

PAPER • OPEN ACCESS

Simulating heat load profiles in buildings using mixed effects models

To cite this article: J Palmer Real *et al* 2021 *J. Phys.: Conf. Ser.* **2069** 012138

View the [article online](#) for updates and enhancements.

You may also like

- [Harvesting big data from residential building energy performance certificates: retrofitting and climate change mitigation insights at a regional scale](#)
João Pedro Gouveia and Pedro Palma
- [Post-evaluation of a ground source heat pump system for residential space heating in Shanghai China](#)
Y Lei, H W Tan and L Z Wang
- [Wood burning habits and its effect on the electrical energy demand of a retrofitted Norwegian detached house](#)
L. C. Felius, M. Thalfeldt, L. Georges et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada

Extended abstract submission deadline: Dec 17, 2021

Connect. Engage. Champion. Empower. Accelerate.
Move science forward



Submit your abstract



Simulating heat load profiles in buildings using mixed effects models

J Palmer Real¹, J Kloppenborg Møller¹, C Rasmussen¹, K B Lindberg², I Sartori² and H Madsen¹

¹ DTU Compute, Asmussens Allé, Building 303B, Lyngby, Denmark

² SINTEF Building and Infrastructure, Oslo, Norway

E-mail: jpre@dtu.dk

Abstract. The landscape of buildings is a diverse one and long-term energy system planning requires simulation tools that can capture such diversity. This work proposes a model for simulating the space-heating consumption of buildings using a linear mixed-effects model. This modelling framework captures the noise caused by the differences that are not being measured between individual buildings; e.g. the preferences of their occupants. The proposed model uses outdoor temperature and space-heating consumption measured at hourly resolution; thus, the model is able to predict the intra-day variations as well as longer effects. Given the stochastic nature of the simulation, the prediction interval of the simulation can be estimated, which defines a region where the consumption of any unobserved building will fall in. A whole year has been simulated and compared to out-of-sample measurements from the same period. The results show that the out-of-sample data is virtually always inside the estimated 90% prediction interval. This work uses data from Norwegian schools, although the model is general and can be built for other building categories. This amount of detail allows energy planners to draw a varied and realistic map of the future energy needs for a given location.

1. Introduction

In order to plan and develop strategies for the future power market, it is necessary to create tools that reliably represent it. Such tools need to be able to predict the energy consumption of the different systems that form the energy landscape. The current tools dedicated to this task are often based on trends based on historical data [1; 2; 3]. As the power sector shifts towards a more flexible framework with high integration of renewable energy sources, it is necessary to re-visit these methods used for long-term forecasting [4].

Buildings take a significant portion of the total energy use [5]; thus, modelling their consumption is a key task in order to develop suitable forecasting tools. From the total energy consumed in buildings, in Europe, the major part is dedicated to space heating [6], and there exist an extensive literature focused on modelling it [7]. A well-known example is the energy signature method (ES); a data driven approach that quantifies a building thermal performance based, mainly, on the outdoor temperature [8]. In general, the ES is a static method, even though their parameters might change over the course of the year [9].

The proposed model in this work aims to capture the dynamic nature of heat consumption; given that, when predicting the energy use of a building, capturing the peaks that take place



during the day is of particular interest. Similarly, it is crucial to see how this pattern changes as buildings become more efficient.

There are factors that impact the heat consumption of a building that can be specially difficult to measure. A clear example is the behaviour of the occupants; predicting it is far from being trivial due to their noisy nature and intrusive measurement set ups [10]. A long-term forecasting tool needs to account for such phenomena to be general enough to represent the existing variety of the building stock. This means capturing the inherent differences from building to building, caused by random unobserved events. In this work, this is done by using a mixed effects model. We depart from the work done in [11], where a linear fixed model was fit to generate an hourly profile of the energy consumption in buildings. Then, a random term is added to the fixed model structure, to account for the individual differences between buildings. This addition reduces model complexity, quantifies the differences between observed buildings and facilitates predicting the consumption from unobserved ones.

The outline of this work is as follows: first, in section 2, the mixed effects model structure is introduced and explained in the context of modelling building energy load; section 3 presents the data used to fit the model; section 4 displays the results using mixed effects; finally, section 5 discusses the main results and presents the following steps.

