

## Self-organized quantification of hormone pulsatility: Separating growth hormone secretion in health and disease

Klaus Prank,<sup>\*</sup> Mirko Kloppstech,<sup>\*</sup> Steven J. Nowlan,<sup>†</sup> Terrence J. Sejnowski,<sup>‡</sup> and Georg Brabant<sup>\*</sup>  
<sup>\*</sup>Department of Clinical Endocrinology, Medical School Hannover, D-30623 Hannover, Germany; <sup>†</sup>Lexicus Inc.,  
San Jose, California 95134; <sup>‡</sup>Howard Hughes Medical Institute and Computational Neurobiology Laboratory,  
The Salk Institute, San Diego, California 92186-5800; <sup>†</sup>Department of Biology, University of California,  
San Diego, La Jolla, California 92093.

### Address correspondence to:

Dr. Klaus Prank  
Dept. of Clinical Endocrinology  
Medical School Hannover  
D-30623 Hannover, Germany

E-mail: ndxdpran@rrzn-user.uni-hannover.de

### Abstract

The pulsatile pattern of growth hormone (GH) secretion was assessed over 24 hours in 10 healthy subjects and in 6 patients with a GH producing pituitary tumor (acromegaly) before treatment with the somatostatin analogue octreotide by sampling blood every 10 min. Time series prediction based on a single feedback neural network has recently been demonstrated to separate the secretory dynamics of parathyroid hormone (PTH) in healthy controls from patients with osteoporosis, a severe bone disease. To reveal possible differences of GH secretory dynamics in healthy controls and patients with acromegaly we tested time series prediction based on a single feedforward neural network and a system of multiple neural networks acting in parallel (*adaptive mixtures of local experts*). Both approaches significantly separated GH dynamics under the various conditions. By performing a self-organized quantification of hormone pulsatility of GH the *adaptive mixtures of local experts* performed significantly better than the single network approach. It thus may represent a potential tool to characterize alterations of the dynamic regulation in hormonal systems associated with diseased states.

### Introduction

Pulsatile hormone secretion has been demonstrated in a large number of different endocrine systems with pulse frequencies ranging from approximately 6 to 140 pulses in 24 hours (Brabant et al. 1992). This dynamic mode of secretion has important effects on the regulation of target cells and organs in health and disease (Brabant et al. 1992).

Time series prediction has been used to capture regularities in the temporal pattern of complex time series particularly in the context of nonlinear dynamical systems and to separate deterministic from random behavior (Sugihara and May 1990, Tsonis and Elsner 1992, Weigend and Gershenfeld 1993). This technique tries to predict the future dynamics of a time series from a number of past values. Differences in the temporal dynamics of a system are then reflected in an altered predictability. Such predictive approaches have been effectively used in several biological systems (Blinowska and Malinowski 1991, Chang et al. 1994, Lefebvre et al. 1993, Schiff et al. 1994, Sugihara 1994, Scott and Schiff 1995) and may be applied for short time series containing a very limited number of data points (Sugihara and May 1990, Tsonis and Elsner 1992). We recently applied neural networks to the prediction of hormone concentration time series to clearly separate the secretory dynamics of parathyroid hormone (PTH) in a group of osteoporotic patients from that in healthy controls, demonstrating a significantly higher predictability of the PTH secretory dynamics in healthy subjects than in the osteoporotic group (Prank et al. 1995). In 24h profiles of PTH serum concentrations we could additionally demonstrate a switching behavior between two dynamic phases, a phase of high and one of low predictability. Based on these results a single neural network predictor might not be optimal for the prediction of time series with different switching dynamics.

### Subjects

Ten male young lean subjects and six patients with clinical and biochemical diagnosis of acromegaly not cured by previous surgical and/or radiation therapies took part in this study. The studies were approved by the local Committee on Medical Ethics, and all subjects and patients gave their informed written consent. Blood samples were taken every 10 minutes over 24 hours starting at 1800 h. Healthy controls were studied under normal nutrition. The 24h GH secretory profile in the acromegalic group was studied without any medication. The study design has been described in more detail previously and some of the data of these studies have been included in the present data set (Riedel et al. 1992, Riedel et al. 1995).

