EFFECTS OF SCALE ON LOAD PREDICTION ALGORITHMS

Axel TIDEMANN
Norwegian University of Science and Technology – Norway
tidemann@idi.ntnu.no

Boye A. HØVERSTAD
Norwegian University of Science and Technology – Norway
hoversta @idi.ntnu.no

Helge LANGSETH
Norwegian University of Science and Technology – Norway
helgel @idi.ntnu.no

Pinar ÖZTÜRK
Norwegian University of Science and Technology – Norway
pinar @idi.ntnu.no

ABSTRACT

Load prediction is an important tool for grid utilities and power companies for managing the power system. This has traditionally mainly been applied at transmission and sub-transmission system levels. However, as the traditional grid evolves into a smart grid, load prediction at smaller scales becomes necessary for efficient management and operation. In this paper we investigate how the qualitative and statistical properties of load time series change as a function of the number of individual loads aggregated in the series, and how these properties influence the predictability of the time series. We study the performance of a traditional autoregressive model, a wavelet-based model, an Echo State Network, and a variation of Case-Based Reasoning (CBR) at the subtransmission (~10000 customers), distribution substation (~150 customers) and single-meter level. For all the four prediction methods, we employ an evolutionary algorithm as a meta-learner to automatically optimize the free parameters for each model-dataset combination. We find that relatively accurate predictions can be made at finer granularity, but care must be taken in choosing, tuning and analyzing the prediction model as the regularity of the consumption patterns decreases.

INTRODUCTION

Load prediction is an important tool for several bodies at the regional and national level. Firstly, authorities and power utilities rely on long and medium term prediction for capacity and maintenance planning and infrastructure development. Secondly, utilities, both grid operators and power generators, use shorter-term load prediction (in Norway, typically down to 24-hour and 1-hour) for production planning and market operations. A variety of short-term load prediction strategies designed for the transmission network level are presented in the literature, see e.g. [1, 2] for reviews.

Already, long- and medium-term load prediction also plays an important role for planning and development in the distribution grid. Additionally, the evolution of the smart distribution grid, combined with modern society's increasing dependence on a reliable supply of high-quality power, significantly increases the need for optimal real-time operation and control also at this level with a view towards, e.g., integration of energy from intermittent sources, substation transformer operation, anomaly detection, and contingency planning.

Finally, one may envision the use for load prediction at the level of a single meter (households, commercial buildings, industry). Electrical devices in modern residential and commercial buildings are increasingly equipped with interfaces for communication and control, and load prediction at each endpoint can improve the utilization and efficiency of these devices. This includes charge scheduling for electrical vehicles, thermal load scheduling, optimal utilization of rooftop solar panels, and improved accuracy in estimating cost of energy not supplied (CENS).

There is thus a growing need for load prediction at several levels of the distribution network. Given the variety of approaches and models available for prediction at the transmission level, we are interested in if and how these models can be adapted to load time series as we get ever closer to the household meter. Moving from transmission level through the distribution grid to the final customer implies at least three important changes: the number of prediction points increases drastically, the resources available decrease correspondingly, and the data becomes significantly less regular and predictable. As a consequence, models that work well at one level of granularity may not work well at another, and vice versa. In this paper we investigate how the load time series change as a function of the number of individual loads aggregated in the series, and how these properties influence the short-term (24-hour) predictability of the data streams. In particular, we study load time series at the subtransmission (~10000 customers), distribution substation (~150 customers) and single-meter level, using four prediction algorithms with different characteristics.

METHODS

The research focus is to study how different predictive models perform at various levels of data granularity. We have used data from two different sources: SINTEF Energy Research [3] and transmission level data from British Columbia (BC Hydro). The SINTEF data was recorded from 2004 through 2006, the BC Hydro dataset covers the period 2004 through 2010. More details on this data set can be found in [4]. The SINTEF data allows us to investigate how models perform on the single-meter level, as well as distribution substation (~150 meters).

A key aspect in this regard is how to parameterize the different models in an optimal way. Instead of manually searching for the optimal parameters, we employ a *genetic algorithm* (GA) to find these parameters [5]. A genetic algorithm tries out different solutions to a problem at random, passing on the most fit (i.e. the individuals that performed best at a certain task) to the next generation. The genome of a "child" in the next generation is created by combining parents' genes (crossover) and by random mutation.

