System Identification & Forecasting

with

Advanced Neural Networks

Principles, Techniques, Applications

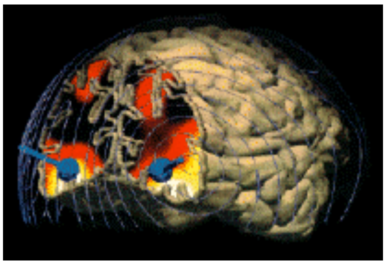
Hans Georg Zimmermann

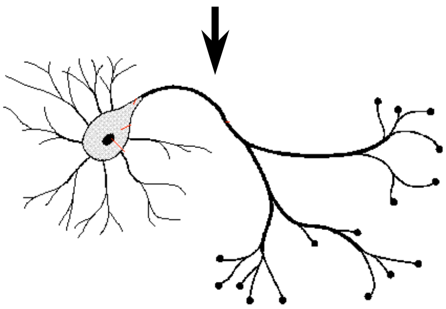
Siemens AG

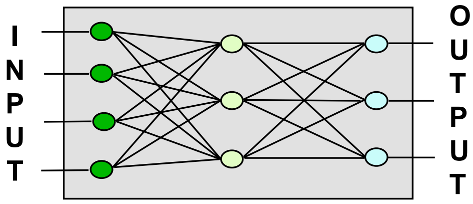
Email : Hans_Georg.Zimmermann@siemens.com

Neural Networks - from Biology to Mathematics

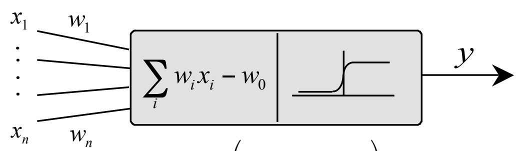
From the modeling of biology to the learning of high dimensional, nonlinear systems







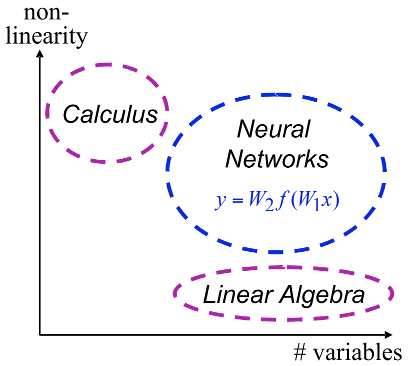
$$y = f\left(W_2 f(W_1 x - w_1) - w_2\right)$$



$$y = f\left(\sum_i w_i x_i - w_0\right)$$

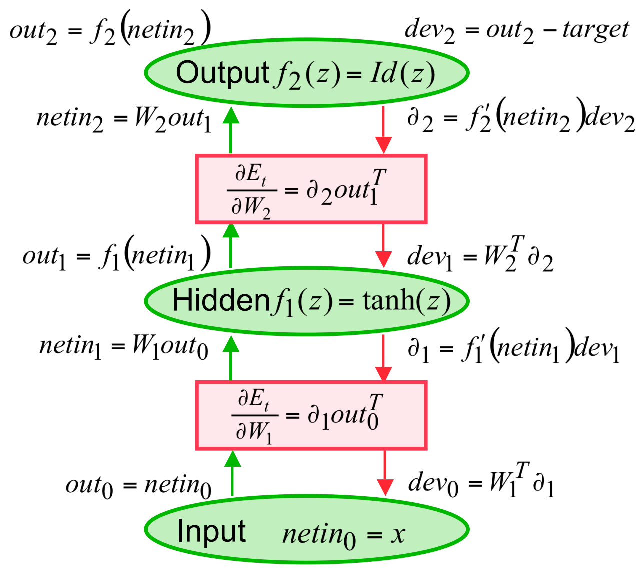
Distinct to linear superpositions of basis functions, NN are *composed substructures*

Mathematical Neural Networks - A Correspondence Principle



$$y = f_2(W_2\, f_1(W_1 x))$$

$$E = \frac{1}{T}\sum_{t=1}^{T} E_t = \frac{1}{T}\sum_{t=1}^{T}(y_t - y_t^d)^2 \rightarrow \min_{W_1, W_2}$$



- NN can approximate any high-dim, nonlinear input-output relationship.
- NN imply a **correspondence** between
  - ***Equations***,
  - ***Architectures***,
  - ***Local Algorithms***.

