Process fault detection through quantitative analysis of learning in neural networks

Martijn van Veelen, Jos A.G. Nijhuis and Lambert Spaanenburg

University of Groningen,
Department of Computing Science
P.O. Box 800,
9700 AV Groningen, the Netherlands
tel. +31 (0)50 363 3939 fax. +31 (0)50 363 3800
email: {martijn,ti}@cs.rug.nl

Abstract—In this paper we demonstrate the use of quantitative measures of learning in neural networks for detection of chronic disturbances. This is a first attempt to provide such a measure in the data-driven black-box engineering practice for detection of unknown non-cataleptic disturbances. Such events are characterized by the presence of unknown influences, but observable in measurable variables. Though such events do not necessarily have an immediate impact on a model’s validity for process measurements, a significant change in the model’s learning behavior can be detected. This is demonstrated by detecting discrepancies in some exemplary toy-problem data sets in approximation, prediction and identification. Though analysis of parameter dynamics appears successful in these problems, future research is required on sensitivity robustness and promptness considering measures of external and internal behavior.

Keywords—neural networks, learning behavior, fault detection,

I. INTRODUCTION

Process fault detection and isolation (FDI) relies on many different areas of research involved in (1) model construction, such as mathematical system theory and computational intelligence, and (2) decision making such as statistical hypothesis testing and classical artificial intelligence providing for theory on rule-based systems. An interesting challenge is offered in the development of new combinations of various achievements in this multi-disciplinary research. This paper presents such a combination.

FDI paradigms have been formulated on basis of mathematical systems theory in the late eighties by Isermann in [7], and covered extensively by Patton, Frank and Clark in [14]. This foundation is focussed on control systems and system identification using the white-box and model-based design approaches, i.e. models are based on laws of physics rather than on data. Basically the models are obtained through linearization, on-line adaptation is typically realized through Kalman-filters [14]. In essence a detection algorithm consists of the following three steps [5]:

1. model selection;
2. computation of residuals called signatures;
3. comparison of signatures and decision making.

Over the last decade, three developments can be identified in the research related to FDI. The first development is the increasing use of neural networks as replacement and supplement in traditionally mathematical systems [12] and the application in detection [9]. The second development is the use of learning methodology for detection [17] also in combination with a traditional white-box model-based approach [16]. A third development is the increasing research into detection for processes other than those related to mathematical control systems and electrical engineering such as data-mining and internet monitoring systems [17]. A result of these developments is that we are now facing two problems which prevent the successful application of the traditional approach to find detection measures through analysis of parameter dynamics.

Problem 1: The disturbances that are to be detected are no longer always known in advance, though this was the primary assumption in most traditional FDI approaches [5].

Problem 2: The nature of both the black-box models and associated data-driven optimization procedures, such as error-backpropagation (EBP) and evolutionary algorithms, differ from that of traditional white-box model-based approaches and associated optimization algorithms, making model parameters and their dynamics hard to interpret.

If the process structure is obscure and can only be observed through data, the question is: can process changes, reflected in the data as chronic disturbances, be revealed by internal and external metrics of learning systems designed for data-driven parameter optimization of non-linear systems. Quantification of learning
behavior in non-linear systems, e.g. neural network with EBP, is not easily realized. Some more or less successful attempts have been reported [1,6,11,13,18]; though quite hairy, parameter dynamics can be described in a probabilistic framework.

In section II we start with the basics of detection algorithms and chronic disturbances. In section III we describe learning systems and discuss how they can serve the purpose of the detection in abstracto. The actual quantitative measures subject of this research are presented in section IV. Experiments and results are discussed in section V.

II. Detecting Chronic Disturbance

While many definitions of faults, abnormalities and novelties exist, they all express the notion of a difference between what is expected and what is observed. What is expected is often defined by a model, what is observed by data. It is commonly assumed that measurements are noisy and models incomplete and uncertain. The term disturbance refers to the combined effect of noise, random factors, unknown effects and modelling uncertainties. For detection a clear separation is required between chronic and cataleptic disturbances, this is realized by a detection metric. Cataleptic disturbances can not be helped by improving a model while unknown effects and modelling uncertainties can. Chronic and time-related disturbances differ essentially from cataleptic disturbances as they are: disturbances related to known quantities.

