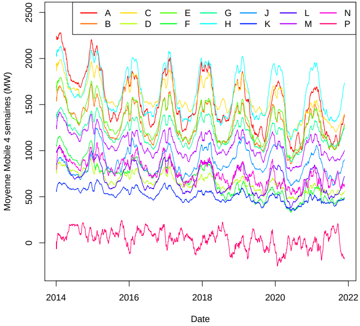
Modèles espace-état pour la prévision adaptative de consommation électrique

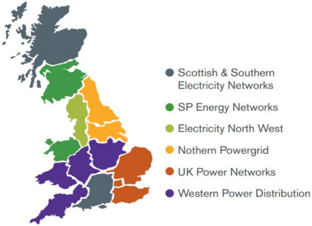
Jethro Browell, Matteo Fasiolo, Yannig Goude, **Joseph de Vilmarest** et Olivier Wintenberger

Journées MAS: 30 Août 2022

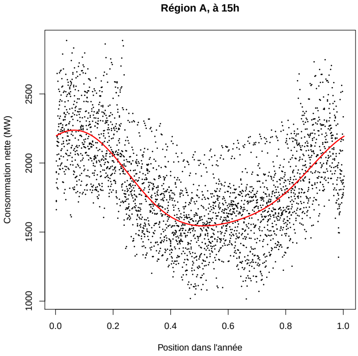
Prévision de séries temporelles

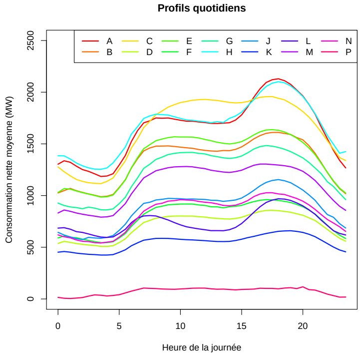
Nous cherchons à prévoir $y_t \in \mathbb{R}$. Décomposition en 14 régions.
Consommation *nette* = consommation - production *non pilotable*.



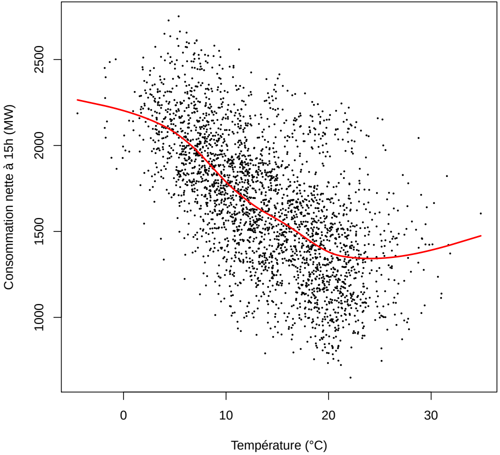


Variables explicatives: calendaires





Variables explicatives: météorologie



Objectif

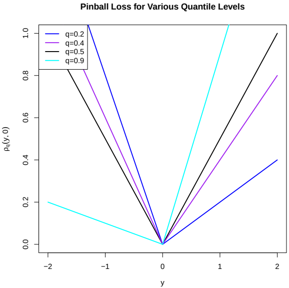
Nous souhaitons prévoir y_t sachant x_t. Dans quel objectif ?

- Prévision **en moyenne**: estimation de $\mathbb{E}[y_t \mid x_t]$.
  Équivalent au minimum de $\mathbb{E}[(y_t - \hat{y}_t)^2 \mid x_t]$.

Objectif

Nous souhaitons prévoir y_t sachant x_t. Dans quel objectif ?

- Prévision **en moyenne**: estimation de $\mathbb{E}[y_t \mid x_t]$.
  Équivalent au minimum de $\mathbb{E}[(y_t - \hat{y}_t)^2 \mid x_t]$.
- Prévision **en probabilité**: estimation de $\mathcal{L}(y_t \mid x_t)$.
  Pour $0 < q < 1$ la prévision $\hat{y}_{t,q}$ satisfait $\mathbb{P}(y_t \leq \hat{y}_{t,q} \mid x_t) = q$.
  Équivalent au minimum de $\mathbb{E}[\rho_q(y_t, \hat{y}_t) \mid x_t]$:



