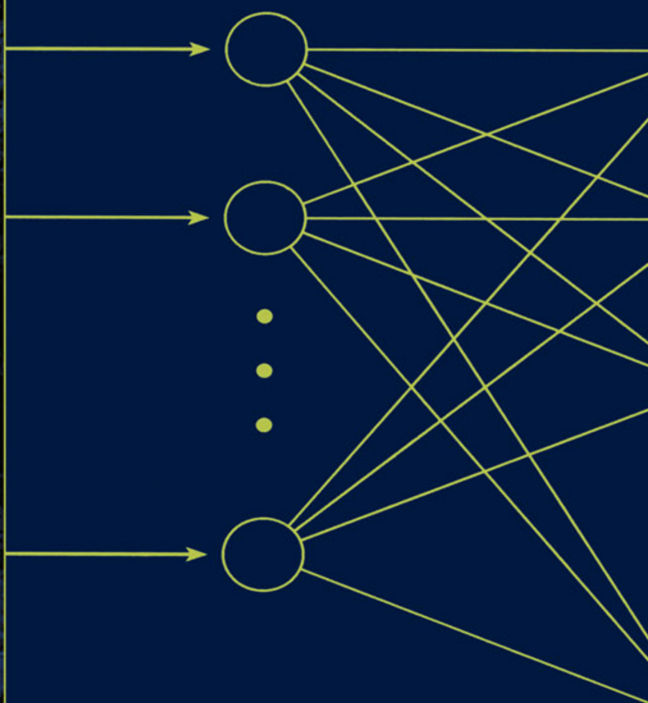
NEURAL NETWORKS

for hydrological modelling

Robert J. Abrahart, Pauline E. Kneale
& Linda M. See



NEURAL NETWORKS FOR HYDROLOGICAL MODELLING

Neural Networks for Hydrological Modelling

ROBERT J.ABRAHART University of Nottingham

PAULINE E.KNEALE University of Leeds

LINDA M.SEE University of Leeds

A.A.BALKEMA PUBLISHERS
Leiden/London/NewYork/Philadelphia/Singapore

Library of Congress Cataloging-in-Publication Data A Catalogue record for this book is available from the Library of Congress

Cover design: Miranda Bourgonjen

Published by: A.A.Balkema Publishers (Leiden, The Netherlands), a member of Taylor & Francis Group plc. http://balkema.tandf.co.uk and www.tandf.co.uk
This edition published in the Taylor & Francis e-Library, 2005.
To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.

ISBN 0-203-02411-7 Master e-book ISBN

ISBN - (OEB Format)
ISBN 90 5809 619 x (Print Edition)

Contents

List of Contributors

Robert J.Abrahart School of Geography, University of Nottingham, Nottingham NG7 2RD, UK

James E.Ball Water Research Laboratory, School of Civil and Environmental Engineering, The University of New South Wales, Sydney, New South Wales, Australia

Christian W.Dawson Department of Computer Science, Loughborough University, Leicestershire, LE11 3TU, UK

Simon A.Corne correspondence c/o Prof Pauline Kneale, School of Geography, University of Leeds, Leeds, LS2 9JT, UK

Michael J.Hall UNESCO-IHE Institute for Water Education, Delft, The Netherlands

Giles M.Foody School of Geography, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

Claire E.Imrie Department of Environmental Science and Technology, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2AZ, UK

Pauline E.Kneale School of Geography, University of Leeds, Leeds, LS2 9JT, UK

Kin Choi Luk correspondence c/o Prof James Ball, Water Research Laboratory, School of Civil and Environmental Engineering, The University of New South Wales, Sydney, New South Wales, Australia

Anthony W.Minns WL|Delft Hydraulics, Delft, The Netherlands

Stan Openshaw correspondence c/o Prof Pauline Kneale, School of Geography, University of Leeds, Leeds, LS2 9JT, UK

George S.Pankiewicz Satellite Applications Section, The Meteorological Office, London Road, Bracknell, RG12 2SZ, UK

Linda M.See School of Geography, University of Leeds, Leeds, LS2 9JT, UK

Asaad Y.Shamseldin School of Engineering, Department of Civil Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT

Henk van den Boogaard WL|Delft Hydraulics, Delft, The Netherlands

Susan M.White Institute of Water and Environment, Cranfield University, Silsoe, Bedfordshire, MK45 5DT, UK

Robert L Wilby Environment Agency, Trentside Offices, Nottingham, NG2 5FA, UK

Preface

This book is intended as an introduction to those who are new to neural network hydrological modelling and as a useful update for those who have been experimenting with different tools and techniques in this area. The scope for applying neural network modelling to hydrological forecasting and prediction is considerable and it is only really in the last five to ten years that it has been tried and tested. The various chapters show that while rainfall runoff forecasting is the main area of research, neural networks are also used in ecological, fisheries, water quality, sediment, groundwater and many other water related applications. The scope is considerable because a neural network works in an equation free environment so that economic, social, hydrological and chemical data can be integrated on an equal basis. Neural networks are often denigrated as black box solutions, but they are sophisticated black boxes, which can produce very useful results. We hope that this book will encourage further users to get involved and experiment.

Each of the chapters has been the subject of an independent review and we are grateful for the many comments and time involved. We are also grateful to the authors for responding to our comments and the reviewers' input and for making the changes requested.

Robert Abrahart	Pauline Kneale	Linda See
Nottingham	Leeds	Leeds

1
Why Use Neural Networks?

PAULINE E.KNEALE, LINDA M.SEE
School of Geography, University of Leeds, UK

ROBERT J.ABRAHART
School of Geography, University of Nottingham, UK

ABSTRACT: Neural networks are one computational methodology for hydrological forecasting. Although widely used in other research and application fields they are employed less by hydrologists than might be expected given the data driven nature of the applied problems to be solved. Neural networks provide a modelling route that can be helpful when there is enough data to link x to y and especially where results are needed in real time. This chapter introduces neural network issues generally, setting them in a wider modelling context and provides a framework link to later chapters which handle neural network topics in detail.

1 INTRODUCTION

Neural networks (NN) are an alternative and complementary set of techniques to traditional models. NN can be thought of as computational pattern searching and matching procedures that permit forecasting without an intimate knowledge of the physical or chemical processes. For the hydrologist this technique has considerable appeal, provided the absence of a detailed process explanation can be borne.

NN rely on the provision of adequate data sets, and where these are available, NN may be programmed to search for patterns within the data. On the basis of this pattern-matching, forecasts are made on independent data sets first for model validation and then for operational purposes. NN are one approach within the broader hydroinformatics framework which emerged in the 1990s as a route to managing information overload in an effective way (Govindaraju & Rao, 2000). Price (2000) recognises four strands to hydroinformatics, the mathematical and physical sciences understanding, the handling of data and the human cultural element. One of the significant strengths of the NN approach is that it can handle all data types.

