7th International Building Physics Conference

IBPC2018

Proceedings

SYRACUSE, NY, USA

September 23 - 26, 2018

Healthy, Intelligent and Resilient
Buildings and Urban Environments

ibpc2018.org | #ibpc2018

Rational selection of experimental readouts for hygric material characterisation

Evy Vereecken[1,*], Hans Janssen[1] and Staf Roels[1]

[1]KU Leuven, Department of Civil Engineering, Building Physics Section, Belgium

[*]*Corresponding email: Evy.Vereecken@kuleuven.be*

ABSTRACT
To reduce the experimental effort and time required for the hygric characterisation of building materials, dynamic measurements in combination with parameter estimation are promising. However, noisy input data as well as limited and noisy output data complicate a reliable estimation of the material properties. A well-considered design of the dynamic experiments is advisable to reduce parameter uncertainty and to resolve non-identifiability issues.
Based on a virtual hygric test case, this paper shows the profile likelihood analysis to be an interesting starting point for the optimal design of hygric experiments. Trajectories of internal observables, such as relative humidity, plotted for the parameter sets along the profile likelihood curve are used to map highly uncertain ranges in the prediction space. This way, a first indication of information-rich (but originally non-measured) readouts that can reduce the parameter uncertainty is achieved. A *profile likelihood sensitivity (PLS) index* is used to quantify the individual uncertainty impact of the model parameters on a dynamic model prediction. Additionally, the *PLS entropy* quantifies the parameters' individual contribution to the *overall* prediction uncertainty. Both quantitative measures are applied to select the optimal readouts for reducing parameter uncertainty. The major advantage of the methodology is that it is a sample-based approach, and hence no gradient or curvature information of the objective function is required. This renders the methodology preferable over classical approaches based on Fisher information especially when dealing with non-linear problems.

KEYWORDS
Hygric properties, dynamic experiments, optimal experimental design, profile likelihood analysis, PLS index and entropy

INTRODUCTION
The reliable evaluation of moisture transfer in building materials is of great importance for the assessment of hygric damage patterns, building energy consumption, etc. In this respect, an accurate description of the moisture storage and transport properties of building materials is essential. Traditionally, these hygric properties are mainly determined based on steady-state measurements, which are tedious and time-consuming though. To reduce the experimental efforts required for such hygric material characterisations, dynamic measurements in combination with inverse parameter estimation show a lot of promise (Vereecken et al., 2018). However, noisy input data together with a limited and noisy output can complicate a reliable estimation of the properties. A well-considered design of the dynamic experiment helps reducing parameter uncertainty and could even resolve non-identifiability issues (i.e. parameters that cannot be uniquely derived from the available experimental output). Optimal Experimental Design (OED) methods can be applied to define such a purposeful information-rich design. Often, these OED methods aim at maximizing an optimality criterion that makes use of the Fisher information matrix (a.o. Dantas et al., 2002; Balsa-Canto et al., 2007), which is a matrix that quantifies the amount of information about an unknown model parameter

comprised in an observable. Unfortunately, the calculation of the Fisher information matrix requires information on the curvature or gradient of the log-likelihood objective function, which can be hard or even impossible to obtain for non-linear problems.
To resolve these issues, this paper explores, for a virtual hygric test case, the applicability of a sample-based OED approach evolved in systems biology research (Raue et al., 2009; Flassig et al., 2015). The method uses the information obtained in a profile likelihood analysis to map highly uncertain ranges in the prediction space. In this way, information-rich (but originally perhaps non-measured) readouts that reduce the parameter uncertainty can be selected. Quantitative measures for the uncertainty impact of the model parameters and their individual contribution to the overall prediction uncertainty are applied to make a rational selection between different potential readouts.

METHODS

The next paragraphs describe the virtual benchmark case, the parameter estimation method, the profile likelihood analysis and how to use this information to make a rational selection of suitable additional readouts in further experimental planning.

