputs, while values of α closer to one tend to lead to networks that are able to recall further into the past [313, 314].

2.3.2.3 Input scaling

It is also possible to uniformly scale $\mathbf{W}_{inp}^{res}$, in order to adjust the extent to which the reservoir neurons are driven by the present input, in accordance with Equation 2.2. The dependence of $\mathbf{W}_{inp}^{res}$ on the input scaling, ι, can be seen in Equation 2.4, where $\mathbf{W}_{inp}^{\prime res}$ is the initial input weight matrix.

$$\mathbf{W}_{inp}^{res} = \iota \times \mathbf{W}_{inp}^{\prime res} \tag{2.4}$$

Finding an optimal value for the input scaling is just as important as finding an optimal value for the spectral radius, since an excessively high value can cause the present vector of inputs to dominate the reservoir and wipe out any useful internal dynamics [315]. The input scaling also significantly affects the behaviour of the reservoir: higher input scaling values lead to increasingly non-linear behaviour [316], while low input scaling leads to linear behaviour [44]. When using a tanh activation function, large input scaling causes reservoir neurons' activations to take values at the extreme points of the tanh curve, meaning that they effectively act in a saturated manner [29].

2.3.2.4 Reservoir size

While the number of input and output neurons is usually determined by the dimensionality of the input data and the desired network output, the number of neurons present within the ESN reservoir (the reservoir size) can be optimised according to the task. Having a sufficiently large reservoir is important, since it has been found that reservoir size is critical to an ESN developing suitable internal dynamics [315]. It has been shown in the past that a larger reservoir leads to both improved performance and greater memory capacity [317]. Indeed, the memory capacity of an ESN is limited by reservoir size, and the number of time steps that an ESN can recall cannot exceed the number of neurons in the reservoir [22]. Furthermore, as reservoir size increases, the variability in trained ESNs decreases, and at a reservoir size of 20,000 neurons, the difference between trained ESNs is usually negligible [318]. A large reservoir is, therefore, desirable in most applications.

Unlike the other ESN parameters, the reservoir size can determine computational tractability. Although the relatively efficient training process for ESNs allows very large reservoirs to be trained relatively quickly, increasing the reservoir size leads to an associated increase in computational cost. It is, therefore, usually recommended that the reservoir should be as big as is computationally affordable, and that the minimum reservoir size for a task should be equal to the number of time steps of memory required [29]. However, this may not be true for some tasks, and the best reservoir size for any given dataset should be determined by analysing test data performance. This thesis examines the use of ESNs in 'real-world' applications, where short training times and near-immediate responses to input data are preferable, so as to provide real-time monitoring. Consequently, there must be some trade off between reservoir size and training time. It was decided that reservoir size would usually be limited to 500 neurons, which is large enough to allow for good performance, but not so large that training and response times are prohibitive.

2.3.2.5 Reservoir connectivity

The reservoir connectivity determines the fraction of the reservoir neurons that are connected together, such that reservoir connectivity of unity leads to every single weight in $\mathbf{W}_{res}^{\prime res}$ being non-zero, and connectivity of zero means that every single weight in $\mathbf{W}_{res}^{\prime res}$ is zero (i.e. there are no internal reservoir connections). While some improved performance has been obtained by optimising the reservoir connectivity (in the work of Song and Feng, for example [319]), it is generally considered that the connectivity has a minimal effect on performance and the richness of reservoir dynamics if the reservoir size is sufficient [315, 320]. The reduced number of connections that results from a decreased reservoir connectivity does have the advantage of allowing for faster ESN processing [29].

2.3.2.6 Reservoir adaptation

Before training begins, intrinsic plasticity can be used to optimise the reservoir. Intrinsic plasticity was first proposed in 2005, and is based on the ability of biological neurons to change their excitability by modifying voltage gated channels [321]. Intrinsic plasticity is an input-driven unsupervised learning rule that allows reservoirs to adapt their internal dynamics to a given task, regardless of the architecture chosen. The rule chosen here was first detailed by Schrauwen et al. in 2008, and was the first intrinsic plasticity rule designed for reservoir neurons with tanh activation functions [322]. This states that if a gain factor, a, and a bias, b, are added to the activation function of each neuron in the reservoir, as shown in Equation 2.5, then a and b can be updated according to Equations 2.6 and 2.7.

$$f_{gen}(x) = f(ax + b) \tag{2.5}$$

$$\Delta b = -\eta\Big(- \frac{\mu}{\sigma^2} + -\frac{y}{\sigma^2}\Big(2\sigma^2 + 1 - y^2 + \mu y\Big)\Big) \tag{2.6}$$

$$\Delta a = \frac{\eta}{a} + \Delta b x \tag{2.7}$$

In Equations 2.6 and 2.7, y is the output activity of the neuron, η is the learning rate (set to 0.001), μ is the desired mean of the distribution of neuron outputs (usually zero) and σ is the desired standard deviation of the distribution of neuron outputs (set to 0.1). The idea is to maximise the amount of information on its inputs that is contained in each neuron's output and to drive the distribution of the neuron outputs so that it is exponential.

The input data is randomly fed into the network for a set number of epochs, which can be adjusted for optimal performance. Like reservoir size, there are practical limits to the number of epochs of reservoir adaptation that can be performed, since adaptation is time consuming for large reservoirs or large datasets. Different numbers of adaptation epochs were trialled for each one of the ESNs used in this thesis.

2.3.2.7 Neuron activation function

The response of an individual ESN neuron for any given input is determined by the neuron activation function chosen, in accordance with Equation 2.2. While a tanh activation function is usually chosen, several other functions can be used. Schrauwen noted that while many different types of function have been used, there is no clear understanding of why some perform better than others in different tasks [30]. Some of the possible options are detailed below.

Firstly, let Z(t) be equal to the weighted sum of the inputs to an ESN neuron at time t, prior to the application of the activation function,

$$\mathbf{Z}(t) = (1-\delta)\mathbf{x}(t-1) + \delta(\mathbf{W}_{inp}^{res}\mathbf{u}(t) + \mathbf{W}_{res}^{res}\mathbf{x}(t-1)), \tag{2.8}$$

such that the response of a neuron, $\mathbf{x}(t)$, using the tanh activation function, is given by Equation 2.9.

$$\mathbf{x}(t) = tanh(\mathbf{Z}(t)) \tag{2.9}$$

The tanh activation function offers linear behaviour for inputs around zero, but will act in an almost binary fashion if the weighted inputs operate at the function's extremities.

One commonly used alternative to tanh is the Fermi activation function, given in 2.10 [322]. Past work has reported success when using this function, particularly when using reservoir adaptation [323, 324]. While it is similar to the tanh function, it is bound to the range [0,1], whereas the tanh function is bound to the range [-1,1]. Consequently, an extremely negative weight sum on inputs will result in the neuron giving an output very close to zero.

$$\mathbf{x}(t) = \frac{1}{1 + e^{-\mathbf{Z}(t)}} \tag{2.10}$$

The Lorentzian function was first proposed for use in neural networks by Giraud in 1995, citing the ease of controlling and understanding the stability of a network using such an activation function [325]. The equation for the Lorentzian function is given in Equation 2.11. Unlike the sigmoid tanh and fermi activation functions, the output of the Lorentzian function actually peaks for a weighted input sum of zero, and will give outputs close to zero for extreme inputs.

$$\mathbf{x}(t) = \frac{1}{1 + \mathbf{Z}(t)^2} \tag{2.11}$$

Similar to the Lorentzian is the radial basis function, given in Equation 2.12. Radial basis functions have long been associated with ANNs, with the first network using a radial basis activation function (radial basis network) being reported in 1988, and are useful for their ability to combine a linear dependence on weights with an ability to model nonlinear relationships [326].

$$\mathbf{x}(t) = e^{-\mathbf{Z}(t)^2} \tag{2.12}$$

The final possible activation function considered here is the triangular basis function, which has previously been successfully implemented in ANNs and is given in Equation 2.13 [327, 328]. It is similar in nature to the Lorentzian and radial basis functions in that it peaks when the sum of the weighted inputs is zero, but differs in that it it decreases linearly towards zero as the sum of the weighted inputs moves further away from zero.

$$\mathbf{x}(t) = 1 - |\mathbf{Z}(t)| \tag{2.13}$$

The five activation functions presented above were all used in this thesis, since each one could potentially subtly affect the behaviour of an ESN. The differences between each function mean that each one could potentially be suited for application to the datasets used in this thesis.

2.3.3 Recent advances in echo state networks

Since their introduction in 2001, ESNs have continued to be the subject of CI research. The broad focus of most of this CI ESN research has been to try to obtain a better analysis of the different properties of an ESN, in order to try and obtain a richer understanding of the different parameters and properties that make ESNs a powerful tool for both regression and classification. In addition, some CI researchers have modified the standard ESN architecture in an attempt to make bespoke networks for a particular task, or in order to integrate a state-of-the-art CI technique. One challenge when attempting to apply ESNs to a problem is optimising the different parameters for a specific task, and different methods for achieving this are also an active area of CI research. Another interesting area of research is in the physical implementation of ESNs using photonics, where laser systems are used to create a real, physical ESN, rather than relying on purely software simulations of their properties. ESNs typically use a simple, linear readout, although other readouts have been suggested in the literature, along with alternatives to the usual ridge regression approach to training. A summary of these most recent developments, along with the corresponding references, is given in

Table 2.1.

Type of Development	References
Property Analysis	[329–335]
Modification to ESN Architecture	[336–344]
Parameter Optimisation	[345–348]
Physical ESNs using Photonics	[349–351]
New Readout	[350, 352]
New Training Regime	[353–355]

Table 2.1: Recent developments in ESNs.

One further area of study is in the unconventional use of ESNs as a static pattern classifier. This approach involves 'clamping' the usually time-varying trained-network inputs by simply repeatedly presenting the same input pattern as part of a short static sequence [356–358]. This unusual ESN modality has been at the core of recently published CI research, some of which has arisen from the work described in this thesis (see Appendix B). Applying the clamped ESN methodology to hyperspectral data obtained from plant samples in a laboratory setting showed that, contrary to perceived wisdom, it is possible to obtain good classification accuracy without having to wait for the reservoir's internal states to settle, providing that the number of input presentations used with unseen test data is the same as that used in training. This interesting work is outside of the scope of this thesis, but is presented in Appendix B.

2.3.4 Typical echo state network applications

ESNs have been applied in a wide range of fields since they were first presented in 2001. In 2012, Lukosevicius gave several examples of the typical applications that ESNs have been used for [359], while a more recent paper by Goudarzi and Teuscher suggested a set of open problems facing reservoir computing in 2016, recommending the use of ESNs for high-dimensional video data [360]. The most recent applications of ESNs are summarised in Table 2.2.

What each of these applications has in common is that they exploit the short-term memory of ESNs for signal processing applications. For example, much of the medicine-based research concerned the processing of electroencephalographic signals for different

Application	References
Audio Processing	[361, 362]
Cryptography	[363]
Data Detrending	[364]
Energy	[365–369]
Industrial	[370, 371]
Medicine	[372–379]
Networking and Communications	[380–385]
Optics	[386]
SHM	[387]
Trajectory Estimation	[388, 389]

Table 2.2: Recent areas of application for ESNs.

purposes, while many of the energy and industrial applications used ESNs for processing longitudinal sensor data. In the case of cryptography, the ESN reservoir was used as an encryption and decryption key for data streams as they were sent and received. While they are a relatively recent addition to the neural network researcher's tool-kit, ESNs are becoming a well-established technique for dealing with temporally and spatially variant data, and are, hence, an appropriate tool for use in the case studies included in this thesis.

2.3.5 Echo State Network approaches to data fusion

Although the use of ESNs for heterogeneous data fusion in SHM and NDT is a novel contribution of this thesis, ESNs have been used to carry out some limited data fusion in the past. This section outlines these past approaches.

2.3.5.1 Human activity classification

One of the areas where ESNs have been most commonly used for data fusion is in human activity classification. This is largely due to the work by researchers from the University of Pisa. In the work published by Bacciu et al., some rooms were fitted with four sensors each, and the radio signal strength between these sensors and sensors strapped to a human subject was measured. ESNs were then used to estimate whether the person was about to change room or not based on their trajectory [390, 391]. Similarly,

Palumbo et al. published work where the strength of signals shared between mobile phone-based sensors and environmental sensors was gathered in order to classify the type of activity being performed by a subject, who would carry the phone about their person. A decision tree was used to determine a broad category for the activity, and then the information would be fed into an ESN in order to find the specific task that was done, such as falling, cycling, standing, walking, lying or bending [392, 393].

Away from the University of Pisa, other researchers have also attempted to classify different human activities by fusing data with ESNs. Scherer et al. combined video and audio data [394], and channels of modulated microphone data [395–397] in order to detect human laughter. Meanwhile, a 100% success rate was achieved when an ESN was used to fuse data gathered by nine separate video recorders in order to recognise when a tennis player had played a forehand shot [398].

2.3.5.2 Trajectory estimation

One of the most common uses of data fusion is target tracking, one part of which is trajectory estimation. ESNs are amongst the many techniques that have been used to accurately predict the trajectory of a target. Li et al. used ESNs as one part of a larger data fusion scheme, which used RADAR data from different sensors for final trajectory estimation [399]. In a similar work, Shao et al. also used ESNs as one part of a wider fusion scheme for vessel tracking. In this case, the ESN was used to predict a control input for a Kalman filter by taking in position data, velocity data and the previous control input [400]. The idea of employing an ESN as part of a larger fusion was also used by Tsai et al., who filtered noisy observation data using the corresponding robot image velocity, in order to produce a noise-free parameter value for a Kalman filter to use to estimate the optimal system state. The ultimate object of this work was visual state estimation, where dynamic moving targets are tracked [401].

2.3.5.3 Robotics

In 2015, Qin and Lei used ESNs to combine twelve sensors embedded on a human to move the arm on an NAO humanoid robot, with outputs corresponding to the behaviour of the arm joints on the robot [402]. The same style of robot had previously been used by Kanoi and Hartland, who embedded the robot with 15 sensors so that an ESN could detect walking instability that could lead to a fall [403].

2.3.5.4 Damage detection and remaining useful life estimation

The damage detection task is not unlike the work carried out in the later chapters of this thesis, in particular Chapter 3. Sun et al. used an ESN to predict the discrete cosine transform composite spectrum entropy for the prognosis of a hydraulic pump [404]. In work performed by Peng et al., ESNs were used to estimate the remaining useful life of aircraft engines, based on 21 engine information sensors and three operating information sensors [405].

2.3.5.5 Prediction of meteorological phenomena

The final field of study that have employed ESNs for data fusion is that of meteorological prediction. An ESN architecture produced by Ruffing and Venayagamoorthy used nine output units to predict solar irradiance at nine future instances, based on extraterrestrial irradiance, relative humidity, temperature, barometric pressure, and wind speed data [406]. Liu et al. used historical wind speed and temperature data as inputs to an ESN that forecast the short-term wind speeds [407].

2.3.6 Echo state network implementation

In this thesis, ESNs were implemented using the Reservoir Computing Toolbox for MATLAB [317].

2.4 Summary

This chapter introduced data fusion, and briefly reviewed some important past works, before then going on to describe ESNs in detail, building on the short introduction in Section 1.2. Different ESN parameters were discussed, along with recent advances in ESNs, typical applications of ESNs and ESN approaches to data fusion in other fields.

CHAPTER 3

Detection of Defects in Steel Reinforcing Bars

3.1 Introduction

This chapter concerns the application of ESNs to the EMAD technique, and then the fusion of EMAD data with cover depth data, using ESNs, in order to more accurately locate and identify defects in rebars.

There are four principle reasons why the fusion of cover depth data with EMAD data should be beneficial. Firstly, each relates to a physical property of the rebar: the presence of defects for the EMAD, and the distance of the rebar from the surface for the cover depth. Secondly, there is likely to be a relationship between cover depth, and rebar deterioration. This could potentially be exploited by a data fusion regime. Thirdly, there is a relationship between cover depth and the amplitude of the signals from the EMAD probe, which is discussed in more detail in this chapter. Finally, the cover depth is actually highly variable in 'real-world' settings, which is a source of noise in the aforementioned relationships.

The first of these reasons relates to the fact that both the EMAD and cover depth data look directly at specific properties of the rebar, namely the magnetic flux of the rebar after energisation and the distance between the rebar and the surface. Several of the other possible candidates for data fusion would give details on factors that may indicate an increased likelihood of corrosion, but do not necessarily relate to the current physical state of the rebar. For example, the sort of information about the probability

of corrosion that could be provided by the half-cell potential technique, or the data on the electrical resistivity of concrete provided by the electrical resistivity method could complement the EMAD probe data, but are determined more by the environmental conditions than the actual rebar itself. Similarly, RABIT, discussed in Chapter 1, used GPR to detect delamination and concrete degradation, both of which could potentially indicate an increased likelihood of corrosion leading to a defect, but which do not give an insight into the actual physical condition of the rebar. The fact that the cover depth is a particular physical property of the rebar means that it can then complement the EMAD technique usefully.

The second reason for the use of the cover depth concerns the relationship between cover depth and deterioration. Pailes noted that areas with a low cover depth will probably experience deterioration sooner than areas where the cover depth is greater [34]. It is, therefore, possible that if an area is found to have a particularly low cover depth, the likelihood that a magnetic anomaly in the EMAD probe data relates to a defect increases. As a result of this, information on the cover depth should increase the confidence in the results of the EMAD probe data.

The third reason is that the cover depth in a 'real-world' setting could vary significantly. For example, the UK Highways Agency standard BD 44/95 states that the minimum depth of cover for a concrete surface that will be affected by de-icing salts is 50.0 mm in the UK for grade 40 concrete [408]. European Standard EN 1992-1-1, approved in 2004, gives Equation 3.1 for the calculation of the minimum appropriate cover depth, C_{min} [409]:

$$C_{min} = max\{C_{min,b}; C_{min,dur} + \Delta C_{dur,\gamma} - \Delta C_{dur,st} - \Delta C_{dur,add}; 10 \text{ mm}\} \qquad (3.1)$$

In Equation 3.1, $C_{min,b}$ is the minimum cover due to bond requirement (equal to the diameter of the rebar for separated rebar that are not part of a bundle), $C_{min,dur}$ is the minimum cover due to environmental conditions (between 40 mm and 65 mm for concrete subject to a cyclic wet and dry environment), $\Delta C_{dur,\gamma}$ is an additive safety

element (0 mm in the UK national annex), $\Delta C_{dur,st}$ is a reduction of minimum cover if stainless steel is used and $\Delta C_{dur,add}$ is a reduction of minimum cover if additional protection is used. In 2015, Ekolu and Solomon surveyed three newly constructed highway bridges in South Africa and found that the average cover depth was 51 mm, with a maximum recorded cover of 90 mm and a minimum of 20 mm [410].

Finally, the magnitude of MFL defect signals is reduced according to distance from the defect. In 1998, Hillemeier et al. found that the peak to peak amplitude of the x component of the magnetic field at a defect, A_{pp} (measured in mT), varies due to concrete cover c (measured in mm) according to Equation 3.2 [411].

$$A_{pp}(n, c) = \bar{A}_{pp}(n)c^{-1.75} \tag{3.2}$$

$\bar{A}_{pp}(n)$ is a fitting parameter that depends on the number of fractures in one rebar. As the concrete cover depth, and, hence, the distance of the scan from the rebar increases, the amplitude of recorded signals will significantly decrease. This phenomenon is shown in Figure 3.1, which depicts the X and Z axis components of the magnetic flux recorded at vertical distances of 42.5 mm and 85.0 mm away from a clean break in a rebar. The peak to peak amplitude of the X axis component reduces from 4769 at 42.5 mm to 1770 at 85.0 mm, while the peak amplitude of the Z axis component reduces from 4849 to 3171. This reduction becomes more pronounced as the distance increases.

This is important, because the data processing algorithm that has previously been used for processing EMAD data in commercial surveys (introduced in Section 3.2.2) considers both the Z axis value and the peak to peak X axis value when producing its final output. This suggests that it may produce the same output for both a minor magnetic anomaly scanned at a close distance, and a major defect recorded from further away. Similarly, an ESN that has only been trained to recognise the high-amplitude signals seen at a close distance, such as the one described in the following section, may not recognise low-amplitude signals from a greater distance as representing defects. In a 'real-world' scenario, this could lead to significant defects being ignored, or repairs taking place to address only minor anomalies. A data fusion technique that could

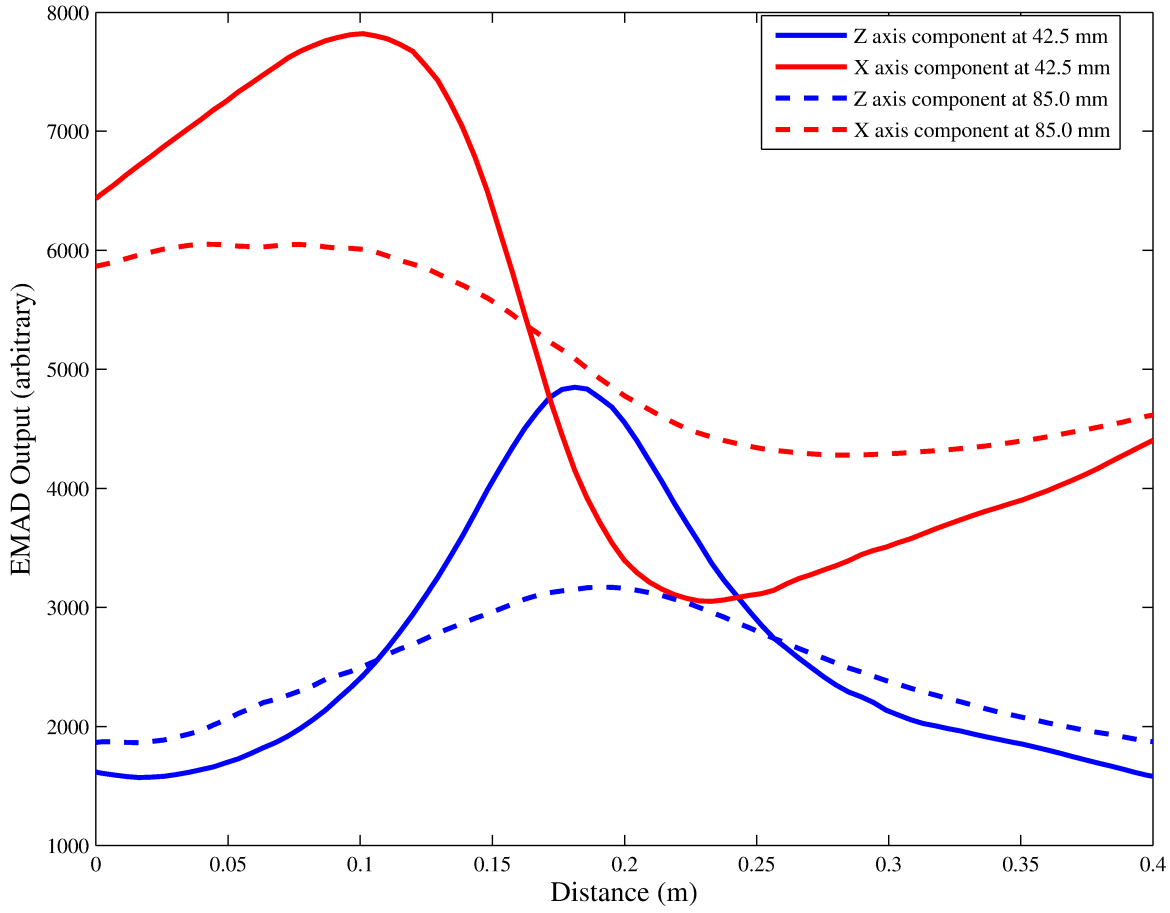


Figure 3.1: The X and Z components of the magnetic flux recorded at 42.5 and 85.0 mm above a defect signal. The peak amplitude of the z component signal and peak to peak amplitude of the x component both reduce as distance from the defect increases.

combine information about the cover depth with recorded EMAD data, would be able to avoid this problem.

The first part of this chapter showcases a preliminary investigation into the ability of ESNs to detect defects using EMAD data, in comparison with the current standard methodology. The remainder of the chapter then details the methodology and results of the attempt to obtain more accurate results by fusing EMAD and covermeter data.

3.2 Preliminary Work: Application of ESNs to the EMAD technique

Before the ESNs could be used for data fusion, their suitability for use with the EMAD technique was tested. To this end, a series of datasets were taken from a reinforced

concrete test bed, which featured manually inserted defects at known locations, using the EMAD technique. These datasets were then used to train and test ESNs, with ESN performance on the test datasets being compared to an Analytical Technique (AT) specifically designed to exploit the characteristic magnetic signature of a defect seen in the EMAD probe data.

3.2.1 The Keele University concrete test bed

The datasets were taken from the Keele University concrete test bed, created by Butcher as part of a PhD project [301]. The test bed was discussed by Butcher et al. in 2013 [45] and later used by Wootton et al. in 2014 [46]. The test bed featured a steel reinforcing mesh with an approximate area of 4.25 m x 1.755 m available for an EMAD survey. This mesh consisted of 5 mm thick rebars separated by 195 mm in both the longitudinal and transverse direction, encapsulated in concrete. Prior to encapsulation, a number of defects were manually inserted in the mesh. These defects took the form of either a clean break in a rebar, the application of corrosion products to a part of the mesh or a combination of both of these. Corrosion was accelerated by immersing the mesh in a sodium chloride solution. A total of twelve defects were initially inserted, but three of the corrosion product defects were lost after encapsulation due to passivation. A scanning area consisting of ten longitudinal lines, each aligned with a longitudinal rebar, was set up, the layout of which, along with the location of the nine remaining defects and the transverse rebars, can be seen in Figure 3.2.

For most of the datasets gathered, the energiser was run along each longitudinal line in turn from left to right, bottom to top, as shown in Figure 3.2. After the whole mesh had been energised, each longitudinal line was scanned with the EMAD probe. Once the data were all gathered, the energiser was then run along each line in the opposite direction and in reverse order, so that the mesh was left with a remanent magnetic field with reversed polarity. This ensured that when the procedure was subsequently repeated a following 16 times, the energiser would definitely be changing the magnetic state of the steel prior to each EMAD scan. If this step was missed, it is possible that the mesh

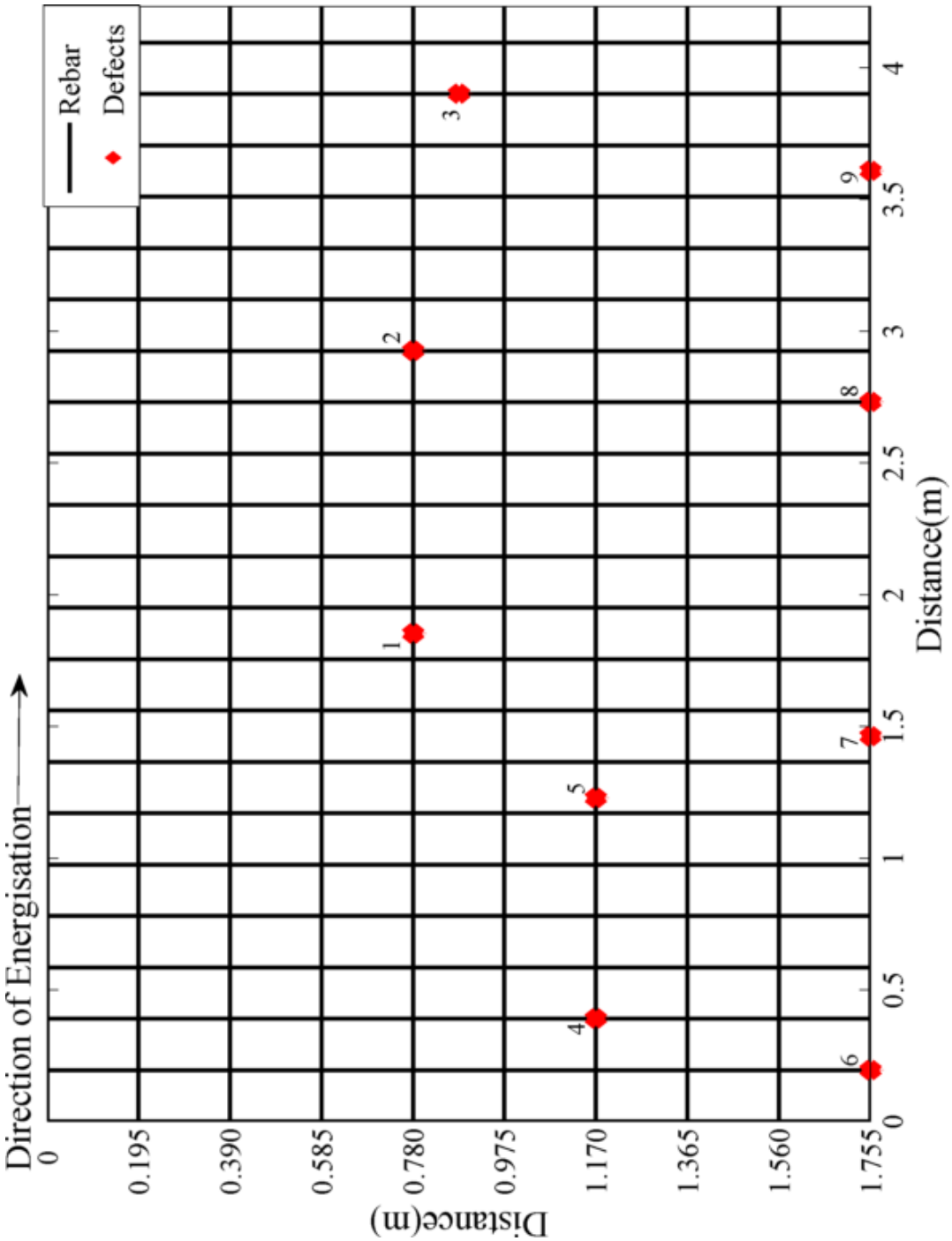


Figure 3.2: The layout of the Keele University concrete test bed. Black lines represent rebar, while red diamonds represent defect locations. The defect locations are numbered, and this numbering system is used for the duration of this section.

would already be saturated due to a previous scan and that the energiser would not change the remanent magnetic field. In each case, the mesh was both energised and scanned at a vertical distance of 41.0 ± 0.5 mm above each rebar. The scanning of the mesh took place intermittently over a two year period, ensuring that no two datasets were the same due to the degradation of the mesh over time and small variations in the energisation and scanning procedure. After the repeats had been completed, a significant bank of datasets for both training and testing the ESNs had been built up.

3.2.2 Analytical technique for processing EMAD data

In past commercial use, EMAD data was processed using a combination of expert interpretation and a procedural algorithm, referred to here as the AT. This AT was tailored to exploit the properties of the characteristic signature of a defect while disregarding noise, with the aim of highlighting areas featuring significant magnetic anomalies indicative of defects. An example of some EMAD data, featuring two defects at known locations, can be seen in Figure 3.3. The two defects can be recognised by their characteristic shape, as the recorded Z component data rises to reach a peak at the centre of the defect, while the recorded X component data exhibits a slope with a negative gradient at this point. For an EMAD scan carried out in a direction opposite to the direction of the rebars remanent magnetization, the Z component data would reduce to a negative peak at the centre of the defect and the X component data would exhibit a slope with a positive gradient. Because of this, the first step in the AT was to differentiate the Z component data with respect to distance and to then integrate it with respect to distance using the trapezium rule. The modulus of the integrated values was then taken. This centred the Z component data about zero and ensured that the direction of the scan would not affect the final output.

A regular ripple can be seen in the data shown in Figure 3.3, most notably in the range 0 – 1.5 m. This ripple was due to the presence of transverse bars rather than any defect and is consequently undesirable noise. Therefore, the next stage of the AT was to smooth both the integrated Z component data and the x component data using a

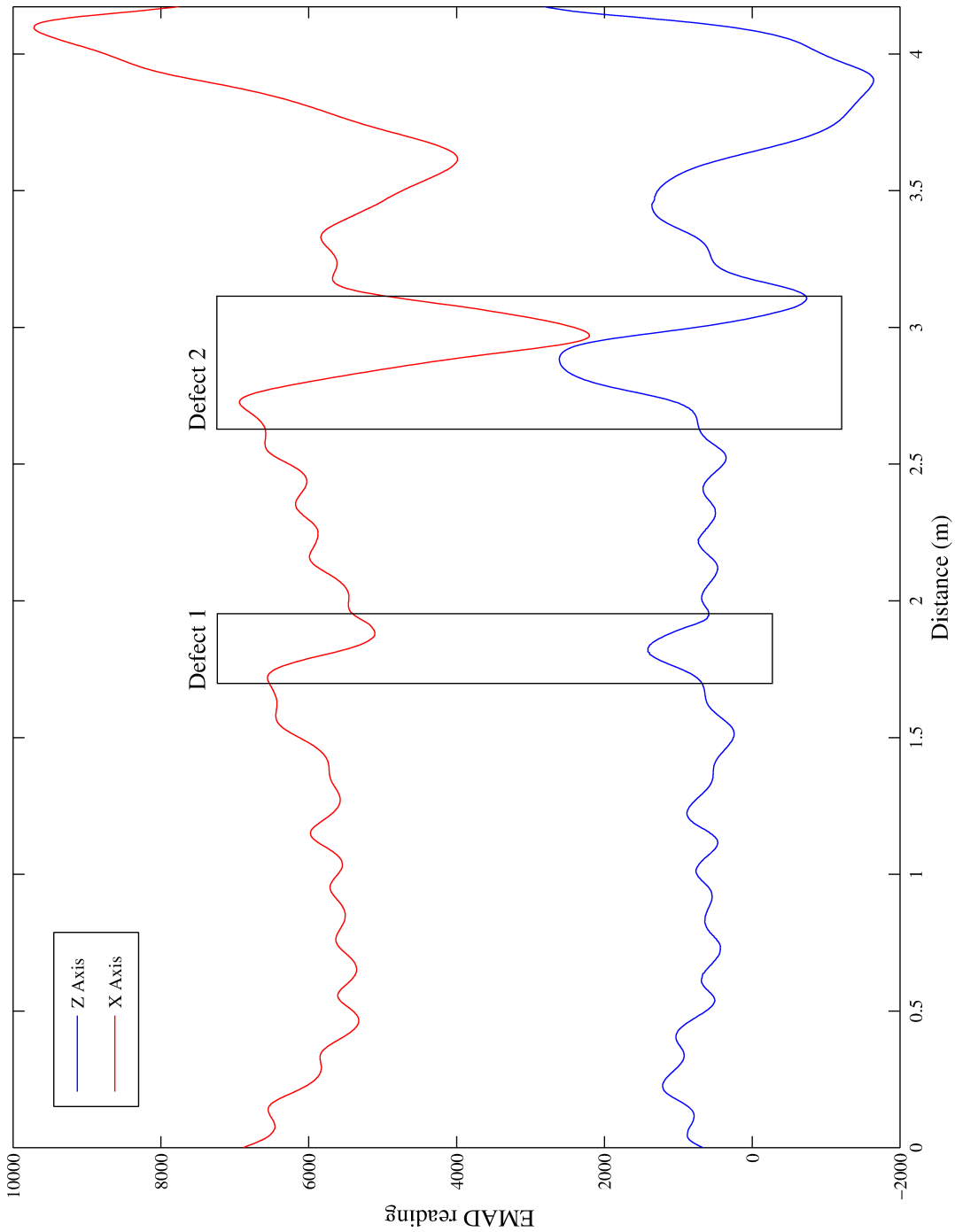


Figure 3.3: A typical set of EMAD data, with the Z axis data in blue and the X axis data in red. The characteristic signatures of the two defects present in this particular dataset are highlighted.

Gaussian filter. The output of a Gaussian filter for the ith point in a dataset is given by Equation 3.3:

$$G_i(x) = \frac{\sum_{n=-\sigma/2}^{n=\sigma/2} x_{i+n} e^{(d_i - d_{i+n})^2}}{\sigma + 1} \tag{3.3}$$

In Equation 3.3, x is the X component data value, d is the distance of the X component data point from the start of the scan line and σ is the number of data points to be smoothed over. When analysing the signal shown in Figure 3.3 by eye, the primary difference between the noisy ripples and the defect signals is the peak to peak height in the x component data. Small, sharp decreases in the X component data are not as significant as the much larger X component signals seen at points of interest. As a further attempt to minimise the impact of noise on the final output of the AT, a peak detection method based on Billauer's peakdet script for locating maxima and minima in noisy data [412] was incorporated into the AT. This method detected turning points in the X component data and then calculated the difference in magnitude of consecutive minima and maxima. An example of the output of the peak detection method for a set of X component data can be seen in Figure 3.4.

Since the characteristic defect signature features a slope in the X component data, the next stage of the AT was to calculate the slope by differentiating the X component data with respect to distance from the start point of the scan line. The final output of the AT, O, for any given data point was then given by Equation 3.4:

$$O = \frac{|P \times G(z_0 \times \frac{dx}{dd})|}{\bar{O}} \tag{3.4}$$

In Equation 3.4, P is the output of the peak detection method at that point, z_0 is the value of the centred Z component data at that point, $\frac{dx}{dd}$ is the value of the differentiated X component data, $\bar{O}$ is the mean of O and G represents the Gaussian smoothing function, which was applied to the product of the centred Z component data and the differentiated X component data. The overall AT can be summarised as follows:

1. differentiate the Z component data with respect to distance and then integrate it

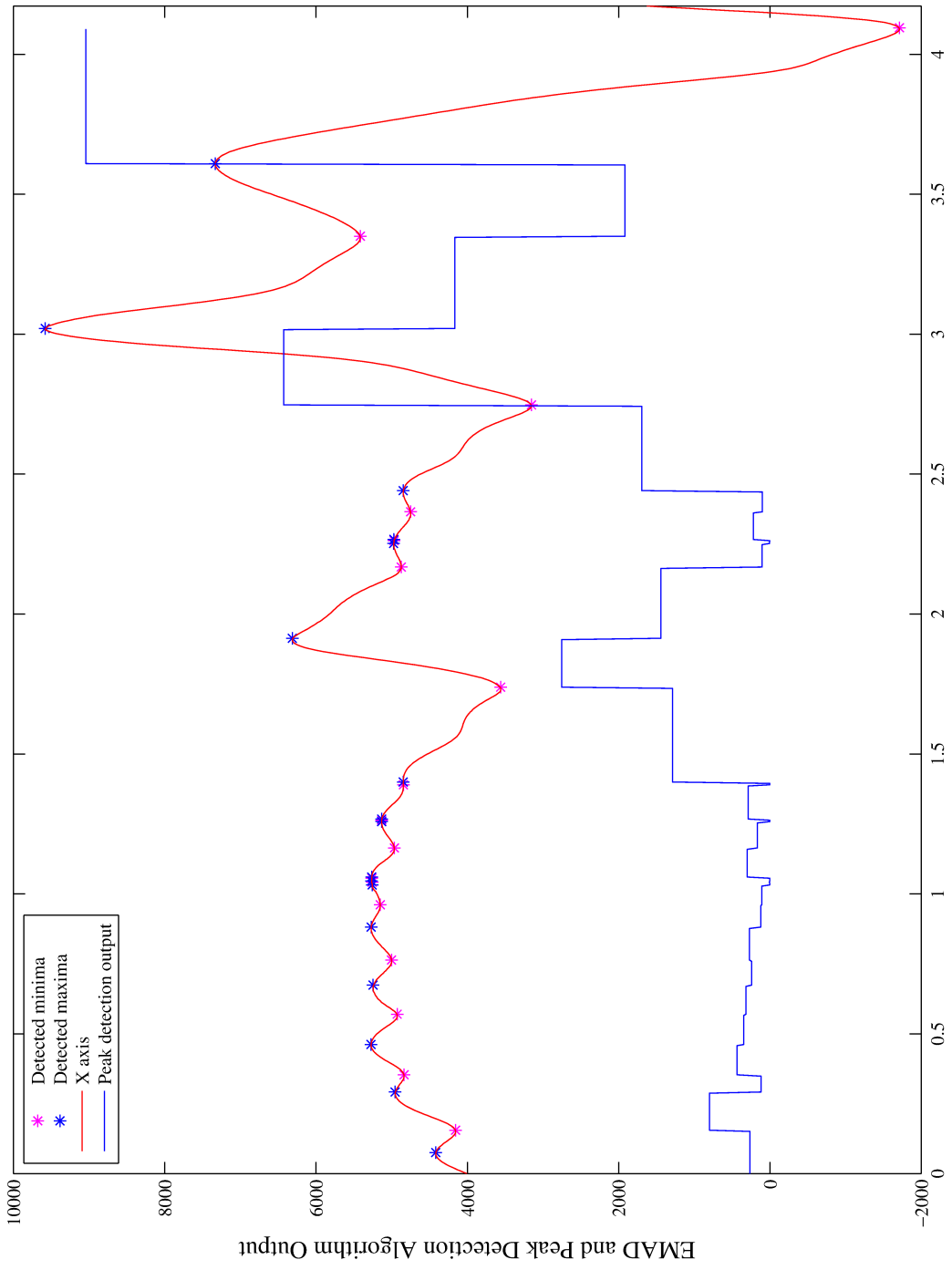


Figure 3.4: The AT's peak detection method. The peak to peak amplitude between adjacent peaks is calculated and then used as part of the AT.

using the trapezium rule, again with respect to distance;

2. take the modulus of this value;

3. apply a Gaussian filter to the X component data;

4. detect turning points in the X component data and calculate the difference in size between adjacent turning points;

5. differentiate the X component data with respect to distance;

6. multiply the differentiated X component data by the centred Z component data and apply a Gaussian filter to the resultant values;

7. multiply these values by the values obtained by the peak detection method and then

8. divide throughout by the mean of the output values.

The resultant output data could then be plotted on a two-dimensional contour plot and a threshold could be applied in order to highlight only the largest anomalies, since a larger output value would indicate a larger magnetic anomaly. An example of such a two-dimensional contour plot can be seen in Figure 3.5.

3.2.3 ESN approach

In order for ESNs to be used here, a suitable training regime had to be formulated. A total of 82 different individual scan lines from datasets gathered over a number of years were used for training. These were then categorised as either a line featuring a known defect, or a line without any known defects. This was so that during training, the ESNs would be presented with an equal number of lines featuring defects and lines without defects, which would prevent overtraining on either. Each scan line was then labelled so that at points without a defect the ESN had a target output of -1 and at points with a defect the ESN had a target output of +1. During training, each scan line was presented to the ESN in turn and 41-fold cross validation was used. Two datasets,

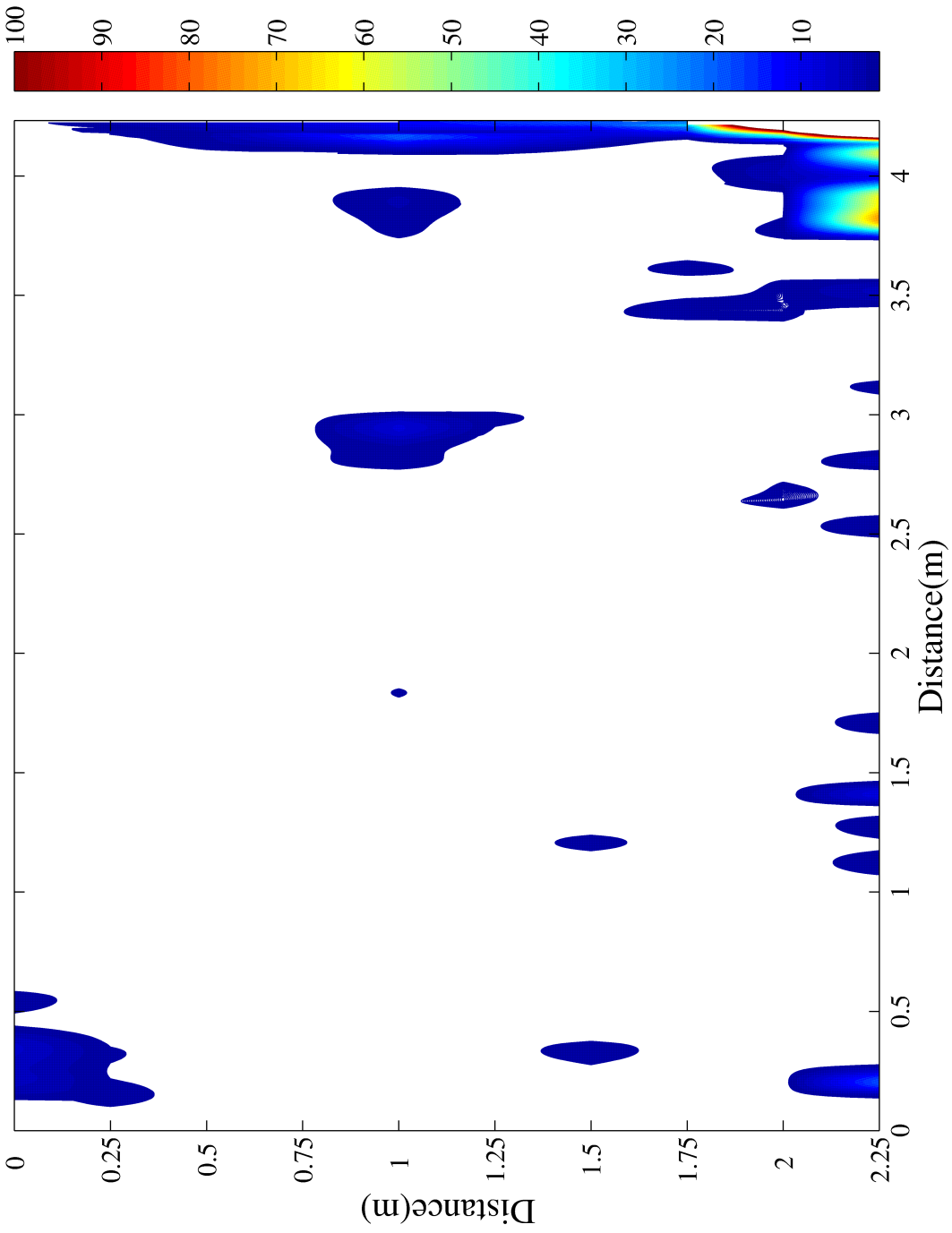


Figure 3.5: An example of a contour plot produced using the AT. Areas that the AT has determined are free of defects are white, while defect signals are coloured according to their magnitude.

Parameter	Range varied over	Optimal value
Spectral Radius	0 - 2	0.99
Input Scaling	0 - 2	1.8
Leak Rate	0 - 1	0.05
Adaptation Epochs	0 - 10	1
Reservoir Size	1 - 500	475
Reservoir Connectivity Factor	0.1 - 1	0.65
Activation Function	Tanh, Lorentzian, Triangular Basis Function, Radial Basis Function, Fermi	Tanh

Table 3.1: ESN parameters, the range that they were varied over in order to find the optimal ESN configuration and the optimal value found for each parameter.

henceforth referred to as dataset A and dataset B, were kept back so that they could be used as unseen testing data.

However, before the ESNs could be compared to the AT, a good set of ESN configuration parameters needed to be found for this problem domain. To this end, a grid search was performed, where ten ESNs would be trained with a certain configuration. These ten ESNs would then be presented with a test dataset and the average performance on the test dataset recorded. One parameter in this configuration would then be adjusted and the process would then be repeated until all of the parameters had been adjusted over a set range. These parameters, the range of the adjustment and the optimal value for each are given in Table 3.1.

500 ESNs with the optimal configuration were then trained, and their performance when presented with the two unseen test datasets compared to the performance of the AT on the same test data. The reason for training 500 ESNs was that the weighted connections between reservoir neurons are randomly generated, and an individual ESN may be particularly well (or poorly) configured, meaning that the performance of that one ESN would not be representative of the typcal performance that one might expect to obtain. The results from training 500 ESNs would be more representative of the typical performance in this problem domain.

Since this is a two-class (‘defect’ and ‘no defect’) problem, performance was assessed

using a Receiver Operator Characteristic (ROC) curve. Analysis of an ROC curve, where sensitivity is calculated against False Positive Rate (FPR) at several different thresholds, provides a richer measure of classification accuracy due to de-coupling of classifier performance from class skew and error costs, while the area underneath this curve, the Area Under Curve (AUC), is equivalent to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [413]. An AUC value of unity is indicative of perfect classification, while a value of 0.5 suggests that the results from the classifier are little better than guessing. In machine learning, the AUC has been found to be a better tool for the analysis of two-class classifier performance than a simple calculation of classifier accuracy [414, 415]. Furthermore, ROC curve analysis has been used for assessing performance in both NDT [416] and SHM [417]. In the 'real-world', the processed EMAD data is normally presented on a thresholded contour plot, so an evaluation method that considers performance at different thresholds is particularly useful. The average AUC for the ESNs was therefore compared to the AUC for the AT on both of the unseen test datasets. The FPR was calculated in accordance with Equation 3.5, while the sensitivity was calculated using Equation 3.6.

$$FPR = \frac{FP}{FP + TN} \tag{3.5}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.6}$$

In Equations 3.5 and 3.6, FP is the number of false positives, TN is the number of true negatives, TP is the number of true positives and FN is the number of false negatives. One practical consideration for use of the technique in the 'real-world' is the consistency of the optimal threshold, which can be found by determining the point on the ROC curve that had the smallest Euclidean distance to the point that would represent perfect classification, (0,1). When the ground truth is known, as in this case, finding the threshold that produces the clearest contour plots for defect location is

Classifier and Dataset	AUC	Optimal Threshold
ESNs on Dataset A	0.9369 (0.0187)	0.0683 (0.0076)
ESNs on Dataset B	0.9367 (0.0163)	0.0666 (0.0061)
AT on Dataset A	0.9578	4.3468
AT on Dataset B	0.9436	0.9276

Table 3.2: The average AUC found by 500 trained ESNs and the AT on datasets A and B, with the standard deviation for the ESNs given in brackets. The optimal threshold, determined by ROC curve analysis, is also given.

trivial. However, in the 'real-world', the ground truth is not usually known, meaning that expert analysis is often required when setting the best threshold. If the calculated threshold is consistent across different datasets, then expert analysis is not as important. It was, therefore, considered that for ESNs to be suitable for analysing data taken with the EMAD, the average optimal threshold should be consistent for both datasets.

Following the ROC analysis, a qualitative analysis of the output of the best performing ESN and the AT was also performed.

3.2.4 Results and Discussion

The two unseen test datasets were presented to the AT and 500 ESNs in turn, with the resulting average AUC for this given in Table 3.5. Note that in this section, known defects are referred to by their label in Figure 3.2.

The values in Table 3.2 represent good performance by both techniques on both of the test datasets. On average, the AT marginally outperformed the ESNs in terms of AUC. However, the main object of the preliminary work was not necessarily to outperform the AT, but to obtain performance that was comparable to that of the AT and, hence, demonstrate that ESNs are a suitable approach to fusing EMAD data with cover depth data. The fact that the ESNs achieved a high value for the AUC suggests that it was able to learn to model the temporal relationship between the X and Z component data. The ability to do this is absolutely key for any approach that seeks to interpret EMAD data. The fact that this was shown with a relatively simple ESN model that had only two input units is interesting, and indicates that a more sophisticated ESN model could potentially outperform the AT.

The results for the optimal threshold, where the ESN can be said to have outperformed the AT, are equally interesting. The difficulty in judging an appropriate threshold when using the AT is made clear by the discrepancy between the optimal threshold for dataset A, 4.3468, and the optimal threshold for dataset B, 0.9276, especially since the two datasets were obtained from the same reinforcing mesh. In direct contrast, the ESNs had, on average, very similar optimal thresholds of 0.0666 for dataset A and 0.0683 for dataset B.

A more subjective comparison between the output of the AT and the best performing individual ESN reveals some interesting details. It should be noted that this individual ESN actually outperformed the AT on its own, giving an AUC of 0.9697 and optimal threshold of 0.0757 when presented with dataset A, and an AUC of 0.9632 and optimal threshold of 0.0886 when presented with dataset B. The optimally thresholded output of each technique when applied to dataset A can be seen in the contour plots in Figure 3.6 for the AT and Figure 3.7 for the ESN. The centres of the ground truth defect locations are marked with red diamonds, while the contours give damage locations indicated by the technique, in accordance with the colour bar on the right.

It is immediately apparent from Figures 3.6 and 3.7 that neither approach works completely perfectly. The ESN managed to detect all of the ground truth defects, but also erroneously indicated the presence of defects at 'clean' locations. The AT indicated slightly fewer false defects, but also failed to detect one of the real defects. This is a significant failing, since this defect, located at approximately (2, 0.78), was a clean break in the steel. Furthermore, the signal for the defect at (3.6, 1.755) is very weak compared to some of the other signals, while the defect at (3.9, 0.975) is lost amongst a number of false positives.

Had these results been recorded as part of a typical 'real-world' survey, where the ground truth would not have been known, using the AT output seen in Figure 3.6 would lead to a very different set of recommendations to using the output of the ESN seen in Figure 3.7. Expert analysis of the output of the AT would probably lead to the following interpretation:

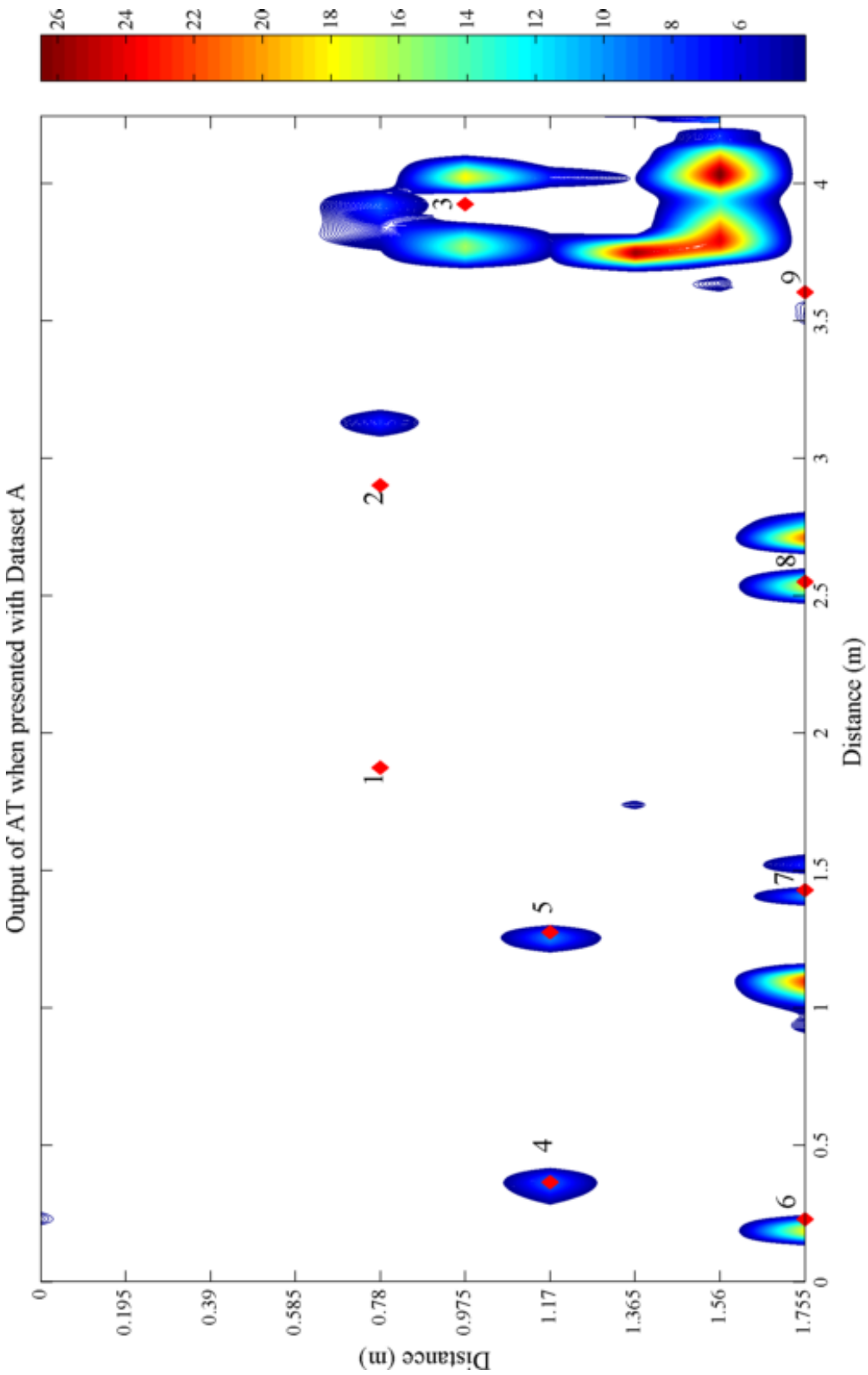


Figure 3.6: A contour plot of the output of the AT when applied to Dataset A, with the threshold chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations.

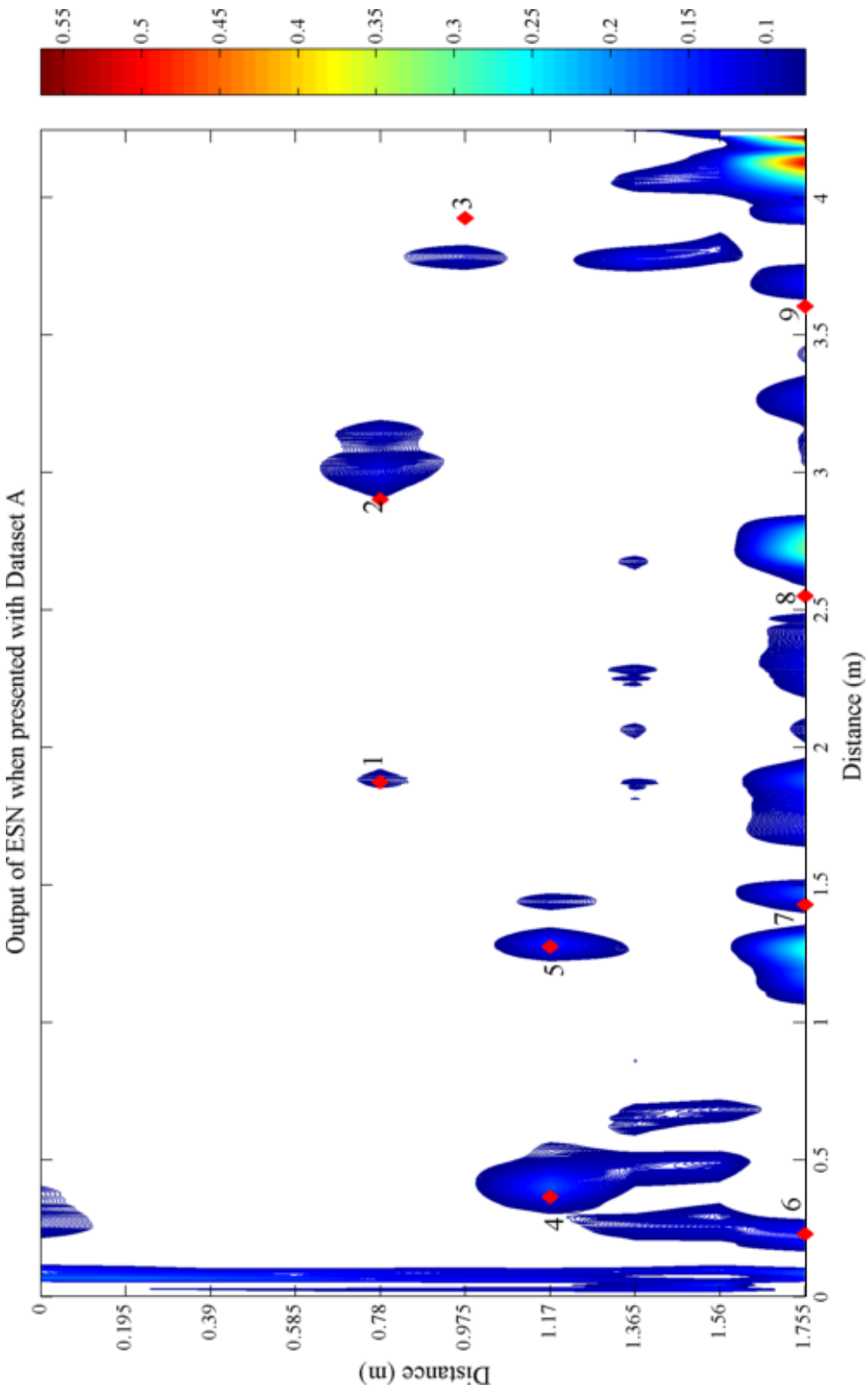


Figure 3.7: A contour plot of the output of the ESN when applied to Dataset A, with the threshold chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations.

1. Large anomalous signals can be seen at (0.2, 1.755), (1.1, 0.1755) and (2.5 - 2.75, 1.755). These are the most likely regions for rebar defects and the surface should be broken out to inspect the condition of the rebar at these points.

2. Smaller, but potentially significant signals can be seen at (0.4, 1.17), (1.3, 1.17), (1.4, 1.755) and (3.3, 0.78). The strongest of these signals is the one located at (1.3, 1.17), and the surface should be broken out here to assess the state of the rebar. If damage is present, the other anomalies should also be examined.

3. There are at least six large, anomalous signals in the region (3.6 - 4.2, 0.585 - 1.755). These are probably just 'end effects' present in the data owing to the end of the energisation region and can be ignored.

4. If large faults are found at the locations given in points 1 and 2, it may also be worth assessing the rebar at (1.75, 1.365).

5. The small signals at (0.25, 0) and (3.5, 1.755) are probably low level noise and can be discounted.

Based on this interpretation, even though the AT technically flagged up eight out of the nine defects, defects 1, 3 and 9 would all be missed, while defect 2 could also be missed due to the large offset between the actual defect location and the AT signal. Furthermore, the surface could potentially be broken out in two areas where there are no defects present. In short, use of the AT would result in most of the defects being investigated, but also in a number of clean areas of rebar being examined and some actual defects being missed, reducing confidence in the efficacy of the EMAD technique.

In contrast, analysis of the output of the ESN seen in Figure 3.7 would probably lead to the following interpretation:

1. Large anomalous signals can be seen at (0.4, 1.17), (1.25, 1.755), (1.45, 1.755) and (2.75, 1.755). These are the most likely regions for rebar defects and the surface should be broken out to inspect the condition of the rebar at these points.

2. Smaller, but potentially significant signals can be seen at (1.3, 1.17), (1.9, 1.755), (2.3, 1.755), (3, 0.78), (3.25, 1.755), (3.7, 1.755) and (3.75, 0.975). The strongest of these signals is the one located at (1.3, 1.17), and the surface should be broken out here to assess the state of the rebar. If damage is present, the other anomalies should also be examined.

3. The large anomalies in the region (0 - 0.15, 0 - 1.755) and (4 - 4.25, 1.56 - 1.755) are probably just 'end effects' present in the data owing to the end of the energisation region and can be ignored.

4. Several anomalies appear in the range (0 - 4.25, 1.375 - 1.56). The regular nature of these anomalies suggests that they are due to magnetic interference from transverse rebar, and can be discounted.

5. If large faults are found at the locations given in points 1 and 2, it may also be worth assessing the rebar at (0.3, 0) and (2.3, 0.78). The signal at (0.25, 1.755) is slightly offset from the regular noise mentioned in point 4 and may also merit further investigation.

Carrying this analysis forwards would lead to the successful location of all of the defects, although there would also five unnecessary instances of the surface being removed and the offset between defects 3 and 8 and the ESN output could result in them being missed. Overall, the ESN would give better results in a 'real-world' survey, but, importantly, would still not be completely reliable.

Analysis of Figures 3.8 and 3.9 leads to the same conclusion. The output of both techniques when applied to dataset B is very similar to their output when applied to dataset A, and the expert analysis of the ESN output in Figure 3.9 would be the same as the expert analysis of Figure 3.7. For the AT, the analysis may slightly differ, as the presence of a very large end effect signal at (3.8, 1.755) makes the signals for all of the genuine defects appear to be significantly smaller. This could lead to defects 2 and 4 also being lost. While the AT does, on average, marginally outperform the ESNs in terms of AUC, the ESN used here would be of more practical use in the 'real-world'.

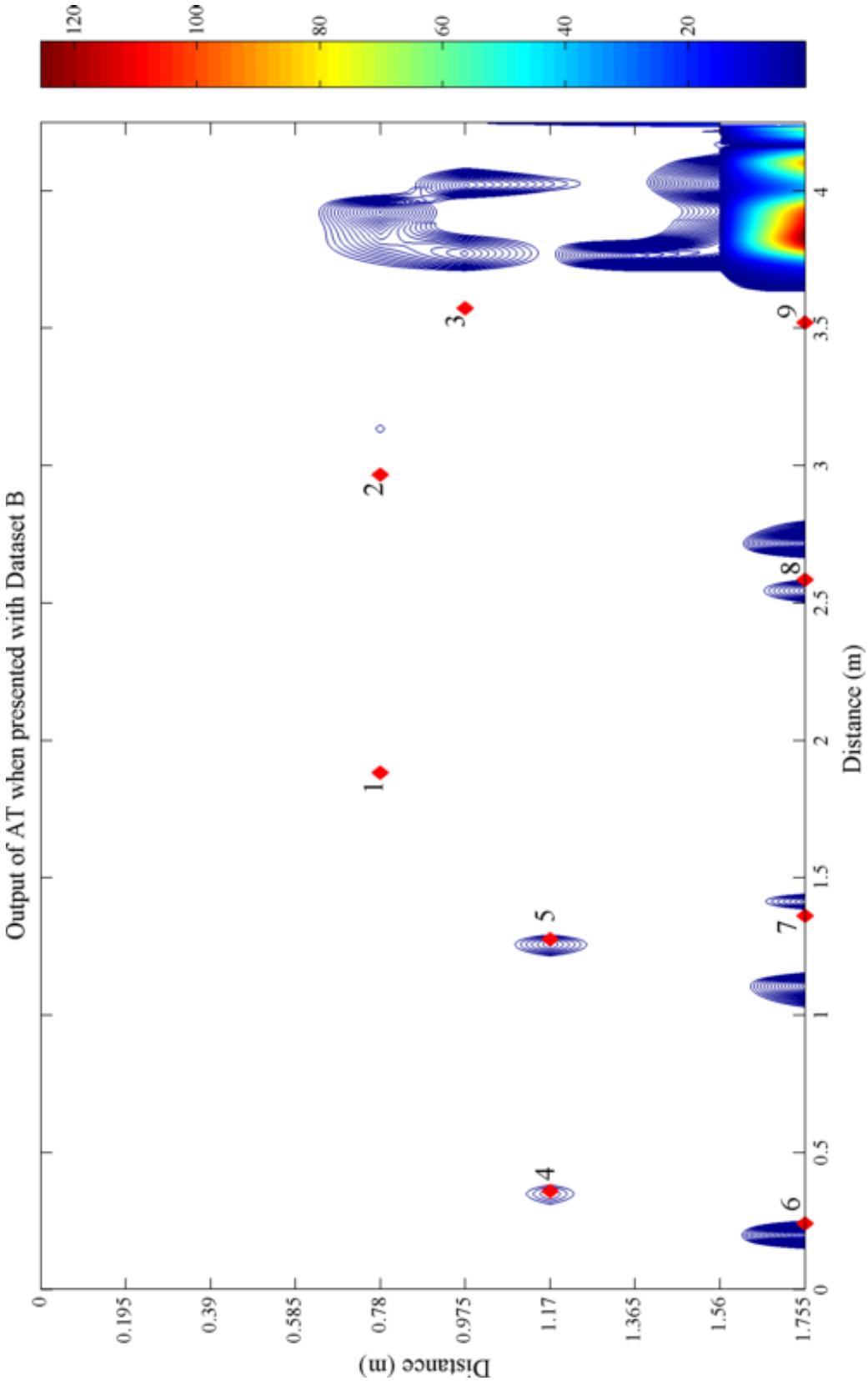


Figure 3.8: A contour plot of the output of the AT when applied to Dataset B, with the threshold chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations.

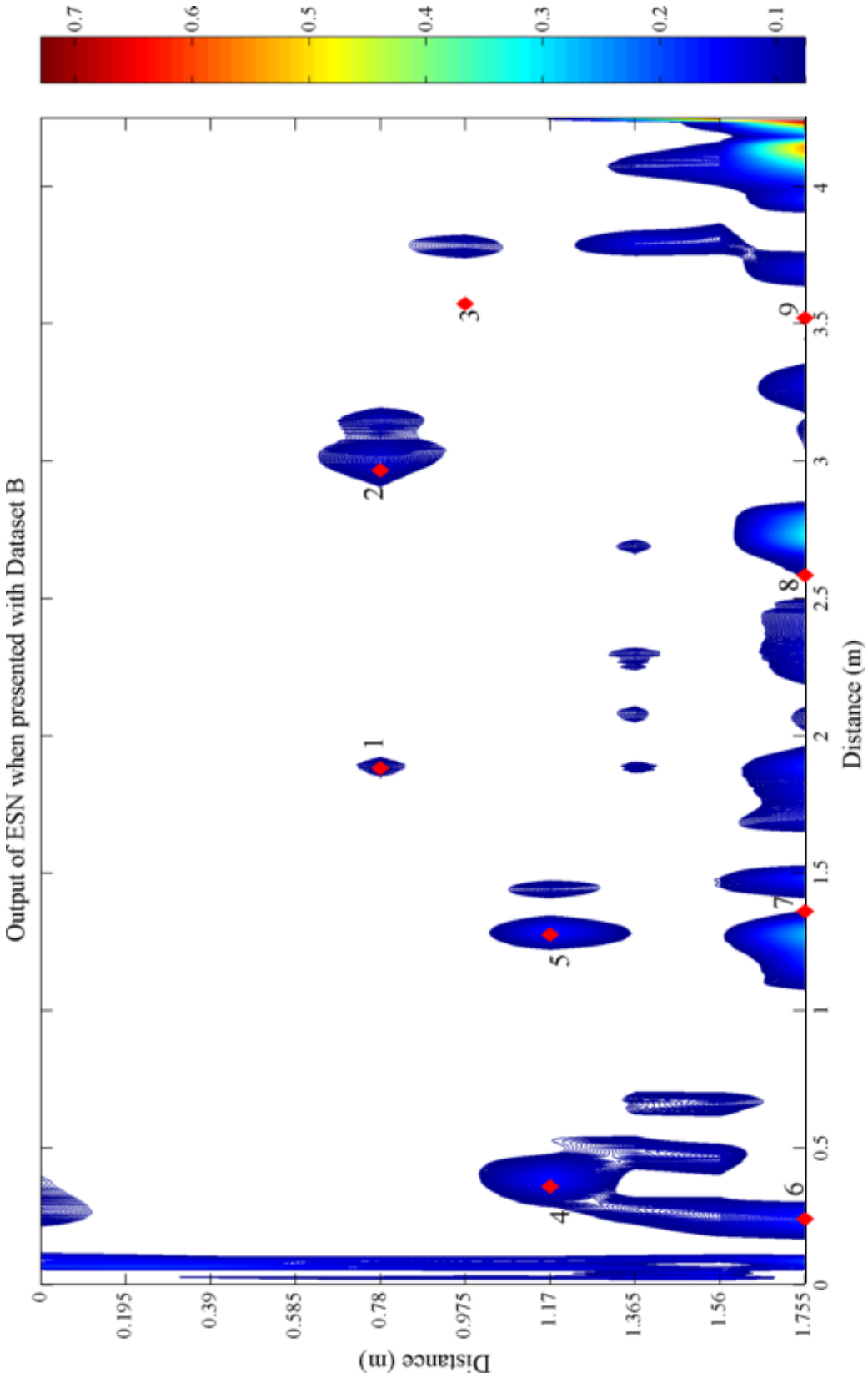


Figure 3.9: A contour plot of the output of the ESN when applied to Dataset B, with the threshold chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations.

3.2.5 Conclusion

When considering the AUC values, the optimal thresholds and the subjective evaluation, two conclusions can be reached. Firstly, ESNs are, indeed, capable of processing EMAD data and can provide competitive performance. Secondly, neither the ESN and AT are wholly able to locate defects perfectly and unambiguously. The work in the remainder of this chapter seeks to address both this problem and the problem of varying cover depths by using the ESN's apparent ability to learn the temporal relationship between the different components of the EMAD data and supplementing it with cover depth data.

3.3 Detecting defects with different cover depths

While both the AT and ESN were able to detect defects well on datasets obtained from the concrete test bed, their ability to detect defects in data obtained at different vertical distances from the rebar was not scrutinised in the previous section. As discussed in the introductory section of this chapter, cover depth is a significant issue, since the rebar in a 'real-world' structure may be encapsulated at varying vertical depths that rarely correspond to the scanning distance of 41.0 mm that was used when scanning the concrete test bed. The following sections show how EMAD data was fused with cover depth data in order to account for this potential variation in depth.

The remainder of this chapter details the work that was done to achieve this.

3.3.1 Data Gathering Methodology

In order to test the ability of ESNs to fuse EMAD and cover depth data, new datasets were obtained, in which systematic changes to the cover were made.

3.3.1.1 Recording concrete cover

Recording the concrete cover presented a challenge, due to the limitations of the commercial covermeters available. For example, the Proceq Profoscope measures cover

to a depth of 185 ± 4 mm for rebar with a diameter greater than 40 mm, and to a depth of 105 mm for a rebar with a diameter of 5 mm, such as the rebar in both the concrete test bed and data fusion mesh. This maximum depth can also be affected by the spacing between the rebar [418]. The Elcometer 331, meanwhile, can detect a 40 mm diameter rebar from a maximum of 200 mm away, and an 8 mm diameter rebar at a distance of 150 mm, with an error of ± 5%. However, there is a requirement that the distance between rebar must be 1.5 × the total cover depth [419]. The Kolectric Research MC8022, meanwhile, can reach 185 mm ± 4 mm for bars of 40 mm diameter and 110 mm ± 2 mm for bars of 6 mm diameter [420]. Furthermore, a covermeter requires careful calibration, especially with regards to bar diameter and range setting[421], and the error in readings can greatly increase depending on the distance between adjacent rebar [422].

