Modelling Electricity Demand in Smart Grids: Data, Trends and Use Cases

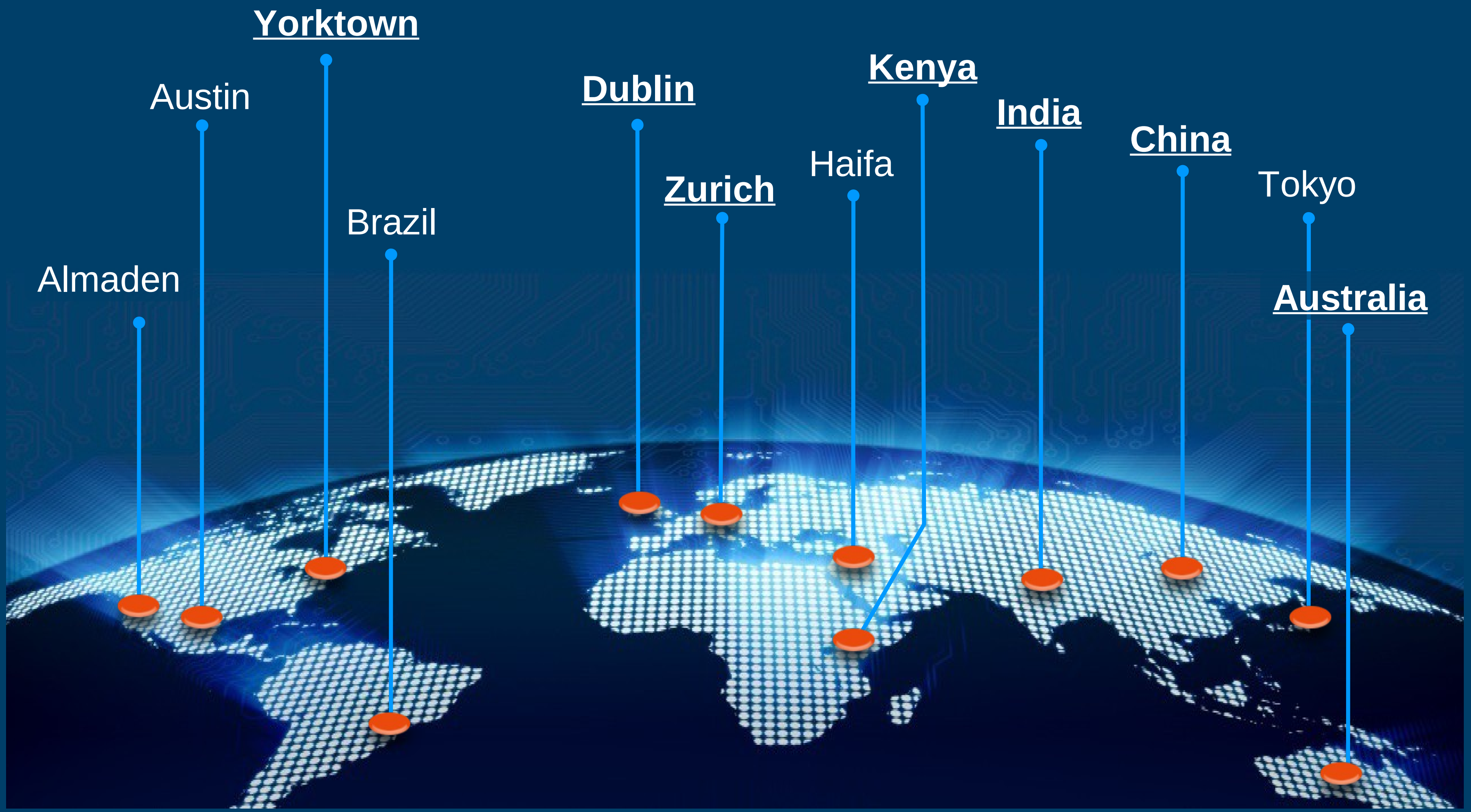
IRSDI-EDF Event

Paris-Saclay, France, 20 October 2017

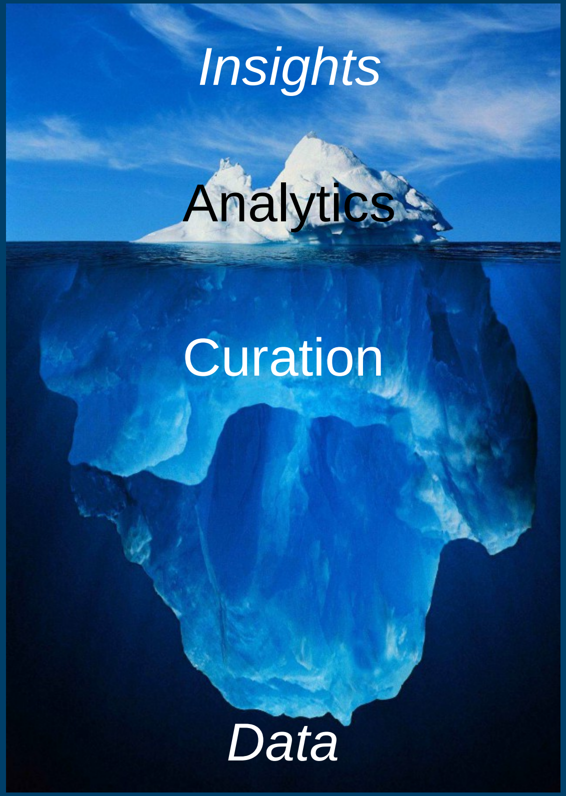
Bei Chen, IBM Research – Ireland

This slide deck has been modified for online publication.

IBM Research labs



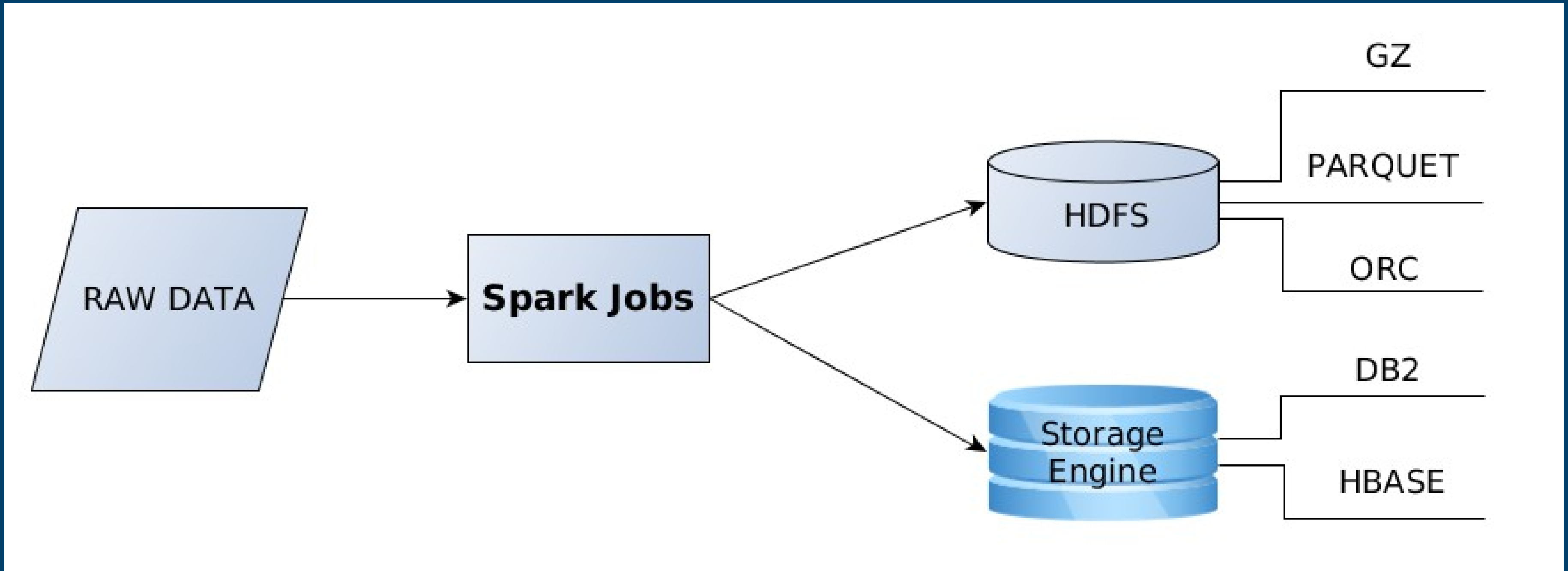
Example Project
Data sources



Challenges:
- Massive amounts of data (TBs)
- Sourced from heterogeneous systems
- Owned by different entities
- Largely unstructured
- Quality issues

Example Project
Data sources

- Spark as framework for data ingestion, filtering and aggregation:



Example Project: *Analytics models*

Regression Model:

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t \qquad [1]$$

where $\{\epsilon_t\}$ is a (stationary) stochastic process with zero mean and unit variance,

$$E[Y_t | X_t = x_t] = \mu(x_t),$$

$$Var[Y_t | X_t = x_t] = \sigma^2(x_t).$$

Given paired observations $(x_t, y_t), t = 1, \ldots, n,$ we assume $\mu(.)$ follows the **Generalized Additive Model**

$$\mu(x_t) = \sum_{i=1}^{p} f_i(x_t) \qquad [2]$$

where $f_i(.)$ are transfer functions. **Note**: GAMs are learned by mgcv packages in R.