The challenge of managing water in its many dimensions and applications calls for techniques which can link a myriad of components, from the complexity of hydraulics and water quality, to financial planning and social agendas. This is a move towards a holistic or integrated approach to modelling. The techniques available to the hydrologist are many and varied, each with their own advantages and drawbacks. The vision of 1970s modellers (Freeze & Harlan, 1969) that forecasting problems would be cracked when computers became powerful enough to handle very complex equations and infinitely large data sets has become a receding, but by no means disappearing, target. Natural environmental variability, the uniqueness of catchments, system chaos and the complexities of scale integration, together with the expense of data acquisition, make the forecasting task challenging. Flood modelling at the basin scale with fine mesh models requires prodigious amounts of computer time, but Beven and Feyen (2002) consider that these goals are coming nearer as visualisation and virtual game technologies advance, so the ambitions for catchment-wide 4D modelling are getting closer. NN do not in any way aim to replace such models but they can provide a very fast forecasting system that is operationally available in very short time frames. NN do not compete with distributed models but rather offer alternative and complementary ways of tackling forecasting problems.

This text is aimed both at those using NN in research for the first time and at those wanting to review recent examples of NN hydrological applications. It is not intended as a manual but should be used as a supportive guide for anyone wanting to experiment with this type of modelling. This chapter introduces some of the basic ideas and background behind the NN approach, particularly for those who are new to this methodology If you are already familiar with NN techniques then skipping to later chapters may be helpful. The sections that follow provide a link to the more detailed materials in the main chapters and to broader applications.

2 THE BASIC IDEAS OF THE NN APPROACH

To understand the basic ideas behind the NN approach, let's look at a simple hydrological example. Imagine that you could access data banks of hydrological information. Suppose that the databases contain stage data recorded every ten minutes at 4 points on a river (C-F), precipitation data from a gauge (A) collated every fifteen minutes from a radar system (B) and weekly soil moisture data (G) (Fig. 1.1). Your first task is to decide which station you want to model. In a conventional approach you might choose to model stage at F using all the data sets including past records from site F as inputs. Alternatively you might (eccentrically) decide to forecast precipitation at B using all the data (Fig. 1.2). The point is that the NN has no knowledge of the spatial relationship between the sites as seen on a map nor any idea about what it is being modelled. The NN only seeks the relationships between the input and output data and then creates its own equations to match the patterns in an iterative manner.

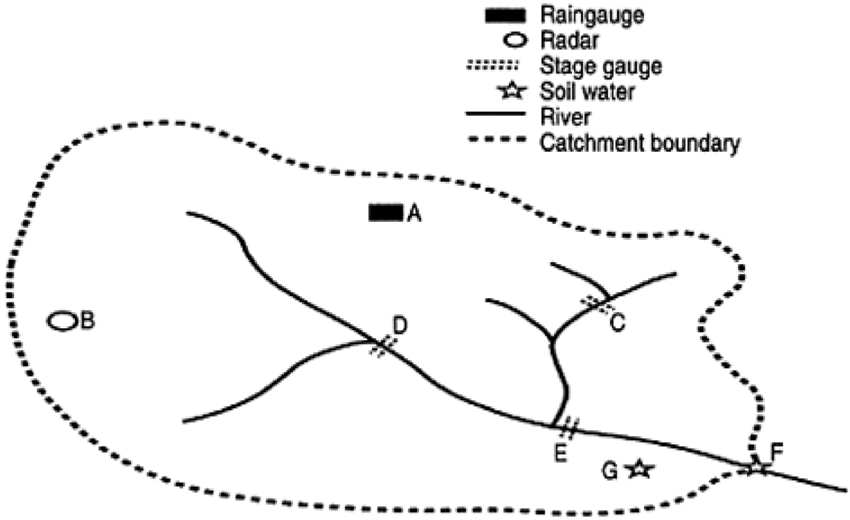


Fig. 1.1. Catchment X.

Input Data	Potential Forecast	Potential Forecast	Potential Forecast
Raingauge A (Continuous)			
Radar B (15 minute)			
Stage C (15 minute)			
Stage D (15 minute)	Stage at F	Stage at C	Radar B
Stage E (15 minute)			
Stage F (15 minute)			
Soil water deficit G (weekly)			

Fig. 1.2. Potential models.

Continuing with this example, a forecaster might choose to start modelling with all the data, and look to eliminate those data sets that are not contributing significantly to the output to find the most parsimonious approach, thereby saving data collection and data processing time. It may be that operationally the best forecasting model for stage at station E is the stage at station E in previous time steps. In forecasting terms this may be the cheapest and most accurate model, but a user might choose a less optimal model that includes an upstream site in real-time forecasting in case there are data transfer problems during a real-time event. The ‘best’ computational solution may not be the ideal practical solution. This is a forecasting approach where there are many decisions to be made by the user.

NN models are variously described as mimicking the parallel-information processes of the brain. However, a typical human brain is thought to contain 10^{11} neurons, each receiving input from an additional 5000 to 15000 neurons. The average worm has approximately 10^3 connections. A NN is likely to have connection numbers in the 10–

1000 range so a NN would be considered to be of sub-worm complexity (Openshaw, 1997). Comparisons of this kind are illuminating at one level but do not inspire immediate confidence in the technical merits of NN as a sophisticated analytical technique. It is important to see why they are so described and evaluate this description (see Chapters 2, 3 and 4).

The brain analogy is helpful for new users. NN are a mathematical representation of a process that operates like nerve cells. Each network is made up of nodes and links, much like the nerve cells and messages in a nervous system. The user defines the architecture of the network and following trial and error runs, this mathematical representation of the NN becomes the model framework. For example, trials may show that the radar data at B (Fig. 1.2) may not be contributing useful additional information so that node would then be removed in further model trials.

Forecasting should follow in three clearly separate stages of NN development, stages that are kept separate to make comparisons as accurate as possible. In 'training mode' the output pattern at say station F (Fig. 1.1) is linked to as many of the input nodes (A-G) as desired and the patterns are defined. In the training phase, part of the total data set is used. Conjusingly, NN scientists may also refer to a 'validation' dataset used at this stage to ensure the model is not overtrained. The data may be temporally contiguous or it may be selected as being representative across the whole period. This can be important if it is thought that there may be systematic change on a catchment across the whole period, arising for example from land use change. This is followed by a 'testing phase' when the model is tested using data sets that were not used in training. If the forecasts are satisfactory then the model may be used in an 'operational' or 'real time mode' to generate live forecasts. These live forecasts are evaluated against real events. Measures of accuracy of a model should ideally refer to forecast performance in the real-time mode or independent validation mode.

Once established the NN can be developed or updated as more data become available. In this sense NN are dynamic in that the operator can adjust and adapt them as change occurs, which makes them potentially very valuable in hydrological operational modes. In this simple hydrological example it would be logical for an operator to update the forecasting networks at the end of each wet season to take account of recent precipitation events and thus give the users additional confidence in the modelling. Because the processing speeds of NN are very high, in practice a model can be updated and redeveloped in real time to take account of new or changing circumstances if required (Abrahart, 2003).

NN may be regarded as data driven techniques but it is argued here that their flexibility in data handling and the ability to solve problems where it is effectively impossible to get primary data, as in groundwater modelling solutions (Ouenes, 2000; Zio, 1997) and with the added complexity of groundwater chemistry (Gumrah *et al.,* 2000), or where processes are highly non-linear and spatially and temporally variant (Islam & Kothari, 2000), makes these techniques well worth exploring. If a distributed modelling solution is not available but the data are, then this may be a useful approach. Certainly many NN applications have been prompted by unsatisfactory results with regression and time series techniques.

