The Science of Prediction

with

Neural Networks

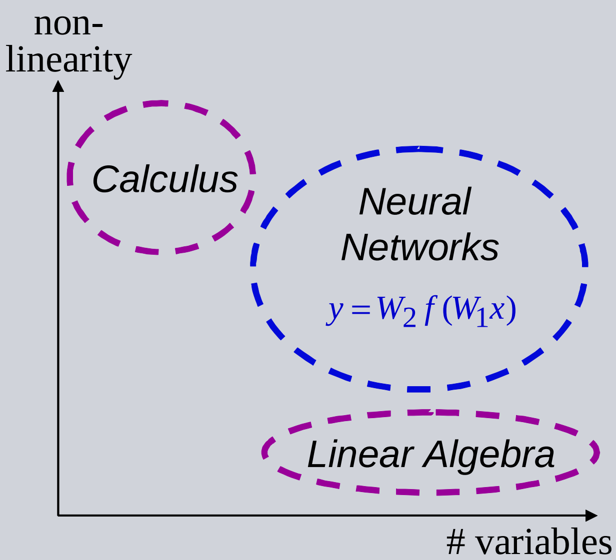
Hans Georg Zimmermann

Siemens AG, Corporate Technology

Email : Hans_Georg.Zimmermann@siemens.com

Mathematical Neural Networks

Complex Systems



Existence Theorem:

(Hornik, Stinchcombe, White 1989)

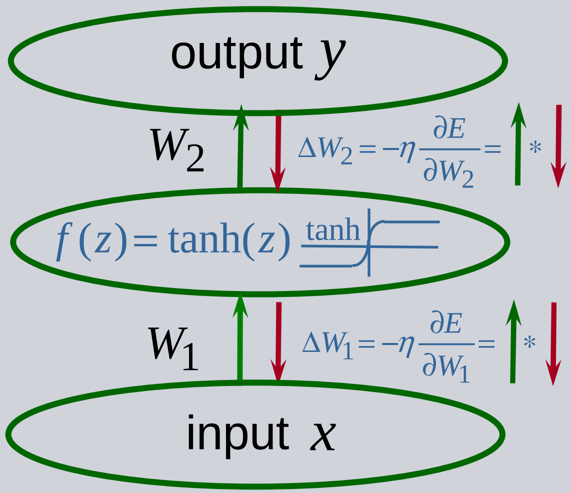
3-layer neural networks can approximate any continuous function on a compact domain.

Nonlinear Regression

Based on data identify an input-output relation

$$y = W_2 f(W_1 x)$$

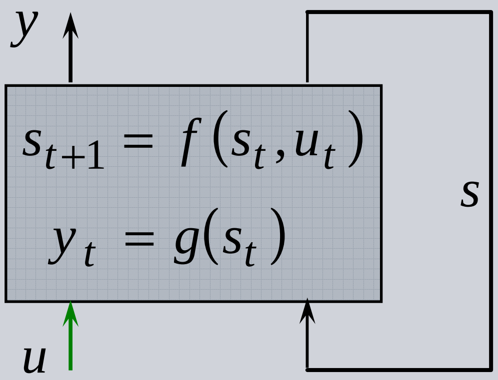
$$E = \sum_{t=1}^{T} \left(y_t - y_t^d\right)^2 \rightarrow \min_{W_1, W_2}$$



Neural networks imply a **Correspondence** of ***Equations***, ***Architectures***, ***Local Algorithms***.

SIEMENS

Modeling of Open Dynamical Systems with Recurrent Neural Networks (RNN)

