

Helsinki University of Technology  
Department of Automation and Systems Technology  
Control Engineering Laboratory  
Professor Ph.D. Heikki Koivo

---

## **Diploma Thesis**

### **Data-Based Modeling of Electroless Nickel Plating**

prepared by

**cand. kyb. Hans-Christian Pfisterer**

Professor Dr. Tech. Heikki Hyötyniemi

Professor Dr.-Ing. Dr.h.c. Michael Zeitz

March 2006

---

# Preface

This diploma thesis is based on the theory of “Neocybernetics”. Professor Heikki Hyöty-niemi was essential in helping and supporting my work. He provided the theoretical framework and upcoming new ideas, which made the work on this thesis very interesting. Robert Tenno and Kalle Kantola, who made a lot of effort in gathering data, summariz-ing and understanding the associated mechanisms of the studied process, provided the industrial and chemical background knowledge about the process for the thesis at hand.

My best thanks also to Professor Heikki Koivo who gave me the opportunity to work, research and teach in his laboratory. The TKK Control Engineering Laboratory’s (CEL) friendly and easy-going atmosphere inspired me during the work and made it feel enjoy-able.

Last but not at all least I would like to thank Professor Michael Zeitz of the Institute for System Dynamics (ISYS) at the University of Stuttgart, for supervising my thesis back in Germany and supporting my stay abroad with organizational help in form of the ERASMUS program.

Everything in this work is written and developed by myself. If not, references to books, reports or homepages are mentioned in particular passages and also generally.

Espoo, 17th of March 2006

Hans-Christian Pfisterer

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Neocybernetics</b>	<b>4</b>
2.1	The Origin of Cybernetics . . . . .	4
2.2	To a New Approach of Cybernetics . . . . .	7
2.3	With Mathematics Towards Elasticity . . . . .	11
2.3.1	Modeling Chemical Systems . . . . .	12
2.3.2	Constraints vs. Degrees of Freedom . . . . .	14
2.3.3	Multivariate Methods . . . . .	17
2.3.4	Restructuring Data . . . . .	18
2.3.5	Neocybernetic Perspective . . . . .	22
<b>3</b>	<b>Multivariate Analysis</b>	<b>26</b>
3.1	Data Preprocessing . . . . .	26
3.2	Principal Component Analysis . . . . .	27
3.3	Linear Model . . . . .	33

---

3.4	Multilinear Regression . . . . .	33
3.5	Principal Component Regression . . . . .	35
3.6	Partial Least Squares . . . . .	37
<b>4</b>	<b>Electroless Nickel Plating</b>	<b>40</b>
4.1	Industrial Background . . . . .	40
4.2	Ni-Au Finishing . . . . .	42
4.3	Model and Control Tasks . . . . .	49
<b>5</b>	<b>Existing Models</b>	<b>52</b>
5.1	Historical Models . . . . .	52
5.2	Electrochemical Model . . . . .	56
<b>6</b>	<b>Plating Process as Cybernetic System</b>	<b>64</b>
6.1	The Way to Practice . . . . .	64
6.2	Available Process Data . . . . .	67
6.3	Towards the Real System . . . . .	71
6.3.1	Used Data . . . . .	71
6.3.2	Applied PCA Analysis . . . . .	72
6.3.3	Selection of Basis Vectors . . . . .	74
6.3.4	About the Model . . . . .	75
6.3.5	PCR Estimation Results . . . . .	77

---

6.4	Final Model . . . . .	83
6.4.1	Applied PLS Analysis . . . . .	83
6.4.2	PLS Estimation Results . . . . .	86
6.5	Discussion . . . . .	91
<b>7</b>	<b>Future Prospects</b>	<b>94</b>
<b>A</b>	<b>Example: Constraints vs. Freedom</b>	<b>96</b>
<b>B</b>	<b>Eigenproblem Properties</b>	<b>98</b>
<b>C</b>	<b>Example: Eigenvectors and Eigenvalues</b>	<b>100</b>
<b>D</b>	<b>Calculated Voltages</b>	<b>102</b>
	<b>References</b>	<b>104</b>

# List of Figures

1.1	Complexity theory pessimism vs. neocybernetic optimism . . . . .	2
2.1	Different levels of abstraction while modeling interaction of particles . . . . .	8
2.2	Schematic illustration of the covariance structure among data . . . . .	16
2.3	Illustration of different time scales . . . . .	24
3.1	Data distributions after different preprocessing operations . . . . .	28
3.2	Illustration of the “natural” and the PCA bases for collinear data . . . . .	30
3.3	Two views of the “directional” information vs. the “undirectional” noise . . . . .	32
3.4	The dependency model $y = f(x)$ refined (PCR) . . . . .	36
3.5	The dependency model $y = f(x)$ refined (PLS) . . . . .	38
4.1	Applications for electroless nickel plating . . . . .	41
4.2	Comparison of electroless and immersion (active power supply) deposition . . . . .	42
4.3	The cross-section of a plated test plate taken with a scanning microscope . . . . .	44
4.4	Two tanks for electroless nickel reaction . . . . .	46

---

4.5	Schematic diagram of the Ni-controller . . . . .	47
4.6	Functionality of a nickel plating model . . . . .	50
5.1	Measured and predicted thickness of the Ni-P alloy . . . . .	61
5.2	Measured and predicted phosphorus content of the Ni-P alloy . . . . .	61
5.3	Measured (M) and predicted (P) hypo- and orthophosphite concentrations	62
6.1	Weighting functions to smoothen values during the plating process . . . . .	67
6.2	Raw process data before synchronization . . . . .	68
6.3	Schematic view of the synchronized data sheet . . . . .	70
6.4	Latent vectors $\theta_i$ (PCA) and the corresponding eigenvalues . . . . .	73
6.5	Alloy thickness: measured data and two compared estimates . . . . .	79
6.6	Alloy phosphorus content: measured data and two compared estimates . .	80
6.7	Hypophosphite concentration: measured data and two compared estimates	81
6.8	Orthophosphite concentration: measured data and two compared estimates	82
6.9	Latent vectors $\theta_i$ (PLS) and the corresponding eigenvalues . . . . .	84
6.10	Alloy thickness: measured data and two compared estimates . . . . .	87
6.11	Alloy phosphorus content: measured data and two compared estimates . .	88
6.12	Hypophosphite concentration: measured data and two compared estimates	89
6.13	Orthophosphite concentration: measured data and two compared estimates	90

”The best material model of a cat is another,  
or preferably the same, cat.”

Norbert Wiener and Arturo Rosenblueth  
*Philosophy of Science* 1945

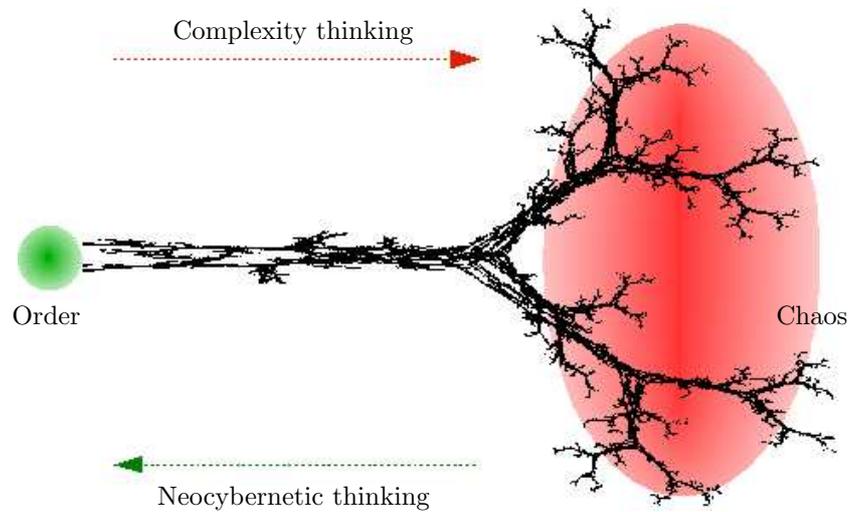
# 1

## Introduction

The Greek philosopher Heraclitus (about 540 B.C.) stated that there is no permanent reality except the reality of change. He was one of the first to figure out the deepest intuitions concerning complex real-life systems:

- *Everything changes, everything remains the same.* Cells are replaced in an organ, material is replaced in a production site — still the functionality and essence remain the same.
- *Everything is based on hidden tensions.* Species compete in ecology, companies in economy — the tension results in balance and diversity.
- *Everything is steered by all other things.* There is no centralized control in the body or in economy — but the interactions result in self-regulation.

Increasingly complex systems are studied because one is increasingly interested in the essence behind complex natural and industrial systems. The requirement to know about nature and life is continuously increasing, and so is the complexity of systems all around



**Figure 1.1:** Complexity theory pessimism vs. neocybernetic optimism

us. Facing these new challenges, the traditional way of understanding complex systems comes to a halt, creating a pessimism about the modeling of complex systems.

A new approach, derived from cybernetics and combined with the power of simple mathematical tools, tackles this dead end and opens new possibilities to regain optimism about comprehension of complex systems. Cybernetics is a special way to look at complex systems and *neocybernetics* is a special way to look at *cybernetic systems*, which makes modeling and understanding easier.

Figure 1.1 illustrates the characterization of complexity theory and neocybernetics among complex structures and systems.

In the present work this new approach is explained and applied to an industrial system. First the details, ideas and advantages of neocybernetics are worked out to understand the motivation behind it. After that the industrial setup, in which the approach will be applied, is introduced and its complex structure presented. The complexity of the process is emphasized by describing the existing models and recent attempts to obtain an accurate model for control purposes. The recently developed model will also act as reference for

---

the model designed in the thesis at hand and therefore be elaborated in more detail. To tie these loose ends together the neocybernetic ideas and its tools will be applied to the process in order to gain a useful model and the results will be presented. A discussion will help to fit the work into the big picture and future prospects with possible and necessary tasks will be outlined.

Neocybernetics grabs the readers attention and the interesting framework provides a base for further work. It will be demonstrated that the thoughts and assumptions hold and meet good and useful results.

## 2

# Neocybernetics

In order to understand the motivation to deal with neocybernetics and its advantages to system understanding, this section introduces the ideas of cybernetics, starting out with the original ideas of the first cyberneticists and coming to the modern uses of the cybernetic thoughts.

## 2.1 The Origin of Cybernetics

Ancient Greek is the language used by the first scientists on this world, who at that time were mostly philosophers rather than scientists. They used it to think about, discuss and write down their immense intuition about nature and science. Modern philosophy, politics and science are based on these ideas written down in ancient Greek.

*Κυβερνήτης* (the ancient Greek word for a steersman, governor, pilot or rudder) was first used by Norbert Wiener in his book “Cybernetics or Control and Communication in the Animal and the Machine” [1] to introduce his new field of science. As mathematician he started to break new ground in robotics, computer control, automation and dynamics,

---

originating from his work about gunnery during World War II. At the intersection of network theory, logic modeling and neurology a field of study for “teleological mechanisms” was popularized. Cybernetics emerged apart from, but touched upon, such established disciplines as electrical engineering, mathematics, biology, neurophysiology, anthropology, and psychology, and hence was supposed to be not only multidisciplinary but rather metadisciplinary, using new terms and methods to describe system behavior and provide tools in order to steer these systems.

Pushed also by Arturo Rosenblueth and Julian Bigelow [2], cybernetics was established among the notable sciences as the science and study of systems and control in an abstracted sense. The emphasis is on the *functional relations* that hold between the different parts of a system, rather than the parts themselves. These relations especially include the transfer of information and circular relations. The main innovation brought about by cybernetics is an understanding of goal-directedness or purpose as resulting from a negative feedback loop which minimizes the deviation between the perceived situation and the desired situation.

With these ideas cybernetics set the fundamentals of control engineering and systems theory. It was suddenly possible to step back from actual matters and problems, implement the complex behavior of any given dynamic structure as an abstract system and figure out the performance according to changes in the flow of information at the borderlines of the system. As said above this can be applied to tasks in any given domain, nontechnical or technical.

Along with new ideas about abstract systems the scientists used already well known mathematics as a tool to tackle the problem of an abstract representation of real life nature. This led to the area of modeling and simulation. Powerful tools for mathematical modeling of dynamic systems were found and arranged, to make a “copy” of the analyzed original system, which should represent the original in appropriate ways. Eventually,

---

holding a good model of an original system, tests and analyses can be carried out in order to understand the interior of a considered system and to understand the reactions to applied signals from the exterior. This finally leads to the implementation of a controller if the goal is the manipulation of particular information in the system or at its border.

A lot of research was done about appropriate modeling. Considering the initiatory quotation of Norbert Wiener and Arturo Rosenblueth, one can assume that this is quite a difficult discussion, which light can be cast on from many different angles. Generally speaking, one could say that the model should not only be as precise as possible, but also as simple as possible. To find a good way through this conflict of aims, people often use a lot of their own ideas, preknowledge and information, which always already affect the resulting model and therefore also simulation results gained while using this model.

In general one wants to estimate the behavior between given inputs and outputs of a system as accurately as possible. Using mathematical tools the model will be a set of equations containing different variables. If one can use some a priori information, one can more easily adjust the equations to fit this particular knowledge and match the known behavior to the behavior of the system model. Said a priori knowledge is mostly based on known dynamics in the system (leading to ordinary differential equations, ODE) and known algebraic constraints (leading to differential algebraic equations, DAE).

Using this “traditional” way of modeling the structure of the considered system has to be known before one starts with the actual modeling approach. To develop the model one also needs an expert on the domain area, in which the studied system is embedded, to get help for the setup of basic equations, describing the inner structure of the behaviour of a system. Expert knowledge is also needed to control and supervise the process of modelling. However, naturally the structures of contemporary man-made systems or natural environments is too complex to be known beforehand and hidden, which causes immense problems for the modeling approach.

---

At this antagonism one can locate the starting points of neocybernetics, to which the next section will give some more substance.

## 2.2 To a New Approach of Cybernetics

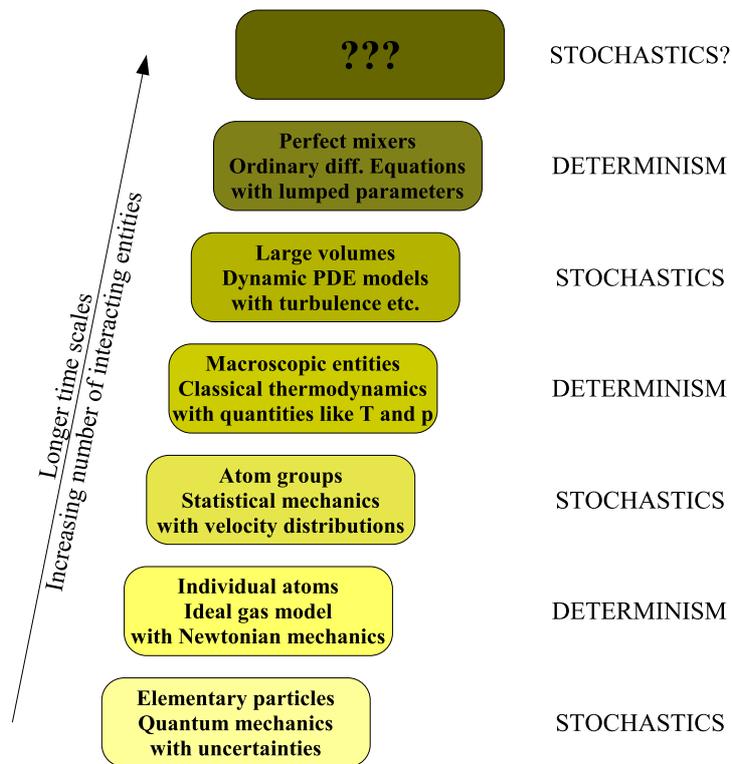
As explained above cybernetics is a special way to look abstractly at complex systems and their dynamics and feedbacks. Further, neocybernetics is a way to look at cybernetic systems. It combines mathematical compactness and expressional power in a consistent framework.

The following principles and ideas can be used as guidelines into the framework of neocybernetics. Complex looking phenomena are seen through neocybernetic eye-glasses, becoming obviously advantageous as a result of iterative refinement processes. The presentation here follows [3].

### **Emergence**

The key concept and also the most important basis of neocybernetics is the idea of *emergence*. Some qualitatively new and unanticipated functionality appears from the complex system after cumulation of simple operations. There is no need to reduce the analysis of higher-level phenomena to their components, like it is done for the traditional reductionistic modeling approaches described above.

To formalize the idea of emergence one can study examples and construct an intuitive understanding of it and after that find common features and cast them in a mathematical framework. To demonstrate different levels of emergence consider Figure 2.1, where emergence takes place between each level and concepts, variables and structures change altogether. In this case the domain is the physics of particle movement.



**Figure 2.1:** Different levels of abstraction while modeling interaction of particles [3]

- First, it seems that stochastic and deterministic expressions for each level are alternating. This is reasonable, because, for example, two successive deterministic levels could easily be merged to a single one.
- Second, it seems that towards higher levels the volumes and time scales increase.

A higher, again stochastic, level of abstraction on top of the presented hierarchy becomes necessary in systems consisting of many ideal mixers, like industrial processes, for example. One can expect another emergence of a hidden behaviour as happened between the lower levels, if performing the step to the next higher level.

---

## Dynamic Balance

The emphasis in cybernetic systems is on the balance rather than on the process itself. In that steady state it is possible to attack the emergent pattern and forget about strongly nonlinear interior processes. Of course this may be a dynamic equilibrium of tensions caused by negative feedback loops, as long as these internal loops can maintain the overall stability. In Section 2.3 the balance in cybernetic systems is discussed and ingrained in mathematics.

But are not most complex systems unstable in at least one mode? How can one assume stability in natural processes? — The motivation is quite simple: If a natural system were unstable, it would have collapsed and ended in explosion or extinction long ago. No one is trying to model all mathematical possible systems — only physically meaningful ones!

## Environment

Neocybernetic systems are assumed to be oriented towards their environment. The underlying idea is that there cannot be a cybernetic system in isolation, thus causing the traditional system theory thinking to collapse. This traditional isolation indeed allows communication with the exterior, but still the system has a well-defined isolating border. A cybernetic system does not only interact with the environment, exchanging information or material, it even more reflects the environment, somehow capturing and mirroring its environment. Recalling Heraclitus' statements one will find the thought of everything carrying its opposite and mirroring its outside. Environment-orientedness gives another motivation for emphasis on balance: only in stable conditions, when fast turbulent phenomena have ceased, something new and fragile can emerge. For mathematics behind environment-orientedness see also Section 2.3.

---

## **Dimensionality**

Using neocybernetic ideas in practice turns environment-orientedness into data-orientedness. It cannot be assumed that the system and its environment are a priori known, but only measurements are available. Defining features out of the environmental information creates a high dimensional complexity instead of the structural complexity according to traditional thinking. All possibly relevant features should be simultaneously captured and made available to the modeling machinery, hoping that it constructs connections among these pieces of data, finding the essence in it. Typically the data is highly redundant, which means that multivariate methods and appropriate mathematical tools are necessary to analyze the data in order to extract neocybernetic models.

Of course, it is only possible to gain the essence out of the given features, if they are actually hidden in the used data. The domain-area semantics should somehow be coded in the measured information.

## **Linearity**

In neocybernetics, the starting point is the assumption of linearity. Generally the assumption can be used to some extent, or it can be ignored, depending on how arguments are favored and how open minded one is to apply these assumptions. However, there are a couple of reasons to assume linearity in cybernetic systems:

- It is well known, that feedbacks “smoothen” nonlinearities. Especially in neocybernetic systems, where balance is emphasized, deviations around the equilibrium can be assumed to be small, leading to a linear approximation. Only systems which are in balance with their environment are studied. Natural systems fulfill this assumption and one would like industrial systems to fulfill it also. Further successful

---

controls nowadays keep the system near its setpoint in balance, regardless of strong environmental disturbances.

- There is also a pragmatic motivation for the linearity assumption. Strong mathematical analysis tools are available for linear systems, no matter what the system dimension is. So it makes sense to avoid nonlinearity as long as it is reasonable, and only introducing nonlinearity if it is absolutely necessary.

Linearity is more like a guiding principle, to be followed as long as possible. Only when there is no other solution one should consider applying nonlinearities. Of course the assumption then can be relaxed, but this should be done only after the basic nature of the cybernetic system is captured.

In the following section the key ideas of neocybernetics will be studied including a mathematical framework, in order to build an understandable background which is not exclusively based upon intuition.

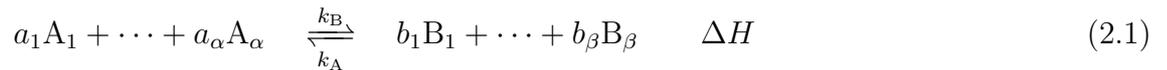
## 2.3 With Mathematics Towards Elasticity

When modeling complex systems the structure is hidden. The objective is *automatic abstraction*, letting the structures automatically emerge. Statistical tools naturally carry out abstraction and give significance only to phenomena that remain consistent over long-term periods. To apply multivariate methods, one has to concentrate on the (thermo)dynamic balances. As it turns out, the domain of chemical systems offers a compact framework for such studies.

---

### 2.3.1 Modeling Chemical Systems

Study a hypothetical example reaction, where there are  $\alpha$  reactants on the left hand side ( $A_i, 1 \leq i \leq \alpha$ ), and the  $\beta$  products on the right hand side ( $B_j, 1 \leq j \leq \beta$ ):



with  $k_A$  and  $k_B$  being reaction speed in forward and backward direction. The symbol  $\Delta H$  denotes the change in enthalpy during the reaction.

Applying the neocybernetic principles, one can “cybernetize” chemical reaction models to achieve a mathematically more compact representation than (2.1).

#### Information Representation

To model complex systems, consisting of various reactions, the data representation has to be extended and different vectors have to be embedded in the same vector space to make them compatible.

The most compact mathematical representation of information is the vector form. Define a vector  $C$  containing all chemical concentrations so that all  $A_i$  and  $B_j$  are represented there. The chemical state can be captured in this vector and individual reactions determine equations in that chemical space. If the coefficients  $-a_i$  and  $b_j$  from (2.1) are collected in the vector  $\mathbf{G}$ , one can express the concentration changes as

$$\Delta C = \mathbf{G}x. \quad (2.2)$$

The scalar  $x$  reveals, *how much* and *in which direction* the reaction has proceeded. The expression (2.2) can be extended to multiple simultaneous reactions, when  $G$  is a matrix containing the individual reaction vectors as columns, and  $\mathbf{x}$  is a vector. If one knows the reaction rates (or the scalars  $x_i$ ), the changes in the chemical contents can be determined.

---

However, the reaction rates and, what is more, the reactions are typically not known. Also (2.2) is not yet what one is looking for. It only captures the *stoichiometric*, more or less *formal balance* among chemicals and does not capture the dynamic balance. One has to study the reaction mechanism closer.