Offline vs Online

- **Offline**: $\hat{y}_t = f_{\hat{\theta}}(x_t)$.
  Exemple: Empirical Risk Minimizer

$$\hat{\theta} \in \arg\min \sum_{t \in \mathcal{T}} \ell(y_t, f_{\hat{\theta}}(x_t))$$

Offline vs Online

- **Offline**: $\hat{y}_t = f_{\hat{\theta}}(x_t)$.
  Exemple: Empirical Risk Minimizer

$$\hat{\theta} \in \arg\min \sum_{t \in \mathcal{T}} \ell(y_t, f_{\hat{\theta}}(x_t))$$

- **Online / Adaptatif**: $\hat{y}_t = f_{\hat{\theta}_t}(x_t)$ avec $\hat{\theta}_{t+1} = \Phi(\hat{\theta}_t, x_t, y_t)$.
  Exemple: Online Gradient Descent

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \gamma_t \frac{\partial \ell(y_t, f_\theta(x_t))}{\partial \theta}\Big|_{\hat{\theta}_t}$$

Modèle initial en deux étapes

- Modèle additif généralisé Gaussien pour la **prévision en moyenne**:

$$y_t = f_1(x_{t,1}) + \ldots + f_d(x_{t,d}) + \varepsilon_t \,, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \,.$$

$f_1, \ldots, f_d$: effets décomposés sur une base de splines:

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

Modèle initial en deux étapes

- Modèle additif généralisé Gaussien pour la **prévision en moyenne**:

$$y_t = f_1(x_{t,1}) + \ldots + f_d(x_{t,d}) + \varepsilon_t \,, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \,.$$

$f_1, \ldots, f_d$: effets décomposés sur une base de splines:

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

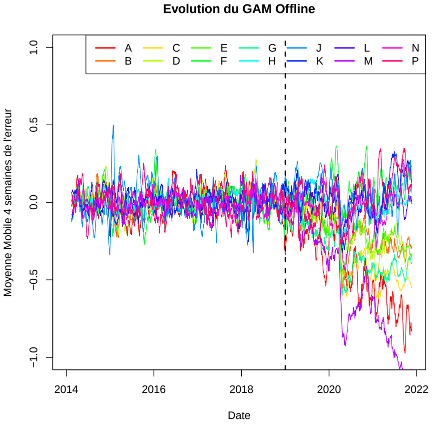
- **Prévision probabiliste**: régression quantile sur les résidus car l'hypothèse Gaussienne n'est pas satisfaite en pratique.

$$\beta_q \in \arg\min_{\beta \in \mathbb{R}^{d_0}} \sum_{t \in \mathcal{T}} \rho_q(y_t - \hat{y}_t, \beta^\top z_t) \,,$$

$$\rho_q(y, \hat{y}_q) = (\mathbb{1}_{y < \hat{y}_q} - q)\,(\hat{y}_q - y) \,.$$

Motivation à l'adaptation

Entraînement: 2014-2018. Test: 2019-2021.



Introduction

Prévision Moyenne

Prévision Probabiliste

Modèle espace-état linéaire Gaussien

- GAM:

$$y_t - \mathbf{1}^\top f(x_t) \sim \mathcal{N}(0, \sigma^2)\,.$$

- Adaptation espace-état:

$$\begin{aligned} &y_t - \theta_t{}^\top f(x_t) \sim \mathcal{N}(0, \sigma_t^2)\,, \\ &\theta_t - \theta_{t-1} \sim \mathcal{N}(0, Q_t)\,. \end{aligned}$$

Modèle espace-état linéaire Gaussien

- GAM:

$$y_t - \mathbf{1}^\top f(x_t) \sim \mathcal{N}(0, \sigma^2)\,.$$

- Adaptation espace-état:

$$\begin{aligned} y_t - {\theta_t}^\top f(x_t) &\sim \mathcal{N}(0, \sigma_t^2)\,, \\ \theta_t - \theta_{t-1} &\sim \mathcal{N}(0, Q_t)\,. \end{aligned}$$

Théorème (R. Kalman and R. Bucy, 1961)