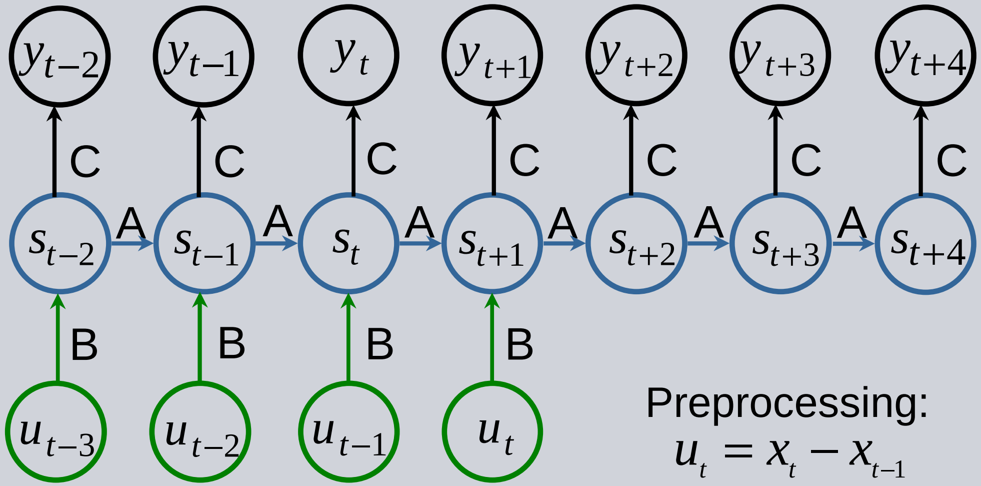


$s_{t+1} = \tanh\left(As_t + Bu_t\right)$	state transition
$y_t = Cs_t$	output equation
$\sum_{t=1}^{T}\left(y_t - y_t^d\right)^2 \rightarrow \min\limits_{A,B,C}$	identification

Finite unfolding in time transforms time into a spatial architecture. We assume, that x_t=const in the future.

The analysis of open systems by RNNs allows a decomposition of its autonomous & external driven subsystems.

Long-term predictability depends on a strong autonomous subsystem.



Preprocessing:
$u_t = x_t - x_{t-1}$

Modeling Dynamical Systems with Error Correction Neural Networks (ECNN)

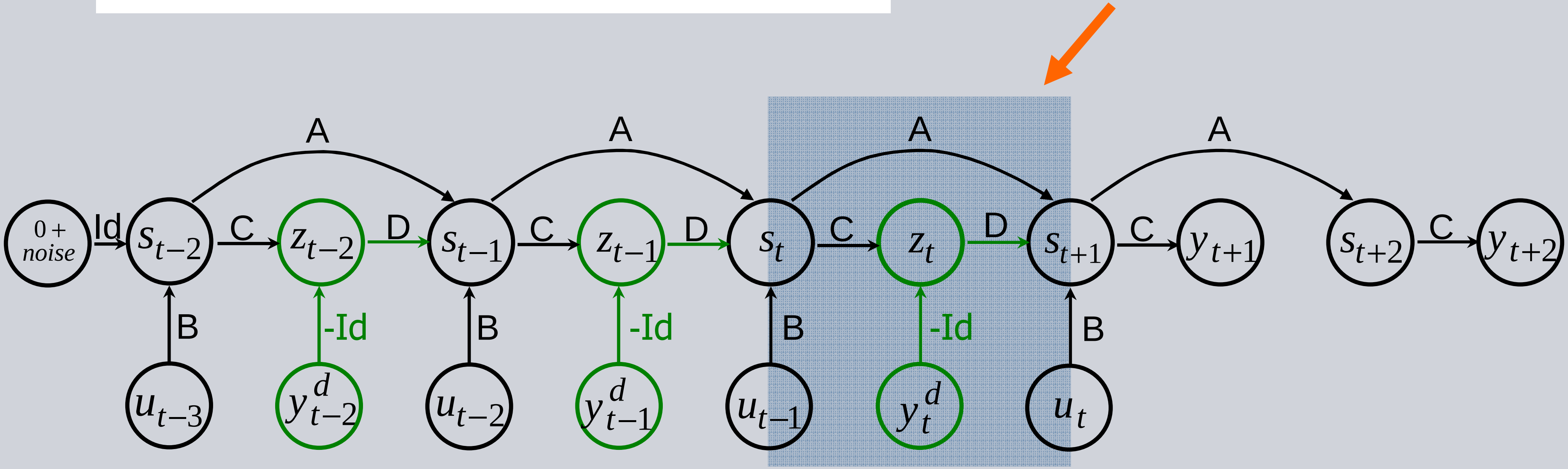
An error correction system considers the forecast error in present time as a reaction on unknown external information.

