Universidad Carlos III de Madrid

Wind Power Forecasting for Efficient System Integration

Ismael Sánchez
Department of Statistics
Universidad Carlos III de Madrid

Wind Power Forecasting for Efficient System Integration

1. Introduction
2. Overview of SIPREOLICO
3. Models for wind power prediction
4. Forecast combination
5. Regional forecasting
6. Probabilistic forecasts

Necessity of Wind Power Predictions

- Wind power is a non-dispatchable energy. Generation cannot be correlated to load.
- In Spanish peninsular System, during periods of low demand, wind power can represent values near 40% coverage of demand.
- The high variability of wind power sometimes generates extreme situations that should be anticipated.
- Wind power forecasting methods become necessary for the proper operation and stability of the system.

Interest in Wind Power Forecasts.

- System Operator:
  - Improve safety and quality of supply
  - Schedule spinning reserve
  - Identification of potential bottlenecks
- Market agents:
  - Adoption of strategies for daily and intraday markets
  - Some of them have a large number of installed wind farms

2-OVERVIEW OF SIPREÓLICO

RED ELÉCTRICA DE ESPAÑA & UNIVERSIDAD DE MADRID · CARLOS III ·

"SIPREÓLICO"

'Short term' statistical prediction tool (up to 48 hours)

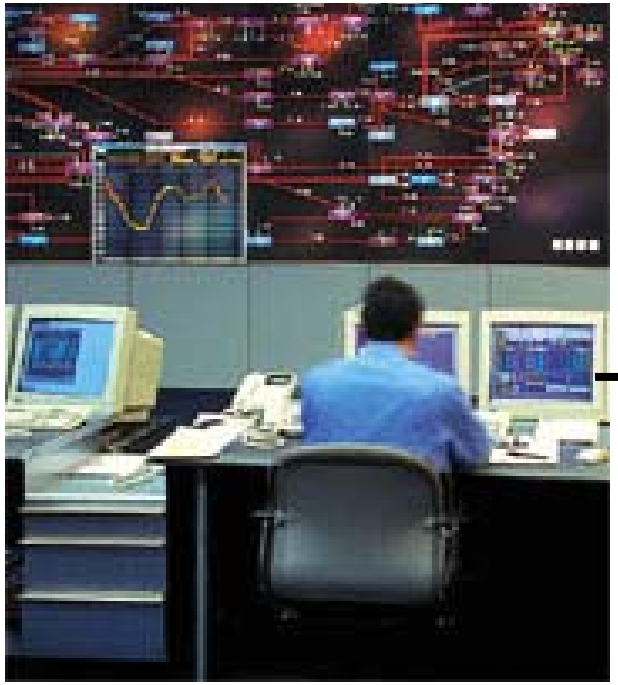
- On-line wind power predictions of individual wind farms
- Every hour, Sipreólico updates the hourly predictions of more than 600 wind farms for the next 48 hours, using on-line production data and meteorological predictions.
- Too many wind farms to allow for individual model building
- The number of wind farms have been increasing

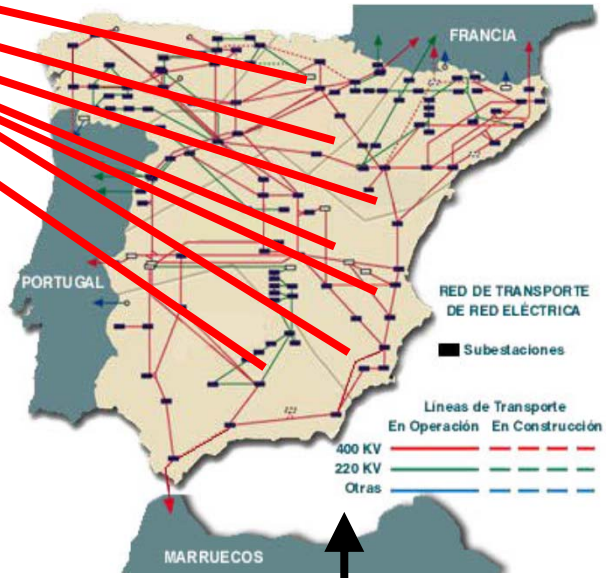
2-OVERVIEW OF SIPREÓLICO

Meteorological Predictions

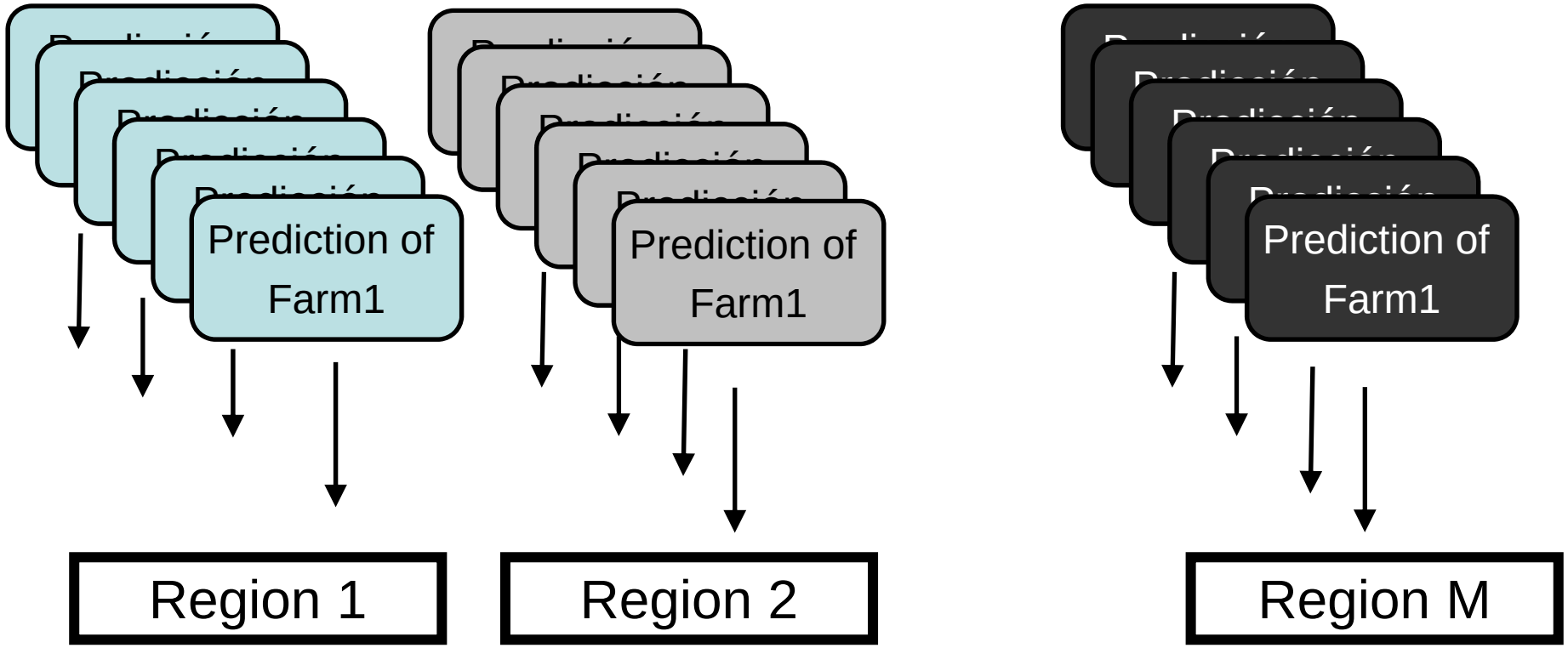
Online production Data

System Operation





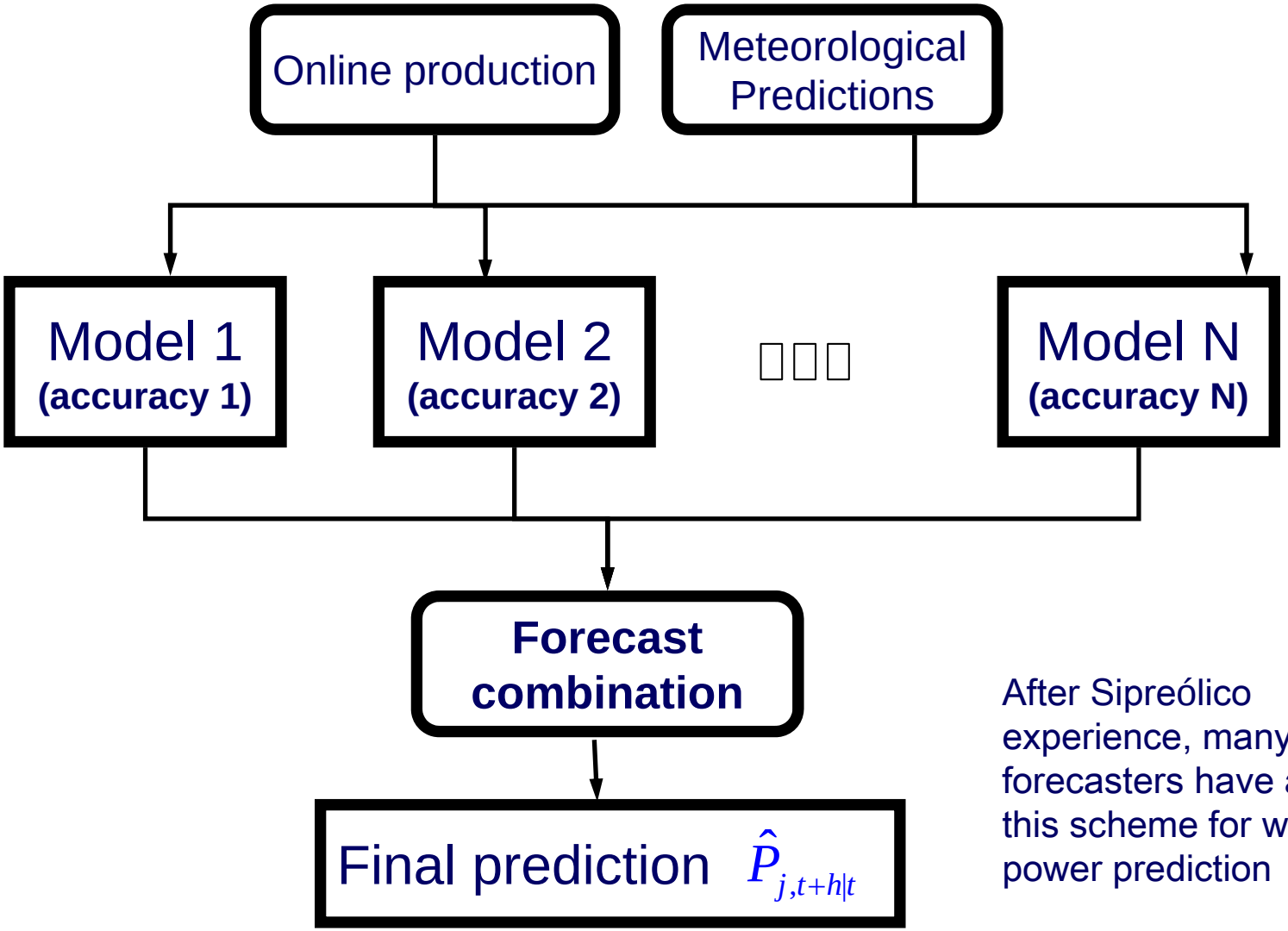
2-OVERVIEW OF SIPREÓLICO



The predictions of the individual wind farms are aggregated according to several criteria (REE), for management purposes.

Also the Total Spanish Peninsular prediction is obtained by adding up the predictions (an alternative approach will be presented).

2-OVERVIEW OF SIPREÓLICO

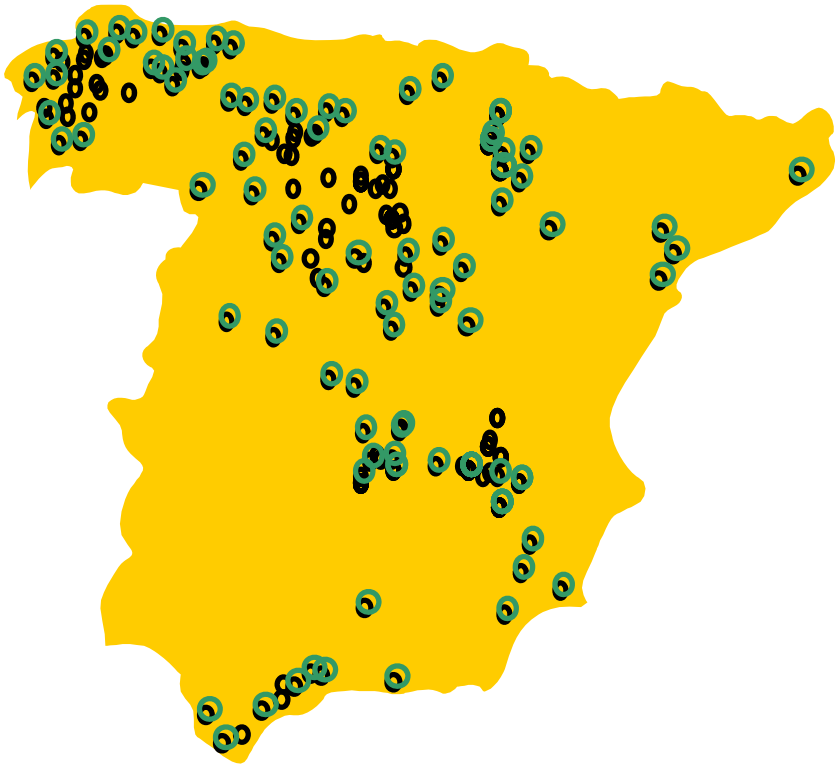


After Sipreólico experience, many forecasters have adopted this scheme for wind power prediction

2-OVERVIEW OF SIPREÓLICO

Meteorological Predictions

- **Wind speed (meters/second) and direction (degrees) at selected points.**
- **Hourly predictions for next 48 hours at more than 100 coordinates.**
- **Should be provided by a meteorological agency.**
- **Wind predictions are, nowadays, of <u>low quality.</u>**
- **REE buy the information from TWO independent meteorological agencies, using different sources of information, to gain accuracy.**



