INTERACTIVE VISUALIZATION AS A TOOL FOR ANALYSING TIME-VARYING AND NON-LINEAR SYSTEMS

Jimmy Johansson * **David Lindgren** **
Matthew Cooper * **Lennart Ljung** **

* *NVIS – Norrköping Visualization and Interaction Studio, Linköping University, Sweden*
** *Division of Automatic Control, Linköping University, Sweden*

Abstract: This paper proposes advanced visualization and interaction techniques as a support for the analysis of system identification data. Non-linear or time-dependent dynamics often leave a significant residual with linear, time-invariant models. The structure of this residual is decisive for the subsequent modelling and by visually analysing the data the modeller may gain a deeper insight into its structure than can be gained using only standard correlation analysis.

Keywords: Interactive visualization, system identification, model validation.

1. INTRODUCTION

System identification is inherently an interactive art. Results from preliminary model building are studied by the user. Based on such studies, decisions about new model structures are taken. The studies are typically of a visual nature, often simple 2-dimensional line plots of correlation functions and residuals. Visualization techniques have gone through a significant development during the past decade. It is an interesting problem to study what such new techniques may offer in terms of improved interaction in system identification.

The purpose of the present contribution is to give some illustrations of how the system identification process can be extended to include sophisticated visualization and interaction techniques. The paper describes how model validation can be aided by the use of interactive 3-dimensional visualization which allows for simultaneous analysis of the dependencies between two variables and time. The study is a result of joint activities between a research group in system identification and one in visualization.

1.1 LTI Models with Non-LTI Systems

Fitting a linear time-invariant (LTI) model to data sampled from a system with nonlinear or time-varying dynamics usually results in a large-magnitude nonlinear or time-varying residual. Unless the modeller is aware from the beginning of the nonlinear or time-variant nature of the system dynamics, the cause of the large residual magnitude will, perhaps, not be evident. Correlation and auto-correlation estimates may reveal that something is wrong but give no clue to whether the bad fit is due to wrong model order, nonlinearities or time-variant dynamics. At this point, it is actually difficult to gain insight into the model discrepancies.

Volume graphics, however, offers a means to go beyond standard summary statistics measures for model validation. By assigning each data point

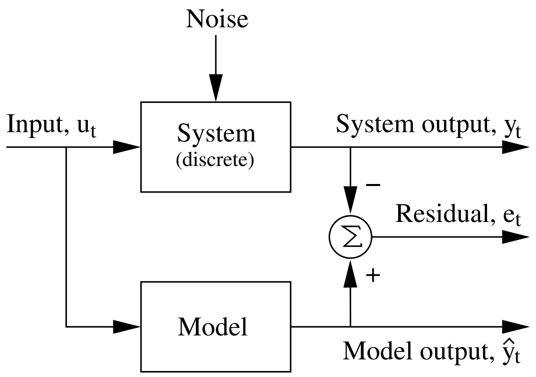


Fig. 1. Model validation. The output y_t of the system, here discrete, is compared to the output of the model, $\hat{y}_t$. The difference between the two makes the residual, e_t. If the residual is dependent on the noise only, not on the input u_t, then there is nothing more to model.

to a semi-transparent volume in a 3-dimensional spatial/temporal space, the modeller may explore and interpret large-scale process data sets (10^4-10^5 sample points).

The process of system identification (Ljung, 1999) can be broken down into selection, estimation and validation of the model. In the first selection step, the task is to select an appropriate model structure. The second step is to estimate model parameters given data sampled from a system and calculate the model residual – the part of the output that is left unmodelled. The last step is then to validate the model. If the residual is satisfactory, nothing else needs to be done. If not, another model structure needs to be tested and the process starts over again. A satisfactory residual is, for instance, a residual small in magnitude and independent of the model/system input.

Particularly in the model validation step, the modeller needs diagnostics tools to support the decision whether to accept the model or try some other structure. At this stage the study of the model residual and the dependency between the residual and the input is decisive.

1.2 Visualization and Interaction Techniques

Being able to visualize data with multiple visualization techniques in multiple views can be powerful when analysing complex data. Volume graphics (Drebin *et al.*, 1988) is one technique for displaying volumetric data as a 2-dimensional image. The volume data can, for example, be the result of sampling an object in 3 dimensions. One common application area of volume graphics is in the medical imaging domain where, for example, data can be constructed from the output of an X-ray Computed Tomography (CT) scanner.

Visualization alone, however, is usually not enough in order to fully understand the data. Of equal importance are also correct preprocessing of the data and the ability to interact and manipulate with the resulting image. For data sets containing long time-series, a focus + context visualization (Doleisch *et al.*, 2003) may be successfully used. Using one view to display the entire time-sequence (the context) and another view for displaying a detailed sub-region (the focus) prevents the user from getting lost in the data while analysing a specific region.

