Human Reproduction Update 1997, Vol. 3, No. 3 pp. 215–234

Neural networks in the analysis of episodic growth hormone release

K.Prank[1], M.Kloppstech and G.Brabant

Abteilung Klinische Endokrinologie, Medizinische Hochschule Hannover, D-30623 Hannover, Germany

TABLE OF CONTENTS

Pulsatile secretion of growth hormone (GH) has been observed in healthy controls as well as acromegalic patients. In healthy adults, highly regulated secretory pulses of GH occur 4–8 times within 24 h. This episodic pattern of secretion seems to be related to the optimal induction of physiological effects at the peripheral level. In contrast to normal subjects, acromegalic patients demonstrate an irregular pattern of excessive GH release. This pattern of secretion is responsible for many systemic effects, such as the stimulation of connective tissue growth, cardiovascular and cerebrovascular disease, diabetes mellitus and arthritis. Standard methods for the analysis of pulsatile patterns of hormone secretion did not consistently separate the temporal dynamics of GH release in healthy controls and acromegalic patients under various study conditions. Using the cutting edge technology of artificial neural networks for time series prediction, we were able to achieve significant separation of both groups under various conditions by means of the predictability of their GH secretory dynamics. Improving the predictive results by using a more refined system of multiple neural networks acting in parallel (adaptive mixtures of local experts), we found that this system performed a self-organized segmentation of hormone pulsatility. It separated phases of secretory bursts and quiescence without any prior knowledge of the form of a GH pulse or a model of secretion. Comparing the predictive results for the GH dynamics with those for computer-simulated stochastic processes, we were able to define the irregular pattern of GH secretion in acromegaly as a random autonomous process. The introduction of neural networks to the analysis of dynamic endocrine systems might help to expand the existing analytical approaches beyond counting frequency and amplitude of hormone pulses.

Key words: acromegaly/prediction/pulse detection/ random secretion/time series

Introduction

Pulsatile hormone secretion has been demonstrated as an ubiquitous phenomenon in a large number of different

[1]To whom correspondence should be addressed. Tel: +49 511 532 3827; Fax: +49 511 532 3825

hormonal systems, with pulse frequencies ranging from approximately 6 to 140 pulses in 24 h (Brabant *et al.*, 1992). This temporal pattern of hormone release has important effects on the regulation of target cell function and structure (Schöfl *et al.*, 1994). In perifused rat hepatocytes, glucose production is closely dependent on the frequency of stimulating glucagon pulses (Weigle and Goodner, 1986). Moreover, our group has demonstrated in single hepatocytes that information encoded in the temporal pattern of an extracellular (adrenergic) stimulus is transferred across the plasma membrane and translated into distinct intracellular information encoded in the frequency and amplitude of Ca^{2+} transients (Schöfl *et al.*, 1993). Other well-characterized examples are the frequency-encoded regulation of pituitary hormones by hypothalamic releasing factors. Thyrotrophin (TSH) secretion depends on the frequency of thyrotrophin-releasing hormone (TRH) stimulation, which differentially regulates transcription of the α- and β-subunits of TSH (Haisenleder *et al.*, 1992). A comparable effect is observed for the synthesis and secretion of luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which are under the frequency-dependent control of gonadotrophin-releasing hormone (GnRH) pulses (Knobil, 1980; Shupnik, 1990). This has important consequences for the treatment of patients with hypothalamic failure of GnRH secretion such as Kallman's syndrome, where only an adaptation of the physiological rhythm of pulsatile GnRH stimulation is able to correct the disorder therapeutically (Santoro *et al.*, 1986).

Currently available methods for the analysis of the temporal pattern of pulsatile hormone secretion are based either on a heuristic description of hormone pulses or on a model-based approach. Heuristic methods have been used to define the parameters of a hormonal pulse, such as its shape, duration and rates of rise and fall, relative to assay noise control (Merriam and Wachter, 1982; Veldhuis and Johnson, 1986). Model-based approaches either make assumptions on the distribution volume and half-life of the hormone under study to estimate the secretory activity of the respective gland from the measured time series of serum concentrations (deconvolution techniques: Prank and Brabant, 1994; Veldhuis *et al.*, 1994), or they describe the data set as the sum of different sine and cosine functions (Fourier transformation). Thus, all of these methods rest on certain assumptions of the behaviour of the system and are used as means to objectively confirm results expected from visual inspection of the time series. However, using standard criteria of analysis, we were unable to detect differences in the pattern of pulsatile GH secretion when comparing healthy controls and patients with acromegaly under somatostatin therapy (Prank, 1995). Visually, differences in pulsatile GH secretion could be clearly distinguished between both groups, even though they had comparable means of the 24 h GH serum concentration profile. Conceptually, an alteration in the timing of GH pulses could be anticipated following surgery in these patients with persistent disease. To overcome these problems, methods have been developed recently to analyse a hormonal time series without prior knowledge of the system.

The approximate entropy (ApEn) statistic has been proposed by Pincus (1991) as an adaptive approach for characterizing regularity of a given time series by a single number. The ApEn statistic was able to reveal differences in pulsatility of a simulated hormone concentration time series where current pulse detection algorithms failed (Pincus and Keefe, 1992). This approach has been successfully applied in separating the irregular temporal dynamics of GH-secreting (Hartman *et al.*, 1994; Prank *et al.*, 1995a) and aldosterone-secreting tumours (Siragy *et al.*, 1995) from physiological secretion.

An alternative approach to the analysis of pulsatile patterns of hormone secretion is based on time series prediction. Using this method, one tries to capture regularities in the temporal pattern of complex time series, particularly in the context of nonlinear dynamic systems, and to separate deterministic from random behaviour (Sugihara and May, 1990; Tsonis and Elsner, 1992; Weigend and Gershenfeld, 1993). This technique predicts 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; Lefebvre *et al.*, 1993; Chang *et al.*, 1994; Schiff *et al.*, 1994a,b; Sugihara, 1994) and may be applied for short time series containing a very limited number of data points (Sugihara and May, 1990; Tsonis and Elsner, 1992). In the recent Santa Fe Institute competition on time series prediction, the best predictions for a number of data sets from various fields were achieved by neural network approaches (Weigend and Gershenfeld, 1993). These approaches have the advantage of broad linear and nonlinear approximation abilities without making specific assumptions of the system under study, in contrast to explicit models that use a set of differential equations.

Artificial neural networks

The use of artificial neural networks, commonly referred to as 'neural networks', has been motivated right from its inception by the recognition that the brain computes in an entirely different way from that of a conventional digital computer. The brain is a highly complex, nonlinear, and parallel computer (information-processing system), which has the capability of performing certain computations by structurally and functionally organizing its neurons (e.g. pattern recognition, perception and motor control) many times faster than the fastest digital computer. Just as plasticity appears to be essential to the functioning of neurons as information-processing units in the human brain, neural networks comparably make use of artificial neurons. A neural network is usually implemented using electronic components or simulated in software on a digital computer. To achieve good performance, neural networks employ a large number of interconnections of simple computing cells, which are referred to as 'neurons' or 'processing units'. Thus, a neural network may be regarded as an adaptive machine which is defined as follows: (i) knowledge is acquired by the neural network through a learning process (training); (ii) interneuron connection strengths known as synaptic weights are used to store the knowledge. The learning process, also referred to as a training process or learning algorithm, modifies the weights of the network in a way to attain a desired objective.

Neural networks are also referred to as neurocomputers, connectionist networks and parallel distributed processors. In the remainder of this article we use the term 'neural network' or, occasionally, 'connectionist approach'.

Benefits of neural networks

It is apparent that a neural network derives its computational power through its parallel distributed structure and through its ability to learn from examples and therefore to generalize to new situations. After being trained (learning process) on a number of examples, neural networks have the ability to sensibly interpolate and extrapolate from the examples in the given data sets. The use of neural networks includes the following useful properties and capabilities:

1. Nonlinearity: A neuron as the central processing unit in neural networks is basically a nonlinear device. Thus, the neural network itself is nonlinear since it is built by interconnected neurons. Nonlinearity is a highly important property, particularly if the underlying physical mechanism responsible for the generation of an input signal (the data to be analysed) is inherently nonlinear: This is the case for most endocrine systems regulated dynamically through feedback and feedforward loops and nonlinear dose–response relationships.
2. Input–output mapping: Supervised learning, also referred to as learning with teacher, is a popular paradigm of neural network learning. In supervised learning we have a list or training set of correct input–output pairs as examples. When we apply one of the training inputs to the network we can compare the network output to the correct output. The connections strengths are then changed in such a way as to minimize the difference between the desired and actual output of the network. The training of the network is performed incrementally until the network reaches a steady state, with no further significant changes in the strengths of the connections. This is achieved by making small adjustments in response to each training pair. Thus, the network learns from examples by constructing an input–output mapping for the data to be analysed. After training the network with known input–output pairs (training patterns), input patterns not used for training the network are tested for generalization (testing process). Thus, learning an input–output mapping is equivalent to extracting regularities from training examples which then are applied to new examples.
3. Adaptive behaviour: Neural networks are capable of adapting their connection strengths (synaptic weights) to changes in the surrounding environment. In particular, a neural network trained to operate in a specific environment can be easily retrained to adapt to minor changes in the operating environmental conditions. An important adaptive feature of neural networks is that they can change their connection strengths in real time, which has major consequences for the analysis of nonstationary systems (i.e. a system whose statistics change with time). The dynamics of most biological systems are nonstationary. Thus, the adaptive feature of working in nonstationary environments allows for the use of neural networks in the analysis of temporal patterns of pulsatile hormone secretion. The architecture of a neural network used for pattern classification, signal processing and control applications, combined with its adaptive capability, make it an ideal tool for use in adaptive pattern classification, adaptive signal processing and adaptive control. To realize the full benefits of adaptive behaviour, the principal time constants of the system should be long enough for the system to ignore spurious disturbances and yet short enough to respond to meaningful changes in the environment (Grossberg, 1988).
4. Fault tolerance: A neural network, implemented in hardware form, e.g. in very-large-scale-integrated (VLSI) technology, has the potential to be inherently

fault tolerant. For example, if a neuron or its connecting links are damaged, recall of a stored pattern is impaired in quality. However, due to the distributed nature of information in the network, the damage has to be extensive before the overall response of the network is degraded significantly. Thus, in principle, a neural network exhibits a graceful degradation in performance rather than catastrophic failure.

5. Hardware implementation in VLSI: The parallel structure and function of a neural network allows for the fast computation of certain tasks compared to conventional digital computers. This same feature makes a neural network ideally suited for implementation using VLSI technology. The particular virtue of VLSI is that it provides a means of capturing truly complex behaviour in a highly hierarchical manner (Mead and Conway, 1980). Thus, it becomes possible to use a neural network as a tool for real-time applications involving pattern recognition and control.
6. Analogy to neurobiology: The structure of a neural network is motivated by analogy with the brain, which is a living proof that fault-tolerant parallel processing is not only physically possible but also fast and powerful. Neurobiologists look to (artificial) neural networks as a research tool for the interpretation of neurobiological phenomena. For example, neural networks have been used to provide insight on the development of premotor circuits in the oculomotor system (responsible for eye movements) and the manner in which they process signals (Robinson, 1992).

