Multi-task Online Learning for Probabilistic Load Forecasting

Onintze Zaballa, Verónica Álvarez and Santiago Mazuelas

ozaballa@bcamath.org

December 18, 2024

Machine learning algorithms for load forecasting

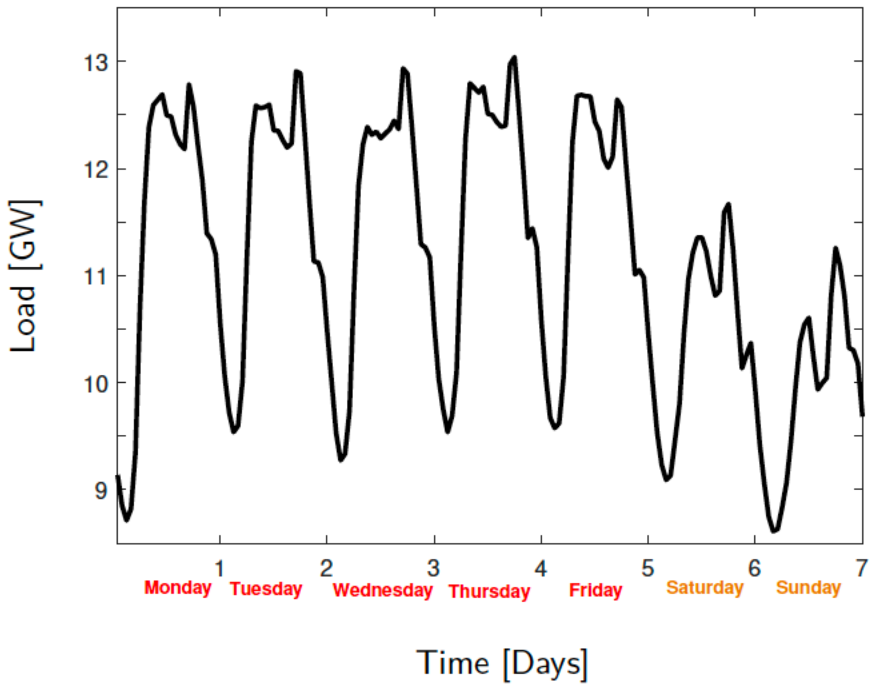
Santiago Mazuelas

Verónica Álvarez

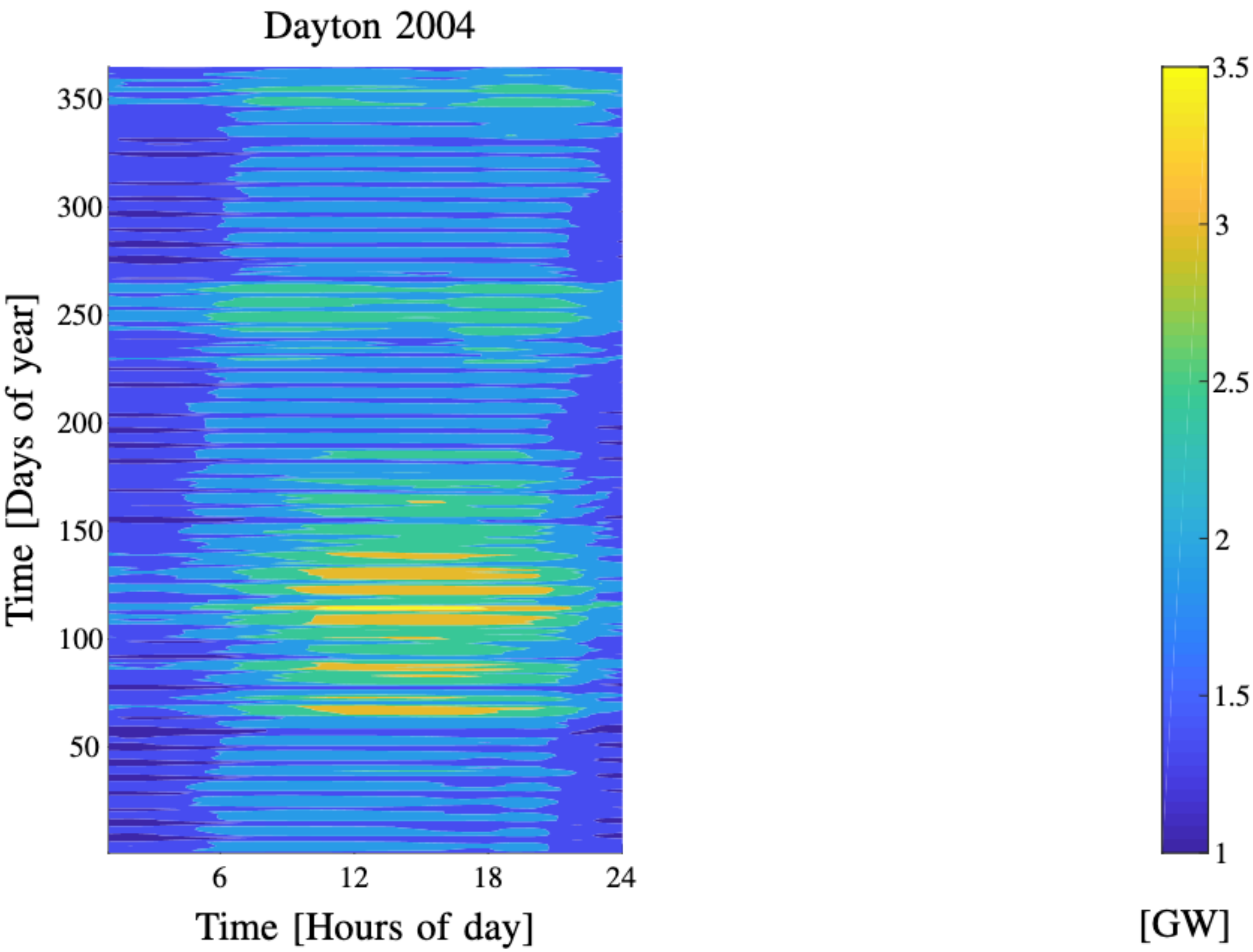
Onintze Zaballa

Load forecasting

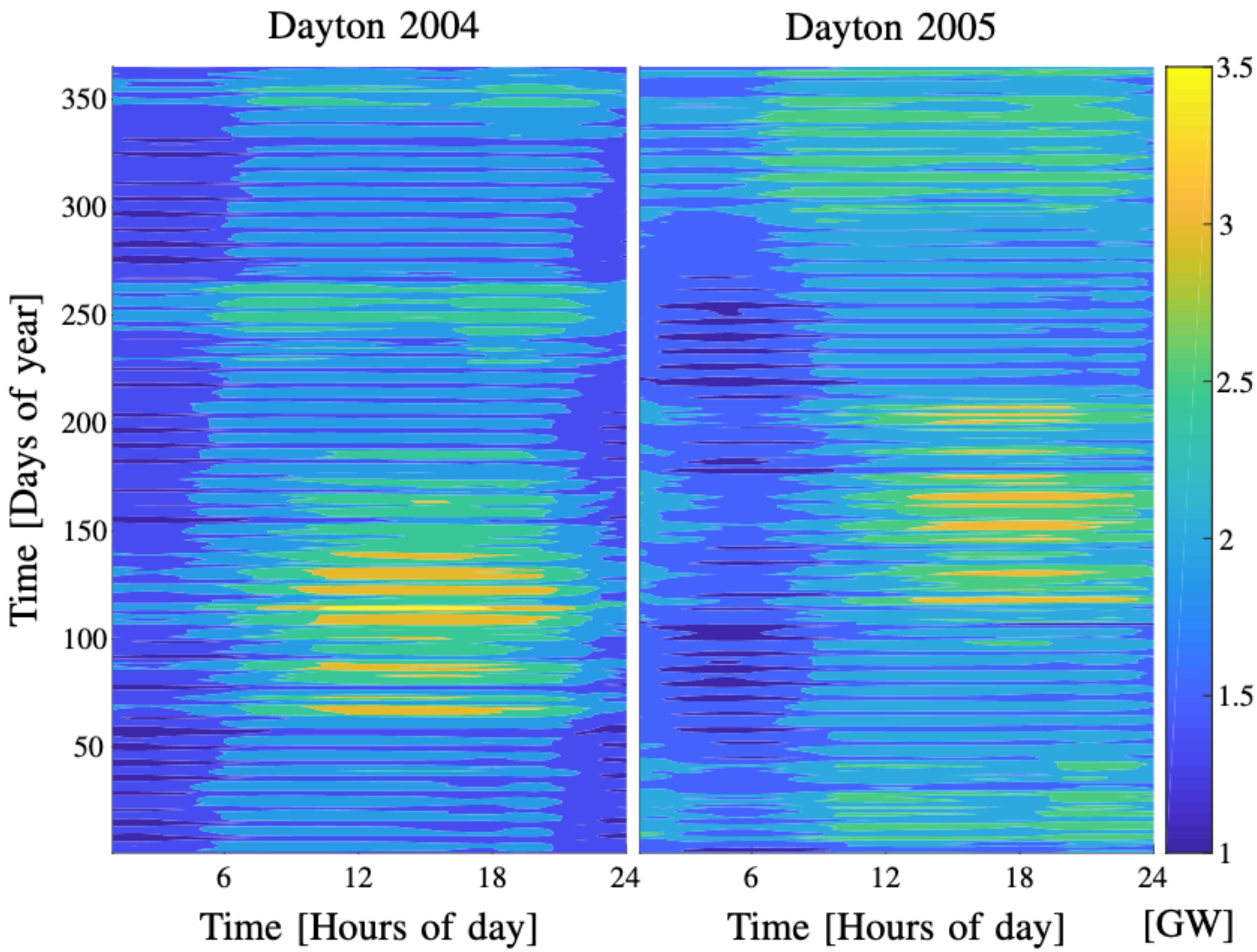
- Load forecasting is important for energy management tasks
- Forecasting methods exploit consumption patterns influenced by multiple factors such as hours of the day, holidays, or weather conditions
- Energy demand is uncertain due to consumer behaviors or unexpected events



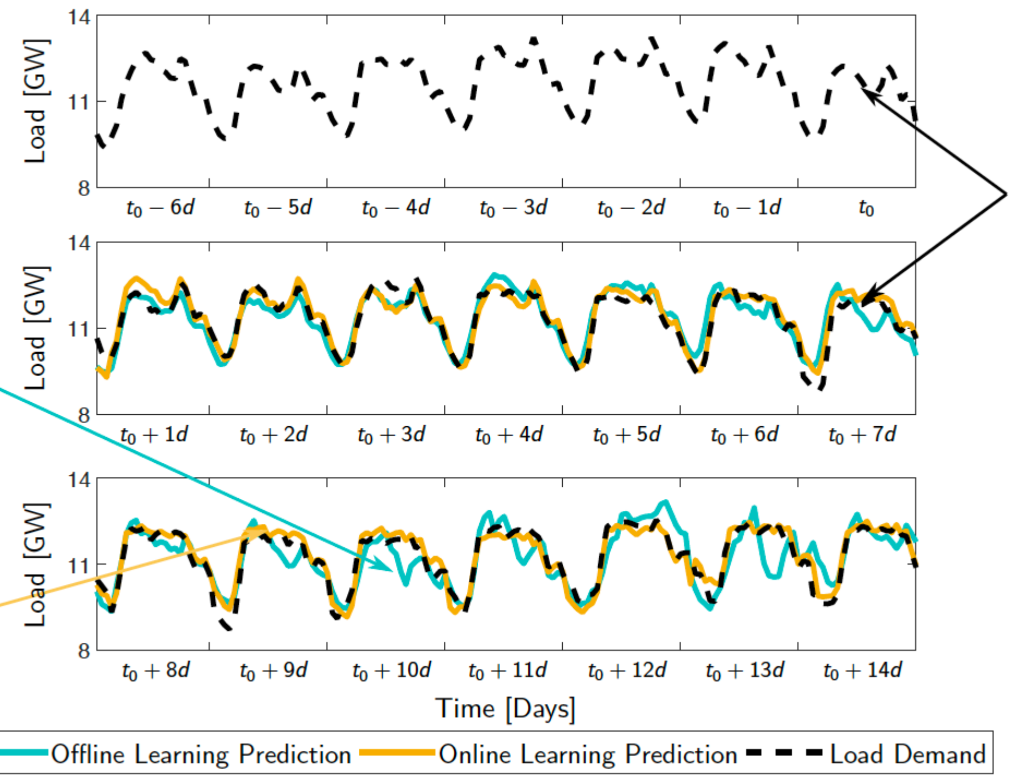
Load consumption patterns change over time



