AN APPROACH FOR A STATISTICAL MODEL FOR THE USER BEHAVIOUR REGARDING WINDOW VENTILATION IN RESIDENTIAL BUILDINGS

Florian Antretter[1], Christine Mayer[1] and Prof. Dr. rer. nat.Ulrich Wellisch[2]
[1]Fraunhofer-Institut für Bauphysik, Holzkirchen, Germany
[2]University of Applied Sciences, Rosenheim, Germany

ABSTRACT

In this paper a user model for the use of windows based on measurements in residential buildings is developed. Detailed literature review provides an overview of existing user models. Most of them base on measurements in office buildings and their validity for residential buildings is questionable. All models are based on different kinds of statistical methods. After detailed assessment of these methods and the analysis of their applicability to the measurements, a new methodological approach is applied, to take the auto-correlation in the longitudinal data into account. The method of Generalized Estimating Equations (Liang and Zeger 1986) for modelling user behaviour regarding natural ventilation is chosen. By using this class of models, the auto-correlation of the non-normally distributed data can be implied in an adequate way. A binary and a continuous response variable regarding the user behaviour is modelled involving various explanatory variables with different scale levels. A model for predicting the probability of window opening and a model for predicting the window open duration is developed. Both statistical models show positive goodness of fit results by a cross-validation using lift plots.

INTRODUCTION

State of the Art

The first observations regarding the use of windows have been held by Dick in 1951 (Dick 1995), Brundett from 1977 (Brundrett 1977) to 1979 and Lyberg in 1982 (Lyberg 1982). They all concluded that the user behaviour is influenced by outdoor temperature. Warren and Perkins investigated in 1984 (Warren and Parkins 1984) the impact of outdoor temperature, solar radiation and wind speed. Outdoor temperature was the driving variable. In an additional survey, they concluded that most of the window activities take place while arriving or leaving the room. The two main reasons for this behaviour were:

- windows were opened to improve air quality
- windows were opened to improve thermal comfort

The first probability model to predict the user behaviour was developed by Fritsch in 1991 (Fritsch et al. 1990). The model was based on measurements carried out in offices in LESO building in Lausanne, Switzerland. The influence of outdoor temperature, wind speed, solar radiation and indoor temperature was reviewed. Again, outdoor temperature was the driving variable. To calculate the next window state, an inverse function of outdoor temperature and time was developed.

In the late 90s the interaction between user behaviour and thermal comfort became more and more interesting and led to several observations in Pakistan (Nicol and Roaf 1996), (Nicol et al. 1999), Great Britain (Raja, Nicol, and McCartney 1998), (Iftikhar A.Raja et al. 2001) and other European countries (McCartney and Nicol 2002). Based on these preceding observations Nicol (Nicol 2001) developed in 2001 a stochastic model by using logistic regression with outdoor temperature as the main variable. In his opinion, outdoor temperature describes the situation of window opening better than indoor temperature although they both have similar correlation coefficients with the user behaviour. He reasoned that outdoor temperature is an input in every building simulation whereas indoor temperature is an output and can already be flawed.

Reiß, Erhorn and Ohl (Reiß, Erhorn, and Ohl 2001) categorized the users regarding their window opening behaviour in rare, medium or frequent by analysing the longitudinal measurements made in 76 apartments. They also discovered the inhomogeneous window operation in different kinds of rooms and the fact that with more available space per occupant the window operation decreases. In buildings with mechanical ventilation, users operate likewise but on lower level than those ones in buildings without mechanical ventilation.

In conflict with preceding opinions, Nicol and Humphreys proceed in ASHRAE Transactions 2004 (Nicol and Humphreys 2004) and Robinson in Windsor Conference (Robinson 2006) that indoor temperature must be the main influencing factor. Otherwise building architecture and method of construction would be neglected.

After that Rijal (Rijal et al. 2007), (Rijal et al. 2008) released an improved model depending on outdoor and indoor temperature, also known as the Humpreys Algorithm. The user behaviour especially in summer times was analysed by Haldi and Robinson (Haldi and Robinson 2008) in 2006. They tested the acceptability of high indoor temperature with the possibility of personal adaption by clothing, activity, cool drinks and environmental adaption by using blinds, windows and fans. The cognition was that those actions are more likely influenced by indoor conditions than outdoor conditions. Yun and Steemers developed an algorithm to predict the occupant behaviour in summertime (Yun and Steemers 2008). The observation was held in offices with and without night ventilation. They note that most window activities take place at arrival or departure. Due to this conclusion, a separate sub-model was developed to simulate user presence depending on time. The result of the sub-model is used as an input for the main window opening model which is based on Markov-Chains. As this model is based on observations made in summertime, they state that the validity of that model in wintertime is doubtful.

