# State Space Model for the Prediction of Energy Consumption

Dexiong Chen, Chia-Man Hung

March 1, 2017

Some statistics on the energy consumptions of buildings:

- 48% of the total energy consumption
- 36% of the total $CO_2$ emission
- 27% of the electric consumption

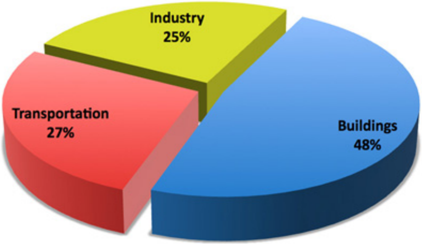


Figure: Energy consumption

Various approaches:

- physics methods
- statistical methods
- neural networks (ANNs)
- support vector machines (SVMs)
- gray model

State space model: simple but extensively used, with theoretical guarantees and confidence intervals.

# Outline

# State Space Model

$$\begin{cases} X_{t+1} = A_t X_t + B_t U_t + \varepsilon_t \\ Y_t = C_t X_t + D_t U_t + \eta_t \end{cases} \tag{1}$$

- the error terms $\varepsilon_t \sim \mathcal{N}(0, P_t)$ and $\eta_t \sim \mathcal{N}(0, Q_t)$ are two independent vector-valued i.i.d. Gaussian sequences.
- $X$ represents state vector sequence, which is not observable.
- $X_0$ is assumed to be $\mathcal{N}(\mu_0, \Sigma_0)$, which is independent of $\varepsilon_t$ and $\eta_t$.
- $\theta_t = (A_t, B_t, C_t, D_t, P_t, Q_t, \mu_0, \Sigma_0)$ are parameters of the model.

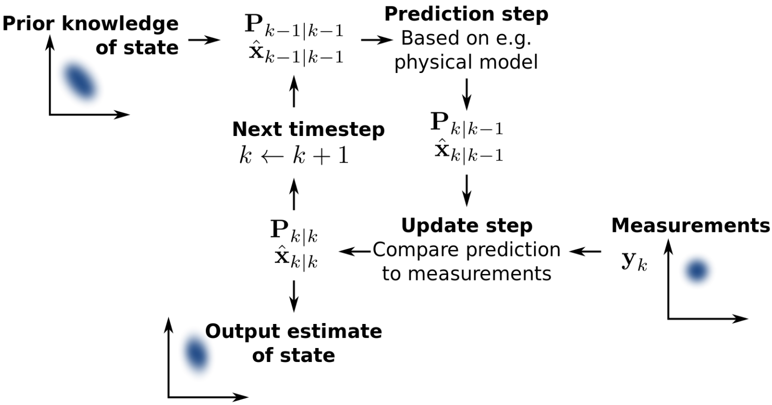


Figure: Scheme for Kalman filter

- $\hat{X}_{t|s} := \mathbb{E}[X_t | Y_{0:s}]$
- $\Sigma_{t|s} := \mathbb{E}[(X_t - \hat{X}_{t|s})(X_t - \hat{X}_{t|s})^T] = \text{Cov}(X_t - \hat{X}_{t|s})$

By induction, we obtain

$$\hat{X}_{t|t-1} = A_t \hat{X}_{t-1|t-1} + B_t U_t \tag{2}$$

$$\Sigma_{t|t-1} = A_t \Sigma_{t-1|t-1} A_t^T + P_t \tag{3}$$

$$K_t = \Sigma_{t|t-1} C_t^T [C_t \Sigma_{t|t-1} C_t^T + Q_t]^{-1} \tag{4}$$

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t (Y_t - C_t \hat{X}_{t|t-1} - D_t U_t) \tag{5}$$

$$\Sigma_{t|t} = [I - K_t C_t] \Sigma_{t|t-1} \tag{6}$$

We intend to compute $\hat{X}_{t|n}$ et $\Sigma_{t|n}$ via a backward recursion.

$$J_t = \Sigma_{t|t} A_t^T \Sigma_{t+1|t}^{-1} \tag{7}$$

$$\hat{X}_{t|n} = \hat{X}_{t|t} + J_t(\hat{X}_{t+1|n} - \hat{X}_{t+1|t}) \tag{8}$$

$$\Sigma_{t|n} = \Sigma_{t|t} + J_t(\Sigma_{t+1|n} - \Sigma_{t+1|t}) J_t^T \tag{9}$$

By analogy, we can compute the one-lag covariance

$$\Sigma_{t,t-1|n} = \Sigma_{t|n} J_{t-1}^T \tag{10}$$

These estimates about hidden states will be used in the E-step of the EM algorithm.