Initial benchmark description

In the benchmark case, the vapour resistance factor and the sorption isotherm of a calcium silicate sample are to be determined based on a dynamic experiment. The material's actual hygric properties (= target values) are shown in Figure 1a-b. A cylindrical sample with a diameter of 8 cm, a height of 4 cm and at an initial relative humidity of 12% is exposed to a dynamic step function in relative humidity (see Figure 1c). The sample's side and bottom are covered with a vapour-tight foil, which yields one-dimensional moisture transport. The mass transfer coefficient at the top of the sample is $3 \cdot 10^{-8}$ s/m + $N(\mu,\sigma^2)$ with $\mu = 0$ and $\sigma = 2 \cdot 10^{-9}$. The moisture transfer in the sample is simulated by use of a control volume model in order to create the virtual experimental output. Initially, the sample's mass change measured with a 1-hour time interval is assumed to be the experimental output. This output is subjected to white measurements noise $\varepsilon \sim N(\mu,\sigma^2)$ with $\mu = 0$ and $\sigma = 0.011$.

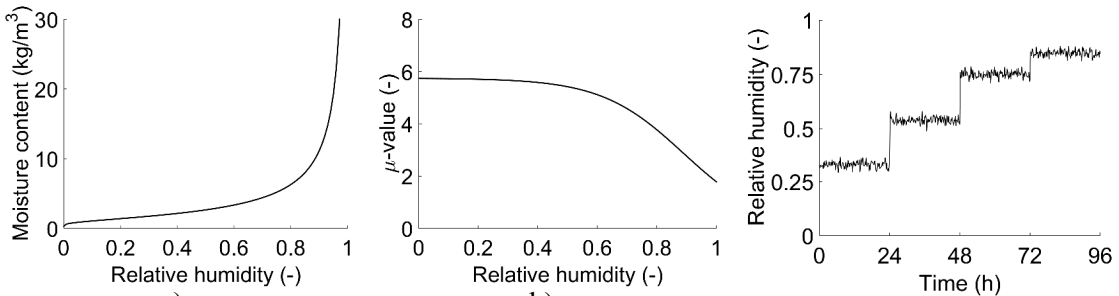


Figure 1. a) Actual sorption isotherm and b) actual vapour diffusion resistance of the calcium silicate sample; c) relative humidity in the environment above the top of the sample.

Parameter estimation

The parameters searched for are given in Table 1, together with the target values (= actual values in the benchmark case) and the lower and upper values of the search space. To estimate the unknown parameters, Bayesian inference is applied. Thereto, the Differential Evolution Adaptive Metropolis algorithm developed by Vrugt (2016) is used. In the estimation process, the moisture transfer in the calcium silicate sample is simulated with the same control volume model as used to simulate the benchmark experiment. A maximum log-likelihood (LL) is pursued, with the log-likelihood expressed as:

$$LL = \frac{-n_m}{2}\log(2\pi) - \sum_{t=1}^{n_m}\log\left(\hat{\sigma}_t\right) - \frac{1}{2}\sum_{t=1}^{n_m}\left(\frac{\tilde{y}_t - y_t}{\hat{\sigma}_t}\right)^2 \quad (1)$$

where n_m the number of measurement points for the sample's mass increase, $\hat{\sigma}_t$ the standard deviation of the observations' measurement noise and $\tilde{y}_t$ and y_t the mass increase in the virtual experimental data set and simulated in the estimation process, respectively. For the priors in the Bayesian inference, a uniform distribution is assumed.

Table 1. Unknown parameters together with the target values and the upper and lower boundaries in the search space. The parameter w_{sat} in the sorption balance is assumed to be known (811.14 kg/m^3). Additionally, w(0.54) is assumed to be measured beforehand (= 2.9 kg/m^3) and is used to infer the parameter k in the sorption balance.