3 SOME ANTECEDENTS

The pattern for classifying hydrological modelling approaches was articulated by Dooge (1977). His three phase black box empirical, lumped and physically-based distributed model distinction is very widely recognised. This lead to an acceptance of an apparent hierarchy in quality of approach with the 'simple black box' considered to be less acceptable than the more mathematically rigorous, theoretically based distributed approach. While this distinction is academically valid, it is not always helpful in practical terms. The advice to use the simplest tool that will do the job is appropriate in practical and operational modelling. If the data are available and the problem is linear then using linear regression is fine. The unit hydrograph and rational formulas survive because they are practical tools that supply useful answers.

While NN are a relatively recent technique for hydrologists they have an established antecedence which Govindaraju and Rao (2000) acknowledge as starting in the 1940s. NN concepts arrived with McCulloch and Pitts' (1943) work but their practical use followed Rumelhart *et al.*'s (1986) development of the back propagation neural network (BPNN) algorithm which lead to a plethora of applications in many subjects. Various text books in the 1990s generated some interest (Masters, 1993; Cruz, 1991) and the first hydrological applications were probably Daniell (1991), French *et al.* (1992) and Hall and Minns (1993). So for hydrologists this is a young technique with a short pedigree. But there has been a rapid uptake and a positive blossoming in conferences and publications. Good generic texts on the subject include Bishop (1995), Haykin (1998) and Picton (2000) but there are many other sources available.

Various authors describe NN models as black box and dismiss them as empirical, and therefore by definition, as inferior. Certainly the calculations are 'set-up' by the modeller but the nature of the relationship between variables is found by the computer (see Chapter 2). So in this sense NN are input-output models. They are therefore vulnerable to the problems of inadequate data and a less than thoughtful forecaster. However, they have the strength when compared for example with ARMA and regression approaches that non-linearity in relationships will be captured (Hsu *et al.,* 1995) and the black box can be looked into in detail if the forecaster wishes (Abrahart *et al.,* 2001; Wilby *et al.,* 2003). The early hydrological literature is dominated by rainfall-runoff forecasting applications, probably because these represent a conceptually straight-forward starting point. There are some lengthy records for both variables for training and validation, and the solutions are evidently non-linear; this theme is well reviewed in this text in Abrahart (Chapter 2), by Minns and Hall (Chapter 9), and in the GIS application of rainfall modelling discussed by Ball and Luk (Chapter 10).

Alternative introductions to NN modelling in hydrological contexts include Maier and Dandy (2001b) who provide a sound introductory overview in the context of cyanobacterium and salinity modelling in River Murray, and Dawson and Wilby (2001). In a Special Issue of *Computers and Operations Research,* Gupta and Smith (2000) cover a significant range of non-hydrological examples, and the business applications considered are of interest to those considering modelling economic and management aspects of water supply and water management.

The hydrological applications from the last seven years fall into a series of broad categories and styles of modelling. There are three main types of NN: backpropagation

(BPNN), radial basis function network (RBFN) and the self-organising feature map (SOFM). Abrahart addresses each style in detail in Chapter 2 and as later chapters will indicate, backpropagation neural networks (BPNN) dominate for forecasts at specified points such as river stage, whereas SOFM mapping algorithms are employed to predict spatial patterns.

Running models with multiple inputs implies the availability of appropriate data sets, a problem for any field-based hydrological work. However where data are captured in remote sensing operations and GIS programmes the NN approach can be very powerful as Foody's Chapter 14 indicates. Gautam *et al.* (2000) have the advantage of a well instrumented catchment at Tono, Japan, providing meteorological, runoff and soil moisture content data for their stream flow forecasts. This is a luxury not available in most areas; however, the results are satisfactory indicating that the NN technique may be of benefit for small catchment forecasts and perhaps in agricultural applications. To forecast soil texture from remotely sensed maps Chang and Islam (2000) use brightness temperature and remotely sensed soil moisture. The soils are classified into six classes. Forecasting the permeability of oil reservoirs, Bruce and Wong (2002) use an evolution NN algorithm to solve a forecasting problem bedevilled by solutions that can be trapped in local minima using backpropagation.

NN are not necessarily run in isolation. In linking NN within their models Maskey *et al.* (2000) for example show how NN can be used with process models to calculate travel times of groundwater pollution plumes in response to well injections and pumping in an experiment to optimise a groundwater clean up programme. The flexibility to use a NN within a broader modelling framework is an attractive use of the technology.

While hydroinfomatics primarily concentrates on aquatic forecasting, for some authors NN technologies assist in the objective inclusion of social and economic dimensions. Jonoski (2002) looks towards a sociotechnological role for the hydrological forecaster where these additional dimensions are an integral part of the modelling process in what they define as Network Distributed Decision Support Systems.

The reported use of NN models is broad and considerable in statistical and engineering applications (Ma *et al.,* 2001; Venkateswaran *et al.,* 2002). Their operational rather than research use is also extensive in a wide range of industries: in mining to identify rocks that can be obstructive (Cabello *et al.,* 2002), converting speech to text (Wang *et al.,* 2002), monitoring wear on machine tools (Scheffer & Heynes, 2001), automating wastewater treatment and chemical monitoring (Zyngier *et al.,* 2001), forecasting sea ice conditions in Canadian waters (El-Rabbany *et al.,* 2002), coffee bean blending (Tominaga *et al.,* 2002), flavour of blackcurrants (Boccorh & Paterson, 2002) and identifying corrosion rates on aircraft parts (Pidaparti *et al.,* 2002).

Govindaraju and Rao (2000) suggest that the adoption of NN techniques by hydrologists has been constrained by the relative newness of the technique, and its position as an empirical methodology in a subject which struggled to get rid of its soft empirical subject image and emerge as an accepted physics-based discipline. Maier and Dandy (2000) reiterate the essential need for thoughtful applications: 'In many applications, the model building process is described poorly, making it difficult to assess the optimality of the results obtained'. Flood and Kartam (1994) also add a relevant observation: 'There is a tendency among users to throw a problem blindly at a NN in the hope that it will formulate an acceptable solution'. Maier and Dandy's (2000) paper would be a great place for many modellers to start. The authors review the issues for modelling with a wide range of practical examples.

Much of this text exemplifies the need for a systematic approach to thinking through the methodological approaches and constraints. Then to apply these approaches to relevant hydrological issues. We would argue that it represents an opportunity to model with greater freedom and speed some of the ‘difficult’ multifaceted problems in hydrology.

However, it is important to point out that NN are not magic boxes. There is an extensive mathematical background and theory that has underpinned their development and for those mathematically inclined this is a rich area of investigation. The NN technique cannot be criticised as theoretically unsupported and therefore unsound. Users can decide to try the NN approach without exploring the mathematics in detail and to take advantage of the plethora of freeware or shareware off-the-shelf packages. This is really no different to users taking some of the more advanced codes in SPSS for partial canonical correlation. *Caveat emptor* always applies, and as the authors of Chapters 2–5 which look at the basics of different types of modelling approaches emphasise, it is vital to understand the data and programming decisions involved. But these are explained in practical terms and are on a par with understanding that 3 samples are not enough for multiple regression with 6 variables and that ANOVA values require significance tests. In other words try it for yourself.