While all of this would be less of an issue in the 'real world', where the concrete cover should be well within the range of the covermeter, it meant that commercially available covermeters were not appropriate for the systematic experimental work that was intended here. Although it is unlikely that the concrete cover in the real world would be 280 mm, it is still important for the ESN to be able to model the full range of signals that could be obtained by the EMAD probe. Consequently, the cover depth data was emulated by manually controlling the distance between the rebar and the EMAD probe and then registering these measurements with each datapoint in the EMAD probe datasets.

While the data were not obtained using an actual covermeter, the cover input stream still reflected a real, physical distance between the EMAD and the steel, and so differs from a dataset in which the EMAD signal attenuation due to separation was just simulated. In order to further reflect the margin of error seen in data obtained by a covermeter, Gaussian white noise was added to the normalised data in the testing set, C2. A signal-to-noise ratio of 51 (or 0.5, if added prior to normalisation), equivalent to variation in the range ± 2 mm, was used.

3.3.1.2 Data fusion testing mesh

A second mesh, termed the 'data fusion mesh', was set up for the purpose of obtaining datasets that could be used for data fusion. The layout of this mesh, along with the location of six manually inserted defects, is given in Figure 3.10. As with the mesh embedded in the concrete test bed in Section 3.2.1, the rebars were 5 mm thick and separated by 195 mm. Unlike the first mesh, this second mesh was not encapsulated in concrete, since different depths of concrete cover were to be artificially realised.

The different levels of cover were simulated using CrazyGadget® plastic paving driveway grids, each of which had approximate dimensions of 500 x 500 x 40 mm. These grids could be clipped together horizontally, and nine were sufficient to cover the full length of the mesh. The grids could also be stacked on top of each other for the simulation of different levels of cover, with a stack of seven giving a cover depth of 280 mm, the limit at which the EMAD probe can detect anomalies. Since both plastic and air are non-magnetic, their magnetic permeability is very close to the magnetic permeability of concrete, which is, in turn, very close to the magnetic permeability of free space, $4\pi \times 10^{-7}$ H m^{-1} [48]. This means that recording the magnetic flux through plastic, rather than concrete, did not affect the behaviour of the magnetic flux. Figure 3.11 shows the second testing mesh with a stack of seven plastic grids in position over two of the rebar, simulating a cover depth of 280 mm. The EMAD probe was wheeled over the top of these grids in order to record the data.

Each line on the data fusion mesh was energised in the direction indicated in Figure 3.10. After energisation, each line was then scanned at seven different cover depths: 42.5 ± 0.5 mm, 85.0 ± 0.5 mm, 124 ± 0.5 mm, 165 ± 0.5 mm, 205 ± 0.5 mm, 251 ± 0.5 mm and 289 ± 0.5 mm. Each scan was performed twice - once by each EMAD device - so as to create two separate datasets. As before, all of the data were normalised between +1 and -1 before being presented to the ESN variants considered here. The end result was two EMAD datasets, with 70 scan lines in each. These are henceforth referred to as datasets C1 and C2.

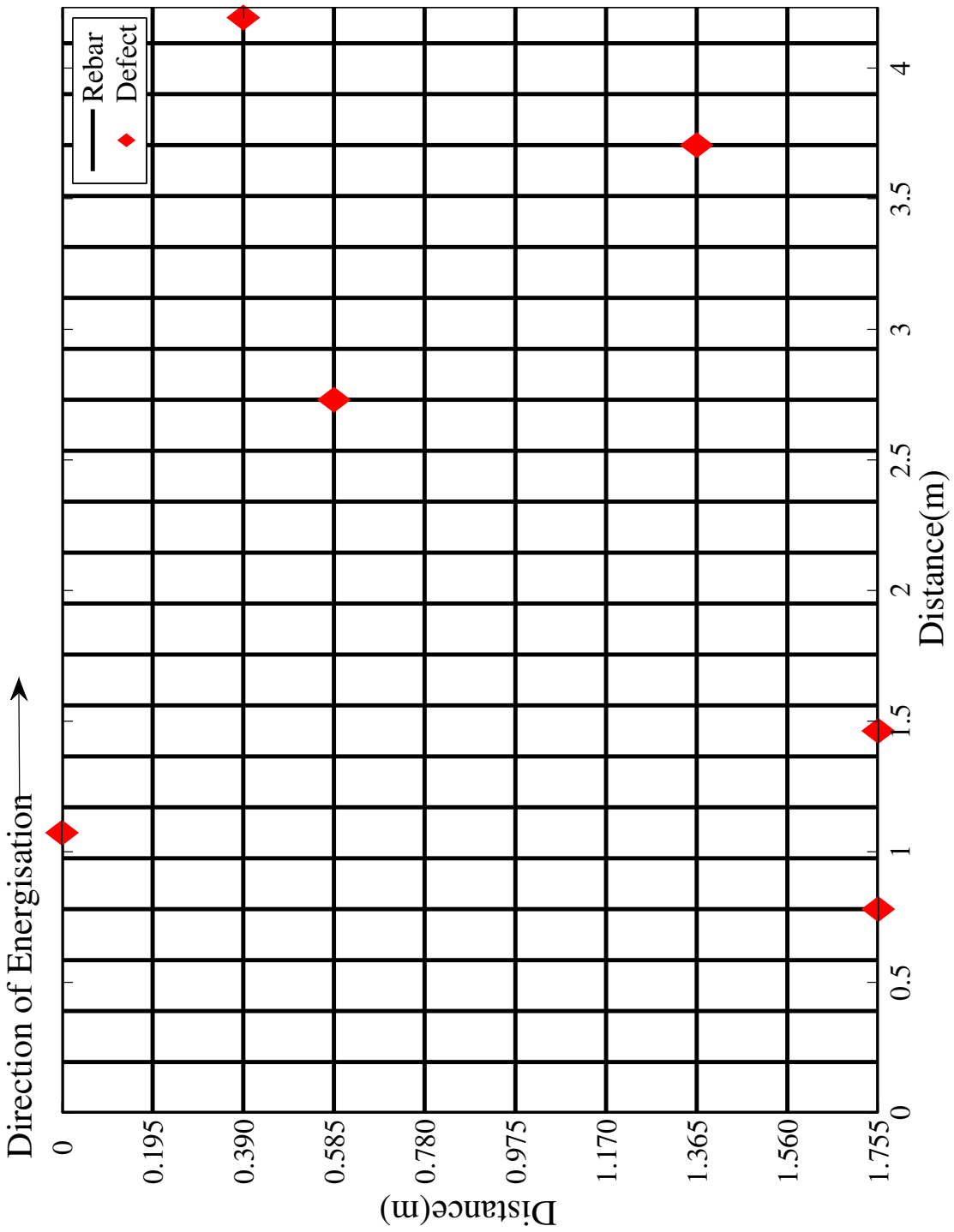


Figure 3.10: The layout of the data fusion testing mesh. Black lines represent rebar, while red diamonds represent defect locations.

Figure 3.11: The data fusion testing mesh, with the plastic grids in position and giving a simulated cover depth of 280 mm.

3.3.2 Application of Computational Intelligence Techniques

A data fusion methodology was formulated and applied to the recorded EMAD and cover depth data. Some alternative ESN approaches were also designed for comparison with this data fusion approach. These were ESN_{GD}, which was trained on data obtained from all cover depths, and ESN_{DS}, a collection of ESNs that were trained to work on data obtained at specific cover depths. The details of these are recorded below.

3.3.2.1 ESN data fusion approach

The data fusion approach that was used here was a relatively simple one. Firstly, each EMAD datum was registered with depth of cover data, such that each datum had a corresponding cover depth value. Then, an ESN was created with three inputs: X axis component of the magnetic flux, Z axis component of the magnetic flux and cover depth. This ESN had one output unit, which was trained to give a value of -1 when no defect was present and +1 when a defect was present. An additional 'bias' input unit was also used, since it was found that repeatedly feeding the network with an input of +1 improved performance. There is some precedent for this in the literature, with

Jaeger suggesting that such 'bias' input units improve training effectiveness when the mean value of the desired output is not zero by increasing the variability of individual neurons' dynamics [297, 303].

Dataset C1 was used for training, which was performed using ridge regression, and then C2 was used as an unseen testing dataset. Although these datasets reflected the same physical reality, they were separately obtained by different EMADs devices and are, hence, different. This meant that the ESN was trained on data from seven different heights, including lines both with and without defects. A further completely unseen dataset obtained from the concrete test bed was also used to evaluate performance.

Before the performance of the data fusion ESN approach could be measured, the best performing ESN configuration had to be found. The parameters that were varied, the range that they were varied over and the optimal values that were found can be seen in Table 3.3.

Parameter	**Range varied over**	**Optimal value**
Spectral Radius	0 - 2	0.1
Input Scaling	0 - 2	1.25
Leak Rate	0 - 1	0.1
Adaptation Epochs	0 - 10	9
Reservoir Size	1 - 500	82
Reservoir Connectivity Factor	0.1 - 1	0.45
Activation Function	Tanh, Lorentzian, Triangular Basis Function, Radial Basis Function, Fermi	Tanh

Table 3.3: The optimal values of the ESN parameters for the data fusion approach, along with the range that these parameters were varied over.

Once the optimal configuration was found, 500 ESNs were trained and their average performance on the testing dataset recorded. Different approaches to analysing the data were then used for comparison with this data fusion technique, and these are detailed in the subsections below.

3.3.2.2 AT and Preliminary ESN

Since it is the usual technique for processing EMAD data, the AT was also applied to C2. If the data fusion approach could not outperform the AT, then it would not have shown any improvement on the existing method. Similarly, the ESNs created in Section 3.2, which had been trained only on data from the concrete test bed, were also tested using C2, and are henceforth referred to as ESN_{Prelim}. This was to investigate whether or not an ESN could detect defects at any cover depth, despite having only seen data obtained with a constant cover depth of 41.0 mm in the preliminary work.

For both the AT and ESN_{Prelim}, the X and Z axis components of the magnetic flux were used as inputs, with a single output indicating whether or not a defect was present.

3.3.2.3 Depth-Specific ESNs (ESN_{DS})

One alternative approach to the problem of different cover is to have a suite of pre-trained ESNs, each of which has been specially trained to process data recorded at a specific cover depth. In the 'real-world', cover measurements could be taken and then the most appropriate ESN selected for the recorded cover depth. A pre-trained ESN is able to deliver an almost instantaneous output when presented with input data, and so it would be possible to conduct an EMAD survey, determine the cover depth and then select the most appropriate ESN from a number of networks that have been trained to work at specific cover depth.

In order to explore this potential scenario, seven separate ESNs were trained, one for each different level of simulated cover. Although there was a relative paucity of training data for these ESNs - one data file from C1 for each of the ten rebar lines scanned at each cover depth - the fact that these ESNs would only ever be exposed to data recorded at one cover depth made up for this, since it would only have to learn how to recognise defect signals, without needing to learn about the variability in amplitude due to cover depth. Each ESN had two input units, corresponding to the X axis and Z axis components of the magnetic flux, and a single output unit, which would give the determined condition of the steel at each point. As with the data fusion ESN, this

Parameter	**ESN**$_{42.5}$	**ESN**$_{85.0}$	**ESN**$_{124}$	**ESN**$_{165}$	**ESN**$_{205}$	**ESN**$_{251}$	**ESN**$_{289}$
Spectral Radius	0.3	0.9	0.8	0.5	0.9	0.9	0.5
Input Scaling	1.9	0.9	0.7	0.7	0.4	0.4	0.1
Leak Rate	0.9	0.6	1.0	0.2	0.2	0.3	0.1
Adaptation Epochs	0	0	5	6	4	10	0
Reservoir Size	300	300	90	120	140	340	20
Reservoir Connectivity Factor	0.2	0.2	0.4	0.6	0.3	0.5	0.99
Activation Function	Radial Basis	Radial Basis	Radial Basis	Radial Basis	Radial Basis	Radial Basis	Lorentz

Table 3.4: The optimal parameters for each different implementation of ESN$_{DS}$. The range varied over for each parameter is the same as in Table 3.3

output unit was trained using ridge regression, and was to give values of -1 when no defect was present, and +1 when a defect was present. Dataset C2 was used as an unseen testing dataset.

The best performing ESN parameters that were determined for each ESN$_{DS}$ are given in Table 3.4. These were determined by performing a grid search, and the range that each parameter was varied over is the same as in Table 3.3.

3.3.2.4 General Depth ESN (ESN$_{GD}$)

A further alternate approach to the problem of different cover depths that does not involve the fusion of data is to train a single ESN on EMAD data from each different cover depth, without providing any information about the cover depth itself. This 'general depth' approach led to ESN$_{GD}$, which was trained on the entirety of one of the two EMAD datasets available. Unlike the ESN used in Section 3.2, ESN$_{GD}$ was exposed

Parameter	Range varied over	Optimal value
Spectral Radius	0 - 2	0.1
Input Scaling	0 - 2	1.3
Leak Rate	0 - 1	0.1
Adaptation Epochs	0 - 10	0
Reservoir Size	1 - 500	70
Reservoir Connectivity Factor	0.1 - 1	0.6
Activation Function	Tanh, Lorentzian, Triangular Basis Function, Radial Basis Function, Fermi	Tanh

Table 3.5: The optimal parameters found for ESN_{GD}

to defect signals recorded at a range of cover depths, and was, therefore, expected to be able to detect the presence of defects at any depth. A potentially confounding factor in this case was that without vital context about the level of cover, it could be unable to discern between low amplitude defects at cover of 280 mm and noise at cover of 42.5 mm. Again, the network was trained on dataset C1 using ridge regression, such that the output unit would give a value of +1 when a defect was present, and -1 when there was no defect. The network had three input units, one each for the recorded X and Z axis components of the magnetic flux, and one which supplied a constant bias value of '+1'. Once the optimal parameters for this had been found, 500 ESNs were trained using this topology and then presented with C2 as an unseen testing dataset. The overall average performance of these 500 ESNs was then used for comparison with the other techniques.

The optimal parameters that were found for ESN_{GD} after a grid search are given in Table 3.5.

3.3.3 Results

The following section details the average AUC and optimal threshold for the AT, ESN_{Prelim}, ESN_{DS}, ESN_{GD} and the fusion ESN at each different cover depth, before finally giving the standard deviation of these thresholds for each technique as a measure of the consistency of the optimal threshold at different cover depths. Only selected

contour plots are referred to in the results for each different cover depth, but the full set of plots is given in Appendix A.

It should be noted that since it would be impractical to plot the output of all 500 repeats of each ESN topology, the contour plots here were plotted by what was determined to be the ESN that gave the performance most typical of that topology. In the case of the fusion ESN, ESN_{Prelim} and ESN_{GD}, this meant calculating the overall average AUC across the entirety of C2 and then selecting the individual ESN whose overall AUC was closest to this average. For ESN_{DS}, this was done for each particular cover depth, rather than for the entirety of C2. The following subsections give the results for each technique at each different cover depth, while a grand results table is given in Table 3.6 for convenience.

3.3.3.1 Performance at cover depth of 42.5 mm

The average AUC and optimal threshold for each technique at the smallest cover depth, 42.5 mm, are given in Table 3.7. Unsurprisingly, given the clarity of defect signals demonstrated in Figure 3.1, performance was very good for each one of the methods. The fusion ESN, ESN_{GD} and $\text{ESN}_{42.5}$ all achieved near-perfect classification accuracy, with ESN_{GD} marginally achieving the best AUC score. It is also worth noting that all three of these were able to outperform the AT.

Figures A.2, A.3 and A.4 all show how well suited the three new ESN approaches were to the dataset. In each case, all of the defects were clearly detected, with only a minimal number of false positives. In contrast, Figure A.1 shows how the AT was able to detect all of the defects, but also introduced several false positives that would make identification of the defects difficult in a 'real-world' survey. Figure A.5 demonstrates the importance of allowing an ESN to be trained on a wide variety of data in order to develop the ability to generalise, as ESN_{Prelim} gave a large number of false positives that seriously obscured the actual defects.

Cover Depth	Classifier	AUC	Optimal Threshold
42.5 mm	AT	0.9882	0.6396
	ESN_{Prelim}	0.9648 (0.0234)	0.709 (0.0086)
	$ESN_{42.5}$	0.9976 (0.0008)	-0.1241 (0.0729)
	ESN_{GD}	0.9984 (0.0002)	-0.4126 (0.0356)
	Fusion ESN	0.9922 (0.0050)	-0.5932 (0.0493)
85.0 mm	AT	0.9742	0.9372
	ESN_{Prelim}	0.8129 (0.0574)	0.0590 (0.0048)
	$ESN_{85.0}$	0.9943 (0.0054)	-0.1203 (0.1547)
	ESN_{GD}	0.9902 (0.0022)	-0.5942 (0.0187)
	Fusion ESN	0.9963 (0.0031)	-0.5466 (0.0339)
124 mm	AT	0.9128	1.4279
	ESN_{Prelim}	0.7769 (0.0660)	0.0567 (0.0040)
	ESN_{124}	0.9726 (0.0156)	-0.2268 (0.0904)
	ESN_{GD}	0.9742 (0.0026)	-0.6766 (0.0126)
	Fusion ESN	0.9857 (0.0023)	-0.5724 (0.0034)
165 mm	AT	0.9034	2.1293
	ESN_{Prelim}	0.7095 (0.0805)	0.0558 (0.0044)
	ESN_{165}	0.9802 (0.0219)	-0.3079 (0.0975)
	ESN_{GD}	0.9749 (0.0019)	-0.6777 (0.0114)
	Fusion ESN	0.9820 (0.0024)	-0.5312 (0.0341)
205 mm	AT	0.8083	0.1240
	ESN_{Prelim}	0.7743 (0.0522)	0.0581 (0.0041)
	ESN_{205}	0.8851 (0.0326)	-0.1832 (0.1002)
	ESN_{GD}	0.9216 (0.0059)	-0.7152 (0.0090)
	Fusion ESN	0.9127 (0.0149)	-0.6233 (0.0479)
251 mm	AT	0.7335	0.0758
	ESN_{Prelim}	0.6945 (0.0628)	0.0568 (0.0043)
	ESN_{251}	0.9296 (0.0340)	-0.2681 (0.0961)
	ESN_{GD}	0.8896 (0.0078)	-0.7347 (0.0090)
	Fusion ESN	0.8596 (0.0147)	-0.5236 (0.0356)
289 mm	AT	0.7504	0.0876
	ESN_{Prelim}	0.6590 (0.0797)	0.0579 (0.0048)
	ESN_{289}	0.8597 (0.0519)	-0.6625 (0.0635)
	ESN_{GD}	0.8531 (0.0126)	-0.7241 (0.0104)
	Fusion ESN	0.6912 (0.0204)	-0.4832 (0.0441)
Test Bed	AT	0.9578	4.3468
	ESN_{Prelim}	0.9369 (0.0187)	0.0683 (0.0076)
	$ESN_{42.5}$	0.8888 (0.0481)	-0.1280 (0.1679)
	ESN_{GD}	0.9441 (0.0069)	-0.3705 (0.0919)
	Fusion ESN	0.9665 (0.0092)	-0.5068 (0.0500)

Table 3.6: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, ESN_{DS}, ESN_{GD} and the fusion ESN across each dataset. The standard deviation for the ESNs is given in brackets.

Classifier	AUC	Optimal Threshold
AT	0.9882	0.6396
ESN_{Prelim}	0.9648 (0.0234)	0.709 (0.0086)
$ESN_{42.5}$	0.9976 (0.0008)	-0.1241 (0.0729)
ESN_{GD}	0.9984 (0.0002)	-0.4126 (0.0356)
Fusion ESN	0.9922 (0.0050)	-0.5932 (0.0493)

Table 3.7: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, $ESN_{42.5}$, ESN_{GD} and the fusion ESN at a cover depth of 42.5 mm. The standard deviations for the ESNs are given in brackets.

3.3.3.2 Performance at cover depth of 85.0 mm

Table 3.8 gives the average AUC and optimal threshold for each method at the second cover depth, 85.0 mm. Most of the techniques maintained a good level of performance, although ESN_{Prelim} faltered when presented with data obtained at a cover depth twice as great as the cover depth used to train it. As before, the fusion ESN, ESN_{GD} and $ESN_{42.5}$ all achieved near-perfect classification accuracy, with the fusion ESN just outperforming the other two. Once again, the AT failed to classify defects as well as the three new ESN architectures.

Classifier	AUC	Optimal Threshold
AT	0.9742	0.9372
ESN_{Prelim}	0.8129 (0.0574)	0.0590 (0.0048)
$ESN_{85.0}$	0.9943 (0.0054)	-0.1203 (0.1547)
ESN_{GD}	0.9902 (0.0022)	-0.5942 (0.0187)
Fusion ESN	0.9963 (0.0031)	-0.5466 (0.0339)

Table 3.8: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, $ESN_{85.0}$, ESN_{GD} and the fusion ESN at a cover depth of 85.0 mm. The standard deviations for the ESNs are given in brackets.

Figure A.7 shows how the fusion ESN was able to distinguish between defects and clean rebar almost unambiguously, introducing only one false positive at (4, 1.755). While ESN_{GD} and ESN_{DS}, shown in Figures A.8 and A.9, were still able to detect all of the defects, they also introduced more false positives than the fusion ESN.

3.3.3.3 Performance at cover depth of 124 mm

The results for the cover depth of 124 mm are given in Table 3.9. In terms of AUC, it can be seen that the fusion ESN again produced the best performance. All three of the proposed ESN topologies continued to impress, and, while the AT still did a reasonable job of detecting defects, there was a clear gulf in performance when compared to the superior CI techniques.

Classifier	**AUC**	**Optimal Threshold**
AT	0.9128	1.4279
ESN_{Prelim}	0.7769 (0.0660)	0.0567 (0.0040)
ESN_{124}	0.9726 (0.0156)	-0.2268 (0.0904)
ESN_{GD}	0.9742 (0.0026)	-0.6766 (0.0126)
Fusion ESN	0.9857 (0.0023)	-0.5724 (0.0034)

Table 3.9: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, ESN_{124}, ESN_{GD} and the fusion ESN at a cover depth of 124 mm. The standard deviations for the ESNs are given in brackets.

Examination of Figures A.12, A.13 and A.14 shows that while the AUC values for the three new ESN architectures seem quite similar, there is a significant difference in terms of the clarity of the resultant contour plots. All of the defects were, once again, successfully detected with good accuracy, but both ESN_{GD} and ESN_{124} added potentially misleading outputs. The false positives seen at (0.8, 0.39), (1.5, 0) and (2.5, 1.755) in the output of ESN_{GD} could all plausibly be mistaken for a real defect signal, while the defect located at (3.75, 1.365) is obscured. While the high leak rate of ESN_{124} (see Table 3.4) gave the best performance, it also led to very fast reservoir dynamics (see the discussion on ESN leak rate in Chapter 2), which resulted in an overall output that appears to be 'rippled'. While most of the defects can be seen clearly, the defects at (3.75, 1.365) and (4.2, 0.39) are somewhat lost amongst adjacent false positives. All three of these methods compare favourably to the AT, shown in Figure A.11, which failed to detect one defect and produced several false positives. Although not unexpected, it should also be noted that by this stage, the output of ESN_{Prelim} seen in Figure A.15 was wholly inadequate for the detection of defects.

3.3.3.4 Performance at cover depth of 165 mm

Despite the EMAD being four times further away from the mesh than the initial 42.5 mm survey, the proposed ESN topologies were able to maintain consistently good performance at a cover depth of 165 mm, as shown in Table 3.10. In fact, ESN_{GD} and ESN_{165} showed marginal improvement compared to the results for cover of 124 mm. However, the fusion ESN continued to outdo all of the other techniques.

Classifier	**AUC**	**Optimal Threshold**
AT	0.9034	2.1293
ESN_{Prelim}	0.7095 (0.0805)	0.0558 (0.0044)
ESN_{165}	0.9802 (0.0219)	-0.3079 (0.0975)
ESN_{GD}	0.9749 (0.0019)	-0.6777 (0.0114)
Fusion ESN	0.9820 (0.0024)	-0.5312 (0.0341)

Table 3.10: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, ESN_{165}, ESN_{GD} and the fusion ESN at a cover depth of 165 mm. The standard deviations for the ESNs are given in brackets.

This is also borne out in the contour plots for the fusion ESN, ESN_{GD} and ESN_{165}, seen in Figures A.17, A.18 and A.19, respectively. All three techniques were still able to detect all of the defects, in spite of the fact that the magnitude of the EMAD signals was greatly reduced. Once again, the fusion ESN achieved this while also producing the fewest full positives and the clearest overall output.

3.3.3.5 Performance at cover depth of 205 mm

Unsurprisingly, increasing the cover depth to almost five times that of the initial 42.5 mm cover depth resulted in worse performance. The biggest difference between the results at a depth of 165 mm and the results for 205 mm is that the AT became unable to clearly distinguish between defect signals and clean rebar. This is demonstrated by the sudden drop in AUC from 0.9034 to 0.8083 seen in Tables 3.10 and 3.11, and the contour plot seen in Figure A.21. Indeed, without prior knowledge of the defect locations, it would be very difficult to accurately determine the location of the defects using Figure A.21. If the new ESN techniques were still able to give a coherent and accurate output at this cover depth, it would show that they have an important advantage over the

typically used AT.

Classifier	AUC	Optimal Threshold
AT	0.8083	0.1240
ESN_{Prelim}	0.7743 (0.0522)	0.0581 (0.0041)
ESN_{205}	0.8851 (0.0326)	-0.1832 (0.1002)
ESN_{GD}	0.9216 (0.0059)	-0.7152 (0.0090)
Fusion ESN	0.9127 (0.0149)	-0.6233 (0.0479)

Table 3.11: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, ESN_{205}, ESN_{GD} and the fusion ESN at a cover depth of 205 mm. The standard deviations for the ESNs are given in brackets.

The AUC values in Table 3.11 show that while the depth specific ESN, ESN_{205}, faltered, the two ESNs trained on data from a variety of different cover depths were able to maintain a strong level of performance, with ESN_{GD} giving the best AUC. Figures A.22 and A.23 indicate that practically speaking, the difference in defect detection ability between the fusion ESN and ESN_{GD} was minute. Both architectures were able to detect all of the defects (although the fusion ESN struggled with the defect at (4.2, 0.39)), and both architectures incorrectly identified the same false positives, such as those at (0.75, 0.39), (1.2, 0.78) and (2.5, 1.755). While the overall picture is more obscure than it was for previous cover depths, both methods allow for better identification of faults than the AT.

3.3.3.6 Performance at cover depth of 251 mm

Table 3.12 shows how, at a cover depth of 251 mm, all of the defect detection approaches began to struggle to detect all defects. For the first time, the AUC for the fusion ESN and ESN_{GD} dropped below 0.9, although there was some improvement in the performance of the depth specific approach, which gave an AUC of 0.9296. Notably, all three of these ESN architectures were again able to outdo the AT, which gave an AUC of 0.7335. In Figure A.26, it can be seen that while most of the defects are detected by the AT, this is mostly because they happen to be in larger regions of false positives, and not because they, specifically, are positively identified as defects.

While ESN_{251} would appear to have been the best approach for a cover depth of 251

Classifier	AUC	Optimal Threshold
AT	0.7335	0.0758
ESN_{Prelim}	0.6945 (0.0628)	0.0568 (0.0043)
ESN_{251}	0.9296 (0.0340)	-0.2681 (0.0961)
ESN_{GD}	0.8896 (0.0078)	-0.7347 (0.0090)
Fusion ESN	0.8596 (0.0147)	-0.5236 (0.0356)

Table 3.12: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, ESN_{251}, ESN_{GD} and the fusion ESN at a cover depth of 251 mm. The standard deviations for the ESNs are given in brackets.

mm based solely on the AUC, Figures A.27 (for the fusion ESN), A.28 (for ESN_{GD}) and A.29 (for ESN_{251}) suggest that this is far from clear cut. The performance of the fusion ESN suffers from the fact that it was unable to identify the defect at (0.9, 0), but the other defects are relatively clearly identified. In contrast, ESN_{GD} successfully detected all of the defects, but almost all of these are hidden amongst larger false positives, making it very difficult to determine the actual defect location. To a lesser extent, the same is true of ESN_{251}. In 'real-world' scenarios, the fusion ESN would, therefore, provide more reliable results.

3.3.3.7 Performance at cover depth of 289 mm

The final cover depth, 289 mm, was particularly challenging, as shown by the results seen in Table 3.13. Even the ESN specifically trained to solely recognise defects at a depth of 289 mm was only able to give an AUC of 0.8597. These results initially suggest that the performance of the fusion ESN was comfortably the worst of all of the techniques, but Figures A.31, A.32, A.33, A.34 and A.35 show that this was not the case.

Classifier	AUC	Optimal Threshold
AT	0.7504	0.0876
ESN_{Prelim}	0.6590 (0.0797)	0.0579 (0.0048)
ESN_{289}	0.8597 (0.0519)	-0.6625 (0.0635)
ESN_{GD}	0.8531 (0.0126)	-0.7241 (0.0104)
Fusion ESN	0.6912 (0.0204)	-0.4832 (0.0441)

Table 3.13: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, ESN_{289}, ESN_{GD} and the fusion ESN at a cover depth of 289 mm. The standard deviations for the ESNs are given in brackets.

For example, Figure A.35 shows how ESN_{Prelim} was unable to distinguish between clean and damaged rebar, and consequently failed to usefully identify any true positives. The signals seen in the contour plot are actually just the baseline of ESN_{Prelim}'s output. Similarly, the detection of defects by the AT in A.31 is almost coincidental, as large swathes of the plot are wrongly identified as positives. This is not the case with the fusion ESN, which failed to detect two of the defects, but did a better job than all of the other techniques in clearly and unambiguously spatially locating those true positives that were successfully identified. For example, while both the fusion ESN and ESN_{289} were able to produce a positive for the defect at (2.6, 0.585), it would be easier to locate this defect using the fusion ESN. In this case, the AUC results alone do not fully reflect the abilities of the different techniques at a cover depth that is at the limit of the EMAD's detection capabilities. All of the techniques struggled to an extent, and only the three new ESN techniques made informed decisions on defect locations.

3.3.3.8 Performance on concrete test bed dataset

Finally, Table 3.14 gives the results for when the different approaches were applied to the concrete test bed dataset, which was gathered at a cover depth of 41.0 mm. If the depth-specific ESN approach was to be used in the 'real-world', it is probable that $ESN_{42.5}$ would be used in this case, since it specialises in data obtained at a similar cover depth. For this reason, $ESN_{42.5}$ was used to represent the ESN_{DS} methodology.

In terms of the AUC, the only one of the new ESN architectures that actually improved on the performance of the AT was the fusion ESN. This is no small feat, and emphasises the value of a data fusion approach with the ESN technique, as the AT was originally developed using a different, older set of data obtained from the concrete test bed. As expected, all three of the new ESN architectures were able to beat ESN_{Prelim}, showing the of superiority the new, more thoroughly trained ESNs over the initial preliminary work.

Figures A.36, A.37, A.38 and A.39 all show that while the fusion ESN and ESN_{GD} were capable of detecting all defects accurately, the longitudinal rebar seen at (0 - 4,

Classifier	AUC	Optimal Threshold
AT	0.9578	4.3468
ESN_{Prelim}	0.9369 (0.0187)	0.0683 (0.0076)
$ESN_{42.5}$	0.8888 (0.0481)	-0.1280 (0.1679)
ESN_{GD}	0.9441 (0.0069)	-0.3705 (0.0919)
Fusion ESN	0.9665 (0.0092)	-0.5068 (0.0500)

Table 3.14: The average AUC and average optimal threshold for the AT, ESN_{Prelim}, $ESN_{42.5}$, ESN_{GD} and the fusion ESN on the concrete test bed dataset. The standard deviation for the ESNs is given in brackets.

1.755) presented a greater challenge. While ESN_{GD} was able to detect all four defects, these were totally obscured amongst a series of large false positives. The fusion ESN was better at unambiguously spatially locating these defects, but still produced some false positives. Only the AT was able to locate these four defects clearly, although its performance on the remainder of the dataset was not as good. $ESN_{42.5}$ failed to detect one defect, and struggled to unambiguously identify others. This highlights the biggest drawback of the ESN_{DS} approach. Having only seen data obtained at a cover depth of 42.5 mm, the results suggest that $ESN_{42.5}$ was sensitive to even a small cover change of 1.5 mm. This suggests that the ESN_{DS} approach might be difficult to use in practice, since the most appropriate ESN available for a given cover depth could be confounded by data obtained by a cover depth that is only slightly different to the one that it saw in training. The ESN approaches that had been trained on data from different cover depths, the fusion ESN and ESN_{GD}, were better equipped to deal with small changes in cover, and subsequently outperformed $ESN_{42.5}$ Overall, the best performance in terms of both AUC and clarity of contour plot was given by the fusion ESN.

3.3.3.9 Calculated AUC threshold consistency

The final performance measure was to calculate the standard deviation of the average optimal thresholds reported for each technique in Tables 3.7 - 3.14. This was to give an indication of how consistent the optimal thresholds were for each technique at different cover depths. A large standard deviation would indicate that the optimal threshold varied quite widely, while a smaller standard deviation would indicate a relatively

consistent optimal threshold level. The standard deviation for the average optimal thresholds at each height is given in Table 3.15.

Classifier	Optimal Threshold Standard Deviation
AT	1.4568
ESN_{Prelim}	0.2287
ESN_{DS}	0.1941
ESN_{GD}	0.1426
Fusion ESN	0.0464

Table 3.15: The standard deviation of the average optimal thresholds for AT, ESN_{Prelim}, ESN_{DS}, ESN_{GD} and the fusion ESN at each different cover depth. A smaller standard deviation indicates a more consistent threshold as cover depth changed.

The standard deviation of the average optimal thresholds for the fusion ESN was over three times smaller than the next smallest standard deviation (given by ESN_{GD}), indicating that the fusion ESN offers the most consistent threshold across different cover depths.

3.4 Discussion

3.4.1 The advantages of data fusion in this case study

The results show that the fusion ESN approach was consistently able to outperform the other techniques in terms of AUC, consistency of the calculated threshold and a subjective evaluation of the clarity of resultant contour plots. Indeed, the fusion ESN had the best AUC for four of the eight different cover depths investigated, the smallest standard deviation in the overall average optimal threshold and was able to detect defects at the very limit of the EMAD's detection capabilities. Furthermore, it was able to not only outperform three comparative ESN architectures, including one that used a bespoke ESN for each different cover depth, but also the AT that has previously been used.

Chapter 2 introduced Bellot's four possible benefits of data fusion [267]. The success of data fusion in this context can be assessed according to each of these four areas.

1. **Representation** Bellot suggested that the final output of the fusion system

should have a greater granularity and greater level of abstraction than that provided by the sensors prior to data fusion. Taken in isolation, the covermeter data cannot provide any detail on the condition of the rebar, while the EMAD data can identify defects after a level of data processing. The output of the data fusion system, in contrast, is a single stream of data that abstracts out both sets of knowledge into an informed judgement on the condition of the rebar. In that sense, the output of the fusion system represents an increase in the granularity and abstraction of the data. Furthermore, the fusion process provides a greater increase in granularity and abstraction than the AT and the non-fusion ESN approaches, which produced a single data stream informed only by the EMAD data, divorced from the context of the cover depth.

2. **Certainty** The results obtained in this chapter show that using fusion would certainly increase the probability of the results being correct, as the fusion ESN was consistently able to outperform the AT, even at the smaller cover depths. Since it was also found that the performance of the fusion ESN was always better or at least comparable to the performance of the other ESN architectures used, it can be said that the fusion ESN also gave greater certainty than the non-fusion based CI techniques. Furthermore, the consistency in the optimal threshold at each different cover depth adds a greater level of certainty, since it means that an expert level of judgement is not required for the accurate application of the threshold.

3. **Accuracy** As with the certainty, the results show that the fusion ESN offered improvements over both the AT and the other ESN architectures in terms of accuracy. This is particularly apparent in the graphs given in Appendix A, where positively identified defects were generally highlighted with greater clarity and without ambiguity by the fusion ESN when compared to other approaches.

4. **Completeness** Ultimately, using only the EMAD for the assessment of reinforced concrete would give a picture of the state of the rebar, but would miss valuable

contextual information on cover depth. By fusing the data, the user gets a picture based on a relationship not only between the different components of the MFL data, but also on the relationship between the magnitude of these signals and the cover depth. This makes it easier to distinguish between low amplitude defect signals at a high cover depth, and noise at a low cover depth, and gives a fuller overall picture of the rebar.

3.4.2 Limitations of the data fusion approach

While the data fusion approach performed well, it should be noted that it only provided the best performance in terms of AUC for cover depths between 85.0 mm and 205 mm. At 205 mm and beyond, the AUC degraded to the levels of the EMAD-only techniques. This is not to say that the fusion ESN's performance was not competitive outside of the range $42.5\,mm < depth < 205\,mm$; instead, it was only as good as the alternative approaches. This can be explained by considering the relationship between signal amplitude and cover depth that was discussed in Section 3.1. At a cover depth of 42.5 mm, the amplitude of the defect signals in the EMAD data is very large, making the defects very easy for all of the techniques to distinguish from clean rebar. For this reason, all of the different approaches were able to produce high AUC values, and the fusion ESN did not confer much of an advantage. At cover depths of 205 mm and greater, the amplitude of the defect signals became increasingly small and, hence, much more difficult to detect. The obscurity of these signals meant that the added context of cover depth did not noticeably offer improved detection in comparison with other methods. Instead, the fusion ESN was most useful for the cover depths at which the defect signals in the EMAD data were still quite clear, but where changes in the amplitude arising from the cover depth made it more challenging for EMAD-only approaches to continue to distinguish between defects and noise.

An exception to this rule can be seen in the results for the concrete test bed dataset, where the cover depth was 41.0 mm. Interestingly, this dataset was much noisier than the others. This is because the data fusion testing mesh consisted of clean rebar with

manually inserted defects, whereas the concrete test bed mesh had been corroding gradually for a number of years in aggressive conditions. Furthermore, the concrete test bed mesh had undergone several transverse energization cycles over its lifetime prior to this case study, which may have interfered with this longitudinal survey. Even when the depth of cover lies outside the range where the fusion ESN offers best performance, benefits from the fusion approach still accrue when the EMAD signals themselves are inherently noisy, such as those captured from the concrete test bed.

3.4.3 Applicability of the data fusion ESN to 'real-world' scenarios

The improved performance offered by the fusion ESN would be worth very little if the technique proved to be unsuitable for use in 'real-world' settings. Both the data fusion testing mesh and the concrete test bed were good approximations of 'real-world' scenarios, since they were real, physical meshes, subject to real environmental conditions and genuine defects. Furthermore, the data were gathered using the full EMAD procedure, rather than simply simulating ideal MFL data. In terms of actual ability to detect defects, all of the new ESN architectures introduced in this chapter were suitable for application to 'real-world' data. However, the ability to detect defects is only one aspect of 'real-world' applicability. With respect to ease of use and the computational resources resources required in real world settings, the pre-trained fusion ESN could be set up on a laptop and provide near instant results for data that is presented in real time.

The on-site ease of use is perhaps best indicated by the consistency of the optimal threshold seen in Table 3.15. One issue with using the AT is that the scale is arbitrary, and the optimal threshold could take almost any positive value. If the ground truth is not known, this makes it difficult to say with complete certainty where the threshold should be for finding defects, especially if the cover depth is also unknown. In direct contrast, the output of the fusion ESN usually fell between -1 and +1, with the standard deviation of the optimal threshold being only 2% of the possible range of output values.

In a real-world survey, an engineer would be able to retain confidence in the accuracy of a contour plot created using a pre-set threshold value, rather than having to try to determine the best threshold value themselves. This does not just mean that the fusion ESN can be applied in the 'real-world', but that it holds a major advantage over the AT.

The fusion ESN is also a viable option in terms of computational requirements. Processing the data gathered at 42.5 mm in dataset C2 with the 'most average' fusion ESN took 0.276 seconds. Although this is slow compared to the 0.0261 seconds required for the AT to process the same data, it is not unacceptably so, especially when the extra fraction of a second was used to produce superior results. It could potentially be argued that costly, licensed software like MATLAB, which was used for all of the data processing in this thesis, is impractical for use on site, but there are open source alternatives for ESN production, such as Oger for Python [423].

All in all, the fusion ESN would be well suited to processing EMAD and covermeter data in the 'real-world'.

3.4.4 Reasons for the effectiveness of the data fusion ESN

3.4.4.1 Cover depth as a scaling factor

The advantage that data fusion conferred upon the fusion ESN can perhaps best be demonstrated by modifying the covermeter data that was used as an input to the network. Figure 3.12 shows how the output of the fusion ESN changed when presented with the same EMAD data, but different covermeter data. The data in question was taken from the scan of one rebar from the data fusion mesh, with a cover depth of 42.5 mm. In the case of the data plotted in blue, the fusion ESN was told that the cover depth was 42.5 mm, whereas the data plotted in red shows how this output changed when the same ESN was told that the cover depth was 289 mm. Similarly, Figure 3.13 shows exactly the same scenario, using data from the same rebar, but gathered at a cover depth of 289 mm. The data plotted in blue again shows the output of the fusion ESN when told that the cover depth was 42.5 mm, while the data in red shows the

output when told that the cover depth was 289 mm.

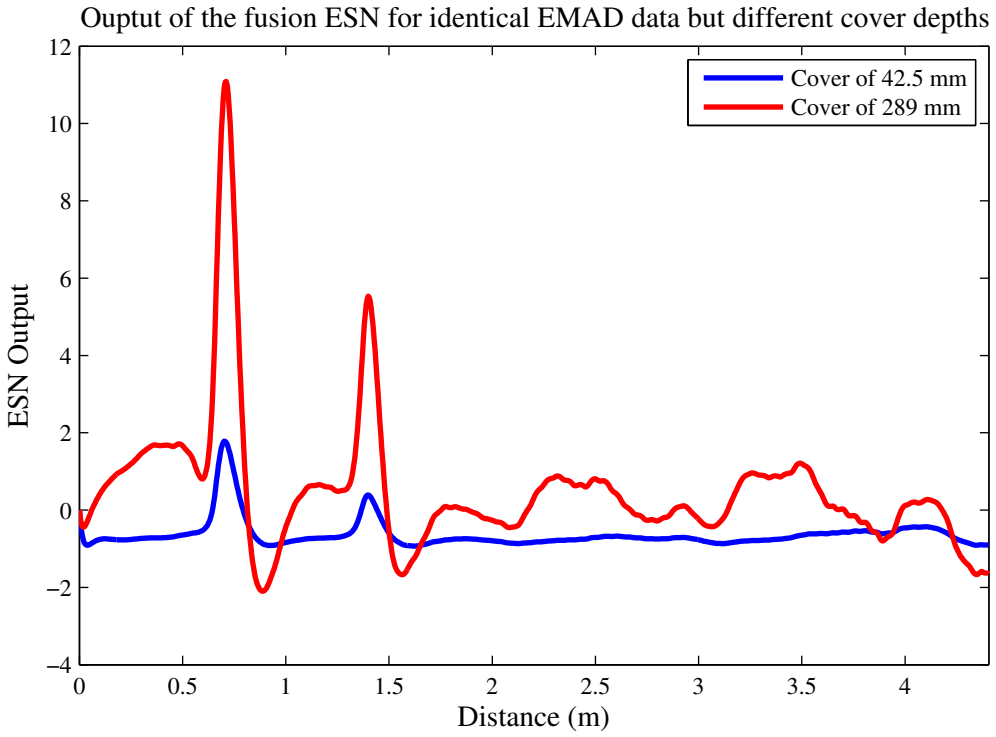


Figure 3.12: The response of the fusion ESN to a set of EMAD data gathered at a cover depth of 42.5 mm when told that the cover depth was 42.5 mm (blue line) and 289 mm (red line). The two large spikes related to actual defects.

In Figure 3.12, it can be seen that telling the ESN that the rebar was 289 mm away instead of 42.5 mm had a significant effect on the output signal. There are two peaks in the output for when the correct cover depth was presented, and these correspond to two real defects. This was to be expected, as the small cover depth resulted in a clean defect signal that had the characteristic peak in the Z axis data and negative gradient in the X axis data. The ESN was trained to recognise that as cover depth increases, these defect signals become weaker, and so when the network saw such a clear defect signal at an apparent cover depth of 289 mm, its output reached a much greater peak value of approximately 11 for the exact same EMAD data. The effect of increasing the cover depth that was input to the network was to massively amplify the overall output.

In Figure 3.13, the opposite effect is achieved. When the ESN was given the correct cover depth of 289 mm, a number of different peaks can be observed in the output data. The first two of these again correspond to real defects, but the later peaks are false positives, caused by the real defects becoming less distinguishable from noise in the

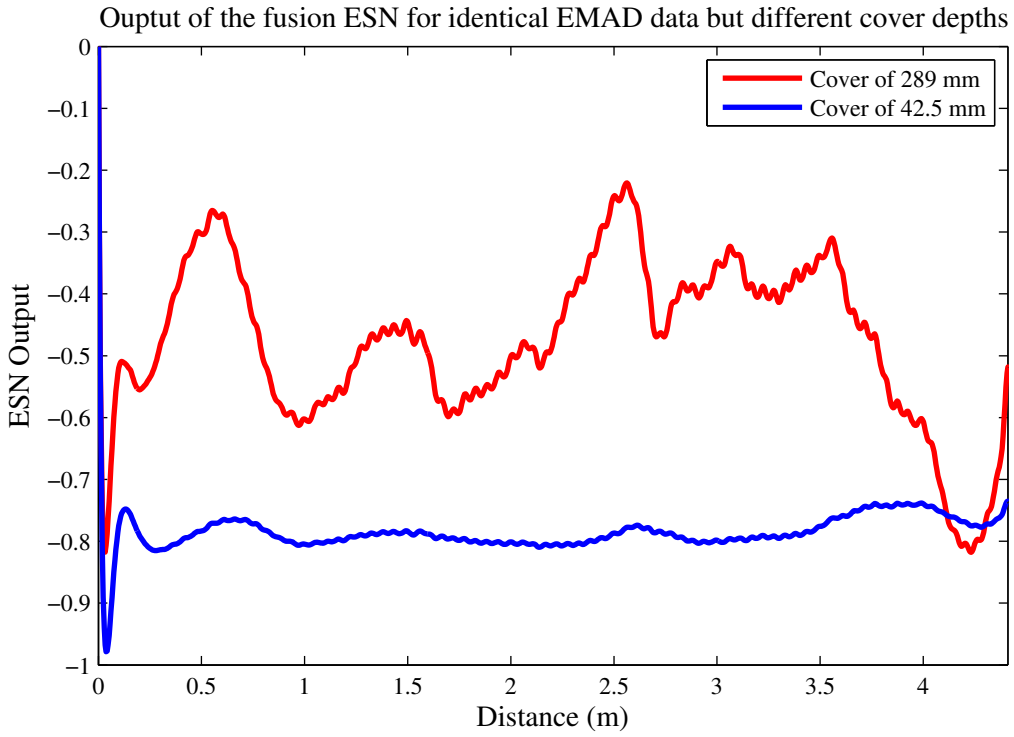


Figure 3.13: The response of the fusion ESN to a set of EMAD data gathered at a cover depth of 289 mm when told that the cover depth was 42.5 mm (blue line) and 289 mm (red line). There are defects present at approximately 0.6 m and 1.5 m.

EMAD data gathered from a greater cover depth. In contrast, there is barely a single peak in the output when the ESN was told that the cover depth was 42.5 mm. The output has a baseline of approximately 0.8, and seldom moves away from that. The ESN had learned that at a greater cover depth, the small, less distinct defect signals could potentially be significant. However, as the cover depth reduced, the likelihood of these smaller signals being indicative of defects also reduced, and they were treated as random noise.

Viewed together, Figures 3.12 and 3.13 show that the ESN learned to use the covermeter data to scale the overall output. This scaling allowed weak defect signals at a greater cover depth to be ascribed the same significance as a stronger defect signal at a small cover depth. It should be noted, however, that the cover depth did not actually affect the ability of the ESN to determine the presence of a defect. Instead, it is this scaling capability that meant that the fusion ESN had a much more consistent average optimal threshold than all of the other techniques evaluated above.

The advantage that this scaling effect conferred upon the fusion ESN compared to

ESN_{GD} is shown in Figure 3.14, which shows how the output of the fusion ESN and ESN_{GD} changed for data gathered from the same rebar at different cover depths. The figure focuses specifically on the region around two defects. It can be seen that at 42.5 mm, both architectures gave very similar outputs, clearly identifying the two defects. When the cover depth increased, the magnitude of the output signals decreased for both architectures. However, this is less pronounced with the fusion ESN, which still produced two clear peaks at the point of the two defects. In comparison, there are still two recognisable peaks in the output of ESN_{GD}, but the magnitude of these peaks is much smaller than those of the fusion ESN. This ability to continue to produce very distinct peaks at defects even when the cover depth increased is what allowed the fusion ESN to outperform ESN_{GD} when the cover depth was 124 mm, and is a direct result of the data fusion process.

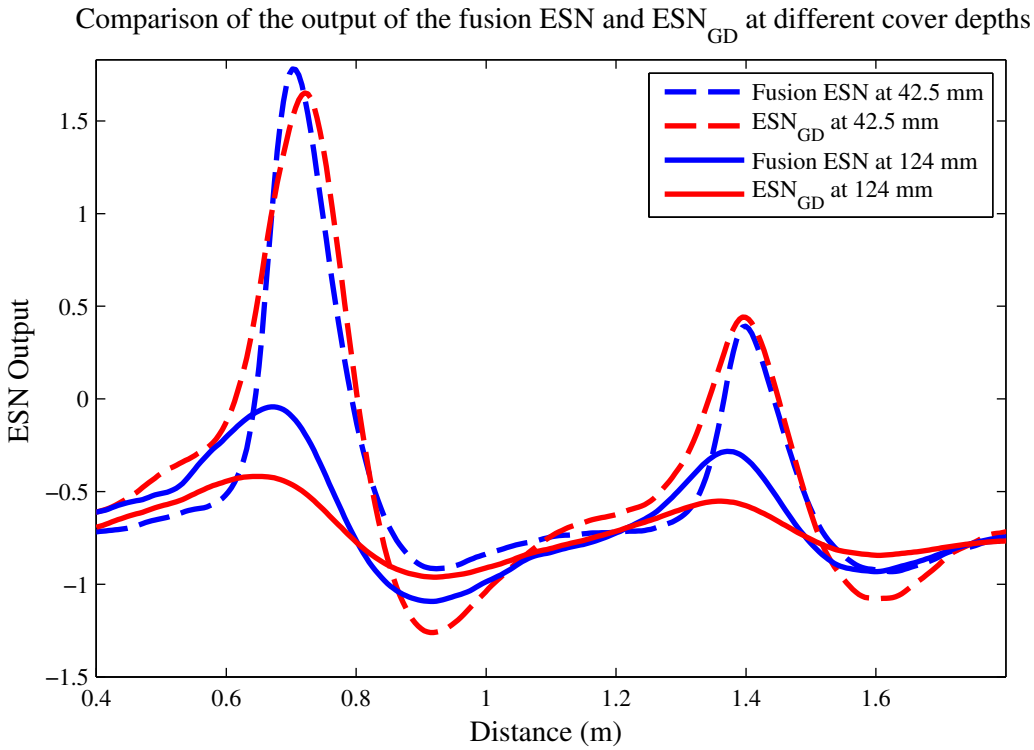


Figure 3.14: A comparison of the output of the fusion ESN and ESN_{GD} at cover depths of 42.5 mm and 124 mm.

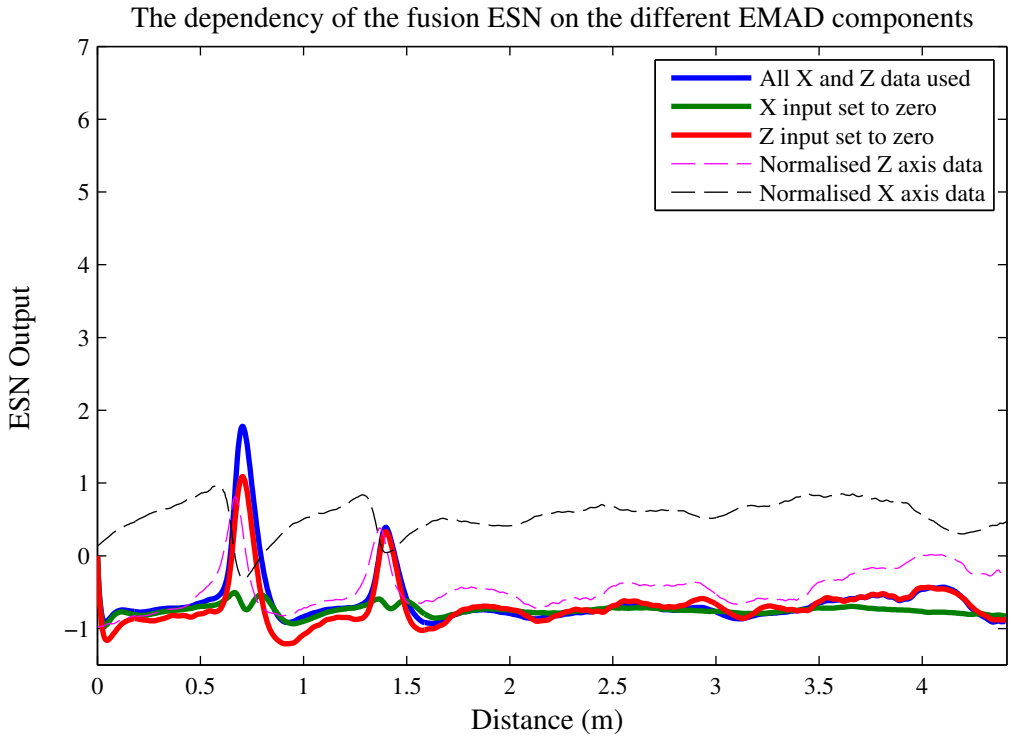


Figure 3.15: The response of the fusion ESN to a set of EMAD data gathered at a cover depth of 42.5 mm when all of the X and Z axis data was used (blue line), when the X axis input was set to zero (green line) and when the Z axis input was set to zero (red line). The original EMAD data are also given.

3.4.4.2 Relationship between ESN output and EMAD data

There is also a more subtle relationship between cover depth, the two components of the EMAD data and the output of the fusion ESN, and this is shown in Figures 3.15 and 3.16. These show the response of the fusion ESN to the data used in Figure 3.12 in six different scenarios. Figure 3.15 shows how the output of the ESN changed when it was presented with all of the data, including the correct cover depth, before then being given the same data but with the X axis component set to a constant value of zero, and then with the Z axis component set to zero. Setting any input to a constant value of zero effectively removes it from the network. The original normalised X and Z axis data are also presented. Figure 3.16 presents the same three scenarios, but with the network told that the cover depth was 289 mm, rather than 42.5 mm, in the cases where the X and Z axis data were removed.

Looking firstly at 3.15, the output of the ESN appears to be largely dependent on the gradient of the X axis data. The output of the ESN when the Z axis was set to

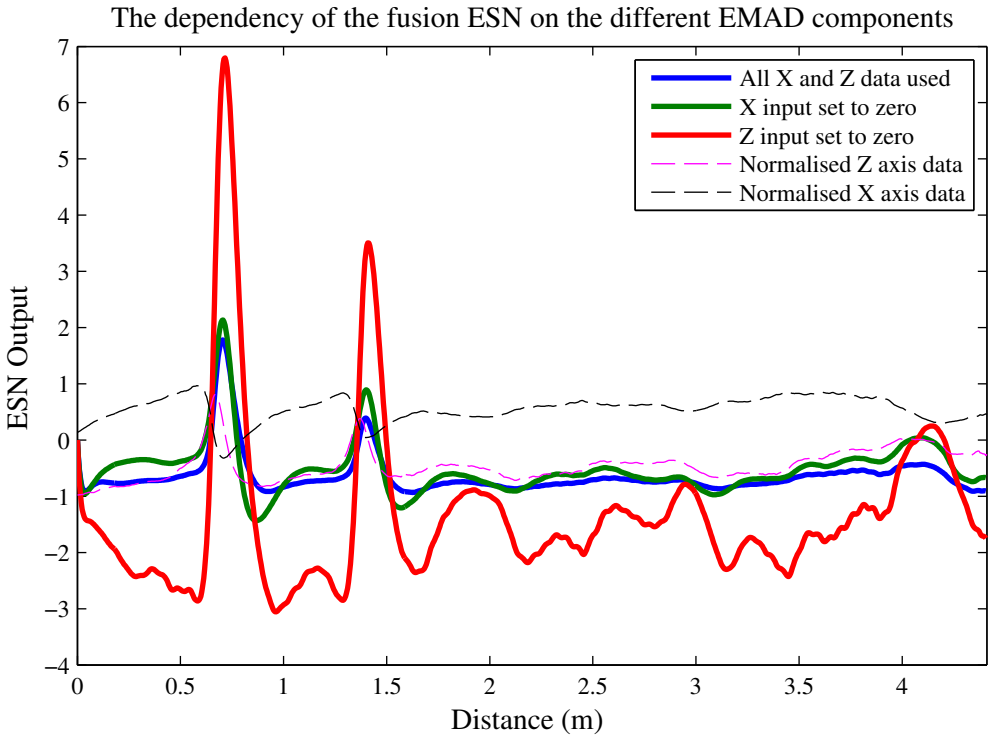


Figure 3.16: The response of the fusion ESN to a set of EMAD data gathered at a cover depth of 42.5 mm when all of the X and Z axis data was used (blue line), when the X axis input was set to zero and cover depth to 289 mm (green line) and when the Z axis input was set to zero and cover depth to 289 mm (red line). The original EMAD data are also given.

zero resembles the output for the full dataset much more closely than when the X axis was set to zero. Furthermore, both the output when all of the data was used and the output for when the Z axis was set to zero give a strong peak at the centre of the slope in the X axis data. In contrast, when the X axis data was removed, there was no peak in the output data. All of this suggests a strong, inverse relationship between the output of the fusion ESN and the gradient of the input X axis data, with the Z axis data having a minimal effect. However, comparing Figure 3.15 with Figure 3.16 reveals a more nuanced relationship that also has some dependency on cover depth.

Perhaps the biggest difference between Figures 3.15 and 3.16 is that in Figure 3.16, both defects were successfully detected when the X axis was set to zero. While the peaks in the output data were still determined solely by the data provided by the EMAD, this result reveals another of the ways in which the data fusion approach was able to benefit the overall network performance. The strong, inverse relationship between the X axis

input and the overall output did not change when the cover depth input changed, as demonstrated by the large peaks in the output when the Z axis was set to zero. When the cover depth was correctly given as 42.5 mm, this relationship meant that defects could not be detected when the X axis was set to zero, since the gradient in the X axis was zero at all points. However, the ESN also learned that as the cover depth increases, the amplitude of the EMAD signals decreases, and so a small gradient in the X axis does not necessarily indicate that no defect is present if there is a sufficiently large corresponding peak in the Z axis. In this case, it did not matter that the gradient of the X axis data was zero, because there was a clear, large peak in the Z axis data that implied the presence of a defect. Hence, the covermeter data is also used by the ESN to moderate what combination of EMAD signals could be indicative of a defect.

Figures 3.15 and 3.16 also give an insight into how the ESN uses the EMAD data to determine its final output. While the biggest factor is the gradient of the X axis data, Figure 3.16 shows how, without any Z axis data, the output becomes very sensitive to small changes in the X axis gradient. Although the two defects are still detected, low-amplitude noise in the data - possibly related to the transverse rebar in the mesh - produced a number of other peaks with a large amplitude relative to the output when all of the correct data was used. The magnitude of the Z axis signal plays a part in whether or not the ESN responds to each change in the overall gradient. For cases where the gradient of the X axis is sufficiently large, the lack of a peak in the Z axis would not prevent the output from suggesting the detection of a defect, but would enhance the overall output signal.

This is all in stark contrast with the impact that the X and Z axis data had on ESN_{GD}. Figure 3.17 shows how the response ESN_{GD} to data gathered at a cover depth of 42.5 mm changed when the X axis data were set to zero and the Z axis data were set to zero. Figure 3.18 shows the same scenario, but when the data were instead gathered at a cover depth of 85.0 mm. In each case, it can be seen that when there was no X axis data, the output simply mimicked the input Z axis data, with a small lag. Similarly, when the Z axis input was set to zero, the output just inverted the X axis data. Unlike

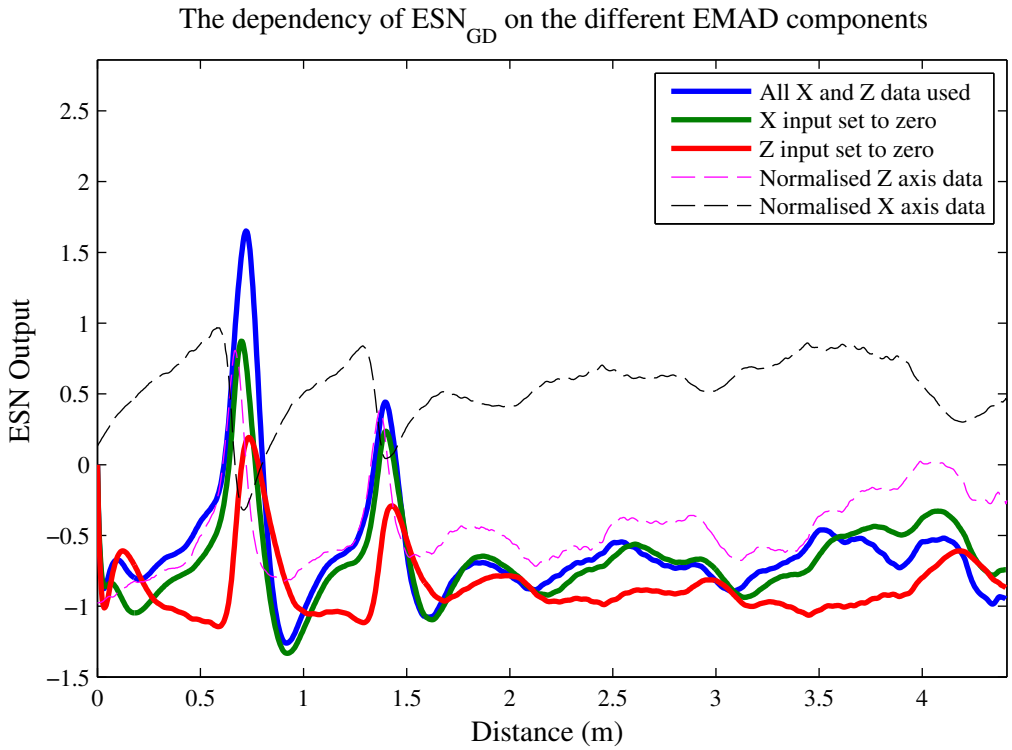


Figure 3.17: The response of ESN$_{GD}$ to a set of EMAD data gathered at a cover depth of 42.5 mm when all of the X and Z axis data was used (blue line), when the X axis input was set to zero (green line) and when the Z axis input was set to zero (red line). The original EMAD data are also given.

the fusion ESN, both figures suggest that ESN$_{GD}$ had a greater reliance on the Z axis data than on the X axis data, since the output when both EMAD inputs were presented closely resembled the output when the X axis input was set to zero. Whereas ESN$_{GD}$ was only able to determine a very simple relationship between the EMAD data and the desired output, the influence of the coveremeter data allowed the fusion ESN to access more subtle relationships, which, in turn, allowed it to outperform ESN$_{GD}$.

3.5 Conclusion

In this chapter, the applicability of data fusion to the NDT of reinforced concrete was demonstrated through the first case study. Initially, it was shown that ESNs are well suited to processing EMAD data, and could provide performance comparable to that of the technique that has previously been used for 'real-world' applications, the AT, when processing data obtained from a real reinforcing mesh that was subject to

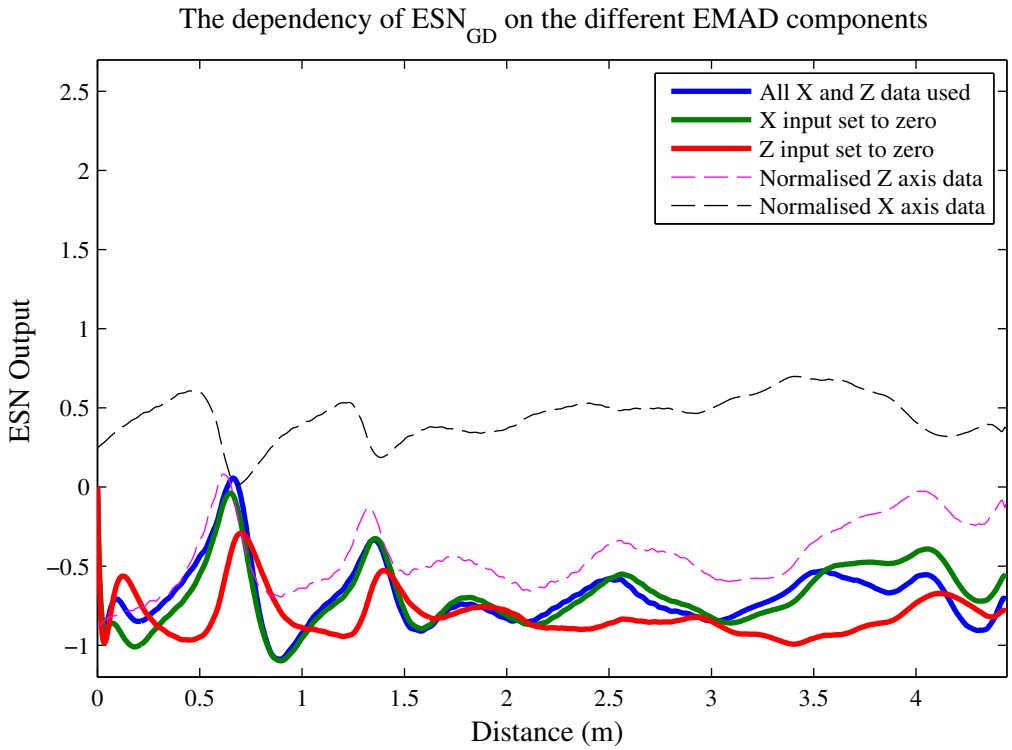


Figure 3.18: The response of ESN_{GD} to a set of EMAD data gathered at a cover depth of 85.0 mm when all of the X and Z axis data was used (blue line), when the X axis input was set to zero (green line) and when the Z axis input was set to zero (red line). The original EMAD data are also given.

corrosive influences. It was then shown that much improved results could be achieved by fusing EMAD data with cover depth data of the sort that might be obtained via coveremeter than by using EMAD data alone. It was found that the fusion ESN outperformed four different EMAD-only approaches, including the AT, in the range $42.5\,mm < depth < 205\,mm$, while providing competitive performance outside of that range. The fusion ESN also proved to be more accurate when processing the noisier data obtained from the Keele University test bed, while providing the most consistent optimal threshold level across all of the data used.

It was found that the data fusion regime used here fulfilled all four of Bellot's four benefits of data fusion (as introduced in Chapter 2), improving data representation, certainty, accuracy and completeness. This completeness was shown with an examination of the response of the fusion ESN to modified inputs. While the location of peaks representing defects in the output data was largely determined from the gradient of the X axis of the EMAD data, the covermeter data served as a scaling factor, damping

the output for small cover depths and amplifying the output for large cover depths. The Z axis of the EMAD data, meanwhile, was used to as a means of determining the response to a large gradient in the X axis. The cover depth was also able to improve the detection capabilities of the fusion ESN in a more nuanced way. For example, it was shown that a large peak in the Z axis data that did not have a corresponding gradient in the X axis data would not give a positive output when the cover depth was small, but would give a positive output when the cover depth was large. In this way, the data fusion approach gave the ESN the ability to modify its response to EMAD signals depending on the cover depth, something that none of the other techniques were capable of. This provided a positive answer to Research Question 1, since the interpretation of the EMAD data was clearly improved by heterogeneous data fusion.

The fusion ESN also demonstrated very good potential 'real-world' applicability. As well as providing the best performance in the range $42.5\,mm < depth < 205\,mm$, the fusion ESN gave the best AUC when applied to the much noisier concrete test bed dataset. This showed how the fusion ESN was adept not just at dealing with relatively clean rebar across several different cover depths, but also at dealing with rebar that has been kept in aggressively corrosive conditions for a number of years. While the presence of defects and the cover depth were both controlled, the data used here were nevertheless obtained using real EMAD devices to scan real, physical rebar, rather than simulated MFL data for an ideal scenario. This all gives good reason to be confident that the fusion ESN would transfer well to 'real-world' applications, where the ground truth may not necessarily be known. The success in the scenarios devised here, along with this potential 'real-world' applicability, shows that, in answer to Research Question 2, heterogeneous data fusion can be applied to quasi-real-world scenarios in NDT.

The consistent optimal threshold obtained using the fusion ESN demonstrated that, in answer to Research Question 3, the data fusion technique could be deployed systematically to reduce the amount of variability that would require expert interpretation. One problem with the AT is that it can be very difficult to determine an optimal threshold for the detection of defects, as shown in Section 3.3.3.9 by the fact that the

AT provided the least consistent optimal threshold values. This means that using the AT in a 'real-world' setting where the ground truth was not known would likely require access to an MFL expert, who would need to use the raw data signals to determine the best threshold to use. This could potentially introduce both lengthy delays and greater cost to the data analysis stage. In contrast, use of the fusion ESN would allow for a threshold to be automatically applied and for real-time access to reliable results. This is an important advantage over the other techniques, and reduces the amount of variability in data interpretation.

In summary, this chapter presented a data fusion approach to the NDT of reinforced concrete that significantly improved upon the AT. In addition to providing a good level of performance, the architecture would be a plausible option for systematic application in 'real-world' surveys. Having worked so well on a real, physical reinforcing mesh, the next stage in this work would be to confirm all of this by performing a survey of a real road or bridge where the ground truth is not known, and then breaking out the concrete to verify the accuracy of the results.

CHAPTER 4

Damage Detection, Localisation, Classification and Assessment in Online Structural Health Monitoring

4.1 Introduction

In online Structural Health Monitoring (SHM) applications, the primary challenge is damage identification in real-time [424]. Any method that would allow a structural engineer to observe and identify damage trends in data would be of great value to industry. This chapter presents the second case study performed as part of this research, which concerned the fusion of different modalities of data extracted from sensors that were embedded in a footbridge by the National Physical Laboratory (NPL) in order to detect damage over the course of a three year monitoring project. A suite of separate ESN approaches, henceforth referred to as ESN_a, ESN_b and ESN_c, was initially applied to data recorded by tilt and temperature sensors, before then also being applied to data recorded by tilt sensors, temperature sensors and strain gauges. The suite consisted of:

- $\textbf{ESN}_{tt}$: a preliminary architecture that simply used temperature sensor data to predict the resultant tilt sensor data.
- $\textbf{ESN}_a$: a data fusion approach that used temperature sensor and strain gauge data as inputs for the purpose of modelling the tilt sensor data. This included $\textbf{ESN}_{asg}$, an architecture that used temperature sensor data to model strain.

- **ESN**$_b$: an ESN that detected specific types of damaging event by finding characteristic temporal signatures in the tilt sensor data.

- **ESN**$_c$: an ESN that used the output of ESN$_a$ to provide a single metric that would describe the condition of the bridge at any given point in the monitoring period.

Worden and Dulieu-Barton [425] gave five different hierarchical levels of damage detection in SHM, building on the four levels previously defined by Rytter [426], with each subsequent level requiring the preceding levels of damage detection:

1. **Detection** a qualitative indication of structural damage.

2. **Localisation** the identification of the probable location of the damage.

3. **Classification** the type of damage is determined.

4. **Assessment** an estimate of the extent of the damage.

5. **Prediction** information on the safety or remaining life of the structure.

The aim of this case study was to use the available sensor data to reach the fourth of these levels of damage detection. When used alongside each other, the data fusion approach used here allowed for damage to be detected and monitored for long term impact, while also allowing for fault locations to be determined. Combining this with two non-fusion based ESN approaches also allowed for the cause of the damage to be classified and an overall measure of the level of damage that the bridge had sustained at any given time to be provided. This synthesis of methods led to the proposal of an overall SHM scheme that has resulted in an outline for a potential SHM toolbox that exploits data fusion and CI to improve sensor interpretation for a quasi-real-world scenario. Figure 4.1 shows which of the different ESN approaches used in this chapter met each of the levels of damage detection.

