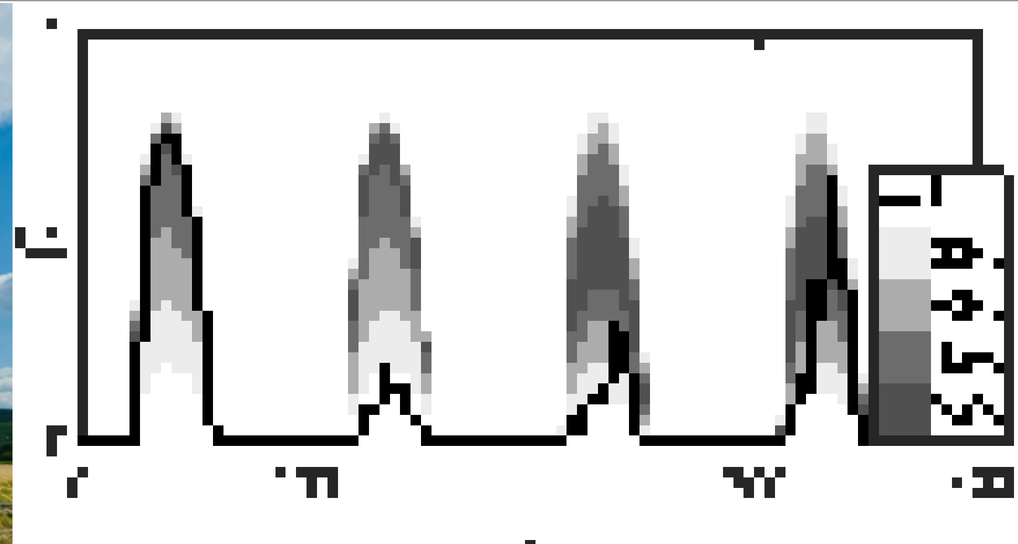
Probabilistic Energy Forecasting

Moritz Schmid

Seminar Energieinformatik WS 2015/16

Institute for Applied Computer Science



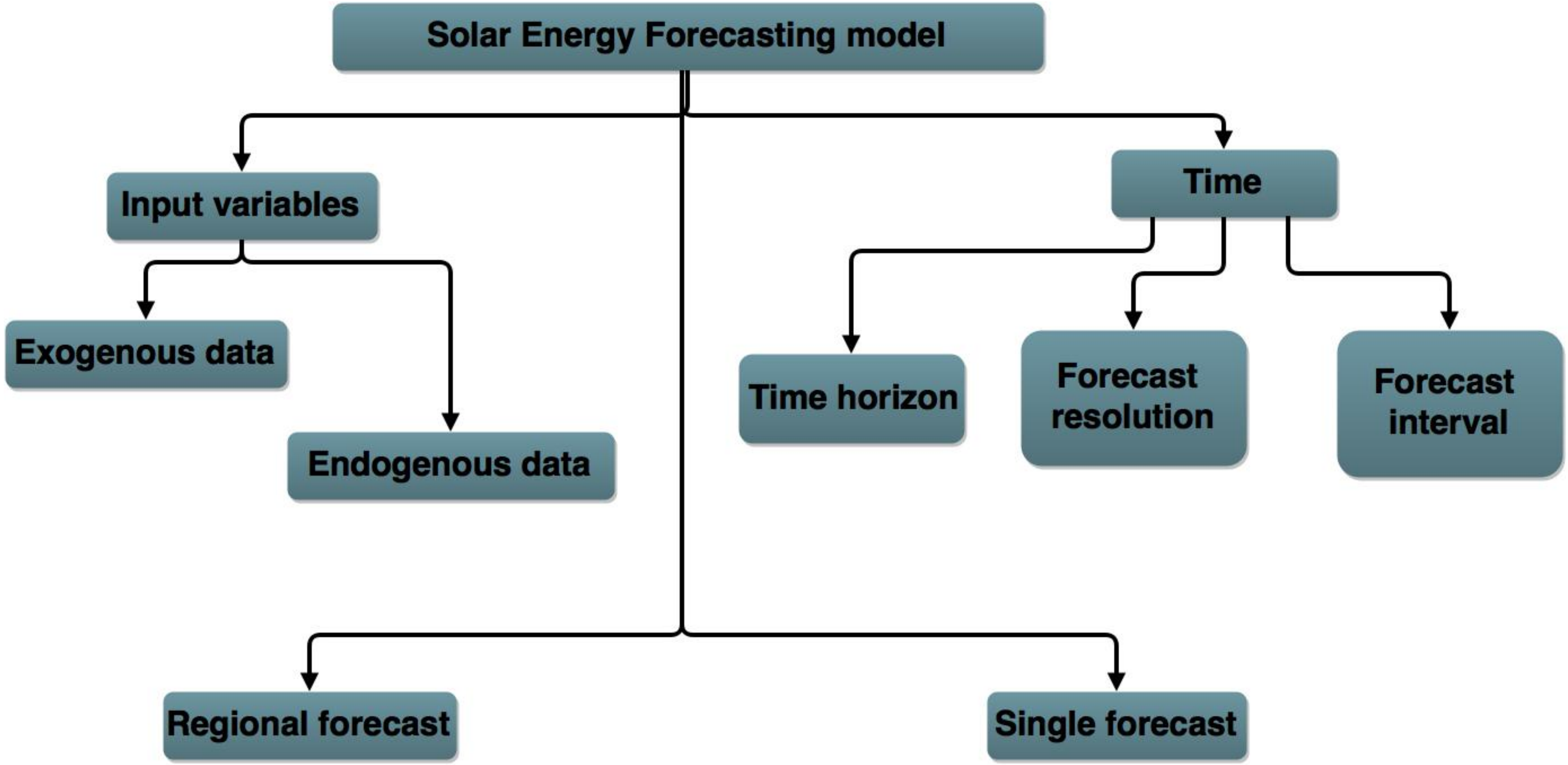
Agenda

- Forecasting challenges
- Renewable energy forecasting
- Electricity price forecasting
- Electric load forecasting
- Probabilistic solar power forecasting model based on k-nearest neighbor and kernel density estimator
- Recap and outlook

Forecasting challenges

- Data cleansing
  - Identifying incomplete and inaccurate data
  - Removing irrelevant data
  - Choosing relevant input variables
- Choosing suitable forecasting techniques
- Integration
  - Scenario generation
  - Modeling
  - Post processing

Renewable energy forecasting



Forecasting techniques

Physical models	Statistical models
■ White box models ■ Also called parametric ■ Use analytical equations ■ E.g. use irradiance forecast	■ Black box models ■ Also called non parametric ■ Direct prediction of power output ■ Use machine learning methods
Hybrid Models	

Statistical forecasting techniques (1)