1.3 Related Work

In recent years, there have been great efforts made to visualize complex process data, for instance by means of scatter plot matrices (Wong and Bergeron, 1996) or by parallel coordinates (Johansson *et al.*, 2004). The applications have mainly been in decision support (Kulhavý, 2003). For the interested reader refer to (de Oliveira and Levkowitz, 2003) for an overview of concepts and taxonomies of visual data exploration.

2. MODEL VALIDATION

A linear discrete time-invariant state space model is used:

$$\begin{cases} x_{t+1} = \boldsymbol{A}x_t + \boldsymbol{B}u_t \\ \hat{y}_t = \boldsymbol{C}x_t. \end{cases} \quad (1)$$

Here, u_t is the input signal (common to the system and model), $\hat{y}_t$ the simulated model output and y_t the sampled output signal of the system to be identified. The model order, n, is the same as the dimensionality of the state space vector x_t. Given n (design parameter), a set $\{u_t, y_t\}_1^N$ with sampled data points and some assumption of the initial state x_0, techniques to estimate the model parameters $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are well-known (state-space subspace system identification), see (van Overschee and de Moor, 1996). Programs for this task are available commercially, for example, the System Identification Toolbox (Ljung, 2003) for use with MATLAB.

Once the model parameters are estimated and $\hat{y}$ simulated according to (1), the model residual is calculated as

$$e_t = y_t - \hat{y}_t, \quad t = 1, 2, \ldots, N. \quad (2)$$

If the model is a correct description of the sampled system and the input is white noise, the residual will depend only on some unknown noise process, not on the input. In fact, if the residual is independent of present and past inputs, nothing more is left to be modelled and the process is complete. Fig. 1. illustrates the signal paths.

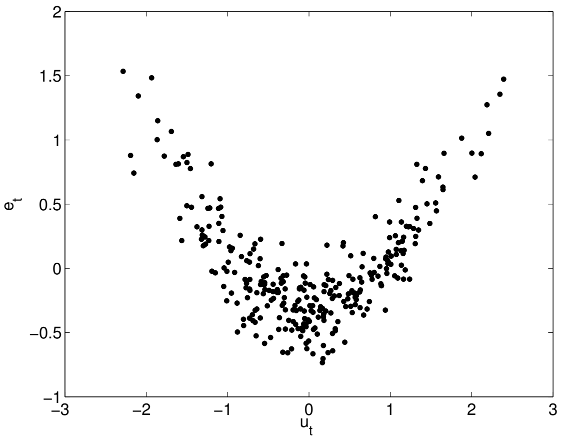


Fig. 2. A simple scatter plot showing samples from two dependent but uncorrelated random variables.

A standard measure of the dependence between two scalar, zero-mean samples is the sample cross covariance,

$$\widehat{R}_{eu}(\tau) = \frac{1}{N-\tau}\sum_{t=1}^{N-\tau} e_{t+\tau}u_t. \tag{3}$$

If some $\widehat{R}_{eu}(\tau)$ is significantly larger than 0, this indicates the model is not sufficient. Plotting $\{e_t\}$ against $\{u_{t+\tau}\}$ in different 2-dimensional scatter plots for $\tau = 0, 1, \ldots$ may, however, reveal dependencies not reflected by $\widehat{R}_{eu}$ at all. Fig. 2 illustrates this by plotting samples from two evidently dependent random variables that are totally uncorrelated, $\widehat{R}_{eu}(0) = 0$.

Non-linear dynamics are always difficult to observe and it is by no means trivial to find a minimum number of variables that reveal the non-linear nature of data. One approach is to search for structure manually by selecting different time-lags of the residual and input and try to visually interpret them. A more systematic approach is proposed in (Lindgren and Ljung, 2005), where linear combinations of the different time-lagged signals are sought by numerical programs. The linear combinations found may then be visualized.

To also visualize time-varying dependencies it will be necessary to use 3-dimensional graphics to display $e_{t+\tau}$, u_t, and t. This gives rise to a number of issues that will be addressed in the subsequent sections.