➡ Extensions to complex systems are straight forward.

Pattern by Pattern Learning - More than a Minimum Search

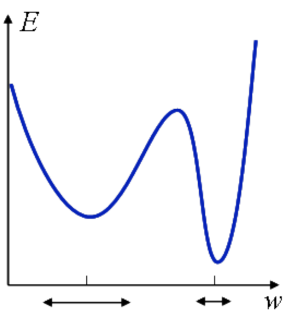
Task: $E = \frac{1}{T}\sum_{t=1}^{T} E_t = \frac{1}{T}\sum_{t=1}^{T}\left(NN(x_t, w) - y_t^d\right)^2 \rightarrow \min_w$

Notation: $g_t = \frac{\partial E_t}{\partial w}$, $g = \frac{1}{T}\sum_{t=1}^{T} g_t$

p-by-p learning: $\Delta w_t = -\eta g_t = -\eta g - \eta(g_t - g)$

steepest descent *stochastic search*

stochastic search
- is easy to compute
- helps to omit local minima
- is a local algorithm, that scales well
- ***implies a curvature penalty for free !!!***



In general, noise on the weights acts as a curvature regularization

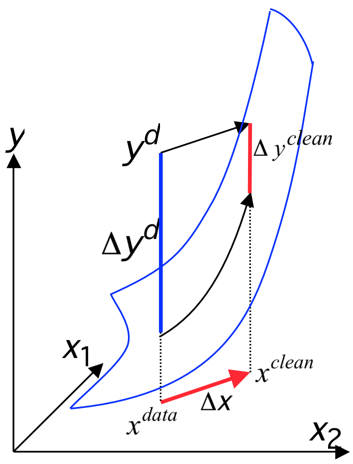
$$\langle E(w)\rangle = \frac{1}{T}\sum_t E(w+\Delta_t) = E(w) + \sum_i \underbrace{\left(\frac{1}{T}\sum_t \Delta_{it}\right)}_{\approx 0} \cdot \frac{\partial E}{\partial w_i} + \frac{1}{2}\sum_i var(\Delta_i)\frac{\partial^2 E}{\partial w_i^2}$$

Pattern by pattern learning $\Delta_t = -\eta\, g_t$ induces a local penalty on *w*

$$\langle E(w)\rangle = \frac{1}{T}\sum_t E(w+\text{Ä}_t) = E(w) + \frac{\eta^2}{2}\sum_i \mathrm{var}(g_i)\frac{\partial^2 E}{\partial w_i^2}$$

The Observer - Observation Dilemma

Geometry of Cleaning



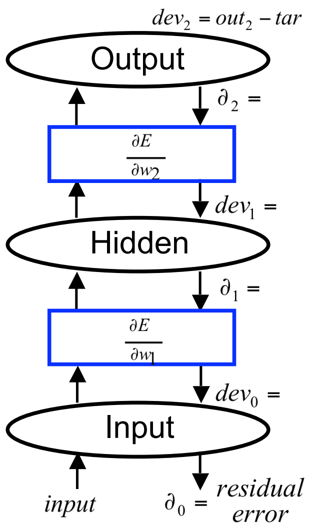
Accepting a correction in x may decrease the error in y significant.

Psychological Dilemma:
How far should observations determine our picture of the world?
&
How far should our picture of the world evaluate observations ?

Technical Dilemma:
How far should observations determine a model ?
&
How far should a model evaluate observations ?

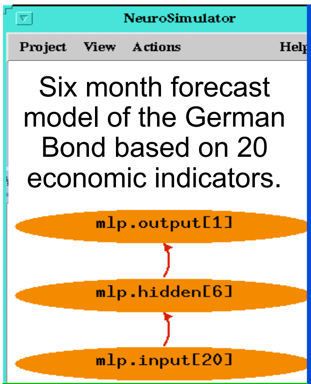
What is the optimal balance between model building and data cleaning ?