In the current application, the individuals are the different predictive models, and the genome specifies the parameters of the model. Each individual is presented with historical load data (the length of which is also specified as part of the genome), which it uses to train the model. Subsequently, a 24-hour period is predicted, and how well the prediction matches the actual load data is how the fitness of each individual is evaluated.

Four different predictive models are evaluated. These are all primarily data-driven models. They all allow for covariates such as weather data, in addition to the load time series. Additional domain knowledge may also be included in various ways, but that is not a main focus in this paper. The models are briefly presented below, along with the parameters that need to be found by the GA.

Autoregressive model (AR)

The AR model is a linear predictor in which each prediction is a weighted sum of previous observations. How many previous values contribute to the output is determined by the *order* (the parameter searched by the GA), and the contribution of the previous values is typically calculated by the least squares method.

Echo State Network (ESN)

An ESN is a recurrent neural network that is characterized by a randomly connected hidden layer where only the output layer weights are trained [6]. The hidden layer is typically larger than a standard recurrent neural network where the backpropagation algorithm is used, and it is therefore referred to as a *reservoir* model. Since only the output layer nodes are trained, the weights can be found by linear regression. The training time is therefore greatly reduced compared to the backpropagation algorithm. There are three inputs to the ESN: the current temperature, the observed load 24 hours ago, and the load in the previous hour. When predicting, the ESN is switched to signal generation mode by feeding the prediction back into the network. To define an ESN, five parameters need to be found by the GA: network size, leakage rate, input scaling, bias scaling and spectral radius (which defines the Echo State property).

Wavelet prediction

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 [7]. The same approach was taken to load prediction, but predictions were only made one hour into the future, instead of 24 hours [8]. The parameter to find by the GA is the number of scales the time series is decomposed to.

Case-Based Reasoning (CBR)

CBR is a method for reusing experiences (cases) to solve new problems [9]. 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, described 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. This is in essence the problem of retrieving and storing time series. Time series are well suited to be stored in R-Trees, however a compact representation of the time series is necessary to build an efficient R-Tree. To achieve this, the Haar wavelet transform is applied. The wavelet coefficients then serve as the indices to the R-Tree (the features in CBR terms). How many coefficients to use is found by the GA.

EXPERIMENTS AND RESULTS

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. Model parameters were evolved on a training set. The resulting 30 parameter sets were then ranked by mean prediction error on a 6-month validation set. Finally, the prediction error on another 6-month test set was recorded. We used root mean square error (RMSE) as error measure for 24-hour prediction. The final prediction error is then represented by the mean RMSE over each period. We chose RMSE over mean absolute percent error (MAPE) since MAPE is not a suitable error measure when working with single-meter predictions, since it is undefined at zero.

Table 1 summarizes the mean prediction error on the test set for the 15 best evolved parameter sets in each experiment (recall that the validation period is used to pick the best 15 out of the 30 evolved parameter sets). In Figure 1, these prediction errors were divided by the highest value in each column, in order to facilitate inter-model comparisons between datasets. We find that the CBR method outperforms the other models at the transmission level, whereas the situation is reversed at the distribution and single-meter levels. Note further that the ESN model has considerably higher standard deviation than the other

models at the distribution and transmission levels.

Example predictions at increasingly finer granularity are shown in Figures 2 through 5. 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 (Fig. 2). 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 (Fig. 3). However, at the single-meter level, there are large variations between data sets. Some are moderately regular (Fig. 4), whereas others are very noisy and difficult to predict (Fig. 5). As the prediction error increases, the differences in "failure mode" between models 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.

Finally, we investigated the relationship between number of aggregated meters and prediction error. For each $n \in [1, 150]$, we created 10 different load time series by summing the consumption from n randomly selected meters and evolved AR predictors for each of these. The result is shown in Figure 6. 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.

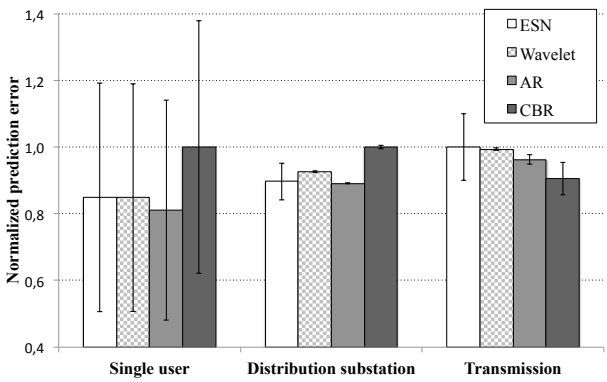

Figure 1. Histogram of prediction errors for different models at different levels of granularity. The error is normalized as described in the text.