Example Project: *Analytics models*

Covariates X_t :

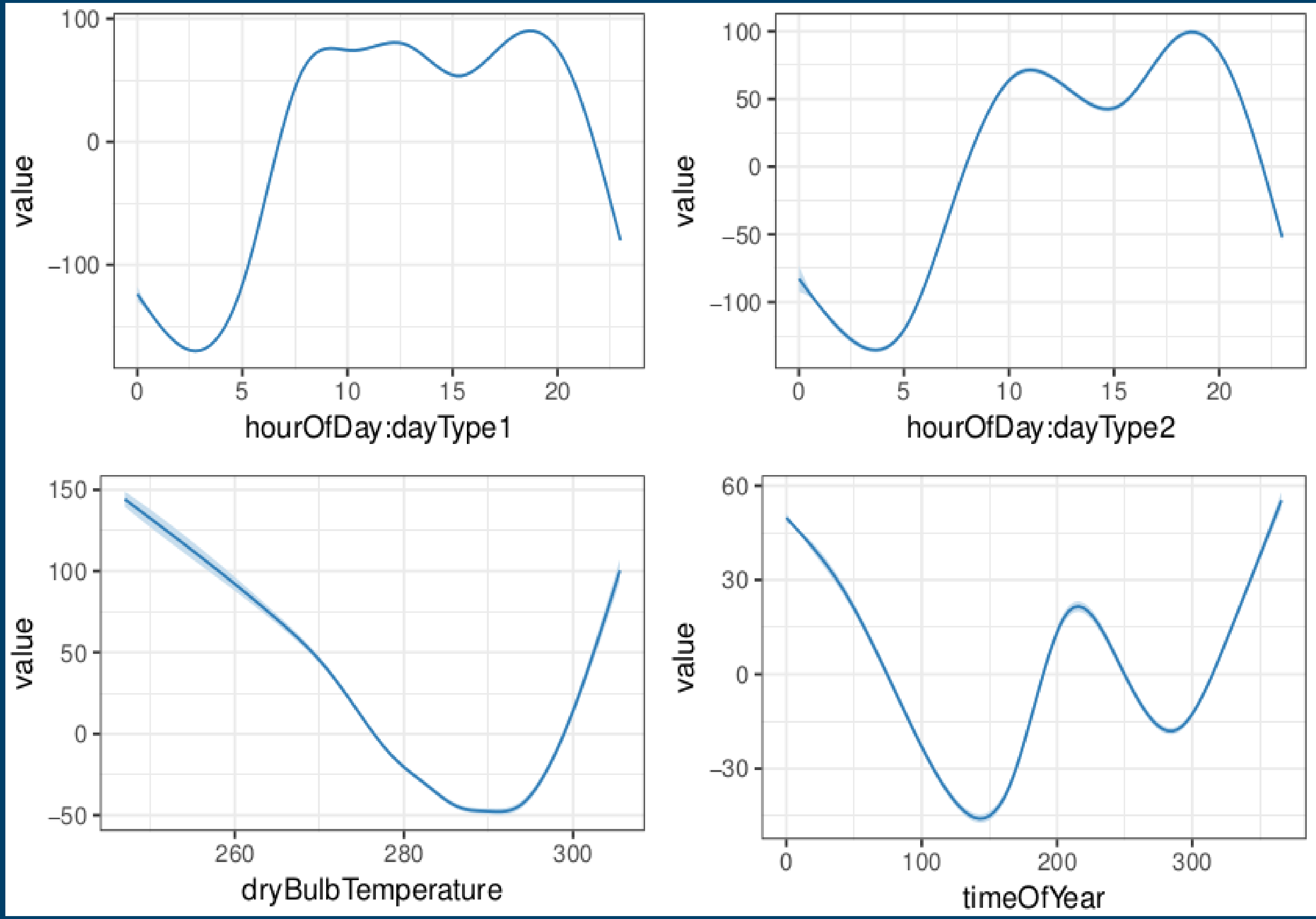
Calendar: TimeOfDay, TimeOfYear, DayType, Season, Special days

Weather: Temperature (lagged, Integrated, Min/Max), irradiance, dew point

PV generation: Data-driven physical models

Real-time measurements:

- 1-16h ahead: DailyMinimum
- 17-24h ahead: Lag 24h
- 25-40h ahead: Lag 48h
- Data anomalies: Fallback models



Example Project: *Analytics models – Uncertainty*

There are **3 main constituents** of *uncertainty* in energy demand forecasts:

1. inherent randomness
2. modeling and estimation error
3. uncertainty in model inputs

Inherent randomness (largest constituent of uncertainty)

Usage of individual electric appliances:

- When? For how long?
- How much electricity do they consume?

} largely “random”
(some daily/seasonal patterns)

Another source of randomness is the noise of metering devices.

Recall regression model:

Conditional mean

Demand → $Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t$ ← “Noise”

Conditional standard deviation

-- ϵ_t represents *inherent randomnesss.*

-- σ^2 accounts for the **uncertainty** depending on X_t , e.g., during *peek demand hours* uncertainty is larger than during low demand hours.

Inherent randomness (largest constituent of uncertainty)

Suppose the **aggregated demand** Y_t is composed of the sum of a large of individual demands $Y_{t,i}$, $i=1,\ldots,k$,

$$Y_{t,i}=\mu_i(X_{t,i})+\sigma_i(X_{t,i})\epsilon_{t,i}.$$

1. Under standard assumptions, $\mu_i(X_{t,i})=O(k)$ and $\sigma_i(X_{t,i})=O(\sqrt{k})$, the **signal-to-noise ratio** increase with k. Empirically, the aggregate demand Y_t can be forecasted with higher accuracy.

2. If k is large, the distribution of ϵ_t can be approximated by normal.

In practice, the convergence to normal is adversely affected by large individual customer where $\epsilon_{t,i}$ *is typically skewed and heavy-tailed.*

Inherent randomness (largest constituent of uncertainty)

Suppose the **aggregated demand** Y_t is composed of the sum of a large of individual demands $Y_{t,i}$, $i=1,\ldots,k$,

$$Y_{t,i}=\mu_i(X_{t,i})+\sigma_i(X_{t,i})\epsilon_{t,i}.$$

1. Under standard assumptions, $\mu_i(X_{t,i})=O(k)$ and $\sigma_i(X_{t,i})=O(\sqrt{k})$, the **signal-to-noise ratio** increase with k . Empirically, the aggregate demand Y_t can be forecasted with higher accuracy.

2. If k is large, the distribution of ϵ_t can be approximated by normal.

In practice, the convergence to normal is adversely affected by large individual customer where $\epsilon_{t,i}$ *is typically skewed and heavy-tailed.*

Inherent randomness (largest constituent of uncertainty)

Given $(x_t, y_t), t=1,\ldots,n,$ we assume $\sigma^2(.)$ also follows a **GAM**.

Uncertainty forecast algorithm (GAM^2):

- **Step 1**. fit a GAM $\hat{\mu}(x_t)$ for the conditional mean. E.g.,

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 DayType_t + \sum_{j=1}^{34} \mathbf{1}(DayType_t = j)\hat{f}_{1,j}(HourOfDay_t) + \hat{f}_2(TimeOfYear_t)$$

$$+ \sum_{l=1}^{4} \mathbf{1}(Season_t = l)\hat{f}_{3,l}(Temperature_t) + \hat{f}_4(Irradiance_t) + \hat{f}_5(Dewpoint_t, HourOfDay_t)$$

$$+ \hat{f}_6(Temperature.lag1_t) + \hat{f}_7(Temperature.lag12_t) + \hat{f}_8(Temperature.lag24_t)$$

$$+ \hat{f}_9(Temperature.Mean.Previous.Day_t) + \hat{f}_{10}(Temperature.Max.Previous.Day_t)$$

Inherent randomness (largest constituent of uncertainty)

- **Step 2**. Calculate the squared empirical residuals $\hat{r}_t^{\,2} = (\hat{y}_t - y_t)^2$ where $\hat{y}_t = \hat{\mu}(x_t)$.
- **Step 3**. Fit a GAM $\hat{\sigma}^2(x_t)$ to $(x_t, \hat{r}_t^2)$.

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \, DayType_t + \sum_{j=1}^{4} \mathbf{1}(DayType_t = j) \hat{f}_{1,j}(HourOfDay_t) + \hat{f}_2(Temperature.Mean.Day_t)$$

  - $DayType_t$: Mon-1, Tue~Thurs-2, Fri-3, Sat-Sun 4.

