On-line Bayesian prediction of hourly electricity demand for the residential sector in Ontario

McGill | Department of Epidemiology, Biostatistics and Occupational Health

Victoire Michal, Paritosh Kumar Roy, Dirk Douwes-Schultz

Biostatistics, McGill University

Société statistique du Canada | Statistical Society of Canada

Introduction

In time series analysis, dynamic linear models (DLMs) are an alternative to the usual auto-regressive integrated moving average (ARIMA) models. In DLMs, the observed process is decomposed as the sum of different time series components such as deterministic or stochastic trends, seasonality and covariate effects. The parameters change smoothly with time, accommodating local structures [4]. Different from ARIMA models, DLMs naturally deal with non-stationary time series and structural changes. On-line one-step-ahead and K-steps-ahead forecasts are obtained through probability distributions that account for the uncertainty in the inference procedure.

Available datasets

Yearly dataset (2003-2016)

- Annual use of electricity per sector (PJ)
- Annual electricity consumption by end-use in the residential sector (PJ)

Hourly dataset (2003-2016)

- Aggregated hourly use of electricity for all sectors (MW/h)
- Precipitation (mm/h)
- Temperature (°C)
- Irradiances surface & toa (W/m^2)
- Snowfall (mm/h)
- Snow depth (cm)
- Cloud coverage (%)
- Air density (kg/m^3)

Notation and additional variables

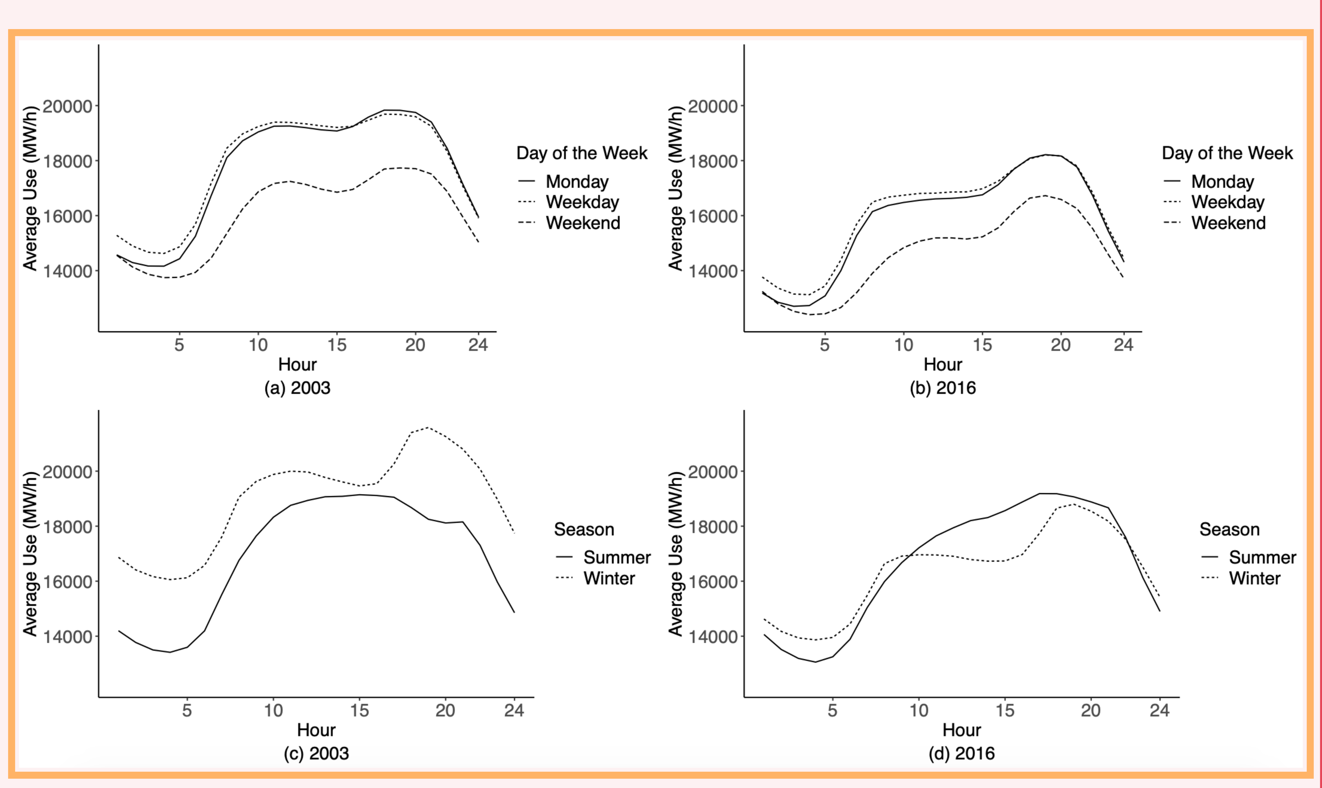
- $\tau = 1, \cdots, 24$: hour of the day
- $t = 1, \cdots, 5114$: day
- $T_t^{(\tau)}$: scaled temperature on day t at hour τ
- $d_{1,t} = 1$ if day t is a weekend day, 0 otherwise
- $d_{2,t} = 1$ if day t is a week day (except a Monday), 0 otherwise