Table 1. Mean test set prediction error (RMSE) of the 15 best predictors evolved for each model, as ranked by evaluation on the validation set.

	Single meter	Distribution substation	Transmission
ESN	0.527	21.56	319
Wavelet	0.527	22.25	317
AR	0.504	21.43	307
CBR	0.621	24.04	289

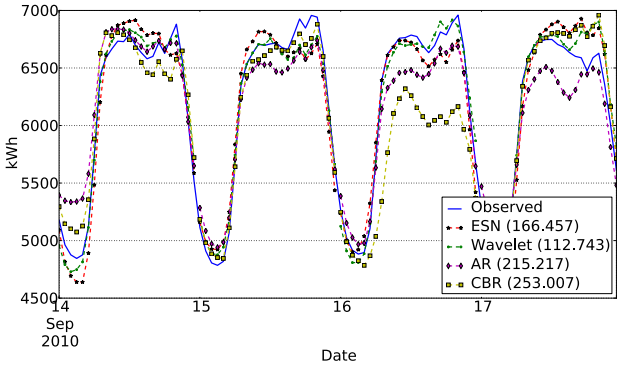

Figure 2. Four consecutive 24-hour predictions at the transmission level. RMSE in parenthesis.

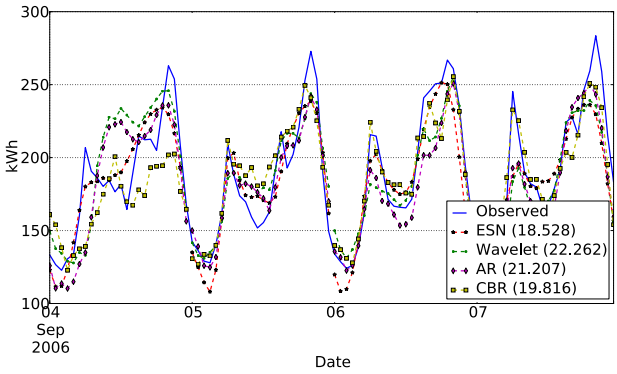

Figure 3. As Fig. 2, at the distribution substation level.

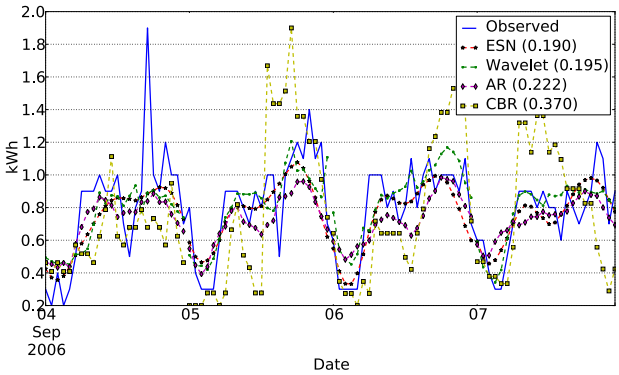

Figure 4. As Fig. 2, at the single-meter level.

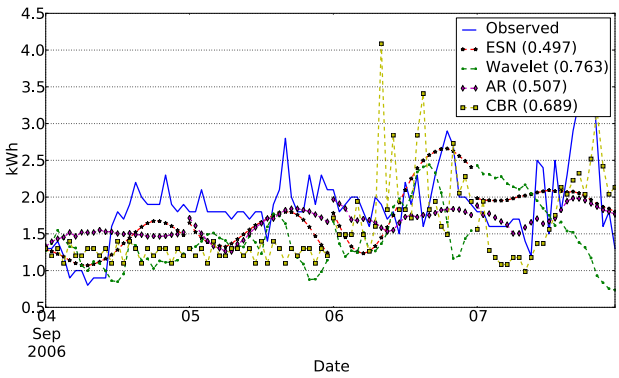

Figure 5. As Fig. 4, but on a less regular load time series.