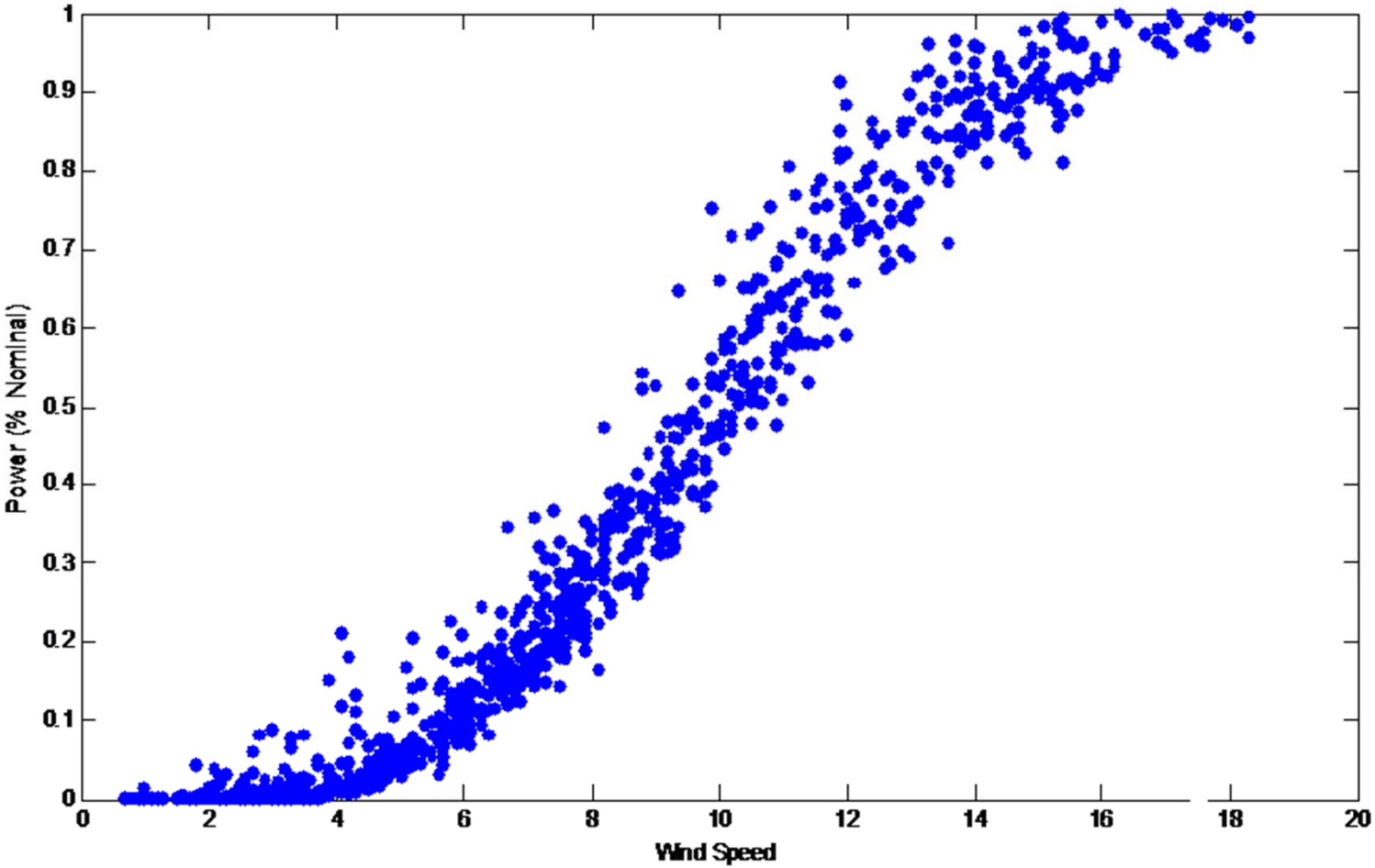
2-OVERVIEW OF SIPREÓLICO

Online production

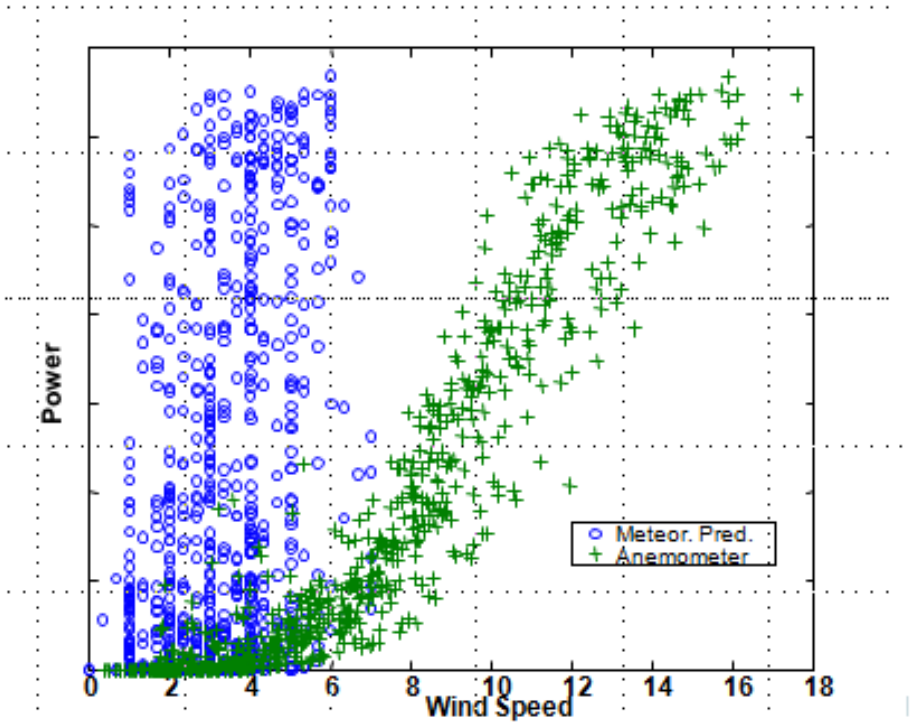
- **More than 600 wind farms**
- **The number of wind farms is constantly increasing: we have to be ready to predict new wind farms (with small datasets)**
- **Too many wind farms: it is not feasible to analyze them one by one**

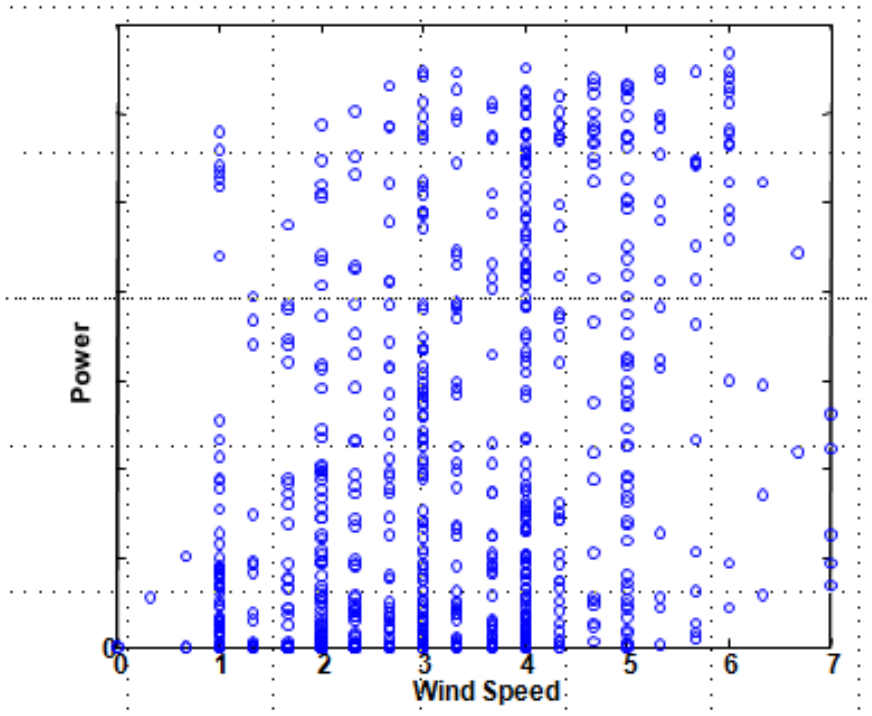
2-OVERVIEW OF SIPREÓLICO

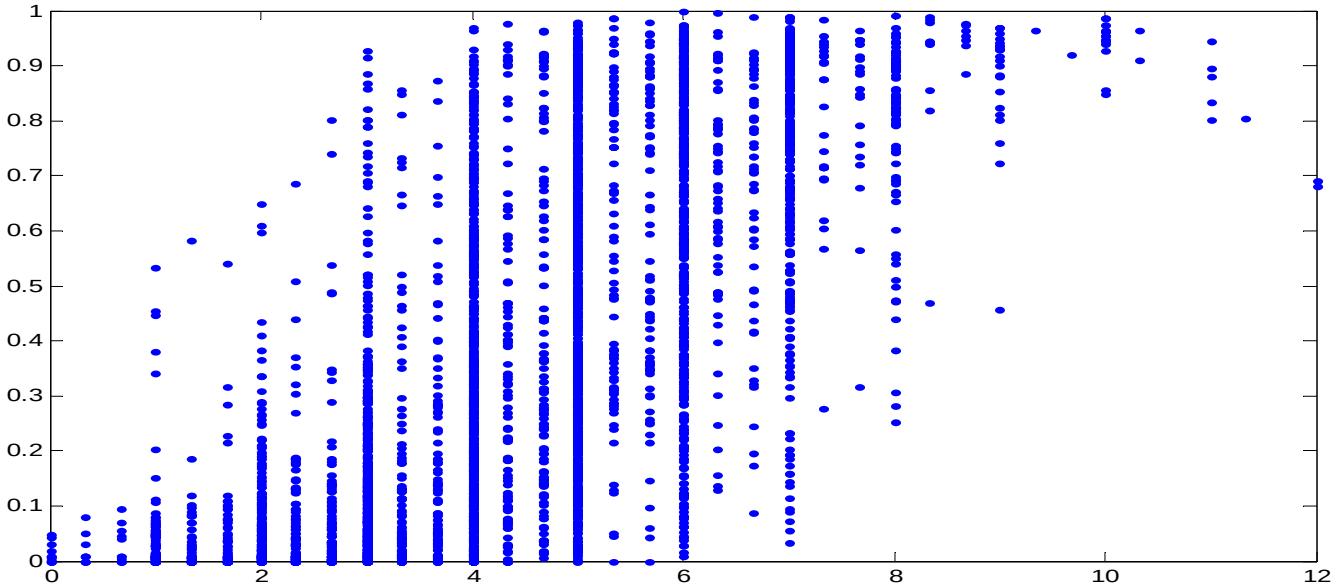
Using real wind, the relationship with contemporaneous power is clear. However, we should use predicted wind

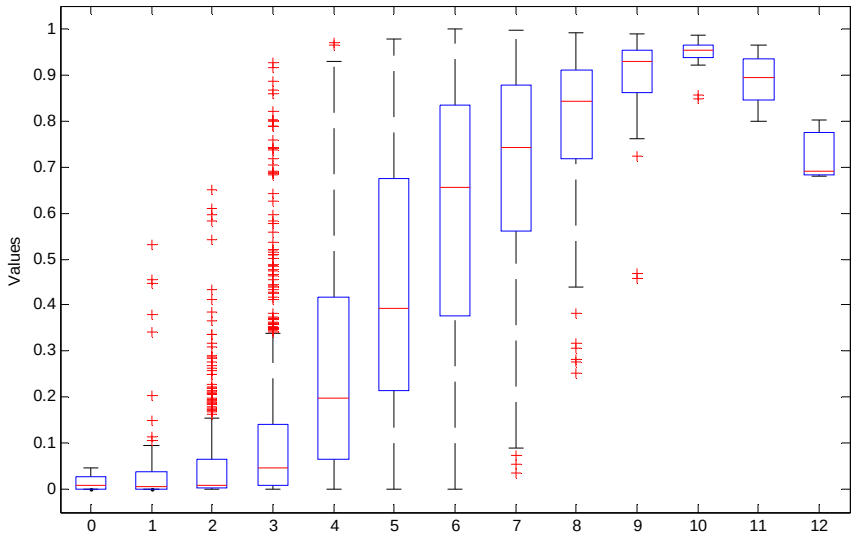


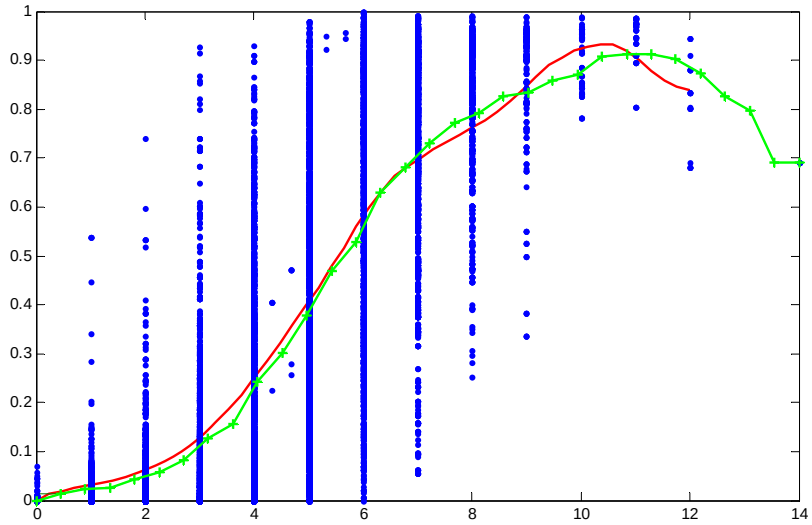
Wind Speed predictions are very poor.
Meteorological prediction for wind power forecasting is a difficult task



