Adaptive Forecasting by State-Space Models

Joseph de Vilmarest, PhD Student
PhD advisors: Olivier Wintenberger, LPSM, Sorbonne Université,
Yannig Goude and Thi Thu Huong Hoang, EDF.

I. Pre-processing
I.1. Data preparation
I.2. Meteorological forecasts
I.3. Baselines

II. Statistical methods
II.1. Linear State-Space Model
II.2. Generalized Additive Model

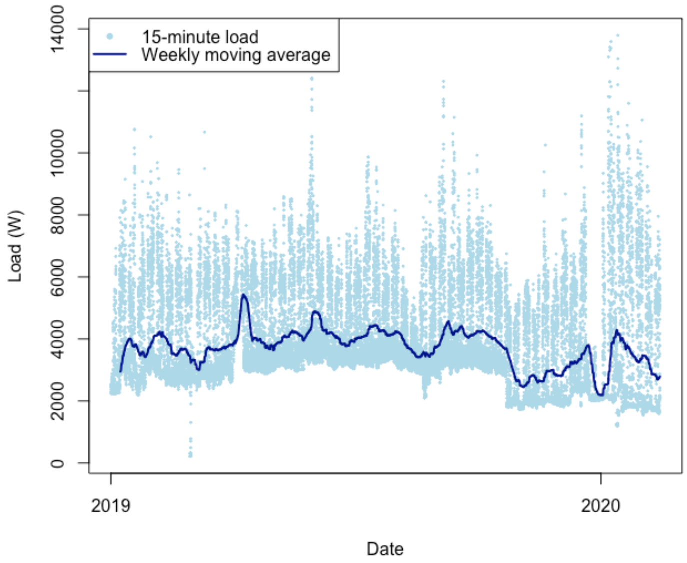
III. Forecasts

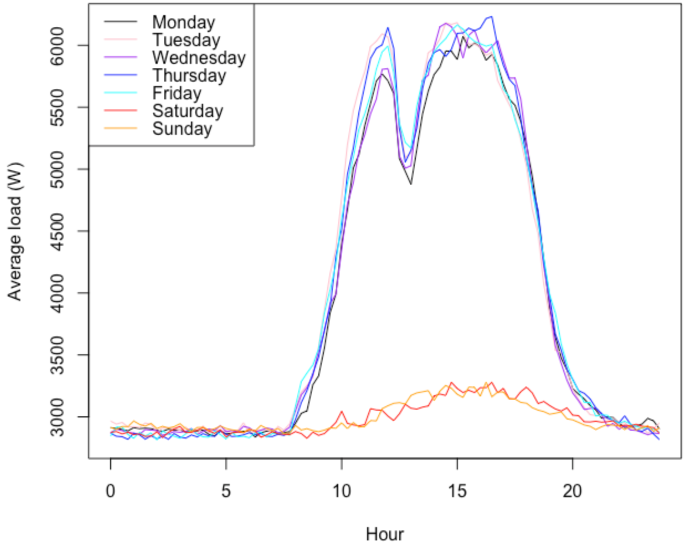
IV. Perspectives
IV.1. Other datasets
IV.2. VIKING: Variational Bayesian Variance Tracking
IV.3. Metric optimization

I. Pre-processing

I.1. Data preparation

Granularity. We consider aggregated load with 15-minute intervals.

