Chapter 7

Conclusions

7.1 Local Model nets

The work undertaken has shown that Local Model Nets have a great deal of potential as a general model structure, suitable for a wide variety of empirical modelling and learning tasks. An important aspect of the architecture is that it forms a link between the world of learning systems such as neural networks and the more conventional world of systems theory and statistical modelling. The local nature of the models means that the trained networks are more transparent, and can easily integrate *a priori* models from the other paradigms. Knowledge about the process being modelled can be used in hybrid local model nets to better cope with high-dimensional systems.

7.1.1 Local Model Net extensions

This thesis extends the understanding of local model networks in several respects:

- The Local Learning methods developed in this work have a regularisation effect which in many cases improve the generalisation of the trained net, as the methods overcome some of the conditioning problems discovered in the parameter estimation phase of training. The nature of the conditioning problems had previously gone unnoticed, and this work demonstrated the dependence of the condition on the level of overlap between models.
- Methods have been developed which use the local nature of the basis functions to allow local error estimates to be interpolated to produce a state-dependent error statistic for the whole network. Such statistics are useful for the validation and interpretation phases of modelling and are useful for on-line application in model-based control and diagnosis systems. The techniques were applied to the rolling mill problem and produced error estimates which corresponded well with the measured residuals on a validation run.

- The effect of the widely used normalisation technique for basis function nets has been analysed, and hitherto undocumented results were discovered. These have a serious effect on the smoothness of the representation and can compromise the local nature of the basis functions. Normalisation still has advantages in that it produces a partition of unity, and makes the model less sensitive to a poor choice of basis functions, but the unexpected side-effects described in Section 3.3.1 should be taken into account when interpreting trained basis function nets, and when developing new learning algorithms.

The thesis also provides descriptions of a number of practical techniques for use in Local Model Networks:

- New Local Learning methods were developed for the parameter estimation phase in Local model nets. These methods are far less computationally expensive than the global ones, and can often result in more interpretable local models. In general, with a reasonable level of basis function overlap, the local optimisation tends to cope better than global learning in high-dimensional, noisy or sparsely populated learning problems. Section 4.4 gives examples of how the local learning methods performed more robustly on noisy test functions, and Chapter 3 demonstrates the better generalisation ability of locally trained local models on the rolling mill application.

  Local learning is also well-suited to heterogeneous local model nets, where a variety of types of local model can be used, each with its own optimisation algorithm. This makes the local model framework much more general than the basic homogeneous linear local model structure, allowing it to integrate a variety of styles of model and methods of knowledge representation.

- The new Multi-Resolution Constructive (MRC) structure identification algorithm for Local Model networks has been developed. This allows the network to fit the basis functions to the data set in a gradual, problem-adaptive manner. The constructive nature of the algorithm speeds up the modelling process by reducing the amount of 'fiddling' needed to produce a good model. It also tends to find more accurate models, as the complexity is increased as needed for the given problem, while overfitting is limited by taking the local density of the training data into account. The fact that trained models are now being trained in minutes as opposed to days with previous learning techniques obviously better supports the interactive nature of the modelling process.

- Active data selection techniques have been developed for local model nets. These speed up the learning process by automatically selecting a training set for the current model structure, where the number of data points needed is reduced, but the most informative data is chosen from the complex areas of the input space. Training tends also to be more robust with respect to the distribution of the training data, as demonstrated in Section 4.3 with synthetic examples and in Chapter 6 on the real applications.

7.1.2 Learning Hierarchies of Models

The novel Learning Hierarchy of Models architecture is a hierarchical extension of the local model networks, which due to the split-like nature of its partitioning mechanism decomposes high dimensional spaces more efficiently than flat Local Model nets.

A new constructive algorithm which automatically grows tree-like structures to fit the target function is described. The soft-splits can be axis-oblique, making the structure more powerful than other more restricted hierarchical structures which are limited to partitioning the input space one variable at a time. Local learning, error estimation and active learning can all be applied in a hierarchical manner.

Use of soft, axis-oblique splits provides the potential for hierarchical structure adaptation, where gradual structure learning occurs simultaneously at several levels of the structure at differing timescales. This is likely to lead to more efficient learning algorithms, which produce more parsimonious models. It is also interesting for on-line use for structure adaptation in time-varying systems. The architecture can also be seen as a fuzzy decision tree, and could apply methods from both decision tree theory and fuzzy systems.

7.1.3 Experimental work

To demonstrate the applicability of the methods to real industrial processes data was taken from an aluminium rolling mill and a robot actuator:

- Data has been taken from an aluminium rolling mill to train a predictive model of the strip thickness using local model nets. This was a real application of the modelling techniques to aspects of a nonlinear, dynamic process which were poorly understood. The Local Model Net techniques described in this thesis have produced the most accurate modelling results known to date on this problem, and further testing with an on-line implementation is planned.