3. IMPLEMENTATION

The application developed exploits the system identification toolbox in MATLAB together with a powerful object-oriented visualization system, AVS/Express (AVS, 2004). This system is based on individual modules and an application is built

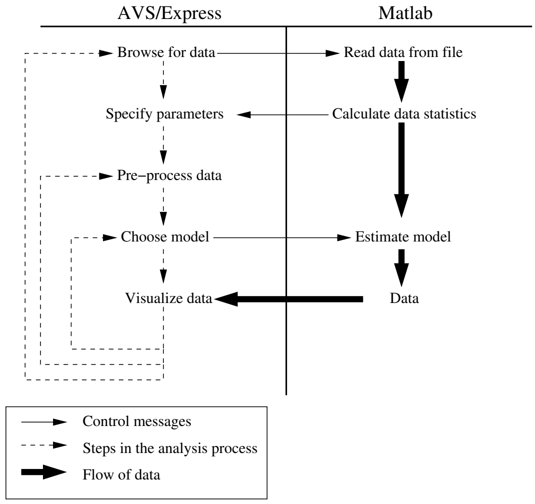


Fig. 3. The work flow between AVS/Express and MATLAB.

by combining these through a visual network editor. MATLAB and AVS/Express have been seamlessly integrated through a common user interface with which the user can interactively preprocess, analyse and visualize data. The application has also been extended from an ordinary desktop PC to a much larger semi-immersive stereoscopic display. This type of visualization creates the psychological illusion of immersion: of being inside the computer-generated environment, rather than viewing it from the outside through a screen.

3.1 Application Overview

A simplified work flow of the application is shown in Fig. 3. The work flow begins with loading data from a file into MATLABs workspace. The user can choose to visualize the original data set but can also perform a number of preprocessing operations on the data. The next step in the analysis process is to choose an appropriate model structure and to perform the estimation. Through the user interface, the user can look at a preview of the estimation result in a 2-dimensional scatter plot. This process is highly interactive so the user can, at any time, go back and change the parameters of the model to further improve the result of the visualization. When the user has identified an area of interest it can be visualized using a volume graphics technique, see Fig. 6 and Fig. 7. To further investigate the data, a cross-section for a particular time instant can be displayed.

To manage the communication between MATLAB and AVS/Express, a communication module has been developed. This module handles all control messages and the entire flow of data between the two systems. Since system identification data may be very large, the data is (as far as possible) handled locally by MATLAB. It is only when the user

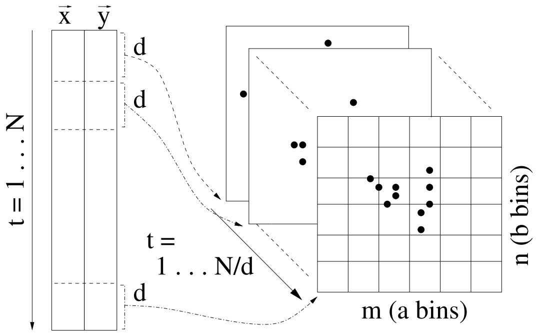


Fig. 4. A subset of the 2-dimensional data is scaled and aggregated. This is done for every subset d, which gives a volume of size $n \times m \times \frac{t}{d}$.

explicitly requests data that it is transferred. Also, it is only the selected subset that is sent, never the entire data set. This makes the processing of very large data sets much easier and the visualization process highly interactive.

3.2 Interactive Volume Graphics

In the analysis process it is possible to perform a volume graphics visualization of a selection of variables. This 3-dimensional visualization technique effectively reveals any time-variations and is an important part of the analysis process.

To make the volume graphics visualization interactive, the volume data needs to be recalculated every time the user makes a change in the parameter settings. How quickly this can be done primarily depends on the volume size but also on the type of platform used. The calculation, described below, and rendering of the visualization of a $128 \times 128 \times 256$ volume on a desktop PC can be done with less than a second of delay. The size of the volume can dynamically be changed to speed up the calculation on a slower PC.

The volume data is build from a number of planes, where each plane contains a subset with d of the N data points, see Figure 4. The different planes thus correspond to different time epochs of the data set. Each plane is, in turn, divided into an a-by-b grid of bins into which the data points are aggregated. The aggregation simply counts the data points that fall into the respective region defined by the grid. This procedure is done for every subset of d data points, which produces a volume of size $a \times b \times \frac{N}{d}$. To make the data points fall within the predefined grid they are first scaled and translated so that they all have values between 0 and 1. This procedure is done for every subset, which produces a volume of size $m \times n \times \frac{t}{d}$.

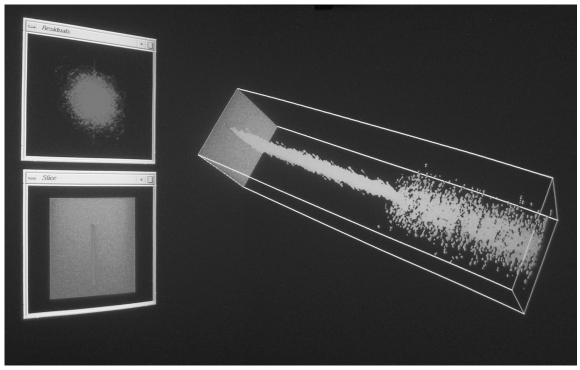


Fig. 5. The application running on a semi-immersive display.