Calculus of Cleaning

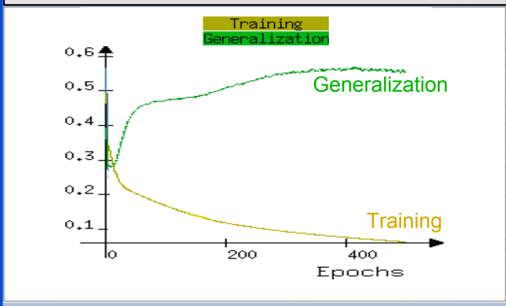


$$input = x^{data} + clean(\partial_0) + noise(\partial_0)$$

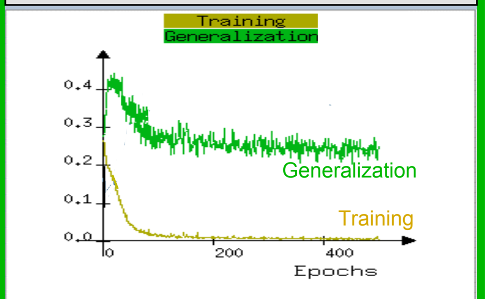
Improved Modeling using the Extended Learning Approach



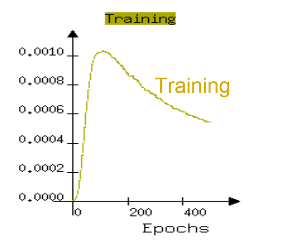
error development during learning by standard backpropagation

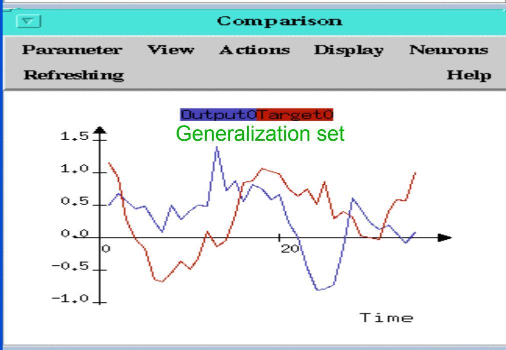


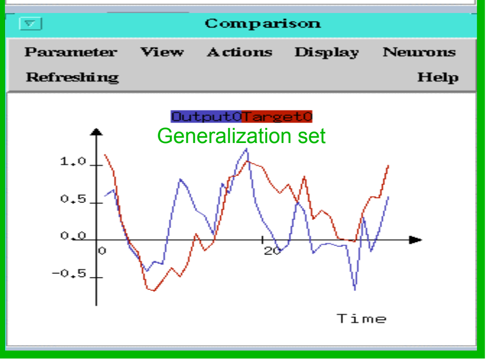
error level during learning by solving the observer - observation dilemma



evaluation of the input data by the model

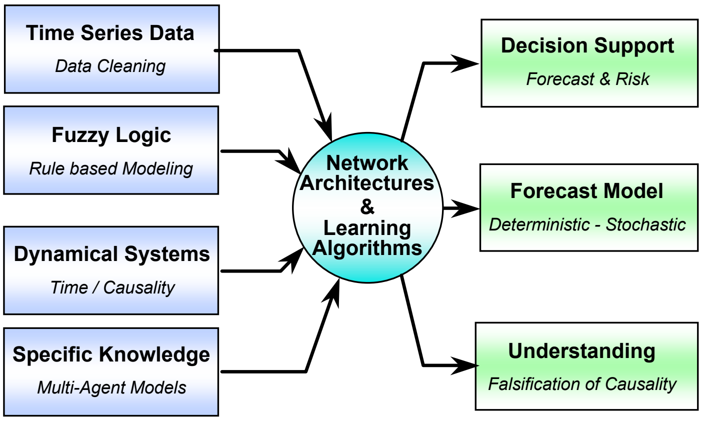






Neural Networks - From Data Mining to Model Building

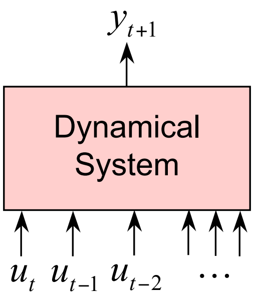
Data alone often do not cover the modeling task. Thus, we merge model building by data, prior knowledge and first principles.



Neural networks (**SENN**) allow systems analysis, forecasting & risk analysis as well as the setup of decision support systems.