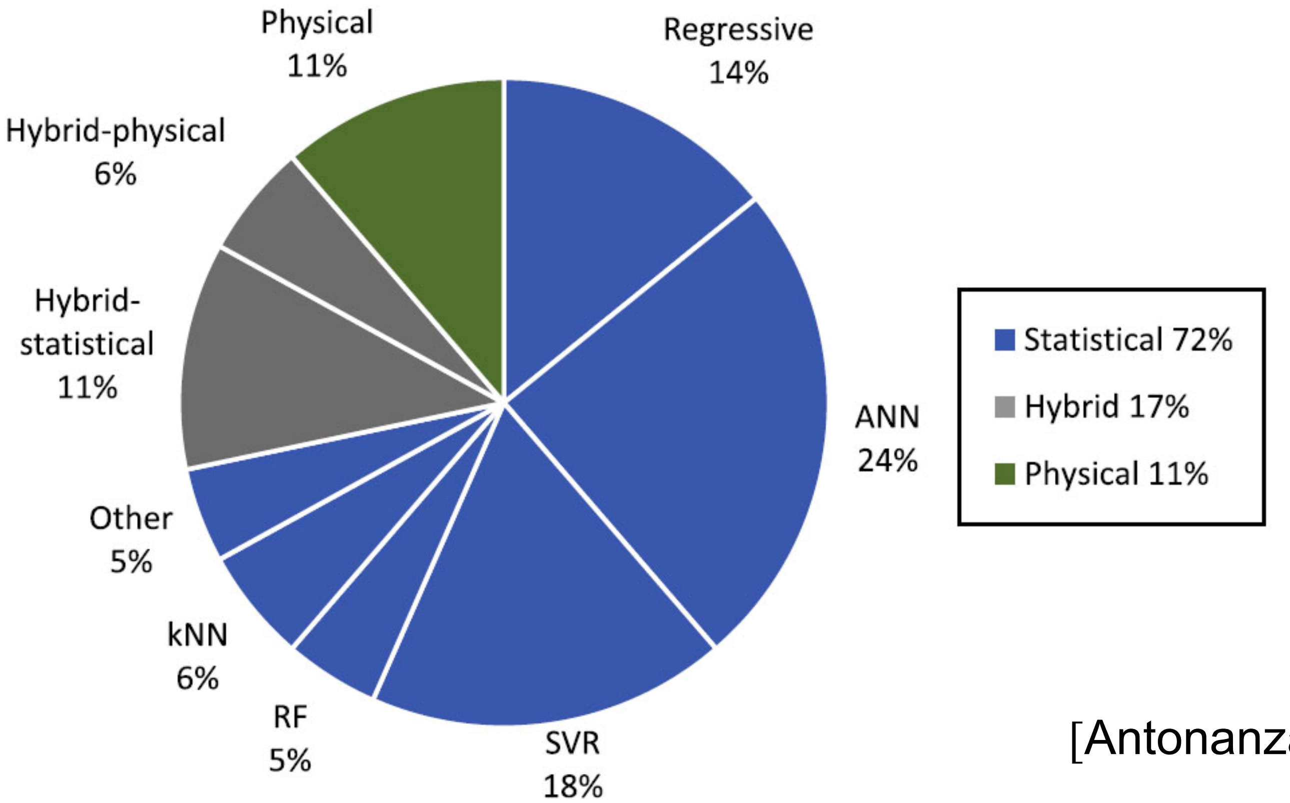
- Regression
  - Estimate the relationship between a dependent and a independent variable
  - Predictor variable: wind speed, …
  - Criterion variable: power output
- Artificial neural networks (ANN):
  - Consists of a group of interconnected neurons
  - Connections have numeric weights
  - All connections together produce an output

Statistical forecasting techniques (2)

- K-nearest-neighbors (kNN):
  - Compares current values with training samples in a feature space
  - K-nearest neighbors with the smallest distance are selected for the predictions

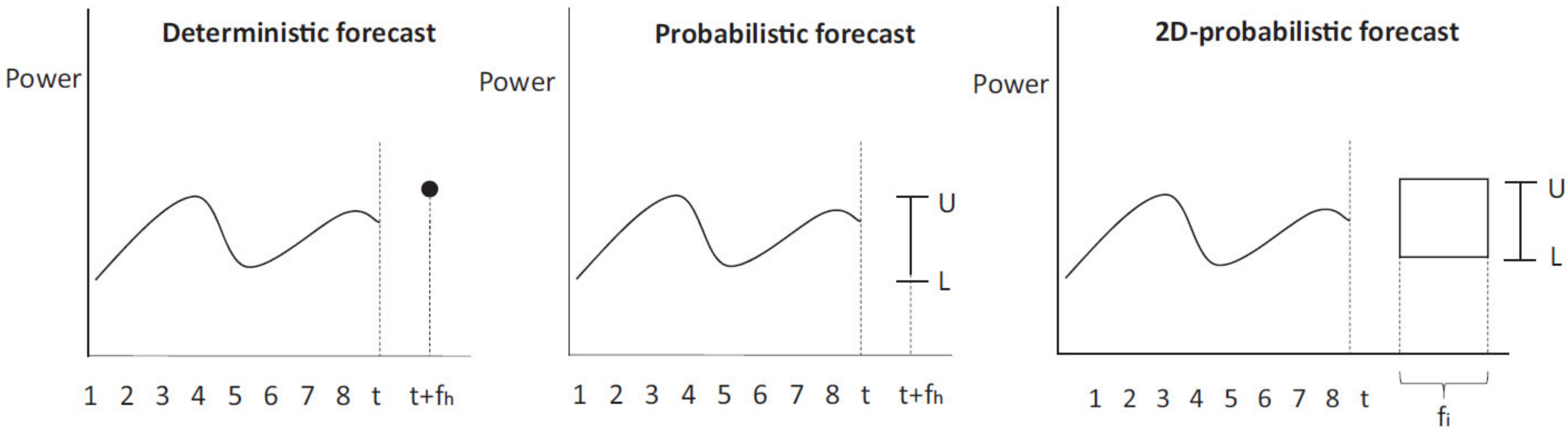
- Support vector machines (SVM):
  - Perform well with non-linear problems
  - Use transformations to keep the complexity of problems low
  - Known as support vector regression machines (SVR) when solving regression problems

Published techniques in solar energy forecasting based on studies



[Antonanzas16]

From deterministic to probabilistic forecasting



[Antonanzas16]

- Probabilistic forecasts
  - Quantify uncertainty
  - Add relevant information about the expected values
  - Assign a probability to each outcome