Objective

Exploratory analysis

- First, the conversion between the hourly MW/h scale to the annual PJ scale (1 PJ $= 3.6/1,000,000$ MW/h) was not exact. For example, summing over the aggregated observed hourly use of 2003 yielded 552.4 PJ whereas the sum over the observed annual sector-specific usage in 2003 was equal to 460.1 PJ.
- In that regard, **the model performance assessments will be calculated solely based on the hourly dataset.** When considering the sector-specific total, the total obtained from the *hourly dataset* will be multiplied by the observed annual proportions of the residential sector from the *annual dataset*.

Figure 1: Yearly average hourly electricity demand for different days of the week and seasons



- The electricity use differs between weekdays and the weekend but also between the seasons. The weekday/season effects also seem to change with time.
- The histograms of the electricity use per hour showed a Gaussian shape.

Analysis plan

- The proportion of the annual residential sector consumption in the aggregated annual usage does not change much over the years with mean equal to 34% and standard deviation equal to 0.9%. Therefore, **the sector-specific annual prediction will be modelled as 34% of the aggregated annual predictions.**
- **Separate DLMs** are fitted to the aggregated electricity usage: one for each hour of the day.
- **On-line predictions** of one-step-ahead, month-step-ahead year-step-ahead electricity demand are obtained. This allows **decision makers** to choose the prediction window they prefer to use.
- Different models were fitted to the data and MAE (based on the one-step-ahead forecasts) was used to choose the best model.
- The final model includes a level, a trend, the scaled **temperature**, an **auto-regressive component** (the observed aggregated use at the same hour on the previous day), three **Fourier form seasonal harmonics**: one for the seasons (spring-summer / fall-winter), two for the weeks and finally, two dummy variables to **distinguish the weekend from the week** with Monday as the baseline.

Dynamic Linear Models: *Aggregated hourly electricity demand*

The **univariate DLM** for the aggregated electricity consumption, $y_t^{(\tau)}$, for each hour $\tau = 1, \cdots, 24$ with day $t = 1, \cdots, 5114$ is parameterised as follows:

$$y_t^{(\tau)} = \boldsymbol{F}_t^{(\tau)} \boldsymbol{\theta}_t^{(\tau)} + \varepsilon_t^{(\tau)}, \qquad \varepsilon_t^{(\tau)} \sim \mathcal{N}(0, V^{(\tau)})$$

$$\boldsymbol{\theta}_t^{(\tau)} = \boldsymbol{G}\boldsymbol{\theta}_{t-1}^{(\tau)} + \boldsymbol{\omega}_t^{(\tau)}, \qquad \boldsymbol{\omega}_t^{(\tau)} \sim \mathcal{N}_{12}(\boldsymbol{0}, \boldsymbol{W}^{(\tau)}),$$

with state vector, $\boldsymbol{\theta}$, design matrix, $\boldsymbol{F}$, and evolution matrix, $\boldsymbol{G}$:

$$\boldsymbol{\theta}_t^{(\tau)} = \left[\mu_{1,t}\ \mu_{2,t}\ \beta_{1,t}\ \beta_{2,t}\ a_{1,t}\ b_{1,t}\ a_{2,t}\ b_{2,t}\ a_{3,t}\ b_{3,t}\ \gamma_{1,t}\ \gamma_{2,t}\right]^{(\tau)\top},$$

$$\boldsymbol{F}_t^{(\tau)} = \left[1\ 0\ T_t^{(\tau)}\ y_{t-1}^{(\tau)}\ 1\ 0\ 1\ 0\ 1\ 0\ d_{1,t}\ d_{2,t}\right],$$

$$\boldsymbol{G} = \text{BDiag}(\boldsymbol{G}_1, \boldsymbol{G}_2, \boldsymbol{G}_3, \boldsymbol{G}_4, \boldsymbol{G}_5, \boldsymbol{G}_6), \quad \text{where} \quad \boldsymbol{G}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{G}_2 = \boldsymbol{G}_6 = \boldsymbol{I}_2,$$

and Fourier form **seasonal harmonics** for the seasons, $\boldsymbol{G}_3$, and the weeks, $\boldsymbol{G}_4$, $\boldsymbol{G}_5$:

$$\boldsymbol{G}_3 = \begin{bmatrix} \cos\left(\frac{2\pi}{180}\right) & \sin\left(\frac{2\pi}{180}\right) \\ -\sin\left(\frac{2\pi}{180}\right) & \cos\left(\frac{2\pi}{180}\right) \end{bmatrix}, \boldsymbol{G}_4 = \begin{bmatrix} \cos\left(\frac{2\pi}{7}\right) & \sin\left(\frac{2\pi}{7}\right) \\ -\sin\left(\frac{2\pi}{7}\right) & \cos\left(\frac{2\pi}{7}\right) \end{bmatrix}, \boldsymbol{G}_5 = \begin{bmatrix} \cos\left(\frac{4\pi}{7}\right) & \sin\left(\frac{4\pi}{7}\right) \\ -\sin\left(\frac{4\pi}{7}\right) & \cos\left(\frac{4\pi}{7}\right) \end{bmatrix}.$$

- The initial auto-regressive component, $y_0^{(\tau)}$, is set to the mean of $y^{(\tau)}$.
- Below, the superscript (τ) is dropped to ease notation.

Forward learning: Kalman Filter[2] for each τ

$$\pi(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) \overset{\text{Evolution}}{\Rightarrow} \pi(\boldsymbol{\theta}_t \mid D_{t-1}) \overset{\text{Updating}}{\Rightarrow} \pi(\boldsymbol{\theta}_t \mid D_t)$$

$$\Downarrow \text{Prediction}$$

$$\pi(y_t \mid D_{t-1})$$

1. Conjugate priors:

$$\boldsymbol{\theta}_0 \sim \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{C}_0) \quad \text{and} \quad V^{-1} \sim \mathcal{G}\left(\frac{n_0}{2}, \frac{S_0}{2}\right) \qquad \text{Information set: } D_t = \{y_t, D_{t-1}\}$$