- The Trallfa robot actuator nonlinearities were also learned successfully. This problem has been used as a test example for a number of other structure identification algorithms, and it is characterised by its low dimensionality, smooth nonlinearities and large training set. Despite its relatively simple nature, the problem demonstrated that the methods could still be competitive on more straightforward, low-dimensional problems.

7.2 Outlook

The training algorithm used to identify the underlying structure of the basis functions is a practical but still relatively *ad hoc* routine, and there is certainly a great deal of progress to be made in the area of structure construction algorithms. User friendly methods for the easy integration of *a priori* knowledge into the network will become more and more important as the demands for accuracy, robustness and transparency increase. There are many opportunities for the development of easy-to-use tools which allow the developer to creatively build engineering knowledge directly into the modelling process, leaving the learning algorithms to cope with the uncertainty in the process, and to warn the user where more information is needed. Tools based on the ideas described in this thesis are likely to benefit the developer in the computer intensive, data driven areas of modelling. Despite the improvements in tools, it is perhaps relevant to quote one of the leading system identification researchers, Lennart Ljung:

> 'Thinking, intuition and insight cannot be made obsolete by automated model construction' (Ljung, 1987)

An immediately practical view of the methods described in this thesis is to see them as computationally- and data-intensive ways of supporting more traditional modelling methods, allowing the engineer to better understand the system by reproducing the behaviour with a learned model, understanding the behaviour and then creating a 'hand built' simplified model which exhibits the essential behaviour of the structurally more complex learned model. This is then more easily understood and validated, and is therefore more likely to be used in practical applications.

Appendix A

Notation

A.1 Notation used

$\mathbf{A}^T$	the transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	the inverse of matrix $\mathbf{A}$
$\mathbf{A}^+$	the pseudoinverse of matrix $\mathbf{A}$
B_j, p	Basis Spline j of order p
$Cov(\mathbf{x})$	the covariance matrix of vector $\mathbf{x}$
t	knot of a basis spline
u	system inputs
$\boldsymbol{\psi}$	information vector inputs
$\mathbf{c}_i$	position of basis function i's centre in the input space
$\tilde{\boldsymbol{\phi}}$	operating point state
y	outputs
$\hat{y}$	estimate of outputs
$\mathcal{D}$	the training set of input-output pairs
$\mathcal{D}_i$	the training set of input-output pairs in receptive field of basis function i
ϕ	basis vector for regression
$\mathbf{\Phi}$	the design matrix for the regression problem
$\mathbf{\Phi}_i$	the design matrix for the local regression problem for local model i
J	cost function
J_V	cost due to model variance
J_B	cost due to model bias
J^*	optimal cost for parameter optimisation
J^{**}	optimal cost for structure optimisation
J_i	cost functional for optimisation of parameters of local model i.
$\mathcal{M}$	the whole model structure

$\mathcal{M}(\boldsymbol{\theta})$	the model structure parameterised by $\boldsymbol{\theta}$
$\mathcal{M}_i$	the model structure of local model i
n_ψ	number of input dimensions to local models
$n_{\mathcal{M}}$	number of basis functions
n_ϕ	dimension of basis vector for regression
n_{cutoff}	number of new units added for ‘stopping window
n_{res}	number of resolution stages in multi-resolution clustering
$n_{\tilde{\phi}}$	dimension of basis function space
n_u	dynamic order of local models in rolling mill model
N	number of training patterns
N_i	number of training patterns for local model i
N_{des}	number of data points desired for a given local model for active selection
N_{min}	minimum number of training patterns in the receptive field of a local model
N_{rand}	number of data points randomly chosen from entire training set during active selection
P_l	the number of parameters in local model l
$\rho(\cdot)$	the basis activation function
$\rho_i(\cdot)$	the basis activation function for basis function i (can be after normalisation)
ρ_{total}	the cumulative activation of a basis function over the whole training set
$\mathcal{R}^n$	Euclidean n-dimensional space
$\mathcal{S}$	the process being modelled
σ	scaling matrix (or factor) for distance function of the basis function
σ_{win}	current window size for the complexity window of Chapter 4
σ_{max}	coarsest (initial) window size for the complexity window of Chapter 4
$\boldsymbol{\theta}$	weights
$\alpha(\cdot)$	error weighting function for weighted least squares problems
$\mathbf{a}$	weighting matrix for distance metric
γ	scaling factor for minimum distance between units
$\mathbf{D}(\boldsymbol{\psi})$	scaling factor for the distance from centre for calculation of σ
λ	reduction factor for window size in iterative cluster algorithm
Q_{min}	minimum angle between two neighbouring centres for covariance calculation
$*\lambda$	weighting of cost function in cost-complexity optimisation
$\mathcal{X}$	the experiment used to gather the training data $\mathcal{D}$
$\mathbf{Y}$	the vector of outputs in training set $\mathcal{D}$
ζ	threshold for basis function activation
$\lvert \cdot \rvert$	Euclidean norm
$\lVert \cdot \rVert$	Matrix norm