At this point, it is important to make the distinction between a centralised and decentralised fusion system. The proposed approach that is provided by the end of the

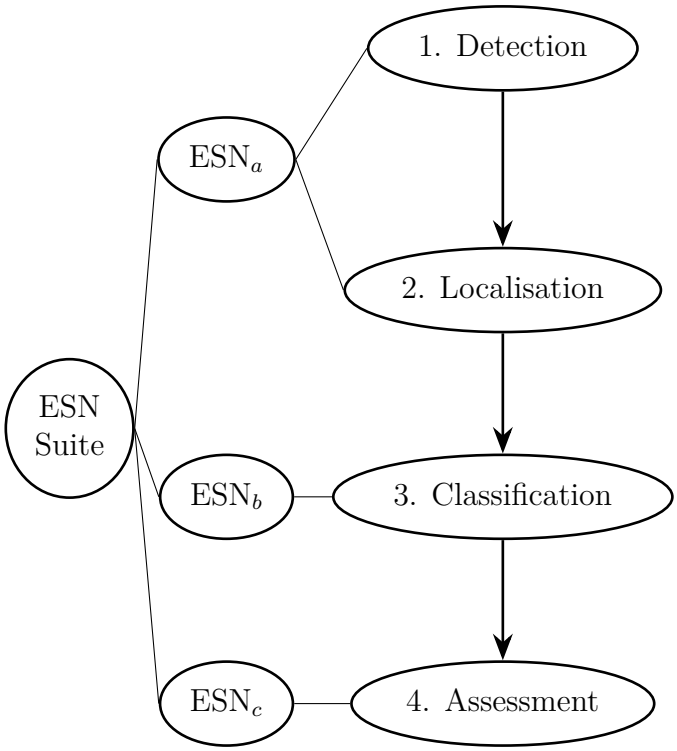


Figure 4.1: A diagram reflecting where each of the ESN approaches fits with respect to the first four levels of the hierarchy of damage detection in SHM.

chapter makes use of a centralised fusion system, where the sensor data is sent to a central processor, which then makes decisions. In a decentralised fusion system, sensors communicate with each other on a local level, making autonomous decisions about the condition of the bridge in their locality and then communicating this to a fusion centre [178]. Since the ESN fusion methodology presented here makes use of all of the sensors on the bridge collectively, it is more appropriate to view it as a centralised system.

The remainder of this chapter is structured slightly differently to the other chapters in this thesis. After the NPL footbridge project is introduced in section 4.2, some preliminary work on modelling the thermal response of tilt sensors is presented. The main focus of this chapter, the fusion of three sensor modalities, is given in section 4.4. Section 4.5 presents two supplementary ESN approaches to assist in the interpretation of the fusion approach, while Section 4.6 then details the proposed overall online SHM system. A discussion is provided in section 4.7 and the chapter is then concluded in section 4.8.

The case study work described in this chapter has also been reported in two original

publications [46, 427], and has extended earlier related work carried out at Keele University [428].

4.2 The NPL Footbridge

4.2.1 Motivation behind the project

The NPL footbridge project was set up as a UK-wide means of developing new sensor technologies and methods for the processing of large time-series datasets from BSNs. The footbridge itself was built in the 1960s, before being taken out of use in 2008 and fitted with a number of sensors. At the time of the project, it was 5 metres high, approximately 20 metres long and weighed 15 tonnes [429]. The datasets that were obtained from the footbridge project have also been made widely available to other research groups, as will be discussed later in this section. At its inception, the three main aims of the project were:

1. To provide test facilities for long-term sensor performance monitoring;

2. To demonstrate the use of different monitoring technologies;

3. To contribute to understanding how to assess the state of infrastructure [430].

Further details on the background and motivation behind the NPL footbridge project have been published elsewhere, and are not repeated here [431, 432].

4.2.2 Sensor modalities

Although a wide variety of sensor modalities, such as crack sensors, accelerometers and RFID sensors were used, the work in this chapter is concerned with the data produced by ten temperature sensors, and eight electrolevel tilt sensors, which consisted of 365 376 data points collected between January 2009 and May 2012, along with the data produced by eight electrical resistance strain gauges collected between February 2009 and February 2011. All of the sensors were provided by ITMSOIL [434]. The

Figure 4.2: The NPL footbridge. The image on the left shows the bridge itself, while the image on the right shows two of the tilt sensors, which are circled in white (reproduced from Kromanis and Kripakaran [433]).

temperature sensors were chosen on the basis that temperature has been found to be the principle cause of bridge deformations [435]. Indeed, it was found that in the case of the NPL footbridge, the vertical displacement at the end of the cantilever due to the average daily temperature variation was 2.7 mm, approximately equivalent to an applied vertical load of 4.5 kN [429]. Strain gauges, meanwhile, are amongst the most commonly used type of sensors used in BSNs [436], and were found to be closely related to temperature, but with a slight lag [429]. Many of the studies discussed in Section 1.4 used accelerometer data to investigate changing natural frequencies in a bridge, but Brownjohn et al. identified that, due to the dependence of the bridge's behaviour on temperature, sensors that focus on movements in the vertical plane are required [435]. For this reason, it was decided that tilt sensors would be better suited to the damage detection approach considered in this chapter. However, it should be noted that past work by Gastineau suggested that multiple tilt sensors are required to show trends across a structure, and that other sensor modalities are required in order to obtain useful information from tilt sensors [437].

Figure 4.3 shows the spatial arrangement of the temperature and tilt sensors on the bridge. Note that tilt sensors 7 and 8 are attached to the two piers of the bridge and that it is a standalone structure, allowing weights to be suspended from the cantilever where sensor 1 is located.

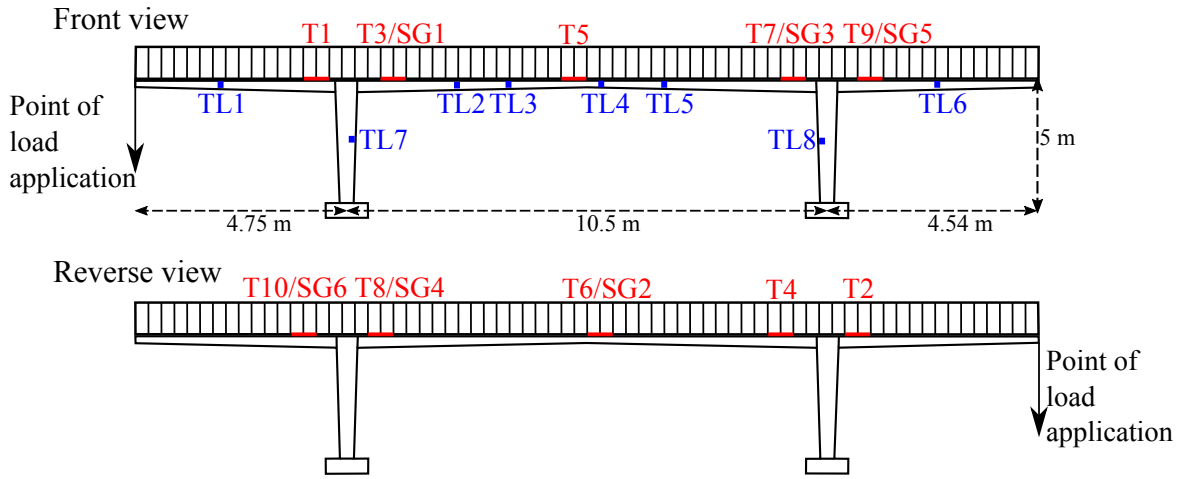


Figure 4.3: The layout of sensors on the NPL footbridge. Tilt (TL) sensor locations are marked in blue, while temperature (T) sensors are marked in red. Some temperature sensors also had embedded strain gauges (SG).

4.2.3 Deliberately damaging interventions

Over the course of the project, several interventions were made by NPL scientists with the intention of simulating damage and recording the response of the bridge to such damage. The most significant of these interventions came in two forms: static tests and fatigue tests. In the case of the former, heavy barrels of water were suspended from the cantilever at the point indicated in Figure 4.3. The fatigue tests, meanwhile, were performed as a way of both simulating damage and assessing the breaking point of the bridge, using an adapted hydraulic system that performed half a million load controlled, 10 kN/s cycles. A creep test was also performed between the 11$^{\text{th}}$ of October 2011 and the 28$^{\text{th}}$ of October 2011. Here, the bridge was kept in a stressed state for a 17 day period. Table 4.1 lists all of the significant interventions, while the full list of damage and repair cycles is given by Livina et al. [438].

4.2.4 Past research

There have been other studies which have analysed the data provided by the NPL footbridge project. Barton and Esward [439] investigated errors in the sensor data, but did not assess long-term damage. Kromanis and Kripakaran [433] used a support vector regression technique to model tilt sensor behaviour with some accuracy, but this

Date	Intervention Type
24th March 2009	Static test
29-30th June 2009	Static test
3rd August 2009	Static test
30th June - 2nd July 2010	Static test
8th October 2010	Static test
18th October 2010	Static test
26th April 2011	Static test
18th - 19th May 2011	Static test
24th June 2011	Static test
27th June 2011	Static test
6th - 7th July 2011	Static test
28th July 2011	Static test
21st August 2011	Static test
24th August 2011	Static test
9th September 2011	Static test
28th September 2011	Fatigue test
29th September 2011	Fatigue test
3rd October 2011	Fatigue test
4th October 2011	Fatigue test
6th October 2011	Fatigue test
7th October 2011	Fatigue test
10th October 2011	Fatigue test
11th October 2011	Fatigue test
11th October 2011	Static test
11th - 28th October 2011	Creep test
28th October 2011	Static test
8th November 2011	Static test

Table 4.1: Significant interventions performed as part of the NPL footbridge project, along with the date when these were performed.

required significant pre-processing and did not give any indication about the condition of the footbridge; the ESN methods used here sought not just to model the sensor behaviour, but to also show long-term damage trends and localise damage. In a later work, Kromanis applied support vector regression and principal component analysis as part of the regression-based thermal response prediction methodology to allow for anomaly detection, and reached the conclusion that short reference periods for data prior to any events cannot lead to high prediction accuracies [440]. As a result of this, the data had to be extrapolated in order to create a sufficiently large training dataset. After going through a process of deseasonalising and detrending the temperature data, Livina et al. managed to detect 'early warning indicators' when significant interventions occurred, but were not able to quantify the damage caused by these events [438]. Worden et al. [441] used cointegration on the tilt sensor data so as to purge environmental effects and detect damage. However, the cointegration procedure required a full yearly cycle of data from January 2009 to February 2010, meaning that the model was based on how the bridge behaved after it had already undergone some damaging events. Furthermore, once the condition of the bridge changed, the data were no longer purged of environmental data and retraining would be required. In the most recent work, it was shown that it was possible to detect the impact of a damaging event, but the location of the damage was not determined [442].

4.3 Preliminary Work: Modelling the Thermal Response of the Tilt Sensors

The ability of neural networks to perform non-linear model estimation without requiring first-principle models or a priori knowledge about the structure has meant that they have long been used for SHM applications [246, 443]. This suggests that ESNs might well be suited to processing the data gathered in this case study, especially in light of the ability of ESNs to exploit potential temporal patterns in datasets.

The initial work simply used a preliminary ESN model, ESN_{tt}, to model the output

of each tilt sensor based on the data recorded by each of the temperature sensors. This temporal relationship is demonstrated in Figure 4.4, which shows the data produced by tilt sensor 1 and temperature sensor 1 (the closest of the ten temperature sensors to tilt sensor 1) on 10$^{\text{th}}$ March 2009, before the first significant intervention occurred. It can be seen that under normal circumstances, the bridge underwent a daily cycle due to change in temperature. Each day, the temperature increased and peaked at around 2 pm, before then gradually reducing, producing a characteristic daily spike. The tilt sensor data followed this, since the tilt of the bridge at that point increased as the temperature increased and peaked as the temperature peaked. As the bridge cooled, the tilt sensor reading gradually reduced again. There is a clear, observable temporal relationship between the two sets of data.

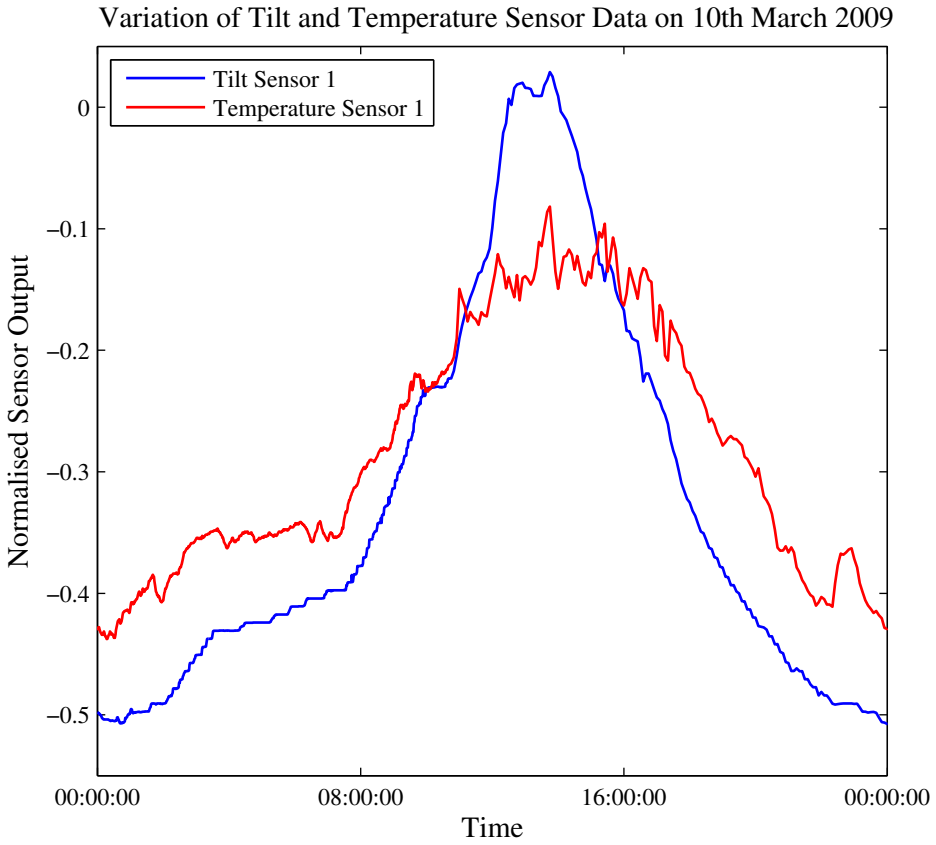


Figure 4.4: The output of tilt sensor 1 and temperature sensor 1 over the course of 10$^{\text{th}}$ March 2009. As the temperature changes, so to does the tilt of the bridge.

It was on this basis that ESN$_{tt}$ was presented with the data from the ten temperature sensors as inputs and was tasked with predicting the output of the eight tilt sensors. The aim of this was to use the temporal relationship between tilt and temperature

sensor readings to model the typical, normal behaviour of the bridge. In doing this, an 'analytical redundancy' [444] approach could be employed in order to generate residuals and, hence, observe deviations between the real and predicted tilt sensor data. It was postulated that if the ESN was able to accurately model the tilt sensor data during the training phase, then a subsequent significant deviation between the output produced by ESN_{tt} and the recorded tilt sensor data would be indicative of anomalous behaviour and possibly damage to the bridge.

4.3.1 Methodology

The approach employed here was to use only the data prior to the first significant intervention on the 29$^{\text{th}}$ June 2009 for training, and to then apply the trained ESNs to the remainder of the data. This approach means that ESN_{tt} was trained on the behaviour of the tilt sensors only under normal conditions. As the portion of the data available for training was limited to the first 12.79% of the data, no sampling was performed and the full dataset up to 29$^{\text{th}}$ June 2009 was instead used. The data were normalised between -1 and +1, with no other pre-processing performed. Training was performed using ridge regression and 75-fold cross validation, with the data presented to ESN_{tt} in batches of one day of data (i.e. all of the data gathered between 00:00 and 23:55 for each full day in the training period). While this data may have been discontinuous, it was still in the time domain. This meant that during training, the ESN would be sequentially presented with the data for one of the full days in the training set, during which time its output weights would be trained, and would then be returned to its initial state (with the newly updated weights) so that the process could be repeated with another randomly selected day. The rationale for doing this, rather than continuously presenting the data from the entire training period in sequence, was that it would allow the ESNs to learn the diurnal cycles of the tilt sensors in a range of different scenarios. This could alleviate the need for a deseasonalisation process prior to data presentation, and would test whether ESNs are sophisticated enough to recognise the diurnal variations as part of a regular pattern of activity, without any seasonal

context.

There is a possibility that the number of training batches available could result in an overtrained network. However, the objective in this preliminary work was to see if it was possible for the ESN to retain the features of the tilt sensors and potentially discriminate subtle damage indicators in the presence of diurnal variability when applied to the rest of the dataset. For this reason, the network was shown as much of this variation as possible from the period prior to the first significant intervention. It is also worth noting that each day in the training dataset is unique, and that the range of temperatures typically seen at the start of the period are not the same as the range of temperatures at the end of the training period. In the first 30 days of the training data, the minimum recorded temperature was -14.42 °C, the maximum recorded temperature was 19.37 °C and the mean temperature was 3.48 °C. In contrast, the final 30 days of the training dataset had a minimum recorded temperature of -1.83 °C, a maximum recorded temperature of 40.15 °C and a mean recorded temperature of 17.74 °C. If the ESNs were to successfully model the tilt sensor behaviour, they would need to know how they behaved in response to a wide range of different temperatures.

A grid search was then performed in order to find the best ESN topology, the results of which are given in Table 4.2. Once this had been found, 100 ESNs with the optimal topology were trained. After training, the 100 ESNs were then presented with the data from the training period in a continuous stream, rather than in batches. The average Pearson Correlation Coefficient (PCC) during the training period for each tilt sensor was recorded, and the output of ESN_{tt} that had the best PCC across all eight tilt sensors was then compared to the real tilt sensor outputs. This was an interesting test, because the ESNs had seen all of the data from this period in batches, but had not previously seen it presented in sequence.

If good correlation between the ESN output and the original tilt sensor data was achieved, the ESNs would then be presented with the temperature sensor data for the entire monitoring period and would be tasked with using these sensors to predict the corresponding output from each tilt sensor. The ESNs would produce an 'ideal' tilt

Parameter	Range varied over	Optimal value
Spectral Radius	0 - 2	0.95
Input Scaling	0 - 2	1.9
Leak Rate	0 - 1	0.2
Adaptation Epochs	0 - 10	0
Reservoir Size	1 - 500	500
Reservoir Connectivity Factor	0.1 - 1	0.5
Activation Function	Tanh, Lorentzian, Triangular Basis Function, Radial Basis Function, Fermi	Tanh

Table 4.2: The optimal parameters for ESN_{tt}.

sensor response to the changes in temperature, and this could then be compared to how the bridge actually behaved. In order to do this, the residual between the output predicted by this ESN and the actual tilt sensor data at each point was calculated and then a moving average of the residual was taken according to Equation 4.1.

$$x = |\frac{\sum_{n-5000}^{n+4999} P_i}{10000} - \frac{\sum_{n-5000}^{n+4999} A_i}{10000}| \tag{4.1}$$

In Equation 4.1, x is the final moving average value, n the number of any given data point, A the actual data point from a tilt sensor and P the data point that was predicted by ESN_{tt} for that sensor. The modulus of each value was taken, since it was the absolute residual that was of interest here. This average error method was employed to address the presence of noise that might otherwise lead to useful information being obscured. 10 000 points, equivalent to approximately one month of data, was found to be an optimal window size for removing the noise without also removing important trends and features.

One drawback of this training and testing methodology is that it allowed no opportunity for testing on data that were from the period when the bridge was undamaged but not previously seen by the ESNs during training. This was an unfortunate consequence of the relative paucity of data available for the undamaged period, which did not allow for a meaningfully large testing dataset to be produced while still providing enough data to train the network with a full range of tilt response to thermal variation. It was

intended that the batch training approach would go a small way to addressing this, since the ESNs would have seen the data as part of small 24 hour windows but not as a continuous sequence, but it must be noted that this testing data cannot be considered truly 'unseen' in the conventional sense.

A threshold value needed to be determined for application to the residual in order to discern between residuals suggestive of damaged and undamaged bridge behaviour. This threshold was tuned according to the small variability seen when ESN_{tt} was continuously presented with the training data. A residual value below the threshold (i.e. within the typical variability seen in the training period) would indicate normal sensor activity at that point, while a residual above the threshold would be indicative of an anomalous event. A number of different threshold values were trialled, and it was found that a value of 0.02 best reflected the difference between normal and anomalous behaviour.

If the residual exceeded 0.02, the full duration of this discrepancy was observed, since a short duration may have been indicative of an initial response to an intervention, whereas a longer time-scale may suggest permanent damage. In this analysis, it was taken that the greater the size of the residual, the greater the level of damage. Although ground truth data was available, only the dates of significant interventions were recorded (see Table 4.1), and the extent of any long lasting damage caused by an intervention was not known.

Since they were located on free-hanging cantilevers, the biggest residual errors would probably be seen in tilt sensors 1 and 6. This might not necessarily indicate significant structural damage to the bridge, but could instead be a reflection of the freedom of movement in the bridge at that point. However, large residual errors in tilt sensors 2, 3, 4 and 5, all of which were located in the central part of the bridge, would be more likely to reflect a major structural change, since they did not have the same freedom of movement.

Tilt Sensor 1	**Tilt Sensor 2**	**Tilt Sensor 3**	**Tilt Sensor 4**	**Tilt Sensor 5**	**Tilt Sensor 6**	**Tilt Sensor 7**	**Tilt Sensor 8**	**Avg.**
0.8791 (0.0118)	0.8059 (0.0117)	0.8647 (0.0106)	0.9723 (0.0039)	0.9614 (0.0043)	0.9558 (0.0050)	0.9666 (0.0032)	0.9388 (0.0062)	0.9181 (0.0053)

Table 4.3: The average PCC for each tilt sensor and 100 trained ESNs, with the standard deviation of these given in brackets.

4.3.2 Results and Discussion

The training regime was designed to allow ESN_{tt} to match the behaviour of the bridge prior to manual interventions, but potentially deviate thereafter due to damaging interventions. The average PCC for each tilt sensor and the 100 trained ESN_{tt} architectures, along with the standard deviation, can be seen in Table 4.3.

It can be seen from the PCC values in Table 4.3 that ESN_{tt} was successfully able to model the output of the tilt sensors prior to the first significant intervention. This confirms that ESNs are capable of learning a diurnal relationship and then applying that to a continuous set of data, without the need to detrend or deseasonalise the data in advance. The average PCC of 0.9181 suggests a strong positive correlation between the original tilt sensor data and the ESN_{tt} prediction. An example of the good correlation between real and predicted tilt sensor data in the training portion can be seen in Figure 4.5, which shows the close match between the real sensor data for tilt sensor 5 and the output of the best performing ESN over the period of 19$^{\text{th}}$ March 2009 to 23$^{\text{rd}}$ March 2009. This is significant, since it shows that ESN_{tt} was able to predict accurately how the bridge should normally behave, strongly implying that any difference between the real and predicted values in the testing portion of the dataset would be due to a change in the state of the bridge, rather than a fault in the predictive capability of ESN_{tt}.

While the value of 0.8059 for the PCC between ESN_{tt} and tilt sensor 2 represents reasonably good correlation, this value is notably lower than that for the other tilt sensors. Further analysis of the output of ESN_{tt} showed that tilt sensor 2 was particularly sensitive to the deliberate interventions. It is possible that the region around tilt sensor

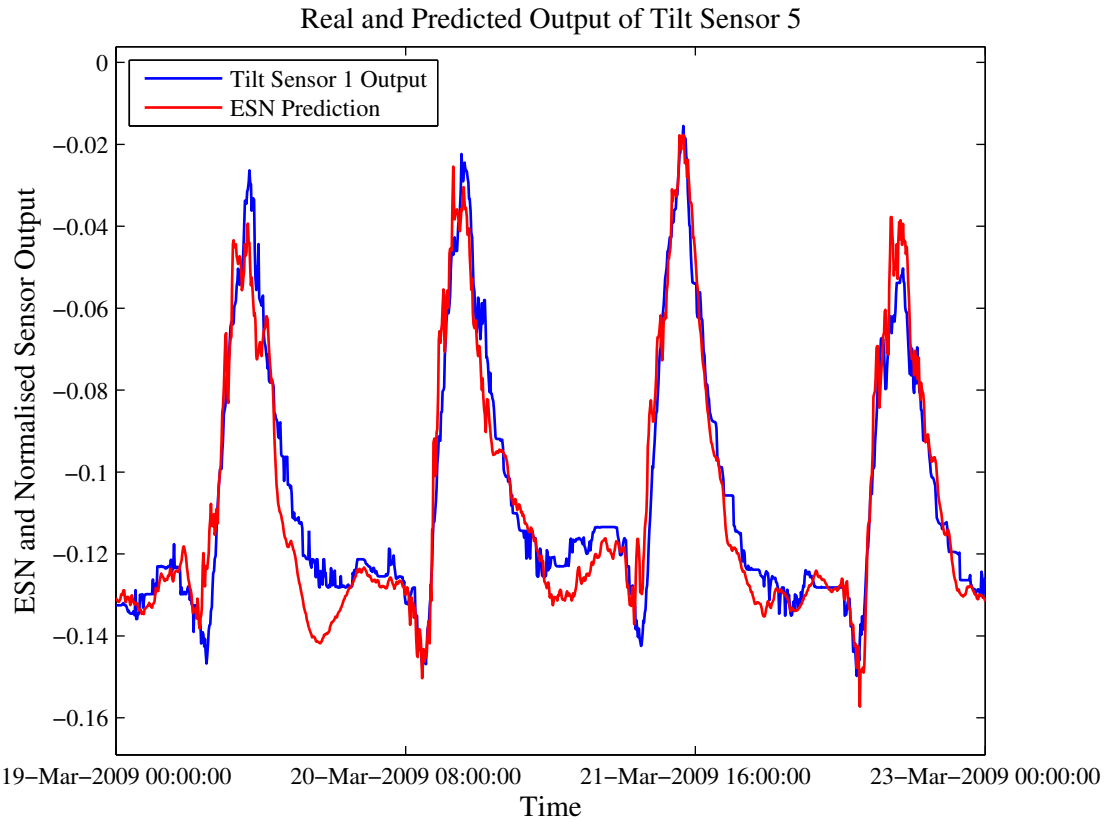


Figure 4.5: The real and best ESN predicted output of tilt sensor 5 over the period of 19$^{\text{th}}$ March 2009 to 23$^{\text{rd}}$ March 2009. ESN$_{tt}$ has learned to model the behaviour of the tilt sensor well.

2 was prone to damage and that there was an underlying structural condition in the bridge at this point that had developed during the 50 years of ordinary use that the bridge underwent prior to the beginning of the sensor monitoring. This could have caused the bridge to start behaving erratically in this region, making it especially difficult for any regression technique to model its behaviour based on temperature alone.

Using Equation 4.1 to analyse the output from ESN$_{tt}$ allowed two key interventions to be identified by inspection of the long-term trends in the data. Figure 4.6 shows a moving average of the residual data for tilt sensors 1 - 3 and 6, during the period July 2009 to May 2010, when the effects of the first key intervention could best be seen. Figure 4.7, meanwhile, shows a moving average of the residual for the same tilt sensors for the period June 2010 to April 2011, when the effect of the second key intervention could best be seen. Each solid vertical red line in these figures represents a significant intervention that occurred at that point in time. A rise in the error level due to an

intervention that is then maintained over a long time period indicates that permanent damage or at least medium term modification to the bridge was caused. A rise in the error level following an intervention that is followed by the error returning quickly to its prior level is indicative of the event affecting the bridge at that point in time, but not causing any lasting damage. If there is no change in the error level, the bridge was probably unaffected at that tilt sensor position.

The first of the two key interventions found using ESN_{tt} occurred in August 2009, when two water tanks suspended from one end of the bridge were filled with water and then emptied. The water tanks were loaded onto the cantilever closest to tilt sensor 1 (see Figure 4.6), and so this sensor exhibited the strongest response to the event itself, something probably exacerbated by the fact that the cantilever was relatively free to move. In the aftermath of this, however, the residuals for tilt sensors 1, 2 and 3 all begin to increase beyond the damage threshold of 0.02. This implies that the bridge had changed state due to the intervention. However, while the residual for tilt sensors 2 and 3 reached a certain value and then began to plateau, the freedom of movement in the cantilever meant that the residual for tilt sensor 1 decreased after peaking in December 2009. This suggests that the cantilever was temporarily affected by the static test, but that it eventually returned to its initial state. Conversely, the fact that the residual error was maintained for tilt sensors 2 and 3 indicates that the bridge did undergo a permanent change of state in that region, potentially indicative of damage. The fact that the error is greater for tilt sensor 2 means that any damage probably occurred closer to tilt sensor 2 than 3.

The residual error for tilt sensor 6 is more difficult to interpret for two reasons. Firstly, the residual is erratic, often moving above or below the damage threshold with no apparent cause; when the residual error for the other three tilt sensors continued to increase in September 2009, the error for tilt sensor 6 began to decrease. In January, the residual error began to grow again without an obvious cause for this change in behaviour. Secondly, the fact that the sensor was located on the cantilever at the opposite end of the bridge to tilt sensor 1, and that both cantilevers were free to move,

suggested that the static tests would elicit a strong response from both cantilevers initially. It is strange that the error for tilt sensor 1 would be maintained for such a long time period, often at a greater level than the error for tilt sensor 2, when tilt sensor 6 behaved so differently. This could be due to external effects other than temperature that made it difficult for the ESN to model tilt sensors 1 and 6 using temperature alone, something that could be addressed with data fusion.

The second of the key interventions came about due to the loading and unloading of two half-full water tanks on the bridge between the 30$^{\text{th}}$ of June and the 2$^{\text{nd}}$ of July 2010, the effects of which were then exacerbated by subsequent interventions on the 8$^{\text{th}}$ and 18$^{\text{th}}$ of October 2010. Figure 4.7 shows the behaviour of tilt sensors 1, 2, 3 and 6 between July 2010 and April 2011, covering the three interventions and their aftermath.

Any damage that the first significant event imparted upon the bridge near tilt sensors 2 and 3 was exacerbated by both the loading events of June/July 2010 and then the static tests of October 2010. The residual can be seen to increase for both sensors after these events, before eventually plateauing. Prior to the June/July loading events, the error for tilt sensor 2 was around 0.03, which had increased to 0.07 by the time it plateaued in April 2011, more than double the initial value. Similarly, the residual error for tilt sensor 3 doubled from 0.02 in June 2010 to 0.04 in April 2011. These results suggested that the area around tilt sensor 2 started in a damaged state due to the first key intervention mentioned previously, and that this damage was increased by further loading events.

The residual error for tilt sensors 1 and 6, meanwhile, still suggested that temperature information alone was not sufficient for modelling free-hanging cantilevers. Tilt sensor 1 exhibited a strong response to all of the events before returning to a low value, indicating that any change to the bridge at that point was temporary. Curiously, the output again peaked during December, once more exceeding the residual generated in the damaged tilt sensor 2 region. The residual for tilt sensor 6 continued to be erratic, breaking the damage threshold at times but ultimately returning to zero by April 2011. Once again, this implies that the recorded temperature measurements were not sufficient on their

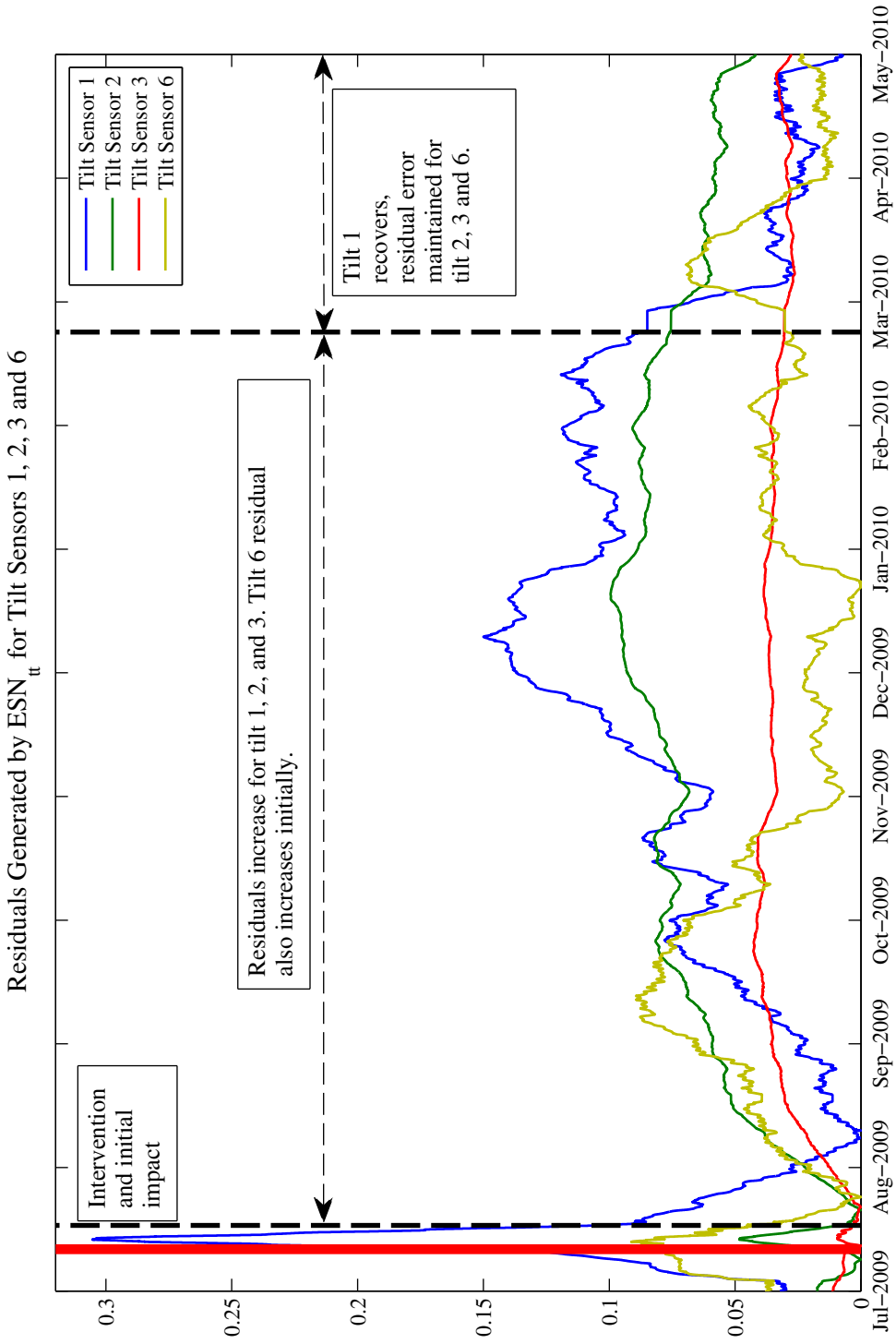


Figure 4.6: Residuals generated for tilt sensors 1, 2, 3 and 6 for the period 24th July 2009 - 15th May 2010. The red vertical line indicates the onset of the intervention.

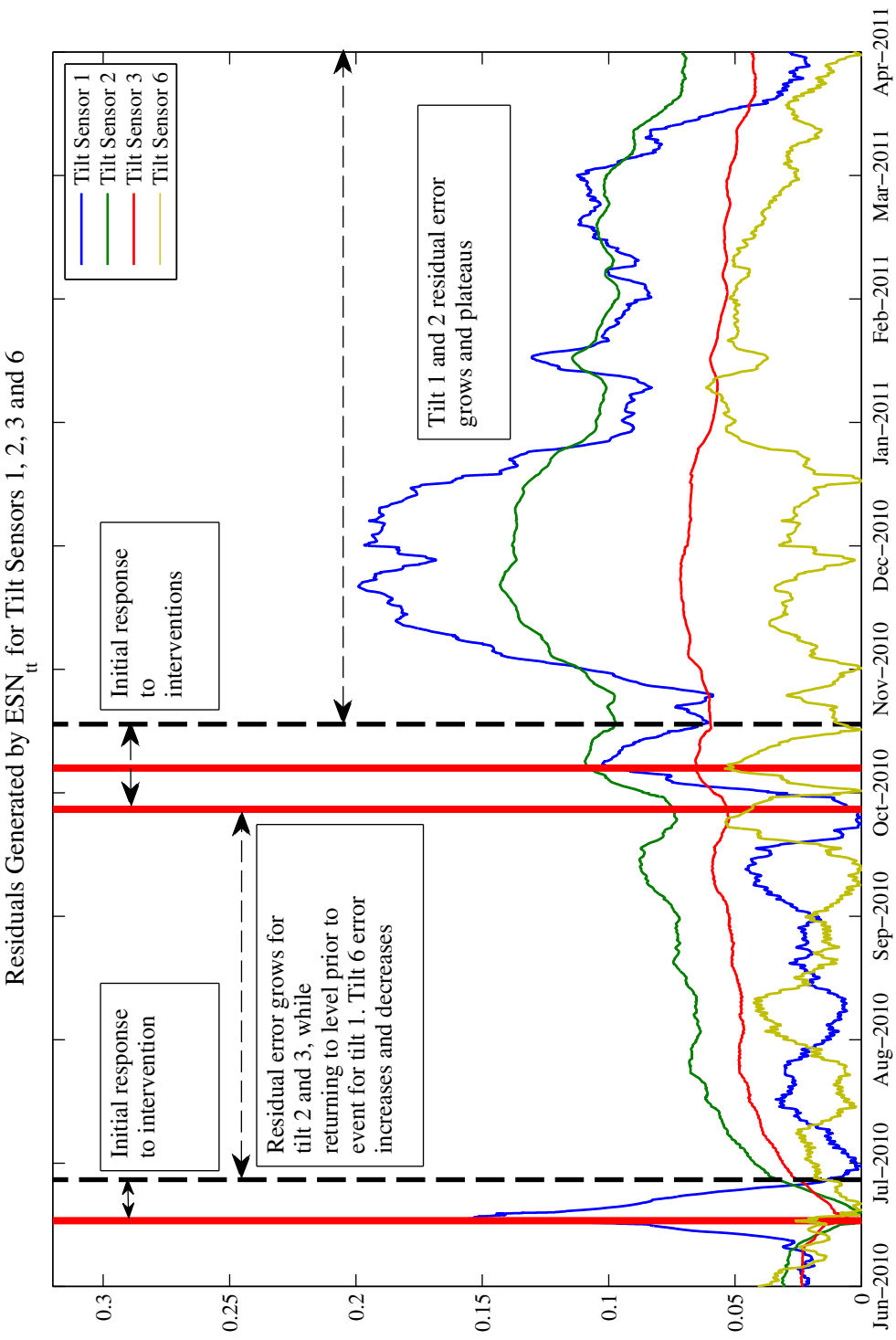


Figure 4.7: Residuals generated for tilt sensors 1, 2, 3 and 6 for the period 14th June 2010 - 10th April 2011. The red vertical lines indicate the onset of an intervention.

own for modelling tilt on free-hanging cantilevers.

4.3.3 Conclusion

There are a few, more general conclusions that can be reached from this initial phase of work on this case study. Firstly, it has been shown that ESNs are well suited to modelling the thermal dependency of the tilt of the bridge at different locations. Prior to the first significant damaging event, good correlation between the real and predicted tilt sensor data was achieved, with an average PCC of 0.9181 achieved across all eight tilt sensors. This was achieved without any need for removing seasonal trends from the data, or synthesising additional training data. Instead, the ESNs were able to embrace the seasonal variations in the data using their ability to model temporal relationships, and this was done using only six months of data. It is, however, important to sound a note of caution at this point, since the network had seen all of the data prior to the first significant damaging event in training, although the data had only previously been seen in batches and were not presented in the continuous fashion used during testing. The fact that the increase in residual begins almost immediately after the end of the training period could potentially be concerning, but can be plausibly explained by the fact that a significant intervention did occur at this point, and that similar behaviour was observed after the second significant loading event. In spite of this, the results suggest that ESNs are a plausible option for online SHM. By not only detecting important events, but also giving some indication of long-term damage and damage locations, the ESNs were able to give a clear picture of the bridge's condition.

Although this preliminary work did not use data fusion in the conventional sense, where input modalities might be combined in order to improve the output results, the exploitation of the relationship between two different sensor modalities showed the potential benefits of mapping between sensor modalities. Looking at sensors in this kind of combination allowed for inferences that could not have been made using just tilt sensor readings or temperature sensor readings alone. The temperature sensors added vital context to the variations in tilt sensor data not just over the course of

a day, but over the course of an entire year. However, the existence of a thermal response by structures is well known, and has been used before (see, for example, [197]). Furthermore, the behaviour of the bridge might not be fully accounted for by its response to temperature, and other external factors could potentially be at play. This could be the reason for the slightly counterintuitive behaviour seen in the residual error for tilt sensors 1 and 6. Sohn previously reported that while temperature is the chief contributor to dynamic changes in a bridge, there are other confounding variables such as wind-induced variation that should be considered [445]: something that seems particularly pertinent for free hanging cantilevers. In a 'real-world' bridge that is in constant use, it is possible that there could be variation in the tilt sensor data based not only on changes in the temperature, but also on the amount of traffic passing over the bridge. Consequently, a data fusion technique would be more useful in the 'real-world' if it could integrate alternative modalities of data that may account for unusual behaviour. The work in the remainder of this chapter makes a novel contribution by not only combining together tilt and temperature sensor data, but by also fusing strain gauge data to obtain better results.

4.4 ESN$_a$: Fusion of temperature, tilt and strain data

The nature of the relationship between tilt sensors, temperature sensors and strain gauges can be seen in Figure 4.8. Although the amplitude of the response is not as great as that of tilt sensor 1, every change in temperature has a corresponding change in the strain on the bridge. This follows, as one would expect variations in the tilt at different points on the bridge to cause variations in the strain at different points. Past research also gave some cause for optimism: it has been suggested that strain gauges can be integrated with environmental sensors for condition monitoring [446], while strain measurements have also been used for modelling acceleration [196, 211] and have been analysed for damage detection [447].

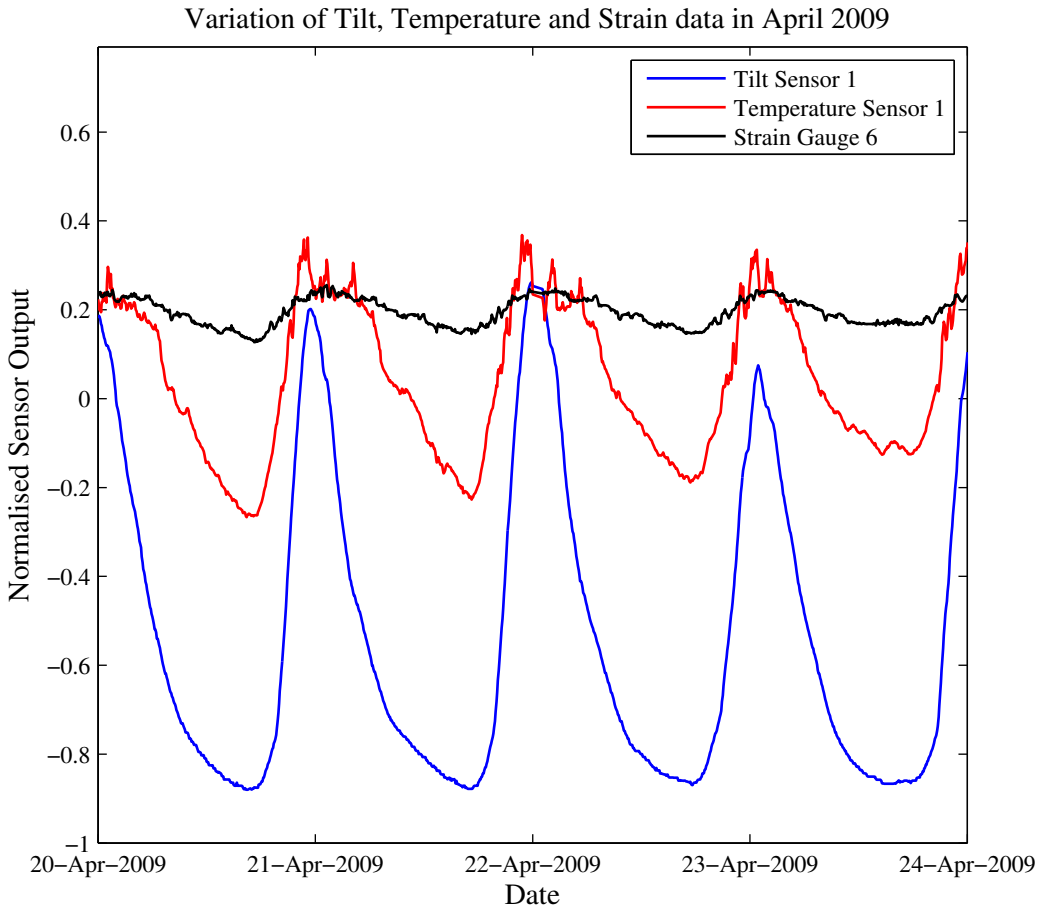


Figure 4.8: The output of tilt sensor 1, temperature sensor 1 and strain gauge 6 between 20$^{\text{th}}$ April and 24$^{\text{th}}$ April 2009. Like the tilt sensor, the strain gauge responds to variations in the temperature of the bridge.

The motivation for including strain data in the data fusion model was that there was a possibility that there may be a relationship between tilt and strain that is independent of temperature. If this was the case, some of the deviations seen in the residuals generated by ESN$_{tt}$ may not necessarily have been indicative of damage. A system that employed different modalities of sensor would be able to take more than just the thermal response of the bridge into account, producing a more accurate model of tilt sensor behaviour. Any subsequent improved performance in terms of PCC during training would increase confidence in the interpretation of the residual error generated by an ESN. It is also possible that including the strain gauge data might address some of the idiosyncrasies seen in the residual error produced by ESN$_{tt}$ for tilt sensors 1 and 6.

4.4.1 Strain gauge data

Although the strain gauges were embedded in the ten temperature sensors, the actual availability of strain gauge data was limited to the six sensors embedded in temperature sensors 3, 6, 7, 8, 9 and 10 (see Figure 4.3). The constant output of -7999 that was given by the strain gauges embedded in temperature sensors 4 and 5 (c.f. typical readings of 0.122 and 0.00434) for extended periods was well beyond the sensors' operational limit of 3000 microstrain [448], suggesting that they had developed stuck faults. The frequency and duration of these faults meant that those two strain gauges gave very little useful information. They were, therefore, excluded from the analysis. The data for the six working strain gauges only covered the period from 1$^{\text{st}}$ February 2009 to 1$^{\text{st}}$ February 2011, compared to 30$^{\text{th}}$ January 2009 to 4$^{\text{th}}$ May 2012 for the other sensors. Consequently, the subsequent analysis in this section looks only at events that took place prior to February 2011.

4.4.2 Modelling strain with temperature

The first part of the ESN approach here was to attempt to model the thermal response of the strain of the bridge. The same residual error approach used in Section 4.3 should make it possible to observe points at which the strain deviated from its expected values, and to possibly observe sensor faults. ESN$_{asg}$ was created in order to do this.

4.4.2.1 Methodology

The ESN$_{tt}$ architecture described in Section 4.3 above was slightly modified so that there were six output units, one for each strain gauge. It was found that optimal performance was obtained with all other ESN parameters kept the same as for ESN$_{tt}$ (see Table 4.2). The portion of data up to the first significant event on 29$^{\text{th}}$ June 2009 was again used for training, despite the fact that the later start time for the strain gauges meant that this training set was smaller than the one used to model the tilt sensor data. Each ESN was trained using 74-fold cross validation.

Strain Gauge 1	**Strain Gauge 2**	**Strain Gauge 3**	**Strain Gauge 4**	**Strain Gauge 5**	**Strain Gauge 6**	**Avg.**
0.9123 (0.0649)	0.9082 (0.0724)	0.9107 (0.0699)	0.9054 (0.0756)	0.9119 (0.0671)	0.9092 (0.0670)	0.9064 (0.0088)

Table 4.4: The average PCC for each strain gauge and 100 trained ESN$_a$ architectures, with the standard deviation of these given in brackets.

100 ESNs using this architecture were trained. The full dataset was then presented to the ESN continuously and Equation 4.1 was used to find the rolling average of the difference between the predicted strain gauge output and the real strain gauge output. It should once again be noted that the paucity of available data meant that all of the undamaged data was presented in batches during training, such that there was no unseen testing data from the undamaged period.

4.4.2.2 Results

As with the tilt sensor data, the training regime was designed so that the ESNs would be able to accurately model the strain gauge data recorded prior to the first manual intervention. Table 4.4 gives the average PCC for each strain gauge and the 100 trained ESN$_{asg}$ architectures, with the standard deviation given in brackets.

The results shown in Table 4.4 are remarkably consistent, with all of the PCC values falling within 0.05 of 0.91. The overall average PCC value of 0.9064 suggests a strong correlation between the ESN$_{asg}$ output and the original strain gauge data, indicating that it was possible to model the strain across the footbridge using the temperature. This, in turn, suggested that the full extent of the data could be fed into ESN$_{asg}$, and a residual generated in order to detect deviations in the strain. The residual for all six strain gauges over the full time period can be seen in Figure 4.9.

Figure 4.9 shows that while all of the strain gauges were in some way affected by the first significant intervention, only strain gauge 2 significantly deviated from the ESN$_{asg}$ prediction. A closer inspection of the raw strain gauge data revealed that in May 2009, the sensor appears to have begun to produce spike and bias faults that caused it to frequently deviate from the ESN$_{asg}$ prediction. An example of this can

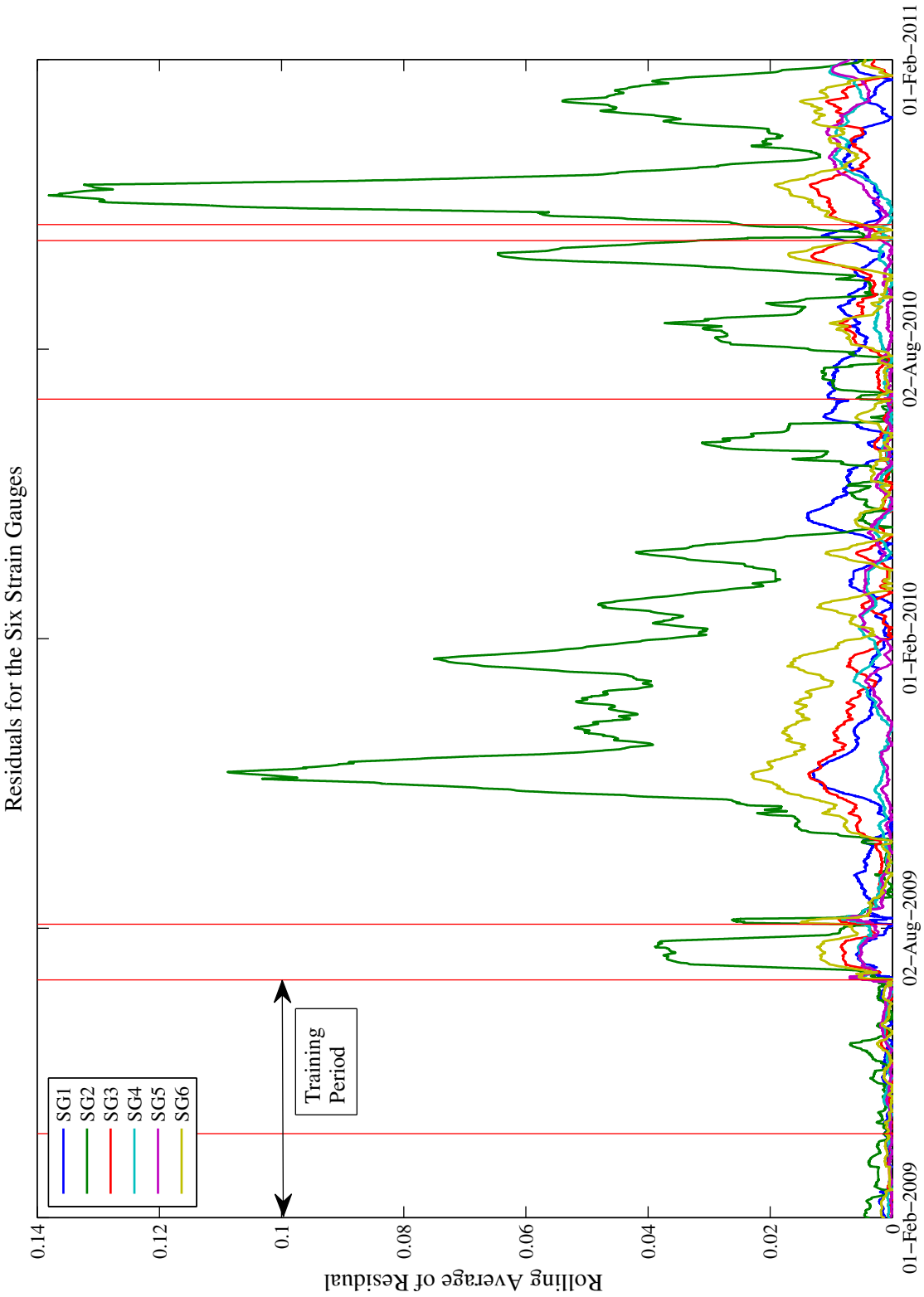


Figure 4.9: Residuals generated for the six strain gauges over the full time period. The red vertical lines indicate the onset of interventions.

be seen in Figure 4.10, which shows how the ESN was able to model the strain gauge data well, up until the point when a spike fault occurred, followed by the emergence of a bias fault. Faults like these were responsible for the large residual seen for strain gauge 2. On a moveable structure like the NPL footbridge, spikes are often caused by loose electrical contacts within the sensor [449]. In this particular case, the size of the residual did not indicate damage to the bridge, but damage to the sensor. The faulty behaviour of strain gauge 2 was taken into account when using the strain gauges to model the behaviour of other sensors.

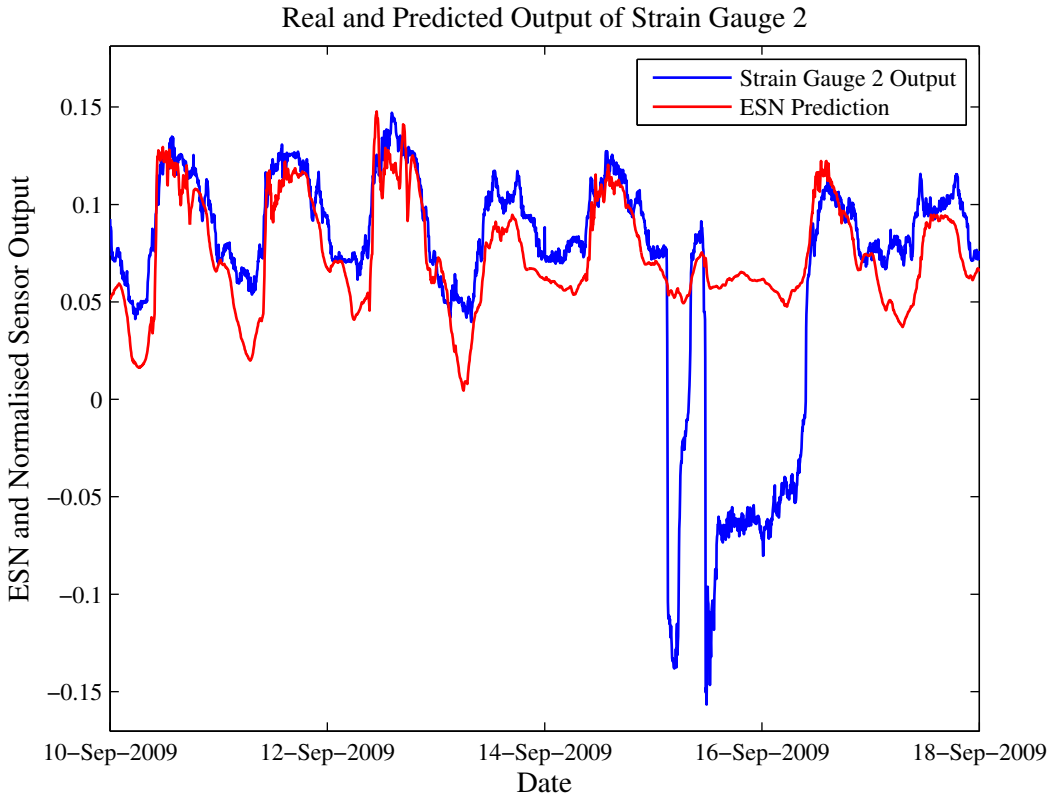


Figure 4.10: The real and predicted output of strain gauge 2 between 10$^{\text{th}}$ September and 18$^{\text{th}}$ September 2009. The ESN models the behaviour of the sensor well, apart from when there are unexpected spike and bias faults on the 15$^{\text{th}}$ of September.

4.4.3 Modelling tilt with strain and temperature

In an online SHM application, where the data are arriving in real-time, it would not be known from the outset that strain gauge 2 had developed a fault. Consequently, it was decided that two separate ESN$_a$ models for predicting the tilt sensor data would be created. The first of these, ESN$_{a16}$ would have 16 input units - the ten temperature

sensors and all six strain gauges - and eight output units (one for each tilt sensor), while the second, ESN$_{a15}$, would have 15 input units - the same as ESN$_{a16}$, but without the faulty strain gauge 2 - and, again, eight output units. The aims of this approach were:

1. To see if there was initially any benefit to using all six strain gauges, and if the ESN$_{a16}$ model was adversely affected by the development of the major faults in one of the input sensors.

2. To see if, upon 'detecting' a sensor fault, it is possible to switch to a pre-trained model that excludes the faulty sensor (ESN$_{a15}$ in this case) and still obtain useful results. [1]

3. To answer Research Question 1 by improving on the sensor interpretation from the preliminary work.

4. To answer Research Question 2 by considering how the ESN$_a$ approach could be applied in a 'real-world' online SHM scenario.

5. To answer Research Question 3 by finding if it is still possible to systematically apply ESN$_a$, even in the presence of faulty sensors.

4.4.3.1 Methodology for data fusion

It was again found that the best ESN parameters for both ESN$_{a16}$ and ESN$_{a15}$ were the same as the optimal parameters for ESN$_{tt}$, given in Table 4.2. The training and testing procedure was also unchanged from the procedure detailed in Section 4.3.1. 100 ESNs of each architecture were trained on discontinuous batches of individual days of data from the training period using 74-fold cross validation, before being continuously presented with the data from the training period during testing. This again meant that the ESNs would not have seen the training data presented in sequence prior to testing,

[1] For a modest number of sensors, the relative ease of training ESNs means that this could be achieved by exhaustively training all possible permutations of sensors and then choosing a network that excludes those sensors that have become faulty. For a bigger, and possibly more realistic, system, a new set of ESNs would have to be trained offline, without the excluded sensor(s), once a sensor fault was detected.

but that there would be no previously unseen data from prior to the first damaging event available for testing. The best performing individual ESN for each architecture was found, and was then used for the production of the graphs and figures seen in Section 4.4.3.2, below. After this, the ESNs were presented with the data from the entire monitoring period so that residuals could be generated for each tilt sensor and potential damage inferred.

4.4.3.2 Results for correlation prior to first significant event

Table 4.5 shows the PCC results when the 100 trained ESN$_{a16}$ architectures were continuously presented with the data from the training period. When these results are compared to those in Table 4.3, one thing that is noticeable is that the PCC value increased for every single tilt sensor, with the overall average rising from 0.9181 to 0.9438. This confirms the hypothesis that there are external factors besides the temperature that affect the behaviour of the bridge, and that some of these are represented in the strain measurements. It also shows an immediate improvement in sensor modelling that has been brought about by data fusion.

Tilt Sensor 1	**Tilt Sensor 2**	**Tilt Sensor 3**	**Tilt Sensor 4**	**Tilt Sensor 5**	**Tilt Sensor 6**	**Tilt Sensor 7**	**Tilt Sensor 8**	**Avg.**
0.9226 (0.0044)	0.8750 (0.0067)	0.8973 (0.0044)	0.9767 (0.0017)	0.9703 (0.0015)	0.9754 (0.0015)	0.9752 (0.0011)	0.9578 (0.0025)	0.9438 (0.0025)

Table 4.5: The average PCC for each tilt sensor and 100 trained ESN$_{a16}$ architectures when all ten temperature sensors and all six strain gauges were used as inputs, with the standard deviation of these given in brackets.

The PCC values from when the 100 trained ESN$_{a15}$ architectures were continuously presented with the data from the training period, meanwhile, are given in Table 4.6. While there is a slight decrease in the PCC for seven of the eight tilt sensors and the overall average, this effect is very small, and the changes were all within one standard deviation. This indicates that even with one of the sensors removed, the data fusion approach still offers an improvement over the preliminary model. This means that in

the case of a detected sensor fault, it would be plausible to simply move to a pre-trained model that excluded the faulty sensor.

Tilt Sensor 1	**Tilt Sensor 2**	**Tilt Sensor 3**	**Tilt Sensor 4**	**Tilt Sensor 5**	**Tilt Sensor 6**	**Tilt Sensor 7**	**Tilt Sensor 8**	**Avg.**
0.9210 (0.0045)	0.8690 (0.0061)	0.8930 (0.0040)	0.9764 (0.0012)	0.9697 (0.0011)	0.9754 (0.0016)	0.9744 (0.0012)	0.9566 (0.0030)	0.9419 (0.0022)

Table 4.6: The average PCC for each tilt sensor and 100 trained ESN$_{a15}$ architectures, with the standard deviation of these given in brackets. Removing the erroneous strain gauge only slightly reduced performance during the training period, and still offered an improvement on the model used in the preliminary work.

The overall effect of the data fusion approach on the ESN prediction for each tilt sensor can best be seen in Table 4.7, which shows the PCC for each tilt sensor prediction produced by the best of the 100 trained ESN$_{tt}$ architectures and the best of the 100 trained ESN$_{a15}$ architectures. The biggest improvement in PCC was for tilt sensor 2.

While the residual errors for some of the tilt sensors were hardly changed by the introduction of the strain data, tilt sensors 1, 2, 3 and 6 were significantly affected, as discussed below in Section 4.4.3.3. This is an interesting result, since these are the sensors that had the poorest PCC for ESN$_{tt}$ and which were in the areas that were identified as being most prone to additional external effects. While the good correlation between the two ESNs for tilt sensors 4 and 5 is easy to explain - ESN$_{tt}$ was able to account for their behaviour very well using only information on the temperature, indicating that this region of the bridge was less prone to external effects - the reason for the good correlation for tilt sensors 7 and 8 is less obvious. No strain gauges were located on the two piers, and so it is probably the case that using strain gauge data did not give ESN$_{a15}$ any additional information about how the tilt should vary on the piers.

4.4.3.3 Results for application to the data from the full monitoring period

When the ESNs were presented with the dataset from the full monitoring period, most of which had not been previously seen by the networks, the effects were significant. Figure 4.11 shows the effects of two of the worst faults in strain gauge 2 on the residuals

Tilt Sensor 1	**Tilt Sensor 2**	**Tilt Sensor 3**	**Tilt Sensor 4**	**Tilt Sensor 5**	**Tilt Sensor 6**	**Tilt Sensor 7**	**Tilt Sensor 8**
0.5960	0.4465	0.7266	0.9388	0.9072	0.8492	0.9562	0.8817

Table 4.7: The PCC for each tilt sensor as predicted by the best performing of the 100 ESN$_{tt}$ architectures and the best performing of the 100 ESN$_{a15}$ architectures over the full extent of the dataset. It can be seen that the prediction for some tilt sensors was largely unchanged, while there were pronounced changes for others.

produced by ESN$_{a16}$ for selected tilt sensors. In each case, it can be seen that a large spike is produced in each residual, which corresponds to the spike in the residual for strain gauge 2. The peak of each spike is well in excess of the damage threshold of 0.02, meaning that the sensor's faults run the risk of being interpreted as the sudden onset of damage in the locations around tilt sensors 1, 2, 3 and 6.

In contrast, Figure 4.12 shows how this problem was resolved by removing the faulty sensor from the inputs to the network. The spikes seen in the residuals generated by ESN$_{a16}$ are totally absent from the residuals generated by ESN$_{a15}$. Furthermore, ESN$_{a15}$ revealed a spike in the residual for tilt sensor 1 that was obscured in the residual from ESN$_{a16}$. While the removal of a faulty sensor from the ESN model marginally reduced the PCC during the training period, the overall performance of ESN$_{a15}$ improved upon that of ESN$_{a16}$ when presented with the full dataset. Artefacts from the faulty sensor were removed without significantly compromising the ESN's ability to predict the tilt sensor values. This means that in an online SHM application, it would plausibly be possible to detect a fault in one of the input sensors, and to simply use a new model (either pre-trained or trained offline once the sensor fault had occurred) that did not include the faulty sensor.

4.4.4 Discussion

The results for ESN$_{asg}$, ESN$_{a16}$ and ESN$_{a15}$ demonstrated a number of useful points. Firstly, as shown by ESN$_{asg}$, it is possible to model the strain of the bridge based on temperature, and to use this model to detect faults in the sensors. Secondly, as shown

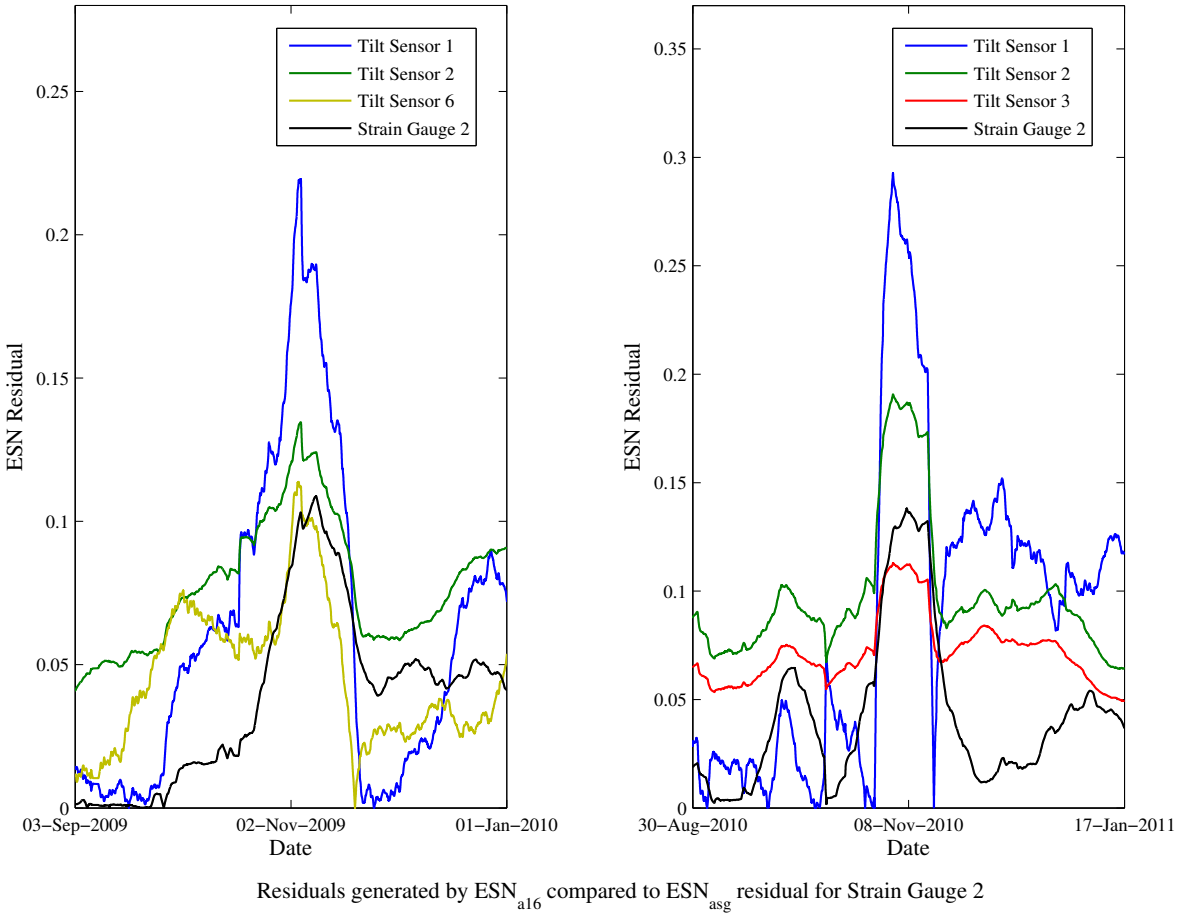


Figure 4.11: A comparison between selected residuals from ESN$_{a16}$ and the residual for strain gauge 2 produced by ESN$_{asg}$. In each plot, it can be seen that the increase in residual for strain gauge 2 caused by sensor fault also results in increases in the residual for the tilt sensors.

by ESN$_{a16}$ and ESN$_{a15}$, it is possible to gain improved models of the tilt of the bridge by incorporating other modalities of sensor as inputs to the ESNs. Thirdly, when sensor faults are detected, faulty sensors can be removed from the inputs with only a minimal drop in network performance.

All of the subsequent discussion in this section refers to the better-performing of the two strain gauge data fusion ESNs: ESN$_{a15}$.

Table 4.7 shows that the biggest differences between the residuals generated by ESN$_{tt}$ and ESN$_{a15}$ were seen in tilt sensors 1, 2, 3 and 6. Figure 4.13 shows how the incorporation of strain gauge data gave residuals for tilt sensors 1 and 6 that were more in line with prior expectations. The period shown in Figure 4.13 is for the full extent of the available strain data. Both sensors still exhibit a large residual when

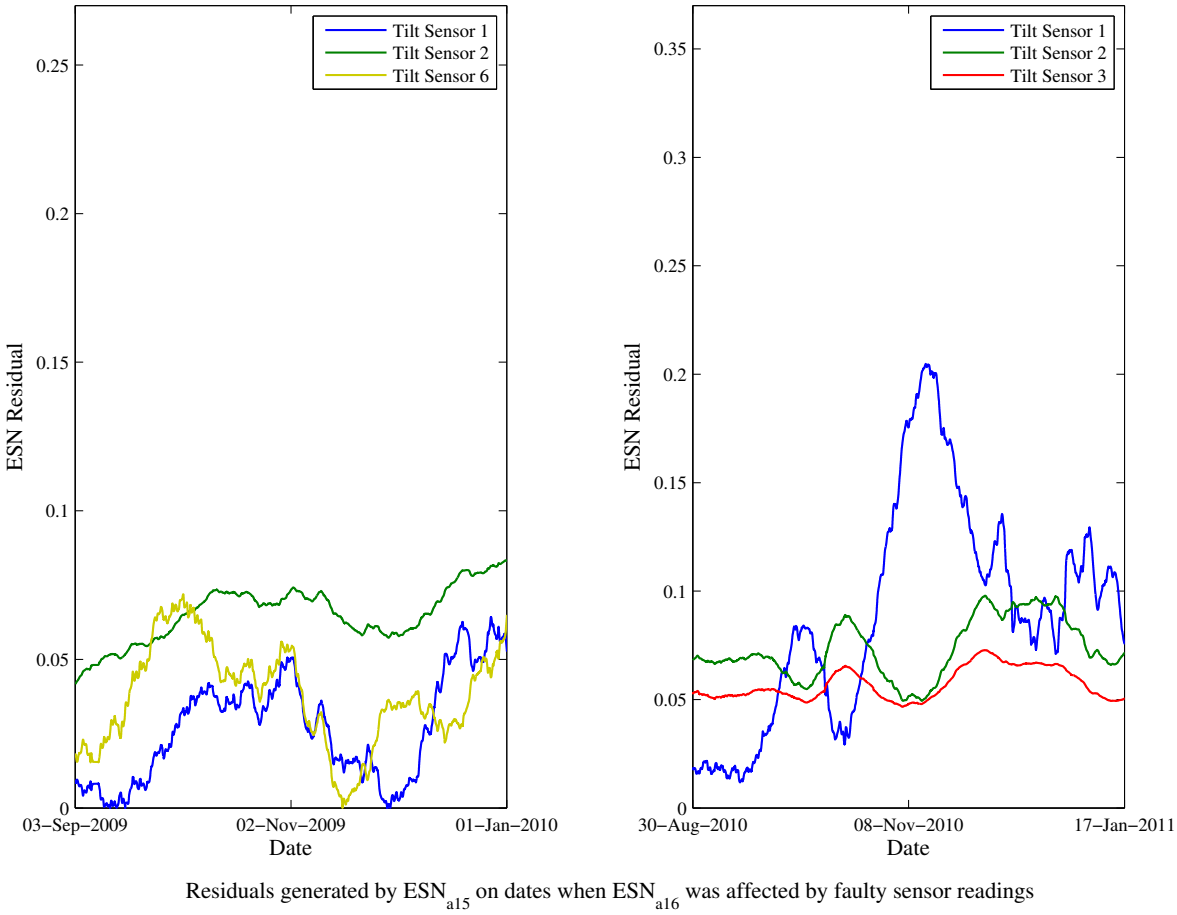


Figure 4.12: The residuals generated by ESN$_{a15}$ for selected tilt sensors on the same dates as the plots in Figure 4.11. It can be seen the removing the faulty strain gauge from the inputs to the network produced removed the effects from the predicted ESN output.

the static tests were performed, but the residual for tilt sensor 1 no longer rises to a peak in December 2009. Instead, the residual returns to zero after the static test, and then varies afterwards due to its location on a free-hanging cantilever. The residual for tilt sensor 6 now amply shows how the effects of the loading events were felt all across the bridge. Rather than simply varying around the damage threshold of 0.02, the new residual for tilt sensor six changes in keeping with the residuals for the other tilt sensors, and comes to be the tilt sensor with the greatest residual at the end of the data. Again, this is to be expected of a free-hanging cantilever, especially one that would be strongly affected by changes in other regions of the bridge.