- **Chebyshev's inequality**: $P(|Y_t - \mu(x_t)| \geqslant \delta) \leqslant \sigma^2(x_t)\delta^{-2}$.
- Assuming $\epsilon_t \sim N(0,1)$. Let $q(\alpha)$ denote the $(1+\alpha)/2$ quantile of $N(0,1)$:

$$P(Y_t \in [\mu(x_t) \pm q(\alpha)\sigma(x_t)]) = \alpha$$

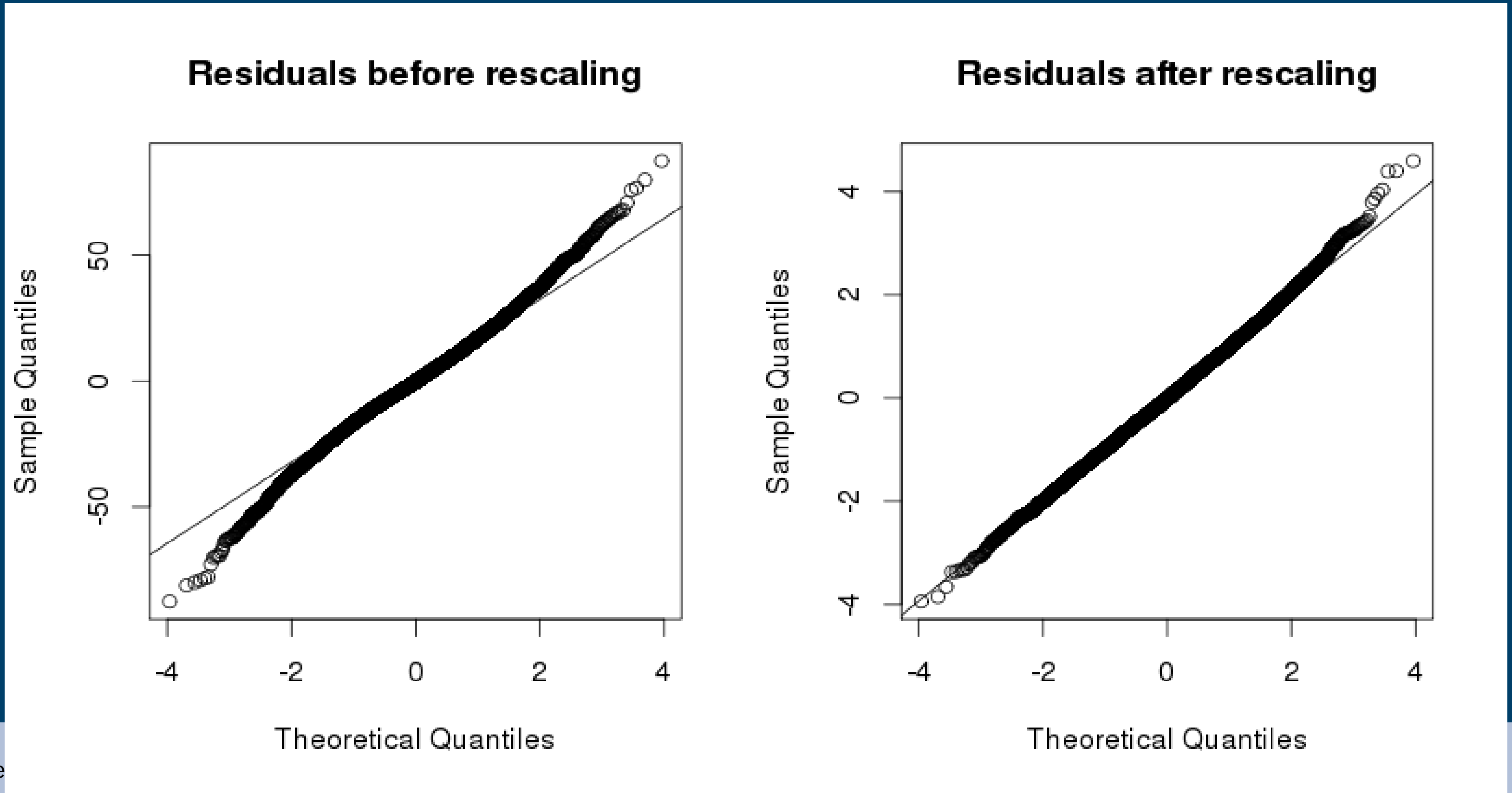
- Substituted by $\hat{\mu}(x_t)$ and $\hat{\sigma}(x_t)$, the 100 α% PI of Y_t: $[\hat{\mu}(x_t) \pm q(\alpha)\hat{\sigma}(x_t)]$.
- When the model assumption is correct, we can show as $n \to \infty$,

$$\hat{\mu}(.) \to \mu(.) \quad \text{and} \quad \hat{\sigma}(.) \to \sigma(.).$$

Inherent randomness (GAM²)

One of the goodness-of-fit criterions we consider *homoscedasticity* and *normality* of the rescaled residuals $(\hat{y}_t - y_t)/\hat{\sigma}(x_t)$.

Example: QQ-plots of $(\hat{y}_t - y_t)$ and $(\hat{y}_t - y_t)/\hat{\sigma}(x_t), t=1,\ldots,n.$



Inherent randomness

- GAM[2] Drawback:
  - 1. $\hat{\sigma}_t^2$ can be negative.
  - 2. the normal assumption may not apply at low aggregation levels.
  - 3. the serial correlation of residuals is complex.

Alternative approach: GAM + boostrap AR-ARCH

- Let us consider hourly model:

$$Y_{t,p} = \mu(X_{t,p}) + \epsilon_{t,p}, \qquad p = 1, \ldots, 24.$$

- where $\mu(.)$ follows GAM with $X_{t,p}$ including *DayType, TimeOfYear, Season, Temperature (lagged, integrated), irradiance, dewpoint*.

Inherent randomness

- Empirical evidence shows $\epsilon_{t,p}$ having features of an AR-ARCH model, e.g., clustered volatility, lepkurtoticity.

For the briefty of notations, we omit the subscript p.

- Recall AR(*p*)-ARCH(*q*) model:

$$\epsilon_t = \sum_{i=1}^{p} \phi_i \epsilon_{t-j} + r_t$$

$$r_t = \sigma_t e_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j r_{t-j}^2$$

where $\phi_{1,..},..,\phi_p$ satisfy $|\phi_j|<1,\ i=1,\ldots,p;$ $\alpha_{0,}\alpha_{1.}..,\alpha_q$ satisfy $\alpha_0>0,\ \ \alpha_j\geq 0,\ j=1,\ldots,q;$ $e_t \sim IID(0,1)$.

Inherent randomness

Algorithm sketch:

- Fit an AR(p)-ARCH(q) model to the empirical residuals $\hat{\epsilon}_t = \hat{y}_t - y_t$, where p and q are selected by AIC. Conduct **residual bootstrap** to the AR-ARCH residuals $\hat{e}_t$ and construct the h-step-ahead bootstrap forecasts $\{r^*_{n+h}\}^B$ and $\{\sigma^{2*}_{n+h}\}^B$, B is the number of bootstrap iteration.
- Construct the empirical cdfs $\hat{F}^*_{r_{n+h}}$ and $\hat{F}^*_{\sigma^2_{n+h}}$ of r_{n+h} and σ^2_{n+h}.
- The $100(1-\alpha)\%$ PIs of Y_{n+h}:

$$[\hat{\mu}(x_{n+h}) + \hat{\epsilon}_{n+h} + I^*_{n+h}(\alpha/2), \hat{\mu}(x_{n+h}) + \hat{\epsilon}_{n+h} + I^*_{n+h}(1-\alpha/2)],$$

where $I^*_{n+h}(p)$ is pth quantiles of $\hat{F}^*_{r_{n+h}}$.

Inherent randomness

- Let F_e and $\hat{F}_{e,n}$ respectively denote the true and empirical cdf of $e_{t.}$
- Assumptions:
  - A1. $Ee_t^4 < \infty$
  - A2. $(Ee_t^4)^{1/2} \sum_{j=1}^{q} \alpha_j < \infty$
  - A3. Let $\theta = c(\alpha_{0,.}..,\alpha_q)$ and $\hat{\theta}_n = c(\hat{\alpha}_{0,.}..,\hat{\alpha}_q)$. Assume $\sqrt{n}\|\hat{\theta}_n - \theta\| = O_p(1)$

- Let be **Mallows metric** defined as

$$d_2(P,Q) = inf\,(E|X - Y^2|)^{1/2},$$

where the infimum is taken over all pairs of (*X,Y*) of r.v. *X* and *Y* respectively distributed according to *P* and *Q*.

Inherent randomness

Lemma. Under assumption A1-A3, as $n \to \infty$,

$$d_2\left(F_e, \hat{F}_{e,n}\right) \to 0$$ in probability.

Theorem. Under assumption A1-A3, as $n \to \infty$,

$$r^*_{n+h} \xrightarrow{d} r_{n+h} \quad \text{and} \quad \sigma^{2*}_{n+h} \xrightarrow{d} \sigma^2_{n+h},$$

For $h = 1, \ldots, H, H \in \mathbb{Z}$.