Herkel, Knapp and Pfafferott developed a user model based on logistic regression (Herkel, Knapp, and Pfafferott 2008), using a sub-model to predict user presence. By analysing the window opening data, they found that the proportion of open windows is similar in autumn and spring, highest in summer and lowest in winter. A change in user behaviour regarding window ventilation is indicated by the first warm day in spring and the first cold day in autumn. Andersen made a survey of occupants in Danish dwellings in 2009 (Andersen et al. 2009). He discovered that besides outdoor temperature also gender, solar radiation, size of the dwelling, kind of ownership, noise and lighting influences the occupant behaviour regarding the use of windows.

Finally Haldi and Robinson (Haldi and Robinson 2009) created a model which implies personal behaviour of the occupants. Based on over seven years of measurements in the LESO building in Lausanne, the users were categorized in three levels: low, medium and high window activity. Three different mathematic methods (logistic regression, Markov-Chains and continuous-time random process) to simulate user behaviour were observed to their ability to present the window opening act. They conclude that a hybrid model combining discrete-time Markov process and continuous-time model predicted the data best.

It is obvious, that the presence or absence of the user is one of the most important variables by simulating user behaviour. To respect this some separate presence models have been developed.

Wang, Federspeil and Rubinstein (Wang, Federspiel, and Rubinstein 2010) created a model based on presence measurements in offices. The presence probability is calculated depending on the time for each calculation step. This model was improved by Page, Robinson and Scatezzini (Page et al. 2008). They take long periods of absence, for example holiday or meetings, into account. The results of these presence models can also be used for estimating inner loads caused by occupants, such as CO_2 emission or moisture production rates, respectively inner heat gains. Furthermore, it could be an input variable to determine the use of other building equipment, such as heaters, fans and air-conditioning.

Summary

Existing models are based on measurements in office buildings. User behaviour in residential buildings has not yet been determined. In these models thermal conditions are the main influencing variable on user behaviour regarding natural ventilation. Whether interior or exterior temperature is the main variable is still under discussion. So far, different mathematical and statistical methods have been used for modelling user behaviour. Most of them do not consider the autocorrelation structure in an adequate way which can lead to misspecifications in the resulting statistical models.

STATISTICAL METHODS

As described in the introduction there are various ways for modelling the user behaviour. The main models methodologies used so far are explained below, with a critical discussion of their applicability for modelling user behaviour regarding window ventilation. Most of the existing models do not imply the auto-correlation structure in the data in an adequate way. For example using the usual statistical significance tests in a logistic regression model where the independence condition for the observations is violated it can lead to false p-values of the parameter tests and therefore to an irregular statistical model. More precisely without an correct treatment of the auto-correlation structure in the data the resulting models can include non-significant influencing variables and project false functional effects to the response variables. For the user model presented in this paper a method is introduced which respects the specific properties of the longitudinal data and treats the autocorrelation structure as a nuisance parameter. Using this technique one arrives a class of semi-parametric regression models, which yield valid significance tests for the development of a statistical model. The result of the statistical model is an multivariate evaluation of the significance of the influencing variables with respect to their effect to window ventilation as well as specific estimating functions for the expected response variables. The developed estimating functions yield a prediction of the probability of window opening and a prediction of the window open duration. In the following we first illustrate generalized linear models, which generalize linear models to the application of our specific response variables and their distributions. Second we introduce another extension of the generalized linear models to the method of generalized estimating equations.

Generalized Linear Models (GLM)

For applying the statistical significance theory in linear models a normal distribution of the response variable is assumed. This limitation is generalized by the method of GLM. In a linear model the expectation of the response is directly given by

$$E(Y_i) = \beta_0 + \beta_1 \times x_{i1} + \beta_2 \times x_{i2} + ... + \beta_p \times x_{ip}$$

where $\beta = (\beta_0, \ldots, \beta_p)$ is the unknown vector of the model parameters, which are to be estimated, x_{ik} is the value of the k-th influencing variable at observation i and Y_i is the response variable of the i-th observation.