Figure 4.14, meanwhile, shows how the residuals for tilt sensors 2 - 5 changed over the same period. A comparison between these sensors is interesting, since they were

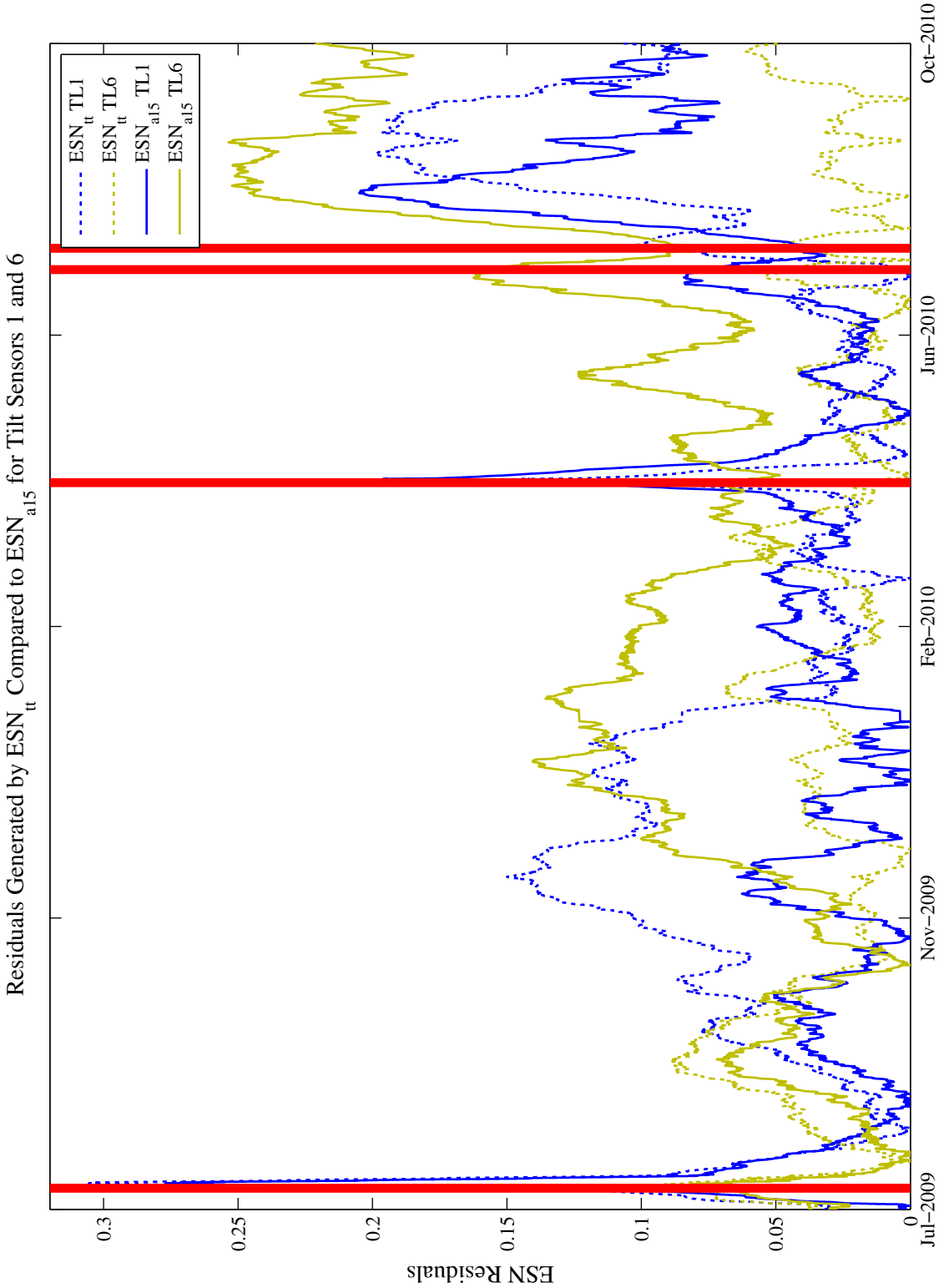


Figure 4.13: Residuals generated for tilt sensors 1, and 6 by ESN$_{tt}$ (dashed lines) and ESN$_{a15}$ (solid lines) during the period 24$^{\text{th}}$ July 2009 - 22$^{\text{nd}}$ January 2010. The red vertical lines indicate the onset of interventions.

identified as the ones that would be best able to indicate any structural damage to the bridge. It can be seen that for the most part, the output of all of the tilt sensors is unchanged. The residuals for tilt sensors 2 and 3 consistently exceeded the damage threshold after each of the static tests, indicating damage to the bridge in that region. In sharp contrast, the residuals for tilt sensors 4 and 5 were consistently below the damage threshold of 0.02, which can be interpreted as suggesting that the bridge was unaffected in this region. The only major change in the residuals can be seen in tilt sensor 2 in the period from around September 2010 to January 2011, where the residual is greatly reduced. This suggests that the damage to the bridge in this region was not quite as bad as ESN_{tt} suggested, but had still been exacerbated by the static tests in June and October 2010. Overall, these results do confirm the conclusion drawn when using ESN_{tt}, in that the interventions caused long-term damage to the bridge in the region around tilt sensors 2 and 3.

When trying to piece together the overall picture of how the bridge was affected by the potentially damaging interventions, it is important to consider how the two piers behaved. Figure 4.15 shows how the residuals for tilt sensors 7 and 8 generated by ESN_{a15} differed from those generated by ESN_{tt}, while also including the residuals generated by ESN_{a15} for tilt sensors 2 and 3 for further comparison. As previously mentioned, there is little difference between the output of ESN_{a15} and ESN_{tt} for tilt sensors 7 and 8, probably owing to the fact that no strain gauges were located on the piers. Tilt sensor 7 was strongly affected by the intervention in July 2009, but this appears to have been a short term effect, since the residual fell below the damage threshold in April 2010 and arrived at a value of 0.009 in January 2011. The residual for tilt sensor 8, meanwhile, increased for almost the entire period following the July 2009 intervention. This could suggest that there was a problem with the piers, or possibly the foundations of the bridge. It is particularly interesting that the residuals for tilt sensors 2 and 3 appear to be strongly linked to the residuals for tilt sensors 7 and 8. Following the static test in July 2009, the residuals for tilt sensors 2 and 3 increased as the residual for tilt sensor 7 increased. As the residual for tilt sensor 7 began to

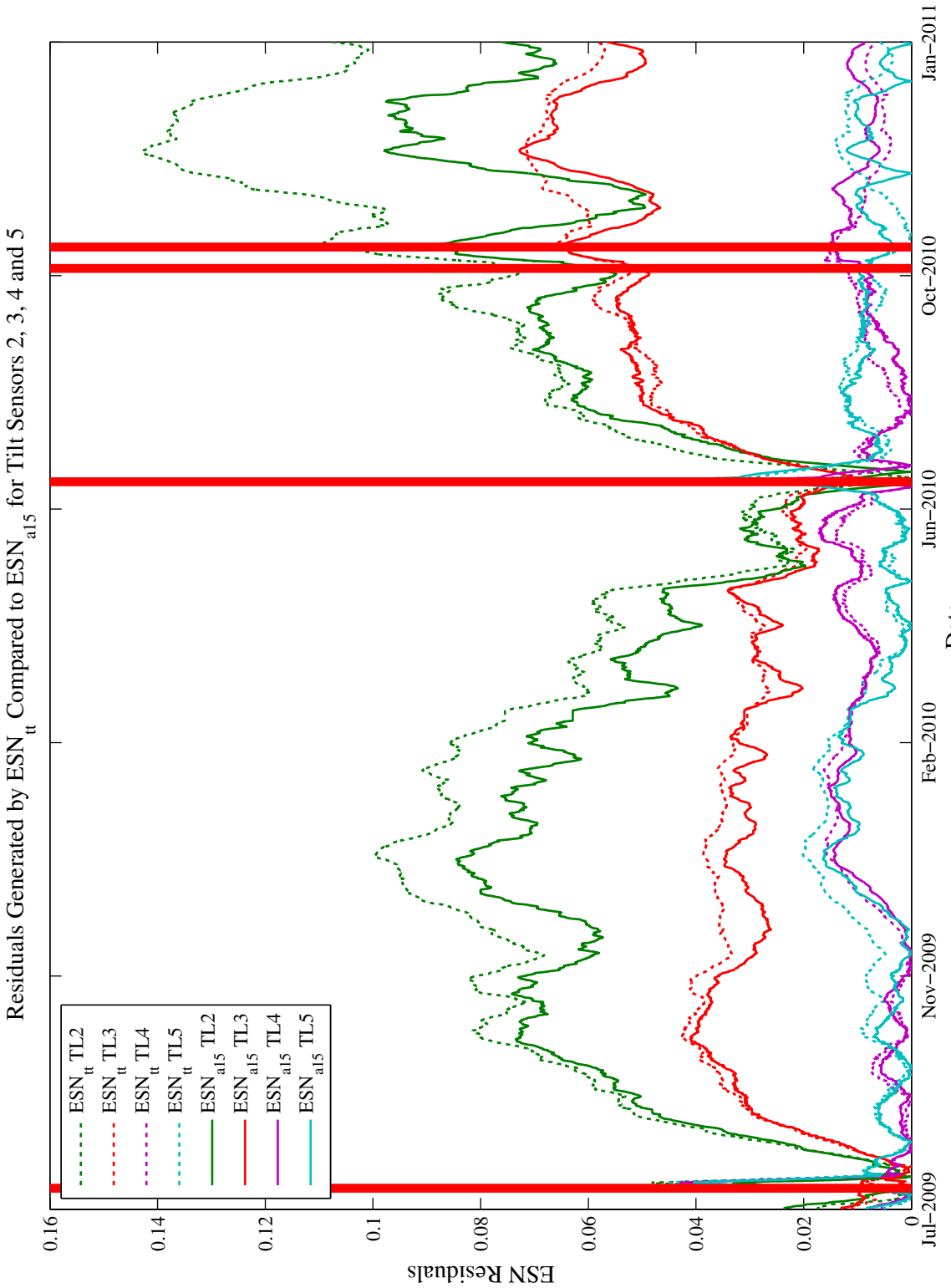


Figure 4.14: Residuals generated for tilt sensors 2, 3, 4 and 5 by ESN$_{tt}$ (dashed lines) and ESN$_{a15}$ (solid lines) during the period 24th July 2010 - 22nd January 2011. The red vertical lines indicate the onset of interventions.

decrease in November 2009, so did the residuals for tilt sensors 2 and 3, although they did experience a sharp spike in January 2010. Similarly, in the aftermath of the June 2010 events, the changes in the residual for tilt sensors 2 and 3 were very similar to the changes in the residual for tilt sensor 8.

Combining all of this information finally allows for the behaviour of the sensors to be explained. In his 2015 thesis, Kromanis suggested that the bridge's foundations experienced differential settlement [440]. Differential settlement is known to cause shears and cracks in structures, something that could only have been made worse by the addition of heavy loads [450]. The results here seem to confirm this, and allow for some interpretation of how and where the bridge became damaged. In July 2009, the water tanks were loaded onto the cantilever closest to tilt sensor 1. The application of this weight caused the bridge to pivot on the first pier, where tilt sensor 7 was located. The initial impact of the resultant force was mostly strongly felt by the bridge in the region around tilt sensor 1, since the cantilever was relatively free to move. Over the subsequent months, it was then tilt sensors 2 and 3 that appeared to show damage. It is likely that the addition of the load, coupled with the differential settlement, caused the pier that tilt sensor 7 was located on to behave in an unexpected fashion. This, in turn, had an effect on the bridge around tilt sensors 2 and 3. It then appears that there was a change in the structure that allowed the pier to return to its usual behaviour, but led to a permanent change in the behaviour of tilt sensors 2 and 3. The fact that the residual for tilt sensor 2 was greater than that for tilt sensor 3 could mean that damage had occurred at a positon closer to tilt sensor 2 than 3. It was reported by Worden et al. that following the intervention, cracking was observed on tilt sensor 7's pier, just above tilt sensor 7 itself, which would explain all of this [441]. Significantly, this damage did not occur at the exact moment of the loading event, but instead developed over several months.

The events of June and October 2010 had a similar effect. This time, it was the pier that tilt sensor 8 was located on that was affected the most by the events, possibly due to torsional effects from the bridge pivoting on tilt sensor 7's pier. This seemed to

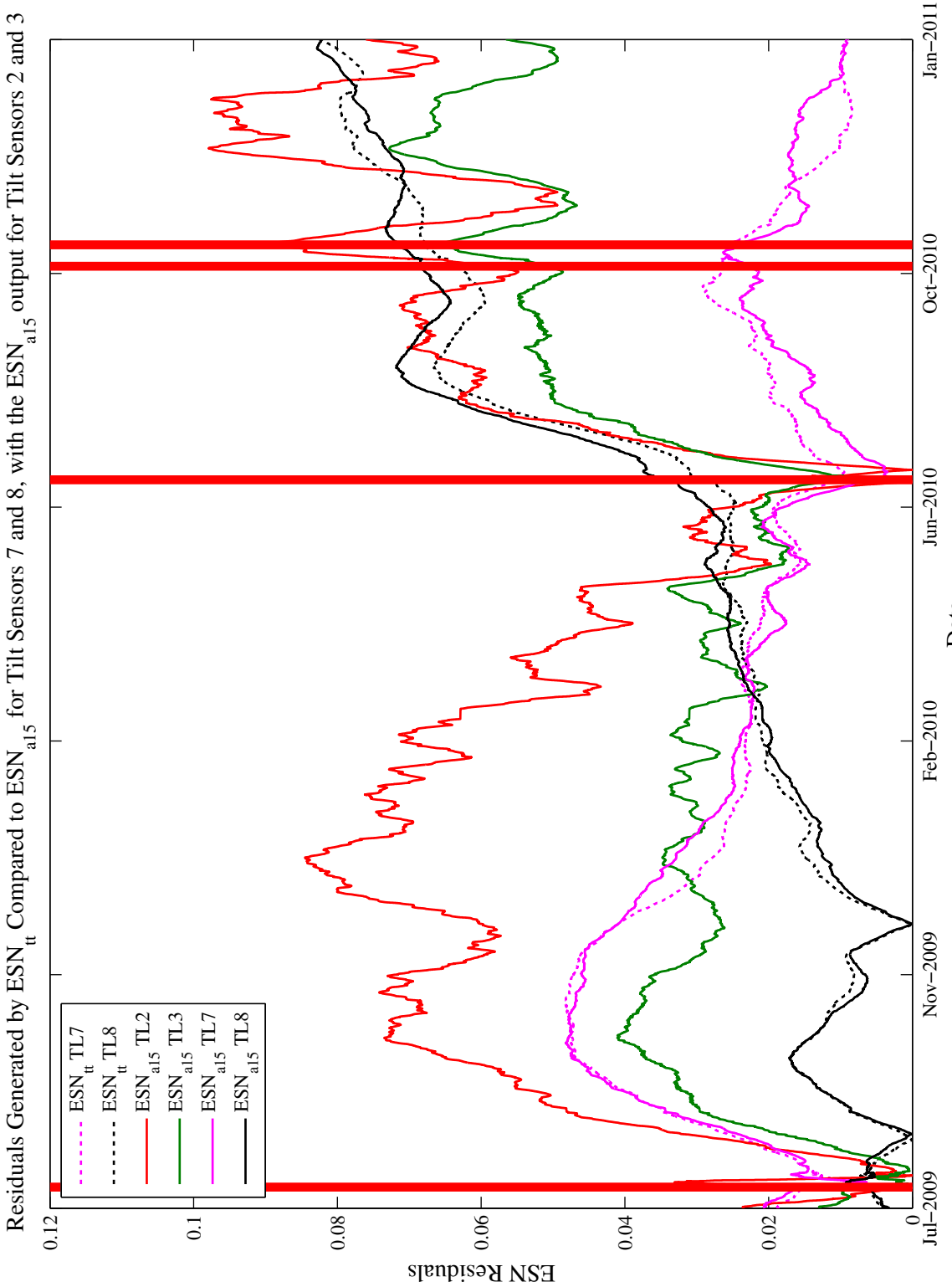


Figure 4.15: Residuals generated for tilt sensors 7 and 8 by ESN$_{tt}$ (dashed lines) and ESN$_{a15}$ (solid lines) during the period 24$^{\text{th}}$ July 2010 - 22$^{\text{nd}}$ January 2011. The ESN$_{a15}$ output for tilt sensors 2 and 3 is also given. The red vertical lines indicate the onset of interventions.

worsen the situation around tilt sensors 2 and 3. By the end of the period covered by the strain gauges, the bridge had been permanently changed, and the region around tilt sensors 2 and 3, in particular, had suffered damage.

Although the overall picture given by these results is somewhat different to that obtained using ESN$_{tt}$, especially for tilt sensors 1 and 6, the improved correlation shown in Table 4.7 gives greater confidence in the accuracy of the results for ESN$_{a15}$. Improving the accuracy of the modelled sensor behaviour during the training period gives greater certainty that the subsequent sensor behaviour predictions would be accurate. The use of strain gauge data fusion here has allowed for more detailed interpretations to be drawn than using just tilt and temperature sensors.

4.4.5 Conclusion

The first part of the ESN$_a$ approach suggested in this section was to use the ESN$_{asg}$ architecture to model the thermal response of the strain gauges. By doing this, it was possible to detect the onset of spike and bias errors in the data from strain gauge 2. It was then shown that combining strain and temperature data produced a more complete model of how the bridge tilted during the time period using ESN$_{a16}$. It was also shown that in the event of an error being introduced to the input to that model, it is possible to compare the output of ESN$_{asg}$ to that of ESN$_{a16}$ in order to see if increases in the residual error in ESN$_{a16}$ were due to genuine damage to the bridge or sensor fault. Finally, ESN$_{a15}$ isolated the faulty sensor with only a slight reduction in performance compared to ESN$_{a16}$. All of this met the first two levels of damage detection, detection and localisation, as set out at the start of this chapter.

While it should again be noted that the small pool of training data available meant that it was not possible to construct a meaningfully large testing dataset, a like-for-like comparison between ESN$_{a15}$ and ESN$_{tt}$ demonstrates the benefits of data fusion in this case. The average PCC when presented with data from the training period rose from 0.9064 to 0.9419, as the strain gauges provided context to the condition of the bridge that was simply not available in the temperature sensor data alone. Using the resultant

residual, it was possible to observe both short and long-term changes in the state of the bridge, allowing for permanent damage to be detected. The results suggested that a fault had developed in the region around tilt sensors 2, 3 and 7, and this was confirmed by a previous physical inspection. This confirmation is important, since it suggests that the ESN interpretation of the sensor data was accurate, something that might be normally tested by assessing correlation when presented with previously unseen testing data. This does not just demonstrate how strain gauges usefully contribute to determining the state of the bridge at any point, but shows that by using CI techniques such as ESNs, it is possible to integrate a variety of different sensors to obtain improved damage detection.

One area of concern is that in this case study, the date of the onset of the interventions was known, which made the interpretation of the results somewhat easier. In the 'real-world' scenarios, this ground truth data would possibly be unknown. It might, therefore, be useful to have further ESN approaches in order to assist in the interpretation of ESN_a. Two new approaches, ESN_b and ESN_c were created with this in mind, and are discussed in the following section.

4.5 Supplementary ESN Approaches

The two complementary approaches that were developed did not rely on data fusion, but did help to interpret the output of ESN_a. They were developed with Worden and Dulieu-Barton's third and fourth levels of damage detection in mind, specifically classification and assessment. ESN_b was created in order to classify every single day in the dataset as either an ordinary day, a day featuring a static test or a day featuring a fatigue test. ESN_c, meanwhile, was created in order to estimate the extent of any damage by using the output of ESN_a to produce a single value that expressed the extent of any damage to the bridge at any given moment.

4.5.1 ESN_b: Classification of Interventions

The potential benefits of being able to classify each day are twofold. Firstly, knowing that a potentially damaging event had occurred would act as a red flag for any engineers monitoring the condition of the bridge, since the output of ESN_a could then be used to observe how and where any damage developed from that point. Secondly, recognising the cause of any long-term change in the bridge would make it easier to look for different types of damage. It was for this reason that ESN_b was tasked with classifying each day as an ordinary day, static test or fatigue test, rather than just as an ordinary day or a day with an intervention.

There was good reason for expecting ESN_b to be able to classify the different types of event. Figure 4.16 shows how the diurnal cycle of the bridge as recorded by tilt sensor 1 could be perturbed by a static test. In the period shown in Figure 4.16, two static tests occurred, one on 24$^{\text{th}}$ June 2011 and one on 27$^{\text{th}}$ June 2011. The signal from tilt sensor 1 on these two days was different from the normal tilt sensor daily cycle for the bridge: in each case, a second spike can be seen shortly after the usual peak in the temperature data. This is due to the weight of the load, applied after 2 pm, causing the bridge to tilt significantly before returning to its normal disposition once the load had been released. This characteristic 'double spike' shape can be seen for all of the 22 static tests that were performed, and was a basis for expecting an ESN technique to characterise the intervention.

Similarly, Figure 4.17 shows how the normal cycle was perturbed by another specific type of intervention, the fatigue test. The fatigue tests were carried out by NPL scientists as a way of simulating damage and assessing the bridge's breaking point. During the period shown in Figure 4.17, fatigue tests occurred on 28$^{\text{th}}$ September, 29$^{\text{th}}$ September, 3$^{\text{rd}}$ October and 4$^{\text{th}}$ October 2011. On these days, the normal daily cycle of the bridge was perturbed as the bridge was kept in a stressed state for several hours, prompting a sharp rise in the value of the tilt sensor readings at the onset of the event and a sharp drop in the value of the tilt sensor readings at the end of the event. This characteristic shape can be seen for all eight fatigue tests that were performed. Again,

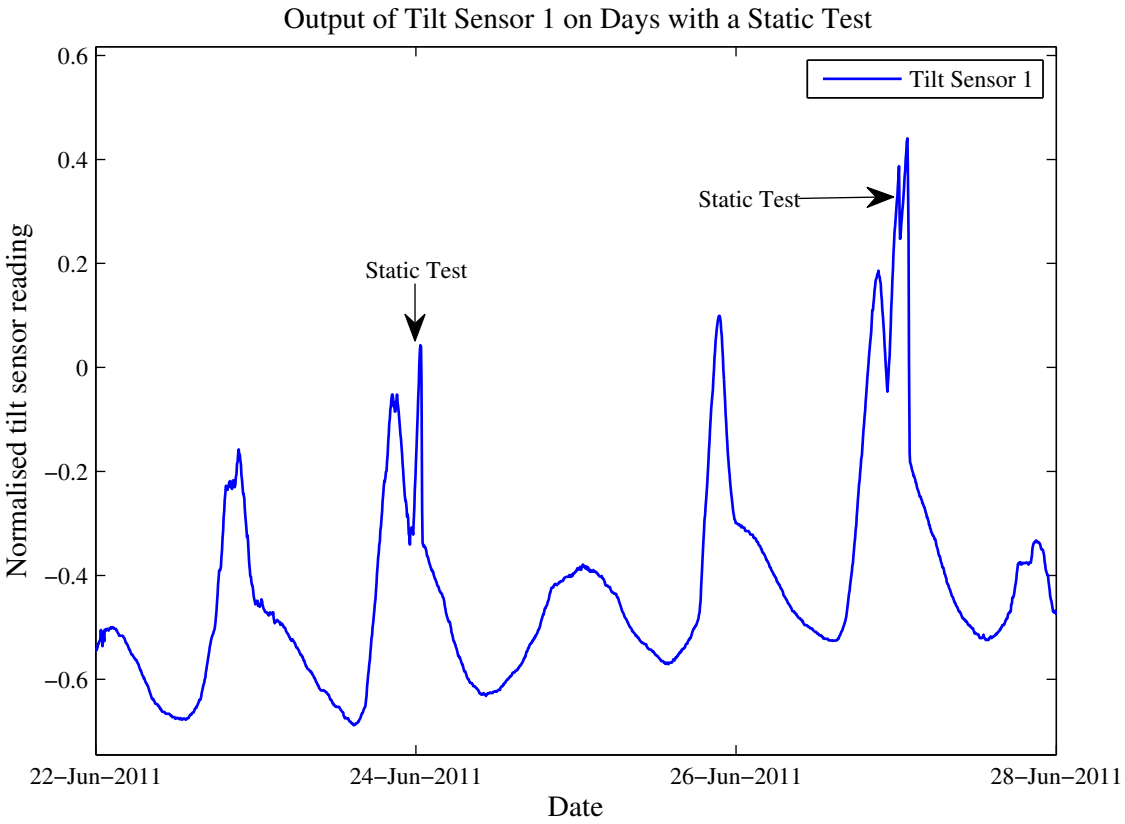


Figure 4.16: The output of tilt sensor 1 between 22nd and 28th June 2011. It can be seen that when a static test was performed, the output of the sensor was perturbed, resulting in a large spike.

this was a basis for expecting the ESN$_b$ technique to be able to characterise this type of intervention.

4.5.1.1 Methodology

Since the two primary types of intervention both had recognisable signatures in the data, an experiment was set up using ESNs to determine automatically the type of intervention using the data from all eight tilt sensors as inputs. The temperature sensors were not used, since they reported only on ambient conditions and did not contain any information about the bridge itself. The strain gauges were also omitted, partly because there was no recognisable trace of any of the events seen in them, and partly because the strain gauge data was only available for a limited time period that excluded most of the interventions.

A new ESN was set up with three output units that would classify each data point

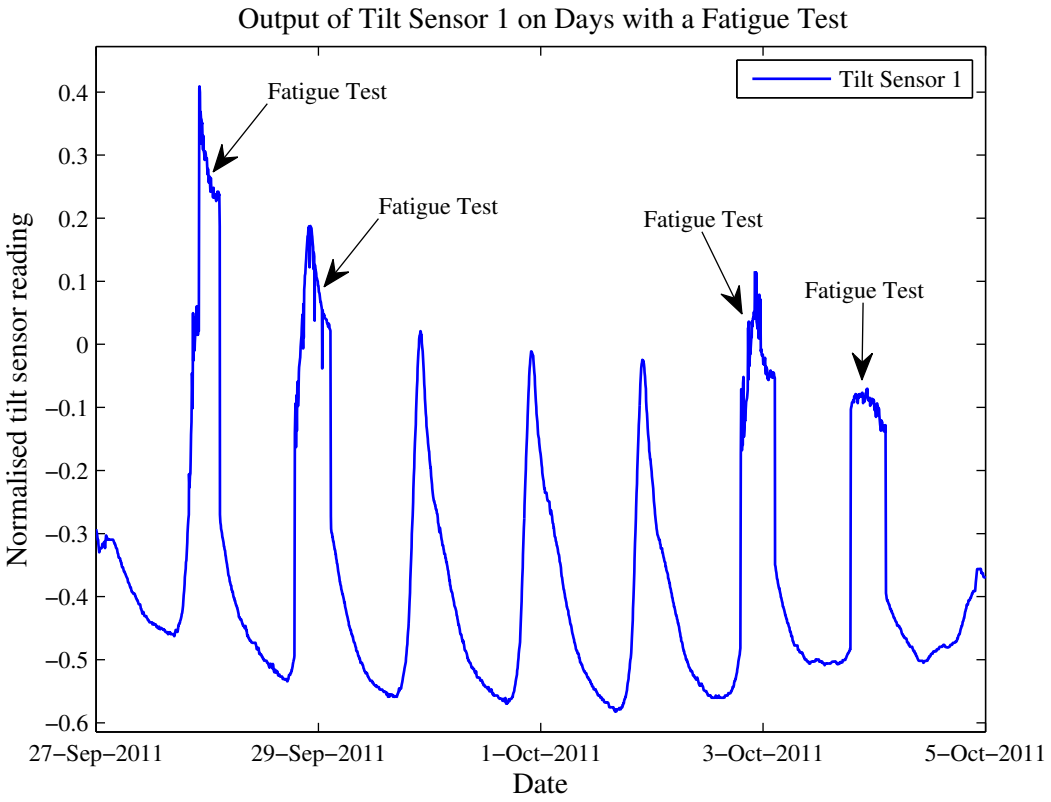


Figure 4.17: The output of tilt sensor 1 between 27$^{\text{th}}$ September and 5$^{\text{th}}$ October 2011. It can be seen that when a fatigue test was performed, the output of the sensor was perturbed, with a sharp rise in the tilt at the onset of the test, and a sharp drop at the end of the test.

as either part of the bridge's normal cycle, a static test or a fatigue test. A training dataset was created using the full data for the 22 days when static tests were performed, the eight days when fatigue tests were performed and 32 randomly selected days when no intervention occurred. It should be noted that there were two days that contained both a static test and a fatigue test, and so there was a total of 60 different samples in the training set. Each of these days was used as an individual training sample, and the networks were trained using 60-fold cross validation. During testing, ESN_b was presented with the entire three year dataset in a continuous stream, and the value produced by the three output nodes was recorded for each datapoint. An ensemble classifier approach was used, such that 100 ESNs were trained and used to obtain an average output for each datapoint during testing. An optimal threshold was then found for each output node, with values above the threshold being classed as a positive, and values below the threshold being classed as negative. For the purposes of analysis, the

outputs were then divided into individual days. If an output node gave a positive value at any time during a given day, then this was classed as a positive prediction. A window of 24 hours was used before and after the known interventions, so as to counteract any delayed classifier response and the possibility of a signature extended over a significant portion of a whole day's data points. Although training the ESNs to recognise all of the available static and fatigue tests is unconventional, the approach has value in demonstrating that the ESNs are still able to detect static and fatigue test signatures when embedded in a longitudinal dataset where the majority of the data pertain to ordinary days, without misclassifying these ordinary days as days of interest.

A standard grid search was employed in order to find the best topology for ESN_b, and this is given in Table 4.8. For most of the applications in this thesis, the large size of the training datasets meant that using a reservoir size greater than 500 neurons was computationally prohibitive. However, ESN_b drew on a much smaller set of training data than the other ESNs, and it was found that a reservoir with 800 neurons gave good performance without too much computational trade-off.

Parameter	**Range varied over**	**Optimal value**
Spectral Radius	0 - 2	1
Input Scaling	0 - 2	0.6
Leak Rate	0 - 1	1
Adaptation Epochs	0 - 10	0
Reservoir Size	1 - 800	800
Reservoir Connectivity Factor	0.1 - 1	0.5
Activation Function	Tanh, Lorentzian, Triangular Basis Function, Radial Basis Function, Fermi	Tanh

Table 4.8: The optimal parameters for ESN_b.

Two separate measures were used to analyse the performance of ESN_b. Firstly, as in Chapter 3, a graph of sensitivity against FPR was again used to calculate the AUC for each output node. Additionally, the Matthews Correlation Coefficient (MCC) [451] was calculated for each output node. The MCC is particularly appropriate here, since it is a good measure of the performance of a classifier that has been applied to an unbalanced

dataset, where there are significantly more instances of one class than another i.e. 32 different potentially damaging events amongst 1120 ordinary days of data. The MCC was calculated using Equation 4.2.

$$MCC = \frac{(TP \times FN) - (FP \times FN)}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \tag{4.2}$$

In Equation 4.2, TP is the number of true positives, FN is the number of false negatives, FP is the number of false positives and TN is the number of true negatives. The output value lies in the range -1 to +1, with +1 indicating perfect classification, 0 suggesting random guesswork and -1 for completely incorrect classification.

4.5.1.2 Results and Discussion

Confusion matrices for the three output classification nodes are given in Tables 4.9 - 4.11. It should be noted that the output of the static test node was not classified on the seven days when only a fatigue test took place, and vice versa. This was done because the abnormal tilt sensor readings caused by one type of event could prompt a positive output on both the static and fatigue test nodes. This is not as problematic as the classifier giving a positive output on an ordinary day, since it may indicate that the bridge is not behaving normally, even if it misidentifies the cause of this anomaly. The values for the AUC and MCC for each node, meanwhile, are given in Table 4.12.

		ESN_b Classification	
		Static	**Normal**
Ground Truth	**Static**	21	1
	Normal	11	1079

Table 4.9: A confusion matrix for the ESN_b static test node.

		ESN_b Classification	
		Fatigue	**Normal**
Ground Truth	**Fatigue**	8	0
	Normal	0	1091

Table 4.10: A confusion matrix for the ESN_b fatigue test node.

		$\mathbf{ESN}_b$ **Classification**	
		Normal	**Event**
Ground Truth	**Normal**	1063	28
	Event	1	28

Table 4.11: A confusion matrix for the ESN_b normal day node.

Measure	**Static Test Node**	**Fatigue Test Node**	**Normal Day Node**
AUC	0.9930	1.0000	0.9930
MCC	0.7867	1.0000	0.6849

Table 4.12: The values of AUC and MCC for each ESN_b classification node

The data in Tables 4.9 - 4.12 show that the ESN_b classification regime was a success, with the fatigue test node providing perfect performance. The static test classification node managed to detect all but one of the static tests, while the normal behaviour classification node misclassified only one damage event as normal behaviour. The high AUC values for all three nodes highlight how successful ESN_b was, while MCC values of 0.7867 and 0.6849 for the static test and normal day nodes, respectively, represent good classification performance.

There are further details that come to light when looking at the days when the static test classification node incorrectly gave a positive response, which are shown in Table 4.13. Of the 11 days seen in Table 4.13, the one that stands out as being particularly anomalous is the false positive on the 15$^{\text{th}}$ March 2010. However, this can be accounted for, as the sensors were switched off between 10$^{\text{th}}$ March 2010 and 15$^{\text{th}}$ March 2010. It is probably the case that the sudden discontinuity in the data led ESN_b to believe that a significant event had occurred, resulting in the false positive. It is also very interesting to note that 5 of the 11 days came within 31 days of events that were determined by ESN_a to have caused permanent damage to the bridge. It is possible that when the bridge underwent significant structural changes due to permanent damage, a series of 'aftershocks' were recorded in the data that were then automatically detected by ESN_b. Consequently, it is possible that a single positive reading would indicate an external factor causing the bridge to behave atypically, but a succession of positive ESN_b readings over the course of a month or so suggests that the initial atypical behaviour has resulted in lasting damage to the bridge. Three false positives occurred 45, 62 and

False Positive Date	Nearest Event Date	Difference (days)
8th May 2009	24th March 2009	45
25th May 2009	24th March 2009	62
24th June 2009	24th March 2009	92
13th August 2009	3rd August 2009	10
16th August 2009	3rd August 2009	13
19th August 2009	3rd August 2009	16
23rd August 2009	3rd August 2009	20
27th August 2009	3rd August 2009	24
15th March 2010	3rd August 2009	224
19th July 2010	2nd July 2010	17
9th August 2010	2nd July 2010	38

Table 4.13: Days when the ESN_b static test classification node produced a false positive, along with the closest static test prior to the date and the difference between the two in days.

92 days after the nearest event, the less significant static test of 24th March 2009, but the duration suggests that they may not be linked to the event itself. Since these false positives were found within the first six months of the monitoring campaign, they could possibly be further evidence of differential settlement in the bridge's foundations.

The results for ESN_b demonstrate the suitability of ESNs for attempting to classify 'real-world' time-series data. It was shown that it is possible to train an ESN on certain patterns and then to have it recognise them in a larger test dataset with very few false positives.

This work suggests that to obtain a fuller picture of the state of a structure, the optimal approach would be to use a classifying ESN such as ESN_b to detect and characterise significant deviations from normal behaviour, but to also task an ESN similar to ESN_a with predicting the typical behaviour of the tilt sensors so that the real and predicted values could be compared for the observation of long-term trends and changes, and the assessment of damage. By doing this, both abnormal events and long-term permanent damage could be detected.

4.5.2 ESN_c: Quantification of Damage

Finally, the third topology, ESN_c, was created with the intention of trying to quantify the overall health of the bridge based on the residual generated by ESN_a. This was

done in order to address the fact that, while ESN_a and ESN_b could be used to detect and characterise long-term damage, a certain level of expert analysis was still required in order to interpret the results. The motivation behind ESN_c was to produce an ESN that gave as its output a single number that would indicate the condition of the bridge at any time, based on current and past sensor readings, with no further interpretation required. This would provide an 'at-a-glance' figure for the current state of the bridge, which could be used as a first point of reference before using the output of ESN_a and ESN_b to examine the bridge in more detail. It could also allow for a long-term view of the bridge's general health to be taken, as changes to the output over time could be observed. The aim of this was that the findings from ESN_c would corroborate the interpretation given by ESN_a. The motivation for this particular approach was in order to meet Worden and Dulieu-Barton's fourth level of damage detection: assessment, where the extent of the damage is estimated.

4.5.2.1 Methodology

The earlier use of ESN_a indicated two main periods when there was a significant divergence between the data recorded by the tilt sensors and the normal behaviour of the bridge predicted by ESN_a. This was something that could also be observed in the output of ESN_{tt}. Since this divergence was taken to be indicative of damage, it was decided that the residual generated by ESN_{tt} would be used to train ESN_c. ESN_{tt} was initially chosen over ESN_a on the basis that that the data available for ESN_{tt} covered a longer time period. Furthermore, it should still be possible to present the network with the residual generated by ESN_a even after being trained on data obtained from ESN_{tt}. It should be noted that the residual used was not subject to the smoothing regime given in Equation 4.1. Data from the period prior to the first significant intervention, when there was a close correlation between the output of ESN_a and the real tilt sensor values, were used to train ESN_c to recognise normal behaviour, while data taken from the first of the periods of significant deviation between the output of ESN_a and the output of the tilt sensors were used to train ESN_c to recognise when the bridge was in

a damaged state. It was intended that the output of ESN_c would clearly identify both of the periods when the bridge was damaged, and the extent of this damage.

ESN_c was set up with eight input units, one for each residual, and one output unit. During training, the ESN was presented with 30 samples of tilt sensor data from the period 25th March 2009 to 27th June 2009 and 30 samples of tilt sensor data from the period 27th September 2009 to 27th February 2010 in batches of one full day (00:00 to 23:55) at a time. The network was trained to give an output of –1 when the bridge was undamaged and +1 when the bridge was damaged. The full dataset was then presented to ESN_c in a continuous stream, as was the case with ESN_a and ESN_b. As before, the data were normalised between -1 and +1, with no other pre-processing performed, and training was performed using ridge regression and 30-fold cross validation. A grid search was performed in order to find the optimal ESN_c topology, the results of which are given in Table 4.14.

Parameter	**Range varied over**	**Optimal value**
Spectral Radius	0 - 2	0.3
Input Scaling	0 - 2	0.5
Leak Rate	0 - 1	0.8
Adaptation Epochs	0 - 10	1
Reservoir Size	1 - 500	200
Reservoir Connectivity Factor	0.1 - 1	0.3
Activation Function	Tanh, Lorentzian, Triangular Basis Function, Radial Basis Function, Fermi	Tanh

Table 4.14: ESN parameters, the range that they were varied over in order to find the optimal ESN configuration and the optimal value found for each parameter.

100 ESNs using the best configuration were trained, and the best performing one was used for the production of the graphs seen in Section 4.5.2.2.

4.5.2.2 Results and Discussion

The output of ESN_c largely confirmed the conclusions reached through the use of ESN_a. Figure 4.18 shows how this output varied over time. As expected, the ESN

gave an output value of approximately -1 throughout the period prior to the first significant event. Following the first significant event, the level of damage gradually increased, before plateauing at an approximate value of +1 in November 2009. Since the network was trained on samples from this time period, this was expected. What is more interesting is the fact that following the interventions of July 2010, the output increased again, plateauing at a value of approximately +2. In keeping with previous conclusions about the condition of the bridge, this suggests that the level of damage had significantly increased due to the interventions. Interestingly, the interventions of 2011 initially prompted an increase in the ESN's output, but the final value at the end of the monitoring period was approximately +2.4, a small overall increase in January 2011.

As a separate test, the pre-trained ESN_c models were presented with the residual generated by ESN_a in order to see how the output would vary if residual derived from a data fusion approach had been used as its input. Figure 4.19 shows a direct comparison between the output of ESN_c when using the residuals generated by ESN_{tt} and ESN_a. It can be seen that very little changed as a result. Indeed, the two outputs have a PCC of 0.9896, indicating that they are extremely similar. Since the chief damage to the structure, in the region around tilt sensors 2, 3 and 7, was apparent in both ESN_{tt} and ESN_a, this result was expected.

Pleasingly, the output of ESN_c is relatively simple to interpret compared to ESN_a. There is only one value to monitor, and a baseline of approximately -1 indicates that there is no damage. A non-zero gradient in the output's baseline suggests that the bridge is, at that moment, undergoing a structural change, while consistent values that are greater than zero indicate damage. It is easy to look at the output of ESN_c at any given moment following the first significant damage event and to determine that the bridge had been damaged. In this way, ESN_c was successful in meeting the goal of having an output that would instantly allow a user to know what condition the bridge was in.

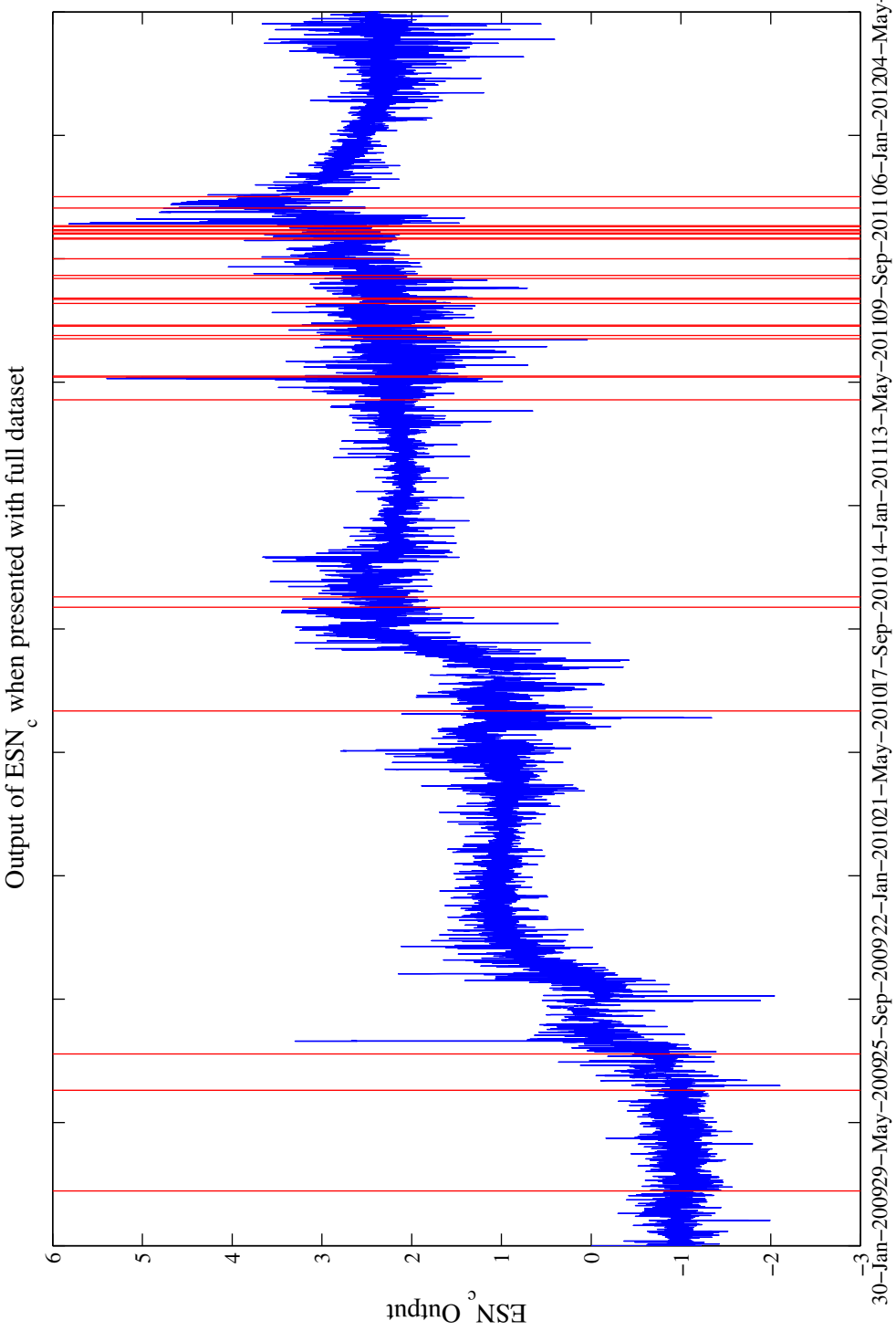


Figure 4.18: The output of ESNc when continuously presented with the residual produced by ESN_{tt} for the full extent of the dataset. The red vertical lines indicate the onset of interventions.

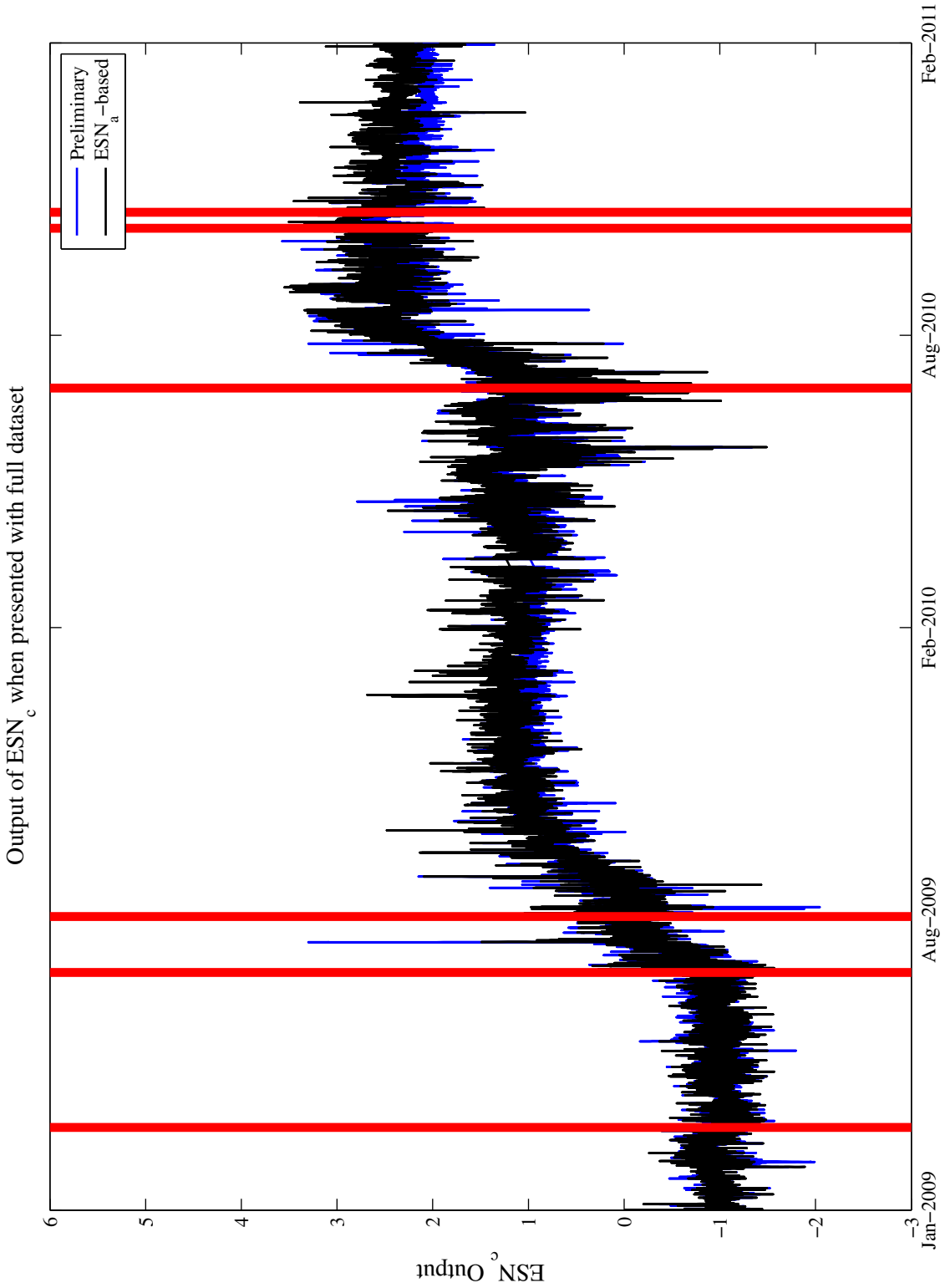


Figure 4.19: The output of ESNc when continuously presented with the residual produced by ESN$_a$ and ESN$_{tt}$ for the full extent of the strain gauge dataset. The red vertical lines indicate the onset of interventions.

4.5.3 Conclusion

This section introduced two supplementary ESN approaches to aid in the interpretation of the data fusion-based ESN_a. The first of these was ESN_b, an ESN designed to attempt to characterise the different types of potentially damaging interventions that were performed. Having been trained to recognise the characteristic signatures of static tests and fatigue tests, ESN_b proved capable of detecting and identifying these interventions when presented continuously with the entire dataset. In addition to this, it was found that a particularly significant intervention may be followed by 'aftershocks' in the dataset, which ESN_b classified as further static tests. The presence of three false positives prior to the first significant event also gave some evidence for the idea of differential settlement in the bridge's foundations.

The picture presented by ESN_c does not provide the same level of detail as ESN_a, which allowed for damage to be located and monitored over a long-term period, but ESN_c does give a single, at-a-glance figure that offers an insight into the condition of the bridge at any given time. The broad findings of ESN_c matched the interpretation derived from ESN_a, namely that the bridge was significantly damaged in June 2009, and that this damage was exacerbated by subsequent static tests in June and July 2010. By the end of the monitoring period, the impact of the static tests had been strongly felt by the bridge, leaving it in a permanently changed state. The same conclusion was reached regardless of whether the residuals generated by ESN_a or ESN_{tt} were used as inputs to ESN_c.

When used together, this suite of ESN approaches meets the first four of Worden and Dulieu-Barton's five levels of damage detection in SHM. The final parts of this chapter now consider how they might be applied in 'real-world' scenarios.

4.6 Proposed fusion-based online structural health monitoring system

The final contribution of this chapter is to combine the four separate ESN approaches (ESN_{asg}, ESN_a, ESN_b annd ESN_c) into a usable monitoring system. The following section suggests how the three ESNs could be used in a potential 'real-world', online SHM system, and how this could potentially be developed into software. Importantly, all of the ESNs used in the proposed system would already have been trained - this means that the outputs and plots displayed by the system are produced at almost the same speed that the real-time input data are fed into them.

Figure 4.20 shows how the ESNs could potentially be built into a single monitoring system that would place all of the necessary information at an engineer's fingertips. The residuals generated by ESN_{asg} are given on in the top left hand corner, while residuals generated by ESN_a are shown on the right. This is so that upon seeing an anomaly in the ESN_a residuals, an engineer could look at whether the anomaly was due to errant sensor behaviour such as that exhibited by SG2, or a genuine fault in the bridge. Should the anomaly prove to be a result of sensor fault, the tick boxes could be used to switch to a data fusion model that excludes particular strain gauges. The current state of the bridge, as determined by ESN_c, is shown in the bottom left corner, while the current state of each ESN_b output node is shown in the bottom right corner. Each set of plots has an adjustable timescale, such that an engineer could focus on the change to the bridge over the current day, or zoom out to look at changes to the bridge over a longer period, fully using the ability of ESN_a to show long-term damage trends. Each plot could be updated in real-time, allowing for a truly online system. By considering the collective output of each of the different ESNs, an engineer could then make an informed judgement on the state of the bridge at any time and act accordingly. For example, if a large residual was suddenly exhibited for one tilt sensor which was not due to a strain gauge fault, the static test node from ESN_b gave a positive output and the output of ESN_c increased, the region around the tilt sensor could then be physically

assessed for damage. Even if there did not appear to be any initial effects, the situation could still be monitored over a longer period with ESN_a and ESN_c.

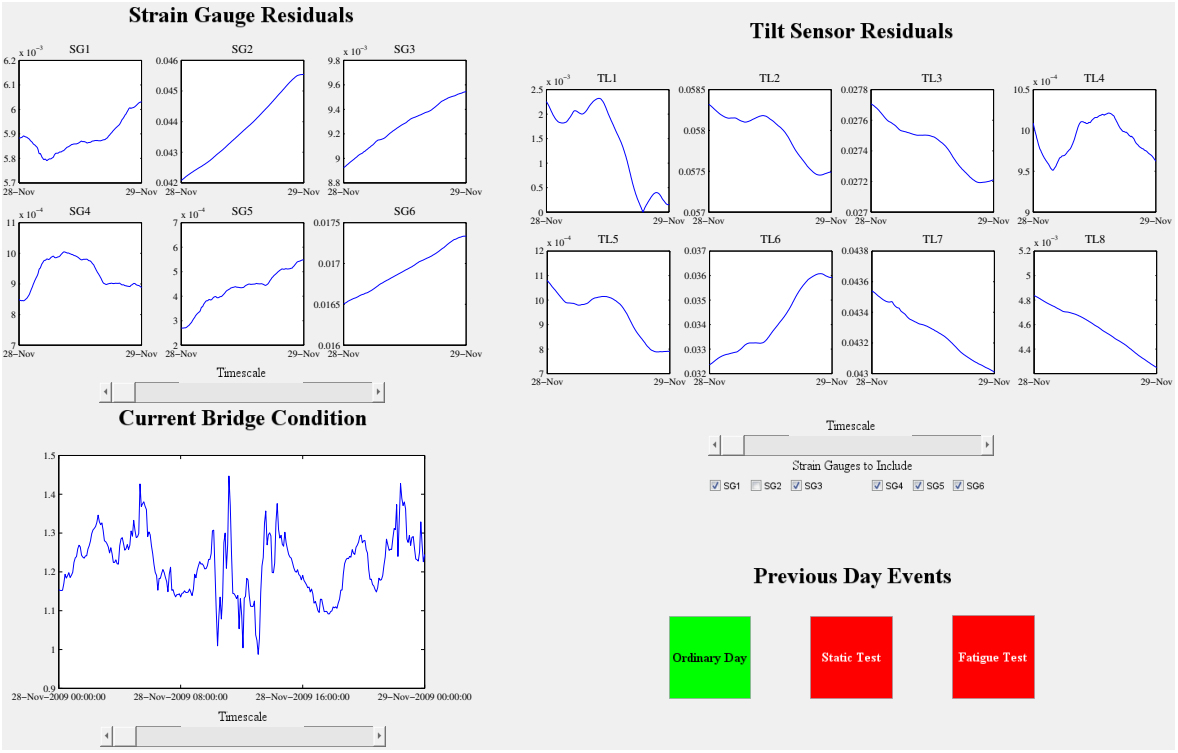


Figure 4.20: An example of how the ESNs could be used simultaneously in a monitoring system.

Although this proposed approach has not actually been developed into software, it does indicate how the work in this chapter could potentially be utilised as part of an online SHM system, and, therefore, how sensor interpretation could be improved using CI techniques for heterogeneous data fusion. A standard procedure for determining whether or not the bridge had been damaged could also be created. For example, this might take the following form if there was a major increase in the residual for one tilt sensor:

1. Observe increase in residual for tilt sensor(s) and go to stage 2.

2. Check strain gauge residuals. If one or more strain gauges are exhibiting errors in keeping with sensor faults, switch to a different ESN model and return to stage 1. Otherwise, go to stage 3.

3. Observe the output of ESN_b over the following 24 hours. If a particular type of

event is detected, dispatch engineers to inspect the bridge in the region around the affected tilt sensor. Otherwise, go to stage 4.

4. Monitor the changes in the tilt sensor residuals and the output of ESN_c over the following months. Should the output of either significantly increase at any point, dispatch engineers to inspect the bridge in the region around the affected tilt sensor. Otherwise, go to stage 5.

5. Continue to monitor the affected residuals and ESN_c output over an extended period of months. If no significant change is seen after six months, return to monitoring the bridge as usual.

Other procedures could be created for other scenarios, such as ESNb detecting a particular type of damage without an immediate response from other sensors.

4.7 Discussion

4.7.1 Levels of damage detection in SHM

At the outset of this chapter, the five different levels of damage detection in SHM given by Worden and Dulieu-Barton [425] were reviewed, and the initial aim was to use an ESN data fusion approach to reach the fourth of these levels. ESN_a was not only shown to have attained the first two levels - detection and localisation - but also allowed for damage trends to be observed over time, meaning that it was possible to consider the long term implications of potentially damaging interventions across different areas of the bridge. This, in turn, means that even if the impact of the damaging event did not appear to have an immediate effect on the structure, subsequent developments could be traced.

The third level of damage detection, classification, was met by ESN_b. Although ESN_b did not utilise data fusion, it was designed to complement the data fusion-based ESN_a. In this particular case, it was able to distinguish between two different types of potentially damaging interventions, along with ordinary days in the dataset. While a

'real-world' structure may not be expected to be routinely subject to deliberate static loading events, the results for an RNN such as ESN_b showed that even when presented with a full, longitudinal dataset in sequence, it was possible to detect any type of event that left a particular temporal signature in the sensor data. Should any other type of damage event leave its characteristic time-series signal in the data recorded by the sensor technology employed on a structure, then the ESN should be able to detect it amongst a continuous stream of data.

ESN_c, an approach based on the residuals generated by ESN_a, met the fourth level of damage detection: assessment. By giving a single number to describe the condition of the bridge at any time, ESN_c was able to estimate the extent of any damage that had occurred. As with ESN_a, ESN_c was not limited to simply giving an instantaneous estimate of damage, but would also allow for an engineer to see how the damage changed and developed over time. By looking at the output for the full dataset that was used here, it was possible to see how the damage to the bridge gradually grew, and worsened due to the impact of subsequent events. Combining this approach with ESN_a and ESN_b in one single interface, such as the one given in Figure 4.20, would allow for each of these four levels of damage detection to be considered in real-time as part of an online SHM approach.

4.7.2 The advantages of data fusion in this case study

The use of data fusion in this chapter also met the four potential benefits of data fusion described by Bellot et al. [267]. Firstly, the representation of the dataset was improved, since the output of ESN_a allowed for a greater level abstraction, with the raw sensor data used to produce a clearer output. Perhaps the best way to demonstrate the way in which the data fusion met the next two possible benefits - certainty and accuracy - is by examining difference in the PCC during the period that was used for training data. The 100 trained ESN_{tt} models, which considered only how the tilt sensors behaved based on the temperature sensors, had an average PCC of 0.9181 across all eight tilt sensors, with a standard deviation of 0.0053. In contrast, the 100 trained ESN_a models

gave an average PCC of 0.9419, with a standard deviation of 0.0022. This improvement in terms of both PCC and standard deviation represents an increase in both certainty and accuracy, and allowed for greater confidence in the ESN predictions for the unseen data. Finally, the data fusion approach gave a more complete picture of the bridge's behaviour, as evidenced by the way in which the first four levels of damage detection were met.

One of the more impressive aspects of the suites of ESNs used here was that a relative paucity of training data was used for each ESN. ESN_a was tasked with predicting the ideal tilt sensor behaviour over a two-year period using only six months of training data, due to the timing of the NPL's first significant intervention after the first six months of monitoring. Despite this, it proved capable of modelling the tilt sensors accurately for both the training and unseen data. ESN_b was tasked with detecting 22 static tests and 8 fatigue tests from over 1100 days of recorded data presented in sequence. Despite having only 36 positive samples in total, it still unambiguously detected all eight fatigue tests, and was able to detect 21 of the 22 static tests with only 11 false positives. ESN_c was only able to use a total of 60 randomly selected days out of the entire dataset. Given that some of the past research into the NPL footbridge synthesised data in order to get around the lack of training data, this is a good result. Furthermore, each ESN was able to do this without having to filter seasonal trends from the data, instead fully embracing these trends and building them into each model. In each case, it is probable that the training regime employed is responsible for this. By presenting the data in groups of one day during training, rather than continuously presenting the entire training period, the ESNs could learn the diurnal variations of the bridge in a range of different scenarios. It did not need to learn how the behaviour changed across different seasons, because it instead learnt how the bridge should be affected by temperature on any given day.

4.8 Conclusion

When the work on ESN_a began, the four aims of its data fusion approach were given:

1. To see if there was initially any benefit to using all six strain gauges, and if the ESN_{a16} model was adversely affected by the development of the major faults in the sensor.

2. To see if, upon 'detecting' a sensor fault, it is possible to switch to a pre-trained model that excludes the faulty sensor (ESN_{a15} in this case) and still obtain useful results.

3. To answer Research Question 1 by improving on the sensor interpretation in the preliminary work.

4. To answer Research Question 2 by considering how the ESN_a approach could be applied in a 'real world' online SHM scenario.

Each one of these aims has been shown to have been substantially met. By using six strain gauges as additional network inputs, the PCC when continuously presented with data prior to the first significant event increased from 0.9181 for the preliminary ESN_{tt} to 0.9438 for ESN_a. By considering multiple sensors, it was also possible to perform sensor validation, finding that one sensor had begun to show spike and bias faults. It was then found that it was still possible to obtain improved performance even when the faulty sensor was removed from the model, with an average PCC of 0.9419 for ESN_{a15}. This improved confidence in the residual generated for the unseen data made it possible not just to use the new results to detect, locate and monitor damage, but to also piece together a fuller picture of what had happened to the bridge, and why. This is something that could not have been done using just ESN_{tt}, and which had not been done previously in the literature. Supplementary ESN approaches then enhanced the interpretation of the sensor data, and a proposed combination of these ESNs for real-time decisions about the state of the bridge was introduced.

In response to Research Question 1, the use of heterogeneous data fusion in ESN_a makes a novel contribution to the field of SHM. Few of the other techniques applied to the NPL footbridge were able to allow for the same level of insight into long-term changes in state of the footbridge, with damage localisation achieved using just six

months of training data. In fact, data fusion has not been used to this effect in the other SHM case studies reviewed in Chapter 1. Indeed, there has been no approach before now that has brought together environmental and structural sensors to model a third type of structural sensor. In this respect, an ESN_a style approach does not just improve upon the sensor interpretation derived from ESN_{tt}, but also upon the sensor interpretation for the SHM of bridges, generally.

By considering multiple sensors both in combination and separately, it was possible to create a sensor validation routine that could distinguish between sensor fault and genuine damage, and to then compensate for sensor faults by switching to a different pre-trained ESN model. As mentioned in Chapter 1, this has not been reported in the literature to date, and is another novel contribution of this chapter.

In terms of Research Question 2, the fact that the data fusion approach worked so well on a 'real-world' structure, rather than on a simulation of a structure, suggests that ESN-based heterogeneous data fusion is a plausible option for 'real-world' online SHM. Although the different interventions were performed manually, the behaviour of the bridge in the meantime was organic and based on the real, naturally varying ambient conditions. Furthermore, the fact that the results were obtained without needing to perform any significant filtering or pre-processing suggests that the methodology might be well suited to online applications. The final section of this chapter presented an example of how an engineer might have all of the information from the different approaches presented in a more easily digestible format that would allow for informed decisions to be made in 'real-time'.

The 'real-world' applicability was also enhanced by the way in which the ESN approaches would reduce the level of expert interpretation required. The ESNs perform much of the data interpretation for an engineer, describing exactly what sort of intervention has occurred, what the overall state of the bridge is (expressed by a single metric), where the damage occurred and how it changed over time. While it is still necessary to make an informed judgement, one need not be an expert sensor analyst to do so. Furthermore, it was possible to create a suggested general systematic approach

to applying the ESN techniques to a structure, again reducing the variability in expert interpretation. In this sense, this chapter was able to meet Research Question 3.

In summary, this chapter introduced a data fusion-based approach to online SHM that offered new insights into the condition of the NPL footbridge and, when combined with supplementary ESN techniques, was able to meet the first four of Bellot's five levels of damage detection in SHM [267], while also detecting and isolating faulty sensors. The results showed that the data fusion approach is suitable for use in 'real-world' scenarios, and the next step might be to use this methodology on a 'real-world' structure that is in regular use, and perhaps has a wider range of available sensor data.

CHAPTER 5

Discussion: Comparison of Case Study Findings

This thesis presented two case studies that illustrated some of the benefits of data fusion in NDT and SHM. The following discussion chapter aims to bring these case studies together and consider, in more general terms, the benefits brought about by fusing heterogeneous sensor modalities using ESNs.

5.1 Summary of case studies

The two case studies looked at applying heterogeneous data fusion to two different aspects of assessing the health of structures, in particular bridges, in civil engineering: NDT and SHM. In each case, the key challenge was damage detection. For the first case study, this meant locating accurately areas of steel reinforcing meshes that had suffered total loss of section due to corrosion. The idea of 'damage' was more nuanced in the second case study, which looked beyond simply observing the occurrence of damaging events, to monitoring, locating, characterising and quantifying long term changes in the state of a bridge, while also considering sensor health. While they were combined by this common thread, the first, NDT-based, case study considered how data fusion could give better results in a one-off survey of a concrete surface, whereas the second case study looked at long term damage trends based on data gathered over three years.

Both case studies looked at improving the interpretation of one widely used set of sensors by incorporating other, widely used sensors in ways that had not previously

been considered. The first case study combined the commercially used EMAD technique with the sort of cover depth data that could be provided by a covermeter. In the second case study, the use of environmental temperature sensors to predict the behaviour of a physical tilt sensor and, hence, generate residuals for novelty detection was not new, but the incorporation of another type of physical sensor, the strain gauge, was. So, too, was the use of these sensors, along with a relatively new RNN technique, to reach the first four of Worden and Dulieu-Barton's five levels of damage detection in SHM [425], to distinguish between sensor fault and structural damage and to infer physical changes in the NPL footbridge.

One further challenge was to reduce the need for NDT or SHM expert interpretation. Whenever EMAD data is used, a threshold must be applied to the output of the data processing system in order to determine what does and does not constitute damage, a task which cannot easily be performed by a first-time user of the equipment. The data fusion approach provided a near-constant optimal threshold across several different cover depths, the standard deviation of which was three times smaller than that for the next most consistent technique. This would serve to almost entirely remove the requirement for an MFL specialist to assess the results before a threshold could be set. In the second case study, the ESN analysis did much of the work that an expert might ordinarily be required to do: it gave a single metric that expressed the current condition of the bridge, it pinpointed the type of damage that had occurred and the residuals generated in the data fusion technique clearly indicated where and how the damage developed over time.

The heterogeneous data fusion approach proved to be successful in both case studies. Fusing EMAD and covermeter data provided more accurate results with a more consistent threshold in the coverage range of $42.5\,mm < depth < 205\,mm$ when compared with models created using EMAD data alone, while still providing competitive performance outside of that range. The fusion ESN also gave the best performance on the noisier concrete test bed dataset. The fusion of temperature, tilt and strain data allowed for actual structural changes and damage types known to be present on

the bridge used in the second case study to be inferred, while the two supplementary ESNs (ESN_b and ESN_c) were capable of performing the tasks that they were given. Through both case studies, heterogeneous data fusion using RNN techniques such as ESNs was confirmed as a plausible option for assessing the condition of structures in civil engineering.

5.2 Advantages of data fusion

In order to look at the precise benefits of data fusion in each case study, it is, once again, helpful to revisit the four benefits of data fusion proposed by Bellot [267].

1. **Representation** The benefits of data fusion for data representation are best seen by considering the improvement in abstraction, that allowed for a richer understanding of the original sensor data. The EMAD, for example, only gives information on the presence of defects in the rebar, while a covermeter only gives the cover depth. The final output of the fusion ESN was a single value that incorporated both the MFL from the rebar and the cover depth, providing a far greater level of understanding than the individual sensors could manage in isolation. Similarly, the temperature sensors and strain gauges embedded on the NPL footbridge do not give any insight into the condition of the bridge, while the tilt sensors are only useful when the ideal behaviour of the sensors is known. The data fusion used a total of 24 different sensors to give eight different values at any given point in time, each one representative of the condition of the bridge at eight different locations. The data abstraction seen in this thesis makes the data fusion approach used here a good fit for extracting useful information from the increasing number of sensors and subsequent high volumes of data available in both SHM and NDT.

2. **Certainty** The fusion ESN that was used to combine the EMAD data with cover depth data proved to give better accuracy in terms of AUC than any non-fusion techniques, having had the highest AUC on more datasets than any of the other

approaches. In a test where the ground truth was not known, this would give a user more confidence in the accuracy of fusion-based results than in any of the other methods. Furthermore, the more consistent optimal threshold level would also allow for a greater level of certainty, as it is less likely that the threshold would be set sub-optimally. For the second case study, the fact that the PCC was greater for the fusion ESN than the non-fusion ESN when continuously presented with the training data increased confidence that the subsequent predicted tilt sensor behaviour would be accurate, which then increased confidence in the conclusions regarding when, where and how damage had developed.

3. **Accuracy** When presented with several different unseen test datasets, it was shown that the fusion ESN highlighted defect locations with greater clarity and less ambiguity when compared to several other techniques, seemingly better able to deal with potentially confounding noisy input signals. With regards to the NPL footbridge, the fusion ESN did not just improve the performance on the continuously presented training data in terms of PCC. The standard deviation associated with each PCC value for the 100 trained fusion ESNs was smaller than the standard deviations for the 100 trained non-fusion ESNs.

4. **Completeness** In both case studies, the results suggest that the chief benefit of data fusion was that it provided models that exploited valuable environmental context. It was shown that the fusion ESN was able to model a complex and nuanced relationship between the two components of the EMAD data and the cover depth, something that was explained in detail in Chapter 3. Without access to this vital information, the non-fusion techniques were not able to respond as well to variations in different datasets. The fusion ESN is, therefore, better equipped to give a more complete picture of the condition of a reinforcing mesh. Similarly, in the second case study, it was known that much of the variation in the behaviour of tilt sensors could be accounted for by considering the ambient temperature, but this could not cover all aspects of the bridge's behaviour. Integrating strain data ensured that the final tilt sensor model that was produced would not just

factor in the thermal response of the bridge, but the response to other factors reflected in the changing strain. In both the NDT and the SHM case study, data fusion gave a more complete picture of the scenario than any of the individual sensors could have done alone.

5.3 Summary

This chapter brought together the two data fusion case studies presented in the thesis, showing how on both occasions, data fusion gave advantages that could not be obtained using individual sensor modalities. The techniques used here gave improved sensor interpretation on two separate quasi-real-world scenarios, holding great promise for future 'real-world' applications.

CHAPTER 6

Conclusion

This final chapter briefly summarises the thesis as a whole, and explains what the chief contributions of this work have been. Practical applications of the thesis are suggested, and then recommendations for possible future work are given.

6.1 Thesis Summary

The first two chapters of this thesis set the scene for the work that was to be done. In Chapter 1, it was shown that, despite the fact that there is an increasing interest in evaluating reinforced concrete by fusing NDT techniques, none of the existing methods could detect specific rebar defects accurately. Furthermore, there had not previously been any attempt to combine the EMAD, or any other MFL approaches, with different sensor modalities in order to achieve this. In terms of the use of ESNs for the SHM of bridges, there were no attempts to use environmental and structural sensors in combination to model a different structural sensor, or to detect both sensor fault and damage in bridges simultaneously. Chapter 2 gave a brief background to data fusion and ESNs, the particular RNN technique used here.

Chapter 3 looked at fusing EMAD and covermeter data together in order to counteract the effect of cover depth on MFL defect signal amplitude. Preliminary work showed that using an ESN trained only on EMAD data gathered from the Keele concrete test bed could provide performance that was comparable to that of the AT. The data

fusion mesh was created, with defects manually inserted at different points and EMAD data were gathered at manually adjusted distances from the mesh, so as to simulate different cover depths. It was found that the fusion ESN outperformed all of the other techniques between cover depths of 85.0 mm and 205 mm, both in terms of accurate defect detection, clarity of output plots and consistency of threshold. An analysis of changes in the fusion ESN output revealed that while the covermeter data did act as a scaling factor, context about the cover depth allowed the ESN to access a more nuanced relationship between the X axis and Z axis components.

The second, and final, case study was presented in Chapter 4, and concerned the detection, localisation, characterisation and clarification of damage in online SHM applications. This used longitudinal data from ten temperature sensors, eight tilt sensors and six strain gauges embedded on the NPL footbridge. It was found that the use of data fusion made it possible to observe the long term effects of interventions in order to determine if any damage was permanent, before considering the location of the tilt sensors that most strongly suggested damage. By modelling the strain based on the temperature, it was possible to find that one strain gauge was faulty, and to then isolate it with minimal performance loss. Two further ESN architectures were successfully used for classifying types of damage and producing a single a metric that expressed the overall condition of the bridge at any given time. Finally, it was shown how the overall suite of ESNs might be incorporated into a single graphical user interface display for ease of interpretation, and a possible strategy for reviewing the sensor data in real-time was given.

The discussion chapter, Chapter 5, considered the two case studies together. Bringing together different modalities of data using ESNs gave the ESN important and useful contextual information to process that would not have been available using just one sensor modality. This gave improvements in terms of data representation, certainty in results, the accuracy of the results and the overall completeness of the picture provided.

6.2 Answers to Research Questions

In Section 1.6, three research questions were posed:

1. Can heterogeneous data fusion lead to improved sensor data interpretation using a relatively new RNN technique from the field of reservoir computing, ESNs?

2. Can any ESN heterogeneous data fusion improvements be demonstrated in quasi-real-world scenarios for:

   (a) NDT, and

   (b) SHM

   rather than relying on simulated data streams?

3. Can systematically applying the properties of ESNs for heterogeneous data fusion affect the extent to which expert interpretation might otherwise be required?

This section considers the ways in which the two case studies answered these two questions.

6.2.1 Research Question 1: improving sensor interpretation with ESN-based heterogeneous data fusion

In each case study, it was shown how heterogeneous data fusion was able to improve upon the existing level of interpretation of sensor data. The fusion ESN for EMAD and covermeter data was compared to the AT, a mathematical approach that is currently used for processing EMAD data. The fusion ESN comprehensively outperformed the AT in terms of AUC for all but one of the datasets, including one obtained from the concrete test bed that the AT was originally designed for. This is a clear, quantifiable improvement on the existing standard. Furthermore, the fusion ESN was also shown to be better than three alternative ESN-based CI attempts at finding defects using EMAD data alone, since the cover depth data contained important context that complemented the

EMAD readings. The heterogeneous data fusion method demonstrated a considerable improvement in automatic sensor interpretation.