EM algorithm allows to estimate the parameters while some hidden variables present in the model.

$$L(\theta) = \int f(x;\theta)\,\lambda(dx) \qquad \ell(\theta) = \log L(\theta) \tag{11}$$

The main idea is to consider an intermediate quantity, referred to as a conditional expectation.

$$\mathcal{Q}(\theta;\theta') = \int \log f(x;\theta) p(x;\theta')\,\lambda(dx) \tag{12}$$

We rewrite $\mathcal{Q}$ as

$$\mathcal{Q}(\theta;\theta') = \ell(\theta) - \mathcal{H}(\theta;\theta'), \tag{13}$$

where

$$\mathcal{H}(\theta;\theta') = -\int \log p(x;\theta)p(x;\theta')\,\lambda(dx) \tag{14}$$

By observing that $\mathcal{H}$ is the Kullback-Leibler divergence, we have the following theorem

**Theoreme**

*Under some assumptions, for any $(\theta,\theta') \in \Theta^2$, we have*

$$\ell(\theta) - \ell(\theta') \geq \mathcal{Q}(\theta;\theta') - \mathcal{Q}(\theta';\theta') \tag{15}$$

EM algorithm:

- Expectation step (E-step): compute $\mathcal{Q}(\theta;\theta')$.
- Maximization step (M-step): choose $\theta^{i+1}$ to maximize the immediate $\theta^{i+1} = \arg\max_{\theta\in\Theta} Q(\theta;\theta^i)$.

We define here

$$\mathcal{Q}(\theta;\theta^i) = \mathbb{E}[\log P(X_{0:n}, Y_{0:n};\theta)|Y_{0:n},\theta^i] \tag{16}$$

Then we can compute explicitly $\mathcal{Q}(\theta;\theta^i)$ and its derivatives with respect to each parameter, then express $\theta^{i+1}$ as a function of $\theta^i$.

# Problem and Data

$$\begin{cases} X_{t+1} = A_t X_t + B_t U_t + \varepsilon_t \\ Y_t = C_t X_t + D_t U_t + \eta_t \end{cases} \tag{17}$$

- $U_t$: exogenous data (temperature, sunshine etc)
- $X_t$: hidden chain
- $Y_t$: energy use measured by sensors

Missing values in $U$: prefilled by using a state space model.

We apply directly the model to the problem

$$\begin{cases} X_{t+1} = AX_t + \varepsilon \\ Y_t = CX_t + DU_t + \eta \end{cases} \tag{18}$$

# Strategy 1: One Regime Model

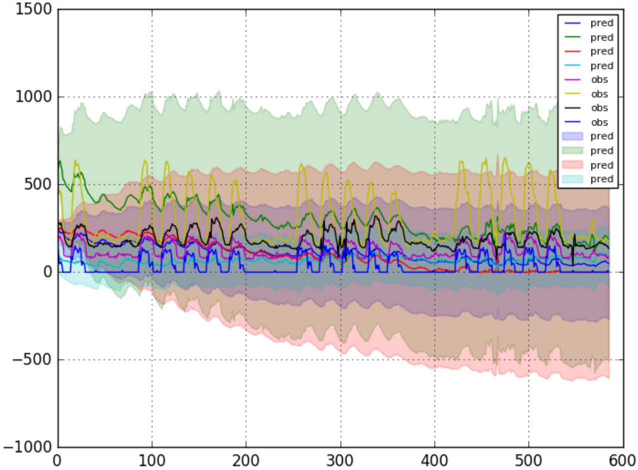


Figure: Building 4

# Strategy 2: Multi-regime Model

$$\begin{cases} X_{t+1} = A_{\sigma(t)} X_t + \varepsilon \\ Y_t = C_{\eta(t)} X_t + D_{\eta(t)} U_t + \eta \end{cases} \tag{19}$$

We divided states into different groups according to their regime. It is a particular case of time-variant state-space model.