2. Method

Mixed effects models allow to quantify the noise introduced by random qualities that are inherent to the modelled system. This section introduces the main concepts, and later focuses on using mixed effects for modelling the heat load of buildings.

2.1. Mixed models

Linear mixed effects models, or linear mixed models, are a generalization of the classical linear model which follows the structure

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\epsilon}; \quad (1)$$

where \mathbf{Y} , \mathbf{X} and \mathbf{Z} are known matrices, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{U} \sim N(\mathbf{0}, \boldsymbol{\Psi})$. In Equation (1), $\boldsymbol{\beta}$ represents the *fixed effects*, while \mathbf{U} are the *random effects*. Therefore, these models follow a hierarchical structure since there is an underlying model structure, defined by \mathbf{X} and \mathbf{Z} , which is affected by a higher-level random variable, \mathbf{U} . Then, an arbitrary observation of \mathbf{Y} , Y_{ij} , has two sources of noise: the random effects $U_i \subset \mathbf{U}$, and the noise of the model, ϵ_{ij} . Hence, given the linear structure of Equation (1), Y_{ij} can be written as

$$Y_{ij} = \sum_{l=1}^L X_{jl}\beta_l + Z_i U_i + \epsilon_{ij}. \quad (2)$$

Notice that the sub-index i denotes a category of observations, which introduces the noise U_i ; whereas j marks the number of available observations. In addition, the term $l \in \{1, \dots, L\}$ represents the number of fixed effects.

When the measurements of the random variable, Y_{ij} , are taken at regular time intervals, the model in Equation (2) can be interpreted as a time-series. In this case, the sub-index j is substituted by t , to denote the time dependency. This formulation, has been extensively used in pharmaco-kinetics, when testing a medicine in a subset of a population [12]. The metabolism of each tested individual might affect the response to the medicine; however, there are latent effects that are common to all subjects. In this simplified example, the latent effects are captured by $\boldsymbol{\beta}$, whereas \mathbf{U} characterizes the variability of the results introduced solely by the individuals that are part of the experiment. For more details and examples on mixed models, see [13].

2.2. Mixed models for buildings

There are factors that affect the energy consumption of buildings that are difficult to identify and measure; such as the preferences of the building users. Such factors cause random differences between the energy behaviour of individual buildings. The cause of those differences can be understood as a random effect. Then, using a mixed model, it is possible to identify the distribution that characterizes the differences between individual buildings. As summarized in Figure 1, estimating the distribution has two different outcomes:

- *Profiling.* Using the observations to estimate the random effects of each particular building from the measured ensemble of buildings, i.e. estimating $\hat{U}_i \forall i \in \{1, \dots, k\}$.
- *Sampling.* Simulating a representative realization of a building that has not been observed, using random effects sampled from the estimated distribution \hat{U} .

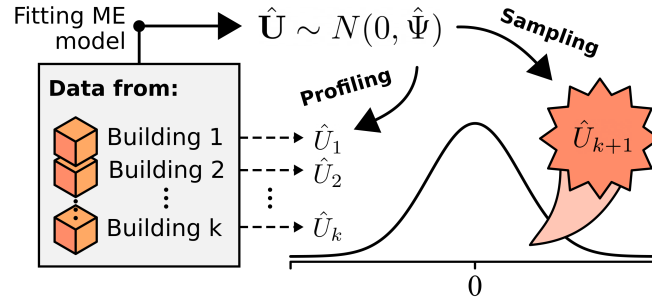


Figure 1: Schematic representation of the outcomes of fitting a mixed effects model

2.2.1. A mixed energy signature. The energy signature (ES) is a method to evaluate the energy performance of buildings. The ES model is a linear model that, using coarse data aggregation, returns the heat loss coefficient (HLC) and the base temperature (T_b). The former coefficient quantifies the energy efficiency of a building, and the latter is the outdoor temperature at which that building is in thermal balance. Thus, both parameters characterize the energy efficiency of a single building. In its simplest form, the ES has the following structure

$$\Phi = \begin{cases} \alpha_0 + \beta_0 T_{\text{out}} + \epsilon & \text{if heating period} \\ \Phi_0 + \epsilon & \text{otherwise} \end{cases}, \quad (3)$$