Information processing units (neurons)

The information processing in a neural network is performed by neurons (information processing units). Three major features of a neuron are listed below and displayed in Figure 1: (i) Neurons are connected by a set of synapses or connecting links to other neurons. These connections are characterized by weights or strengths, i.e. a signal x_j at the input of synapse j that is connected to neuron k is multiplied by the synaptic weight w_{kj}. The weight w_{kj} is positive if the associated synapse is excitatory, it is negative if the synapse is inhibitory. (ii) The weighted inputs are summed up. This operation is a linear combiner. (iii) An activation function, also referred to as squashing function, limits the amplitude of the output of a neuron. Typically, the normalized amplitude range of the output of a neuron is written as the closed unit interval [0,1] or, alternatively, [–1,1].

The model neuron shown in Figure 1 also includes an externally applied threshold θ_k that has the effect of lowering the net input of the activation function. On the other hand, the net input of the activation function may be increased by employing a bias term rather than a threshold. The bias is the negative of the threshold.

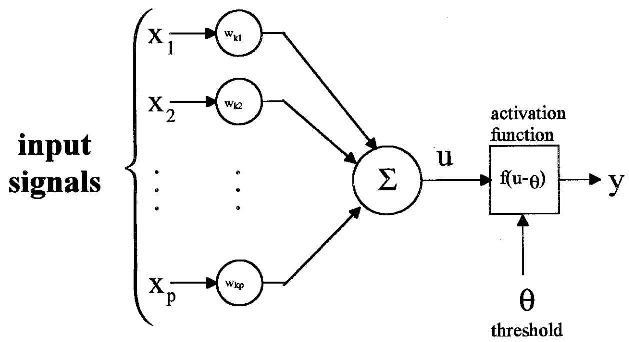

Figure 1. Model of an artificial neuron (information processing unit).

In mathematical terms, we may describe a neuron k by writing the following pair of equations:

$$u_k = \sum_{j=i}^{p} w_{kj} x_j$$

and

$$y_k = \varphi(u_k - \theta_k)$$

where $x_1, x_2, \ldots x_p$ are the input signals; $w_1, w_2, \ldots w_p$ are the synaptic weights of neuron k; u_k is the linear combiner output; θ_k is the threshold; $\varphi(\bullet)$ is the activation function and y_k is the output signal of the neuron.

Types of activation functions

The activation function $\varphi(\bullet)$ defines the output of a neuron in terms of the activity level at its input. We may identify three basic types of activation functions: (i) threshold function (Figure 2a), (ii) piecewise-linear function (Figure 2b) and (iii) sigmoid function. The sigmoid function is by far the most common form of activation function used in the construction of artificial neural networks. It is defined as a strictly increasing function that exhibits smoothness and asymptotic properties. An example of the sigmoid is the logistic function, defined by

$$\varphi(x) = \frac{1}{1 + e^{-ax}}$$

where a is the slope parameter of the sigmoid function. By varying the parameter a, we obtain sigmoid functions of different slopes, as displayed in Figure 2c. In contrast to the threshold function which assumes either the value of 0 or 1, a sigmoid function assumes a continuous range of values between 0 and 1.

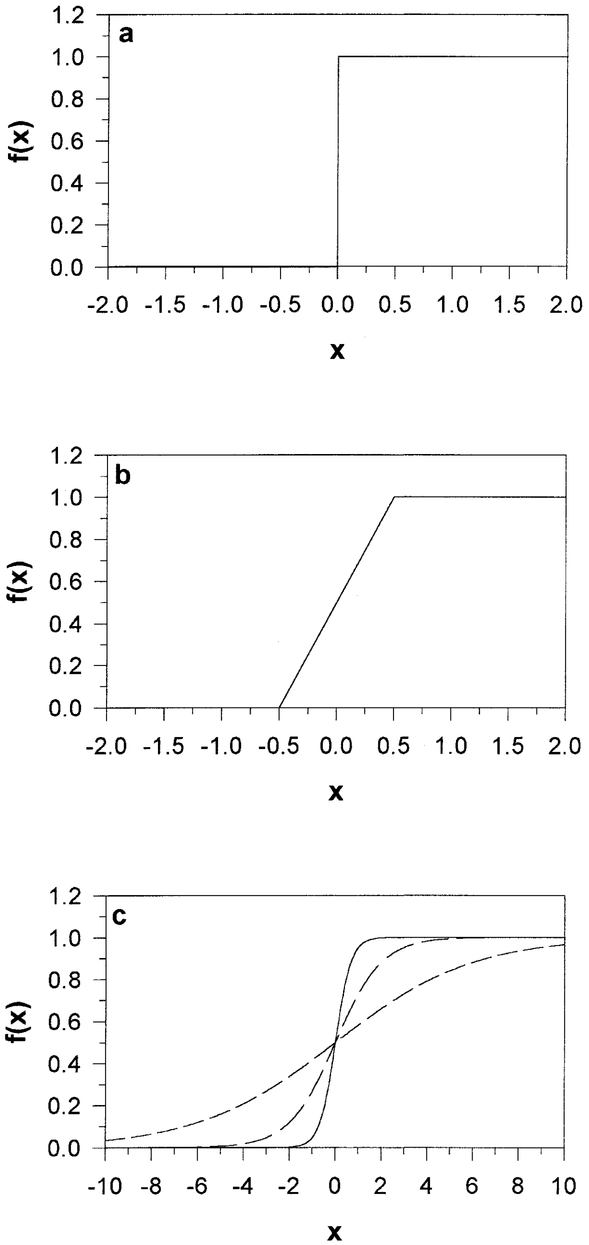

Figure 2. Activation functions: (**a**) threshold, (**b**) piecewise linear, (**c**) sigmoid for three different slopes. See text for more details.

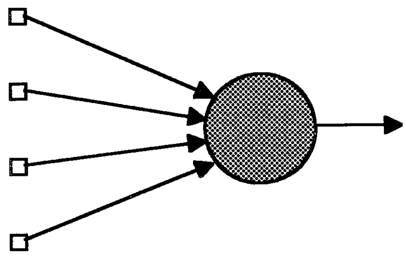

Figure 3. Single-layer feedforward neural network.

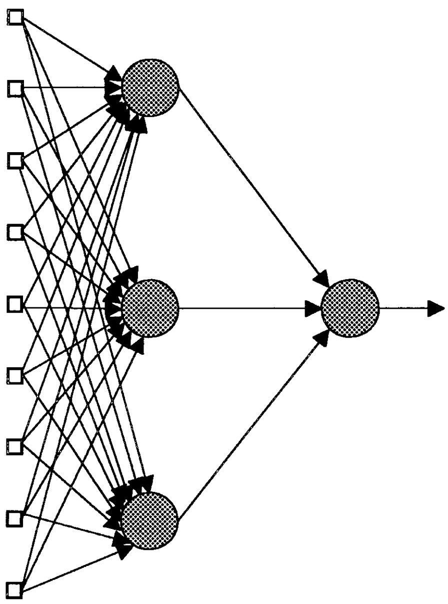

Figure 4. Multi-layer feedforward network with nine input units, three hidden units and one output unit.

Network architectures

Single-layer feedforward networks

A layered neural network is a network of neurons (processing units) organized in the form of layers. In the simplest form of a layered network, we just have an input layer of source nodes that is connected to an output layer of neurons, but not vice versa. Such a neural network is referred to as a single-layer network, since the input layer of source nodes is not counted, because it does not perform any computation (Figure 3). One example for a single-layer neural network is a linear associative memory. Such a network associates an output pattern with an input pattern, and information is stored in the network through modifications of the synaptic weights.

Multilayer feedforward networks

Multilayer feedforward neural networks are characterized by one or more hidden layers, whose computation nodes are correspondingly called hidden neurons or hidden units. The function of the hidden neurons is to intervene between the external input and the network output. By adding one or more hidden layers, the network is enabled to extract higher-order statistics. Thus, the network acquires a global perspective despite its local connectivity by virtue of the extra set of synaptic connections and the extra dimension of neural interactions (Churchland and Sejnowski, 1992). The ability of hidden neurons to extract higher-order

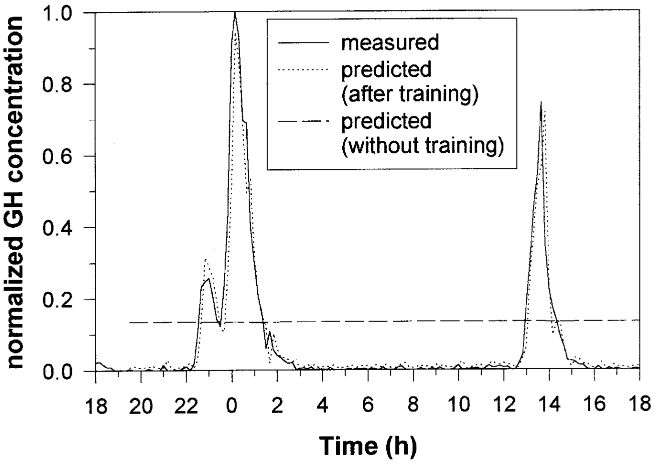

Figure 5. Prediction (10 min ahead) of a 24 h growth hormone (GH) concentration time series by a neural network as displayed in Figure 4. The predicted GH time series are displayed after training the network and without training (random initial weights).

statistics is particularly valuable when the size of the input layer is large.

The input units (neurons) of the input layer supply the second layer (hidden layer) with the input information. The outputs signals of the second layer are used as inputs to the third layer, and so on. Typically, the output signals of the preceding layer are the input signals of the current layer. The set of output signals of the neurons in the output layer (final layer) of the network represents the overall response of the network to the activation pattern supplied by the input units in the input (first) layer. Figure 4 displays the architecture of a multilayer feedforward neural network for the case of a single hidden layer. This network architecture is referred to as a 9–3–1 neural network, since it contains nine input units, three hidden units and one output unit. To illustrate this notation more generally, a feedforward network with i input units, h_1 units in the first hidden layer, h_2 units in the second layer, and o neurons in the output layer is referred to as an i-h_1-h_2-o network. Figure 5 demonstrates the presentation of input information from a time series of hormone concentrations to the network illustrated in Figure 4. Here, a short temporal window of a 24 h GH concentration time series (90 min window corresponds to nine input values, since the data were sampled at 10 min intervals) is presented as input information to the network.

Learning process (training)

A neural network learns about its environment through an iterative process of adjustments applied to its synaptic weights and thresholds. Ideally, the network becomes more knowledgeable about its environment after each iteration of the learning process. Learning in the context of neural networks might be viewed as follows: 'Learning is a process by which the free parameters of a neural network are adapted through a continuous process of stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place'.

Basically, three different classes of learning paradigms exist: supervised learning, reinforcement learning and self-organized (unsupervised) learning. As its name implies, supervised learning is performed under the supervision of an external 'teacher'. Reinforcement learning involves the use of a 'critic' that evolves through a trial-and-error process. Unsupervised learning is performed in a self-organized manner in which no external teacher or critic is required to change the synaptic strengths in the network. Here, we consider only supervised learning, which was used for the neural network analysis reviewed in this article.

