Adaptive Forecasting by Kalman Filter

Application to Electricity Consumption during Spring 2020

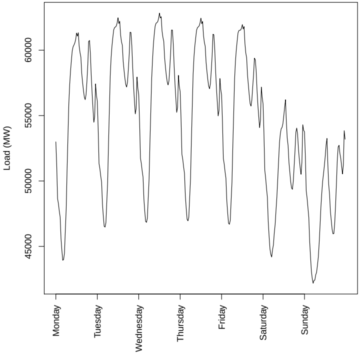
Joseph de Vilmarest

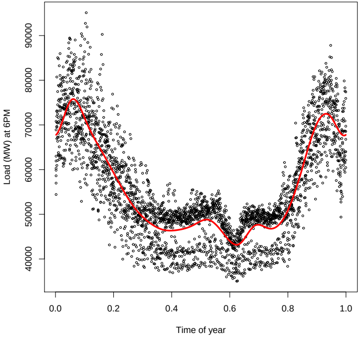
Olivier Wintenberger, LPSM, Sorbonne Université
Yannig Goude and Thi Thu Huong Hoang, EDF R&D

JDS June 9th, 2021

Electricity load description

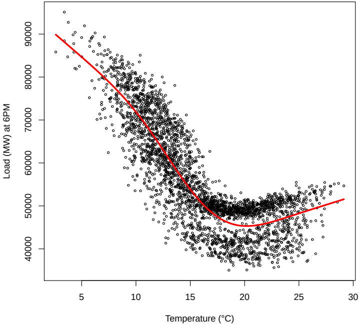
French electricity load (Réseau et Transport d'Électricité) from January 1st 2012 to June 7th 2020 (half-hour period).





Dependence to temperature

French temperature (MétéoFrance) : 32 cities every 3h.



Generalized Additive Model

For each time of day (half-hour) we model the load as

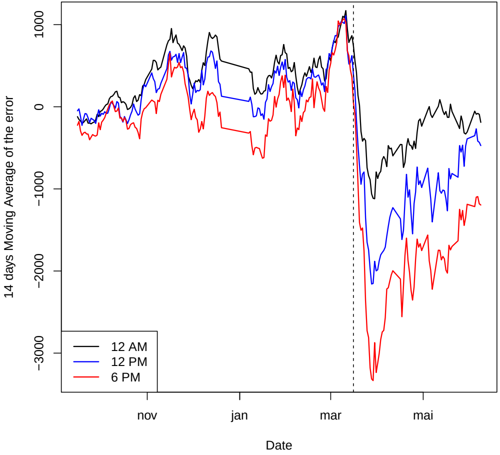
$$y_t = f_1(X_t^1) + ... + f_d(X_t^d) + \varepsilon_t \,,$$

$$f_j(x) = \sum_{k=1}^{m_j} \beta_{j,k} B_{j,k}(x) \,.$$

The different variables are

- Calendar variables (day of the week, time of year),
- Temperature, exponential smoothing variants,
- Lagged load (1 day ago, 1 week ago).

Drift of the model



State-Space model

Fixed model :

$$y_t = \sum_{j=1}^{d} f_j(X_t^j) + \varepsilon_t \,.$$

State-Space model

Fixed model :

$$y_t = \sum_{j=1}^{d} f_j(X_t^j) + \varepsilon_t \, .$$

Adaptive model :

$$\begin{aligned} y_t &= \theta_t^\top f(X_t) + \varepsilon_t \, , \\ \theta_{t+1} &= \theta_t + \eta_t \, , \end{aligned}$$

(ε_t), (η_t) : i.i.d. gaussian noises of variances σ^2, Q.

Kalman Filter

We estimate

$$\hat{\theta}_t = \mathbb{E}[\theta_t \mid X_1, y_1, ..., X_{t-1}, y_{t-1}]\,,$$
$$P_t = \mathbb{E}[(\theta_t - \hat{\theta}_t)(\theta_t - \hat{\theta}_t)^\top \mid X_1, y_1, ..., X_{t-1}, y_{t-1}]\,.$$

Then
$y_t \mid X_1, y_1, ..., X_{t-1}, y_{t-1} \sim \mathcal{N}(\hat{\theta}_t^\top f(X_t), \sigma^2 + f(X_t)^\top P_t f(X_t))$.

