Partial least squares PLS1 vs. PLS2 – optimal input/output modeling in a compound industrial drying oven

Maths Halstensen[1]   Ulrich Hundhausen[2]   Kim H. Esbensen[3]

[1]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway, Norway, maths.halstensen@usn.no
[2]Norwegian Institute of Wood Technology (NTI), Oslo, Norway, uhun@treteknisk.no
[3]KHE Consulting, Copenhagen, Denmark, khe.consult@gmail.com

Abstract

A feasibility study was carried out to assess the possibility of developing prediction models for monitoring drying conditions of wood coatings in one of Europe's largest and most modern coating plants for exterior cladding. These models were based on data from real-time Process Analytical Technology (PAT) sensors, measuring airflow and air direction, temperature and relative humidity). The study revealed that the information from the PAT sensors gave sufficient input to accurately model the complex drying conditions and their interrelations. Modelling was carried out using both Principal Component Analysis (PCA) and PLS-regression in both its PLS1 and PLS2 manifestations. In addition, the diagnostic prediction performance RMSEP between PLS1 and PLS2 models were not significantly different. This is advantageous for an industrial implementation concerning re-calibration operations: PLS1 requires 40 separate calibrations whereas PLS2 requires only one, because PLS1-R is a regression of a singular output variable (y-variable) and PLS2-R of several simultaneous, correlated output variables. While a single calibration based on PLS2 will take approximately one hour, the PLS1 approach will take more than a week.

Keywords: process monitoring; exterior wood coatings; wood cladding; oven drying; multivariate modeling; PLS1-R, PLS2-R

1 Introduction

Wooden cladding is the most common façade material for residential buildings in Norway, most often painted with multi-coat opaque high-build (>60-100 µm dry film thickness) to very high-build systems (>100 µm dry film thickness). Traditionally cladding has been painted by hand after installation, but the last years have seen a strong trend towards industrial coating application in Scandinavia (Hundhausen et al., 2016). Industrial application has the advantage of offering customers cladding that does not need immediate maintenance after mounting; moreover, industrial coating takes place under controlled conditions, which is of special importance as water-based coating systems today dominate the marked driven by environmental requirements for reduced amounts of volatile organic compounds (Elliott and Glass, 2000). While in the furniture and joinery industry UV-curable coatings are widely used, wood products for exterior applications are finished as physically dried coatings. Thus air velocity, relative humidity (RH) and ambient air temperature are the three decisive factors in convection drying (Nienhuis, 2014) causing water to flow from the lower to the upper part of coating films (Trent, 1992; Beetsma, 1996; Vanderhoff, 1966). Deficient drying can cause *blocking*, i.e. coated parts stick together after stacking which results in considerable transport damages, or can significantly affect coating adhesion, which promotes coating defects like paint *flaking*. In addition, the dry-to-recoat stage in industrial drying is critical, as it is one of the major factors controlling the overall batch production rate. Dry-to-recoat is the time interval needed before a second coat can be applied without risks for further defects, such as blistering or loss of adhesion (Koleske, 2012). In short, controlling drying is of overall importance regarding process capacity and yield.

The Norwegian cladding producer Gausdal Bruvoll SA has recently established one of the largest and most modern coating plants for industrial cladding in Europe (Hundhausen et al., 2018a). Controlled fast drying is crucial but technically challenging as different "orders" (customer specified batches of boards) often with different coatings pass through the approximately 70 m long multi-level dryer with a feed speed of 18 m/min (Hundhausen et al., 2016, 2018b). Reliable on-line drying monitoring is highly desirable to increase production yields by avoiding *coatings defects*.

A feasibility study has been carried out to assess the possibility of developing prediction models for monitoring drying conditions of exterior wood coatings in 8 out of 9 possible trajectories in Gausdal Bruvoll's dryer. Can drying conditions air flow, transport direction, air temperature and relative humidity recorded through a set of real-time Process Analytical Technology (PAT) (Bakeev, 2010) sensors be acquired accurately enough for process operators to run the drying facility reliably with a view of process optimality across many types of paints and coating combinations?

The first research objective in the present study is to assess the real-time on-line information from a set of

strategically placed PAT sensors as input variables for modeling.