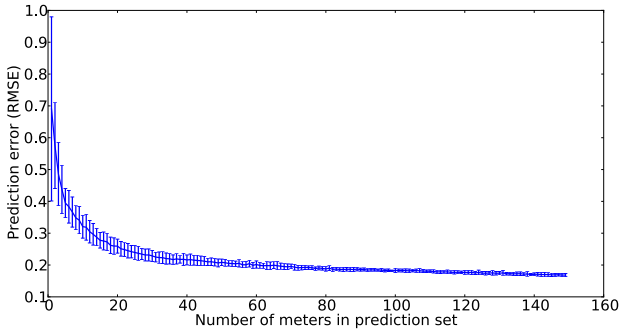

Figure 6. Prediction error (mean and standard deviation) as a function of the number of meters accumulated in the predicted time series.

DISCUSSION

When investigating the performance of the various models, we see that the CBR model has very low prediction errors on the transmission level, but higher errors than the other models on the less regular data sets at the distribution level. The reason for this can be seen in Figures 4 and 5: the CBR model makes predictions at the full extent of peaks and troughs, whereas the other models degrade towards a regression to the daily mean. The evaluation is dependent on the error measure in use. If the goal was to accurately predict the magnitude of the daily peak load, rather than to minimize the overall distance between predicted and observed values, a predictor that flattens its estimate as the noise level increases would be of little value, and a CBR model might be preferable. From an economical perspective, this might be of greater interest for power companies. If a power company underestimates the energy consumption of a user, it must supply the customer with power bought on the more expensive intra-day market.

We also note that the ESN model has considerably higher standard deviation than the other models at the distribution substation and transmission levels (Fig. 1). This is partly due to the large number of parameters involved in tuning this model, which results in a large search space for the genetic algorithm, and in part due to the nature of the model, with feedback signals that under unfortunate circumstances tend to dominate the model dynamics in a detrimental way. In this regard, it seems that the ESN is prone to overtraining, and care must be taken especially at the transmission level, where the regularity of the time series allows for too much specialization to the training data.

Finally, we wanted to examine the relationship between complexity of the load data and prediction performance. Figures 4 and 5 illustrate the stochastic nature of single meter time series, which is in stark contrast to the smoothness and regularity at the transmission level (Fig. 2). To investigate this issue, Figure 6 shows that the AR model is able to make relatively accurate predictions of time series composed of as little as 30-40 energy consumption meters. The figure also reveals how much more difficult it is to predict load power consumption as the number of aggregated users decrease.

Our experiments reveal how the performance of the models changes as the granularity of the data changes. For instance, the CBR model has the best performance at transmission level and the worst performance on the single meter level. In other words, models traditionally employed for load prediction at the transmission level might not be equally suitable for distribution level load prediction.

An important part of prediction at finer levels of granularity, is the tuning of the model in accordance with the characteristics of the data. The use of a genetic algorithm automates this process.

REFERENCES

[1] D.W. Bunn and E.D. Farmer, 1985, ”Review of short-term forecasting methods in the electric power industry”, *Comparative Models for Electrical Load Forecasting*, 13–30.

[2] H. K. Alfares and M. Nazeeruddin, 2002, ”Electric load forecasting: literature survey and classification of methods”, *Intl. Journal of Systems Science*, 33(1):23–34

[3] A. Z. Morch, I. Graabak, and N. Feilberg, 2009, “Results of Monitoring of AMR Systems in Norway: Analysis of Metered ata and Definition of the Performance Parameters”, *20th Intl. Conf. on Electricity Distribution*.

[4] B. Høverstad, A. Tidemann, H. Langseth, 2013, “Effects of Data Cleansing on Load Prediction Algorithms”, *Proc. IEEE Symposium Series on Computational Intelligence (to appear)*.

[5] N. Amjady, F. Keynia, and H. Zareipour, 2010, “Short-term load forecast of microgrids by a new bilevel prediction strategy”, *IEEE Trans. on Smart Grid*, 1(3):286–294.

[6] H. Jaeger and H. Haas, 2004, “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication”, *Science*, 304(5667):78–80.

[7] O. Renaud, J.L. Starck, and F. Murtagh, 2005, “Wavelet-based combined signal filtering and prediction”. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(6):1241–1251.

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

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