2. One-step-ahead distributions:

$$\begin{aligned} \boldsymbol{\theta}_t \mid D_{t-1} &\sim \mathcal{T}_{n_{t-1}}(\boldsymbol{a}_t, \boldsymbol{R}_t) \\ V^{-1} \mid D_{t-1} &\sim \mathcal{G}\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right) \\ y_t \mid D_{t-1} &\sim \mathcal{T}_{n_{t-1}}(f_t, Q_t) \\ \boldsymbol{\theta}_t \mid D_t &\sim \mathcal{T}_{n_t}(\boldsymbol{m}_t, \boldsymbol{C}_t) \\ V^{-1} \mid D_t &\sim \mathcal{G}\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{a}_t &= \boldsymbol{G}\boldsymbol{m}_{t-1}, & \boldsymbol{R}_t &= \boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}^\top + \boldsymbol{W}_t \\ f_t &= \boldsymbol{F}_t^\top \boldsymbol{a}_t, & Q_t &= \boldsymbol{F}_t^\top \boldsymbol{R}_t \boldsymbol{F}_t + S_{t-1} \\ n_t &= n_{t-1} + 1, & S_t &= S_{t-1} + \frac{S_{t-1}}{n_t}\left(\frac{(y_t - f_t)^2}{Q_t} - 1\right) \\ \boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{R}_t \boldsymbol{F}_t \frac{y_t - f_t}{Q_t}, & \boldsymbol{C}_t &= \frac{S_t}{S_{t-1}}\left(\boldsymbol{R}_t - \frac{1}{Q_t}\boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{F}_t^\top \boldsymbol{R}_t^\top\right) \end{aligned}$$

Unknown state variance, $\boldsymbol{W}$: Discount factor[4]

- $\boldsymbol{W}_t$ controls the stochastic variation in the Kalman Filter recursions: given D_{t-1}, $\boldsymbol{W}_t$ increases the *uncertainty* from the ideal $\boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}^\top$ to the realistic $\boldsymbol{R}_t$.
- *Discounting*: impose $\boldsymbol{R}_t = \frac{1}{\delta}\boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}^\top \Rightarrow \boldsymbol{W}_t = \frac{1-\delta}{\delta}\boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}^\top$, where $\delta \in (0, 1]$ is the discount factor. Usually, $\delta \in [0.9, 0.995]$. If $\delta = 1$, then the uncertainty about $\boldsymbol{\theta}_t = \boldsymbol{G}\boldsymbol{\theta}_{t-1}$ is zero. In this study, δ was set to 0.995.

3. K-steps-ahead distributions:

$$\begin{aligned} \boldsymbol{\theta}_{t+k} \mid D_t &\sim \mathcal{T}_{n_t}(\boldsymbol{a}_t(k), \boldsymbol{R}_t(k)) \\ y_{t+k} \mid D_t &\sim \mathcal{T}_{n_t}(f_t(k), Q_t(k)) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{a}_t(0) &= \boldsymbol{m}_t, & \boldsymbol{R}_t(0) &= \boldsymbol{C}_t \\ \boldsymbol{a}_t(k) &= \boldsymbol{G}\boldsymbol{a}_t(k-1), & \boldsymbol{R}_t(k) &= \boldsymbol{G}\boldsymbol{R}_t(k-1)\boldsymbol{G}^\top + \boldsymbol{W}_{t+1} \\ f_t(k) &= \boldsymbol{F}_{t+k}^\top \boldsymbol{a}_t(k), & Q_t(k) &= \boldsymbol{F}_{t+k}^\top \boldsymbol{R}_t(k) \boldsymbol{F}_{t+k} + S_t \end{aligned}$$

Annual month-step and year-step totals

- To compare the predictions to the observed values, the sums of the predictions in each year will be used. The **posterior predictive distributions** will be approximated by