Supervised learning

Let $d_k(n)$ be the desired response or target response for neuron k at time n, and $y_k(n)$ be the actual response of this neuron. The response $y_k(n)$ is produced by a stimulus $x(n)$ applied to the input of the network in which neuron k is embedded. The input vector $x(n)$ and the desired response $d_k(n)$ for neuron k constitute a particular training example presented to the network at time n. It is assumed that this example and all other examples presented to the network are generated by an environment that is probabilistic in nature, but the underlying probability distribution is unknown.

Typically, the actual response $y_k(n)$ of neuron k is different from the desired response $d_k(n)$. Hence, we may define an error signal as the difference between the target response $d_k(n)$ and the actual response $y_k(n)$, as shown by

$$e_k(n) = d_k(n) - y_k(n)$$

The ultimate purpose of error-correction learning is to minimize a cost function based on the error signal $e_k(n)$, such that the actual response of each output neuron in the network approaches the target response for that neuron in some statistical sense. Indeed, once a cost function is selected, error-correction learning is strictly an optimization problem. A criterion commonly used for the cost function is the mean-square-error criterion, defined as the mean-square value of the sum of squared errors:

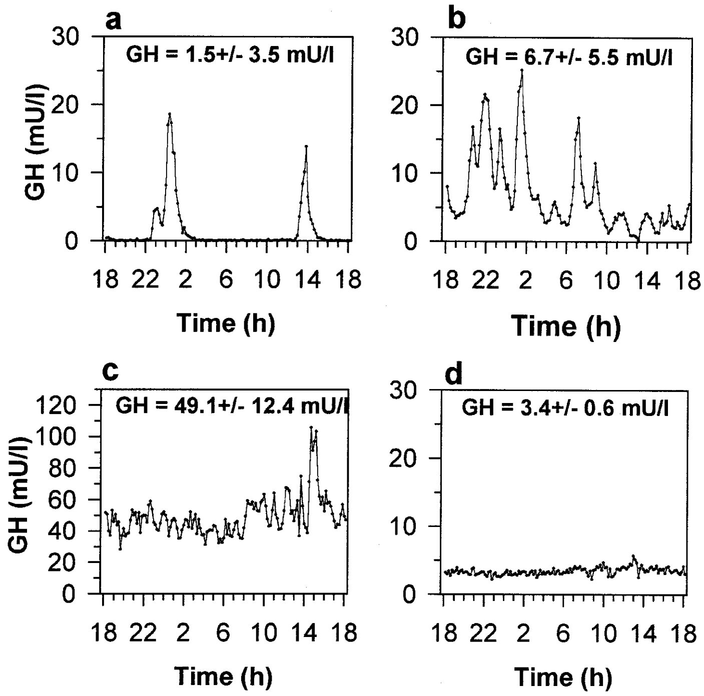

Figure 6. Representative 24 h growth hormone (GH) serum concentration time series (10 min sampling interval) with attached mean 24 h GH serum concentration ± SD. Individual healthy controls cannot be separated from acromegalic patients based on the 24 h mean GH concentrations. Healthy control: (**a**) basal condition and (**b**) after 3 days fasting; acromegalic patient: (**c**) basal condition and (**d**) under octreotide treatment.

$$J = E\left[\frac{1}{2}\sum_k e_k^2(n)\right]$$

where E is the statistical expectation operator, and the summation is over all neurons in the output layer of the network. Minimization of the cost function J with respect to the network parameters leads to the so-called method of gradient descent (Widrow and Stearns, 1985; Haykin, 1991). Using the instantaneous value of the sum of squared errors as the criterion of interest:

$$\zeta(n) = \frac{1}{2}\sum_k e_k^2(n)$$

the network is then optimized by minimizing $\varsigma(n)$ with respect to the synaptic weights of the network. Thus, according to the error-correction learning rule (or delta rule, as it is sometimes called) the adjustment $\Delta w_{kj}(n)$ made to the synaptic weight w_{kj} at time n is given by

$$\Delta w_{kj} = \eta e_k(n) x_j(n)$$

where η is a positive constant that determines the rate of learning (Widrow and Hoff, 1960). That means the adjustment made to a synaptic weight is proportional to the product of the error signal (measured with respect to some desired response at the output of that neuron) and the input signal of the synapse in question. Note that this input signal is the same as the output signal of the presynaptic neuron that feeds the neuron in question. Error-correction learning relies on the error signal $e_k(n)$ to compute the correction $\Delta w_{kj}(n)$ applied to the synaptic weight $w_{kj}(n)$ of neuron k. The new update value of the synaptic weight $w_{kj}(n + 1)$ is computed as follows:

$$w_{kj}(n+1) = w_{kj} + \Delta w_{kj}(n)$$

The choose of the learning rate η has a major impact on the performance of the learning process. If η is small, the learning process proceeds smoothly, but it may take a long time for the neural network to converge to a stable solution. If, on the other hand, η is large, the rate of learning is accelerated, but there exists the possibility that the learning process may diverge and the system therefore becomes unstable.

A plot of the cost function J versus the synaptic weights characterizing the neural network consists of a multidimensional surface referred to as an error-performance surface or simply error surface. The objective of the error-correction learning algorithm is to start from an arbitrary point on the error surface (determined by the initial values assigned to the synaptic weights) and then move toward a global minimum, in a step-by-step fashion. However, this is not always attainable, because it is possible for the algorithm to get trapped at a local minimum of the error surface and therefore never be able to reach a global minimum.

Application of neural networks to clinical medicine

Nonlinear processes and dynamics as they normally occur in biology may not be analysed best by classical linear methods. The application of artificial neural network as nonlinear information processing methods has drawn much attention recently. In a long series of papers in the journal *Lancet* (Baxt, 1995; Cross *et al.*, 1995; Dybowski and Gant, 1995; Baxt and Skora, 1996; Dybowski *et al.*, 1996) it was demonstrated that the use of neural networks allows for the identification of multidimensional nonlinear relationships in clinical data where other forms of analysis fail to improve diagnostic accuracy (Baxt, 1995). The applications range from diagnosis of acute myocardial infarction (Baxt and Skora, 1996) and outcome prediction in critically ill patients (Dybowski *et al.*, 1996) to analysis of biomedical signals such as images [radiographs, positron-emission tomographic (PET) scans, nuclear magnetic resonance (NMR) scans and perfusions scans] and time series of electrocardiogram (ECG) and electroencephalograph (EEG) data (references are listed in Baxt, 1995).

At about the same time, our group demonstrated the superior performance of artificial neural networks in the diagnosis of endocrine diseases compared to classical methods of linear analysis (Prank *et al.*, 1995b, 1996a,b).

Application of neural networks and approximate entropy to the analysis of endocrine time series

We recently applied neural networks to endocrine time series prediction to separate clearly the secretory dynamics of parathyroid hormone (PTH) in a group of osteoporotic patients from those in healthy controls; we were able to demonstrate a significantly higher predictability of the PTH secretory dynamics in healthy subjects than in the osteoporotic group (Prank *et al.*, 1995b).

Here, we focus on the adaptive capabilities of a modular neural network system, referred to as adaptive mixtures of local experts (Jacobs *et al.*, 1991), for the prediction of 24 h time series of GH serum concentration in acromegalic patients and in normal controls. In contrast to classical pulse detection methods, this approach is model and scale independent, and in this respect it is similar to the ApEn regularity statistic. The network system was trained on time series of pooled data from 24 h GH rhythms from a reference group of healthy controls as well as from each subgroup. Its predictive abilities were then tested on an additional group of healthy controls as well as on acromegalic patients. To avoid possible differences in the analytical results simply caused by the higher mean GH pulse amplitude in acromegalic patients, we also compared the physiologically enhanced GH secretion in fasting normals with the pharmacologically reduced secretion in acromegalic patients treated with a somatostatin analogue (octreotide). Using this approach, we found that the neural network system had performed a self-organized quantification of hormone pulsatility (SOPUL) by learning time series prediction without any prior knowledge of the form of a pulse or the model of secretion. It significantly separated the GH secretory dynamics in acromegaly under basal conditions and under octreotide treatment from those in normal controls in basal and fasting state (Prank *et al.*, 1996a).

To investigate the hypothesis of random episodic GH release in acromegaly due to reduced control through the hypothalamus, we analysed time series of GH concentration in normal subjects and in acromegalic patients with respect to their predictability. Comparing the predictive results with those obtained from statistically matched random surrogate data, we found evidence for GH secretion as a random autonomous process in acromegaly. The secretory pattern of GH release was not significantly distinguishable from a variety of stochastic processes (Prank *et al.*, 1996b).

Acromegalic patients and healthy controls

The analytical results presented in this article were obtained from data measured in 10 young lean male controls and six patients (three males and three females) with clinical and biochemical diagnosis of acromegaly that had not been cured by previous surgical and/or radiation therapies. The studies were approved by the local Committee on Medical Ethics, and all participants gave their informed written consent. Blood samples were taken via a central venous catheter every 10 min over 24 h starting at 1800 h. Five healthy controls were studied twice, under basal metabolic conditions and following 3 days of fasting to enhance GH secretion (Riedel *et al.*, 1995), with an interval of 3–6 weeks between studies (Figure 6, normal GH secretion). The 24 h GH secretory profile in the acromegalic group was studied under two different conditions: first, basal condition without any medication, and then 4 weeks later under continuous s.c. pump infusion of 300 μg somatostatin analogue (SMS)/day (Figure 6). On the fourth day of infusion, blood sampling was started over 24 h. The study design and assay characteristics have been described in more detail previously (Riedel *et al.*, 1992, 1995; Prank *et al.*, 1995b).

24 h time series of growth hormone concentrations

The 24 h GH profiles appear visually to differ between healthy controls and acromegalic patients under either study condition (Figure 6): the healthy control is easily separated from the acromegalic patient by a lower mean 24 h GH serum concentration under basal conditions. This is not the case for the comparison between healthy subjects under fasting conditions (Figure 6b) and acromegalic patients under octreotide treatment (Figure 6d).

Pulse detection analysis

The temporal pattern of pulsatile GH secretion was analysed using the heuristic approach of the PULSAR algorithm (Merriam and Wachter, 1982) with the threshold criteria $G(1) = 4.4$, $G(2) = 2.6$, $G(3) = 1.92$, $G(4) = 1.46$, $G(5) = 1.13$, and the CLUSTER program (Veldhuis and Johnson, 1986) with a 2×2 cluster size for peak respectively nadir and a *t*-statistic of 2.32/1.0 for upstroke and downstroke. Additionally, we performed a waveform-independent deconvolution technique (DESADE; Prank and Brabant, 1994) based on a one-compartment model and a single half-life of 26 min (Faira *et al.*, 1989; Buchfelder *et al.*, 1994).

