Data-Based Modeling of Electroless Nickel Plating

Hans-Christian Pfisterer*
*Control Engineering Laboratory
PL 5500
02015 TKK
Finland
h.pfisterer@gmx.de

Heikki Hyötyniemi†
†Control Engineering Laboratory
PL 5500
02015 TKK
Finland
heikki.hyotyniemi@tkk.fi

Abstract

This work connects the neocybernetic theory of elastic systems to modeling of a complex industrial process. It turns out that using simple mathematics combined with multilinear tools leads to linear models that can be used for monitoring and control. As an application example, electroless nickel plating is studied.

1 Introduction

What has artificial intelligence to do with modeling of chemical processes? It seems that such a very technical application domain would have little in common with the AI approaches. Still, in all complex systems the challenges are similar. One would need to understand *emergence* to escape from the individual low-level behaviors to the level of relevant functionalities.

In the work at hand the industrial process of *electroless nickel plating* is studied. This chemical system is very complex since many chemicals are involved and the process has a lot of variables being observed and controlled in order to reach good plating results. What is more, for commercial reasons not all compounds and reactions have been disclosed. It is very difficult also for a domain expert to understand the behavior of the system.

In this paper, a very simple, data-based model for control purposes of the nickel alloy thickness and its phosphorus content is presented. This model was achieved using basic ideas and mathematical tools based on the theory of *neocybernetics.*

2 Neocybernetics

Since models of industrial systems and observed natural systems get more and more complex and very complicated to understand the traditional way of mathematical modeling with dynamic nonlinear differential equations and constraints leads to an unmanageable mess of functions and equations. Furthermore, one has to be a domain expert to understand the interconnections between variables in order to write the mathematical equations for the description of the behaviors in the first place.

Neocybernetics, introduced by H. Hyötyniemi (2006), offers a new approach to get reasonable system models and helpful system understanding when looking at a system from a different angle. Since industrial systems are monitored extensively the data can be used to find connections among the variables and possibly associate some variables with others. The system structure is hidden in this data and it will emerge when manipulated in an appropriate way. Using the right mathematical tools and a holistic view of the system this modeling machinery is also available for non-experts of the particular domain. Here information about the behavior of the system is retrieved directly from the measurement data.

2.1 Key points

To see a system through neocybernetic eyeglasses some assumptions about the system have to be made. There are some basic principles (see H. Hyötyniemi (2006) and H.-C. Pfisterer (2006)):

- *Emergence:* Some unanticipated functionality can appear after the cumulation of some simple operations.
- *Dynamic Balance:* The emphasis is on systems in some kind of balance rather than on the process itself.
- *Linearity pursuit:* The system behavior is considered to be linear as long as nonlinearity is not absolutely necessary.

- *Environment:* Neoybernetic systems are assumed to be oriented towards their environment, they reflect and mirror their environment.

All these principles and assumptions are reasonable and in many natural and industrial systems they can be fulfilled. The linearity assumption needs a further explanation (see Section 5), especially when modeling chemical systems.

2.2 Elasticity

The system is assumed to be in thermodynamic balance. The changes in the environment are seen as disturbances causing tensions in the system, pushing the system away from the balance. According to the Le Chatelier principle (H. Hyötyniemi, 2006), the system yields, but after the pressure vanishes, the original balance is restored. In a sense, the neocybernetic ideas are a functionalization of the intuitions concerning complex *homeostatic* systems.

2.3 Degrees of freedom

As mentioned above the hidden structure of the system will emerge when the modeling concentrates on the remaining degrees of freedom in behaviors rather than on constraints and restrictions. For simple systems this is not reasonable — for very complex systems where many variables are strongly connected and many constraints have to be considered this approach helps to avoid an unmanageable mess of equations. When concentrating on the non-constrained degrees of freedom, the emphasis is on phenomena that are not explicitly seen in any formulas; one could speak of *emergent models.*

Now the system is kept in dynamic balance and one searches for the structure of covariation that is still left in the variable space. A model based on the degrees of freedom approach is as useful as a traditional one but has the advantage of being simple and understandable. With this "inverse thinking" not all constraints and chemicals and reactions need to be known, as long as it can be assumed that the internal interactions and feedbacks always manage to keep the system in balance. More information about degree of freedom modeling can be found in H. Hyötyniemi (2006) and H.-C. Pfisterer (2006).