Inherent randomness

- Empricial coverage probability:

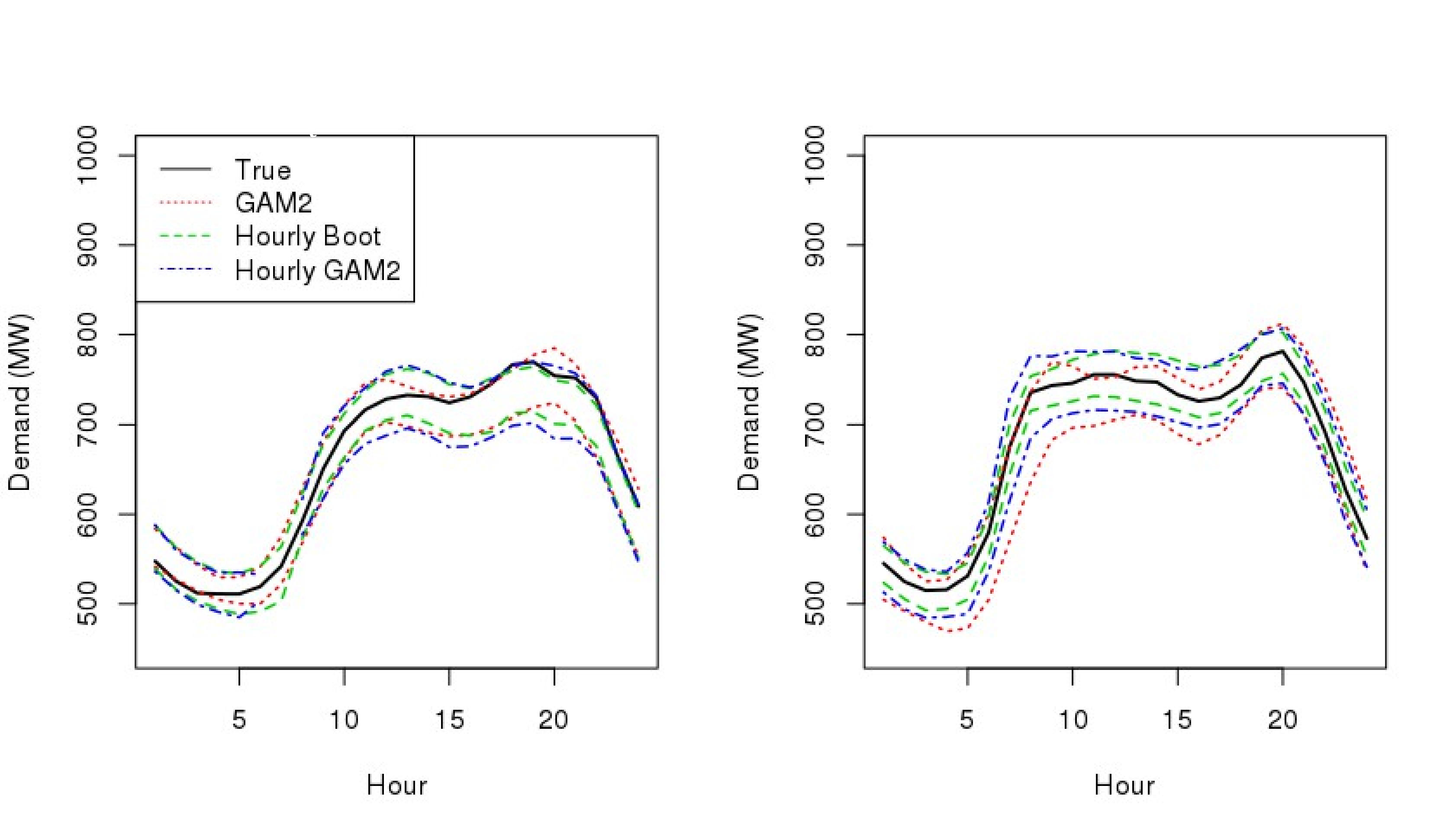
$$C(\alpha, n_t) = n_t^{-1} \sum_{t \in n_t} 1_{(\Phi_t(\alpha/2) \leqslant y_t \leqslant \Phi_t(1-\alpha/2))},$$

where α is the significance level; n_t is the size the test set; $[\Phi_t(\alpha/2), \Phi_t(1-\alpha/2)]$ is the $100(1-\alpha)$% PIs.

Example:

- State hourly energy demand 48-hour-ahead forecasting.
  - Training: 2012-09-01 ~ 2014-06-30
  - Testing: 2014-07-01 ~ 2014-10-31
- Method for comparison:
  - GAM[2]
  - Hourly model with bootstrap AR-ARCH
  - Hourly model with GAM[2]

Inherent randomness



	Coverage (95%)
GAM^2	88.7%
H.Boot	91.8%
$H.GAM^2$	94.7%

- Hourly GAM^2 has the best coverage.
- Hourly Boot can be used as a back up method when $\hat{\sigma}_t^2$ is negative using GAM^2.

Modeling and estimation errors

Given $\hat{\sigma}(.)$ and x_t, the expected value of the squared residual is

$$E[(\hat{y}_t - y_t)^2 | x_t, \hat{\mu}(.)] = (\hat{\mu}(x_t) - \mu(x_t))^2 + \sigma^2(x_t)$$

Hence, $\hat{\sigma}^2(.)$ does not only account for $\sigma^2(.)$ but also $(\hat{\mu}(.) - \mu(.))^2$, which is smaller *in-sample* than *out-of-sample*.

Uncertainty forecast algorithm modification

1. Partition {1,2,...,n} into two disjoint sets *T* and *V*, where *V* is a **held-out** set for **rescaling**.

2. Fit a model $\hat{\sigma}^2(.)$ on the training data $(x_t, \hat{r}_t^2)$ with $t \epsilon T$.

3. Fit a multiplicative correction $\hat{c}^2$ to the model $\hat{\sigma}^2(.)$ on the held-out training data $(x_t, \hat{r}_t^2)$ with $t \epsilon V$.

- The correction is used, on data outside the training set, to rescale the residuals $\hat{y}_t - y_t$ by $\hat{c}\hat{\sigma}(x_t)$ instead of $\hat{\sigma}(x_t)$.
- The efficiency can improved by **randomly sampling** the points to be included in the validation set and by repeating the procedure for different training-validation splits.

Uncertainty in model inputs

Weather covariates:

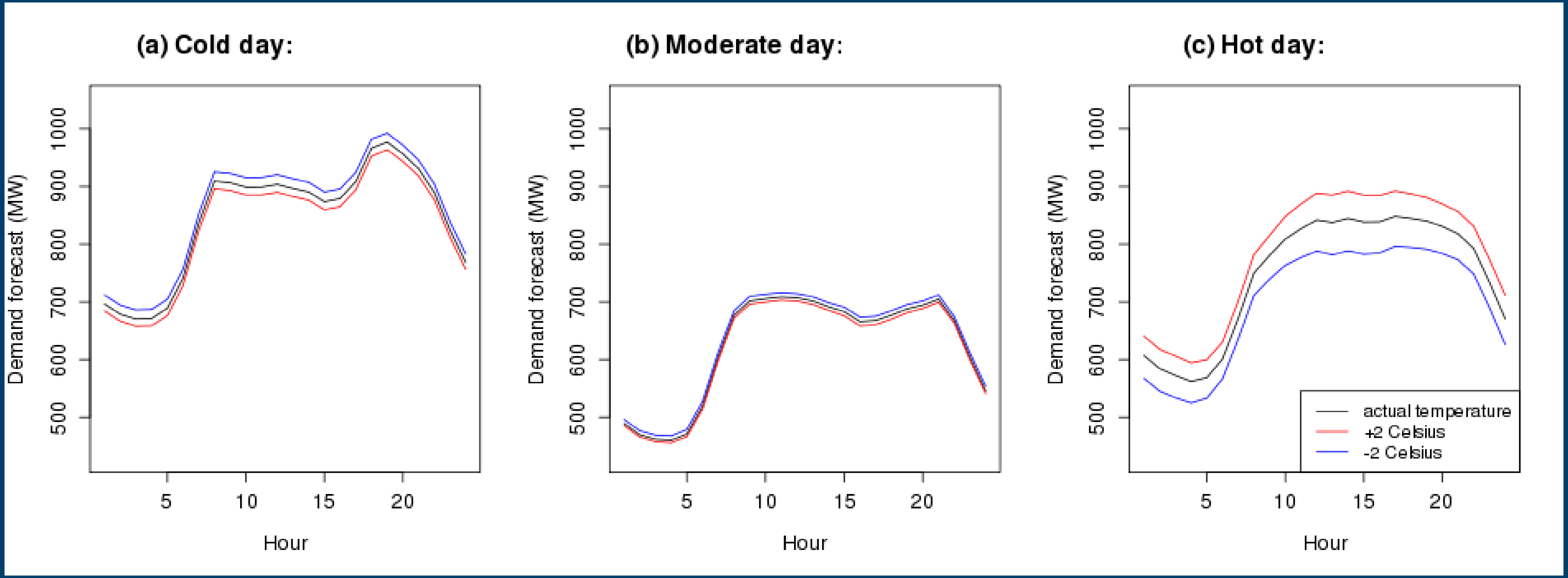
- Temperature (daily max/min/mean, lagged values, etc.)
- Dew point, irradiance

Two ways to account for uncertainty in weather inputs:

(1) Use **weather forecasts** instead of actual weather measurements to train the demand model. Thereby, the model will be less sensitive to weather forecasting errors and treat the uncertainty in the weather inputs as part of the inherent randomness discussed before.

(2) If probabilistic weather forecasts are available, train the model on actual weather measurements and obtain an ensemble of demand forecasts based on sampled weather inputs.

Uncertainty in model inputs



Simulated energy demand forecasts using actual temperature (black), actual temperature +2°C (red) and actual temperature – 2°C (blue), respectively on a cold day (left), a moderate day (middle) and a hot day (right).