Using these three different classical approaches for the detection of discrete hormone pulses, we could not consistently find significant differences in the number of GH pulses within 24 h between the control and patient groups under either study condition (Table I, part A). The mean GH pulse amplitude was significantly higher in healthy controls under fasting conditions than in acromegalic patients under SMS treatment when all three pulse detection algorithms were applied, whereas the GH pulse amplitude was comparable in healthy controls under basal conditions and in acromegalic patients under octreotide treatment (Table I, part B). Only the deconvolution technique DESADE revealed a significantly higher pulse amplitude in acromegalic patients compared to healthy controls under basal conditions.

Table I. Classical pulse detection analysis using the heuristic methods CLUSTER and PULSAR and the model-based-approach DESADE

Group 1 versus group 2	CLUSTER	PULSAR	DESADE
A. Mean number ± SD of GH pulses in 24 h			
Acromegaly, basal versus	13.0 ± 5.7	14.5 ± 8.3	11.0 ± 1.7
control, basal	10.8 ± 4.2	9.0 ± 2.8	9.6 ± 1.5
	NS	NS	NS
Acromegaly, basal versus	13.0 ± 5.7	14.5 ± 8.3	11.0 ± 1.7
control, fasting	13.0 ± 1.9	14.8 ± 3.7	10.8 ± 4.6
	NS	NS	NS
Acromegaly, SMS versus	14.8 ± 3.5	5.5 ± 2.9	8.5 ± 2.9
control, basal	10.8 ± 4.2	9.0 ± 2.8	9.6 ± 1.5
	NS	NS	NS
Acromegaly, SMS versus	14.8 ± 3.5	5.5 ± 2.9	8.5 ± 2.9
control, fasting	13.0 ± 1.9	14.8 ± 3.7	10.8 ± 4.6
	NS	$P = 0.002$	NS
B. Mean ± SD GH pulse amplitude (mU/l)			
Acromegaly, basal versus	18.7 ± 21.5	15.8 ± 13.4	26.0 ± 19.5
control, basal	2.3 ± 1.5	2.5 ± 1.3	4.0 ± 1.9
	NS	NS	$P = 0.04$
Acromegaly, basal versus	18.7 ± 21.5	15.8 ± 13.4	26.0 ± 19.5
control, fasting	5.4 ± 1.5	5.1 ± 1.0	9.6 ± 3.3
	NS	NS	NS
Acromegaly, SMS versus	2.1 ± 1.6	2.8 ± 2.1	4.2 ± 3.7
control, basal	2.3 ± 1.5	2.5 ± 1.3	4.0 ± 1.9
	NS	NS	NS
Acromegaly, SMS versus	2.1 ± 1.6	2.8 ± 2.1	4.2 ± 3.7
control, fasting	5.4 ± 1.5	5.1 ± 1.0	9.6 ± 3.3
	$P = 0.006$	$P = 0.046$	$P = 0.029$

SMS = somatostatin.

Time series prediction using artificial neural networks

Measured time series are the basis for the characterization of an observed system and for predicting its future behaviour. A number of new approaches such as state–space reconstruction and neural networks promise insights that traditional approaches to these very old problems cannot provide. The desire to predict the future of a time series and understand the past drives the search for laws that explain the behaviour of observed phenomena.

The general availability of powerful computers (starting around 1980) permitted the usage of more complex algorithms for the analysis of time series. One development was the emergence of the field of machine learning, typified by neural networks, that can adaptively explore a large space of potential models. With the shift in artificial intelligence from rule-based methods towards data-driven methods, the field was ready to apply itself to time series.

In an approach to predict the future behaviour of a system under study, the next value $x(t_i)$ of the time series generated by this system is predicted from m previous values using an appropriate mathematical model f:

$$x(t_i) = f[x(t_{i-1}), x(t_{i-2}), ...x(t_{i-m})] + \epsilon_i$$

where $x(t_{i-1})$ is the preceding value, and ε_i correspond to noise or fitting error. In our study, we used feedforward neural networks where the model f represents network architectures with linear (equivalent to autoregressive models $\{f[x(t_{-1}), ... x(t_{i-m})] = \Sigma\ \alpha_k x(t_{i-k})\}$) or nonlinear (sigmoidal) activation functions. Using such a model, overfitting can be controlled by applying regularization functions and cross-validation. Overfitting relates to the problem of fitting the neural network model to noise in addition to the signal. These network approaches have been shown previously to be effective predictors for other real-world short, noisy time series (Lapedes and Farber, 1987; Weigend *et al.*, 1990; Nowlan and Hinton, 1992; Weigend and Gershenfeld, 1993; Prank *et al.*, 1995a,b, 1996a,b).

The network models were trained (fitting the model f to the data by minimizing some error of misfit) to predict one time step (10 min) ahead using data from the past. The value predicted one time step into the future:

$$\hat{x}(t_i) = f[x(t_{i-1}), x(t_{i-2}), ...x(t_{i-m})] + \epsilon_i$$

was iterated back to the input of the network to predict two steps ahead:

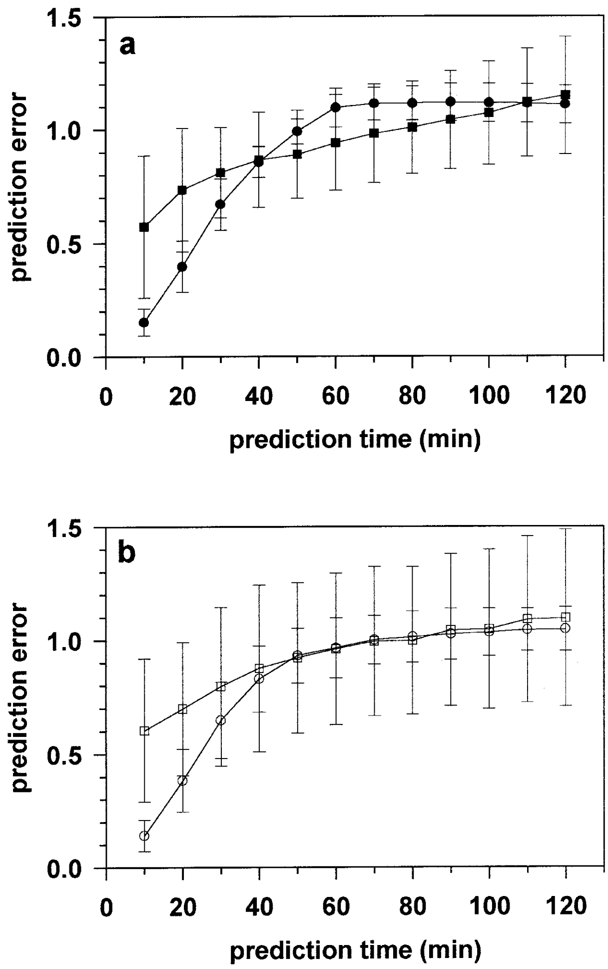

Figure 7. Prediction error average relative variance (arv) as a function of the prediction time. A single feedforward neural network was trained for each subgroup on pooled data using the leave-one-out technique. The testing arv is displayed as mean ± SD for (**a**) healthy controls in the basal state (l) versus acromegalic patients in the basal state (n), and (**b**) healthy controls in a fasting state (¡) versus acromegalic patients under octreotide treatment (o).

$$\hat{x}(t_i) = f\left[\hat{x}(t_i), x(t_{i-1}), ...x(t_{i-m+1})\right] + \epsilon_i$$

This procedure can be repeated to predict any given number of time steps into the future. To decorrelate the prediction error estimate from the variance of the respective time series we used the average relative variance (arv):

$$arv = \frac{[(x_i - \hat{x}_1)^2]}{\sigma^2(x_i)}$$

where the angle brackets denote the mean over all predictions and $\sigma^2(x_i)$ denotes the variance of the measured variable. This measure was used to evaluate the predictive performance of one- and multiple-step ahead predictions during training and testing (Figure 7).

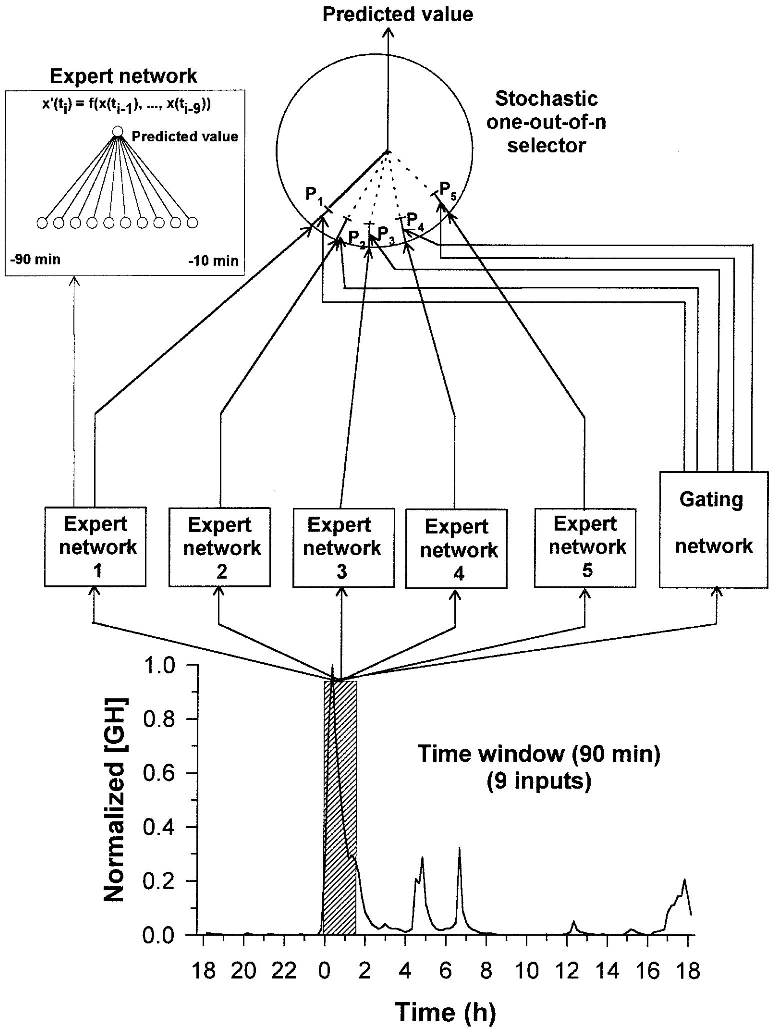

Figure 8. Schematic architecture of adaptive mixtures of local experts used for time series prediction. Each local expert is a feedforward neural network which all have the same architecture (nine input units, one output unit, and linear activation functions) and receive the same input. The gating network 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 jth output of the gating network. In this example, the input pattern consists of 10 normalized ([0,1]) values of a GH serum concentration time series (90 min input window) which are used to predict the next value. These local experts are trained in parallel to perform time series prediction one time step ahead using pooled data from a reference group of (n = 5) healthy controls. During training the experts organize themselves in a way such that each expert specializes to predict best distinct parts of the time series. The trained network system is then tested on additional untrained GH time series from healthy controls (n = 5) and acromegalic patients (n = 6). 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.