where Φ is the heat load, T_{out} is the outdoor temperature. The independent term, α_0 , represents unmodelled heat losses; β_0 is the HLC, and $T_b = \alpha_0/\beta_0$; lastly, Φ_0 represents the residual heat load during periods where there is no weather dependence, i.e. $T_{\text{out}} > T_b$. Working with multiple buildings, it is fair to assume that there will be differences in their parameters, $\{\alpha_0, \beta_0, \Phi_0\}$, due to un-measured differences across the building population. Then, the model in Equation (3) can be extended to a mixed effect formulation to capture those differences as random effects. Hence, the heating regime of Equation (3) becomes

$$\Phi_i = \underbrace{\alpha + \beta T_{\text{out}}}_{\text{Fixed effects}} + \underbrace{U_{i,0} + U_{i,1} T_{\text{out}}}_{\text{Random effects}} + \epsilon, \quad (4)$$

where $U_{i,0} \sim N(0, \sigma_0)$ and $U_{i,1} \sim N(0, \sigma_1)$. Re-writing Equation (4) into

$$\Phi_i = \alpha + U_{i,0} + (\beta + U_{i,1}) T_{\text{out}} + \epsilon, \quad (5)$$

it can be notice that, the unmodelled heat losses now contain the random effects $U_{i,0}$; similarly, the heat loss coefficient, depends on $U_{i,1}$. Then, using the model described in Equation (5), it is possible to retrieve the heating performance parameters from each of the buildings of the population. In other words, for each building, it is possible to obtain its heat loss coefficient, $\beta_i = \beta + U_{i,1}$, and its base temperature, $T_{b,i} = (\alpha + U_{i,0})/\beta_i$.

2.2.2. The simulation model. The objective of this work is creating a stochastic simulation tool, that predicts the hourly consumption of buildings given the weather conditions. The model needs to be dynamic and be able to predict consumption for the whole year. The proposed model for simulating the heat load is

$$\begin{aligned} \Phi_{i,t} = & \\ \text{Fixed hour effects} \rightarrow & \left[\sum_{j=1}^{24} \varrho_j I_{\{t \in j\}} + W_t \sum_{j=1}^{24} \rho_j I_{\{t \in j\}} \right] I_{\{t \in \Omega_{\text{WD}}\}} + \\ \text{Fixed seasonal intercepts} \rightarrow & \theta_1 + \theta_2 W_t + \\ \text{Fixed weather effects} \rightarrow & \theta_3 \Delta T_{i,t} + \\ \text{Random effects} \rightarrow & U_{i,2} + U_{i,3} W_t + U_{i,4} \Delta T_{i,t} + \\ \text{Residuals} \rightarrow & \epsilon_{i,t} = \varphi_1 \epsilon_{i,t-1} + \xi_{i,t}, \end{aligned} \quad (6)$$

where, $I_{\{\cdot\}}$ is the indicator function, which equals to 1 when the condition in $\{\cdot\}$ is fulfilled and equals 0 otherwise. $S_t = I_{\{T_{b,i} < T_{\text{out},t}\}}$, $W_t = I_{\{T_{b,i} > T_{\text{out},t}\}}$ and $\Delta T_{i,t} = (T_{b,i} - T_{\text{out},t}) I_{\{T_{b,i} > T_{\text{out},t}\}}$; lastly, Ω_{WD} is the subset of work days. Hence, the model in Equation (6) uses the previously introduced $T_{b,i}$, to discern between heating and non-heating season. The model accounts for a fixed hourly schedule, which depends on the season. Additionally, each season has a fixed heating baseline. Then, the weather effects are introduced through the variable $\Delta T_{i,t}$, that is positive during heating season, and zero otherwise. Furthermore, the variables $\{S_t, W_t, \Delta T_{i,t}\}$ have a random effect over the heat load; i.e. the seasonal heating intercept and the effects of the weather might vary from building to building. Finally, given the hourly sampling, an auto-correlation term has been added in the residuals. Thus, in the model described in Equation (6), the estimated fixed parameters are $\beta = \{\rho_1, \dots, \rho_{24}, \varrho_1, \dots, \varrho_{24}, \theta_1, \theta_2, \theta_3, \varphi_1\}$.