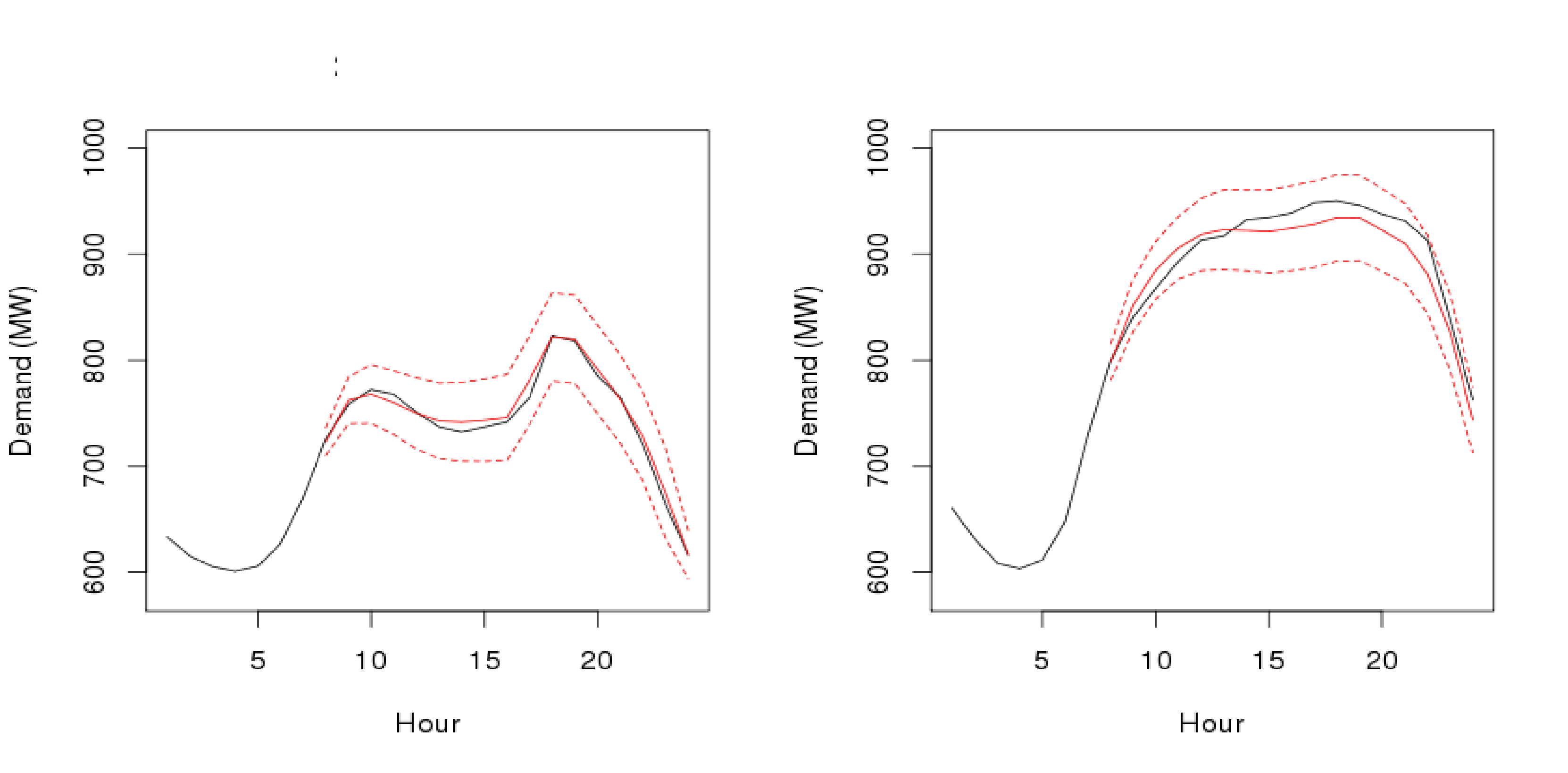
Reduce uncertainty using real-time infomation

- Learn **different models** for every forecasting horizon.
- Use the **most recent** available demand data.

Specifically, given observations $y_t, y_{t-1}, \ldots,$ in the case of 24-hour-ahead forecasting,

- The idea is to train different models $\hat{\mu}_h(\cdot)$ for each of h=1,2,...,24 hours ahead.
- When computing the forecast $\hat{y}_{t+h} = \hat{\mu}(x_{t+h}),$ those models take into account the most recent measurement y_t.
- For small h, this yields significant gains in modeling accuracy.
- Similarly, train 24 different models for the conditional variance.

Reduce uncertainty using real-time infomation



Real-time demand forecast and associated uncertainty on Feb 1st (left) and Jul 1st, 2014 (right).

Automatic forecasting for Distribution Substations

- Forecasting at lower aggregation levels requires handling **large amount** of data and models.
- It is **infeasible** to depend on manual work.
- We present a *Energy Forecasting System* for **automatically** processing and forecasting energy demand at different aggregation levels.

Automatic forecasting for Distribution Substations

Challenges of demand forecasting at substation level:

- *Inherent randomness*: higher volatility, individual customer's impact can be significant.
- *Distributed generation*: the effect caused by renewable is highly noticeble. e.g., demand inversion.
- *Heterogenous profiles*: difference in data patterns due to the characteristics of consumers within the substation, e.g., residential, industry.
- *Load shift*: changes in the network configuration or connectivity lead to significant level shifts.
- *Data issues*: data are more contaminated by messurement errors and noise, e.g., constants, zeros, missings.

Automatic forecasting for Distribution Substations

Residential Vs. **Industrial** substations.