Without giving the complete list of definitions actually required to express detection formally, definition I is sufficient backbone for our discussion. Essential is the mapping of a model’s validity for observed data to signatures, e.g. residuals, and the detection metric which maps signatures to a decision space where chronic and cataleptic disturbances can be separated easily, figure 1.

definition I (detection algorithm)

Given a model $M_W$ with parameter set $W$ intended to describe data samples $\xi = ((u, y) t)_{t \in T}$ from information source $I$, a detection algorithm $DA$ is a test described by test quantity $d_t(M)$ mapping samples $\xi$ to a decision space $D$ which contains a subspace $D_0$ associated with the hypothesis $H_0$ stating that the discrepancies $\xi_t(M)(\xi)$ in the model w.r.t. the observed sample $\xi$ differs chronically from expected disturbances.

Though the detection algorithm operates on samples - series of measurements or data-points - and not on individual data-points, on-line early detection may require a response based on only one or a few data-points. Reliability demands however sufficient data-points to attain confidence in the detection. Therefore we limit our discussion to sample-based detection.

![Fig. 2. Areas of interest.](image)

Detection problems are characterized by the type of model used to fit the data and the type of disturbance investigated. Figure 2 shows the various types of problems investigated in the experiments discussed in section V. However, before discussing the results on the use of learning behavior in the detection algorithm we will first introduce learning systems and discuss how they can be used for detection.

III. Learning Systems in Detection

The dynamic parameter variations during on-line parameter estimation have been recognized to be sensitive to abnormal process behavior by Baskiotis and Rault, Isermann and Patton [4, 14]. White-box models allow for a more or less direct transformation from these dynamics to sensitive decision variables either by eigen-structure analysis as in the parity-space approach [5] or by transforming dynamics through semantical interpretation into process coefficients and non-measurable quantities such as energy use as described in [7]. However, with typical non-linear black-box modelling and step-wise learning algorithms, such transformations may be hard to find. The main causes are the redundant information storage and stochastic nature of the learning process [17]. Clearly development of quantitative measures for learning pro-
cesses in general is required for an effective detection algorithm based on neural networks. The development of such measures demands a flexible definition of the learning system, that goes beyond leastsquares optimization and Kalman-filtering. Such a definition must include model structure, model configuration, optimization algorithm and stopping criterion. Not to much surprise our definition bears close resemblance to that of a state machine:

**definition II** (learning system)
A *learning system* is a five-tuple \((M, w, v, L, C)\) with \(M\) a model structure which together with configuration (parameter set) \(w\) defines a relation \(M_w\) on variables \(v = \{v_1, \ldots, v_n\}\), \(L\) a learn function mapping configuration and variables to configuration changes \(\Delta w\), and \(C\) a mapping from samples \(\xi\) to a variable expressing if the learning is in a (final) state, i.e. a stopping criterion.

A stable and accurate description of states in learning systems reflects the behavior of the model w.r.t. the modelled data which in turn reflects the behavior of the underlying process. It is assumed that learning behavior in an already optimized model reflects the internal turmoil or change in the modelled process as a result of unknown influences, figure 3. An accurate description of the learning behavior of a learning system in relation to a specific information source characterizes an unknown and presumable only internal state of that information source. Obscure influences forcing a state change in the modelled process, perhaps to an undesirable or faulty state, will be revealed by any accurate metric of the learning behavior. Therefore metrics for learning behavior can be used for process fault detection or, more generally speaking, for the detection of chronic disturbances. Finding suitable, stable and *sensitive*, metrics \(C\) of learning behavior is hard but possible as is demonstrated in the next sections.

![Diagram](image)

**Fig. 3.** Changes in the process underlying the observed data are assumed to be reflected in the learning behavior of a learning system.

**IV. Metrics for Learning Systems**

Description of learning processes and equilibria - stable states - is hard, yet some rather successful attempts exist based on statistical physics [6] and neurodynamics [1]. A thorough analysis has been performed by Leen, Orr and Moody to describe behavior in equilibria [11] and in transients [13]. The internal behavior in terms of parameters is described in parameter drift: the distribution of the direction of parameter change; and diffusion: the relative distribution of parameter fluctuations in every direction.