Forecasting by Pattern Recognition versus State Space Modeling

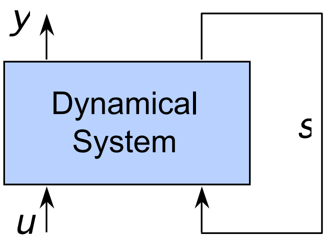
$$y_{t+1} = h(u_t, u_{t-1}, u_{t-2}, \dots)$$



$$y_{t+1} = V \tanh\left(W \begin{bmatrix} u_t \\ u_{t-1} \\ \dots \end{bmatrix} + w_0 \right) + v_0$$

The pattern recognition approach includes a limited understanding of time, while the state space approach models time recursively in form of a memory.

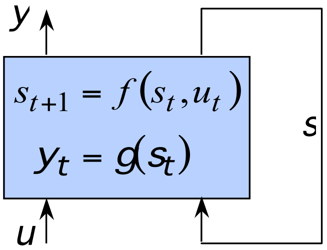
$$s_{t+1} = f(s_t, u_t)$$
$$y_t = g(s_t)$$



$$s_{t+1} = \tanh(As_t + Bu_t + c)$$
$$y_t = Cs_t$$

System identification is solved by minimizing the error function $\frac{1}{T}\sum_{t=1}^{T}\left(y_t - y_t^d\right)^2 \to \min.$

From Temporal Equations to Neural Network Architectures



$$s_{t+1} = \tanh(As_t + Bu_t + c)$$ state transition

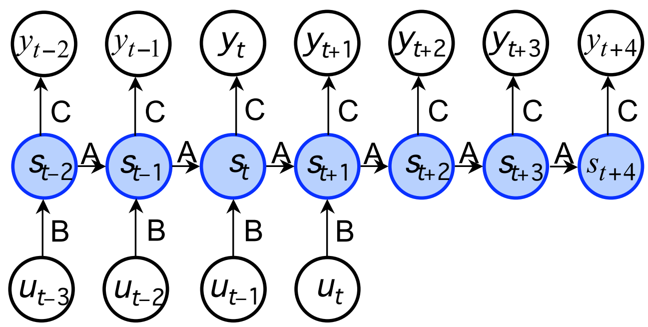
$$y_t = Cs_t$$ output equation

$$\sum_{t=1}^{T} (y_t - y_t^d)^2 \to \min_{A,B,C}$$ identification

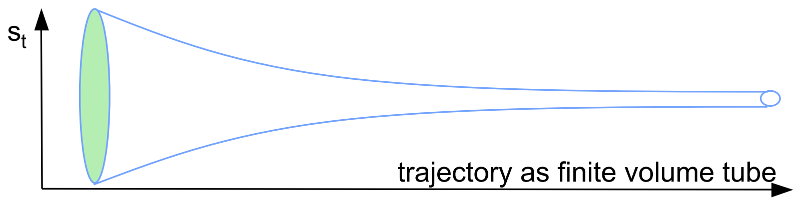
Finite unfolding in time transforms time into a spatial architecture.

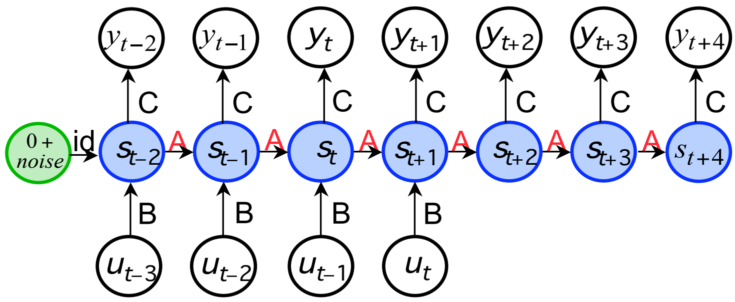
Open systems are superpositions of autonomous & external driven subsystems.

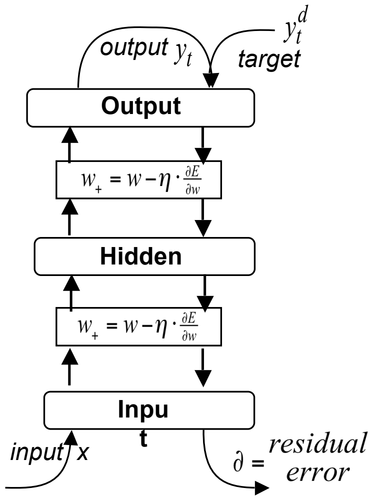
Long-term predictability depends on the extraction of a strong autonomous subsystem.