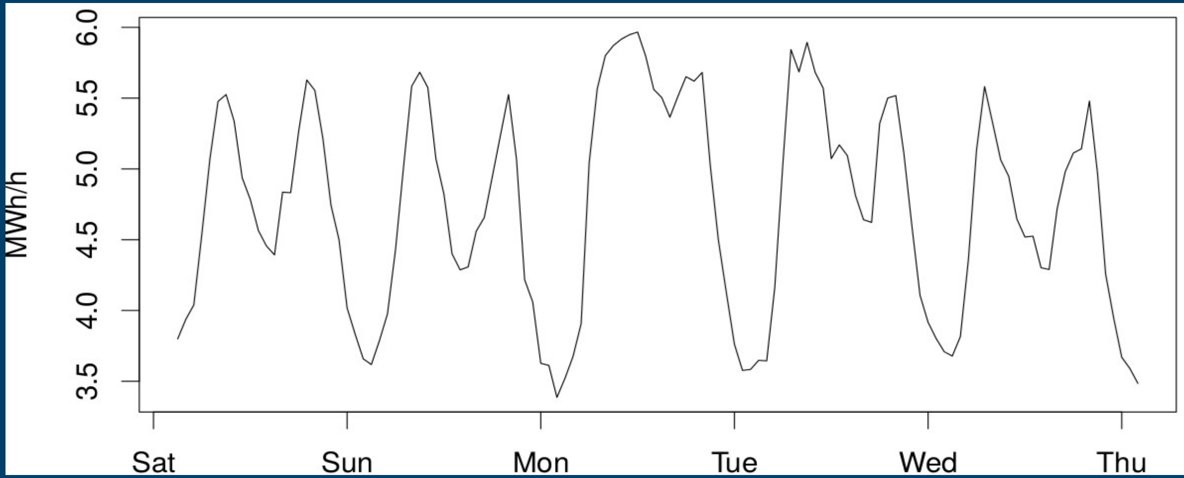


Fig. 1 Residential load with solar

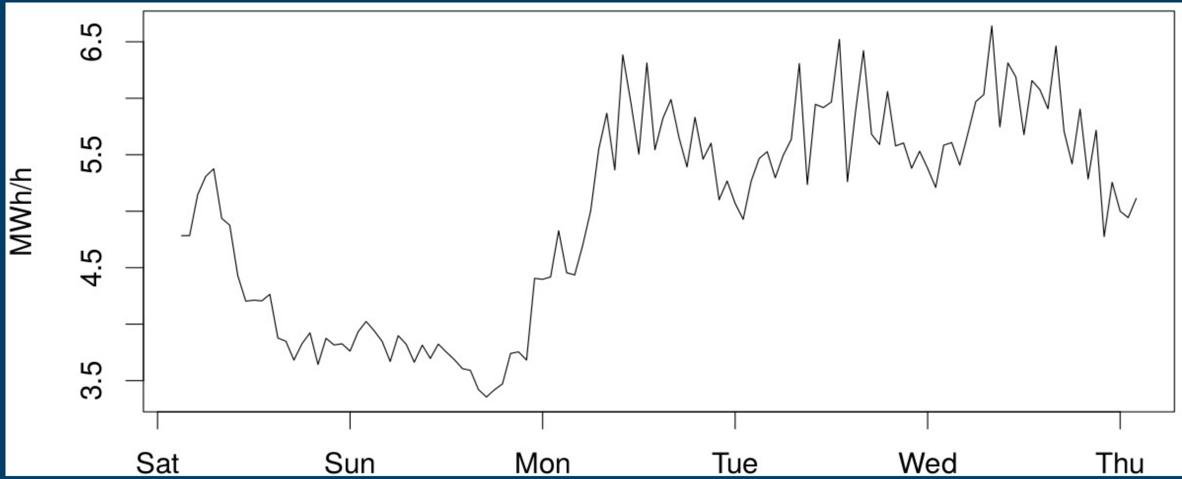
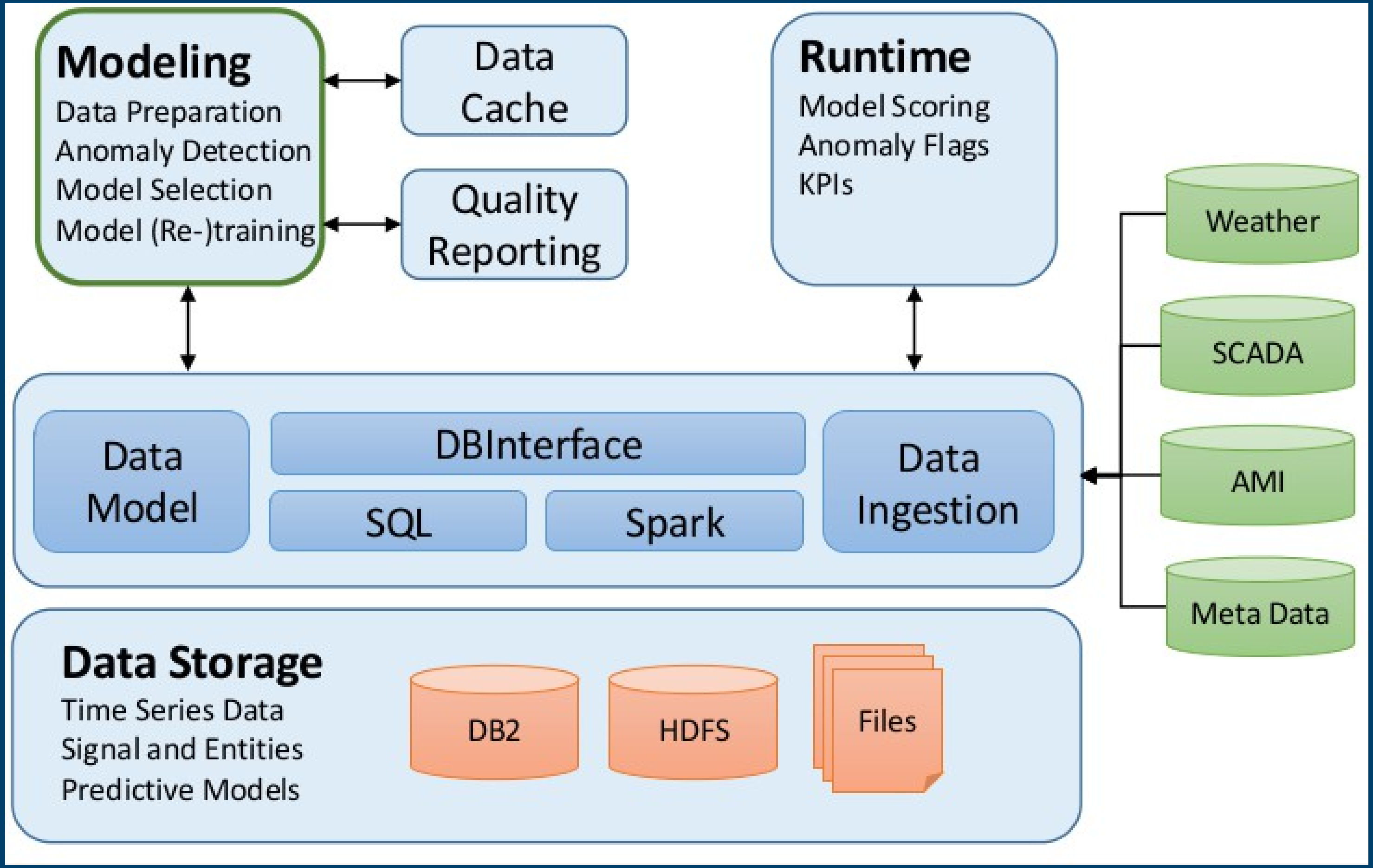
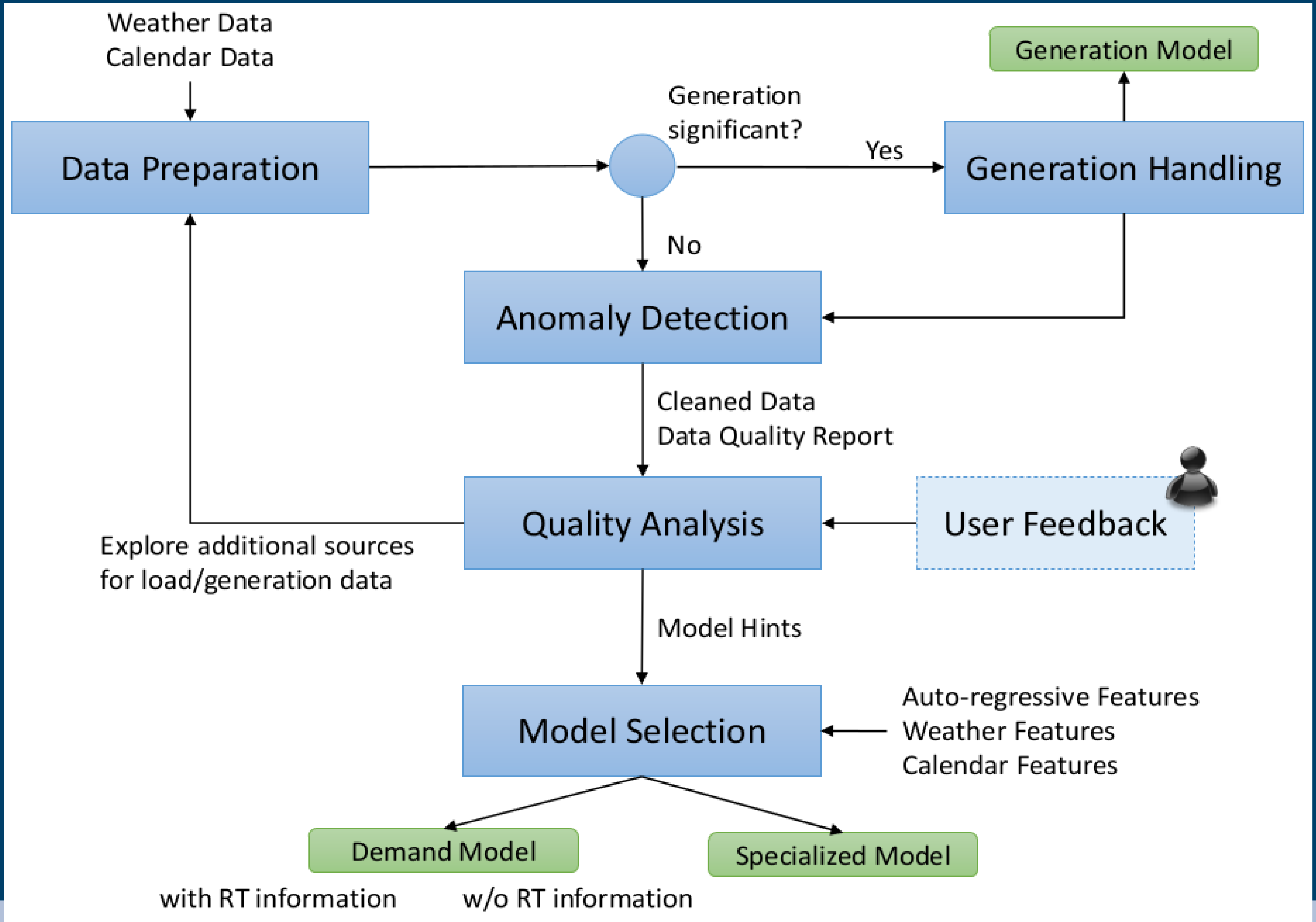


Fig. 2 Industrial load without solar

Automatic forecasting: System Architecture



Automatic forecasting: Workflow and Output



Automatic forecasting: Anomaly handling

Outlier detection – a combination of rule- and model-based method.

- **S1**. Apply static rules to remove implausible values, e.g., negative values and constant segments.
- **S2**. Fit **GAM** to the remaining data.
- **S3**. Compute the **absolute values** of the model residuals.
- **S4**. Remove all values where the residuals are beyond a given quantile of the *normal distribution*, e.g, $\alpha = 0.9999$.
- **S5**. Go back to **S2** until convergence, i.e., no further values being removed.

Automatic forecasting: Anomaly handling

Load shift, equivalently, *multiple change point detection*.

- **S1**. Aggregate the **hourly** electrical loads to **daily** total demand.
- **S2**. Fit a simple **GAM** model to the daily demand (including average daily temperature, dewpoint and irradiance).
- **S3**. Compute the model residuals and record the **explained deviance** *d*.
- **S4**. Apply the **Pruned Exact Linear Time** (PELT) algorithm [2] for detection of multiple change points in the residuals.
- **S5**. Identify the segment which most significantly **deviates in mean** from the rest of the data, and remove the corresponding days.
- **S6**. Go back to **S2** until no data is left, or the **explained deviance** *d* exceeds a given threshold, e.g., $d=0.8$.