Kalman Filter

We estimate

$$\hat{\theta}_t = \mathbb{E}[\theta_t \mid X_1, y_1, ..., X_{t-1}, y_{t-1}]\,,$$
$$P_t = \mathbb{E}[(\theta_t - \hat{\theta}_t)(\theta_t - \hat{\theta}_t)^\top \mid X_1, y_1, ..., X_{t-1}, y_{t-1}]\,.$$

Then
$y_t \mid X_1, y_1, ..., X_{t-1}, y_{t-1} \sim \mathcal{N}(\hat{\theta}_t^\top f(X_t), \sigma^2 + f(X_t)^\top P_t f(X_t))$.

Theorem (Kalman and Bucy, 1961)

If the data-generating process is the state-space model of parameters σ^2, Q, we have

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \frac{P_t f(X_t)}{\sigma^2 + f(X_t)^\top P_t f(X_t)}(y_t - \hat{\theta}_t^\top f(X_t))\,,$$
$$P_{t+1} = P_t - \frac{P_t f(X_t) f(X_t)^\top P_t}{\sigma^2 + f(X_t)^\top P_t f(X_t)} + Q\,.$$

The optimization equivalence

It is equivalent to write

$$
\begin{aligned}
P_{t|t} &= P_t - \frac{P_t f(X_t) f(X_t)^\top P_t}{\sigma^2 + f(X_t)^\top P_t f(X_t)}\,, \\
\hat{\theta}_{t+1} &= \hat{\theta}_t - \frac{P_{t|t}}{\sigma^2} \left(\frac{\partial}{\partial \theta} (y_t - \theta^\top f(X_t))^2 \right) \Big|_{\hat{\theta}_t}\,, \\
P_{t+1} &= P_{t|t} + Q\,.
\end{aligned}
$$

Thus the Kalman Filter (and the Extended version as well) is a second-order Stochastic Gradient algorithm on the quadratic loss.

Choice of Q, σ^2 by maximizing the likelihood

$$y_t \mid X_1, y_1, ..., X_{t-1}, y_{t-1} \sim \mathcal{N}(\hat{\theta}_t^\top f(X_t), \sigma^2 + f(X_t)^\top P_t f(X_t)) \,.$$

Choice of Q, σ^2 by maximizing the likelihood

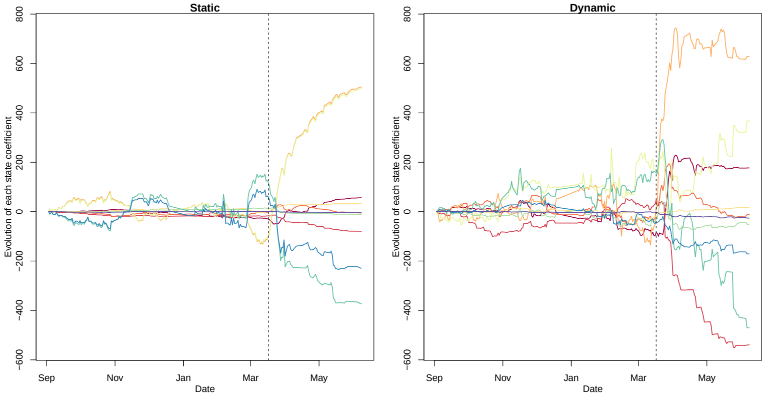
$$y_t \mid X_1, y_1, ..., X_{t-1}, y_{t-1} \sim \mathcal{N}(\hat{\theta}_t^\top f(X_t), \sigma^2 + f(X_t)^\top P_t f(X_t)) \,.$$

The log-likelihood is then

$$\sum_{t=1}^{n} \left(-\frac{1}{2} \log(2\pi(\sigma^2 + f(X_t)^\top P_t f(X_t))) - \frac{1}{2} \frac{(y_t - \hat{\theta}_t^\top f(X_t))^2}{\sigma^2 + f(X_t)^\top P_t f(X_t)} \right) \,,$$

with $\hat{\theta}_t, P_t$ depending on $\hat{\theta}_1, P_1, \sigma^2, Q$.

Coefficient evolution at 6 PM



Static : $\theta_{t+1} = \theta_t$ ($Q = 0$).