Load consumption patterns change over time



Online learning



Consumptions change from a two-peak pattern to a flatter pattern

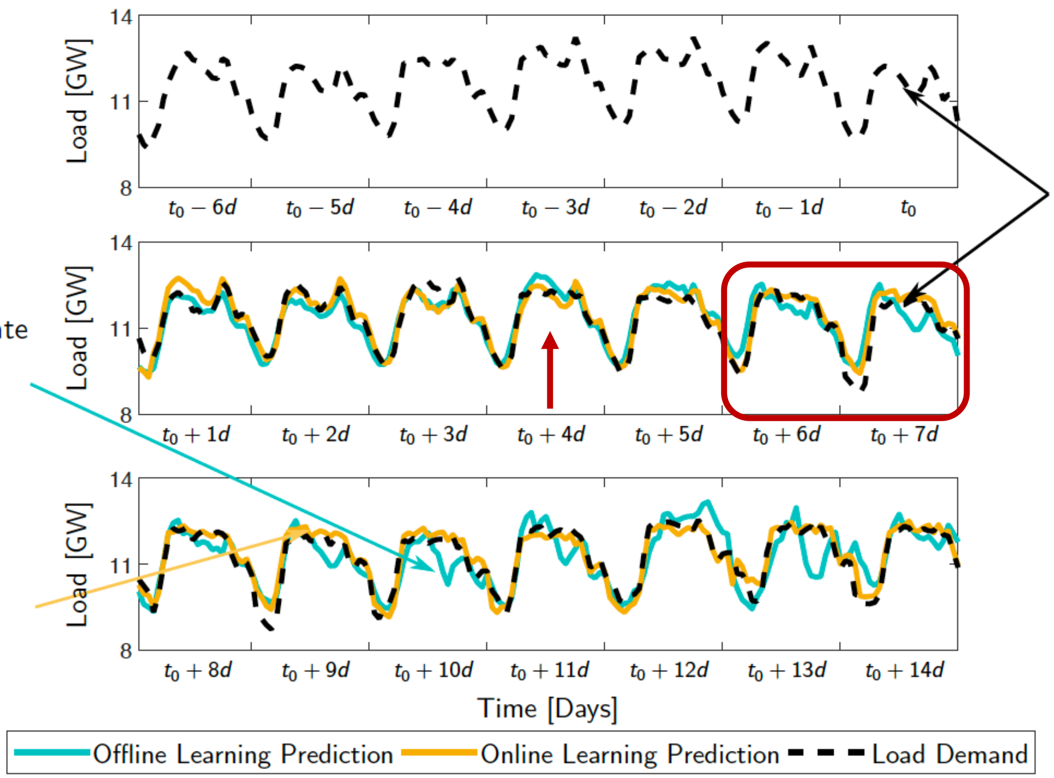
Methods based on offline learning cannot accommodate to such changes

Methods based on online learning correctly adapt models after few days

Online learning

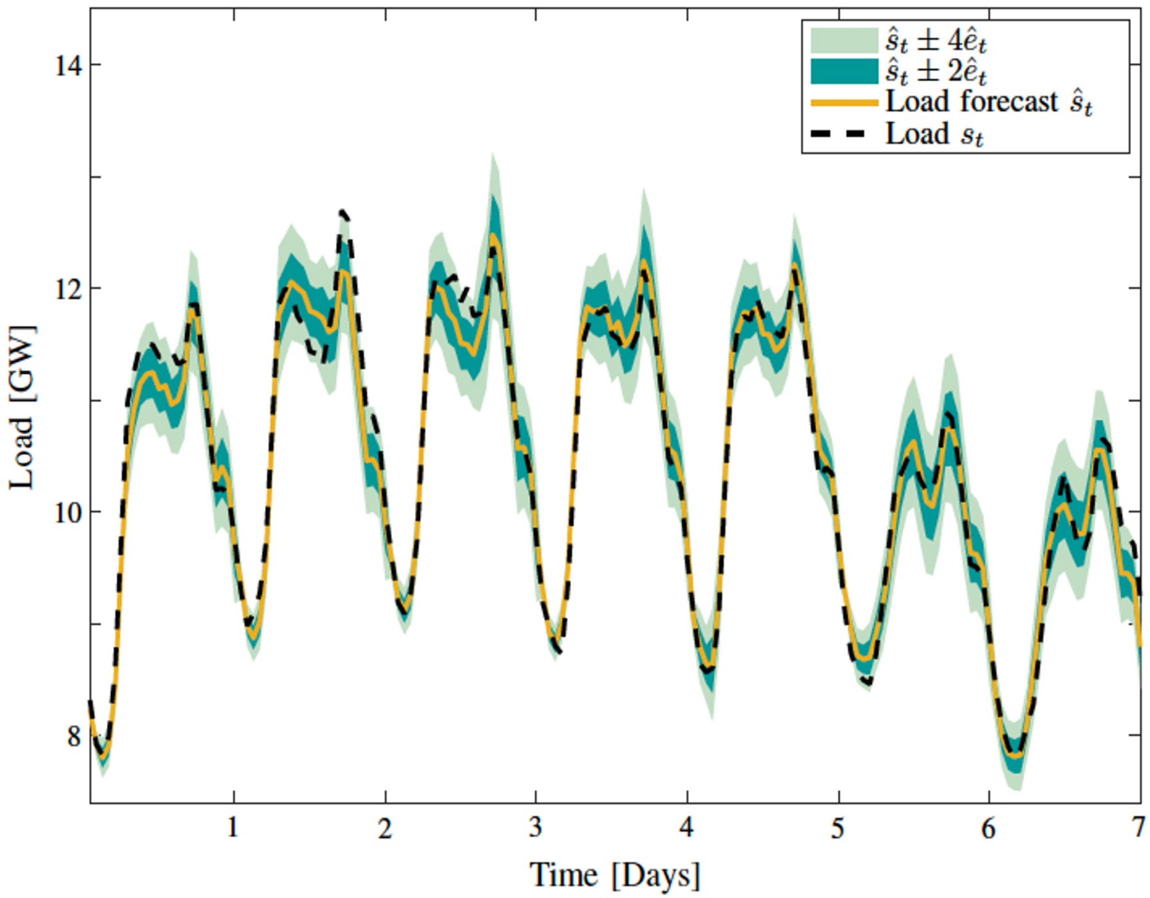
Methods based on offline learning cannot accommodate to such changes

Methods based on online learning correctly adapt models after few days



Consumptions change from a two-peak pattern to a flatter pattern

Probabilistic forecasting techniques



Outline

- Introduction and motivation
- Problem formulation
- Models and theoretical results
- Extension to high-dimensional models
- Experimental results
- Conclusions

Outline

- Introduction and motivation
- Problem formulation

Problem formulation

Forecasting algorithms obtain a vector of L forecasts corresponding with future loads

$$\hat{\mathbf{y}} = [\widehat{s}_{t+1}, \widehat{s}_{t+2}, ..., \widehat{s}_{t+L}]^\top$$

given an instance vector of past loads and observations related to future loads

$$\mathbf{x} = \left[s_{t-m}, ..., s_t, \mathbf{r}_{t+1}^\top, ..., \mathbf{r}_{t+L}^\top\right]^\top, m < t$$

where t is the time when the load is measured.

Problem formulation

Forecasting algorithms obtain a vector of L forecasts corresponding with future loads

$$\hat{\mathbf{y}} = \left[\widehat{s}_{t+1}, \widehat{s}_{t+2}, ..., \widehat{s}_{t+L}\right]^\top$$

given an instance vector of past loads and observations related to future loads

$$\mathbf{x} = \left[s_{t-m}, ..., s_t, \mathbf{r}_{t+1}^\top, ..., \mathbf{r}_{t+L}^\top\right]^\top, m < t$$

where t is the time when the load is measured.

The calendar information

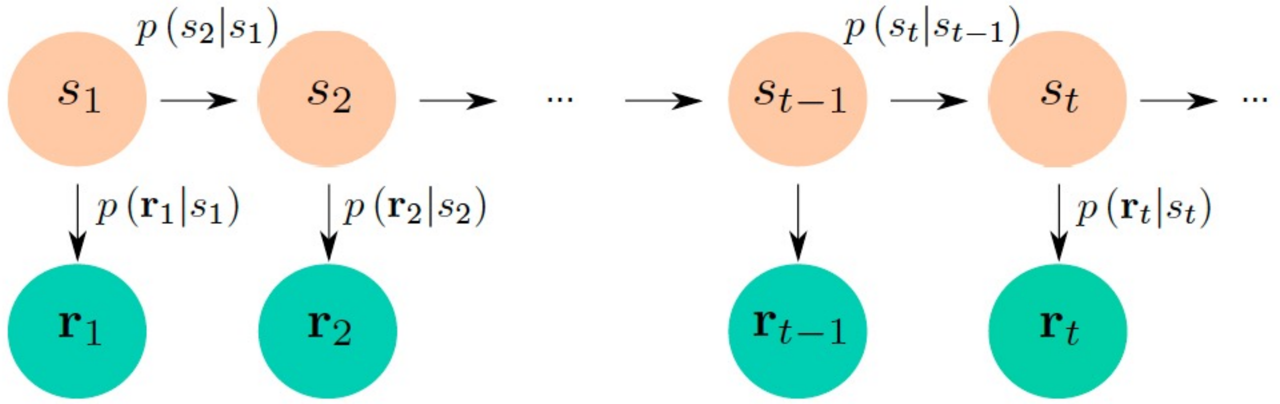
$$c\left(t\right) \in \{1, 2, ..., C\}$$

includes details such as hour of day and holidays.

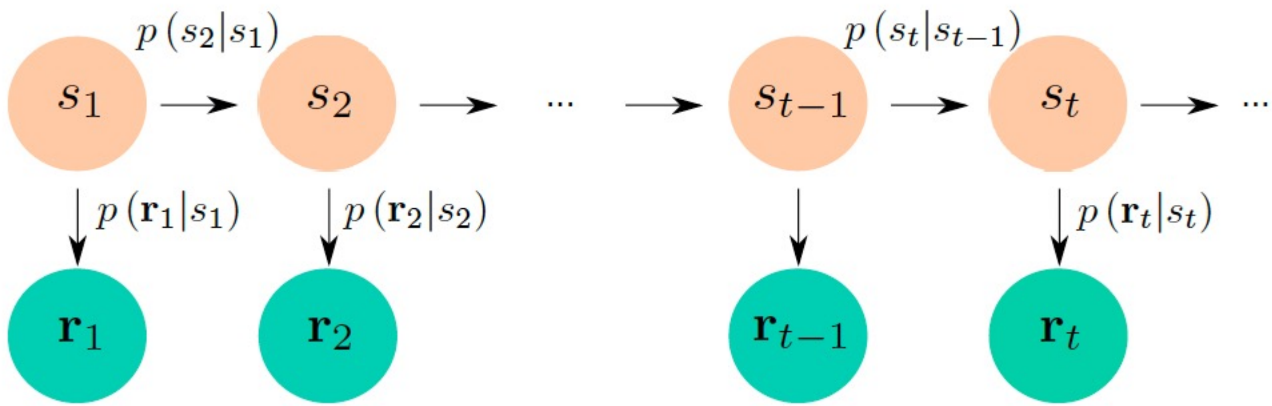
Outline

- Introduction and motivation
- Problem formulation
- Models and theoretical results

Probabilistic Load Forecasting Based on Adaptive Online Learning



Probabilistic Load Forecasting Based on Adaptive Online Learning



Model relationship consecutive loads

$$p(s_t|s_{t-1}) = N\left(s_t; \mathbf{u}_s^{\mathrm{T}} \boldsymbol{\eta}_{s,c}, \sigma_{s,c}\right)$$

Model relationship loads and observations

$$p(\mathbf{r}_t|s_t) \propto p(s_t|\mathbf{r}_t) = N\left(s_t; \mathbf{u}_r^{\mathrm{T}} \boldsymbol{\eta}_{r,c}, \sigma_{r,c}\right)$$

Learning HMM parameters

The HMM describing is characterized by parameters

$$\Theta = \{\boldsymbol{\eta}_{s,c}, \sigma_{s,c}, \boldsymbol{\eta}_{r,c}, \sigma_{r,c} : c = 1, 2, ..., C\}$$

Parameters Θ can be obtained by maximizing the weighted log-likelihood given by

$$L_i\left(\boldsymbol{\eta}, \sigma\right) = \sum_{j=1}^{i} \lambda^{i-j} \log N(s_{t_j}; \boldsymbol{u}_{t_j}^{\top} \boldsymbol{\eta}, \sigma)$$

with $i \leq n$, $\lambda \in (0,1)$ and $s_{t_1}, s_{t_2}, ..., s_{t_n}$ and $\boldsymbol{u}_{t_1}, \boldsymbol{u}_{t_2}, ..., \boldsymbol{u}_{t_n}$ sequences of loads and corresponding feature vectors with the same calendar type.