Automatic forecasting: Anomaly handling

Transfer learning. Assume the load shift occurs at t_{LS} ,

- **S1**. Fit a **GAM** to the data before t_{LS}.
- **S2**. Calculate the predictions $\hat{y}_t = f(x_t), \quad t = t_{LS}+1, t_{LS}+2, \ldots, n.$

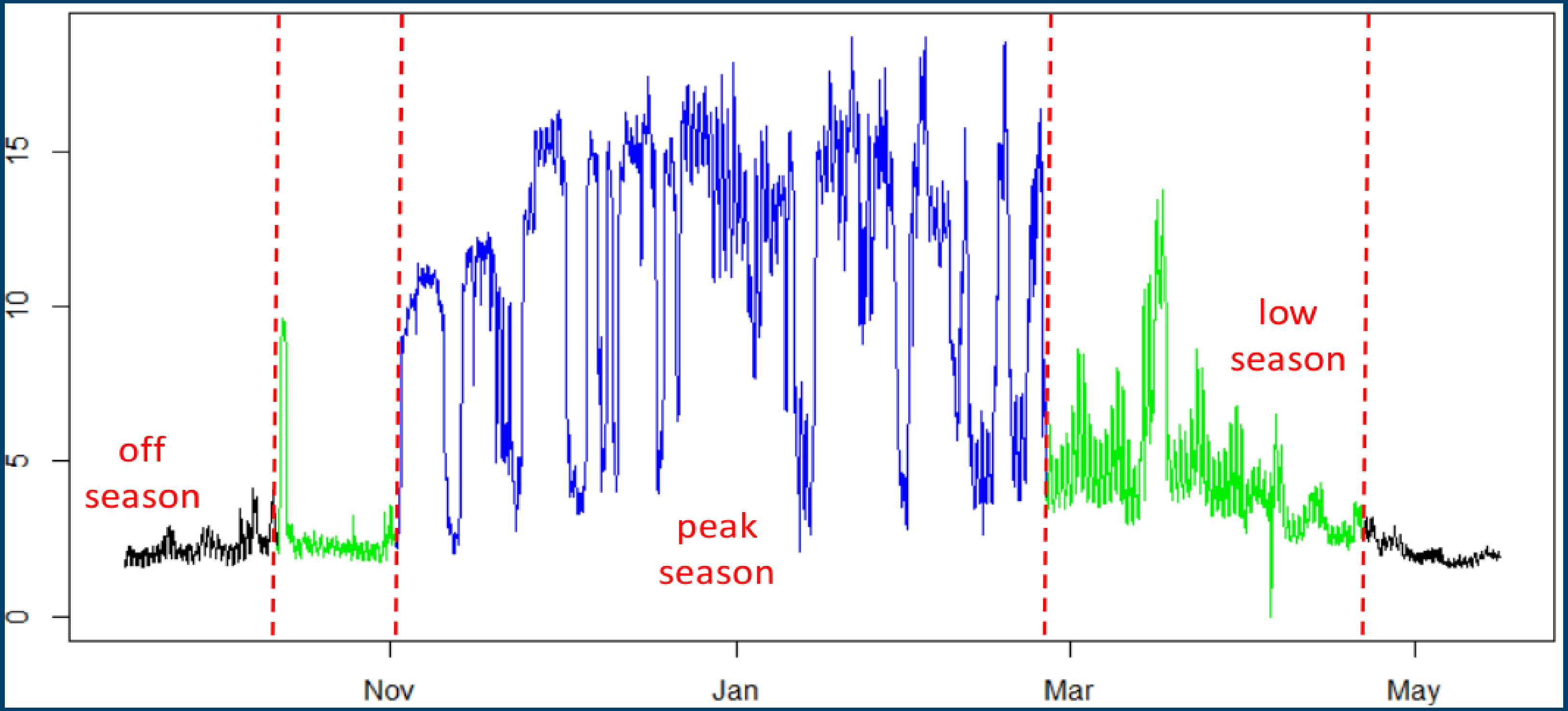
  The predictions are regarded as the **regular demand values** without the load shift.
- **S3**. Learn **transfer model** $y_t = \tilde{f}(\tilde{x}_t)$, where $\tilde{x}_t$ includes $\hat{y}_t$.
- **S4**. Calculate $\tilde{y}_t = \tilde{f}(\tilde{x}_t), \; t = 1, 2, \ldots, t_{LS}$.
- **S5**. In the data, replace the values y_t before $t = t_{LS}$ by $\tilde{y}_t$.

Automatic forecasting: Feature selection

- Select **features** from a pre-defined set using the **backward elimination** (greedy approach).
- The data is divided into a training set S_T and a verfication set S_V.
  - **S1**. Fit **GAM** on S_T using the entire feature set. Compute model quality. e.g., RMSE, on S_V.
  - **S2**. Progressively remove each feature from the **GAM equation**, fit a GAM and compute the model quality (e.g., RMSE). Select the equation with the **highest** improvement over the previous one.
  - **S3**. Repeat **S2** until convergence, i.e., no further improvement can be achieved.

Automatic forecasting: Specialized model

- In some cases, **meta information** is available from the domain experts.
- Example: Substations with **snowmaking**. Special features are added, e.g., peak season, derived temperature effect and day types.



Automatic forecasting: Case study

- **Data**: hourly energy demand 2013-01-01 ~ 2016-03-31 from **substations**.
- Training: 2013-01-01 ~ 2016-02-29
- Testing: 2016-03-01 ~ 2016-03-31
- Forecasting quality measure: normalized RMSE (NRMSE).
- **Comparison**:
  - 1. baseline model with the entire feature set
  - 2. baseline model with anomaly handling
  - 3. automatic model selection
  - 4. both anmaly handling and automatic model selection