While novelty detection has been performed in order to detect damage in a number of previous studies into the SHM of bridges (see Chapter 1), Cross has previously observed that this is insufficiently informative, and does not distinguish between sensor fault, extreme weather or actual damage [192]. The data fusion methodology used here moved this work beyond simple novelty detection, and allowed for a more complete picture of the bridge's behaviour and its present state to be constructed. The task changed from simply detecting when damage may have occurred to monitoring any developments in this damage, and locating where the damage had occurred. CI was further employed to develop this interpretation of the different sensors, allowing for certain types of damage to be determined, and the overall state of the bridge to be reflected in a single metric. By looking at the interdependent relationships between multiple sensor modalities, it was possible to broach new territory in sensor validation, successfully distinguishing between genuine sensor fault in strain gauges and genuine damage to the bridge. The fusion ESN was at the heart of a methodology that provided a more comprehensive picture of the bridge at any given point in the monitoring period.

6.2.2 Research Question 2: 'real-world' applicability in NDT and SHM

The two projects considered in the NDT case study and SHM case study could be described as quasi-real-world scenarios. In each case, the subject of the investigation was a real, physical structure that was assessed using real sensors, with the exception of the covermeter data, although this still reflected the physical reality of the cover depth. None of the data used at any point were generated as part of a simulation or model of a scenario. When defects or damage patterns were seen, these corresponded to actual, physical faults that had occurred. For the NDT case study, the structure was a steel reinforcing mesh that was kept in controlled conditions that made it prone to corrosion, while the SHM case study considered a concrete footbridge that had

undergone regular use for almost 50 years. Where the two differed from truly 'real-world' scenarios was in that they were taken out of regular use so that they could be deliberately damaged, thereby providing reliable ground truth data. This enabled the ability of the heterogeneous data fusion techniques to detect known damage, such that they could be used in 'real-world' scenarios where the extent of any damage was not known. However, the success of the data fusion approaches used here shows great potential for their application to truly 'real-world' structures.

The ease of applying a pre-trained ESN to the data in both case studies suggested that the technique would be particularly suitable for real-time analysis on NDT, potentially eliminating the need for additional costly closures, and online SHM. Aside from simply normalising the data between -1 and +1, no significant pre-processing or filtering was required, meaning that it would be straightforward to present the data to a fusion ESN. This, again, suggests great promise for practical application of the work in a truly 'real-world' setting.

6.2.3 Research Question 3: reducing expert interpretation with systematic application

The data fusion also offered improvements in terms of ease of interpretation, making the monitoring devices themselves potentially more useful in the 'real-world'. In both case studies, using data fusion reduced the level of expert interpretation required in order to get the most useful information out of the sensors. The consistent threshold level for the fusion ESN when processing EMAD and cover depth data totally removed the need for an MFL expert to analyse the data and set the optimal threshold. The fusion ESN made it easy to see long term trends in the tilt sensor data on the NPL footbridge, so that damage could be located and monitored over time, while it also led to a single output metric that expressed the overall condition of the bridge at that particular time.

In each case study, the data fusion approach allowed for a systematic approach to data interpretation that reduced the need for expert interpretation. In Chapter

3, the fusion ESN produced the most consistent optimal threshold, meaning that a pre-trained network could be applied to any set of EMAD and covermeter data and a set threshold applied in order to discern the presence of defects. In contrast, the less consistent EMAD-only techniques would still require expert knowledge of MFL in order to inspect the raw data and decide where the threshold should be set. An alternative option for using these techniques in the 'real-world' would be to take a sample of any anomalous signals from the EMAD and to visually inspect the rebar at these locations by removing the concrete cover. The threshold could then be set based on which signals corresponded to real defects. However, this would require destructive inspection and could potentially be impractical if, for example, it would necessitate road closure and repairs. Either of these options would cause a significant delay between conducting the survey and obtaining usable results. The systematic approach provided by the fusion ESN would prevent all of these problems, and the short time required to apply a pre-trained ESN to a dataset would mean that results could be provided almost as soon as the survey was completed, without requiring extensive expert interpretation.

Similarly, it was in Chapter 4 that a constant threshold of 0.02 was applied to the modulus of the residual generated for tilt sensors by ESN_a using temperature sensors and strain gauges in order to detect potential damage to the NPL footbridge. ESN_b, meanwhile, was shown to be able to highlight the onset of particular events. Both of these would reduce the level of expert interpretation required in order to initially detect potentially damaging events and then to determine if the bridge was subsequently damaged. Section 4.6 showed how these two ESNs, along with ESN_c could be applied systematically in order to detect, localise, classify and assess damage by presenting all of the relevant data in a potential, easy to use graphical user interface form and gave a procedure for using this data. While it was not possible to completely eliminate the need for some expert judgement in this process, the overall demand for this expertise was reduced.

6.3 Thesis Contribution

The work presented in this thesis contained a number of novel contributions that are of benefit to both science and industry. Firstly, while ESNs have previously been applied in both NDT and SHM, and have also been used in data fusion, this is the first time that they have been used to fuse together heterogeneous data in the context of SHM and NDT. It was shown that the recurrent ESNs' short-term memory made them adept at dealing with the sort of data often provided by structural inspections, which is usually either temporally or spatially continuous in nature, variable in signal intensity and in the presence of noise. It was shown in both Chapters 3 and 4 that this meant that ESNs were very good at detecting characteristic fault signatures from amongst noisy data, while Chapter 4 also showed that ESNs are very good at modelling the relationship between different sets of data, such that it is possible to predict the value read from a sensor at any given time. ESNs have been shown to be a good option for applications in the fusion of data in SHM and NDT.

Chapter 3 presented the first instance of heterogeneous data fusion being used for the accurate detection of defects in steel reinforcing meshes, and the first time that MFL data had been fused with any other data modality. The fusion ESN that was created outperformed both the AT and the alternative ESN training methodologies that were employed. The work offered a clear improvement on the data processing methods that are currently employed, and it was shown that this could potentially be packaged so that it could be used in 'real-world' surface inspections. Overall, the chapter gave a new method for using MFL to detect accurately defects in rebar, such that the effect of different cover depths on the EMAD data was accounted for.

The suite of ESNs used in Chapter 4 approached the SHM in bridges problem in a new way. While the data produced by the NPL footbridge has been studied in the past, no other approach was able to detect, locate, characterise and quantify the damage using such a short training period. It has not previously been possible to do all of this on a bridge while also observing sensor faults, and isolating faulty sensors.

The principle of combining different modalities of sensor, including environmental and physical sensors, to predict the behaviour of another physical sensor, was introduced and found to be more effective than simply modelling a physical sensor based on an environmental sensor. The chapter also recommended a new procedure for using the suite of ESNs to monitor a bridge, and how each ESN should be used to make inferences about a structure. While residuals have been used for novelty detection, the use of residuals to monitor long term damage trends in conjunction with a damage classifier and a bridge health quantifier is new.

A summary of these contributions is given in Table 6.1.

Contribution	**Chapter**	**Case Study**
ESNs for data fusion in NDT and SHM	3 and 4	NDT and SHM
Fusion of heterogeneous data for rebar defect location	3	NDT
Combination of MFL with covermeter	3	NDT
ESN 'suite' approach	4	SHM
Simultaneous detection, localisation, classification and assessment in SHM of bridges	4	SHM
Fusion of environmental and physical sensors to model other physical sensors	4	SHM

Table 6.1: The overall contributions of the thesis, and where these can be found

6.4 Practical Applications of Thesis Findings

In each case study, an attempt was made to envisage how the end result might be taken out into the 'real-world' and applied in a civil engineering setting. The intention here was to ensure that the work was not just a theoretical extension of data fusion, but something that would allow for a genuine practical impact.

The most obvious practical application of this work would be to use the case study presented in Chapter 3 to improve the quality of results provided in MFL (particularly EMAD) surveys of reinforced concrete surfaces. By simply including covermeter measurements in the data analysis, a considerable improvement on the current AT can be obtained. The improvement in accuracy and the increased ease-of-use owing to the

consistent threshold level would help to make the EMAD technique a more appealing option than some of the more commonly employed methods for this task, such as half-cell potential. This, in turn, would be of benefit to the overall civil infrastructure, as scans of reinforced concrete surfaces could find faults in the rebar before those faults caused a serious incident such as a bridge collapse.

The NPL footbridge case study from Chapter 4 could be similarly useful for the timely detection of damage in bridges. The NPL footbridge work provides a proof of concept of certain principles: how it is possible to detect and localise damage by fusing sensor modalities together, how this damage can then be classified and quantified and how physical and environmental sensors can be usefully brought together. While another bridge may be on a bigger physical scale than the footbridge, and may have different types of sensor, these principles can be applied in order to obtain key information on the condition of the bridge at any time. To assist in this, the chapter presented a suggested approach to using the suite of ESNs and a procedure for inspection of a monitored bridge. Future and related directions for the application of these principles are given in the future work section below.

6.5 Future Work

The most obvious direction is to now move away from quasi-real-world scenarios and use the methods presented here to inspect and monitor structures that are currently in use. One particularly interesting challenge would be to apply the suite of ESNs from Chapter 4 to a structure such as the Tamar suspension bridge, which has been the subject of other studies in the past (see Chapter 1). The Tamar suspension bridge is much larger than the NPL footbridge (643 metres in length compared to 20 metres in length), has a greater variety of sensors in situ and the fact that the bridge is still in use today would allow for the work in Chapter 4 to be extended. The principles developed could be used for the combination of wind speed anemometers, temperature sensors, strain gauges and traffic loading in order to determine deck vertical displacement and, potentially, detect significant changes in the bridge's state. Application of a suite of

ESNs to any such comprehensively monitored, regularly used bridge where the ground truth was unknown would further test the practical applicability of these methods.

Similarly, future work relating to Chapter 3 could involve the application of the heterogeneous data fusion approach to surveys of 'real-world' structures with naturally developed defects. Future EMAD surveys should use a covermeter to determine the cover depth associated with each reading, so that data fusion can be performed. It would be particularly helpful to survey structures where it is possible to break out the concrete afterwards in order to confirm the accuracy of the results.

While the work on NDT in this thesis considered only the EMAD technique and covermeters, there are several other options available for the NDT of reinforced concrete, as detailed in Chapter 1. Just as the further work on the SHM study considered different sensor modalities, it is also worth investigating the possibility of fusing alternative NDT methods. For example, the EMAD technique might complement existing approaches to providing an overall picture of the condition of concrete surfaces, such as VOTERS, OSSCAR/BetoScan and RABIT, presenting interesting data fusion possibilities. The aim of the work in Chapter 3 was to combine two sensor modalities to more accurately locate defects. For a more holistic investigation into a reinforced concrete surface, one worthwhile avenue of research would be in combining EMAD data with the likes of ultrasonic, GPR and thermography data to produce a single metric expressing the health of the surface at any given point, much like ESN_c in Chapter 4.

There is one more obstacle that stands in the way of a full practical application of the work from both case studies which should be addressed in any future work. The development of bespoke industrial software for the two case studies is beyond the scope of this thesis, which focused instead on the heterogeneous data fusion techniques themselves, but is something that would need to be addressed if the approaches in this thesis are to be widely used in industry. Both Chapters 3 and 4 proposed how this might be realised for each case study, and an attempt could be made to produce viable software that could be used on site in real-time.

It was found in Chapter 4 that ESNs could be taught to recognise and reproduce

diurnal variations in tilt sensor data, such that generating residuals then allowed for unusual tilt sensor behaviour to be observed. This meant that there was no necessity to detrend or deseasonalise the data prior to novelty detection. However, given the effectiveness of approaches such as cointegration [187, 188, 192, 441], one possible avenue of future research could be to see if the ESN data fusion approach to SHM could be improved by combining it with a deseasonalisation technique.

Finally, while the benefits of heterogeneous data fusion in NDT and SHM have been demonstrated, the work done in this thesis also confirmed that ESNs are a very appealing option for longitudinal signal processing tasks. Appendix B presents the results of a separate investigation into ESN properties when applied to a static pattern classification task using an input clamping method. This work drew some new conclusions, including the fact that it was possible to obtain good results without the need for reservoir or output unit stabilisation, while also showing that ESNs could be applied successfully beyond the build environment to plant species classification tasks using high-dimensional hyperspectral data. One interesting direction for future work would be to continue this research into the distinctive properties and behaviour of ESNs, and to perhaps even apply the techniques shown in this thesis to heterogeneous data fusion in other settings, such as the natural environment. Hyperspectral data could, for example, be fused with lidar data on canopy height, in order to more accurately discriminate between plant species and plant vitality in the 'real-world'.

A table summarising the proposed recommendations for future work is given in Table 6.2.

Recommendation 1	Application of the suite of ESNs from Chapter 4 to the online SHM of a large, more comprehensively monitored bridge that is undergoing constant use.
Recommendation 2	Performing of surveys using the EMAD and covermeter on reinforced concrete surfaces that have naturally corroded over several years, with break-outs used to assess the accuracy of the results
Recommendation 3	Use of the ESN data fusion approach to combine sensors that look at different aspects of a reinforced concrete surface, such as the EMAD, ultrasonics and GPR to produce a value that gives an overall picture of the surface's condition at any given point.
Recommendation 4	The development of bespoke software to allow for real-time data processing on site using the fusion ESNs.
Recommendation 5	A combination of ESN data fusion and deseasonalisation techniques when processing SHM data.
Recommendation 6	Continued research into the distinctive properties of ESNs.
Recommendation 7	Application of ESN-based heterogeneous data fusion in the natural environment.

Table 6.2: Recommendations for future work based on this thesis.

CHAPTER 7

References

[1] World Economic Forum. The global competitiveness report 2016-2017. Online, 2017. http://reports.weforum.org/global-competitiveness-index/.

[2] World Economic Forum. The global competitiveness report 2015-2016. Online, 2016. https://reports.weforum.org/global-competitiveness-report-2015-2016/.

[3] World Economic Forum. The global competitiveness report 2014-2015. Online, 2015. http://reports.weforum.org/global-competitiveness-report-2014-2015/.

[4] World Economic Forum. The global competitiveness report 2013-2014. Online, 2014. http://reports.weforum.org/the-global-competitiveness-report-2013-2014/.

[5] World Economic Forum. The global competitiveness report 2012-2013. Online, 2013. http://reports.weforum.org/global-competitiveness-report-2012-2013/.

[6] G. Osborne and D. Alexander. Investing in Britain's future. Policy Paper Cm 8669, HM Treasury and Infrastructure UK, 2013.

[7] C. Matthews and G. Dalton. Highways agency annual report and accounts 2014 to 2015. Corporate report HC112, Highways England and Highways Agency, 2015.

[8] American Society of Civil Engineers (ASCE). 2017 report card for America's infrastructure, 2017. http://www.infrastructurereportcard.org/wp-content/uploads/2017/01/Bridges-Final.pdf.

[9] C. Duncan, S. Landau, D. Cutler, B. Alstadt, and L. Petraglia. Integrating transportation and economic models to assess impact of infrastructure investment. *Transportation Research Record: Journal of the Transportation Research Board*, 2297:145–153, 2012.

[10] U.S. Department of Transportation Federal Highway Administration. Deficient bridges by state and highway system 2014. Online, 2015.

[11] J. Taber. Bolts, cables examined as officials investigate nipigon bridge collapse. *The Globe and Mail*, January 2016.

[12] Associated Engineering. Nipigon river bridge independent technical review. Report, Ontario Ministry of Transportation, September 2016.

[13] B. Kelman and A. Rumer. I-10 will reopen friday after bridge collapse. *The Desert Sun*, July 2015.

[14] B. Kelman, C. Atagi, J. Marx, and S. Roth. Collapsed I-10 bridge given an A rating just last year. *The Desert Sun*, July 2015.

[15] S. Samb. Bridge collapse in Guinea kills 65. *Reuters*, March 2007.

[16] S. Hao. I-35W bridge collapse. *Journal of Bridge Engineering*, 15(5):608–614, 2010.

[17] K. Rofidal. I-35W Collapse - Day 4 - Operations & Scene. Online - Wikimedia Commons [Public Domain], August 2007.

[18] P. A. Vanniamparambil, F. Khan, R. Carmi, S. Rajaram, E. Schwartz, M. Bolhassani, A. Hamid, A. Kontsos, and I. Bartoli. Multiple cross validated sensing system for damage monitoring in civil structural components. In *Structural Health Monitoring 2013: A Roadmap to Intelligent Structures: Proceedings of the Ninth International Workshop on Structural Health Monitoring, September 10–12, 2013*, 2013.

[19] F. E. White. Data fusion lexicon. Technical report, Data Fusion Panel, Joint Directors of Laboratories, Technical Panel for C3, 1991.

[20] H. Hu and J. Q. Gan. Sensors and data fusion algorithms in mobile robotics. Technical Report CSM-422, University of Essex, 2005.

[21] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.

[22] H. Jaeger. Short term memory in echo state networks. Technical report, Fraunhofer Institute for Autonomous Intelligent Systems, 2002.

[23] W. Maass, T. Natschlager, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, nov 2002.

[24] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. The unified reservoir computing concept and its digital hardware implementations. In *Proceedings of the 2006 EPFL LATSIS Symposium*, pages 139–140, 2006.

[25] L. Busing, B. Schrauwen, and R. Legenstein. Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons. *Neural Computation*, 22(5):1272–1311, May 2010.

[26] M. Lukosevicius and H. Jaeger. Reservoir computing approach to recurrent neural network training. *Computer Science Review*, 3(3):127 – 149, August 2009.

[27] D. C. Montgomery, E. A. Peck, and C. G. Vining. *Introduction to Linear Regression Analysis*. Wiley, 1982.

[28] X. Dutoit, B. Schrauwen, J. Van Campenhout, D. Stroobandt, H. Van Brussel, and M. Nuttin. Pruning and regularization in reservoir computing. *Neurocomputing*, 72:1534–1546, 2009.

[29] M. Lukosevicius. A practical guide to applying echo state networks. In G. Montavon, G. B. Orr, and K-R. Muller, editors, *Neural Networks: Tricks of the Trade*,

volume 7700 of *Lecture Notes in Computer Science*, chapter 27, pages 659–686. Springer Berlin Heidelberg, 2nd edition, 2012.

[30] B. Schrauwen, D. Verstraeten, and J. Van Campenhout. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th European Symposium on Artificial Neural Networks*, pages 471–482, 2007.

[31] D. E. Ellsworth and K. Ginnado. Guide for visual inspection of structural concrete building components. Technical report, US Army Corps of Engineers: Construction Engineering Research Laboratory, 1991.

[32] J. J. Moughty and J. R. Casas. A state of the art review of modal-based damage detection in bridges: Development, challenges, and solutions. *Applied Sciences*, 7(5), 2017.

[33] R. Arndt and F. Jalinoos. NDE for corrosion detection in reinforced concrete structures - a benchmark approach. In *NDTCE'09, Non-Destructive Testing in Civil Engineering*, 2009.

[34] B. M. Pailes. *Damage identification, progression, and condition rating of bridge decks using multi-modal non-destructive testing*. PhD thesis, Rutgers University, 2014.

[35] M. Raupach, K. Reichling, J. Broomfield, J. Gulikers, U. Schneck, M. Serdar, and I. Pepenar. Inspection strategies for reinforcement corrosion surveys. *Materials and Corrosion*, 64(2):111 – 115, 2013.

[36] Kyosti Tuutti. *Corrosion of steel in concrete*. PhD thesis, Lund University, 1982.

[37] J. P. Broomfield. *Corrosion of Steel in Concrete: UnderstaUnder, Investigation and Repair*. Taylor & Francis, 1998.

[38] A. Poursaee. *Corrosion of Steel in Concrete Structures*, chapter 2, pages 19–33. Woodhead Publishing, 2016.

[39] M. F. Montemor, A. M. P. Simoes, and M. G. S. Ferreira. Chloride-induced corrosion on reinforcing steel: from the fundamentals to the monitoring techniques. *Cement and Concrete Composites*, 25(4–5):491 – 502, 2003. Concrete Durability.

[40] P. Lambert. *Corrosion and Passivation of Steel in Concrete*. PhD thesis, University of Aston, 1983.

[41] A. Steffens, D. Dinkler, and H. Ahrens. Modeling carbonation for corrosion risk prediction of concrete structures. *Cement and Concrete Research*, 32:935 – 941, 2002.

[42] M. Hocking, L. North, A. Wright, and P. Haycock. Non-destructive detection of corrosion in steel reinforcement. *Concrete Engineering*, 10:39–41, 2006.

[43] J. B. Butcher, M. Lion, C. R. Day, P. W. Haycock, M. J. Hocking, and S. Bladon. A low frequency electromagnetic probe for detection of corrosion in steel-reinforced concrete. In *Concrete Solutions*, pages 417–424. CRC Press, June 2009.

[44] J. B. Butcher, D. Verstraeten, B. Schrauwen, C. R. Day, and P. W. Haycock. Reservoir computing and extreme learning machines for non-linear time-series data analysis. *Neural Networks*, 38:76 – 89, 2013.

[45] J. B. Butcher, C. R. Day, P. W. Haycock, D. Verstraeten, and B. Schrauwen. Defect detection in reinforced concrete using random neural architectures. *Computer-Aided Civil and Infrastructure Engineering*, 29(3):191–207, 2014.

[46] A. J. Wootton, C. R. Day, and P. W. Haycock. Echo state network applications in structrual health monitoring. In *Proceedings of the 53rd Annual Conference of The British Institute of Non-Destructive Testing (NDT 2014)*, pages 289 – 300, 2014.

[47] G. Sawade and H.-J. Krause. Magnetic flux leakage (MFL) for the non-destructive evaluation of pre-stressed concrete structures. In *Non-Destructive Evaluation of Reinforced Concrete Structures: Non-Destructive Testing Methods*, chapter 11. Elsevier, 2010.

[48] H. C. Rhim and O. Buyukozturk. Electromagnetic properties of concrete at microwave frequency range. *ACI Materials Journal*, 95(3):262 – 271, 1998.

[49] H. Diederich and T. Vogel. Detection of reinforcement bar fractures by measuring the remanent and active magnetic field. In *The 10th fib International PhD Symposium in Civil Engineering*, 2014.

[50] F. N. Kusenberger and J. R. Barton. Detection of flaws in reinforcing steel in prestressed concrete bridge members. Technical report, Southwest Research Institute, 1981.

[51] F. N. Kusenberger, A. S. Lozano, and W. B. Tarver, Jr. Magnetic inspection of reinforcing steel using sensor array, 07 1985.

[52] J. Makar and R. Desnoyers. Magnetic field techniques for the inspection of steel under concrete cover. *NDT & E International*, 34(7):445 – 456, 2001.

[53] Y. Shi, C. Zhang, R. Li, M. Cai, and G. Jia. Theory and application of magnetic flux leakage pipeline detection. *Sensors*, 15(12):31036–31055, 2015.

[54] H. Scheel and B. Hillemeier. Capacity of the remanent magnetism method to detect fractures of steel in tendons embedded in prestressed concrete. *NDT & E International*, 30(4):211 – 216, 1997.

[55] B. Hillemeier, C. Flohrer, and A. Schaab. Die zerstorungsfreie ortung von spannstahlbruchen in spannbeton-deckentragern. *Beton- und Stahlbetonbau*, 84(10):268–270, 1989.

[56] A. Walther and B. Hillemeier. Schnelle zerstörungsfreie ortung von spannstahlrissen in querspanngliedern von spannbetonbrücken. In *Structural Faults + Repair - 2008*, 2008.

[57] S. Knapp and B. Hillemeier. Application of line scanner in remanent and active field compared with the big magnet impulse magnetization. In F. Biondini and

D. M. Frangopol, editors, *Bridge Maintenance, Safety, Management, Resilience and Sustainability*, 2012.

[58] A. Walther. The magnetic flux method on prestressed building structures. In *NDT-CE 2015*, 2015.

[59] K. Szielasko, S. Youssef, A. Sourkov, J. Kurz, S. Pushkarev, and R Birringer. Magnetic flux leakage detection of corrosion damage in prestressed concrete poles. In *Electromagnetic Nondestructive Evaluation (XVIII)*, 2015.

[60] A. Ghorbanpoor. Magnetic-based NDE of steel in prestressed and post-tensioned concrete bridges. In *Proceedings of SPIE 3400, Structural Materials Technology III: An NDT Conference*, pages 343–347, 1998.

[61] B. Fernandes, D. Nims, and V. Devabhaktuni. Computer aided modeling of magnetic behavior of embedded prestressing strand for corrosion estimation. *Journal of Nondestructive Evaluation*, 32(2):124–133, 2013.

[62] B. Fernandes, D. Nims, and V. Devabhaktuni. Comprehensive MMF-MFL inspection for corrosion detection and estimation in embedded prestressing strands. *Journal of Civil Structural Health Monitoring*, 4(1):43–55, 2014.

[63] T Wolf and T. Vogel. Detection of breaks in reinforcing bars with the magnetic flux leakage method. In *Proceedings of the 4th International Conference on Structural Health Monitoring of Intelligent Infrastructure: 22 - 24 July 2009 Zurich, Switzerland*, 2009.

[64] T. Wolf and T. Vogel. Detection of reinforcement breaks: Laboratory experiments and an application of the magnetic flux leakage method. In *Structural Faults and Repair - 2012*, 2012.

[65] H. Diederich and T. Vogel. Evaluation of reinforcing bars using the magnetic flux leakage method. *Journal of Infrastructure Systems*, 23(1):B4016001, 2017.

[66] M. Goktepe. Non-destructive crack detection by capturing local flux leakage field. *Sensors and Actuators A: Physical*, 91:70 – 72, 2001.

[67] D. Perin and M. Goktepe. Inspection of rebars in concrete blocks. *International Journal of Applied Electromagnetics and Mechanics*, 38(2):65–78, 2012.

[68] T. C Jensen, S. E. Wendt, J. N. Gray, C. C. H. Lo, F. J. Margetan, D. J. Eisenmann, N. Nakagawa, and L. H. Brasche. Feasibility study for detection and quantification of corrosion in bridge barrier rails. Technical report, Iowa State University, 2013.

[69] H.-J. Krause, W. Wolf, W. Glaas, E. Zimmermann, M. I. Faley, G. Sawade, R. Mattheus, G. Neudert, U. Gampe, and J. Krieger. SQUID array for magnetic inspection of prestressed concrete bridges. *Physica C: Superconductivity*, 368(1-4):91 – 95, 2002.

[70] S. Huang and S. Wang. *New Technologies in Electromagnetic Non-destructive Testing*, chapter Magnetic Flux Leakage Testing, pages 185–222. Springer Singapore, Singapore, 2016.

[71] T. A. Bubenik, J. B. Nestleroth, R. J. Eiber, and B. F. Saffell. Magnetic flux leakage (MFL) technology for natural gas pipeline inspection. Topical report, November 1992. Technical report, Battelle Memorial Institute, Columbus, OH (United States), Nov 1992.

[72] A. P. Pereira Fulco, J. D. D. Melo, C. A. Paskocimas, S. N. de Medeiros, F. L. de Araujo Machado, and A. R. Rodrigues. Magnetic properties of polymer matrix composites with embedded ferrite particles. *NDT & E International*, 77:42 – 48, 2016.

[73] A. Mohamed, M. S. Hamdi, and S. Tahar. A machine learning approach for big data in oil and gas pipelines. In *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on*, pages 585–590, Aug 2015.

[74] A. A. Carvalho, J. M. A. Rebello, L. V. S. Sagrilo, C. S. Camerini, and I. V. J. Miranda. MFL signals and artificial neural networks applied to detection and classification of pipe weld defects. *NDT & E International*, 39(8):661 – 667, 2006.

[75] P. Karuppasamy, A. Abudhahir, M. Prabhakaran, S. Thirunavukkarasu, B. P. C. Rao, and T. Jayakumar. Model-based optimization of MFL testing of ferromagnetic steam generator tubes. *Journal of Nondestructive Evaluation*, 35(1):1–9, 2015.

[76] X. Hu, G. Shen, C. Lu, and D. Liu. Study of evaluation method of the function of magnetic flux leakage testing instrument for large atmosphere storage tanks. In *Structural Health Monitoring and Integrity Management*, pages 409–413. CRC Press, May 2015.

[77] S. K. Burke, M. E. Ibrahim, and G. R. Hugo. Principles and application of magnetic rubber testing for crack detection in high-strength steel components: II. residual-field inspection. Technical report, Defence Science and Technology Organisation, Fishermen's Bend (Australia) Maritime Division, 2014.

[78] S. K. Verma, S. S. Bhadauria, and S. Akhtar. Review of nondestructive testing methods for condition monitoring of concrete structures. *Journal of Construction Engineering*, 2013:1 – 11, 2013.

[79] J. Helal, M. Sofi, and P. Mendis. Non-destructive testing of concrete: A review of methods. *Electronic Journal of Structural Engineering*, 14(1):97 – 105, 2015.

[80] R. F. Stratfull. The corrosion of steel in a reinforced concrete bridge. *CORROSION*, 13(3):43–48, 1957.

[81] B. Elsener, C. Andrade, J. Gulikers, R. Polder, and M. Raupach. Half-cell potential measurements-potential mapping on reinforced concrete structures. *Materials and Structures*, 36(7):461–471, 2003.

[82] ASTM Standard C876-15: Standard test method for corrosion potentials of uncoated reinforcing steel in concrete, 2015.

[83] Z. H. Zou, J. Wu, Z. Wang, and Z. Wang. Relationship between half-cell potential and corrosion level of rebar in concrete. *Corrosion Engineering, Science and Technology*, 51(8):588–595, 2016.

[84] V. Leelalerkiet, J.-W. Kyung, M. Ohtsu, and M. Yokota. Analysis of half-cell potential measurement for corrosion of reinforced concrete. *Construction and Building Materials*, 18(3):155 – 162, 2004. 3rd Kumamoto International Workshop on Fracture, Acoustic Emission and NDE in Concrete (KIFA-3).

[85] B. Assouli, G. Ballivy, and P. Rivard. Influence of environmental parameters on application of standard ASTM C876-91: Half cell potential measurements. *Corrosion Engineering Science and Technology*, 43(1):93 – 96, 2008.

[86] S. Sathiyanarayanan, P. Natarajan, K. Saravanan, S. Srinivasan, and G. Venkatachari. Corrosion monitoring of steel in concrete by galvanostatic pulse technique. *Cement and Concrete Composites*, 28(7):630 – 637, 2006.

[87] B Elsener, O Klinghoffer, T Frolund, E Rislund, Y Schiegg, and H Böhni. Assessment of reinforcement corrosion by means of galvanostatic pulse technique. In *Proceedings of International Conference on Repair of Concrete Structures-From Theory to Practice in a Marine Environment*, pages 391–400, 1997.

[88] J. A. Gonzalez, A. Cobo, M. N. Gonzalez, and S. Feliu. On-site determination of corrosion rate in reinforced concrete structures by use of galvanostatic pulses. *Corrosion Science*, 43(4):611 – 625, 2001.

[89] Y. T. Dou, B. H. Hao, J. Xie, B. Meng, M. L. Dong, and A. L. Zhang. The study to the corrosion of reinforcing steel in concrete by using galvanostatic pulse technique. In *Advances in Civil and Structural Engineering III*, volume 501 of *Applied Mechanics and Materials*, pages 916–919. Trans Tech Publications, 4 2014.

[90] M. G. Grantham and J. Broomfield. The use of linear polarisation corrosion rate measurements in aiding rehabilitation options for the deck slabs of a reinforced

concrete underground park. *Construction and Building Materials*, 11(4):215 – 224, 1997.

[91] C. Andrade and C. Alonso. Test methods for on-site corrosion rate measurement of steel reinforcement in concrete by means of the polarization resistance method. *Materials and Structures*, 37(9):623–643, 2004.

[92] S. Feliu, J. A. González, J. M. Miranda, and V. Feliu. Possibilities and problems of in situ techniques for measuring steel corrosion rates in large reinforced concrete structures. *Corrosion Science*, 47(1):217 – 238, 2005.

[93] W. Liu, R. Hunsperger, M. Chajes, and E. Kunz. An overview of corrosion damage detection in steel bridge strands using TDR. In *Proceedings of the 2nd International Symposium on TDR for Innovative Applications*, 2001.

[94] W. Liu, R. G. Hunsperger, K. Folliard, M. J. Chajes, J. Barot, D. Jhaveri, and E. Kunz. Detection and characterization of corrosion of bridge cables by time domain reflectometry. In *Proc. SPIE*, volume 3587, pages 28–39, 1999.

[95] W. Morris, E.I. Moreno, and A.A. Sagues. Practical evaluation of resistivity of concrete in test cylinders using a wenner array probe. *Cement and Concrete Research*, 26(12):1779 – 1787, 1996.

[96] R. B. Polder. Test methods for on site measurement of resistivity of concrete — a RILEM TC-154 technical recommendation. *Construction and Building Materials*, 15(2–3):125 – 131, 2001. Near Surface Testing of.

[97] W. Lopez and J. A. Gonzalez. Influence of the degree of pore saturation on the resistivity of concrete and the corrosion rate of steel reinforcement. *Cement and Concrete Research*, 23(2):368 – 376, 1993.

[98] S. Feliu, J. A. Gonzalez, S. Feliu, and C. Andrade. Relationship between conductivity of concrete and corrosion of reinforcing bars. *British Corrosion Journal*, 24(3):195–198, 1989.

[99] W. Morris, A. Vico, M. Vazquez, and S. R. de Sanchez. Corrosion of reinforcing steel evaluated by means of concrete resistivity measurements. *Corrosion Science*, 44(1):81 – 99, 2002.

[100] G.K. Glass, C.L. Page, and N.R. Short. Factors affecting the corrosion rate of steel in carbonated mortars. *Corrosion Science*, 32(12):1283 – 1294, 1991.

[101] F. Soldovieri, R. Persico, E. Utsi, and V. Utsi. The application of inverse scattering techniques with ground penetrating radar to the problem of rebar location in concrete. *NDT & E International*, 39(7):602 – 607, 2006.

[102] D. R. Huston, N. V. Pelczarski, B. Esser, and K. R. Maser. Damage detection in roadways with ground penetrating radar. In *Proc. SPIE*, volume 4084, pages 91–94, 2000.

[103] A. Tarussov, M. Vandry, and A. De La Haza. Condition assessment of concrete structures using a new analysis method: Ground-penetrating radar computer-assisted visual interpretation. *Construction and Building Materials*, 38:1246 – 1254, 2013. 25th Anniversary Session for ACI 228 – Building on the Past for the Future of NDT of Concrete.

[104] A. Giannopoulos. Modelling ground penetrating radar by GprMax. *Construction and Building Materials*, 19(10):755 – 762, 2005. Non Destructive Testing: Selected papers from Structural Faults and Repair 2003.

[105] M. Solla, H. Lorenzo, and V. Perez-Gracia. *Non-Destructive Techniques for the Evaluation of Structures and Infrastructure*, chapter 5, pages 89 – 112. CRC Press, 2016.

[106] M. I. Hasan and N. Yazdani. An experimental study for quantitative estimation of rebar corrosion in concrete using ground penetrating radar. *Journal of Engineering*, 2016:1 – 8, 2016.

[107] N. Martino, K. Maser, R. Birken, and M. Wang. Quantifying bridge deck corrosion

using ground penetrating radar. *Research in Nondestructive Evaluation*, 27(2):112–124, 2016.

[108] S. Sherratt, L. North, P. Haycock, S. Hoon, and N. Cassidy. A ferromagnetic resonance probe. In *Concrete Solutions*, pages 431 – 437. CRC Press, June 2009.

[109] S. L. Sherratt. *Characterisation of iron oxide corrosion product using impulse ferromagnetic resonance*. PhD thesis, Keele University, 2010.

[110] J. Dobson and P. Cawley. Independent component analysis for improved defect detection in guided wave monitoring. *Proceedings of the IEEE*, 104(8):1620–1631, Aug 2016.

[111] B. L. Ervin, D. A. Kuchma, J. T. Bernhard, and H. Reis. Monitoring corrosion of rebar embedded in mortar using high-frequency guided ultrasonic waves. *Journal of Engineering Mechanics*, 135(1):9 – 19, 2009.

[112] U. Amjad, S. Kumar Yadav, and T. Kundu. Detection and quantification of diameter reduction due to corrosion in reinforcing steel bars. *Structural Health Monitoring*, 14(5):532–543, 2015.

[113] S. Sharma and A. Mukherjee. Monitoring corrosion in oxide and chloride environments using ultrasonic guided waves. *Journal of Materials in Civil Engineering*, 23(2):207 – 211, 2011.

[114] M. D. Beard, M. J. S. Lowe, and P. Cawley. Inspection of steel tendons in concrete using guided waves. *AIP Conference Proceedings*, 657(1):1139–1147, 2003.

[115] N. J. Carino. *Concrete Construction Engineering Handbook*, chapter 19, pages 1 – 68. CRC Press, 1997.

[116] B. P. C. Rao. Eddy current testing: Basics. *Journal of Non Destructive Testing & Evaluation*, 10(3):7 – 16, 2011.

[117] D. Huston, J. Cui, D. Burns, and D. Hurley. Concrete bridge deck condition assessment with automated multisensor techniques. *Structure and Infrastructure Engineering*, 7(7-8):613–623, 2011.

[118] D. Huston, N. Gucunski, A. Maher, J. Cui, D. Burns, and F. Jalinoos. Bridge deck condition assessment with electromagnetic, acoustic and automated methods. In *Proceedings of the 6th International Workshop on Structural Health Monitoring*, 2007.

[119] D. Huston, N. Gucunski, J. Cui, D. Burns, A. Maher, and F. Jalinoos. Multisensor and automated measurement of bridge deck condition. In *World Forum on Smart Materials and Smart Structures Technology*, pages –. CRC Press, July 2008.

[120] D. R. Huston, J. Cui, D. Burns, D. Hurley, and R. Arndt. Multiple sensor subsurface condition assessment of reinforced concrete bridge decks. In *Bridge Maintenance, Safety and Management*, pages 127–127. CRC Press, July 2010.

[121] F. Jalinoos, R. Arndt, D. Huston, and J. Cui. Periodic NDE for bridge maintenance. In *Proceedings of Structural Faults and Repair Conference*, 2010.

[122] C. Kohl, M. Krause, C. Maierhofer, and J. Wostmann. 2D- and 3D-visualisation of NDT-data using data fusion technique. *Materials and Structures*, 38(9):817–826, 2005.

[123] D. Breyssea. Recent developments in analyses techniques for non-destructive testing and assessment of concrete properties. *EPJ Web of Conferences*, 12:03001, 2011.

[124] A. Taffe, T. Kind, M. Stoppel, and H. Wiggenhauser. OSSCAR – development of an On-Site SCAnneR for automated non-destructive bridge testing. In *Concrete Repair, Rehabilitation and Retrofitting II*, 2008.

[125] H. Wiggenhauser, M. Stoppel, G. Dobmann, J. Kurz, M. Raupach, and K. Reichling. BETOSCAN - an instrumented mobile robot system for the diagnosis of

reinforced concrete floors. In *Concrete Repair, Rehabilitation and Retrofitting II*, pages 255–256. CRC Press, October 2008.

[126] N. Gucunski, A. Maher, and H. Ghasemi. Condition assessment of concrete bridge decks using a fully autonomous robotic NDE platform. *Bridge Structures*, 9(2, 3):123–130, 2013.

[127] R. Birken, G. Schirner, and M. Wang. VOTERS: Design of a mobile multi-modal multi-sensor system. In *Proceedings of the Sixth International Workshop on Knowledge Discovery from Sensor Data*, SensorKDD '12, pages 8–15, New York, NY, USA, 2012. ACM.

[128] Y. Lu, H. Liu, M. L. Wang, and R. Birken. Complementary pavement subsurface assessment using mobile acoustic subsurface sensing and ground penetrating radar systems. In *27th Annual Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP)*, 2014.

[129] M. Li, R. Vilbig, D. Busuioc, R. Birken, and M. Wang. Novel antenna designs for compact ground penetrating radar systems and in-traffic air-coupled applications. In *Symposium on the Application of Geophysics to Engineering and Environmental Problems 2013*, pages 181–190, 2013.

[130] D. Vines-Cavanau, D. Busuioc, R. Birken, and M. Wang. Millimeter-wave nondestructive evaluation of pavement conditions. In *Proc. SPIE*, volume 8347, pages 83472B–83472B–8, 2012.

[131] S. Ghanta, S. S. Shamsabadi, J. Dy, M. Wang, and R. Birken. A hessian-based methodology for automatic surface crack detection and classification from pavement images. In *Proc. SPIE*, volume 9437, pages 94371Z–94371Z–11, 2015.

[132] S. Ghanta, R. Birken, and J. Dy. Automatic road surface defect detection from grayscale images. In *Proc. SPIE*, volume 8347, pages 83471E–83471E–12, 2012.

[133] R. Birken, J. Zhang, and G. Schirner. *System-level design of a roaming multi-modal*

multi-sensor system for assessing and monitoring civil infrastructures, volume 2, chapter 6, pages 172 – 203. Elsevier, 2014.

[134] J. Zhang, H. Qiu, S. S. Shamsabadi, R. Birken, and G. Schirner. SIROM3 – a scalable intelligent roaming multi-modal multi-sensor framework. In *2014 IEEE 38th Annual Computer Software and Applications Conference (COMPSAC)*, pages 446–455, July 2014.

[135] J. Zhang, H. Qiu, S. S. Shamsabadi, R. Birken, and G. Schirner. WiP abstract: System-level integration of mobile multi-modal multi-sensor systems. In *2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, pages 227–227, April 2014.

[136] S. S. Shamsabadi, M. L. Wang, and R. Birken. PAVEMON: A GIS-based data management system for pavement monitoring based on large amounts of near-surface geophysical sensor data. In *Symposium on the Applicaton of Geophysics to Engeineering and Environmental Problems*, 2014.

[137] S. S. Shamsabadi. Design and implementation of PAVEMON: a GIS web-based pavement monitoring system based on large amounts of heterogeneous sensor data. Master's thesis, Northeastern University, 2014.

[138] J. H. Kurz, H. Rieder, M. Stoppel, and A. Taffe. Control and data acquisition of automated multi-sensor systems in civil engineering. In *NDTCE'09 Non-Destructive Testing in Civil Engineering*, 2009.

[139] P. Cotic, Z. Jaglicic, E. Niederleithinger, M. Stoppel, and B. Bosiljkov. Image fusion for improved detection of near-surface defects in NDT-CE using unsupervised clustering methods. *Journal of Nondestructive Evaluation*, 33(3):384–397, 2014.

[140] G. Dobmann, J. H. Hurz, A. Taffe, and D. Streicher. *Development of automated non-destructive evaluation (NDE) systems for reinforced concrete structures and other applications*, chapter 3, pages 30 – 62. Elsevier, 2010.

[141] C. Boller, J. Kurz, S. Pushkarev, N. Diersch, and S. Feistkorn. Combination for non-destructive testing: improving the field of application in buildings through characterization and fusion of the proceedings. Technical report, Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR) in the Federal Office for Building and Regional Planning, 2013.

[142] H. M. La, R. S. Lim, B. B. Basily, N. Gucunski, J. Yi, A. Maher, F. A. Romero, and H. Parvardeh. Mechatronic systems design for an autonomous robotic system for high-efficiency bridge deck inspection and evaluation. *IEEE/ASME Transactions on Mechatronics*, 18(6):1655–1664, Dec 2013.

[143] N. Gucunski, A. Maher, B. Basily, H. La, R. Lim, H. Parvardeh, and S.-H. Kee. Robotic platform RABIT for condition assessment of concrete bridge decks using multiple NDE technologies. *Journal of the Croatian Society of Non-Destructive Testing*, 3(4):5 – 12, 2013.

[144] N. Gucunski, S.-H. Kee, H. La, B. Basily, and A. Maher. Delamination and concrete quality assessment of concrete bridge decks using a fully autonomous RABIT platform. *Structural Monitoring and Maintenance*, 2(1):19 – 34, 2015.

[145] H. M. La, N. Gucunski, S.-H. Kee, and L. V. Nguyen. Data analysis and visualization for the bridge deck inspection and evaluation robotic system. *Visualization in Engineering*, 3(1):1–16, 2015.

[146] N. Gucunski, J. Yi, B. Basily, T. Duong, J. Kim, P. Balaguru, H. Parvardeh, A. Maher, and H. Najm. Concrete bridge deck early problem detection and mitigation using robotics. *Proc. SPIE*, 9437:94370P–94370P–12, 2015.

[147] J. Kim, N. Gucunski, and K. Dinh. Similarities and differences in bare concrete deck deterioration curves from multi NDE technology surveys. In *Proc. SPIE*, volume 9805, pages 98052H–98052H–9, 2016.

[148] H. M. La, T. H. Dinh, N. H. Pham, Q. P. Ha, and A. Q. Pham. Automated

robotic monitoring and inspection of steel structures and bridges. arXiv preprint, May 2017. arXiv:1705.04888.

[149] T. Le, S. Gibb, H. M. La, L. Falk, and T. Berendsen. Autonomous robotic system using non-destructive evaluation methods for bridge deck inspection. arXiv preprint, April 2017. arXiv:1704.04663.

[150] C. Volker and P. Shokouhi. Clustering based multi sensor data fusion for honeycomb detection in concrete. *Journal of Nondestructive Evaluation*, 34(4):1–10, 2015.

[151] C. Volker and P. Shokouhi. Multi sensor data fusion approach for automatic honeycomb detection in concrete. *NDT & E International*, 71:54 – 60, 2015.

[152] L. Jue-Long, L. Hai-Rui, Q. Zhi-Gen, X. Jian-Chun, and W. Wen-Tao. The research on damage prediction of corroded reinforced concrete based on data fusion. In Z. Dimitrovova, editor, *11th Conference on Vibration Problems*, 2013.

[153] Y. F. Wen, C. Z. Cai, X. H. Liu, J. F. Pei, X. J. Zhu, and T. T. Xiao. Corrosion rate prediction of 3C steel under different seawater environment by using support vector regression. *Corrosion Science*, 51(2):349 – 355, 2009.

[154] L. Sadowski. Non-destructive investigation of corrosion current density in steel reinforced concrete by artificial neural networks. *Archives of Civil and Mechanical Engineering*, 13(1):104 – 111, 2013.

[155] P. A. Vanniamparambil, M. Bolhassani, R. Carmi, F. Khan, I. Bartoli, F. L. Moon, A. Hamid, and A. Kontsos. A data fusion approach for progressive damage quantification in reinforced concrete masonry walls. *Smart Materials and Structures*, 23(1):015007, 2014.

[156] F. Khan, I. Bartoli, S. Rajaram, P. A. Vanniamparambil, A. Kontsos, M. Bolhassani, and A. Hamid. Acoustics and temperature based NDT for damage assessment of concrete masonry system subjected to cyclic loading. *Proc. SPIE*, 9063:90630B–90630B–10, 2014.

[157] J. Hola, L. Sadowski, and K. Schabowicz. Nondestructive identification of delaminations in concrete floor toppings with acoustic methods. *Automation in Construction*, 20(7):799 – 807, 2011.

[158] T. Gorzelanczyk, J. Hola, L. Sadowski, and K. Schabowicz. Methodology of nondestructive identification of defective concrete zones in unilaterally accessible massive members. *Journal of Civil Engineering and Management*, 19(6):775–786, 2013.

[159] S. Yaghi. Integrated remote sensing technologies for condition assessment of concrete bridges. Master's thesis, Concordia University, 2014.

[160] C. Kohl, M. Krause, C. Maierhofer, K. Mayer, J. Wostmann, and H. Wiggenhauser. 3D-visualisation of NDT data using a data fusion technique. *Insight - Non-Destructive Testing and Condition Monitoring*, 45(12):800–804, 2003.

[161] C. Kohl and D. Streicher. Results of reconstructed and fused NDT-data measured in the laboratory and on-site at bridges. *Cement and Concrete Composites*, 28(4):402 – 413, 2006. Non-Destructive Testing.

[162] J.P. Balayssac, S. Laurens, G. Arliguie, and I. Fortune. SENSO: a French project for the evaluation of concrete structures by combining non destructive testing methods. In *RILEM Symposium on On Site Assessment of Concrete, Masonry and Timber Structures - SACoMaTiS 2008*, pages 289 – 297, 2008.

[163] Z.-M. Sbartai, D. Breysse, M. Larget, and J.-P. Balayssac. Combining NDT techniques for improved evaluation of concrete properties. *Cement and Concrete Composites*, 34(6):725 – 733, 2012.

[164] H. L. Chavez-Garcia, E. M. Alonso-Guzman, W. Martinez-Molina, M. Graff, and J. C. Arteaga-Arcos. Prediction of the static modulus of elasticity using four non destructive testing. *Revista de la construccion*, 13:33 – 40, 04 2014.

[165] W. Martinez-Molina, A. A. Torres-Acosta, J. C. Jauregui, H. L. Chavez-Garcia, E. M. Alonso-Guzman, M. Graff, and J. C. Arteaga-Arcos. Predicting concrete

compressive strength and modulus of rupture using different NDT techniques. *Advances in Materials Science and Engineering*, 2014:1 – 15, 2014.

[166] D. R. Huston, J. Cui, D. Burns, D. Hurley, and R. Arndt. Multiple sensor subsurface condition assessment of reinforced concrete bridge decks. In *Bridge Maintenance, Safety, Management and Life-Cycle Optimization: Proceedings of the Fifth International IABMAS Conference, Philadelphia, USA*, 2010.

[167] O. Moselhi, M. Ahmed, and A. Bhowmick. Multisensor data fusion for bridge condition assessment. *Journal of Performance of Constructed Facilities*, 31(4):04017008, 2017.

[168] J. Cui, D. Huston, and R. W. Arndt. Data fusion for multiple sensor nondestructive evaluation on concrete bridge deck. In *TRB 92nd Annual Meeting Compendium of Papers*, 2013.

[169] K. R. Maser. Integration of ground penetrating radar and infrared thermography for bridge deck condition evaluation. In *NDTCE'09, Non-Destructive Testing in Civil Engineering*, 2009.

[170] Elcometer. Elcometer CoverMaster^{TM} Software. Online, 2016. Can be downloaded from http://www.elcometer.co.uk/en/concrete-inspection/covermaster-software/elcometer-covermaster-software.html.

[171] D. Clayton and M. Hileman. Light water reactor sustainability: Nondestructive evaluation for concrete research and development roadmap. Technical report, Oak Ridge National Laboratory for the U.S. Department of Energy, 2012.

[172] K. E. D. Wapenaar. TNO early research program 2015 - 2018: Annual plan 2015. Technical report, Netherlands Organisation for Applied Scientific Research (TNO), 2014.

[173] C. R. Farrar and K. Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1851):303–315, 2007.

[174] D. Balageas. *Introduction to Structural Health Monitoring*, chapter 1, pages 13–43. ISTE, 2010.

[175] J. P. Lynch and K. J. Loh. A summary review of wireless sensors and sensor networks for structural health monitoring. *The Shock and Vibration Digest*, 38(2):91 – 128, 2006.

[176] H. Jain, A. Rawat, and A. K. Sachan. A review on advancement in sensor technology in structural health monitoring system. *Journal of Structural Engineering and Management*, 2(3):1 – 7, 2015.

[177] K. Dragos and K. Smarsly. A comparitive review of wireless sensor nodes for structural health monitoring. In *Proceedings of the 7th International Conference on Structural Health Monitoring of Intelligent Infrastructure*, 2015.

[178] R. Viswanathan and P. K. Varshney. Distributed detection with multiple sensors i. fundamentals. *Proceedings of the IEEE*, 85(1):54–63, Jan 1997.

[179] M. Bhuiyan, J. Wu, G. Wang, and J. Cao. Sensing and decision-making in cyber-physical systems: The case of structural event monitoring. *IEEE Transactions on Industrial Informatics*, 12(6):2103 – 2114, 2016.

[180] D. Hughes and N. Correll. Distributed machine learning in materials that couple sensing, actuation, computation and communication. arXiv preprint, 2016. arXiv:1510.03800.

[181] S. D. Glaser, H. Li, M. L. Wang, J. Ou, and J. Lynch. Sensor technology innovation for the advancement of structural health monitoring: a strategic program of us-china research for the next decade. *Smart Structures and Systems*, 3(2):221 – 244, 2007.

[182] H. Sohn, M. Dzwonczyk, E. G. Straser, A. S. Kiremidjian, K. H. Law, and T. Meng. An experimental study of temperature effect on modal parameters of the alamosa canyon bridge. *Earthquake Engineering & Structural Dynamics*, 28(8):879–897, 1999.

[183] J. M. Ko, K. K. Chak, J. Y. Wang, Y. Q. Ni, and T. H. T. Chan. Formulation of an uncertainty model relating modal parameters and environmental factors by using long-term monitoring data. In *Proc. SPIE*, volume 5057, pages 298–307, 2003.

[184] J. Gu, M. Gul, and X. Wu. Damage detection under varying temperature using artificial neural networks. *Structural Control and Health Monitoring*, -:e1998–n/a, 2017. e1998 STC-15-0016.R3.

[185] Y. Yu and J. Ou. Development of a kind of multi-variable wireless sensor for structural health monitoring in civil engineering. In *Proc. SPIE*, volume 5765, pages 158–166, 2005.

[186] F. Xiao, J. L. Hulsey, and R. Balasubramanian. Fiber optic health monitoring and temperature behavior of bridge in cold region. *Structural Control and Health Monitoring*, Early View Article:e2020–n/a, 2017.

[187] E. J. Cross, K. Worden, and Q. Chen. Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. The Royal Society, 2011.

[188] E.J. Cross and K. Worden. Approaches to nonlinear cointegration with a view towards applications in SHM. *Journal of Physics: Conference Series*, 305(1):012069, 2011.

[189] K. Worden, E. J. Cross, and J. M. W. Brownjohn. Switching response surface models for structural health monitoring of bridges. In S. Koziel and L. Leifsson, editors, *Surrogate-Based Modeling and Optimization: Applications in Engineering*, pages 337–358. Springer New York, New York, NY, 2013.

[190] B. Peeters, J. Maeck, and G. de Roeck. Dynamic monitoring of the Z24-bridge: separating temperature effects from damage. In *Proceedings of the European*

COST F3 Conference on System Identification and Structural Health Monitoring, 2000.

[191] B. Peeters and G. De Roeck. One-year monitoring of the Z24-bridge: environmental effects versus damage events. *Earthquake Engineering & Structural Dynamics*, 30(2):149–171, 2001.

[192] E. Cross. *On Structural Health Monitoring in Changing Environmental and Operational Conditions*. PhD thesis, University of Sheffield, 2012.

[193] E. J. Cross, K. Y. Koo, J. M. W. Brownjohn, and K. Worden. Long-term monitoring and data analysis of the Tamar bridge. *Mechanical Systems and Signal Processing*, 35(1-2):16 – 34, 2013.

[194] I. Laory, T. N. Trinh, I. F. C. Smith, and J. M. W. Brownjohn. Methodologies for predicting natural frequency variation of a suspension bridge. *Engineering Structures*, 80(0):211 – 221, 2014.

[195] N. de Battista, R. Westgate, K. Y. Koo, and J. M. W. Brownjohn. Wireless monitoring of the longitudinal displacement of the Tamar suspension bridge deck under changing environmental conditions. In *Proc. SPIE*, volume 7981, pages 79811O–79811O–15, 2011.

[196] T. Kijewski-Correa, M. Haenggi, and P. Antsaklis. Wireless sensor networks for structural health monitoring: A multi-scale approach. In *Structures Congress 2006*, pages 1–16, 2006.

[197] R. Kromanis and P. Kripakaran. SHM of bridges: characterising thermal response and detecting anomaly events using a temperature-based measurement interpretation approach. *Journal of Civil Structural Health Monitoring*, 6(2):237–254, 2016.

[198] C. Kramer, C. A. M. de Smet, and G. de Roeck. Z24 bridge damage detection tests. In *1999 IMAC XVII - 17th International Modal Analysis Conference*, 1999.

[199] J. M. W. Brownjohn. Structural health monitoring of the Tamar bridge. In *Industrial Safety and Life Cycle Engineering*, chapter 22, pages 465 – 490. VCE Vienna Consulting Engineers ZT GmbH, 2013.

[200] K. Y. Koo, J. M. W. Brownjohn, D. I. List, and R. Cole. Structural health monitoring of the Tamar suspension bridge. *Structural Control and Health Monitoring*, 20(4):609–625, 2013.

[201] N. Dervilis, E. J. Cross, R. J. Barthorpe, and K. Worden. Robust methods of inclusive outlier analysis for structural health monitoring. *Journal of Sound and Vibration*, 333(20):5181 – 5195, 2014.

[202] N. Dervilis, I. Antoniadou, E. J. Cross, and K. Worden. A non-linear manifold strategy for SHM approaches. *Strain*, 51(4):324–331, 2015.

[203] N. Dervilis, H. Shi, K. Worden, and E. J. Cross. Exploring environmental and operational variations in SHM data using heteroscedastic gaussian processes. In Shamim Pakzad and Caicedo Juan, editors, *Dynamics of Civil Structures, Volume 2: Proceedings of the 34th IMAC, A Conference and Exposition on Structural Dynamics 2016*, pages 145–153, Cham, 2016. Springer International Publishing.

[204] F. Flammini, A. Gaglione, N. Mazzocca, and C. Pragliola. DETECT: a novel framework for the detection of attacks to critical infrastructures. In S. Martorell, C. G. Soares, and J. Barnett, editors, *Safety, Reliability and Risk Analysis: Theory, Methods and Applications*, pages 105 – 112. CRC Press, 2008.

[205] F. Flammini, A. Gaglione, F. Ottello, A. Pappalardo, C. Pragliola, and A. Tedesco. Towards wireless sensor networks for railway infrastructure monitoring. In *Electrical Systems for Aircraft, Railway and Ship Propulsion*, pages 1–6, Oct 2010.

[206] A. Smyth and M. Wu. Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring. *Mechanical Systems and Signal Processing*, 21(2):706 – 723, 2007.

[207] X.-H. Zhang. *Multi-sensing and multi-scale monitoring of long-span suspension bridges*. PhD thesis, Hong Kong Polytechnic University, 2013.

[208] H. Jo. *Multi-scale structural health monitoring using wireless smart sensors*. PhD thesis, University of Illinois at Urbana-Champaign, 2013.

[209] S. H. Sim, B. F. Spencer, Jr., F. Asce, and T. Nagayama. Multimetric sensing for structural damage detection. *Journal of Engineering Mechanics*, 137(1):22 – 30, 2011.

[210] J. W. Park, S.-H. Sim, and H. J. Jung. Displacement estimation using multimetric data fusion. *IEEE/ASME Transactions on Mechatronics*, 18(6):1675–1682, Dec 2013.

[211] J.-W. Park, S.-H. Sim, and H.-J. Jung. Wireless displacement sensing system for bridges using multi-sensor fusion. *Smart Materials and Structures*, 23(4):045022, 2014.

[212] F. Bruschetta, D. Zonta, C. Cappello, R. Zandonini, M. Pozzi, B. Glisic, D. Inaudi, D. Posenato, M. L. Wang, and Y. Zhao. Fusion of monitoring data from cable-stayed bridge. In *2013 IEEE Workshop on Environmental Energy and Structural Monitoring Systems (EESMS)*, pages 1–6, Sept 2013.

[213] D. Zonta. Sensor data analysis, reduction and fusion for assessing and monitoring civil infrastructures. In M. L. Wang, J. P. Lynch, and H. Sohn, editors, *Sensor Technologies for Civil Infrastructures*, volume 56 of *Woodhead Publishing Series in Electronic and Optical Materials*, chapter 2, pages 33 – 66. Woodhead Publishing, 2014.

[214] D. Sun, V. C. S. Lee, and Y. Lu. An intelligent data fusion framework for structural health monitoring. In *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, pages 49–54, June 2016.

[215] S. Cho, J.-W. Park, R. P. Palanisamy, and S.-H. Sim. Reference-free displacement

estimation of bridges using kalman filter-based multimetric data fusion. *Journal of Sensors*, 2016:1–9, 2016.

[216] C. D. Zhang and Y. L. Xu. Structural damage identification via multi-type sensors and response reconstruction. *Structural Health Monitoring*, 15(6):715–729, 2016.

[217] R. V. Kulkarni, A. Forster, and G. K. Venayagamoorthy. Computational intelligence in wireless sensor networks: A survey. *IEEE Communications Surveys Tutorials*, 13(1):68–96, First 2011.

[218] C. R. Farrar, G. Park, D. W. Allen, and M. D. Todd. Sensor network paradigms for structural health monitoring. *Structural Control and Health Monitoring*, 13(1):210–225, 2006.

[219] S. K. Yung. Local validation of sensor signals. In *IEE Colloquium on Fault Diagnosis and Control System Reconfiguration*, pages 7/1–7/8, May 1993.

[220] A. Abhinav. Sensor failure mode detection and self-validation. Master's thesis, University of Cincinnati, 2008.

[221] J. Gertler. Analytical redundancy methods in fault detection and isolation. In *Proceedings of IFAC/IAMCS symposium on safe process*, volume 1, pages 9 – 21, 1991.

[222] J. Ma, J. Q. Zhang, and Y. Yan. Wavelet transform based sensor validation. In *IEE Colloquium on Intelligent and Self-Validating Sensors (Ref. No. 1999/160)*, pages 10/1–10/4, Jun 1999.

[223] S. Wellington. *Algorithms for Sensor Validation and Multisensor Fusion*. PhD thesis, Nottingham Trent University, 2002.

[224] M. Abdelghani and M. I. Friswell. Sensor validation for structural systems with multiplicative sensor faults. *Mechnical Systems and Signal Processing*, 21:270–279, 2007.

[225] S. J. Lee, H. Sohn, and J.-W. Hong. Time reversal based piezoelectric transducer self-diagnosis under varying temperature. *Journal of Nondestructive Evaluation*, 29(2):75–91, 2010.

[226] T.-H. Yi, H.-B. Huang, and H.-N. Li. Development of sensor validation methodologies for structural health monitoring: A comprehensive review. *Measurement*, 109:200 – 214, 2017.

[227] V. Giurgiutiu. *Structural Health Monitoring with Piezoelectric Wafer Active Sensors*. Structures and Fracture ebook Collection Series. Elsevier Science, 2007.

[228] I. Buethe, B. Eckstein, and C.-P. Fritzen. Model-based detection of sensor faults under changing temperature conditions. *Structural Health Monitoring*, 13(2):109–119, 2014.

[229] C. P. Fritzen and I. Buethe. Sensor performance assessment based on a physical model and impedance measurements. In *Damage Assessment of Structures X*, volume 569 of *Key Engineering Materials*, pages 751–758. Trans Tech Publications, 10 2013.

[230] B. L. Grisso and D. J. Inman. Temperature corrected sensor diagnostics for impedance-based SHM. *Journal of Sound and Vibration*, 329(12):2323 – 2336, 2010. Structural Health Monitoring Theory Meets Practice.

[231] K. R. Mulligan, N. Quaegebeur, P.-C. Ostiguy, P. Masson, and S. Létourneau. Comparison of metrics to monitor and compensate for piezoceramic debonding in structural health monitoring. *Structural Health Monitoring*, 12(2):153–168, 2013.

[232] T. G. Overly, G. Park, K. M. Farinholt, and C. R. Farrar. Piezoelectric active-sensor diagnostics and validation using instantaneous baseline data. *IEEE Sensors Journal*, 9(11):1414–1421, Nov 2009.

[233] S. Park, G. Park, C.-B. Yun, and C. R. Farrar. Sensor self-diagnosis using a modified impedance model for active sensing-based structural health monitoring. *Structural Health Monitoring*, 8(1):71–82, 2009.

[234] K. R. Mulligan, N. Quaegebeur, P. Masson, L.-P. Brault, and C. Yang. Compensation of piezoceramic bonding layer degradation for structural health monitoring. *Structural Health Monitoring*, 13(1):68–81, 2014.

[235] D. Liang, L. Wu, Z. Fan, and Y. Xu. Self-diagnosis and self-reconfiguration of piezoelectric actuator and sensor network for large structural health monitoring. *International Journal of Distributed Sensor Networks*, 2015:1 – 16, January 2015.

[236] Q. Zhu and Q. Mao. Time-domain sensor validation for PVDF array. In *The 21st International Congress on Sound and Vibration*, 2014.

[237] Michael I. Friswell and Daniel J. Inman. Sensor validation for smart structures. In *Proc. SPIE*, volume 4073, pages 150–161, 2000.

[238] H.-B. Huang, T.-H. Yi, and H.-N. Li. Sensor fault diagnosis for structural health monitoring based on statistical hypothesis test and missing variable approach. *Journal of Aerospace Engineering*, 30(2):B4015003, 2015.

[239] G. Kerschen, P. De Boe, J.-C. Golinval, and K. Worden. Sensor validation using principal component analysis. *Smart Materials and Structures*, 14(1):36, 2005.

[240] M. Sharifi. *Sensor fault diagnosis using principal component analysis*. PhD thesis, Texas A&M University, 2009.

[241] M. R. Hernandez-Garcia and S. F. Masri. Application of statistical monitoring using latent-variable techniques for detection of faults in sensor networks. *Journal of Intelligent Material Systems and Structures*, 25(2):121–136, 2014.

[242] Q. Ling, Z. Tian, Y. Yin, and Y. Li. Localized structural health monitoring using energy-efficient wireless sensor networks. *IEEE Sensors Journal*, 9(11):1596–1604, Nov 2009.

[243] L. Jiang, D. Djurdjanovic, J. Ni, and J. Lee. *Sensor Degradation Detection in Linear Systems*, pages 1252–1260. Springer London, London, 2006.

[244] D. A. Pereira and A. L. Serpa. Bank of H_∞ filters for sensor fault isolation in active controlled flexible structures. *Mechanical Systems and Signal Processing*, 60-61:678 – 694, 2015.

[245] D. A. Pereira and A. L. Serpa. Fault detection of sensors and actuators in smart structures. In *Proceedings of COBEM 2011*, 2011.

[246] K. Smarsly and K. H. Law. Decentralized fault detection and isolation in wireless structural health monitoring systems using analytical redundancy. *Advances in Engineering Software*, 73:1 – 10, 2014.

[247] K. Smarsly and Y. Petryna. A decentralized approach towards autonomous fault detection in wireless structural health monitoring systems. In V. Le Cam, L. Mevel, and F. Schoefs, editors, *EWSHM - 7th European Workshop on Structural Health Monitoring*, Nantes, France, July 2014. IFFSTTAR, Inria, Université de Nantes.

[248] P. Sun, Z. Wu, H. Yang, Z. Ming, and X. Guo. Research methods of sensors validation based on naive bayesian classifier. In *2015 11th International Conference on Computational Intelligence and Security (CIS)*, pages 235–238, Dec 2015.

[249] C. Lo, M. Liu, and J. P. Lynch. Distributive model-based sensor fault diagnosis in wireless sensor networks. In *2013 IEEE International Conference on Distributed Computing in Sensor Systems*, pages 313–314, May 2013.

[250] K. Worden. Auto-associative neural networks and principal component analysis. In *IMAC-XXI: Conference & Exposition on Structural Dynamics*, 2003.

[251] V. Avasarala, J. R. Celaya, K. Goebel, and N. Eklund. Sensor validation in non-destructive evaluation using clustering. In *Proc. SPIE*, volume 6167, pages 61671F–61671F–8, 2006.

[252] G. Cazzulani, S. Cinquemani, and M. Ronchi. A sliding-mode based observer to identify faults in FBG sensors embedded in composite structures. In *Proc. SPIE*, volume 9803, pages 98033S–98033S–7, 2016.

[253] J. Harmouche, C. Delpha, and D. Diallo. Incipient fault amplitude estimation using KL divergence with a probabilistic approach. *Signal Processing*, 120:1 – 7, 2016.

[254] Z. Li, B. H. Koh, and S. Nagarajaiah. Detecting sensor failure via decoupled error function and inverse input–output model. *Journal of Engineering Mechanics*, 133(11):1222–1228, 2007.

[255] C. Lo, J. P. Lynch, and M. Liu. Distributed model-based nonlinear sensor fault diagnosis in wireless sensor networks. *Mechanical Systems and Signal Processing*, 66–67:470 – 484, 2016.

[256] A. R. Mohan Rao, V. Kasireddy, N. Gopalakrishnan, and K. Lakshmi. Sensor fault detection in structural health monitoring using null subspace–based approach. *Journal of Intelligent Material Systems and Structures*, 26(2):172–185, 2015.

[257] A. R. M. Rao, S. K. Kumar, and K. Lakshmi. A sensor fault detection algorithm for structural health monitoring using adaptive differential evolution. *International Journal for Computational Methods in Engineering Science and Mechanics*, 15(3):282–293, 2014.

[258] V. A. Tran, S. T. Quek, and W. H. Duan. Sensor validation in damage locating vector method for structural health monitoring. *International Journal of Structural Stability and Dynamics*, 11(01):149–180, 2011.

[259] Z. Liu and N. Mrad. Validation of strain gauges for structural health monitoring with bayesian belief networks. *IEEE Sensors Journal*, 13(1):400–407, Jan 2013.

[260] H. Xiaowei, Z. Peng, D. Pinyi, C. Weimin, and X. Fei. Research on bridge sensor validation based on correlation in cluster. *Mathematical Problems in Engineering*, 2016:1–10, 2016.

[261] J. Kullaa. Detection, identification, and quantification of sensor fault in a sensor network. *Mechanical Systems and Signal Processing*, 40(1):208 – 221, 2013.

[262] H.-B. Huang, T.-H. Yi, and H.-N. Li. Canonical correlation analysis based fault diagnosis method for structural monitoring sensor networks. *Smart Structures and Systems*, 17(6):1031–1053, 2016.

[263] H.-B. Huang, T.-H. Yi, and H.-N. Li. Bayesian combination of weighted principal-component analysis for diagnosing sensor faults in structural monitoring systems. *Journal of Engineering Mechanics*, 143(9):Early Access, 2017.

[264] M. A. Hamid, M. Abdullah-Al-Wadud, and M. M. Alam. A reliable structural health monitoring protocol using wireless sensor networks. In *14th International Conference on Computer and Information Technology (ICCIT), 2011*, pages 601–606, Dec 2011.

[265] X. Liu, J. Cao, S. Tang, and P. Guo. Fault tolerant complex event detection in WSNs: A case study in structural health monitoring. *IEEE Transactions on Mobile Computing*, 14(12):2502–2515, Dec 2015.

[266] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28 – 44, 2013.

[267] D. Bellot, A. Boyer, and F. Charpillet. A new definition of qualified gain in a data fusion process: application to telemedicine. In *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997)*, volume 2, pages 865–872 vol.2, July 2002.

[268] D. L Hall and J. Llinas. An introduction to multisensor data fusion. In *Proceedings of the IEEE*, volume 85, 1997.

[269] D. L Hall and A. Steinberg. Dirty secrets in multisensor data fusion. Technical report, Pennsylvania State University Applied Research Laboratory, 2001.

[270] A. N. Steinberg, C. L. Bowman, and F. E. White. Revisions to the JDL data fusion model. In Belur V. Dasarathy, editor, *Proceedings of SPIE*, volume 3719, pages 430–441. SPIE, 1999.

[271] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, and F. White. Revisiting the JDL data fusion model ii. In P. Svensson and J. Schuber, editors, *Proceedings of the Seventh International Conference on Information Fusion (FUSION 2004)*, pages 1218–1230, 2004.

[272] W. Elmenreich. An introduction to sensor fusion. Research Report 47/2001, Technische Universitat Wien, Institut fur Technische Informatik, Treitlstr. 1-3/182-1, 1040 Vienna, Austria, 2001.

[273] B. Khaleghi, A. Khamis, and F. Karray. Multisensor data fusion: A data-centric review of the state of the art and overview of emerging trends. In *Devices, Circuits, and Systems*, pages 15–34. CRC Press, August 2015.

[274] Federico Castanedo. A review of data fusion techniques. *The Scientific World Journal*, 2013:1–19, 2013.

[275] J. Dong, D. Zhuang, Y. Huang, and J. Fu. Advances in multi-sensor data fusion: Algorithms and applications. *Sensors*, 9:7771–7784, 2009.

[276] H. Durrant-Whyte and T. C. Henderson. *Springer Handbook of Robotics*, chapter 25, pages 585–610. Springer-Verlag Berlin Heidelberg, 2008.

[277] D. Smith and S. Singh. Approaches to multisensor data fusion in target tracking: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 18(12):1696–1710, Dec 2006.

[278] R. C. Luo and M. G. Kay. Multisensor integration and fusion in intelligent systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5):901–931, Sep 1989.

[279] B. Helifa, M. Feliachi, I. K. Lefkaier, F. Boubenider, A. Zaoui, and N. Lagraa. Characterization of surface cracks using eddy current NDT simulation by 3D-FEM and inversion by neural network. *Applied Computational Electromagnetics Society Journal*, 31(2):187 – 194, 2016.

[280] S. Sambath, P. Nagaraj, N. Selvakumar, S. Arunachalam, and T. Page. Automatic detection of defects in ultrasonic testing using artificial neural network. *International Journal of Microstructure and Materials Properties*, 5(6):561–574, 2010.

[281] M Sutcliffe and J Lewis. Automatic defect recognition of single-v welds using full matrix capture data, computer vision and multi-layer perceptron artificial neural networks. *Insight - Non-Destructive Testing and Condition Monitoring*, 58(9):487–493, 2016.

[282] O. Abdeljaber, O. Avci, N. T. Do, M. Gul, O. Celik, and F. N. Catbas. *Quantification of Structural Damage with Self-Organizing Maps*, pages 47–57. Springer International Publishing, Cham, 2016.

[283] S. Chatterjee, S. Sarkar, S. Hore, N. Dey, A. S. Ashour, and V. E. Balas. Particle swarm optimization trained neural network for structural failure prediction of multistoried rc buildings. *Neural Computing and Applications*, 0:1–12, 2016.