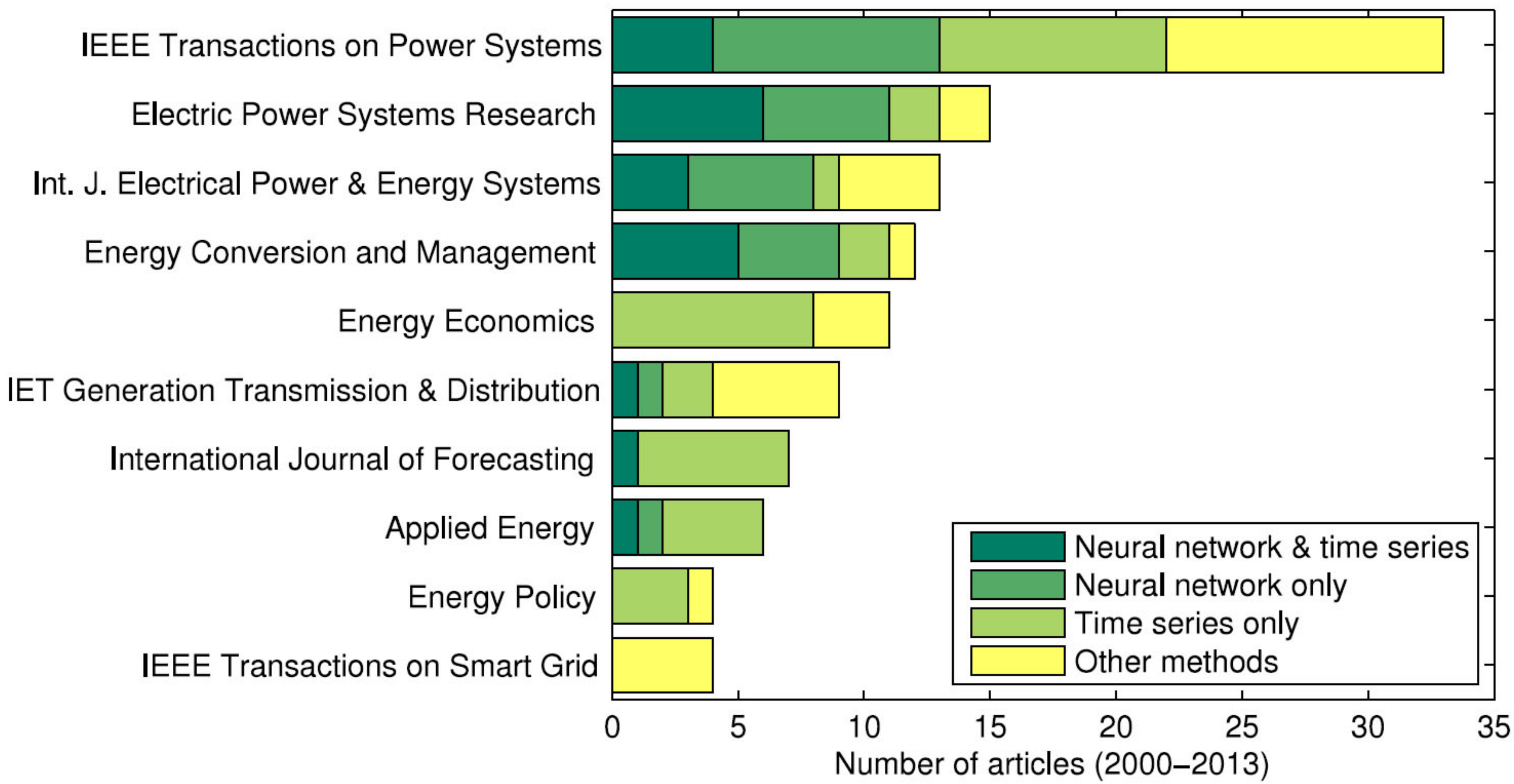
Obtaining probabilistic forecasts

- Parametric approach
  - Forecast values follow a known distribution
  - Goal: determine the parameters describing it
- Non parametric approach
  - Makes no assumption about a distribution
  - Goal: approximate the actual distribution with training data
  - E.g. Combining Quantile regression models, point forecasting models
- Both parametric and nonparametric approaches are feasible and commonly used

Confidence and prediction interval

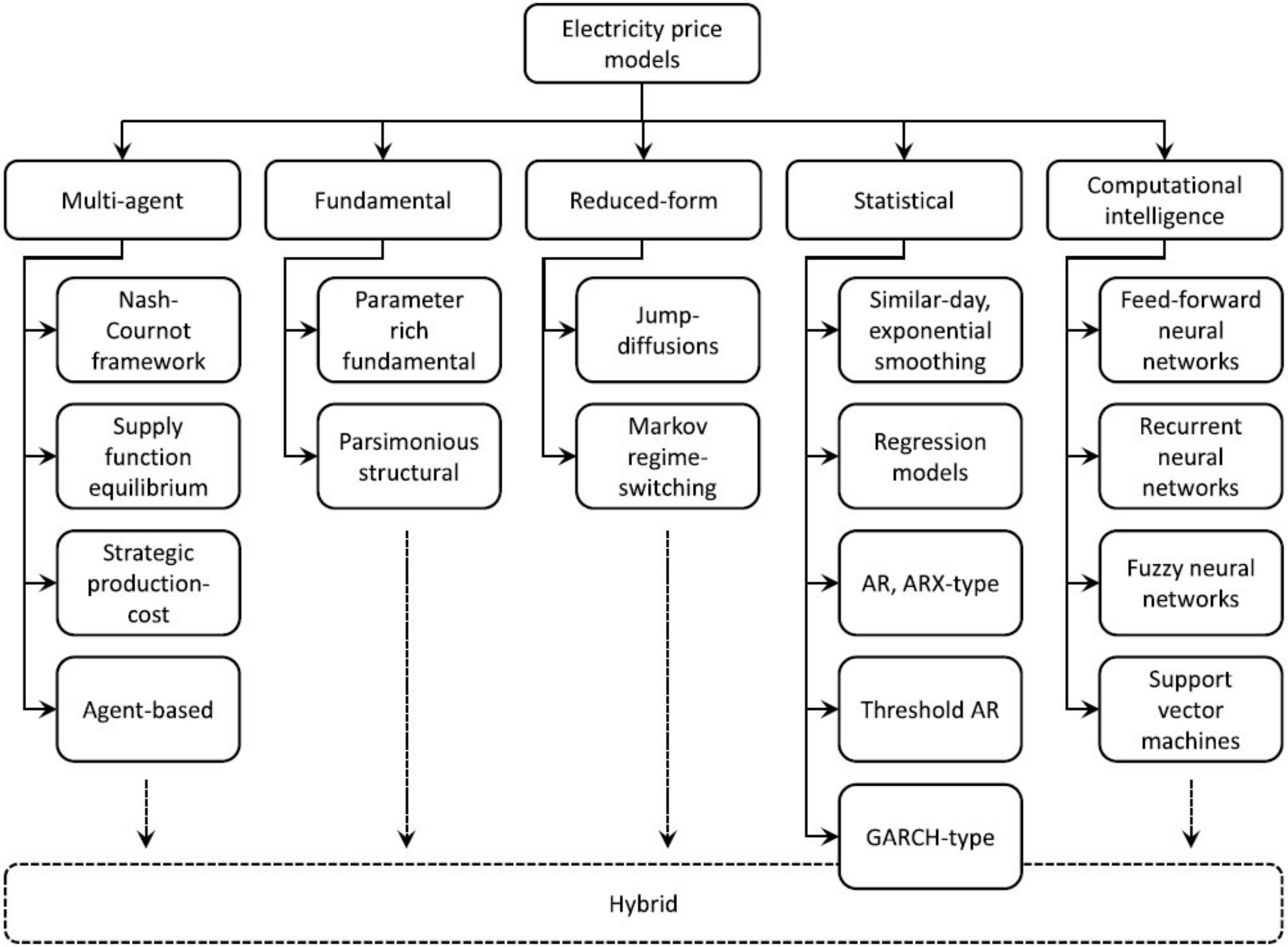
Confidence interval	Prediction interval
▪ Contains the parameter of interest with a specific probability ▪ Possible sample set: $\{X_1, ..., X_n\}$ ▪ Is smaller than prediction interval	▪ Contains a random variable **yet to be observed** ▪ Estimates the value of the next sample variable, X_{n+1} ▪ Account for uncertainty of population mean and data scatter

Electricity price forecasting(1)



[Hong16]

