Forecasting Solar Radiation with Online Ensemble Learning

Jean Thorey[1,2], Vivien Mallet[2],
Christophe Chaussin[3], Laurent Descamps[4], Philippe Blanc[5]

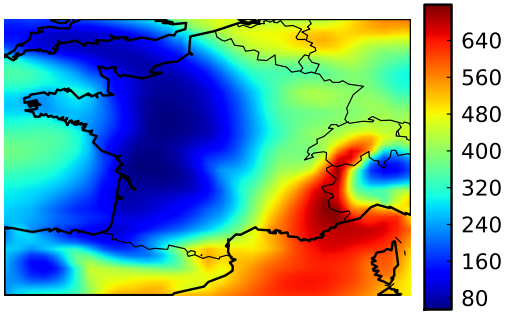
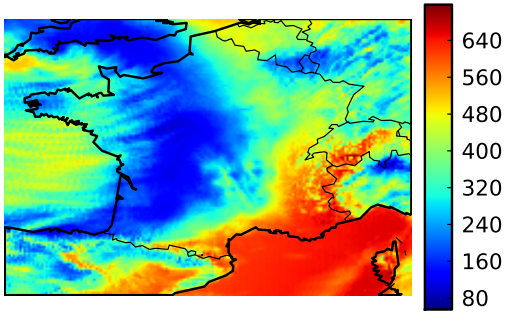
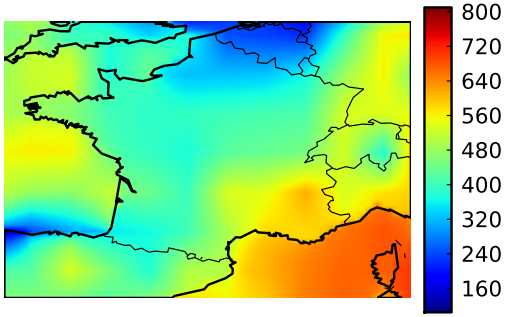
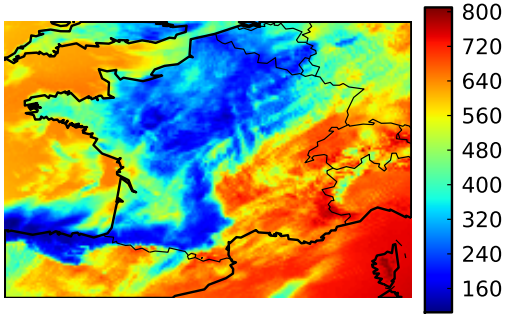
[1] CEREA, Joint Laboratory of École des Ponts ParisTech - EDF R&D
[2] INRIA [3] EDF R&D [4] Météo-France
[5] Center for Energy and Processes, MINES ParisTech

ICEM 2015

Outline

Typical days

The variable of interest is Global Horizontal Irradiance (GHI in $\mathrm{W\,m^{-2}}$).

	Forecast	Verification (satellite)
Example 1		
Example 2		

→ Need for multiple forecasts.
→ Spatial resolution.

Meteorological forecasting

Forecast horizon from several hours to a few days : we use Numerical Weather Predictions (NWP).

Many model runs form an ensemble of forecasts (different atmospheric models, parameterizations, perturbations).

