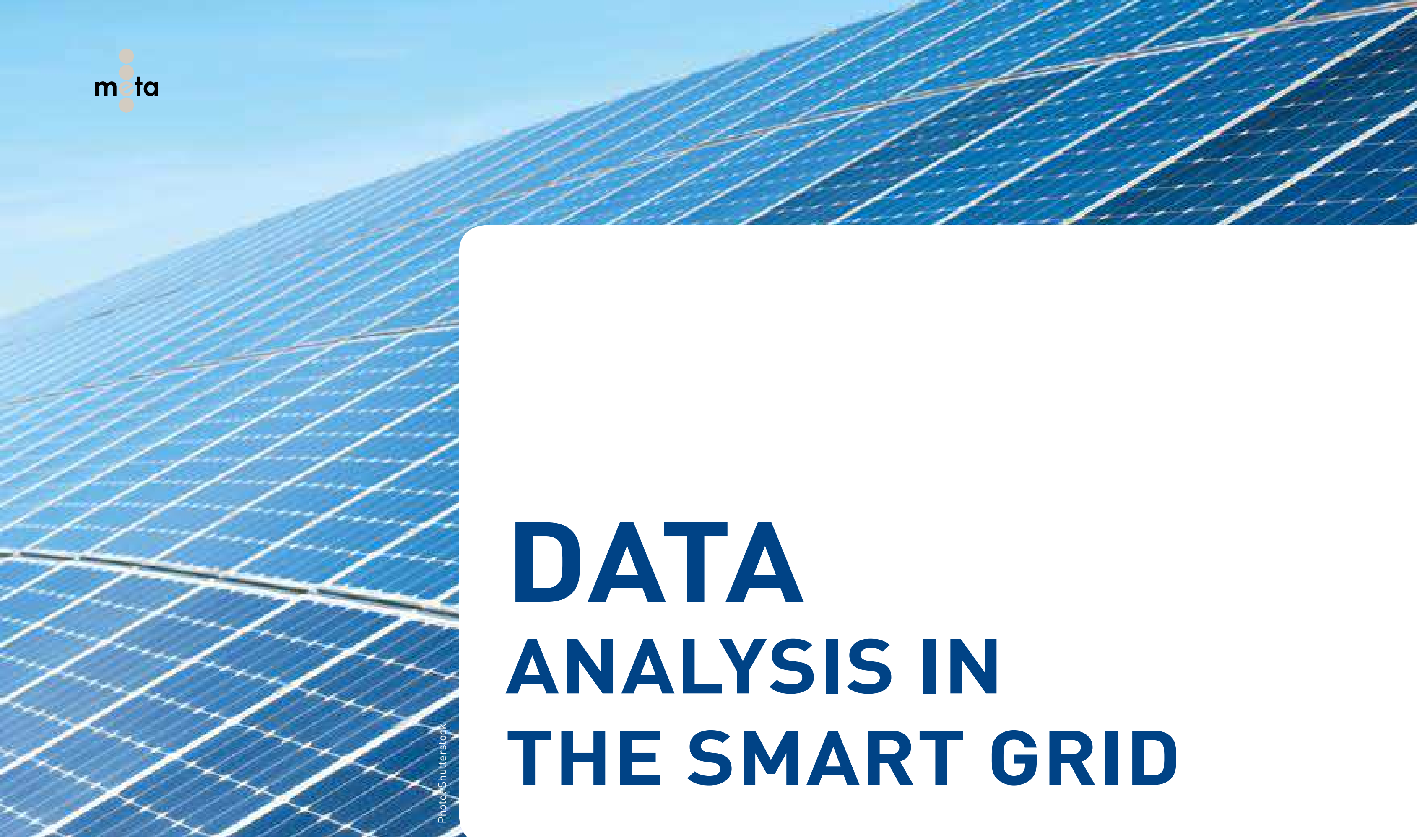

DATA
ANALYSIS IN THE SMART GRID

AUTHORS

Boye A. Høverstad
Post doctoral fellow
Dept. of Computer and
Information Science, NTNU

Axel Tidemann
Post doctoral fellow
Dept. of Computer and
Information Science, NTNU

ABSTRACT
The next decade will see a complete transformation of the way we use, produce and relate to electricity: roof-mounted solar panels, electric cars and smart devices that pre-heat (or cool) your house when power is cheap are just some examples of the expected changes. The electrical grid infrastructure must undergo dramatic changes to support this development, with the rollout of smart meters with two-way communication capabilities as one of the most visible first steps. As part of the Next Generation Control Centres for Smart Grids research project, we are looking for ways to take advantage of AMI data beyond simple metering and billing.