4 WHERE DO I FIND THEM? NN PLATFORMS

Individual chapters in the book direct you to specific software sources, while this section provides a brief overview of the sites available. There are a very substantial number of companies and web sites offering NN software and a range of product support packages. The most useful starting point might be ; a users site that is updated monthly. As it says: ‘its purpose is to provide basic information for individuals who are new to the field of neural networks’. There are software programs to download via ftp sites, for use on multiple platforms. Table 1.1 provides a short starting list of websites that you might check out, while later chapters point users to particular software packages.

You can also find NN embedded within data mining software such as Clementine or IBM’s Intelligent Data Miner. Data mining is a popular term in

Table 1.1. NN software suppliers and web sites, a starting list.

Software name and company	Web sites
Free or Share ware	
Ainet—Freeware Neural Network	www.ainet-sp.si/
GENESIS and PGENESIS 2.2	http://www.bbb.caltech.edu/GENESIS
KarstenKutza—	http://www.geocities.com/CapeCanaveral/1624/
NEURALFUSION—	http://www.neuralfusion.com/

PDP Plus, MIT Press	http://www.cnbc.cmu.edu/PDP++/PDP++.html
SNNS, Stuttgarter Neural Network Simulator, University of Tuebingen, Germany	http://www-ra.informatik.uni-tuebingen.de/SNNS
Commercial packages	
Brain Maker, California Scientific Software Company	www.calsci.com
Cortex-Pro	www.reiss.demon.co.uk/webctx/detail.html
IBM Neural Network Utility, IBM Company	nninfo@vnet.ibm.com
NeuralWorks Professional II Plus, NeuralWare Inc.	http://www.neuralware.com/
Neuro Genetic Optimizer (NGO), Bio Comp Systems Inc.	www.bio-comp.com/pages/neuralnetworkoptimizer.htm
Neuro Shell Predictor, Ward Systems Group Inc.	www.wardsystems.com
NeuroSolutions v3.0, Neuro Dimension, Inc.	http://www.nd.com/
QNET v2000	www.qnetv2k.com
STATISTICA: Neural Networks version 4.0, Statsoft Inc.	http://www.statsoft.com/
Neural Connection, SPSS Inc.	http://www.spss.com/

the business world for all techniques that can be used to turn large amounts of data into useful information, of which NN are only one example. Clearly any package needs evaluation and for the novice the array of software available is confusing. The hydrological NN literature is not awash with citations of software used; some users will have written their own programmes but given the availability of packages this seems as unnecessary today as writing a program to calculate regression. A starter suggestion is the SNNS, Stuttgarter Neural Network Simulator which is well documented and user friendly.

Rather than re-inventing program codes for backpropagation it would seem to be more useful for hydrological forecasters to develop a suitable suite of quality testing procedures. Kneale and See (2003) testing Time Delay Neural Network (TDNN) forecasts use ten tests to compare hydrograph forecast accuracy. It is critical that the tests chosen include those normally used in hydrological model evaluation, such as the Nash and Sutcliffe (1970) index. This permits users to evaluate the forecasts in a consistent and objective manner and compare them to results obtained from traditional hydrological forecasting procedures.

5 SUMMARY

Essentially NN are one of many tools at the disposal of the hydrological researcher. The user defines the independent and dependent variables and has all the normal modelling problems of locating suitable data sets to develop, test and validate the models.

One major advantage the NN approach has over traditional input-output modelling is that it makes fewer demands on the data. Unlike multiple regression, where the constraints of normality in the data distributions are often simply ignored, NN do not make assumptions about the statistical properties of a data set. Data for different variables can be of all types and available on different time or spatial scales. This allows for a flexible approach to data collection and model development. In management models for example weather related, soil dynamic, crop development and agricultural management information can be used as inputs using parameters that are recorded on hourly, weekly, monthly, m^2, hectare and currency scales.

A second major advantage is that in searching for patterns and links in the data sets there are no assumptions of linearity. NN are non-linear pattern identification tools, which is why they are potentially so attractive for tackling the non-linear problems of hydrology.

The powerful potential of NN models to solve ‘hard computational problems’ including those where the underlying ecological relations are not understood was cited by Lek and Guegan (1999). There is a wealth of understanding of hydrological processes at a range of scales from laboratory to hillslope and catchment. But it is not always clear how to write the equations to link processes that are understood at the m^2 scale so that they scale up to the basin scale. NN search for the patterns in the data and therefore have the potential to create the equations that describe the processes operating on the catchment under study As with all modelling an ill-specified NN will generate inadequate to useless forecasts. A good hydrological understanding of the relevant field processes is a pre-requisite of good modelling. That together with enough understanding of the NN to have the confidence to eliminate inessential variables and so define, through experimentation, the most parsimonious but efficient model. The relationship that a NN defines must be sought again in data for different catchments, the chosen model reflects the complex interactions within the specified data sets. However the final selection of parameters, model architectures and training times for any model will be helpful guidance for forecasters applying the NN approach in comparable catchments, speeding up the development of future models.

The potential speed of model development is a factor that most NN users find attractive. Forecasting algorithms are available from a range of web and shareware sources. Data acquisition is part of every modelling process but the forecaster then moves into model development and testing. Our experience of river stage modelling is that computational run times are a matter of minutes and validation and independent forecasting is effectively instantaneous (Kneale *et al.,* 2000). A forecaster should not find this element of the hydrologist’s toolbox more difficult to apply than partial Canonical Correspondence Analysis, a GIS system, an ARMA model or complex process-based software applications.

It may be that the role of NN is as part of a larger modelling framework, where the NN is one element in a data handling and management tool. Most of this text is concerned with the application of NN to solving specific hydrological problems with the NN as the primary technique, but this is just one potential role. The considerable scope for links to GIS models is made explicit in Foody's Chapter 14. There is a dominance of rainfall runoff applications which are explored more fully in various chapters. NN were developed to mirror biological activities, their non-linear flexibility makes them very attractive for forecasting complex multi-disciplinary hydrological problems like crop and fish stock management, pesticide leaching and runoff from hill-slopes, and groundwater pollution and abstraction interactions (Freissinet *et al.,* 1999; Tansel *et al.,* 1999; Morshed & Powers, 2000; Tingsanchali & Gautam, 2000).

Where the NN fits in the mosaic of techniques for the hydrologist is still uncertain but we hope these chapters will encourage each reader to see its relevance in a range of applications and to try the techniques.

REFERENCES

Abebe, A.J., Solomatine, D.P. & Venneker, R.G.W. 2000. Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. *Hydrological Sciences Journal* 45(3):425–436.

Abrahart, R.J. 2003. Neural network rainfall-runoff forecasting based on continuous resampling. *Journal of Hydroinformatics* 5(1):51–61.

Abrahart, R.J., See, L. & Kneale, P. 2001. Applying saliency analysis to neural network rainfall-runoff modelling. *Computers and Geosciences* 27: 921–928.

Berry, M.J. & Linoff, G. 1997. *Data Mining Techniques: For Marketing, Sales and Customer Support.* Chichester: Wiley.

Berry, M.J. & Linoff, G. 2000. *Mastering Data Mining: The Art and Science of Customer Relationship Management.* Chichester: Wiley.

Beven, K. & Feyen, J. 2002. The future of distributed modelling—Special Issue editorial. *Hydrological Processes* 16(2):169–172.

Bishop, C.M. 1995. *Neural Networks for Pattern Recognition.* Oxford: Clarendon Press.

Boccorh, R.K. & Paterson, A. 2002. An artificial neural network model for predicting flavour intensity in blackcurrant concentrates. *Food Quality and Preference* 13(2):117–128.