A.2 Abbreviations

ANN	Artificial Neural Network
ANOVA	ANalysis Of VAriance
ARMAX	Auto-Regressive Moving Average model with eXternal variable
ARX	Auto-Regressive model with eXternal variable
ART	Adaptive Resonance Theory
ASMOD	Adaptive Splines MODelling
BF	Basis Function
BP	Back-Propagation
CART	Classification And Regression Trees
CMAC	Cerebellar Model Articulation Controller
EBF	Ellipsoidal Basis Function
FIR	Finite Impulse Response
GCV	Generalised Cross-Validation
GSS	Generalised Smoothing Splines
HME	Hierarchical Mixtures of Experts
HSOL	Hierarchical Self-Organising Learning
ID3	Iterative Dichotomiser 3
k-N-N	k-Nearest-Neighbour
LHM	Learning Hierarchy of Models
LMN	Local Model Network
LMS	Least Mean Squared
LSA	Local Search Algorithm
LTU	Linear Threshold Unit
LVQ	Learning Vector Quantisation
MARS	Multiple Adaptive Regression Splines
MLP	Multi-Layer Perceptron
NARMAX	Non-linear Auto-Regressive Moving Average model with eXternal variable
PCA	Principle Component Analysis
PDP	Parallel Distributed Processing
PDF	Probability Distribution Function
PID	Proportional Integral Derivative
PLS	Partial Least Squares
PNN	Probabilistic Neural Network
PVM	Predictive Value Maximisation
RBF	Radial Basis Function
RCE	Restricted Coulomb Energy
RK	Reproducing Kernels
SOM	Self-Organising Map

SVD	Singular Value Decomposition
TDNN	Time-delay neural network
VI-Net	Validity Index Net

References

Åström, K. J. (1987). Adaptive feedback control. Proc. IEEE **75**, 185–217.

Aizermann, M., Braverman, E., and Ronzonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. Automatika i Telemekhanika **25**, 147–169.

Albus, J. S. (1972). *Theoretical and experimental aspects of a cerebellar model.* PhD thesis, University of Maryland.

Albus, J. S. (1975a). Data storage in the cerebellar model articulation controller (CMAC). Trans. ASME. Jnl. Dyn. Sys. Meas. and Control **63**, 228–233.

Albus, J. S. (1975b). A new approach to manipulator control: The cerebellar model articulation controller (CMAC). Trans. ASME. Jnl. Dyn. Sys. Meas. and Control **63**, 220–227.

Alpaydin, E. (1991). GAL: Networks that grow when they learn and shrink when they forget. Technical Report 91-032, Int. Comp. Sci. Inst., Berkeley.

Anderson, J. A. and Rosenfeld, E., editors (1988). *Neurocomputing: Foundations of Research.* MIT Press, Cambridge.

Anderson, J. A. and Rosenfeld, E., editors (1990). *Neurocomputing 2: Directions for Research.* MIT Press, Cambridge.

Andrews, H. C. (1983). *Introduction to Mathematical Techniques in Pattern Recognition.* Robert E. Krieger.

Atkeson, C. G. (1990). Memory-based approaches to approximating continuous functions. In *Workshop on Nonlinear Modelling and Forecasting, Santa Fe Institute.* Addison-Wesley.

Back, A. D. and Tsoi, A. C. (1991). FIR and IIR synapses, a new neural network architecture for time series modeling. Neural Computation **3**, 375–385.

Bakshi, B. R. and Stephanopoulos, G. (1993). Wave-Net: a Multiresolution, Hierarchical Neural Neural Network with Localized Learning. AIChE Journal **39**.

Banan, M. R. and Hjelmstad, K. D. (1992). Self-organization of architecture by simulated hierarchical adaptive random partitioning. In *IJCNN, Baltimore*, volume III, pages 823–828.

Barnes, C. et al. (1991). Applications of neural networks to process control and modelling. In *Artificial Neural Networks, Proceedings of 1991 Internat. Conf. Artif. Neur. Nets*, volume 1, pages p321–326.

Barron, A. R. and Barron, R. L. (1988). Statistical learning networks: a unifying view. In *Computer Science and Statistics: Proc. of 21st Interface*, pages 192–203.

Bellman, R. E. (1961). *Adaptive Control Processes.* Princeton University Press, Princeton, NJ.

Benaim, M. (1994). On Functional Approximation with Normalised Gaussian Units. Neural Computation **6**, 319–333.