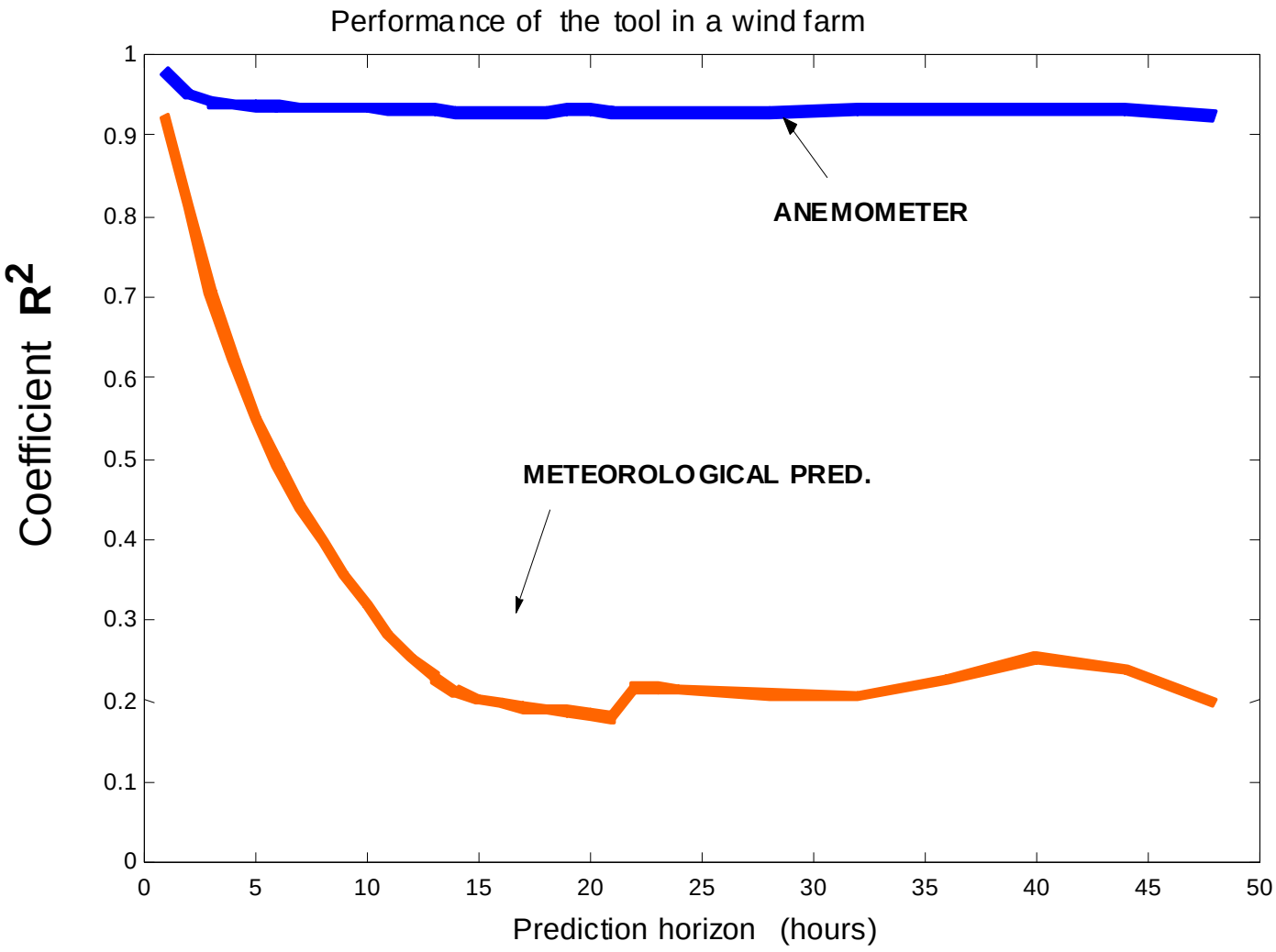



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12

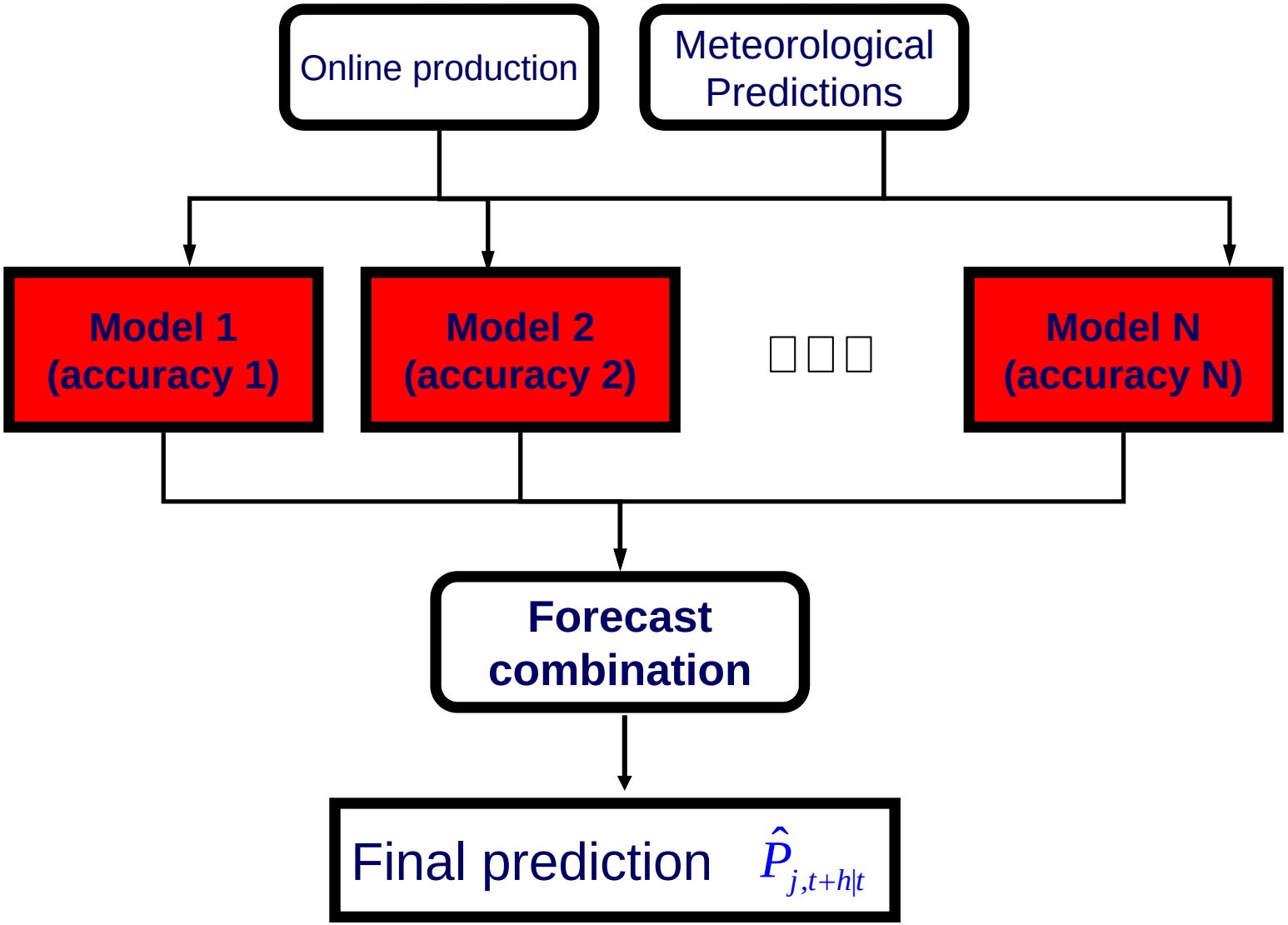

Values
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
8
9
10
11
12


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14

2-OVERVIEW OF SIPREÓLICO



3-MODELS FOR WIND POWER PREDICTION



3-MODELS FOR WIND POWER PREDICTION

Model 1 (accuracy 1) | **Model 2 (accuracy 2)** | □□□ | **Model N (accuracy N)**

Alternative models are running in parallel

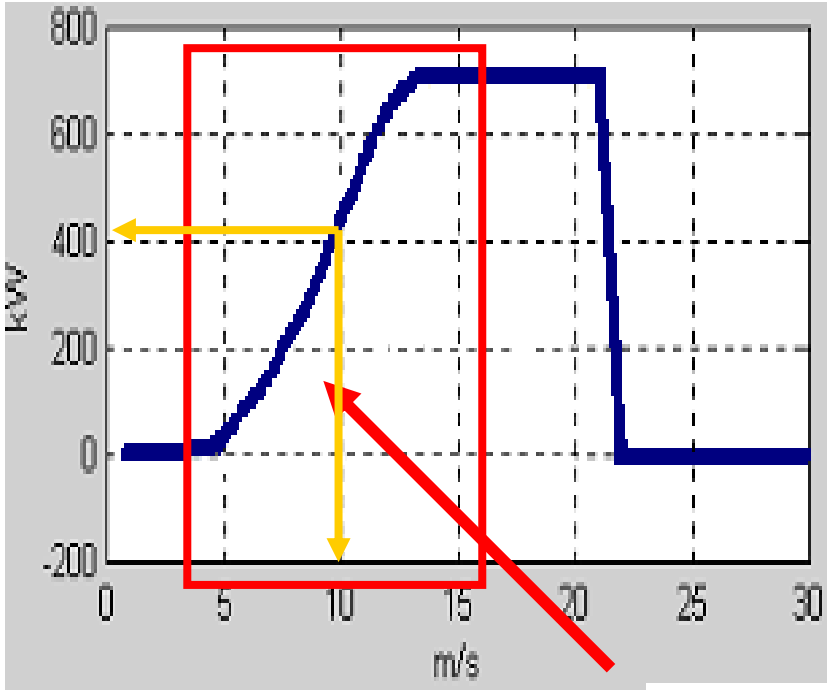
- Flexibility to adapt to different wind farms
- Flexibility to predict from short to long term
- Flexibility to situations with small or large data sets
- Adaptive and recursive estimation

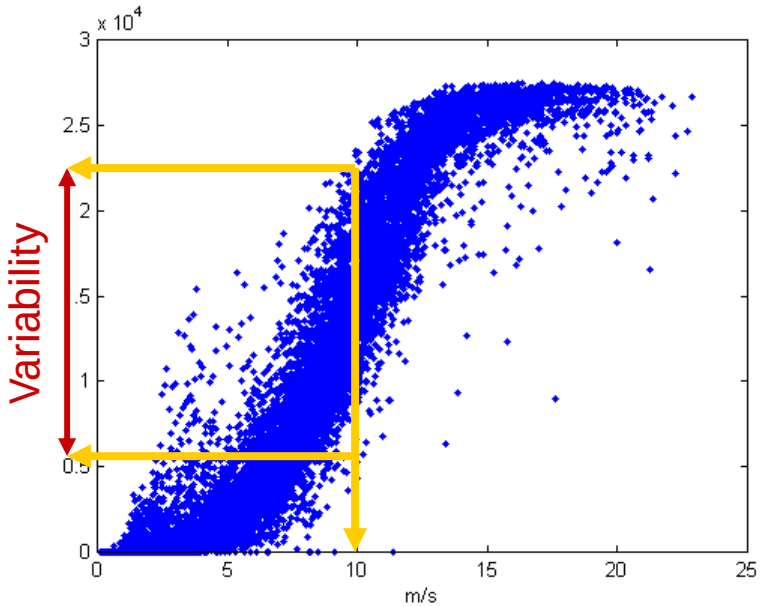
- Autoregressive models using ***on-line measurements***
- Parametric dynamic models with wind speed and direction
- Nonparametric (local regression) models with wind speed and direction

3-MODELS FOR WIND POWER PREDICTION

Main foundation for wind power models:

Theoretical power curve for a wind turbine





$$p = \frac{1}{2}\rho A v^3$$

Parametric Models

Model 3

Model 2

Model 1 Model 4 …

$$P_{t+H} = f\left\{P_t, P_{t-1}, P_{t-2}, \ldots; \hat{V}_{t+H}; \sin(\hat{\phi}_{t+H}), \cos(\hat{\phi}_{t+H})\right\}$$

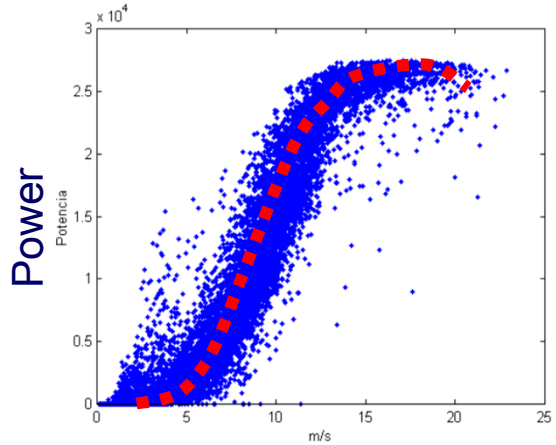
Useful for 1-4 hours ahead: Persistence, AR(2)

Predicted wind speed

$$\hat{V}_{t+h|t}, \hat{V}^2_{t+h|t}, \hat{V}^3_{t+h|t},$$

Predicted wind direction

3-MODELS FOR WIND POWER PREDICTION- Parametric ones



Some examples…

$$p_{t+h} = \alpha_0 + \alpha_1 \hat{v}_{t+h|t} + \alpha_2 \hat{v}^2_{t+h|t} + a_t$$

Wind speed predictions for t+h available at time t

$$p_{t+h} = \alpha_0 + \alpha_1 \hat{v}_{t+h|t} + \alpha_2 \hat{v}^2_{t+h|t} + \beta_1 \sin\left(\frac{2\pi\hat{\phi}_{t+h|t}}{360}\right) + \beta_2 \cos\left(\frac{2\pi\hat{\phi}_{t+h|t}}{360}\right) + a_t$$

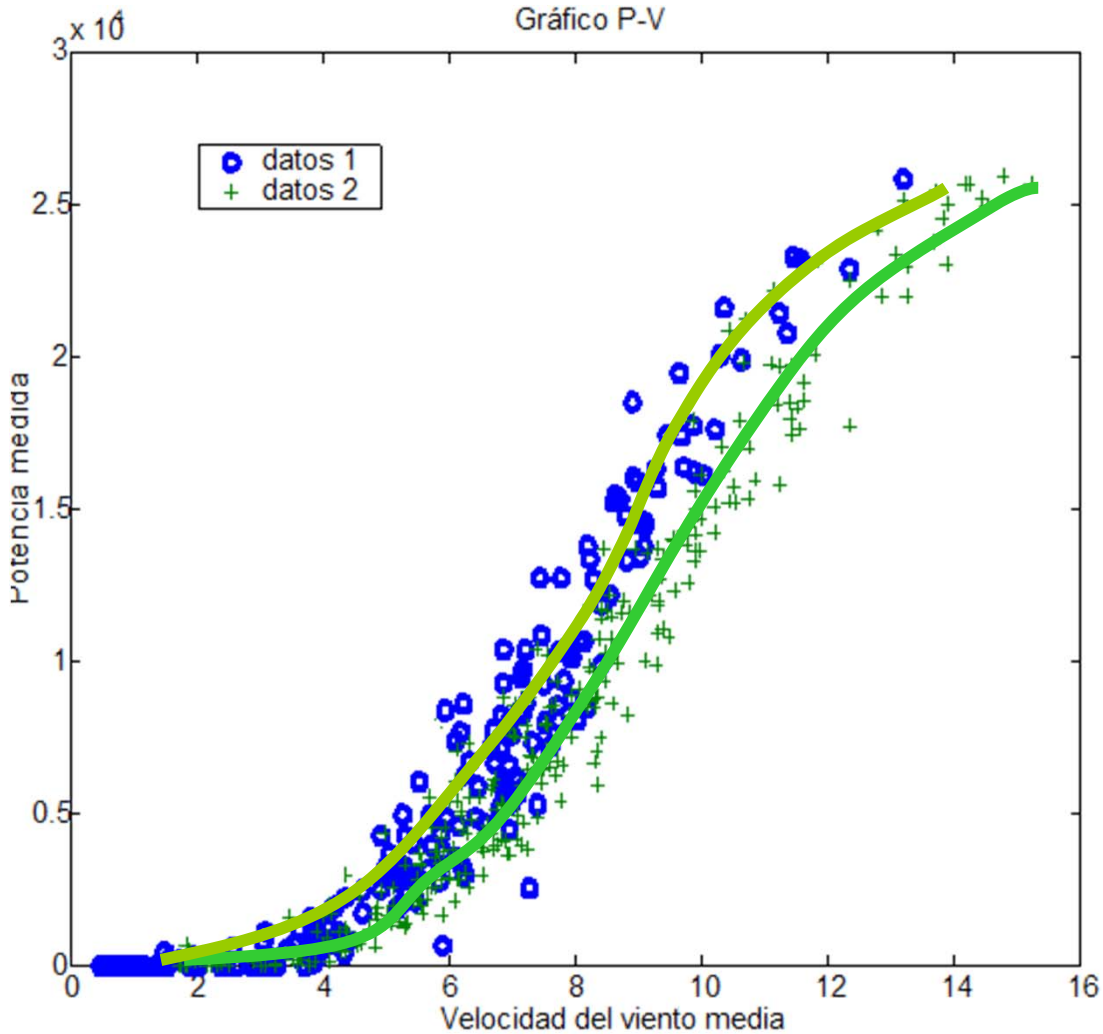
Wind speed directions for t+h available at time t

ARX: $$p_{t+h} = \alpha_0 + \phi_1 p_t + \phi_2 p_{t-1} + \alpha_1 \hat{v}_{t+h|t} + \alpha_2 \hat{v}^2_{t+h|t} + a_t$$

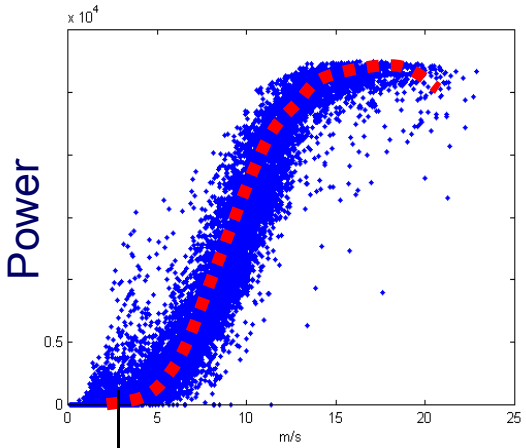
3-MODELS FOR WIND POWER PREDICTION- Parametric ones