The second research objective is strictly data analytical. Partial Least Squares 1 (PLS1) regression modelling is compared with its PLS2 alternative (Marten and Næs, 1989; Esbensen and Swarbrick, 2018). PLS1-R is regression of a singular output variable (y-variable); PLS2-R is regression of several simultaneous, correlated output variables (y-variables). The rationale for evaluating both modelling strategies depends on the inherent complex industrial oven conditions: If PLS2 prediction results are comparable to the standard PLS1 results; this will save a significant amount of the work for model updating and/or maintenance during uptime industrial operations. PLS2 requires only one calibration operation while PLS1 in this case would require 40 separate calibrations. If the single calibration based on PLS2 takes one hour, the PLS1 approach will take more than a week, so PLS2 is a much more efficient approach and a significant time saver in the industrial context, which needs to have a strong focus on minimal downtime.

2 Materials & methods

2.1 The hot air dryer

The hot-air-dryer has a length of 73 m and a width of 10 m. It is horizontally divided into 3 drying and 1 cooling zone, and is vertically stacked with three floors (Figure 1 and 2). While the zones are separated by hanging flaps, the floors are not separated, making the zones the operative unit in the drying process. It is up to the operators to make effective use of the vertical floors; there are significant advantages in planning the compound drying of several batches of boards (termed 'orders') intelligently. This requires a reliable PAT-based 'prediction of drying-end-results' facility, for which optimal monitoring of the compound drying will be critical.

The physical drying of the waterborne paint coating is based on the parameters air T, air speed, RH and drying time. The air is blown downwards onto the coating through tubes (Fig. 2).
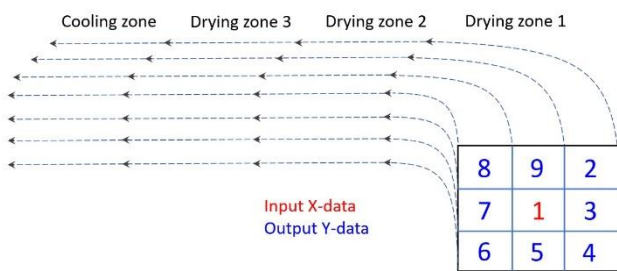

Figure 1. Simplified illustration of the nine drying trajectories for wood cladding. In the example indicated with red color, data from trajectory 1 was used as the input X data in calibration of the prediction model. The drying

conditions in all eight other trajectories were predicted based on the two alternative regression models PLS1 and PLS2.

As indicated with red color in Figure 1 the data from trajectory 1 was used as the input X data in calibration of the regression models. The drying condition described by the 40 output variables covering all the other eight trajectories were predicted based on the two alternative regression-modelling strategies PLS1 and PLS2, and compared see section 4.2.
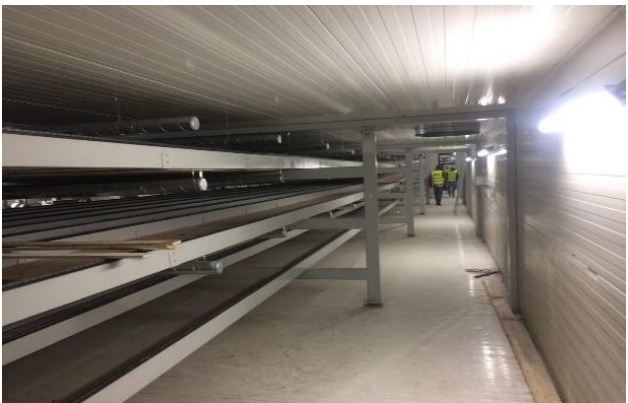

Figure 2. The dryer in the coating plant, showing the upper, middle and lower floors in one drying zone. Note the air ducts close to the ceiling, which supply the drying air to all levels.

Partial Least Squares 1, PLS1 (single output) and Partial Least Squares 2, PLS2 (multiple y-outputs) were used to calibrate regression models for the drying conditions.