Learning HMM parameters

The HMM describing is characterized by parameters

$$\Theta = \{\boldsymbol{\eta}_{s,c}, \sigma_{s,c}, \boldsymbol{\eta}_{r,c}, \sigma_{r,c} : c = 1, 2, ..., C\}$$

Parameters Θ can be obtained by maximizing the weighted log-likelihood given by

$$L_i\left(\boldsymbol{\eta}, \sigma\right) = \sum_{j=1}^{i} \boxed{\lambda^{i-j}} \log N(s_{t_j}; \boldsymbol{u}_{t_j}^{\top} \boldsymbol{\eta}, \sigma)$$

with $i \leq n$, $\lambda \in (0, 1)$ and $s_{t_1}, s_{t_2}, ..., s_{t_n}$ and $\boldsymbol{u}_{t_1}, \boldsymbol{u}_{t_2}, ..., \boldsymbol{u}_{t_n}$ sequences of loads and corresponding feature vectors with the same calendar type.

Online learning of HMM parameters

Theorem

Let $\boldsymbol{\eta}_i^* \in \mathbb{R}^K$, $\sigma_i^* \in \mathbb{R}$ parameters that maximize the weighted log-likelihood.
If parameters $\boldsymbol{\eta}_i \in \mathbb{R}^K$, $\sigma_i \in \mathbb{R}$ are given by recursions

$$\boldsymbol{\eta}_i = \boldsymbol{\eta}_{i-1} + \frac{\mathbf{P}_{i-1}\mathbf{u}_{t_i}}{\lambda + \mathbf{u}_{t_i}^\mathsf{T}\mathbf{P}_{i-1}\mathbf{u}_{t_i}}(s_{t_i} - \mathbf{u}_{t_i}^\mathsf{T}\boldsymbol{\eta}_{i-1}),$$

$$\sigma_i = \sqrt{\sigma_{i-1}^2 - \frac{1}{\gamma_i}\left(\sigma_{i-1}^2 - \frac{\lambda\left(s_{t_i} - \mathbf{u}_{t_i}^\mathsf{T}\boldsymbol{\eta}_{i-1}\right)^2}{\lambda + \mathbf{u}_{t_i}^\mathsf{T}\mathbf{P}_{i-1}\mathbf{u}_{t_i}}\right)},$$

with

$$\mathbf{P}_i = \frac{1}{\lambda}\left(\mathbf{P}_{i-1} - \frac{\mathbf{P}_{i-1}\mathbf{u}_{t_i}\mathbf{u}_{t_i}^\mathsf{T}\mathbf{P}_{i-1}}{\lambda + \mathbf{u}_{t_i}^\mathsf{T}\mathbf{P}_{i-1}\mathbf{u}_{t_i}}\right), \qquad \gamma_i = 1 + \lambda\gamma_{i-1},$$

for any $i \leq n$ such that the matrix $\sum_{j=1}^{i}\lambda^{i-j}\mathbf{u}_{t_j}\mathbf{u}_{t_j}^\mathsf{T}$ is not singular.

Probabilistic forecasts

Theorem

If $\{s_t, \mathbf{r}_t\}_{t\geq 1}$ is a HMM characterized by parameters Θ. Then, for $i = 1, 2, \ldots, L$

$$p\left(s_{t+i} | s_t, \mathbf{r}_{t+1}, \ldots, \mathbf{r}_{t+i}\right) = N\left(s_{t+i}; \hat{s}_{t+i}, \hat{e}_{t+i}\right)$$

where $\hat{s}_{t+i}$ and $\hat{e}_{t+i}$ can be computed by the following recursions

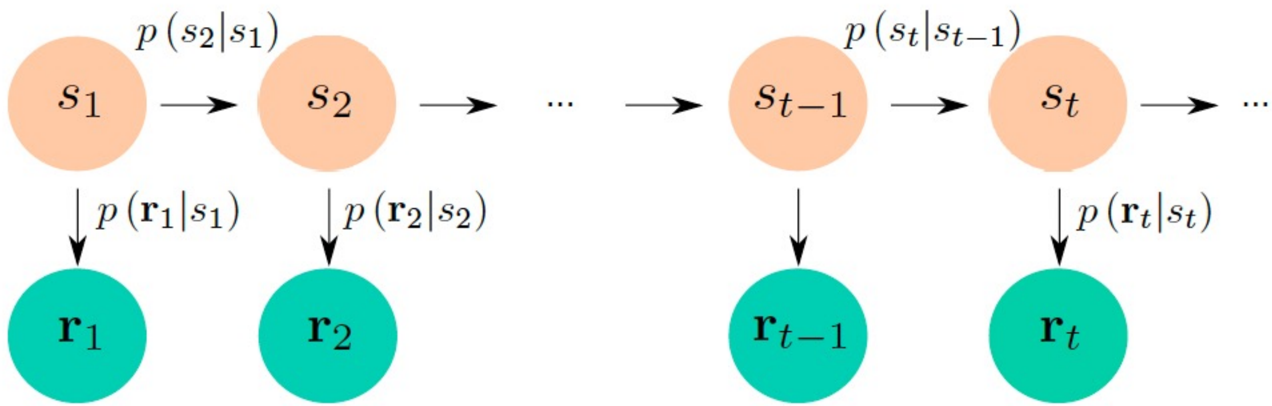
$$c = c\left(t+i\right), \hat{\mathbf{u}}_s = \left[1, \hat{s}_{t+i-1}\right]^{\mathrm{T}}, \mathbf{u}_r = u_r\left(\mathbf{r}_{t+i}\right)$$

$$\hat{s}_{t+i} = \frac{\hat{\mathbf{u}}_s^{\mathrm{T}} \boldsymbol{\eta}_{s,c} \sigma_{r,c}^2 + \mathbf{u}_r^{\mathrm{T}} \boldsymbol{\eta}_{r,c} \left(\sigma_{s,c}^2 + \left([0,1]\,\boldsymbol{\eta}_{s,c}\right)^2 \hat{e}_{t+i-1}^2\right)}{\sigma_{r,c}^2 + \sigma_{s,c}^2 + \left([0,1]\,\boldsymbol{\eta}_{s,c}\right)^2 \hat{e}_{t+i-1}^2}$$

$$\hat{e}_{t+i} = \sqrt{\frac{\left(\sigma_{s,c}^2 + \left([0,1]\,\boldsymbol{\eta}_{s,c}\right)^2 \hat{e}_{t+i-1}^2\right) \sigma_{r,c}^2}{\sigma_{r,c}^2 + \sigma_{s,c}^2 + \left([0,1]\,\boldsymbol{\eta}_{s,c}\right)^2 \hat{e}_{t+i-1}^2}}$$

for $i = 1, 2, \ldots, L$, and $\hat{s}_t = s_t$, $\hat{e}_t = 0$.

Probabilistic Load Forecasting Based on Adaptive Online Learning



Model relationship consecutive loads

$$p(s_t|s_{t-1}) = N\left(s_t; \mathbf{u}_s^{\mathrm{T}} \boldsymbol{\eta}_{s,c}, \sigma_{s,c}\right)$$

Model relationship loads and observations

$$p(\mathbf{r}_t|s_t) \propto p(s_t|\mathbf{r}_t) = N\left(s_t; \mathbf{u}_r^{\mathrm{T}} \boldsymbol{\eta}_{r,c}, \sigma_{r,c}\right)$$

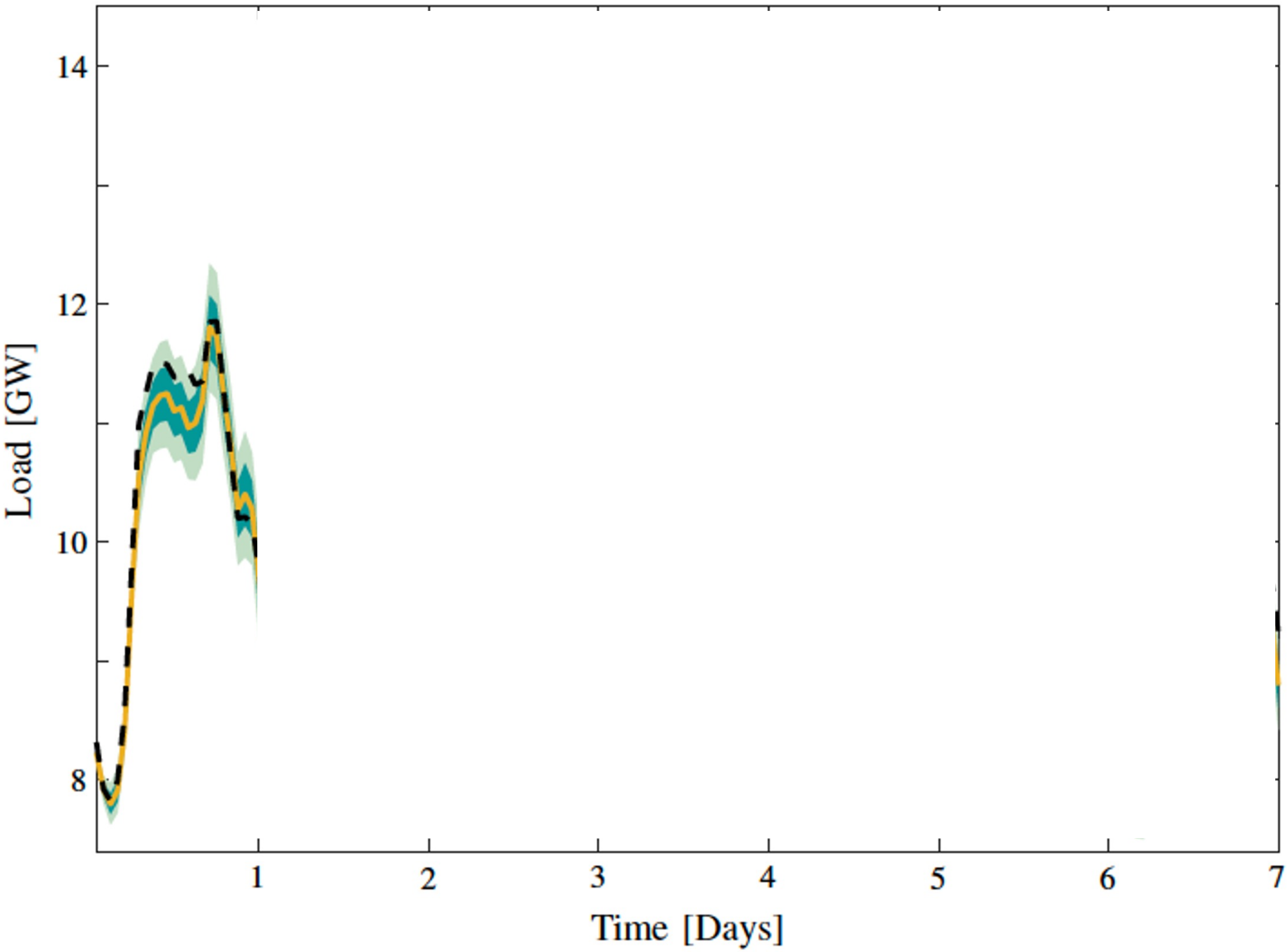
Online learning

$$\boxed{\boldsymbol{\eta}_i} = \boxed{\boldsymbol{\eta}_{i-1}} + \frac{\mathbf{P}_{i-1}\mathbf{u}_{t_i}}{\lambda + \mathbf{u}_{t_i}^{\mathrm{T}} \mathbf{P}_{i-1} \mathbf{u}_{t_i}} \left(s_{t_i} - \mathbf{u}_{t_i}^{\mathrm{T}} \boldsymbol{\eta}_{i-1}\right)$$