Table II. Architecture of the best performing neural network predictor. The networks were trained within each subgroup using a 'leave-one-out technique'

Group	Input units	Hidden units	Output units	Activation function
Acromegaly, basal (n = 6)	5	8	1	sigmoidal
Acromegaly, SMS (n = 6)	7	5	1	sigmoidal
Control, basal (n = 10)	7	10	1	sigmoidal
Control, fasting (n = 5)	7	0	1	sigmoidal

SMS = somatostatin.

Mixture of experts

Divide-and-conquer strategies are effective methods for solving complex problems by dividing them into simpler problems whose solutions can be combined to yield a solution to the complex problem (Nowlan, 1990; Jacobs *et al.*, 1991; Jordan and Jacobs, 1992, 1994; Cacciatore and Nowlan, 1994; Pawelzik *et al.*, 1996). The 'mixture of experts' architecture (Nowlan, 1990; Jacobs *et al.*, 1991) was proposed as an approach to learning a task decomposition in parallel in a neural network context. This architecture 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 (Figure 8). 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 decompose a problem effectively to yield higher classification performance than a single network (Nowlan, 1991). The mixtures of experts also generalizes better from limited amounts of training data.

Learning and generalization of single feedforward neural networks

Prior to training, each individual GH serum concentration time series was normalized to the interval [0,1] to exclude any bias caused by different mean GH pulse amplitudes in the acromegalic and control groups and to account for nonlinearities in the sigmoidal activation functions. First, we evaluated the predictive performance of a large variety of single feedforward neural networks with linear as well as nonlinear (sigmoidal) activation functions. The network size ranged from 6 to 24 input units, corresponding to a time window of between 1 and 4 h, 0 to 12 hidden units, and 1 output unit. In our computer experiments we used a regularization technique (Nowlan and Hinton, 1992) to control overfitting during training (Prank *et al.*, 1995b). The weights were updated by a conjugate gradient descent method. For linear activation functions, this method is formally equivalent to a least-squares linear fit. The different network architectures were trained to predict the next GH concentration value of a time series from its past m values, where m is the number of input units, using pooled GH time series data from the control and acromegalic groups under either condition (basal and fasting or SMS treatment) applying a 'leave-one-out technique' or jacknife procedure, where the time series being tested for prediction is left out of the training procedure. The predictive performance of the architectures was evaluated on validation time series which had not been used for training previously. The best predicting neural network architecture (Table II) within each subgroup was then selected for statistical evaluation (Table III). The simulations of single predictive neural networks were performed on a Sun SPARCstation 20 using customized C-code.

Table II displays the architecture of the best predictive single feedforward neural network for each group under either study condition. In most experiments, nonlinear networks with hidden units and sigmoidal activation functions demonstrated higher short-term predictability than linear networks without hidden units. Using the best predictor for each subgroup, we found that the 24 h GH serum concentration profile was significantly more predictable in healthy controls than in acromegalic patients (Figure 7 and Table III).

Table III. Predictive performance of the best single neural network predictor as listed in Table II and best adaptive mixture of local experts predictor (Figure 9). The average relative variance (arv) denotes the prediction error one time step ahead

Group 1 versus group 2	Single network	Mixture of experts
Acromegaly, basal versus	0.51 ± 0.29	0.24 ± 0.13
control, basal	0.14 ± 0.07	0.07 ± 0.03
	$P = 0.002$	$P = 0.01$
Acromegaly, basal versus	0.51 ± 0.29	0.24 ± 0.13
control, fasting	0.12 ± 0.04	0.09 ± 0.06
	$P = 0.02$	$P = 0.02$
Acromegaly, SMS versus	0.54 ± 0.22	0.21 ± 0.11
control, basal	0.14 ± 0.07	0.07 ± 0.03
	$P = 0.0001$	$P = 0.01$
Acromegaly, SMS versus	0.54 ± 0.22	0.21 ± 0.11
control, fasting	0.12 ± 0.04	0.09 ± 0.06
	$P = 0.003$	$P = 0.03$

Learning and generalization of adaptive mixtures of local experts

The one-step ahead prediction error (arv) was evaluated in a variety of experiments for each group using adaptive mixtures of local experts. In our simulations we varied the

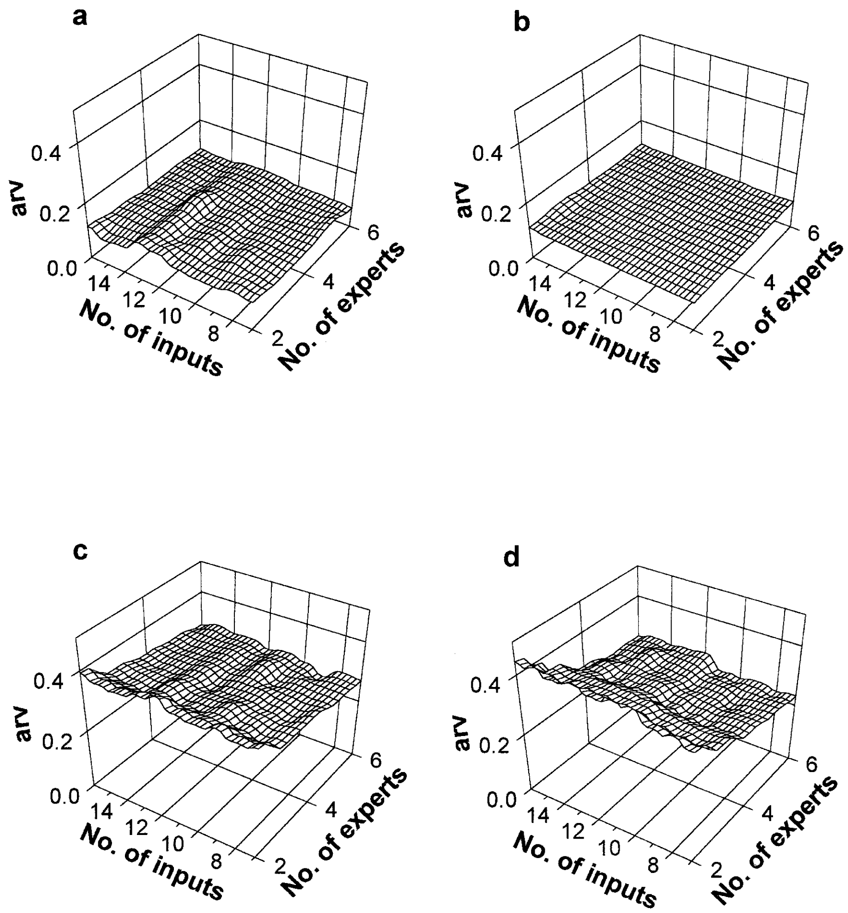

Figure 9. One step-ahead prediction error average relative variance (arv) as a function of the number of input units and number of experts. Each system was trained and tested to perform a one-step ahead prediction using a leave-one-out technique where all but one time series were used for training. Finally, the validation time series was used for testing the predictive performance. (**a**) Healthy controls, basal state; (**b**) healthy controls, fasting state; (**c**) acromegalic patients, basal state; (**d**) acromegalic patients, octreotide treatment.

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. The size of the gating network ranged from 7 to 16 input units, 0 to 5 hidden units, and we used linear as well as nonlinear (sigmoidal) activation functions. The best predictive mixture of experts system was selected for each group based on its one-step prediction testing error (Figure 9). The respective system was then used to perform some of the further analysis. The performance of each best predictive mixture of experts system was compared to that of the corresponding single neural network predictor (Table IV). Additionally, we trained the best predictive mixture of experts system using pooled reference data from healthy controls ($n = 5$). This system consisted of five linear experts with nine 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 (Figure 10). The number of the respective expert selected to predict a given value of the time series was recorded versus time (Figure 11). Finally, a frequency distribution of the expert selection was computed for each subgroup (Table V). Adaptive mixtures of local experts were simulated with customized C-code on a Sun SPARCstation 20.

Extending the single neural network approach to a modular connectionist system (Figure 8), healthy controls were easily separated from acromegalic patients by means of different compositions of adaptive mixtures of local experts evaluated to predict best in the respective group (Figure 9). The best predictive performance of the 24 h GH serum concentration time series in healthy controls was obtained using an adaptive mixture of five local experts (linear activation functions, no hidden units) with nine input units each (90 min input window) and a gating network consisting of nine input units and three hidden units with sigmoidal activation functions (Figure 9a), whereas the 24 h GH profile in acromegalic patients was best predicted using an adaptive

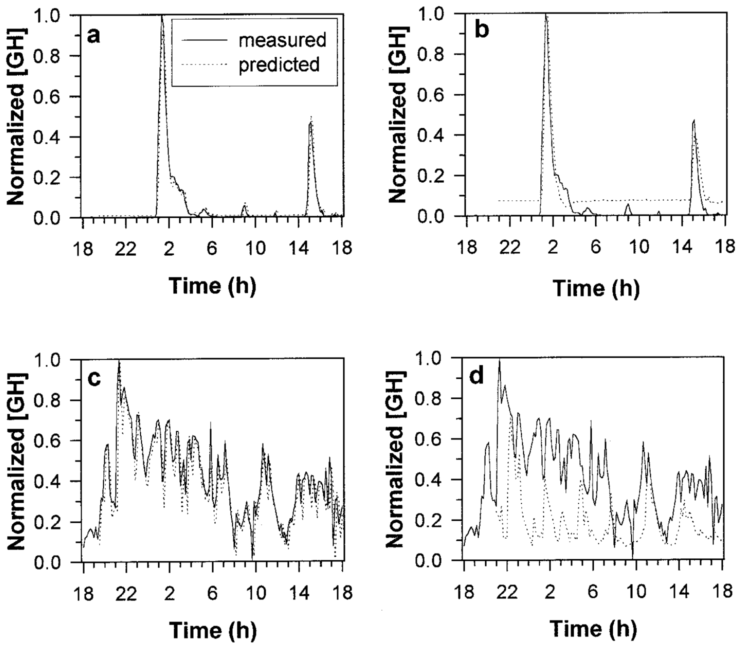

Figure 10. One- and 10-step ahead predictions using an adaptive mixture of five local experts with nine input units each (nonlinear gating network with three hidden units; see Figure 8). Solid line = original time series of normalized growth hormone (GH) serum concentration; dotted line = predicted time series. Healthy control: (**a**) one step ahead (10 min) prediction, (**b**) 10 steps ahead (100 min) prediction; acromegalic patient: (**c**) one step ahead (10 min) prediction, (**d**) 10 steps ahead (100 min) prediction.

mixture of six local experts (linear activation functions, no hidden units) with nine input units each (90 min input window) and a gating network consisting of nine input units and four hidden units with sigmoidal activation functions (Figure 9c). The one-step ahead prediction error (arv) was found to be significantly lower in healthy controls under basal (Figure 9a) and fasting (Figure 9b) conditions than in acromegalic patients under basal conditions (Figure 9c) and SMS treatment Figure 9d) for all parameter values used in our experiments (Table III). The pulsatile features of the 24 h GH profile in healthy controls were well preserved predicting even 100 min (10 steps) into the future (Figure 10b). In contrast, the acromegalic patients demonstrated huge differences between the measured and the predicted GH baseline concentration as well as a large temporal delay between the measured and the predicted GH serum concentration time series (Figure 10d). The predictive performance of this modular approach was significantly superior compared to a single neural network predictor for each of the subgroups (Table IV) and significantly separated these groups by means of the prediction error (Table III).