Bruce, A.G. & Wong, P.M. 2002. Permeability prediction from well logs using an evolutionary neural network. *Petroleum Science and Technology* 20(3–4):317–331.

Cabello, E.E., Sanchez, M.A. & Delgado, J. 2002. A new approach to identify big rocks with applications to the mining industry. *Real-Time Imaging* 8(1):1–9.

Chang, D-H. & Islam, S. 2000. Estimation of soil physical properties using remote sensing and artificial neural network. *Remote Sensing of Environment* 74(3):534–544.

Corne, S. Kneale, P.E., Openshaw S. & See L. 1998. The use and evaluation of artificial networks in flood forecasting. *Proceedings 33rd MAFF Annual Conference.*

Cruz, C. 1991. *Understanding Neural Networks: A Primer.* Cutter Information Corporation, Arlington, MA: 74(3).

Daniell, T.M. 1991. Neural networks—applications in hydrology and water resources engineering. *Proceedings of the International Hydrology and Water Resources Symposium.* National Conference Publication 91/22 Institute of Engineering, Barton, Australia, 3:797–802.

Dawson, C.W. & Wilby, R.L. 2001. Hydrological modelling using artificial neural networks. *Progress in Physical Geography* 25(1):80–108.

Dooge, J.C.I. 1977. Problems and methods of rainfall-runoff modeling. In T.A.Cirani, U.Maione & J.R.Wallis (Eds.) *Mathematical Models for Surface Water Hydrology* (pp. 71–108). Chichester: John Wiley.
El-Rabbany, A., Auda, G. & Abdelazim, S. 2002. Predicting sea ice conditions using neural networks. *Journal of Navigation* 55(1):137–143.
Flood, I. & Kartam, N. 1994. Neural networks in civil engineering I Principles and understanding. *Journal of Computing in Civil Engineering* 8(2):131–148.
Freeze, R.A. & Harlan, R.L. 1969. Blueprint for a physically-based, digitally simulated hydrologic response model. *Journal of Hydrology* 9:237–258.
Freissinet, C., Vauclin, M. & Erlich, M. 1999. Comparison of first-order analysis and fuzzy set approach for the evaluation of imprecision in a pesticide groundwater pollution screening model. *Journal of Contaminant Hydrology* 37(1–2):21–43.
French, M.N., Krajewski, W.F. & Cuykendall, R.R. 1992. Rainfall forecasting in space and time using a neural network. *Journal of Hydrology* 137:1–31.
Gautam, M.R., Watanabe, K. & Saegusa, H. 2000. Runoff analysis in humid forest catchment with artificial neural network. *Journal of Hydrology* 235(1–2):117–136.
Govindaraju, R.S. & Rao, A.R. (Eds.) 2000. *Artificial Neural Networks in Hydrology.* Dordrecht: Kluwer Academic.
Gumrah, F., Oz, B., Guler, B. & Evin, S. 2000. The application of artificial neural networks for the prediction of water quality of polluted aquifer. *Water Air and Soil Pollution* 119(1–4):275–294.
Gupta, J.N.D. & Smith, K.A. 2000. Neural networks in business, Special Issue. *Computers & Operations Research* 27:1017–1247
Hall, M.J. & Minns, A.W. 1993. Rainfall-runoff modelling as a problem in artificial intelligence: experience with a neural network. *Proceedings of the Fourth National Hydrological Symposium,* Cardiff, 5.51–5.57.
Haykin, S.S. 1998. *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice Hall.
Hong, Y-S. & Rosen, M.R. 2001. Intelligent characterisation and diagnosis of the groundwater quality in an urban fractured-rock aquifer using an artificial neural network. *Urban Water* 3(3):217–228.
Hsu, K.L., Gupta, H.V. & Sorooshian, S. 1995. Artificial neural network modelling of the rainfall runoff process. *Water Resources Research* 31:2517–2530.
Islam, S. & Kothari, R. 2000. Artificial neural networks in remote sensing of hydrologic processes. *Journal of Hydrologic Engineering* 5(2):138–144.
Jonoski, A. 2002. *Hydroinformatics as Sociotechnology—Promoting Individual Stakeholder Participation by Using Network Distributed Decision Support Systems.* Rotterdam: Balkema.
Kneale, P.E. & See, L. 2003. Forecasting river stage with artificial neural networks. In G.Clarke & J.C.H Stillwell, (Eds.), *Applied GIS and Spatial Analysis.* Chichester: John Wiley.
Kneale, P.E., See, L., Cameron, D., Kerr, P. & Merrix, R. 2000. Using a prototype neural net forecasting model for flood predictions on the Rivers Tyne and Wear. *Proceedings of the British Hydrological Society, Seventh National Hydrology Symposium,* 3.45–3.49.
Lek, S. & Guegan, J.F. 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* 120(2–3):65–73.
Ma, Q.C., Wang, J.T.L., Shasha, D. & Wu, C.H. 2001. DNA sequence classification via an expectation maximization algorithm and neural networks: A case study. *IEEE Transactions on Systems Man and Cybernetics Part C—Applications and Reviews* 31(4):468–475.
Maier, H.R. & Dandy, G.C. 2001a. Neural network based modelling of environmental variables: A systematic approach. *Mathematical and Computer Modelling* 33(6–7):669–682.
Maier, H.R. & Dandy, G.C. 2001b. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software* 15(1):101–124.
Maskey, S., Dibike, Y.B., Jonoski, A. & Solomatine, D. 2000. Groundwater model approximation with artificial neural network for selecting optimum pumping strategy for plume removal. In

O.Schleider & A.Zijderveld (Eds.), *AI Methods in Civil Engineering* (pp. 45–58). Rotterdam: Cottbus

Masters, T. 1995. *Neural, novel & hybrid algorithms for time series prediction.* New York: John Wiley & Sons.

McCulloch, W.S. & Pitts, W. 1943. A logic calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5:115–133.

Morshed, J. & Powers, S.E. 2000. Regression and dimensional analysis for modeling two-phase flow. *Transport In Porous Media* 38(1–2):205–221.

Nash, J.E. & Sutcliffe, J.V. 1970. River flow forecasting through conceptual models, I, A discussion of principles. *Journal of Hydrology* 10:282–290.

Ouenes, A. 2000. Practical application of fuzzy logic and neural networks to fractures reservoir characterization. *Computers and Geosciences* 26(8): 953–962.

Picton, P. 2000. *Neural networks* (2nd ed.). Basingstoke: Palgrave.

Pidaparti, R.M., Jayanti, S. & Palakal, M.J. 2002. Residual strength and corrosion rate predictions of aging aircraft panels: Neural network study. *Journal of Aircraft* 39(1):175–180.

Pokrajac, D. 2001. Groundwater in urban areas. *Urban Water* 3(3):155.

Price, R.K. 2000. Hydroinformatics and urban drainage: an agenda for the beginning of the 21st century. *Journal of Hydroinformatics* 2:85–86.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. 1986. Learning internal representations by error propagations. In D.E.Rumelhart & J.L.McClelland, (Eds.), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* 1 (pp. 318–362). Cambridge, MA: MIT Press.

Scheffer, C. & Heyns, P.S. 2001. Wear monitoring in turning operations using vibration and strain measurements. *Mechanical Systems and Signal Processing* 15(6):1185–1202.