Bently, J. L. (1975). Multidimensional binary search trees used for associative searching. Communications of the ACM **18**, 509–517.

Billings, S. A. (1980). Identification of nonlinear systems–a survey. IEE Proc.,Pt D **127**, 272–285.

Billings, S. A. and Chen, S. (1989). Extended model set, global data and threshold model identification of severly non-linear systems. Int. J. Control **50**, 1897–1923.

Billings, S. A. and Voon, W. S. G. (1987). Piecewise linear identification of non-linear systems. Int. J. Control **46**, 215–235.

Bishop, C. (1991). Improving the generalization properties of Radial Basis Function neural networks. Neural Computation **3**, 579–588.

Bishop, C. M. (1994). Training with noise is equivalent to Tikhonov Regularization. Submitted for publication.

Bottou, L. and Vapnik, V. (1992). Local learning algorithms. Neural Computation **4**, 888–900.

Breiman, L. et al. (1984). *Classification and Regression Trees.* Wadsworths & Brooks, Monterey, Ca.

Bridgett, N. et al. (1994). Associative memory network construction algorithms.

Broomhead, D. S. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. Complex Systems **2**, 321–355.

Brown, M. and Harris, C. (1994). *Neurofuzzy Adaptive Modelling and Control.* Prentice Hall, Hemel-Hempstead, UK.

Carlin, M. (1992). Radial Basis Function networks and nonlinear data modeling. In *Neuro–Nimes 92*, pages 623–633.

Chen, S. and Billings, S. A. (1989). Representation of non-linear systems: The NARMAX model. Int. J. Control **49**, 1013–1032.

Chen, S. and Billings, S. A. (1992). Neural networks for non-linear dynamic system modelling and identification. Int. J. Control **56**, 319–346.

Chen, S. and Billings, S. A. (1994). Neural networks for nonlinear dynamic system modelling and identification. In J., H. C., editor, *Advances in Intelligent Control*, pages 85–112, London. Taylor and Francis.

Chen, S., Billings, S. A., and Grant, P. M. (1990). Non-linear system identification using neural networks. Int. J. Control **51**, 1191–1214.

Chen, S., Cowan, C. F. N., and Grant, P. M. (1991). Orthogonal Least Squares learning algorithm for Radial Basis Function networks. Trans. IEEE on Neural Networks **2**.

Cleveland, W. S., Devlin, S. J., and Grosse, E. (1988). Regression by local fitting. Journal of Econometrics **37**, 87–114.

Cohn, D. (1994). Neural network exploration using optimal experiment design. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*, San Franciso, CA. Morgan Kaufmann Publishers.

Cohn, D., Atlas, L., and Ladner, R. (1990). Training connectionist networks with queries and selective sampling. In Touretzky, D. S., editor, *Advances in Neural Information Processing Systems 2*. Morgan Kaufmann Publishers, San Mateo, CA.

Cohn, D., Ghahramani, Z., and Jordan, M. I. (1994). Active learning with statistical models. Submitted to NIPS'94.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. Numerical Math. **31**, 317–403.

Dasarathy, B. (1990). *Nearest Neighbour Pattern Classification Techniques.* IEEE Computer Society Press.

Deprettre, E. F., editor (1988). *SVD and Signal Processing: Algorithms, Analysis and Applications.* North Holland.

Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis.* John Wiley & Sons.

Fahlmann, S. E. and Lebiere, C. (1990). The cascade-correlation learning architecture. In Touretzky, D. S., editor, *Advances in Neural Information Processing Systems 2*, pages 524–532. Morgan Kaufmann Publishers, San Mateo, CA.

Fedorov, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics , 179–188.

Foss, B. A. and Johansen, T. A. (1993). On local and fuzzy modelling. In *3rd Int. Conf. on Industrial Fuzzy Control and Intelligent Systems*, Houston, Texas.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. Annals of Statistics **19**, 1–141.

Fritzke, B. (1994). Supervised learning with growing cell structures. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*, pages 255–262. Morgan Kaufmann Publishers, San Franciso, CA.

Fu, K. S. (1970). Learning control systems – review and outlook. Trans. IEEE on Automatic Control **16**, 210–221.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the Bias/Variance Dilemma. Neural Computation **4**, 1–58.

Girosi, F., Jones, M., and Poggio, T. (1993). Priors, Stabilizers and Basis Functions: from regularization to radial, tensor and additive splines. MIT AI Memo 1430, MIT.

Golub, G. H. and van Loan, C. F. (1989). *Matrix Computations.* Johns Hopkins University Press.

Goodwin, G. C. and Payne, R. L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press, New York.

Gu, C. et al. (1990). The computation of GCV function through Householder tridiagonalization and application to the fitting of interaction spline models. SIAM J. Matrix Analysis , 457–480.