Combining the responses of all observations Y_i in the vector Y the expectation $E(Y)$ can than be written as

$$E(Y) = \mu = X\beta$$

with X being the design matrix of the linear model including the influencing variables.

In the theory of GLM the response variable is assumed to be generated from a particular distribution in the exponential family (generalizing the assumption of a normal distribution in the ordinary linear model to distributions such as the binomial distribution and the Poisson distribution). The structure equation of the expected response is extended by a link-function which connects the expectation of the response variable to the linear predictor $\eta_i = x_i\beta$, where x_i is the vector of the influencing variables of the i-th observation.

The relationship between the expectation of a response $E(Y_i) = \mu_i$ and the linear predictor $\eta_i = x_i\beta$ is defined by a monotone, invertible function g, called link-function. With the existing inverse link-function g^{-1} in a GLM the following model equations are valid.

$$g(\mu_i) = x_i\beta \qquad \text{link-function}$$

$$\mu_i = g^{-1}(x_i\beta) \qquad \text{inverse link-function}$$

Depending on the application (i.e. the kind of the response and the underlying distribution), an adequate form of the link-function can be used. Using a link-function the range of the response variable can be adapted to the specific situation. Following Table 1 shows the natural link-function and the related inverse link-function for a binomial distributed response variable. The resulting specific GLM is well known as the Logistic Regression Model.

TABLE 1. The binomial link- function

distribution	Link-function	Inverse function/ Expected value
Binomial	$\eta_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$	$\mu_i = \frac{e^{(\eta_i)}}{1+e^{(\eta_i)}}$

Logistic Regression

The Logistic Regression model allows the prediction of the probability of an event with respect to the influence variables. Because of the logarithmic link, the range of the predicted response variable is the unit interval $]0,1[$. In this case, the predicted response corresponds to the estimated probability that the event occurs. The following graph shows the transformation by the link-fuction (with different model parameters) of the linear predictor into the unit interval.

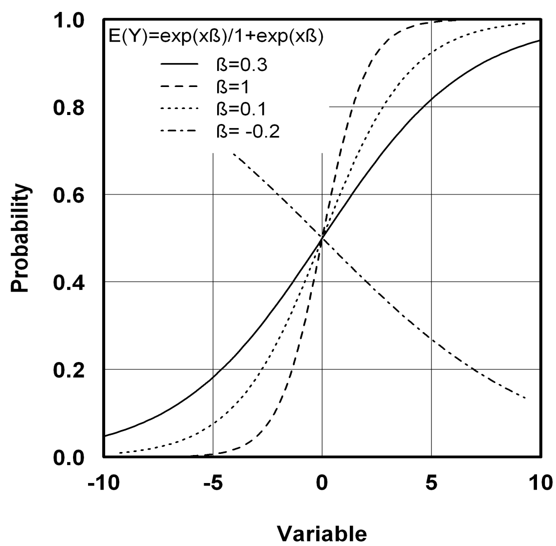

Figure 1. Logistic Regression with different parameters

The stochastically independence of the single observations is a serious limitation of GLM. However window opening data has a strong time dependency. Any observation is really indexed by the room and the time step of the measurement, too. Therefore a stochastic independence between all readings (observations) cannot be assumed. In other words GLM are a type of regression models, which are very appropriate to deal with our response variables and their distributions. But taking no notice of the underlying dependency structure in the data leads to an irregular model building process by using the inappropriate significance tests of GLM. Another adaptation of the model class is needed.

Generalized Estimating Equations (GEE)

The window opening measurements have a very specific structure. The data may be considered as independent observations with respect to the different rooms or buildings. There is however an obvious time dependency between the observations at one room or building with respect to the single time steps. The data situation is, therefore, as follows: many short time series (within correlated observations by the time variable) where the observations are independent according to different time series. In the statistical science, one denotes such a data structure as longitudinal data. To consider this special nature of the data, the measured window states of one day in one room have to be clustered. The model development can be made by using Generalized Estimating Equations, which are introduced by Liang and Zeger (Liang and Zeger 1986) in 1986. GEE are extensions of GLM, just with the feature of considering correlation between single measurements. The method of GEE is a semi-parametric approach, which yields to estimating equations for the regression parameters without a full specification of the joint distribution of the observations. In this way the GEE are an extension of the usual Likelihood-based estimating equation in GLM. By using GEE the correlation structure is treated as a nuisance term and is included as a so called working correlation assumption (vgl. Liang and Zeger 1986) in the estimating algorithm. GEE yields consistent estimators of the regression parameters and the significance tests are correctly adjusted to the situation of correlated data. The consistency of the estimators and the applicability of the significance tests still remain valid if the assumed working correlation structure is incorrect. The focus of modelling by GEE is (similarly as in the GLM context) the modelling of the mean dependence structure between the response variable and the influencing variables. The result of a GEE analysis (including the selection of the influencing variables by the statistical significance theory and the final model choose) is a valid statistical model which describes the relationship of the response variable and the influencing variables.