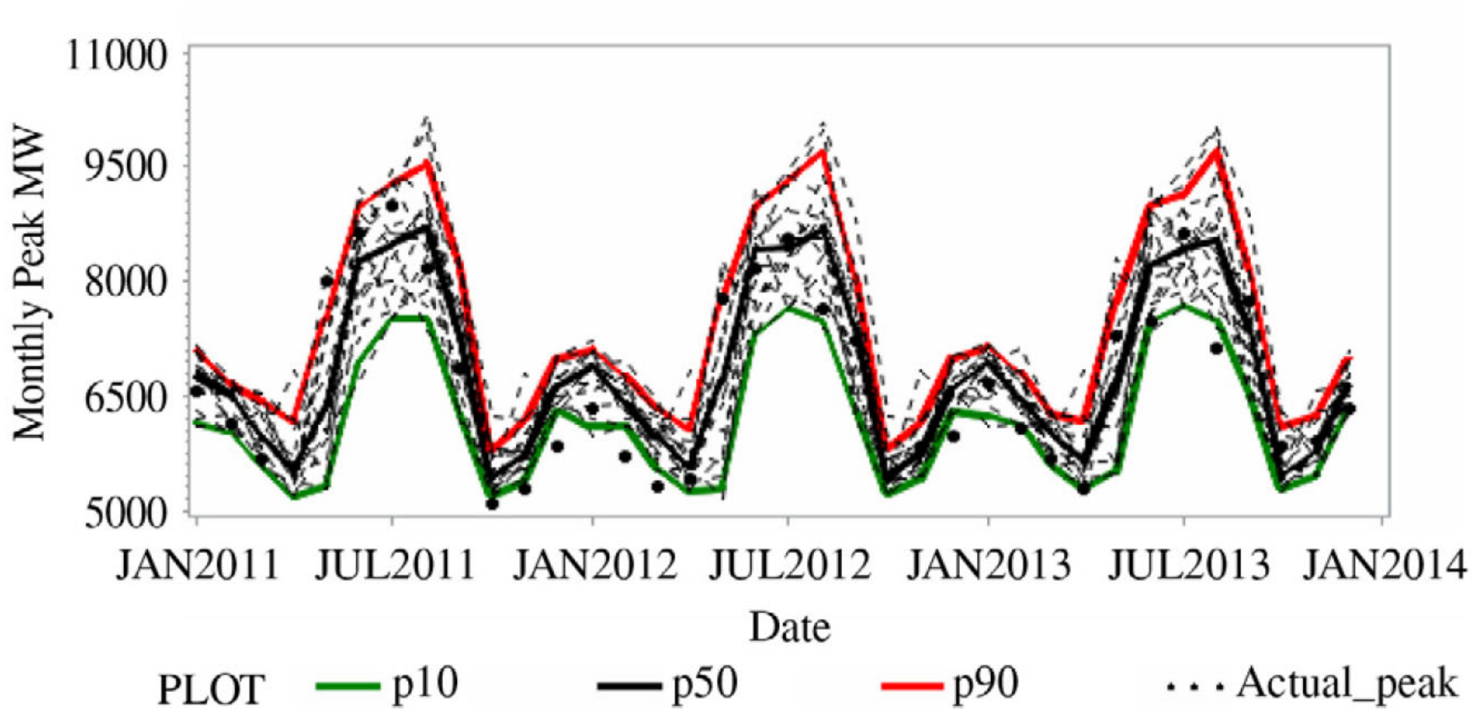
Electricity price forecasting(2)



[Hong16]

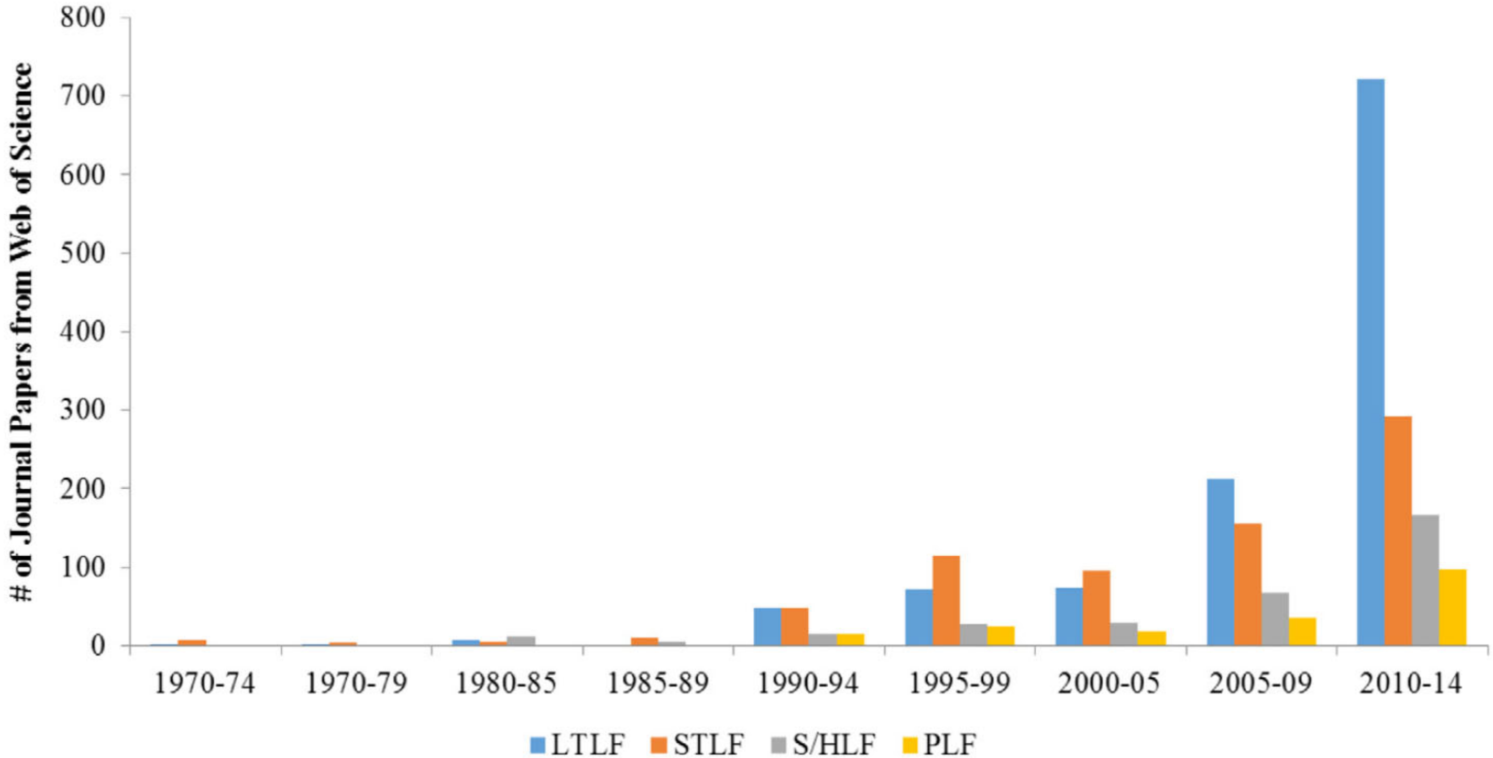
Electric load forecasting

- Probabilistic monthly peak load forecast for a US utility



[Hong16]