3. Description of the data set

The data set used in this work contains hourly measurements of the heat load of 33 different Norwegian schools. The heating has been split into electric heating and space heating. In this work, only the space heating is used. In addition, the outdoor temperature for the different schools is available. For each building, the measurements span from 1 to 3 years with no gaps in the data. Other general information about all buildings is known, this information contains details like location, efficiency label and built area. In the following results, all schools have the *Regular* efficiency label, which means that they do not comply with the TEK10 efficiency standards and above. Lastly, to normalize the data, the heat load of the schools has been divided by the area of each building, i.e. $[\Phi_{i,t}] = kWh/m^2$.

In order to validate the results of the stochastic simulation, the data has been split into two sets: one for training and one for testing. The training set contains data from 25 buildings. In order to have a balanced data set for training, one year has been chosen arbitrarily for each of the 25 buildings. The final training set contains data from 2009, 2010, 2011, 2012 and 2017. On the other hand, the test set contains data from 8 different buildings. The data from the test set is all from 2010, to ensure consistency with the time stamps and weather data.

4. Results

As explained in the method section, mixed effects models can be used to study the differences between the measured buildings (profiling), or to simulate the behaviour of a new unobserved building (sampling). In this section, both outcomes are presented: first, the results of fitting model in Equation (5) are shown, which allows computing the variable $T_{b,i} \forall i \in \{1, \dots, 33\}$. Then, the model in Equation (6) has been fit using the training set and the simulated consumption has been compared to the test data.

4.1. Profiling energy performance

All buildings in this work have the same energy efficiency label, implying that their response to weather conditions should be similar. Using the model in Equation (5) it has been possible to retrieve energy performance parameters, $\{\beta_i, T_{b,i}\}$, from all individual buildings and compare the population of available buildings. The results can be seen in Figure 2, where it can be noticed that, the base temperature, as well as the HLC, vary significantly from building to building. For reference, a global T_b has been computed fitting the classical fixed effects ES, using data from all schools simultaneously. Notice that, even though numerous buildings signature lie close to the global one, the differences between individual buildings vary significantly, with the HLC of some of the worst performers doubling the HLC of the best ones.

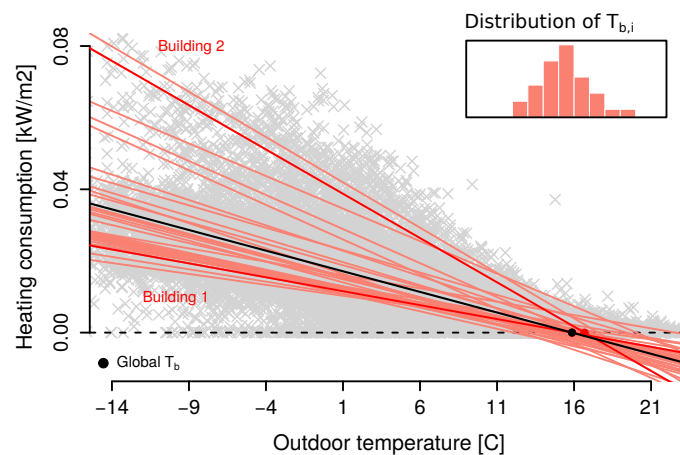


Figure 2: Energy signature for the total building population of regular schools. Two arbitrary buildings have been highlighted in red. It can be seen that the performance parameters change significantly from building to building.

4.2. Sampling energy profiles

The previous section showed that mixed models allow us to assess the differences across individual buildings that are present in the building population. This section, shows the results simulating the consumption of an unobserved school. This simulation is based on the model in Equation (6), that has been fit using data from 25 schools.

Figure 3 shows the model prediction given the temperatures of the month of February 2010; where, it can be noticed that the model simulation follows closely the trend of the test data. As expected, this trend shows clear peaks during the work days and a flatter trend during the weekends. The simulation under-predicts the highest peaks taking place in the morning of work days. When the morning peaks are captured, the valleys at night are over-predicted, which highlights the difficulty of capturing sudden changes in heat consumption.

Since the simulation is stochastic, the prediction interval (PI) of the simulation has been computed. As expected, this prediction interval shows a constant width during the whole time-series due to the assumption of normally distributed noise. It is easy to see that, in this February example, the prediction interval includes practically all test data points.