2.2 Sensors on the move

An innovative element in acquiring simultaneous data with which to perform a first model of the complex drying oven conditions was to send a recording '*sensor box*' through the entire drying cycle. The box includes a temperature and relative humidity (RH) sensor (S-THB-M002 Sensor, Onset Computer Corporation, Bourne, MA, USA) and three hot wire anemometers (T-DCI-F350-W5B3, Onset Computer Corporation) measuring the air flow horizontally in transport direction, horizontally perpendicular to the transport direction and vertically to the transport direction. All sensors have a maximum update rate of 400 ms. The measurements were recorded with a sampling rate of 1 Hz using a multi-channel data logger (Hobo H22, Onset Computer Corporation).
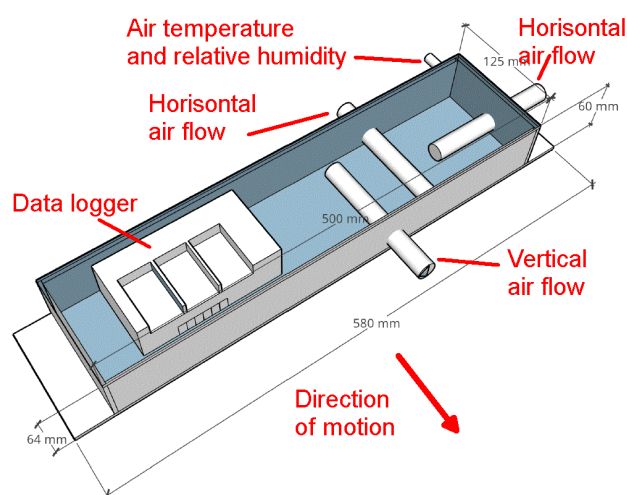
Figure 3. The moving "sensor box" equipped with *temperature/RH sensors and three hot-wire anemometers, measuring* along – and across the contemporary box velocity vector plus vertical air velocity.

2.3 Multivariate analysis and regression modelling

PLS1 regression is well described in the chemometric literature (Marten and Næs, 1989; Esbensen and Swarbrick, 2018) and numerous other background references herein. The PLS1 model is designed to maximize the effective correlations between the X-space (X-variables: PAT sensor recordings) and a singular y-variable. Here one also finds a description of the superior PLS modelling vs. prediction features in relationship to PCR (Principal Component Regression) and MLR (Multiple Linear Regression). The PLS1 approach has been the overwhelmingly most popular approach within process chemometrics because of the optimal modelling ability when dealing with one y-variable only.

Originally, in the early history of chemometrics, it was considered an advantage to be able to employ a PLS-regression modelling facility also in the case of a full multivariate Y-space (a multivariate set of y-variables), in cases where such a setup was required. The PLS2 modeling/prediction/validation situation is quite similar to PLS1, only the [X-space]-[Y-space] correlation modelling makes use of PLS components in both variable spaces simultaneously. In this fashion, the individual y_i variables are modelled based on the same foundational X-Y model. This only succeeds in the case in which all influential y-variables are strongly correlated – otherwise it is the overwhelming experience that modelling the individual y_i variables results in optimal prediction behavior – which is the reason for the dominant position of PLS1.

There only exist a few successful PLS2 application cases in the literature, e.g. Esbensen et al. (2001) who interconnected an acoustic chemometric approach using 'clamp-on' acoustic sensors (X) with laser velocimetry data in a radial pipeline profile. Their final feasibility demonstration successfully employed PLS2 inter-calibration of the acoustic monitoring data (X) with the laser velocimetric reference profiles (Y) using 39 y-variables.

A higher-level treatment of PLS2 method, including two practical application examples, can be found in (Stoccero et al., 2019).

With PLS1, the Y-data consist of a singular vector holding the 'reference values' used in calibration of the model. In PLS2, there are multiple y-output variables, so Y is a matrix in this case. Thus with PLS two alternative models are available, which is different from e.g. Linear regression and principal component regression (PCR) where only X is modelled; here no equivalent to PLS exists. The model for X and Y can be seen in equation (1) and (2) below.
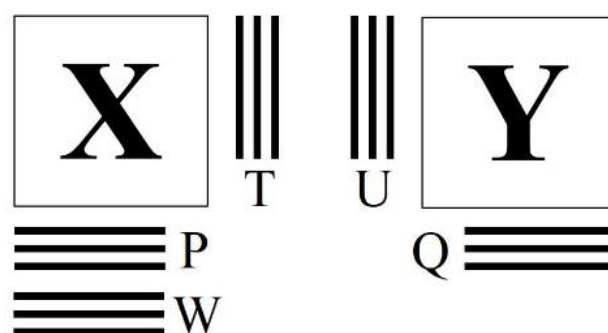

Figure 4. PLS data matrices (X and Y) and the resulting scores (T and U), loadings (P and Q) and Loading weights (W).

$$X = T \cdot P^T + E \qquad (1)$$
$$Y = U \cdot Q^T + F \qquad (2)$$

In the PLS approach the covariance between t and u are maximized which connects the two models.