	Vapour resistance $\mu(RH) = 1/(a+b\cdot\exp(c\cdot RH))$			Sorption isotherm $w(RH) = w_{sat}\cdot[1+(k\cdot\ln(RH))^{1/(1-m)}]^{-m}$	Noise balance	Mass transfer coefficient
Parameters	a	b	c	m	σ_M	ß
Target value	0.174	$2.806\cdot10^{-4}$	7.229	0.4321	0.011	$3\cdot10^{-8}$
Lower boundary	0.05	10^{-10}	4	0.1	10^{-5}	10^{-9}
Upper boundary	0.8	10^{-1}	30	0.8	1	10^{-6}

Profile likelihood analysis

To assess the identifiability and the confidence of the estimated model parameters, the profile likelihood analysis developed by Raue et al. (2009) is applied. In this approach, the profile likelihood PL curve for a model parameters θ_i is calculated by:

$$PL(\theta_i) = \max_{\theta_{j\neq i}} L(\theta_j) \quad (2)$$

with L the likelihood and θ_j the other parameters searched for in the estimation process. Hence, the PL curve for a parameter θ_i can be constructed by a re-optimisation of the parameters $\theta_{j\neq i}$ for fixed values of θ_i within an interval in the search space. In the current study, the knowledge on the likelihood of the candidate solutions analysed in the Bayesian inference procedure is used to draw the PL curves. A frequentist optimisation method as applied in (Vereecken et al., 2018) is furthermore used to create additional points to check and to complete the PL curves. The profile likelihood-based confidence region (CR) is defined by:

$$\left\{\theta_i \mid -2\log\left(PL(\theta_i)\right) \leq -2\log\left(\hat{L}\right) + \chi^2_\alpha\right\} \quad (3)$$

with χ^2_α the α^{th} quantile of the χ^2_{df}-distribution, with $df = 1$ degree of freedom and $\hat{L}$ the maximum likelihood estimate. The borders of this confidence region define the confidence interval (CI). A finite CI indicates a practically identifiable parameter, which means that - based on the available data - for θ_i a limited range of values with a likelihood not significantly different from $\hat{L}$ can be determined. A (semi-)infinite CI with a variable likelihood level is associated with practical non-identifiability, whereas an infinite CI with a constant likelihood

level below the desired threshold indicates that the parameter is both practically and structurally non-identifiable and, thus, that a unique parameterisation is theoretically not possible for the model.

Optimal experimental design

To define an OED, insight on the impact of the parameter's uncertainty on model predictions is required. To analyse this impact, in the current approach, the set of parameters along the PL curves in the confidence region are used (in case of a semi- or infinite confidence region the set of parameters along the PL curve in a region of a few orders of magnitude around the maximum likelihood estimate (MLE) can be used). The spread of the model prediction trajectories drawn for this set of parameters contains information on the parameter's uncertainty impact; a large spread indicates a large impact of the uncertainty of the specific parameter on the specific model prediction. To quantify the individual uncertainty impact of a model parameter θ_i on a model prediction $p_i(t_k)$, the profile likelihood sensitivity (PLS) as defined by Flassig et al. (2015) can be used:

$$s_i(t_k) = \left(\frac{\max\left(\left\{p_i(t_k)\right\}\right) - \min\left(\left\{p_i(t_k)\right\}\right)}{\left\langle \hat{p}(t) \right\rangle_t} \right)^2 \tag{4}$$

where $max/min(\{p_i(t_k)\})$ define the maximum/minimum for a model prediction p_i at time t_k and where the denominator corresponds to the time average of the model prediction for the MLE of the parameters. The overall uncertainty of the set of model parameters (n_θ) for a specific model prediction over a period defined by n_t time-points can be obtained as:

$$s_{tot} = \sum_{i=1}^{n_\theta} \sum_{k=1}^{n_t} s_i(t_k) \tag{5}$$

To reduce the uncertainty on a model prediction, and thus to reduce the uncertainty of the parameters that produce this prediction uncertainty, an additional readout that maximizes the PLS index (s_i or s_{tot} depending on the interest in a single parameter θ_i or in the total set of parameters, respectively) should be selected. When a reduction of the uncertainty of a set of model parameters it pursued, it is furthermore of importance to look for a design with a more or less equal contribution of these parameter uncertainties to the PLS index. Shannon's entropy gives information on this contribution and can be calculated as (Flassig et al., 2015):

$$J_{tot} = \sum_{k=1}^{n_t} J_k = \sum_{k=1}^{n_t} \sum_{i=1}^{n_\theta} -\frac{s_i(t_k)}{\sum_{i=1}^{n_\theta} s_i(t_k)} \log\left(\frac{s_i(t_k)}{\sum_{i=1}^{n_\theta} s_i(t_k)} \right) \tag{6}$$

The larger Shannon's entropy the more homogenous the PLS index of the individual parameters (s_i) contributes to the total PLS index (s_{tot}).