Time-varying relationships are observed
(it is usual with long series)

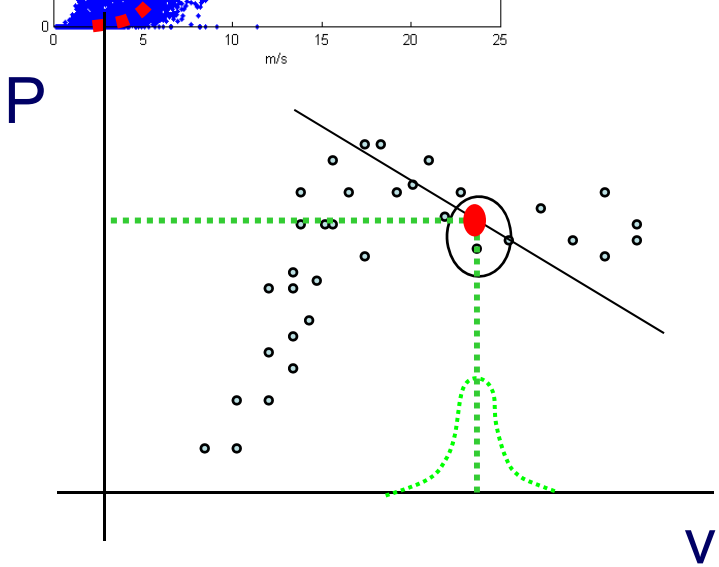
Time-varying estimation of parameters



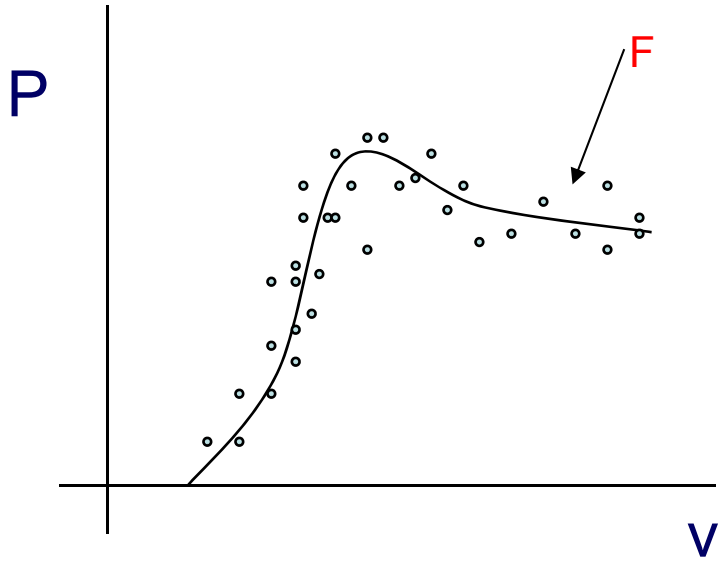
3-MODELS FOR WIND POWER PREDICTION- Non-parametric models



Nonparametric models: several methodologies



For each point we fit a local line weighting the data acording to the distance to the point

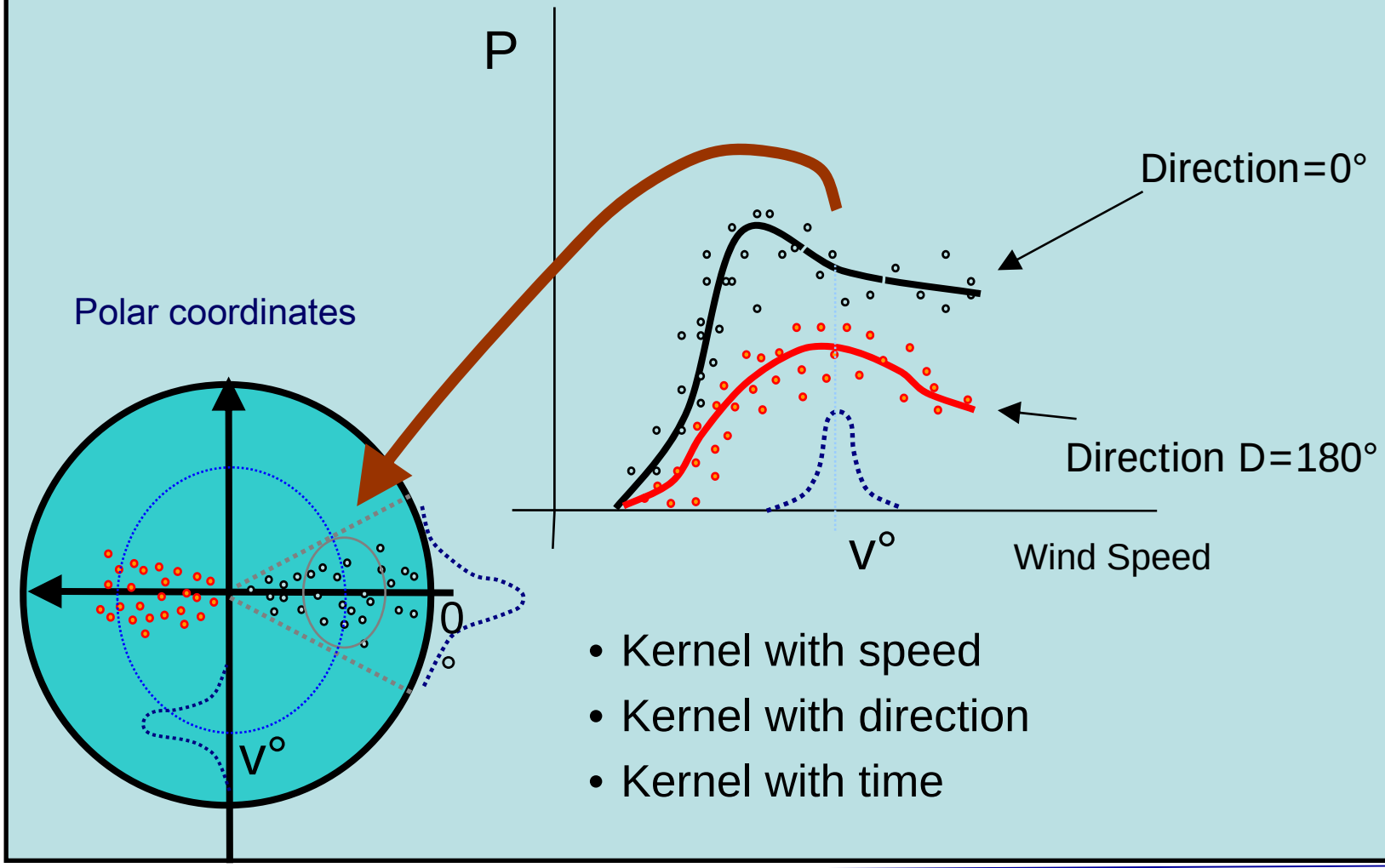


We repeat for all the desired points

$$p_{t+h} = \alpha_0 + \phi_1 p_t + \phi_2 p_{t-1} + \alpha_1 F_t + a_t$$

mixed model

Nonparametric models



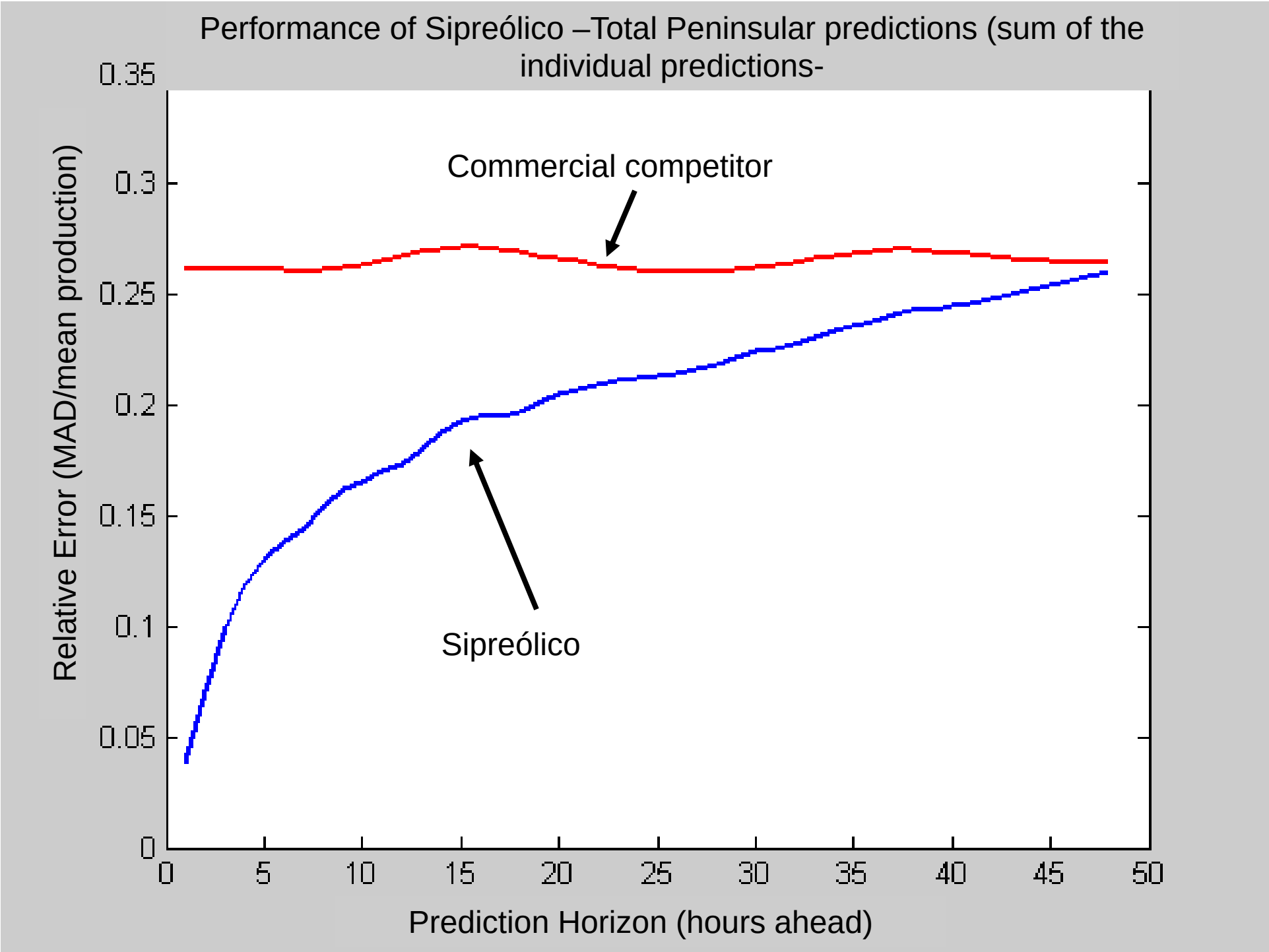
- Kernel with speed
- Kernel with direction
- Kernel with time

3-MODELS FOR WIND POWER PREDICTION- Non-parametric models

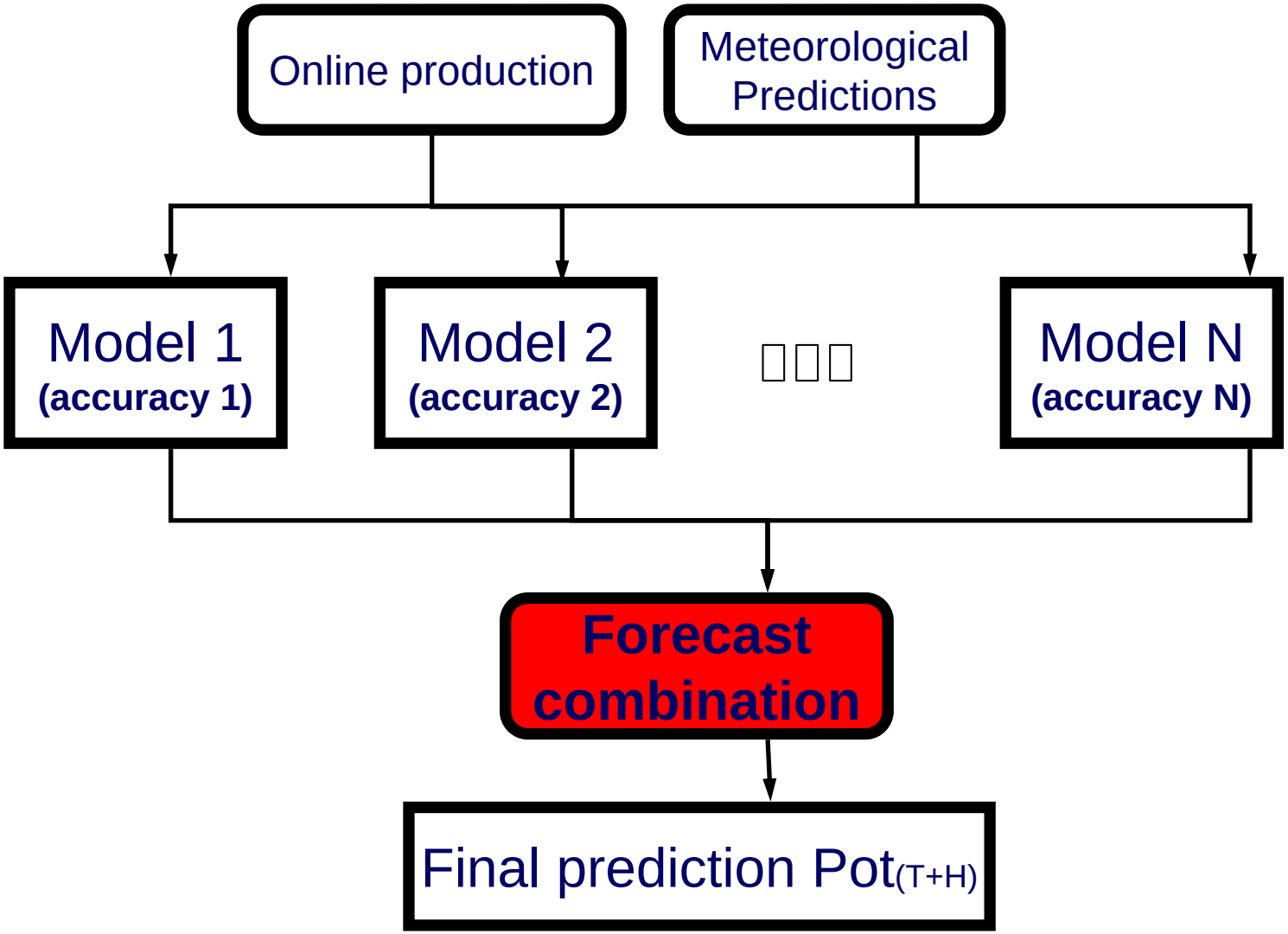
- Parametric models need smaller data sets
- Nonparametric models are the most accurate if data set is large enough
- The combination of the competing models assigns weights according to their accuracy

The best models

- Persistence and autoregressive models at 1-4 hours (requires on-line data)
- Dynamic ARX parametric models at 4-10 hours (requires on-line data)
- Models with only wind for more than 10 hours ahead
- Parametric models if the wind farm is new (small data set)
- Nonparametric models with several months of data


Performance of Sipreólico –Total Peninsular predictions (sum of the individual predictions-
Commercial competitor
Sipreólico
Relative Error (MAD/mean production)
Prediction Horizon (hours ahead)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
35
40
45
50

4-FORECAST COMBINATION



The combination model

Let p_{t+h} be the future power

Let $p^{(1)}_{t+h|t},\ldots,p^{(K)}_{t+h|t}$ be the alternative lead h predictors

Let $e^{(1)}_{t+h|t},\ldots,e^{(K)}_{t+h|t}$ be the alternative prediction errors

$$\hat{p}^{C}_{t+h|t} = \hat{\alpha}_{0,t} + \sum_{k=1}^{K} \hat{\alpha}_{k,t}\hat{p}^{(k)}_{t+h|t}$$

$\hat{\alpha}_{k,t}(h)$ is time-varying, horizon-dependent, and estimated on-line using some recursive algorithm

4-FORECAST COMBINATION

$$\hat{p}^{C}_{t+h|t} = \hat{\alpha}_{0,t} + \sum_{k=1}^{K} \hat{\alpha}_{k,t} \hat{p}^{(k)}_{t+h|t}$$

There are two different combination strategies (Yang, 2004)

1.Combination for improvement

All the alternative predictions have some amount of independent information.

