Short-term wind turbine power forecasting

Lucie Montuelle

LPMA - Université Paris Diderot

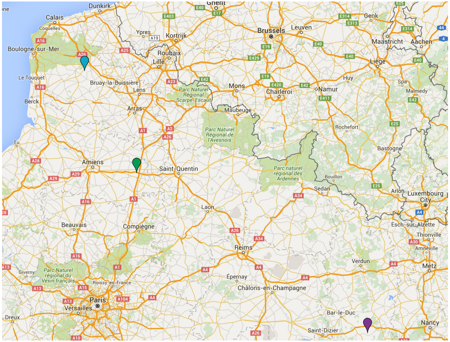
Motivation

Sharp prediction is important for :

- optimal energy selling on markets
- maintenance planning
- developing wind farms

Focus on :

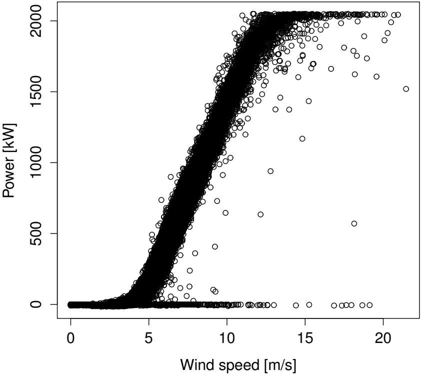
- 3 farms in the North and East of France
- 4 to 6 turbines per farm
- 8.2 to 12.3 MW per farm
- 4 years of 10 minutes data points $\approx$ 210 250 points/turbine



Available data set

Every 10 minutes

- Power
- 2 wind speed measures
  + the average
- Wind direction
- Temperature
- State of the turbine (on/off/starting)



How can this amount of data be efficiently used for production forecast?

Several procedures

- Persistence: $\hat{Y}_t = Y_{t-1}$.
- Linear regression: $\hat{Y}_t = a_0 + a_1 W_t$
- Logistic regression: $\hat{Y}_t = \frac{C}{1+\exp(a_0+a_1 W_t)}$
- Polynomial logistic regression: $\hat{Y}_t = \frac{C}{1+\exp(a_0+a_1 W_t+a_2 W_t^2+a_3 W_t^3)}$
- LASSO
- CART
- Bagging+CART
- Random Forest
- Support Vector Machine
- k-Nearest Neighbors

CART : Classification And Regression Tree

[Breiman et al., 1984]

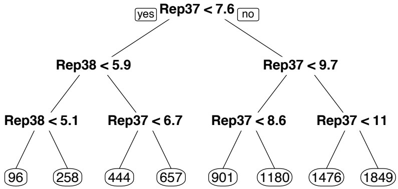
Grow a binary tree :

- choose the variable that provides the *best* cut
- criterion : intra-leaf variance

Prune the tree to avoid over-fitting :

- cross-validation
- minimize MSE

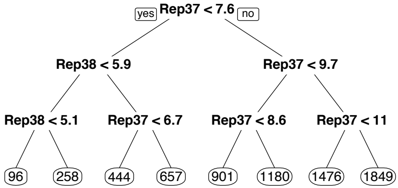
Prediction : mean of the leaf



Random Forest

[Breiman, 2001]

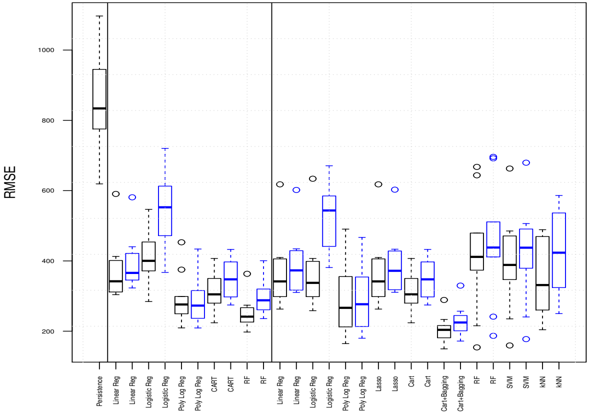
- Bootstrap one sample per tree to be grown
- Grow a binary tree :
  - choose randomly *m* variables
  - among theses variables, calculate the *best* split
  - criterion : minimize the variance of the children
- Prediction of each tree : mean of the leaf
- Average the trees