In practical applications of learning systems only two states are considered: the initial state usually obtained by random initialization, and the final state where the system is not required to continue to learn as the trained model \(M_w\) meets the designer’s criterion. The model’s output error is usually the basic ingredient of a stop-criterion for learning systems, see equation (1). In practice these states are described by the model’s output error such as the RMSE. However existence of local minima and complexity of the relation to learn cause the existence of many more states; more elaborate different mappings \(C\) exist to express and describe these different states.

\[
C(w, \xi) \equiv \sum_{(v_{in}, v_{out}) \in \xi} ||M_w(v_{in}) - v_{out}|| < \varepsilon \tag{1}
\]

Detection metrics, definition I, define by subspace \(D_0 \equiv \{\xi| -C(w, \xi)\}\) the samples for which \(C\) does not hold. However, it is questionable whether chronic disturbances cause the stop-criterion not to hold. Exogenous or externally observable behavior of the learning can be measured in several ways. Typically residual analysis is performed to determine if the model can be improved; more elaborate analysis can be performed on dynamical models using extrapolation. Such *external residue-based metrics* are not studied here: artificial data is constructed so that they are bound to fail. Externally observable learning behavior is the number of epochs required to achieve the stopping criterion. We study models \(M_w\) which have already been optimized for a specific sample \(\xi\), i.e. \(w^*\) is an equilibrium and \(C(w^*, \xi)\) invariant for learning \(w^*\) with \(\xi\).

- If a sample \(\xi^?\) contains chronic disturbances, learning affects the model which may cause the stopping criterion to fail: \(C(w^*, \xi) \neq C(w^?, \xi^?)\); as a result it will take some epochs of training with non-disturbed data before the stopping criterion is re-achieved, this is the *recovery time*.

Learning is intended to optimize the model parameters using a training sample; therefore the gradients reflect the learning system’s response to the training.
sample. Many internal metrics for learning behavior are based on gradients, such as neuron sensitivity, saliency and redundancy. Such metrics are at the heart of a successful network structure and learning rate optimization [3,8,10]. Use of learning rate and momentum is required for stable parameter updates but damps the gradients; for better resolution the gradients are used instead of parameter updates. We will first look at the gradients themselves, as metrics for chronics disturbances. Two detection test are used here:

- Boundary test: fraction of gradients \( L(w^*, \xi^*) \) of possibly disturbed samples \( \xi^* \) outside the 2\( \sigma \) boundary, which is the standard deviation of gradients in response to reference sample \( \xi \).
- Kolmogorov-Smirnov test: giving the probability that the distribution of gradients \( L(w^*, \xi^*) \) is equal to that of \( L(w^*, \xi) \) [15].

Gradients are not expected to be stable metrics as there is still problem 2 from the introduction: sample \( \xi \) and model structure \( M \) do not determine an unique configuration \( w \); stochastical components such as random initialization, random sample selection and jittering improve the learning process significantly [2].

But they cause non-uniqueness of trajectories in the parameter space. Moreover stability of the model’s exogenous behavior, as can be formulated by a stopping criterion, does not require stability of the model configuration during learning. While the gradients will contain very specific information from which chronic disturbances can be detected, they can not always trusted for stable detection. To find a suitable projection of the gradients to a stable detection test, we must first understand how learning systems behave in equilibria. One interesting way to look at learning processes in black-box models is to view them from a control perspective:

A learning process has the goal of steering the models parameters to a “best” configuration by feeding the proper signals through the selected model structure.

This can is realized by: (1) selecting a proper model architecture, possibly augmenting it during training; (2) selecting patterns available through samples of the process, compute errors from them by which the parameters can be driven, and (3) selection of suitable learning rates. Again, random effects such as random sample selection can significantly improve the learning process.

An equilibrium can now be recognized in terms of a control problem: the learning system can not improve the model configuration in any way, offering the measurements from the available sample - in any order - does not result in a “better” model. When parameters do fluctuate, as is often the case in neural networks, while further learning does not improve performance, it may be concluded that any improvement on (part of) the sample w.r.t. to a subset of model parameters implies a degradation for one or more model parameters on (another part of) the sample [2]. Thus, an equilibrium in the learning process is characterized by the observation that model parameters can not be steered independently with the available data, i.e. either parameter adaptations are coupled statically, or a parameter adaptation at one learn cycle in the learning epoch is cancelled by the combined effects of the other learning cycles in the same learning epoch (dynamic coupling).

While the dynamic coupling of gradients grows very quickly with network size and epoch length, it will not easily provide a scalable detection metric. We study here the discriminative power of static relations between gradients. As chronic disturbances are related either to time or to measurement variables, they are expected to cause significant change in the couplings of gradients if not even in the gradients themselves.

V. Experiments

In the previous section we have argued that metrics for learning behavior can be used for detection of chronic disturbances, and subsequently reveal the existence of unknown influences on the process behavior. In the discussion on learning processes we indicate that any sub-optimal parameter configurations is characterized by either or both static and dynamic couplings in gradients. Moreover, it is expected that change in the dynamic couplings of the gradients indicates a non-catalectic or chronic disturbances in the sample used for learning. The experiments presented here have been conducted in order to find numerical support for this idea.