We should use them all in a linear combination to merge all the information

↓

Regression-like procedures

2.Combination for adaptation

We are only interested in selecting the best predictor. But the best predictor is time varying.

↓

Adaptive selection

1. Combination for improvement

Regression-based Combination (Bates and Granger, 1969…)

$$\hat{p}^{C}_{t+h|t} = \hat{\alpha}_{0,t} + \sum_{k=1}^{K} \hat{\alpha}_{k,t} \hat{p}^{(k)}_{t+h|t}$$

Statistical methods:

- 'Optimal combination' by Bates & Granger, based on the covariance matrix of the prediction errors
- Dynamic regression with/without restrictions in the value of the coefficients
- Kalman Filter

Example: Golagh -Irish wind farm-

9 competing forecasters P1 to P9. Two years of hourly data

Minimum		Deviation from minimum								
H	RMSE	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
1	0.04	0.13	0.05	0.10	0.00	0.05	0.11	0.13	0.14	0.05
4	0.13	0.03	0.01	0.01	0.00	0.01	0.01	0.03	0.04	0.01
7	0.14	0.03	0.01	0.00	0.00	0.01	0.01	0.02	0.03	0.01
10	0.15	0.03	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.00
13	0.15	0.03	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.00
16	0.16	0.03	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.00
19	0.17	0.03	0.01	0.00	0.01	0.00	0.01	0.02	0.02	0.00
22	0.17	0.03	0.02	0.01	0.01	0.00	0.01	0.02	0.02	0.01
25	0.17	0.03	0.02	0.00	0.01	0.00	0.01	0.02	0.03	0.01
28	0.18	0.03	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.01
31	0.18	0.03	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.01
34	0.19	0.03	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.01
41	0.19	0.04	0.01	0.01	0.01	0.01	0.00	0.03	0.02	0.01

best predictors

Alternative Combination-for-improvement methods (regression-based)

H	Minimum RMSE	Deviation from minimum C_1	C_2	C_3	C_4
1	0.04	-0.03	-0.03	-0.03	0.02
2	0.10	-0.00	-0.01	-0.01	0.02
4	0.13	0.01	0.00	0.00	0.00
7	0.14	0.01	0.01	0.01	-0.00
10	0.15	0.01	0.01	0.01	-0.00
13	0.15	0.01	0.01	0.01	-0.00
16	0.16	0.02	0.01	0.01	-0.01
19	0.17	0.04	0.02	0.01	-0.01
22	0.17	0.06	0.02	0.01	-0.01
25	0.17	0.07	0.02	0.02	-0.00
28	0.18	0.07	0.02	0.02	-0.01
31	0.18	0.05	0.02	0.02	-0.00
34	0.19	0.05	0.01	0.01	-0.00
41	0.19	0.05	0.01	0.02	-0.00

Example: Golagh -Irish wind farm-

H	Minimum RMSE	Deviation from minimum C_1	C_2	C_3	C_4
1	0.04	-0.03	-0.03	-0.03	0.02
2	0.10	-0.00	-0.01	-0.01	0.02
4	0.13	0.01	0.00	0.00	0.00
7	0.14	0.01	0.01	0.01	-0.00
10	0.15	0.01	0.01	0.01	-0.00
				0.01	-0.00
				0.01	-0.01
				0.01	-0.01
				0.01	-0.01
				0.02	-0.00
28	0.18	0.07	0.02	0.02	-0.01
31	0.18	0.05	0.02	0.02	-0.00
34	0.19	0.05	0.01	0.01	-0.00
41	0.19	0.05	0.01	0.02	-0.00

The success depends on the covariance structure of the competing predictions

2. Combination for adaptation

$$\hat{p}^{C}_{t+h|t} = \sum_{k=1}^{K} \hat{\alpha}_{k,t} \hat{p}^{(k)}_{t+h|t}$$

$$\sum_{i=1}^{K} \hat{\alpha}_{k,t} = 1$$

Adaptive Exponential Combination **(AEC)** **(Sánchez 2008)**

- Tend to select the best predictor. Weights tend to 0 or 1
- Time-varying version of Yang (2004)

Example: Golagh -Irish wind farm-

	Minimum	Deviation from minimum				
H	RMSE	C_1	C_2	C_3	C_4	AEC
1	0.04	-0.03	-0.03	-0.03	0.02	0.00
2	0.10	-0.00	-0.01	-0.01	0.02	-0.00
4	0.13	0.01	0.00	0.00	0.00	0.00
7	0.14	0.01	0.01	0.01	-0.00	0.00
10	0.15	0.01	0.01	0.01	-0.00	0.00
13	0.15	0.01	0.01	0.01	-0.00	-0.00
16	0.16	0.02	0.01	0.01	-0.01	0.00
19	0.17	0.04	0.02	0.01	-0.01	-0.00
22	0.17	0.06	0.02	0.01	-0.01	-0.00
25	0.17	0.07	0.02	0.02	-0.00	-0.00
28	0.18	0.07	0.02	0.02	-0.01	-0.00
31	0.18	0.05	0.02	0.02	-0.00	-0.00
34	0.19	0.05	0.01	0.01	-0.00	-0.00
41	0.19	0.05	0.01	0.02	-0.00	-0.00

What is the best combination method?

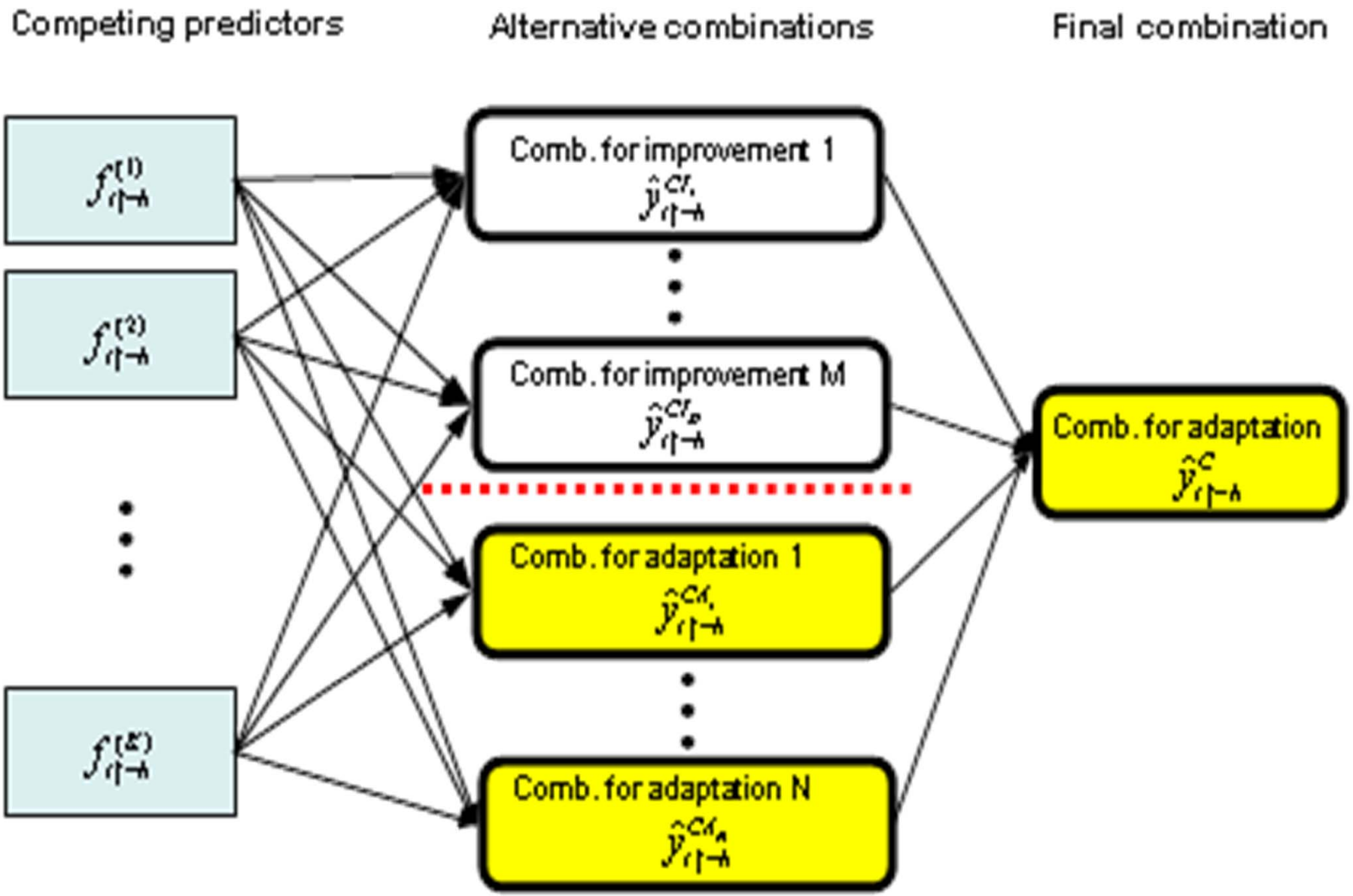
Combination Strategy: 2 steps

Step 1: Use alternative combination methods, with AEC among them

↓

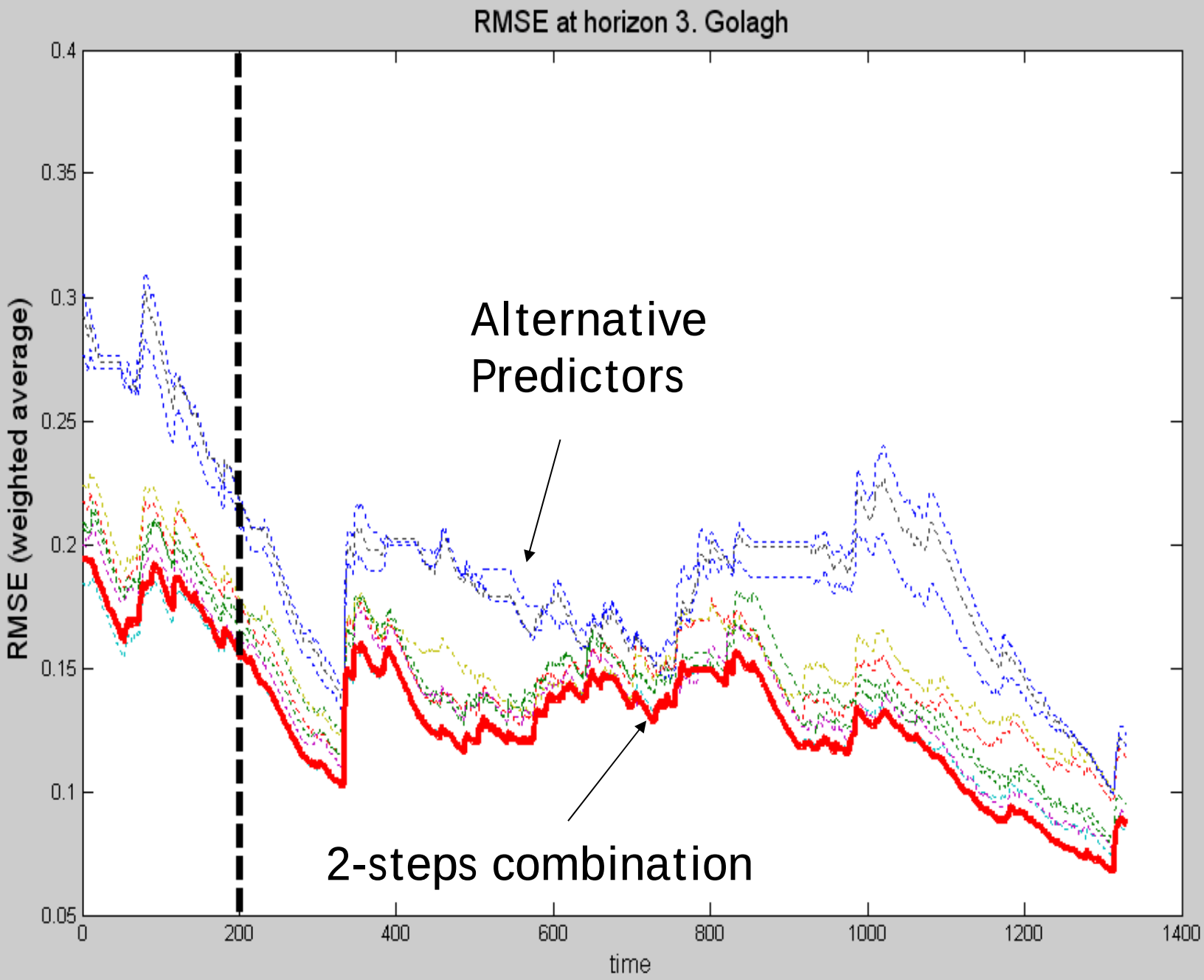
Alternative combined predictions

Step 2: Combine the combined predictions using AEC (Comb. for adaptation)


Competing predictors
Alternative combinations
Final combination
Comb. for improvement 1
Comb. for improvement M
Comb. for adaptation 1
Comb. for adaptation N
Comb. for adaptation

STEP 1 STEP 2

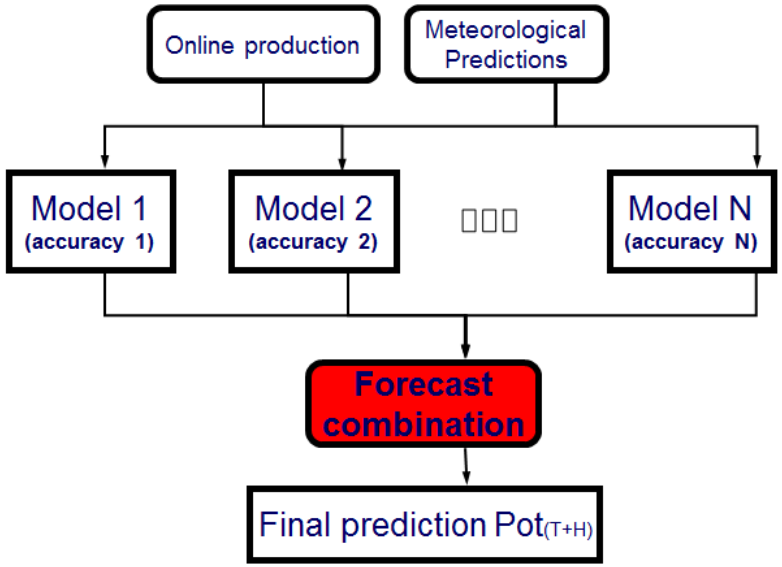
H	Minimum RMSE	Deviation from minimum C_1	C_2	C_3	C_4	AEC	AEC
1	0.04	-0.03	-0.03	-0.03	0.02	0.00	-0.03
2	0.10	-0.00	-0.01	-0.01	0.02	-0.00	-0.01
4	0.13	0.01	0.00	0.00	0.00	0.00	0.00
7	0.14	0.01	0.01	0.01	-0.00	0.00	-0.00
10	0.15	0.01	0.01	0.01	-0.00	0.00	-0.00
13	0.15	0.01	0.01	0.01	-0.00	-0.00	-0.01
16	0.16	0.02	0.01	0.01	-0.01	0.00	0.00
19	0.17	0.04	0.02	0.01	-0.01	-0.00	-0.01
22	0.17	0.06	0.02	0.01	-0.01	-0.00	-0.01
25	0.17	0.07	0.02	0.02	-0.00	-0.00	-0.00
28	0.18	0.07	0.02	0.02	-0.01	-0.00	-0.01
31	0.18	0.05	0.02	0.02	-0.00	-0.00	-0.00
34	0.19	0.05	0.01	0.01	-0.00	-0.00	-0.01
41	0.19	0.05	0.01	0.02	-0.00	-0.00	-0.02


RMSE at horizon 3. Golagh
RMSE (weighted average)
time
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
200
400
600
800
1000
1200
1400
Alternative
Predictors
2-steps combination

What do we combine?