### Time series prediction

In the present study we used feedforward neural networks to predict future values of the time series of GH serum concentrations. This form of a predictive model was chosen since it is relatively easy to control over-fitting using regularization functions and cross-validation. The next value  $x(t_i)$  of each GH concentration time series was predicted from  $m$  past values  $x(t_i) = f(x(t_{i-1}), x(t_{i-2}), \dots, x(t_{i-m})) + \epsilon_i$ , where  $x(t_{i-1})$  is the preceding value, and  $\epsilon_i$  corresponds to noise or fitting error. The function  $f$  represents feedforward neural networks with linear or nonlinear (sigmoidal) activation functions. The neural networks were trained (fitting the model  $f$  to data by using a conjugate gradient descent technique) to predict 10 min. into the future within either group. Prior to training the original data sets were normalized to a [0,1] interval to account for the nonlinearities in the activation functions and different mean GH levels between the healthy controls and acromegalic patients. The training within each group was performed using a "leave one out technique" where the time series being tested for prediction is left out of the training procedure. To predict multiple time steps ahead the value  $x(t)$  predicted one time step ahead was iterated back to predict two steps into the future. This procedure can be iteratively repeated to predict any given number of time steps ahead. We used the average relative variance (*arv*) as prediction error estimate to decorrelate it from the variance of the respective time series:

$arv = \langle (x_i - x'_i)^2 \rangle / \sigma^2(x_i)$ , where the angle brackets denote the mean over all predicted values and  $\sigma^2(x_i)$  denotes the variance of the measured time series. The simulations were performed on a Sun SPARCstation 20 using customized C-code. A large variety of neural network architectures was explored. We found a 7 input units, 10 hidden units, 1 output unit network with sigmoidal activation functions as the best predictor within the control group and a 5 input units, 8 hidden units, 1 output unit network with sigmoidal activation functions for the acromegalic group.

### Mixture of Experts

Divide-and-conquer strategies are effective methods for solving complex problems by dividing it into simpler problems whose solutions can be combined to yield a solution to the complex problem (Jacobs et al. 1991, Jordan and Jacobs 1994, Nowlan 1990, Pawelzik et al. 1996). The mixture of experts is a modular neural network system composed of several different "expert" networks plus a gating network that decides which of the experts should be used for each input (Fig. 1). During the training procedure the experts compete to generate the desired output for each input pattern, and adapt to a particular input pattern in proportion to their performance relative to the other expert networks. When an expert has less error than the weighted average of the errors of all experts, its responsibility for that case is increased, and when it does worse, its responsibility is decreased. In this way, individual experts specialize for specific subsets of the input pattern. The experts are therefore local since the weights in one expert are decoupled from the weights in other experts. At the same time, the gating network learns to select the best performing expert for a given case by adjusting the mixing proportions of the experts. Simulations on a complex vowel classification task have shown that adaptive mixtures of local experts are able to effectively decompose a problem to yield higher classification performance than a single network (Nowlan 1991). The mixtures of experts also generalizes better from limited amounts of training data.

In our simulations we varied the number of local experts between 2 and 6 and the input window size between 7 and 16 data values. Each expert consisted of a feedforward linear network without hidden units. We trained the best predictive mixture of experts system using pooled reference data from 5 healthy controls. This system consisted of 5 linear experts with 9 input units each and a gating network with 3 hidden units and sigmoidal activation functions. It was tested for its predictive ability on the remaining control and acromegalic GH time series. The number of the respective expert selected to predict a given value of the time series was recorded versus time. Finally, a frequency distribution of the expert selection was computed for each subgroup. Adaptive mixtures of local experts were simulated with customized C-code on a Sun SPARCstation 20.

### Results

The one-step ahead prediction error (*arv*) was found to be significantly ( $p < 0.002$ ) lower in healthy controls ( $arv=0.14 \pm 0.07$ ) than in acromegalic patients ( $arv=0.51 \pm 0.29$ ). This was reflected in correctly predicting the pulsatile features of the 24h GH profile in healthy controls for 100 minutes ahead (Fig. 2b), in contrast to a significant predictability of only 10 min. in acromegaly (Fig. 2c). Compared to a single neural network predictor the predictive performance of the modular approach was superior for both groups resulting in a significantly smaller one-step ahead prediction error (controls:  $arv = 0.07 \pm 0.03$ ,  $p < 0.01$ ; acromegaly:  $arv=0.24 \pm 0.13$ ,  $p < 0.01$ ). Analyzing the temporal pattern of the selected local expert for each predicted GH value we found that the multiple network approach had performed a self-organized segmentation of the pulsatile secretory pattern (SOPUL) in healthy controls and acromegalic patients reflected in a switching between distinct expert networks. The selection pattern of the local experts easily separates acromegalic patients healthy controls. During the training phase, each expert network specialized to predict certain parts of a GH pulse. Expert 1 focussed its performance on predicting the decrease of a hormonal pulse whereas expert 5 was best in predicting the increase of a GH pulse and the GH baseline. Although experts 2 to 4 contributed significantly to the predictive performance of this modular approach they were selected in only about 1 % of the cases in healthy controls and in about 5 to 8 % of the cases in the patient group. The frequency distributions of these most probable experts significantly differ in acromegalic patients from those in healthy controls.