From Unknown Initial States to Finite Volume Trajectories







- The cleaning calculus allows to estimate the uncertainty of the unknown initial state s^0.
- By noise on the initial state we stiff the model against the unknown s^0:

$$Input_t^i = 0 + \Delta s^{0,i}_{\tau_1}(\partial) - \Delta s^{0,i}_{\tau_2}(\partial)$$ double local start noise

- Matrix A becomes a contraction to squeeze out the initial uncertainty.

Modeling Dynamical Systems with Error Correction Neural Networks

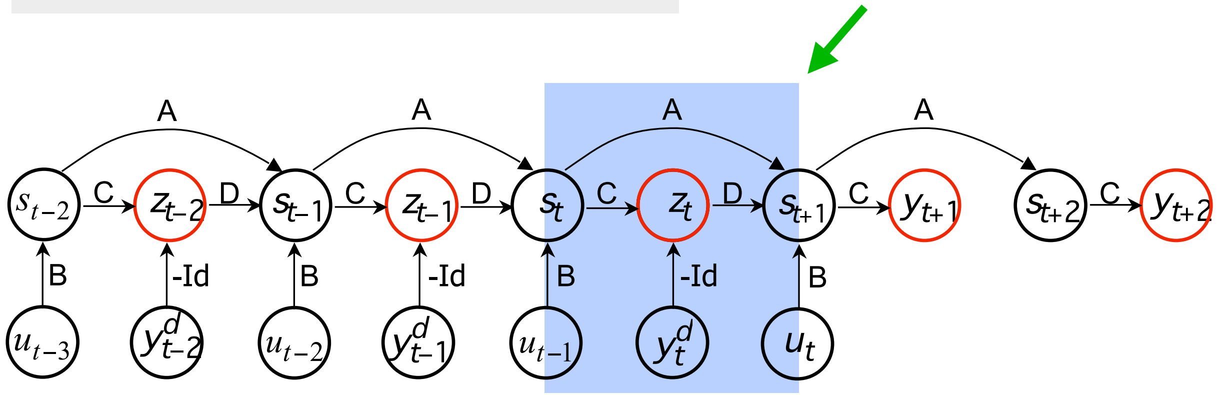
An error correction system considers the forecast error in present time as a reaction on unknown external information.

In order to correct the forecasting this error is used as an additional input, which substitutes the unknown external information.

$$s_{t+1} = f\left(s_t, u_t, (y_t - y_t^d)\right)$$

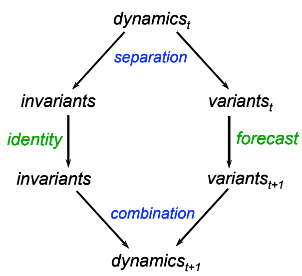
$$y_t = g(s_t)$$

$$\sum_{t=1}^{T} \left(y_t - y_t^d\right)^2 \rightarrow \min_{f,g}$$



Selected Topics of Interest

Variance-Invariance Separation



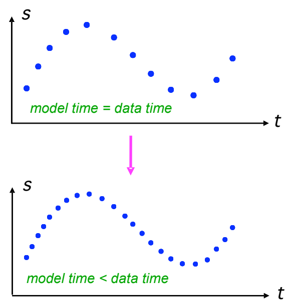
By invariants, the dynamics is constrained to move on a manifold in the phase space.

Examples of invariants: energy & impuls conservation, arbitrage freedom in finance.

For measured dynamical systems the existence of invariants is not a priori evident.

If we can identify invariants, the dimensionality of the forecasting problem can be lowered.

From Data Time to Model Time

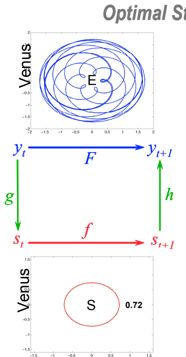


By coarsing the time grid we would suppress information.

A consistent time grid refinement can be derived from the a priori assumption of uniform causality.

If we understand time as a recurrency, uniform causality implies that iterated steps along the finer time grid have to match the measurements of the original time grid.