RESULTS

The OED methodology is applied to define the best position to measure the relative humidity in the sample. Four possible designs are compared: a RH-sensor at 1, 2, 3 or 4 cm from the top of the sample. Figure 2 shows the PL curves for the model parameters that define the

hygric properties, and this for the case when the experimental output is limited to the sample's mass increase (and thus without a RH-sensor). The four parameters are practically identifiable, as is the case with the other model parameters (not shown) as well. Based on the parameter sets along the PL curves below the 95% threshold, the trajectories of different model predictions are drawn. As an example, Figure 3a shows the trajectories for RH_4 (= RH at the bottom of the sample), and this for the model parameter *m*. The individual PLS indices as a function of time are given in Figure 3b, and this for the model prediction RH_4. Finally, Figure 3c shows for the four designs the criterion space with the total PLS index and entropy. The maximum PLS index and entropy is obtained for RH_4. Hence, of the four designs this is the best option to choose as an additional readout. Overall, the largest reductions of the confidence intervals are obtained when including data on RH_4, as indicated in Table 2. Note also that some CI's become slightly wider after including a RH-sensor, which might be the result of the extra noise term σ_{RH} included in the parameter estimation process.

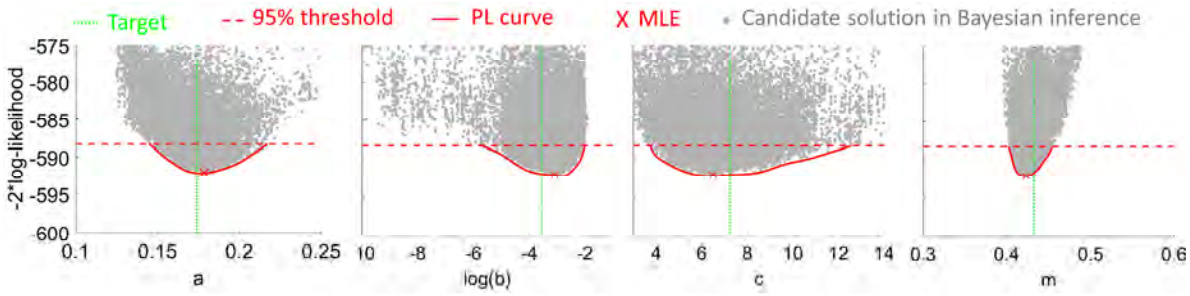
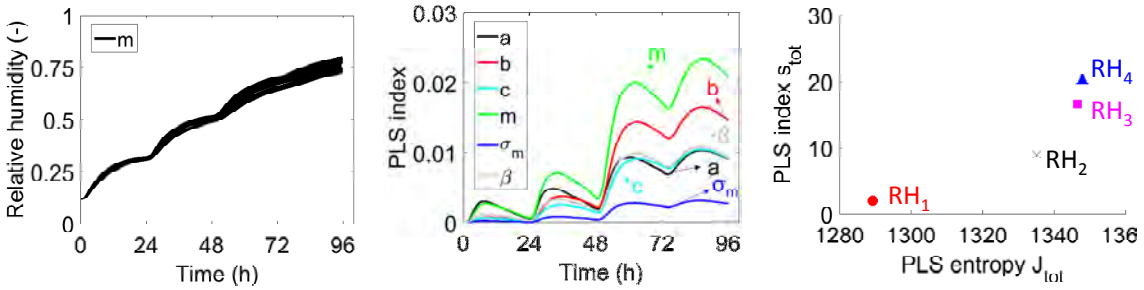


Figure 2. Profile likelihood curves for the model parameters *a, log(b), c* and *m*.



a) b) c)

Figure 3. a) Trajectories of RH_4 along the PLS curve for the parameter *m*, b) PLS indices of the different model parameters for prediction RH_4 and c) criterion space.

Table 2. Upper and lower bounds of the CI's for the initial experiment and the experimental designs. The narrowest CI's are indicated in bold.