<u>PLS1 algorithm</u>
Let $X_f = X$ and $y_f = y$ For f = 1, 2,…, A perform steps 2 to 6

1. $w_f = X^T y_f / \lVert X^T y_f \rVert$ (normalized to length 1)
2. $t_f = X \ w_f$
3. $q_f = t^T y_f (t_f^T t_f)^{-1}$
4. $p_f = X_f^T t_f (t_f^T t_f)^{-1}$
5. $b = u^T t (t^T t)^{-1}$
6. $X_{f+1} = X_f - t_f p_f^T$ and $y_{f+1} = y_f - b_f t_f^T$
7. $f=f+1$

Repeat 1 through 7 until f = A (the optimal number of components found by *validation*)

The compact regression coefficients for the Y variable can then be found like this:

$$\hat{b} = W(W^T X^T X W)^{-1} W^T X^T y \qquad (3)$$

Predictions based on new X-data can then be calculated:

$$\hat{y} = X\hat{b} + b_0 \qquad (4)$$

The PLS2 algorithm is similar but has a few more steps:

PLS2 algorithm

Let $X_f = X$ and $Y_f = Y$ For f = 1, 2,…, A perform steps 2 to 6

1. As the initial u_f vector, use any column in Y

2. $w_f = X^T u_f / |X^T u_f|$ (normalized to length 1)

3. $t_f = X w_f$

4. $q_f = Y^T t_f / |Y^T t_f|$ (normalized to length 1)

5. $u_f = Y q_f$

6. $p_f = X^T t (t^T t)^{-1}$

7. $b = u^T t (t^T t)^{-1}$

8. $X_{f+1} = X_f - t_f p_f^T$ and $Y_{f+1} = Y_f - b t_f q_f^T$

9. $f = f+1$

Repeat 1 through 9 until f=A (the optimal number of components found by validation)

The compact regression coefficients for the Y variables can again be found like this:

$$\hat{B} = W(W^T X^T X W)^{-1} W^T X^T Y \qquad (5)$$

Predictions based on new X-data can then be calculated:

$$\hat{Y} = X\hat{B} + B_0 \qquad (6)$$

In evaluating the regression model, the root mean squared error of prediction RMSEP, offset, slope and correlation coefficient are commonly used as model quality indices. Besides these, visual evaluation of the relevant score plots, loading weights plots, explained variance plots also provide useful information for calibrating and development of the prediction model (Marten and Næs, 1989; Esbensen & Swarbrick, 2018).

The root mean squared error of prediction is calculated as:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{i,\text{predicted}} - y_{i,\text{reference}})^2}{n}} \qquad (7)$$

3   Experimental

3.1 Air dryer experiments

The sensor box was mounted on a cladding board of 5518 mm length in three positions (left, middle, right) to record the climate across the width of the dryer. The position on both sides was 168 mm from the board end (Figure 5), and the box center was positioned 2759 mm from both ends. This was done for each of the three floors, giving nine runs (Figure 1 and Figure 2). After each run, the data were downloaded and deleted from the logger. The run-through time in each of the three drying zones was also recorded to assign the measurements to the zones. No other boards were in the dryer during the experiment.
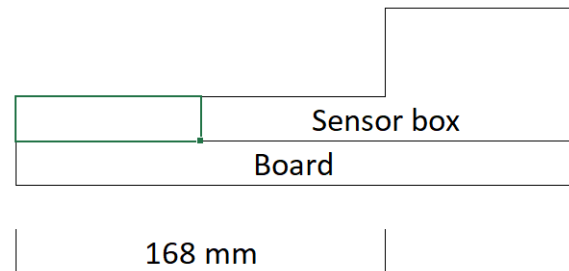

Figure 5. Positiong of the board and the box on the right side towards the gangway, trajectory 2 in Figure 2.

4   Results &discussion

4.1 PCA

The sensor recordings are represented as variables in the PCA X matrix, with the successive recording time points serve as objects. To get an overview of the full data set and how samples and variables relate to each other; PCA was used as the initial analysis method. All variables from all nine trajectories were included in this analysis.