Haas, R. and Murray-Smith, R. (1993). Fuzzy/Neuro-Kombinationen – Konzepte und Perspektiven. Technischer Bericht, Daimler-Benz Research, Berlin. E-mail:haas@Dbresearch-berlin.de, murray@DBresearch-berlin.de.

Harris, C., Moore, C. G., and Brown, M. (1993). *Intelligent Control: Aspects of Fuzzy Logic and Neural Nets.* World Scientific.

Hartman, E. and Keeler, J. D. (1991). Predicting the future: Advantages of semilocal units. Neural Computation **3**, 566–578.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Monographs on Statistics and Applied Probability 43. Chapman and Hall, London.

Haykin, S. (1991). *Adaptive Filter Theory, 2nd ed.* Prentice-Hall.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation.* Macmillan.

Hebb, D. O. (1949). *The Organization of Behavior.* Wiley, New York. Partially reprinted in (Anderson and Rosenfeld, 1988).

Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation.* Addison-Wesley, Redwood City.

Hlavácková, K. and Neruda, R. (1993). Radial Basis Function networks. Neural Network World **1**, 93–101.

Holden, S. B. (1994). *On the Theory of Generalization and Self-Structuring in Linearly Weighted Connectionist Networks.* PhD thesis, Cambridge University.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, USA **79**, 2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Sciences, USA **81**, 3088–3092.

Hrycej, T. (1992). *Modular Learning in Neural Networks: A Modularized Approach to Neural Network Classification*. John Wiley and Sons.

Hunt, K. J. et al. (1992). Neural networks for control systems: a survey. Automatica **28**, 1083–1112.

Hutchinson, J. M. (1994). *Radial Basis Function Approach to Financial Time Series Analysis*. PhD thesis, Massachusetts Institute of Technology, Dept. EECS.

Isaksson, A. J., Ljung, L., and Strömberg, J.-E. (1991). On recursive construction of trees as models of dynamical systems. In *30th Conf. on Decision & Control, Brighton*, pages 1686–1687.

Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. IEEE Transactions on Systems, Man, and Cybernetics **1**, 364–378.

Jacobs, R. A. et al. (1991). Adaptive mixtures of local experts. Neural Computation **3**, 79–87.

Jang, J. S. R. and Sun, C. T. (1993). Functional Equivalence between Radial Basis Function networks and Fuzzy Inference Systems. Trans. IEEE on Neural Networks **4**, 156–158.

Johansen, T. A. and Foss, B. A. (1992a). A NARMAX model representation for adaptive control based on local models. Modelling, Identification and Control **13**, 25–39.

Johansen, T. A. and Foss, B. A. (1992b). Nonlinear local model representation for adaptive systems. In *Proceeding of the Singapore Int. Conf. on Intelligent Control and Instrumentation*, volume 2, pages 677–682.

Johansen, T. A. and Foss, B. A. (1992c). Representing and learning unmodeled dynamics with neural network memories. In *Proceedings of the American Control Conference, Chicago, Il.*, pages 3037–3043.

Johansen, T. A. and Foss, B. A. (1993). State-space modeling using operating regime decomposition and local models. In *Preprints 12th IFAC World Congress, Sydney, Australia, 19-23 July. Extended paper in Technical Report 93-40-W, Department of Engineering Cybernetics, Norwegain Institute of Technology, Trondheim.*

Johansen, T. A. and Foss, B. A. (1994a). A dynamic modeling framework based on local models and interpolation – combining empirical and mechanistiv knowledge and data. Submitted to Computers and Chemical Engineering.

Johansen, T. A. and Foss, B. A. (1994b). Identification of non-linear system structure and parameters using regime decomposition. To be presented at the IFAC Symposium on System Identification, Copenhagen.

Jones, R. D. et al. (1989). Function approximation and time series prediction with neural networks. Technical Report 90-21, Los Alamos National Lab., New Mexico.

Jordan, M. I. and Jacobs, R. A. (1991). Hierarchies of adaptive experts. In Moody, J. E., Hanson, S. J., and Lippmann, R. P., editors, *Advances in Neural Information Processing Systems 4*. Morgan Kaufmann Publishers, San Mateo, CA.

Jordan, M. I. and Jacobs, R. A. (1993). Hierarchical Mixtures of Experts and the EM algorithm. Technical Report 9301, MIT, Computational Cognitive Science.

Kavli, T. (1992). *Learning Principles in Dynamic Control*. PhD thesis, University of Oslo.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by Simulated Annealing. Science **220**, 671–680.

Kodratoff, Y. and Michalski, R. S., editors (1990). *Machine Learning: an artificial intelligence approach (Vol. 3)*. Morgan Kaufmann, Los Altos.