WHAT IS THE SMART GRID?
The technology of the electrical infrastructure, a crucial part of any developed society, has remained largely unchanged for the last 100 years or so. Although the networks have been scaled up in accordance with the increase in consumption of electrical power, the principles for controlling power flow and management have not seen the same development as for instance the computer and telecom industries. This is expected to change with the introduction of the smart grid; an evolution of today's electricity grid into a more dynamic system where real-time monitoring, control and communication between intelligent devices at all levels of electricity production, distribution and consumption enables optimal use of both existing infrastructure and energy from renewable sources.

Broadly speaking, the electrical grid is divided into the transmission and distribution levels. The transmission level carries high-voltage electricity over long distances, and is fairly well instrumented and subject to automated real-time control. The distribution level receives electricity from the transmission grid and delivers it to end users. Compared to the transmission level, this part of the grid is not monitored to the same extent, and the control abilities of the system operators are limited. This is the reason why you have to call the power company and tell them whenever something is wrong (e.g. you lose power) – they simply don't know. Furthermore, the distribution operators face several challenges in the years to come: there will be an increase in distributed generation (e.g. roof mounted solar panels, wind turbines, small-scale hydro-electric power plants), a change in consumption patterns (e.g. charging of electrical cars, tankless water heaters, heat pumps), in addition to an increase in general demand. On top of all this, modern electric and electronic equipment is more dependent on power quality than the incandescent bulbs of yore.

As building new infrastructure is very expensive, it is desirable to address these challenges by optimizing the existing grid capabilities. This is envisioned to happen by transforming the grid into a smart grid, where digital technology enables a more efficient use of existing infrastructure (Lightner and Widergren, 2010). As part of implementing the smart grid, smart meters are being deployed to end users. These record and transmit power consumption at regular intervals (typically every hour or 15 minutes), and can also receive information (e.g. price and control signals). The rollout of smart meters will become an important part of an advanced metering infrastructure (AMI), and it is essential that their capabilities and the potential of this new infrastructure is exploited to full extent.

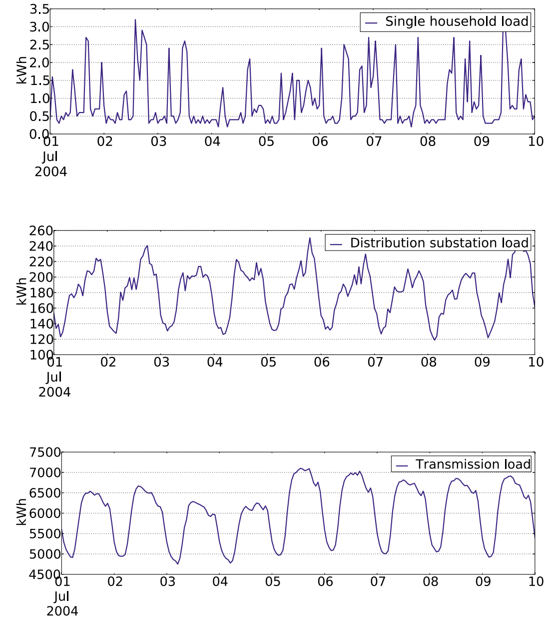

Figure 1. Sample time series of power load data at end user, distribution substation, and transmission level.

DISTRIBUTION SYSTEM OPERATION

The project Next Generation Control Centres for Smart Grids is a long-term effort to investigate how the distribution grid control centres can be modernized, and better suited to meet the challenges discussed above. More specifically, our group is interested in how applications that make use of sensor data can contribute to efficient grid operation.

Efficient and robust operation of the power grid mandates situation awareness. This implies awareness of both the current state of the system, and the expected situation in the imminent future. Operators in the distribution control room monitor the grid in order to avoid voltage band violations, ensure safe operation during planned outages (typically due to maintenance or construction work), and efficient handling and restoration of power in the event of unexpected disruptions. Moreover, it is desirable to avoid prolonged overloading of equipment, as this will lead to increased failure rates.

With the planned increase in instrumentation, more data flows into the control centre. However, the nature of the distribution grid prevents it from being instrumented to the same extent as the transmission grid; its topology is vastly more complex, and it is not economically feasible to install sensors and remote control units at each substation. The installation of smart meters will not provide the same data fidelity and control possibilities. But the AMI data will clearly enable a more automated surveillance of the distribution grid, and our goal is to make use of this data to fill some of the gap that is due to missing instrumentation. As part of this work, we have used NOTUR resources to develop a platform that performs load forecasting and data cleansing at the level of distribution substations, based on AMI data. Currently, we are investigating how this can be extended to the end users. This will enable state estimation and situation awareness at the distribution level, without requiring the same level of instrumentation as on the transmission grid.