[284] Z. Dworakowski, T. Stepinski, K. Dragan, A. Jablonski, and T. Barszcz. *Ensemble ANN Classifier for Structural Health Monitoring*, pages 81–90. Springer International Publishing, Cham, 2016.

[285] P. Molchanov, S. Gupta, K. Kim, and K. Pulli. Multi-sensor system for driver's hand-gesture recognition. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–8, May 2015.

[286] R. Kiros, R. Salakhutdinov, and R. Zemel. Multimodal neural language models. In *Proceedings of the 31st International Conference on Machine Learning*, 2014.

[287] F. Alidoost and H. Arefi. Knowledge based 3d building model recognition using convolutional neural networks from lidar and aerial imageries. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3:833–840, 2016.

[288] Z. Deng, C. Mao, and X. Chen. Deep self-organizing reservoir computing model for visual object recognition. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 1325–1332, July 2016.

[289] Z. K. Malik, A. Hussain, and Q. J. Wu. Multilayered echo state machine: A novel architecture and algorithm. *IEEE Transactions on Cybernetics*, 47(4):946–959, 2016.

[290] G. Wen, H. Li, and D. Li. An ensemble convolutional echo state networks for facial expression recognition. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 873–878, Sept 2015.

[291] N. Lopes, B. Ribeiro, and J. Goncalves. Restricted boltzmann machines and deep belief networks on multi-core processors. In *2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, June 2012.

[292] D. L. Ly, V. Paprotski, and D. Yen. Neural networks on GPUs: Restricted boltzmann machines. Technical report, University of Toronto, 2008.

[293] S. Zhang, H. Yang, and Z. Yin. Multiple deep convolutional neural networks averaging for face alignment. *Journal of Electronic Imaging*, 24(3):033013, 2015.

[294] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[295] L. R. Medsker and L. C. Jain, editors. *Recurrent Neural Networks: Design and Applications*. CRC Press, 2001.

[296] H. Cardot and R. Bone, editors. *Recurrent Neural Networks for Temporal Data Processing*. InTech, 2011.

[297] H. Jaeger. A tutorial on training recurrent neural networks, covering BPPT,

RTRL, EKF and the "echo state network" approach. Technical report, German National Research Center for Information Technology, 2013.

[298] J. L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[299] R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. In S. Dasgupta and D. McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1310–1318, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

[300] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang. Phoneme recognition using time-delay neural networks. In Alex Waibel, , and Kai-Fu Lee, editors, *Readings in Speech Recognition*, pages 393 – 404. Morgan Kaufmann, San Francisco, 1990.

[301] J. B. Butcher. *Reservoir computing with high non-linear separation and long-term memory for time-series data analysis*. PhD thesis, Keele University, 2012.

[302] S. Hochreiter and J. Schmidhuber. Lstm can solve hard long time lag problems. In *Advances in neural information processing systems*, pages 473–479, 1997.

[303] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks - with an erratum note. Technical report, Fraunhofer Institute for Autonomous Intelligent Systems, 2010.

[304] H. Jaeger. A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach. Technical Report 4, Fraunhofer Institute for Autonomous Intelligent Systems (AIS), 2013.

[305] H. Jaeger, M. Lukosevicius, D. Popovici, and U. Siewert. Optimization and applications of echo state networks with leaky- integrator neurons. *Neural Networks*, 20(3):335 – 352, 2007. f.

[306] D. Verstraeten. *Reservoir Computing: computation with dynamical systems*. Phd thesis, Ghent University, 2009.

[307] E.A. Antonelo, B. Schrauwen, and D. Stroobandt. Event detection and localization for small mobile robots using reservoir computing. *Neural Networks*, 21(6):862 – 871, 2008. Computational and Biological Inspired Neural Networks, selected papers from ICANN 2007 17th International Conference on Artificial Neural Networks (ICANN).

[308] A. J. Wootton, S. L. Taylor, C. R. Day, and P. W. Haycock. Optimizing echo state networks for static pattern recognition. *Cognitive Computation*, 9(3):391–399, 2017.

[309] M. H. Tong, A. D. Bickett, E. M. Christiansen, and G. W. Cottrell. Learning grammatical structure with echo state networks. *Neural Networks*, 20(3):424 – 432, 2007.

[310] G. K. Venayagamoorthy and B. Shishir. Effects of spectral radius and settling time in the performance of echo state networks. *Neural Networks*, 22(7):861 – 863, 2009.

[311] M. Buehner and P. Young. A tighter bound for the echo state property. *IEEE Transactions on Neural Networks*, 17(3):820–824, May 2006.

[312] I. B. Yildiz, H. Jaeger, and S. J. Kiebel. Re-visiting the echo state property. *Neural Networks*, 35:1 – 9, 2012.

[313] S. Ganguli, D. Huh, and H. Sompolinsky. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48):18970–18975, 2008.

[314] D. Verstraeten, J. Dambre, X. Dutoit, and B. Schrauwen. Memory versus non-linearity in reservoirs. In *2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2010.

[315] D. Koryakin, J. Lohmann, and M. V. Butz. Balanced echo state networks. *Neural Networks*, 36:35 – 45, 2012.

[316] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar. Information processing capacity of dynamical systems. *Scientific Reports*, 2:514–, 2012.

[317] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):391 – 403, 2007.

[318] H. Jaeger. Long short-term memory in echo state networks: Details of a simulation study. Technical Report 27, Jacobs University, 2012.

[319] Q. Song and Z. Feng. Effects of connectivity structure of complex echo state network on its prediction performance for nonlinear time series. *Neurocomputing*, 73(10-12):2177 – 2185, 2010. Subspace Learning / Selected papers from the European Symposium on Time Series Prediction.

[320] C. Gallicchio and A. Micheli. Architectural and markovian factors of echo state networks. *Neural Networks*, 24(5):440 – 456, 2011.

[321] J. Triesch. A gradient rule for the plasticity of a neuron's intrinsic excitability. In W. Duch, J. Kacprzyk, E. Oja, and S. Zadrozny, editors, *Artificial Neural Networks: Biological Inspirations – ICANN 2005: 15th International Conference, Warsaw, Poland, September 11-15, 2005. Proceedings, Part I*, pages 65–70, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[322] B. Schrauwen, M. Wardermann, D. Verstraeten, J. J. Steil, and D. Stroobandt. Improving reservoirs using intrinsic plasticity. *Neurocomputing*, 71(7-9):1159 – 1171, 2008. Progress in Modeling, Theory, and Application of Computational Intelligence: 15th European Symposium on Artificial Neural Networks 2007.

[323] J. J. Steil. Online reservoir adaptation by intrinsic plasticity for backpropagation-decorrelation and echo state learning. *Neural Networks*, 20:353–364, 2007.

[324] J. Boedecker, O. Obst, N. M. Mayer, and M. Asada. Initialization and self-organized optimization of recurrent neural network connectivity. *HFSP Journal*, 3(5):340 – 349, 2009.

[325] B.G. Giraud, A. Lapedes, L. C. Liu, and J. C. Lemm. Lorentzian neural nets. *Neural Networks*, 8(5):757 – 767, 1995.

[326] D. S. Broomhead and D. Lowe. Radial basis function, multi-variable function interpolation and adaptive networks. Technical report, Royal Signals and Radar Establishment, 1988.

[327] S. K. Lahiri and K. C. Ghanta. Development of an artificial neural network correlation for prediction of hold-up of slurry transport in pipelines. *Chemical Engineering Science*, 63(6):1497 – 1509, 2008.

[328] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering*, 55(2):512–518, Feb 2008.

[329] S. M. A. Pahnehkolaei, A. Alfi, and J.A. T. Machado. Dynamic stability analysis of fractional order leaky integrator echo state neural networks. *Communications in Nonlinear Science and Numerical Simulation*, 47:328 – 337, 2017.

[330] S. Basterrech. Empirical analysis of the necessary and sufficient conditions of the echo state property. arXiv preprint, March 2017. arXiv:1703.06664.

[331] A. S. Charles, D. Yin, and C. J. Rozell. Distributed sequence memory of multi-dimensional inputs in recurrent neural networks. *Journal of Machine Learning Research*, 18:1 – 37, 2017.

[332] L. Livi, F. M. Bianchi, and C. Alippi. Determination of the edge of criticality in echo state networks through fisher information maximization. *IEEE Transactions on Neural Networks and Learning Systems*, PP(99):1–12, 2017.

[333] A. Prater. Comparison of echo state network output layer classification methods on noisy data. arXiv preprint, March 2017. arXiv:1703.04496.

[334] C. Gallicchio and A. Micheli. Echo state property of deep reservoir computing networks. *Cognitive Computation*, 9(3):337–350, 2017.

[335] K. Nikiforou, P. A. M. Mediano, and M. Shanahan. An investigation of the dynamical transitions in harmonically driven random networks of firing-rate neurons. *Cognitive Computation*, 9(3):351–363, 2017.

[336] W. Yao, Z. Zeng, and C. Lian. Generating probabilistic predictions using mean-variance estimation and echo state network. *Neurocomputing*, 219:536 – 547, 2017.

[337] C. Sheng, J. Zhao, and W. Wang. Map-reduce framework-based non-iterative granular echo state network for prediction intervals construction. *Neurocomputing*, 222:116 – 126, 2017.

[338] S. Zhong, X. Xie, L. Lin, and F. Wang. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. *Neurocomputing*, 238:191 – 204, 2017.

[339] K. Wu, Y. Zhu, Q. Li, and G. Han. Algorithm and implementation of distributed ESN using spark framework and parallel PSO. *Applied Sciences*, 7(4):353, Apr 2017.

[340] S. Nichele and M. S. Gunderson. Reservoir computing using non-uniform binary cellular automata. arXiv preprint, February 2017. arXiv:1702.03812.

[341] S. Ortin and L. Pesquera. Reservoir computing with an ensemble of time-delay reservoirs. *Cognitive Computation*, 9(3):327–336, 2017.

[342] S. Nichele and M. S. Molund. Deep reservoir computing using cellular automata. arXiv preprint, March 2017. arXiv:1703.02806.

[343] H. Palangi. *Deep Learning for Sequence Modelling: Applications in Natural Languages and Distributed Compressive Sensing*. PhD thesis, University of British Columbia, 2017.

[344] N. M. Mayer and Y.-H. Yu. Orthogonal echo state networks and stochastic evaluations of likelihoods. *Cognitive Computation*, 9(3):379–390, 2017.

[345] H. Wang, C. Ni, and X. Yan. Optimizing the echo state network based on mutual information for modeling fed-batch bioprocesses. *Neurocomputing*, 225:111 – 118, 2017.

[346] N. Chouikhi, B. Ammar, N. Rokbani, and A. M. Alimi. PSO-based analysis of echo state network parameters for time series forecasting. *Applied Soft Computing*, 55:211 – 225, 2017.

[347] F. M. Bianchi, L. Livi, C. Alippi, and R. Jenssen. Multiplex visibility graphs to investigate recurrent neural network dynamics. *Scientific Reports*, 7:44037–, February 2017.

[348] F. Xue, Q. Li, H. Zhou, and X. Li. Reservoir computing with both neuronal intrinsic plasticity and multi-clustered structure. *Cognitive Computation*, 9(3):400–410, 2017.

[349] J. Bueno, D. Brunner, M. C. Soriano, and I. Fischer. Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback. *Optics Express*, 25(3):2401–2412, Feb 2017.

[350] M. Antonik, P.and Haelterman and S. Massar. Online training for high-performance analogue readout layers in photonic reservoir computers. *Cognitive Computation*, 9(3):297–306, 2017.

[351] A. Katumba, M. Freiberger, and J. Bienstman, P.and Dambre. A multiple-input strategy to efficient integrated photonic reservoir computing. *Cognitive Computation*, 9(3):307–314, 2017.

[352] J. Dolinsky, K. Hirose, and S. Konishi. Readouts for echo-state networks built using locally regularized orthogonal forward regression. *Journal of Applied Statistics*, 0:1–23, 2017.

[353] Z. Song-Lin and L. Xue. Adaptive forgetting factor echo state networks for time series prediction. *International Journal of Intelligent Systems Technologies and Applications*, 16(1):80–93, 2017.

[354] J. Saadat, P. Moallem, and H. Koofigar. Training echo estate neural network using harmony search algorithm. *International Journal of Artificial Intelligence*, 15(1):163–179, 2017.

[355] S. Lokse, F. M. Bianchi, and R. Jenssen. Training echo state networks with regularization through dimensionality reduction. *Cognitive Computation*, 9(3):364–378, 2017.

[356] L. A. Alexandre and M. J. Embrechts. Reservoir size, spectral radius and connectivity in static classification problems. In Cesare Alippi, Marios Polycarpou, Christos Panayiotou, and Georgios Ellinas, editors, *Artificial Neural Networks - ICANN 2009*, volume 5768 of *Lecture Notes in Computer Science*, pages 1015–1024. Springer Berlin Heidelberg, 2009.

[357] L. A Alexandre, M. J. Embrechts, and J. Linton. Benchmarking reservoir computing on time-independent classification tasks. In *2009 International Joint Conference on Neural Networks (IJCNN)*, pages 89–93, June 2009.

[358] R. F. Reinhart and J. J. Steil. Attractor-based computation with reservoirs for online learning of inverse kinematics. In *European Symposium on Artificial Neural Networks (ESANN) 2009*, 2009.

[359] M. Lukosevicius, H. Jaeger, and B. Schrauwen. Reservoir computing trends. *KI - Kunstliche Intelligenz*, 26(4):365–371, 2012.

[360] A. Goudarzi and C. Teuscher. Reservoir computing: Quo vadis? In *Proceedings of the 3rd ACM International Conference on Nanoscale Computing and*

Communication (NANOCOM'16), pages 13:1–13:6, New York, NY, USA, 2016. ACM.

[361] L. Keuninckx, J. Danckaert, and G. Van der Sande. Real-time audio processing with a cascade of discrete-time delay line-based reservoir computers. *Cognitive Computation*, 9(3):315–326, 2017.

[362] S. Ntalampiras. A transfer learning framework for predicting the emotional content of generalized sound events. *The Journal of the Acoustical Society of America*, 141(3):1694–1701, 2017.

[363] E. Ramamurthy, C. Bauckhage, K. Buza, and S. Wrobel. Using echo state networks for cryptography. arXiv preprint, April 2017. arXiv:1704.01047.

[364] E. Maiorino, F. M. Bianchi, L. Livi, A. Rizzi, and A. Sadeghian. Data-driven detrending of nonstationary fractal time series with echo state networks. *Information Sciences*, 382–383:359 – 373, 2017.

[365] G. Shi, D. Liu, and Q. Wei. Echo state network-based Q-learning method for optimal battery control of offices combined with renewable energy. *IET Control Theory & Applications*, 11:915–922(7), April 2017.

[366] S. Morando, S. Jemei, D. Hissel, R. Gouriveau, and N. Zerhouni. ANOVA method applied to proton exchange membrane fuel cell ageing forecasting using an echo state network. *Mathematics and Computers in Simulation*, 131:283 – 294, 2017. 11th International Conference on Modeling and Simulation of Electric Machines, Converters and Systems.

[367] M. Badoni, B. Singh, and A. Singh. Implementation of echo-state network based control for power quality improvement. *IEEE Transactions on Industrial Electronics*, 64(7):5576–5584, 2017.

[368] M. Dorado-Moreno, L. Cornejo-Bueno, P.A. Gutierrez, L. Prieto, C. Hervas-Martinez, and S. Salcedo-Sanz. Robust estimation of wind power ramp events with reservoir computing. *Renewable Energy*, 111:428–437, 2017.

[369] Z. Zheng, S. Morando, M.-C. Pera, D. Hissel, L. Larger, R. Martinenghi, and A. B. Fuentes. Brain-inspired computational paradigm dedicated to fault diagnosis of PEM fuel cell stack. *International Journal of Hydrogen Energy*, 42(8):5410 – 5425, 2017.

[370] J. Park, B. Lee, S. Kang, P. Y. Kim, and H. J. Kim. Online learning control of hydraulic excavators based on echo-state networks. *IEEE Transactions on Automation Science and Engineering*, 14(1):249–259, Jan 2017.

[371] E. A. Antonelo, C. A. Flesch, and F. Schmitz. Reservoir computing for detection of steady state in performance tests of compressors. University of Luxembourg Open Repository, April 2017. http://hdl.handle.net/10993/30753.

[372] C. Prahm, A. Schulz, B. Paasen, O. Aszmann, B. Hammer, and G. Dorffner. Echo state networks as novel approach for low-cost myoelectric control. In A. ten Telje, J. H. Holmes, L. Sacchi, and C. Popow, editors, *Proceedings of the 16th Conference on Artificial Intelligence in Medicine (AIME 2017)*, 2017.

[373] H. H. Kim and J. Jeong. Representations of directions in EEG-BMI using winner-take-all readouts. In *2017 5th International Winter Conference on Brain-Computer Interface (BCI)*, pages 121–122, Jan 2017.

[374] Z. H. Khan, N. Hussain, and M. I. Tiwana. Classification of EEG signals for wrist and grip movements using echo state network. *Biomedical Research*, 28:1095 – 1102, 2017.

[375] K. Szalisznyo, D. Silverstein, M. Teichmann, H. Duffau, and A. Smits. Cortico-striatal language pathways dynamically adjust for syntactic complexity: A computational study. *Brain and Language*, 164:53 – 62, 2017.

[376] L. Bozhkov, P. Koprinkova-Hristova, and P. Georgieva. Reservoir computing for emotion valence discrimination from EEG signals. *Neurocomputing*, 231:28 – 40, 2017. Neural Systems in Distributed Computing and Artificial Intelligence: Selected papers from the PAAMS 2015 conference.

[377] A. Souahlia, A. Belatreche, A. Benyettou, and K. Curran. Blood vessel segmentation in retinal images using echo state networks. In *2017 9th International Conference on Advanced Computational Intelligence (ICACI)*, 2017.

[378] I. Batzianoulis, S. El-Khoury, E. Pirondini, M. Coscia, S. Micera, and A. Billard. EMG-based decoding of grasp gestures in reaching-to-grasping motions. *Robotics and Autonomous Systems*, 91:59 – 70, 2017.

[379] S. Ayyagari. *Reservoir Computing approach to EEG-based Detection of Microsleeps*. PhD thesis, University of Canterbury, 2017.

[380] M. Chen, W. Saad, and C. Yin. Virtual reality over wireless networks: Quality-of-service model and learning-based resource management. arXiv preprint, March 2017. arXiv:1703.04209.

[381] M. Chen, W. Saad, C. Yin, and M. Debbah. Echo state networks for proactive caching in cloud-based radio access networks with mobile users. *IEEE Transactions on Wireless Communications*, PP(99):1–1, 2017.

[382] M. Chen, W. Saad, and C. Yin. Echo state networks for self-organizing resource allocation in LTE-U with uplink-downlink decoupling. *IEEE Transactions on Wireless Communications*, 16(1):3–16, Jan 2017.

[383] Y. Zhao, H. Gao, N. C. Beaulieu, Z. Chen, and H. Ji. Echo state network for fast channel prediction in ricean fading scenarios. *IEEE Communications Letters*, 21(3):672–675, March 2017.

[384] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong. Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience. *IEEE Journal on Selected Areas in Communications*, PP(99):1–1, 2017.

[385] X. Ruan, C. Li, W. Yang, G. Cui, H. Zhu, Z. Zhou, Y. Dai, and X. Shi. Blind sequence detection using reservoir computing. *Digital Signal Processing*, 62:81 – 90, 2017.

[386] Y. Tan, Z. Zhou, X. Ruan, Y. Dai, H. Zhu, and Y. Zhang. Signal detection for polarization multiplexing QAM optical coherent receivers using ESN. In *Proc. SPIE*, volume 10244, pages 1024402–1024402–6, 2017.

[387] E. A. Antonelo, E. Camponogara, and B. Foss. Echo state networks for data-driven downhole pressure estimation in gas-lift oil wells. *Neural Networks*, 85:106 – 117, 2017.

[388] Z. Peng, J. Wang, and D. Wang. Distributed containment maneuvering of multiple marine vessels via neurodynamics-based output feedback. *IEEE Transactions on Industrial Electronics*, 64(5):3831–3839, May 2017.

[389] L. Liu, D. Wang, Z. Peng, and T. Li. Modular adaptive control for LOS-based cooperative path maneuvering of multiple underactuated autonomous surface vehicles. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(7):1–12, 2017.

[390] D. Bacciu, P. Barsocchi, S. Chessa, C. Gallicchio, and A. Micheli. An experimental characterization of reservoir computing in ambient assisted living applications. *Neural Computing and Applications*, 24(6):1451–1464, 2014.

[391] D. Bacciu, C. Gallicchio, A. Micheli, S. Chessa, and P. Barsocchi. Predicting user movements in heterogeneous indoor environments by reservoir computing. In *Proceedings of the IJCAI Workshop on Space, Time and Ambient Intelligence (STAMI)*, pages 1–6, 2011.

[392] F. Palumbo, P. Barsocchi, C. Gallicchio, S. Chessa, and A. Micheli. Multisensor data fusion for activity recognition based on reservoir computing. In *Evaluating AAL Systems Through Competitive Benchmarking*, pages 24–35. Springer, 2013.

[393] F. Palumbo, C. Gallicchio, R. Pucci, and A. Micheli. Human activity recognition using multisensor data fusion based on reservoir computing. *Journal of Ambient Intelligence and Smart Environments*, 8(2):87 – 107, 2016.

[394] S. Scherer, F. Schwenker, N. Campbell, and G. Palm. Multimodal laughter detection in natural discourses. In H. Ritter, G. Sagerer, R. Dillmann, and M. Buss, editors, *Human Centered Robot Systems*, volume 6 of *Cognitive Systems Monographs*, pages 111–120. Springer Berlin Heidelberg, 2009.

[395] S. Scherer, M. Oubbati, F. Schwenker, and G. Palm. Real-time emotion recognition from speech using echo state networks. In Lionel Prevost, Simone Marinai, and Friedhelm Schwenker, editors, *Artificial Neural Networks in Pattern Recognition*, volume 5064 of *Lecture Notes in Computer Science*, pages 205–216. Springer Berlin Heidelberg, 2008.

[396] S. Scherer, M. Oubbati, F. Schwenker, and G. Palm. Real-time emotion recognition using echo state networks. In E. Andre, L. Dybkjaer, W. Minker, H. Neumann, R. Pieraccini, and M. Weber, editors, *Perception in Multimodal Dialogue Systems*, volume 5078 of *Lecture Notes in Computer Science*, pages 200–204. Springer Berlin Heidelberg, 2008.

[397] S. Scherer, V. Fritzsch, F. Schwenker, and N. Campbell. Demonstrating laughter detection in natural discourses. In *Interdisciplinary Workshop on Laughter and other Interactional Vocalisations in Speech*, 2009.

[398] B. Bacic. *Echo State Network for 3D Motion Pattern Indexing: A Case Study on Tennis Forehands*, pages 295–306. Springer International Publishing, Cham, 2016.

[399] M. Li, B. Lv, W. Dong, and D. Wang. Model of multi-sensor data fusion and trajectory prediction based on echo state network. In *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE)*, volume 1, pages 338–341, Aug 2010.

[400] H. Shao, N. Japkowicz, R. Abielmona, and R. Falcon. Vessel track correlation and association using fuzzy logic and echo state networks. In *2014 IEEE Congress on Evolutionary Computation (CEC)*, pages 2322–2329, July 2014.

[401] C.-Y. Tsai, X. Dutoit, K.-T. Song, H. Van Brussel, and M. Nuttin. Visual state estimation using self-tuning kalman filter and echo state network. In *IEEE International Conference on robotics and Automation, 2008. ICRA 2008.*, pages 917–922, May 2008.

[402] L. Qin and B. Lei. Distributed multiagent for NAO robot joint position control based on echo state network. *Mathematical Problems in Engineering*, 2015:7, 2015.

[403] R. Kanoi and C. Hartland. Fall detections in human walk patterns using reservoir computing based control architecture. In *5th National conference on Control Architecture of Robots*, 2010.

[404] J. Sun, H. Li, and B. Xu. Prognostic for hydraulic pump based upon DCT-composite spectrum and the modified echo state network. *SpringerPlus*, 5(1):1–17, 2016.

[405] Y. Peng, H. Wang, J. Wang, D. Liu, and X.. Peng. A modified echo state network based remaining useful life estimation approach. In *IEEE Conference on Prognostics and Health Management (PHM)*, pages 1–7, June 2012.

[406] S. M. Ruffing and G. K. Venayagamoorthy. Short to medium range time series prediction of solar irradiance using an echo state network. In *15th International Conference on Intelligent System Applications to Power Systems, 2009. ISAP '09.*, pages 1–6, Nov 2009.

[407] D. Liu, J. Wang, and H. Wang. Short-term wind speed forecasting based on spectral clustering and optimised echo state networks. *Renewable Energy*, 78:599 – 608, 2015.

[408] Highways Agency Standard BD 44/95 Volume 3, Section 4, Part 14: the assessment of concrete highway bridges and structures, 2008.

[409] En 1992-1-1: Eurocode 2: Design of concrete structures - part 1-1: General rules and rules for buildings.

[410] O. S. Ekolu and F. Solomon. Statistical analysis of concrete cover in new highway bridges. *IOP Conference Series: Materials Science and Engineering*, 96(1):012081, 2015.

[411] B. Hillemeier and H. Scheel. Magnetic detection of prestressing steel fractures in prestressed concrete. *Materials and Corrosion*, 49(11):799–804, 1998.

[412] E. Billauer. peakdet: Peak detection using MATLAB. Online, 2012.

[413] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861 – 874, 2006. ROC Analysis in Pattern Recognition.

[414] A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[415] J. Huang and C. X. Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3):299–310, 2005.

[416] R. Jarvis, P. Cawley, and P. B. Nagy. Performance evaluation of a magnetic field measurement NDE technique using a model assisted probability of detection framework. *NDT & E International*, 91:61 – 70, 2017.

[417] C. Liu, J. Dobson, and P. Cawley. Efficient generation of receiver operating characteristics for the evaluation of damage detection in practical structural health monitoring applications. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 473(2199), 2017.

[418] Proceq. *Profoscope Operating Instructions*, 2013.

[419] Elcometer. *Elcometer 331 Concrete Covermeter Model THD: Operating Instructions*, 2012.

[420] Kolectric Research. *MC8022 Covermeter*, 2014.

[421] R Barnes and T. Zheng. Research on factors affecting concrete cover measurement. *NDT.net - The e-Journal of Nondestructive Testing*, December 2008.

[422] J. R. Hoki. Analysis of selected factors affecting concrete cover measurements on bridge decks. Master's thesis, Brigham Young University, 2011.

[423] D. Verstraeten, B. Schrauwen, P. Brakel, M. Fiers, S. Dieleman, F. O'Donnell, and D. Pecevski. Oger, 2012. `http://mloss.org/software/view/419/`.

[424] R. Ohayon and M. Bernadou. Model-based structural damage detection: Present status and challenges. In *Tenth International Conference on Adaptive Structures and Technologies*, 2000.

[425] K. Worden and J. M. Dulieu-Barton. An overview of intelligent fault detection in systems and structures. *Structural Health Monitoring*, 3(1):85–98, 2004.

[426] A. Rytter. *Vibrational Based Inspection of Civil Engineering Structures*. PhD thesis, Denmark, 1993. Ph.D.-Thesis defended publicly at the University of Aalborg, April 20, 1993 PDF for print: 206 pp.

[427] A. J. Wootton, C. R. Day, and P. W. Haycock. An echo state network approach to structural health monitoring. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, July 2015.

[428] J. B. Butcher, T. K. Kyriacou, C. R. Day, and P. W. Haycock. Detection of anomalies present in reinforced concrete bridge data using NARMAX and echo state networks. Unpublished, composed for an internal meeting with the NPL, 2012.

[429] E. N. Barton and B. Zhang. Details of temperature compensation for strain measurements on NPL bridge - demonstrator for SHM. *Applied Mechanics and Materials*, 24 - 25:173–178, 2010.

[430] E. Barton, C. Middleton, K. Koo, L. Crocker, and J. Brownjohn. Structural finite element model updating using vibration tests and modal analysis for NPL foot-

bridge - SHM demonstrator. *Journal of Physics: Conference Series*, 305(1):012105, 2011.

[431] E. Barton. SHM demonstrator at NPL: Two years of monitoring experience and future challenges. Online, May 2011.

[432] Footbridge monitoring project (SHM) - background. Online.

[433] R. Kromanis and P. Kripakaran. Predicting thermal response of bridges using regression models derived from measurement histories. *Computers & Structures*, 136:64 – 77, 2014.

[434] ITMSOIL. Online, 2016. http://www.itmsoil.com/.

[435] J. M. W. Brownjohn, K.-Y. Koo, A. Scullion, and D. List. Operational deformations in long-span bridges. *Structure and Infrastructure Engineering*, 11(4):556–574, 2015.

[436] C. R. Farrar and K. Worden. *Structural Health Monitoring: A Machine Learning Perspective*, chapter 2, pages 17 – 44. John Wiley & Sons, Ltd, 2013.

[437] A. Gastineau, T. Johnson, and A. Schultz. Bridge health monitoring and inspections - a survey of methods. Technical Report MN/RC 2009-29, Minnesota Department of Transportation, 2009.

[438] V. Livina, E. Barton, and A. Forbes. Tipping point analysis of the NPL footbridge. *Journal of Civil Structural Health Monitoring*, 3:1–8, 2013.

[439] E. Barton and T. Esward. The origins of measurement uncertainty in SHM - NPL footbridge case study. In *6th European Workshop on Structural Health Monitoring*, 2012.

[440] R. Kromanis. *Structural Performance Evaluation of Bridges: Characterizing and Integrating Thermal Response*. PhD thesis, Universty of Exeter, 2015.

[441] K. Worden, E. Cross, and E. Barton. Damage detection on the NPL footbridge under changing environmental conditions. In *6th European Workshop on Structural Health Monitoring*, 2012.

[442] K. Worden, T. Baldacchino, J. Rowson, and E. J. Cross. Some recent developments in SHM based on nonstationary time series analysis. *Proceedings of the IEEE*, 104(8):1589–1603, Aug 2016.

[443] R. J. Patton, J. Chen, and T. M. Siew. Fault diagnosis in nonlinear dynamic systems via neural networks. In *International Conference on Control, 1994. Control '94.*, volume 2, pages 1346–1351, March 1994.

[444] R. J. Patton. Fault detection and diagnosis in aerospace systems using analytical redundancy. *Computing Control Engineering Journal*, 2(3):127–136, May 1991.

[445] H. Sohn. Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1851):539–560, 2007.

[446] T. Hongell, I. Kivela, and I. Hakala. Wireless strain gauge network - best-hall measurement case. In *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1–6, April 2014.

[447] Y. Bao, H. Li, Y. An, and J. Ou. Dempster–shafer evidence theory approach to structural damage detection. *Structural Health Monitoring*, 11(1):13–26, 2012.

[448] Soil Instruments. *ST2 Vibrating Wire Arc Weldable Strain Gauge*, 2017.

[449] C. Lo, J. P. Lynch, and M. Liu. Reference-free detection of spike faults in wireless sensor networks. In *Resilient Control Systems (ISRCS), 2011 4th International Symposium on*, pages 148–153, Aug 2011.

[450] A. W. Skempton. The bearing capacity of clays. In *Selected Papers on Soil Mechanics*. Thomas Telford Ltd, 1984.

[451] B. W. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2):442 – 451, 1975.

[452] M. J. Embrechts, L. A. Alexandre, and J. D. Linton. Reservoir computing for static pattern recognition. In *European Symposium on Artificial Neural Networks (ESANN) 2009*, 2009.

[453] C. Emmerich, R. Reinhart, and J. Steil. Recurrence enhances the spatial encoding of static inputs in reservoir networks. In K. Diamantaras, W. Duch, and L. S. Iliadis, editors, *Artificial Neural Networks - ICANN 2010*, volume 6353 of *Lecture Notes in Computer Science*, pages 148–153. Springer Berlin Heidelberg, 2010.

[454] C. Edwards and S. L. Taylor. A survey and strategic appraisal of rhododendron invasion and control in woodwood areas in Argyll and Bute. Technical report, Perth Conservancy, Forestry Commision Scotland, 2008.

[455] F. Williams, R. Eschen, A. Harris, D. Djeddour, C. Pratt, R. S. Shaw, S. Varia, J. Lamontagne-Godwin, S. E. Thomas, and S. T. Murphy. The economic cost of invasive non-native species on Great Britain. Technical report, CABI, 2010.

[456] M. A. Cho, I. Sobhan, A. K. Skidmore, and J. de Leeuw. Discriminating species using hyperspectral indices at leaf and canopy scales. In *ISPRS 2008 : Proceedings of the XXI congress: Silk road for information from imagery: the International Society for Photogrammetry and Remote Sensing*, pages 369–376, 2008.

[457] K. S. He, D. Rocchini, M. Neteler, and H. Nagendra. Benefits of hyperspectral remote sensing for tracking plant invasions. *Diversity and Distributions*, 17(3):381–392, 2011.

[458] S. L. Taylor, R. A. Hill, and C. Edwards. Characterising invasive non-native rhododendron ponticum spectra signatures with spectroradiometry in the laboratory and field: Potential for remote mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 81:70 – 81, 2013.

[459] L. Wang. Invasive species spread mapping using multi-resolution remote sensing data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:135–142, 2008.

[460] R. Gaulton, G. Olaya, E. D. Wallington, and T. J. Malthus. Continuous cover forestry sensing in the UK? Quantifying forest structure using remote sensing. In *Proceedings of ForestSAT Conference*, 2005.

[461] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 International Joint Conference on Neural Networks (IJCNN)*, volume 2, pages 985–990, July 2004.

[462] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[463] K. Bache and M. Lichman. UCI machine learning repository, 2013.

[464] H. O. Sillin, R. Aguilera, H.-H. Shieh, A. V. Avizienis, M. Aono, A. Z. Stieg, and J. K. Gimzewski. A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing. *Nanotechnology*, 24(38):384004, 2013.

APPENDIX A

All contour plots from Chapter 3

This appendix presents contour plots for the five approaches - the AT, the fusion ESN, ESN_{GD}, ESN_{DS} and ESN_{Prelim} - used in Chapter 3. Since each ESN architecture was produced 500 times and the results averaged, the contour plots for these techniques were produced by the individual ESN that produced the AUC closest to the overall average AUC for that architecture, as detailed in Chapter 3.

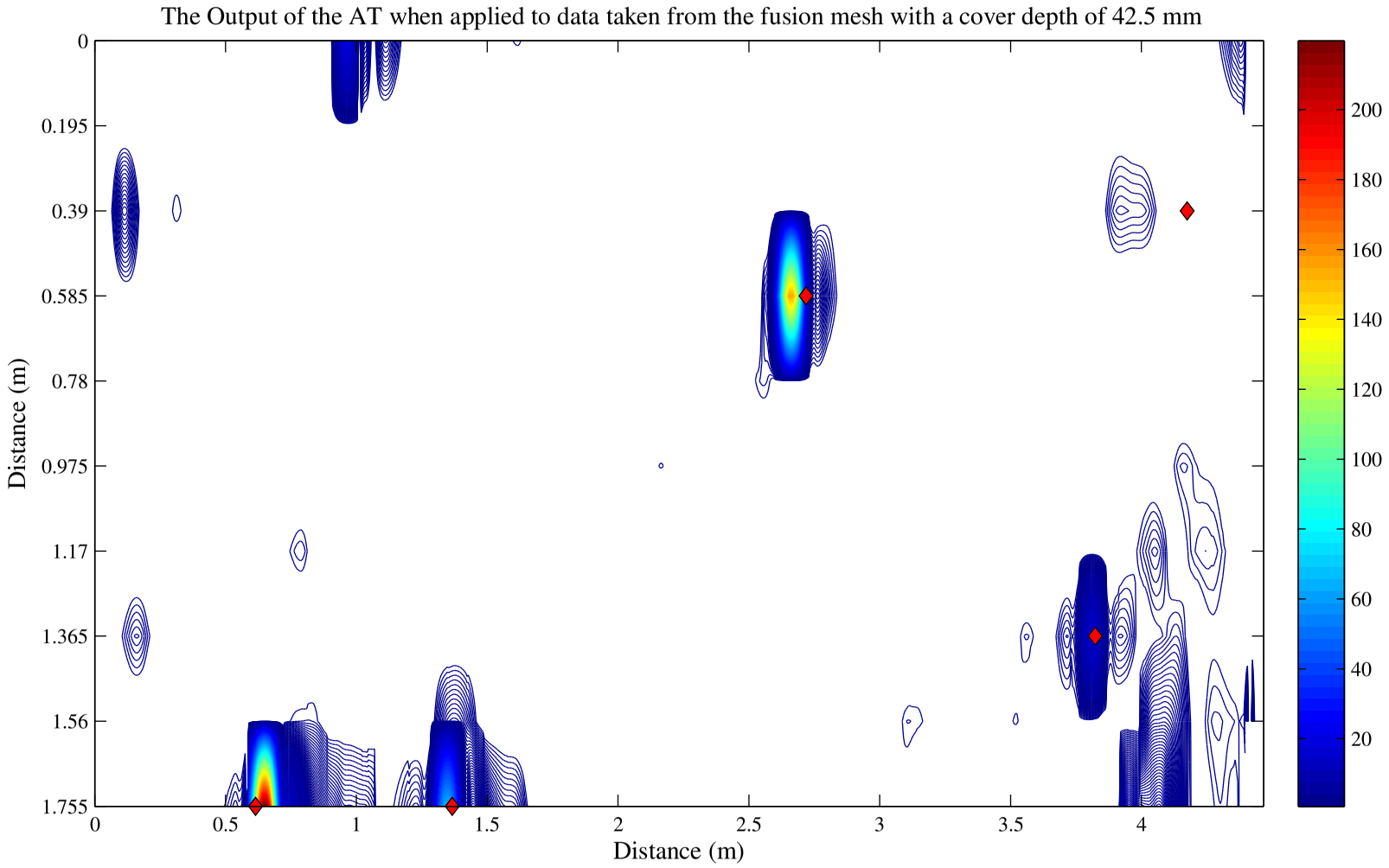


Figure A.1: A contour plot of the output of the AT when applied to data from dataset C2 at a cover depth of 42.5 mm, with the threshold (0.6396) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9882.

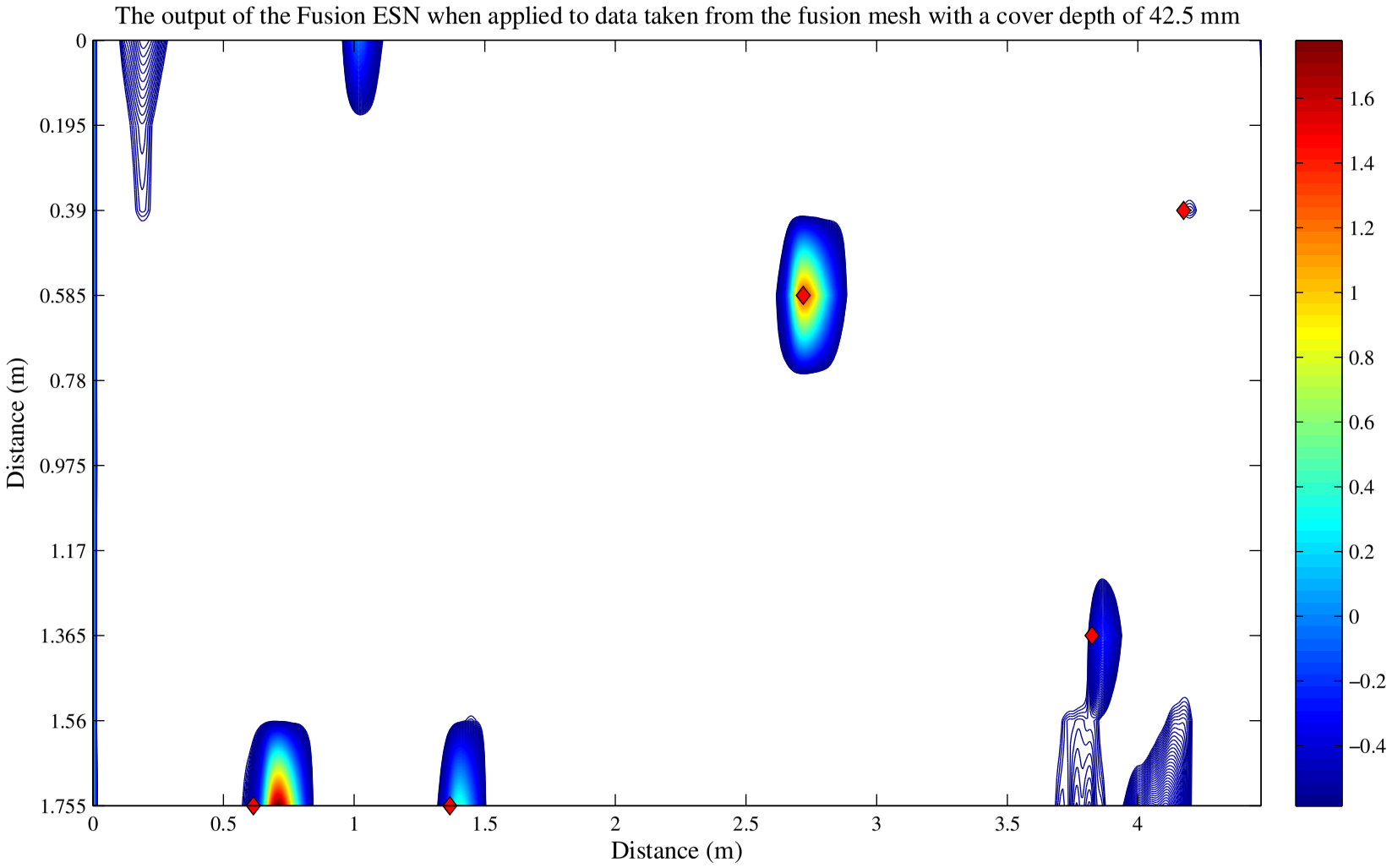


Figure A.2: A contour plot of the output of the fusion ESN when applied to data from dataset C2 at a cover depth of 42.5 mm, with the threshold (-0.5959) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9962.

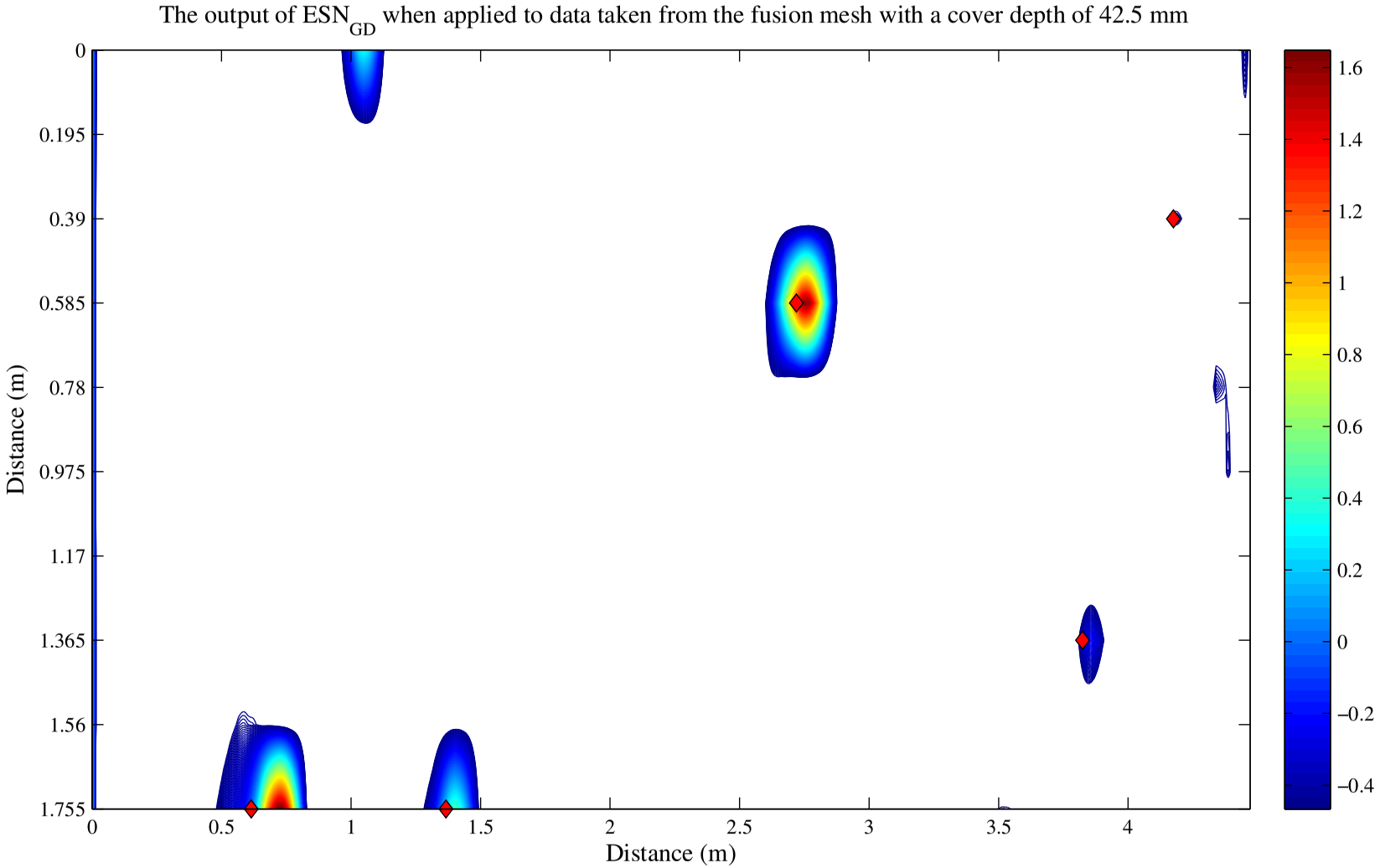


Figure A.3: A contour plot of the output of ESN$_{GD}$ when applied to data from dataset C2 at a cover depth of 42.5 mm, with the threshold (-0.4465) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9984.

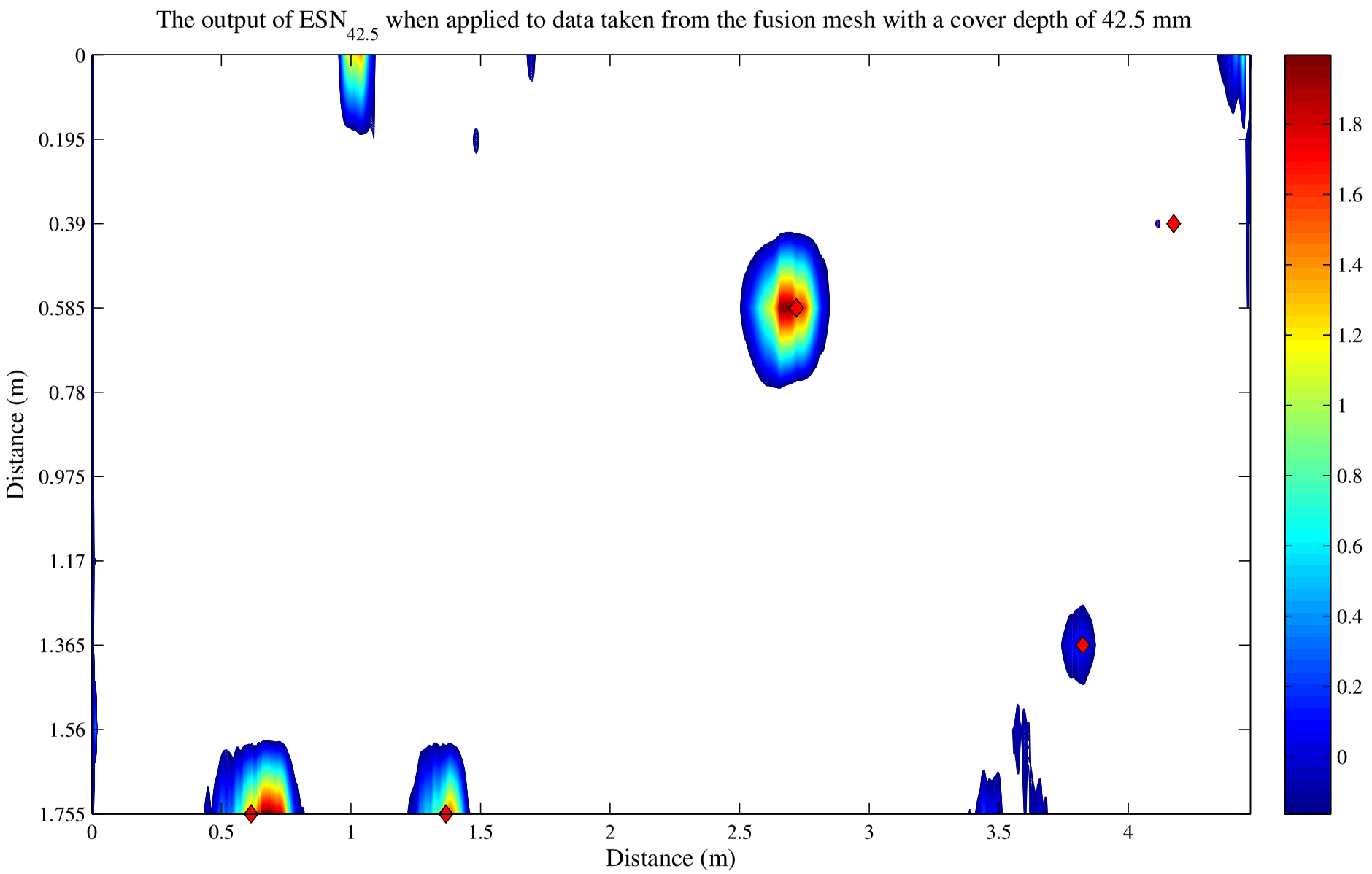


Figure A.4: A contour plot of the output of $ESN_{42.5}$ when applied to data from dataset C2 at a cover depth of 42.5 mm, with the threshold (-0.1432) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9976.

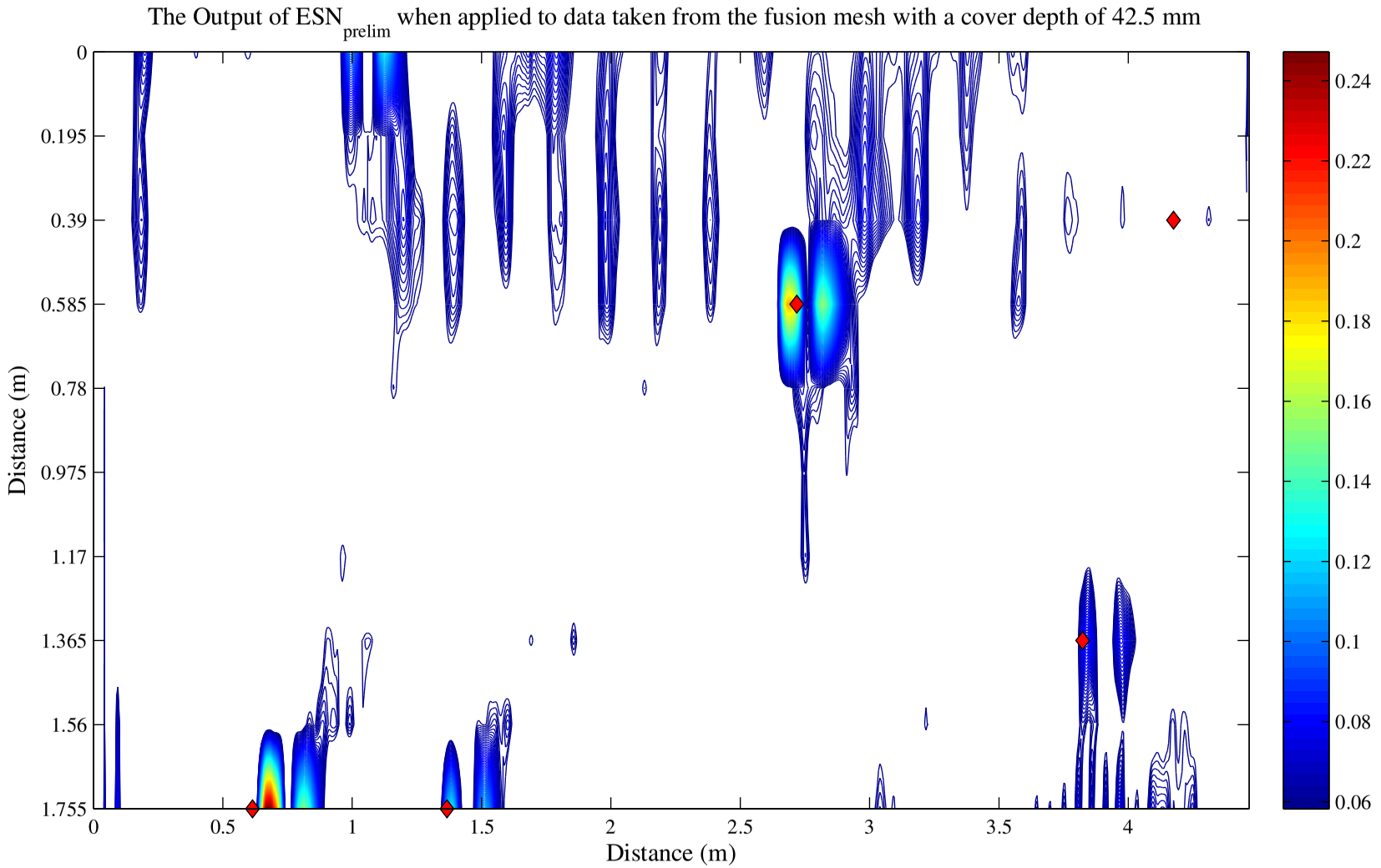


Figure A.5: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 42.5 mm, with the threshold (0.0602) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9914.

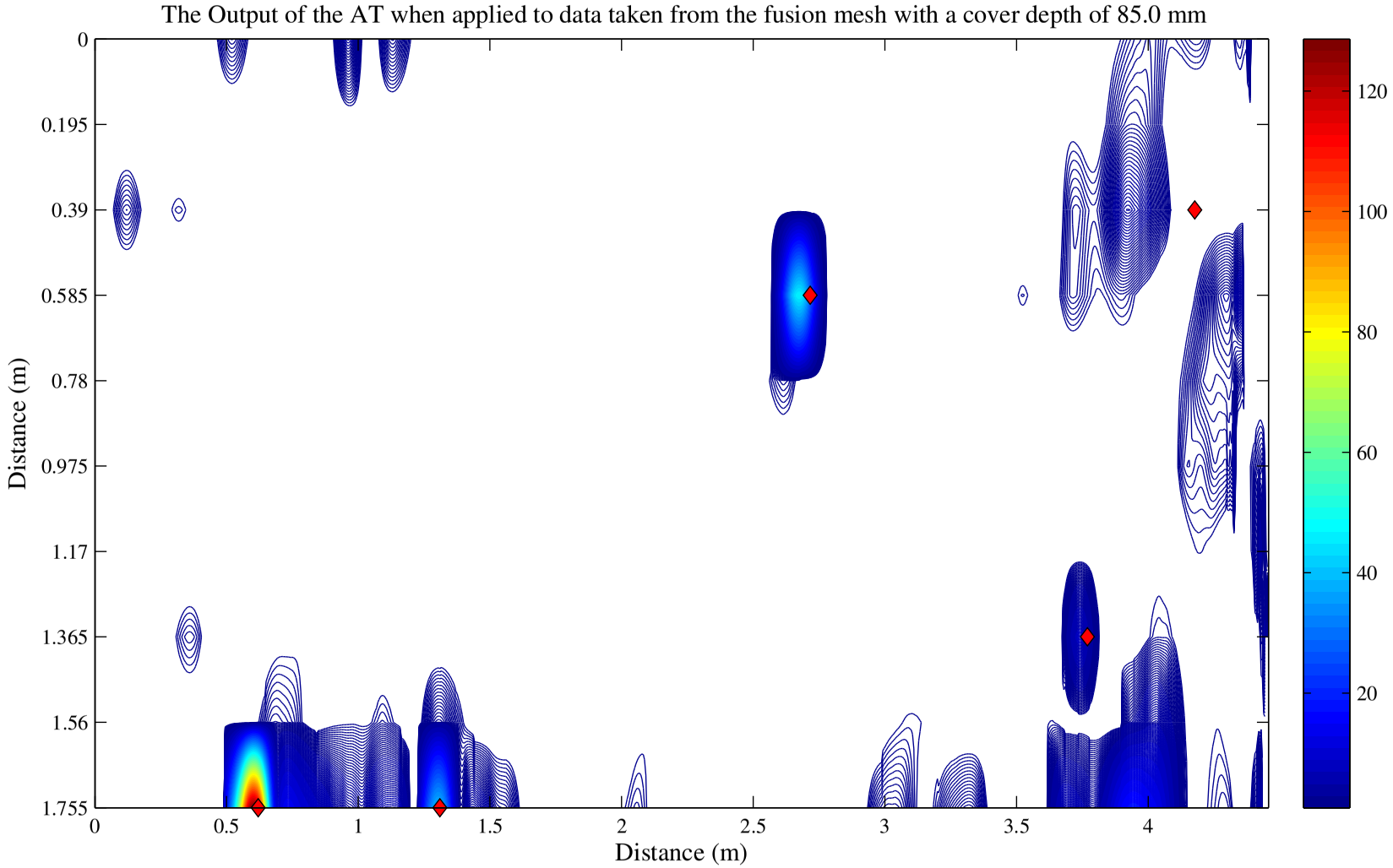


Figure A.6: A contour plot of the output of the AT when applied to data from dataset C2 at a cover depth of 85.0 mm, with the threshold (0.9372) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9742.

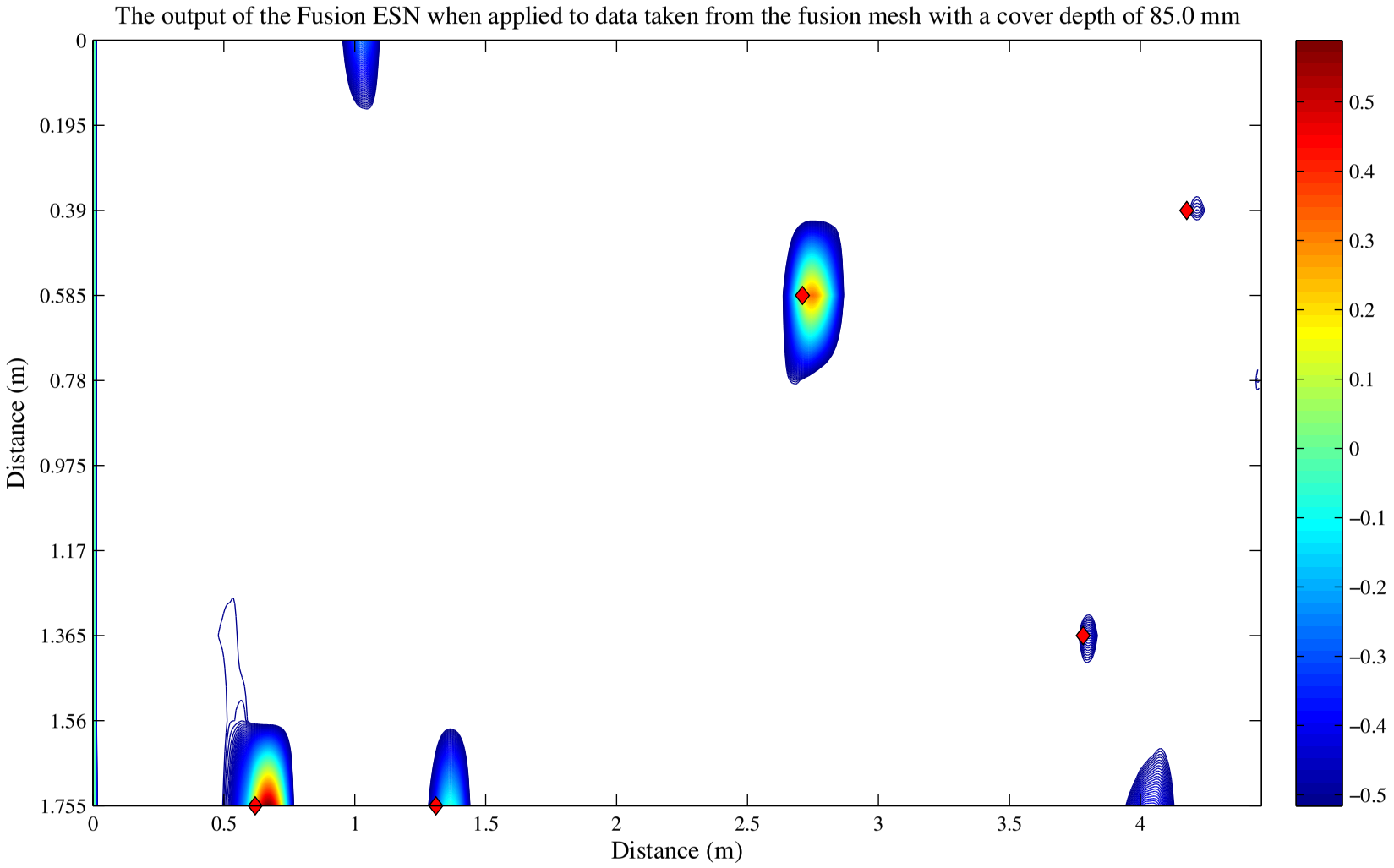


Figure A.7: A contour plot of the output of the fusion ESN when applied to data from dataset C2 at a cover depth of 85.0 mm, with the threshold (-0.4960) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9976.

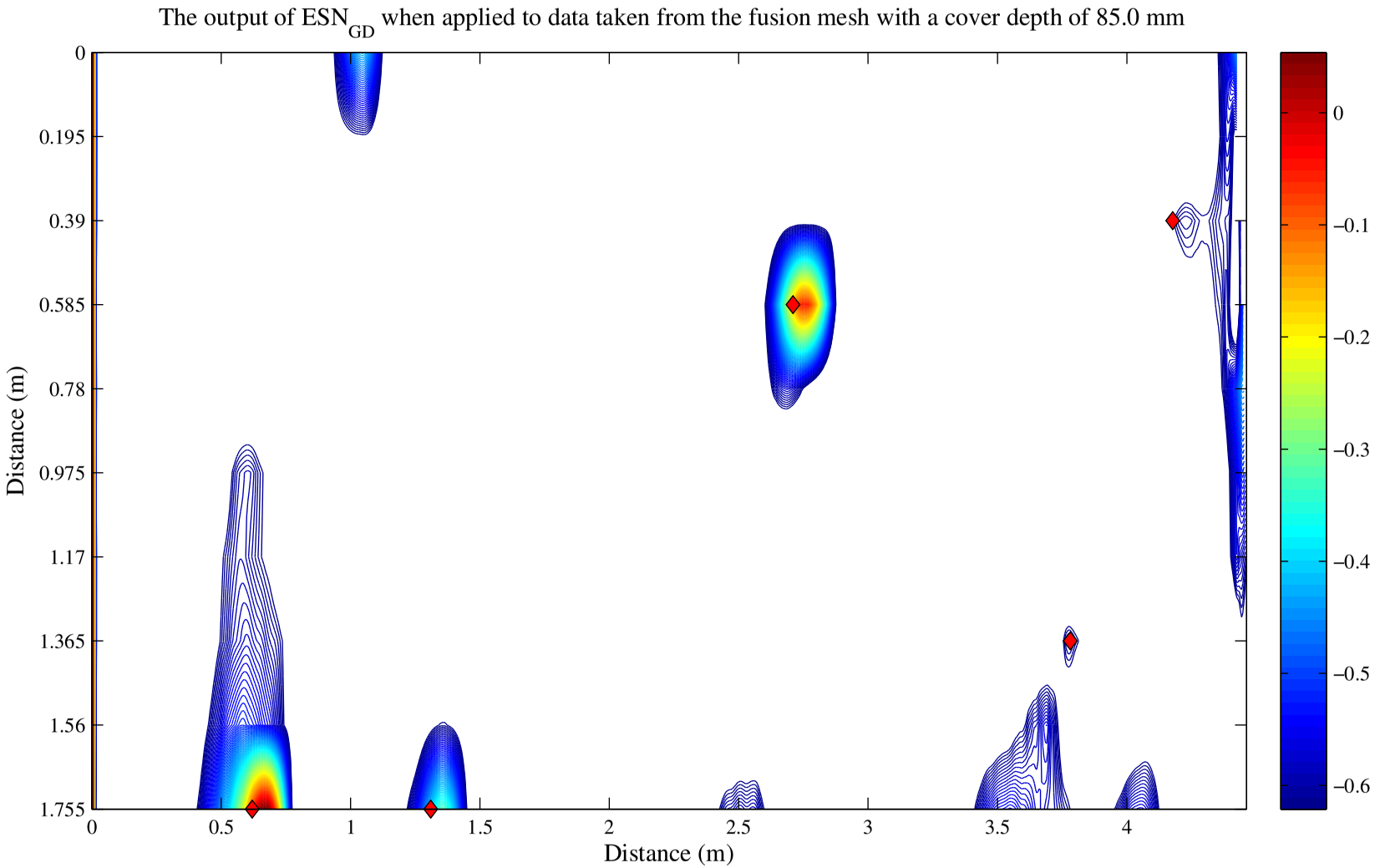


Figure A.8: A contour plot of the output of $ESN_{85.0}$ when applied to data from dataset C2 at a cover depth of 85.0 mm, with the threshold (-0.6012) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9857.

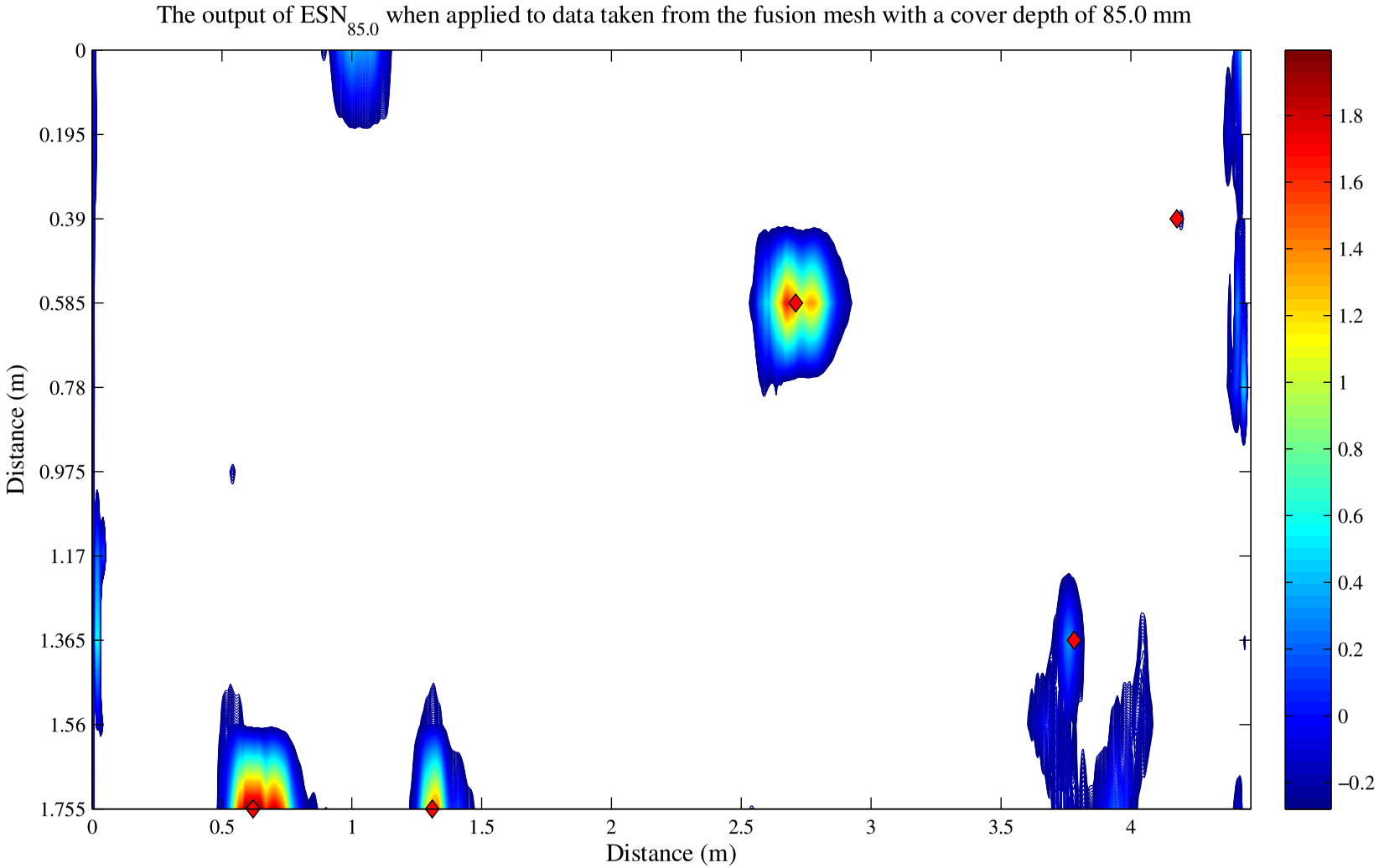


Figure A.9: A contour plot of the output of $ESN_{85.0}$ when applied to data from dataset C2 at a cover depth of 85.0 mm, with the threshold (-0.2595) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9943.

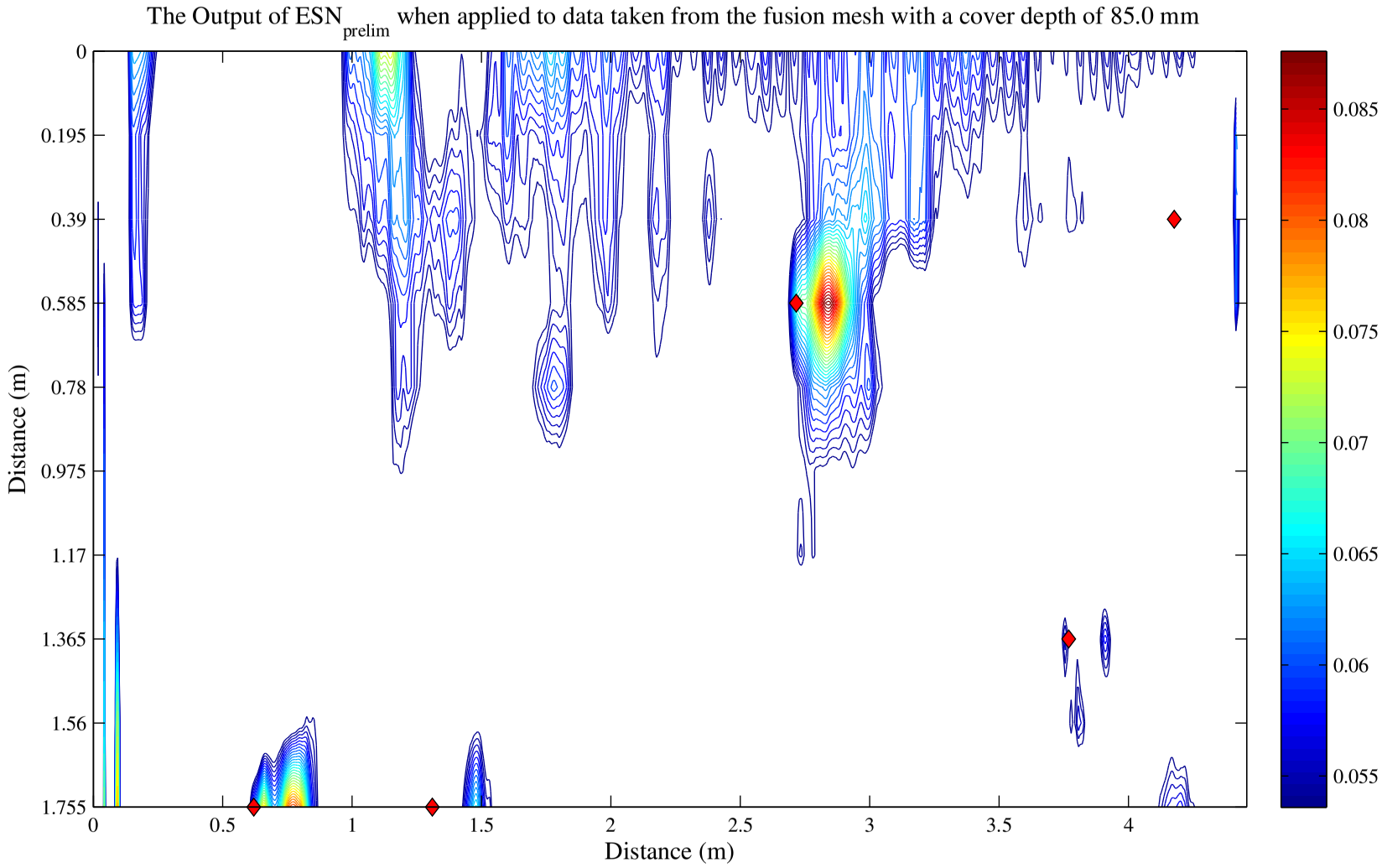


Figure A.10: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 85.0 mm, with the threshold (0.0556) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9429.