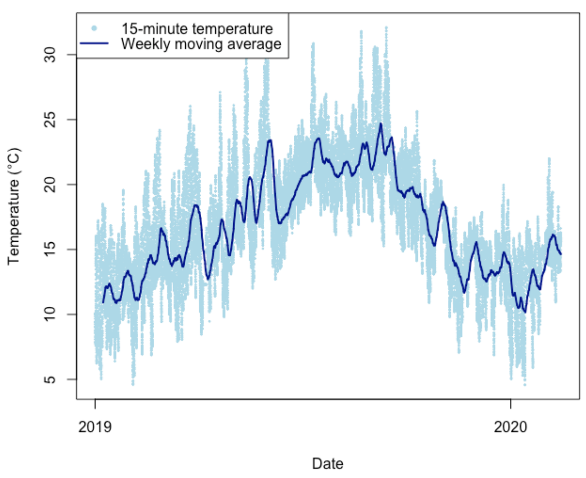


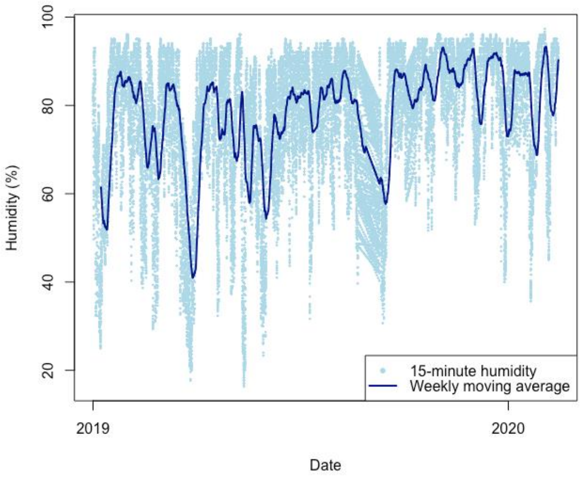


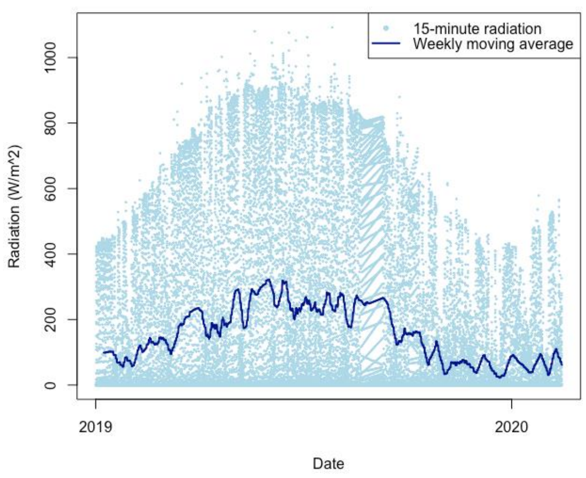
I. Pre-processing

I.1. Data preparation

Exogenous variables. We use the humidity and radiation from *weather_data* and the temperature from *building_sensor*.
Missing values. We use linear interpolation per time of day.







I. Pre-processing
I.2. Meteorological forecasts

Meteorological forecasts. We use autoregressive models with different parameters at each time of day:

$$z_t = \sum_l \alpha^l_{h(t)} z_{t-l} + \varepsilon_t, \qquad 0 \leq h(t) < 96.$$

Temperature and humidity: last 10 days at the same time of day, along with the last 10 available values.
Radiation: last 2 days at the same time of day.

I. Pre-processing
I.3. Baselines

We validate during 01-01-2020 / 02-09-2020.

	MAE (W)	RMSE (W)	Metric (W)
Last load	1509	2739	3703
Lag 1 day	1179	2074	2813
Lag 1 week	968	1748	2252
Average of lags	**929**	**1563**	**2179**
Week profile	1330	1635	2616

II. Statistical methods

II.1. Linear State-Space Model

We first consider a linear model. We optimize the coefficients separately for each of the 96 different times of day.

$$Load_t = \alpha_1 Temp_t + \alpha_2 Temps99_t + \alpha_3 Humidity_t + \alpha_4 Radiation_t$$

$$+\alpha_5 LoadLast_t + \sum_{i=1}^{7} \beta_i \mathbb{1}_{DayType_t=i} LoadDay_t$$

$$+ \sum_{i=1}^{7} \gamma_i \mathbb{1}_{DayType_t=i} LoadWeek_t + \sum_{i=1}^{7} \delta_i \mathbb{1}_{DayType_t=i} + \varepsilon_t$$

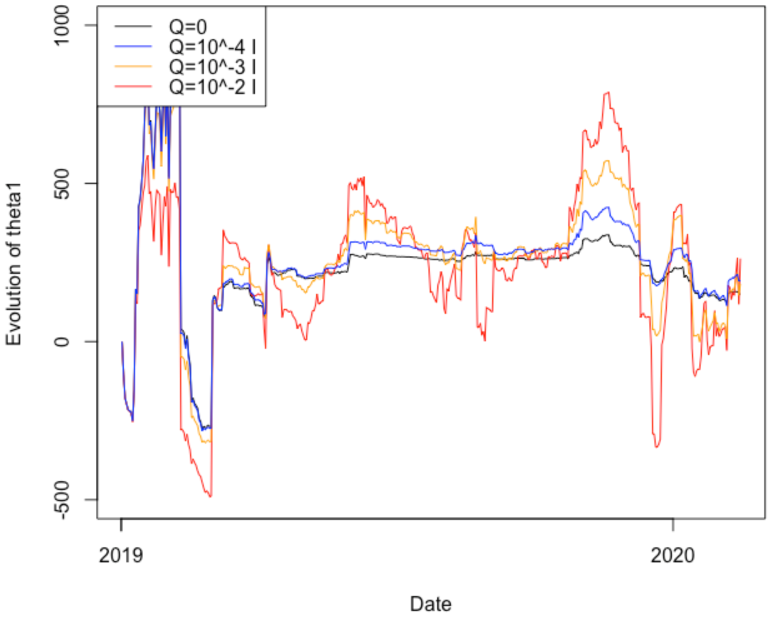
II. Statistical methods
II.1. Linear State-Space Model

We concatenate the covariates of the linear model into a vector x_t and we define the following state-space model:

$$Load_t = \theta_t^\top x_t + \varepsilon_t$$
$$\theta_{t+1} = \theta_t + \eta_t$$

ε_t, η_t are gaussian noises of variances σ^2, Q.