Kohonen, T. (1990). The self-organizing map. Proceeding of the IEEE **78**, 1464–1480.

Kolodner, J. (1993). *Case Based Reasoning.* Morgan Kaufmann.

Kramer, M. A. (1993). Diagnosing dynamic faults using modular neural nets. IEEE Expert .

Kramer, M. A., Thompson, M. L., and Phagat, P. M. (1992). Embedding theoretical models in neural networks. In *Proc. American Control Conference, Chicago, Il.*, pages 475–479.

Kuipers, B. and Åström, K. (1994). The composition and validation of heterogeneous control laws. Automatica **30**, 233–249.

Kurcova, V. (1992). Universal approximation using feedforward neural networks with Gaussian Bar units. In *Proc. ECAI'92, Vienna*, pages 193–197.

Lane, S. H., Handelman, D. A., and Gelfand, J. J. (1991). Higher order CMAC neural networks - theory and practice. In *Proc. American Control Conference, Boston, USA*, pages 1579–1585.

Lee, S. and Kil, M. R. (1991). A gaussian potential function network with hierarchically self-organizing learning. Neural Networks **4**, 207–224.

Leonard, J. A., Kramer, M. A., and Ungar, L. H. (1992). A neural network architecture that computes its own reliability. MIT Industrial Liason Report 3-7-92, MIT, Dept. of Chemical Engineering.

Leontaritis, I. J. and Billings, S. A. (1985). Input-output parametric models for non-linear systems. Int. J. Control **41**, 303–344.

Ljung, L. (1987). *System Identification — Theory for the User.* Prentice-Hall, Englewood cliffs, New Jersey, USA.

Lowe, D. (1994). Non local Radial Basis Functions for forecasting and density estimation. In *IEEE Intern. Conf. on Neural Networks*, volume II, pages 1197–1198, Florida.

Mason, J. C. and Parks, P. C. (1992). Selection of neural network architectures: Some approximation theory guidelines. In *In: K. Warwick, G. W. Irwin, K. J. Hunt (Eds), Neural networks for control and systems*, pages 151–180. Peter Peregrinus.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics **5**, 115–133.

Mesarovic, M. D., Macko, D., and Takahara, Y. (1970). *Theory of Hierarchical, Multilevel Systems.* Academic Press.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M., editors (1983). *Machine Learning: an artificial intelligence approach.* Morgan Kaufmann, Los Altos.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M., editors (1986). *Machine Learning: an artificial intelligence approach (Vol. 2).* Morgan Kaufmann, Los Altos.

Miller, W. T., Sutton, R. S., and Werbos, P. J., editors (1990). *Neural Networks for Control.* MIT Press, Cambridge, MA.

Minsky, M. L. and Papert, S. A. (1969). *Perceptrons.* The MIT press, Cambridge, Mass.

Moody, J. and Darken, C. (1989). Fast-learning in networks of locally-tuned processing units. Neural Computation **1**, 281–294.

Murray-Smith, R. (1992). A Fractal Radial Basis Function network for modelling. In *Inter. Conf. on Automation, Robotics and Computer Vision, Singapore*, volume 1, pages NW–2.6.1–NW–2.6.5. E-mail:murray@DBresearch-berlin.de.

Murray-Smith, R. (1994). Local Model Networks and Local Learning. In *Fuzzy Duisburg, '94*, pages p404–409. E-mail:murray@DBresearch-berlin.de.

Murray-Smith, R. and Gollee, H. (1994). A constructive learning algorithm for local model networks.

In *Proc. IEEE Workshop on Computer-intensive methods in control and signal processing, Prague, Czech Republic*, pages 21–29. E-mail:murray@DBresearch-berlin.de.

Murray-Smith, R., Neumerkel, D., and Sbarbaro-Hofer, D. (1992). Neural Networks for Modelling and Control of a Non-linear Dynamic System. In *IEEE Symposium on Intelligent Control, Glasgow*, pages p404–409. E-mail:murray,neumerk@DBresearch-berlin.de.

Murray-Smith, R. and Thakar, S. (1993). Combining Case Based Reasoning with neural networks. In *AAAI'93 Workshop on AI in Service and Support, Washington DC*. E-mail:murray,thakar@DBresearch-berlin.de.

Neumerkel, D., Murray-Smith, R., and Gollee, H. (1993). Modelling dynamic processes with clustered time-delay neurons. In *Proc. International Joint Conference on Neural Networks, Nagoya, Japan*. E-mail:neumerk,murray,gollee@DBresearch-berlin.de.

Noble, B. and Daniel, J. W. (1988). *Applied linear algebra*. Prentice–Hall Int., 3rd edition.

Omohundro, S. M. (1987). Efficient algorithms with neural network behavior. J. Complex Systems **1**, 273–347.