Optimal State Space Reconstruction



The observed dynamics y_t may be measured in a non-optimal coordinate system. This can result in a very inefficient description

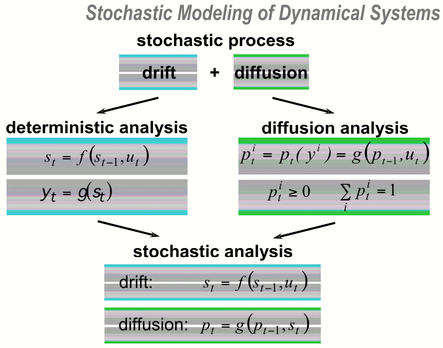
$$y_{t+1} = F(y_t).$$

In economics the measurement of a dynamics is tricky, because the units themselves change over time.

Find coordinate transformations g, h, such that the transformed trajectory s_t is easier to forecast.

Technically, the trajectory s_t should have a lower curvature than the original one.

Stochastic Modeling of Dynamical Systems in Discrete Time



The accumulated impact of the past is modeled in form of an internal state vector

The increasing forecast uncertainty is represented by a diffusion process.

By postprocessing:

$$y_t = \sum_i p_t^i y^i$$

$$\sigma_t^2 = \sum_i p_t^i \cdot (y^i - y_t)^2$$

Neural Networks in Mathematics

Significance

Method ↔ Content

Diff. Equations ↔ Physics

Neural Networks ↔ Complex Systems

Study of large, nonlin. Syst. in Nature & Economics

Insights

- NN offer a unique correspondence between Equations - Architectures - Local Algorithms.
- Interesting topics in nonlin. science are not an extension of linear models.
- Small systems are not a guideline for the analysis of large systems.

Functional Equations

$$f(x+y) = f(x) \cdot f(y)$$

1dim, nonlinear functional equation

$$f(\lambda x + \mu y) = \lambda f(x) + \mu f(y)$$

high dim, linear functional equation

$$f(\lambda + \mu, x) = f(\lambda, f(\mu, x))$$

state transition functional equation

Interrelationships

Statistics

Dyn. Systems Approximation

Control Neural Networks Optimization

Lie Theory Econometrics

Differential Geometry

Chaos

Selected References I

1. *Zimmermann, H.G.:* ***Neuronale Netze als Entscheidungskalkül,*** In: Neuronale Netze in der Ökonomie; Eds.: Rehkugler, H., Zimmermann, H.G., p. 1-87; Franz Vahlen; 1994; ISBN 3-8006-1871-0
2. *Neuneier, R.; Zimmermann, H.G.:* ***How to Train Neural Networks,*** In: Neural Networks: Tricks of the Trade; Eds.: Orr G. B., Müller K. R.; p. 373-423; Springer, 1998; ISBN 3-540-65311-2
3. *Zimmermann, H.G.; Neuneier, R.:* ***Neural Network Architectures for the Modeling of Dynamical Systems,*** In: A Field Guide to Dynamical Rec. Networks; Eds.: Kolen J. F., Kremer S. C.; IEEE Press; 2000; ISBN 0-780-35369-2
4. *Zimmermann, H.G.; Neuneier, R.; Grothmann, R.:* ***Modeling of the German Yield Curve by Error Correction Neural Networks***. In: Leung, K., Chan, L.-W. and Meng, H., Eds. Proceedings IDEAL 2000, p. 262-7, Hong Kong.
5. *Zimmermann, H.G.; Neuneier, R.; Grothmann, R.:* **Active Portfolio Management based on Error Correction Neural Networks,** In: Proceedings of Neural Information Processing (NIPS), Vancouver, Canada, Dec. 2001.
6. *Zimmermann, H.G.; Neuneier, R.; Grothmann, R.:* **Modeling of Dynamical Systems by Error Correction Neural Networks,** In: Modeling and Forecasting Financial Data, Techniques of Nonlinear Dynamics; Eds. Soofi, A. and Cao, L., Kluwer Academic Pub., 2002; ISBN 0-792-37680-3.
7. *Zimmermann, H.G.; Neuneier, R.; Grothmann, R.:* ***Undershooting: Modeling Dynamical Systems by Time Grid Refinements***, In: Proceedings of European Symposium on Artificial Neural Networks (ESANN) 2002, Belgium.
8. *Zimmermann, H.G.; Tietz, Ch.; Grothmann, R.:* **Yield Curve Forecasting by Error Correction Neural Networks & Partial Learning,** In: Proceedings of European Symposium on Artificial Neural Networks (ESANN) 2002, Belgium.
9. *Zimmermann, H.G.; Mueller A.; Erdem C. and Hoffmann R.:* **Prosody Generation by Causal-Retro-Causal Error Correction Neural Networks**. In: Proceedings Workshop on Multi-Lingual Speech Communication, ATR, 2000.
10. *Siekmann, S.; Neuneier, R.; Zimmermann, H.G.; Kruse R.:* ***Neuro Fuzzy Systems for Data Analysis,*** In: Computing with Words in Information / Intelligent Syst. 2; Eds.: Zadeh L. A., Kacprzyk J.;p. 35-74; Physica Verlag; 1999
11. *Zimmermann, H.G.; Neuneier, R.; Grothmann, R.:* ***Multi Agent Market Modeling of FX-Rates*** In: Advances in Complex Systems (ACS); Special Issue Vol. 4, No. 1; Hermes Science Publ. (Oxford) 2001.
12. *Zimmermann, H.G.; Neuneier, R.; Grothmann, R.:* **An Approach of Multi-Agent FX-Market Modelling based on Cognitive Systems,** In: Proc. of the Int. Conference on Artificial Neural Networks (ICANN); Springer, Vienna 2001.