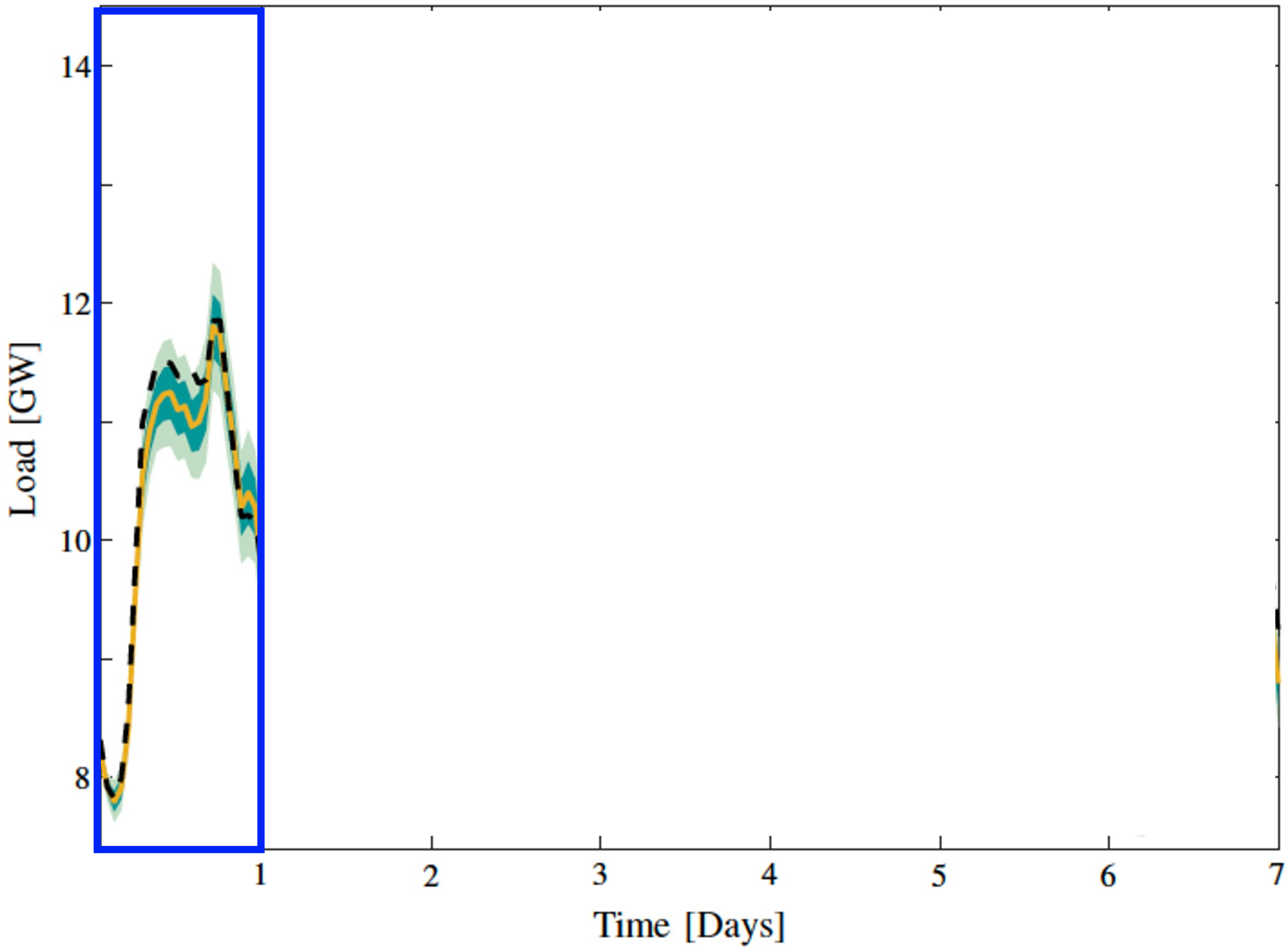
Probabilistic forecasts

$$p(s_{t+i}|s_t, \mathbf{r}_{t+1}, \ldots, \mathbf{r}_{t+i}) = \boxed{N\left(s_{t+i}; \hat{s}_{t+i}, \hat{e}_{t+i}\right)}$$