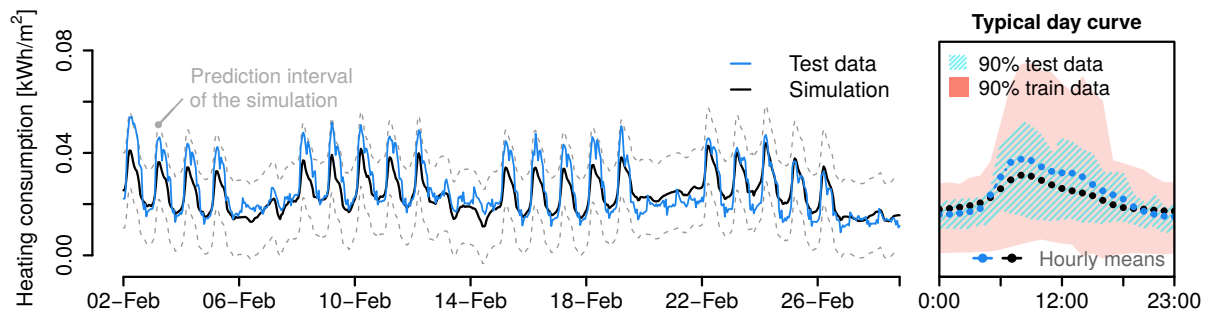


Figure 3: Comparison of the simulated data and a test set for the month of February 2010.

In addition, Figure 3 includes the daily profile of the prediction and the test data. It can be observed that the typical day curve of the simulation is lower than the testing data during the working hours. This damped trend in the simulation might be due to the range difference between the training and testing set. As it can be seen that the hourly range of data points of the training set is significantly wider than test set.

This model is fitted with data from all year round, so it has been possible to simulate heat consumption for every month. Figure 4 shows the typical daily consumption for every month of 2010. The simulated curve and the test curve follow a similar pattern during the colder months. Notice that the daily baseline consumption decreases during the summer, where only the hour effects are present. During June, July and August the test consumption is virtually zero, and the simulation still shows a low periodic hourly pattern. Nevertheless, Figure 4 also includes the percentage of test data points that fall inside the 90% prediction interval of the simulation. It can be seen that, for the whole year, practically all test data falls inside the expected region.

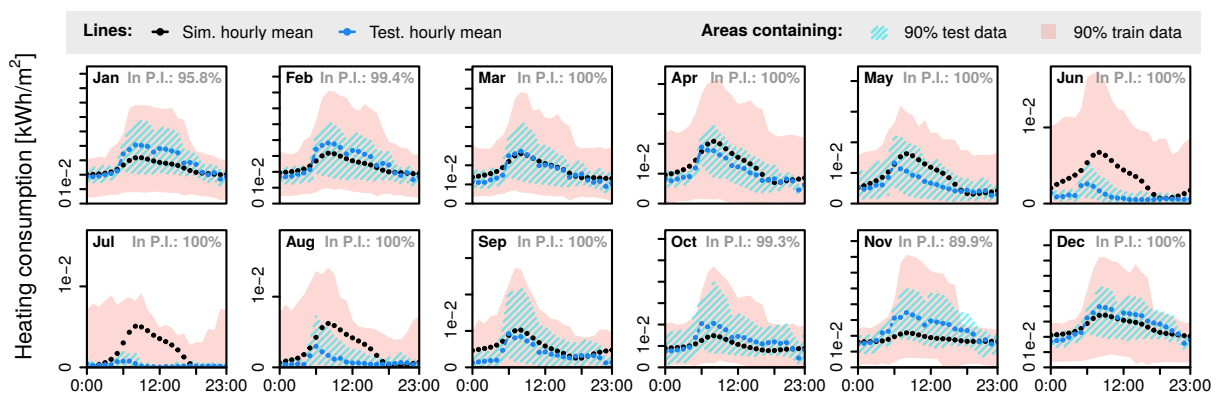


Figure 4: Typical days for every month of the year 2010. Every month includes the % of test data that falls inside the 90% prediction interval, denoted by "In P.I.".

5. Discussion

The results in this work show the potential of mixed effects models to be used to forecast long-term energy consumption of buildings. These models are a natural extension of fixed effects models, that have been proven successful in past work. Mixed effects are able, not only to generate a representative prediction of the heat consumption in buildings, but also they estimate the inherent uncertainty of the simulation due to non-measured events.