Self-organized quantification of pulsatility

Analysing the temporal pattern of the selected local expert for each predicted GH value, we found that the adaptive mixtures of local experts had performed a self-organized quantification of pulsatility (SOPUL) in healthy controls and acromegalic patients that was reflected in a switching between distinct expert networks (Figure 11). The selection pattern of the local experts easily separated acromegalic patients under both study conditions from healthy controls in basal state and after fasting (Figure 11). During the training phase, each expert network specialized to predict certain parts of a GH pulse. Expert 1 focused 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 2–4 contributed significantly to the predictive performance of this modular approach, they were selected in only ~1% of the cases in healthy controls and in ~5–8% of the cases in the patient group. The frequency distributions of these most probable experts were significantly different in acromegalic patients from those in healthy controls (Table V). Expert 1 was chosen 2.5–4.4% in normal controls under basal and fasting conditions compared to 12.7–19.7% in the acromegalic group under basal conditions and SMS treatment, which separated the controls from acromegalic patients with high confidence.

Table IV. One-step prediction error (average relative variance, arv) for the best predictive single neural network compared to the best adaptive mixtures of experts (Figure 9). The arv is given as mean ± SD

Group	arv (single network)	arv (mixture of experts)	Single network versus mixture of experts
Acromegaly, basal	0.51 ± 0.29	0.24 ± 0.13	$P = 0.01$
Acromegaly, SMS	0.54 ± 0.22	0.21 ± 0.11	$P = 0.005$
Control, basal	0.14 ± 0.07	0.07 ± 0.03	$P = 0.006$
Control, fasting	0.12 ± 0.04	0.09 ± 0.06	$P = 0.04$

Table V. Frequency distribution of the two most probable local experts 1 and 5 and approximate entropy (ApEn) statistic ($m = 1$, $r = 0.20$, SD, $n = 144$)

Group 1 versus group 2	Expert 1 (%)	Expert 5 (%)	ApEn
Acromegaly, basal versus	12.7 ± 6.7	82.2 ± 9.7	1.36 ± 0.29
control, basal	2.5 ± 1.6	96.0 ± 2.1	0.35 ± 0.11
	$P = 0.009$	$P = 0.01$	$P = 0.001$
Acromegaly, basal versus	12.7 ± 6.7	82.2 ± 9.7	1.36 ± 0.29
control, fasting	4.4 ± 1.2	94.4 ± 1.8	0.85 ± 0.23
	$P = 0.02$	$P = 0.02$	$P = 0.01$
Acromegaly, SMS versus	19.7 ± 9.6	73.0 ± 12.6	1.38 ± 0.35
control, basal	2.5 ± 1.6	96.0 ± 2.1	0.35 ± 0.11
	$P = 0.003$	$P = 0.003$	$P = 0.001$
Acromegaly, SMS versus	19.7 ± 9.6	73.0 ± 12.6	1.38 ± 0.35
control, fasting	4.4 ± 1.2	94.4 ± 1.8	0.85 ± 0.23
	$P = 0.007$	$P = 0.005$	$P = 0.01$

SMS = somatostatin analogue.

The approximate entropy algorithm as a regularity measure

The approximate entropy (ApEn) statistic is as a complementary adaptive approach for the quantification of the temporal pattern of hormone pulses (Pincus and Keefe, 1992; Hartman *et al.*, 1994; Prank *et al.*, 1995a; Siragy *et al.*, 1995). The ApEn statistic measures the logarithmic likelihood that parts of a time series that are similar for m observations remain close on the next incremental comparisons. We choose a relative filter value r as 20% of the respective SD of each individual GH time series, as proposed by Hartman *et al.* (1994) and Siragy *et al.* (1995), to decorrelate the ApEn value from the SD of the time series under study. For the purpose of greater reproducibility, the window size for calculating ApEn was chosen as $m = 1$ due to the limitation of $N = 144$ data points in each GH time series, comparable to previously published data (Siragy *et al.*, 1995). The ApEn was calculated for each group under both conditions and statistically compared between groups (Table V). We found significantly lower values for the ApEn measure (Table V) in healthy controls under basal conditions and under fasting conditions than in acromegalic patients under either study condition, indicating greater irregularity in acromegalic GH secretion.

Random release of growth hormone in acromegaly

In contrast to normal subjects, acromegalic patients demonstrate an irregular pattern of fluctuations of GH serum concentrations (Figure 12; GH secretion in acromegaly) over 24 h (Riedel *et al.*, 1992; Hartman *et al.*, 1994). This pattern of high concentrations of GH is responsible for many of the systemic effects seen in acromegalic patients, such as the stimulation of connective tissue growth, cardiovascular and cerebrovascular disease, diabetes mellitus and arthritis (Daughaday, 1995).

Higher values for the irregularity of GH release in acromegalic patients compared to healthy controls as revealed by the ApEn measure led to the hypothesis that this irregular pattern of GH concentration fluctuations is caused by random, pulsatile GH secretion from the pituitary due to decreased control by the normal orderly secretion of growth hormone-releasing hormone (GHRH) and growth hormone-inhibiting hormone (somatostatin) from the hypothalamus (Hartman *et al.*, 1994).

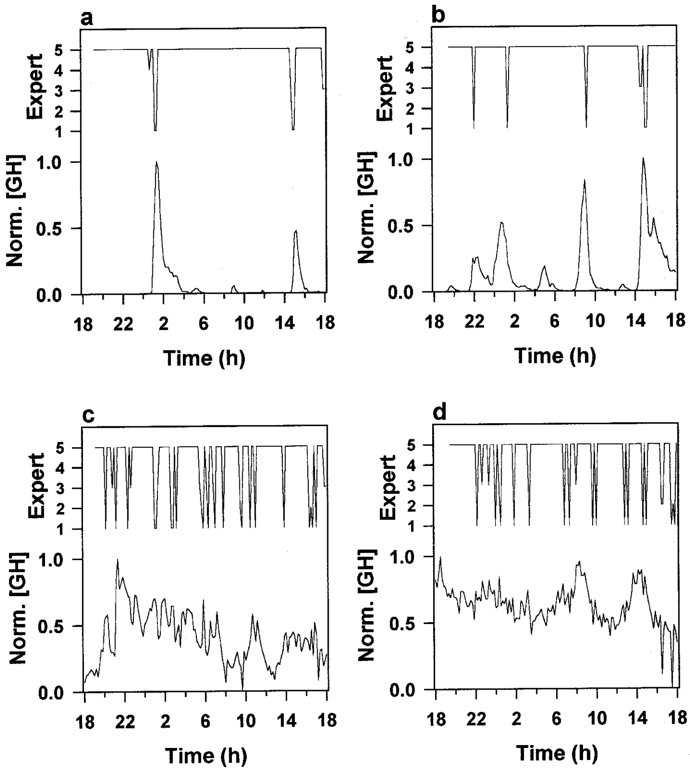

Figure 11. Self-organized quantification of pulsatility (SOPUL). Normalized original growth hormone (GH) time series and selection pattern of local experts used for prediction. Healthy control: (**a**) basal state, (**b**) fasting state; acromegalic patient: (**c**) basal state, (**d**) octreotide treatment.

Table VI. Prediction length of 24 h growth hormone (GH) time series versus random surrogate data[a]

Group	Original	Phase randomized	Gaussian scaled	Fourier shuffled
Healthy subjects	30 ± 7	20 ± 0	10 ± 0	16 ± 5
(*n* = 10)		*P* < 0.05	*P* < 0.005	*P* < 0.01
Patients	20 ± 9	17 ± 5	15 ± 5	13 ± 5
(*n* = 6)		*P* = 0.175	*P* = 0.076	*P* = 0.102

[a]Ten realizations of three types of random surrogate data (as shown in Figure 13) were constructed for each original time series of GH blood concentration. One- and multiple-step-ahead predictions were performed for each individual 24 h GH time series as well as for 10 surrogate data sets of each type, using the same network architecture trained with the 'leave-one-out method'. The prediction length is given in minutes (mean ± SD) when the prediction error average relative variance (arv) exceeds a value of 0.5. This prediction criterion is a measure of the predictability of the system dynamics by the neural network. A prediction error arv of 1.0 may be achieved simply by always guessing the mean of the time series under study as the predicted value. A predictor that meets our criterion for the prediction length is on average twice as good as a predictor that completely ignores the dynamics of time series. Probability values are based on a two-sided paired *t*-test. A significant difference between the predictability of the original GH concentration time series and the corresponding surrogate data sets is assumed for values of *P* < 0.05.

To separate random from deterministic behaviour, the technique of time series prediction has been effectively used in a variety of biological time series containing only a very limited number of data points. In these studies, the degree to which a time series can be predicted is itself a measure of its deterministic structure. A large variety of neural network architectures was explored to determine the best predictor within the group of healthy subjects and the group of patients with excessive GH secretion.

Linear correlation can give a time series, even a stochastic one, a degree of predictability that mimics determinism when making use of linear or nonlinear prediction algorithms. To rule out a mimicking of determinism for stochastic systems, we made extensive use of surrogate random data sets (Chang *et al.*, 1994; Sugihara, 1994) as statistical controls for the test for determinism. Surrogate data refer to data sets that preserve certain linear statistical properties of the experimental data.

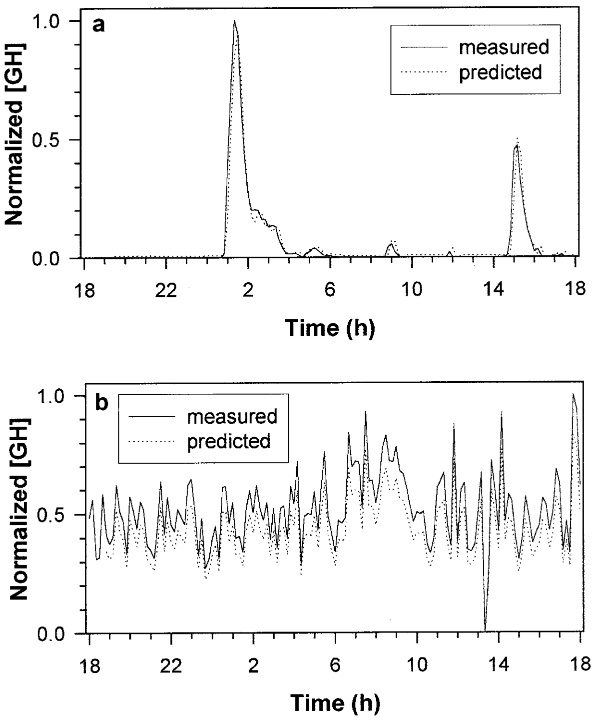

Figure 12. Original and 10-min-ahead-predicted 24 h time series of growth hormone (GH) concentrations. (**a**) Healthy subject; (**b**) patient with acromegaly. Blood samples were taken at 10 min intervals. Acromegalic patients were without any medication for 4 weeks and had not been cured by previous surgical and/or radiation therapies.