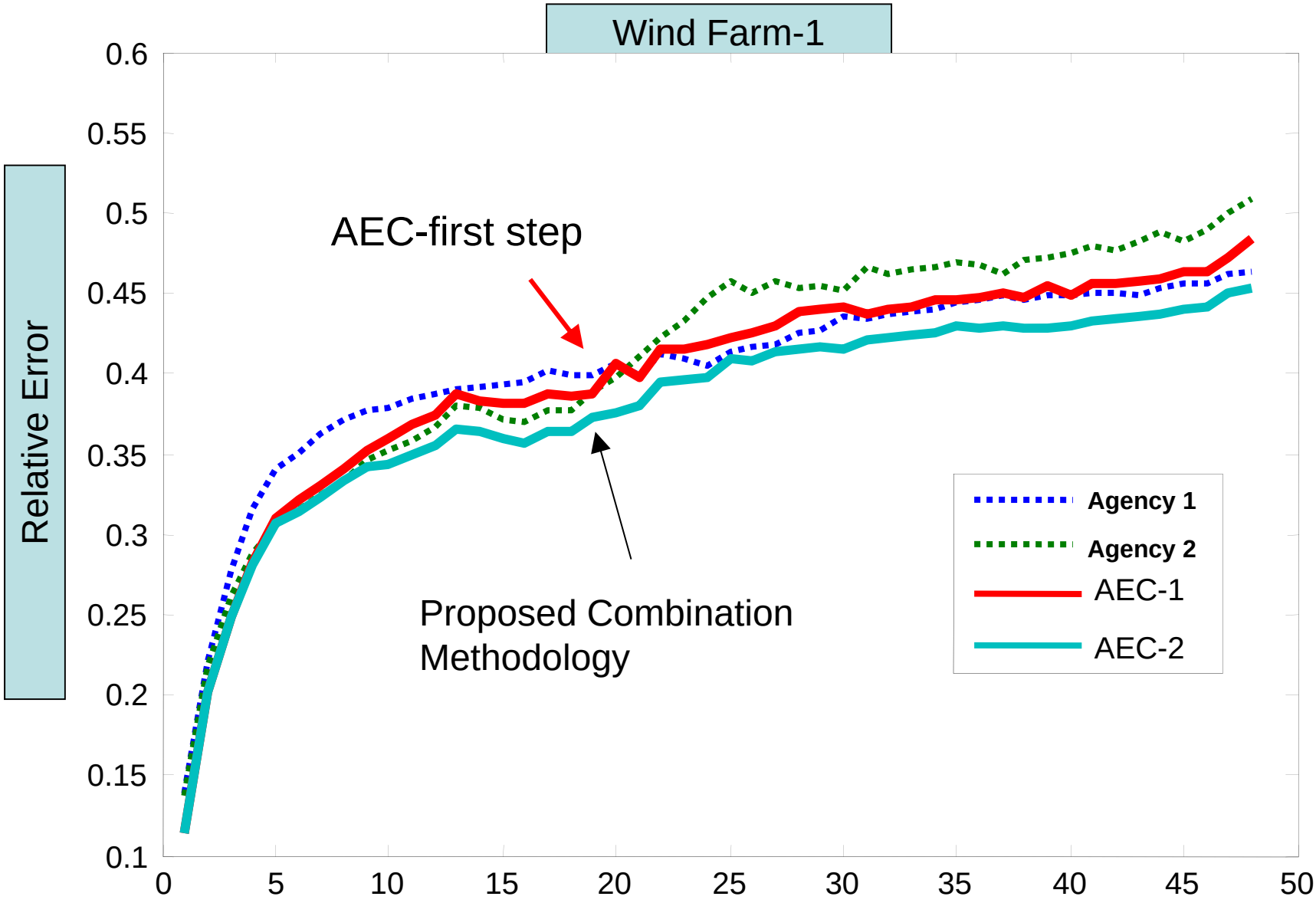
EVERYTHING!!!

- For a given wind farm and a meteorological agency: we combine the predictions of the competing models



- We repeat this with two alternative meteorological agencies, and combine the two predictions

4-FORECAST COMBINATION



5-REGIONAL FORECASTING

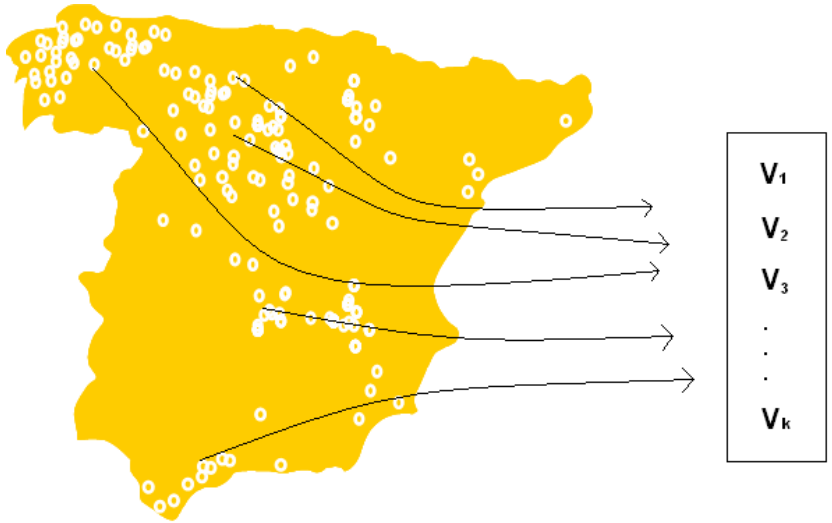
Problem: we have M wind farms and want to predict their aggregated power production

Should just sum the M individual predictions, or is there anything different that provides better performance?

Aggregated Wind Power Forecasting Methods:

- We predict the aggregate without computing the individual predictions.
- Faster than making individual forecasts for each wind farm and add them
- No need of information about the characteristics of each wind farm
- Accuracy increases with the size of the region and the number of wind farms in the region, due to spatial smoothing effects

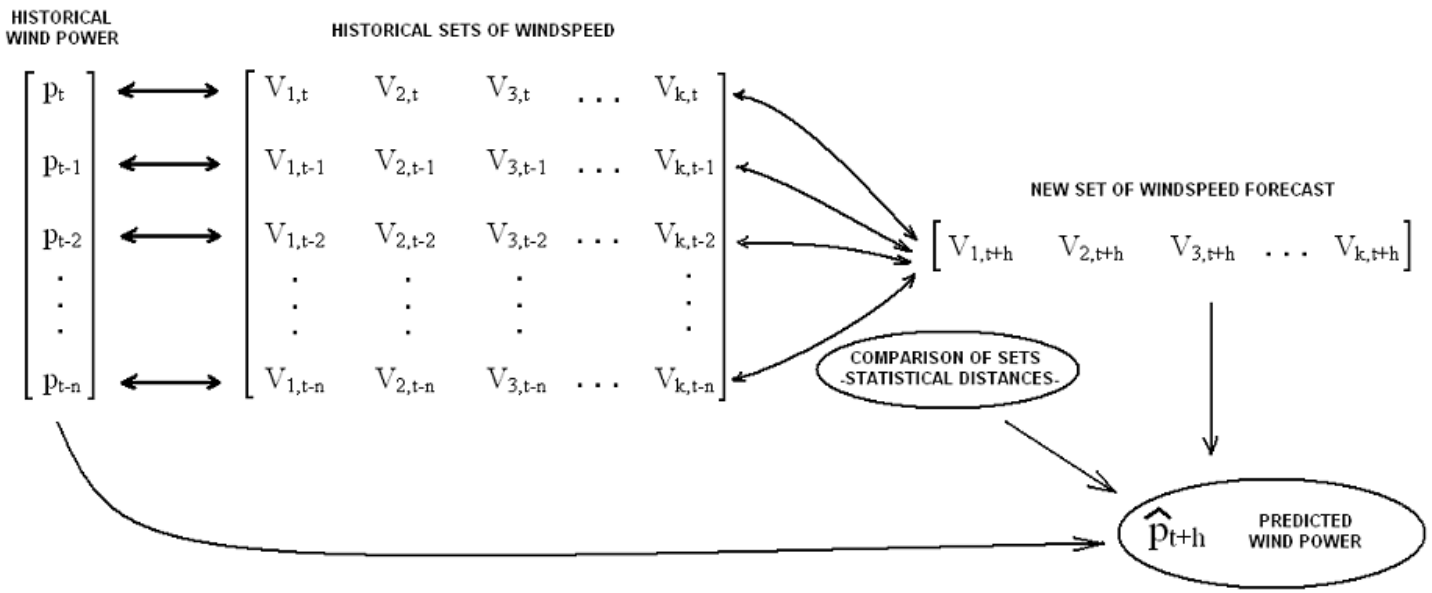
5-REGIONAL FORECASTING



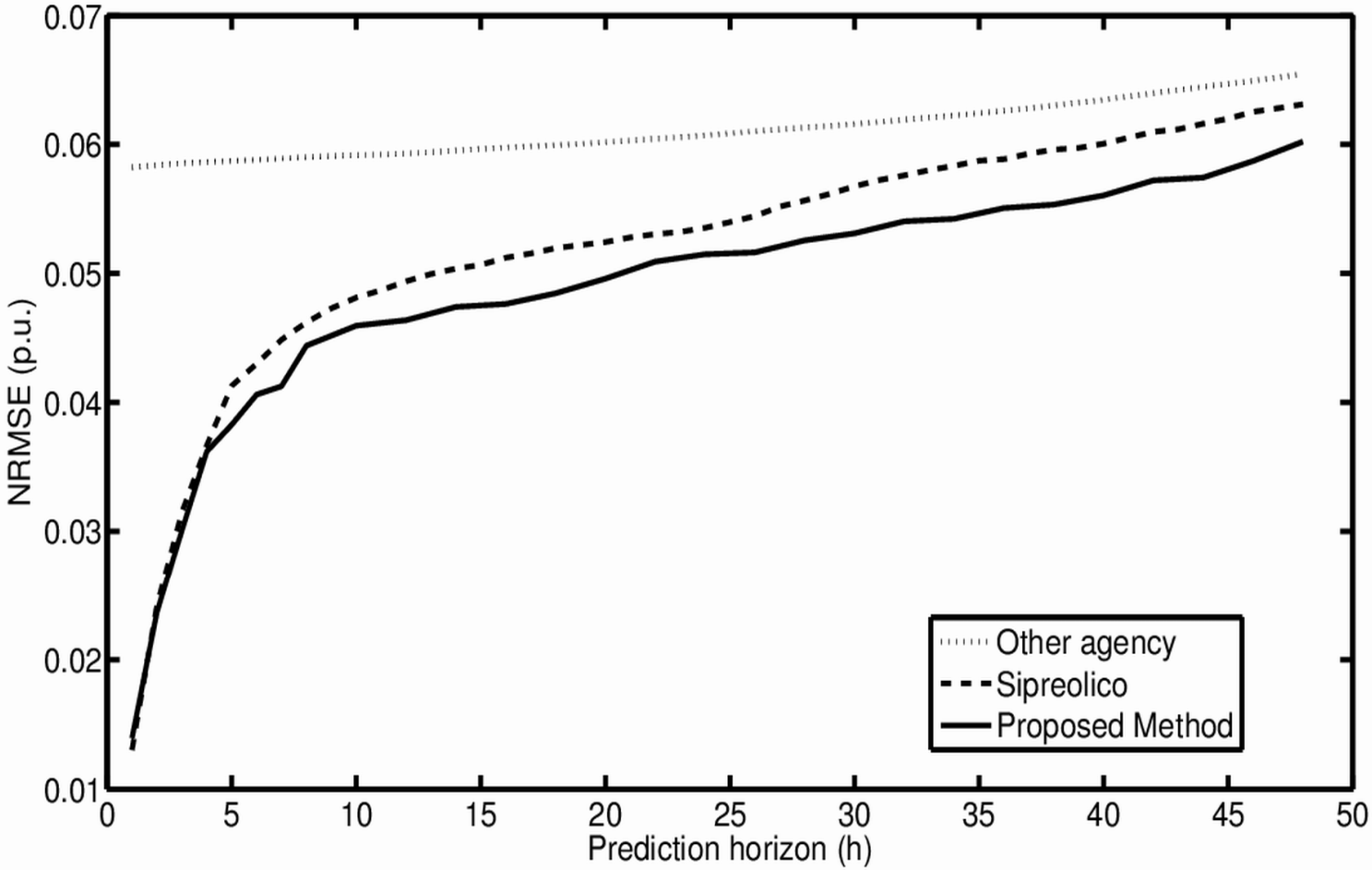
- Selection of a set of *K* coordinates in the region
- Historical data set of *N* wind speed forecasts for the K selected coordinates

  $\boldsymbol{v}_i = [v_{1,i}, \ldots, v_{K,i}], \quad i=1,\ldots,N$
- Historical data of the *N* corresponding total Wind Power measured in the region

  $P_i, \quad i=1,\ldots,N$



5-REGIONAL FORECASTING



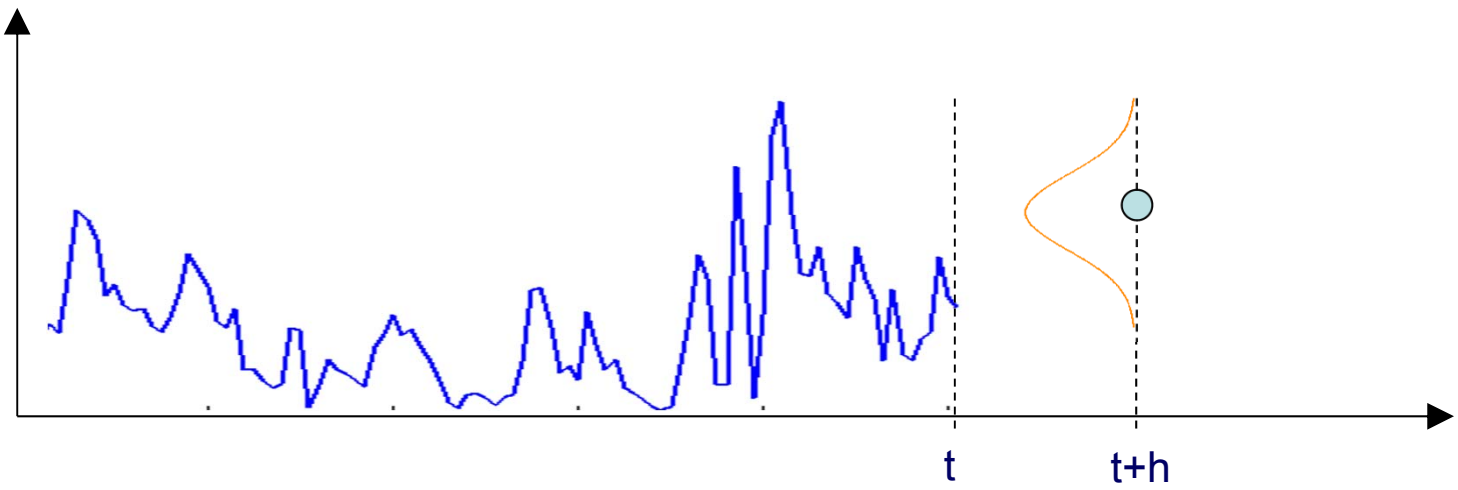
6-PROBABILISTIC FORECASTS

The goal: on-line estimation of a parametric density function of wind power predictions.