The PCA score plot in figure 6 shows a marked depiction of the score value development with the three different drying zones (blue, red and green) clearly
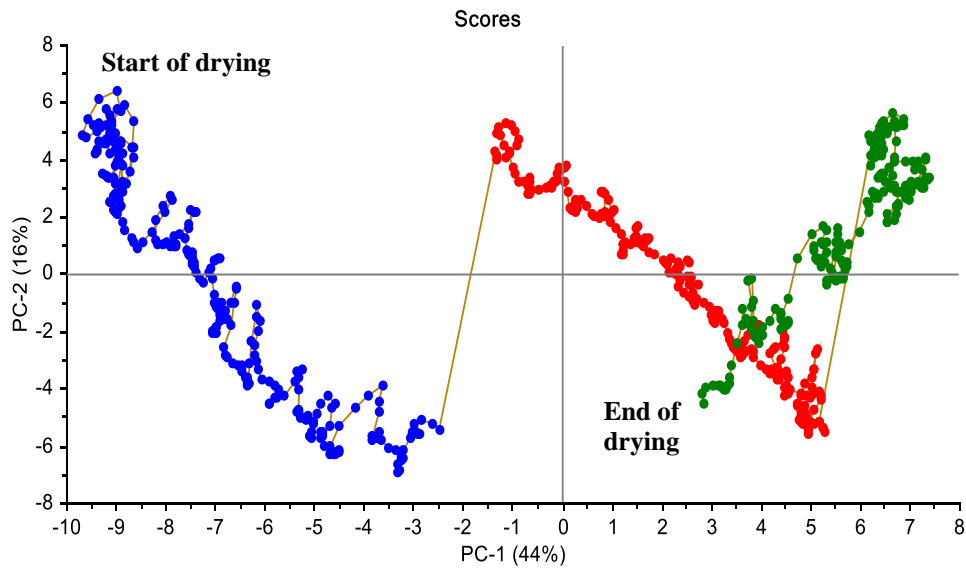
Figure 7. PCA – score plot of data from all nine drying trajectories as revealed by a full run-though in the drying oven of the "sensor box". Blue=drying zone 1, Red=drying zone 2 and Green=drying zone 3.
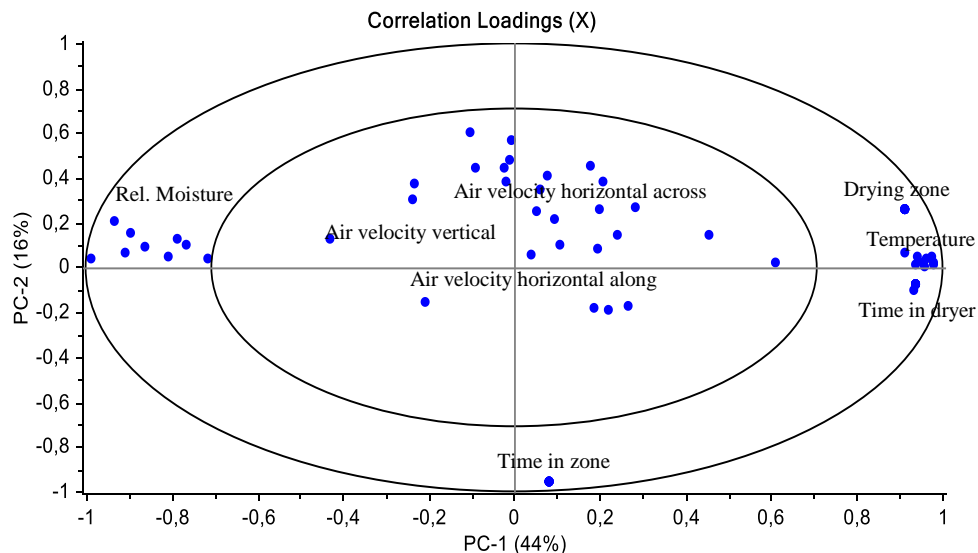

Figure 6. Correlation loadings corresponding to the score-plot in figure 6 (above) showing how the 8 variables from 9 drying trajectories relate to each other along PC1 and 2.

separated. No outliers can be seen in the data so all the samples and variables will be included in the next data modelling steps.

In the corresponding PCA loading plot, Fig. 7, all variables related to air velocity (vertical, horizontal along and horizontal across) are located close to the origin meaning their variance and covariance are less influential than the other variables occupying positions near the perimeter of this plot. These latter are the variables most influential on the model (variables that display a strong systematic inter-correlation), while the variables close to the origin contain a significant amount of irregular structure (noise).