Different kinds of auto-correlation can be considered by GEE. The correlation can be unstructured, equal in every clustered data set or the correlation can be a function of the time distance between the single measurements in a clustered data set.

The estimated correlation structure of the data is presented in Table 2. The resulting matrix shows the correlation coefficients between opening duration in every single hour of a day. To illustrate the correlation trend through the day, the colour gets darker with higher coefficients. The figure shows clearly, that window openings are highly correlated when they are close in time. The correlation values become smaller with rising time distance. During the hours of the night, more previous hours correlate with the actual hourly window state.

TABLE 2. Auto-correlation matrix for window openings on a 24 hour basis

	1	2	3	4	5	6	7	8	9	0	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1.00	0.93	0.89	0.87	0.86	0.84	0.81	0.74	0.66	0.60	0.55	0.51	0.49	0.47	0.47	0.45	0.44	0.45	0.45	0.46	0.48	0.51	0.55	0.59
2	0.93	1.00	0.97	0.95	0.93	0.92	0.88	0.80	0.69	0.61	0.55	0.51	0.49	0.46	0.46	0.44	0.44	0.44	0.44	0.45	0.46	0.49	0.53	0.58
3	0.89	0.97	1.00	0.99	0.97	0.96	0.91	0.83	0.71	0.61	0.55	0.50	0.48	0.45	0.45	0.43	0.42	0.43	0.43	0.43	0.45	0.48	0.51	0.56
4	0.87	0.95	0.99	1.00	0.99	0.97	0.93	0.85	0.72	0.62	0.55	0.50	0.47	0.45	0.45	0.43	0.42	0.42	0.42	0.43	0.44	0.47	0.50	0.56
5	0.86	0.93	0.97	0.99	1.00	0.99	0.94	0.86	0.73	0.62	0.55	0.51	0.47	0.45	0.45	0.43	0.42	0.42	0.42	0.43	0.44	0.47	0.50	0.55
6	0.84	0.92	0.96	0.97	0.99	1.00	0.96	0.88	0.74	0.63	0.56	0.51	0.47	0.45	0.45	0.43	0.42	0.42	0.42	0.42	0.44	0.47	0.50	0.55
7	0.81	0.88	0.91	0.93	0.94	0.96	1.00	0.91	0.76	0.65	0.57	0.52	0.48	0.45	0.45	0.43	0.42	0.43	0.43	0.44	0.45	0.48	0.51	0.56
8	0.74	0.80	0.83	0.85	0.86	0.88	0.91	1.00	0.86	0.71	0.63	0.56	0.52	0.49	0.49	0.48	0.46	0.47	0.47	0.47	0.48	0.50	0.52	0.56
9	0.66	0.69	0.71	0.72	0.73	0.74	0.76	0.86	1.00	0.85	0.72	0.64	0.59	0.56	0.56	0.54	0.53	0.53	0.52	0.52	0.53	0.54	0.54	0.56
10	0.60	0.61	0.61	0.62	0.62	0.63	0.65	0.71	0.85	1.00	0.85	0.73	0.67	0.63	0.63	0.60	0.58	0.57	0.57	0.56	0.56	0.56	0.56	0.57
11	0.55	0.55	0.55	0.55	0.55	0.56	0.57	0.63	0.72	0.85	1.00	0.87	0.76	0.70	0.70	0.65	0.62	0.61	0.60	0.59	0.59	0.58	0.57	0.56
12	0.51	0.51	0.50	0.50	0.51	0.51	0.52	0.56	0.64	0.73	0.87	1.00	0.87	0.77	0.77	0.71	0.67	0.65	0.64	0.62	0.61	0.60	0.57	0.55
13	0.49	0.49	0.48	0.47	0.47	0.47	0.48	0.52	0.59	0.67	0.76	0.87	1.00	0.88	0.88	0.78	0.73	0.69	0.67	0.65	0.62	0.60	0.57	0.54
14	0.47	0.46	0.45	0.45	0.45	0.45	0.45	0.49	0.56	0.63	0.70	0.77	0.88	1.00	1.00	0.88	0.79	0.73	0.71	0.67	0.64	0.61	0.57	0.54
15	0.47	0.46	0.45	0.45	0.45	0.45	0.45	0.49	0.56	0.63	0.70	0.77	0.88	1.00	1.00	0.88	0.79	0.73	0.71	0.67	0.64	0.61	0.57	0.54
16	0.45	0.44	0.43	0.43	0.43	0.43	0.43	0.48	0.54	0.60	0.65	0.71	0.78	0.88	0.88	1.00	0.89	0.80	0.75	0.70	0.67	0.63	0.58	0.54
17	0.44	0.44	0.42	0.42	0.42	0.42	0.42	0.46	0.53	0.58	0.62	0.67	0.73	0.79	0.79	0.89	1.00	0.89	0.81	0.74	0.69	0.65	0.59	0.55
18	0.45	0.44	0.43	0.42	0.42	0.42	0.43	0.47	0.53	0.57	0.61	0.65	0.69	0.73	0.73	0.80	0.89	1.00	0.90	0.80	0.73	0.67	0.61	0.57
19	0.45	0.44	0.43	0.42	0.42	0.42	0.43	0.47	0.52	0.57	0.60	0.64	0.67	0.71	0.71	0.75	0.81	0.90	1.00	0.89	0.79	0.71	0.64	0.58
20	0.46	0.45	0.43	0.43	0.43	0.42	0.44	0.47	0.52	0.56	0.59	0.62	0.65	0.67	0.67	0.70	0.74	0.80	0.89	1.00	0.88	0.76	0.67	0.60
21	0.48	0.46	0.45	0.44	0.44	0.44	0.45	0.48	0.53	0.56	0.59	0.61	0.62	0.64	0.64	0.67	0.69	0.73	0.79	0.88	1.00	0.87	0.75	0.65
22	0.51	0.49	0.48	0.47	0.47	0.47	0.48	0.50	0.54	0.56	0.58	0.60	0.60	0.61	0.61	0.63	0.65	0.67	0.71	0.76	0.87	1.00	0.87	0.73
23	0.55	0.53	0.51	0.50	0.50	0.50	0.51	0.52	0.54	0.56	0.57	0.57	0.57	0.57	0.57	0.58	0.59	0.61	0.64	0.67	0.75	0.87	1.00	0.87
24	0.59	0.58	0.56	0.56	0.55	0.55	0.56	0.56	0.56	0.57	0.56	0.55	0.54	0.54	0.54	0.54	0.55	0.57	0.58	0.60	0.65	0.73	0.87	1.00