First we want to establish that learning in equilibria might indeed imply couplings between gradients which are characteristic for the learning sample. The approximation of the XOR, which is known to cause internal conflicts, should reveal these couplings between gradients. Moreover the internal conflict should be easily revealed relating conflicting directions of parameter change to different input-domains, subsection A. As parameter changes are bound to affect the model performance it is verified that recoverytime is indeed a potential detection metric.

While gradient couplings are characteristic for equilibria in general and not for learning problems alone, the gradient couplings are investigated in the less problematic case of approximating a sine in subsection B. Estimating the couplings through cross-correlation of different gradients, first numerical support for the possibility of detecting chronic disturbances using inter-
nal learning behavior is found. It is also verified that variation in the gradients does not exclusively respond to chronic disturbance, and can therefore not be considered a suitable detection metric.

Process models are in general dynamic. Therefore the various proposed detection metrics are used for detection with dynamic data in subsection C. The first problem is detection of both statically related as dynamic disturbances in a sine wave which is to be predicted one step ahead. This toyproblem bears already close relation with data encountered in practice from dynamical systems, as in this example the dynamic disturbance bears no direct relation with the input variables and may therefore be considered an unknown influence. Even more interesting is the multivariate dynamical process Volterra-Lotka, here it is used to demonstrate that even gradients of connections from different inputs in a sub-optimal model are related in very specific ways. While the proposed detection relies on the presence of couplings between gradients, these couplings should be present in large amounts. An estimate of the distribution of correlations between all available gradients in a reasonable good Volterra-Lotka model, indicates sufficient amounts of related gradients. Whether this indeed promises sensitive detection will be investigated in the near future.

A. Exclusive OR

The exclusive-OR, equation (2), is known to be problematic for gradient-based learning. Many learning runs do not converge to a solution. Causes are assumed to be: (1) the fact that the subspaces of the XOR are not linearly separable; and (2) discontinuity of the first derivative in 0.

\[
\text{XOR}(v_1, v_2) = \begin{cases} 
0 & \text{if } \text{sign}(v_1) = \text{sign}(v_2) \\
1 & \text{if } \text{sign}(v_1) \neq \text{sign}(v_2)
\end{cases}
\] (2)

Leaving out one of the four quadrants in the training data sampled from the XOR-function will solve the problem of separability, thus any optimized configuration is an instable equilibrium in the learning process and will display the internal conflict which is observable as orthogonal gradients, i.e. the parameter is adapted in opposite directions, and coupling of parameter adaptations when trained with XOR-data.

To test this, several multilayer Perceptrons (linear output and 5, 7 or 9 hidden neurons) were trained with XOR-data sampled on an equidistant grid of 6×6 points: \(v_1[n] = 2/5(n \div 6) - 1, v_1[n] = 2/5(n \mod 6) - 1\), Gaussian output noise is added to obtain the sample \(\xi\):

\[
(v_1[n],v_5[n]) = \text{XOR}(v_1[n],v_2[n]) \pm \mathcal{N}(0,0.05)\) \(0 \leq n \leq 36\) (3)

The gradients are measured in those networks that actually have converged to a solution \(w^*\) with stopping criterion \(C(w^*,\xi) \equiv \text{RMSE}(\xi) < 0.13\). Figure 4 shows the typical gradient-gradient (\(\partial \mathcal{E}(v)/\partial w^*_{ij}, \partial \mathcal{E}(v)/\partial w^*_{pq}\)) plot for two connections \(w^*_{ij}\) and \(w^*_{pq}\) from input to hidden neurons for a sample \(\xi\) from an equidistant input-grid of 16×16 points. Each point in this scatterplot was obtained by computing the local gradient using EBP, for every data-point in the sample, for two connections from the input to a hidden neuron of an MLP; one gradient \(\partial \mathcal{E}(v)/\partial w^*_{ij}\) determines the position on the horizontal axis the other \(\partial \mathcal{E}(v)/\partial w^*_{pq}\) the position on the vertical axis. The gradients reveal the conflict in both ways: by coupling and by opposite sign. It may be concluded that the error-function \(\mathcal{E}(w_{ij},w_{pq})\) has a saddle-point in \(w^*\) where these gradients are measured. Figure 5 shows the dependencies in the different quadrants.