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

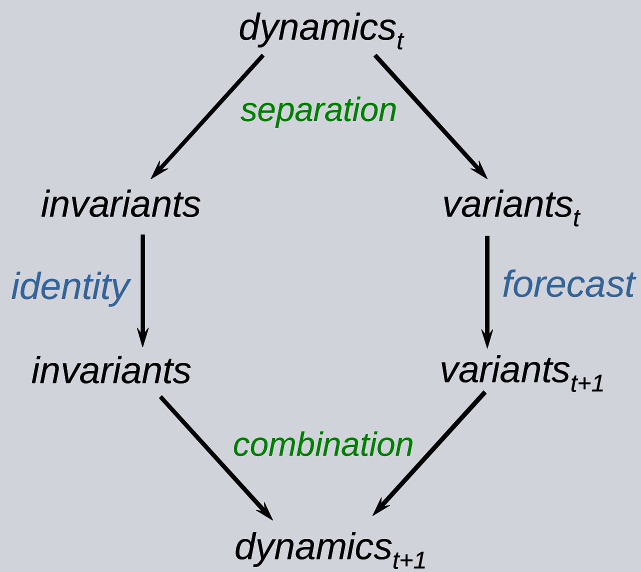
$$s_{t+1} = f\left(s_t, u_t, \left(y_t - y_t^d\right)\right)$$

$$y_t = g(s_t)$$

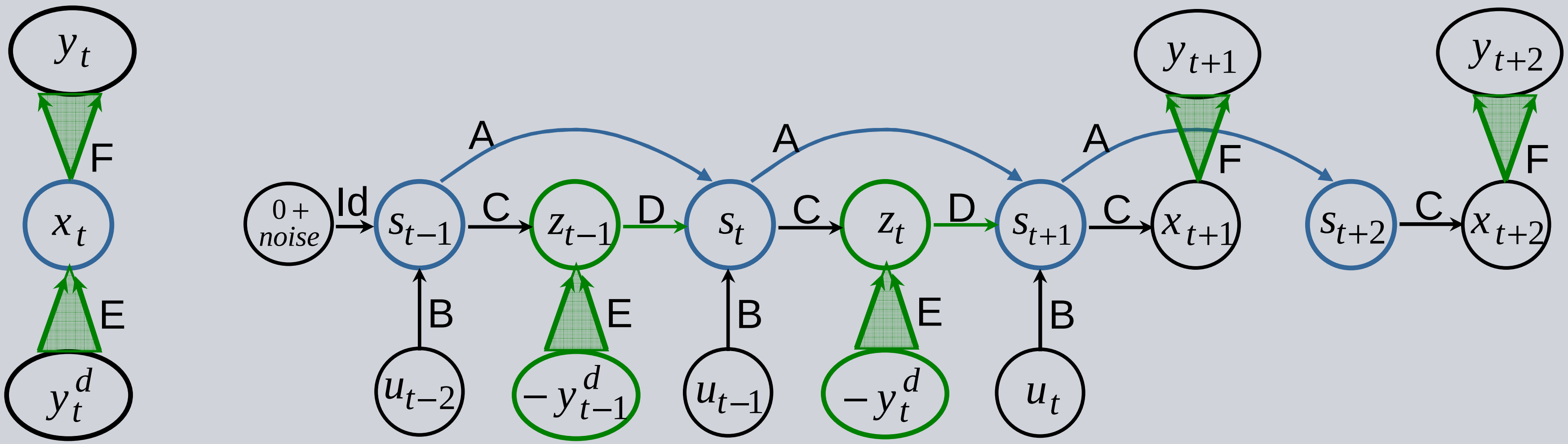
$$\sum_{t=1}^{T} \left(y_t - y_t^d\right)^2 \to \min_{f,g}$$