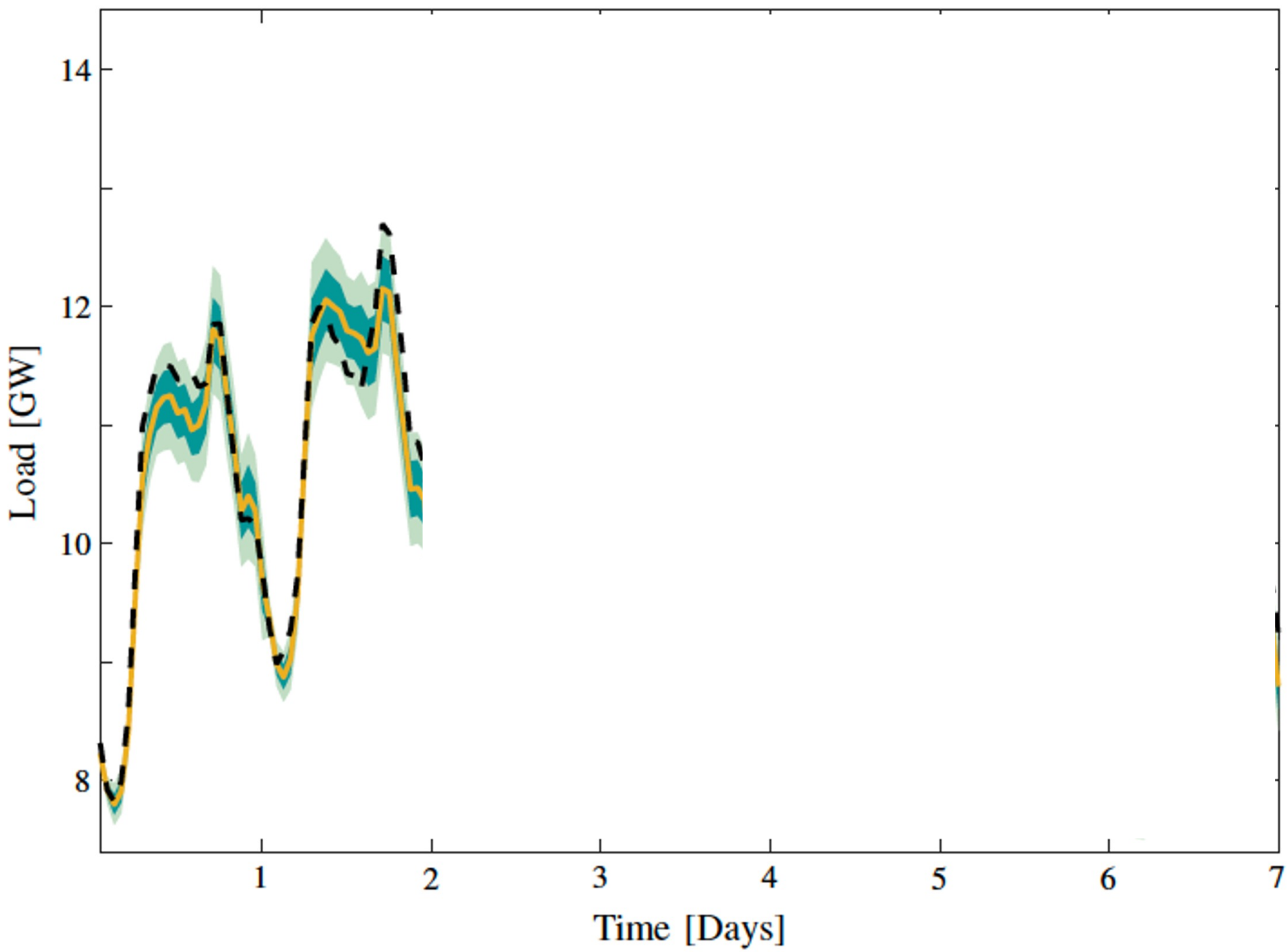
Load forecasting in real-time



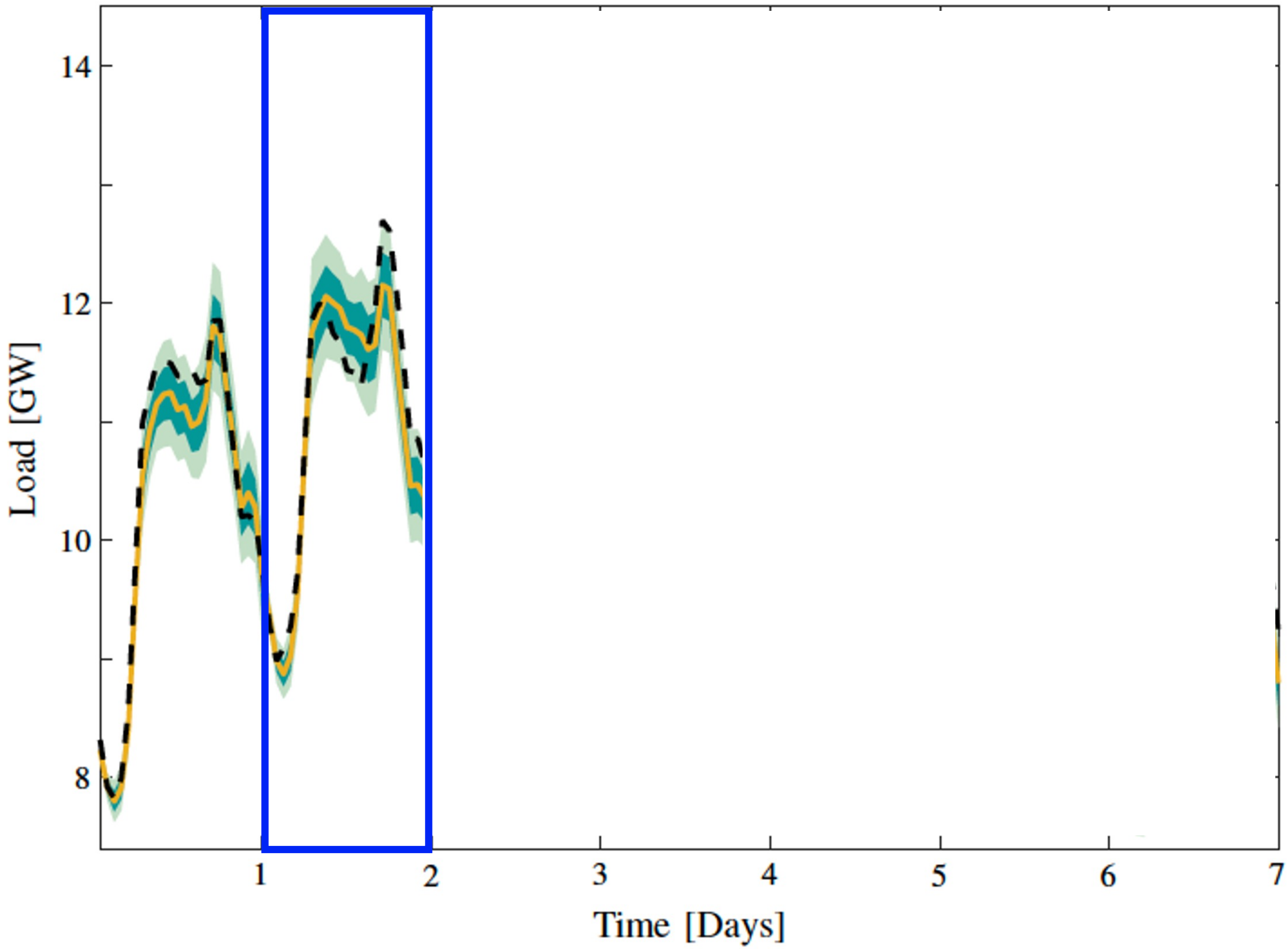
Load forecasting in real-time



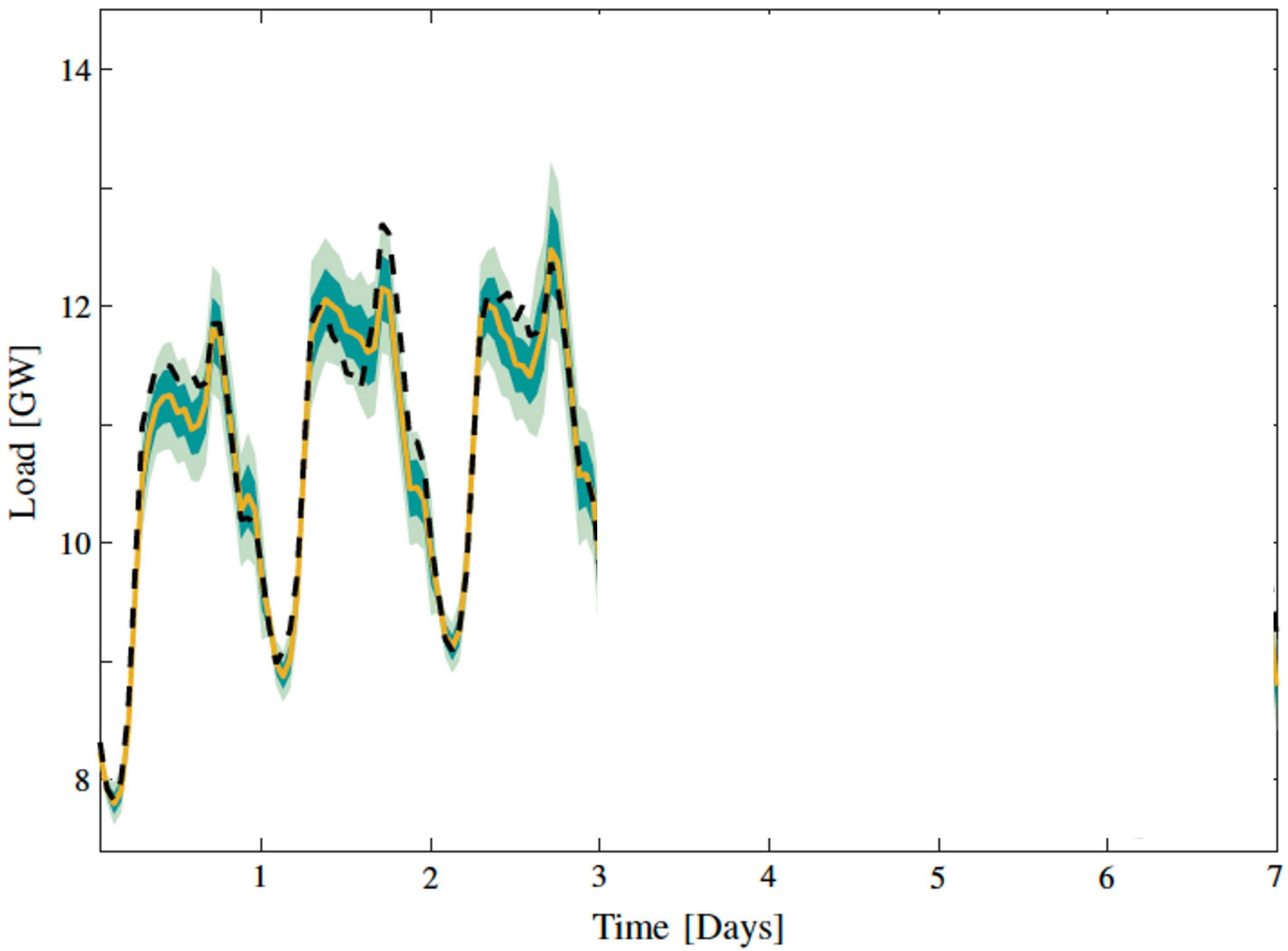
Load forecasting in real-time



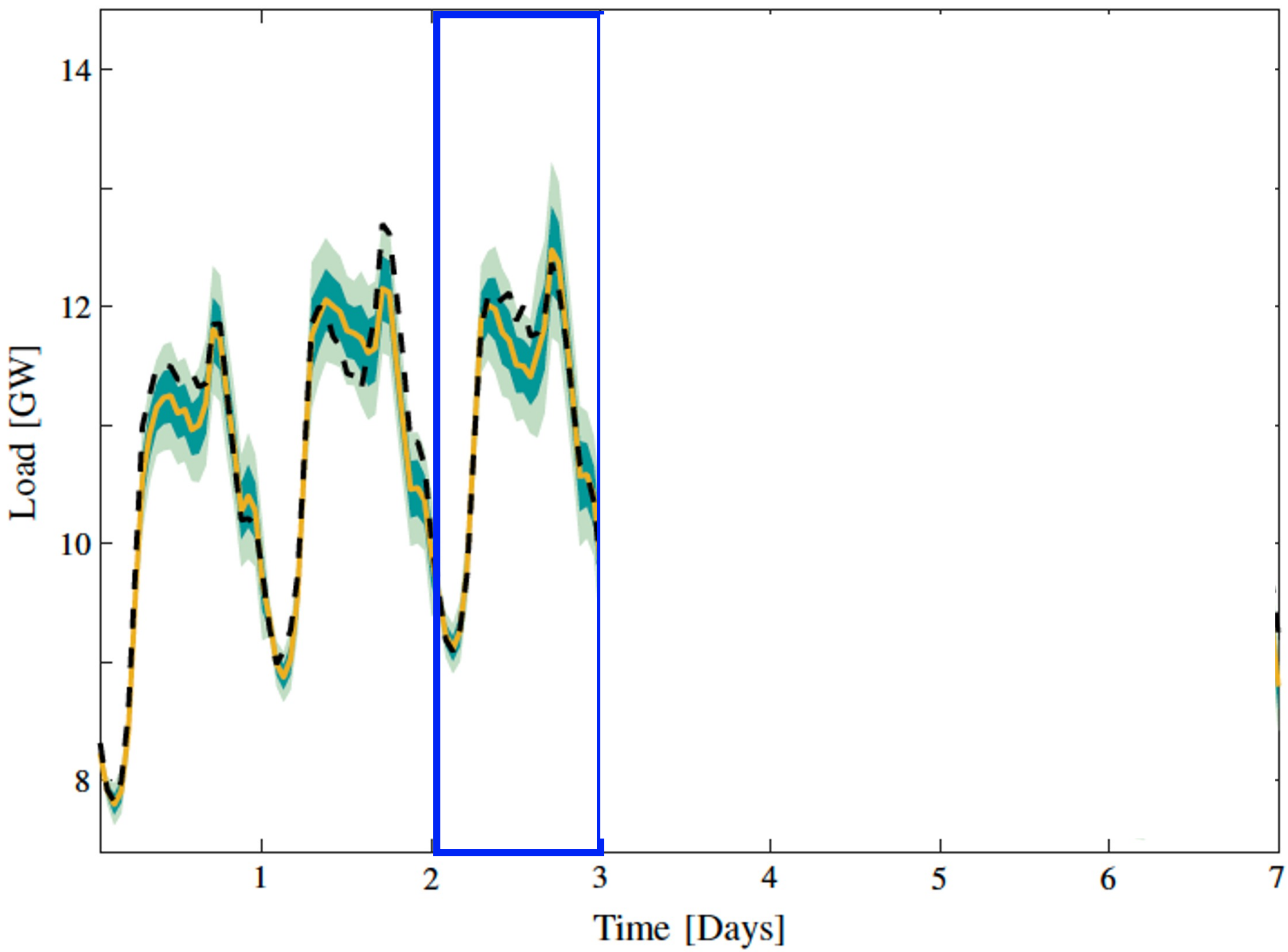
Load forecasting in real-time



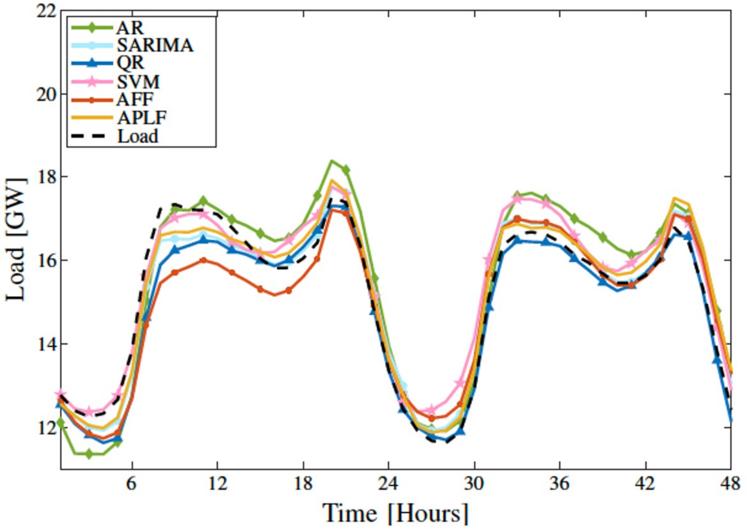
Load forecasting in real-time

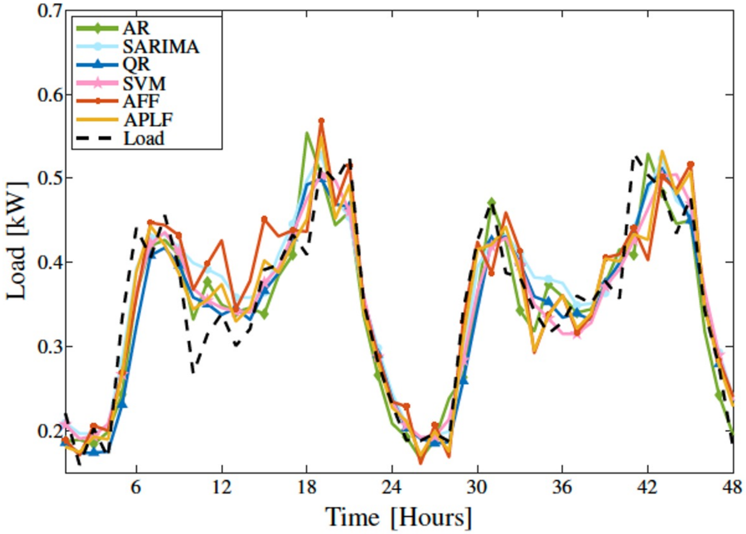


Load forecasting in real-time

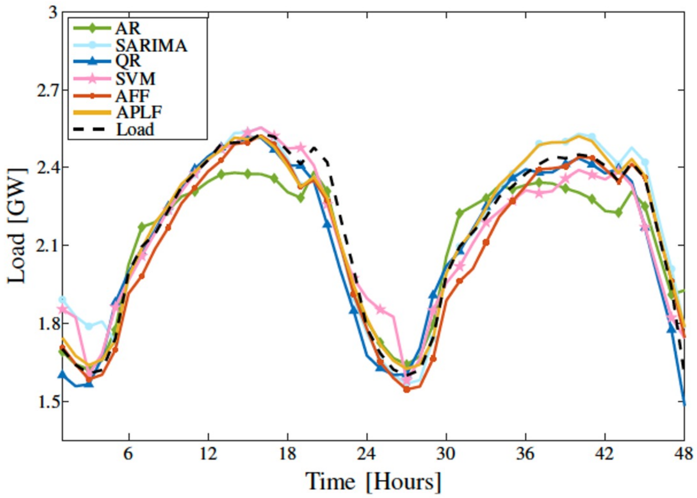


Forecasting in different regions





New England



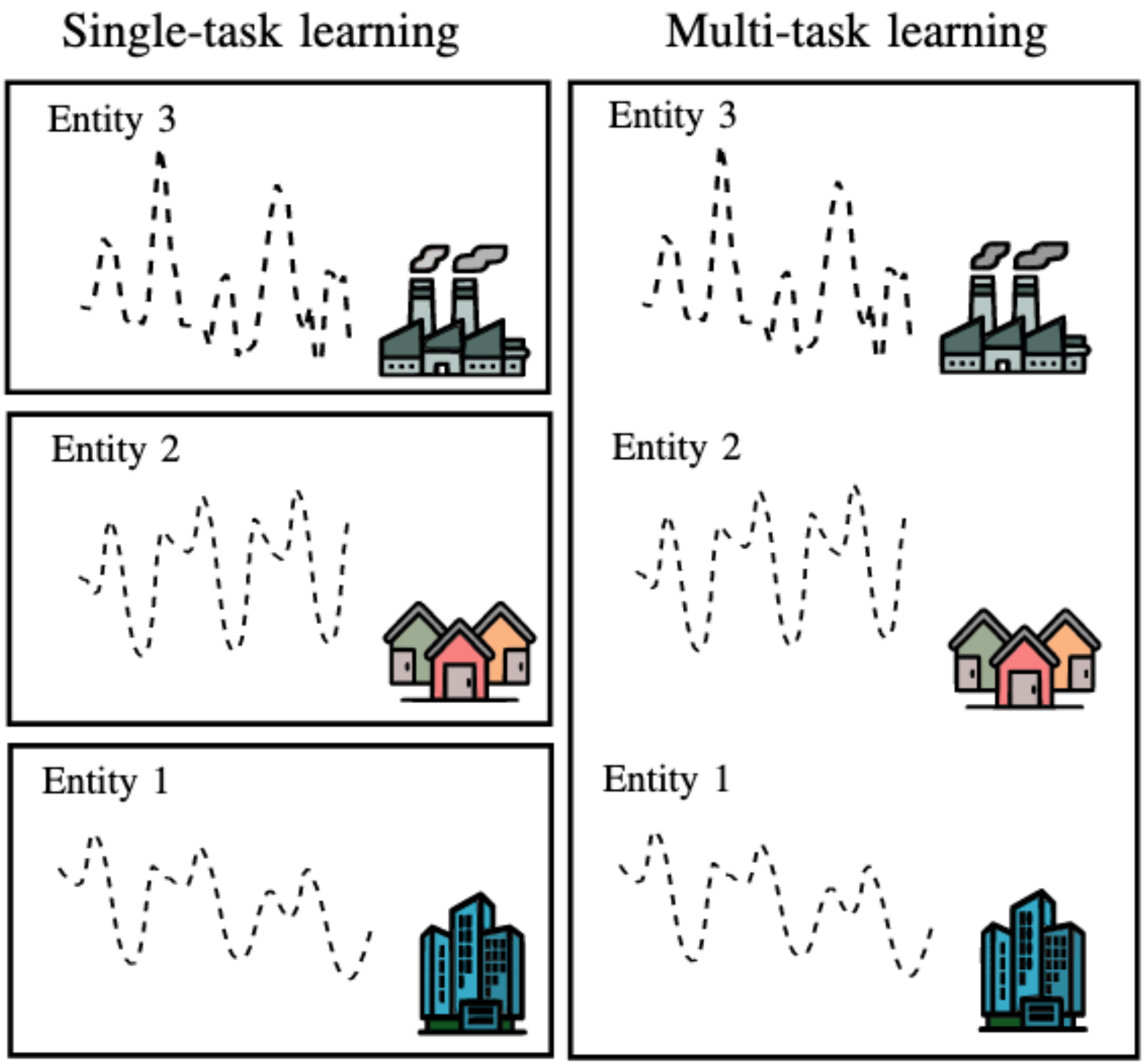
100 buildings

Dayton

Outline

- Introduction and motivation
- Problem formulation
- Models and theoretical results
- Extension to high-dimensional models