![Fig. 4. Saddle point causing the XOR-problem observed in gradient-gradient.](image)

![Fig. 5. Gradient-vs-gradient conditioned to quadrant reveals the dependency of the conflict and of the coupling to the input-domain.](image)
As the converged XOR models are in such delicate equilibria, they are expected to be easily forced out of the equilibrium $w^*$ by learning any chronic-disturbed sample $\xi^*$, and as a consequence failing the stopping criterion $C(w^*, \xi^*)$.

When the stopping criterion is no longer valid, retraining the model with the “good” data $\xi$ is required. Recovery time – the number of epochs to achieve the stopping criterion – is expected to be a suitable metric for the amount of chronic disturbance in the sample $\xi^*$ when such delicate equilibria exist as in the XOR-model. This was tested using 25 converged XOR-models, using several types of disturbances w.r.t equation (3). Figure 6 shows the average recovery time. The disturbances are: (1) gauss - original sample; (2) gauss-II - $N(0, 0.1)$ replaces $N(0, 0.05)$; (3) noise - $v_3[n] = U[-1, 1]$ replaces entire function; (4) block - if $n \div 6 < 3$ then 0.1 else -0.1 replaces $N(0, 0.05)$; (5) sawtooth - 0.1($n \mod 3$) – 0.1 replaces $N(0, 0.05)$.

Fig. 6. Results in average recovery-times for different types of disturbances.

The random disturbances require hardly any recovery time, while the chronic – input related – disturbances require significantly more epochs to re-achieve the stopping criterion. While this strongly supports the idea that chronic disturbances can be detected from learning behavior, the equilibrium will hardly ever be as delicate that such small disturbances immediately cause the stopping criterion to fail. In the next subsections we present less problematic data samples to approximate.

B. Function approximation : sine

Though approximation of a sine function by a MLP is not as problematic as the XOR, it should also reflect the gradient dependencies in equilibria. MLPs with linear outputs are trained with a sample $\xi$ as defined in equation (4). The sample consists of 1024 points, and we have trained 25 cross-validation models, with stopping criterion $C \equiv \text{RMSE}(\xi_{\text{test}}) < 0.06$ for 10 subsequent epochs.

$$v|v_2 = \sin(2\pi v_1) + N(0, 0.05) \land v_1 \in U[-1, 1]$$ (4)

The relation between gradients indeed appears to be very characteristic for the learning system w.r.t. the learned mapping in the equilibrium. Moreover the characterization is robust in separating random deviations from chronic disturbances: the random disturbances blur the scatterplot slightly while the chronic disturbances actually cause a different relation between gradients. Figure 8 shows fraction of differences in gradient correlations larger than 0.03, measured using a sample equivalent to equation (4) and several disturbances, all the chronic disturbances having equal or smaller amplitudes than the random disturbances!

Mapping these gradient relations will, even using relatively simple measures such as crosscorrelation, grow very quickly $O(n! \times m)$ with the number of con-
connections \( n \) and the number of data-points \( m \) in the test sample \( \xi^2 \) and even faster than \( O(n! \times m^2) \) if we were also to include dynamical dependencies. It should be verified that the detection can not be based on simpler measurement such as the gradients themselves. Using the 25 cross-validation models we have also performed the boundary test and the Kolmogorov-Smirnov test as described in section IV. Figure 9 shows the results. The boundary test has a very low resolution, using only 10% of the scale \([0 -1]\), the random disturbances can be relatively more out of bound while the relation between gradients is not significantly affected, as is shown in figure 7.

While the Kolmogorov-Smirnov test performs well in separating different distributions of the gradients, figure 9b, also the uniformly distributed and second Gaussian distribution are easily considered to be different from the reference gradients as computed on Gaussian disturbance (the top bar which has estimated probability 1.0 of being equal to the reference gradients). The use of distribution comparison tests is questionable when looking for chronic disturbances, as changes in the distribution are also caused by cataleptic (random) disturbances.

The chronic disturbances do not necessarily cause the stopping criterion to fail when they are used for training, as can be observed in figure 10. However in these cases the average recovery times over the 25 crossvalidation models are significantly longer for the chronic disturbances.