DATA

The model described in this paper bases on data from 17 residential buildings in Germany. The Fraunhofer-Institute for Building Physics in Stuttgart held the monitoring between 1996 and 2005. All of the buildings were part of different research projects for determining energy use and energy saving in residential buildings. Seven of them are on passive house standard and two of them are so called three-litre buildings. Each building was monitored for about 2 years. The following hourly measurements were recorded:

- Exterior conditions (exterior temperature, air humidity, wind speed)
- Interior conditions (interior temperature, air humidity)
- Duration of open window (seconds per hour summarized for all windows in the room)

Additional available information is time of the measurement, kind of the room (bathroom, living room, kitchen, sleeping), air tightness of the building, size of the room and if there is a mechanical ventilation in the room.

Before data could be used for modelling, a binary value of the window state was produced. It was set to 1, if the sum of measured window openings of all windows in a room is bigger than one. That means if one window was opened in that certain hour, the value was set to 1, otherwise to 0.

Important for modelling user behaviour are conditions just before a window is opened. In the case of room temperature, the value of the previous hour is used for modelling. This approach avoids falsification in cases where e.g. high ventilation rates already resulted in low room temperatures (which would result in a causally false model predicting a high probability of open windows at low temperatures), etc.

For meaningful validation of the model at the end, the data were divided into two parts: modelling data and validation data. It was important not to part clusters, i.e. measurements from the same day in the same room, as described in the modelling chapter.