3 Multivariate tools

Using the above mentioned key points and utilizing strong mathematical tools a practical modeling machinery can be set up. After applying the modeling tools a black box between known variables and unknown ones to be estimated can be filled with life. For these purposes the known data (column vectors of data samples) is collected in a matrix X and the unknown variables in a target space formed by the matrix Y. Here, it is assumed the k sample vectors of length n and m are stored as rows in the matrices X and Y, respectively.

Principal Component Analysis (PCA)

The information in terms of covariations in the data is captured in the (unscaled) correlation matrix

$$R = X^T X.$$

A lot of hidden information can be revealed by eigenvalues λ_i of this matrix, and by their corresponding eigenvectors θ_i as can be read in H. Hyötyniemi (2001). The eigenvectors are orthogonal and point in the direction of maximum variation in the original dataset. The amount of variety of each particular direction is given by the corresponding eigenvalue. Directions of most variety are assumed to carry most information about the system and it is reasonable to take these into account in the model; hence the model is built on the degrees of freedoms found in the correlation structure of the data.

Large eigenvalues stand for directions of essential information, small eigenvalues stand for directions which most probably contain redundant information or only measurement noise. When the data is projected onto the subspace Z of dimension $N \leq n$ by the mapping matrix $F1$, redundancy and measurement noise in the data can be reduced.

Multilinear Regression (MLR)

To find the connection to the target space Y a regression is applied. In order to achieve good results the base of the starting space should be orthogonal and no redundancy should be there. Using the dataspace Z these prerequisites are fulfilled and the MLR algorithm can find a mapping $F2$ from Z to Y. Hence a combination of both procedures, now called Principal Component Regression (PCR), can explain the behavior in Y by the information already given in X. This is illustrated in Figure 1.

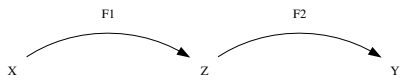

Figure 1: Data spaces and projections with PCR

Partial Least Squares (PLS)

This algorithm includes a view into the target space Y already while building the mapping $F1$. Not only the dataspace X is scanned for the essential information (collected in $Z1$) but also in Y the key information is extracted (to $Z2$) and only then an overall mapping is obtained which now bridges input and output. This steps and transformations are illustrated in Figure 2.

Figure 2: Data spaces and projections with PLS

After introducing the industrial process and some specific variables these tools will be applied and their results shown in Section 6.

4 Electroless nickel plating

The process of electroless nickel plating is a step of surface finishing. Printed wiring boards (PWBs) consist of a epoxy laminate base and a copper layer in which electric circuits are etched. Witout protection the copper would oxidize very fast, especially when coming in contact with humidity. In order to enhance the lifetime of the PWB and to improve its mechanical properties the PWB is coated with a gold layer of about $0.01\mu m$. But since the copper would diffuse in the gold and form again an oxidizeable compound the two layers are separated by a nickel layer of around $4\mu m$. Figure 3 shows a cut through a plated PWB and explains the layers in detail; since the gold layer is so thin it can not be seen here.

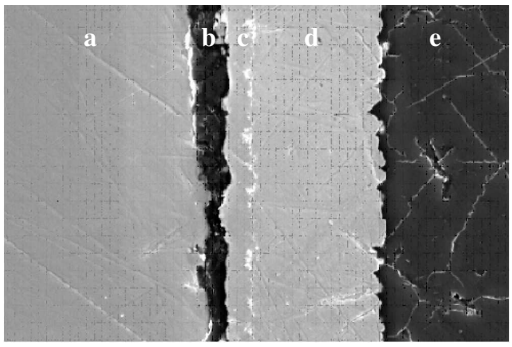

Figure 3: Cut through a plated substrate. Layers: (a) a metallic fastener and (b) a piece of conductive plastic to fix the sample, (c) Ni-P layer, (d) copper layer, (e) the base of PWB (epoxy laminate) (K. Kantola, 2004)

The use of the word "electroless" is misleading; it emphasizes the difference between the method where an external power supply is connected to the substrate which is to be plated and the approach of an active substrate that deposits nickel atoms from ions on its surface. Here nickel is provided in form of ions in an aqueous solution, beside the reducer hypophosphite that provides the necessary electrons for the transformation from nickel ions to real atoms. Due to the used reducer the alloy has a phosphorus content of up to 15 weight percent. The catalytically activated substrate is immersed into the bath and an alloy of nickel atoms and some phosphorus can be built, because nickel itself is also catalytically active itself. There are electric currents in the bath caused by the conversion of ions to atoms, so the process is not really electroless.