4.2 PLS1 and PLS2 modelling results

Each variable in trajectory 2-9 was selected as an independent y-variable and was modelled by PLS1, which resulted in 40 separate PLS1 models. PLS2 was used to calibrate all the variables at once, which resulted in one PLS2 model also with 40 outputs.

Selection of training vs. validation data sets: The training data set was every second sample and the validation set was the other half of the data from the sensor box. Each PLS1 and PLS2 model was validated by this validation strategy (Esbensen and Swarbrick 2018). Results from the PLS1 and PLS2 models are reported and compared in Table 1 below.

Figure 8a shows the score-plot of the first two principal components (t1 vs t2). As can be seen it is possible to completely separate the samples from the three drying zones. No outliers were observed in the score plots t1-t2 nor in the score plots for the higher components.

The loading weights plot shows in figure 8b shows that all variables contribute to the model for the first three components. Although the variable "airflow horizontal across" shows relatively low loading values the prediction performance of the model does not improve if this variable is deleted.

Overfitting of regression models resulting from using too many components in the model is always a risk (Esbensen and Swarbrick 2018). As shown in the residual validation variance plot in figure 8c it is not immediately obvious how many components is optimal since the curve is decreasing for all the eight possible components. Since the model is validated based on independent data (test set validation), and there are 40 output variables to "satisfy" at once, we decided to use three components in the final prediction model, without loss of generality.
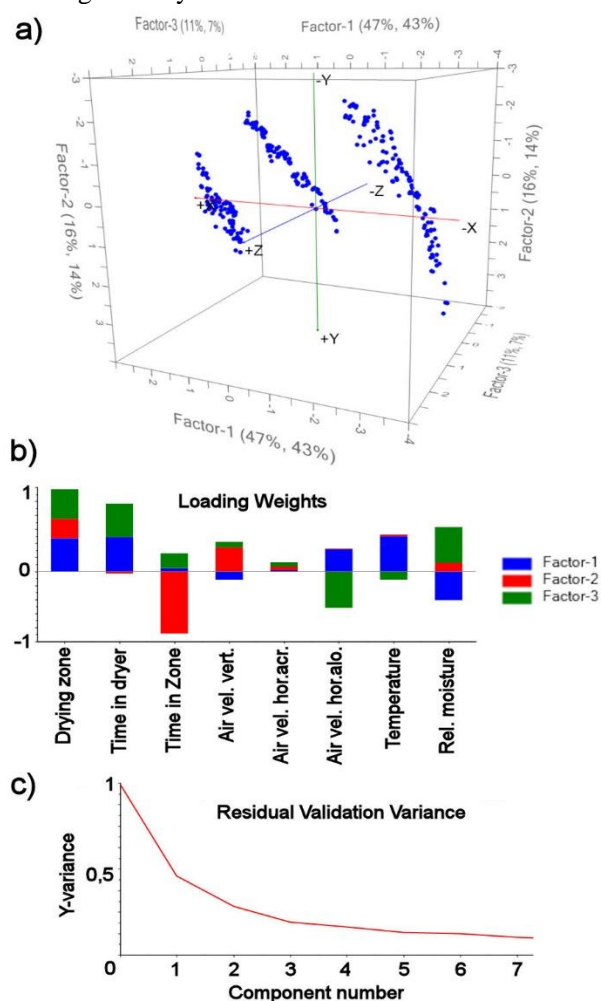

Figure 8. PLS2 model, a) Score plot t1-t2-t3, b) Loading weights for component 1, 2 and 3, c) Residual validation variance.

Figure 9 shows the predicted and reference data for a randomly selected trajectory, number 4. Figure 9a-c shows that the predicted air velocities can outline the main drying trends in the data well, but there is also a significant amount of irregular scatter along these trends, which we believe is due to the less influential air velocity data from the interior, very complex between-track and between-floors crosswinds set up by the ducted drying air supplied from the air ducts along the ceiling of the oven only.

We find these first foray predictions quite acceptable and maybe even better that the reference measurements recorded with the sensor box which contains a significant amount of random noise in themselves.