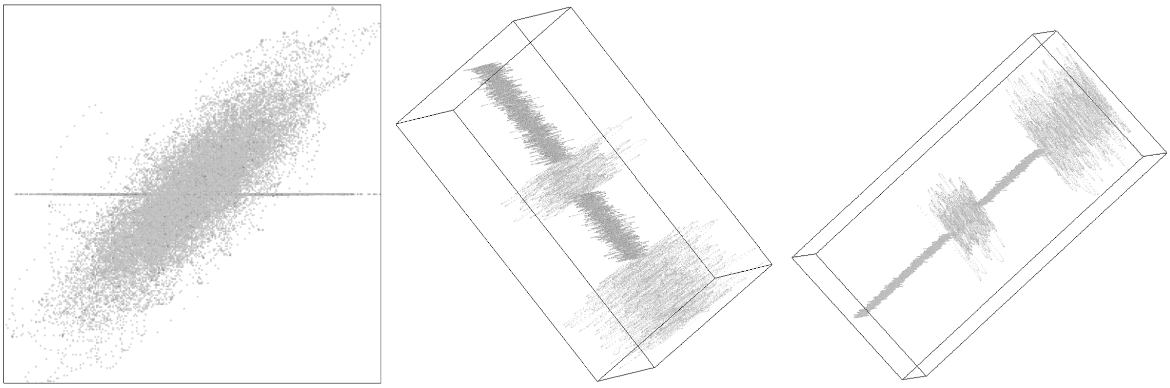
4. RESULTS

Volume graphics provides the modeller with useful insight into the structure of data sampled from a time-varying system. The residual of an LTI state space model has been combined with the model input and time to construct a spatial/temporal 3-dimensional volume. Integrating this 3-dimensional volume visualization with several 2-dimensional visualizations creates an interactive platform for exploratory analysis of such data. The understanding and experience gained from interacting with data through this platform has proven to be a valuable tool in the system identification process.

The results of extending the visualization from a desktop PC to a much larger, semi-immersive, stereoscopic display have so far been promising. An example of this is shown in Fig. 5. The data comes from a simulated linear feedback system with variable feedback gain. The 2-dimensional scatter plot in the upper left view shows the relationship between u_t and e_t. This gives a quick overview of the data and it can be seen that there seems to be two or more separate regions but there is no information about where in time these regions are. Visualizing the same variables using a volume graphics visualization (right view) gives the necessary temporal information. To further investigate the relationship between these variables, the cross-section from the clip plane is displayed in the lower left view. The plane can also be animated through the entire time sequence.

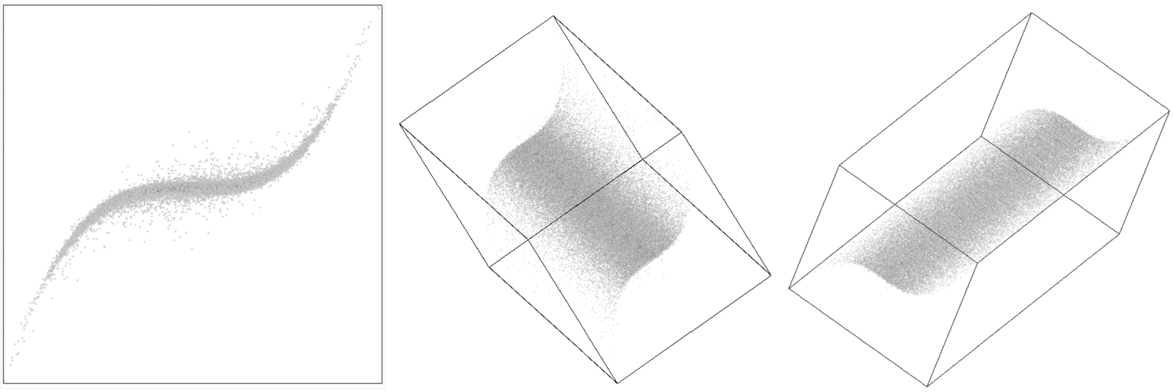
Fig. 6 shows the use of interactive volume graphics where time-varying dynamics is detected. This is the same data as in Fig. 5 but with slightly different settings. Here it is clearly seen that the system dynamics is divided into four separate temporal regions. The axes in Fig. 6 are time, t, model input, u_t, and the residual, e_t.