Si le modèle espace-état est satisfait pour des variances connues, et si $\theta_1 \sim \mathcal{N}(\hat{\theta}_1, P_1)$, alors $\theta_{t+1} \mid (x_s, y_s)_{s \le t} \sim \mathcal{N}(\hat{\theta}_{t+1}, P_{t+1})$ pour

$$P_{t|t} = P_t - \frac{P_t f(x_t) f(x_t)^\top P_t}{f(x_t)^\top P_t f(x_t) + \sigma_t^2}\,, \qquad P_{t+1} = P_{t|t} + Q_{t+1}\,,$$

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \frac{P_{t|t}}{\sigma_t^2}\Big(f(x_t)(\hat{\theta}_t^\top f(x_t) - y_t)\Big)\,.$$

Le filtre de Kalman, un algorithme de gradient

$$P_{t|t} = P_t - \frac{P_t f(x_t) f(x_t)^\top P_t}{f(x_t)^\top P_t f(x_t) + \sigma_t^2}\,, \qquad P_{t+1} = P_{t|t} + Q_{t+1}\,,$$
$$\hat{\theta}_{t+1} = \hat{\theta}_t - \frac{P_{t|t}}{\sigma_t^2}\Big(f(x_t)(\hat{\theta}_t^\top f(x_t) - y_t)\Big)\,.$$

1. **Statique**: $Q_t = 0, \sigma_t^2 = 1$. Alors $P_t = O(1/t)$.
2. **Dynamique** avec variances constantes: $Q_t = Q, \sigma_t^2 = \sigma^2$. Alors $P_t = O(1)$.
3. **Dynamique** avec variances adaptatives[1].

[1] J. de Vilmarest, O. Wintenberger (2021), Viking: Variational Bayesian Variance Tracking, *arXiv:2104.10777*

Variances constantes

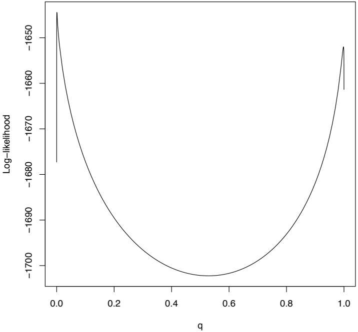
$$
\begin{aligned}
y_t - \theta_t^\top f(x_t) &\sim \mathcal{N}(0, \sigma^2)\,, \\
\theta_t - \theta_{t-1} &\sim \mathcal{N}(0, Q)\,.
\end{aligned}
$$

[1] D. Obst, J. de Vilmarest, Y. Goude (2021), Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France, *IEEE Transactions on Power Systems*

Variances constantes

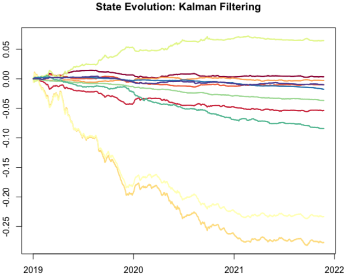
$$
\begin{aligned}
y_t - \theta_t^\top f(x_t) &\sim \mathcal{N}(0, \sigma^2)\,, \\
\theta_t - \theta_{t-1} &\sim \mathcal{N}(0, Q)\,.
\end{aligned}
$$

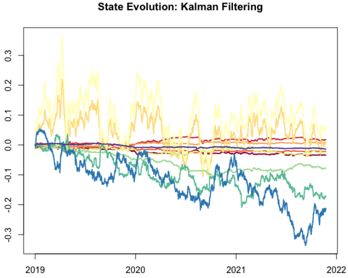
- Log-vraisemblance non convexe. Pas de garantie d'optimalité.
- Q diagonale[1]. Optimisation par *iterative grid search*.



[1] D. Obst, J. de Vilmarest, Y. Goude (2021), Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France, *IEEE Transactions on Power Systems*

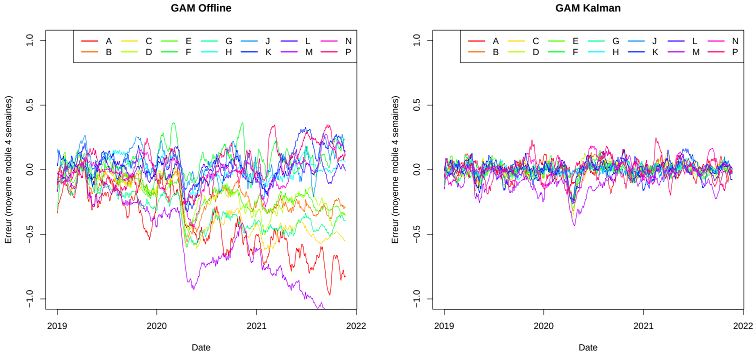
Évolution des coefficients





Gauche: cadre statique avec $\theta_{t+1} = \theta_t$.
Droite: cadre dynamique où $\theta_{t+1} - \theta_t \sim \mathcal{N}(0, Q)$.