Selected References II

13. *Zimmermann, H.G.; Neuneier, R.*, Tietz, Ch. and Grothmann, R.: ***Market Modeling based on Cognitive Agents*,** In: Proceedings of the Int. Conference on Artificial Neural Networks (ICANN), Madrid, 2002.

14. *Zimmermann, H.G. and Neuneier, R.:* ***The Observer – Observation Dilemma in Neuro-Forecasting***, In: Advances in Neural Information Processing Systems, Vol. 10, MIT Press, 1998.

15. *Zimmermann, H.G. and Neuneier, R.:* ***Combining State Space Reconstruction and Forecasting by Neural Networks.*** In: G. Bol, G. Nakhaeizadeh and K. H. Vollmer (Eds.): Datamining und Computational Finance, 7. Karlsruher Ökonometrie-Workshop, Wissenschaftliche Beiträge 174, Physica-Verlag, 1999, pp. 259-267.

16. *Zimmermann, H.G. and Neuneier, R.:* ***Modeling Dynamical Systems by Recurrent Neural Networks***, in: Ebecken, N. and Brebbia, C.A. (Eds.) Data Mining II, WIT Press, 2000, pp. 557-566.

17. *Zimmermann, H.G., Neuneier, R. and Grothmann, R.:* ***Multi-Agent Modeling of Multiple FX-Markets by Neural Networks****. IEEE Transactions on Neural Networks Special Issue 12(4):1-9, 2001*.

18. *Zimmermann, H.G.:* ***Optimal Asset Allocation for a Large Number of Investment Opportunities.*** In: Proceedings Forecasting Financial Markets (FFM), London 2002.

19. *Zimmermann, H.G., Grothmann, R. and Tietz Ch.:* ***ATM Cash Management based on Error Correction Neural Networks***, in: Fichtner, W.; Geldermann, J.: Einsatz von OR-Verfahren zur techno-ökonomischen Analyse von Produktionssystemen. Verlag Peter Lang, Frankfurt, 2003

20. *Grothmann, R.:* ***Multi-Agent Market Modeling based on Neural Networks***, Ph.D. Thesis, University of Bremen, Germany, E-Lib, 2002, http://elib.suub.uni-bremen.de/publications/dissertations/E-Diss437_grothmann.pdf

21. Zimmermann, H.G, Grothmann, R., Schäfer, A.M., Tietz, Ch.: Dynamical Consistent Recurrent Neural Networks, in: Prokhorov, D.: Proc. of the Int. Joint Conference on Neural Networks (IJCNN), Montreal 2005.

22. Zimmermann, H.G, Grothmann, R., Schäfer, A.M., Tietz, Ch.: Identification and Forecasting of Large Dynamical Systems by Dynamical Consistent Neural Networks, in: Haykin, S.; Principe, J.; Sejnowski, T. and Mc Whirter, J. [Eds.]: New Directions in Statistical Signal Processing: From Systems to Brain, MIT Press, forthcomming 2006.

23. *Tietz, Ch.:* ***SENN User Manual****,* Siemens AG, 2004.