### Thermodynamic Balance

According to (2.1) it takes  $a_1$  molecules of  $A_1$  and so on for one unit reaction to take place. These molecules have to be sufficiently near to each other and the probability for that is proportional to the number of such molecules in a volume unit. This molecular density is revealed by concentration (mole/liter). Assuming that the locations of molecules are independent from each other, the probability for them being found within the range is proportional to the product of their concentrations. Because, in addition, the reverse reaction according to the molecules of the right hand side of (2.1) takes place, the rate of change for the concentration of the chemical  $A_i$ , for example, can be expressed as

$$\frac{dC_{A_1}}{dt} = -k_B C_{A_1}^{a_1} \dots C_{A_\alpha}^{a_\alpha} + k_A C_{B_1}^{b_1} \dots C_{B_\beta}^{b_\beta}. \quad (2.3)$$

In equilibrium there holds  $dC_{A_1}/dt = 0$ , and one can define a constant characterizing the thermodynamic equilibrium [3]

$$K = \frac{k_B}{k_A} = \frac{C_{B_1}^{b_1} \dots C_{B_\beta}^{b_\beta}}{C_{A_1}^{a_1} \dots C_{A_\alpha}^{a_\alpha}}. \quad (2.4)$$

### Linearity Objective

One of the neocybernetic objectives is that of linearity (see Section 2.2). The expression (2.4) is far from linear — indeed, it is purely multiplicative. The linearity can be reached with simple syntactic tricks. Taking logarithms on both sides there holds

$$\log K' = b_1 \log C_{B_1} + \dots + b_\beta \log C_{B_\beta} - a_1 \log C_{A_1} + \dots - a_\alpha \log C_{A_\alpha}. \quad (2.5)$$

---

To get rid of constants and logarithms, it is also possible to differentiate the expression:

$$0 = b_1 \frac{\Delta C_{B_1}}{\bar{C}_{B_1}} + \dots + b_\beta \frac{\Delta C_{B_\beta}}{\bar{C}_{B_\beta}} - a_1 \frac{\Delta C_{A_1}}{\bar{C}_{A_1}} + \dots - a_\alpha \frac{\Delta C_{A_\alpha}}{\bar{C}_{A_\alpha}}, \quad (2.6)$$

where  $\Delta C_i/\bar{C}_i$  are deviations from nominal values, divided by those nominal values, meaning that it is *relative changes* that are of interest. The differentiated model is only valid in the vicinity of the nominal value.

Assume that vector  $\mathbf{z}$  contains all relevant variables including relative changes in all chemical concentrations. This means that the vector  $\mathbf{\Gamma}_i$  represents a single reaction, and it can contain various zeros if the corresponding chemicals are not contributing the reaction  $i$ . Collecting the vectors  $\mathbf{\Gamma}_i$  as columns of the matrix  $\mathbf{\Gamma}$  one can write

$$0 = \mathbf{\Gamma}^T \mathbf{z}, \quad (2.7)$$

where one row now represents one chemical reaction.

The key point to observe here is that analysis of complicated reaction networks can be avoided, one only needs to study the levels of concentrations, not changes in them. No matter what has caused the chemical levels, only the *prevailing tensions* in the system are of essence. Nothing mathematically special is being done, but when seen from the appropriate point of view, new conceptual tools for modeling of complex systems can be available.

The following subsection will give an insight of these new views, which were also previously mentioned in the introducing sections of this chapter.

### 2.3.2 Constraints vs. Degrees of Freedom

It can be claimed that the *freedom-oriented* way of modeling is just as natural as the *constraints-oriented* approach. A closer look is needed to understand the meaning of this idea.

---

Traditional models are based on *constraints*. The system properties and the connection between interior variables are captured by a set of equations of the general form

$$0 = f(\mathbf{z}), \tag{2.8}$$

where  $f$  is some function (scalar or vector-valued) of the variables  $\mathbf{z}$ . The chemical model in (2.7) consists of various independent equations or constraints in matrix form. Even the generally known form of linear models  $\mathbf{y} = F^T \mathbf{x}$  can be cast in that framework: Written as  $0 = F^T \mathbf{x} - I \mathbf{y}$  and using the definitions

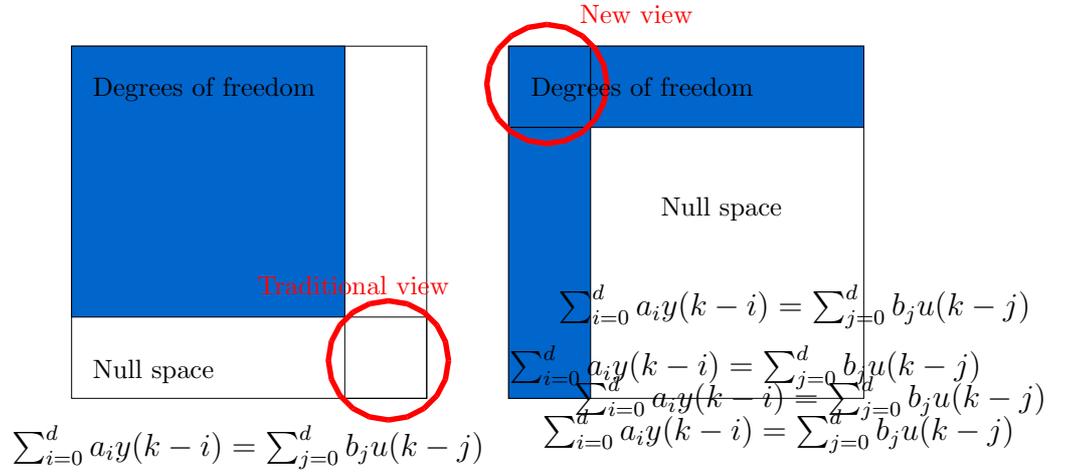
$$\Gamma = \begin{pmatrix} F \\ -I \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \tag{2.9}$$

one achieves the form of (2.7). Note that such models are not unique, since  $\Gamma$  can be scaled freely without changing the validity of the equations.

To better understand the structure of models presented in the constraints-oriented form, study a single-output system, where  $y_i$  is a scalar and  $\mathbf{\Gamma}_i$  is a vector.  $y_i = \mathbf{\Gamma}_i^T \mathbf{x}$  defines an *one-dimensional null-space* in the high-dimensional space of  $\mathbf{z}$  and because the inner product  $\mathbf{\Gamma}_i^T \mathbf{z}$  between the data and the vector  $\mathbf{\Gamma}_i^T$  equals zero, this vector defines a vector that is *orthogonal to this subspace*. A mathematical example in a three-dimensional data space to visualize the freedoms-oriented model structures can be found in Appendix A.

Mathematically speaking, if there are  $\mu$  separate variables, there are  $\mu$  degrees of freedom in the data space, but each (linear) constraint, describing the coupling between variables, decreases the number of degrees of freedom by one. Hence  $\nu$  constraints leave  $\mu - \nu$  degrees of freedom. In the directions of the constraints there is no variability, constituting a null-space within the data space — all the remaining variation is concentrated in  $\mu - \nu$  degrees of freedom.

The key point here is that essentially the same dependencies among variables can be captured in terms of degrees of freedom as with constraints. When the number of constraints increases, the most economical representation changes: the simplest model with



**Figure 2.2:** Schematic illustration of the covariance structure among data when there are few constraints (on the left) and when there are many constraints (on the right)

the least parameters is not the constraint-oriented model but the freedoms-oriented model. Whereas the constraint-oriented modeling approach becomes an unmanageable mess, the freedoms-oriented models become clearer as the data dimension increases. The higher the number of variables is, the more appropriate the pattern-based representation seems to become. Furthermore it turns out that freedoms-oriented models are *more intuitive* than the constraints-oriented models.

Figure 2.2 shows schematically why said change of the system representation makes sense if the number of constraints increases. It becomes clear that on the right hand side in a dataset with many constraints the traditional view will end up in a model of a complex set of equations, while the complexity of the new view is only dependent of the few degrees of freedom.

The *emergence*, which was already discussed and introduced in Section 2.2 among other ideas of neocybernetics, appears here in a more mathematical framework. The example in Appendix A shows the emergence of an exponential behaviour which was not seen before if only looking at the given constraints. Generally speaking there are the multivariate statistical methods that directly attack the degrees of freedom and emergence of

---

behaviour, abstracting away structural details. Thus those methods will be introduced in the following subsection.

### 2.3.3 Multivariate Methods

The data, that is collected from some process, can be preprocessed in different ways, but the preprocessing should be done very carefully, because it affects the resulting model immensely. A standard approach is mean-centering, where the subtraction of the mean in the data makes this mean equal to zero, and normalization to unit variance by dividing the data by its variance, in order to make the variation in different variables equally “visible”. The used way of data preprocessing will be specified in Section 6.1. Whatever the data processing steps, the original data  $\mathbf{z}$  after preprocessing will be denoted as  $\zeta$ .

Since all measurement values are inaccurate, the assumption of (2.7) cannot hold and must be relaxed, and one has to extend the original model to

$$\mathbf{e} = \Gamma^T \mathbf{z}, \tag{2.10}$$

when  $\mathbf{e}$  represents the error. Traditionally one wants to minimize the sum of squared errors over a set of measurement data:

$$\Gamma = \arg \min_{\Gamma} \{E\{\mathbf{e}^T \mathbf{e}\}\}. \tag{2.11}$$

To avoid the trivial solutions of  $\Gamma_i = 0$  and get a well-conditioned optimization task, the additional restriction of  $|\Gamma_i| = 1$  for all  $i$  is needed. When searching for the freedoms instead of the constraints, the objective is exactly opposite:

$$\varphi = \arg \max_{\varphi} \{E\{\xi^T \xi\}\}, \tag{2.12}$$

again with the restriction of  $|\varphi_i| = 1$  for all  $i$ . Here,  $\varphi$  and  $\xi$  are used to emphasize the difference to  $\Gamma$  and  $\mathbf{e}$ . Defining  $\xi = \varphi^T \zeta$  instead of (2.10), it is now the “error”  $\xi$  that is to be maximized and  $\varphi$  is the axis along which this maximum variation in data is reached.

---

Applying the objective (2.12), it is assumed that variation in data represents information, whereas traditionally this variation is only seen as noise. And, specifically, it is *covariation* among variables that carries information: covariations reveal the underlying *common causes* that are reflected in measurements.

The solution for the problem (2.12) is given by *Principal Component Analysis*, or *PCA* (see [4], [5], [6], etc). The following Chapter 3 will give more details to PCA and even more useful tools, but the basic results can be summarized as follows:

The degrees of freedom can be analyzed studying the covariance matrix  $E\{\zeta\zeta^T\}$ . The variability is distributed in the data space along the eigenvector directions of this matrix, variance in the eigenvector direction  $\theta_i$  given by the corresponding eigenvalue  $\lambda_i$ :

$$E\{\zeta\zeta^T\}\theta_i = \lambda_i\theta_i. \quad (2.13)$$

It turns out that the eigenvectors are orthogonal, so that the principal component directions can be used as well-conditioned subspace basis vectors in a mathematically efficient way. If the variables are selected appropriately, there is no reason why a mathematical machinery could not capture the same phenomena that are followed by biological or chemical machinery. This strong assumption will be confirmed by the thesis at hand.

### 2.3.4 Restructuring Data

The formulation of the multivariate structured data in (2.7) must again be studied and extended to get a better feeling about the variables in and around the system and to make the next step in neocybernetic thoughts. The vector  $\mathbf{z}$  in (2.7) is divided in two parts.

$$\mathbf{z} = \begin{pmatrix} \mathbf{u} \\ \bar{\mathbf{x}} \end{pmatrix}.$$

---

Vector  $\mathbf{u}$  of dimension  $m$  describes the *environmental* conditions, and vector  $\bar{\mathbf{x}}$  of dimension  $n$  contains system specific *internal* variables, characterizing the equilibrium state of the system. Distinguishing between the two different parts, assume that one can rewrite the characterizing constraints of (2.7) to

$$A\bar{\mathbf{x}} = B\mathbf{u}. \quad (2.14)$$

Having the same number of constraints as internal variables makes the internal system well-defined and the matrix  $A$  square. According to environment-orientedness one can assume a linear dependency between  $\bar{\mathbf{x}}$  and  $\mathbf{u}$ . Assuming that  $A$  is invertible, one can explicitly solve the unique linear function from the variables of the environment into the system state

$$\bar{\mathbf{x}} = A^{-1}B\mathbf{u}, \quad (2.15)$$

and so define an explicit mapping matrix from  $\mathbf{u}$  to  $\bar{\mathbf{x}}$  as

$$\phi^T = A^{-1}B. \quad (2.16)$$

However, the main motivation for the formulation (2.14) is that one can formally extend the static model into a dynamic formulation. Assuming that the data structures are selected appropriately, so that  $-A$  is stable (eigenvalues having only negative real parts), one can define

$$\frac{d\mathbf{x}}{\gamma dt} = -A\mathbf{x} + B\mathbf{u}. \quad (2.17)$$

The parameter  $\gamma$  can be used to adjust the time scale. The steady state of (2.17) equals (2.15), so that  $\lim_{t \rightarrow \infty} \mathbf{x} = \bar{\mathbf{x}}$  for constant  $\mathbf{u}$ . Because of linearity, this steady state is unique.

The extension to this dynamic model is justified, because there *must* exist such an inner structure beyond the surface. The seemingly static dependencies of (2.7) have to be

---

basically dynamic equilibria systems (see 2.17) so that the equality in the equation can be restored after any disturbance. The actors, or molecules in case of a chemical model, do not know the “big picture” of the whole reaction system, and it is the interactions between the molecules that provide for the tensions resulting in the tendency towards balance.

Thinking of the “mindless” actors in the system, the only reasonable explanation for the common distributed behaviours is *diffusion*. It is the concentration gradients that only are visible at the local scale of a chemical system. Interpreting (2.17) as a (negative) gradient, there has to exist a criterion which is being minimized. By integration of (2.17) with respect to  $\mathbf{x}$  it is found that

$$\mathcal{J}(\mathbf{x}, \mathbf{u}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{x}^T B \mathbf{u} \quad (2.18)$$

gives a mathematical “pattern” that also characterizes the system. Such an optimization-oriented system view combines the two ways of seeing systems: the criterion itself represents the *pattern view*, whereas the optimization process represents the *process view* (see [3] and [7]).

Looking somewhat closer at (2.15), one can conclude that it is not only the environment changing the system by having a mapping  $\phi^T$  into the system, but there has also to be an inverse mapping. This inverse mapping  $\varphi$  characterizes the effect of the internal system variables  $\mathbf{x}$  to their environment  $\mathbf{u}$ . Taking the starting point (2.7) into account, it becomes clear that every variable may influence every other variable, so there are clearly both directions of effect possible. One can speak of *pancausality*, or even of a *holistic view*. Even Herclitus stated, that “all is one”.

This two way assumption blurs the traditional way of distinguishing clearly between a system and its environment, creating a strict system border. Here no clear distinction between the system itself and its environment can be seen. Further, the environment is a combination of cybernetic systems and subsystems, so the vector  $\mathbf{u}$  represents the net

---

effect of all accompanying systems and subsystems. The deep connection between the mappings  $\varphi$  and  $\phi^T$  is a key issue when trying to capture the behaviours of cybernetic systems.

## Elastic Systems

Using the above presented mathematical background and ideas, one can assume that cybernetic systems are *elastic systems*. Considering the mechanical domain for an example, the following setup can be used to match the above equations. The vector  $\mathbf{u}$  could denote *forces* acting in a mechanical system (like a spring), and  $\mathbf{x}$  denotes the resulting *deformations*. Further,  $A$  is interpreted as the *elasticity matrix* and  $B$  is the *projection matrix*, mapping the forces onto the deformation axis.  $A$  must be symmetric and must be positive definite to represent stable structures sustaining external stresses.

Then it turns out, that (2.18) is the difference between the potential energies stored in the mechanical system. The *principle of minimum potential (deformation) energy* states that a structure under pressure ends in minimum of this criterion, trying to exhaust the external force with minimum of internal deformations. For example, a marble placed in a bowl will move to the bottom and rest there or a tree branch full of snow will bend down into a new position because of this new external force. These are stable configurations, so called equilibria [8]. Hence, cybernetic systems are elastic and always end up in a state of balance. For a mechanical system of a steel plate these phenomena were studied in the neocybernetic framework by Sailer [9].

The same criterion can be seen to characterize all cybernetic balance systems, also from different domains than the mechanical one. It does not matter what the domain is, and what the physical interpretation of the “forces”  $\mathbf{u}$  and of the “deformations”  $\mathbf{x}$  is, the structure of system behaviour remains intact. A pressed system bends away, but when the pressure is released, the original state is restored. Indeed, in chemical environments,

---

this principle is known as the *Le Chatelier principle*: if a chemical system at equilibrium experiences a change in concentration, temperature or total pressure the equilibrium will shift in order to minimize that change [10].

To summarize the above ideas and observations:

Neocybernetic systems are identical with *elastic systems* — systems that are characterized by dynamic equilibria rather than static equivalences.

### 2.3.5 Neocybernetic Perspective

The following discussions are only relevant if thinking about biologically adapting systems. Nevertheless they are part of the neocybernetic theory and therefore shortly presented.

The effect of environmental pressures on a system can easily be quantified. Just as in the case of a potential field, it is the product of the force and the displacement that determines the change in potential energy. Regardless of physical units one can similarly interpret the product  $\bar{x}_i u_j$  in terms of *energy transferred from the environment into the system* by this particular pair of variables. It must be remembered that there is not only the effect from the external environment into the internal system — there is symmetric interaction that takes place. It is the matrices  $\phi^T$  and  $\varphi$  that characterize the energy transfers. It is not only so that  $\mathbf{u}$  should be seen as the “force” and  $\bar{\mathbf{x}}$  as the “effect”, but  $\bar{\mathbf{x}}$  can be seen as action and  $\mathbf{u}$  as reaction as well.

The transferred energies are also effectively divided by time, so that it is some kind of power that is transferred. Hyötyniemi [3] names this “emergent level energy” as *emergy* and it works as “information energy”, which is the prerequisite for emergence of information structures.

---

## Goal of Evolution

Going on in this line of ideas and argumentation, one can step to evolution theory and discuss about its goals. If the fitness criterion for evolution were the “maximum number of offspring”, there would be only bacteria on earth [3]. On the other hand, the “blind watchmaker” [11] theory with only random optimization simply cannot be the mechanism beyond evolution.

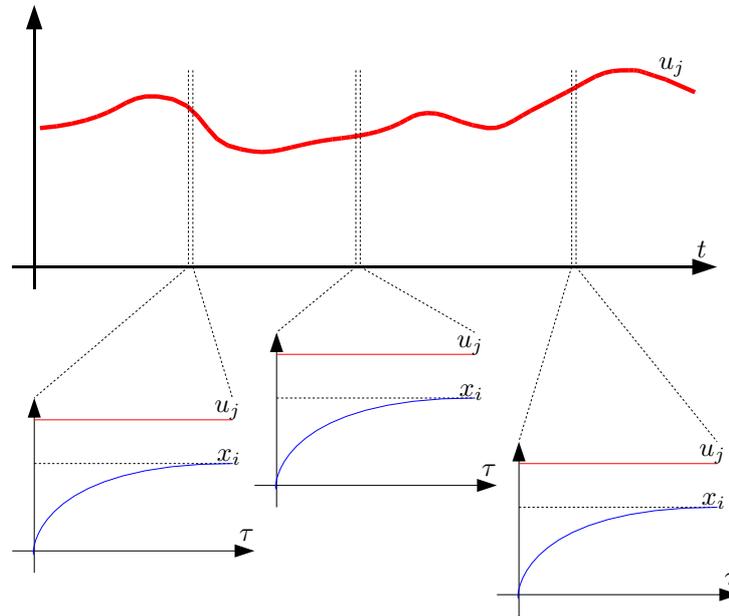
Neocybernetic environment-orientedness suggests the idea of a kind of *best match with environment*. Applying the above discussion about energy transfer, an intuitively appealing fitness criterion would be

Maximize the average amount of energy that is being transferred between the system and the environment [3].

No matter what the physical manifestation of the environmental variables is in any special case, a surviving system interprets them as resources, and uses them as efficiently as possible. It is not predetermined what should be done with the extracted energy — there are various options. This makes it possible, that the evolutionary process proceeds in many different ways. The relevance of the changes and behaviour is later evaluated by the evolutionary selection.

All changes among this process typically affect all elements of the mappings  $\phi^T$  and  $\varphi$  — but all of them only a little. This high number of discrete parameters is more or less projected to low dimensional “emergent patterns”. What is more, the *local* optimizations are independent of each other — this makes the optimization a parallel process, relatively fast and robust (see also [9]). The time scale in this stochastic optimization is much longer than in (2.17).

It is reasonable to assume that the process of adaption finds a fixed state [3], but it needs to be recognized that the adaption of the system is completely local for any element



**Figure 2.3:** Schematic illustration of two different time scales. The dynamics of  $u$  (scale  $t$ ) are much slower than the dynamics of  $x$  (scale  $\tau$ )

in the matrices  $\phi^T$  and  $\varphi$  even though the assumed goal of the process is evolutionary improvement and presented in a collective matrix format. The net effects can still be far from trivial.

### Complex Structures

Above, the balances of  $\mathbf{x}$  were studied among a fixed environment  $\mathbf{u}$ . To reach interesting results, the neocybernetic principles can be exploited. Specifically, see Figure 2.3. The environment  $\mathbf{u}$  changes now on the wider scale, denoted by  $t$ , but stays also stationary and behaves like a system around its dynamic equilibrium. On the system time scale  $\tau$ , which is far faster than the environment's scale, the balance is restored quickly, assuming the environment to be fixed for a short moment — a “balance model of balances”. A truly cybernetic model is a *second-order balance model*, or even a *higher-order balance model*.

As an overall conclusion it can be observed, that within the neocybernetic framework

---

local learning has globally meaningful results. From the functional point of view new interpretations for cybernetic systems are available:

**First-order cybernetic system** finds balance under external pressure, pressures being compensated by internal tension. It implements *minimum (observed) deformation energy* in the system.

**Second-order cybernetic system** adapts internal structures to better match the observed environmental pressures, towards maximum experienced stiffness. It implements *minimum average observed deformation energy* in the system.

**Higher-order cybernetic system** adapts the external structures of the system to better match the observed environmental structures. Evolutionary optimal environment, or system of systems, only contains higher-order cybernetic systems. It implements *maximum average transfer of energy* through the environment.

As it is shown in [3] and [9], local maximization of transferred energy result in global results: As a whole, the system spans a principal subspace of the resource variations, thus optimizing exploitation.

This subsection gave an outlook on the neocybernetic line of thought, in order to give an idea of the powerful intuitions that are behind cybernetic sytem thinking. However, the system considered in the thesis at hand is more simple and ranks in the field of first-order cybernetic systems.

With this background about the cybernetic history and about the neocybernetic ideas with its anticipations the theoretical framework of this thesis is given. Also the strong motivation to apply these ideas to real systems and expect good results is pointed out. The corresponding mathematical tools which are used later to apply these ideas will be described in the following Chapter 3.

# 3

## Multivariate Analysis

In the previous chapter it became clear that multivariate tools are the ideal instrument to approach the problems addressed by neocybernetics. Hence this chapter introduces the later applied methods and gives examples to understand their functionality.

Before one can apply mathematical tools to find patterns in a huge set of variables, the data set has to be preprocessed. The way of preprocessing will be described and afterwards the tools of analysis will be shown.

### 3.1 Data Preprocessing

The role of variable scaling during data preprocessing is to make the relevant features optimally visible in the data. Study an example [6]: assume that there are temperature values in the range of 100 °C to 200 °C and associated pressure values in the range from 100000 Pa to 200000 Pa. The variation range (ignoring units) in temperature is 100 and in pressure 100000. In the mathematical analysis the role of temperature will be neglected because the variation range is so narrow if variations are emphasized as the later described tools do.