Automatic forecasting: Case study – NRMSE comparison

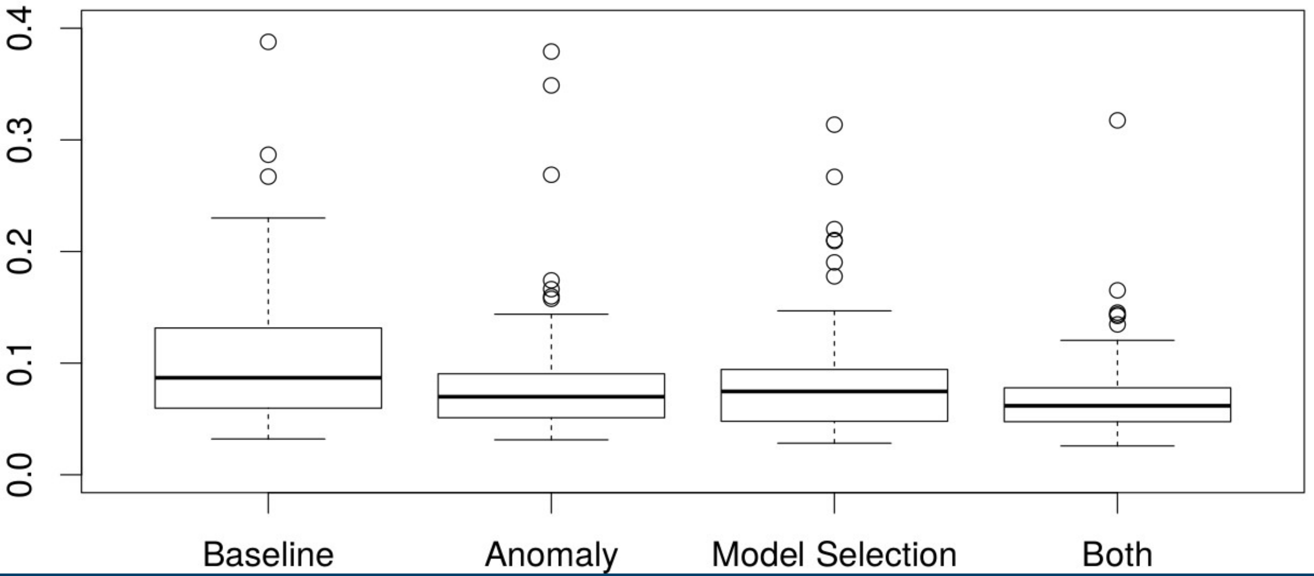


Fig. 3
NRMSE without autoregressive

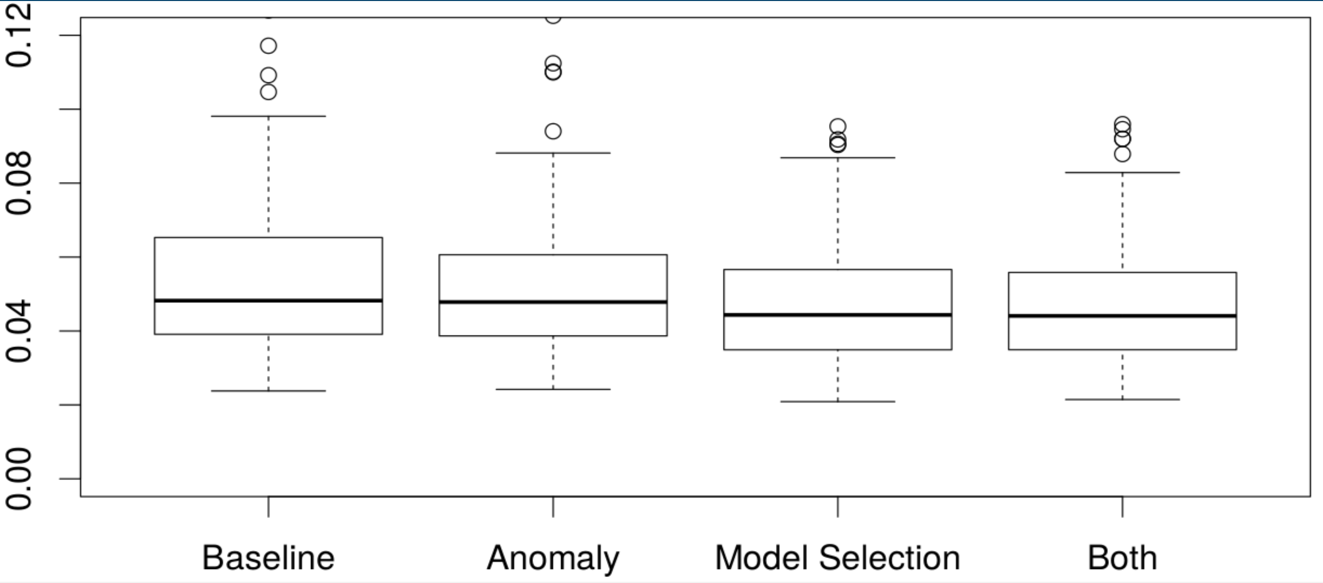
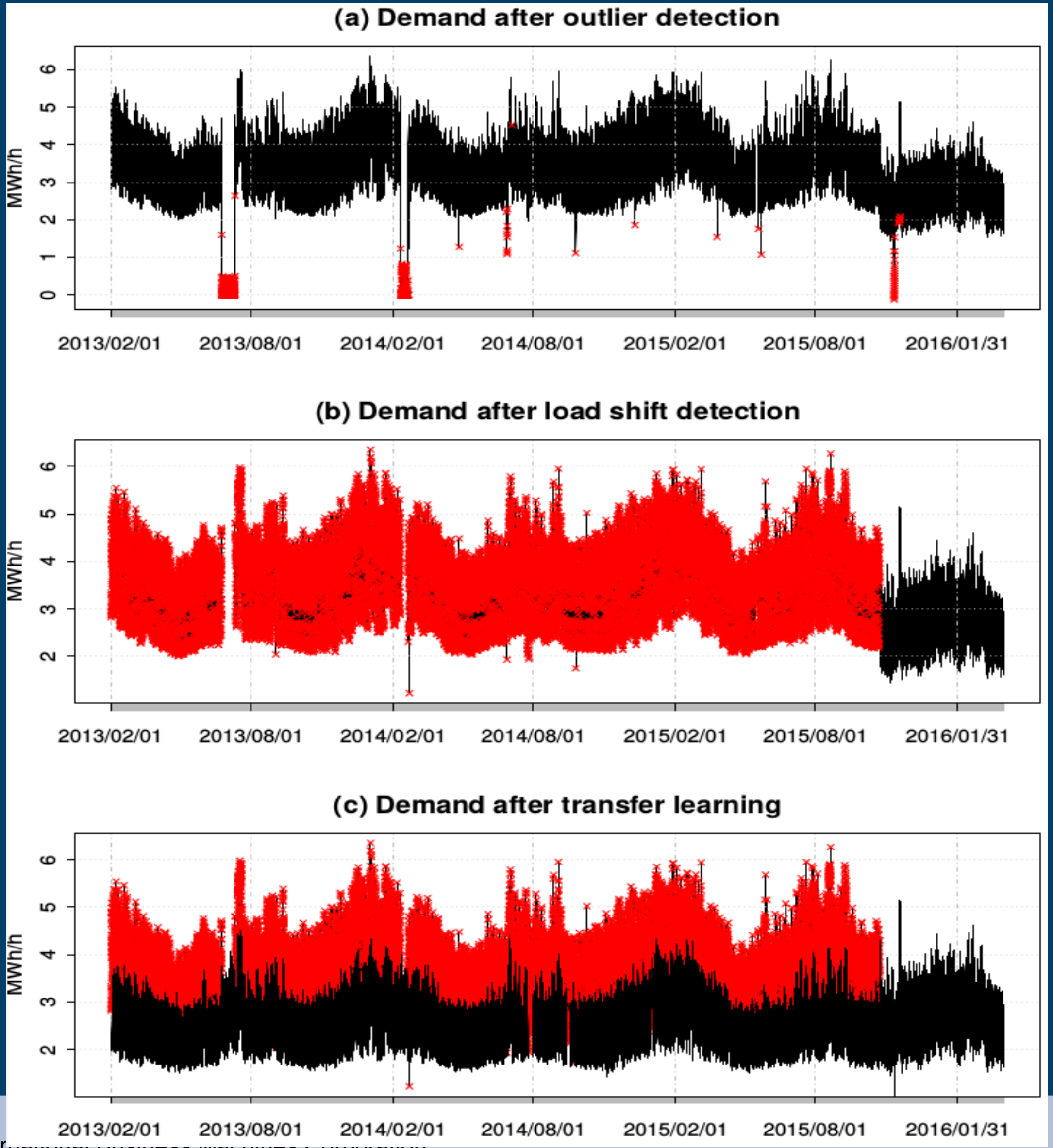


Fig. 4
NRMSE with autoregressive

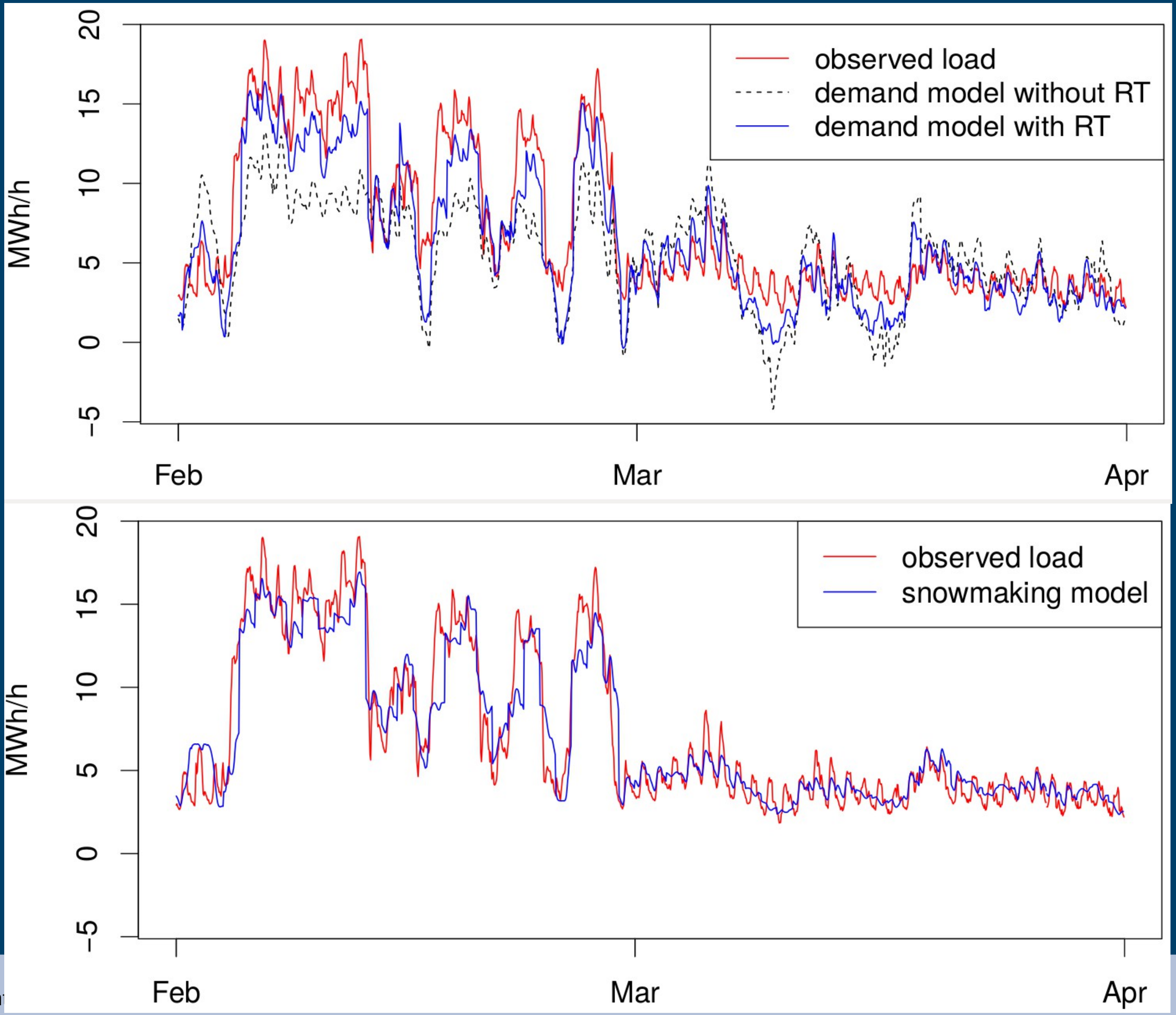
Automatic forecasting: Case study – Anomaly handling



Substation C

Method	NRMSE
baseline	23%
anomaly handling	6.5%
anomaly handling & automatic model selection	5.3%

Automatic forecasting: Case study – Specialized model



Substation S

Method	NRMSE
Baseline without AR	19%
Baseline with AR	11%
Snow-Making model	8%

Example Project
Use cases

- Long-term planning:
  - Data and analytics services
  - Analysis of bulk and subsystem issues
  - Analysis of non-transmission alternatives
- Short-term planning / operations:
  - Renewable integration
  - Contingency analysis
  - Outage planning
- Peak energy management:
  - Reduction of regional network charges
  - Mitigation of congestion events

Reference

[1] T.J. Hastie, R.J. Tibshirani (1990): Generalized Additive Models. London, UK Chapman & Hall.

[2] S.N. Wood (2006): Generalized Additive Models An Introduction with R. Boca Raton, FL, USA Chapman & Hall.

[3] A.Ba, M.Sinn, Y.Goude, P.Pompey (2012) Adaptive learning of smoothing functions Application to electricity load forecasting. NIPS, 2519–2527.

[4] J.Fan, Q.Yao (1998): Efficient Estimation of Conditional Variance Functions in Stochastic Regression. Biometrika 85(3), 645-660.

[5] T.K. Wijaya, M. Sinn, and B. Chen (2015): Forecasting uncertainty in electricity demand. Proc. AAAI-15 Workshop.


Thank
You