Then the celebrated Kalman Filter yields recursive estimation of θ_t.



II. Statistical methods
II.2. Generalized Additive Model

Generalized Additive Models are semi-parametric models that have been applied to predict the electricity load at a more aggregated level.

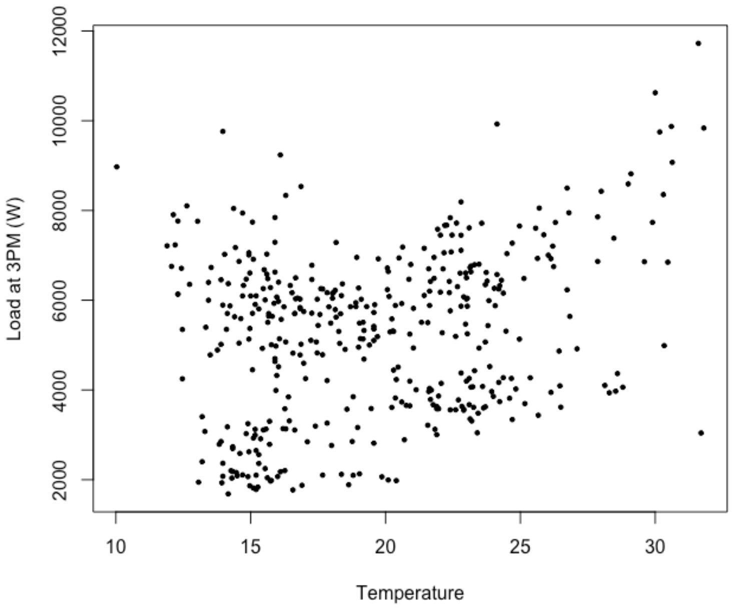
$$Load_t = s(Temp_t) + \alpha LoadLast_t + \sum_{i=1}^{7} \beta_i \mathbb{1}_{DayType_t=i} LoadDay_t + \sum_{i=1}^{7} \gamma_i \mathbb{1}_{DayType_t=i} LoadWeek_t + \sum_{i=1}^{7} \delta_i \mathbb{1}_{DayType_t=i} + \varepsilon_t$$

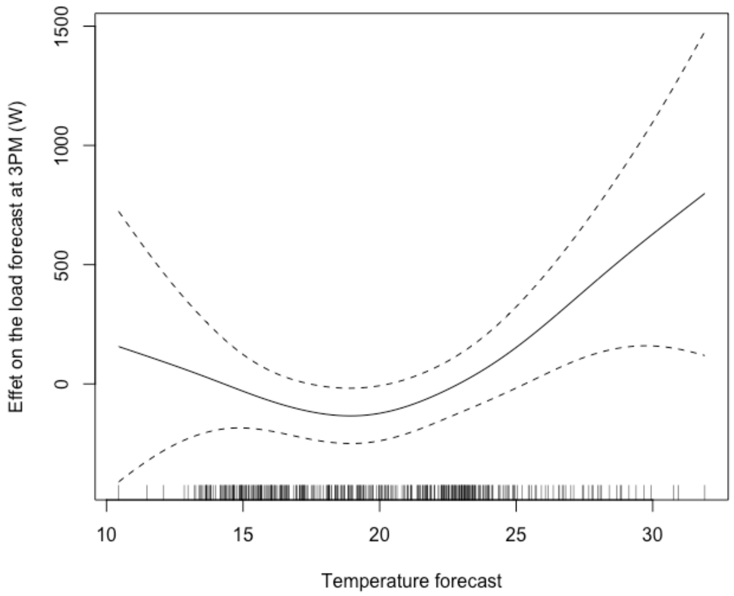
The effect of the temperature is decomposed on a spline basis and the optimisation is realized by penalized least-squares. See the R package *mgcv*.

IEEE PES Power & Energy Society®

IEEE

II. Statistical methods
II.2. Generalized Additive Model





II. Statistical methods
II.2. Generalized Additive Model

We obtain an adaptive model thanks to the state-space formulation.
We freeze the GAM effects, transforming the explanatory variables x_t into a feature vector $f(x_t)$.