Expert 1 was chosen in 2.5 % in normal controls compared to 12.7 % in acromegaly, which separated the controls from acromegaly with high confidence ( $p < 0.01$ ).

### Discussion

Using time series prediction with a single feedforward neural network we were able to separate the secretory dynamics of GH of healthy subjects from acromegalic patients. We previously used the same analytical approach successfully to predict PTH serum levels and to separate healthy controls from osteoporotic patients where standard methods of time series analysis as computing the mean, the variance, or the power spectrum failed (Prank et al. 1995). However, such a single neural network predictor has distinct disadvantages when applied to the prediction of time series with nonuniform local dynamics as described for episodic hormone release in most endocrine systems with phases of secretory bursts and quiescence (Brabant et al. 1992). This may elicit strong interference effects in single feedforward neural networks and result in slow learning and poor generalization. To circumvent these problems we used adaptive mixtures of local experts in the present study. The modular neural network architecture reduces interference by a mixture of several distinct expert neural networks which are chosen by a gating network to optimally adapt for each training and testing case. By this approach each experimental group was predicted much better than using the single neural network. Each expert specializes on distinct parts of a time series. The switching between local experts may be used to perform a self-organized segmentation of the pulsatile pattern of hormone secretion (SOPUL). We show that in acromegalic patients the expert specialized on predicting the decrease of a GH pulse contributed significantly more to the overall prediction of the GH time series than in healthy controls. Thus, histograms of selected expert networks may be used to separate temporal rhythms in health and disease.

In contrast to standard methods for the analysis of pulsatile patterns of secretion (Prank and Brabant 1994) our SOPUL approach does not require any previous knowledge of the physiological characteristics of the system. Therefore it may reflect better the "true" pattern of endocrine information transfer than the approaches published so far which all rely on certain assumptions on the nature of a pulse. The approach discussed here may help to expand standard techniques for the analysis of hormonal time series beyond counting pulse frequency and amplitude. Our approach using multiple adaptive neural networks may form an analytical basis to test the potential importance of temporal coding in the endocrine system.

### Acknowledgments

K. P. and G. B. were supported by DFG. S. J. N. and T. J. S. were supported by Howard Hughes Medical Institute.

### References

- Blinowska, K. J., and M. Malinowski. 1991. *Biol. Cybern.* 66, 159.
- Brabant, G. et al. 1992. *Trends Endocrinol. Metab.* 3, 183.
- Chang, T., et al. 1994. *Biophys. J.* 67, 671.
- Jacobs, R. A. et al. *Neural Comp.* 3, 79.
- Jordan, M. I., and R. A. Jacobs. 1994. *Neural Comp.* 6, 181.
- Lefebvre, J. H. et al. 1993. *Chaos* 3:267.
- Nowlan, S. J. 1990. *Technical Report CRG-TR-90-5*, Department of Computer Science, University of Toronto.
- Nowlan, S. J. 1991. *CMU-CS-91-126*, School of Computer Science, Carnegie Mellon University, Pittsburgh, Ph.D. thesis. 110-130.
- Nowlan, S. J., and G. E. Hinton. 1992. *Neural Comp.* 4, 473.
- Pawelzik, K. et al. 1996. *Neural Comp.* In press.
- Prank, K., and G. Brabant. 1994. In *Methods Neurosci.* vol. 20. J. E. Levine, ed. Academic Press, San Diego. 377.
- Prank, K. et al. 1995. *J. Clin. Invest.* 95, 2910.
- Riedel, M. et al. *Clin. Endocrinol. (Oxf.)* 37: 233.
- Riedel, M. et al. 1995. *Metabolism* 44, 605.
- Schiff, S. J. et al. 1994. *Biophys. J.* 67, 684.
- Scott, D. A., and S. J. Schiff. 1995. *Biophys. J.* 69, 1748.
- Sugihara, G., and R. M. May. 1990. *Nature (Lond.)* 344, 734.
- Sugihara, G. 1994. *Phil. Trans. R. Soc. Lond. A* 348, 477.
- Tsonis, A. A., and J. B. Elsner. 1992. *Nature (Lond.)* 358, 217.
- Weigend, A. S., and N. A. Gershenfeld. 1993. *Time Series Prediction: Forecasting the Future and Understanding the Past*. SFI studies in the sciences of complexity, Proc. Vol. XV, Addison-Wesley, Reading. 643 pp.