Tansel B., Jordahl C. & Tansel, I. 1999. Mapping of subsurface contaminant profiles by neural networks. *Civil Engineering and Environmental Systems* 16(1):37–50.

Thuraisingham, B. 1998. *Data Mining: Technologies, Techniques, Tools, and Trends.* New York: CRC Press.

Tingsanchali, T. & Gautam, M.R. 2000. Application of tank, NAM, ARMA and neural network models to flood forecasting. *Hydrological Processes* 14(14):2473–2487.

Tominaga, O., Ito, E., Hanai, T., Honda, H. & Kobayashi, T. 2002. Sensory modeling of coffee with a fuzzy neural network. *Journal of Food Science* 67(1):363–368.

Venkateswaran, S., Rai, M.M., Govindan, T.R. & Meyyappan, M. 2002. Neural network modeling of growth processes. *Journal of the Electrochemical Society* 149(2):G137–G142.

Wang, W.J., Liao, Y.F. & Chen, S.H. 2002. RNN-based prosodic modeling for mandarin speech and its application to speech-to-text conversion. *Speech Communication* 36(3–4):247–265.

Wilby, R.L., Abrahart, R.J. & Dawson, C.W. 2003. Detection of conceptual model rainfall-runoff processes inside an artificial neural network. *Hydrological Sciences Journal* 48(2):163–181.

Zio, E. 1997. Approaching the inverse problem of parameter estimation in groundwater models by means of artificial neural networks. *Progress in Nuclear Energy* 31(3):303–315.

Zyngier, D., Araujo, O.Q.F. & Lima, E.L. 2001. Hybrid estimator for a wastewater treatment process. *Latin American Applied Research* 31(5):507–511.

2
Neural Network Modelling: Basic Tools and Broader Issues

ROBERT J.ABRAHART
School of Geography, University of Nottingham, UK

ABSTRACT: The purpose of this chapter is to define and illustrate the basic terms and concepts involved in neural network modelling. The main neurohydrological modelling tools used to date are introduced. The chapter also includes an insight for new users into the scope and function of potential neural network hydrological modelling applications with respect to the broader hydrological picture.

1 INTRODUCTION

This chapter discusses the main elements in the neural network (NN) toolbox; it also addresses the 'what' and 'when' of NN hydrological modelling. Section 2 contains a brief introduction to the mechanisms and procedures involved—which includes a discussion on architectures and learning; while Section 3 contains a detailed description of the most popular tools that have been used in the field of water related research. Sections 2 and 3 are intended to complement one another and are designed to impart the minimum amount of information that would be required to understand the various operations and processes that are adopted in neurohydrological modelling. There are several respected sources that can be consulted for a more authoritative and comprehensive discussion on generic NN modelling items or issues of interest. Bishop (1995) and Masters (1995) are good academic texts; each book contains a copious amount of in-depth material. Reed and Marks (1999) is oriented towards the developer and practitioner. It describes selected techniques in sufficient detail, such that real-world solutions could be implemented, and technical issues or operational problems could be resolved. Section 4 illustrates the range of different hydrological possibilities and potentials that exist in which to develop and implement a neural solution. Section 5 highlights the numerous opportunities and benefits that are on offer and further strengthens the argument for increased research into the provision of data-driven models. Sections 4 and 5 are thus intended to bolster appeal and to encourage uptake amongst interested parties; the exploration and testing of unorthodox strategies and alternative mindsets can indeed be a rewarding experience that leads to fresh insights and discoveries.

2 WHAT IS A NEURAL NETWORK?

NN are structures which forecast and predict through pattern matching and comparison procedures. NN tools are, in most cases, non-linear adaptive information processing structures that can be ‘described mathematically’ (Fischer, 1998). NN can exist as real-time hard-wired mechanisms, software simulators, optical processors and specialized neurocomputing chips (Taylor, 1993) and their computational elements are generic. NN software simulation programs, written in a standard high-level language, are the most common form.

There are a number of commercial and public domain simulators that users can select from, depending upon their preferred computer platform, and the sophistication offered in such packages provides a significant attraction. Catalogues of established software and shareware can be found on the World Wide Web e.g. NEuroNet (2001) or Sarle (2002). See Table 1.1 for a more comprehensive list. It is both an advantage and a potential drawback that users can download and install powerful NN products and packages with little or no real effort e.g. Stuttgart Neural Network Simulator (SNNS Group, 2003). Trained NN solutions can also be converted into dedicated 3GL (Third Generation Language e.g. C++) functions, for amalgamation into home-grown software products, or linked to commercial applications using a run-time connection based on standard software libraries (e.g. DLL). This is a major advantage for users, and especially new users wishing to experiment with the technique, but all users must be clear about the pros and cons of this modelling procedure.

2.1 *Network architecture*

NN are constructed from two basic building blocks: processing units (also referred to as elements or nodes or neurons) and weighted connections (also referred to as arcs or edges or links). These components and their respective organisation, into a set of interconnected items, form the ‘network architecture’.

Maren (1991) has suggested that the architectural configuration can be described at three basic levels and this framework is used to explain the com ponents here:

(a) ***Microstructure.*** The characteristics of each processing unit in a network.
(b) ***Mesostructure.*** The manner in which a network is organised, including such features as the number of layers, the pattern of connections, and the flow of information.

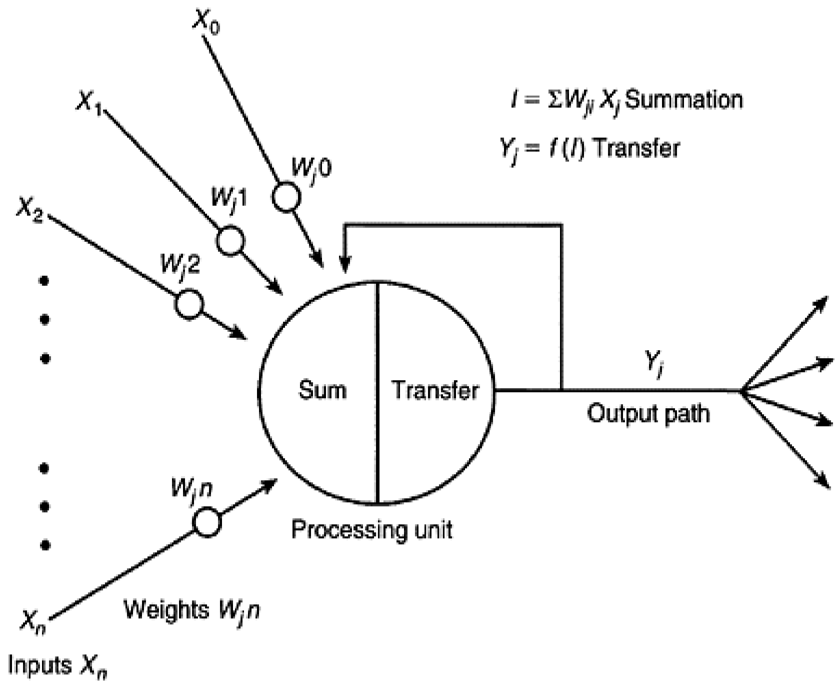


Fig. 2.1. The microstructure of a neural network model in terms of processing units.

(c) ***Macrostructure.*** The manner in which a number of 'networks' are linked together, interacting with each other to build a more complex solution, for more demanding tasks.