---

Assume there is a set of two variables, which should be analyzed. Figure 3.1 shows the way from the original data distribution, sketched with an oval in the two-dimensional dataspace. Calculating the means of all (two) directions and subtracting them from the data values centers the distribution around the origin. Furthermore, normalization to unit variance carries out the actual step of information “equalization” as described in the example above. Another step could be data whitening, which removes all covariance from the data vectors. So if  $\mathbf{x}$  would be centered data, and  $\tilde{\mathbf{x}}$  the according whitened vector, the covariance matrix of  $\tilde{\mathbf{x}}$  equals the identity matrix:

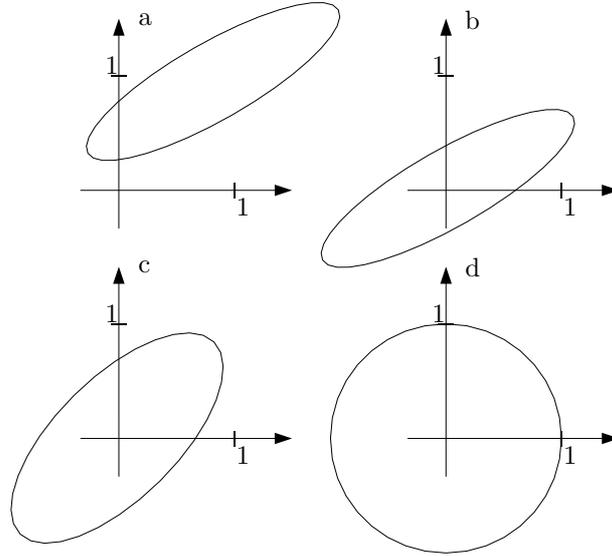
$$\mathbb{E}\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I}. \quad (3.1)$$

For the data used later in Section 6, only the basic and necessary operations of centering and normalization (corresponding to (b) and (c) in Figure 3.1) were carried out, to achieve a good starting point for the multivariate tools on one hand, but to keep the data as original as possible on the other hand. The algorithms that are introduced later in this chapter can not find any structure or pattern in whitened data, since they are based on the assumption that all data information can be found in its covariance [6].

## 3.2 Principal Component Analysis

How to handle now this high-dimensional dataspace and to find the underlying behaviour among the variables? There are multivariate tools, which are very powerful and help to find solutions for this problem. *Principal Component Analysis* or *PCA*, as already mentioned in Subsection 2.3.3, can find patterns in a large dataspace, highlight differences and similarities in it and give a structured view of the freedoms in the data. Whereas noise is (assumed to be) purely random, consistent correlations between variables hopefully reveal something about the real system structure [4], [6], [12].

The other main advantage of PCA is that once one has found these patterns in the data



**Figure 3.1:** Data distributions after different preprocessing operations. The *original* distribution (a) is first *centered* (b), then *normalized* (c) and eventually *whitenened* (d)

one can compress the data, for example, by reducing the number of dimensions without losing much information. This technique is also often used in image compression, where one has to deal with high-dimensional data.

Assume that  $\theta_i$  is the direction of maximum variance one is searching for in the dataset  $X$  with  $k$  measurement samples. Points in  $X$  can be projected onto this one-dimensional subspace by  $Z_i = X\theta_i$ . The (scalar) variance of the projections can be calculated as  $E\{z_i^2(k)\} = \frac{1}{k} \cdot Z_i^T Z_i = \frac{1}{k} \cdot \theta_i^T X^T X \theta_i$ . Of course the length of the vector  $\theta_i$  must be restricted somehow, for example, to  $\theta_i^T \theta_i = 1$ . This means one is facing a constrained optimization problem with

$$\begin{cases} f(\theta_i) &= \frac{1}{k} \cdot \theta_i^T X^T X \theta_i \\ g(\theta_i) &= 1 - \theta_i^T \theta_i. \end{cases} \quad (3.2)$$

Using Lagrange multipliers, the optimum solution  $\theta_i$  has to obey

$$\frac{dJ(\theta_i)}{d\theta_i} = \frac{d}{d\theta_i} (f(\theta_i) - \lambda_i \cdot g(\theta_i)) = 0 \quad (3.3)$$

---

or

$$2\frac{1}{k} \cdot X^T X \theta_i - 2\lambda_i \theta_i = 0, \quad (3.4)$$

giving

$$\underbrace{\frac{1}{k} \cdot X^T X}_{R} \theta_i = \lambda_i \theta_i. \quad (3.5)$$

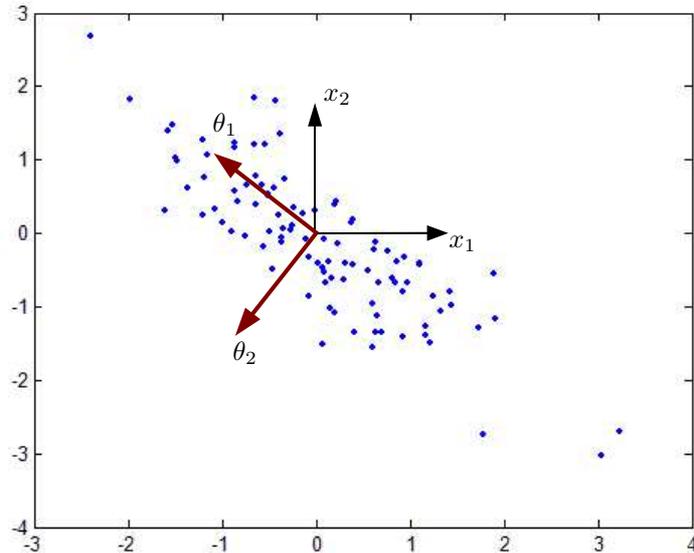
The variance maximization problem has turned into a much more familiar *eigenvalue problem*, with the searched vector  $\theta_i$  being an eigenvector of the matrix  $R = \frac{1}{k} \cdot X^T X$ . The eigenvectors of the data covariance matrix are called *principal components*.

It will be assumed that the eigenvectors are always normalized to unit length. Since the covariance matrix  $R$  is symmetric, it can be verified that

- the eigenvectors  $\theta_i$  are *orthogonal*, and because of the assumed unit length also *orthonormal* and
- the eigenvalues  $\lambda_i$  are always *real* and *non-negative*.

See also [6] and Appendix B for further proof. These characteristics are crucial — because of orthogonality the eigenvectors are uncorrelated, and the basic vectors corresponding to the maximum variance can be extracted without disturbing the analysis in other directions.

Assume again, that the dataspace  $X$  has two dimensions ( $n = 2$ ). Figure 3.2 shows the distribution of the example data after centering and normalizing to unit variance and plotted in black into the data points its original “natural” basis. The data is designed to be collinear, what means that there is at least one linear dependency among the variables. It is actually obvious from the figure that there must be an underlying freedom for the variables in one somewhat diagonal direction and a constraint in the other direction. This guess will be verified by PCA analysis.



**Figure 3.2:** Illustration of the “natural” and the PCA bases for collinear data

The covariance matrix (scaling factor  $k$  omitted) of this in MATLAB generated data is

$$R = X^T X = \begin{pmatrix} 1.0083 & -0.8412 \\ -0.8412 & 0.9962 \end{pmatrix}. \quad (3.6)$$

Note that the variances of the two data vectors can be found as diagonal elements of  $R$ . Beside some small numerical errors due to MATLAB algorithms they equal 1. The eigenvectors of this matrix are

$$\theta_1 = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \theta_2 = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad (3.7)$$

and the corresponding eigenvalues

$$\lambda_1 = 1.8435 \quad \lambda_2 = 0.1611 \quad (3.8)$$

if ordered in descending order of the numeric value of the eigenvalues. The directions of the eigenvectors are also plotted into the data in Figure 3.2 in red color. One can see that

---

an initial guess of two different characteristics of freedoms/constraints can be found and shown.

If one uses the eigenvectors as new basis for the data, then denoted as  $Z$ , the mapping  $Z = X\Theta$  with

$$\Theta = \left( \theta_1 \mid \dots \mid \theta_n \right) \tag{3.9}$$

needs to be performed. These new variables are mutually uncorrelated and the eigenvalues  $\lambda_i$  directly reveal the variances of the new variables, broke down into the different directions [6]. The variability in  $X$  is not changed, but only redistributed in  $Z$ , because the covariance matrix of the data after projection  $\frac{1}{k} \cdot Z^T Z$  has the same eigenvalues as  $R$  and it holds

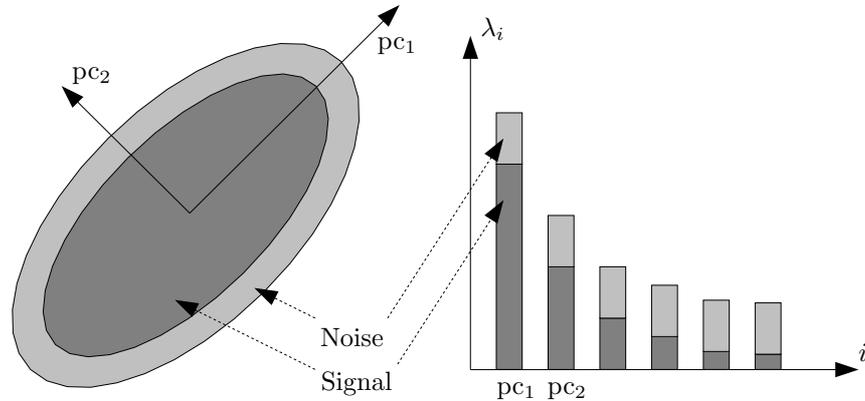
$$\text{var}\{z_1\} + \dots + \text{var}\{z_n\} = \lambda_1 + \dots + \lambda_n = \text{var}\{x_1\} + \dots + \text{var}\{x_n\}. \tag{3.10}$$

If the dimension is to be reduced, the optimal approach is to drop out those variables that carry least information. If an  $N < n$  dimensional basis is to be used instead of the full  $n$  dimensional one, construct  $\Theta_N$  as

$$\Theta_N = \left( \theta_1 \mid \dots \mid \theta_N \right), \tag{3.11}$$

and perform the mapping  $Z = X\Theta_N$ . It turns out that the eigenvalues of  $R$  give a straightforward method for estimating the significance of PCA basis vectors — the amount of neglected data variance when  $\theta_i$  is dropped is  $\lambda_i$ .

Furthermore, the noise, which actually remains in the projected and reordered data  $Z$ , can be reduced by omitting some of the least important latent vectors. Figure 3.3 shows very clearly that the signal/noise-ratio becomes increasingly worse for each principal component, because there is no more actual information in the directions with lower  $\lambda_i$ . However, the noise is assumedly randomly distributed among data and uncorrelated, thus present in every principal component with the same significance. Uncorrelated noise has a covariance



**Figure 3.3:** Two views of the “directional” information vs. the “undirectional” noise. On the left 6 dimensional data was projected to the first two principal components, on the right the corresponding PCA eigenvalues. Relatively, most of the noise seems to be concentrated in the direction of lowest overall variation

matrix of  $q \cdot I_n$ , and adding this noise matrix to the data covariance matrix simply shifts *all* eigenvalues of the final covariance matrix up the amount  $q$  (see the light grey parts in Figure 3.3).

Back to the example of Figure 3.2, one can see that the ratio between the carried information of the two variables, hence the ratio of the two eigenvalues, is  $\lambda_1/\lambda_2 = 11.4443$ . This means that one variable carries around 11 times more information than the other one. The basis vector  $\theta_1$  is much more important as compared to  $\theta_2$ . One may also express the amount of carried information in percent:

$$w_{\theta_i} = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \quad w_{\theta_1} = 91.96\% \quad w_{\theta_2} = 8.04\%. \quad (3.12)$$

When a reduced basis with only one vector (obviously  $\theta_1$ ) is applied, all the blue data points are projected onto this vector direction and only the deviations from the line  $x_2 = -x_1$  are assumed to be noise and neglected. What is more, the *data collinearity is avoided altogether*.

---

Also the equality in (3.10) holds

$$\sum_{i=1}^n \lambda_i = 2.0046 = \sum_{i=1}^n \text{var}\{x_i\} = \text{trace}\{R\}. \quad (3.13)$$

This example explains the principles and very useful features of Principal Component Analysis, especially when searching for degrees of freedom among data and trying to get rid of noise in very correlated and redundant data.

### 3.3 Linear Model

Again the  $n$  variables  $\mathbf{X}_i$  and their  $k$  measurements are collected in the matrix  $X$  with dimension  $k \times n$ . The aim of regression (and also of the thesis at hand) is the computation of a simple and manageable model between some “input” variables and some “output” variables. Hence there has to be a second set of measurements that holds the variables to be estimated. These  $m$  variables  $\mathbf{Y}_i$  are collected in the matrix  $Y$  of the dimension  $k \times m$ . It is assumed like it is mostly the case that there are much more measurements available than variables, hence  $k \gg n$ . One would like to find a matrix  $F$  so that

$$Y = X \cdot F \quad (3.14)$$

would simply hold. Finding a “good” matrix  $F$  is the main task here. There are  $n \cdot m$  free parameters in this model, and the optimum is searched for in this parameter space.

### 3.4 Multilinear Regression

Assume  $m = 1$  so that there is only one output  $\mathbf{Y}_i$ . This reduces (3.14) to

$$\mathbf{Y}_i = X \cdot \mathbf{F}_i \quad (3.15)$$

---

and the matrix  $F$  to a vector  $\mathbf{F}_i$ . Generally no exact solutions can be found here, since  $X$  is not invertible. To find an approximation, the model needs to include modeling errors, as

$$\mathbf{Y}_i = X \cdot \mathbf{F}_i + \mathbf{E}_i, \quad (3.16)$$

where  $\mathbf{E}_i$  is an  $k \times 1$  vector containing the reconstruction error for each measurement  $k$ . To simultaneously minimize all the errors in (3.16), one can write  $\mathbf{E}_i = \mathbf{Y}_i - X \cdot \mathbf{F}_i$  and *minimize the sum of error squares*. This sum can be expressed as

$$\begin{aligned} \mathbf{E}_i^T \mathbf{E}_i &= (\mathbf{Y}_i - X \cdot \mathbf{F}_i)^T (\mathbf{Y}_i - X \cdot \mathbf{F}_i) \\ &= \mathbf{Y}_i^T \mathbf{Y}_i - \mathbf{Y}_i^T X \mathbf{F}_i - \mathbf{F}_i^T X^T \mathbf{Y}_i + \mathbf{F}_i^T X^T X \mathbf{F}_i. \end{aligned} \quad (3.17)$$

This scalar can be derivated with respect to the parameter vector  $\mathbf{F}_i$

$$\frac{d(\mathbf{E}_i^T \mathbf{E}_i)}{d\mathbf{F}_i} = 0 - X^T \mathbf{Y}_i - X^T \mathbf{Y}_i + 2X^T X \mathbf{F}_i. \quad (3.18)$$

Because the second derivative is always positive, and because quadratic functions only have one single extremum, the extremum in (3.18) is a minimum and also unique. Setting (3.18) to zero gives the optimal parameters:

$$-2X^T \mathbf{Y}_i + 2X^T X \mathbf{F}_i = 0 \quad (3.19)$$

resulting in

$$\mathbf{F}_i = (X^T X)^{-1} X^T \mathbf{Y}_i. \quad (3.20)$$

With this, the estimate for  $\mathbf{Y}_i$  is found as

$$\hat{\mathbf{Y}}_{\text{est}, i} = \mathbf{F}_i^T X_{\text{est}}. \quad (3.21)$$

As it can be seen later, it is necessary to split the available data  $X$  in a part for model estimation (training set) and another separate part for model validation. This is why  $X$

---

and  $\mathbf{Y}_i$  here are denoted here as  $X_{\text{est}}$  and  $\mathbf{Y}_{\text{est}, i}$ . The accent  $\hat{\mathbf{Y}}_{\text{est}, i}$  is used to distinguish between model estimated data and real measurement data  $\mathbf{Y}_{\text{est}, i}$ .

If there are various output signals (as it is in the nickel plating case), so that  $m > 1$ , the above calculations can be carried out for each variable separately and collected together as

$$\begin{cases} \mathbf{F}_1 &= (X^T X)^{-1} X^T \mathbf{Y}_1 \\ &\vdots \\ \mathbf{F}_m &= (X^T X)^{-1} X^T \mathbf{Y}_m. \end{cases} \quad (3.22)$$

This set of formulas can be rewritten in a compact matrix form, so that

$$F = \left( \mathbf{F}_1 \mid \dots \mid \mathbf{F}_m \right) = (X^T X)^{-1} X^T \cdot \left( \mathbf{Y}_1 \mid \dots \mid \mathbf{Y}_m \right). \quad (3.23)$$

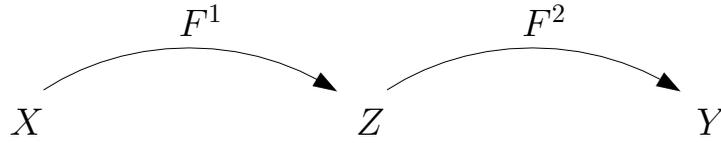
This results in a *Multilinear Regression (MLR)* model from  $X$  to estimated  $Y$  as

$$F = (X^T X)^{-1} X^T Y. \quad (3.24)$$

The MLR solution is optimal and exact in the sense of the least squares criterion. However, one has to be careful using this regression method without proper reflecting. Trying to explain noisy data too exactly may make the model sensitive to individual noise realizations, or the algorithm may collapse altogether. Also a non-orthogonal basis of the data space corrupts the MLR algorithm. The different coordinates “compete” against each other (heuristically speaking), often resulting in excessive numeric values. More applications and analysis about MLR are carried out in [6].

### 3.5 Principal Component Regression

The reader may have noticed that the approach to tackle all problems of redundant data and non-orthogonal bases was already introduced. Principal Component Analysis offers



**Figure 3.4:** The dependency model  $y = f(x)$  refined (PCR)

not only the possibility to find the most important directions among data and to get rid of noise directions, even more it provides this new “filtered” information in an orthogonal, even orthonormal basis. Using PCA as redundancy elimination it can serve as basis for the regression algorithm. Using a new subspace basis  $\Theta$ , derived by PCA, one can implement the above explained MLR, which can thus be called *Principal Component Regression (PCR)*.

The overall regression model construction becomes a two phase process, implementing the following tasks

1. Determine the basis  $\Theta$
2. Construct the mapping  $F^1 = \Theta(\Theta^T\Theta)^{-1}$
3. Calculate the “latent variables”  $Z = XF^1$
4. Construct the second-level mapping  $F^2 = (Z^T Z)^{-1}Z^T Y$
5. Eventually estimate  $\hat{Y}_{\text{est}} = X_{\text{est}}F = X_{\text{est}}F^1F^2$

The mappings that are calculated in this algorithm are visualized in Figure 3.4.

Here  $Z$  names the internal coordinates on the basis  $\Theta$ . Obviously this algorithm works fine for every basis  $\Theta$ , and, as discussed above, even better and without numerical instability if  $\Theta$  is orthogonal and  $Z$  holds mostly real information as compared to noise. If one derives  $\Theta$  using PCA, this basis will be orthogonal and optimal in this sense.

---

Since  $\Theta$  is also even orthonormal, the construction of  $F^1$  in step 2 is even more straight forward. In this case there holds  $\Theta^T\Theta = I_N$ , thus  $F^1 = \Theta$ . However, for unnormalized bases the *pseudoinverse* [13]  $\Theta^\dagger = (\Theta^T\Theta)^{-1}\Theta^T$  has to be used.

Two methods were introduced that offer the best possible solutions to well-defined compact problems. Combined, they are a powerful tool not only for neocybernetic ideas of freedoms-oriented modeling but also for the simplicity pursuit — the overall model calculation for PCR including PCA and MLR

$$F = F^1F^2 = (\Theta^T X^T X \Theta)^{-1} \Theta^T X^T Y \quad (3.25)$$

seems to be easy managable.

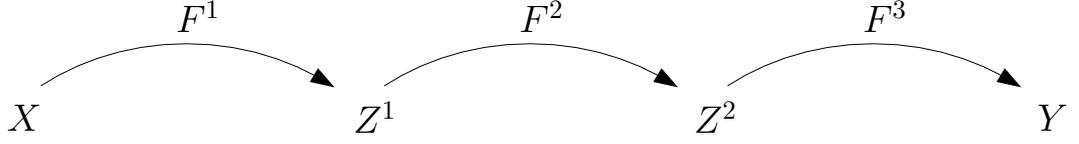
### 3.6 Partial Least Squares

The main disadvantage of PCR is the fact that it is exclusively concentrating on the input variables. Keeping some of the first principal components for the regression to  $Y$  solves the problem of possible multicollinearity in  $X$  — but they are chosen to explain  $X$  rather than  $Y$ . And so, nothing guarantees that the principal components, which “explain”  $X$ , are relevant for  $Y$ . The next step is to connect the output variables in the analysis and synthesis of the model, since not only the variance among an input dataset  $X$  is to be captured, but also the correlation between  $X$  and some output data  $Y$ . Among these strategies the *Partial Least Squares (PLS)*<sup>1</sup> seems to be most known among practicing engineers, emphasizing correlations rather than variance.

The process becomes slightly more complex than in the case of PCA. Not only the input block  $X$  has to be restructured, but the internal structure of  $Y$  is also searched for. The procedure becomes such that  $X$  is projected on a  $X$ -oriented subspace spanned by the

---

<sup>1</sup>Sometimes also known as *Projection onto Latent Structure* [6].



**Figure 3.5:** The dependency model  $y = f(x)$  refined (PLS)

basis vectors  $\theta_i$ . After that, data is projected on the  $Y$ -oriented subspace spanned by the vectors  $\phi_i$ , and only after that, the final projection onto the  $Y$  space is carried out. Figure 3.5 shows the refined model and projection structure.

The PLS model is ususally constructed in yet another way as described here [6], but it is extremely uninstructive and implicit. A practicing engineer does not have to grasp the unpenetrable algorithmic presentation of the PLS ideas. Also here the available toolboxes were used along with the given explanation.

The objective now is to find the basis vectors  $\theta_i$  and  $\phi_i$  so that the correlation between the projected data vectors  $X\theta_i$  and  $Y\phi_i$  is maximized while the lengths of the basis vectors remain constant. Again one faces a constrained optimization problem

$$\begin{cases} f(\theta_i, \phi) = \frac{1}{k} \cdot \theta_i^T X^T \cdot Y \phi_i \\ g_1(\theta_i) = 1 - \theta_i^T \theta_i \\ g_2(\phi_i) = 1 - \phi_i^T \phi_i, \end{cases} \quad (3.26)$$

with two separate constraints  $g_1$  and  $g_2$ . Defining Langrange multipliers  $\eta_i$  and  $\mu_i$  gives

$$\frac{1}{k} \cdot \theta_i^T X^T \cdot Y \phi_i - \eta_i(1 - \theta_i^T \theta_i) - \mu_i(1 - \phi_i^T \phi_i) \quad (3.27)$$

and setting the differentiation with respect to  $\theta_i$  and  $\phi_i$  to zero gives a pair of equations

$$\begin{cases} \frac{1}{k} \cdot X^T Y \phi_i - 2\mu_i \theta_i = 0 \\ \frac{1}{k} \cdot Y^T X \theta_i - 2\mu_i \phi_i = 0. \end{cases} \quad (3.28)$$

Solving to  $\theta_i$  and  $\phi_i$  gives once more the familiar formulation of an eigenvalue problem:

$$\begin{cases} \frac{1}{k^2} \cdot X^T Y Y^T X \theta_i = 4\mu_i \eta_i \cdot \theta_i \\ \frac{1}{k^2} \cdot Y^T X X^T Y \phi_i = 4\mu_i \eta_i \cdot \phi_i. \end{cases} \quad (3.29)$$

---

Again the significance of the vectors  $\theta_i$  (for the  $X$  block) and  $\phi_i$  (for the  $Y$  block) is revealed by the corresponding eigenvalues  $\lambda_i = 4\eta_i\mu_i$ .