Combining Variance - Invariance Separation with ECNN

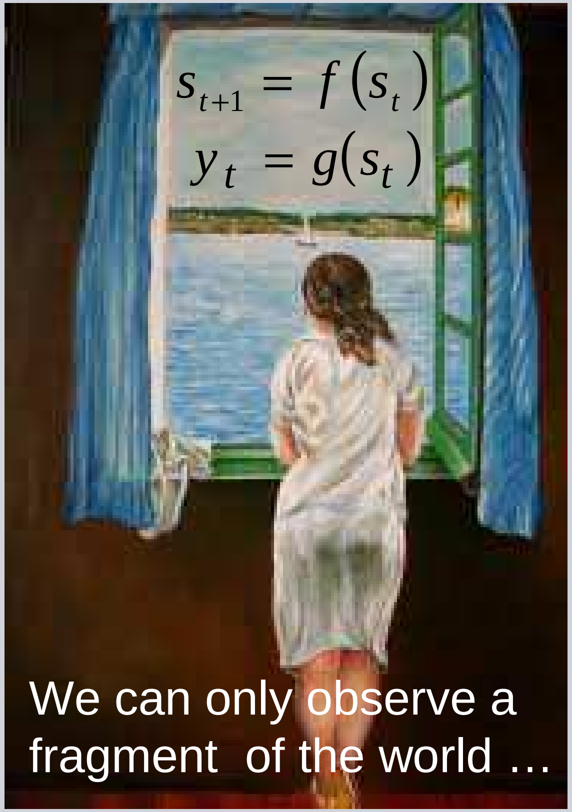


- The bottleneck autoassociator solves the variance - invariance decomposition.
- The Error Correction Neural Network solves the transformed temporal problem.
- The sub-networks are implicitly coupled by shared weights.



 CT RTC BAM

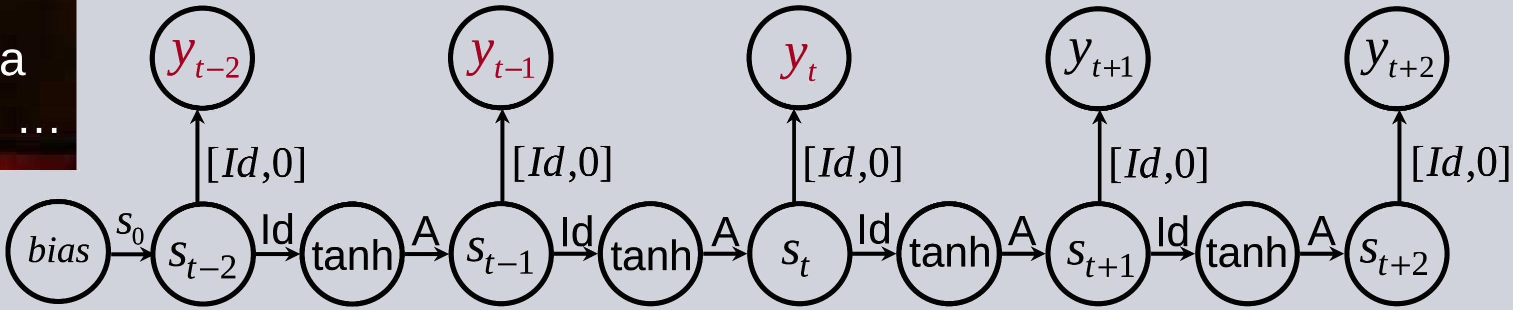
Modeling Closed Dynamical Systems with Recurrent Neural Networks