The bath consists of many more chemicals which help to keep the bath stable and in a desired status. The composition of these chemicals is either not exactly known or a well kept industry secret which makes the traditional modeling of the bath behavior very difficult or even impossible. Furthermore one has to be a chemical domain expert to understand the connection between the chemicals and their behavior especially when connected to the substrate and its expanding nickel alloy.

The characteristics of the bath are observed continuously and as many variables as possible are measured. Hence online measurements of the current bath nickel concentration, bath temperature and pH value are available and with good controllers kept along desired values to keep the bath constitution constant. The substrate and its characteristics are measured in a laboratory when the plating is finished. Summarizing the dataspaces X and Y (Section 3) basically contain the following information:

X: Nickel concentration, pH value, ammonia concentration, ammonia and nickel sulfate pumping, plating area and temperature.

Y: Alloy thickness, phosphorus content, hypophosphite concentration and orthophosphite concentration.

The original dataset X is further prepared and expanded as explained in the following section. The information about hypophosphite and orthophosphite is not really a plate characteristic and for that reason not further studied here.

5 Linearity

The system is in a state of thermodynamic balance. This state is kept constant by a very strong and accurate control mechanism provided by the bath surrounding. Balance also is one of the neocybernetic key points and so is very important to this approach. The thermodynamic equilibrium can be described by the constant

$$K = \frac{C_{\mathrm{B}_1}^{b_1} \dots C_{\mathrm{B}_\beta}^{b_\beta}}{C_{\mathrm{A}_1}^{a_1} \dots C_{\mathrm{A}_\alpha}^{a_\alpha}}.$$

This constant is highly nonlinear dependent on the concentrations of the different chemicals in the solution. But applying a logarithm on both sides and differentiating the expression leads to

$$\begin{aligned} 0 \quad = \quad & b_1 \frac{\Delta C_{\mathrm{B}_1}}{\bar{C}_{\mathrm{B}_1}} + \dots + b_\beta \frac{\Delta C_{\mathrm{B}_\beta}}{\bar{C}_{\mathrm{B}_\beta}} \\ & - a_1 \frac{\Delta C_{\mathrm{A}_1}}{\bar{C}_{\mathrm{A}_1}} + \dots - a_\alpha \frac{\Delta C_{\mathrm{A}_\alpha}}{\bar{C}_{\mathrm{A}_\alpha}}, \end{aligned}$$

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. For more information about this equations see H.-C. Pfisterer (2006) and H. Hyötyniemi (2006).

Even more the whole system can be represented in a linear way when using variables in a proper way. Nonlinearity can by avoided by appropriate preprocessing of the variables (by employing relative changes among the data). What is more, the dynamic nature can be put in the static form when the variables are selected in a clever way. The following points explain how to use information in order to stay in the linear domain even if the real system is nonlinear and can also be found in H. Hyötyniemi (2006):

- *Temperature:* According to the Arrhenius formula, reaction rates k are functions of the temperature, so that $k \propto \exp(c/T)$. When this is substituted, 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 $\mathrm{pH} = -\log c_{\mathrm{H}^+}$. Because this is a logarithm of a concentration variable, one can directly include the changes in the pH value among the variables as $\mathbf{z}_{\mathrm{pH}} = \Delta\mathrm{pH}$.
- *Voltage:* In electrochemical reactions one may characterize the "concentration of electrons". It turns out that according to the Butler-Volmer theory the amount of free electrons is exponentially proportional to the voltage. Hence, after taking logarithms, the "electron pressure" can be characterized by $\mathbf{z}_{\mathrm{e}^-} = \Delta U$.
- *Physical phenomena:* It is evident that phenomena that are originally linear like diffusion can directly be integrated in the model, assuming that appropriate variables (deviations from a nominal state) are included among the variables.

In practice, some reactions are not in balance; rather, there can be a constant flux of some chemical exiting from the reacting solution. The above balance expressions can be written also for such "dissipative" reactions, and linear models are again applicable around the equilibrium flux. This means that also the rates of change need to be included among the variables that characterize the dynamic state of the process.

Some of the variables to be included in the regression model are integrals over time — for example, in the coating process, the layer thickness is the integral of the nickel reduction rate. Because of the model linearity, such integrals can be transferred from the model output to the input — meaning that the layer thickness can be modeled as the integrals of the variables are included among the variable values themselves among the data.