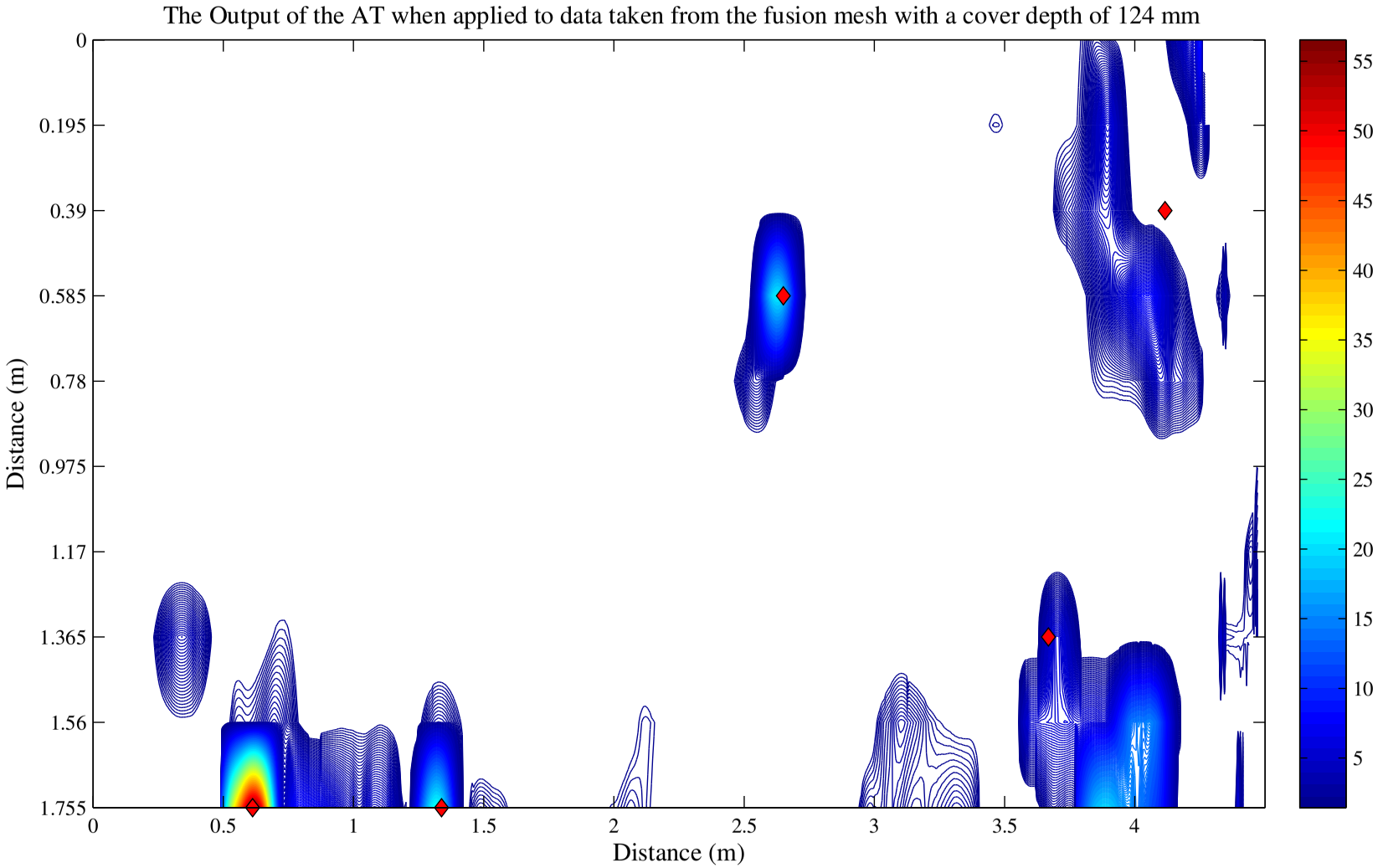


Figure A.11: A contour plot of the output of the AT when applied to data from dataset C2 at a cover depth of 124 mm, with the threshold (1.4279) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9128.

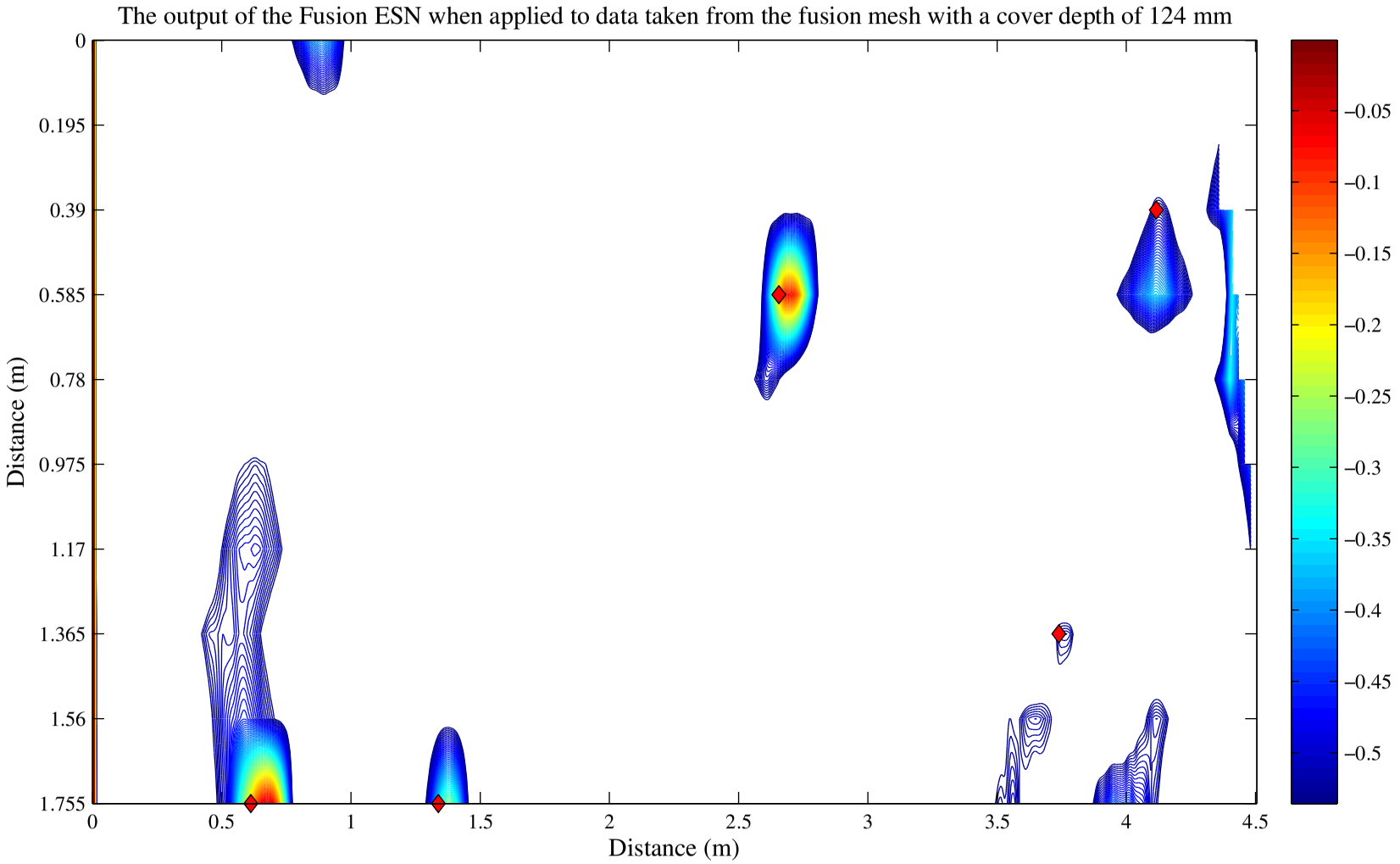


Figure A.12: A contour plot of the output of the fusion ESN when applied to data from dataset C2 at a cover depth of 124 mm, with the threshold (-0.5157) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9871.

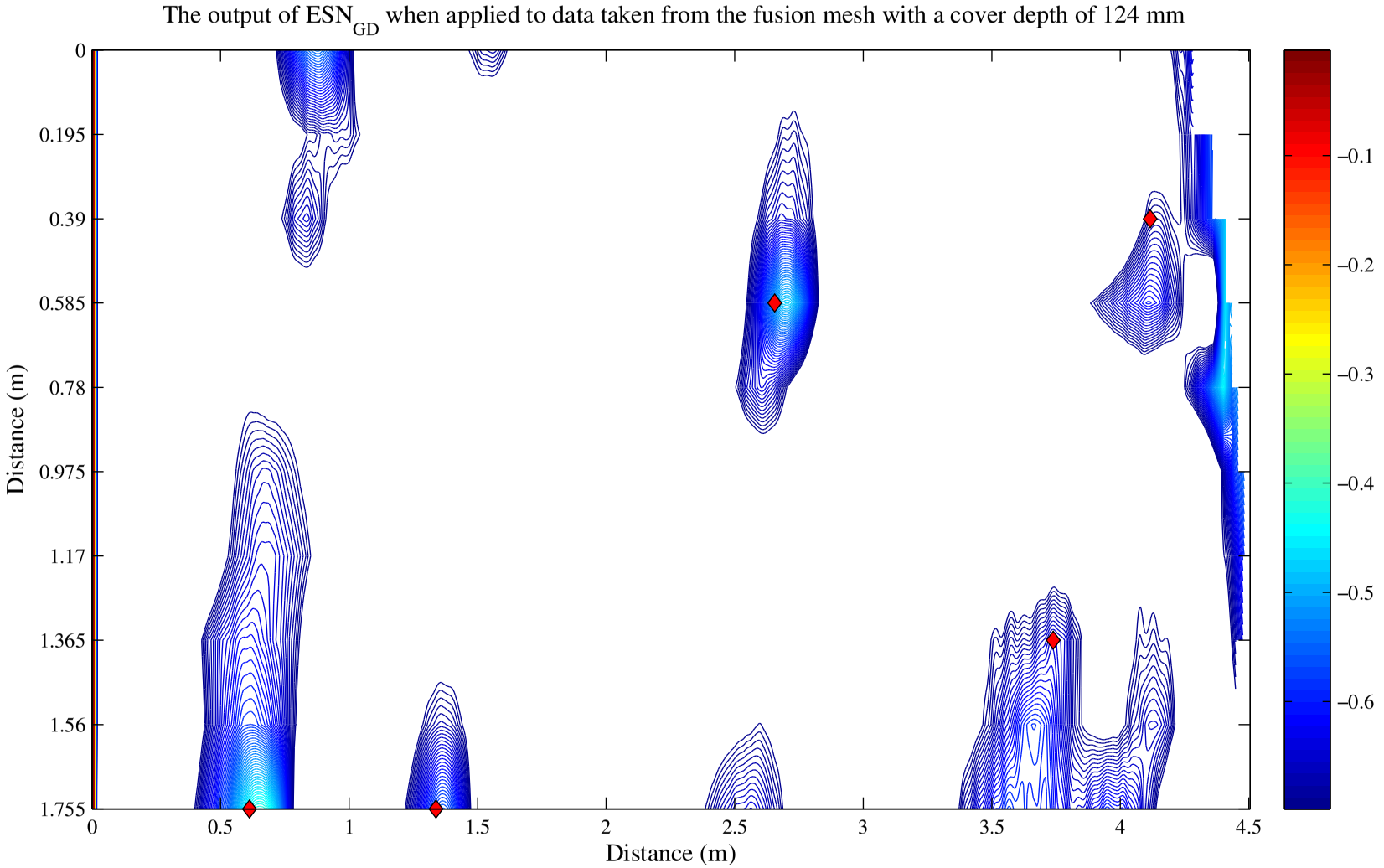


Figure A.13: A contour plot of the output of ESN_{GD} when applied to data from dataset C2 at a cover depth of 124 mm, with the threshold (-0.6785) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9726.

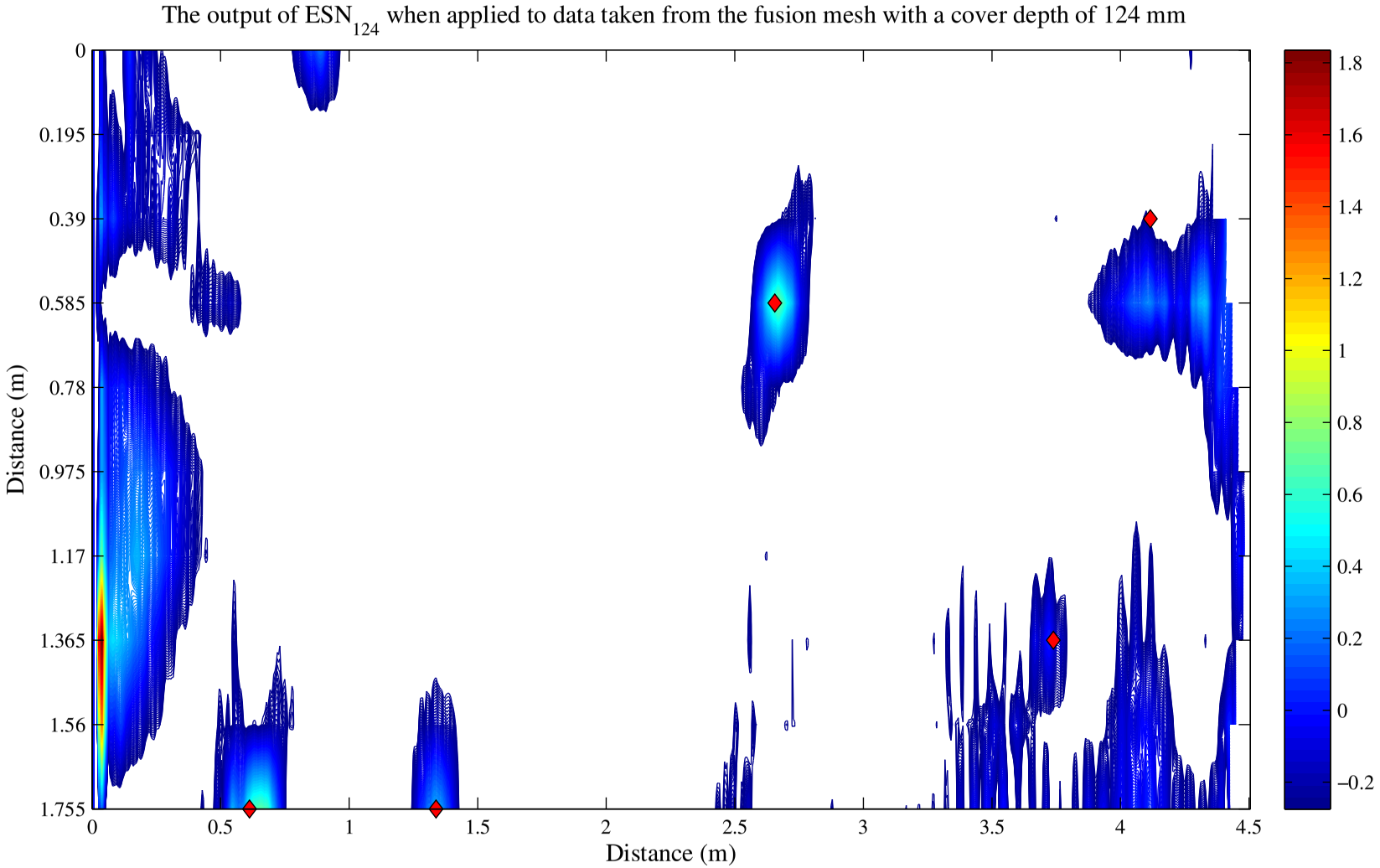


Figure A.14: A contour plot of the output of ESN_{124} when applied to data from dataset C2 at a cover depth of 124 mm, with the threshold (-0.2548) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9776.

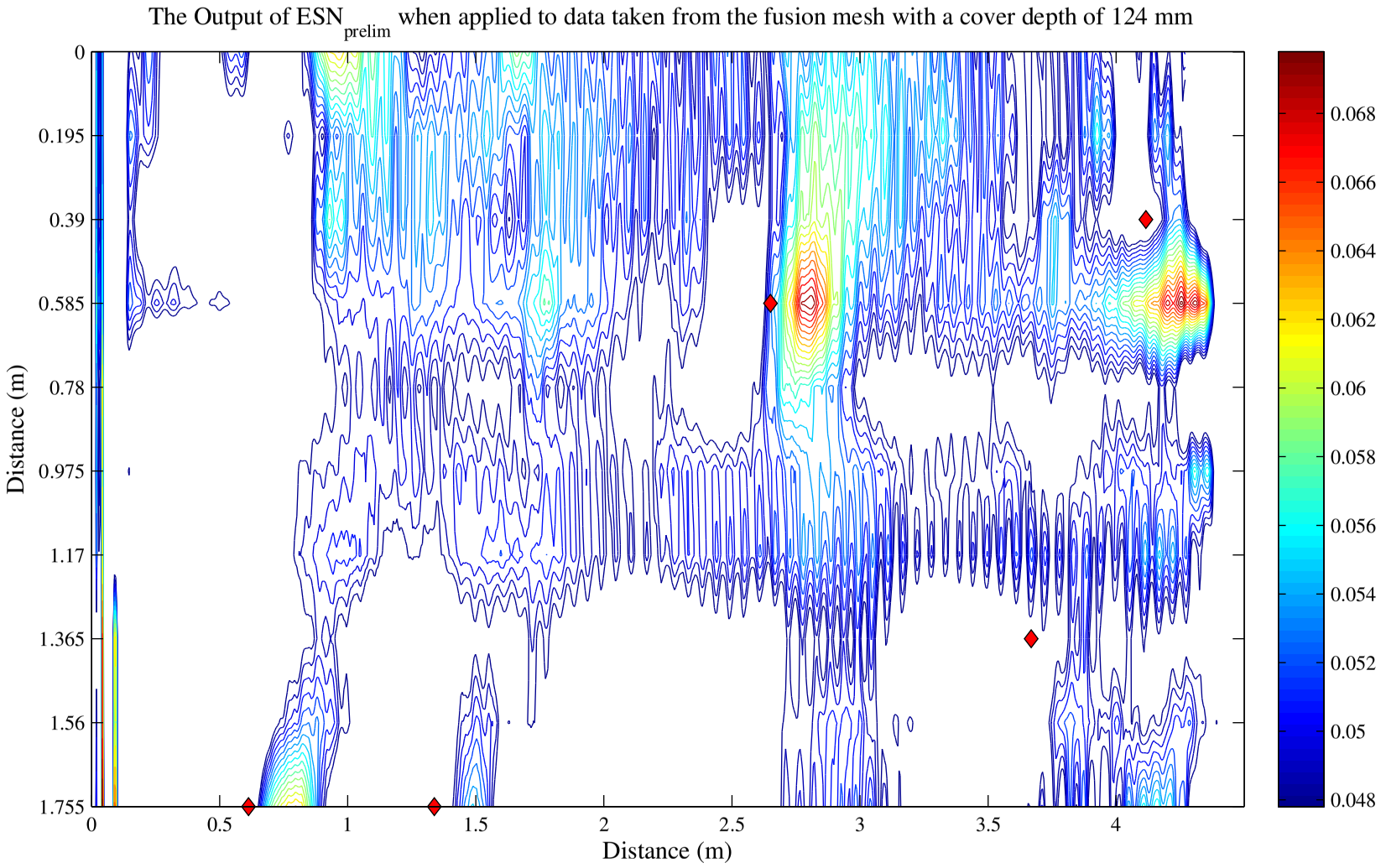


Figure A.15: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 124 mm, with the threshold (0.0478) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.7895.

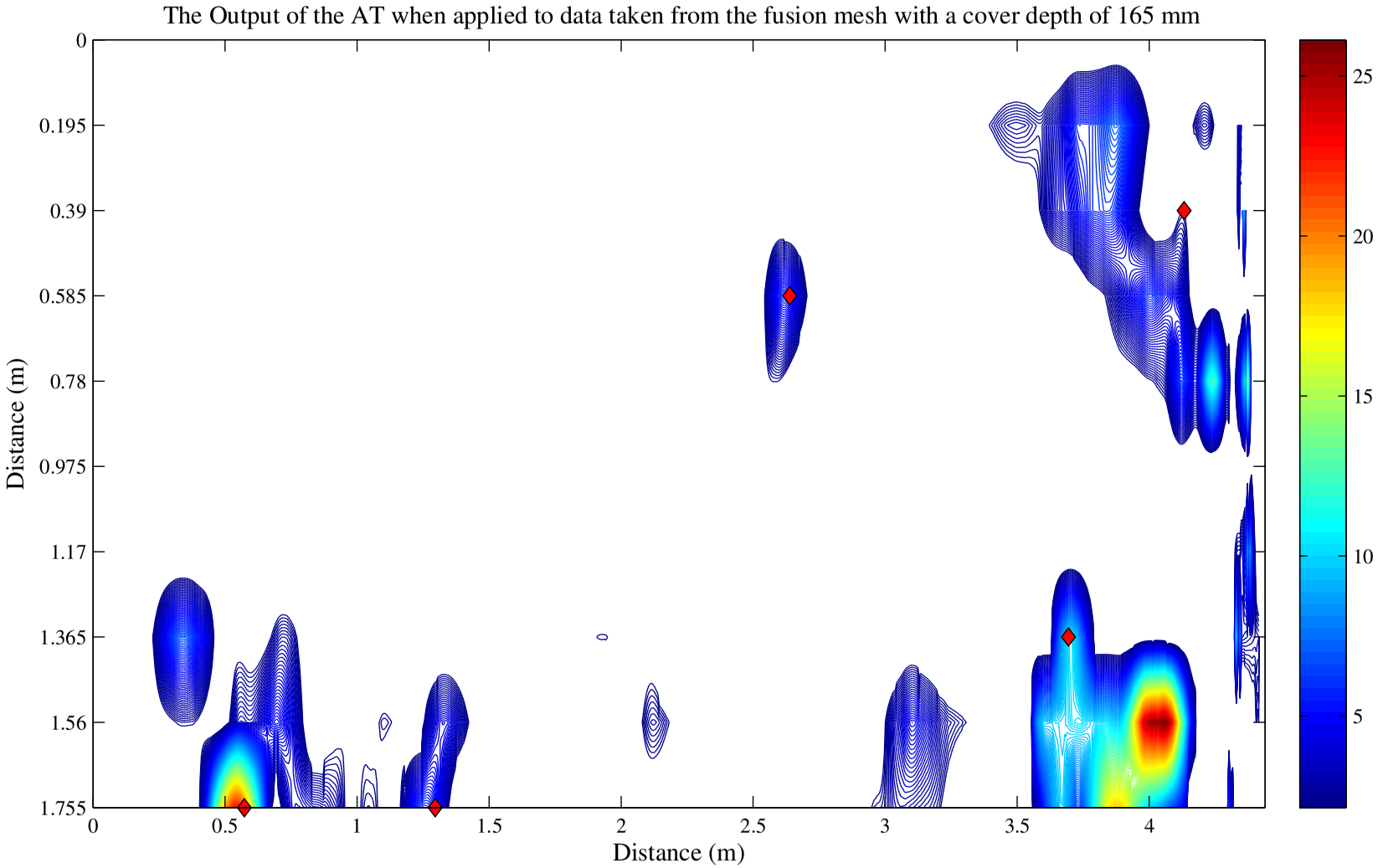


Figure A.16: A contour plot of the output of the AT when applied to data from dataset C2 at a cover depth of 165 mm, with the threshold (2.1293) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9034.

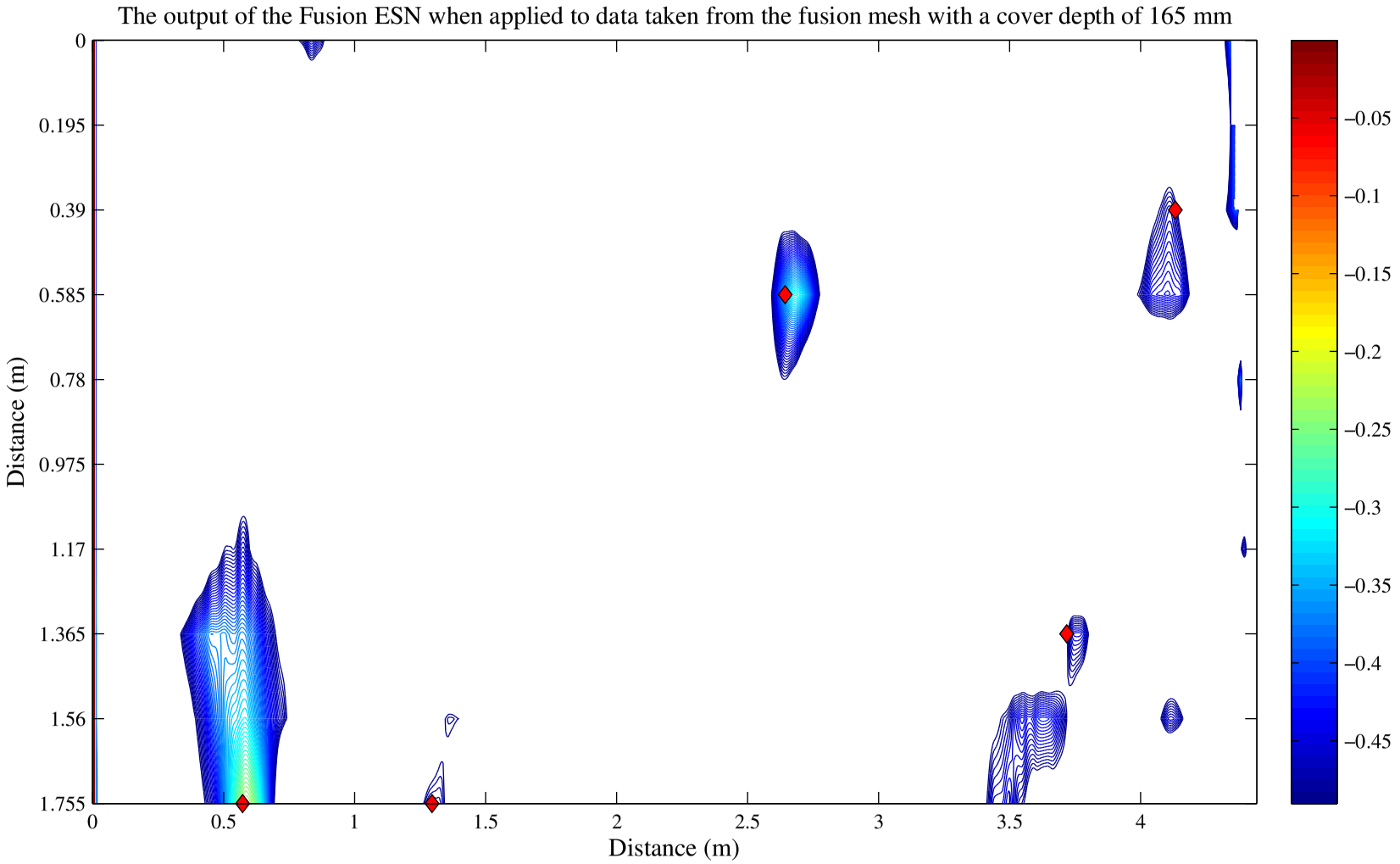


Figure A.17: A contour plot of the output of the fusion ESN when applied to data from dataset C2 at a cover depth of 165 mm, with the threshold (-0.5820) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9279.

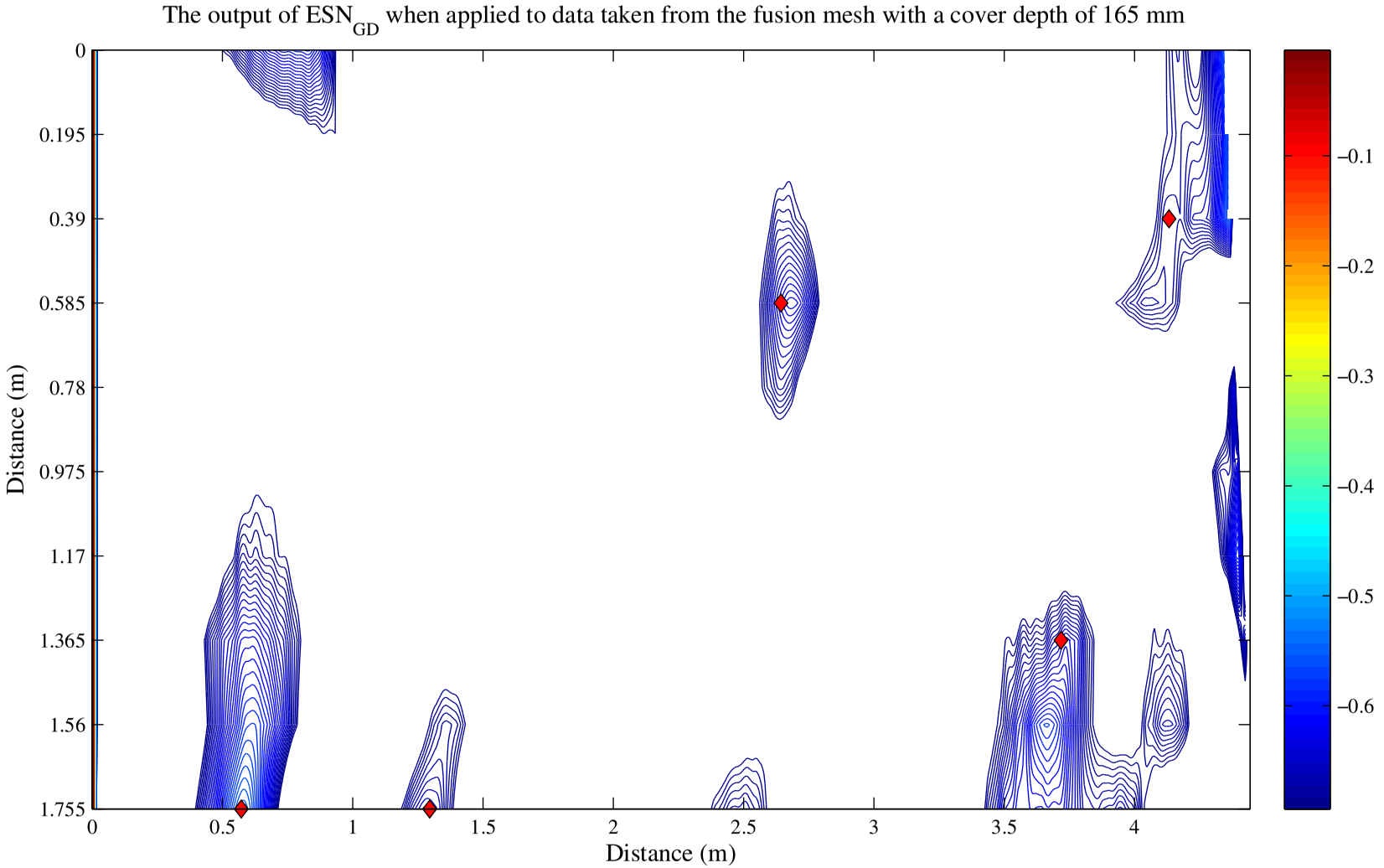


Figure A.18: A contour plot of the output of ESN$_{GD}$ when applied to data from dataset C2 at a cover depth of 165 mm, with the threshold (-0.6739) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9741.

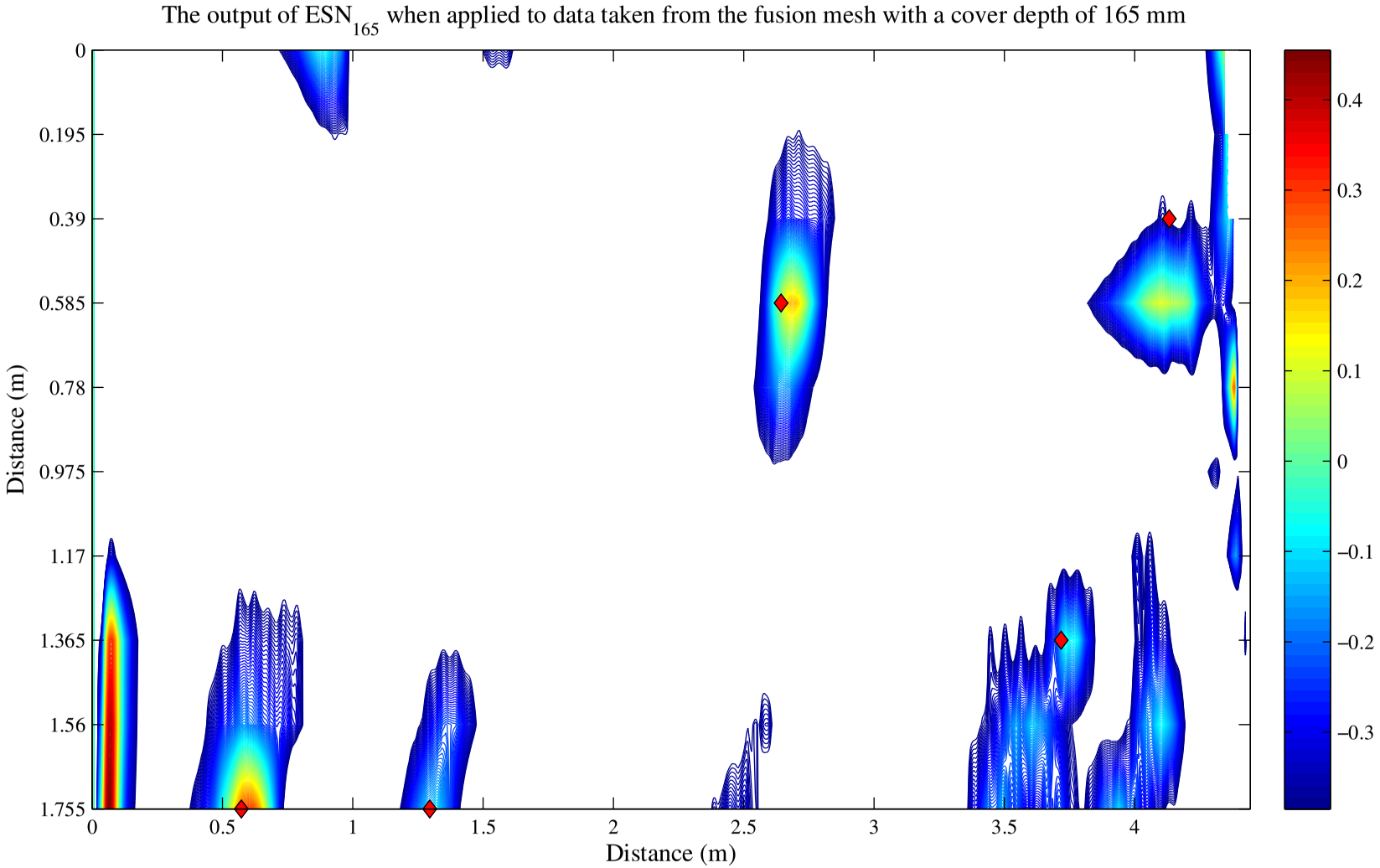


Figure A.19: A contour plot of the output of ESN_{165} when applied to data from dataset C2 at a cover depth of 165 mm, with the threshold (-0.3651) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9801.

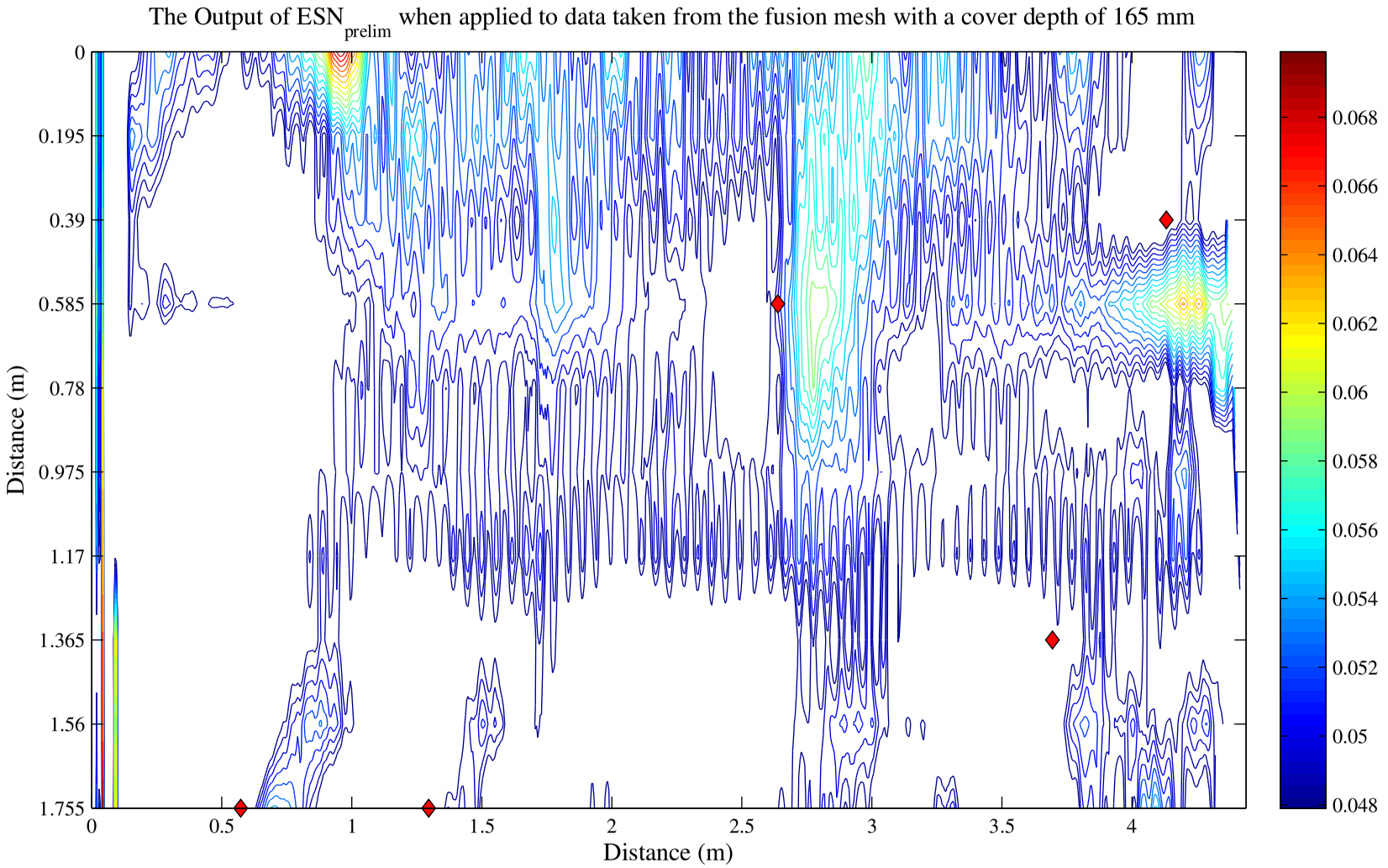


Figure A.20: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 165 mm, with the threshold (0.0479) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.7501.

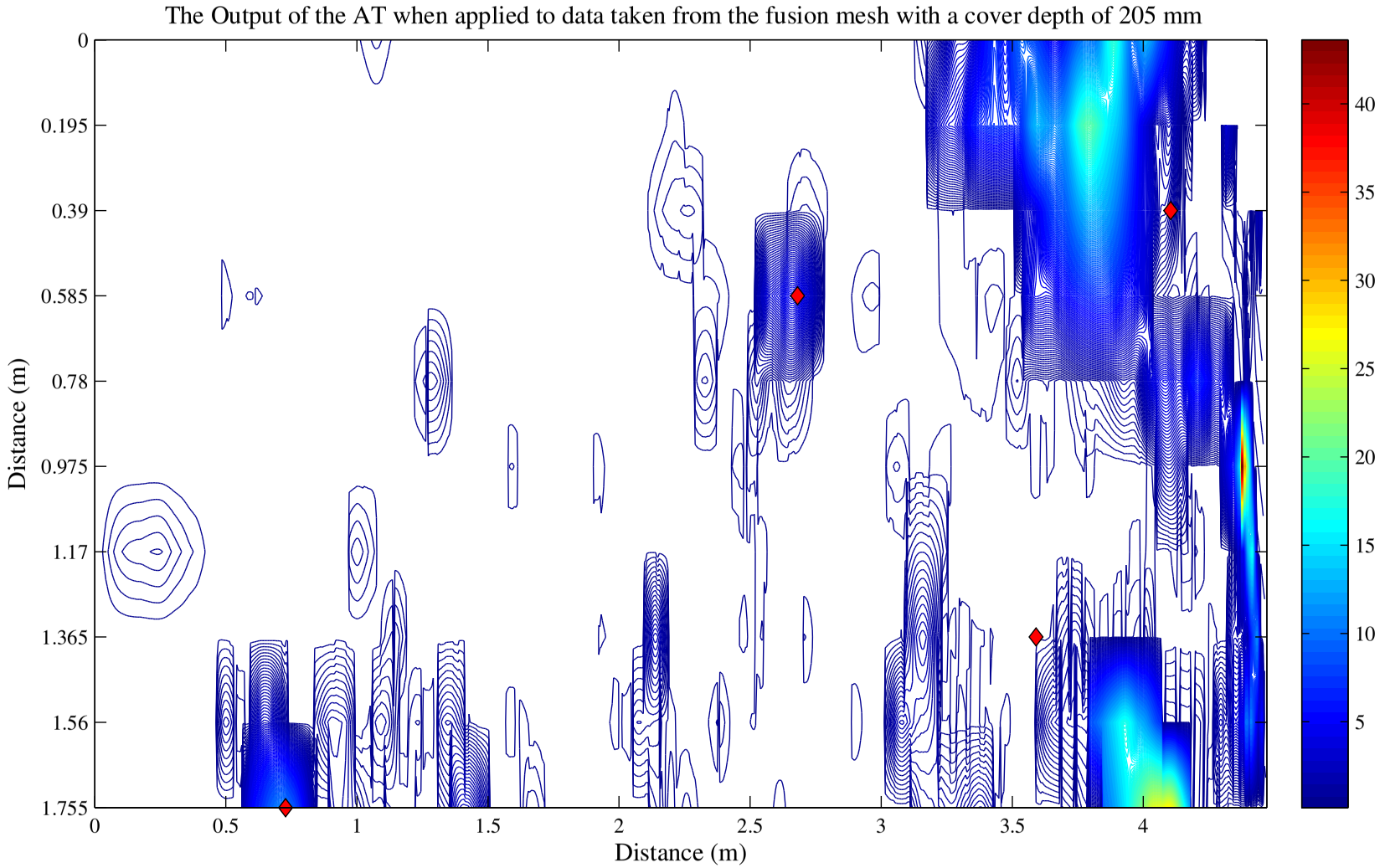


Figure A.21: A contour plot of the output of the AT when applied to data from dataset C2 at a cover depth of 205 mm, with the threshold (0.1240) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8083.

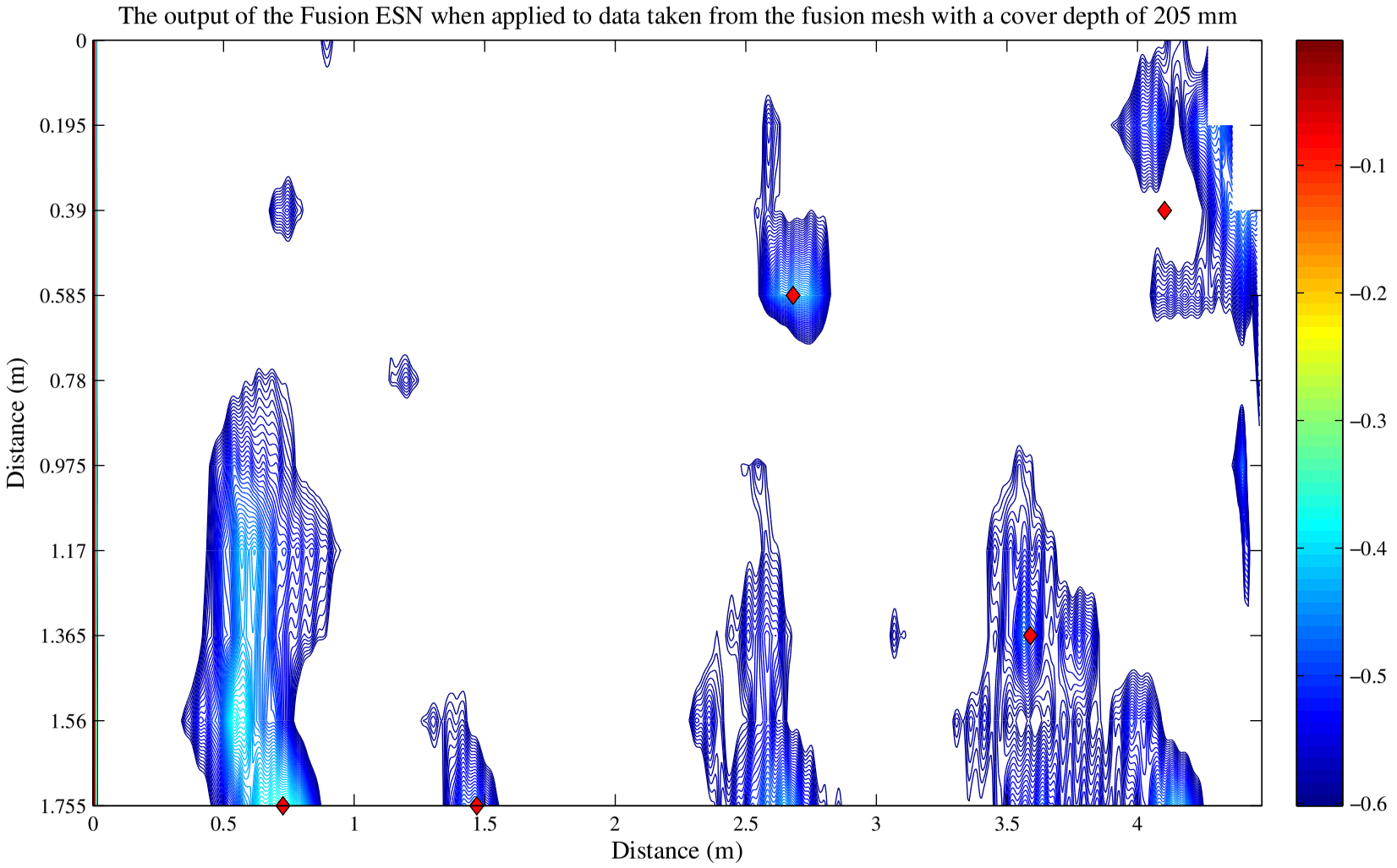


Figure A.22: A contour plot of the output of the fusion ESN when applied to data from dataset C2 at a cover depth of 205 mm, with the threshold (-0.5820) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9279.

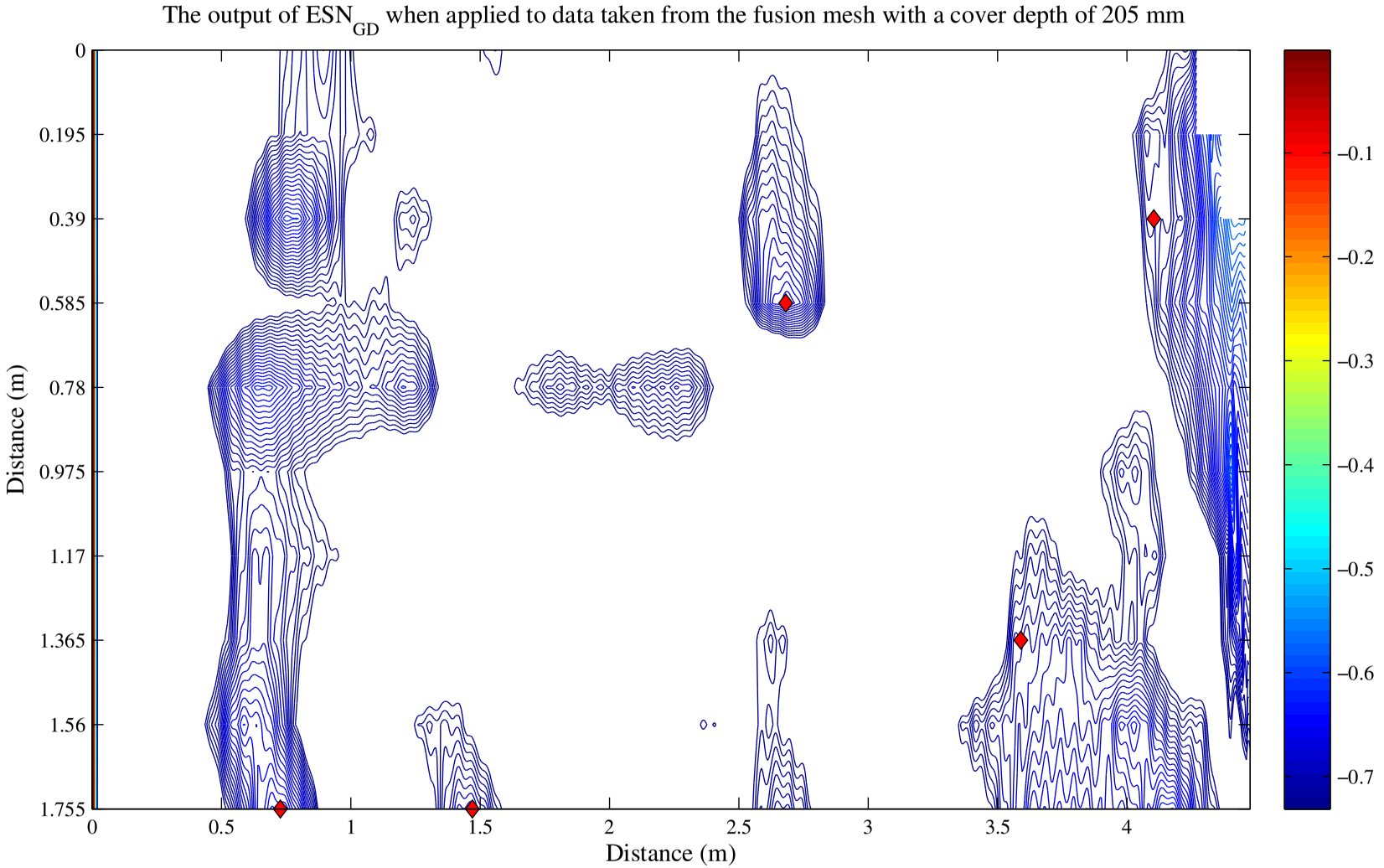


Figure A.23: A contour plot of the output of ESN$_{GD}$ when applied to data from dataset C2 at a cover depth of 205 mm, with the threshold (-0.7112) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9203.

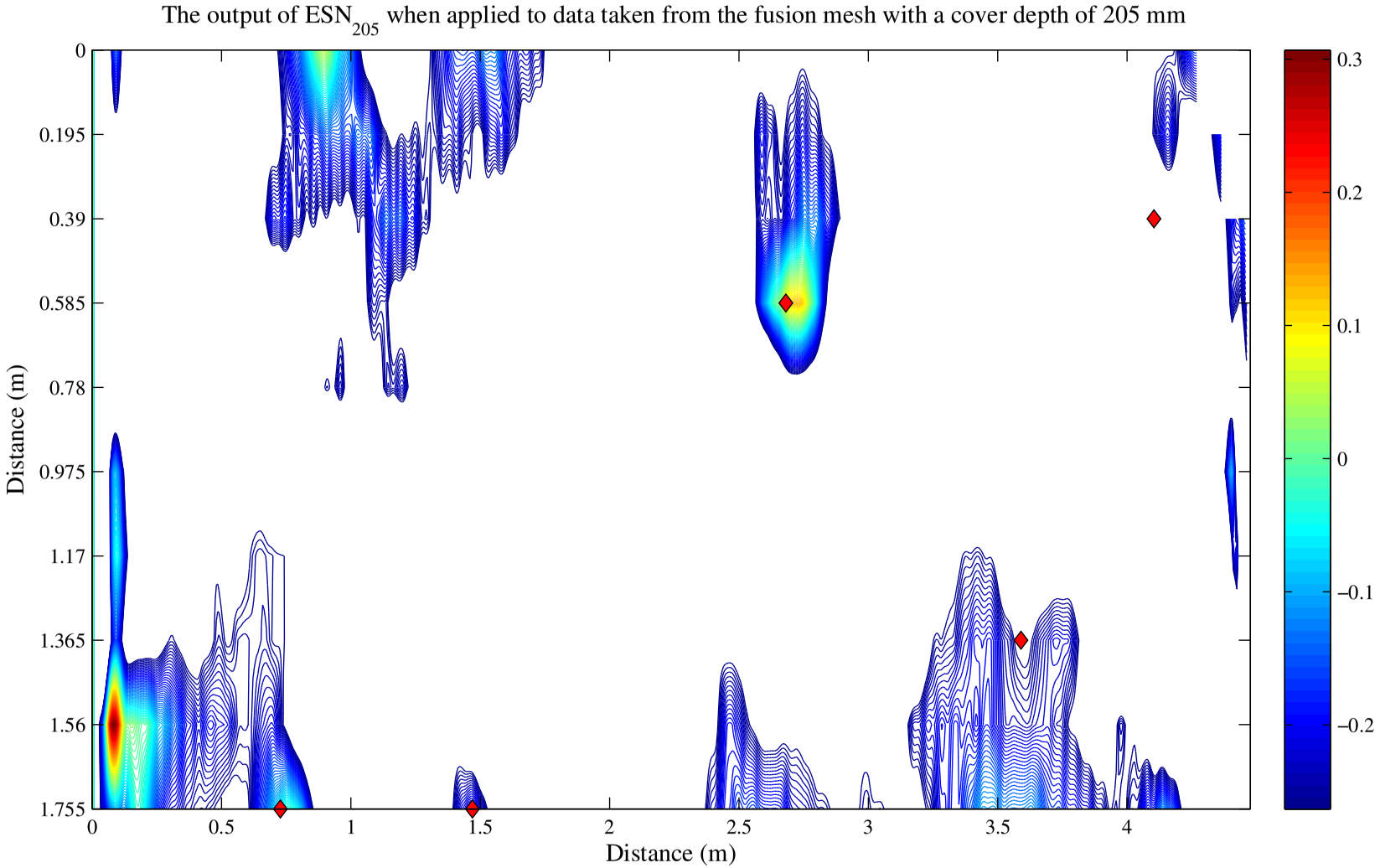


Figure A.24: A contour plot of the output of ESN_{205} when applied to data from dataset C2 at a cover depth of 205 mm, with the threshold (-0.2432) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8851.

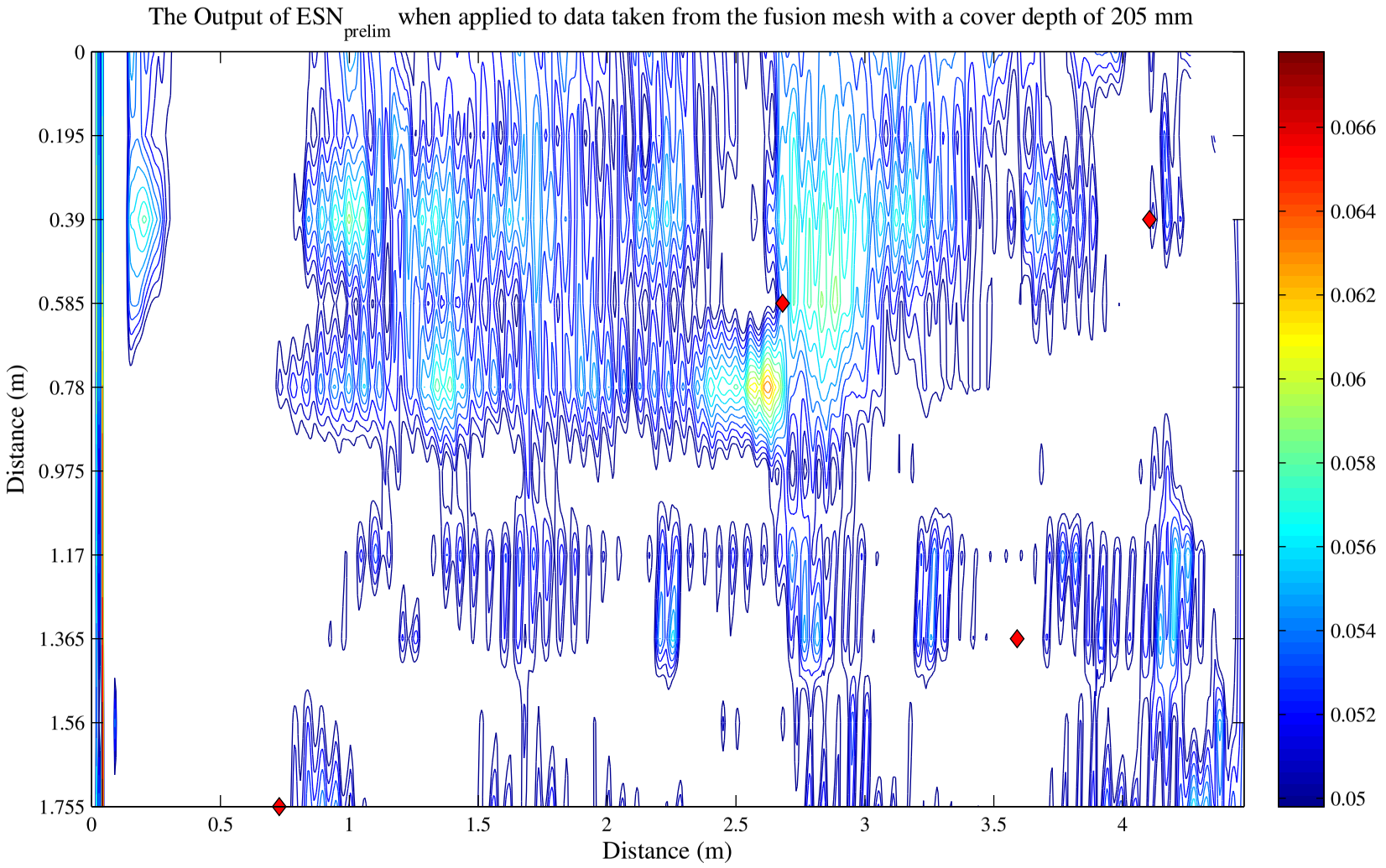


Figure A.25: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 205 mm, with the threshold (0.0498) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8518.

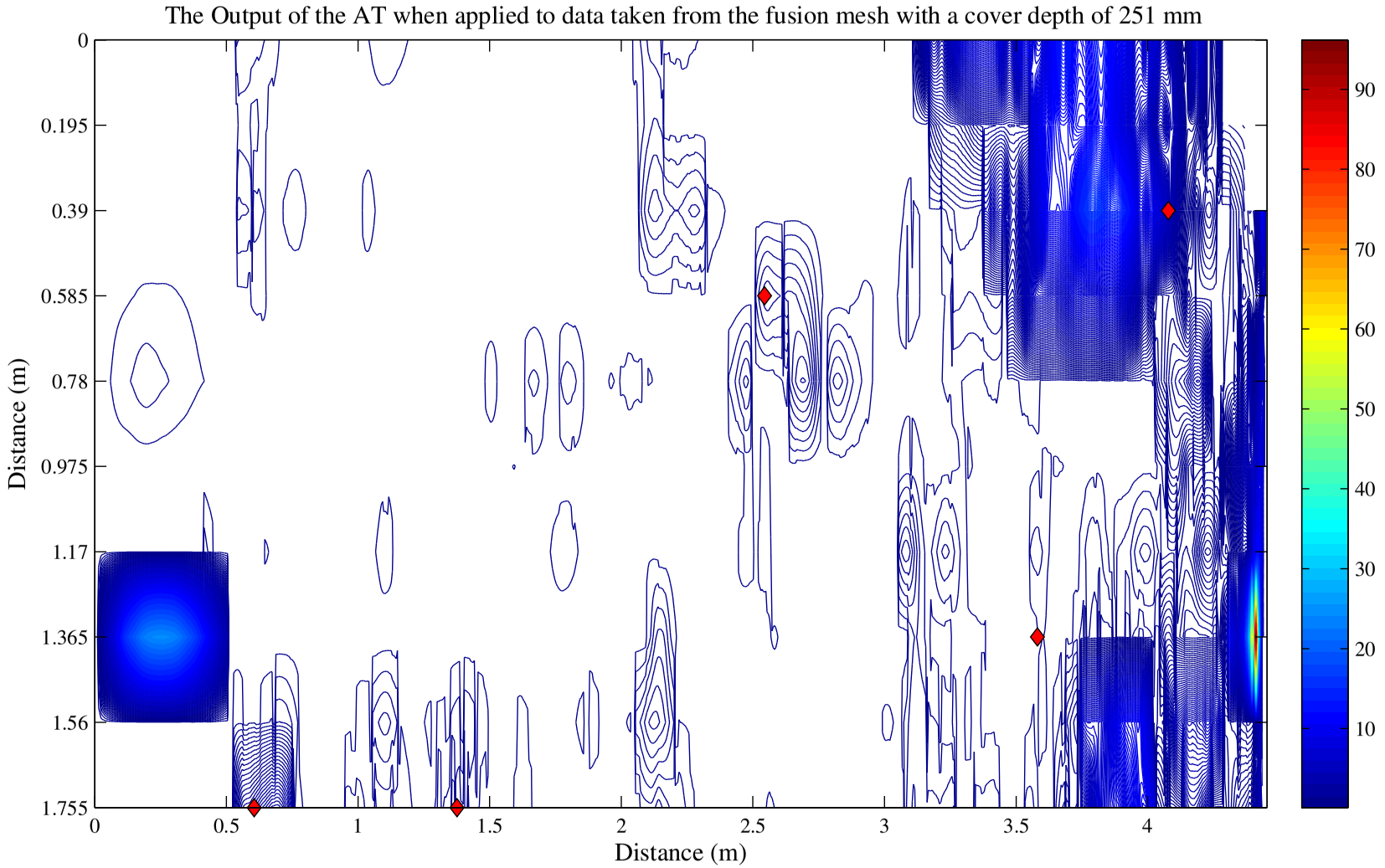


Figure A.26: A contour plot of the output of the AT when applied to data from dataset C2 at a cover depth of 251 mm, with the threshold (0.0758) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.7335.

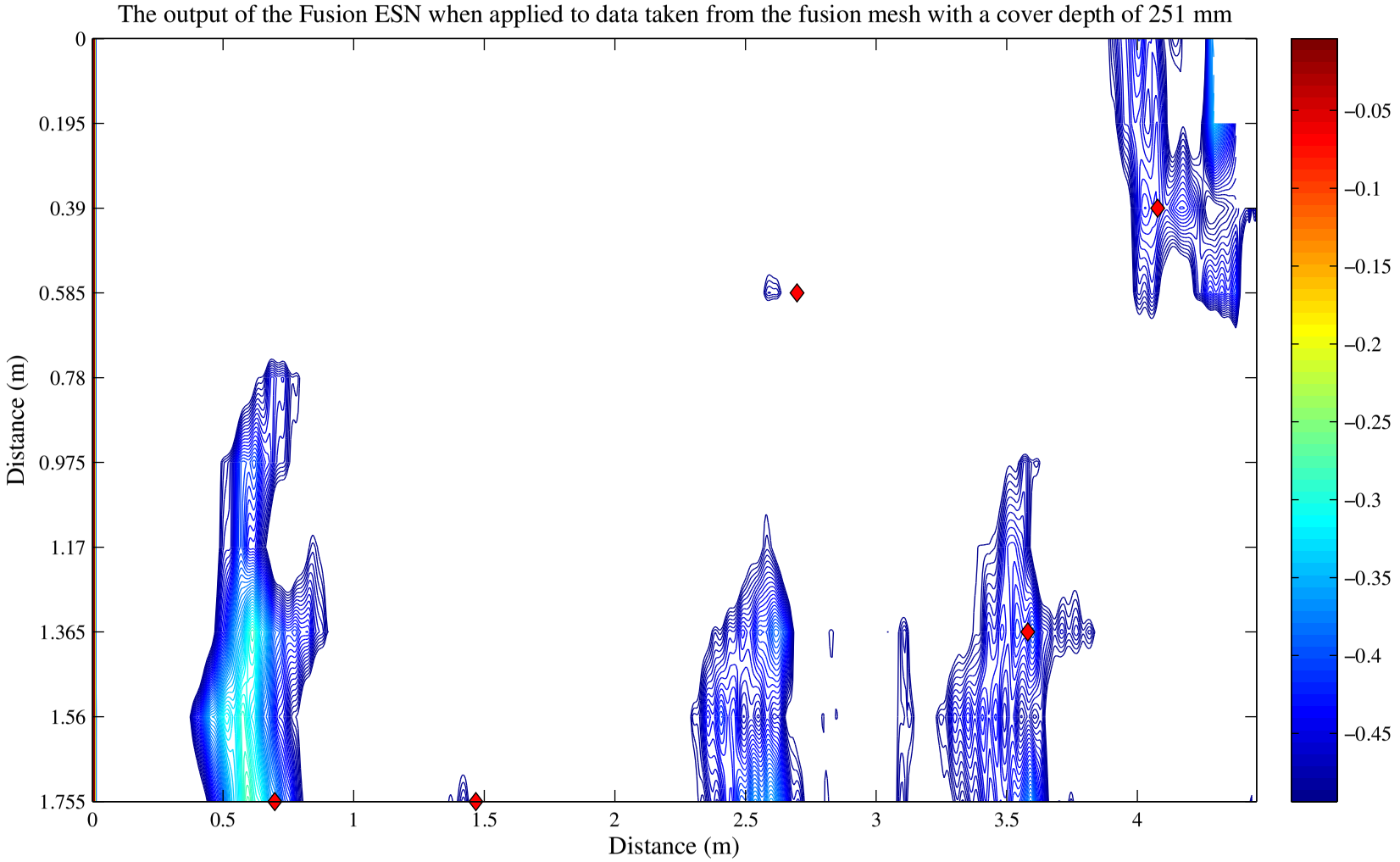


Figure A.27: A contour plot of the output of the fusion ESN when applied to data from dataset C2 at a cover depth of 251 mm, with the threshold (-0.4739) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8576.

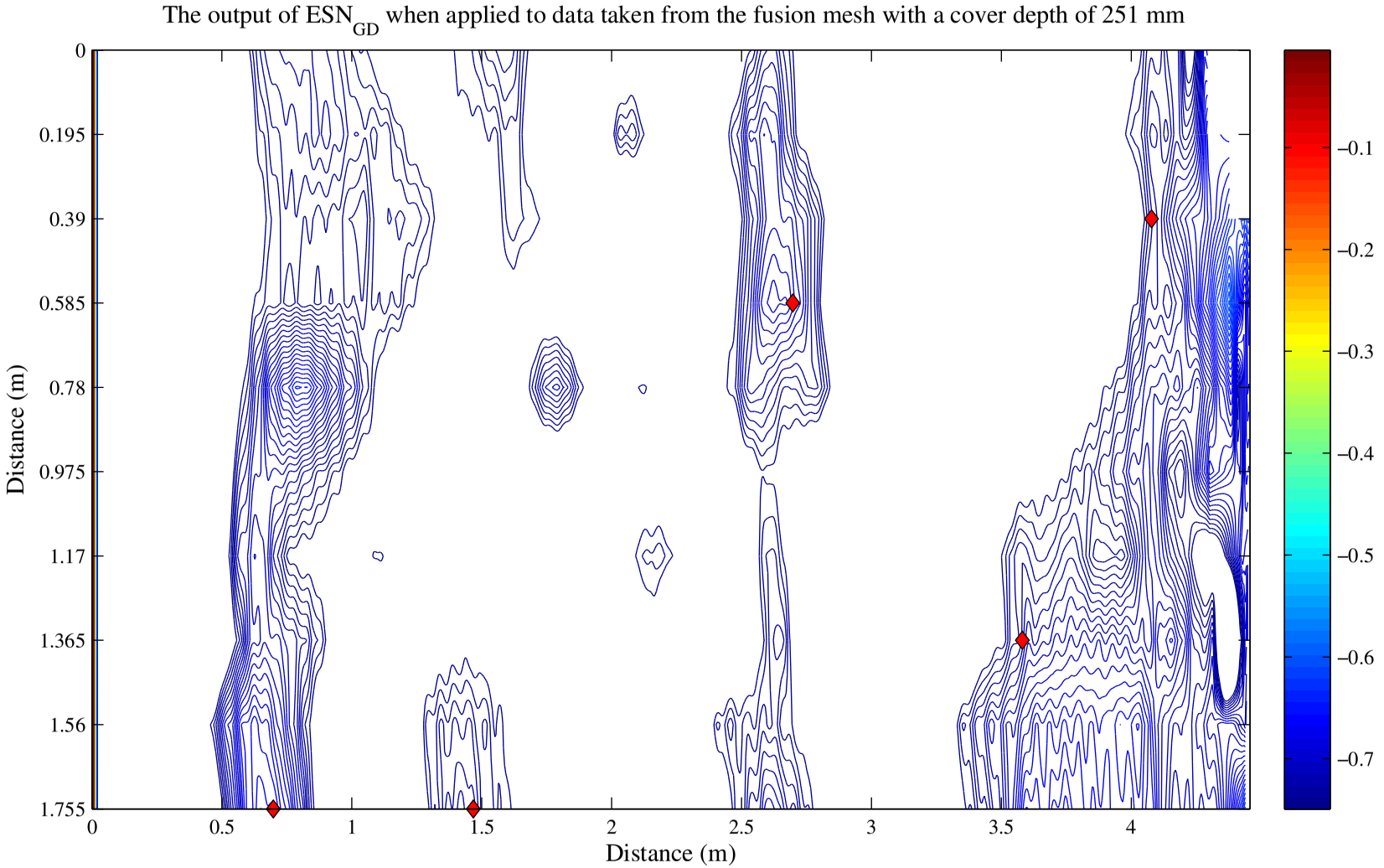


Figure A.28: A contour plot of the output of ESN$_{GD}$ when applied to data from dataset C2 at a cover depth of 251 mm, with the threshold (-0.7294) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8956.

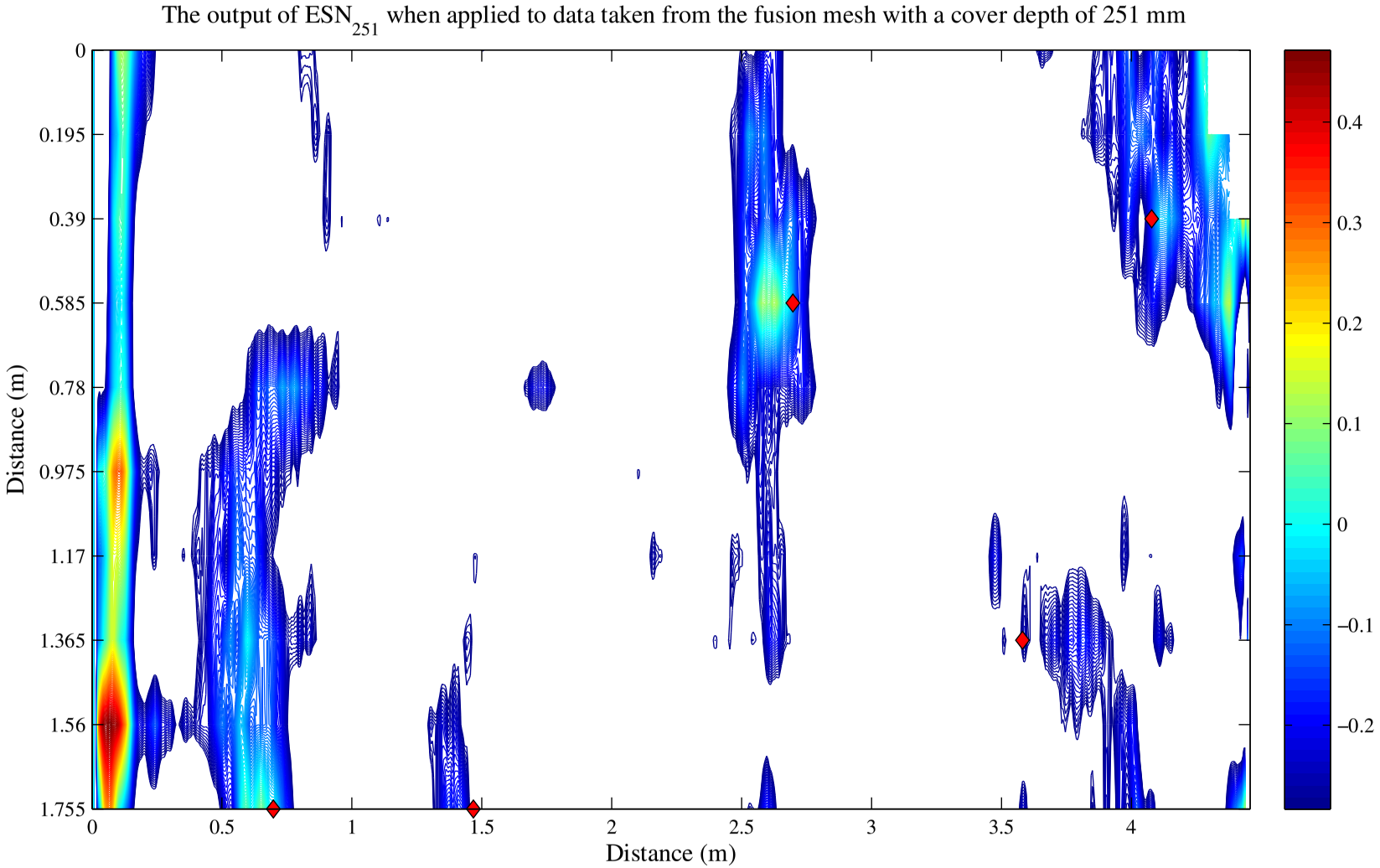


Figure A.29: A contour plot of the output of ESN_{251} when applied to data from dataset C2 at a cover depth of 251 mm, with the threshold (-0.2634) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9293.