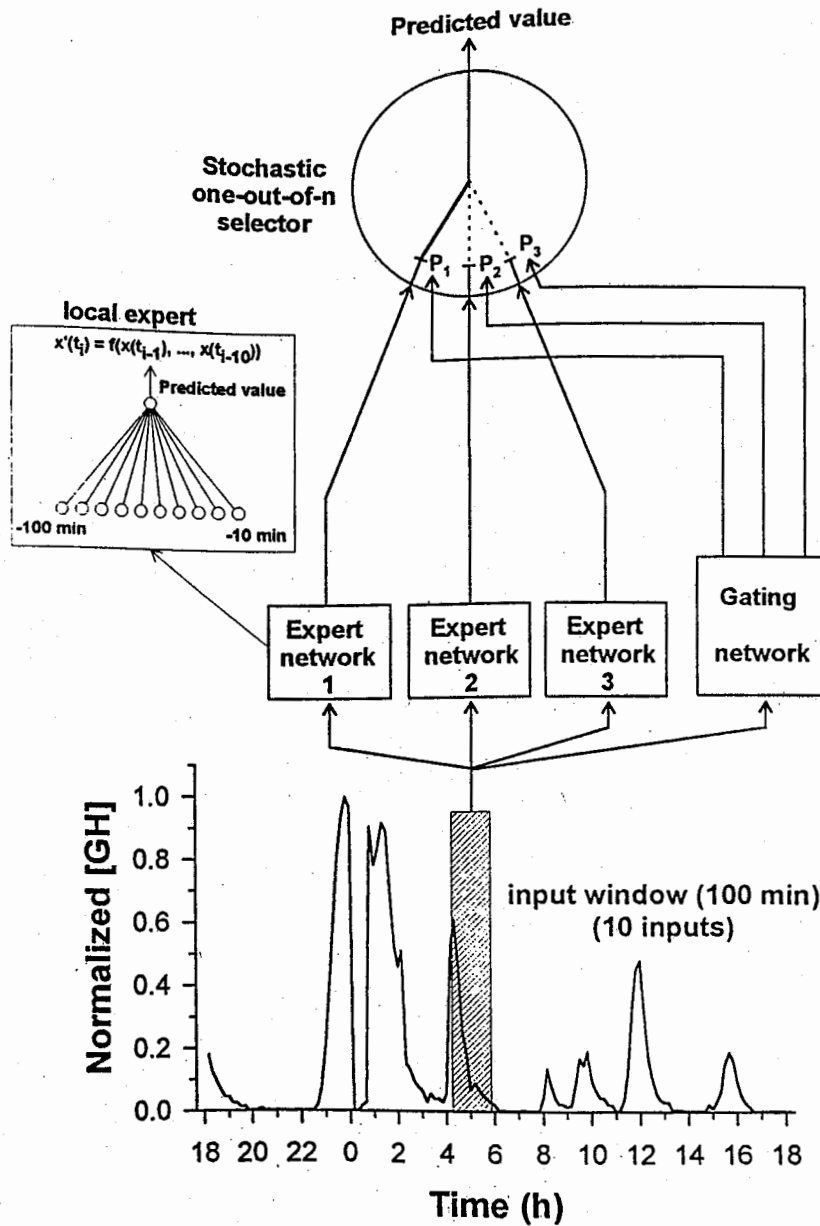


Figure 1. Schematic architecture of adaptive mixtures of local experts used for time series prediction. Each local expert is a feedforward neural network which all have the same architecture and receive the same input. The gating network is also feedforward and receives the same input as the expert networks. It has normalized outputs  $p_j = \exp(x_j) / \sum_i \exp(x_i)$ , where  $x_j$  is the total weighted input received by output unit  $j$  of the gating network. The selector acts like a multiple input, single output stochastic switch. The probability that the switch will select the output from expert  $j$  is  $P_j$ , which is the  $j^{\text{th}}$  output of the gating network. These local experts are trained in parallel to perform time series prediction one time step ahead. During testing the weighted average of the expert outputs (with weights assigned by the gating network) is used as the output of the mixture of experts.