Multi-task learning techniques

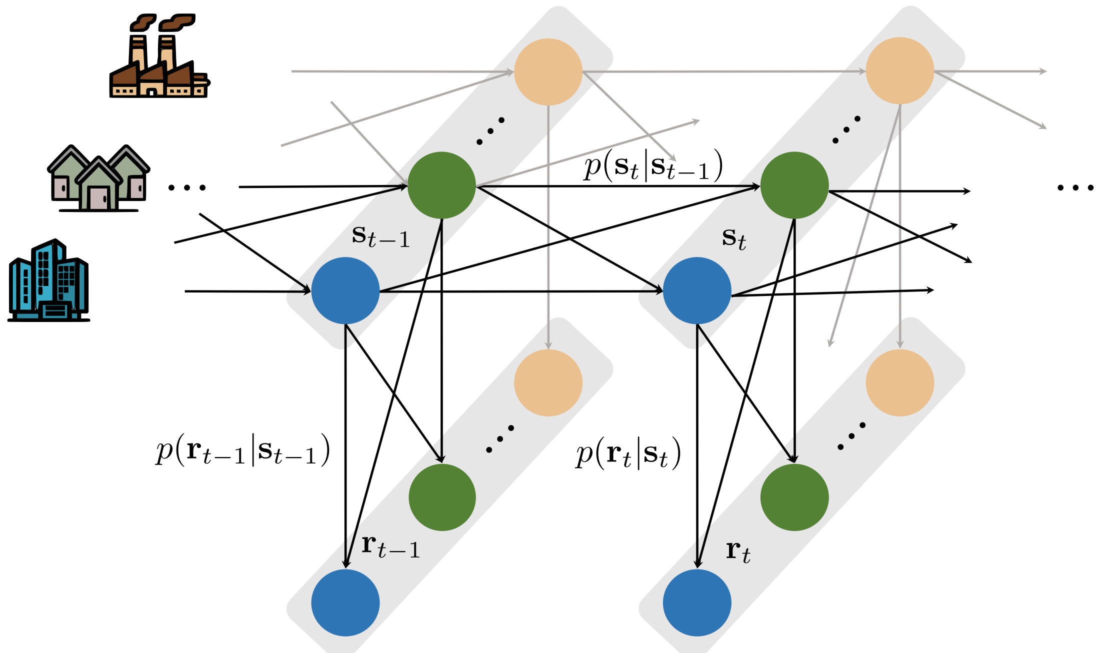


Multi-task Hidden Markov model

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c}) \qquad \text{Loads}$$

$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c}) \qquad \text{Observations}$$

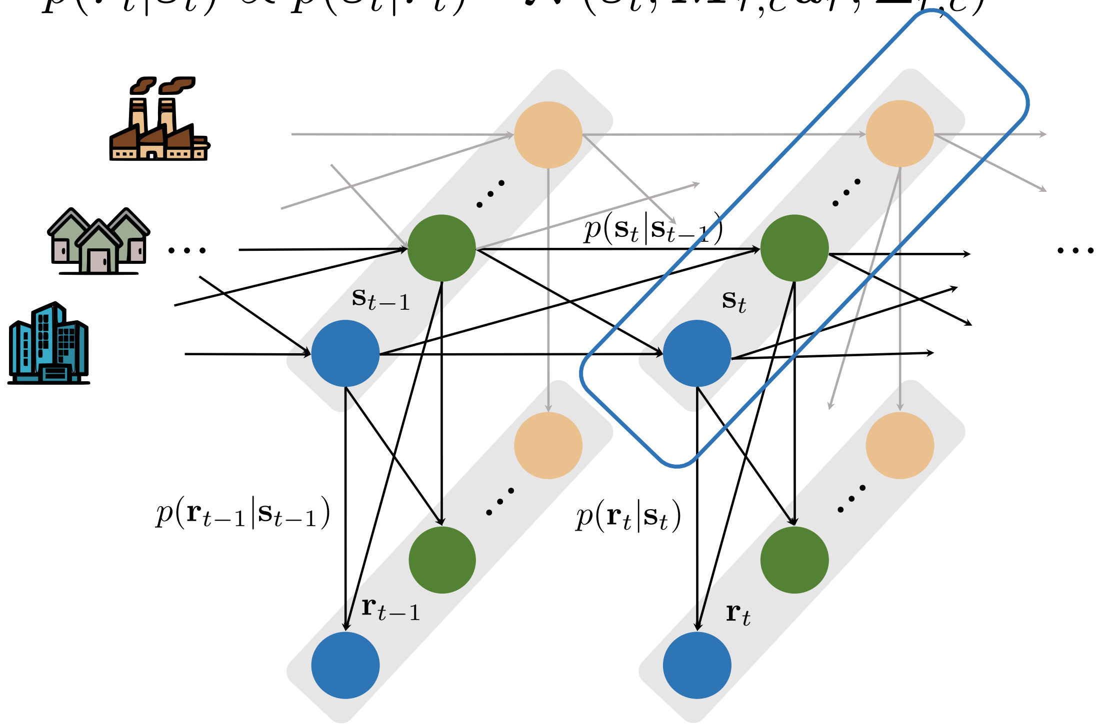


Multi-task Hidden Markov model

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c}) \quad \text{Loads}$$

$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c}) \quad \text{Observations}$$

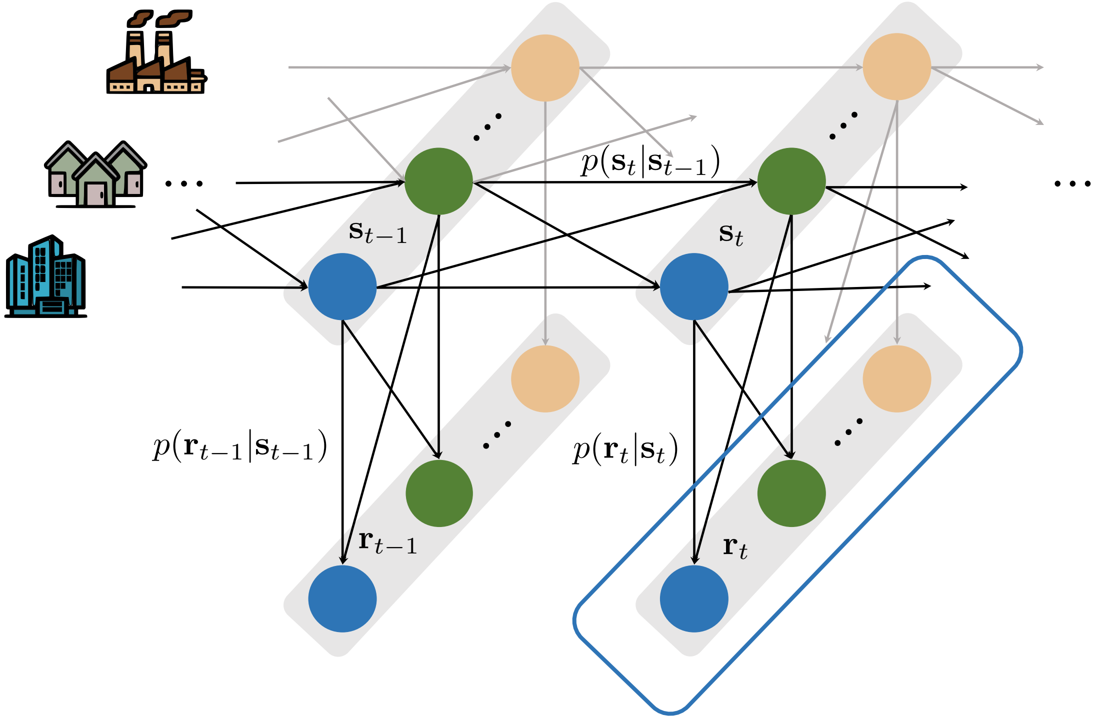


Multi-task Hidden Markov model

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c}) \qquad \text{Loads}$$

$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c}) \qquad \text{Observations}$$

