A new look at the dynamic covariance structure of various approaches for batch process modelling

J. Camacho, J. Picó
Department of Systems Engineering and Control.

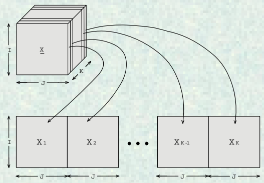
A. Ferrer
Department of Applied Statistics, Operations Research and Quality.

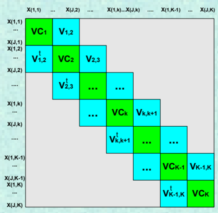
Technical University of Valencia.
Cno. de Vera s/n, 46022, Valencia, Spain

The data set collected from a batch process is three-way: a set of variables are measured at different sampling times during the processing of a batch, and this is repeated for a number of batches. Therefore, the data matrix has to be conveniently rearranged in a number of two-way matrices to apply bilinear models as PCA and PLS. This can be done in several ways. The aim of this poster is to analyze the effect of using one way or another from the study of the covariance matrix. This analysis helps us to determine how dynamics are built in the models, which parameters in the models are related with a single sampling time and which are averages of several/all sampling times, which information –if any- is discarded after unfolding or dividing in several sub-models, and in which cases this is convenient.

ARRANGEMENT IN TWO-WAY | COVARIANCE MATRIX | FEATURES

BATCH WISE



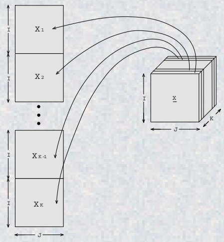


a) Both the instantaneous relationships of the variables (the variances and instantaneous cross-covariances in matrices VC_k) and the dynamic relationships (the auto-covariances and lagged cross-covariances in matrices $V_{k-n,k}$) are modelled.

b) Changing dynamics are modelled by placing sub-matrices VC_k and $V_{k-n,k}$ for two different values of k in a different part of the covariance matrix.

c) The number of parameters in the covariance matrix, each of them fitted from I samples, is: $\frac{(JK+1)JK}{2}$

VARIABLE WISE



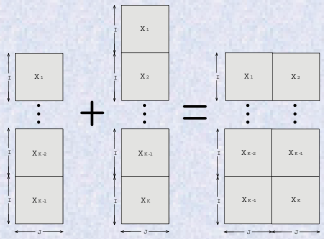
$$1/k \cdot (VC_1 + \dots + VC_k + \dots + VC_K)$$

a) Only instantaneous relationships (matrices VC_k) of the variables are modelled.

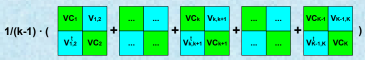
b) The covariance matrix results from the average of the covariance matrices of the single sampling times. Therefore, the former captures averaged information of the process.

c) The number of parameters in the covariance matrix, each of them fitted from IK samples, is: $\frac{(J+1)J}{2}$

BATCH DYNAMIC



Batch dynamic model with 1 lagged measurement vector



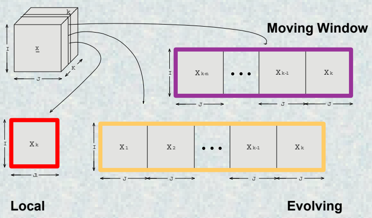
a) Instantaneous relationships (matrices VC_k) and dynamic relationships up to a certain number of lags are modelled.

b) The more lagged measurement vectors (LMVs) included, the more similar to batch-wise the unfolding is and the more capacity to model changing dynamics.

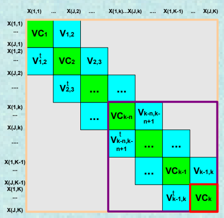
c) The less LMVs included, the more similar to variable-wise the unfolding is and the more the number of sampling times averaged.

d) The number of parameters,each of them fitted from I(K-LMVs) samples, is: $\frac{(J(1+LMVs)+1)J(1+LMVs)}{2}$

K MODELS



Covariance matrices of the sub-models of sampling time k



a) Since there is one model for every sampling time, changing –dynamic or instantaneous- relationships are modelled.

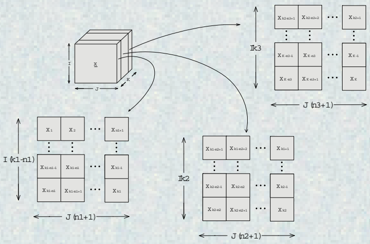
b) All the dynamics are included in the Evolving models. No dynamics are included in Local models. Moving Window models are an intermediate case.

c) The number of parameters, each of them fitted from I samples, is:

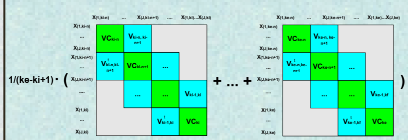
$$\frac{1}{2}\sum_{k=1}^{K}(p_k+1)p_k$$

where p_k ranges from J (Local) to Jk (Evolving).

MULTI PHASE



Covariance matrix of a single phase from sampling time ki to ke with n LMVs



a) Changing dynamics are modelled using several sub-models (phases).

b) Dynamics are included adding LMVs.

c) This is a general parametrization, where any of the preceding modelling approaches are included.

d) The number of parameters ranges from those of a Variable-wise model to those of an Evolving model, each of them fitted from IK to I samples.

Take the example of data with -almost- no dynamic information but where the relationships among the variables change throughout the batch processing. The Batch-wise unfolding may be used to model these data, but then most of the parameters of the covariance matrix (those inside matrices of the form $V_{k-n,k}$) will be noisy. Variable-wise cannot be used, since the relationships are changing. The Local models is the appropriate structure –i.e. arrangement in two-way.

Now assume the instantaneous relationships are constant throughout the batch duration, again with no dynamic information. The variable-wise is by far the most parsimonious –and so preferable- solution.

Finally, take the case when the initial part of the process is of utmost importance in its evolution. In such a case, the initial measurements have to be taken into account during the whole batch and thus, the Batch-wise unfolding may be adequate.

CONCLUSION: The most appropriate structure depends on the current process data. Therefore, this structure should be flexible enough to adapt to the features of the process. The Multi-phase modelling approach provides such flexibility.