Omohundro, S. M. (1991). Bumptrees for efficient function, constraint and classification learning. In Lippmann, R. P., Moody, J. E., and Touretzky, D. S., editors, *Advances in Neural Information Processing Systems 3*, pages 693–699. Morgan Kaufmann Publishers, San Franciso, CA.

Pantaleón-Prieto, C. J., de María, F. D., and Figueiras-Vidal, A. (1993). On training RBF networks. In *Neural Networks and their Industrial & Cognitive Applications, Nimes, France*, pages 279–288.

Pao, Y.-H. (1992). Functional link net computing. Computer , 76–79.

Park, J. and Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. Neural Computation **3**, 246–257.

Park, J. and Sandberg, I. W. (1993). Approximation and Radial-Basis-Function networks. Neural Computation **5**, 305–316.

Parker, D. (1985). Learning-logic. Tech. Report. TR-47, Center for Computational Research in Economics and Management Science, MIT, MA.

Platt, J. (1991). A Resource-Allocating Network for function interpolation. Neural Computation **3**, 213–225.

Plutowski, M. (1994). *Selecting Training Exemplars for Neural Network Learning*. Ph.D. Thesis, University of California, San Diego, USA.

Poggio, T. and Girosi, F. (1990). Networks for approximation and learning. In *Proceedings of the IEEE*, volume 78, pages 1481–1497.

Pottmann, M., Unbehauen, H., and Seborg, D. E. (1993). Application of a general multi-model approach for identification of highly nonlinear processes - a case study. Int. J. Control **57**, 97–120.

Powell, M. J. D. (1987). Radial Basis Functions for multivariable interpolation: A review. In *Algorithms for Approximation*, pages 143–167, Oxford. Clarendon Press.

Press, W. H. et al. (1988). *Numerical Recipes (C): The Art of Scientific Computing*. Cambridge Press, UK.

Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis*. Academic Press.

Quinlan, J. (1992). Learning with continuous classes. In *Australian AI Conf*, pages 343–348.

Quinlan, J. (1993). *C4.5 Programs for Machine Learning*. Morgan Kaufmann.

Raipala, J. and Koivo, H. N. (1992). Self-generating radial base network in fault diagnosis. In *Inter.*

Conf. on Automation, Robotics and Computer Vision, Singapore, volume 1, pages NW–3.2.1–NW–3.2.5.

Rich, E. (1988). *Artificial Intelligence*. McGraw-Hill.

Roberts, S. and Tarassenko, L. (1994). A probabilistic resource allocating network for novelty detection. Neural Computation **6**, 270–284.

Röscheisen, M., Hofmann, R., and Tresp, V. (1992). Neural control for rolling mills: Incorporating domain theories to overcome data deficiency. In Moody, J. E., Hanson, S. J., and Lippmann, R. P., editors, *Advances in Neural Information Processing Systems 4*, pages 659–666, San Mateo, CA. Morgan Kaufmann.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review **65**, 386–408.

Rosenblatt, F. (1962). *Principles of Neurodynamics*. Spartan, New York.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing*. MIT Press, Cambridge, Mass.

Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition,* Vol. 1: *Foundations*. MIT Press, Cambridge, Mass.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of Research and Development **3**, 210–229.

Samuel, A. L. (1967). Some studies in machine learning using the game of checkers. part II–recent progress. IBM Journal of Research and Development **11**, 601–617.

Sanger, T. D. (1991a). A tree-structured adaptive network for function approximation in high-dimensional spaces. IEEE Trans. on Neural Networks **2**, 285–293.

Sanger, T. D. (1991b). A tree-structured algorithm for reducing computation in networks with separable basis functions. Neural Computation **3**, 67–78.

Sanner, R. M. and Slotine, J.-J. (1992). Gaussian networks for direct adaptive control. IEEE Trans. Neural Networks **3**, 837–863.

Sbarbaro, D. (1992a). *Connectionist Feedforward Networks for Control of Nonlinear Systems*. Ph.D. Thesis, Department of Mechanical Engineering, Glasgow University, Glasgow, Scotland.

Sbarbaro, D. G. (1992b). A comparative study of different learning algorithms for gaussian networks. In *IFAC symposium on Intelligent Components and Intruments for Control Applications. Malaga, Spain*, pages 301–305.

Shamma, J. S. and Athans, M. (1991). Gain scheduling: Potential hazards and possible remedies. In *Proceedings American Control Conference, Boston, Ma.*, pages 516–251.

Shavlik, J. W. and Ditterich, T. G. (1990). *Readings in Machine Learning*. Morgan Kaufmann, San Mateo.