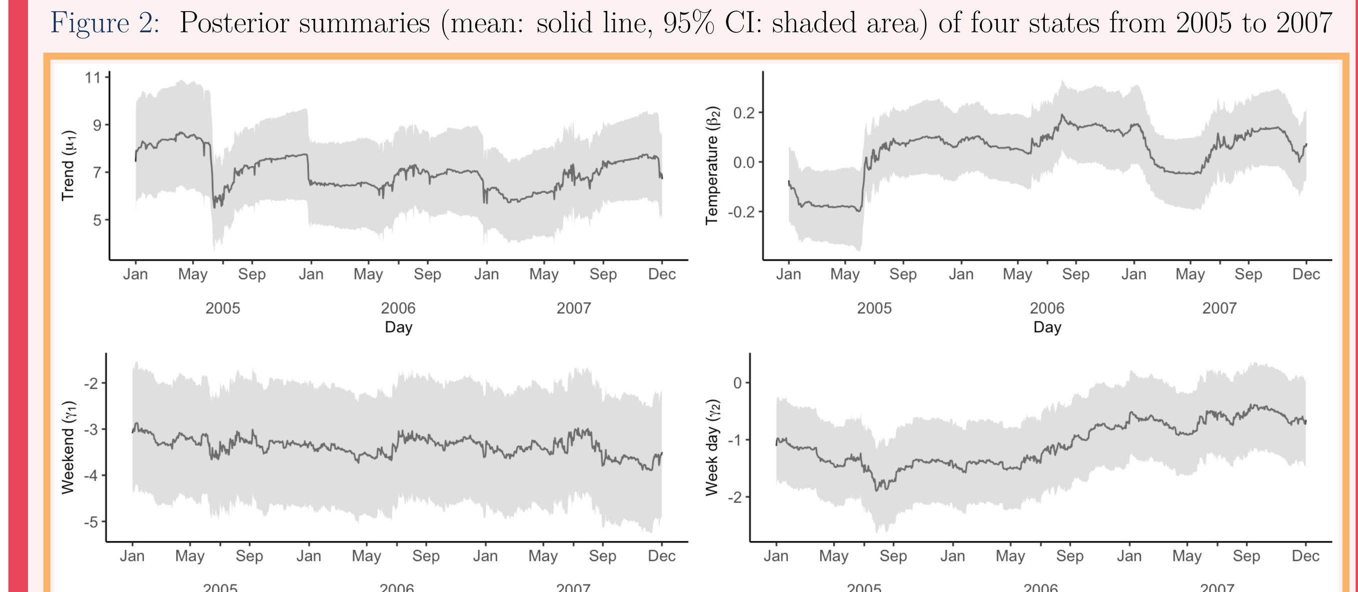
$$\sum_{t,\tau} y_t^{(\tau)} \sim \mathcal{N}\left(\sum_{t,\tau} f_t^{(\tau)}, \sum_{t,\tau} Q_t^{(\tau)}\right) \quad \text{and} \quad \sum_{k,\tau} y_{t+k}^{(\tau)} \sim \mathcal{N}\left(\sum_{k,\tau} f_t(k)^{(\tau)}, \sum_{k,\tau} Q_t(k)^{(\tau)}\right),$$

in the one-step and K-steps cases, respectively. These are not exact distributions as the independence assumption between $\varepsilon_t^{(\tau)}$ and $\boldsymbol{\omega}_t^{(\tau)}$ is violated in the summation [3]. The normal approximation comes from the large degrees of freedom in the $\mathcal{T}$ distributions.
- The **month-step-ahead** predictions can start from 2003/02/01 based on January 2003, then 2003/03/01 based on January and February 2003, adding one month of information at each step.
- The **year-step-ahead** predictions can start from 2004/01/01 based on 2003, then 2005/01/01 based on 2003 and 2004, adding one year of information at each step.

Results: *Effect of some model parameters*

- The residuals did not present any remaining particular structure.
- The *aggregated* electricity use is **decreasing with time**.

Figure 2: Posterior summaries (mean: solid line, 95% CI: shaded area) of four states from 2005 to 2007



Results: *Predicted electricity use*

Aggregated estimated hourly electricity usage

Figure 3: Posterior summaries of the on-line fitting with the one-step-ahead (left), month-step-ahead (middle) and year-step-ahead (right) forecasts, for τ = 12pm.
Last observations used: 2003/01/31 & 2016/11/30 (month-step); 2003/12/31 & 2015/12/31 (year-step)
(Posterior means: solid gray line, 95% CI: shaded area, observed values: solid black line and solid circles)