	Mass only	Mass + RH_1	Mass + RH_2	Mass + RH_3	Mass + RH_4
a	[0.145; 0.22]	[0.15; 0.2]	[0.154; 0.2]	**[0.16; 0.19]**	[0.142; 0.178]
log(b)	[-6; -2]	[-6; -2]	[-6.8; -2.18]	[-5.8; -3]	**[-4.0; -2]**
c	[3.7; 14]	[4; 13.5]	[5; 15]	[5.6; 13]	**[3.4; 8.7]**
m	[0.405; 0.455]	[0.426; 0.442]	**[0.428; 0.438]**	**[0.428; 0.438]**	**[0.426; 0.436]**
σ_M	[-2; -1.88]	**[-2; -1.94]**	[-1.99; -1.87]	[-1.99; -1.87]	[-1.99; -1.88]
σ_{RH}	/	[-1.84; -1.72]	**[-1.875; -1.77]**	**[-1.875; -1.77]**	[-1.83; -1.72]
ß	[-7.825; -6.85]	[-7.8; -6.5]	[-7.75; -6.9]	[-7.75; -6.9]	**[-7.74; - 6.93]**

DISCUSSION

A sample-based approach that starts from the information in a profile likelihood analysis is used to make a decision on the optimal position of a RH-sensor in a hygric experiment. A RH-

sensor at the bottom of the sample is found to be most efficient in reducing the confidence region. However, it should be kept in mind that such a reduction is not necessarily always accompanied by a closer agreement with the target values.
One of the advantages of the presented sample-based approach is that it is based on information that is already at hand when analysing the practical identifiability based on the PL curves. Additionally, the method yields information on the uncertainty induced by the individual parameters and their contribution to the overall prediction uncertainty. This way, one can also focus on specific model parameters. A disadvantage of the sample-based approach might be its limitation to the readouts in experimental planning. In its current form no direct information on the best input conditions is achieved. Ultimately, as with all OED methods, the OED process is an iterative process that requires a priori knowledge. In the presented approach, a first experiment should be performed before a rational selection of further readouts can be made, whereas in the classical OED approaches initial parameter assumptions have to be made.

CONCLUSIONS

The profile likelihood analysis is shown to be a highly valuable starting point for the optimal design of hygric experiments. Based on the trajectories of internal observables plotted for the parameter sets along the profile likelihood curve uncertain ranges in the prediction space are mapped. A quantitative measure of the parameters' uncertainty (PLS index) and their contribution to the overall uncertainty (PLS entropy) enable a rational selection of highly informative readouts. The major advantage of the methodology is that it is a sample-based approach, which makes that no gradient or curvature information of the objective function is required. This way, especially when dealing with non-linear problems, the methodology is preferable over classical approaches based on Fisher information.

ACKNOWLEDGEMENT

Evy Vereecken is a postdoctoral fellow of the Research Foundation (FWO) – Flanders, Belgium (FWO project 12J5216N). This financial support is gratefully acknowledged.

REFERENCES

Balsa-Canto E., Rodriguez-Fernandez M., Banga J.R., 2007. Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. *Journal of Food Engineering* 82, 178-188.

Dantas L.B., Orlande H.R.B., Cotta R.M. 2002. Estimation of dimensionless parameters of Luikov's system for heat and mass transfer in capillary porous media. *International Journal of Thermal Sciences* 41, 217-227.

Flassig R., Migal I., van der Zalm E., Rihko-Struckmann L., and Sundmacher K. 2015. Rational selection of experimental readout and intervention sites for reducing uncertainties in computational model predictions. *BMC Bioinformatics*, 16, 13p.

Raue A., Kreutz C., Maiwald T., Bachmann J., Schilling M., Klingmüller U., Timmer J. 2009. Structural and practical identifiability analysis of partially observed dynamical models and exploiting the profile likelihood. *Bioinformatics*, 25(15), 1923-1929.

Vereecken E., Roels S., and Janssen H. 2018. Inverse hygric property determination based on dynamic measurements and swarm-intelligence optimisers. *Building and Environment*, 131, 184-196.

Vrugt J. 2016. Markov Chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, 75, 273-316.