Analyses of the data and modelling was executed with R (GNU R 2010).

MODELLING

For modelling with Generalized Estimation Equations in R the package geepack can be used. It was developed by Hakeloh and Højsgaard in 2005 (Halekoh, Højsgaard and Yan 2006).

Geepack requires various inputs. First, the coherence of response and explaining variables has to be defined (for example quadratic or cubic). The distribution of the response variable has to be selected. For the open probability, the binomial distribution was chosen. For the open duration, we choose, according to the underlying quasi-continuous measuring scale, the Poisson distribution. As mentioned above the data has to be clustered for considering the auto-correlation in the data. An analysis of the auto-correlation of the data shows high correlation between the hours of one day (Table 2). Therefore, our decision was to cluster the values on a daily basis assuming that the window opening behaviour on successive days is not correlated, but successive hours of one day are. An additional vector specifying these clusters for each day is added. Following code was used for clear nomination: House – room – date. Each of the data clusters consist therefore of 24 values of each variable. Days with less than 24 valid measurements were removed. With this so-called ID, the autocorrelation in each of the clusters can be considered. It is also important to order the values in each cluster (from 00:00 to 23:00 o'clock) and to specify the correlation type. In this case we use the time depended AR1 structure, which assumes that the observations are related to their own past values through a first order autoregressive process.

The result output gives an estimated parameter for each variable and the belonging p-value of the corresponding parameter tests. To determine the most influencing variables for modeling the response variable, the model was developed stepwise with every combination of the covariates. The intention was to find variables which are not important for predicting the response variable. The p-value describes the statistical significance of a single variable with respect to the influence on the response variable; the smaller the p-value, the higher the influence. If a p-value is higher than 0.05 it is common to regard the influence as pretty low and the variable is not necessary for the modeling. So covariates with a p-value of 0.05 or higher are not included in the final model.

Not taking the auto-correlation into account and performing a univariate logistic regression would result in a p-value for wind speed (2-6 m/s) for the binary response variable of 0.011 and the variable would be incorporated in the model. Using GEEs the p-value is 0.343 and the variable would not be part of the model – so the auto-correlation has an effect on the proper selection of influencing variables.

RESULTS AND DISCUSSION

Binary response variable

For modelling open probability, the binary window state is used. After stepwise modelling with all variables, following final model was developed for predicting the probability that any window in a room is open in the respective hour:

$$\mu = \frac{e^{(-1{,}76+0{,}10AT+0{,}007AF-0{,}074RT+0{,}011TZ-0{,}387W-0{,}517B+0{,}46S)}}{1+e^{(-1{,}76+0{,}10AT+0{,}007AF-0{,}074RT+0{,}011TZ-0{,}387W-0{,}517B+0{,}46S)}}$$

The significant variables are

AT – outdoor temperature
AF – outdoor humidity
RT – room temperature
TZ - time of the day
Kind of room (W = living room, B = bathroom, S = sleeping, kitchen is included in the intercept)

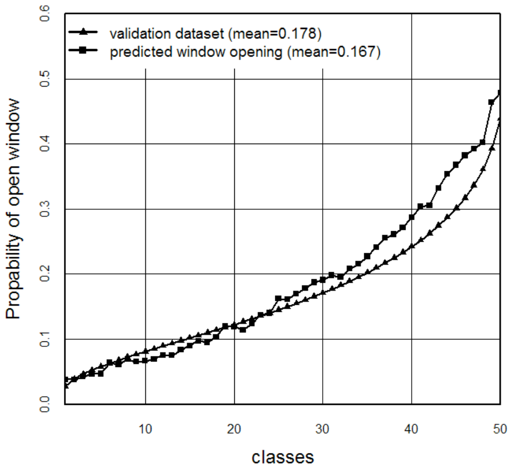

Figure 2. Lift plot for binary response variable

A lift plot in figure 2 shows the validity of the model on the validation dataset, which is not used for development of the model. For this graphical analysis, the model was applied on the validation data set. The predicted open probability for every time unit and the related measured window state were graded by the predicted values. Afterwards the values were merged into 50 equal classes and the mean value was calculated. By comparing values from the predicted and from the measured open probability respectively window state, the validity of the developed model is shown.