Fitting a mixed effects version of the energy signature, has showed that the range of energy performance varies significantly, even though all schools have the same efficiency label. This result highlights the importance of working with stochastic simulations, given the wide variety of energy performance in the building stock.

There is still room for improvement in the current version of the model. Although all test values fall inside the prediction interval, the simulation mean shows damped peaks, when compared to the test data. In addition, practically 100% of test data falls inside the 90% prediction interval; which hints that such interval should be narrower. Similarly, it can be seen that the interval is symmetric and constant. However, a more realistic model would have a prediction interval that: i) is asymmetric since consumption can only take positive values; and ii) is wider when consumption is expected to be higher. These limitations come from the assumption that the modelled data follows a gaussian distribution. In the future, different distribution families will be used, to take into account such issues. In addition, the dependence of the weather can be improved, to make the base temperature variable over the year, and then skip the need for fitting first the mixed effects energy signature. Despite the aforementioned issues, the results are promising enough and the next steps well defined to pursue further this methodology with a more complete model. Ultimately, given the generality of modelling with mixed effects, the work presented here can be extended to other building categories to simulate a broader energy landscape.

Acknowledgments

The authors would like to thank the Flexbuild project for providing the motivation, data and foundation to build this work on. In addition, funding from the Centre for IT-Intelligent Energy Systems in Cities (CITIES) and Flexible Energy Denmark (FED) have made possible this research.

References

- [1] Carvallo J P, Larsen P H, Sanstad A H and Goldman C A 2018 *Energy Policy* **119** 410–422 URL <https://ideas.repec.org/a/eee/enepo1/v119y2018icp410-422.html>
- [2] Chen T and Wang Y C 2012 *International Journal of Electrical Power Energy Systems* **43** 454–464 ISSN 0142-0615 URL <https://www.sciencedirect.com/science/article/pii/S0142061512002645>
- [3] Boßmann T, Elsland R, Lickert F and Wietschel M 2013 The german load curve in 2050: Structural changes through energy efficiency measures and their impacts on the electricity supply side
- [4] Lindberg K, Seljom P, Madsen H, Fischer D and Korpås M 2019 *Utilities Policy* **58** 102–119 ISSN 0957-1787 URL <https://www.sciencedirect.com/science/article/pii/S0957178719300116>
- [5] Pérez-Lombard L, Ortiz J and Pout C 2008 *Energy and Buildings* **40** 394–398 ISSN 0378-7788 URL <https://www.sciencedirect.com/science/article/pii/S0378778807001016>

- [6] Taylor P G, d'Ortigue O L, Francoeur M and Trudeau N 2010 *Energy Policy* **38** 6463–6474 ISSN 0301-4215 energy Efficiency Policies and Strategies with regular papers. URL <https://www.sciencedirect.com/science/article/pii/S0301421509003280>
- [7] Fouquier A, Robert S, Suard F, Stéphan L and Jay A 2013 *Renewable and Sustainable Energy Reviews* **23** 272–288 ISSN 1364-0321 URL <https://www.sciencedirect.com/science/article/pii/S1364032113001536>
- [8] Hammarsten S 1987 *Applied Energy* **26** 97–110 ISSN 0306-2619 URL <https://www.sciencedirect.com/science/article/pii/0306261987900122>
- [9] Rasmussen C, Bacher P, Calì D, Nielsen H and Madsen H 2020 *Energies* **13** ISSN 1996-1073
- [10] Wolf S, Møller J K, Bitsch M A, Krogstie J and Madsen H 2019 *Energy and Buildings* **183** 672–683 ISSN 0378-7788 URL <https://www.sciencedirect.com/science/article/pii/S0378778818320887>
- [11] Lindberg K, Bakker S and Sartori I 2019 *Utilities Policy* **58** 63–88 ISSN 0957-1787 URL <https://www.sciencedirect.com/science/article/pii/S0957178719300128>
- [12] T K, C S, MG S and M E 2012 *Clinical Pharmacokinetics* **51** 512–515
- [13] Madsen H and Thyregod P 2011 *Introduction to general and generalized linear models* Texts in statistical science (CRC Press) ISBN 9781420091557