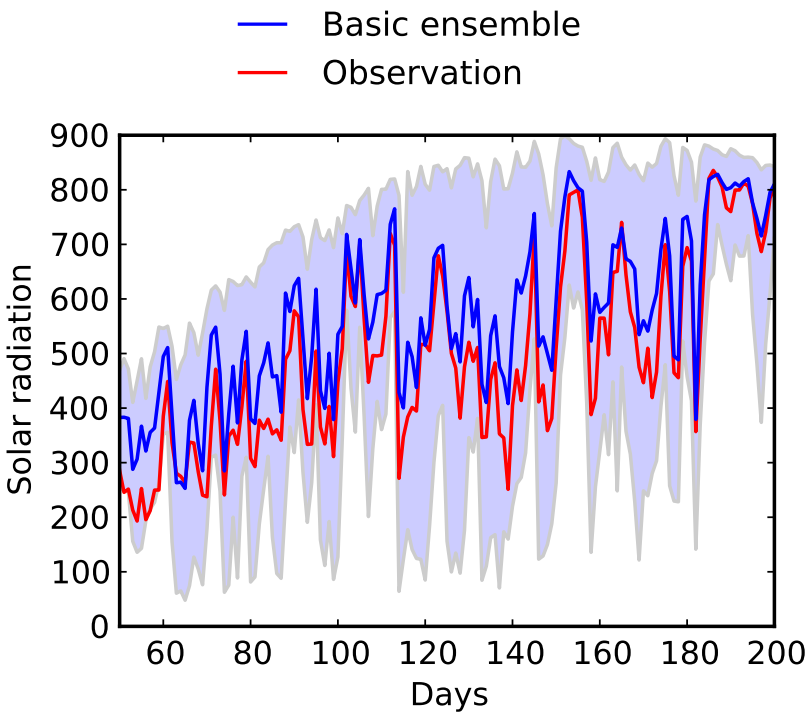
Current Research

Study ensemble forecasting with statistics and online learning tools.
Application: solar forecasting.

Ensemble of forecasts

Solar radiation data has a high variability and a low predictability.

The ensemble of forecasts is designed to catch the uncertainty of the observations.



The experiment

Goal: estimate the average GHI between 0600 and 1200 UTC, on the high resolution grid.

Horizon: $12\,\mathrm{h}$, with possible extensions.

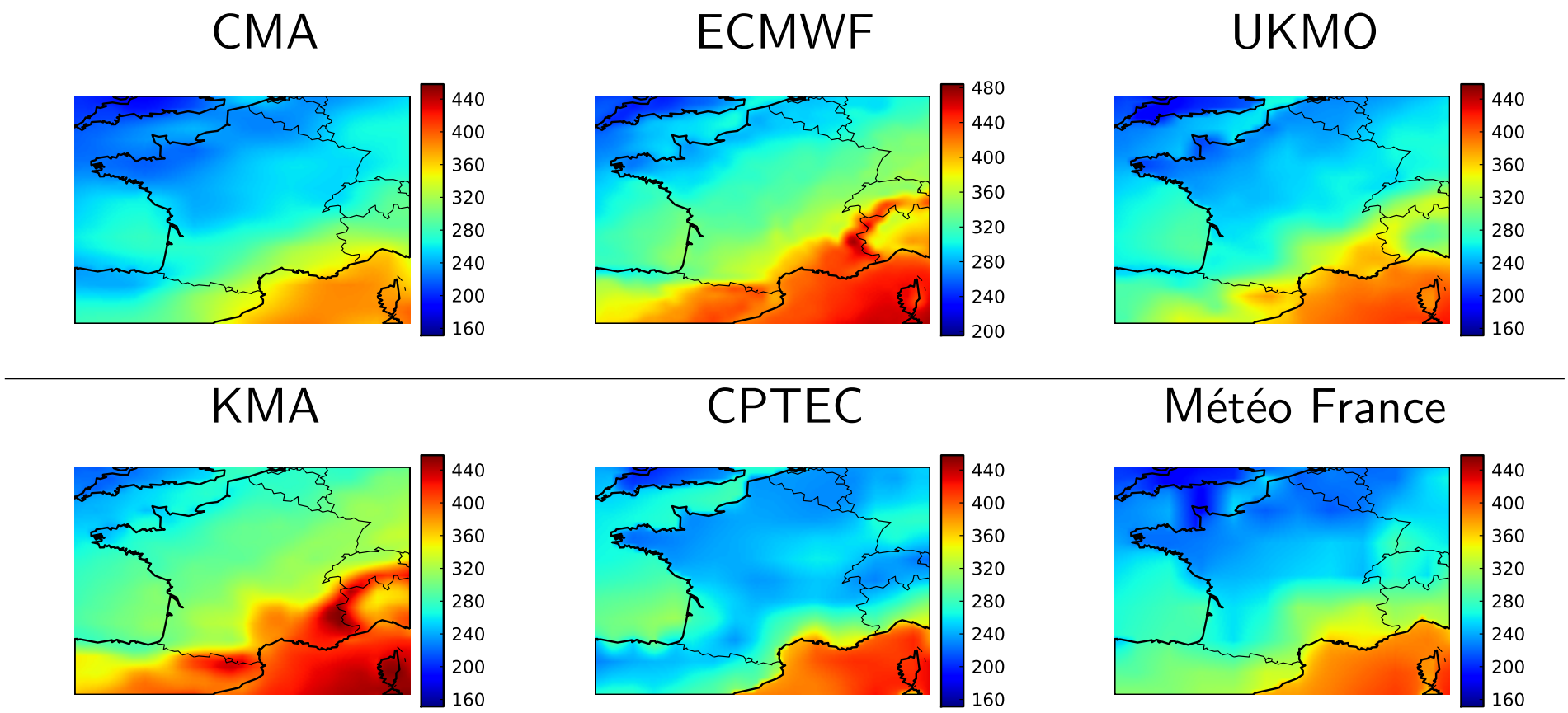
Verification data: real-time satellite-derived estimations, resolution of $\sim 10\,\mathrm{km}$, averaged from the database HelioClim 3.