Shorten, R. and Murray-Smith, R. (1994). On Normalising Basis Function networks. In *4th Irish Neural Networks Conf., Univ. College Dublin.*

Sjöberg, J., Hjalmarsson, H., and Ljung, L. (1994). Neural networks in system identification. Technical Report LiTH-ISY-I-1622, Dept. of Electrical Engineering, Linköping University, Sweden. Ftp address:130.236.24.1 cd pub/Reports/1994.

Sjöberg, J. and Ljung, L. (1992). Overtraining, regularization, and searching for minimum in neural networks. In *Proc. IFAC Symposium on Adaptivve Systems in Control and Signal Processing,* *Grenoble, France.*, pages 669–674.

Sjöberg, J., T.McKevey, and Ljung, L. (1993). On the use of regularization in system identification. In *Proc. 12th World Congress IFAC, Sydney Australia.*, volume 7, pages 381–386.

Skeppstedt, A., Ljung, L., and Millnert, M. (1992). Construction of composite models from observed data. Int. J. Control **55**, 141–152.

Sklansky, J. (1966). Learning systems for automatic control. IEEE Trans. on Automatic Control **11**, 6–19.

Söderström, T. and Stoica, P. (1989). *System Identification.* Prentice Hall, Englewood Cliffs, NJ.

Specht, D. F. (1991). A general regression neural network. Trans. IEEE on Neural Networks **2**.

Steinbuch, K. (1961). Die lernmatrix. Kybernetik **1**, 36–45.

Steinbuch, K. (1963). *Automat und Mensch.* Springer Verlag.

Stokbro, K., Umberger, D. K., and Hertz, J. A. (1990). Exploiting neurons with localized receptive fields to learn chaos. Complex Systems **4**, 603–622.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. J. Royal Statistical Soc. B **36**, 111–133.

Strömberg, J.-E., Gustafsson, F., and Ljung, L. (1991). Trees as black-box model structures for dynamical systems. In *European Control Conference, Grenoble*, pages 1175–1180.

Sugeno, M. and Kang, G. T. (1988). Structure identification of fuzzy model. Fuzzy Sets and Systems **26**, 15–33.

Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its applications for modeling and control. IEEE Trans. on Systems, Man and Cybernetics **15**, 116–132.

Thrun, S. B. (1992). The role of Exploration in Learning Control. In *Handbook of Intelligent Control: Neural Fuzzy and Adaptive Approaches.* Van Nostrand Reinhold.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of Ill-posed problems.* Winston, Washington DC.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press. Oxford Statistical Science Series 6.

Tsypkin, Y. Z. (1971). *Adaptation and Learning in Automatic Systems.* Academic Press, New York.

Tsypkin, Y. Z. (1973). *Foundations of the Theory of Learning Systems.* Academic Press, New York.

Vaccaro, R. J., editor (1991). *SVD and Signal Processing II: Algorithms, Analysis and Applications.* Elsevier.

van der Veen, A.-J., Deprettre, E. F., and Swindlehurst, A. L. (1993). Subspace-based signal analysis using singular value decomposition. IEEE Proceedings **81**, 1277–1308.

Wahba, G. (1990). Spline models for observation data. In *Regional Conference Series in Applied Mathematics*, Philadelphia, PA. SIAM.

Wahba, G. (1992). Multivariate function and operator estimation, based on smoothing splines and reproducing kernels. In Casdagli, M. and Eubank, S., editors, *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity*, volume XII. Addison-Wesley.

Wang, L.-X. (1994). *Adaptive Fuzzy Systems and Control: Design and Stability Analysis.* Prentice Hall.

Warwick, K., Irwin, G. W., and (Eds), K. J. H., editors (1992). *Neural networks for control and systems.* Peter Peregrinus.

Wassermann, P. D. (1993). *Advanced methods in neural computing.* Van Nostrand Reinhold, New

York.

Weiss, S. M. and Kulikowski, C. A. (1991). *Computer Systems that Learn.* Morgan Kaufmann, San Mateo, California.

Werbos, P. J. (1974). *Beyond regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* PhD thesis, Masters Thesis, Harvard University.

Werntges, H. W. (1993). Partitions of unity improve neural function approximation. In *Proc. IEEE Int. Conf. Neural Networks*, pages 914–918, San Francisco, CA. Vol. 2.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. In *1960 IRE WESCON Convention Record*, volume 4, pages 96–104. IRE, New York.

Wiener, N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine.* MIT Press.

Xu, L., Jordan, M. I., and Hinton, G. (1994). An alternative model for mixtures of experts. Submitted to NIPS'94.

Żbikowski, R. (1994). *Recurrent Neural Networks: Some Control Problems.* Ph.D. Thesis, Department of Mechanical Engineering, Glasgow University, Glasgow, Scotland.

Zhao, Y. (1992). *On Projection Pursuit Learning.* PhD thesis, MIT, Dept. of Mathematics and the AI Lab.

Index