Figure 2.1 illustrates the standard organization of an individual processing unit—which is the microstructure. Each processing unit can have numerous incoming connections, that arrive from other processing units, or from the 'outside world' $X_1 \ldots X_n$. The 'outside world' could be raw input data, or outputs produced from another forecasting model, that exports data to the NN. The connections function as unidirectional paths that conduct signals or data, and transmit their information in a predetermined direction. These are the user-defined 'input connections' and there is no upper limit on their number. There is also a program default input, termed bias, that is a constant X_0=1. Each processing unit first computes an intermediate value that comprises the weighted sum of all its inputs $I=\sum W_{ji} X_j$. This value is then passed through a transfer function $f(I)$, which performs a non-linear 'squashing operation'. The user can opt for default transfer functions or in certain software packages define their own—the standard options being logistic, sigmoid, linear, threshold, gaussian and hyperbolic tangent—with the selection of an appropriate transfer function being dependant upon the nature of each specific problem and its proposed solution. Shamseldin *et al.* (2002) explored the application of several different transfer functions to the amalgamation of multi-model hydrological forecasts and found that in most cases a logistic function provided the best results and an arctan function produced the worst results. Each processing unit can have numerous output connections, that lead to other processing units or to the 'outside world', and again there is no restriction on their number. Each

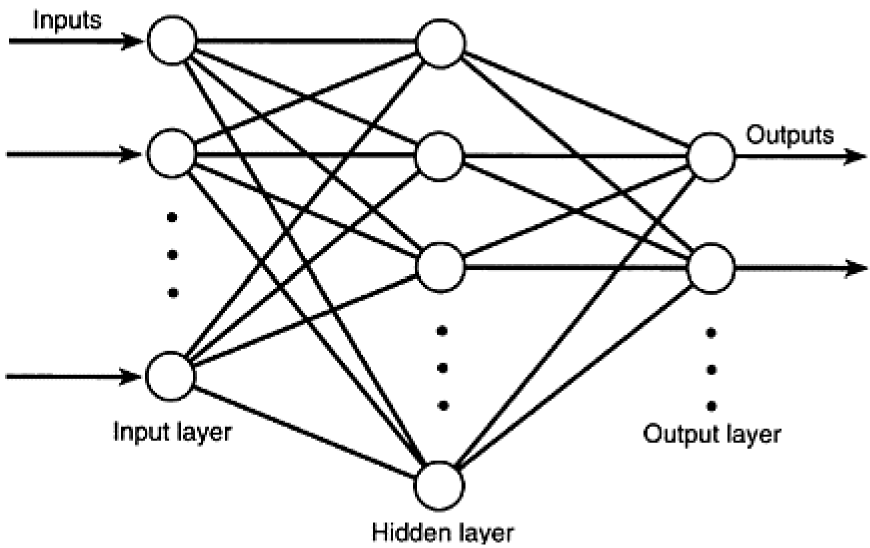


Fig. 2.2. The mesostructure of a neural network model in terms of processing units.

output connection carries identical copies of each numerical output, or signal, which is the state, or activation level, of that processing unit Y_j. The weights are termed 'connection parameters'. It is these weights that are adjusted during the learning process, to determine the overall behaviour of the neural solution, and that in combination generate the so-called 'network function'.

Figure 2.2 illustrates the standard organisation of a network architecture—which is the mesostructure. The basic structure consists of a number of processing units, arranged in a number of layers, and connected together to form a network. Data enters the network through the input units (left). It is then passed forward, through successive intermediate hidden layers, to emerge from the output units (right). The outer layer, where information is presented to the network, is called the input layer and contains the input units. These units disperse their input values to units in the next layer and serve no other function or purpose. The layer on the far side, where processed information is retrieved, is called the output layer and contains the output units. The layers in between the two outer layers are called hidden layers, being hidden from direct contact with the outside world, and contain the hidden units. Full connection is said to exist if each node in each layer is connected to all nodes in each adjacent layer. To avoid confusion the recommended method for describing a NN is based on the number of hidden layers. Figure 2.2 thus depicts a one-hidden-layer feedforward architecture with no feedback loops. However, it is also possible to have connections that transfer information backwards from output units to input units, from output units to hidden units, or from a unit to itself. These are termed partial-recurrent networks (PRNN)—see Van den Boogaard (Chapter 7) and Ball and Luk (Chapter 10). If the internal connections circulate information from each node to all other nodes then it is a recurrent network.

The use of storage tanks and chronological updating procedures is a familiar concept to the hydrologist and such items comprise an integral part of most conceptual models and distributed modelling solutions. Thus far, however, in direct contrast most published NN hydrological modelling applications have been based on static models that contain no explicit consideration of time, previous events, antecedent conditions or state-space evolution—with no attempt being made to account for the complex interaction that should in fact occur between sequential representations of different but related input-output ‘snapshots’. It is therefore argued that feedback loops could perhaps be used to address this issue, through the addition and circulation of dedicated variables that change or update specific factors in response to previous computations, and thus provide a dynamic and responsive solution that is better suited to modelling hydrological processes.

The number of processing units in the input and output layers is fixed according to the number of variables in the training data and is specific to each individual problem depending on the number of predictors and predictands. But the selection of an optimal number of hidden layers and hidden units will in all cases depend on the nature of the application. Intuition suggests that ‘more is better’—but there are limits on the extent to which this is true. In certain instances a small(er) number of hidden units is advantageous. The number of hidden units and layers is important, since a larger architecture will extend the power of the model to perform more complex modelling operations, but there is an associated trade-off between the amount of training involved and the level of generalisation achieved. The use of large hidden layers can also be counterproductive since an excessive number of free parameters encourages the overfitting of the network solution to the training data, and so reduces the generalisation capabilities of the final product (Fig. 2.3). The other question that needs to be addressed is the number of hidden layers and the relative organisation of their hidden units. Practical methods to establish an ‘optimum’ set of hidden features range from best guess (e.g. Cheng & Noguchi, 1996) or trial and error (e.g. Shamseldin, 1997) to the application of sophisticated computational solutions e.g. cascade correlation which is a constructive algorithm (Imrie *et al.,* 2000; Lekkas *et al.,* 2001); weight or node based pruning which is a destructive algorithm (Abrahart *et al.,* 1999); or evolution-based approaches using a dedicated genetic algorithm package (Abrahart *et al.,* 1999). In the first instance inexperienced users might opt for one hidden layer with the number of hidden units equal to the number of inputs. More experienced users might match the number of hidden units to an anticipated number of empirical functions.