$$s_{t+1} = A\tanh(s_t) \quad , s_0$$ state transition

$$y_t = [Id, 0] s_t$$ output equation

$$\sum_{t=1}^{T} \left(y_t - y_t^d\right)^2 \to \min_{A, s_0}$$ identification

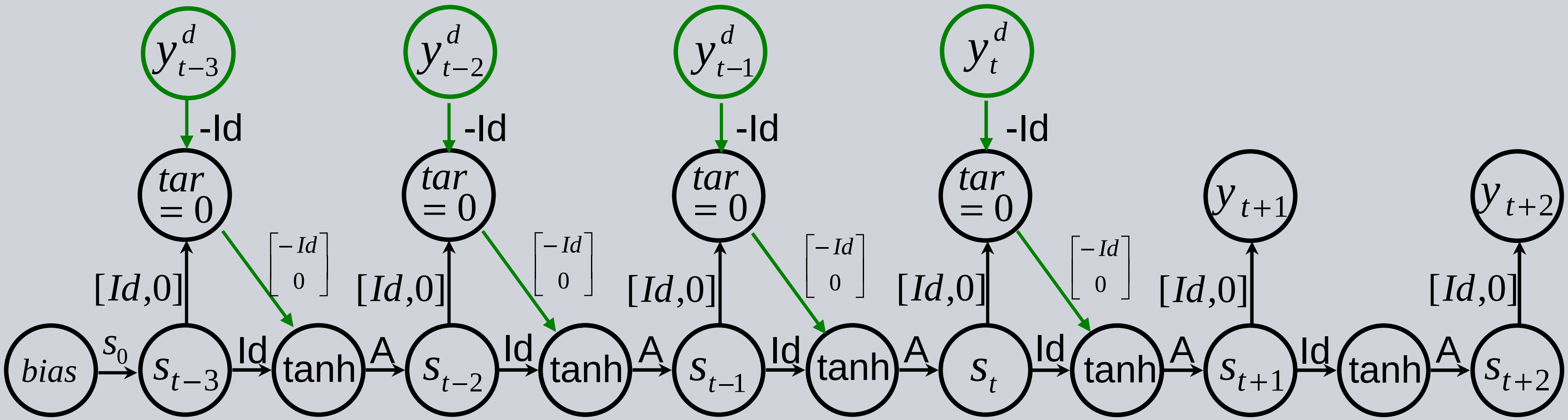


… but to understand the dynamics of the observables, we have to reconstruct at least a part of the hidden states of the world. Forecasting is based on observables and hidden states.

SIEMENS

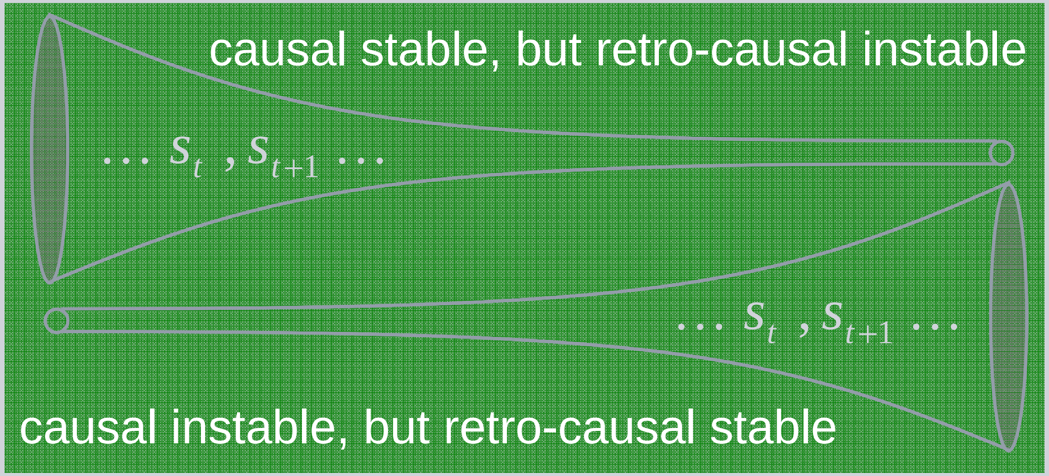
The Identification of Dynamical Systems in Closed Form

Embed the original architecture into a larger architecture, which is easier to learn. After the training, the extended architecture has to converge to the original model.