Such data serve to test statistically whether our experimental results could be accounted for by a linear stochastic process.

In our study, we made use of three different types of surrogate data which all have their specific merits: phase-randomized, Gaussian-scaled and Fourier-shuffled surrogates. It has been demonstrated that using only phase-randomized surrogates one can produce spurious identifications of deterministic structure (Rapp *et al.*, 1994). Phase-randomized surrogate data have a power spectrum, mean variance and autocorrelation identical to the original experimental time series, but a different amplitude distribution (Figure 13a and d). This sort of surrogate data is used to test the null hypothesis that the results can be explained by using data from a Gaussian linear stochastic process. Other types of surrogate data are necessary since this procedure does not preserve the histogram of the signal amplitudes. The Gaussian-scaled surrogates are used to test the null hypothesis that the results can be explained by data from a nonlinear static transformation of a linear stochastic process (Figure 13b and e). The Fourier-shuffled surrogates in some cases more closely approximate the power spectrum than the Gaussian-scaled surrogates while preserving the amplitude distribution (Figure 13c and f).

Using time series prediction in conjunction with surrogate data as statistical controls, we found no evidence of deterministic temporal structure in the 24 h time series of GH concentration in the group of patients (Table VI). In contrast, the 24 h GH time series in healthy subjects were predictable significantly further into the future compared to their corresponding surrogate data sets, a sign of highly regulated deterministic secretion (Table VI). Our findings support the suggestion of random secretion of GH in acromegaly. Whether the regulatory mechanisms controlling GH secretion operate stochastically or deterministically in the normal and diseased situation must be a fundamental feature of any theoretical model that can explain how this hormonal subsystem processes information. Furthermore, a realistic model for the regulation of GH secretion has to describe the conditions under which such hormonal control mechanisms might switch between stochastic and deterministic behaviour.

Conclusions

In the current study, a visual inspection of the 24 h GH serum concentration time series revealed differences between the healthy subjects and acromegalic patients under basal conditions and after fasting or octreotide treatment. The secretory pattern in healthy controls could be characterized by more regular GH pulses, in contrast to an irregular pattern in acromegalic patients. These visual differences were not detected by a variety of well-established pulse detection algorithms used according to standard criteria. In contrast to these techniques, the ApEn statistic clearly separated healthy controls from patients bearing a GH secreting tumour, regardless of whether they were studied under basal conditions, following fasting or during treatment with octreotide. ApEn as a scale- and model-independent regularity measure revealed higher values for pulsatile GH secretion in acromegalic patients, suggesting a higher irregularity of the 24 h GH profiles. Our data confirmed and expanded previous results demonstrating that the ApEn statistic is particularly useful for distinguishing the disturbed secretory dynamics of GH in patients with mild acromegaly (Hartman *et al.*, 1994), e.g. those in early stages of the disease and those with residual or recurrent disease after radiation and/or surgery, when the mean 24 h GH serum concentration in diseased states is not significantly different from the normal range.

It has been postulated that a higher irregularity of pulsatile GH secretion in acromegalic patients might be an indication of reduced hypothalamic control of regular GH secretion (Hartman *et al.*, 1994). Our results on the

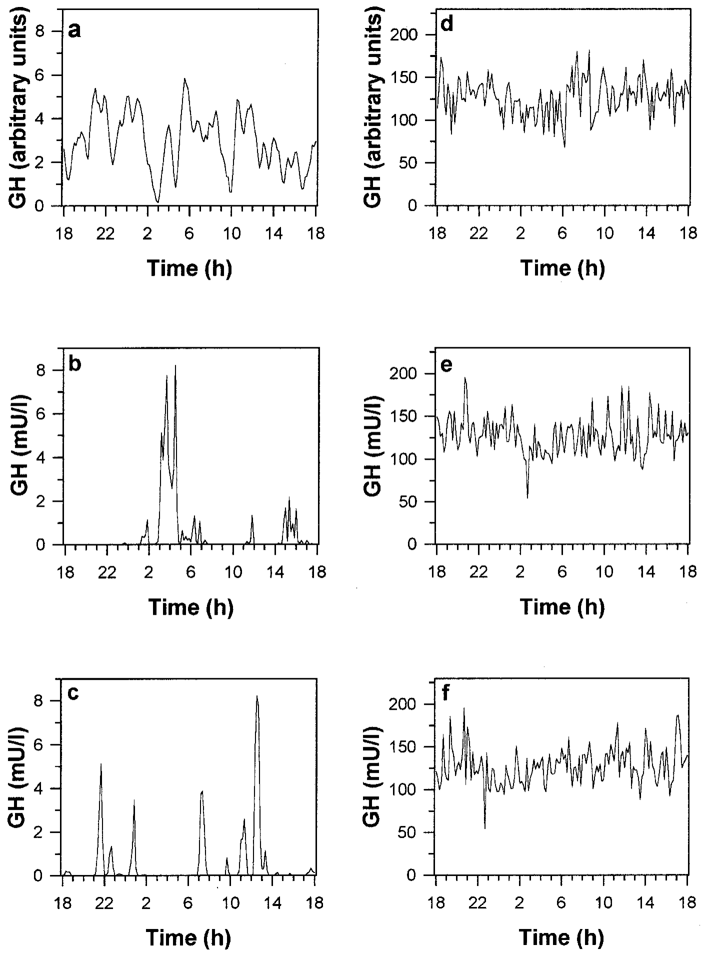

Figure 13. Three types of random surrogate data sets corresponding to original growth hormone (GH) time series in Figure 12. Healthy subject: (**a**) phase-randomized, (**b**) Gaussian-scaled, (**c**) Fourier-shuffled; patient with acromegaly: (**d**) phase-randomized, (**e**) Gaussian-scaled, (**f**) Fourier-shuffled. Ten random surrogate data sets of each type were generated for each experimental 24 h GH time series. The construction of these types of surrogates is given in more detail in Chang *et al.* (1994). A full listing of the MATLAB (The Mathworks, Inc., Natick, MA, USA) source code can be found in Schiff *et al.* (1994b).

prediction of time series by neural networks expanded this concept. In contrast to the published heuristic and model-based algorithms, no assumptions were made on the physiological regulation of the endocrine system under study. The use of several specialized expert networks allowed for an optimal local adaptation of the predicted GH serum concentrations to the actually measured values. By this approach, acromegalic patients under different therapeutic regimens were clearly distinguishable, with mean GH concentrations comparable to healthy controls under fasting. The 24 h temporal pattern of pulsatile GH secretion in healthy controls was significantly better predicted than that of acromegalic patients. A modular system of adaptive mixtures of local expert networks significantly improved the predictive power compared to a previous technique where we used a single neural network for predicting the 24 h PTH profile in osteoporotic patients and healthy controls (Prank *et al.*, 1995b). If a single feedforward neural network is trained for time series prediction, strong interference effects might occur due to different local secretory dynamics, resulting in slow learning and poor generalization. This interference was reduced using several distinct expert neural networks plus a gating network to decide which expert to choose for each training and testing case.

The expert networks specialize in predicting distinct parts of a time series in a self-organized way, without any prior knowledge of the physiological characteristics of the system. Thus, time series analysis is expanded beyond counting frequency and amplitude of hormone pulses. This approach may eventually lead to new and more general conclusions regarding the temporal architecture of hormonal stimulation and its impact on target organs. Results from a number of in-vitro studies have suggested that the response of a target cell to a given hormonal stimulus may vary depending on the timing of the previous stimulus (Knobil, 1980; Weigle and

Goodner, 1986; Shupnik, 1990; Haisenleder *et al.*, 1992; Schöfl *et al.*, 1993).

Recently, the question of temporal versus mean rate coding has been raised for the nervous system. If a mean rate code is operative, the average number of spikes within a given temporal integration window will carry all the information about the message. Rearranging spike time patterns without changing the number of spikes should have no effect if neurons encode visual information only by a mean rate code (Barlow, 1972). However, neocortical neurons are capable of precise spike timing (Mainen and Sejnowski, 1995), and temporal coincidences have been observed between pairs of cortical neurons (Abeles, 1991; Engel *et al.*, 1992). The importance of temporal coding is not solely based on its increased efficiency over mean rate coding since multiple components can carry separate information that a single component code would confound. Although it cannot be proved yet whether a temporal code is actually used in the processing of visual information, temporal codes have several advantages over mean rate coding. To date, experimentalists in the field of endocrinology have mostly focused on the impact of mean rate coding by pulsatile hormone stimulation on cellular responses rather than on a possible impact of temporal coding (Knobil, 1980; Weigle and Goodner, 1986; Shupnik, 1990; Haisenleder *et al.*, 1992; Schöfl *et al.*, 1993). Similarly, currently available methods for the analysis of an endocrine time series mainly concentrate on mean rate coding. Thus, in contrast to the system of multiple adaptive neural networks described here, previous approaches to the analysis of pulsatile hormone secretion might have missed some biologically important information.

Both the ApEn statistic and the adaptive mixture of local experts presented in this study may be capable of extracting temporal information from a hormonal rhythm. Analysis of the selection pattern of the local experts chosen to predict a distinct value of the GH time series revealed that the adaptive mixture of experts system had performed a self-organized quantification of hormone pulsatility, reflected in a switching between local experts. In normal controls the frequency distribution of the selected expert networks indicated that the local expert which focused on predicting the GH baseline was chosen in almost every case. In acromegalic patients the expert best predicting the decrease of a GH pulse contributed significantly more to the prediction of the GH time series than in healthy controls. Thus, these histograms of selected expert networks may be used to separate temporal rhythms in health and disease in different endocrine systems. Furthermore, it may be possible in the future to reduce significantly the number of data points necessary for the analysis of such a hormonal rhythm. Currently available methods of discrete pulse detection are only applied in an experimental set-up where a serum concentration time series of the hormone under study is measured over a long period, usually 24 h. Due to this limitation these approaches are not useful as a diagnostic tool on a routine basis in clinical endocrinology. In preliminary studies we have significantly reduced the amount of data necessary to separate disturbed pulsatile GH secretion in acromegaly from physiological secretion (Prank *et al.*, 1996b). Further studies are needed using more frequent blood samples to assess systematically the possibility of increasing the sampling interval (subsampling) and reducing the total length of the measured time series while still allowing for a significant separation of pathophysiologically disturbed from physiological patterns of pulsatile hormone secretion. The methods described here open new avenues for the study of time-dependent regulation in endocrine systems.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft under grant Br 915/4–2.

References

Abeles, M. (1991) *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press, Cambridge.

Barlow, H.B. (1972) Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, **1,** 371–394.

Baxt, W.G. (1995) Application of artificial neural networks to clinical medicine. *Lancet*, **346,** 1135–1138.

Baxt, W.G. and Skora, J. (1996) Prospective validation of artificial neural network trained to identify acute myocardial infarction. *Lancet*, **347,** 12–15.

Blinowska, K.J. and Malinowksi, M. (1991) Non-linear and linear forecasting of the EEG time series. *Biol. Cybern.*, **66,**159–165.