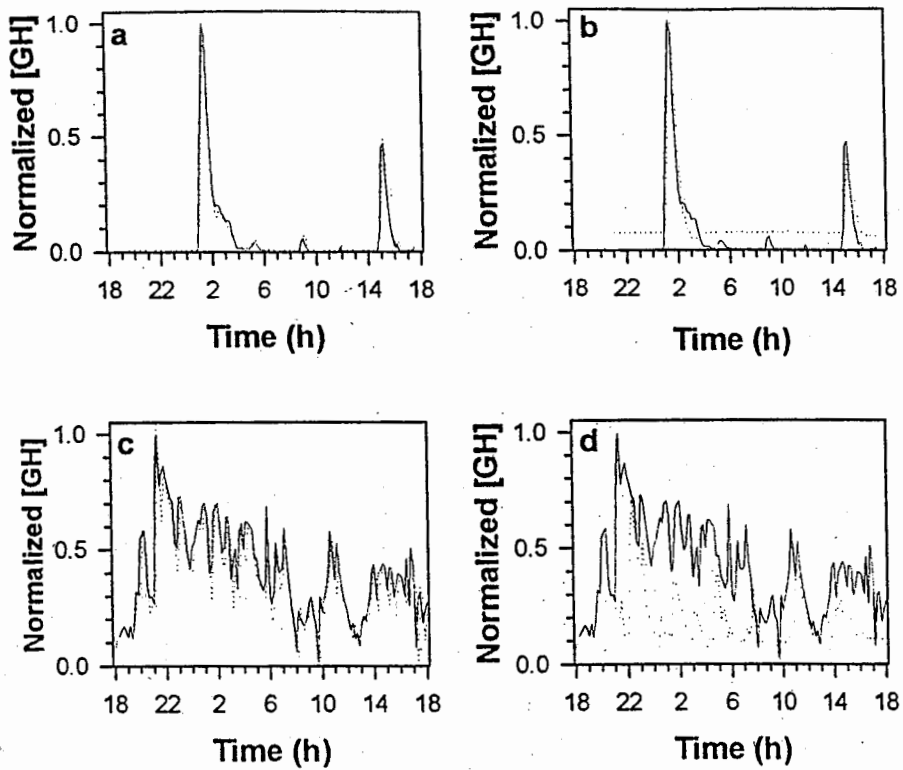


Figure 2. One and ten-step ahead prediction using an adaptive mixture of five local experts with nine input units each (nonlinear gating network with three hidden units). Solid line: original time series of normalized GH serum concentration, dotted line: predicted time series. Healthy control: (a) one step ahead (10 minutes) prediction, (b) ten steps ahead (100 minutes) prediction; acromegalic patient: (c) one step ahead (10 minutes) prediction, (d) ten steps ahead (100 minutes) prediction.

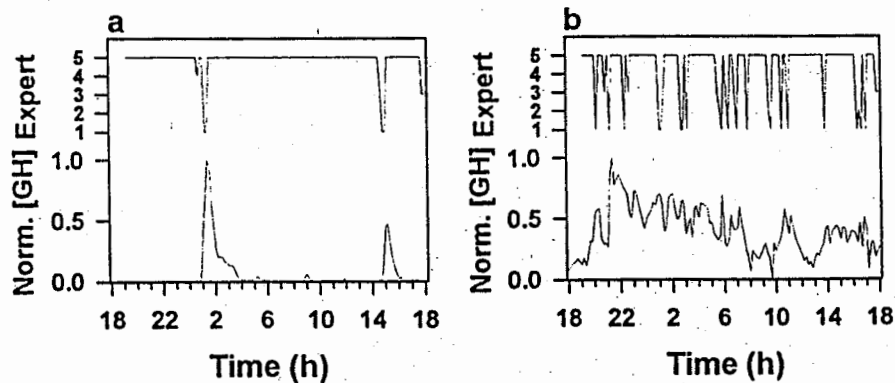


Figure 3. Self-organized segmentation of the pulsatile pattern of secretion (SOPUL). Normalized original GH time series and selection pattern of local experts used for prediction. (a) healthy control (b) acromegalic patient.