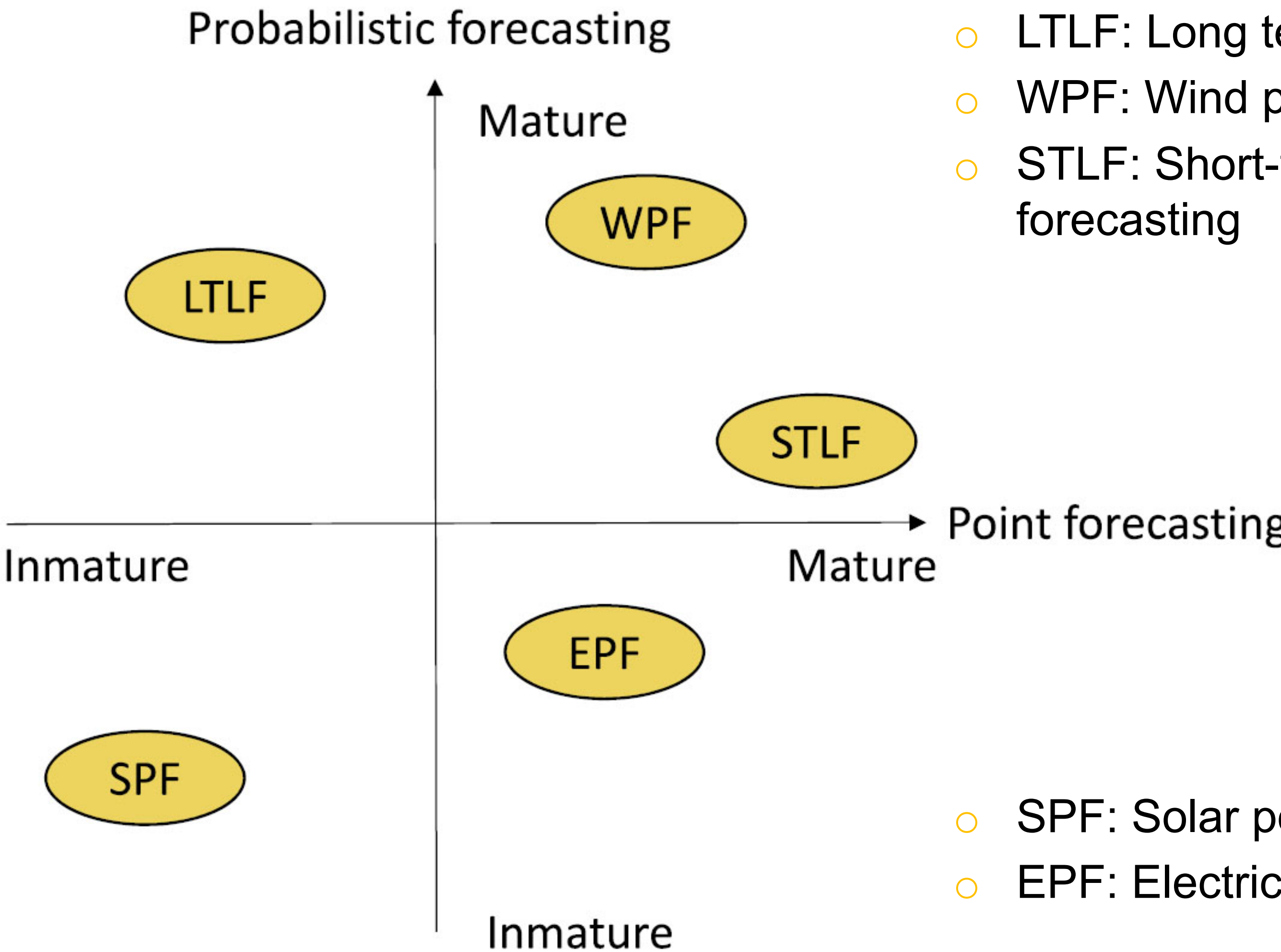
Number of journal papers in load forecasting



[Hong16]

PLF: Probabilistic Load Forecasting

Forecasting maturity in the energy sector



- LTLF: Long term load forecasting
- WPF: Wind power forecasting
- STLF: Short-term load forecasting
- SPF: Solar power forecasting
- EPF: Electricity price forecasting

[Hong16a]

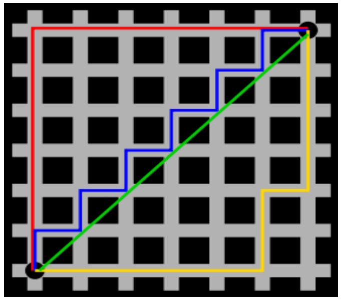
Probabilistic solar power forecasting model based on k-nearest neighbor and kernel density estimator

- Developed by Y. Zhang, J. Wang
- Goal: Improve decision making in power system operations
- Model approach:
  1. Using k-nearest neighbor algorithm to find days with similar weather conditions in historical dataset
  2. Calculating point forecast based on averaged corresponding power values
  3. Determining optimal weighting factors
  4. Deriving probability density by applying a kernel density estimator method

K-NN algorithm (1)

- Task: Find the k-closest training examples in the feature space
- Steps:
  - Calculating the distance between testing and training data
  - Choosing k nearest neighbors with the smalles distance
    - Weighted manhattan distance:

$$D[X, Y] = \sum_{i=1}^{n} w_i |x_i - y_i| ,$$



[Tax16
]

X ,Y : two instances from the training and testing data sets

x_i, y_i : input variables

n : number of input variables

w_i : weight assigned to the i-th input variable

K-NN algorithm (2)

- The instances X_1, ..., X_k with the smallest distances are the k nearest neighbors
- Point forecast is made based on the corresponding power observations p_1, ..., p_K
- Calculation by using a weighted exponential function

$$\hat{p} = \sum_{k=1}^{K} \delta_k p_k = \frac{\sum_{k=1}^{K} e^{-d_k} \cdot p_k}{\sum_{k=1}^{K} e^{-d_k}}.$$

δ_k : Weight associated with the instance X_k

p_k: Power output associated with the instance X_k

d_k: Distance associated with the instance X_k

Determining optimal weighting factors

- Use sum of square error measure to assess prediction performance

$$SSE = \sum_{i=1}^{m} (p^i - \hat{p}^i)^2$$

p^i : observed power output
p^{*i}: predicted power output

- Coordinate decent algorithm determines optimal weighting factors of the manhattan distance
- The algorithm minimizes of the sum of square error

Algorithm 1: Coordinate Descent

```
Loop until convergence
    For n = 1,2 ..., r
        ŵ_n = argmin_{w_n} SSE(w_1, ..., w_{n-1}, w_n, w_{n+1}, ... w_r)
    End For
End Loop
```

[Zhang14]

Kernel density estimator

- Derive probability density

$$\hat{f}(p) = \frac{1}{Kh}\sum_{k=1}^{K} \delta^k G\left(\frac{p-p^k}{h}\right) = \frac{1}{Kh}\sum_{k=1}^{K} \frac{e^{-d^k} G\left(\frac{p-p^k}{h}\right)}{\sum_{k=1}^{K} e^{-d^k}}$$