Because  $X^T Y Y^T X$  and  $Y^T X X^T Y$  are symmetric, the orthogonality properties, which are so crucial for PCR, apply to these eigenvectors. In practice, the basis vectors  $\phi_i$  are redundant and they need not to be explicitly calculated. The “complexity” of PLS is an illusion and only presented in this way in order to reach conceptual comprehensibility, since between  $Z^1$  and  $Z^2$  there is no additional information loss or compression introduced. The key point here is that the basis vectors  $\theta_i$  in (3.29) are derived by using also the output block  $Y$ , hence bridging the input to the output.

The main advantage of PLS as compared to PCR is not only involving the output values that one is actually interested in, but also the reduced dimension of the final linear model. Because the rank of a product of matrices cannot exceed the rank of the multiplied matrices, there will be only  $\min\{n, m\}$  non-zero eigenvalues. That is why the PCR approach may give higher dimensional models than PLS.

This chapter demonstrated the multivariate tools used later in the thesis at hand. Together with neocybernetic ideas these tools provide a powerful framework for modeling complex systems.

The following Chapter 4 will give an introduction to a very complicated chemical process, the nickel plating of printed wiring boards. Its industrial background and relevance is introduced and emphasized. The importance of a very accurate model is pointed out. This leads to Chapter 5, which describes shortly the existing models, developed in the last century and also recently at the TKK Control Engineering Laboratory.

To tie these two loose ends together, Chapter 6 will show how neocybernetic ideas lead to a model of the process. This simple model will be compared to the recent existing one and the results will be presented.

# 4

## Electroless Nickel Plating

After introducing the framework and its ideas this chapter will describe the actual industrial process, to which the neocybernetic thoughts should be applied. The complicated and chemically still mostly unknown process of the surface finishing process *electroless nickel plating* will be presented and embedded in its industrial background. The presentation follows [14]. At the end of the chapter the motivation to model this process and its important current parameters are emphasized.

### 4.1 Industrial Background

There are many uses for electroless nickel plating in industry, because it provides some very accurate features compared to other surface finishing and plating methods. That is why this coating method is widely used, for example, for hydraulic cylinders, valves, gears, electronic conductors and plugs etc [14]. One application is PWB (printed wiring board) manufacturing, where it is used as an oxidation barrier between the copper electric circuits itself and the gold finishing. Figure 4.1 shows some applications as itemized above, especially a PWB at the right hand side.

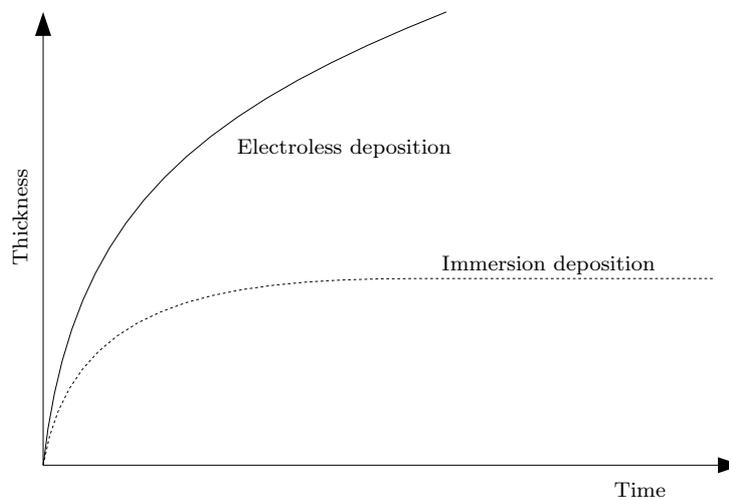


**Figure 4.1:** Some different applications for electroless nickel plating [15]

In this method the deposited nickel is not provided by an anode (electro nickel plating) but by a specific nickel salt. This salt is solved in an aqueous solution where the nickel then is in ion form. A substrate, a piece which should be plated, serves as a cathode in a bath of different components, including the nickel ion solution. Therefore the substrate surface has to be catalytically active or before the process somehow activated. Fortunately nickel itself is catalytically active, which makes it possible to accumulate nickel onto nickel and in that way form very thick alloys.

During the plating process, the electrons, which reduce the nickel ions to actual nickel, are not provided by an external current source as anode, but by a reducing agent which makes the bath itself anodic against the catalytic surface of the substrate. This is the reason for the name “*electroless* nickel plating”. However it is somehow misleading, even without real electrodes the electric current still exists. Hence, *autocatalytic deposition of nickel* and *chemical reduction of nickel* are other scientific and somewhat clearer names for the process [14], [16].

In comparison to electroplating, where an external active power supply provides the needed electrons, the electroless nickel deposit provides very good protection against corrosion, it has high abrasion resistance and good adhesion, which makes it an excellent surface finishing. Its hardness is greater and the thickness distribution far more uniform. The thickness of the deposition is not bounded, because the surface is always active, even



**Figure 4.2:** Comparison of electroless and immersion (active power supply) deposition [17]

if there are already some layers of nickel atoms on it (see Figure 4.2). However, the electroless plated nickel alloy contains a small amount of the used reducing agent. Physical properties of the surface, like hardness of the deposit, ductility, inner stress, resistance, soldering and corrosion are affected by this small content [14].

In 90% of industrial electroless nickel plating lines sodium hypophosphite ( $\text{NaH}_2\text{PO}_2$ ) is used as reducing agent [14]. Due to that the deposit contains up to 15 weight percent of phosphorus. A detailed list of bath components and their attributes will be given in Section 5.1. All the above mentioned properties of the deposit can be affected by the difference in the actual phosphorus content of the layer. Because of its great influence there is a great interest to control the phosphorus content besides the alloy thickness itself properly.

## 4.2 Ni-Au Finishing

This section goes closer to the plating process itself, to used material and important variables. Glass fiber reinforced epoxy laminate, which is laminated with a thin copper

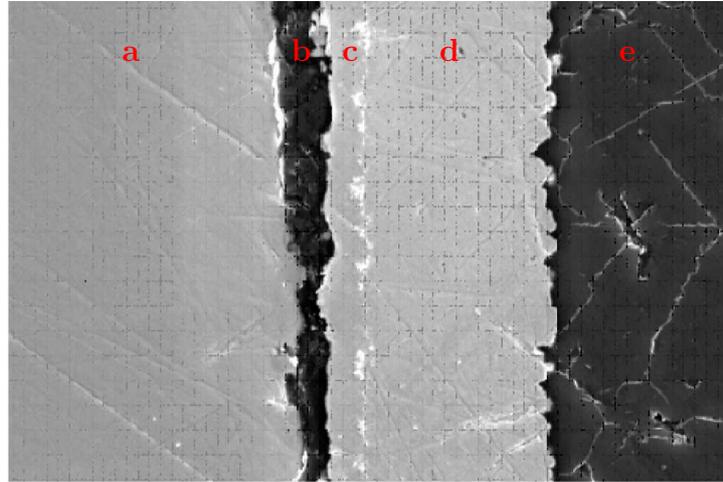
---

layer, is the base for PWB manufacturing. Often also multilayer wiring boards are used with up to 10 layers. Into the copper layer(s) the wanted layout of the circuit board is formed using a “print-and-etch”-process [14]. The copper layer, without any protection against environment, would oxidize very fast, so it needs to be protected from oxidation, especially when used as a keyboard, for example, for mobile phones or other applications which bring the boards in contact with water or humid skin [18].

To protect the copper layer from oxidation it is necessary to coat the copper with a very thin (0.05 – 0.15  $\mu\text{m}$ ) gold (Au) layer. However, gold cannot be plated directly onto copper, because the copper would diffuse into the gold and create an again easily oxidizing compound. That is why a nickel-phosphorus (Ni-P) layer (less than 5  $\mu\text{m}$ ) is added between copper and gold layer. So the gold layer protects both underneath layers from oxidation and the nickel layer protects the diffusion of copper into gold. The nickel layer is added by using the electroless nickel plating technique. However, the additional layer is not only used as a barrier between copper and gold, but also because of its above already mentioned mechanical properties [14], [18].

Figure 4.3 shows the different layers on the base of a PWB after the Ni-Au finishing. From right to left one can see the copper layer, into which the electronic circuits of the PWB were etched, and the nickel layer. However, the anti-oxidizing gold layer (between (c) the nickel layer and (b) the fastener tool) is so thin that it can not be distinguished on the picture. It can be seen that the surface of the nickel layer is not even at all, which might influence thickness measurements and make them quite noisy. Also the phosphorus content measurement is very challenging, since the phosphorus is distributed unevenly over the nickel layer, which makes it necessary to take multiple measurements.

The described plating processes are very crucial in the production of PWB’s and it is not possible to measure the characteristics of the Ni-P layer during plating, which complicates quality control a great deal. The quality inspections take place randomly when the plating



**Figure 4.3:** The cross-section of a plated test plate taken with a scanning microscope. Layers: (a) a metallic fastener and (b) a piece of conductive plastic to fix the sample, (c) Ni-P layer, (d) copper layer, (e) the base of PWB (epoxy laminate) [14]

process is already over. Errors have radical consequences to the whole batch or even to the customer, if an error in a particular batch is not detected.

### Desired Values

The most important values to be controlled during the plating process are the alloy thickness and its phosphorus content. It is possible to derive *desired values* for these variables from requirements to their functionality. The nickel deposit thickness should be between 2.5 and 5  $\mu\text{m}$ . If the layer is thinner it does not prevent diffusion of copper and nickel into gold and if it is thicker it would harm solderability and corrosion resistance of the whole alloy. The phosphorus content should be between 7 and 10 weight percent. The corrosion resistance improves with increased phosphorus contents, solderability improves with decreased phosphorus. It is quite difficult to find the optimum to this antagonism. To prevent both copper and nickel layer from oxidation, the gold layer should have a thickness between 0.05 and 0.15  $\mu\text{m}$  [14].

---

## Process Steps

In literature there are many alternative processes suggested for the creation of a Ni-Au line, depending on needed volume, used equipments and wanted final properties. This work is based on the production results and measurements which were collected during the research for [14] and [19].

After loading the raw copper PWB's a crane takes them from bath to bath. After each functional bath the boards are rinsed. The first baths are cleaning baths, clearing the surface first from grease, dust and remains of organic solders, then etching away last dirt and metal oxides, preparing the surface for activation. The "auro dip" removes the cleaning acid and the "pre dip" acidifies the surface with sulphuric acid in preparation for the next bath.

Because copper is catalytically passive, nickel would not deposit on it. To have the desired reaction in the nickel bath, the PWD surface is now activated using dilute palladium chloride solution. The outermost copper atoms are thereby replaced by palladium atoms that form a catalytic surface. The "post dip" follows.

Now the PWBs are ready for the nickel bath. Despite intensive studies, this particular step, the electroless nickel decomposition reaction mechanism, is still not clearly understood [14], [17]. In Figure 4.4 one can see the two tanks responsible for nickel plating, one of them with immersed PWBs, the other one idle. After the nickel bath and rinsing the gold layer is added by a bath of potassium gold cyanide, then the finished PWD is dried and stored.

## Important Parts of the Nickel Bath

Focusing on the tank of the nickel bath there are some important parts to be mentioned. The system is controlled by two computers and a PLC (programmable logic controller).

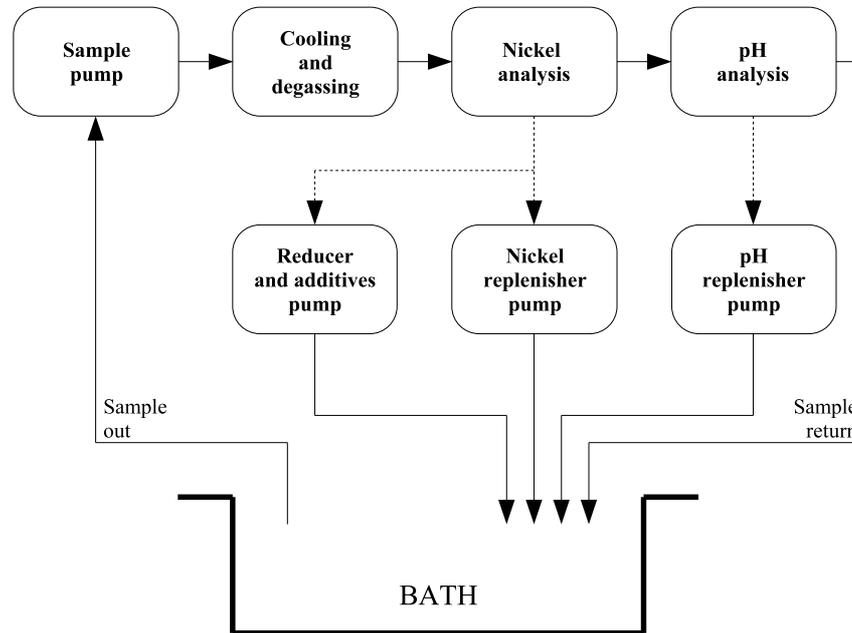


**Figure 4.4:** Two tanks for electroless nickel reaction. Behind there are two bars of PWDs immersed for deposition [14]

This system stores the measurements, activates the crane and shows the information to the operator for supervising. The tank itself has to be large enough to be economically profitable and strong enough to stand the weight of the bath and the reaction going on inside without wasting expensive reactants by getting deposited onto the tank walls. The bath is heated by electric immersion heaters to a temperature between 80 and 90 °C. The agitation of the bath is secured by air agitation, to guarantee that there are always fresh reactants on the substrate surface and the by-products of the reaction are moved away. To keep contaminants from the substrate or the air out of the bath the solution is also continuously filtered by a bag filter.

## Measurements

If it were possible to measure the thickness of the deposition and its phosphorus content on-line and during the process itself, there would be no need for a model to estimate or predict these variables for control purposes. Therefore, earlier tests were made to develop



**Figure 4.5:** Schematic diagram of the Ni-controller [14]

an online measurement device, for example, on a QCM (quartz crystal microbalance) detector [14], which vibrates at a certain frequency in the nickel bath. The QCM sensor is getting covered with nickel like the substrate itself and changes its frequency while being plated. However, with QCM it was not possible to measure the alloy's phosphorus content besides its thickness, so the experiments were not continued [14].

Therefore the present-day control strategy in a modern nickel plating line is based on the assumption that *the deposition rate does not change if there is no change of circumstances*. According to this assumption the thickness is controlled only by immersion time. The only interest in measurements is to keep the state of the bath constant. Thickness and phosphorus content are measured only seldom in a laboratory. The idea to use the measurements to build up a model of the plating process requires a closer look at the measurement system, to understand where the used data comes from.

Temperature, pH and nickel concentration are measured with a device called Ni-controlle

---

[14]. It also adds replenishers to the bath if one of the latter differs from a given set value. Nickel sulphate is added when there is a loss of nickel in the bath detected and ammonia when the pH value drops. Furthermore the MTO (metal turnover), which is defined as the relation of pumped nickel to the nickel concentration of the fresh bath, is calculated and can be seen as an indicator for the bath's age. The schematic diagram of the controller and its functionality are shown in Figure 4.5. One can see that after preparing the sample the controller is taking the actual measurements and calculates the needed amount of replenishers, which are then added by specific pumps.

Looking somewhat deeper inside the above given description one can find a control loop. The current values of variables are *measured*, afterwards *compared* to desired values and a *correction* is made in order to keep the variable to its set point. These control loops are examples for the modern control machinery keeping the bath and its variables in balance. Meanwhile these controllers are so effective, that the *dynamic balance* of the system is well maintained (see Section 2.2).

The described measurement sequence takes some time, from 3 to 15 minutes, depending on the amount of replenishers to be added. Because of this delay the measurements can not strictly be called "online". But since the time constant of the whole chemical process in the bath itself is much higher [14], the measurements can be considered as on-line.

The measured data then has to be recorded for later use. For this purpose a data acquisition system was built up which records, for example, from the Ni-controller the time stamp when the sample is taken, and with it the actual nickel concentration, pH value, and the temperature of the sample. There is also a log file where the movements of the substrate by the crane are recorded in order to save information about the load of the bath, the duration of plating, and the time of immersion of each PWB. Nowadays most thickness and phosphorus content control mechanisms are based on this information about immersion time. The third source of data is the laboratory, where sometimes measurements of the

---

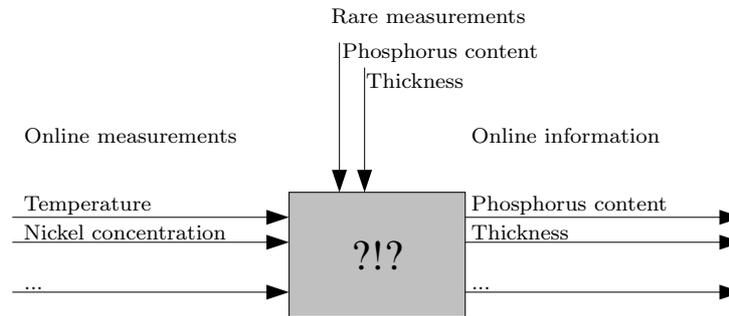
alloy thickness and its phosphorus content are taken and recorded with the time stamp when the examined PWB was actually plated. A computer program was developed [14] to combine the different sources of data into one by time stamps synchronized source, which now holds information, for example, about load, temperature, pH value and nickel concentration arranged among a continuous timeline. There will be more information about the used data in Section 6.2.

Of course it has to be considered, that sometimes the measurements are not accurate because of erroneous sensors or they are simply not available. Especially thickness and phosphorus content (seldom measured in comparison to other variables) naturally cause a great deal of time stamps without information about the actual thickness. This is the basic reason to consider the possibility of a thickness model, to get information about the plating process continuously and not only from laboratory measurements.

### 4.3 Model and Control Tasks

Control of thickness of the PWBs being plated in state of the art Ni-Au lines is based on the *assumption*, that with the state of the bath (maintained by the control loops described above) also the deposition rate stays constant. Therefore, in this case it is only necessary to control the time, how long a plate is immersed into the bath to reach the desired thickness [20].

This approach is not satisfying at all. It is not known to the operator how the bath conditions change, how the plate in the bath behaves, how the alloy is developing, or how the quality of the PWB is changing. This can lead to a great deal of waste after examining the thickness of the deposit and its quality from a random sample, let alone the undetected waste, which causes much more economical damage to the producing factory and its customers.



**Figure 4.6:** Functionality of a nickel plating model

A technical model of the process, which is providing on-line information about the actually immersed plate, using information that is available through online measurements throughout the plating process, would give the chance to close the outer control loop. Thickness, for example, could be predicted or estimated and according to that, the immersed substrate can be *emersed in time*, before the thickness differs from a desired value. No regular laboratory measurements would be needed, the thickness is known and the finished product can be delivered. Only very rarely could laboratory measurements be used to keep the quality and check the accuracy of the model. This principle is illustrated in Figure 4.6.

Thus the aim is to get online information about the actual state of the substrate and use it to secure quality and save money. For that, only simply available data should be used, thus making the expensive measurements in a laboratory needless and the production and its control economical.

As one will realize while reading the following chapters, this task is anything but easy. The process going on in the bath is rather unknown and the few known things about it are mostly a well kept secret of the producing companies. These companies are not providing the full information about their plating lines in order to keep their position on the economic market of surface finishing. Not even the chemical behaviour can be summarized in simple reaction equations [17].

---

Many different reactants are used to keep certain characteristics of the bath under control. Along with these reactants, activators and inhibitors many form new, mostly unknown, bindings between themselves and the main substances. These connections can be seen as additional constraints to the process variables, which must be taken into account while modeling the process with the traditional methods and tools. This carries on the motivation to use a freedoms-oriented approach (see Subsection 2.3.2) to model characteristics of the nickel plating process.

The next Chapter 5 will describe existing models and their behaviour in predicting the important parameters and variables of the process. Chapter 6 will then look at the process from another point of view.

# 5

## Existing Models

For a long time, chemists tried to discover what is happening during a nickel deposition. This chapter describes the experiences with understanding and modeling the behaviour of an electroless nickel plating process, starting in the 19th century up to present day.

### 5.1 Historical Models

In 1844, Wurtz observed that nickel cations were reduced by hypophosphite anions. However, Wurtz only obtained black powder. The first bright depositions of nickel-phosphorous alloys were developed in 1911 by Breteau. These baths decomposed spontaneously and formed deposits on every surface immersed into the solution, also on the container walls. In 1946, Brenner and Riddell published a paper about the electroless nickel deposition in proper conditions [17].

The chemical and physical properties of the electroless nickel coating depend on its composition, which depends on the operation conditions of the nickel bath. Typically the constituents of the bath are

- 
- A source of nickel ions
  - A reducing agent
  - pH-control acid
  - Suitable complexing agents
  - Stabilizers/inhibitors
  - Energy

The preferred, and in this case used, nickel source is nickel sulphate ( $\text{NiSO}_4$ ). The purpose of the reducing agent is to provide electrons to the nickel ion and reduce it to a nickel atom. This would theoretically mean that the deposit layer would be of pure nickel, which is almost (99% and more) achieved when using Hydrazine. However, in the described plating line sodium hypophosphite ( $\text{NaH}_2\text{PO}_2 \cdot \text{H}_2\text{O}$ ) is used. In this case, the alloy contains not only nickel, but also up to 15 weight percent of phosphorus. As said above, this influences the properties of the deposition significantly. Ammonia ( $\text{NH}_3$ ) is used to control the pH value of the bath.

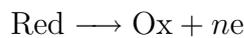
The complexing agent is a compound that stabilizes the pH of the bath and prevents precipitation of nickel salts. The agent makes the nickel complex more passive to make the deposition possible only on the substrate surface. It changes the reaction rate but keeps the bath and the ongoing reaction stable and controllable. To prevent the bath from suddenly decomposing and building salts everywhere in the bath and not only on the surface of the substrate, stabilizers are used. The stabilizer might make it possible to accelerate the deposition because it lowers the possibility of a decomposition of the bath. Its concentration is quite critical, because too much of it can stop the whole plating reaction, so its use and concentration are mostly a well kept business secret. Energy is given to the bath in form of heat. The temperature of the bath is kept around  $80^\circ\text{C}$ , a

---

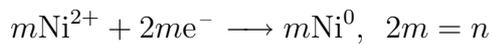
lower temperature would slow down the deposition process, a higher one would lead the bath into an unstable state [14], [17].

Electroless nickel deposition can be seen, in a very elementary manner, as the sum of two chemical reactions — an oxidation reaction liberating electrons and a nickel reaction consuming them:

Oxidation of reducing agent



Reduction of nickel ion



---

Overall or sum reaction



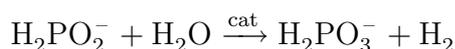
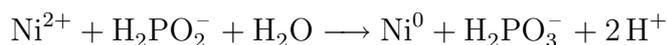
Experimentally observed reaction characteristics indicate that the reaction has to be much more difficult, since these stoichiometric equations fail to describe all the phenomena that are observed during plating [17].

Before discussing chemical equations, which should explain the process of plating more precisely, it might be informative to recall certain characteristics of the process the mechanism must explain [17]. Most important are the first two in the following list:

- The reduction of nickel is always accompanied by the evolution of hydrogen gas.
- The deposit contains not only nickel but also phosphorus from the reducing agent.
- The reduction reaction takes place on certain metals and also on the depositing metal itself.
- Hydrogen ions are generated as by-product.

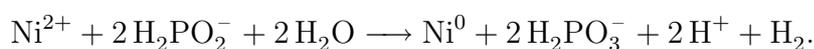
- 
- The molar ratio of nickel deposited to reducing agent consumed is usually equal to or less than 1.