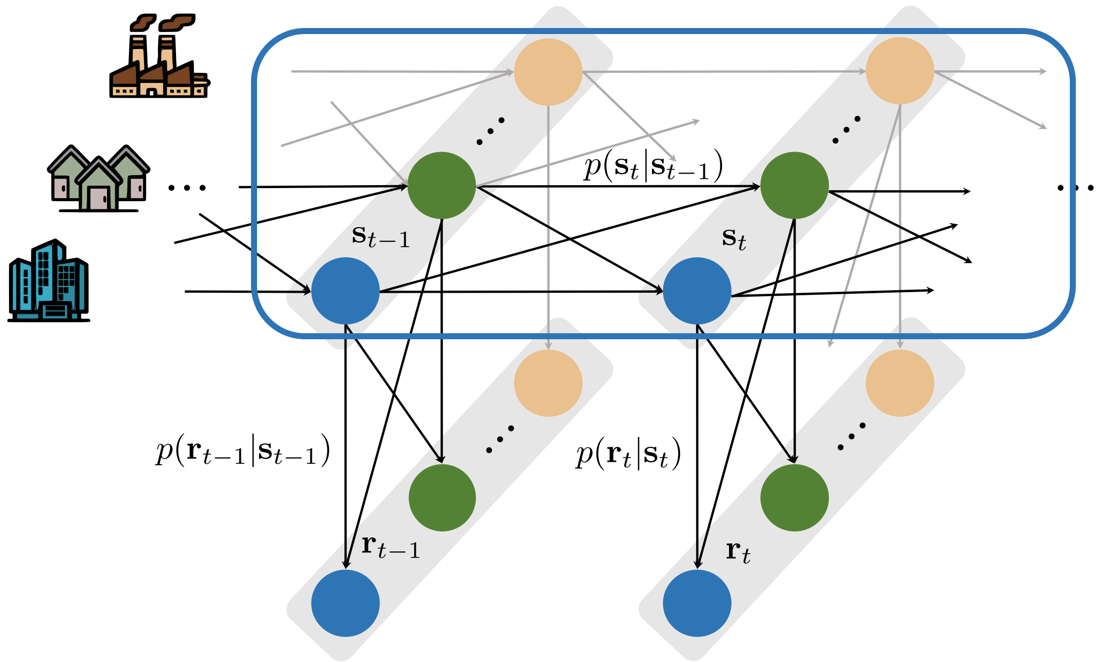


Multi-task Hidden Markov model

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c})$$ Loads

$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c})$$ Observations

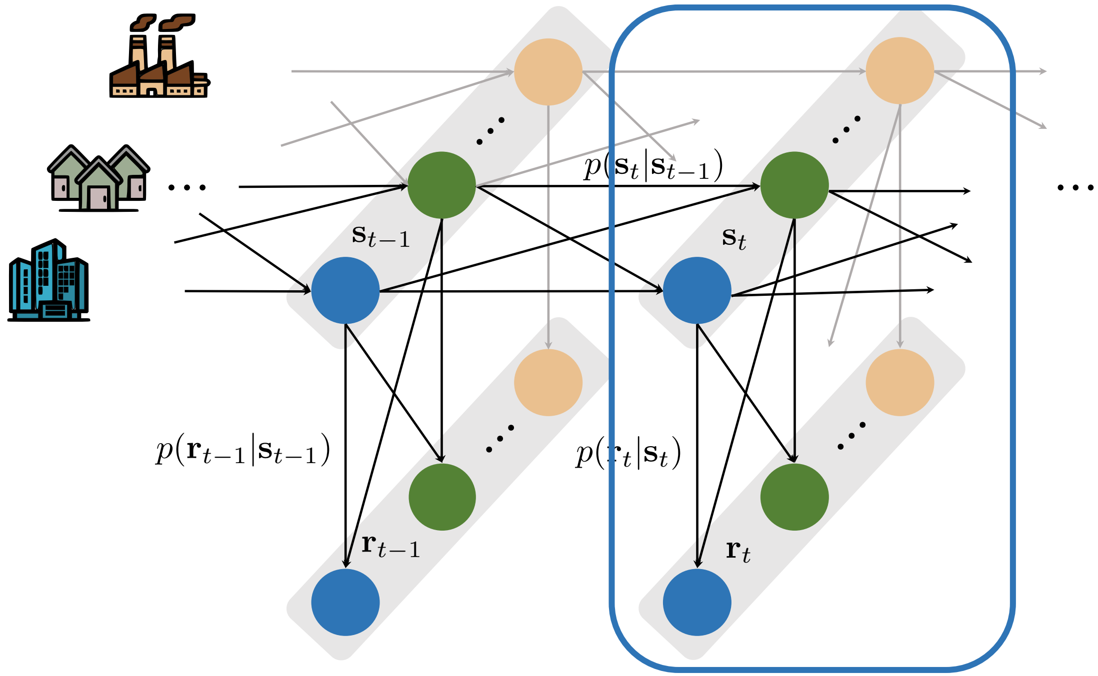


Multi-task Hidden Markov model

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c})$$ Loads

$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c})$$ Observations

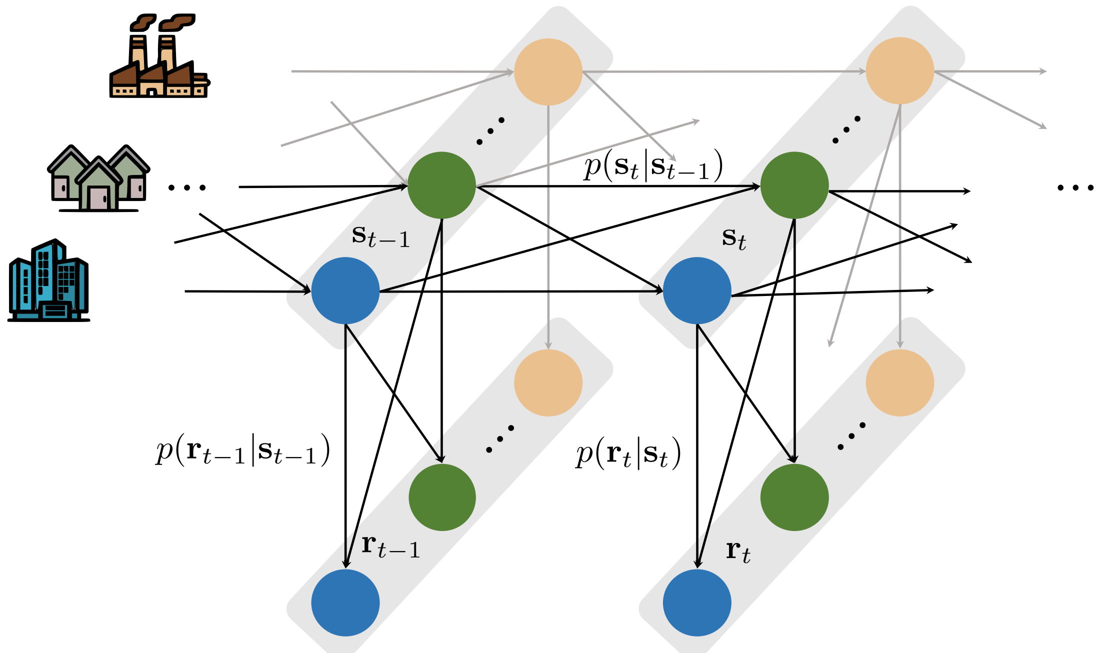


Multi-task Hidden Markov model

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boxed{\boldsymbol{M}_{s,c}}\boldsymbol{u}_s, \boxed{\boldsymbol{\Sigma}_{s,c}}) \qquad \text{Loads}$$

$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boxed{\boldsymbol{M}_{r,c}}\boldsymbol{u}_r, \boxed{\boldsymbol{\Sigma}_{r,c}}) \qquad \text{Observations}$$



The calendar variable describes time factors that may affect load demand

$$c(t) \in \{1, ..., C\}$$

Online learning of multi-task HMM parameters

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c})$$
$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c})$$

Parameters $\boldsymbol{M}_{s,c}$, $\boldsymbol{\Sigma}_{s,c}$, $\boldsymbol{M}_{r,c}$ and $\boldsymbol{\Sigma}_{r,c}$ are given by recursions

$$\boldsymbol{M}_i = \boldsymbol{M}_{i-1} + \frac{(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})\boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}}{\lambda + \boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}\boldsymbol{u}_{t_i}}$$

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i-1} - \frac{1}{\gamma_i}\left(\boldsymbol{\Sigma}_{i-1} - \frac{\lambda^2(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})^\top}{(\lambda + \boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}\boldsymbol{u}_{t_i})^2}\right)$$

Online learning of multi-task HMM parameters

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c})$$
$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c})$$

Parameters $\boldsymbol{M}_{s,c}$, $\boldsymbol{\Sigma}_{s,c}$, $\boldsymbol{M}_{r,c}$ and $\boldsymbol{\Sigma}_{r,c}$ are given by recursions

$$\boldsymbol{M}_i = \boxed{\boldsymbol{M}_{i-1}} + \frac{(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})\boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}}{\lambda + \boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}\boldsymbol{u}_{t_i}}$$

$$\boldsymbol{\Sigma}_i = \boxed{\boldsymbol{\Sigma}_{i-1}} - \frac{1}{\gamma_i}\left(\boxed{\boldsymbol{\Sigma}_{i-1}} - \frac{\lambda^2(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})^\top}{(\lambda + \boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}\boldsymbol{u}_{t_i})^2}\right)$$

Online learning of multi-task HMM parameters

Hidden Markov Model characterized by the following Gaussian distributions:

$$p(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{s,c}\boldsymbol{u}_s, \boldsymbol{\Sigma}_{s,c})$$
$$p(\boldsymbol{r}_t|\boldsymbol{s}_t) \propto p(\boldsymbol{s}_t|\boldsymbol{r}_t) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{M}_{r,c}\boldsymbol{u}_r, \boldsymbol{\Sigma}_{r,c})$$

Parameters $\boldsymbol{M}_{s,c}$, $\boldsymbol{\Sigma}_{s,c}$, $\boldsymbol{M}_{r,c}$ and $\boldsymbol{\Sigma}_{r,c}$ are given by recursions

$$\boldsymbol{M}_i = \boxed{\boldsymbol{M}_{i-1}} + \frac{\boxed{(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})}\boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}}{\lambda + \boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}\boldsymbol{u}_{t_i}}$$

$$\boldsymbol{\Sigma}_i = \boxed{\boldsymbol{\Sigma}_{i-1}} - \frac{1}{\gamma_i}\left(\boxed{\boldsymbol{\Sigma}_{i-1}} - \frac{\lambda^2\boxed{(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})(\boldsymbol{s}_{t_i} - \boldsymbol{M}_{i-1}\boldsymbol{u}_{t_i})^\top}}{(\lambda + \boldsymbol{u}_{t_i}^\top \boldsymbol{P}_{i-1}\boldsymbol{u}_{t_i})^2}\right)$$

Multi-task probabilistic forecasting

Using the most recent learned parameters of the HMM we can provide probabilistic forecasts

$$p(\boldsymbol{s}_{t+i}|\boldsymbol{s}_t, \boldsymbol{r}_{t+1:t+i}) = \mathcal{N}(\boldsymbol{s}_{t+i}; \widehat{\boldsymbol{s}}_{t+i}, \widehat{\boldsymbol{E}}_{t+i}) \quad i = 1, 2, ..., L$$