LOAD FORECASTING AND DATA CLEANSING

Load forecasting is performed on several time-scales, depending on their intended use. Long- and medium-term forecasting (10 to 50 years ahead) is required for capacity planning and maintenance scheduling. The importance of short-term forecasting (up to a week ahead) increased after deregulation of the power market, allowing competition between the buying and selling parties. Short-term load forecasting is important for power producers and vendors, as well as for planned outages to ensure grid stability, and gains even more importance as the grid is transformed to a smart grid. Short-term load forecasting is routinely performed on the transmission level, but is to a lesser extent employed at the distribution level. This is particularly relevant in Norway, where we have numerous small distribution system operators (compared to the rest of Europe), coupled with a rapidly increasing penetration of both small-scale hydro-electric power and electric vehicles (EVs).

In our recent work, we have focused on 24-hour load forecasting, with cleansing as an integral part of this process. In order to compare different forecasting methods, we have developed a framework for parameter optimization. When using any given method, you normally have to set some parameters (also dubbed "magic numbers"). Setting these parameters often depend on experience and intuition; when faced with different methods and datasets this becomes a hurdle for unbiased comparison. We have a hands-off approach: our system performs an explorative search in the parameter space of each model. This search is performed using an evolutionary algorithm.

An evolutionary algorithm (EA) works by trying out many different solutions to a problem. In the initial population, all the different solutions are generated randomly. The solutions that have the lowest prediction error (which in the EA terminology are known as the individuals that are most fit) are passed on to the next generation, where new offspring are created from the best individuals by genetic crossover (combining parents' genes) and mutation (randomly altering a gene). In problems where the parameter space is too large to be searched exhaustively, and the structure of the problem is too complex to be solved in polynomial time, EAs have been shown to find good approximative solutions (Eiben and Smith, 2003).

We have used NOTUR resources to run extensive simulations, mainly on the vilje supercomputer. The EA search is computationally expensive but easily parallelized, since each individual in the population can be evaluated independently. This was done using MPI. Depending on the complexity of the forecasting algorithm and the size of the population in the EA, each fitness evaluation was parallelized over a number of cores on the same node using OpenMP.

The choice of dataset is an important factor for the parameterization of a model; if you change the dataset, the model parameters that yield optimal performance are likely to change as well. In order to investigate how forecasting methods originally applied at the transmission level would scale down to the distribution grid, we optimized load forecasters on different levels of data granularity: end user, distribution substation level (~150 end users) and transmission. The signal characteristics for each of these levels vary greatly (Figure 1).

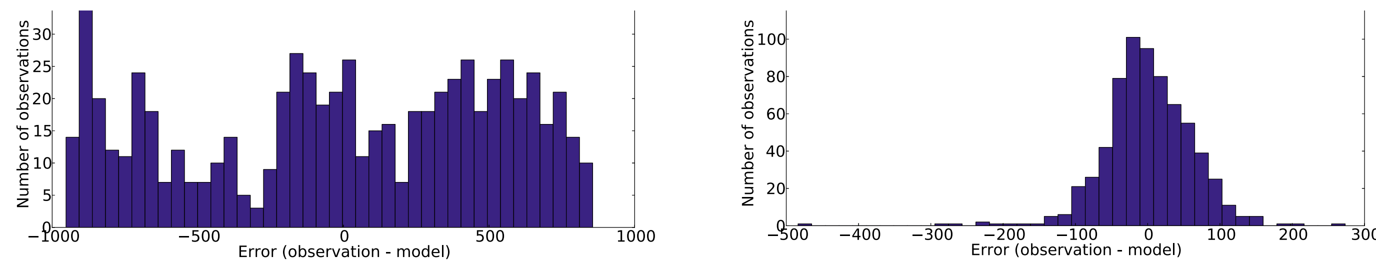

Figure 3. Histogram of distances from observed data to expected "true data", when cleansing either raw data (left) or preprocessed data where cycles are subtracted (right). The distribution takes on a Gaussian shape after preprocessing.

Four different forecasting models were evaluated: an autoregressive (AR) model, an echo-state network (ESN), a wavelet-based predictor, and finally a case-based reasoning approach (CBR). These are all data-driven models that take advantage of covariates such as weather data, in addition to the load time series. Moreover, they must all be described with a set of parameters that need to be found by the GA.