In the literature nickel deposition using hypophosphite was sometimes presented as




---

Overall



However, the presented mechanism fails to account for the phosphorus component in the alloy. Further the rate of deposition would be proportional to the concentration of the reactants. This was experimentally disproved, when Gutzeit showed that the rate is independent of nickel ion concentration.

Since the publication of the above equations in 1946 four more mechanisms were proposed, trying to explain more of the above listed characteristics [17]:

- 1 **Atomic hydrogen mechanism** Does not explain certain phenomena like the reduction of hydrogen or phosphorus deposition and why the utilization of hypophosphite is always less than 50 percent.
- 2 **Hydride transfer mechanism** More accurate than (1), but uses the unlikely existence of  $\text{H}^-$ -ions as an intermediate reducing agent.
- 3 **Hydroxyd mechanism** Based on experimental data and still better than (1) and (2), but also uses some material (atomic hydrogen), whos existence is very unlikely.
- 4 **Electrochemical mechanism** Based on electrochemistry (compared to 1-3, which are purely chemical). Implies that the nickel ion concentration should have an effect on the rate of deposition — the converse is true.

---

This summary should give the reader some idea of how difficult and unknown the process at hand actually is. For more than 150 years scientists tried to describe all phenomena, which was obviously not possible up to now.

None of the mentioned mechanisms is precisely right, although the latter two are quite close to reality. Thus researchers believe that the process cannot be fully chemical, but is controlled by an electrochemical mechanism. This is also the most commonly used approach [14], [17].

Because of this the latter mechanism (4) was also used as a basis for the development of the recent model at the TKK Control Engineering Laboratory by Kantola [14] and Tenno [19]. This model will be the basis of comparison of the results, which the model presented in this work (Chapter 6) is producing. For this reason, Kantola's model is shortly introduced in the following section and some results are presented.

## 5.2 Electrochemical Model

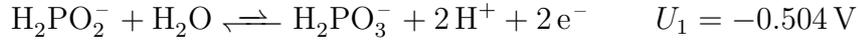
Based on chemistry and electrochemistry a mathematical model was derived, which predicts quite accurately the critical alloy characteristics like thickness and phosphorus content. In this section the development of the model will be sketched. The detailed description can be found in [14].

### Process Chemistry

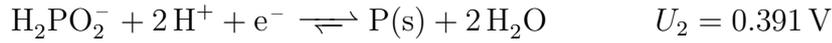
As a fundament for this model it is assumed that the process follows an electrochemical mechanism, as mentioned above (4); the reactions of the mechanism (4) now go both ways [14]. The chemistry can be described as:

---

Anodic reaction: hypophosphite oxidation



Cathodic reactions: phosphorus deposition, hydrogen evolution, nickel deposition



with

s = solid

g = gas

$U$  = voltage.

These equations account for the evolving hydrogen gas ( $\text{H}_2(\text{g})$ ) and also for phosphorus ( $\text{P(s)}$ ) on the surface of the substrate.  $U_{1-4}$  are the normal potentials of each reaction at 25°C. The voltages are necessary to combine the chemical equations with the mixed potential theory, which is a good basis for mathematical modelling and was done in [14] and [19].

The used mechanism only takes into account the main reagents. It does not describe the reactions of stabilizers or accelerators, mainly because the exact chemistry for them is not known and also because the model should stay as simple as possible.

Since chemical equations only describe stoichiometric correlations, there is still mathematics needed, to derive equations, which take dynamics into account and the actual interesting values like the alloy thickness or the phosphorus content. Also the influence of chemicals, which are not represented in the chemical equations, are studied there.

---

## Mathematical Implementation

Kantola used mathematical equations from an electrochemical cell model, which was originally developed for batteries [19], as basis for the mathematical model. There the Butler-Volmer equation was used to model the four current densities  $i_{1-4}$  according to potentials of the 4 electrochemical reactions (page 57). Every anodic/cathodic reaction creates an electric potential, which in turn causes a current  $i$  that can be calculated using the Butler-Volmer equation. The two free parameters in each equation were calculated from the experimental data, because there are no accurate approximations available. Here, assumed preknowledge about the process and real process information from data is mixed to achieve a fully functional model.

Using the assumption of electrical neutrality, which states that the sum of all anodic and cathodic current densities is zero, and the Arrhenius equation the free parameter of current potentials can be calculated. The other free parameter, the anodic apparent current coefficient, is gathered from the best fit to experimental data. Also basic formulas for the potential of the bath and voltages (Nernst equation) and for the rate constant (experimentally found) are used to close gaps in the current density equations.

The deposition formation itself was studied as the superposition of two independent processes — a nickel deposition and a phosphorus deposition:

- **Nickel deposition** rate is proportional to the current density  $i_4$  and also dependent on the molecular weight of nickel, the Faraday's constant and nickel's density, which are all well known.
- **Phosphorus deposition** rate follows the same principle. It is proportional to the current density  $i_2$  and dependent on the molecular weight of phosphorus and also on the Faraday's constant.

---

Thus the alloy deposition rate is the sum of the two above presented deposition rates. The thickness of each partial deposition and the overall deposition can be calculated by integrating the corresponding deposition rate over plating time. Normally the plating time is between 15 and 25 minutes. The phosphorus content is expressed as a weight percent, thus nickel and phosphorus are weighted with the corresponding densities.

In this model the concentration dynamics are essential. The concentration of the reagents, which is kept constant by measuring the actual amount of a reagent and adding fresh substance, and the by-products of the ongoing reactions have a great effect on the plating reaction, so a model was developed to describe their dynamics.

Equations for the concentration of hypophosphite, which works as a reducer in the reaction, and nickel were derived. Also for produced hydrogen, which lowers the pH value, and for ammonia, which is added to compensate the lowering, models were calculated. Sulphate, ammonium, ammonium sulphate, orthophosphite, and hydrogen gas were also cast in equations.

The main purpose (as already discussed in Section 4.3) of the model was to find relations between the measurable values of the bath and the thickness and the phosphorus content of the nickel deposition. However, the presented model is more wide-ranging and explains the process in a broader manner.

## **Results**

The results of the model are presented and compared with a set of experimental data. This data was derived and gathered during active experiments, which means that during the production of PWBs the circumstances in the bath were actively changed. The desired value for nickel concentration was changed between 95 and 105% of its reference value, which is  $6 \text{ mol/dm}^3$ . The other changed parameter was the pH value, which was varied from 4.72 to 4.79. This was done to get feasible changes in measured data, to distinguish

---

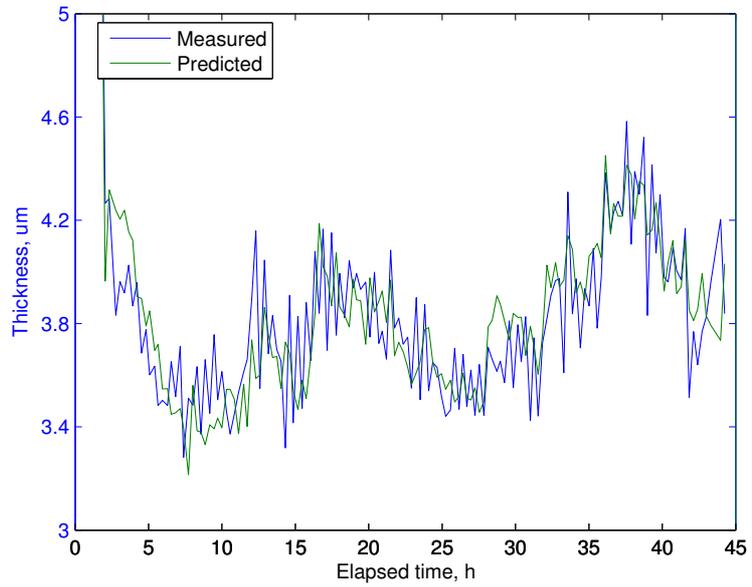
between real trends and measurement noise. Because of the Ni-controller (Section 4.2) the properties of the bath followed the changing desired values very well. Figures, evaluations and interpretations about the results are directly taken from [14].

Figure 5.1 shows the *Ni-P alloy thickness*, where measurement and model prediction are plotted in one graph. As can be seen, the model predicts the thickness of the deposition quite well, especially the trend behaviour is very accurate, considering the uneven surface of the alloy (shown in Figure 4.3) and the resulting uncertainty of the measurement. The large values of the first measurement are caused by two test runs to initialize the bath.

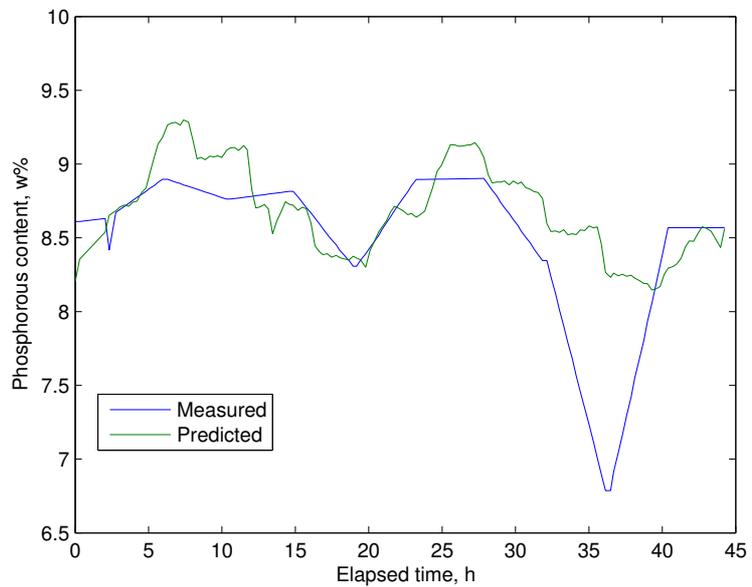
In Figure 5.2 the Ni-P deposition *phosphorus content* is presented. Again the measurements and the prediction results are plotted in one figure. The measurements were taken according to the capacity of the laboratory, thus the sample rate is varying from 2 to 5 hours. The “measured values” between the real measurements were interpolated using a linear model. Partly the model follows the measured values quite well. The large difference around hour 36 can be explained by an error in the measurement or in the plating process itself, especially errors in the preceding baths before the nickel coating bath.

To control deposition properties and achieve good and steady quality it is also important to monitor the properties of the bath, for example, the concentrations. Since the nickel concentration and also the pH value of the nickel bath is measured online, and also corrected to a desired value with the Ni-controller, it is not very important to estimate these values. However, measurements of hypophosphite and orthophosphite concentration, which also influence the deposition process, are only measured seldom in a laboratory, hence it would be practical to get online estimates for these concentrations.

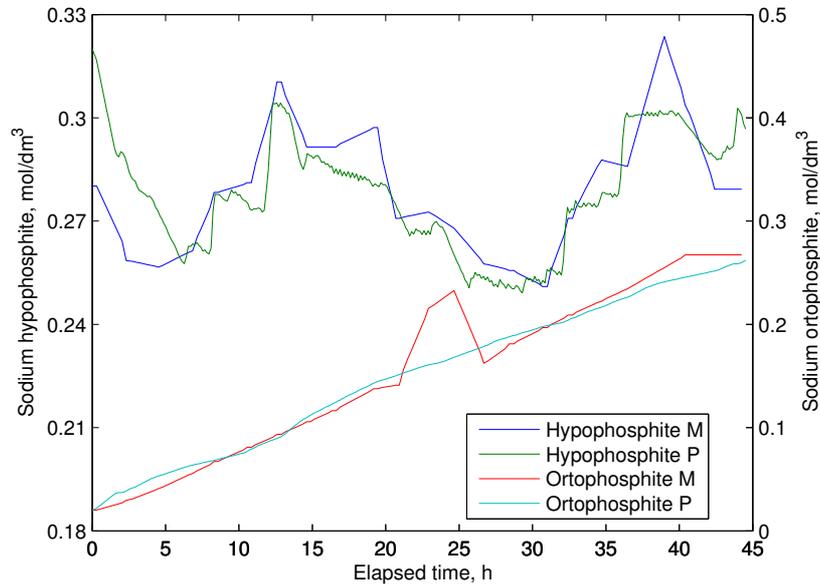
Figure 5.3 shows the measurements of *hypo- and orthophosphite concentration* and the prediction of the model. Again the measurements are interpolated linearly. The model prediction follows the measurements quite well, and the good trend behaviour confirms again the use of correct variables and parameters. The deviation between hours 20 and



**Figure 5.1:** Measured and predicted thickness of the Ni-P alloy



**Figure 5.2:** Measured and predicted phosphorus content of the Ni-P alloy



**Figure 5.3:** Measured (M) and predicted (P) hypo- and orthophosphite concentrations

25 in this experiment refers to an error during the laboratory measurements, because the trend fits perfectly and there is no additional sign for a change in orthophosphite concentration at this time.

## Conclusions

The sketched model predicts the critical alloy characteristics like deposition thickness and its phosphorus content satisfactorily. Because of the lack of measurements, it is not always possible to evaluate the model accuracy.

Not only the presented variables can be estimated by this model, but also the full process can be studied and evaluated. Kantola presents estimates for all concerned concentrations, pH value, current densities and deposition speeds, thus even for values, which are measured online and hence not that important. Because the later introduced model concentrates on the deposition thickness and its phosphorus content, and also on the hypo-

---

and orthophosphite concentration, only these results were specified here.

This model can still be adjusted, for example, by adapting its parameters according to previous online measured values. This might minimize the influence of errors and optimize the quality of the estimates. In general this model is very accurate. It has to be kept in mind that the data was collected during an active experiment, when process characteristics were changed on purpose. During normal plating the process is kept as constant as possible, which means that the model will be even more accurate than it was during the experiments.

# 6

## Plating Process as Cybernetic System

After introducing the mostly intuitive ideas of neocybernetics, the corresponding mathematical tools and presenting the setup of the industrial process, this chapter will tie these ends together and give a another solution for the modeling problem. This solution will surprise in its simplicity along with its power. Mathematics will help on the *way to practice* and eventually the *model* with its estimation *results* will be shown.

### 6.1 The Way to Practice

On the way to a real application, theoretical derivations always have to be adapted and adjusted. The measurement vector  $\mathbf{z}$  in (2.7) needs to be further studied to make it possible to capture all *inner tensions* in chemical systems. For example, the following extensions can be implemented without ruining the linear structure among the variables [3]:

- 
- **Temperature.** According to the Arrhenius formula (also used for the model in Section 5.2), reaction rates  $k$  are functions of the temperature, so that  $k \propto \exp(c/T)$ . When this is substituted, for example, in (2.4) and when logarithms and differentiations are carried out, the model remains linear if one augments the data vector and defines an additional variable  $\mathbf{z}_T = \Delta T/\bar{T}^2$ .
  - **Acidity.** The pH value of a solution is defined as  $\text{pH} = -\log c_{\text{H}^+}$ . Because this is a logarithm of a concentration variable, one can directly include the changes in the pH value among the variables as  $\mathbf{z}_{\text{pH}} = \Delta \text{pH}$ .
  - **Voltage.** In electrochemical reactions one may characterize the “concentration of electrons”. It turns out that according to the Butler-Volmer theory (also used in 5.2) the amount of free electrons is exponentially proportional to the voltage. Hence, after taking logarithms, the “electron pressure” can be characterized by  $\mathbf{z}_{e^-} = \Delta U$ .
  - **Physical phenomena.** It is evident that phenomena that are originally linear lie diffusion can directly be integrated in the model, assuming that appropriate variables (deviations from a nominal state) are included among the variables.

The vector  $\mathbf{z}$  is the measurement vector, carrying *all* possible quantities that affect the process and the system behaviour — internal system variables and external environmental variables alike. To get the actual *data vector* the vectors first have to be appropriately scaled and preprocessed. In practice, speacially if the relationships between units are not clear, it can be motivated to carry out data normalization and mean centering to make data items more compatible.

There is still the motivation to extend the available data using intuition and the neocybernetic guidelines. The more information the data vector  $\mathbf{z}$  carries in the end the higher are the chances to find the common pattern among this data.

---

## Integrals

The relative changes in the momentary deposition layer growth rate are linear functions of the other state variables. As mentioned in Chapter 4, this is the basic assumption for the status quo in controlling the alloy thickness, simply by changing the immersion time. Hence, the overall relative change is reached when one integrates the momentary rate over the plating time.

Equation (2.3) specifies the change in a concentration and hence the unidirectional flow of material. If one considers the nickel concentration, for example, it makes sense to integrate over this flow in order to get information about the actually consumed nickel. This can be done also for other variables of the process.

In addition, because of the linearity of this mapping model  $F$ , the integration can be moved “through” the model. The linearity pursuit in Section 2.2 motivates also the use of this integral. If one includes the integrals of relative changes among the variables, a linear model should be capable of capturing the layer changes around the nominal cumulation rates.

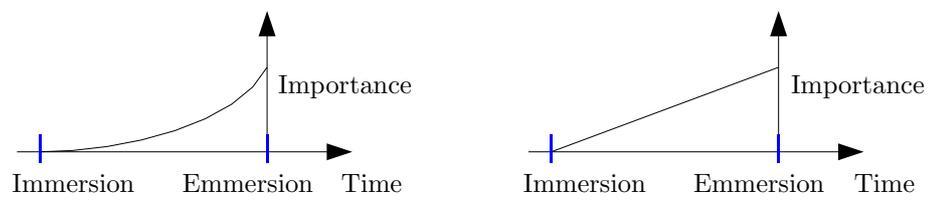
There is also a more pragmatic reason for including the integrals to the dataspace. Without the integrals one would exclusively use the data at a time stamp, where all the information (also the seldom measured alloy thickness etc.) is available, and discard the measurements taken in-between. But since the plate is immersed for some time before this very moment of immersion, it clearly makes sense to use all the values affecting the plate and its deposition layer during this time. It is somehow necessary to transport the information from between the complete samples into such a sample and so extend it. This is achieved by using the integrals.

---

## Smoothing

Yet another way to avoid the omission of the variables in-between complete samples is the use of a smoothing function among past values. This means the values among the plating time of an immersed plate are weighted from very small importance in the beginning of the actual plating process up to a prime importance in the moment of plate emmersion. This method also helps to reduce measurement noise in the data before it is taken to the modeling machinery, since building a kind of average balances the data level and cuts noise away.

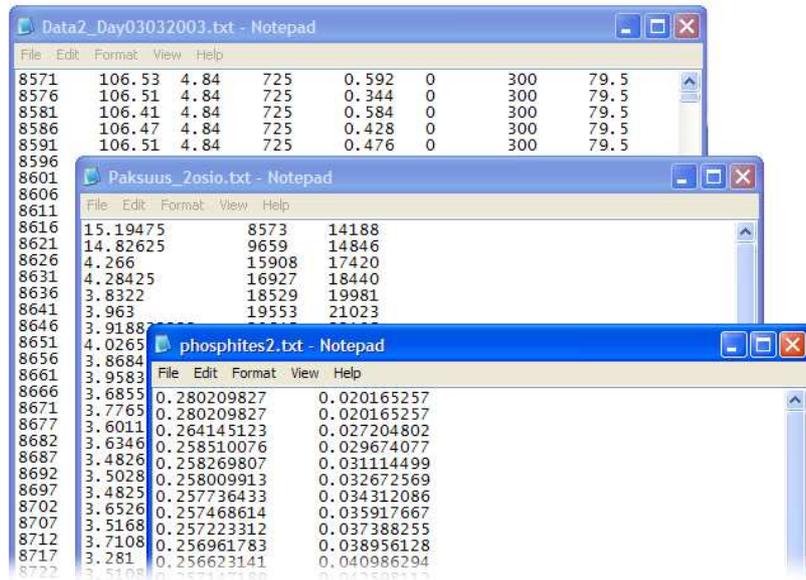
Here one can use a linear or exponential weighting. The principle is shown in Figure 6.1.



**Figure 6.1:** Weighting functions to smoothen values during the plating process

## 6.2 Available Process Data

During the research for the reports [14], [17] and [19] process data was collected. Active experiments were performed to get enough data for modeling, estimating and validation. The Ni-controller provided “online” measurements for the concentration of nickel, the pH value of the solution and the temperature of the samples. Along with this, added ammonia and nickel sulfate are recorded. The controller of the crane, which is moving the substrate from one bath to the other, records the actual plated area (further denoted as plating area), and with it the time of immersion and the time between immersion and emmersion (further denoted as plating time). The sequence of measurements was taken during almost 88 hours. So called online measurements are available around every



**Figure 6.2:** Raw process data before synchronization

5 seconds, what makes around 62800 samples; measurements of plate characteristics (for example alloy thickness) were available on average every 0.4 hours, making around 220 samples.

Figure 6.2 shows examples of the given files. Most of the important information is collected by the Ni-controller, whose data can be seen in the very back — values like time stamp, starting from an unspecified value of 8571, nickel reference value and pH value from the left side and plating area and temperature to the right can be seen. The second file shows specific information about the substrate from the laboratory, giving values for the alloy thickness, immersion and emmersion time. Another file provides the phosphorus content, also taken in the laboratory. The file in front shows measurements of hypo- and orthophosphite concentration, which are taken in a laboratory at the same time, when deposition thickness and phosphorus content were measured. Of course the measurement takes some time, but the discovered values and characteristics are here already stored with the time stamp, when the plate was taken out of the bath.

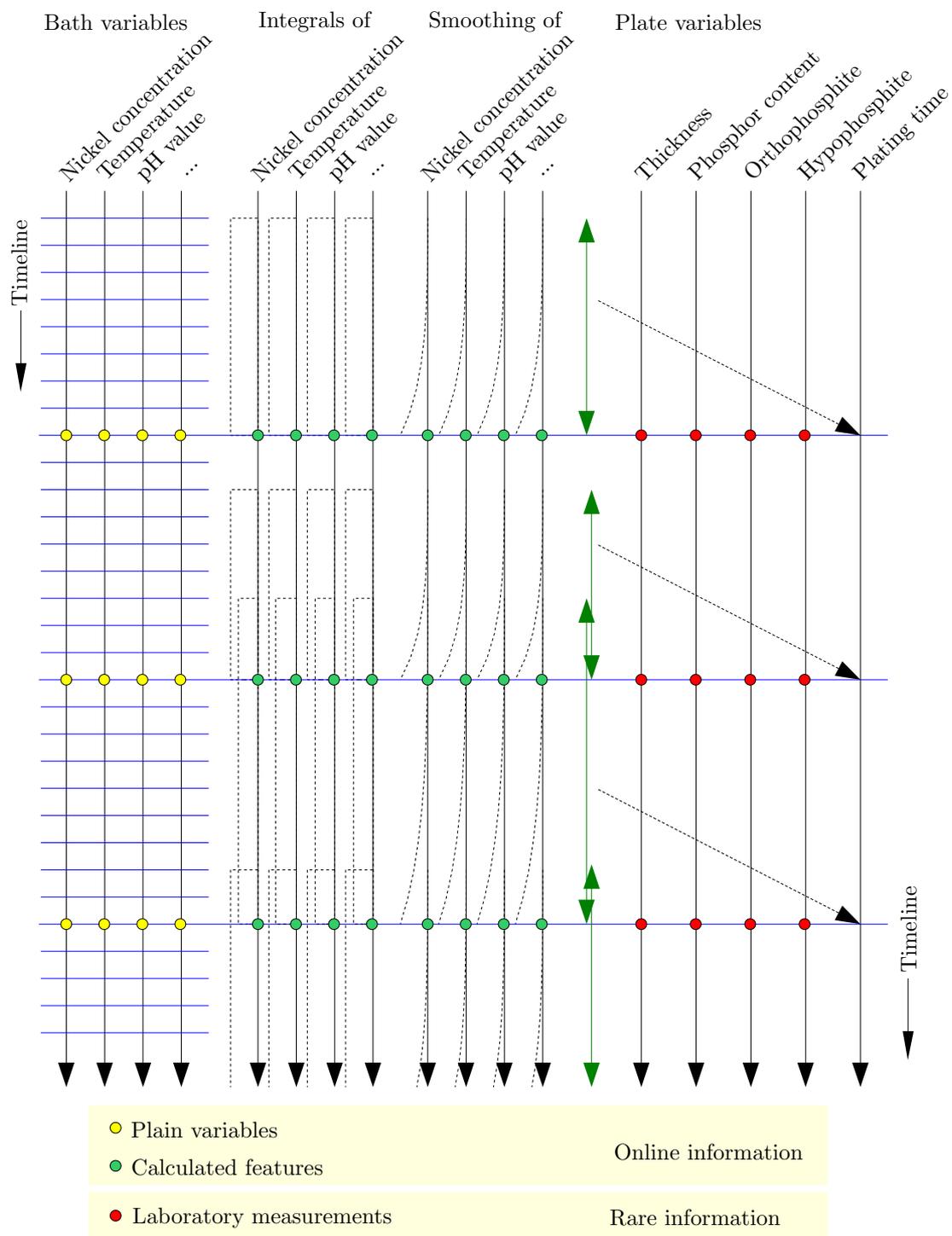
---

All the mentioned data were now collected in a single EXCEL-file, and all time stamps were matched to get a consistent timeline with the available information at each time stamp. The starting time of the measurements was set to zero. This made it possible to handle the data, keep the overview and eventually to import the process data into MATLAB for further use.