C. Prediction examples: sine and predator-prey

In the last two experiments we come to show that the suggested quantifications of learning dynamics are also suitable for detecting chronic disturbances in process data.

\[
 v_n = \sin \left( \frac{2\pi f}{T_S} \right) + \mathcal{N}(0, 0.01) \text{ with } T_S = 32 \]  

The first experiment consists of the one-step-ahead prediction of a sine function, equation (5), using 1024 data points with \( f = 75/64 \). The model contains 7 delay elements and 10 hidden neurons with linear output, 25 cross-validation models have been trained to stopping on the stable train error \( C \equiv |RSE(w_n, \xi_{\text{train}}) - RSE(w_{n+1}, \xi_{\text{train}})| < 10^{-4} \) for 10 subsequent epochs. The final RSE on the test sets \( \xi_{\text{test}} \) test is \( 1.46 \cdot 10^{-2} \pm 3.2 \cdot 10^{-3} \).

The disturbances are:

1. \texttt{gauss} - original sample;
2. \texttt{noise} - as in equation (5);
3. \texttt{double} - \( \mathcal{N}(0, 0.02) \) replaces \( \mathcal{N}(0, 0.01) \);
4. \texttt{sawtooth} - \( 0.02/T_S(n \mod T_s) - 0.01 \) replaces \( \mathcal{N}(0, 0.01) \);
5. \texttt{shift} - \( f' = 1.05f \) in equation (5);
6. \texttt{sin} - \( 0.01 \sin(n\pi/T_s) \) replaces \( \mathcal{N}(0, 0.01) \).
The results for the boundary test on the gradients is shown in figure 11. The additive chronic disturbances sawtooth and sin are not significantly out of bound (around and below 5 % as expected for random disturbances using $2\sigma$ boundary), while the internally changed process shift causes the gradients to go out of bound indicating a chronic disturbance. However the double disturbances would also suggest a chronic disturbance if this metric would be used, thus it lacks robustness.

The differences in correlation test do not suffer from this robustness problem, as can be observed in the simulation results in figure 12. The largest observed fraction different correlations in the 25 crossvalidation models over all connections barely tops the minimal observed fraction different gradient-correlations for the sawtooth and does not even come close to the minimum fraction different in correlation for the shifted sine!

The question remains whether these disturbances are also observable by inspecting the recovery times. Using the earlier mentioned stopping criterion, the differences between the recovery times in all 25 crossvalidation models are not larger than 1 epoch. As small chronic disturbances do change the model’s gradient dependencies sufficiently to create havoc, this test has apparently missed the detection for hitherto unknown reasons.

The Volterra-Lotka system, better known as the predatorprey system was identified with an MLP with 5 delays per input one step ahead prediction of one of the two inputs. The properties of 25 models, that converged sufficiently, confirm what we argued in this paper. One of the interesting questions is whether the gradients of the connections from the input delay-lines of the different inputs of the Volterra-Lotka system reveal as much mutual dependency as is the case in the static and single variable systems described in the previous subsections. Figure 13 shows some gradients scatterplots, of such connections, chosen randomly. A dependency is clearly present, indicating potential detection.

The robustness and sensitivity of gradientdependency based detection relies on the presence of many related gradients. Awaiting further simulation results for the detection performance of the metric in the VolterraLotka models, the presence of the gradient dependencies can already be investigated. Figure 14 shows the distribution of the correlation coefficients for all connections to hidden neurons. Sufficient dependency seems to be present for detection.

VI. Conclusions and Future Research

Apparently the difference in relation between gradients measured bluntly with the cross-correlation already reveals small chronic disturbances while being invariant to random disturbances of equal size and robust w.r.t. increasing random disturbances. While it may well be expected that it is troublesome to find suitable metrics for blackbox datadriven learning systems, our experiments demonstrate some promising candidates using learning dynamics. The role of redundancy and nonlinearity in neural networks is expected to play a crucial role in the sensitivity and robustness for the gradient dependencies as metric for
chronic disturbances. The learning rate and momentum, in all the presented experiments, fixed at 0.7 and 0.2 resp. can be optimized to increase sensitivity of the recoverytime, which is also feasible by reducing the size of the reference samples. Moreover the parameters of the detection tests used here, such as the limit difference in correlation, have been chosen ad hoc, here 0.05. These parameters regulate the tradeoff between sensitivity and robustness. We have applied cross-correlation for the dependency measure. As this takes only linear dependency into account, the replacement by, for instance, mutual information will probably make for a further increase of the discriminative power.

We have shown that obscured chronic process perturbations, reflected in the data as chronic disturbances, can be revealed by internal metrics of learning systems designed for data-driven parameter optimization of nonlinear models. The reported results indicate that gradient dependencies are a promising method for tracing chronic process changes. This paves the way for robust process fault detection through quantitative analysis of learning behavior.

REFERENCES