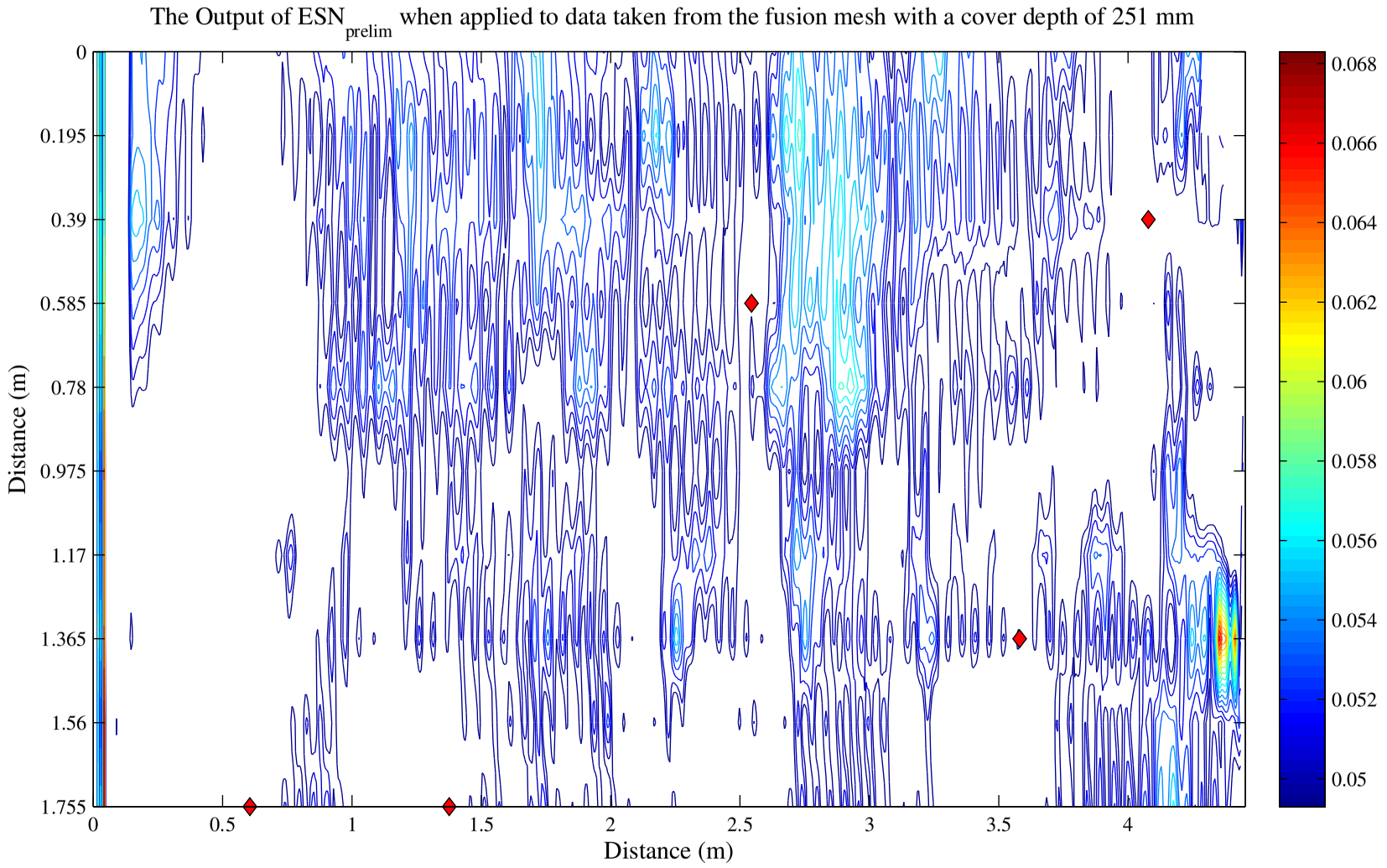


Figure A.30: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 251 mm, with the threshold (0.0493) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.7730.

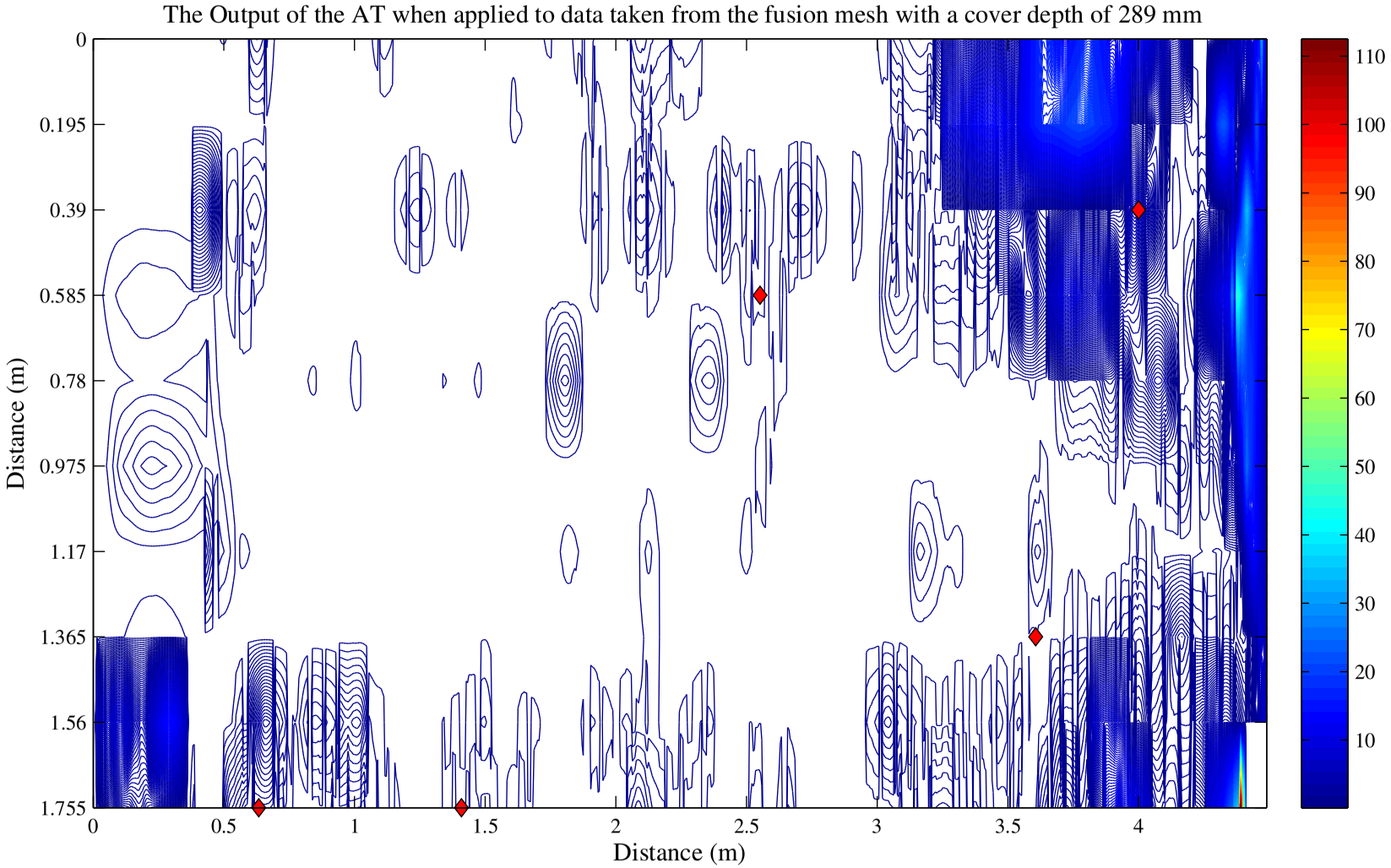


Figure A.31: A contour plot of the output of the AT when applied to data from dataset C2 at a cover depth of 289 mm, with the threshold (0.0876) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.7504.

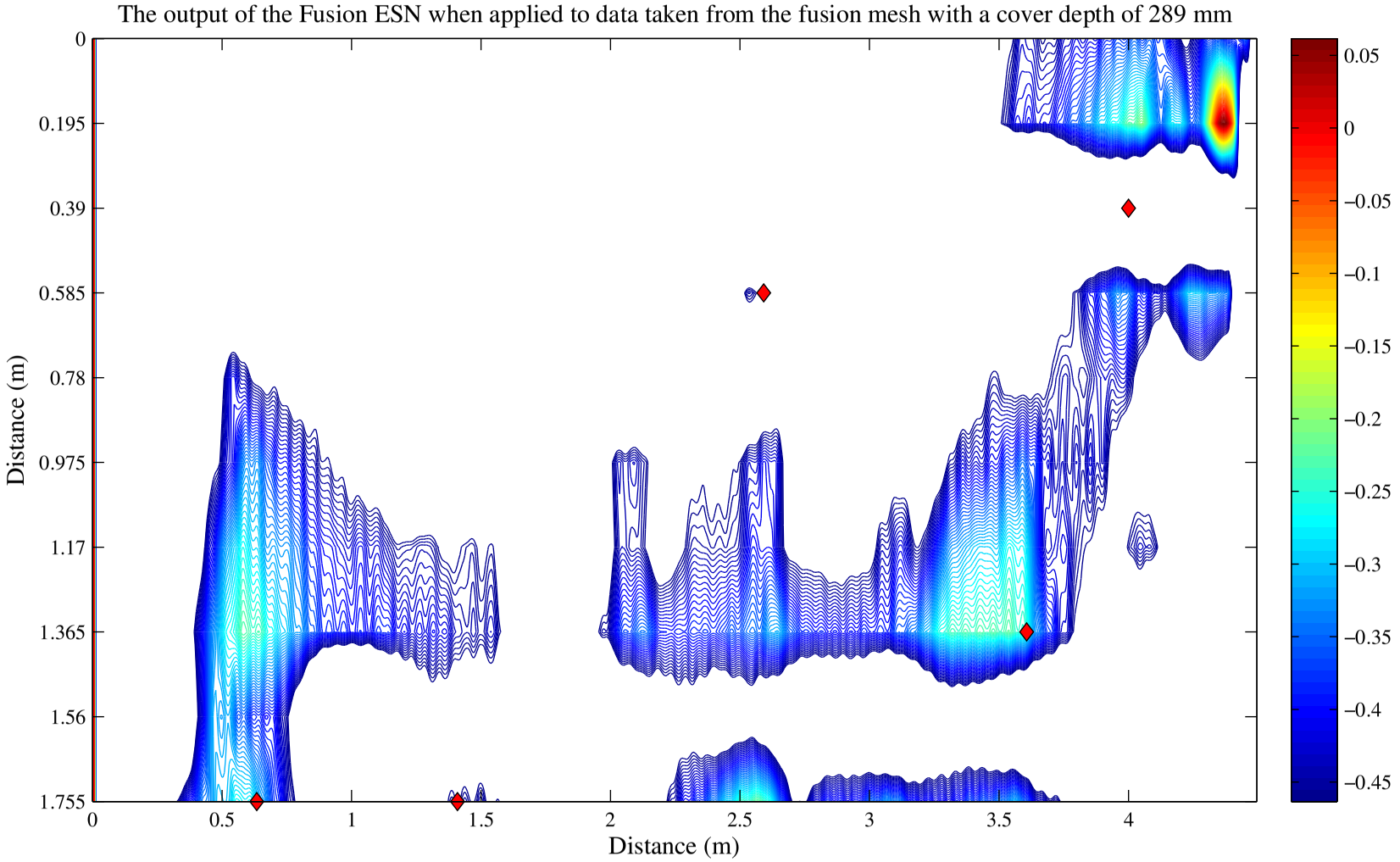


Figure A.32: A contour plot of the output of the fusion ESN when applied to data from dataset C2 at a cover depth of 289 mm, with the threshold (-0.4435) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.6717.

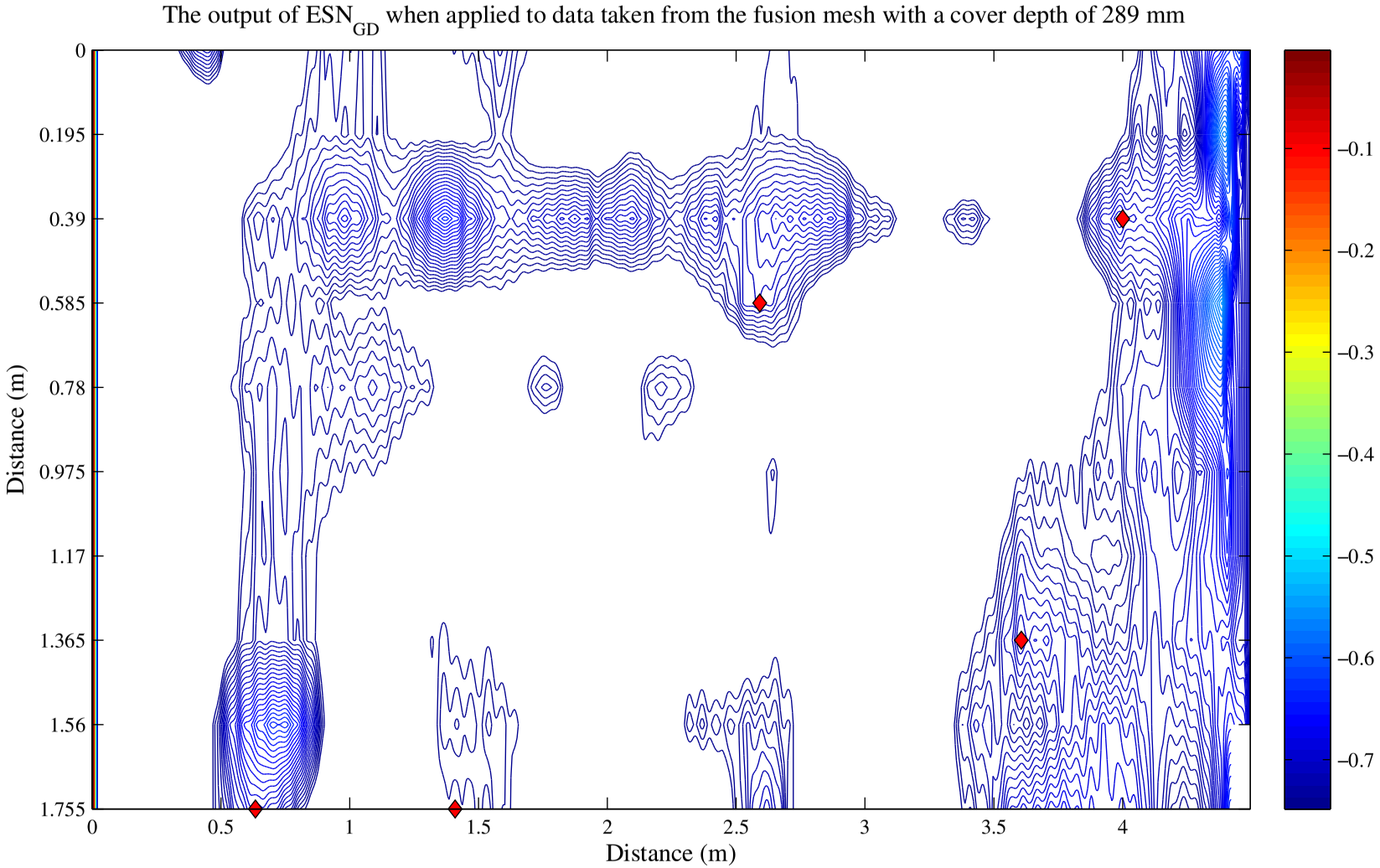


Figure A.33: A contour plot of the output of ESN$_{GD}$ when applied to data from dataset C2 at a cover depth of 289 mm, with the threshold (-0.7285) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8632.

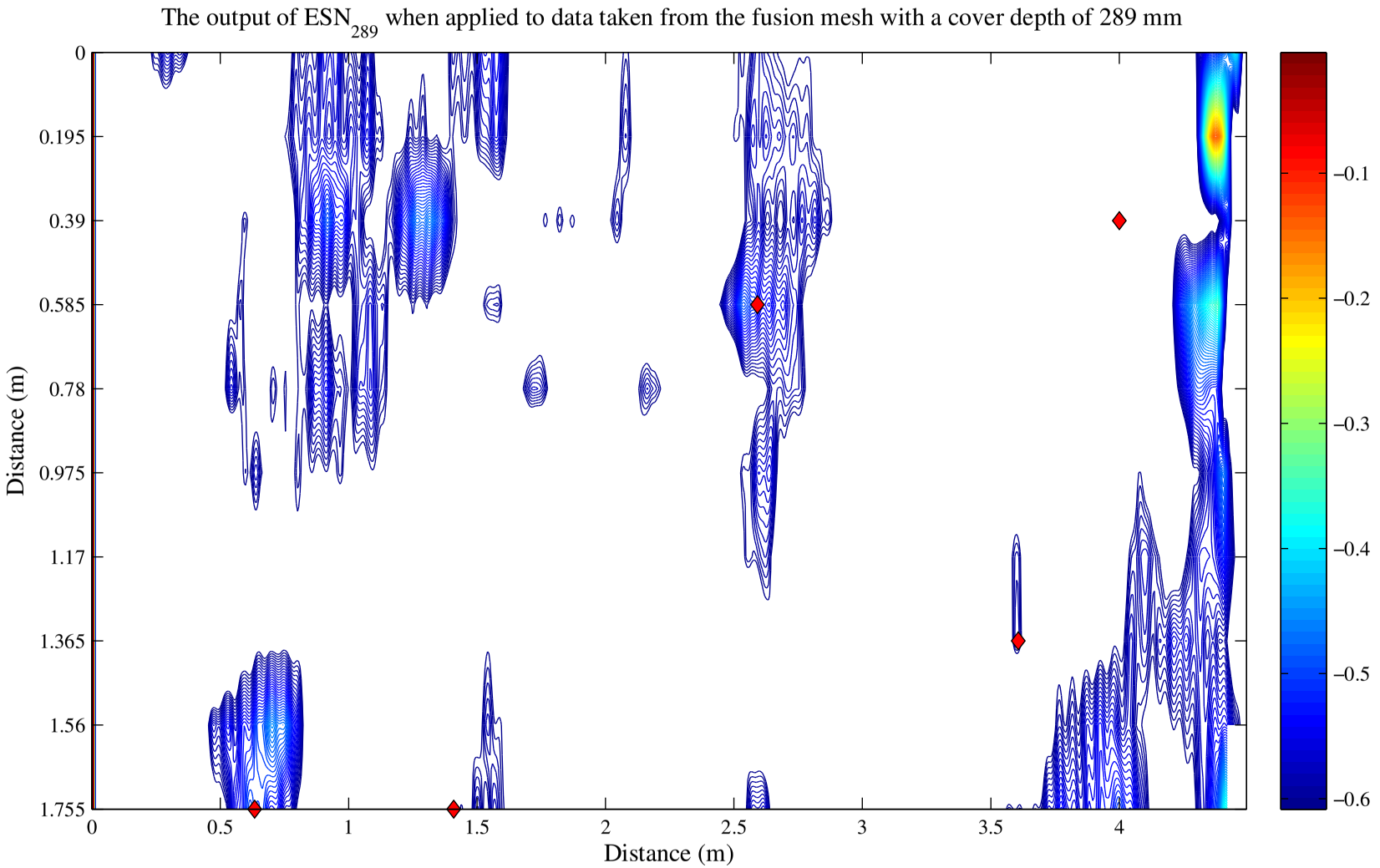


Figure A.34: A contour plot of the output of ESN_{289} when applied to data from dataset C2 at a cover depth of 289 mm, with the threshold (-0.5885) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8597.

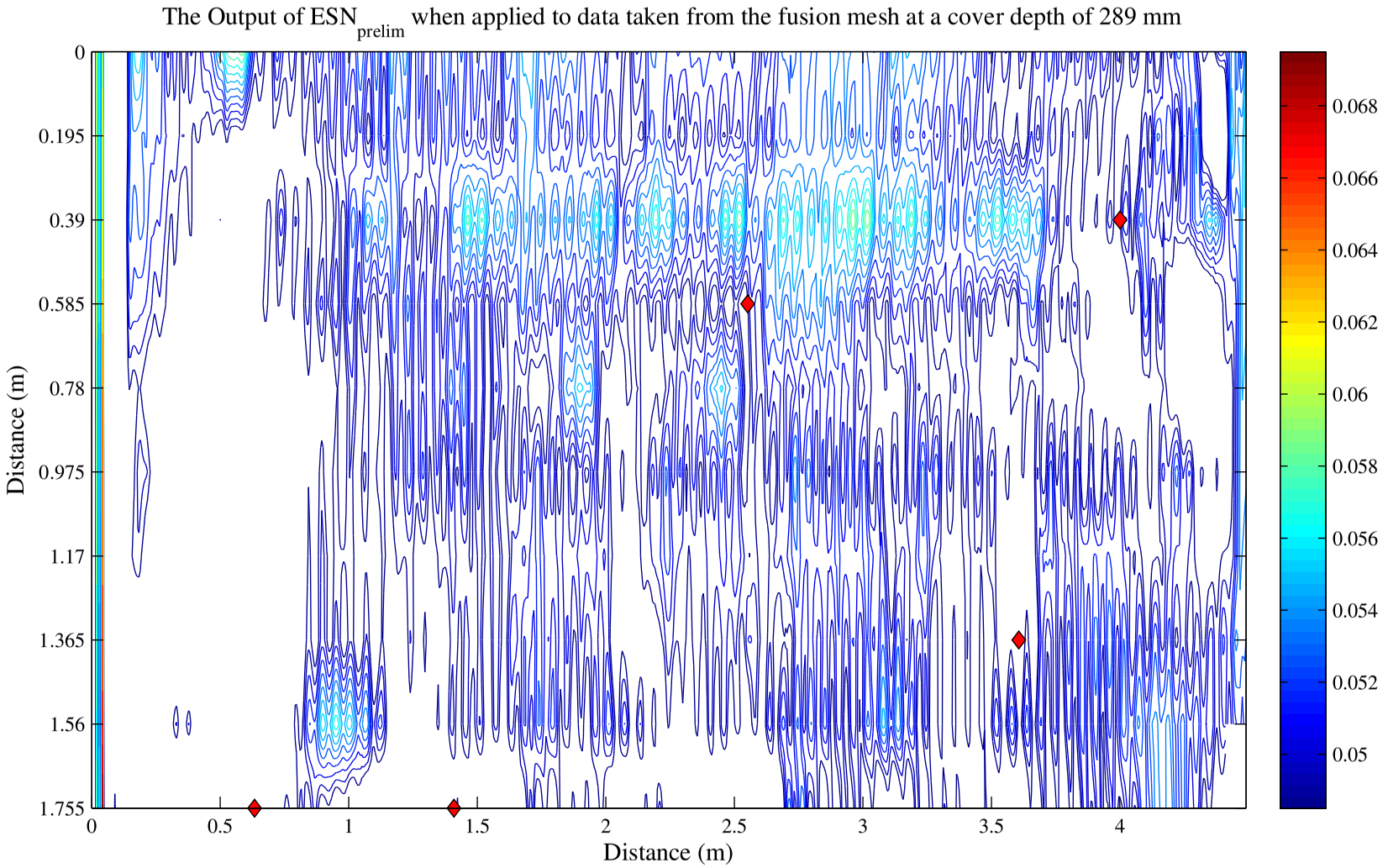


Figure A.35: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 289 mm, with the threshold (0.0485) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.7507.

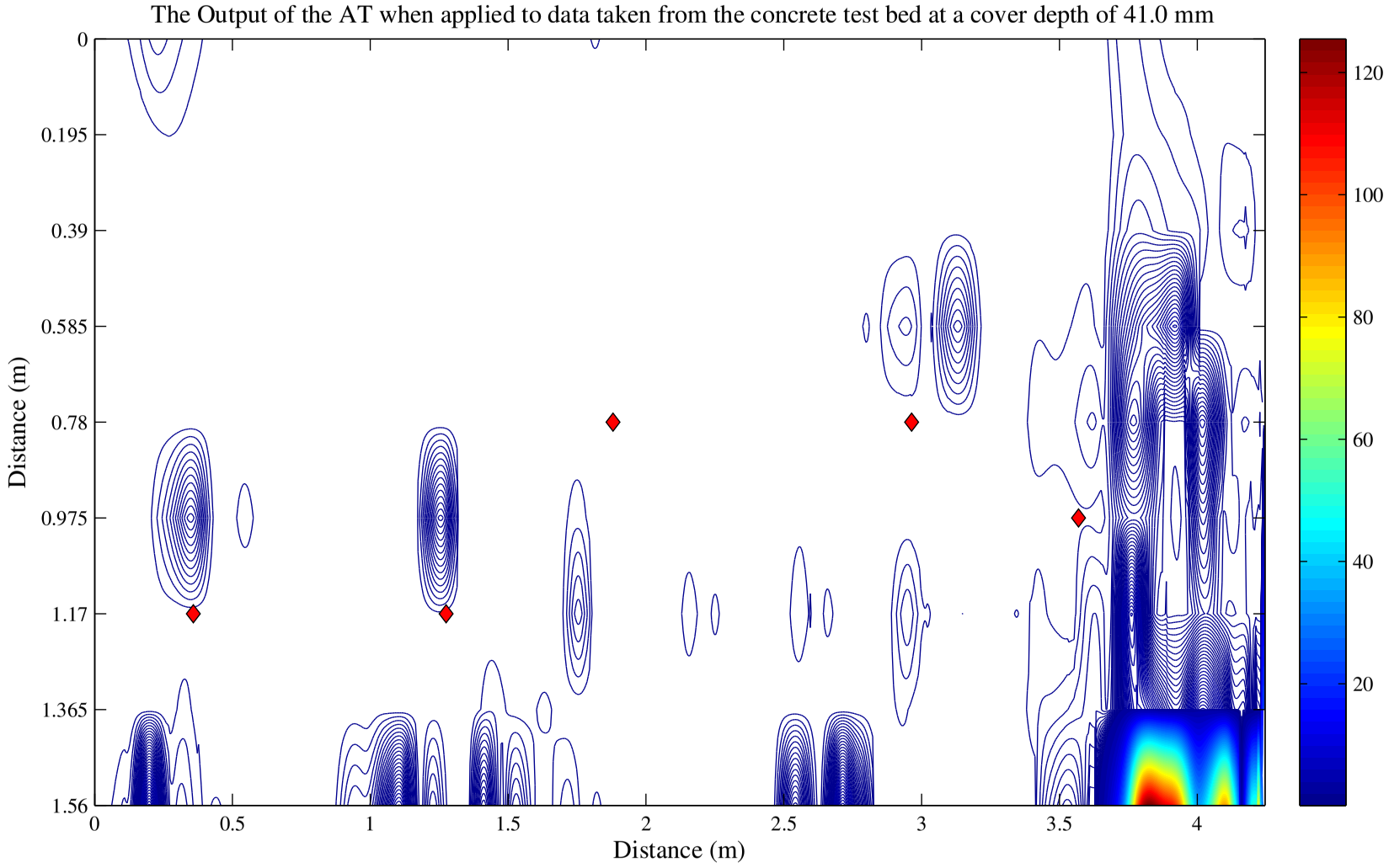


Figure A.36: A contour plot of the output of the AT when applied to data from the concrete test bed at a cover depth of 41.0 mm, with the threshold (4.3468) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9578.

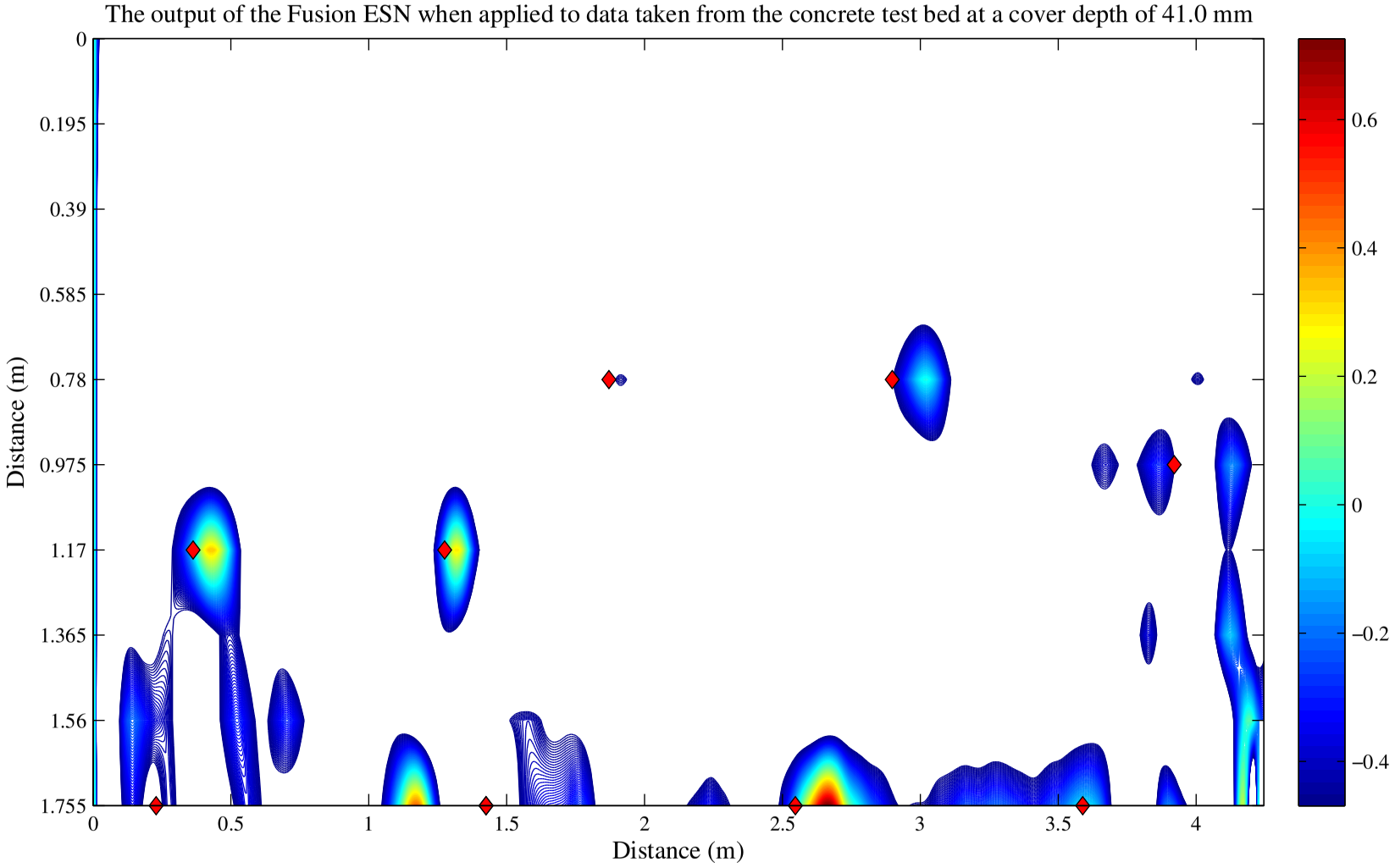


Figure A.37: A contour plot of the output of the fusion ESN when applied to data from the concrete test bed at a cover depth of 41.0 mm, with the threshold (-0.4496) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9679.

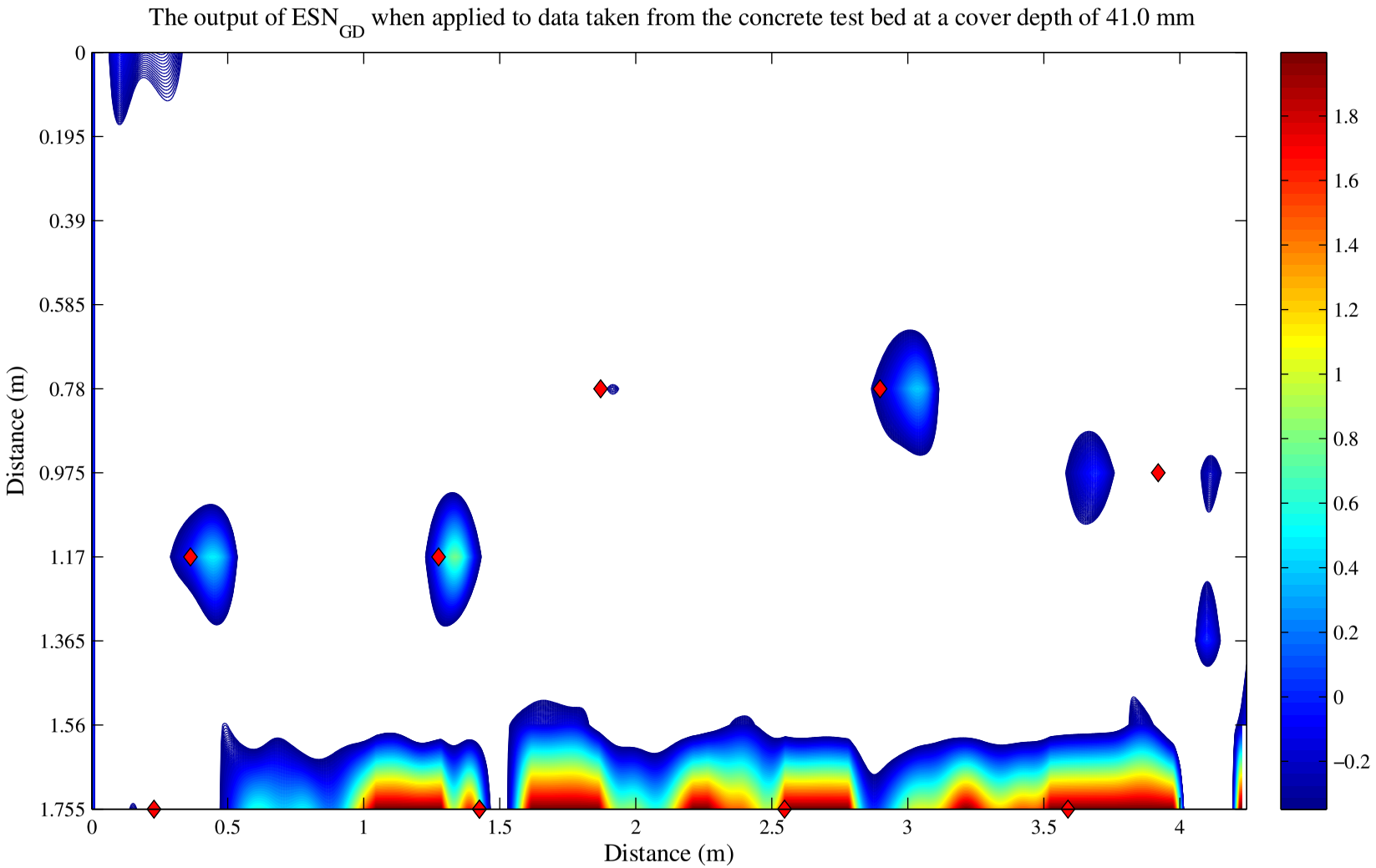


Figure A.38: A contour plot of the output of $ESN_{42.5}$ when applied to data from dataset C2 at a cover depth of 41.0 mm, with the threshold (-0.3284) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9363.

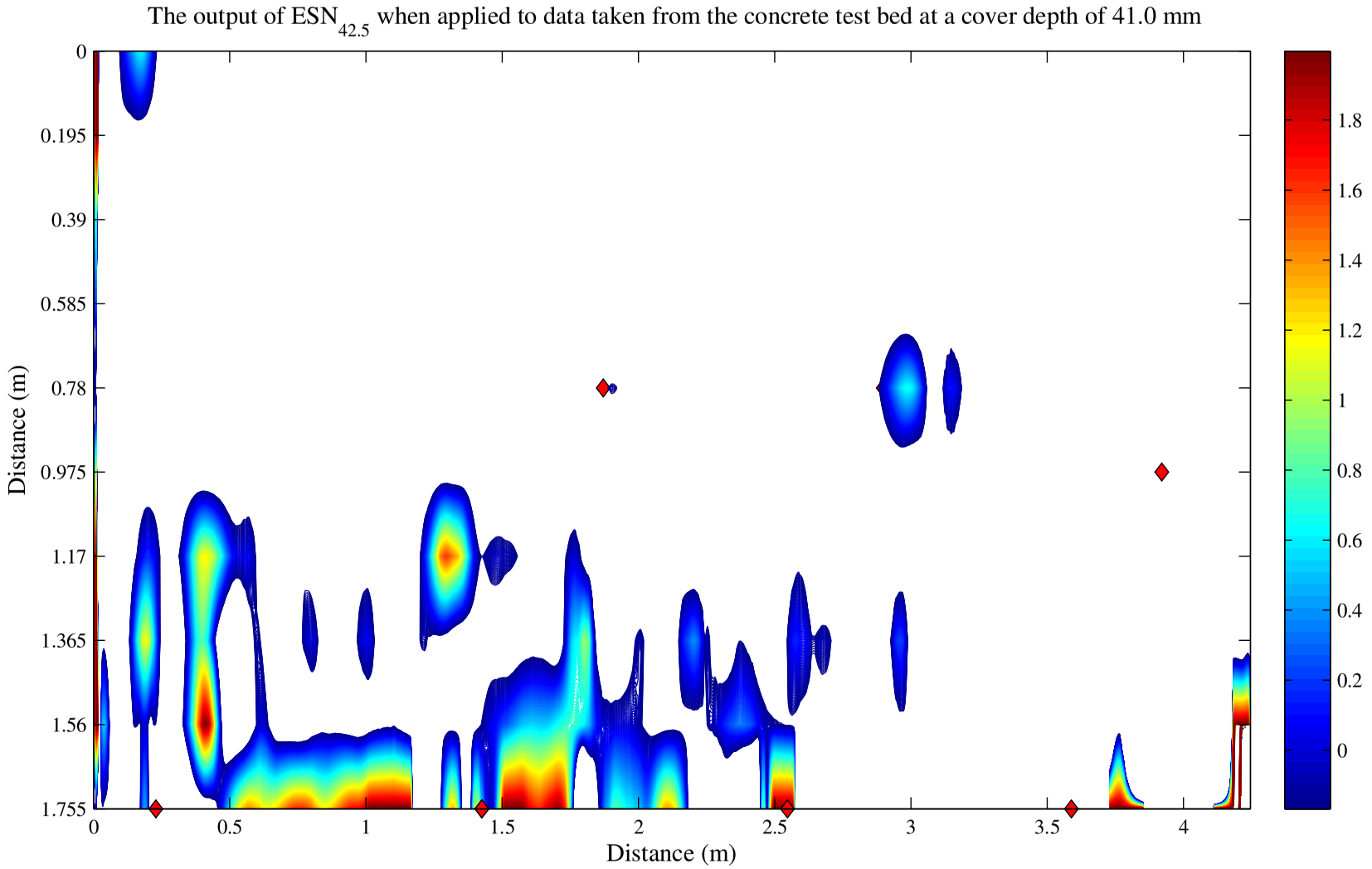


Figure A.39: A contour plot of the output of ESN_{GD} when applied to data from dataset C2 at a cover depth of 41.0 mm, with the threshold (-0.1483) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.8729.

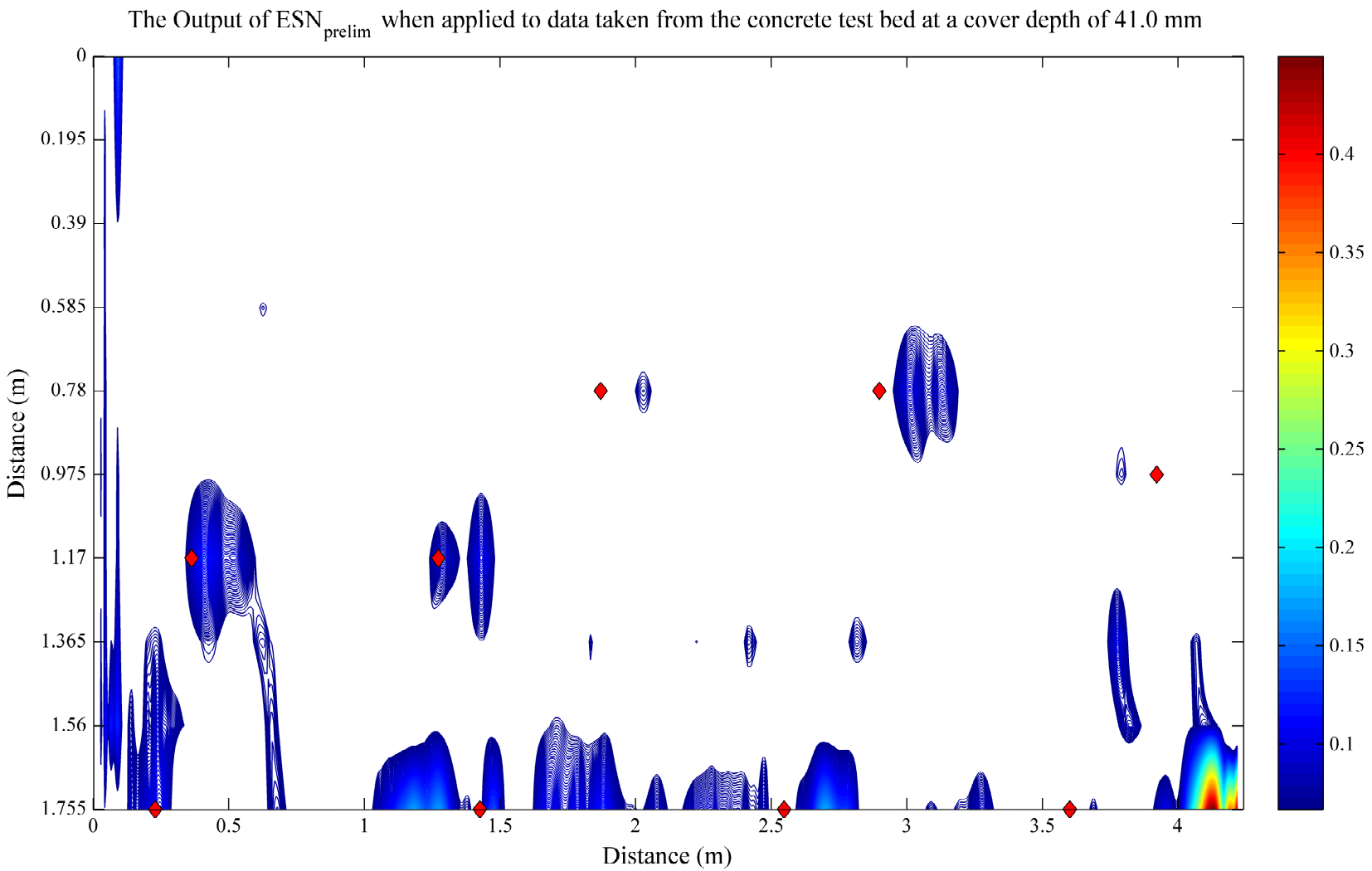


Figure A.40: A contour plot of the output of ESN_{Prelim} when applied to data from dataset C2 at a cover depth of 41.0 mm, with the threshold (0.0668) chosen by analysis of an ROC curve. White areas indicate that no defect was found, while coloured areas show the presence of defects. Red diamonds indicate the actual defect locations. The value of the AUC for this plot was 0.9660.

APPENDIX B

Optimizing Echo State Networks for Static Pattern Recognition

This appendix presents the results of a recently published investigation into the internal reservoir dynamics of ESNs [308].

B.1 Introduction

Two different ESN architectures were applied to a static pattern recognition task in order to classify the hyperspectral signatures of five different plant species, and compared with three other common CI techniques. This work took place in the context of the monitoring of the natural environment, as opposed to the built environment, and only considered one sensor modality, meaning that it is outside the scope of the thesis. However, the work is included in this appendix due to the insights it gives into the internal workings of ESNs, and the promising avenues of future CI research that it opens up.

B.1.1 Static Pattern Recognition

Static pattern recognition is a common problem in machine learning, where classifiers are trained to recognize a combination of attributes as belonging to particular classes and there is no spatial or time-series relationship between consecutive input patterns. An ESN static clamping regime has been used for this sort of task in the past [357, 358,

452, 453], and were applied here in a similar fashion in order to explore how such input pattern clamping might affect overall classification accuracy.

The motivation for using ESNs that were optimized for static pattern recognition was that conventional static pattern recognisers, such as MLPs, will produce exactly the same response when presented with a static input pattern, regardless of how many times the pattern is presented and regardless of what might have been presented in the recent past. In contrast, a recurrent architecture, after its initial response, is likely to follow a trajectory towards an attractor in the high dimensional reservoir space that might not have been reached by a single static pattern presentation. This could allow the weights connecting the reservoir units to the output units to be better able to classify the input patterns.

B.1.2 The Problem of Invasive Rhododendrons

Rhododendrons (*Rhododendron ponticum*) are an invasive plant species that needs managing in many parts of the UK [454]. In 2010, it was reported that the annual cost of controlling rhododendron is in excess of £8.5 million [455]. Since the size of most survey areas makes visual inspection impractical, airborne or satellite-based remote sensing techniques for locating them are being developed, including the use of optical spectroradiometry reflectance data [456–459]. Reflectance data sampled under laboratory conditions at 1869 wavelengths that had been collected from evergreen rhododendron leaves and the leaves of species also present in the shrub layer of UK woodlands were analysed using ESNs and alternative widely used CI techniques, each of which was optimized to deliver their best performance with this dataset. It was hoped that successful application to laboratory data would make the techniques used here viable candidates for application to 'real-world' remote sensing datasets, such as those already gathered at Coed-y-Brenin, Kiedler Forest District, Glen Affric and the Queen Elizabeth Forest Park [460]. More detail on the information contained within the plant leaf dataset can be found in Section B.1.3.

The aim of the study was to see if the short term memory developed by the ESN

reservoir and the statically clamped inputs could deliver improved overall classification accuracy when compared to other typical classification techniques. This differs from a past work on the optimization of ESNs for static pattern recognition in that it considers a wider range of ESN parameters [356]. Furthermore, while past works allow the reservoir units to settle, the possibility of instead allowing the output units to settle, or of letting neither settle, is considered here.

B.1.3 Leaf Sample Spectroradiometry Dataset

A leaf sample spectroradiometry dataset was used to measure the classifier performance. The dataset contained 282 real world leaf samples, each with 1869 attributes, corresponding to the absolute optical reflectance of the leaf sample in 10 nm wavelength bands, and was gathered by Taylor et al. [458]. The leaf set samples consisted of various species: 126 rhododendron, 48 beech (*Fagus sylvatica*), 36 holly (*Ilex aquifolium*), 36 ivy (*Hedera helix*) and 36 cherry laurel samples (*Prunus laurocerasus*). This was done in order to ensure that any classifier was able to distinguish between rhododendron and four other species commonly found in the UK, three of which are native. The classifiers were presented with 24 randomly selected samples from each species for training and then the rest of the data were used during the testing phase.

Although there were five different plant species to identify, particular importance was placed on the detection of the rhododendron samples, since this was the target invasive species under consideration as an important vector for plant infections. Therefore, two key performance aspects of the classifiers were the percentage of rhododendron samples correctly identified and the percentage of samples of other species that were misclassified as rhododendron.

B.2 Methodology

As an extra test of the importance of the ESN's short term memory in static pattern classification, Extreme Learning Machines (ELMs) were also trained and applied to the

data. An ELM can be thought of as an ESN without the possibility of recurrence, with no interconnectivity in the reservoir (i.e. a spectral radius of zero) and, therefore, no short term memory. While this could be implemented using the reservoir computing toolbox, it was found that the training regime used in Huang's ELM implementation for MATLAB [461] gave better performance. Using this implementation, it was found that using a triangular basis activation function in a 1000 unit hidden layer produced excellent results that were considerably better than the results obtained using other activation functions or different numbers of ELM hidden layer units. So as to provide further performance comparison, two other widely used classification methods were applied to the same data: MLPs, which were implemented using the neural network toolbox for MATLAB, and Support Vector Machines (SVMs), which were implemented using libsvm for MATLAB [462]. For each of these classifier approaches, a similar grid-search optimization method was used, so as to provide the optimal set of parameters for each approach. The best performing MLP had a hidden layer with eight units and the best performing SVM was found to use a linear kernel function. Since the weighted connections in ELMs, MLPs and ESNs are randomly generated at initialisation, the results given for these three methods are the average performance of 1000 separately trained networks for each technique. Similarly, the performance of SVMs is strongly dependent on the training data, so 1000 random training datasets were generated and the average performance of the SVM is reported below. The classification accuracy was calculated using Equation B.1.

$$Accuracy = 100\left(1 - \frac{m}{n}\right) \tag{B.1}$$

In B.1, m is the number of misclassified samples and n is the total number of samples. As an additional performance measure, the MCC was calculated for each output class from each classifier. All of the data were normalised between +1 and -1 prior to presentation to any of the classifiers.

Prior to the use of the leaf sample spectroradiometry dataset, the ESN approach was validated using two well-known benchmark datasets from the UCI machine learning

repository [463], namely the Iris dataset and the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The former consists of 150 samples, each with four different attributes. There are three possible classes for the data to be sorted into and 50 samples for each class. Each of the classifiers was presented with 25 samples of each class for training and tasked with correctly classifying the remaining 25 samples during testing. The breast cancer dataset is somewhat more challenging, containing 569 samples with 30 features each, with two possible classes. There were 357 samples for the first class and 212 for the second class. Fifty samples from each class were used for training the classifiers, which were then tasked with classifying the remaining data in the testing phase.

B.2.1 ESN Optimization for Static Pattern Recognition

This section details how the adjustment of the ESN parameters (as detailed in Chapter 2) was observed to affect the performance of ESNs when given a static pattern recognition task. The tuned parameters here were the spectral radius, the leak rate, the reservoir size and the input scaling.

In order to find the best static pattern recognition values for the parameters, the ESNs were presented with data from the leaf sample spectroradiometry dataset. This required an ESN with 1869 input units – one for each wavelength reflectance attribute – and five output units, each representing a different plant species classification. An attempt was made to reduce the dimensionality of the dataset using principle component analysis; this did not lead to improved performance in preliminary investigations. The overall performance of the network was then thoroughly tested using 20-fold cross validation and it was on this basis that the optimal settings were chosen.

The presence of recurrent connections in the reservoir of a conventional ESN means that the ESN can have a memory of past inputs that will affect the final output unit activations. Static data, by definition, have only one set of input data for each sample and so the most common approach is to keep the input clamped until the reservoir state has settled [357, 358, 452, 453]. The approach used here, though, was to maintain the clamped inputs until the output unit states were settled. The output node that

produced the highest value at the end of a sequence of repeated pattern presentations was considered the 'winner' for that set of clamped inputs. During each clamped sequence of input pattern presentations, the output of each output node gradually converged and eventually settled on a particular value. The input pattern clamping regime effectively prevents the ESN learning a relationship between two consecutive input patterns.

B.2.1.1 Leak rate

It was observed that the higher the leak rate, the smaller the number of repeated input pattern presentations required for each output unit to settle on a value. Once the outputs had settled, they continued to produce this value for every subsequent repeated input pattern presentation. The sooner the outputs converge, the fewer repeated input pattern presentations are required.

An example of the ESN output unit activations for a high leak rate (0.9) and a low leak rate (0.1) can be seen in Figure B.1. Although the size of the leak rate did not radically affect the classification performance of the ESN during cross validation, so long as there were sufficient repeated input pattern presentations, the best classification performance came with a leak rate of 0.3820. Figure B.1 shows that for input patterns that are difficult to classify (for example, rhododendron often misclassifies as cherry laurel), having a lower leak rate allows the ESNs to distinguish more clearly between input patterns from those classes. For the outputs shown in Figure B.1, both ESNs were repeatedly presented with the same input data taken from a cherry laurel sample. However, in Figure B.1B, the final output activation for the laurel output node is almost the same as the rhododendron output node's activation, showing that the network has difficulty distinguishing between these two classes. In Figure B.1A, the ESN has a lower leak rate and the output unit classifications are correctly separated. It was generally observed that there was a clearer distinction between classes for a slower reservoir of this sort. A leak rate of 0.3820 provided a good balance, offering sufficiently slow dynamics and also requiring a suitably small number of repeated presentations; classification

performance did not improve when the number of input pattern presentations was more than 60.

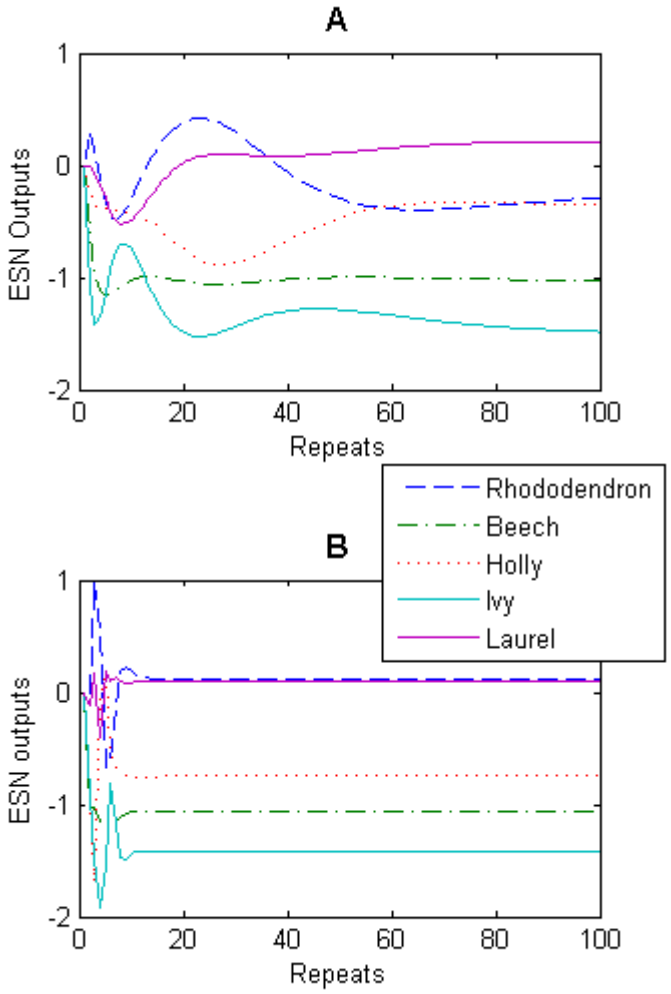


Figure B.1: ESN outputs over 100 input pattern presentations for the same set of input data. Each coloured line represents values produced by a different output node, with each node representing a different class. The ESN in A has a very low leak rate (0.1) such that the output varies for some time as the repeats progress, while the ESN in B has a high leak rate (0.9) and thus very fast dynamics. While the output of most units is roughly the same in both cases, the holly output unit settles on a much lower output of -0.7, rather than -0.4. Note also that in B, it can be difficult to distinguish between the rhododendron and laurel outputs, which give very similar values.

B.2.1.2 Reservoir size

It was found that a relatively small reservoir consisting of just 55 neurons provided the best classification performance on the cross validation data. This is a curious result and led to an unconventional application of ESNs in this work. One of the basic ideas of

ESNs is to project an input into a higher-dimensional space – the reservoir – allowing for the data to become more linearly separable [297, 464]. Consequently, one might assume that the very high-dimensional input used here would necessitate a reservoir containing in excess of 1869 neurons, which could be computationally unwieldly. The use of a lower-dimensional representation is more akin to the recent work by Bozhkov et al., where ESNs were used for the dimensionality reduction of static input patterns [376]. While that work featured combinations of the equilibrium states of reservoir neurons being fed into classification techniques such as SVMs and decision trees, the high-dimensional input data here effectively underwent dimensionality reduction by being fed into the low-dimensional reservoir, with the output weights then trained to learn the relationship between this representation and the desired output. In this sense, the static reservoir acts almost as a feature selection tool and classifier.

B.2.1.3 Spectral radius

While the ESN used here is unconventional in the sense that the inputs are static, it can still be said that the spectral radius has influence over the extent to which the ESN outputs vary with each repeated input presentation.

Figure B.2 shows the output of two ESNs processing the same clamped input pattern, but with different spectral radii: 0.99 in A and 0.01 in B. Comparing the plots in A and B, it is clear that the number of repeated input pattern presentations required for the output unit activations to settle is similar, if slightly longer for the smaller spectral radius. Prior to that point, though, the behaviour of the output units is significantly influenced by the network's memory of previous inputs. When the spectral radius was increased past a value of unity, chaotic outputs and poor network classification performance were observed. Based on classification accuracy during the cross-validation testing, the best value for the spectral radius was found to be quite low at 0.181. This suggests that for static classification tasks, a network should be strongly driven by the clamped input pattern, and that reservoir feedback is less important. Consequently, it is better to use a network with more stable dynamics. Importantly, it should be

noted that that some feedback is required (i.e. the best spectral radius was greater than zero and performance was generally good for all spectral radii less than unity, as can be seen by the range of spectral radii evaluated in Figure B.3). A non-zero spectral radius is therefore one of the properties of the ESN that is desirable for a static pattern classification task.

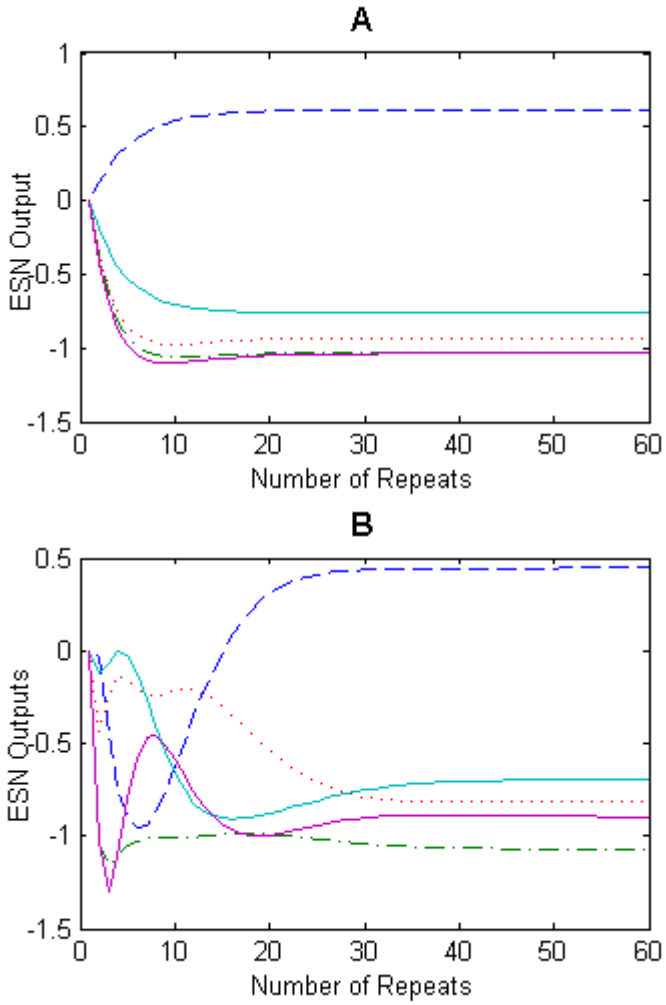


Figure B.2: An ESN with a spectral radius of 0.01 (A) (i.e. a small short-term memory) and an ESN with a spectral radius of 0.99 (B) (i.e. a long short-term memory). The lines representing each class are the same as in Figure B.1

B.2.1.4 Input scaling

Figure B.4 shows how the performance of the ESN varied as input scaling was varied. Classification performance during cross-validation peaked with an input scaling value of 0.17. This means that the reservoir units would have been operating in the linear

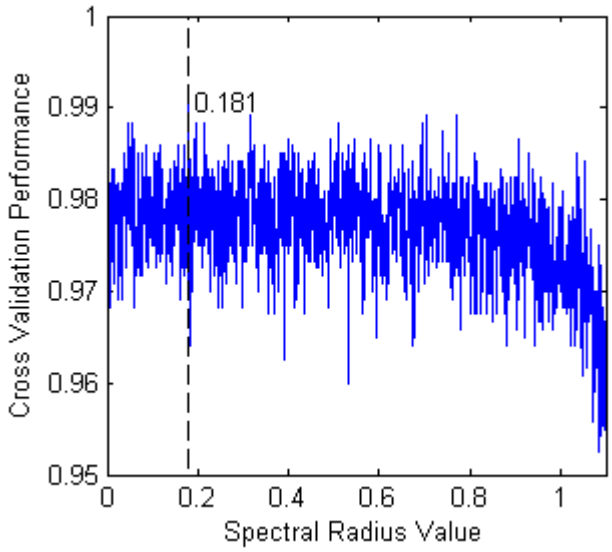


Figure B.3: The effect of the spectral radius on ESN performance. The final optimal value, 0.181, is marked by the dashed line.

part of their tanh activation function.

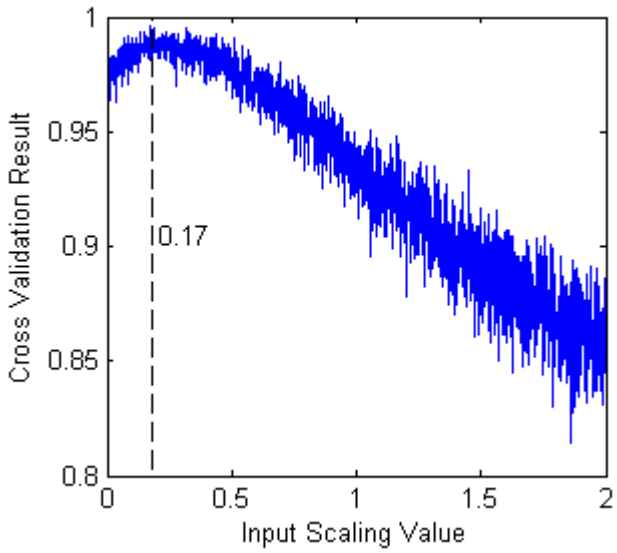


Figure B.4: The effect of the input scaling value on the performance of the network. The optimal input scaling, 0.17, is marked on the graph with the dashed line.

B.2.1.5 Interaction between reservoir and output neurons

One of the most instructive details found during the optimization process concerned the relationship between the interaction of the ESNs' reservoir neurons and their output neurons. Compared to the work of Alexandre et al. [357], whose focus was on reservoir

behaviour, Figure B.5A shows the outputs of twelve randomly selected reservoir neurons for each presentation of a randomly chosen clamped input pattern that was clamped at the input units. Figure B.5B shows the corresponding output activations from each of the five output units during the same input pattern clamping. The vertical dashed line on each plot marks the point at which output units and the selected reservoir units had plateaued, respectively. The lag between the number of pattern presentations required for the reservoir and the output units to plateau indicates that the best results may not necessarily achieved by harvesting values after reservoir stabilisation, as suggested by Alexandre et al. [357], but after output unit stabilisation has occurred instead. Furthermore, although the output unit activations had all plateaued after around 24 input pattern presentations during training, the best cross validation classification results were obtained with 60 input pattern presentations, suggesting that the output weights of the ESN need to be trained on a settled reservoir. However, only 24 input pattern presentations are required when a trained network is presented with unseen test data.

B.2.1.6 Input clamping

Earlier work has noted that reservoir neurons need to be 'clamped' until full reservoir stabilisation has been achieved for good classification performance [357, 358, 452, 453]. However, while a grid search found that using 60 repeated input pattern presentations was optimal, it was observed that good performance could also be achieved by as few as two input pattern presentations during both the training and testing period. Using only two input presentations is equivalent to obtaining an output without allowing the reservoir to settle at all, suggesting that for good (if not optimal) performance, input 'clamping' may be almost unnecessary. This is a counter-intuitive result that goes against the conventional wisdom.

Interestingly, the fact that the reservoir is still in an unstable state for the first ten input presentations means that the number of input presentations in testing must not exceed that used in training, unless the number used in training is greater than the

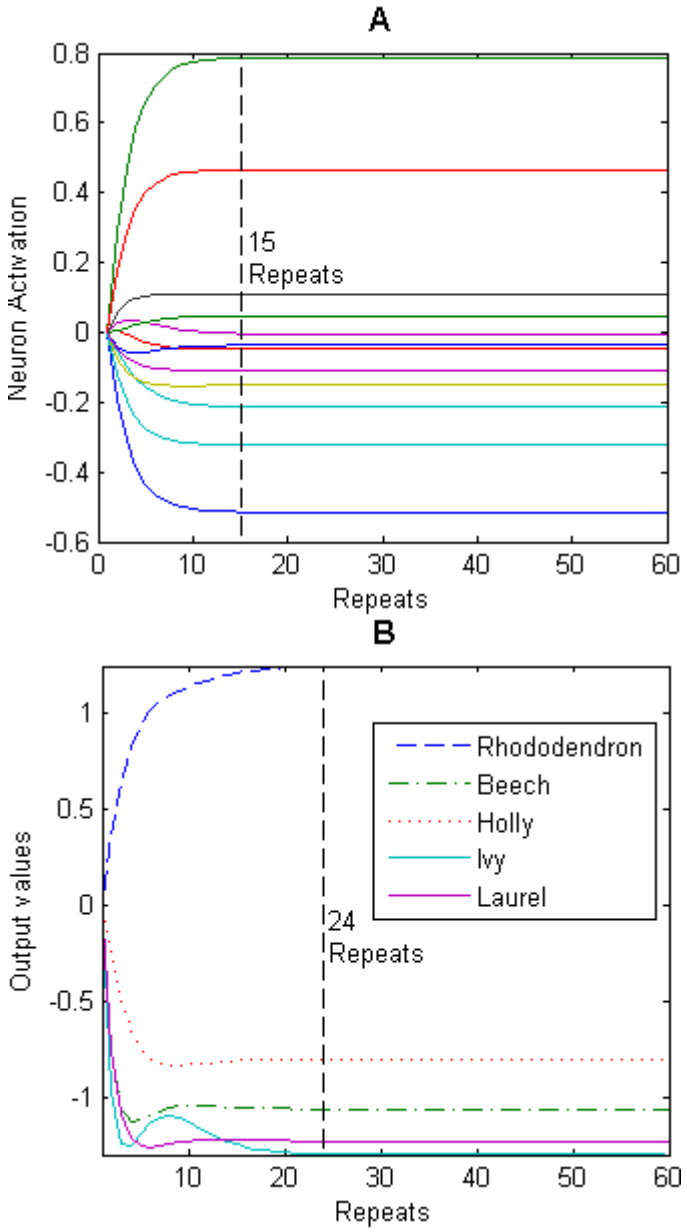


Figure B.5: The activations of twelve randomly selected reservoir neurons (A) and the corresponding outputs of the five output units (B). On each graph, the point at which the data had all plateaued is marked with a dashed line.

extent of the unstable state. Figure B.6 shows how the output from each node varied with each repeated input pattern. The purple line represents the cherry laurel output node, which should give the highest output value for this test pattern in question. This plot demonstrates why it is important that the number of testing repetitions should not exceed the number of training repetitions when using this approach. Although the cherry laurel node has the highest output after the second presentation, beyond this point a different output node starts to produce much higher values, causing that to become the winner after 25 testing presentations were used. It was usually the case that for every sample presented during testing, each ESN trained on two presentations had one output node consistently produce much higher output values than all of the others, although this was not necessarily the correct winning output node. It seems to be the case that the dynamics of the reservoir and corresponding output weights of the network had been tuned so that the correct output was obtained after the first two presentations, but these weights were not configured to continue produce the correct output when the reservoir dynamics were driven beyond this number of presentations, possibly due to the short term memory of the network.

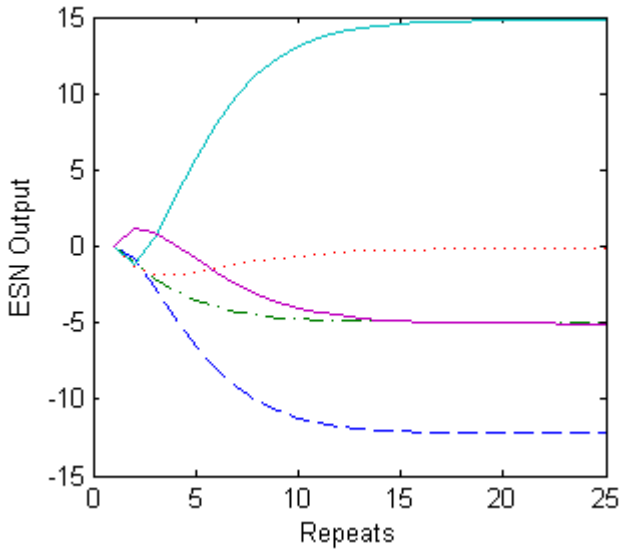


Figure B.6: The effect of presenting more than two presentations to a network trained on only two presentations. The lines representing each class are the same as in Figure B.5

Following this result, two separate training methods were used on ESNs with

the optimized architecture outlined above. ESNs subject to the first training regime (henceforth referred to as ESN-Op) were presented with the optimal 60 input repeats during both training and testing, while ESNs subject to the second training regime (henceforth referred to as ESN-2R) were given two input repeats during training and testing. By doing this, it would be possible to determine exactly how necessary reservoir stabilisation is for good performance.

B.3 Results

Table B.1 gives the performance of each technique on the benchmark datasets. These results show that on average, both ESN architectures achieved better performance than all of the other classifiers, with the exception of the MLP. This confirms that the input clamping methodology can provide good performance on static data when compared to other widely-used classifiers.

Technique	**Dataset Classification Accuracy**		
	Iris	**WDBC**	**Average**
ESN_{Op}	96.37%	95.17%	95.77%
ESN_{2R}	97.69%	94.90%	96.30%
ELM	91.20%	99.96%	95.58%
SVM	94.67%	92.75%	93.71%
MLP	98.63%	95.99%	97.31%

Table B.1: Benchmark dataset results.

Table B.2 shows that, when applied to the leaf sample spectroradiometry dataset, the two ESN training regimes used here both produce good test data classification performance comparable to the ELM and SVM techniques. The MLP, in contrast, was the worst performing technique in all but one category. While it was adept at dealing with the relatively straightforward benchmark datasets, the more challenging real-world dataset proved to be too complex for the MLP. The ELM was the best at distinguishing between non-rhododendron samples, having achieved perfect classification for these during testing. Despite this, the two ESNs were found to be better at correctly identifying rhododendron samples and also gave slightly better overall classification

accuracy. The SVM produced good results, but was outperformed by both ESN training regimes in all but three categories. Given that the focus of this task was on detecting rhododendron and that, although they did not classify the other species perfectly, the ESNs misclassified very few of the non-rhododendron samples as false positives, the ESNs offered the best performance of the CI techniques used for this particular problem. Curiously, the SVM and ELM gave very similar overall performance, with the SVM proving to be marginally better at rhododendron detection, but with the ELM superior at classifying the other plant species. It would seem that while other classifiers can offer good performance, the recurrent reservoir in an ESN adds certain discriminatory advantages that are not available to other methods.

Test	**Technique**				
	ESN$_{Op}$	**ESN$_{2R}$**	**ELM**	**SVM**	**MLP**
Accuracy	96.71% (0.016)	96.49% (0.017)	96.05% (0.007)	96.07% (0.027)	71.66% (0.223)
Rhod. MCC	0.9330	0.9291	0.9202	0.9214	0.5675
Rhod. Identified	95.34%	95.01%	93.72%	94.42%	61.53%
Beech Identified	99.90%	99.92%	100.00%	100.00%	0.23%
Holly Identified	99.84%	99.71%	100.00%	99.66%	85.13%
Ivy Identified	98.88%	98.70%	100.00%	99.66%	85.13%
Laurel Identified	96.63%	96.78%	100.00%	96.28%	91.67%
Beech Misidentified as Rhod.	0.09%	0.07%	0.00%	0.00%	2.19%
Holly Misidentified as Rhod.	0.05%	0.12%	0.00%	0.97%	4.36%
Ivy Misidentified as Rhod.	0.91%	0.95%	0.00%	0.24%	8.53%
Laurel Misidentified as Rhod.	3.01%	2.82%	0.00%	3.09%	2.30%

Table B.2: Leaf sample spectroradiometry results. The standard deviation for the classification accuracy is given in brackets.

B.4 Discussion

Looking at ESN-Op and ESN-2R, it would appear that there is very little difference between the two techniques in terms of the overall accuracy. ESN-Op marginally outperformed ESN-2R on seven out of the eleven performance measures, including the critical classification accuracy, percentage of rhododendrons identified and rhododendron MCC measures, but the difference was minimal. So long as the number of repeated

input pattern presentations used during testing is kept the same as the number of clamped presentations used during training, this result suggests that there is no need to clamp the reservoir neurons until reservoir stabilisation has occurred. This is contrary to results reported in the literature elsewhere, where it was proposed that the activations of the reservoir units needed to be allowed to settle in order to provide the best ESN performance [357, 358, 452, 453]. The most significant contribution of the work in this appendix, therefore, is that it is possible to obtain accurate results while the reservoir is still unstable, so long as the number of repeated input presentations used during testing is the same as during training. The need for algorithms to determine the point at which the reservoir units can be deemed to have stabilised are thus rendered unnecessary.

B.5 Conclusion

ESNs were optimized for static pattern classification and the effect of altering spectral radius, leak rate, input scaling, reservoir size and input pattern repeats on the behaviour of both reservoir and output units was investigated. The performance of these was then tested by presenting them with a challenging classification dataset and comparing them to MLP, ELMs and SVMs approaches.

Two different ESN training regimes were employed and in each case it was found that the optimized ESNs outperformed the other classification techniques. Only the MLP produced better classification accuracy on the benchmark datasets, although the MLP was later unable to deal with the more complex leaf sample spectroradiometry dataset. The fact that the ELM and SVM achieved very similar performance implies that the recurrent connections within the reservoir are important for providing optimal classification performance, as does the fact that the optimal ESN spectral radius was greater than zero. Contrary to results reported in the literature, it was found that it is possible to obtain similar classification capability by presenting the input pattern only twice, even though the ESN reservoir might still be in a highly unstable state at this stage. This idiosyncratic network behaviour needs further investigation, in particular how the spectral radius affects final output designations and the function of

the unusually small reservoir.