The AR model is a linear predictor in which each prediction is a weighted sum of previous observations. An ESN is a recurrent neural network that is characterized by a large number of randomly connected nodes in the hidden layer, where only the links to the output layer are trained (Jaeger and Haas, 2004). The wavelet transform is a multiresolution technique that allows a prediction model to represent structure in the load data at different time scales, e.g. looking for daily, weekly and seasonal patterns. We employ the redundant Haar transform, which yields continuous signals at each scale of the transform. By using a selection of the values found in each scale, linear regression can be used to make a predictor based on the wavelet transform (Renaud et al., 2005). The same approach has been employed for load forecasting, but predictions were only made one hour into the future, instead of 24 hours (Benaouda et al., 2006). CBR is a method for reusing past experiences (cases) to solve new problems (Aamodt and Plaza, 1994). CBR does not have an explicit model, and it is therefore well suited in domains that are difficult to model mathematically. Cases are stored in a database, indexed by features. When the system is used to solve a new problem, the system tries to find a similar case by matching features. The most similar case is then reused. For the current domain, the CBR system searches for similar load profiles in order to predict the next 24 hours.

In order to investigate and compare the performance of the different prediction models, 30 evolutionary runs with different random seeds were performed for each dataset-model combination, and the best model was chosen by evaluation on a test set. As expected, predictability changes in proportion to the number of meters aggregated. The transmission level time series is fairly regular, and the predictions are correspondingly accurate. The time series at the distribution substation level is less regular than at the transmission level, but all the models still produce fairly accurate predictions. However, at the single-meter level, there are large variations between data sets. Some are moderately regular and predictable, whereas others are very irregular and difficult to predict. As the prediction error increases, differences in how the models fail become apparent. In particular, we observe that the CBR model degrades less gracefully than the other models, and that the autoregressive model converges toward the daily mean.

After observing that the load at a distribution substation can be forecasted fairly accurately, while end user loads are much harder, we wanted to investigate the relationship between number of aggregated meters and prediction error. We thus created different load time series by summing the consumption from 1 to 150 randomly selected meters and evolved predictors for each of these. The result is shown in Figure 2. When the time series consists of a sum of less than about 20 meters, prediction error and deviation rise sharply. However, fairly accurate predictions can be made with as few as 30–40 meters.

It is evident from Figure 1 that a power load time series follows a very characteristic cycle: in the morning and afternoon the consumption is high. This is known as the peak hours. On the other hand, comparatively little power is used at nighttime. The à priori knowledge of these cycles is something we wanted to exploit. As a first step, we extended an existing approach to load cleansing (Chen et al., 2010) by first subtracting the cycles, then evolving the parameters for cleansing, before the cycles were added back and tomorrow's load was forecasted. The effect of subtracting the cycles is that the remaining noise on the time series tends to have a normal distribution, and thus in line with the theoretical foundation of the cleansing approach (Figure 3). Figure 4 exemplifies the difference between the two approaches. The original algorithm has a higher tendency to cut peaks and fill troughs, which is undesirable for the subsequent use in forecasting.

A natural extension to this work is to perform the forecasting itself after subtracting the cycles. Again, we used the evolutionary framework to compare this approach to forecasting algorithms that have been designed specifically to deal with the double seasonality (i.e. daily and weekly cycles) present in load time series (manuscript in preparation). The results show a great improvement in forecasting accuracy when applying this simple transform. Example forecasts are shown in Figure 5.

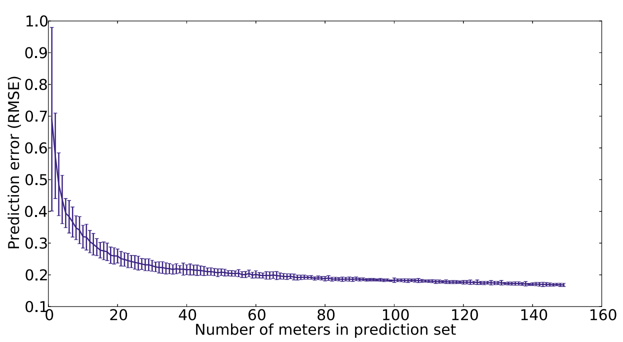

Figure 2. Forecast error (mean and standard deviation) as a function of the number of meters accumulated in the forecasted time series.

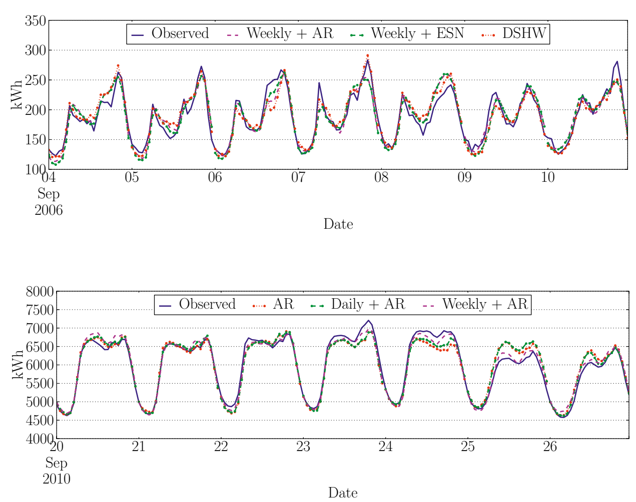

Figure 5. Sample forecasts made by some of the forecasting models under study. Distribution substation data on top, transmission network data on bottom.