Hence apart from the plain data also some features can be included in the dataset X. Since the measurement sample is only taken at the precise time when the plate is taken out of the bath it makes sense to include more information about the whole time the plate is immersed. A weighted integral over the plating time solves this problem, hence the dataset X is extended with integrals of the plain variables which doubles the dimension of X. The mentioned connection between layer thickness and nickel reduction rate is another reason for using integrals in the input structure. Furthermore an old set of data can be included by weighting the old data differently in comparison to the recent information. This smoothing of data again adds featured data to X.

A schematic view and a summary of all the mentioned variables in the input structure can be seen in Figure 4.

6 Modeling the plating process

When using the described setup the first step is PCR (Section 3). Figure 5 shows the eigenvalues λ_i of the correlation matrix of X in descending order. Since the dimension of the used dataspace is $n = 18$ there are as many eigenvalues, each one representing the

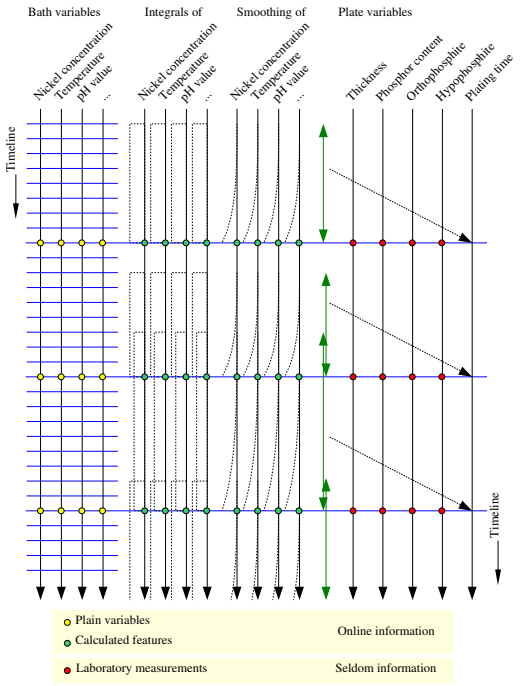

Figure 4: Schematic view of input data set along the timeline of the plating process

amount of information kept in the corresponding direction θ_i. It can be seen clearly that the last directions do not carry any information; that shows the redundancy in the data. Since the noise is assumed to be uncorrelated it is evenly distributed in all directions, hence some directions carry mostly noise and no real information. In the end it is reasonable to use only the first 7 directions for a new dataspace Z from which a final mapping to Y should be obtained. This first $N = 7$ vectors carry already 91% of the full information.

The alloy thickness and its phosphorus content can be accurately estimated with these 7 vectors. Using PLS as an data preparing algorithm the dataspace can even be reduced to a dimension of $N = 2$ and obtain an even better estimate. The results of the model gained by PLS analysis are presented here. For more results and discussions see H.-C. Pfisterer (2006).

Figure 6 shows the process data of the alloy thickness divided in three parts. The first one was used for model estimation along the above described way. The other two parts are validation sets. Blue information is the real data, measured in the laboratory and stored by a data acquisition system. Green data is a model estimation, obtained with a traditional complex electrochemical model by K. Kantola (2004) and used as a comparison to validate the linear model at hand. In

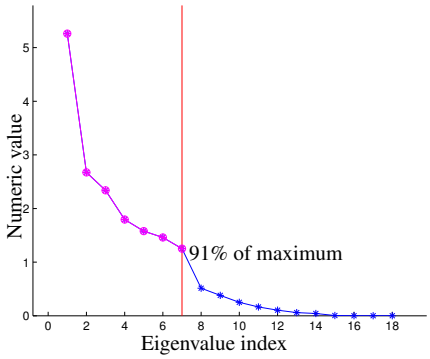

Figure 5: Latent variables θ_i (PCA) and the numerical values of the corresponding eigenvalues (equals importance of information among data)

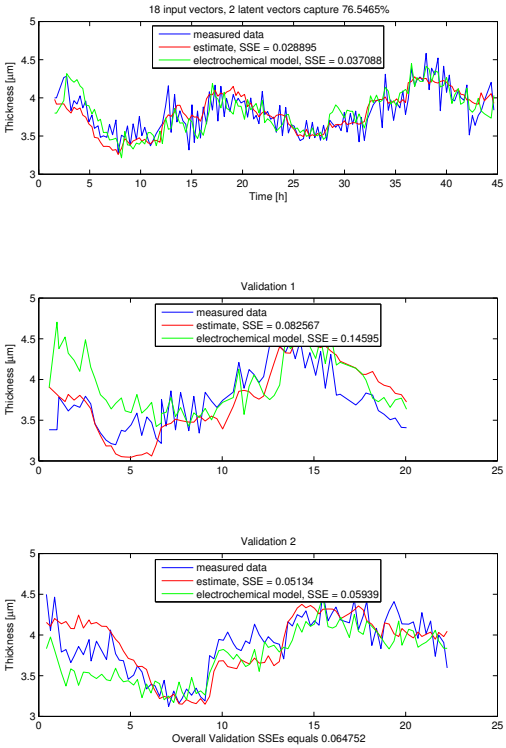

Figure 6: Alloy thickness: measured data and two compared estimates

red color the data estimation for the alloy thickness by the linear model can be seen and compared to real data and the other model estimation.