The plots of the predicted temperature and relative moisture content in figure 9d-e shows that the predictions follow the reference data very closely, which is promising for the eventual decision of implementing this modelling strategy in the industrial production.
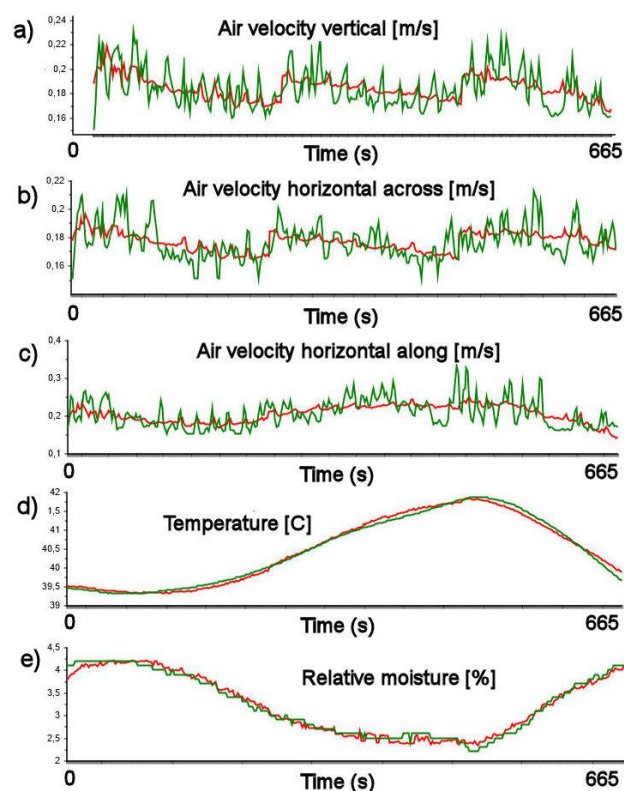

Figure 9. Predicted drying condition in trajectory 4 a) air velocity vertical b) air velocity horizontal across c) air velocity horizontal along d) Temperature e) Relative moisture content

A comparison of PLS 1 and PLS2 prediction results is shown in table 1. In all trajectories, the RMSEP values are similar for the two alternative modelling methods.

Table 1. PLS1 and PLS2 prediction comparison

Comparison of PLS1 and PLS2 model predictions			
Trajectory	Name of predicted variable	PLS1 RMSEP	PLS2 RMSEP
2	Air velocity vertical [mm/s]	10,34	10,38
2	Air velocity horizontal across [mm/s]	9,16	9,17
2	Air velocity horizontal along [mm/s]	34,60	34,63
2	Temperature [C]	0,04	0,04
2	Relative moisture [%]	0,15	0,15
3	Air velocity vertical [mm/s]	9,94	9,95
3	Air velocity horizontal across [mm/s]	7,93	7,96
3	Air velocity horizontal along [mm/s]	21,23	21,27
3	Temperature [C]	0,04	0,04
3	Relative moisture [%]	0,096	0,097
4	Air velocity vertical [mm/s]	12,28	12,33
4	Air velocity horizontal across [mm/s]	10,15	10,16
4	Air velocity horizontal along [mm/s]	28,59	28,62
4	Temperature [C]	0,085	0,085
4	Relative moisture [%]	0,09	0,09
5	Air velocity vertical [mm/s]	12,73	12,99
5	Air velocity horizontal across [mm/s]	8,95	9,06
5	Air velocity horizontal along [mm/s]	24,13	24,23
5	Temperature [C]	0,03	0,03
5	Relative moisture [%]	0,08	0,08
6	Air velocity vertical [mm/s]	13,04	13,19
6	Air velocity horizontal across [mm/s]	10,36	10,56
6	Air velocity horizontal along [mm/s]	17,98	18,15
6	Temperature [C]	0,06	0,06
6	Relative moisture [%]	0,09	0,09
7	Air velocity vertical [mm/s]	11,18	11,21
7	Air velocity horizontal across [mm/s]	10,73	10,91
7	Air velocity horizontal along [mm/s]	22,98	23,05
7	Temperature [C]	0,08	0,08
7	Relative moisture [%]	0,09	0,09
8	Air velocity vertical [mm/s]	11,24	11,56
8	Air velocity horizontal across [mm/s]	16,85	17,18
8	Air velocity horizontal along [mm/s]	17,12	17,33
8	Temperature [C]	0,06	0,06
8	Relative moisture [%]	0,20	0,20
9	Air velocity vertical [mm/s]	10,12	10,17
9	Air velocity horizontal across [mm/s]	7,14	7,62
9	Air velocity horizontal along [mm/s]	30,29	30,27
9	Temperature [C]	0,06	0,06
9	Relative moisture [%]	0,34	0,34

These results show it is possible to predict drying conditions in the dryer at GB with a satisfactory first foray quality. There were no significant differences for the diagnostic RMSEP values between PLS2 and PLS1 models. This indicates the possibility of establishing an effective maintenance schedule for such *in-production* models.