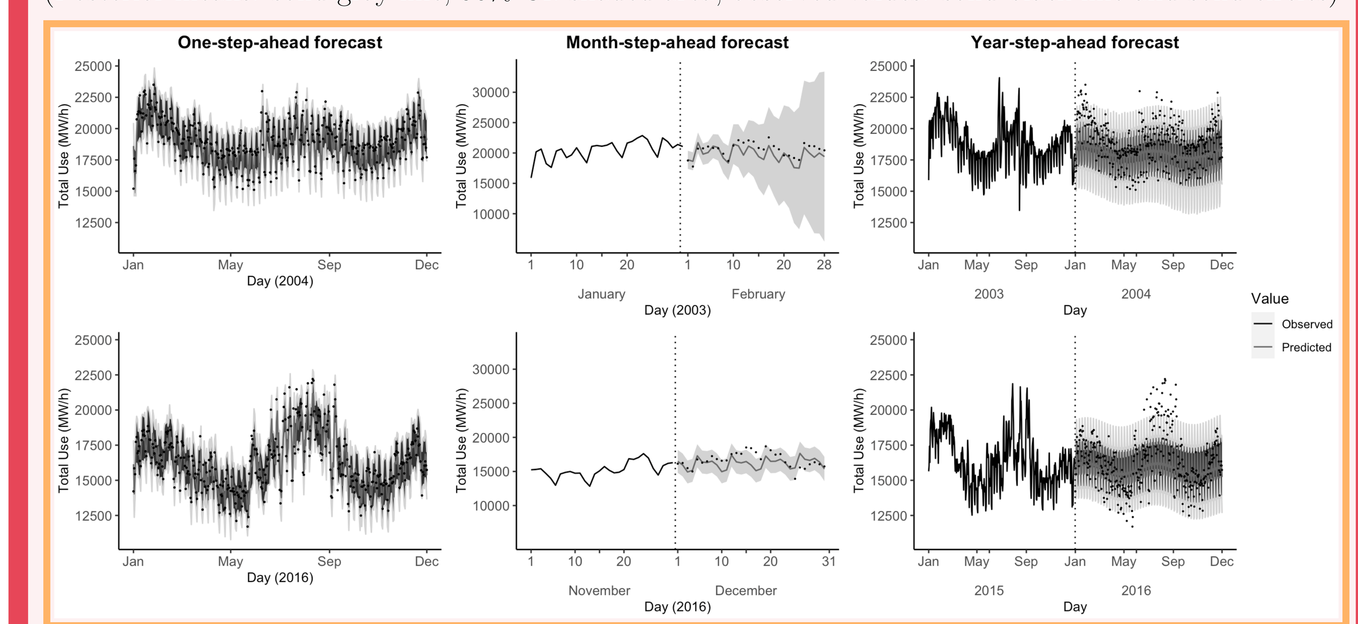
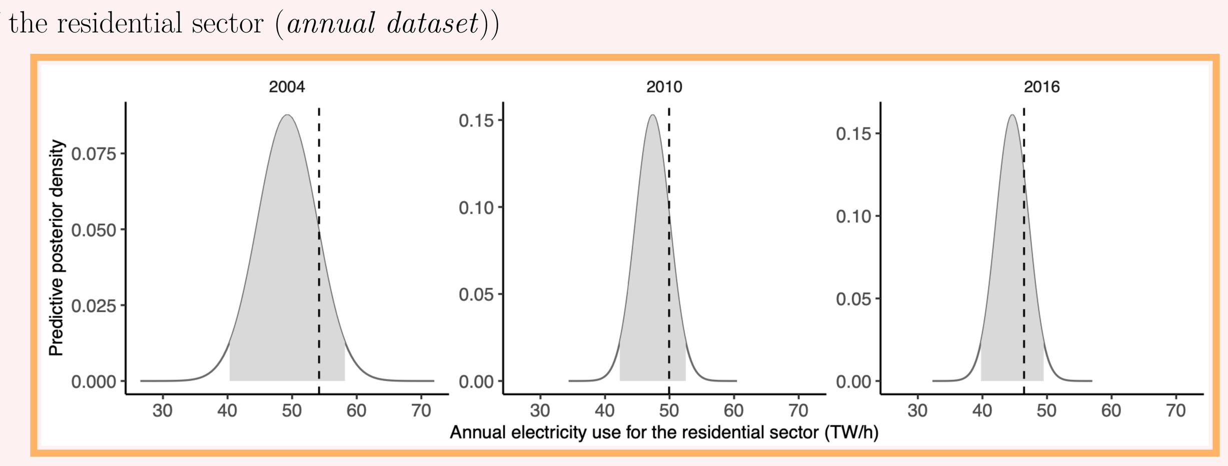