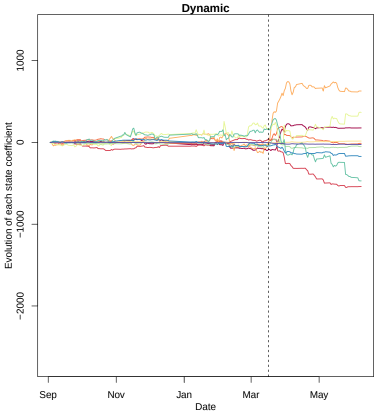
Break at COVID lockdown

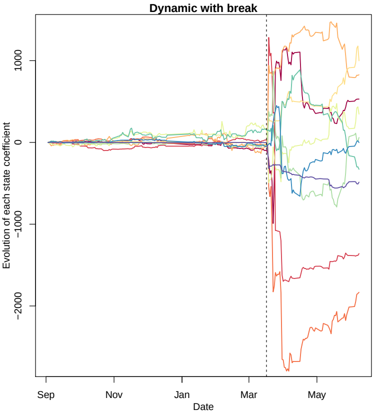
$$y_t = \theta_t^\top f(X_t) + \varepsilon_t \,,$$
$$\theta_{t+1} = \theta_t + \eta_t \,,$$

(ε_t), (η_t) : i.i.d. gaussian noises of variances σ^2, Q.

- Would it be better to consider Q_t ?
- First test : $Q_t = Q$ except $Q_T \gg Q$ with T the lockdown beginning. We test $Q_T = P_1$.

Coefficient evolution at 6 PM





Work in progress : dynamic variances

We model the variances as dynamic latent variables : $\sigma_t^2 = f(a_t)$ and $Q_t = g(b_t)$ (for instance exponential or quadratic function) :

$$a_0 \sim \mathcal{N}(\hat{a}_0, s_0)\,, \qquad b_0 \sim \mathcal{N}(\hat{b}_0, \Sigma_0)\,,$$

$$a_t - a_{t-1} \sim \mathcal{N}(0, \rho_a)\,, \qquad b_t - b_{t-1} \sim \mathcal{N}(0, \rho_b I)\,.$$

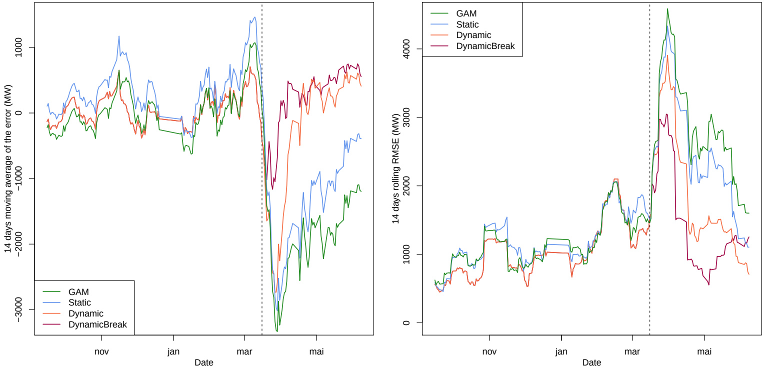
We study the Variational Bayesian approach : we approach the posterior distribution of θ_t, a_t, b_t with a simple factorized distribution.

Summary of the methods

We have described several methods :

- Generalized Additive Model : $y_t \sim \mathcal{N}(\theta^\top f(X_t), \sigma^2)$, $\theta = \mathbf{1}$.
- Static setting : same model with θ learned incrementally.
- Dynamic : $y_t \sim \mathcal{N}(\theta_t^\top f(X_t), \sigma^2)$ with $\theta_{t+1} - \theta_t \sim \mathcal{N}(0, Q)$.
- Dynamic with break : $\theta_{t+1} - \theta_t \sim \mathcal{N}(0, Q_t)$ with $Q_t = Q$ except $Q_T \gg Q$.
- Dynamic variances.

Rolling performances at 6 PM



Aggregate RMSE

	Before 03/16	03/16 - 04/15	04/16 - 06/07
Base GAM	1085 MW	2961 MW	1753 MW
Static	1077 MW	2923 MW	1588 MW
Dynamic	**979 MW**	2351 MW	1002 MW
DynamicBreak	-	**1902 MW**	**854 MW**

Conclusion

- D. Obst, J. de Vilmarest and Y. Goude : *Adaptive Methods for Short-Term Electricity Load Forecasting During COVID-19 Lockdown in France* (IEEE Transactions on Power Systems).
- We applied Kalman Filter to adapt additive models and also neural networks to win a competition (*Day-Ahead Electricity Demand Forecasting : Post-COVID Paradigm*).
- Work in progress : adaptive estimation of σ_t^2, Q_t.
  J. de Vilmarest and O. Wintenberger : *Recursive Estimation of State-Space Noise Covariance Matrix by Approximate Variational Bayes* (working paper arXiv 2104.10777).