The calculations for the mentioned integrals and smoothed values were carried out in MATLAB. After including this additional information to the data, the dimension of the data vector is very high. As mentioned in Section 2.2 the structural complexity of traditional system thinking turns into high-dimensionality when thinking in cybernetic ways.

To visualize the dataset, which now can be used for all further processing steps and also as basis for a freedoms-oriented modeling approach, Figure 6.3 shows a schematic view of the synchronized data. The timeline (black) is sketched top down and along it the measurement time stamps (blue). On the left hand side one can see the online measurements for the bath characteristics like nickel concentration and temperature, followed by the calculated integrals and smoothed older values. On the right hand side the seldom measured variables of the plate characteristics, like thickness, are arranged among the timeline, apparently having a very different “sampling time” between them as compared to the continuous measurements. The information about plating time (green) is taken from a time stamp with a complete dataset and used for integration and smoothing.

The overall aim is to find a model emerging from the given data and use only online measurements, which are available online or computable during the actual process. The model will be needed to estimate values for the characteristics of the substrate to get rid of expensive and longsome laboratory measurements. This variables to be estimated are marked in Figure 6.3 with red dots.



**Figure 6.3:** Schematic view of the synchronized data sheet

---

## 6.3 Towards the Real System

After demonstrating the mathematical tools in Chapter 3 and preparing the industrial process in the previous section, one may now apply the instruments.

From now on the data space  $X$  denotes the online measurements of the bath characteristics that are considered as input data and described in Section 6.2. The other part of the measurements, like the characteristics of the plated substrate, are collected in the output block  $Y$ .

### 6.3.1 Used Data

After many talks and discussions it turned out that the measurements of the following six variables are trustworthy and could add new information to the data set  $X$ . They were already indicated in Figure 6.3.

- Nickel concentration (here as deviation from its reference value, which is  $6 \text{ mol/dm}^3$ . There is no difference to the use of absolute values — the information carried in the data is the same, since one should concentrate anyway only on deviations around an equilibrium)
- pH value
- Ammonia concentration
- Ammonia and nickel sulfate pumping
- Plating area and
- Temperature.

---

These variables are the first six elements added as  $k \times 1$  vectors. The next six vectors are filled with the corresponding integrals of the variables over plating time. These are followed by another six variables, created from the smoothed data of the corresponding variables, both likewise indicated in Figure 6.3. Linear smoothing was used at first, where values are weighted from importance 0 to 1 from the time of plate immersion up to its emersion. This  $X$  block now has the dimension  $k \times 18$ .

Along with this the output block  $Y$  consists of the four measurements of

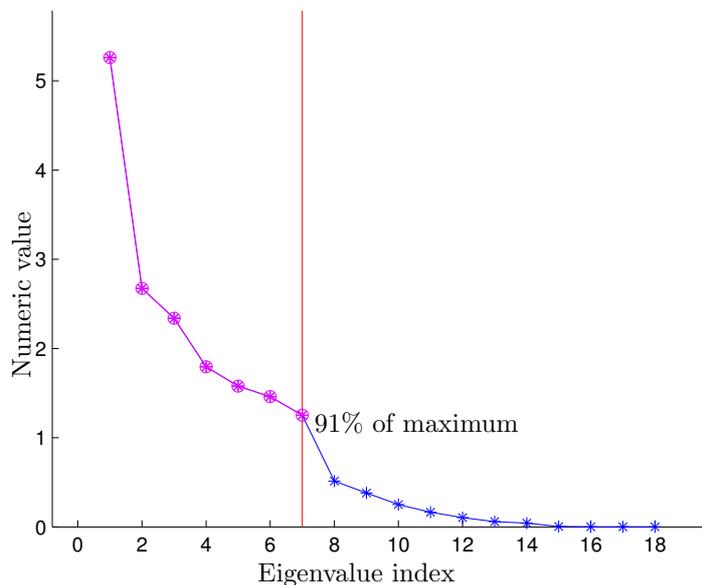
- Nickel alloy thickness
- Alloy Phosphorus content
- Concentration of Hypophosphite and
- Concentration of Orthophosphite,

hence has the dimension  $4 \times k$ .

### 6.3.2 Applied PCA Analysis

Figure 6.4 shows the output of the PCA analysis of this dataset. Since the covariance matrix is square and of the same dimension as the respective data, there are also 18 eigenvectors  $\theta_i$  and corresponding eigenvalues  $\lambda_i$ . The eigenvalues are again ordered from left to right among their numerical values. One can easily choose the new latent basis  $\Theta$  from all the 18 eigenvectors and project the data space  $X$  to another 18-dimensional space  $Z$ .

But the essence here is the visualization of the distribution of variance and/or noise in the data that is revealed by the order of eigenvalues in Figure 6.4. Of course, nobody can tell the actual difference between system structure and noise — there would be no need



**Figure 6.4:** Latent vectors  $\theta_i$  (PCA) and the numerical values of the corresponding eigenvalues (equals importance of information among data)

for all the work at hand. But according to Figure 3.3 it makes sense to discard some of the least important principal components in order to reduce not only the data dimension but also the amount of captured noise.

Somehow intuition knows to take the first seven principal components into account and to discard the rest. An “unnatural” drop in the numerical values after the seventh component could have a meaning, deviding the components carrying real information and the components reflecting almost only noise. Indeed, the first seven principal components were taken into account for modeling and they capture around 91% of the maximum information in the data when information is interpreted as variation.

To give an example, the base  $\Theta_7$ , defined by the selected  $N = 7$  eigenvectors, and the corresponding eigenvalues are presented in Appendix C.

---

### 6.3.3 Selection of Basis Vectors

The question, however, is how to determine the dimension of the latent basis? For normalized data (as it is used here) it holds

$$\sum_{i=1}^n \lambda_i = n, \tag{6.1}$$

and a crude approximation is to include only those first  $N$  latent vectors  $\theta_i$  in the model for which there holds  $\lambda_i > 1$  [6]. Those directions carry “more than the average amount” of the total information. However, the overall behaviour of the eigenvalue scope should be taken into account. Plotted in descending order there might be a significant drop between some of them — this may suggest where to put the model order.

As a general rule, it can be argued that the directions of the largest eigenvalues are the most important, whereas the effects of noise are pushed to the later principal components. However, analysis of the later components may also reveal some peculiarities in the system structure, and this information should not be automatically rejected.

If the first principal component dominates excessively, it may be reasonable to check whether the data preprocessing has been successful: for instance, if the data is not mean-centered, it is this mean that dominates among all other information rather than the true data variation. Especially if the numerical data values are far from the origin, as was the case in the pressure example in Subsection 3.1, the large mean dominates. The absolute minimum eigenvalue is zero, meaning that the set of measurements is linearly dependent or there are too few measurements, so that  $k < n$ . Note that the PCA type of data modeling can still be carried out, whereas simple MLR would collapse [6].

If there exists eigenvectors with exactly equal eigenvalues in the covariance matrix, the selection of the eigenvectors is not unique. This is specially true for whitened data; PCA can find no structure in whitened data [6].

---

The reader may notice that after data preprocessing, the selection of the basis vectors for the model is the second crucial step. Appropriate preprocessing, as well as selection of principal components, effects the model essentially and therefore needs some experience and also many experiments. As it turns out though, even the first guess of taking seven principal components is very good — changing the number of used components indeed affects the model, but only slightly. All this also applies for the use of PLS and its results, which will be presented later.

### 6.3.4 About the Model

According to the algorithm presented in 3.5, the mapping  $F^1$  should be calculated to derive the latent variables  $Z$ . Since the new base  $\Theta$  is orthonormal,  $F^1$  equals  $\Theta$  and is of the dimension  $18 \times 7$ . Therefore, the intermediate dataspace  $Z$  has the dimension seven.

Multilinear regression (see Section 3.4) finds the second-level mapping  $F^2$  in minimizing the sum of squared errors. This matrix projects the 7-dimensional space  $Z$  onto the 4-dimensional output space  $Y$ , and hence has the dimension  $7 \times 4$ .

The linear model  $F$  in (3.14) to estimate the output data is now the product of both matrices, giving

$$F = F^1 F^2 \tag{6.2}$$

of the dimension  $18 \times 4$ .

The matrix  $F$  holds for every input variable the information about its importance for the output variables. A linear combination of the input sample directly gives the estimate of the actual output value. For the estimate of the variable thickness, for example, this is

---

done with  $\mathbf{F}_1$ , the first column of the mapping  $F$ .

$$\mathbf{F}_1^T = \begin{pmatrix} 0.2174 & 0.1742 & -0.0176 & 0.0016 & -0.0439 & -0.1406 & \dots \\ 0.2121 & 0.1733 & 0.0072 & 0.0385 & -0.0076 & 0.1124 & \dots \\ 0.2133 & 0.1733 & 0.0148 & 0.0032 & 0.0104 & 0.2129 & \dots \end{pmatrix}. \quad (6.3)$$

The vector is already sorted in groups of six, to match the accoring values in the input data.  $F_{1,1}$  to  $F_{1,6}$  are weights for the plain variables of the input sample.  $F_{1,7}$  to  $F_{1,12}$  refer to the values formed by the integral over plating time and  $F_{1,13}$  to  $F_{1,18}$  connect the smoothened values to the output.

It turns out that information about the *nickel concentration* in the bath and the *pH value* have highest significance in all its three versions (plain, integral, smoothed). Also *temperature* carries a great deal of information. Least significant is the actual pumping, which is becoming more important with its integral, but is still far from relevant. Nickel concentration, pH value and temperature carry in all its specifications more than 90% of the information for the estimate of thickness.

The actual plated area, which is the loading of the bath, obviously does not affect the thickness of the plated substrate itself. Plating area is projected into the thickness estimate respectively by the fifth value of each group in (6.3),  $F_{1,5}$ ,  $F_{1,11}$  and  $F_{1,17}$ . This makes sense, since the nickel is distributed evenly in the bath and new nickel sulfate is added to the bath continuously to maintain the available amount of nickel ions. It does not matter, how many plates are immersed at the same time, their alloy will have (almost) the same thickness.

This is already a very important and interesting result. If it turns out that this data-based method of modeling and estimating is feasible for the process of nickel plating, the linear model  $F$  could shed light on the necessary measurements. Since good and reliable

---

measurements are always very expensive, one could learn from the model which values are really needed and which could be omitted without harming the estimation too much. Eventually the model was used to estimate process variables. To take the already started line, the derived base of 7 vectors and the resulting mapping  $F$  was used for the following results.

### 6.3.5 PCR Estimation Results

Throughout the following pages the style of presenting results will be the same. The data was separated in three parts, the first one used to estimate the model  $F$  and the other two parts to validate the model. This means that the results along the estimation dataset can be very good, because the algorithm of MLR minimized the error between measurement and estimate exactly for that data. A good result here might not tell anything about the quality of the model, because even adding a random variable to the dataspace  $X$  would improve the result here. That is why there is validation, which uses the same model  $F$ , but the validation data *was not seen by the modeling machinery before*.

Where there was data available there is also the result of the electrochemical model (Section 5.2) presented. This model was also calibrated using the first dataset and validated with the other two parts. Since Kantola [14] used the measurements the same way a direct comparison is possible. Original measurements (blue), electrochemical model estimates (green) and data-based model estimates (red) are plotted among the time in one plot.

The measure of fit is carried out mainly visually, but also with the *sum of squared errors*

$$SSE = \frac{1}{k} \cdot (\hat{\mathbf{Y}}_{\text{est}, i} - \mathbf{Y}_{\text{est}, i})^T (\hat{\mathbf{Y}}_{\text{est}, i} - \mathbf{Y}_{\text{est}, i}). \quad (6.4)$$

The overall SSE is calculated from both validation experiments as weighted sum and taken as the most important criterion for the quality of the final model.

---

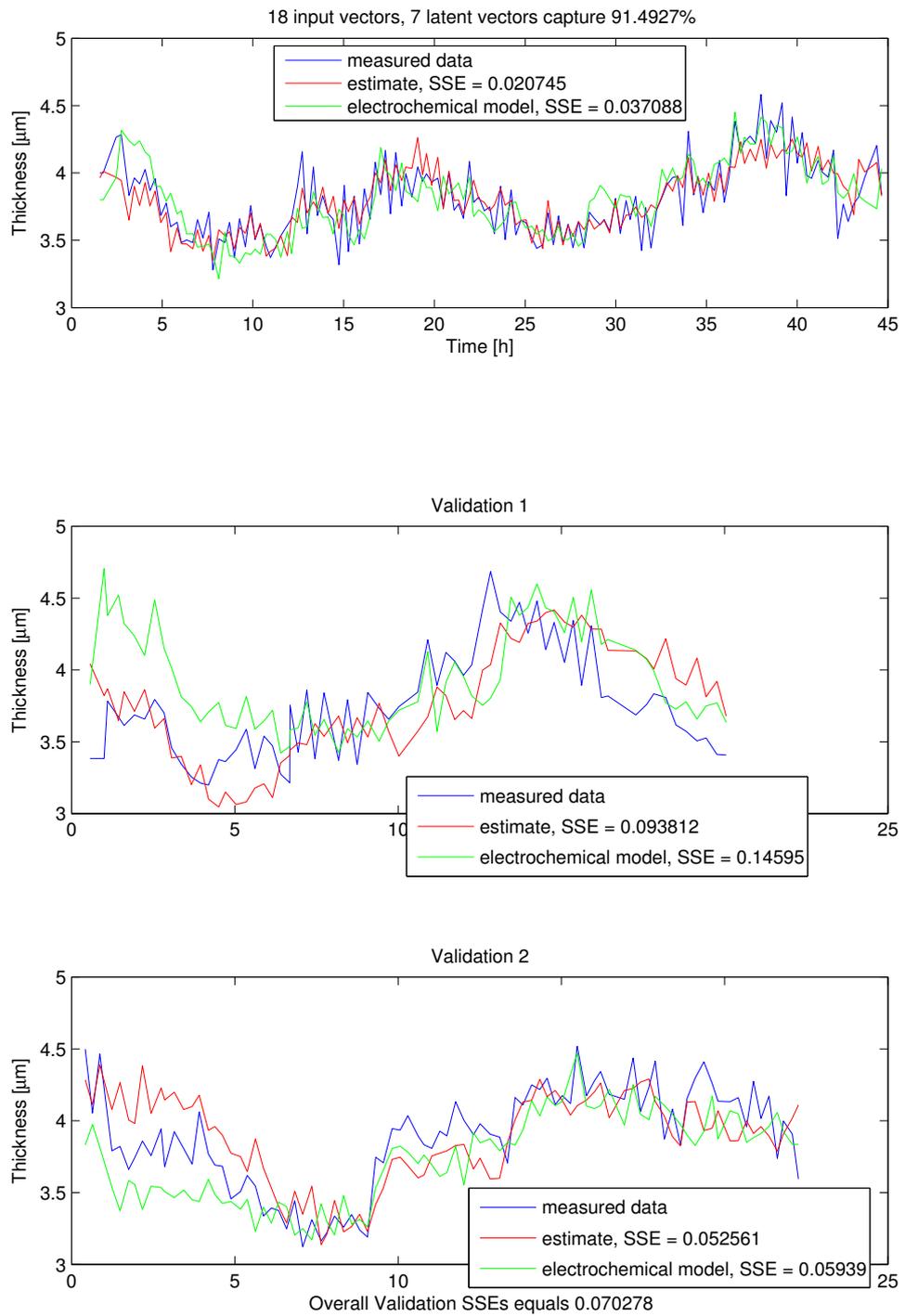
Figure 6.5 shows a described plot, in this case for the *thickness of the nickel alloy*. The first subplot shows that the electrochemical model follows the data very well. Also the linear model  $F$  produces estimates, which are very close to the real data, and according to the SSE-criterion, even closer. However, as said above, this information does not tell enough about a model's accuracy. The two following subplots show more measured data and again the prediction of the two models. In the validation set 1 the linear model  $F$  shows better behaviour than the electrochemical model, in validation set 2 it is very similar.

The *alloy phosphorus content* in measurement and prediction is shown on Figure 6.6. Again the model accuracy ranks around the existing electrochemical model. The model finds the data level, but in this case cannot predict variations very well. In some spots and areas, the data-based model is more accurate than the electrochemical one.

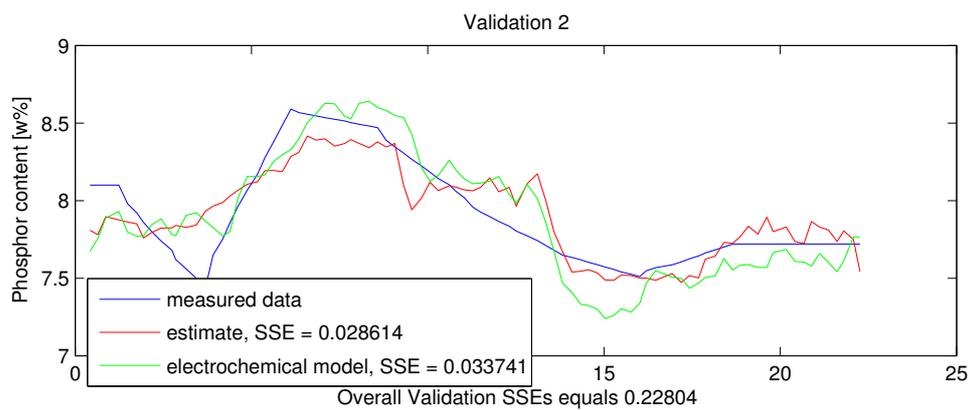
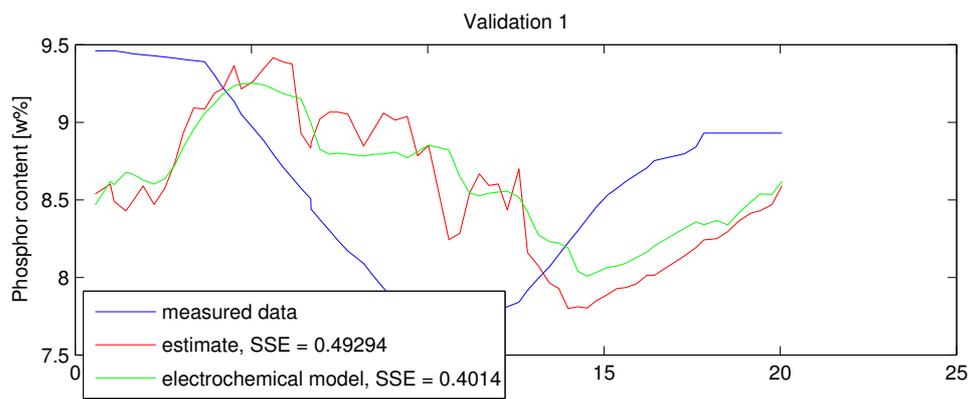
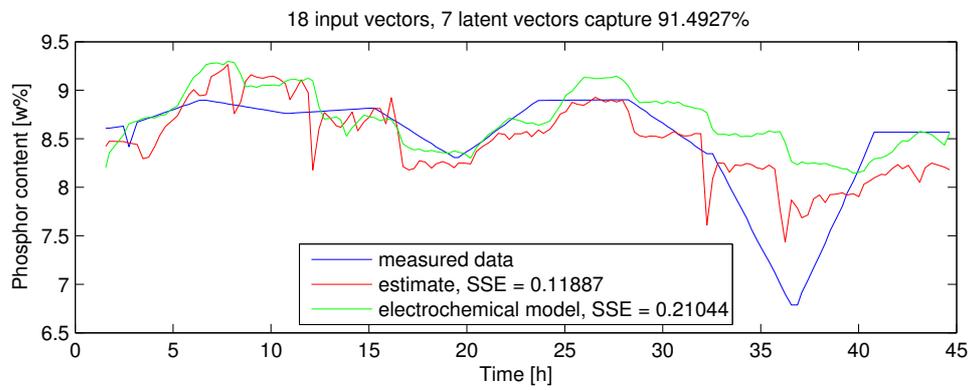
There are some notable key observations:

- The result of the linear data-based model is very good (considering the complicated process it models and estimates).
- Especially variations of level and steps are modeled accurately.
- If the estimate differs from the measurement, the electrochemical model differs as well (keeping in mind how complex the structure of the latter model is).
- There is still room for improvement using other multivariate tools.