Correction du biais



Performance

$$RMSE = \sqrt{\frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}}(y_t - \hat{y}_t)^2}\,, \qquad MAE = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}}|y_t - \hat{y}_t|$$

	2019		2020		2021	
Forecast	nRMSE	nMAE	nRMSE	nMAE	nRMSE	nMAE
Persistence (7 days)	0.691	0.589	0.710	0.599	0.737	0.639
Persistence (2 days)	0.767	0.686	0.755	0.668	0.736	0.668
Offline GAM	0.356	0.327	0.485	0.453	0.635	0.601
Incremental offline GAM (yearly)	-	-	0.407	0.376	0.387	0.378
Incremental offline GAM (daily)	0.338	0.307	0.370	0.344	0.377	0.365
Kalman GAM (Static)	0.337	0.307	0.374	0.347	0.380	0.368
Kalman GAM (Dynamic)	**0.324**	**0.292**	**0.328**	**0.301**	**0.332**	**0.307**

Introduction

Prévision Moyenne

Prévision Probabiliste

Prévisions quantiles par filtre de Kalman

Le filtre de Kalman fournit $\hat{\theta}_t, P_t$ tel que $\theta_t \mid (x_s, y_s)_{s<t} \sim \mathcal{N}(\hat{\theta}_t, P_t)$ et $y_t - \theta_t^\top f(x_t) \sim \mathcal{N}(0, \sigma^2)$.

Prévisions quantiles par filtre de Kalman

Le filtre de Kalman fournit $\hat{\theta}_t, P_t$ tel que $\theta_t \mid (x_s, y_s)_{s<t} \sim \mathcal{N}(\hat{\theta}_t, P_t)$ et $y_t - \theta_t^\top f(x_t) \sim \mathcal{N}(0, \sigma^2)$.

- **Si le modèle est satisfait**:

$$y_t \sim \mathcal{N}(\hat{\theta}_t^\top f(x_t), \sigma^2 + f(x_t)^\top P_t f(x_t)) \,.$$

Prévisions quantiles par filtre de Kalman

Le filtre de Kalman fournit $\hat{\theta}_t, P_t$ tel que $\theta_t \mid (x_s, y_s)_{s<t} \sim \mathcal{N}(\hat{\theta}_t, P_t)$ et $y_t - \theta_t^\top f(x_t) \sim \mathcal{N}(0, \sigma^2)$.

- **Si le modèle est satisfait**:

$$y_t \sim \mathcal{N}(\hat{\theta}_t^\top f(x_t), \sigma^2 + f(x_t)^\top P_t f(x_t))\,.$$

- **En pratique**: prévision en moyenne, puis régression quantile sur les résidus $y_t - \hat{\theta}_t^\top f(x_t)$.
  $\rightarrow$ peut-on obtenir une régression quantile adaptative ?

Régression quantile adaptative

Régression quantile *offline*:

$$\beta_q \in \arg\min_{\beta \in \mathbb{R}^{d_0}} \sum_{t \in \mathcal{T}} \rho_q(y_t - \hat{y}_t, \beta^\top z_t)\,.$$

Online Gradient Descent avec pas de gradient $\alpha > 0$:

$$\beta_{t+1,q} = \beta_{t,q} - \alpha \frac{\partial \rho_q(y_t - \hat{y}_t, \beta^\top z_t)}{\partial \beta}\Big|_{\beta_{t,q}}\,,$$

avec $\frac{\partial \rho_q(y_t - \hat{y}_t, \beta^\top z_t)}{\partial \beta}\Big|_{\beta_{t,q}} = (\mathbb{1}_{y_t < \hat{y}_t + \beta_{t,q}^\top z_t} - q)\; z_t$ (hors cas dégénéré).