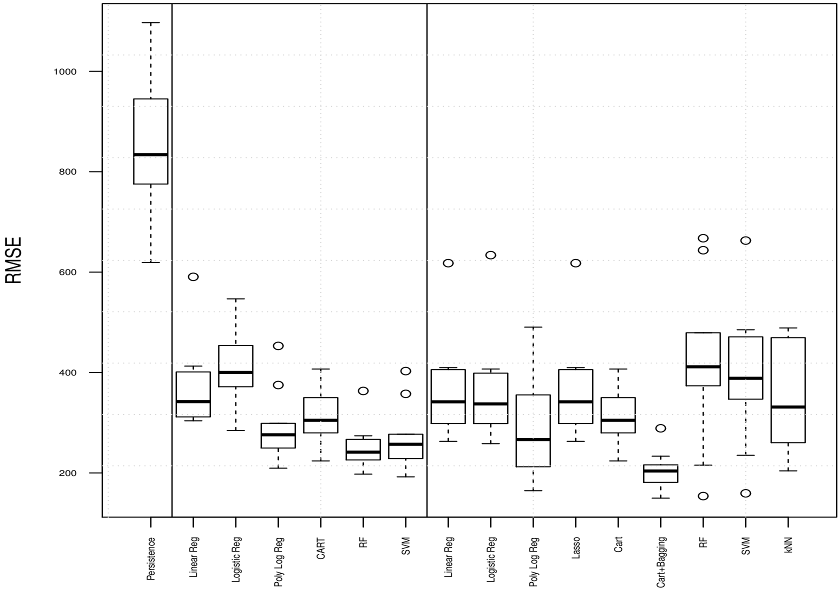
Real-time forecast

Procedures :

- Linear, logistic, polynomial logistic regression
- LASSO
- CART
- Bagging+ CART
- Random Forest
- SVM
- kNN



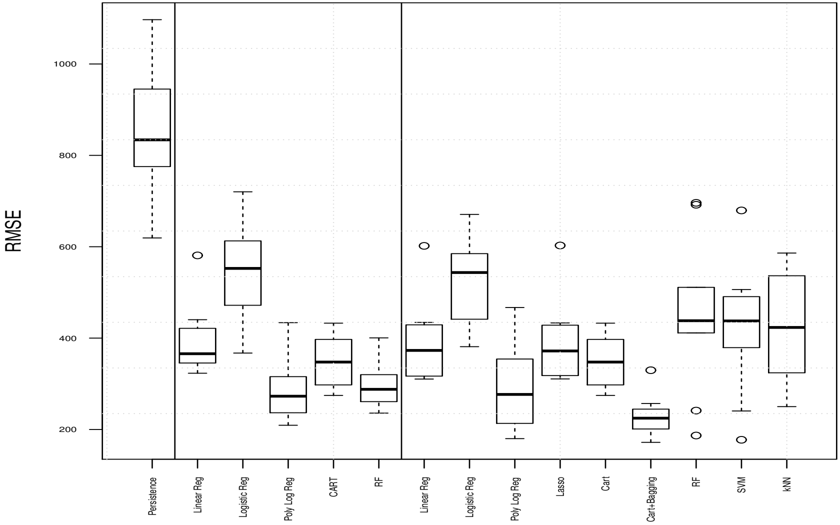
Real-time forecast with local measures



Procedure performances with local measures

	Method	Mean of RMSE	Sd of RMSE	% of IP
	Persistence	855.52	141.14	6.96
using wind speed only	Linear Regression	373.61	86.91	3.04
	Logistic Regression	404.86	76.74	3.29
	Polynomial Log. Reg.	290.36	73.87	2.36
	CART	314.46	57.74	2.56
	CART-Bagging (=RF)	**250.52**	**46.52**	**2.04**
	SVM	269.94	64.21	2.19
using all variables	Linear Regression	364.21	102.39	2.96
	Logistic Regression	362.76	107.58	2.95
	Polynomial Log. Reg.	292.57	100.53	2.38
	LASSO	364.21	102.39	2.96
	CART	314.46	57.74	2.56
	CART-Bagging	**203.50**	**39.72**	**1.65**
	RF	425.78	161.53	3.46
	SVM	382.16	134.34	3.11
	kNN (k=2)	355.29	109.96	2.89

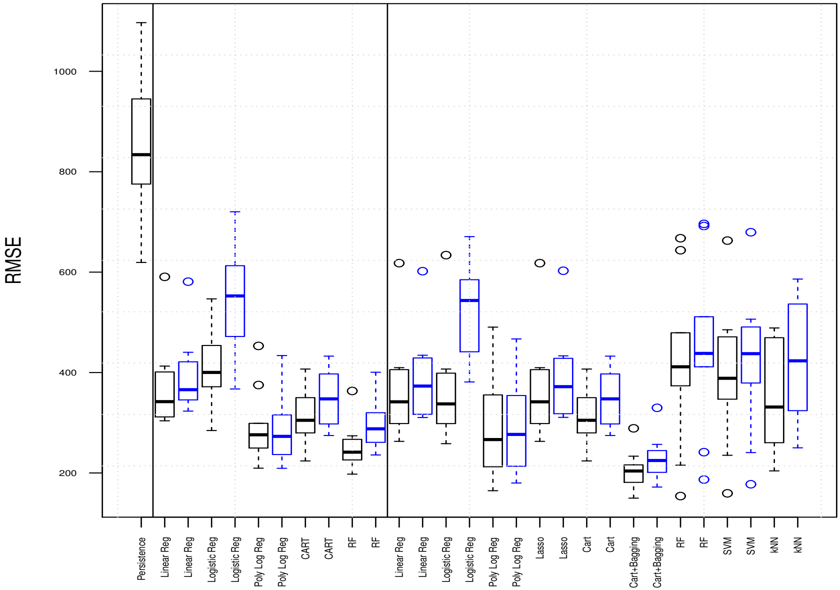
Real-time forecast with the mean of wind speed measures