G(): kernel function, h: bandwidth

- The predictive density of the power output is converted into 99 quantiles
- for a specific quantile a:

$$q_a = \hat{F}^{-1}\left(\frac{a}{100}\right)$$

F^{*-1} : inversion of the cumulative distribution function

Application to a real world use case

- Data from the European Centre for medium range weather forecasts

Variable	Description	Abbreviate	Unit
VAR78	Total Column Liquid Water	TCLW	kg/m^2
VAR79	Total Column Ice Water	TCIW	kg/m^2
VAR134	Surface Pressure	SP	Pa
VAR157	Relative Humidity at 1000 mbar	R	%
VAR164	Total Cloud Cover	TCC	0-1
VAR165	10 Meter U Wind Component	10U	m/s
VAR166	10 Meter V Wind Component	10V	m/s
VAR167	2 Meter Temperature	2T	K
VAR169	Surface Solar Radiation Down	SSRD	J/m^2
VAR175	Surface Thermal Radiation Down	STRD	J/m^2
VAR178	Top Net Solar Radiation	TSR	J/m^2
VAR228	Total Precipitation	TP	m

- Input variables to the model selected via forward feature selection:

Stage	Feature Subset	Feature	Score Q
1	HOUR/VAR169	VAR169	0.013913
2	HOUR/VAR169/VAR79	VAR79	0.013708
3	HOUR/VAR169/VAR79/VAR78	VAR78	0.013410
4	HOUR/VAR169/VAR79/VAR78/VAR157	VAR157	0.013395

Deterministic forecasting results(1)

- Evaluation measure for the point forecast:

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(p^i - \hat{p}^i)^2}$$

p^i: observed power output

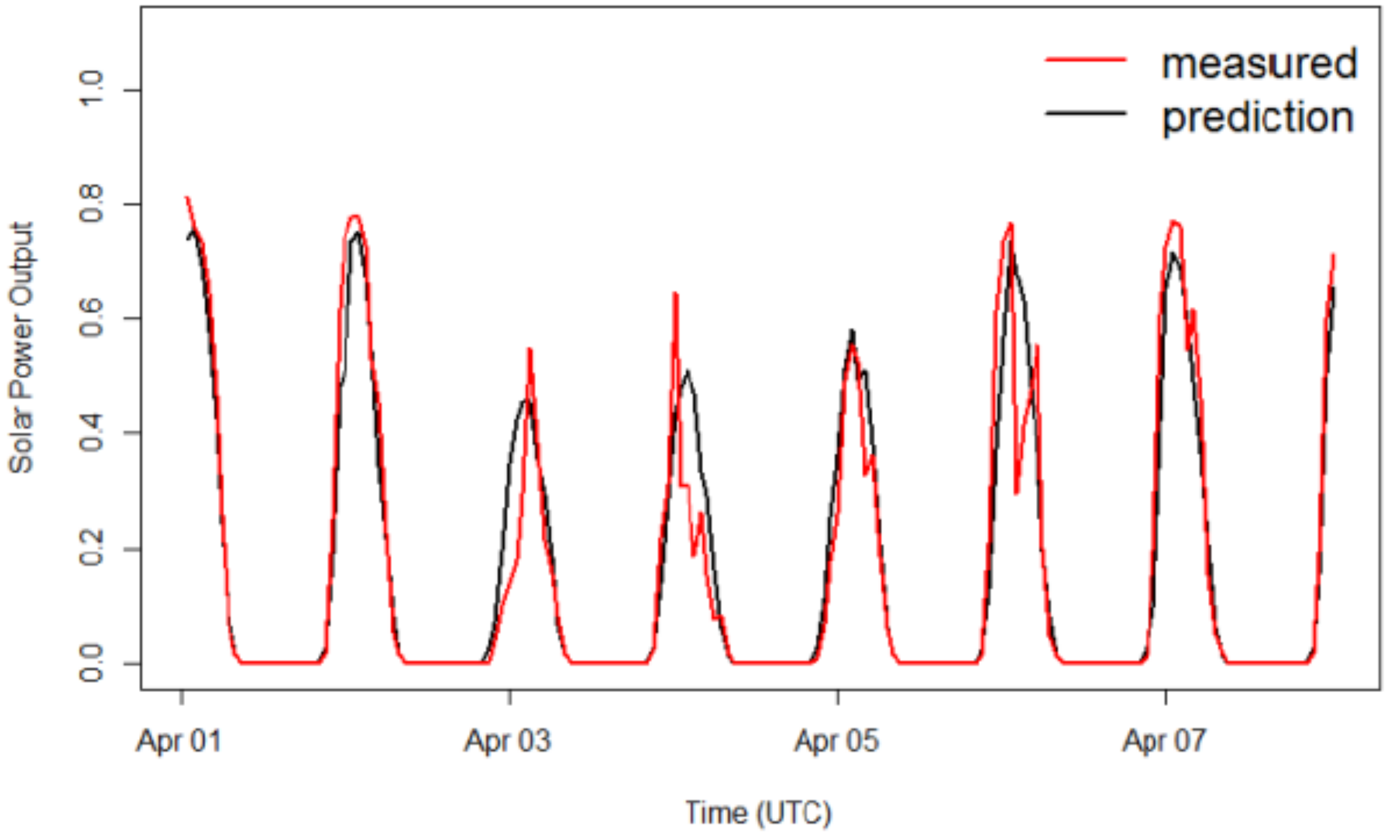
p^{*i} : predicted power output

- Evaluation results of point forecast:

Task	Period	RMSE		
		Farm #1	Farm #2	Farm #3
1	Apr 2013	0.088132	0.077563	0.075661
2	May 2013	0.071035	0.079000	0.068553
3	May 2013	0.076475	0.079952	0.083946
4	Jul 2013	0.101122	0.098185	0.092904
5	Aug 2013	0.089117	0.104147	0.126994
Average		0.085176	0.087769	0.089611
Standard Variance		0.011770	0.012438	0.022795

Deterministic forecasting results(2)

- Point forecasting of solar power output from April 1st to April 7th 2013



[Zhang14]

Probabilistic forecasting results(1)

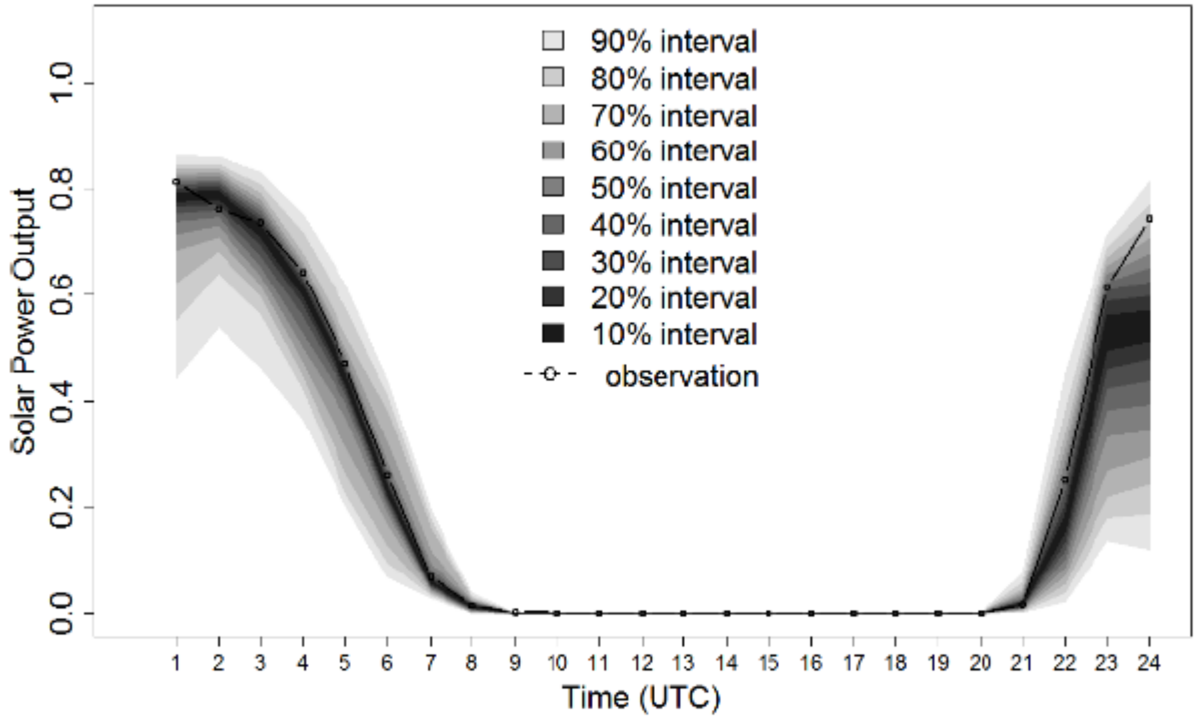
- Evaluation measure for probabilistic prediction:

$$Q(q_a, p^i) = \begin{cases} \left(1 - \frac{a}{100}\right)(q_a - p^i) & \text{if } p^i < q_a \\ \frac{a}{100}(p^i - q_a) & \text{if } p^i \geq q_a \end{cases}$$

- Evaluation results of probabilistic prediction:

Task	Period	Score Q		
		Farm #1	Farm #2	Farm #3
1	Apr 2013	0.013646	0.013064	0.013299
2	May 2013	0.010751	0.012373	0.010904
3	May 2013	0.011650	0.013272	0.013477
4	Jul 2013	0.015350	0.016171	0.014242
5	Aug 2013	0.013778	0.016880	0.021629
Average		0.013035	0.014352	0.014710
Standard Variance		0.001831	0.002027	0.004064

Probabilistic forecasting results(2)



[Zhang14]

Probabilistic forecasting of solar power output on April 1st 2013

Recap and outlook

- Benefits of accurate forecasts
- Benefits of probabilistic forecasting
- Forecasting techniques and input variables
- Probabilstic forecasting is a hot topic but is still in its infancy
- Especially for probabilistic solar energy forecasting there is room for improvements

Literature (1)

[Hong16] T. Hong, S. Fan: Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting 32 (2016) 914–938

[Hong16a] T. Hong, P. Pinson, S. Fan, H. Zareipour, A.Troccoli, R. Hyndman, Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond, International Journal of Forecasting 32 (2016) 896–913

[Antonanzas16] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de-Pison, F. Antonanzas-Torres: Review of photovoltaic power forecasting, Solar Energy 136 (2016) 78–111

[Tax16] https://en.wikipedia.org/wiki/Taxicab_geometry

[Zhang14] Y. Zhang, J. Wang: GEFCom2014 Probabilistic Solar Power Forecasting based on k-Nearest Neighbor and Kernel Density Estimator

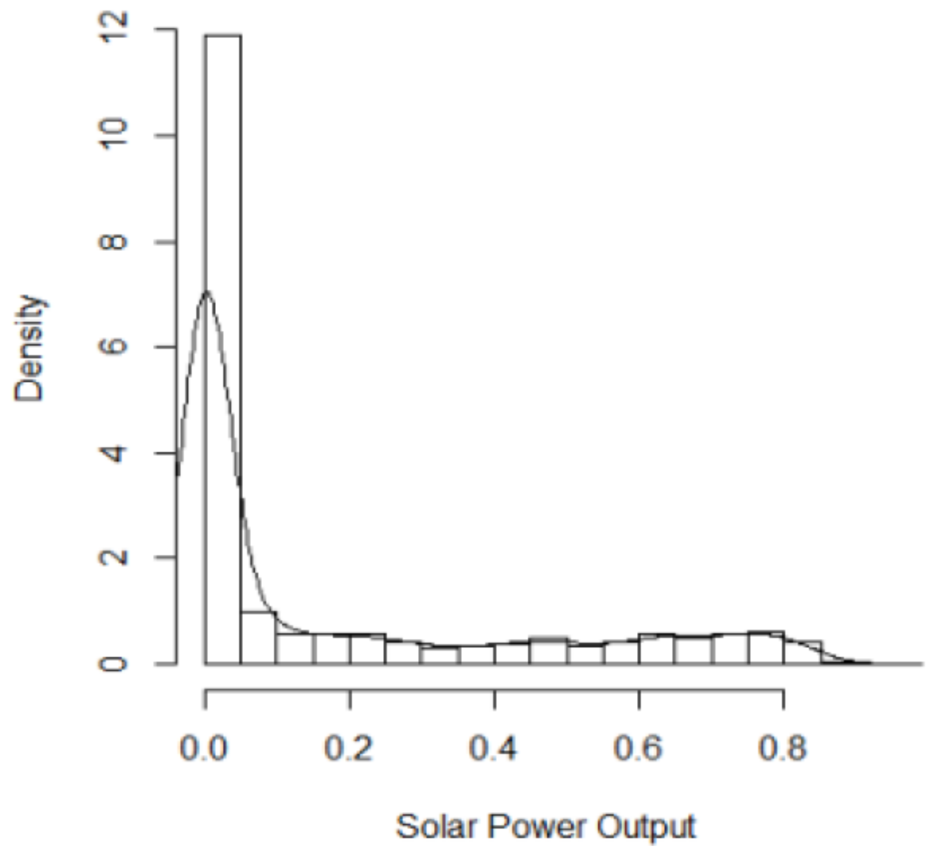
Literature (2)

[Wer14] F. Weron: Electricity price forecasting: A review of the state-of-the-art with a look into the future, International Journal of Forecasting 30 (2014) 1030–1081