Multi-task probabilistic forecasting

Using the most recent learned parameters of the HMM we can provide probabilistic forecasts

$$p(\boldsymbol{s}_{t+i}|\boldsymbol{s}_t, \boldsymbol{r}_{t+1:t+i}) = \mathcal{N}(\boldsymbol{s}_{t+i}; \widehat{\boldsymbol{s}}_{t+i}, \widehat{\boldsymbol{E}}_{t+i}) \quad i = 1, 2, ..., L$$

Multi-task probabilistic forecasting

Using the most recent learned parameters of the HMM we can provide probabilistic forecasts

$$p(\boldsymbol{s}_{t+i}|\boldsymbol{s}_t, \boldsymbol{r}_{t+1:t+i}) = \mathcal{N}(\boldsymbol{s}_{t+i}; \widehat{\boldsymbol{s}}_{t+i}, \widehat{\boldsymbol{E}}_{t+i}) \quad i = 1, 2, ..., L$$

where $\widehat{\boldsymbol{s}}_{t+i}$ and $\widehat{\boldsymbol{E}}_{t+i}$ can be computed using the following recursions

$$\widehat{\boldsymbol{s}}_{t+i} = \boldsymbol{W}_1(\boldsymbol{W}_1 + \boldsymbol{W}_2)^{-1}\boxed{\boldsymbol{M}_{r,c}\boldsymbol{u}_r} + \boldsymbol{W}_2(\boldsymbol{W}_1 + \boldsymbol{W}_2)^{-1}\boxed{\boldsymbol{M}_{s,c}\widehat{\boldsymbol{u}}_s}$$

$$\widehat{\boldsymbol{E}}_{t+i} = \boldsymbol{W}_2(\boldsymbol{W}_1 + \boldsymbol{W}_2)^{-1}\boldsymbol{W}_1$$

with

$$\boldsymbol{W}_1 = \boldsymbol{\Sigma}_{s,c} + \boldsymbol{M}_{s,c}\boldsymbol{N}\widehat{\boldsymbol{E}}_{t+i-1}(\boldsymbol{M}_{s,c}\boldsymbol{N})^\top$$

$$\boldsymbol{W}_2 = \boldsymbol{\Sigma}_{r,c}$$

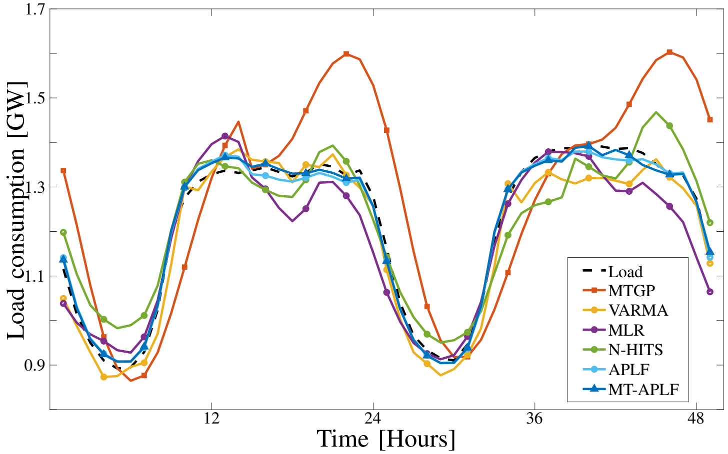
$$c = c(t+i), \widehat{\boldsymbol{u}}_s = [1, \widehat{\boldsymbol{s}}_{t+i-1}]^\top, \boldsymbol{u}_r = u_r(\boldsymbol{r}_{t+i})$$

$$\widehat{\boldsymbol{s}}_t = \boldsymbol{s}_t, \widehat{\boldsymbol{E}}_t = \boldsymbol{0}$$

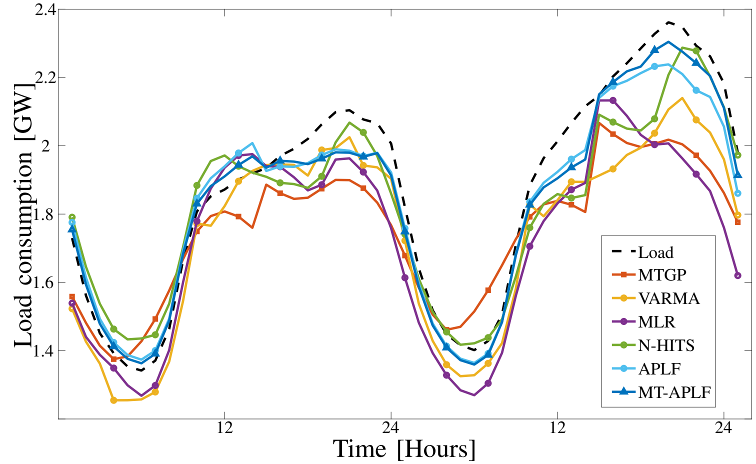
Outline

- Introduction and motivation
- Problem formulation
- Models and theoretical results
- Extension to high-dimensional models
- Experimental results

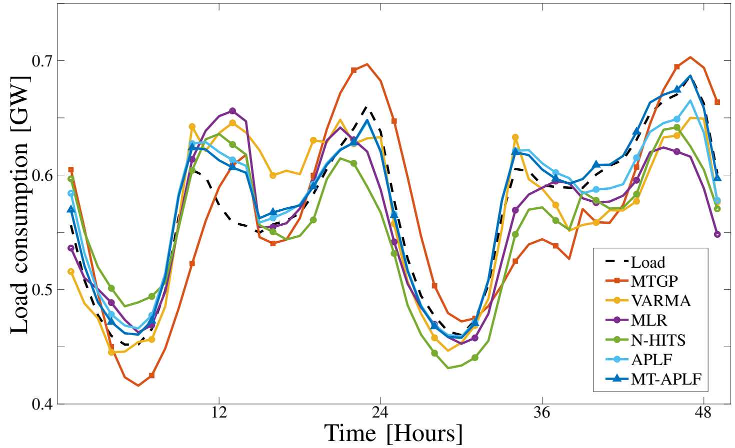
Prediction accuracy in 3 regions of New England



Maine



Vermont



Western Massachusetts

Results

		APLF		MTGP		VAR		MLR		N-HITS		MT-APLF	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
GEFCom2017	Entity 1	5.04	0.25	6.33	0.28	7.07	0.34	11.01	0.62	7.52	0.35	**4.57**	**0.22**
	Entity 2	3.18	**0.05**	4.55	0.07	5.28	0.09	6.87	0.13	4.77	0.07	**3.04**	**0.05**
	Entity 3	4.53	0.19	6.15	0.24	6.98	0.29	8.92	0.40	6.72	0.25	**3.90**	**0.16**
	Entity 4	4.29	0.08	5.95	0.0 1	6.97	0.13	9.83	0.21	6.52	0.11	**3.73**	**0.07**
	Entity 5	4.41	0.06	5.96	0.07	6.57	0.09	10.99	0.16	6.83	0.08	**4.04**	**0.05**
	Entity 6	5.15	0.12	6.34	0.14	7.03	0.17	12.75	0.35	7.16	0.16	**4.43**	**0.10**
	Entity 7	4.64	**0.03**	5.64	0.04	6.48	0.05	8.66	0.07	6.05	0.04	**4.37**	**0.03**
	Entity 8	4.86	0.13	6.43	0.16	7.18	0.19	9.63	0.29	6.29	0.16	**4.06**	**0.10**
	TOTAL	4.51	0.11	5.92	0.14	6.70	0.17	9.83	0.28	6.48	0.15	**4.02**	**0.10**
NewEngland	Entity 1	9.80	0.13	11.34	0.14	12.91	0.16	16.15	0.23	10.47	0.14	**9.29**	**0.12**
	Entity 2	5.58	0.09	6.81	0.10	7.44	0.11	11.50	0.20	7.32	0.11	**5.11**	**0.08**
	Entity 3	11.67	0.07	14.01	0.08	19.68	0.10	31.51	0.15	16.30	0.09	**11.12**	**0.06**
	Entity 4	5.52	0.23	6.95	0.29	7.63	0.32	12.72	0.61	7.33	0.31	5.34	**0.22**
	Entity 5	7.52	0.08	9.07	0.09	9.62	0.11	15.79	0.20	9.54	0.10	**7.00**	**0.07**
	Entity 6	6.16	0.12	7.63	0.15	8.85	0.17	14.81	0.34	8.36	0.16	**6.00**	**0.12**
	Entity 7	8.11	0.22	9.21	0.23	10.32	0.26	18.31	0.50	8.95	0.23	**7.36**	**0.19**
	Entity 8	4.71	0.16	6.03	0.21	6.93	0.24	11.41	0.42	6.42	0.22	4.47	**0.15**
	TOTAL	7.38	0.14	8.88	0.16	10.42	0.18	16.53	0.33	9.34	0.17	**6.96**	**0.13**
PJM	Entity 1	4.08	0.82	6.39	1.13	7.94	1.42	6.16	1.11	5.52	1.04	**3.78**	**0.74**
	Entity 2	5.25	0.95	8.47	1.24	9.20	1.15	7.21	0.97	6.41	0.99	**4.48**	**0.69**
	Entity 3	5.20	0.14	10.05	0.25	9.46	0.22	7.16	0.17	6.94	0.18	**4.65**	**0.11**
	Entity 4	5.10	0.20	7.44	0.27	9.20	0.34	7.32	0.27	6.79	0.27	**4.72**	**0.18**
	Entity 5	5.16	1.04	8.16	1.41	6.46	1.15	6.36	1.15	6.19	1.18	**4.51**	**0.83**
	Entity 6	4.26	0.09	5.43	0.11	6.44	0.12	6.41	0.13	5.48	0.11	**4.12**	**0.08**
	Entity 7	8.78	0.18	12.79	0.23	24.64	0.48	22.51	0.44	10.45	0.20	**8.34**	**0.17**
	Entity 8	4.01	0.44	10.37	1.00	6.62	0.63	5.80	0.59	5.34	0.54	**3.66**	**0.37**
	TOTAL	5.23	0.48	8.64	0.70	9.99	0.69	8.62	0.60	6.64	0.55	**4.78**	**0.40**

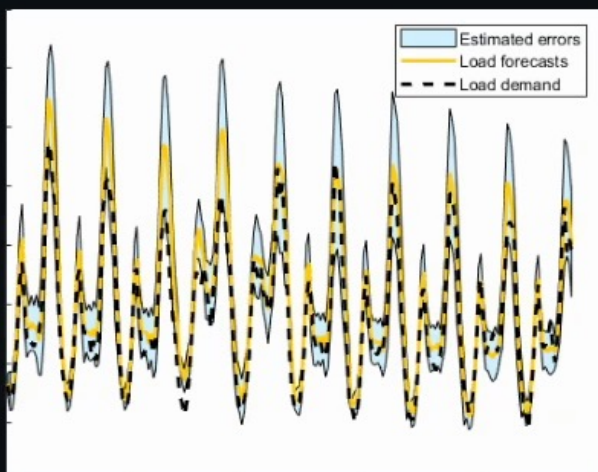
Github with the code and documentation

README License

Probabilistic Load Forecasting based on Adaptive Online Learning (APLF)

License MIT | Made with MATLAB | Made with Python | Ask me anything

This repository contains code for the paper Probabilistic Load Forecasting based on Adaptive Online Learning [1]. We use the implementation details described in the paper.



Implementation of the method

APLF folder contains the Python file, a Jupyter notebook and a Matlab folder that contains all the Matlab scripts required to execute the method:

Outline

- Introduction and motivation
- Problem formulation
- Models and theoretical results
- Extension to high-dimensional models
- Experimental results
- Conclusions

Conclusions

- Proposed multi-task learning method can **learn the relationship** among multiple entities, **adapt to changes** in consumption patterns and **asses load uncertainties**.
- Parameters of the model are updated using a simple **recursive algorithm**, and the method provides **probabilistic predictions** with the most recent parameters
- Experimental results demonstrate that the proposed method achieves **higher performance** compared to existing multi-task learning techniques for load forecasting.