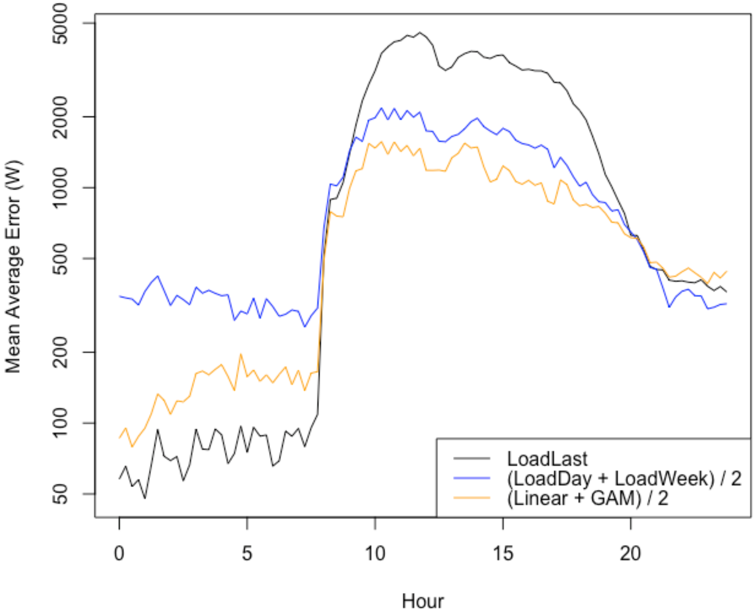
$$Load_t = \theta_t^\top f(x_t) + \varepsilon_t$$
$$\theta_{t+1} = \theta_t + \eta_t$$

Similarly to the linear model we take $Q = 10^{-4} I$ (the covariance matrix of η_t).

III. Forecast

	MAE (W)	RMSE (W)	Metric (W)
Offline linear	721	1292	1761
Kalman linear ($Q = 10^{-4}$)	693	1263	1681
Offline GAM	710	1246	1733
Kalman GAM ($Q = 10^{-4}$)	686	1206	1674
Average of both Kalman	681	1222	1656

III. Forecast



Final forecasts:

- **12AM to 7:45AM**: last load available.
- **8AM to 8PM**: average of adaptive linear and GAM.
- **8:15PM to 11:45PM**: average of the daily and weekly lags.

This decreases the validation metric by 2%.

IV. Perspectives
IV.1. Other datasets

This forecasting procedure is in the vein of previous works to forecast the load at larger scales (country, city):

- Obst, D., de Vilmarest, J. and Goude, Y. (2021): Adaptive Methods for Short-Term Electricity Load Forecasting During COVID-19 Lockdown in France. *IEEE Transactions on Power Systems*.
- First place with Y. Goude at Day-Ahead Electricity Demand Forecasting Competition: Post-Covid Paradigm, hosted by *IEEE DataPort*.

It would be interesting to look at the building data during covid ?

IV. Perspectives

IV.2. VIKING: Variational Bayesian Variance Tracking

The choice of the variances in the state-space model is difficult. Here I simply used $Q = qI$. Our current work with O. Wintenberger consists in treating the variances as latent variables which are estimated jointly with the state:

$$
\begin{aligned}
y_t - \theta_t^\top x_t &\sim \mathcal{N}(0, \exp(a_t)) \,, \\
\theta_t - \theta_{t-1} &\sim \mathcal{N}(0, diag(\exp(b_t))) \,, \\
a_t - a_{t-1} &\sim \mathcal{N}(0, \rho_a) \,, \\
b_t - b_{t-1} &\sim \mathcal{N}(0, \rho_b I) \,.
\end{aligned}
$$

The inference relies on the Variational Bayes approach: we estimate the joint posterior distribution with a factorized one of the form:

$$\mathcal{N}(\hat{\theta}_{t|t}, P_{t|t})\mathcal{N}(\hat{a}_{t|t}, s_{t|t})\mathcal{N}(\hat{b}_{t|t}, \Sigma_{t|t})$$

The best factorized distribution is obtained minimizing the Kullback-Leibler divergence.

IV. Perspectives

IV.3. Metric optimization

Kalman Filters are designed to optimize the quadratic loss and may be seen as stochastic gradient algorithm on that loss. We used the competition metric only for validation.

There are a few ways to change that:

- Based on the metric, consider one unique model instead of one per 15 minutes,
- Use aggregation of models (not uniform average) with diverse forecasting methods where the metric is specified.

Questions ?