In Fig. 7 another important discrepancy of a model is detected, namely non-linear dynamics. The (real-life) silver data is sampled from an electronic circuit. This data set has been studied



(a) (b) (c)

Fig. 6. Volume graphics visualization of time-varying dynamics. A linear time-varying data set seen from three different angles (a,b,c). The three axes are time, t, input, u_t, and residual, e_t.



(a) (b) (c)

Fig. 7. Volume graphics visualization of non-linear dynamics. Three different angles (a,b,c) of the silver data. The axes represent time, t, output, y_t, and a particular linear combination of time lagged versions of the input and output.

before, see (NOLCOS, 2004) and in theory it obeys the non-linear differential equation

$$m\frac{d^2y(t)}{dt} + d\frac{dy(t)}{dt} + ay(t) + by(t)^3 = u(t). \quad (4)$$

The sampling interval is $0.0016384s$. The axes in Fig. 7 represent time, t, output, y_t, and a particular linear combination of time lagged versions of the input and output. This linear combination is found by an optimisation program. The criteria for the optimisation is how well a third-order polynomial fits the data for different linear combinations, see (Lindgren and Ljung, 2005).

For the modeller it is is necessary to get insight of both time-variations and non-linearities in order to proceed to develop more adequate models.

Colour versions of the figures presented in this section can be found at (Colour, 2005).

5. CONCLUSIONS

Advanced visualization techniques as a support for system identification have a great potential and the understanding and experience gained from interacting with data through our application has proven to be a valuable tool. Future plans and ideas include completion of the platform with additional model structures: for instance non-linear Wiener-Hammerstein models and adaptive time-variant models.

Further study on non-linear dynamics and how volume graphics and projection techniques can facilitate the pursuit of adequate model structures are also interesting.

Finally, merging the interactive volume graphics with projection based methods known from the area of exploratory data analysis may provide new insights into the process of model validation.

6. ACKNOWLEDGEMENTS

This work is part of the VISIMOD project (VISIMOD, 2005), supported by the Swedish Foundation for Strategic Research.

A special thanks to Prof. Anders Ynnerman at NVIS, Linköping University for ideas during the initial phase of the project.

REFERENCES

AVS (2004). Advanced visual systems. http://www.avs.com/software/soft_t/avs-xps.html.

Colour (2005). Colour plates. http://nvis.itn.liu.se/publications/IFAC2005/.

de Oliveira, Maria Cristina Ferreira and Haim Levkowitz (2003). From visual data exploration to visual data mining, a survey. *IEEE Transactions on Visualization and Computer Graphics* **9**(3), 378–394.

Doleisch, Helmut, Martin Gasser and Helwig Hauser (2003). Interactive feature specification for focus+context visualization of complex simulation data. In: *Proceedings of the symposium on Data visualisation 2003*. Eurographics Association. pp. 239–248.

Drebin, Robert A., Loren Carpenter and Pat Hanrahan (1988). Volume rendering. In: *Proceedings of the 15th annual conference on Computer graphics and interactive techniques*. ACM Press. pp. 65–74.

Johansson, Jimmy, Robert Treloar and Mikael Jern (2004). Integration of unsupervised clustering, interaction and parallel coordinates for the exploration of large multivariate data. In: *Proceedings of the 8:th IEEE International Conference on Information Visualisation*. pp. 52–57.

Kulhavý, Rudolf (2003). A developer's perspective of a decision support system. *IEEE Control System Magazine* pp. 40–49.

Lindgren, David and Lennart Ljung (2005). Nonlinear dynamics identified by multi-index models. In: *Proceedings of the 16th IFAC World Congress*. Prague, Czech Republic.

Ljung, Lennart (1999). *System Identification: Theory for the User*. 2:e edition ed.. Prentice-Hall. Upper Saddle River, N.J. 07458, USA.

Ljung, Lennart (2003). *System Identification Toolbox User's Guide*. version 6 ed.. The MathWorks, Inc.. Natick, MA. USA.

NOLCOS (2004). Special session on identification of nonlinear systems: The silver box study. In: *Proceedings of NOLCOS 2004 – the IFAC Symposium on Nonlinear Control System*.

van Overschee, Peter and Bart de Moor (1996). *Subspace identification for linear systems : theory, implementation, applications*. Boston: Kluwer.

VISIMOD (2005). Visimod homepage. http://www.itn.liu.se/ jimjo/IFAC/.

Wong, Pak Chung and R. Daniel Bergeron (1996). Multiresolution multidimensional wavelet brushing. In: *Proceedings of the 7th conference on Visualization '96*. IEEE Computer Society Press. pp. 141–ff.