Given a historic sequence of predictions and actual values, we want to fit a conditional density function for the predictions:

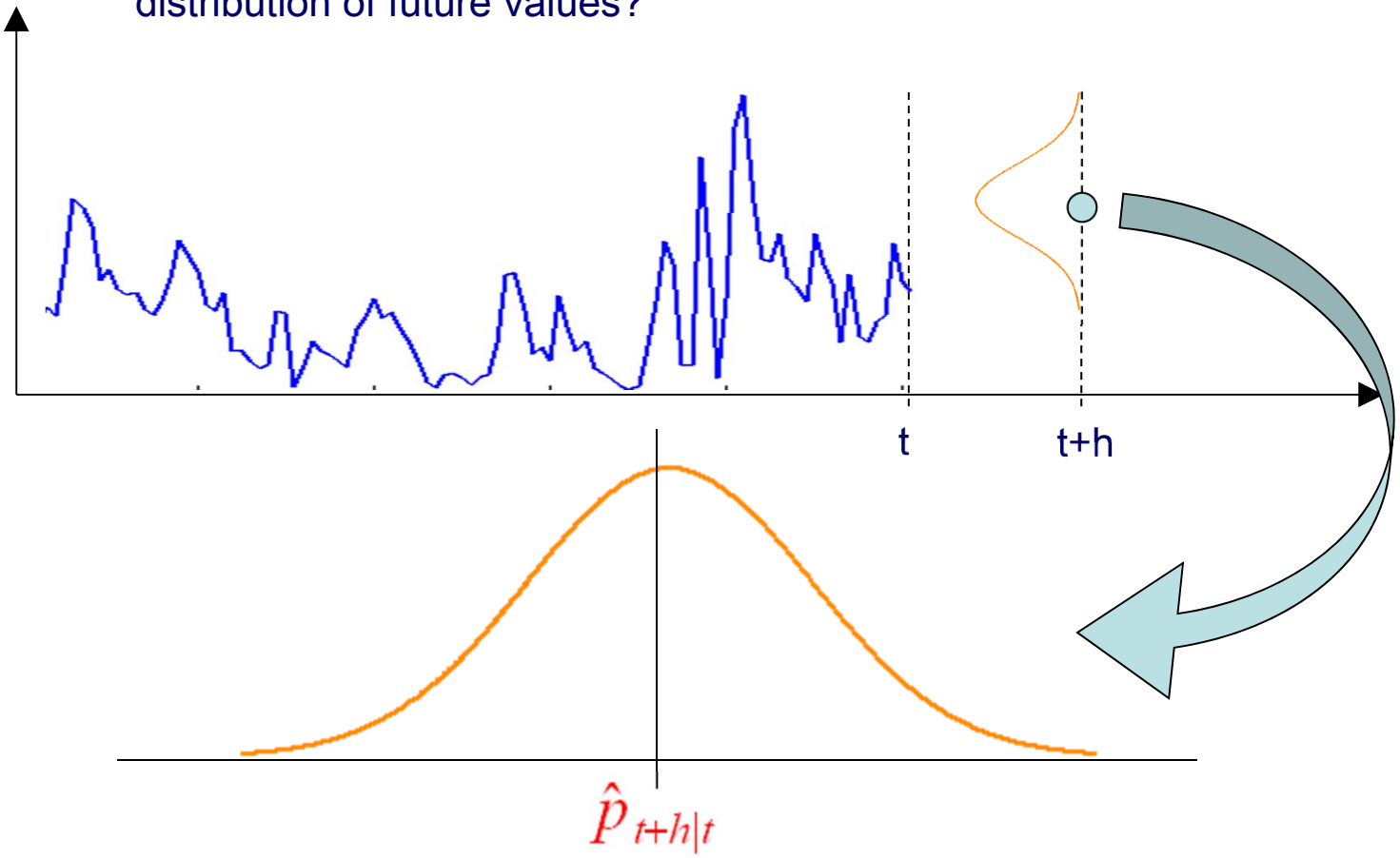
Given the prediction $\hat{p}_{t+h|t}$ of the value p_{t+h}

¿What is the distribution of future values?



Description of the problem

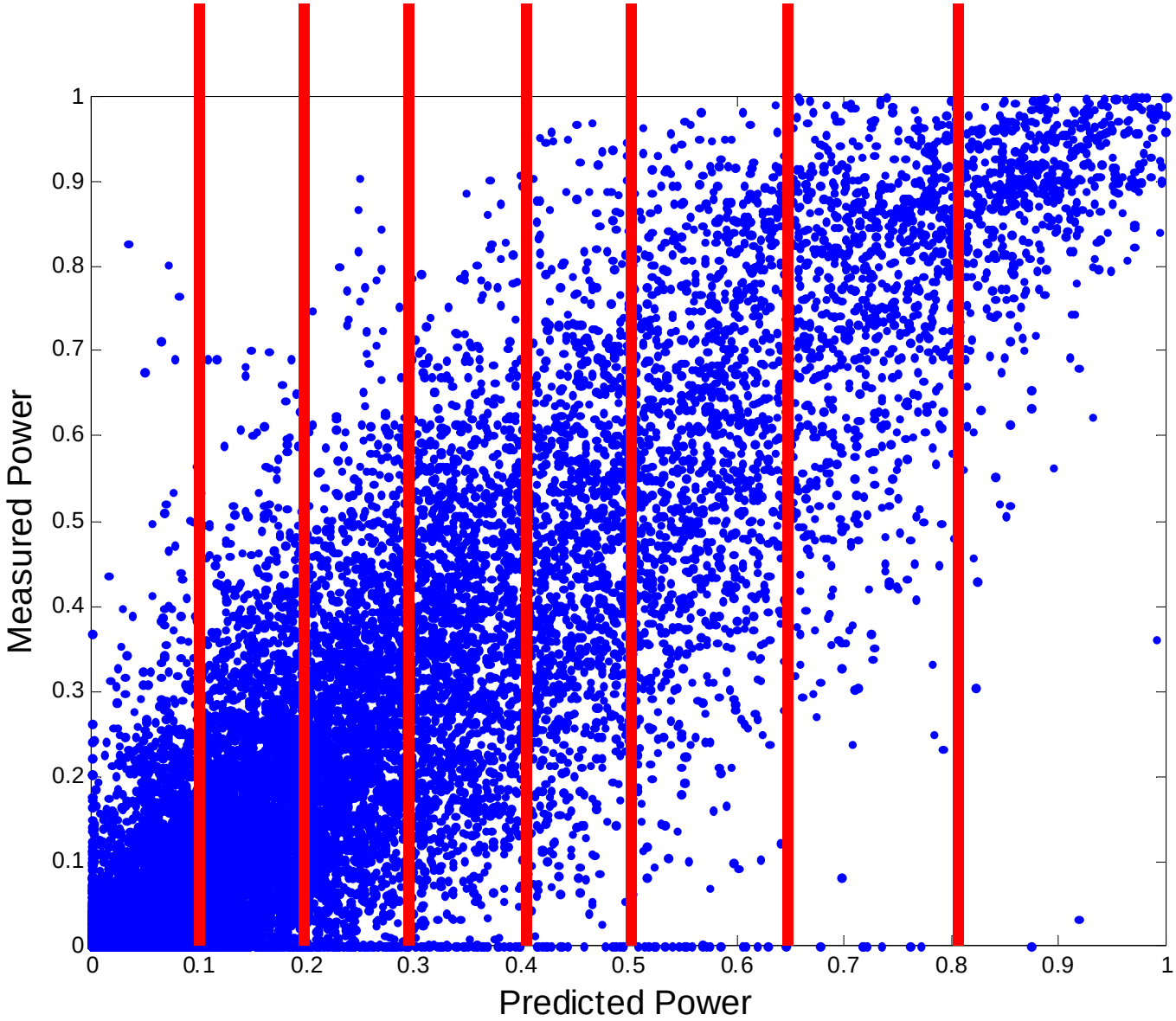
Using only the time series of observed and predicted values ¿What is the distribution of future values?



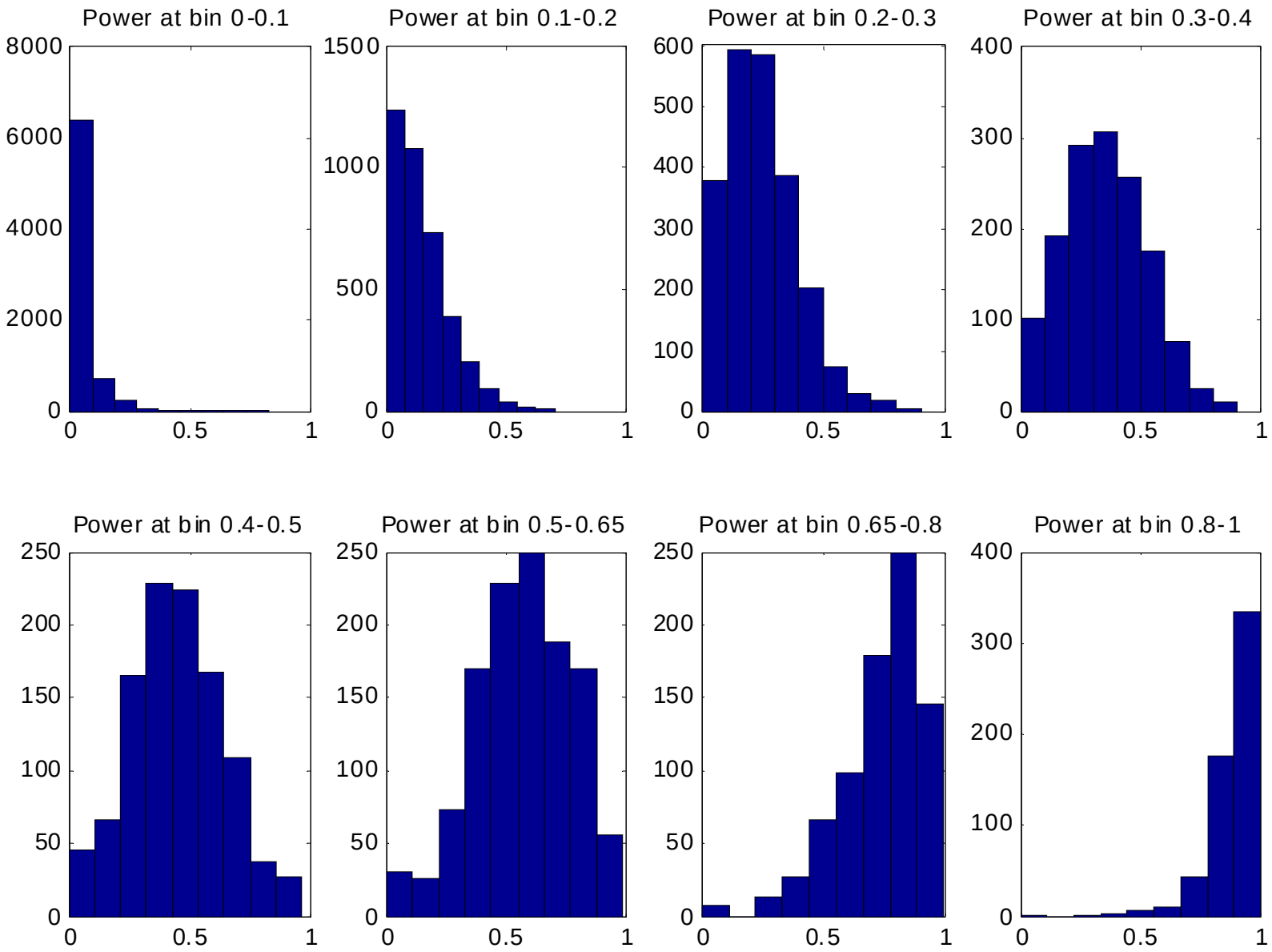
Distribution of future values at horizon H conditioned to the value of the prediction:

$$f(p_{t+h}|\hat{p}_{t+h|t})$$

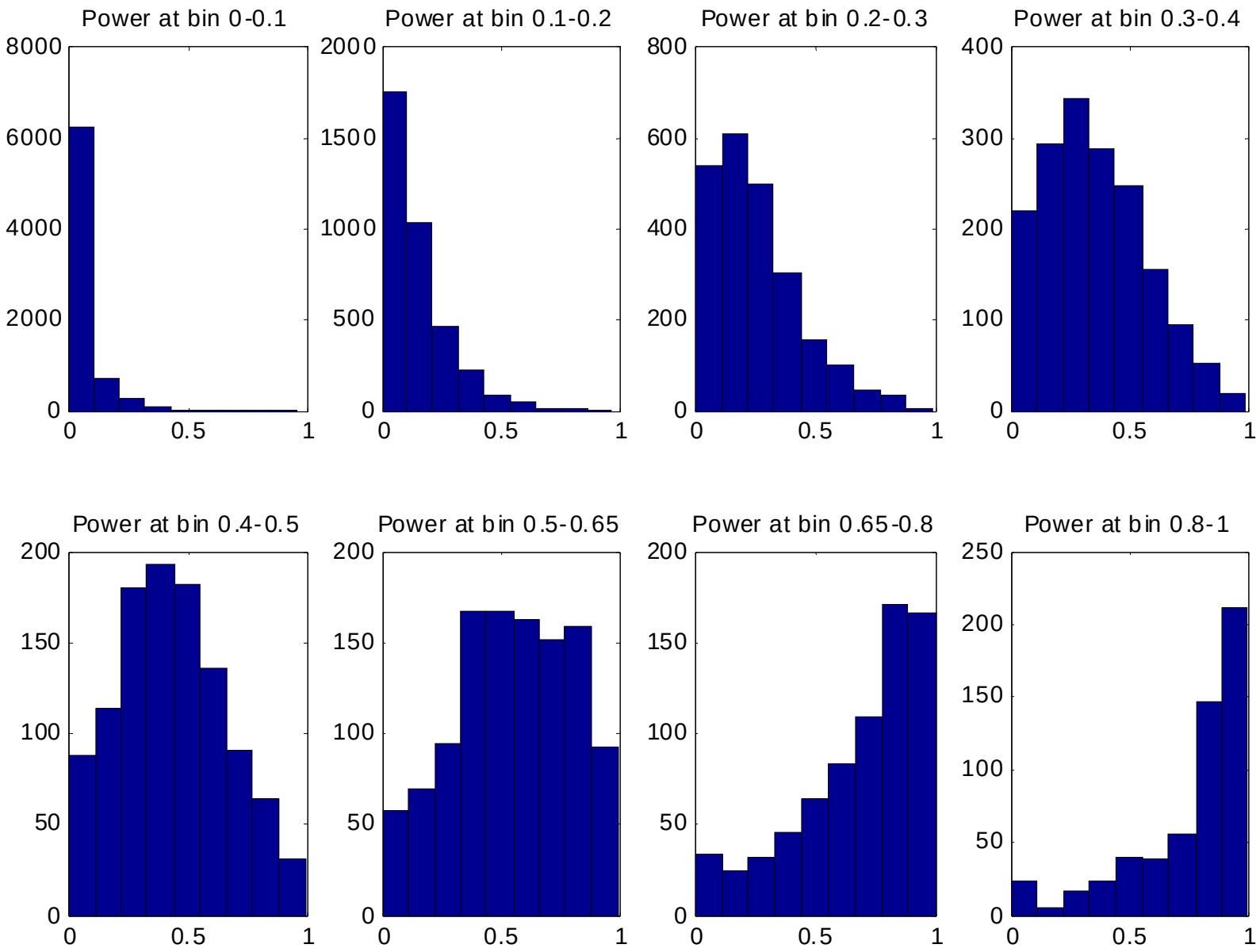
Distribution of data in bins of (normalized) predictions