Choix du pas de gradient par agrégation d'experts

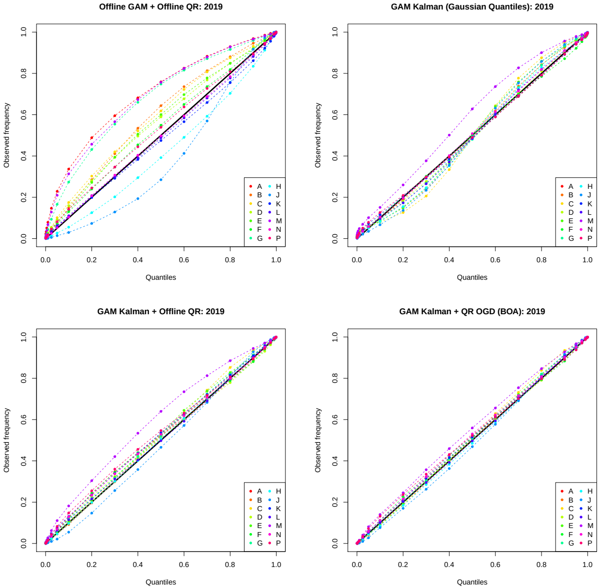
- Nous utilisons différents pas de gradients α_k, typiquement 10^k.
- Création d'experts $\hat{y}_{t,q}^{(k)}$ correspondant aux α_k.
- Agrégation d'experts: Bernstein Online Aggregation[2]:

$$\hat{y}_{t,q} = \sum_k p_t^{(k)} \hat{y}_{t,q}^{(k)} \,,$$

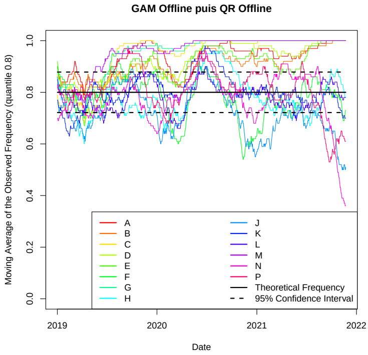
où $p_t^{(k)}$ est estimé récursivement.

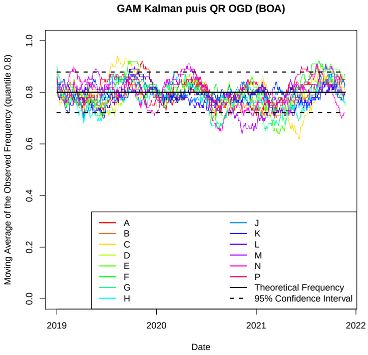
[2] O. Wintenberger (2017), Optimal learning with Bernstein online aggregation, *Machine Learning*

Calibration



Calibration au cours du temps





Métrique

Numériquement nous utilisons le *continuous ranked probability score*[3]:

$$CRPS(F, y) = \int_{-\infty}^{+\infty} (F(x) - \mathbb{1}_{y \leq x})^2 dx = 2 \int_0^1 \rho_q(y, F^{-1}(q)) dq \,.$$

Version discrète:

$$RPS\big((\hat{y}_{q_1}, \ldots, \hat{y}_{q_I}), y\big) = \sum_{i=1}^{I} \rho_{q_i}(y, \hat{y}_{q_i})(q_{i+1} - q_{i-1}) \,,$$

[3] T. Gneiting and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *Journal of the American statistical Association*

Performances

	2019	2020	2021
Offline Method	0.231	0.338	0.454
GAM Kalman (Gaussian Quantiles)	0.212	0.217	0.222
GAM Kalman + Offline QR	**0.206**	**0.214**	**0.217**
Offline GAM + QR OGD (10^{-3})	0.218	0.270	0.293
Offline GAM + QR OGD (10^{-2})	0.207	0.221	0.218
Offline GAM + QR OGD (10^{-1})	0.250	0.248	0.293
Offline GAM + QR OGD (BOA)	0.204	0.211	0.216
GAM Kalman + QR OGD (10^{-2})	0.205	0.204	0.212
GAM Kalman + QR OGD (BOA)	**0.202**	**0.201**	**0.209**

Conclusion

- Utilisation d'un modèle espace-état pour la prévision en moyenne. Similaire à un algorithme de descente de gradient. Analogie dans le cas probabiliste: Online Gradient Descent.
- L'évolution de la consommation d'électricité est bien capturée par les modèles espace-état: tests sur différents pays, différentes échelles, différents objectifs.