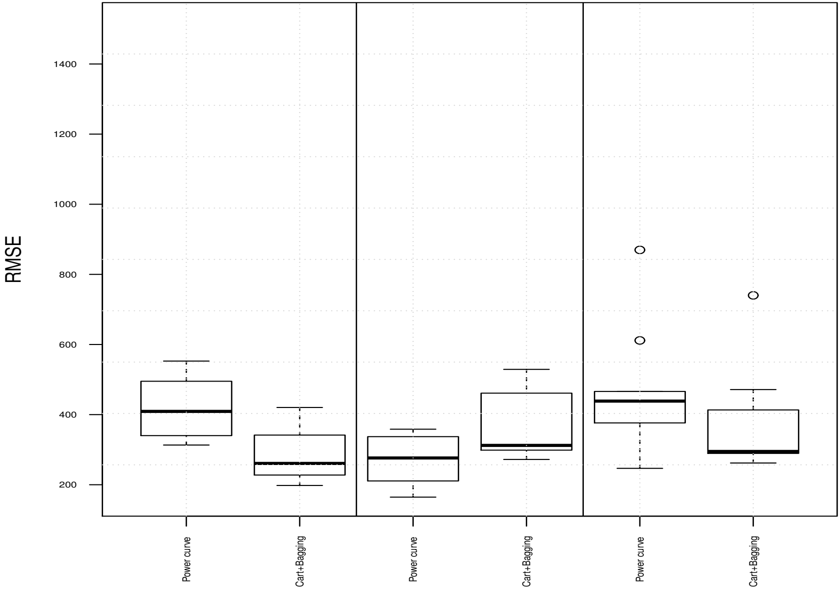
Procedure performances with the mean of wind measures

	Method	Mean of RMSE	Sd of RMSE	% of IP
	Persistence	855.52	141.14	6.96
using wind speed only	Linear Regression	393.09	77.25	3.20
	Logistic Regression	541.37	103.15	4.40
	Polynomial Log. Reg.	**288.28**	**75.23**	**2.34**
	CART	349.17	53.20	2.84
	CART+Bagging (=RF)	293.26	48.96	2.38
using all variables	Linear Regression	387.71	89.73	3.15
	Logistic Regression	524.30	92.58	4.26
	Polynomial Log. Reg.	297.16	92.79	2.42
	LASSO	387.44	89.86	3.15
	CART	349.17	53.20	2.84
	CART + Bagging	**228.75**	**43.35**	**1.86**
	RF	447.77	161.84	3.64
	SVM	424.15	143.02	3.45
	kNN	428.05	125.84	3.48

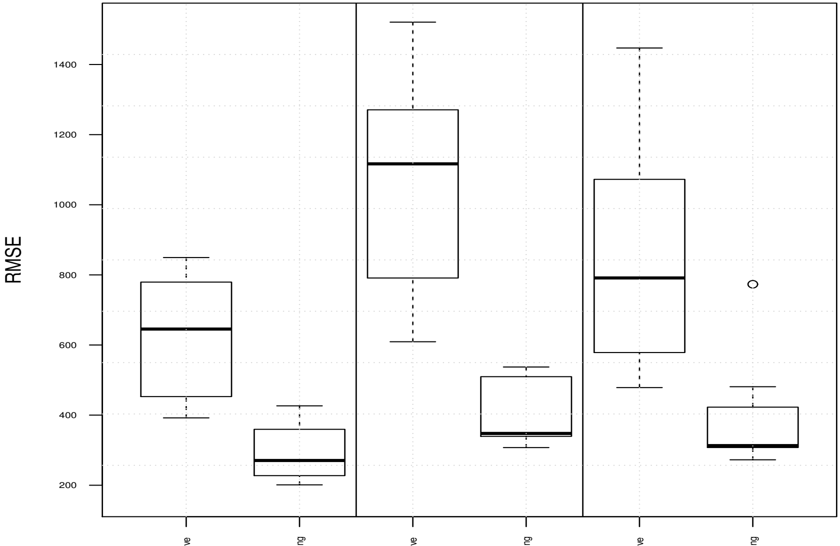
Real-time forecast



Farms comparison with local measures



Farms comparison with the mean of wind speed measures



Conclusion and perspectives

- Among parametric methods, polynomial logistic regression performs well
- Well-calibrated Random Forests outperform other procedures and seem robust

On-going work:

- use of weather forecasts
- forecast with a *blind* time

Thank you for your attention

Références I

L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 0885-6125. doi: 10.1023/A:1010933404324. URL `http://dx.doi.org/10.1023/A%3A1010933404324`.

L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.