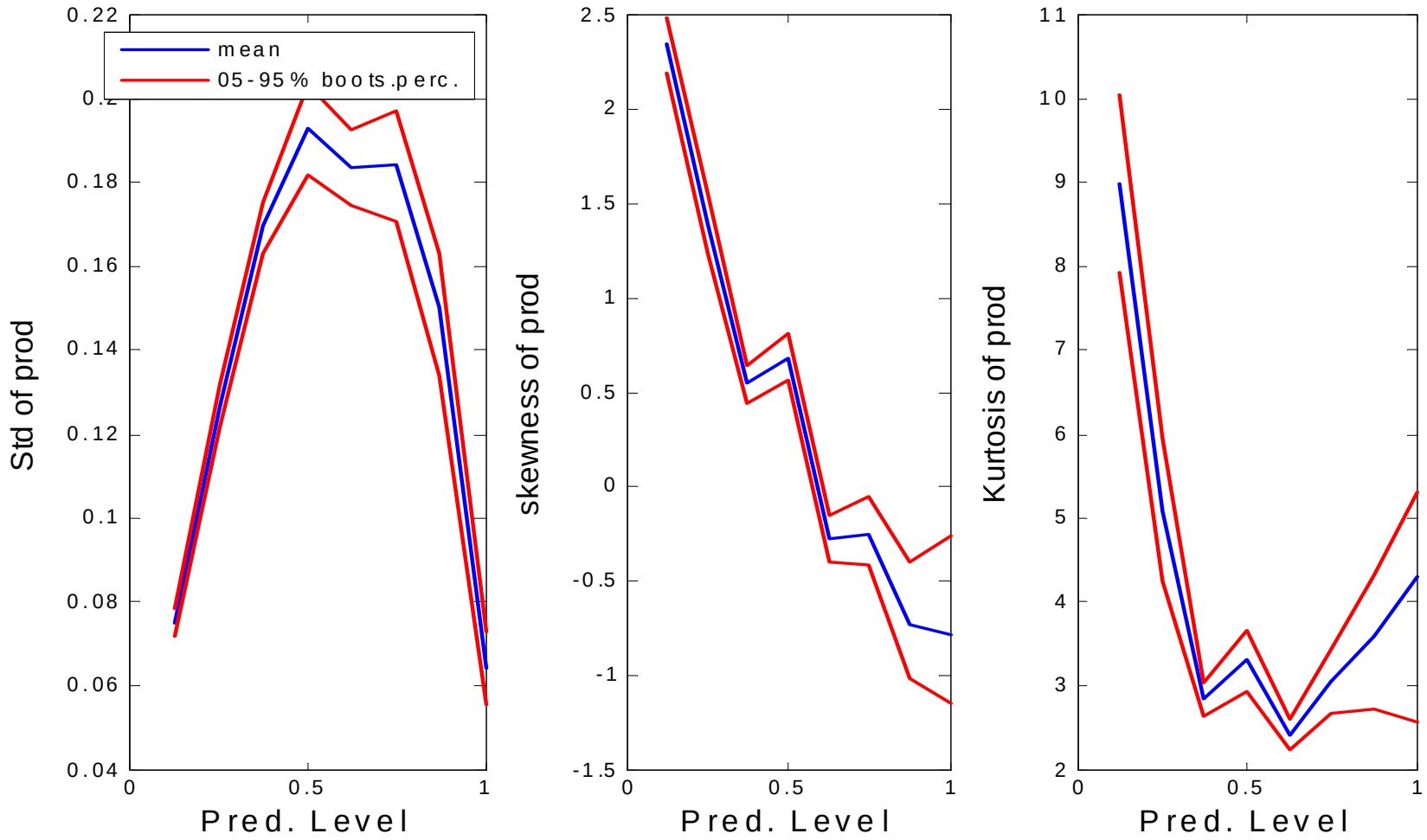
Example with a wind farm, H=1



Example with a wind farm, H=24



6-PROBABILISTIC FORECASTS



Recursive estimation of conditional moments

The moments are easily estimated conditioned to the power

→ To estimate parametric distribution by the Moments Method

Conditioned variance

$$e^2_{t+h|t} = \alpha_{0t} + \alpha_{1t}\hat{p}_{t+h|t} + \alpha_{2t}\hat{p}^2_{t+h|t} + \alpha_{3t}\hat{p}^3_{t+h|t} + w_{t+h}$$

$$\hat{\sigma}^2_{t+h|t} = \hat{\alpha}_{0t} + \hat{\alpha}_{1t}\hat{p}_{t+h|t} + \hat{\alpha}_{2t}\hat{p}^2_{t+h|t} + \hat{\alpha}_{3t}\hat{p}^3_{t+h|t}$$

6-PROBABILISTIC FORECASTS

Conditioned skewness

$$e^3_{t+h|t} = \beta_{0t} + \beta_{1t}\hat{p}_{t+h|t} + \beta_{2t}\hat{p}^2_{t+h|t} + \beta_{3t}\hat{p}^3_{t+h|t} + u_{t+h}$$

$$\hat{m}^{(3)}_{t+h|t} = \hat{\beta}_{0t} + \hat{\beta}_{1t}\hat{p}_{t+h|t} + \hat{\beta}_{2t}\hat{p}^2_{t+h|t} + \hat{\beta}_{3t}\hat{p}^3_{t+h|t}$$

$$\hat{\zeta}_{t+h|t} = \frac{\hat{m}^{(3)}_{t+h|t}}{\hat{\sigma}^3_{t+h|t}}$$

Conditioned kurtosis

$$e^4_{t+h|t} = \gamma_{0t} + \gamma_{1t}\hat{p}_{t+h|t} + \gamma_{2t}\hat{p}^2_{t+h|t} + \gamma_{3t}\hat{p}^3_{t+h|t} + v_{t+h}$$

$$\hat{m}^{(4)}_{t+h|t} = \hat{\gamma}_{0t} + \hat{\gamma}_{1t}\hat{p}_{t+h|t} + \hat{\gamma}_{2t}\hat{p}^2_{t+h|t} + \hat{\gamma}_{3t}\hat{p}^3_{t+h|t}$$

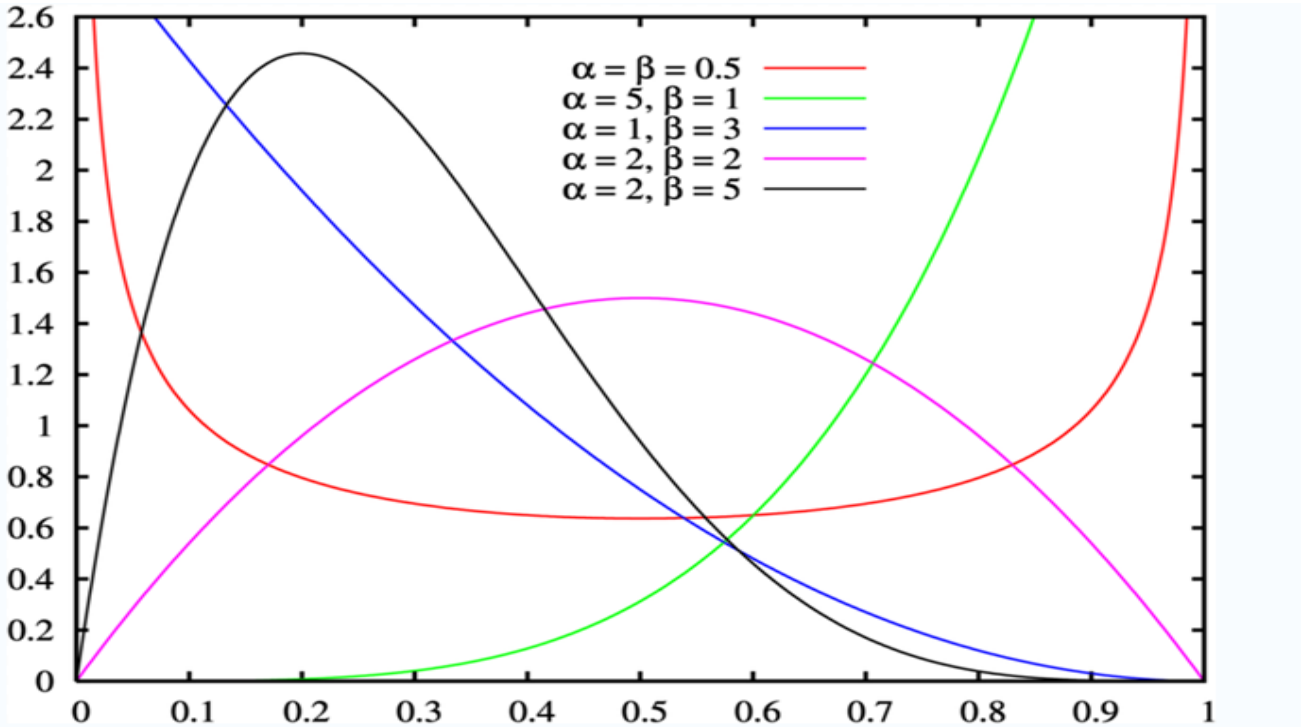
$$\hat{\kappa}_{t+h|t} = \frac{\hat{m}^{(4)}_{t+h|t}}{\hat{\sigma}^4_{t+h|t}}$$

With a set of estimated ‘moments’ we can propose a predictive distributions. For instance, using the ‘Maximum Entropy Method’

6-PROBABILISTIC FORECASTS

For aggregate data (Total Peninsular) we can use the Beta Distribution

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du}$$



Fitting a beta distribution

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du}$$

We estimate the parameters in a recursive and adaptive way

$$\alpha = \mu\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right)$$

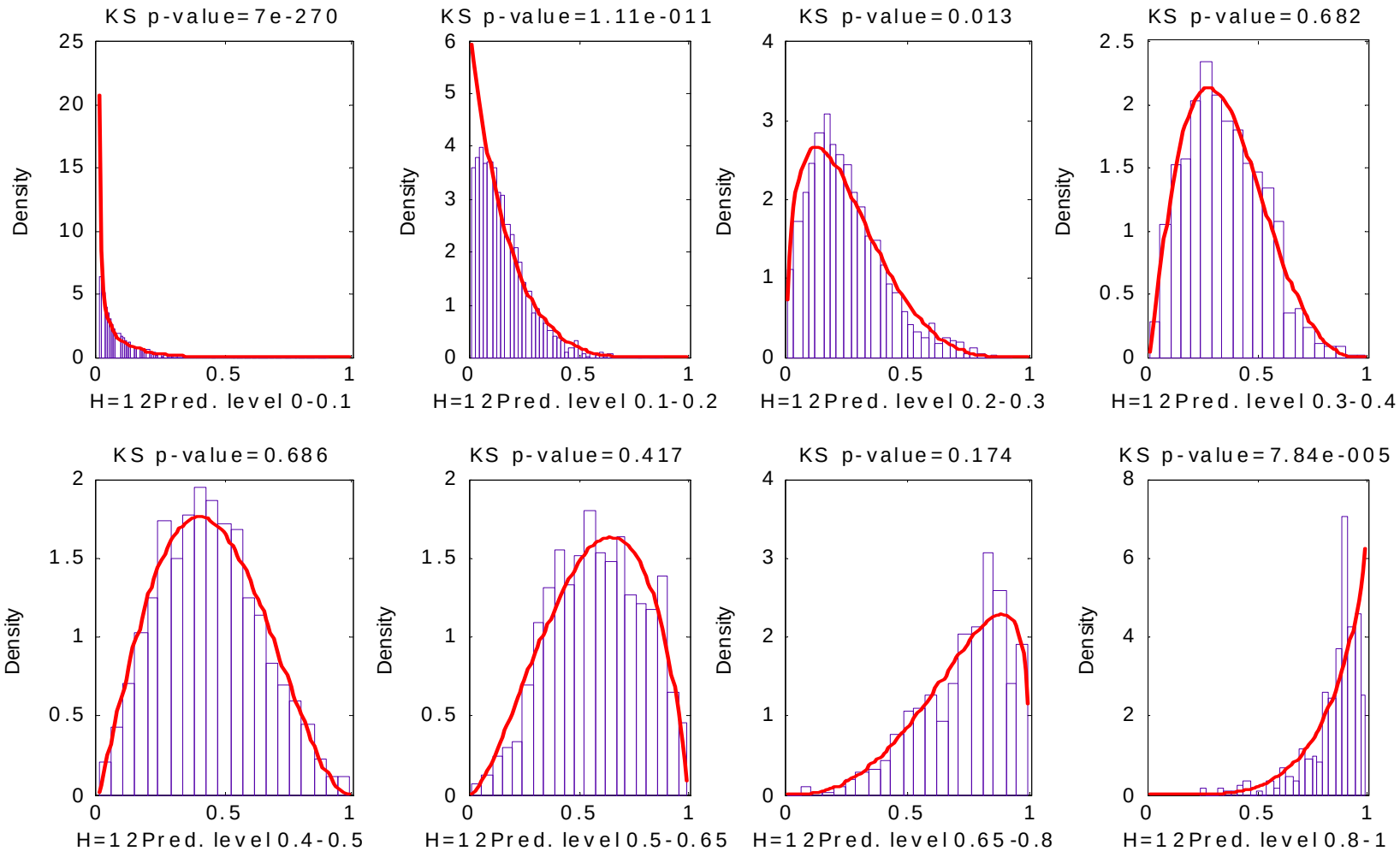
$$\beta = (1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right)$$

Estimation of conditioned mean: $\hat{p}_{t+h|t}$

Estimation of conditioned variance: estimated variance at **t** conditioned on $\hat{p}_{t+h|t}$

6-PROBABILISTIC FORECASTS

Example with a wind farm…
Good overall fitting. Lower performance at the bounds



Conclusions:

- Fitting a conditional parametric model to wind power predictions is not that easy. It should be very adaptive.
- Distributions should have a [0,1] support.
- Solution should be amenable for recursive estimation.
- Predictive distribution can be based on the recursive estimation of conditional moments