- An instance choice of regimes: “night”, “day”, “day-to-night”, “night-to-day” and “weekend”.
- Remark: when applying the EM algorithm for this model, we compute the empirical average respectively over each regime period to estimate parameters belonging to this regime.

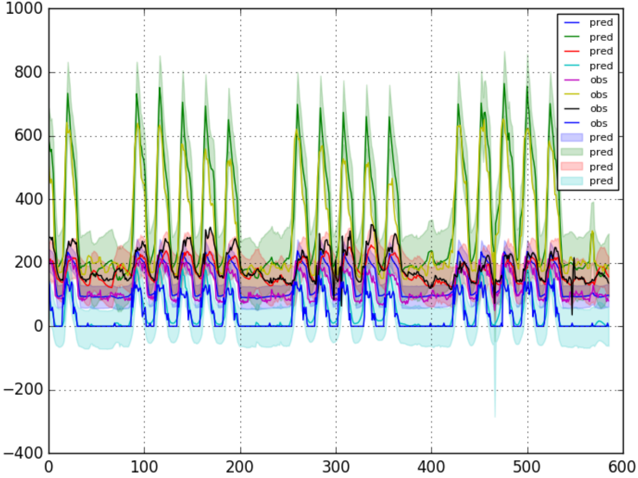


Figure: Building 4

We replace $U_t$ by $\tilde{U}_t = (U_t, U_{t-1})$ in the previous model,

$$\begin{cases} X_{t+1} = A_{\sigma(t)} X_t + \varepsilon \\ Y_t = C_{\eta(t)} X_t + D_{\eta(t)} \tilde{U}_t + \eta \end{cases} \tag{20}$$

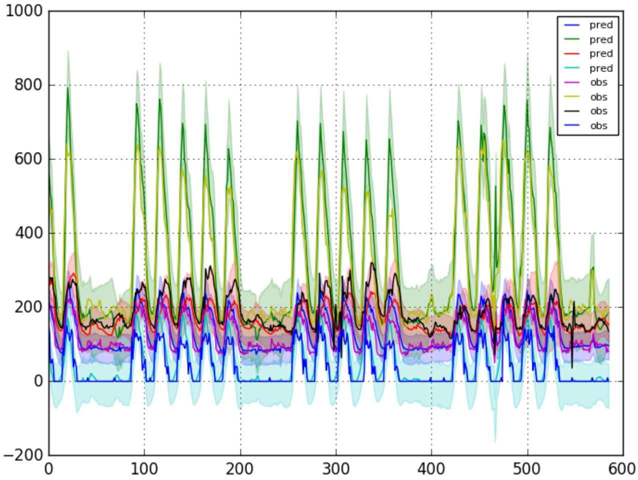


Figure: Building 4

# Comparisons

| building | strategy 1 | strategy 2 | strategy 3 |
|---|---|---|---|
| 1 | 35693 | 8130 | 8043 |
| 2 | 22737 | 5323 | 4568 |
| 3 | 17371 | 8825 | 6491 |
| 4 | 48977 | 6007 | 5032 |

Table: The mean square error on each building using different strategies.

# Conclusion

## Contributions

- investigate and develop a linear state-space model for the problem of energy consumption prediction.
- elaborate two variant from the original state-space model: multi-regime model and two-lag multi-regime model.
- three strategies reflect the track of improvement in our exploration

## Futur work

- prediction accuracy on the test data is not as good as that on the validation data. Solution: model the exogenous observations by another state space model, and estimate the parameters simultaneously for thermal and exogenous model.
- extend to non-linear model, using extended or unscented Kalman filter and EM algorithm.
- adaptive learning of regime function, without using fixed regime periods.
- more historical exogenous observations can be taken into account in the model.