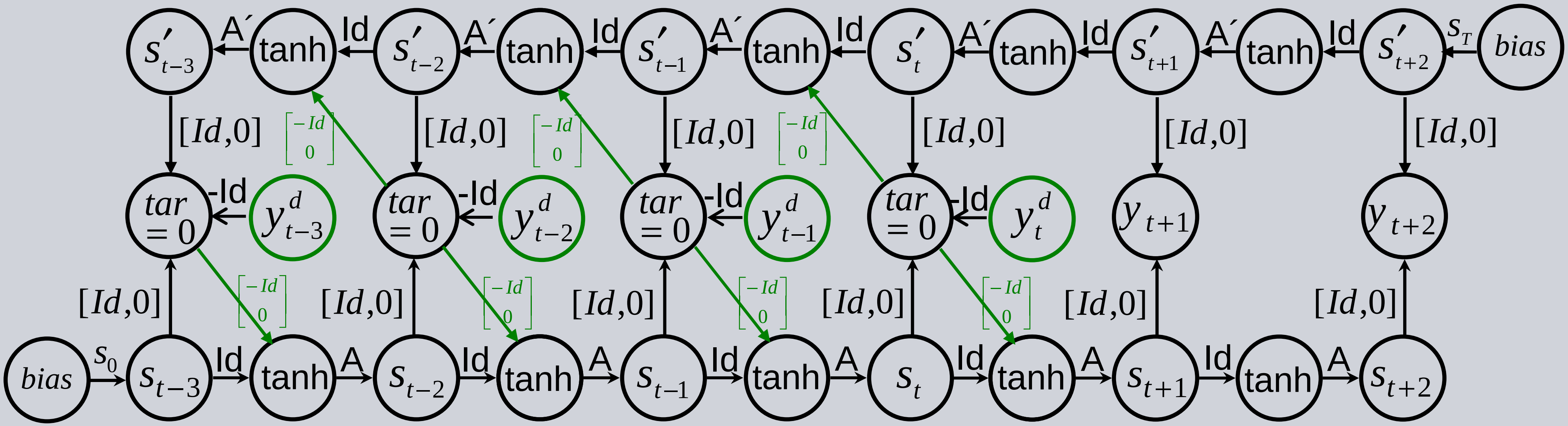


The essential task is NOT to reproduce the past observations, but to identify related hidden variables, which make the dynamics of the observables reasonable.

Decomposition of the Dynamics in Stable & Instable Parts of CRCNN

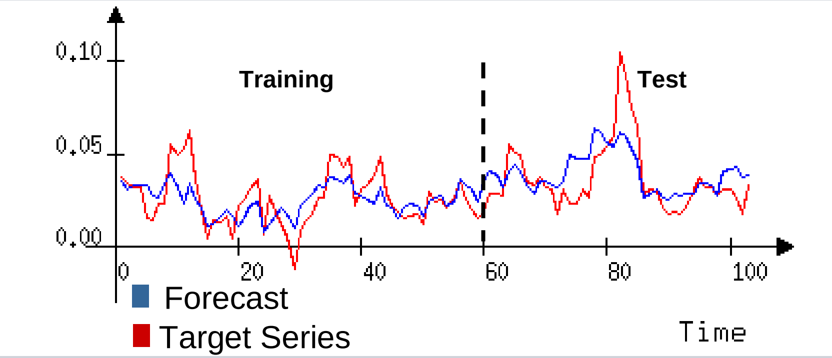


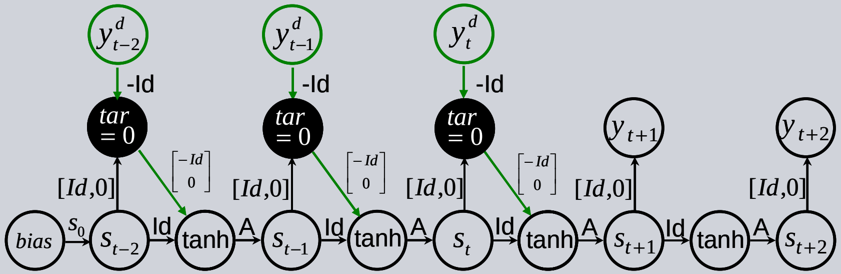
The tubes describe possible ensembles of instable trajectories.
If we have a causal instable (chaotic) sub dynamics, it is retro-causal stable and vice versa. Learning automatically should enforce a decomposition such that the causal and the retro-causal branch are both stable.

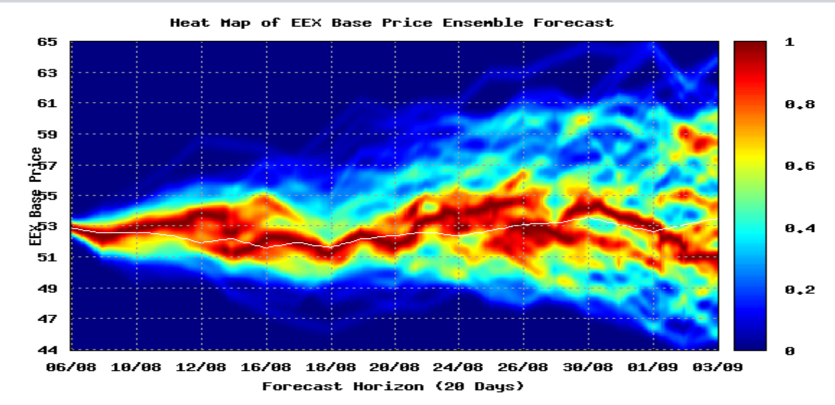


SIEMENS

Approaches to Model Uncertainty in Forecasting







1 Measure uncertainty as volatility (variance) of the target series. The underlying forecast model is a constant. Thus sin(ωt) can be highly uncertain!??