2.2 *Learning considerations*

NN ‘learning’ is defined as ‘deliberate or directed change in the knowledge structure of a system that allows it to perform better or later repetitions of some given type or task’ (Fischler & Firschein, 1987). Specific information on a particular topic or task is thus encoded in order for the solution to produce a suitable response on subsequent occasions. The two most common types of learning are supervised and unsupervised: the difference between them is that supervised learning requires each input pattern to have an associated output pattern. In

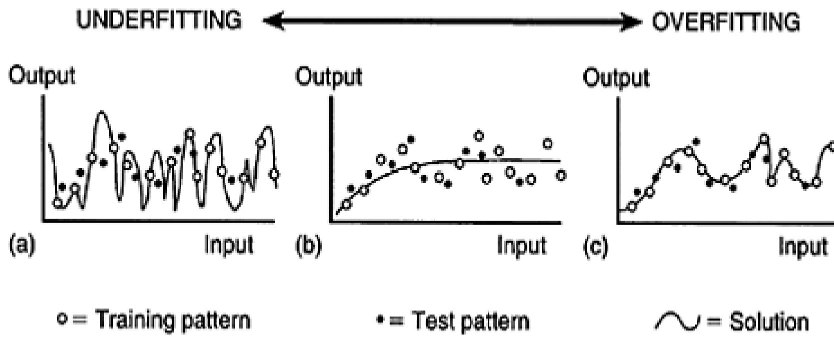


Fig. 2.3. The training trilemma
(adapted from Flood & Kartam, 1994).

supervised training the model input might be discharge data collected at one or more upstream gauges with the output being forecast discharge at a downstream station. Cameron *et al.* (2002), for example, used a combination of river stage at two upstream stations and two local variables to estimate future river stage at a downstream station. In unsupervised training the output is in most cases a set of clusters; for instance river level series can be partitioned into different categories of event (Abrahart & See, 2000); rainfall and river series records can be partitioned to establish combined clusters that span the total input space (Hsu *et al.,* 2002); catchments can be clustered into homogeneous categories that possess similar geomorphological and climatological characteristics (Hall *et al.,* 2002).

Each combination of input and output data is referred to as a training pair and the complete set of training pairs is the training set. The training period for the presentation of an entire training set is one epoch. The goal of training is to minimise the output error, which is achieved through the use of different algorithms that 'search the error surface' and 'descend the gradient'. Inputs (predictors) are passed through the network to become the outputs (predictands) and through the learning process the internal connection weights are modified in response to computed error—the equation that specifies this change is termed the 'learning law' or 'learning rule'. There are a large number of different learning methods and the learning process is often complex, with numerous options, variables, and permutations to choose from.

The learning process is continued until such time as an acceptable solution is arrived at. This is accomplished through numerous repeated iterations of data presentation and weight updates, until such time as an acceptable pre-specified stopping condition is met, and the underlying function has been 'discovered'. However, it is important to ensure that the network does not become over-familiarised with the training data, and thus lose its power to generalise to unseen data sets. Figures 2.3 and 2.4 illustrate the basic problem of underfitting (undertraining) and overfitting (overtraining). The data set used in this process may be referred to as a 'validation' set.

If a neural solution has insufficient complexities, or has been underfitted, it will fail to detect the full signal in a complicated data set. If the neural solution is too complex, or has been overfitted, it will fit the noise as well as the signal. To differentiate between these opposing situations in an effective manner

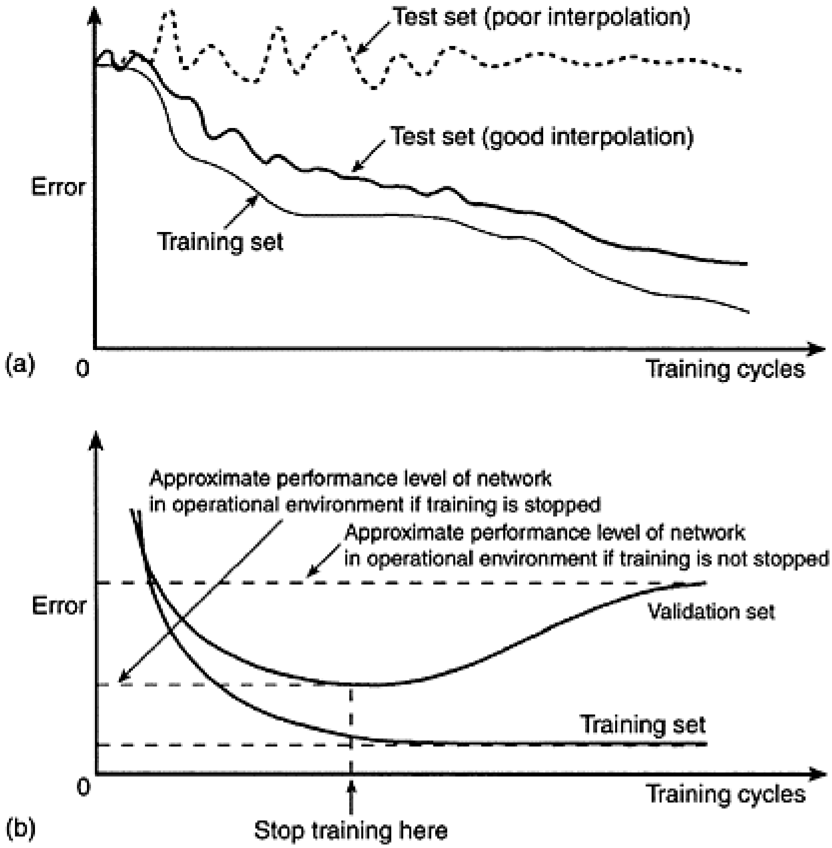


Fig. 2.4. Two possible scenarios for a plot of network error against training cycles. In each case overfitting arises when the solution learns the exact nuances of each individual case in the training data such that the final product has limited or no real interpolation capabilities (a) after Flood and Kartam

(1994) (b) after Caudhill and Butler (1992).

is problematic and continuous assessment would be required throughout the different stages of construction and development. Several techniques are avail able to prevent overfitting:

(a) **Jitter:** addition of artificial noise to the input data during training that will produce a smoother final mapping between inputs and outputs e.g. Abrahart and White (2001).
(b) **Weight Decay:** addition of an extra term to the error function that penalises large weights in order to create a smoother final mapping between inputs and outputs—but no hydrological modelling investigation of this method has been reported.
(c) **Early Stopping:** use of split-sample validation, cross-validation or bootstrapping techniques to determine that point at which a sufficient degree of learning has taken place. For a comparison between continuous cross-validation and continuous bootstrapping applied to discharge forecasting see Abrahart (2003).
(d) **Structural Control:** restrict the number of hidden units and weighted connections such that a limited number of free parameters is available during the 'fitting process'. Each hidden node in each solution will attempt to + represent a discrete input-output association; so in the case of discharge forecasting simple functions such as 'quickflow' and 'baseflow' will be assigned to specific hidden nodes, whereas more complex entities such as 'soil moisture switches', would be assigned to one or more of the unclaimed units. Wilby *et al.* (2003) illustrate the inner workings of this mechanism, in a series of river-level forecasting experiments, in which a conceptual model is cloned with a number of neural solutions.

3 MAIN CATEGORIES OF MODEL

Neural networks are often promoted as a one-stop-shop but *caveat emptor* applies; users must recognise that there are several important decisions that must be taken to select an appropriate class of model. Certain forms of solution might be better suited to modelling specific hydrological functions or processes—although this notion is still quite novel and extensive testing will be required before indicative outcomes could be converted into a set of definitive guidelines. Different types of solution can nevertheless be differentiated in terms of:

(a) node characteristics i.e. properties of the processing units;
(b) network topologies i.e. the pattern of connections; and
(c) the learning algorithm and its associated parameters.

The number of possible combinations and permutations that could be implemented is enormous and to perform a detailed and comprehensive analysis is impractical. However, for hydrological modelling purposes, the three most common tools are:

(a) BPNN—backpropagation neural network;
(b) RBFN—radial basis function network;