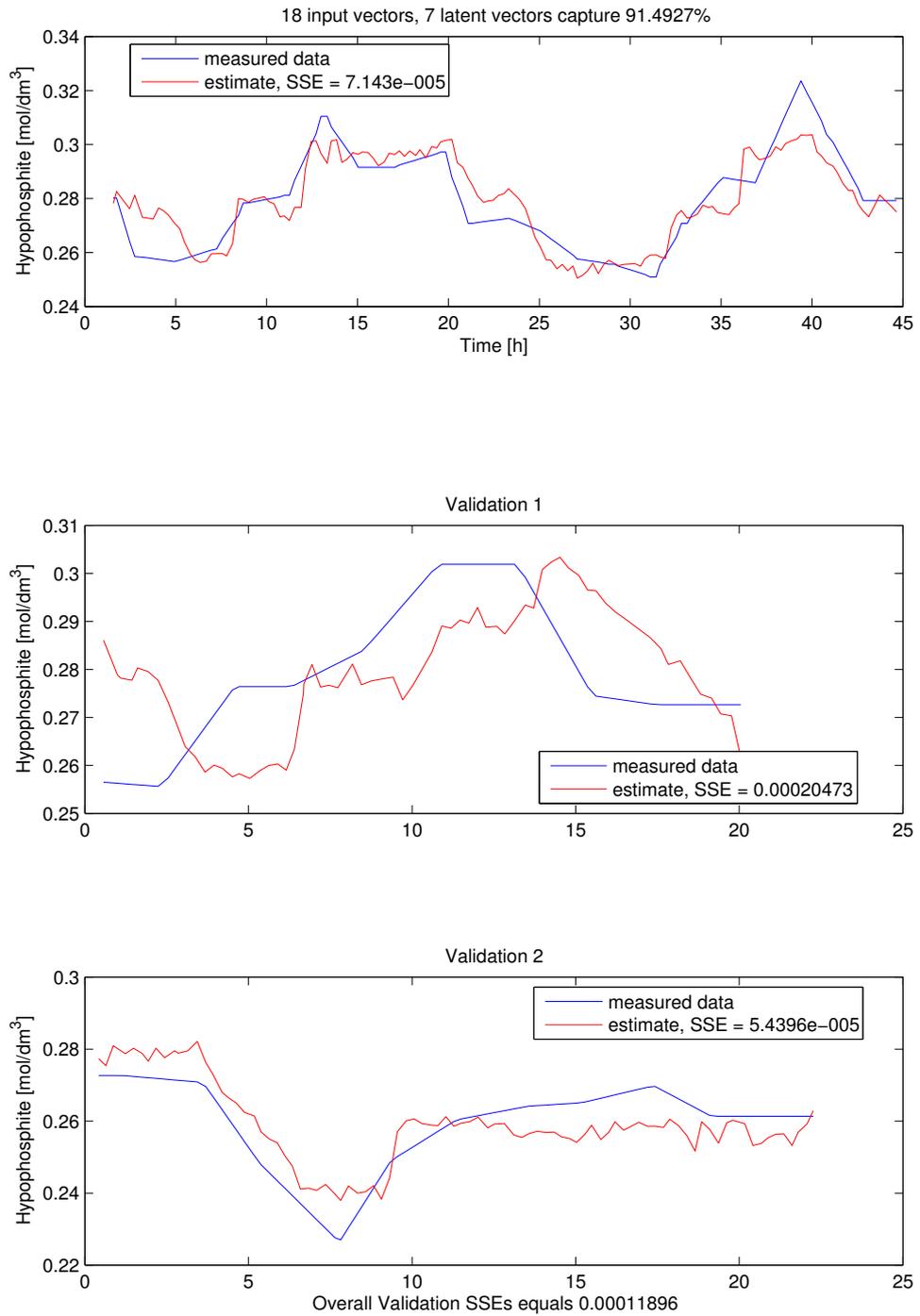
Figure 6.7 shows the estimate of the bath characteristic *hypophosphite concentration*. Here no estimates from the electrochemical model were available. Clearly the estimation worked quite well and the model predicts the hypophosphite concentration accurately. Looking at the second subplot one might expect a timeshift problem: the model seems



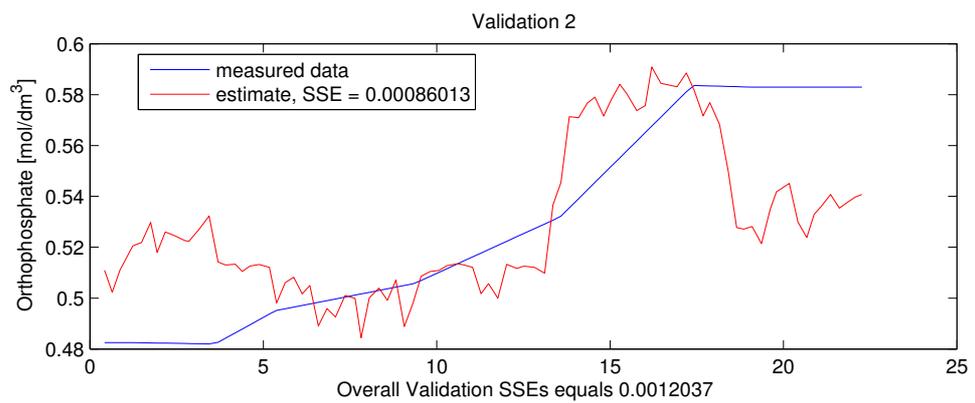
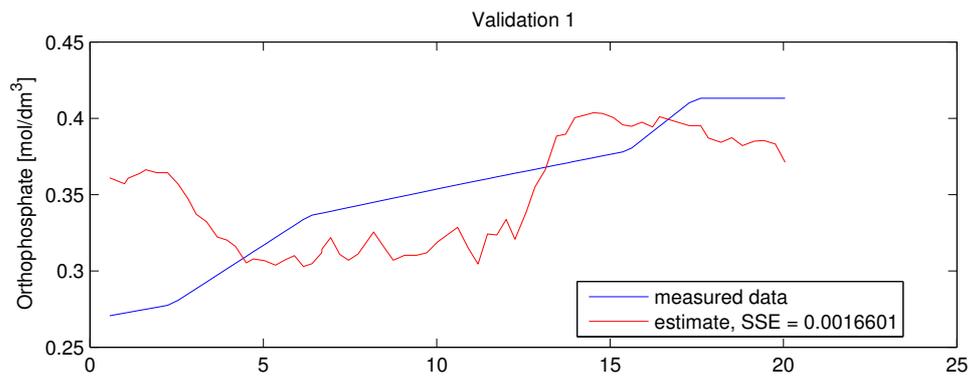
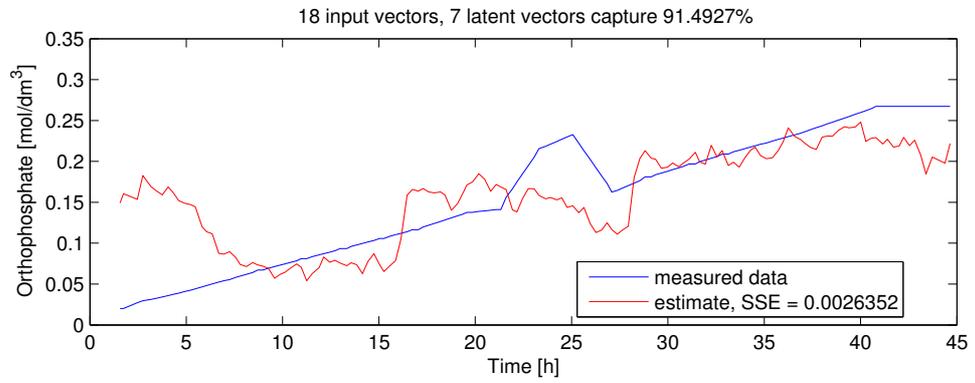
**Figure 6.5:** Alloy thickness: measured data and two compared estimates



**Figure 6.6:** Alloy phosphorus content: measured data and two compared estimates



**Figure 6.7:** Hypophosphite concentration: measured data and two compared estimates



**Figure 6.8:** Orthophosphate concentration: measured data and two compared estimates

---

to predict variations around 2 hours too late. The validation set 2 is very well estimated, the model follows even the large step off the average around hour 8.

The other critical bath characteristic, *orthophosphite concentration*, can also not be compared to other model data. However, the databased model fails almost completely in explaining variations in orthophosphite concentration. The average level is estimated well enough, but the continuous increase is not reflected in the model estimate.

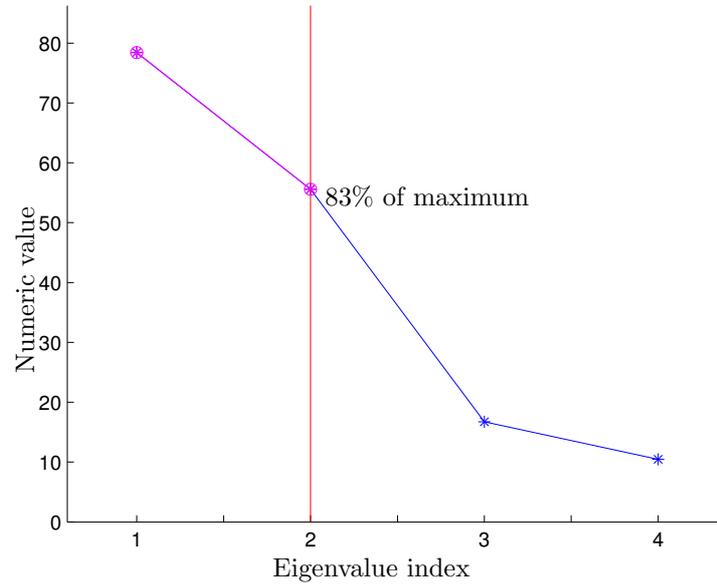
The following Section 6.4 shows the improvement which is still possible. PLS will be applied and the input space  $X$  will be varied to achieve better results and minimize the deviation from original data.

## 6.4 Final Model

In the previous section the first multivariate methods were applied and the first results were shown. To improve the behaviour of the model one can still apply PLS and make some experiments in the set of input variables.

### 6.4.1 Applied PLS Analysis

Using the same data setup as above, the PLS procedure offers four latent vectors. As explained in Section 3.6 the PLS algorithm produces generally lower dimensional models than PCA analysis, because the utilization of the output dataset lowers the rank of the correlation matrices. Figure 6.9 shows the result of the PLS analysis of the 18-dimensional input set and the 4-dimensional output set. The numerical values of the eigenvalues are larger than in the PCA case, because the multiplication  $Y^T X X^T Y$  produces very large matrix entries.



**Figure 6.9:** Latent vectors  $\theta_i$  (PLS) and the numerical values of the corresponding eigenvalues (equals importance of information among data)

It is possible, to choose only two latent vectors to capture 83% of the information hidden in the variation of the input data. The basis  $\Theta_2$  would produce an intermediate dataspace  $Z$  of the small dimension  $k \times 2$  containing all necessary information. This remarkable reduction when compared to the PCA approach is feasible, because now the algorithm *knows* what kind of features it should search for among the input information.

After selecting the two most significant vectors as a new basis for the dataspace  $X$ , it is projected using this basis into  $Z$  of dimension  $k \times 2$ . The mapping to the final space  $Y$  has the dimension  $2 \times 4$ , hence the overall linear model  $F$  apparently again has the dimension  $18 \times 4$ .

As an example the first vector  $\mathbf{F}_1$  of the linear model is presented in the same format as

---

above to give a clue about its composition:

$$\mathbf{F}_1^T = \begin{pmatrix} 0.2005 & 0.1994 & 0.0041 & -0.0083 & 0.0063 & -0.0206 & \dots \\ & 0.1976 & 0.1992 & 0.0130 & 0.0326 & 0.0202 & 0.0458 & \dots \\ & & 0.1966 & 0.1970 & 0.0190 & 0.0356 & 0.0235 & 0.0487 \end{pmatrix}. \quad (6.5)$$

Again nickel concentration in the bath and pH value have the largest significance. Obviously it is no longer necessary to use the information about temperature in the bath. The importance of temperature is now on the level of the other “unuseful” measurements, like the loading of the bath and the currently plated area, which both do not effect the thickness of the nickel layer.

It might also be interesting to look at the importance of the input data for the output variable phosphorus content. This information is stored in the second column of  $F$ :

$$\mathbf{F}_2^T = \begin{pmatrix} -0.1545 & -0.1985 & 0.0010 & 0.0091 & 0.0005 & 0.0185 & \dots \\ -0.1516 & -0.1982 & -0.0046 & -0.0273 & -0.0096 & -0.0402 & \dots \\ -0.1507 & -0.1963 & -0.0100 & -0.0289 & -0.0113 & -0.0428 \end{pmatrix}. \quad (6.6)$$

Here also, nickel concentration and pH value ( $F_{2,1}$  and  $F_{2,2}$ ) account mainly for the phosphorus content in the nickel alloy. It is notable that the information about these two variables comes into the model negatively. This means that with higher nickel concentration and higher pH value the phosphorus content decreases.

Industry is not interested in spreading information about their already working processes in nickel plating. However, the dependencies between a change in bath characteristics and the resulting changes in plate attributes are known from many experiments and the long experience of operators. According to consultations with domain experts and their information the described behaviour of the linear model and the patterns it reveals obviously match this experience and this fact corroborates the accuracy of the model  $F$ .

---

## 6.4.2 PLS Estimation Results

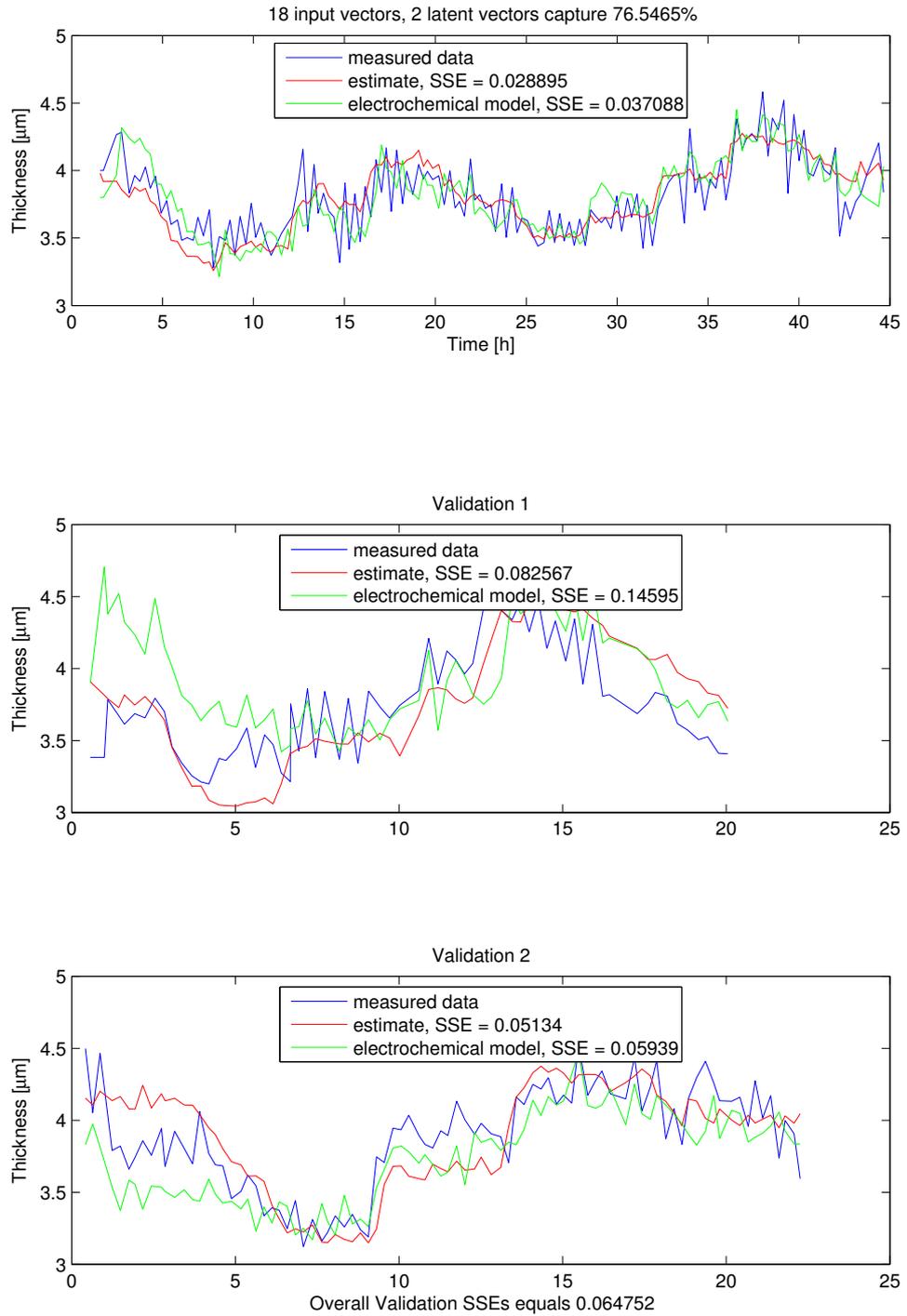
How did the results change using a different method? The model based on PCA analysis generated a good model, which already could compete with a highly complex electrochemical model. Now, using PLS, the gap between input and output is bridged and so the result improved.

The model is now based on a subspace of  $X$ , which has only the dimension of  $k \times 2$ . Note that with two selected latent vectors all important information of  $X$  is captured instead of using seven latent vectors as it was necessary after the PCA analysis. The results are presented in the same framework as earlier.

Figure 6.10 shows the *Ni-P alloy thickness* and estimates from the PLS based model and the electrochemical model. Again the simple linear model beats the constraints-based model visually and also by means of SSE. The linear model is definitely better if talking about level: validation set 1 shows that the electrochemical model needs longer to adapt to the real data level and to reach reasonable values. The first impression shows no improvement as compared to the results of Section 6.3.5, but in means of SSE this approach could enhance the estimate.

The *phosphorus content* in measured data, along with linearly and electrochemically estimated data is given in Figure 6.11. The phosphorus content is not better estimated if it comes only to SSE. The linear estimate increases its sum of squared error a great deal in validation set 2 if compared to the PCA case. However, a visual examination and comparison will reveal a better and smoother behaviour not only in comparison to real data, but also in comparison to the electrochemical model.

Generally the values of the linear estimate are now smoother than those from the PCA based model. For this see the estimation dataset for phosphorus content in Figure 6.11 and compare it to the estimation of this variable in the previous case (Figure 6.6).



**Figure 6.10:** Alloy thickness: measured data and two compared estimates

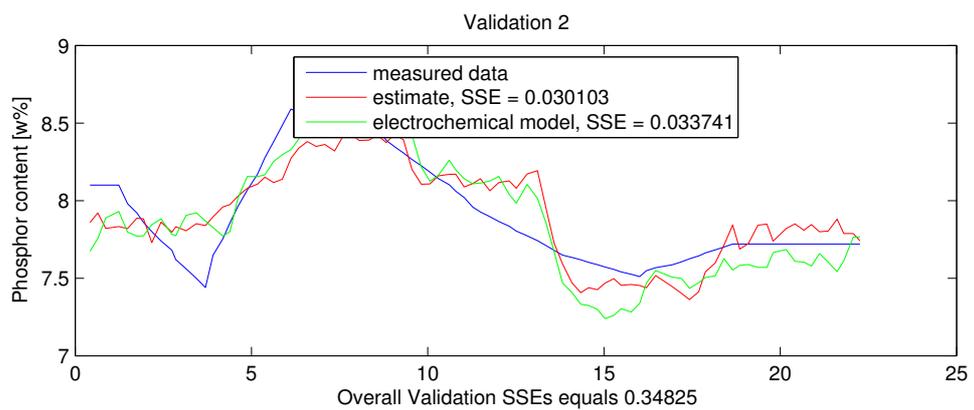
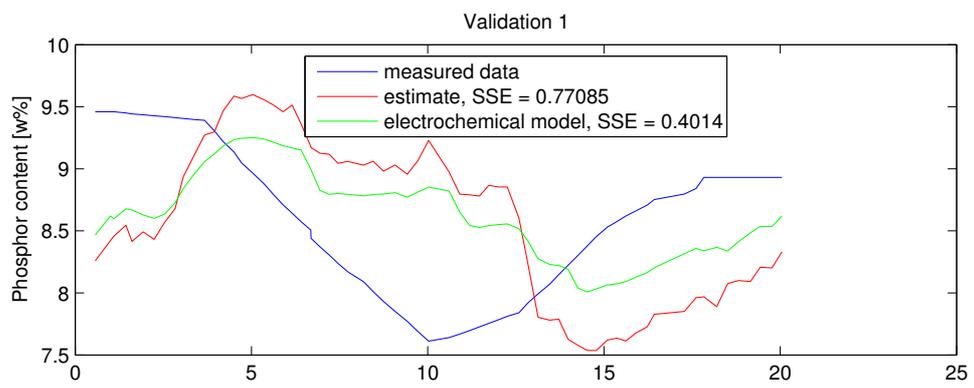
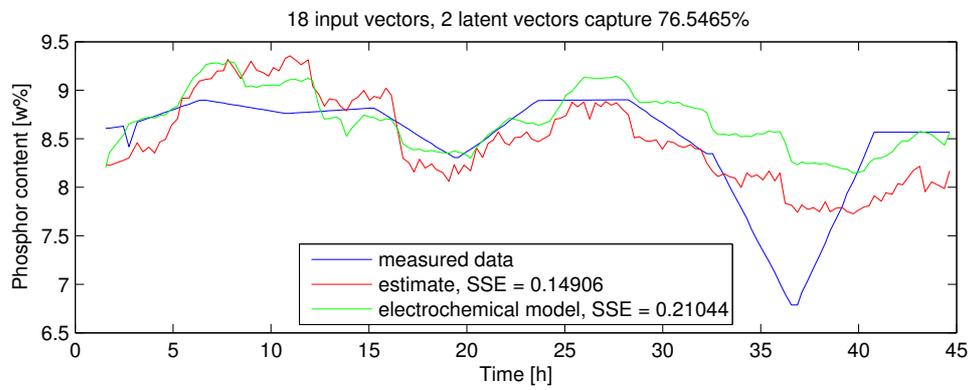
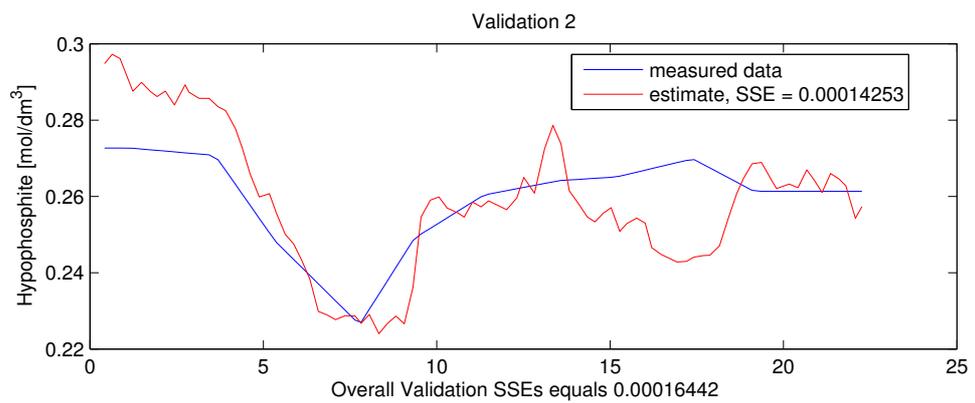
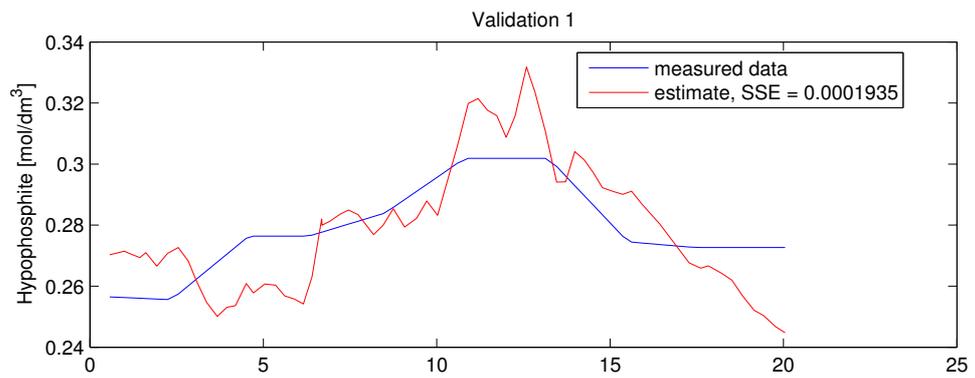
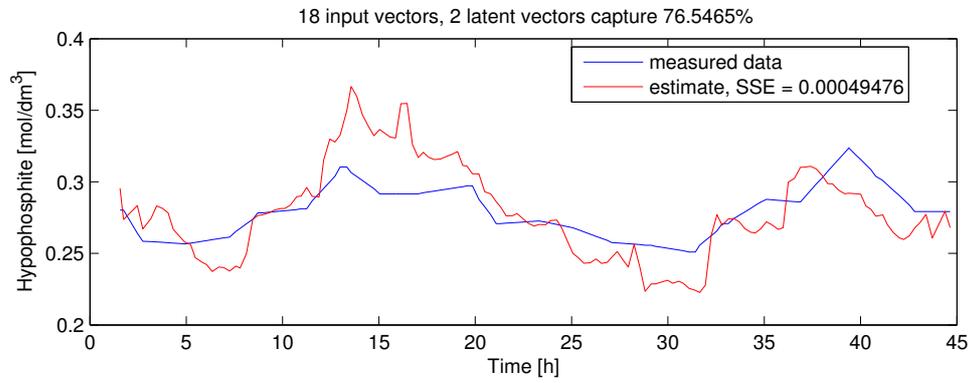
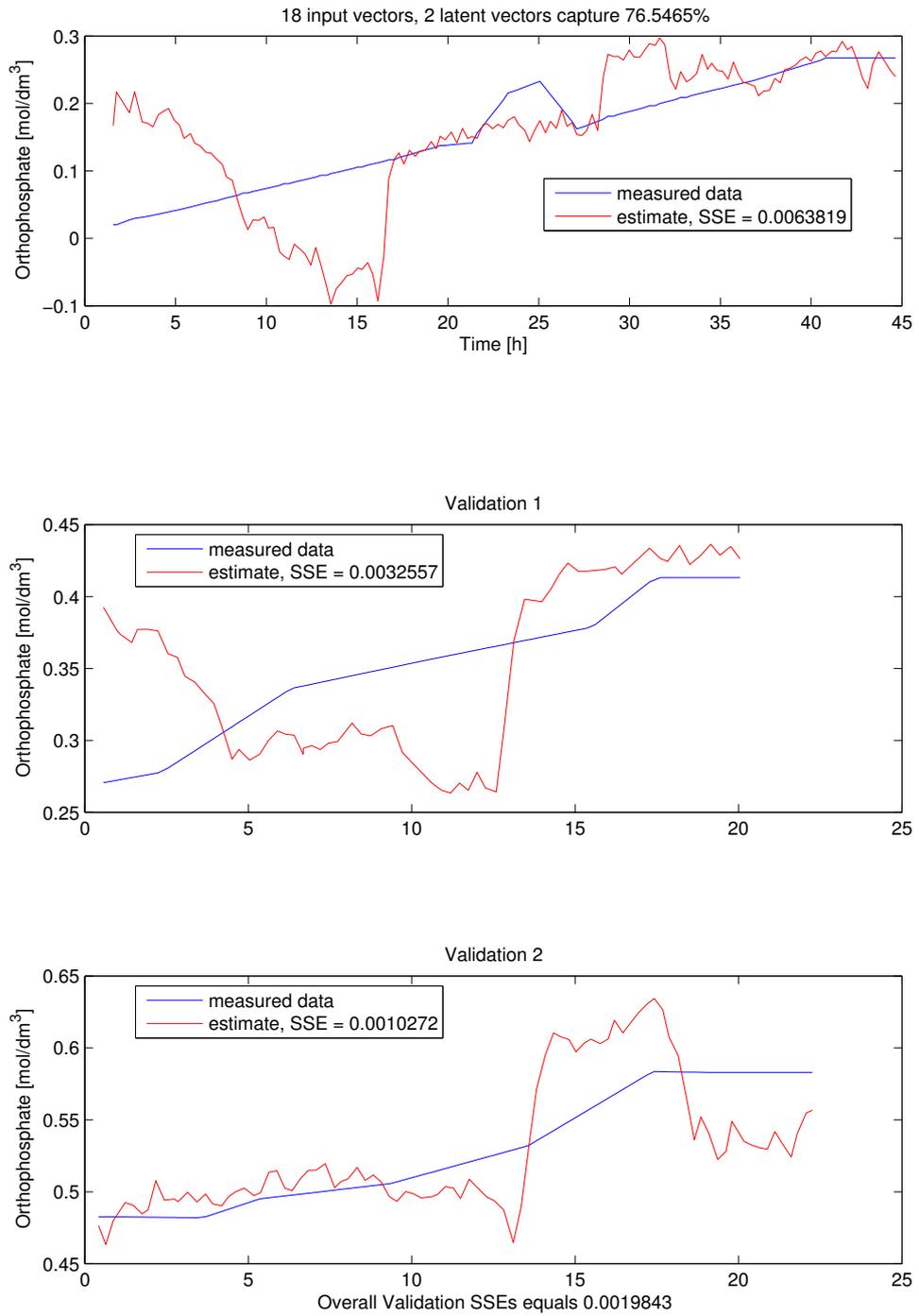