The squares show modelled open probability ordered by size. Triangles show belonging measured window states. Both lines are having approximately the same course and are close together. For example that means that under circumstances where rarely windows have been opened, the model predicts correctly very low open probabilities.

Continuous response variable

For modelling the open duration, the same technique for the selection of the model was used as described for the binary response variable. After stepwise modelling with all variables, following final model was developed for predicting the duration of the opening of a window:

$$\mu = e^{(6{,}1+0.1AT-0{,}27RT+0{,}007AF-0{,}02WI+0{,}19W-0{,}28B+0{,}58S)}$$

WI – wind speed

In this case, wind speed was found to be more important than time of the day. The lift plot (FIG. 3) for the continuous model particularly shows a very good fit for short window openings (1-10 minutes).

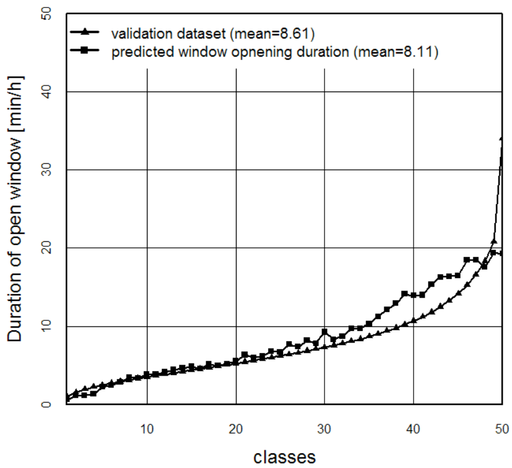

Figure 3. Lift plot for continuous target variable

For higher ventilation time the model underestimates the reality. The assumed Poisson distribution of the response variable could be a reason for this weakness of fit.

CONCLUSIONS

This paper describes the development of two statistical models for predicting window open behaviour in residential buildings. They are the first user model based on Generalized Estimation Equations and respect auto-correlation of the data in an adequate way as a nuisance term. The auto-correlation of the dataset proved to be important and has to be recognized. The validation of the models by cross-validation on an extra test data set and the resulting lift plots show good results. The developed methodology is applicable to different auto-correlated processes, which are user influenced in a building, e.g. blind operation, and of course to window opening in office buildings.

Significant variables regarding window opening by the user in residential buildings are indoor and outdoor temperature, outdoor humidity, time of the day and wind speed. The kind of room plays – contrary to office buildings, where usually only one room type is considered – an important role.

Still, the developed models can not be implemented into building simulation straightforward. Some enhancements to the models and further research are necessary.

For the development of the model in this paper the open duration of all windows in a room were added, so that windows that are rarely or never opened do not falsify the model. With this treatment predicted open duration of over 60 minutes per hour in one room are possible. This handling could be improved. The deviances in the lift plot of the continuous model are probably caused by using an inappropriate Poisson distribution of the response variable. This assumption needs verification, too. As GEE are an extension of GLM, the underlying linear dependency is not always given, e.g. dependency with exterior temperature, where a decrease in opening activities is found above a certain exterior temperature whereas the linear model assumes a continuous increase. A next step could be to develop sub-models for regions where linear dependency can be assumed or an extension of the modelling by adequate transformations to nonlinear functional influences of the covariates to the linear predictor. User presence was not measured and therefore this information was not available for the modelling. A combination of the methodology with a presence model of building inhabitants would allow a more precise window model, but besides it would allow to determine all other user actions as well as interior loads caused by the presence of people in more detail.

The basis of every model is the dataset. The dataset for this model has a time step of one hour. Describing an action (window opening and closing) which lasts just a few minutes on this basis is difficult. A dataset with smaller time step is needed to define exact circumstances at the moment the window was opened or closed. The window state for this model was measured with a simple contact resistance. This allows no differentiation between fully opened and tilted window states. For further research, a measurement method is recommended that allows assessing the area available for air exchange. This is the main input variable for a ventilation model.

Further interesting variables should be considered in future measurements, e.g. occupancy and CO_2 concentration as interior factors, noise and odour pollution as outdoor factors.

For a generalisation of the results, measurements in apartments and multifamily houses are required. To determine the influence of mechanical ventilation buildings with and without this technical adjustment should be assessed. More information about the residents would be interesting as well, for example sex, smoker – non smoker, age, working full- time or not. If more information about residents is available, different models for certain user types can be developed.