Brabant, G., Prank, K. and Schöfl, C. (1992) Pulsatile patterns in hormone secretion. *Trends Endocrinol. Metab.*, **3,** 183–190.

Buchfelder, M., Hofmann, B., Prank, K. *et al.* (1994) Plasma half-life of pituitary hormones determined by perioperative measurements in secreting adenomas. *The Endocrine Society, 76th Annual Meeting*, Anaheim, 1586. (Abstr.)

Cacciatore, T.W. and Nowlan, S.J. (1994) Mixtures of controllers for jump linear and non-linear plants. In Cowan, J., Tesauro, G. and Alspector, J. (eds), *Advances in Neural Information Processing Systems 6.* Morgan Kaufman, San Mateo, CA, pp. 719–726.

Chang, T., Schiff, S.J., Sauer, T. *et al.* (1994) Stochastic versus deterministic variability in simple neuronal circuits: I. Monosynaptic spinal cord reflexes. *Biophys. J.*, **67,** 671–683.

Churchland, P.S. and Sejnowski, T.J. (1992) *The Computational Brain*. MIT Press, Cambridge, MA.

Cross, S.S., Harrison, R.F. and Kennedy, R.L. (1995) Introduction to neural networks. *Lancet*, **346,** 1075–1079.

Daughaday, W.H. (1995) Growth hormone, insulin-like growth factor, and acromegaly. In DeGroot, L. J. (ed.), *Endocrinology.* Saunders, Philadelphia, pp. 303–329.

Dybowski, R. and Gant, V. (1995) Artificial neural networks in pathology and medical laboratories. *Lancet*, **346,** 1203–1207.

Dybowski, R., Weller, P., Chang, R. and Gant, V. (1996) Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet*, **347,** 1146–1150.

Engel, A., König, P., Kreiter, A. *et al.* (1992) Temporal coding in the visual cortex – new vistas on integration in the nervous system. *Trends Neurosci.*, **15,** 218–226.

Faira, A.C.S., Veldhuis, J.D., Thorner, M.O. and Vance, M.L. (1989) Half-life of endogenous growth hormone (GH) disappearance in normal man after stimulation of GH secretion by GH-releasing hormone and suppression with somatostatin. *J. Clin. Endocrinol. Metab.*, **68,** 535–541.

Grossberg, S. (1988) Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, **1,** 17–61.

Haisenleder, D.J., Ortolano, G.A., Dalkin, A.C. *et al.* (1992) Differential actions of thyrotropin (TSH)-releasing hormone pulses in the expression of prolactin and TSH subunit messenger ribonucleic acid in rat pituitary cells in vitro. *Endocrinology*, **130,** 2917–2923.

Hartman, M.L., Pincus, S.M., Johnson, M.L. *et al.* (1994) Enhanced basal and disorderly growth hormone secretion distinguish acromegalic from normal pulsatile growth hormone release. *J. Clin. Invest.*, **94,** 1277–1288.

Haykin, S. (1991) *Adaptive Filter Theory*, 2nd edn. Prentice-Hall, Englewood Cliffs, NJ.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J. and Hinton, G.E. (1991) Adaptive mixtures of local experts. *Neural Comp.*, **3,** 79–87.

Jordan, M.I. and Jacobs, R.A. (1992) Hierarchies of adaptive experts. In Moody, J., Hanson, S. and Lippman, R. (eds), *Advances in Neural Information Processing Systems 4*. Morgan Kaufman, San Mateo, CA, pp. 985–993.

Jordan, M.I. and Jacobs, R.A. (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Comp.*, **6,** 181–214.

Knobil, E. (1980) The neuroendocrine control of the menstrual cycle. *Recent Prog. Horm. Res.*, **36,** 53–88.

Lapedes, A.S. and Farber, R.M. (1987) *Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling*. Technical Report No. LA-UR-87–2662, Los Alamos National Laboratory, CA.

Lefebvre, J.H., Goodings, D.A., Kamath, M.V. and Fallen, E.L. (1993) Predictability of normal heart rhythms and deterministic chaos. *Chaos*, **3,** 267–276.

Mainen, Z.F. and Sejnowski, T.J. (1995) Reliability of spike timing in neocortical neurons. *Science*, **268,** 1503–1506.

Mead, C.A. and Conway, L. (eds) (1980) *Introduction to VLSI Systems*. Addison-Wesley, Reading, MA.

Merriam, G.R. and Wachter, K.W. (1982) Algorithms for the study of episodic hormone secretion. *Am. J. Physiol.*, **243,** E310–E318.

Nowlan, S.J. (1990) *Competing Experts: An Experimental Investigation of Associative Mixture Models*. Technical Report CRG-TR-90–5, Department of Computer Science, University of Toronto, Canada.

Nowlan, S.J. (1991) *Soft Competitive Adaptation: Neural Network Learning Algorithms Based on Fitting Statistical Mixtures*. CMU-CS-91–126, School of Computer Science, Carnegie Mellon University, Pittsburgh, Ph.D. thesis, pp. 110–130.

Nowlan, S.J. and Hinton, G.E. (1992) Simplifying neural networks by soft weight-sharing. *Neural Comp.*, **4,** 473–493.

Pawelzik, K., Kohlmorgen, J.M. and Müller, H.R. (1996) Annealed competition of experts for a segmentation and classification of switching dynamics. *Neural Comp.*, **8,** 340–356.

Pincus, S.M. (1991) Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA*, **88,** 2297–2301.

Pincus, S.M. and Keefe, D.L. (1992) Quantification of hormone pulsatility via an approximate entropy algorithm. *Am. J. Physiol.*, **262,** E741–E754.

Prank, K. (1995) Nonlinear dynamics in endocrine systems. *Exp. Clin. Endocrinol.*, **102** (Suppl. 1), S2 (Abstr.).

Prank, K. and Brabant, G. (1994) Estimating thyrotropin secretory activity by a deconvolution procedure. *Methods Neurosci.*, **20,** 377–389.

Prank, K., Riedel, M., von zur Mühlen, A. and Brabant, G. (1995a) Separating the secretory dynamics of growth hormone (GH) in acromegaly from normal subjects: quantification of regularity in pulsatile hormone secretion by the measure of approximate entropy. *Exp. Clin. Endocrinol.*, **102** (Suppl. 1), V25 (Abstr.).

Prank, K., Nowlan, S.J., Kloppstech, M. *et al.* (1995b) Time series prediction of plasma hormone concentration: evidence for differences in predictability of parathyroid hormone secretion between osteoporotic patients and normal controls. *J. Clin. Invest.*, **95,** 2910–2919.

Prank, K., Kloppstech, M., Nowlan, S.J. *et al.* (1996a) Self-organized segmentation of time series: separating growth hormone secretion in acromegaly from normal controls. *Biophys. J.*, **70,** 2540–2547.

Prank, K., Kloppstech, M., Nowlan, S.J. *et al.* (1996b) Random episodic release of growth hormone in humans. *Physiol. Rev. Lett.*, **77,** 1909–1911.

Rapp, P.E., Albano, A.M., Zimmerman, I.D. and Jimenez-Montano, M.A. (1994) Phase-randomized surrogates can produce spurious identifications of non-random structure. *Phys. Lett. A*, **192,** 27–33.

Riedel, M., Günther, T., von zur Mühlen, A. and Brabant, G. (1992) The pulsatile GH secretion in acromegaly: hypothalamic or pituitary origin? *Clin. Endocrinol.*, **37,** 233–239.

Riedel, M., Hoeft, B., Blum, W.F. *et al.* (1995) Pulsatile growth hormone (GH) secretion in lean and obese men: differential metabolic regulation of GH release during energy restriction. *Metabolism*, **44,** 605–610.

Robinson, D.A. (1992) Signal processing by neural networks in the control of eye movements. In *Computational Neuroscience Symposium*. Indiana University-Purdue University at Indianapolis, pp. 73–78.

Santoro, N., Filicori, M. and Crowley, W.F. (1986) Hypogonadotropic disorders in men and women: diagnosis and therapy with pulsatile gonadotropin-releasing hormone. *Endocr. Rev.*, **7,** 11–23.

Schiff, S.J., Jerger, K., Chang, T. *et al.* (1994a) Stochastic versus deterministic variability in simple neuronal circuits: II. Hippocampal slice. *Biophys. J.*, **67,** 684–691.

Schiff, S.J., Sauer, T. and Chang, T. (1994b) Discriminating deterministic versus stochastic dynamics in neuronal activity. *Integrative Physiol. Behav. Sci.*, **29,** 246–261.

Schöfl, C., Brabant, G., Hesch, R.D. *et al.* (1993) Temporal pattern of α_1-receptor stimulation regulate amplitude and frequency of calcium transients. *Am. J. Physiol.*, **265,** C1030–C1036.

Schöfl, C., Prank, K. and Brabant, G. (1994) Mechanisms of cellular information processing. *Trends Endocrinol. Metab.*, **5,** 53–59.

Shupnik, M.A. (1990) Effects of gonadotropin-releasing hormone on rat gonadotropin gene transcription in vitro: requirement for pulsatile administration for luteinizing hormone β-gene stimulation. *Mol. Endocrinol.*, **4,** 1444–1450.

Siragy, H.M., Vieweg, W.V.R., Pincus, S. and Veldhuis, J.D. (1995) Increased disorderliness and amplified basal and pulsatile aldosterone secretion in patients with primary aldosteronism. *J. Clin. Endocrinol. Metab.*, **80,** 28–33.

Sugihara, G. (1994) Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. Lond. A*, **348,** 477–495.

Sugihara, G. and May, R.M. (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344,** 734–741.

Theiler, J., Eubank, S., Longtin, A. *et al.* (1992) Testing for nonlinearity in time series—the method of surrogate data. *Physica*, **58D**, 77–94.

Tsonis, A.A. and Elsner, J.B. (1992) Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature*, **358,** 217–220.

Veldhuis, J.D. and Johnson, M.L. (1986) Cluster analysis: a simple, versatile, and robust algorithm for endocrine pulse detection. *Am. J. Physiol.*, **250,** E486–E493.

Veldhuis, J.D., Moorman, J. and Johnson, M.L. (1994) Deconvolution analysis of neuroendocrine data: waveform-specific and waveform- independent methods and applications. *Methods Neurosci.*, **20,** 279–325.

Weigend, A.S. and Gershenfeld, N.A. (eds) (1993) Time series prediction: forecasting the future and understanding the past. *SFI Studies in the Sciences of Complexity, Proceedings*, vol. XV. Addison-Wesley, Reading.

Weigend, A.S., Huberman, B.A. and Rumelhart, D.E. (1990) Predicting the future: a connectionist approach. *Int. J. Neur. Sys.*, **1,** 193–209.

Weigle, D.S. and Goodner, C.J. (1986) Evidence that the physiological pulse frequency of glucagon secretion optimizes glucose production by perifused rat hepatocytes. *Endocrinology*, **118,** 1606–1613.

Widrow, B. and Hoff, M.E. Jr. (1960) Adaptive switching circuits. *IRE WESCON Convention Record*, 96–104.

Widrow, B. and Stearns, S.D. (1985) *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Received on November 18, 1996; accepted on March 21, 1997