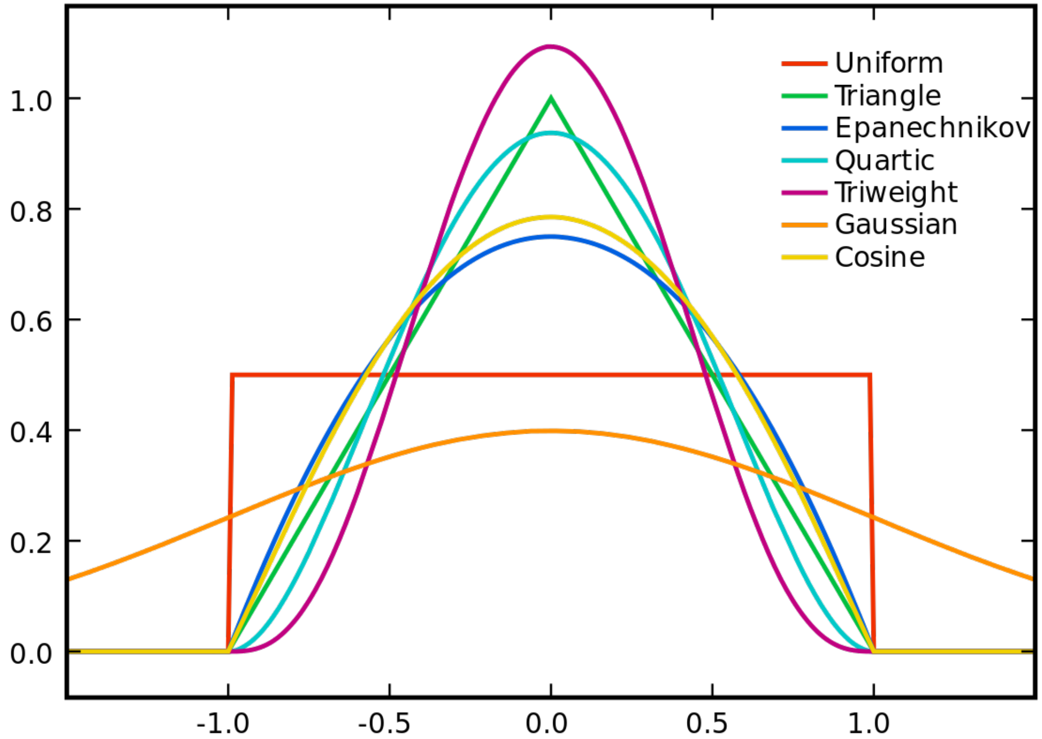
Back up: Coordinate decent algorithm

$$\underset{w_1,w_2,\dots,w_r}{\operatorname{argmin}} \text{SSE} = \underset{w_1,w_2,\dots,w_r}{\operatorname{argmin}} \sum_{i=1}^{m} [p^i - \hat{p}^i(w_1, w_2, \dots, w_r)]^2$$

Histogram of solar power output



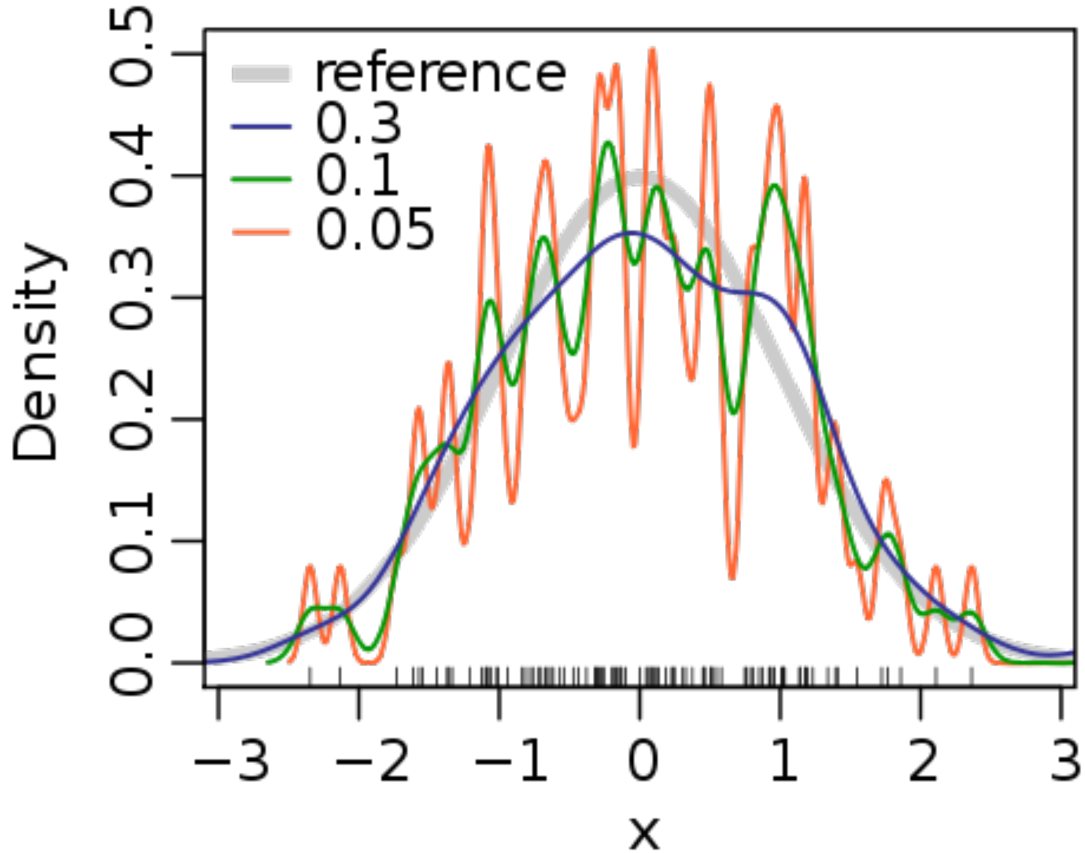
Commonly used kernel functions



Gaussian Kernel

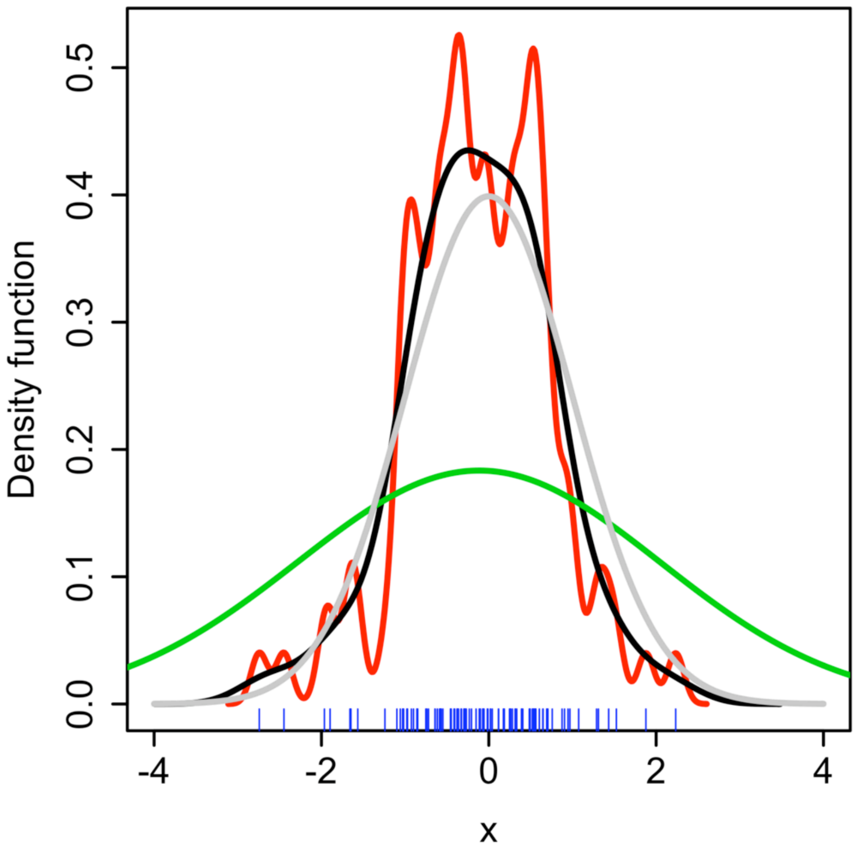
$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

Kernel density estimation of 100 normally distributed random numbers



different smoothing bandwidths are used

Kernel density estimation of 100 normally distributed random numbers



Grey: true density (standard normal)

Red: KDE with h=0.05

Black: KDE with h=0.337

Green: KDE with h=2

Persistence model

- $P_p(t + f_h) = P(t)$

- f_h: forecast horizon
- P: power output
- Applicable if no change of conditions between t and $t + f_h$