Table 1: Total use of electricity **in TW/h**: observed sum and sums of predictions (standard deviation)

Year	Total Observed	Aggregated One-step predictions	Aggregated Month-step predictions	Aggregated Year-step predictions
2003	151.7	152.0 (74.40)	-	-
2004	153.4	153.1 (7.02)	150.5 (7.49)	144.9 (13.39)
2005	157.0	157.1 (6.06)	155.8 (6.36)	148.1 (9.31)
2006	151.1	151.5 (5.80)	153.8 (6.12)	152.0 (8.67)
2007	152.2	151.9 (5.73)	150.6 (6.02)	139.9 (8.28)
2008	148.7	148.9 (5.71)	148.2 (6.00)	145.1 (8.39)
2009	139.2	139.3 (5.46)	137.6 (5.71)	147.3 (8.04)
2010	142.2	141.8 (5.31)	140.3 (5.58)	139.4 (7.68)
2011	141.5	141.5 (5.23)	141.9 (5.49)	132.9 (7.64)
2012	141.3	141.1 (5.21)	142.2 (5.46)	135.9 (7.50)
2013	140.7	140.6 (5.21)	139.1 (5.45)	138.4 (7.50)
2014	139.8	139.9 (5.17)	141.1 (5.41)	147.7 (7.46)
2015	137.0	136.9 (5.10)	137.2 (5.34)	140.4 (7.35)
2016	137.0	136.6 (5.13)	133.7 (5.35)	131.3 (7.28)
MAE (TW/h)		0.015	0.119	0.464
MAE (MW/h)		15335.1	118705.5	464245.3

Annual estimated use of electricity for the *residential sector*

Figure 4: Approximated posterior predictive densities of the residential annual consumption for 2004 (left), 2010 (middle) and 2016 (right), based on the year-step-ahead hourly predictions
(Shaded area: 95% CI, dashed line: observed sum (*hourly dataset*) multiplied by the observed proportion of the residential sector (*annual dataset*))



- Clearly, the electricity demand for the *residential sector* is **decreasing with time**.

Conclusions and Discussion

- DLMs allow for flexible and interpretable inference through the decomposition of the electric demand at hour τ and day t into components such as a mean trend, a dynamic regression and several seasonal effects.
- DLMs do not rely on the assumption of stationarity of the time series and naturally account for local structures and structural changes.
- Conjugate analysis provides closed form posterior distributions. Hence, the fitting procedure is fast, regardless of the size of the data.
- Measures of *uncertainty* are easily obtained following this Bayesian framework.
- Although only the hourly aggregated electricity use was modelled, the hourly residential demand can be estimated as 34% of the aggregated hourly prediction.
- The effect of **temperature** changes with time, being greater in the summer. The **day of week** and seasonal components for the **week** and the **seasons** also have time-varying effect on the demand.
- An alternative, to account for correlations within a day, would be to consider a multivariate DLM for $\boldsymbol{y}_t = \left[y_t^{(1)} \cdots y_t^{(24)}\right]^\top$ with Normal and Inverse-Wishart conjugate priors, as in [1].
- Sensitivity to the choice of the discount factor needs further investigation. Among possible solutions are Markov chain Monte Carlo methods.

References

[1] H. S. Migon and L. C. Alves, *Multivariate dynamic regression: modeling and forecasting for intraday electricity load*, Applied Stochastic Models in Business and Industry, 29 (2013), pp. 579–598.

[2] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic Linear Models with R*, Springer, 2009.

[3] A. M. Schmidt and D. Gamerman, *Temporal aggregation in dynamic linear models*, Journal of Forecasting, 16 (1997), pp. 293–310.

[4] M. West and J. Harrison, *Bayesian forecasting and dynamic models*, Springer Science & Business Media, 2006.

Acknowledgements

Special thanks to Alex Schmidt (Biotatistics, McGill University) who introduced us to DLMs.