AMI DATA ANALYSIS AND STATE ESTIMATION

As mentioned earlier, with more instrumentation of the distribution grid, it becomes possible to get a better grasp of the grid status. In particular, knowledge of the instantaneous voltage and current at each endpoint of the grid would enable the grid operators to perform distribution system state estimation. This is in principle the same state estimation that is routinely performed on the transmission level, however the data required to solve the equations are lacking. More specifically, using the AMI data directly is insufficient because the frequency of measurements is too low for real-time use. This problem is exacerbated by the fact that modern loads such as EVs and tankless water heaters draw high currents from the grid over relatively short time intervals, thus making the system more complex.

In short, there is a lack of observability both of the present and the expected future at the end user level. As shown in the previous section, our results indicate that traditional load forecasting algorithms are relatively successful at forecasting the load for the coming 24 hours down to the level of distribution substations, but they are practically unable to predict the highly irregular load time series that are observed at the level of the end user. Our current research deals with the possibility of refining this picture somewhat, by combining predictions at the level of the substation with smart meter readings at each endpoint and a single high-frequency reading at the substation. By using a variation of single channel source separation (Yilmaz et al., 2004), we will attempt to identify the different loads making up the load at each endpoint. This may make it easier to project substation forecasts down to each end user, as well as enabling long-term forecasts under different scenarios such as significantly higher EV penetration, or increase in distributed generation.

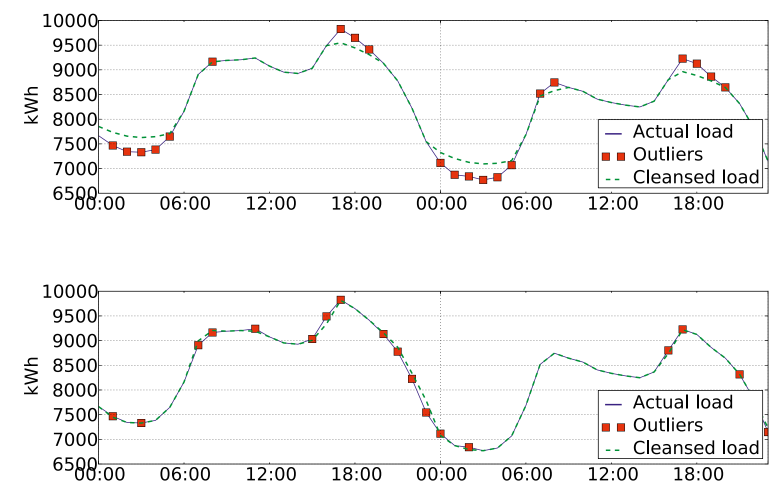

Figure 4. Examples of cleansed load data, using the original algorithm (top) and after subtracting cycles (bottom). The original algorithm has a higher tendency to cut peaks and fill troughs.

Literature

E.M. Lightner and S.E. Widergren, "An Orderly Transition to a Transformed Electricity System," IEEE Transactions on Smart Grid, 1(1):3–10, 2010.

A.E. Eiben and J.E. Smith, Introduction to evolutionary computing, Berlin: Springer Verlag, 2003.

H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication", Science, 304(5667):78–80, 2004.

O. Renaud, J.L. Starck, and F. Murtagh. "Wavelet-based combined signal filtering and prediction". IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 35(6):1241–1251, 2005.

D. Benaouda, F. Murtagh, J.L. Starck, and O. Renaud, "Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting", Neurocomputing, 70(1):139–154, 2006.

A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches", AI communications, 7(1):39–59, 1994.

J. Chen, W. Li, A. Lau, J. Cao, and K. Wang. "Automated load curve data cleansing in power systems", IEEE Transactions on Smart Grid, 1(2):213–221, 2010.

B. A. Høverstad, A. Tidemann and H. Langseth, "Effects of data cleansing on load prediction algorithms," IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG), 2013.

A. Tidemann, B. A. Høverstad, H. Langseth and P. Öztürk, "Effects of scale on load prediction algorithms," 22nd International Conference on Electricity Distribution (CIRED), 2013.