Figure 6.11: Alloy phosphorus content: measured data and two compared estimates



**Figure 6.12:** Hypophosphite concentration: measured data and two compared estimates



**Figure 6.13:** Orthophosphate concentration: measured data and two compared estimates

---

For the *concentrations of hypophosphite and orthophosphite* in Figure 6.12 and Figure 6.13 there is also no significant improvement. Still the estimates are sometimes too late, sometimes trying to model variations where according to the measurements nothing had happened. Not even the PLS approach could find a pattern among the input variables that reveals the behaviour of these concentrations.

Obviously this method is more feasible to estimate plate characteristics than bath characteristics. Since the aim was to model and estimate the behaviour of the alloy characteristics, however, this modeling approach gives a very accurate alternative to a complex constrained based model.

To summarize, the following Section 6.5 will give a discussion about the obtained results, variations in the model, and its features.

## 6.5 Discussion

The conclusion is rather simple. Using neocybernetic ideas, despite its simplicity one can achieve very accurate results. The very simple model for this complicated process turned out to be equivalent and even better than the complex model. Freedoms-oriented thinking helped to find the underlying patterns in the available data and made it easy to estimate hidden characteristics. It took much more effort to obtain this result on the traditional way of constraints-oriented thinking.

Furthermore, this model also tells which measurements are necessary and which variables have no effect on the characteristics of the plated substrate. This may help to save money while monitoring the process and its variables by avoiding expensive measurements. It was, for example, possible to use only measurements about nickel concentration, pH value, and temperature, and still only  $N = 4$  latent vectors  $\theta_i$  could capture 98% of information. The difference to the presented results is marginal.

---

Attempts to improve the presented results to a greater extent mostly failed or did not add significant enhancement. It turned out that some modifications and increments of the input data space of online available variables improved the overall result, and some did not. The best results are presented in this work by means of best possible fit.

The biggest step was taken by adding the integral information. Using not only the values of the actual time stamp, but also all available information during the plating time, helped to improve the sum of squared error. There was also visually improvement detectable, especially in the estimation for phosphorus content.

Adding the smoothed values did not improve the result significantly. The information added by this approach was obviously already provided by the integral feature. In addition PCR as well as PLS also remove noise from data, so the smoothing did not help in this task. Maybe a different way of smoothing, for instance with an exponential “forgetting” function, might improve the overall behaviour.

Since the process is based on electrochemical reactions, it might be helpful to include information about the actual flow of current or the actual potential in the bath. These values are unfortunately not measured accurately, hence the measurements were not used in this data-based modeling approach. However, it is possible to calculate these voltages, using the Nernst equation (see Appendix D). Also this method could not improve the estimation result crucially, neither by means of SSE nor visual examination.

This is reasonable as the reader may have noticed. Adding information, which is calculated out of already included information, makes no sense. The algorithms PCR and PLS are searching for underlying patterns in the dataspace, and a combination of data added to this space cannot give more information. One just adds more redundancy, which is filtered out again when using the new basis  $\Theta$ . Really measured information about the voltage in the bath could help to improve the result, because it might add fresh information to the algorithms.

---

All these improvements and degradations are in the range of very small deviations. Indeed, the sum of squared error for the thickness estimation improved by 12% when using PLS instead of PCR while even reducing the information in terms of variation to two instead of seven dimensions. All other attempts though produced only a difference of few percent.

Generally it is very remarkable that if there is deviation between the model estimation and the measured values that not only the model at hand produced this difference, but also the electrochemical model had difficulties to estimate correctly. This might be another hint for errors in the measurement data, caused by mistakes during the actual measurements. But there are also some parts where the linear model deviates in the different direction than the electrochemical model — this might be caused by the dynamic behaviour this electrochemical model implements.

# 7

## Future Prospects

The work and research at the Control Engineering Laboratory provided me an insight into new ways of thinking. It was interesting to read about neocybernetics and follow the advancements of its ideas. They give a new background in system thinking and reveal a new area of modeling complex systems. This thesis tried to encompass the understandable key points and arouse interest in the reader.

The industrial process of nickel plating was no less interesting than the neocybernetic theory. It provided an insight into a new chemical domain and gave me the opportunity to enter a new scientific field. Furthermore, the way of combining these two scientific fields kept me intrigued. All assumptions and expectations given before any data analysis or system modeling were in the end confirmed and even exceeded. The model proves that neocybernetic ideas can be applied to this kind of chemical process with success.

Through simplicity and linearity a way was found to model the complicated nickel plating process. A linear model for monitoring and estimating plate characteristics was presented and explained. It was as accurate, if not better, than a model derived through traditional thinking — this underlines the accuracy of all assumptions. The model is easy to understand, easy to use, and easy to adapt. Recall the introducing quote of Norbert

---

Wiener; the best way is not always the modeling approach closest to the cat, closest to the system itself. Stepping back and using simple tools can even produce better and more understandable results.

The important tasks of estimating the *Ni-P alloy thickness* and its *phosphorus content* were fulfilled. The other unknown variables were not modeled accurately enough — but also other models failed to estimate the concentrations of hypophosphite or orthophosphite accurately.

There is still work to do. Since there is a great deal of data available and the results vary slightly among the use of data and the proposed additional features, the model has to be fixed before further use. The following must be defined: which data should be used in the end and which measurements are crucial and absolutely necessary for the model.

In the next step the model can be applied to the real process and its accuracy can be tested by comparing the estimated variables to the real characteristics of the plated substrate. The model could be opened for adaptation among measurements if it turns out that the parameters in the mapping  $F$  are not properly calculated. Also a combination of the linear model and the electrochemical model [14] might improve the behaviour of either one.

The solution presented in the thesis at hand along with the solution presented by Kantola in [14] might give companies the possibility to control plate characteristics during the actual plating process. This would reduce expenses during the process and lower the price of production. Furthermore, it would make the process and its products more accurate and increase quality.

The future will tell if this approach can be used to satisfy all these expectations and if it will make the step from a scientific work to an actual application on a real-life process.

# Appendix A

## Example

### Constraints vs. Freedom

Assume that the available variables are measurements of some time-domain signal  $y$ , so that samples are indexed as  $y(\kappa)$ ,  $y(\kappa - 1)$ , etc. Further assume that there are three variables that are connected by a model

$$\begin{cases} y(\kappa) = ay(\kappa - 1) \\ y(\kappa + 1) = ay(\kappa). \end{cases} \quad (\text{A.1})$$

The difficulty here is that one does not know beforehand whether some of the variables are redundant. The data vectors are three-dimensional:

$$\mathbf{v}(\kappa) = \begin{pmatrix} y(\kappa - 1) \\ y(\kappa) \\ y(\kappa + 1) \end{pmatrix}. \quad (\text{A.2})$$

In this case the constraint vectors without normalization are

$$\Gamma = \begin{pmatrix} a & 0 \\ -1 & a \\ 0 & -1 \end{pmatrix}. \quad (\text{A.3})$$

---

The whole data space  $\mathcal{S}$  is spanned by the constraints and the degrees of freedom:

$$\mathcal{S} = \left( \Gamma \mid \Theta \right). \quad (\text{A.4})$$

In this example apparently the constraints span a two-dimensional null-space in the three-dimensional variable space. The remaining dimension, which is the degree of freedom, can be solved by orthogonalization, for example applying the Gram-Schmidt procedure according to [21]:

$$\left( \begin{array}{cc|c} a & 0 & 1 \\ -1 & a & 0 \\ 0 & -1 & 0 \end{array} \right) \longrightarrow \left( \begin{array}{cc|c} \frac{a}{\sqrt{1-a^2}} & \frac{a^2}{\sqrt{(1+a^2)(1+a^2+a^4)}} & \frac{1}{\sqrt{1+a^2+a^4}} \\ \frac{-1}{\sqrt{1-a^2}} & \frac{a^3}{\sqrt{(1+a^2)(1+a^2+a^4)}} & \frac{a}{\sqrt{1+a^2+a^4}} \\ 0 & -\frac{(1+a^2)}{\sqrt{(1+a^2)(1+a^2+a^4)}} & \frac{a^2}{\sqrt{1+a^2+a^4}} \end{array} \right), \quad (\text{A.5})$$

where the 3 vectors are now not only orthogonal but also orthonormal. This means according to (A.4) that the model  $\Theta = \theta$  becomes

$$\theta = \begin{pmatrix} 1 \\ a \\ a^2 \end{pmatrix} / \sqrt{1+a^2+a^4}. \quad (\text{A.6})$$

The ‘‘axis of freedom’’ clearly has an *exponential outlook* in the data space. This is an exact correspondence with the time-domain behaviour of the system that is characterized in (A.1). The degrees of freedom determine ‘‘behavioural fragments’’, so that actual observations can be constructed from combinations of them.

If the data vector  $\mathbf{v}$  would be much larger, also the vector of the model would become larger. But the pattern of the exponential outlook will remain easily visible, whereas in the huge data vector nothing can be seen.

# Appendix B

## Eigenproblem Properties

The formulation of the eigenvalue problem (3.5) in Section 3 can be studied closer to extract some very useful properties. These properties are the basis for all the helpful applications of PCR and PLS.

The covariance matrix  $R$  turns out to be symmetric, because elements  $R_{ij}$  and  $R_{ji}$  have equivalent expressions:

$$R_{ij} = \frac{1}{k} \cdot X_i^T X_j = \frac{1}{k} \cdot X_j^T X_i = R_{ji}. \quad (\text{B.1})$$

Next, multiply the eigenvalue problem (3.5) from the left with the vector  $\theta_i^T$ :

$$\frac{1}{k} \cdot \theta_i^T X^T \cdot X \theta_i = \lambda_i \cdot \theta_i^T \theta_i. \quad (\text{B.2})$$

This expression consists essentially of two dot products that can be interpreted as squares of vector lengths. Because these quantities must be real and non-negative, and because  $k$  is positive integer, the eigenvalue  $\lambda_i$  is always *real* and *non-negative*.

Again, multiply the eigenvalue problem (3.5) from the left side, this time with another vector  $\theta_j^T$ :

$$\theta_j^T R \theta_i = \lambda_i \cdot \theta_j^T \theta_i. \quad (\text{B.3})$$

---

Because  $R$  is symmetric ( $R = R^T$ ) there must hold

$$\theta_j^T R = (R^T \theta_j)^T = (R \theta_j)^T = \lambda_j \theta_j^T, \quad (\text{B.4})$$

so that from (B.3) one has

$$\lambda_j \cdot \theta_j^T \theta_i = \lambda_i \cdot \theta_j^T \theta_i \quad (\text{B.5})$$

or

$$(\lambda_i - \lambda_j) \cdot \theta_j^T \theta_i = 0. \quad (\text{B.6})$$

For  $\lambda_i \neq \lambda_j$  this can only hold if  $\theta_j^T \theta_i = 0$ . Eventually this means that the eigenvectors of  $R$  are *orthogonal*.

Since MATLAB was used for the calculations and the implemented MATLAB functions automatically return normalized vectors, said eigenvectors are always even *orthonormal*.

# Appendix C

## Example

## Eigenvectors and Eigenvalues

In Section 6.3 a selection of 7 vectors was chosen. To give an example, these vectors and the corresponding eigenvalues are presented here.

The eigenvalues are

$$\begin{aligned} \left( \lambda_1 \quad \dots \quad \lambda_N \right) = \\ \left( 5.2617 \quad 2.6728 \quad 2.3393 \quad 1.7939 \quad 1.5778 \quad 1.4594 \quad 1.2524 \right), \end{aligned} \tag{C.1}$$

and the new base for the dataspace  $X$  is

$$\Theta_7 = \left( \theta_1 \quad | \quad \dots \quad | \quad \theta_7 \right) =$$

---


$$\begin{pmatrix}
 -0.2874 & 0.3244 & 0.1536 & 0.3197 & -0.1391 & 0.0419 & 0.0450 \\
 0.2355 & 0.3876 & -0.3505 & -0.0228 & 0.0072 & 0.0721 & 0.0422 \\
 -0.2556 & -0.0130 & -0.1840 & -0.3423 & -0.3305 & -0.0250 & 0.1052 \\
 -0.1168 & -0.2221 & -0.3506 & 0.3086 & 0.0961 & -0.1733 & 0.0213 \\
 -0.2914 & 0.0608 & -0.0957 & -0.1517 & 0.4397 & 0.1104 & -0.0898 \\
 0.0206 & 0.0806 & -0.0067 & -0.0386 & -0.1850 & -0.0031 & -0.8395 \\
 -0.2920 & 0.3239 & 0.1625 & 0.3066 & -0.1282 & 0.0497 & 0.0414 \\
 0.2377 & 0.3856 & -0.3493 & -0.0128 & 0.0097 & 0.0819 & 0.0414 \\
 -0.2992 & 0.0030 & -0.2008 & -0.3217 & -0.3428 & -0.0700 & 0.1263 \\
 -0.1350 & -0.1964 & -0.4063 & 0.3670 & 0.0422 & -0.2087 & 0.0406 \\
 -0.3139 & 0.1141 & -0.0772 & -0.1773 & 0.4454 & -0.0243 & -0.1050 \\
 0.0826 & 0.2140 & 0.1092 & -0.0983 & -0.0580 & -0.6586 & -0.2946 \\
 -0.2813 & 0.3465 & 0.1939 & 0.2737 & -0.1015 & 0.0600 & 0.0481 \\
 0.2497 & 0.3789 & -0.3371 & 0.0071 & 0.0383 & 0.0706 & 0.0389 \\
 -0.2744 & 0.0487 & -0.1746 & -0.2896 & -0.2985 & -0.0424 & 0.0726 \\
 -0.1560 & -0.1589 & -0.3227 & 0.3084 & -0.1056 & -0.1922 & -0.1585 \\
 -0.3026 & 0.1326 & -0.0575 & -0.1692 & 0.4107 & -0.0493 & -0.0860 \\
 0.1001 & 0.1581 & 0.1632 & -0.0663 & 0.1279 & -0.6391 & 0.3336
 \end{pmatrix} . \tag{C.2}$$

# Appendix D

## Calculated Voltages

The process of nickel plating is considered to be based on underlying electrochemical reactions. Hence, there are variations in potential and small flows of current in the bath. Since these variables are not measured very accurately, it might be possible to calculate them from available data.

According to [14] the Nernst equation gives the possibility to calculate the equilibrium potential from the bath concentrations and temperature:

$$u_1 = U_1 + \Delta U_1 + \kappa(\log c_5 - \log c_1 - 2\text{pH}) \quad (\text{D.1})$$

$$u_2 = U_2 + \Delta U_2 + 2\kappa(\log c_1 - 2\text{pH}) \quad (\text{D.2})$$

$$u_3 = U_3 - 2\kappa\text{pH} \quad (\text{D.3})$$

$$u_4 = U_4 + \Delta U_4 + \kappa \log c_4, \quad (\text{D.4})$$

---

with

$$\kappa^{-1} = 2 \frac{F}{RT} \log_{10} e$$

$$F = \text{Faraday's constant, } 96487 \text{ C/mol}$$

$$R = \text{universal gas constant, } 8.3145 \text{ J/(mol K)}$$

$$T = \text{Temperature [K]}$$

$$c_1 = \text{hypophosphite concentration [mol/dm}^3\text{]}$$

$$c_4 = \text{nickel concentration [mol/dm}^3\text{]}$$

$$c_5 = \text{orthophosphite concentration [mol/dm}^3\text{]}$$

$$\Delta U_1 = 10^{-6}(T - 25 \text{ }^\circ\text{C})360 \text{ } \mu\text{V}/^\circ\text{C}$$

$$\Delta U_2 = 10^{-6}(T - 25 \text{ }^\circ\text{C})300 \text{ } \mu\text{V}/^\circ\text{C}$$

$$\Delta U_4 = 10^{-6}(T - 25 \text{ }^\circ\text{C})60 \text{ } \mu\text{V}/^\circ\text{C}.$$

Since there is no online information about the hypophosphite concentration  $c_1$  from the measurements, one is only able to calculate  $u_3$  and  $u_4$ . For each time stamp the voltages were calculated and eventually added to the input data set, increasing its dimension to  $k \times 20$ . The obtained variables could add more information about the actual state of the bath on which the estimates for the plate characteristics will be based.

# References

- [1] N. Wiener. *Cybernetics, or Control and Communication in the Animal and the Machine*. MIT Press / The Massachusetts Institute of Technology, 1948.
- [2] A. Rosenblueth, N. Wiener, and J. Bigelow. Behavior, Purpose and Teleology. Technical report, Harvard Medical School and Massachusetts Institute of Technology, 1943.
- [3] H. Hyötyniemi. Neocybernetics in Biological Systems. Last update: February 24, 2006. See: <http://www.control.hut.fi/Research/cybernetics/>.
- [4] J. E. Jackson. *A User's Guide to Principal Components*. Wiley-Interscience, 2003.
- [5] S. T. Roweis. EM Algorithms for PCA and SPCA. Technical report, California Institute of Technology, Computation and Neural Systems, 1997.
- [6] H. Hyötyniemi. Multivariate Regression. Technical report, Helsinki University of Technology, Control Engineering Laboratory, 2001.
- [7] H. A. Simon. Sciences of the Artificial. Technical report, MIT Press / The Massachusetts Institute of Technology, 1996.
- [8] Wikipedia. [http://en.wikipedia.org/wiki/Minimum\\_total\\_potential\\_energy\\_principle](http://en.wikipedia.org/wiki/Minimum_total_potential_energy_principle).

- 
- [9] M. Sailer. Distributed Control of a Deformable System — Analysis of a Neocybernetik framework. Master's thesis, Helsinki University of Technology, Control Engineering Laboratory, 2006.
- [10] Wikipedia. [http://en.wikipedia.org/wiki/Le\\_Chatelier's\\_Principle](http://en.wikipedia.org/wiki/Le_Chatelier's_Principle).
- [11] Wikipedia. [http://en.wikipedia.org/wiki/The\\_Blind\\_Watchmaker](http://en.wikipedia.org/wiki/The_Blind_Watchmaker).
- [12] L. I. Smith. A Tutorial on Principle Component Analysis, 2002.
- [13] K. Mayberg, P. Vachenauer. *Höhere Mathematik I*. Springer-Verlag Berlin Heidelberg, 5th edition, 1999.
- [14] K. Kantola. Modelling of Electroless Nickel Plating for Control Purposes. Master's thesis, Helsinki University of Technology, Control Engineering Laboratory, 2004.
- [15] New Brunswick Plating, Inc. <http://www.nbplating.com/capabilities.htm>.
- [16] M. Schlesinger, M. Paunovic. *Modern Electroplating*. Wiley & Sons, 2000.
- [17] G. O. Mallory, J. B. Hajdu. *Electroless Plating: Fundamentals and Applications*. American Electroplaters and Surface Finishers Society, 1990.
- [18] J. Harju. Characterization of Electroless Nickel Plating Process. Master's thesis, Helsinki University of Technology, Control Engineering Laboratory, 2002.
- [19] R. Tenno, H. Koivo. State Estimation and Bath Control for the Electroless Nickel-Plating Process. Technical report, Helsinki University of Technology, Control Engineering Laboratory, 2003.
- [20] F. J. Nuzzi. Automatic Control of Electroless Nickel Plating. 1983.
- [21] I. N. Bronstein, K. A. Semendjajew, G. Musiol. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 5th edition, 2000.

- 
- [22] R. Tenno, K. Kantola, H. Koivo, A. Tenno. Model Identification for Electroless Nickel Plating Through Holes Board Process. Technical report, Helsinki University of Technology, Control Engineering Laboratory, 1996.