For an implementation in building simulation, the air exchange between zones or between a zone and the exterior is required. The information if a window is open or not needs to be coupled with a model that allows assessing the amount of air exchanged because of that certain window state. This means the window-opening model needs to be coupled with other models predicting the resulting air exchange.

To sum it up, Generalized Estimating Equations were successfully used to model user behaviour regarding window ventilation in residential buildings. The methodology can be applied to other processes with auto-correlated data. Implementation in building simulation requires some enhancements of the developed models and coupling with other models to predict the resulting air exchange.

ACKNOWLEDGEMENTS

This project was funded by the German Federal Ministry of Economics and Technology.

REFERENCES

Andersen RV, Toftum J, Andersen KK, and Olesen BW. 2009. Survey of occupant behaviour and control of indoor envrionment in Danish dwellings. Energy and Buildings 41: 11-16.

Brundrett RW. 1977. Ventilation: a behavioural aproach., 289-298.

Dick T. 1995. Ventilation research in occupied houses., 306-326.

Fritsch R, Kohler A, Nygard-Ferguson M, and Scartezzini JL. 1990. A Stochastic Model of User Behaviour Regarding Ventilaton., 173-181.

GNU R. 2010. Language and environment for statistical computing and graphics.

Haldi F, and Robinson D. 2008. On the behaviour and adaption of office occupants. Building and Environment 43: 2163-2177.

-----. 2009. Interactions with window openings by office occupants., 2378-2.

Halekoh U, Højsgaard S, and Yan J. 2006. The R Package geepack for Generalized Estimating Equastions. Journal of Statistical Software 15 (2).

Herkel S, Knapp U, and Pfafferott J. 2008. Towards a stochastic model regarding the manual control of windows in office buildings. Building and Environment 43: 588-6.

Iftikhar A.Raja, Nicol F, McCartney KJ, and Humphreys MA. 2001. Thermal comfort: use of controls in naturally ventilated buildings., 235-442.

Liang K-Y, and Zeger SL. 1986. Longitudinal Data Analysis Using Generalized Linear Models., 13-22.

Lyberg. 1982. Energy losses due to airing by occupants. at Dubling.

McCartney KJ, and Nicol F. 2002. Developing an adaptive control algorithm for Europe., 623-635.

Nicol F. 2001. Characterising occupant behaviour in buildings: Towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans. Paper presented at Seventh IBPSA Conference, Rio de Janeiro, Brazil.

Nicol F, and Humphreys MA. 2004. A stochastic approach to thermal comfort - occupant behaviour and energy use in buildings. In ASHRAE Transactions, 554-568.

Nicol F, Raja IA, Allaudil A, and Jamy GN. 1999. Climatic variations in comfortable temperatures: the Pakistan project., 261-279.

Nicol F, and Roaf S. 1996. Pioneering new indoor temperature standards: the Pakistan project., 169-174.

Page J, Robinson D, Morel N, and Scartezzini JL. 2008. A generalised stochastic model for the simulation of occupant presence. Energy and Buildings 40: 83-9.

Raja IA, Nicol F, and McCartney KJ. 1998. Natural ventilated builsings: use of controls for changing indoor climate., 391-394.

Reiß J, Erhorn H, and Ohl J. 2001. Klassifizierung des Nutzerverhaltens bei der Fensterlüftung. HLH Heizung Lüftung/Klima Haustechnik 52 (8): 22-26.

Rijal HB, Tuohy P, Humphreys MA, Nicol F, Samuel A., and Clarke J. 2007. Using results from field surveys to predict the effect of open windows on themal comfort and energy use in buildings. Energy and Buildings 39: 823-836.

Rijal HB, Tuohy P, Nicol F, Humphreys MA, Samuel A., and Clarke J. 2008. Development of an adaptive window-opening algorithm to predict the thermal comfort, energy use and overheating in buildings. Journal of Building Perfromance Simulation 1 (1): 17-30.

Robinson D. 2006. Some trends and research needs in energy and comfort prediction. Paper presented at Windsor Conferece 2006.

Wang D, Federspiel CC, and Rubinstein F. 2010. Modeling occupancy in single person offices. Energy and Buildings 37: 121-1.

Warren PR, and Parkins LM. 1984. Window-opening behaviour in office buildings., 89-101.

Yun GY, and Steemers K. 2008. Time dependent occupant behaviour models of window control in summer. Building and Environment 43: 1471-1482.