5 Conclusion

A feasibility study was carried out to assess the possibility of developing prediction models for monitoring drying conditions of wood coatings in one of Europe's largest and most modern dryers for exterior wood cladding. It was fully possible to perform detailed multivariate data models of the complex drying conditions and their interrelations using both Principal Component Analysis (PCA) and PLS-regression in both its PLS1 and PLS2 manifestations.

Since there were no significant differences for the diagnostic prediction performance RMSEP quality index between PLS2 and PLS1 models, PLS2 can be used to model the dryer at Gausdal Bruvoll, which indicates a positive prospect for very efficient data model updating and maintenance routine for future *in-production* PLS2 models. However, there remains a significant amount of industrial calibration work before a final evaluation can be performed.

References

K. A. Bakeev, *Process Analytical Technology, 2nd edition*. *Wiley*, Chichester, United Kingdom, 2010, ISBN: 978-0-470-72207-7, 2010. doi:10.1002/9780470689592

J. Beetsma. Alkyd Paints: From the ease of organic solvents to the difficulties of water. Proceedings of *XXIIth FATIPEC Conference*, Ease of Organic Solvents to the Difficulties of Water, 2: 157-164, 1995.

P. T. Elliot and J.E. Glass, *Applied Polymer Science: 21st Century, Elsevier Science*, 563-588, 2000. doi:10.1016/B978-0-08-043417-9.X5000-4

K. H. Esbensen, M. Waskaas, I. H. Matveyev, K. E. Wolden, J. G. Lode, T. T. Lied, and M. Halstensen. Feasibility of EMF-induced pipewall friction reduction by PLS2 inter-calibration of acoustic chemometrics and reference laser velocimetry. *J. Chemometrics*, 15: 241-263, 2001. doi: 10.1002/cem.684

K. H. Esbensen and B. Swarbrick, *Multivariate Data Analysis – An introduction to Multivariate Analysis, Process Analytical Technology and Quality by design*, 6th Ed., *CAMO Publishing*, 2018. ISBN 078-82-691104-0-1

U. Hundhausen, D. Kraniotis, S. Charisi, and M-L. Sortland. Quality control in coating lines for exterior cladding. From proceedings of *PRA's 10th International Woodcoatings Congress*, Amsterdam, The Netherlands, 2016. ISBN: 9783866302815

U. Hundhausen and M. Slabohm. In-line quality control of the film thickness in industrial coating processes for exterior cladding, Proceedings from the *International Research Group on Wood Preservation, Johannesburg, South Africa*, 18-40, 2018.

U. Hundhausen, M. Slabohm, and P. Meinlschmidt. Industrial coating of wood cladding: Inline control of board temperature, film thickness, and microfoam. Proceedings from *PRA's 11th International Woodcoatings Congress*, Amsterdam, The Netherlands, 23-24 October 2018. ISBN: 9783748600039

J. V. Koleske (Editor). *Paint and coating testing manual:15th Edition of the Gardner-Sward Handbook*. West Conshohocken, PA, USA, *ASTM International*, 2012. ISBN: 978-0-8031-7017-9. doi: 10.1520/MNL17-2ND-EB

H. Martens and T. Næs. *Multivariate Calibration*, *Wiley*, Chichester, UK, 1989 reprint 1994. ISBN-10: 0471930474, ISBN-13: 978-0471930471

J. G. Nienhuis, Review on drying and curing techniques of coatings. In proceedings from Cost E18 Final congress, Paris, 2004, 1-7, www.virtual.vtt.fi/virtual/proj6/coste18/nienhuisdryingpaper.pdf.

M. Stocchero, E. Locci, E. d'Aloja, M. Nioi, E. Baraldi, and G. Giordano. PLS2 in Metabolomics. Metabolites, 9(3):51, 2019. doi:10.3390/metabo9030051

A. Van Trent. *Turbidity Study of the Process of Film Formation of Polymer Particles in Drying Thin Film of Acrylic Latices.* PhD Thesis, University of Delft, The Netherlands, 1992.

J. W. Vanderhoff, E. B. Bradford, and W. K. Carrington. The transport of water through latex films. *Journal of Polymer Science*, *Polymer Symposia,* 41: 155-174, 1973. doi: 10.1002/polc.5070410116