2 Build a forecast model. The error is interpreted as uncertainty in form of additive noise. The width of the uncertainty channel is constant over time.

3 Describe uncertainty as a diffusion process (random walk). The diffusion channel widens over time, e.g. scaled by the one-step model error.

For large systems 2 & 3 fail: We have to learn to zero error → the uncertainty channel disappears.

4 One large model doesn't allow to analyze forecast uncertainty, but an ensemble forecast shows the characteristics of an uncertainty channel: Given a finite set of data, there exist many perfect models of the past data, showing different future scenarios caused by different estimations of the hidden states.

On Model Building …

Start with a simple = universal framework

e.g.
- neural networks
- others …

Add reasonable a priori structure

e.g.
- monotonic input-output relations
- diversity / similarity analysis
- dynamical systems
- dynamics on manifolds
- linearity

Add data

e.g.
- temporal / cross sectional
- continuous / ordered / nominal

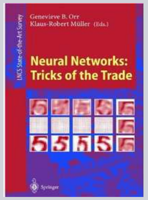
Interpret the a posteriori model causality

on past data we can detect correlations only, their interpretation as causality is an intellectual insight, supported by prediction

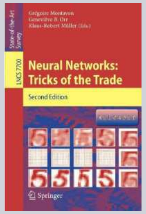
SIEMENS

Selected References

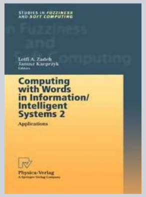
Zimmermann, H.G.: Neuronale Netze als Entscheidungskalkül, In: Neuronale Netze in der Ökonomie; Eds.: Rehkugler, H., Zimmermann, H.G., p. 1-87; Franz Vahlen; 1994; ISBN 3-8006-1871-0



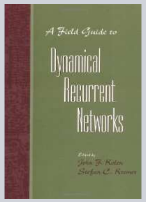
Neuneier, R.; Zimmermann, H.G.: How to Train Neural Networks, In: Neural Networks: Tricks of the Trade; Eds.: Orr G. B., Müller K. R.; p. 373-423; Springer, 1998; ISBN 3-540-65311-2



Zimmermann, H.G.; Tietz, C.; Grothmann, R.: Forecasting with Recurrent Neural Networks, In: Neural Networks: Tricks of the Trade, **Second Edition**; Eds.: Montavon, G; Orr G. B., Müller K.R.; p. 687-707; Springer, 2012; ISBN 978-3-642-35288-1

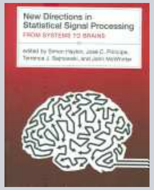


Siekmann, S.; Neuneier, R.; Zimmermann, H.G.; Kruse R.: Neuro Fuzzy Systems for Data Analysis, In: Computing with Words in Information / Intelligent Syst. 2; Eds.: Zadeh L. A., Kacprzyk J.;p. 35-74; Physica Verlag; 1999



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

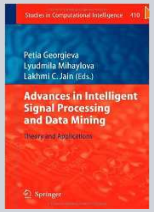
Zimmermann, H.G.; Neuneier, R.; Grothmann, R.: Modeling of Dynamical Systems by Error Correction Neural Networks, In: Modeling and Forecasting Financial Data, Techniques of Nonlinear Dynamics; Eds. Soofi, A. and Cao, L., Kluwer, 2002



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.; et. al. [Eds.]: New Directions in Statistical Signal Processing: From Systems to Brain, MIT, 2006.

Zimmermann, H.-G., Grothmann, R., Tietz Ch and von Jouanne-Diedrich, H: Market Modeling, Forecasting and Risk Analysis with Historical Consistent Neural Networks, in: Proceedings of the International Conference on Operations Research, Munich 2010, Springer, 2011

Zimmermann, H.-G., Grothmann, R., Tietz, Ch.: Forecasting market prices with causal-retro-causal neural networks. In: Lüthi, H.-J. et. al (eds.): OR Proceedings 2011: Selected Papers of the Int. Conference on Operations Research (OR 2011), Zurich, Springer, 2012



Zimmermann, H.G, Grothmann, R., Tietz, Ch.: Historical Consistent Neural Networks: New Perspectives on Market Modelling, Forecasting and Risk Analysis, in: Georgieva P. et al. (Eds): Advances in Intelligent Signal Processing and Data Mining: Theory and Applications, Springer 2012