The same setup of data was used to estimate the phosphorus content of the nickel alloy (Figure 7). In red color again the result of the linear model at

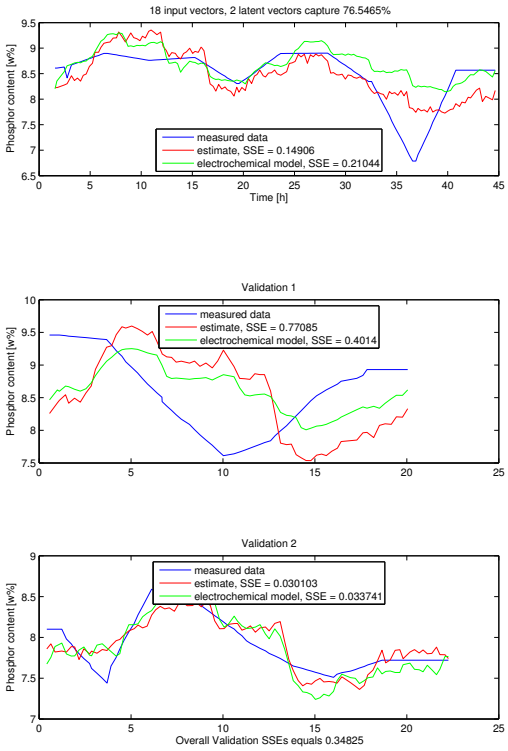

Figure 7: Alloy phosphorus content: measured data and two compared estimates

hand along with real data (blue) and results of another model (green).

The linear model provides a very accurate estimation of the alloy thickness and an accurate estimation of its phosphorus content. This is no surprise in the first parts of either dataset, since they were used to build the model and the model is tailored by the algorithm especially for them. But the correct estimates in the other parts confirm and validate the good performance of the linear model. The power and the correct combination of strong mathematical tools produced a very accurate result which can keep up and even beat complicated and inscrutable models. Since this model is a linear combination of measured data it is also possible to see the importance of each measured variable for the characteristic of each output variable.

It is interesting to see that if the estimate of the linear model is wrong also the estimate of the complex model is wrong. This is a hint for missing information that is not available for the modeling machinery in either case or some mistakes in the measurements, hence no fault of the linear model. Further it should be pointed out that there is an improvement when using PLS instead of PCR and also when including integrals in the input data vector. However, smoothed old data can not improve the results.

The overall result combined with the simplicity of its design leads a way to a good control machinery for nickel plating processes. The model can easily be used and adjusted by process workers and it can provide the necessary information for a controller in form of an observer for the plate characteristics. The information about these characteristics is already hidden in the anyway measured bath variables and it was just necessary to reveal it and make it emerge.

Acknowledgements

Many thanks to Prof. Dr.-Ing. Dr.h.c. Michael Zeitz from the Institute for System Dynamics (ISYS) of the University of Stuttgart who made the paper at hand possible and gave his support and advice.

References

H.-C. Pfisterer. Data-Based Modeling of Electroless Nickel Plating. Master's thesis, Helsinki University of Technology, Control Engineering Laboratory, 2006.

H. Hyötyniemi. Multivariate Regression. Technical report, Helsinki University of Technology, Control Engineering Laboratory, 2001.

H. Hyötyniemi. Neocybernetics in Biological Systems. Technical report, Helsinki University of Technology, Control Engineering Laboratory, 2006.

J. E. Jackson. *A User's Guide to Principal Components*. Wiley-Interscience, 2003.

J. Harju. Characterization of Electroless Nickel Plating Process. Master's thesis, Helsinki University of Technology, Control Engineering Laboratory, 2002.

K. Kantola. Modelling of Electroless Nickel Plating for Control Purposes. Master's thesis, Helsinki University of Technology, Control Engineering Laboratory, 2004.

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.