Forecasts: TIGGE ensembles, resolution of $\sim 25\,\mathrm{km}$, available online.

Center	Origin	Number of members	Run
CMA	China	14	2400
ECMWF	UE	50	2400
UKMO	UK	23	2400
KMA	Korea	23	2400
CPTEC	Brazil	14	2400
Météo France	France	34	1800
Total		158	

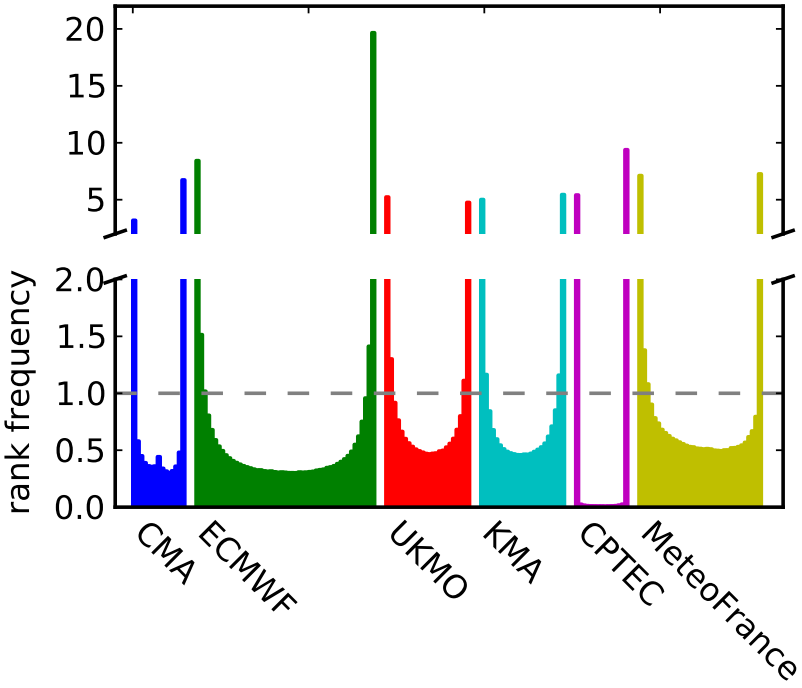
$\rightarrow$ Combine TIGGE forecasts.

Forecasts from various meteorological centers

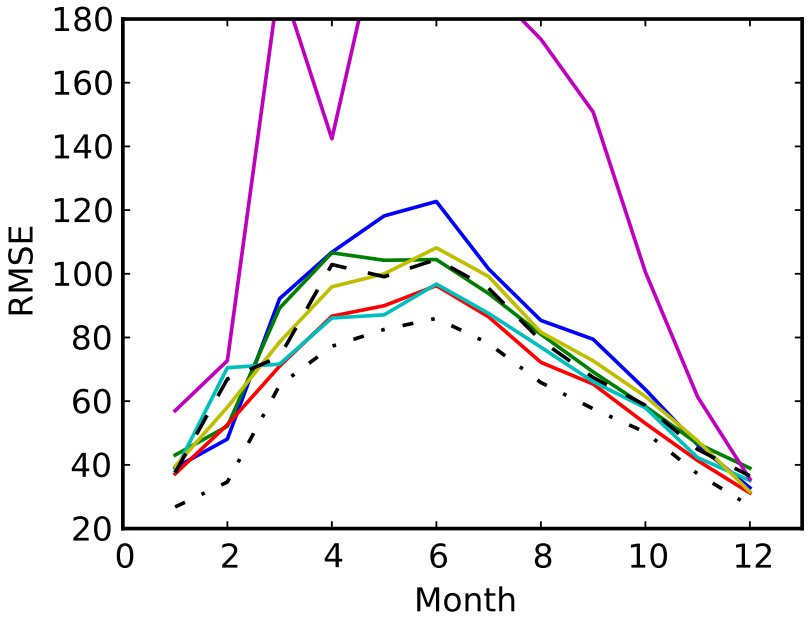


Annual average of the forecasts in $\mathrm{W\,m^{-2}}$ for year 2012.

TIGGE ensembles



Rank diagrams.
They show underdispersed ensembles.



Monthly scores.
— TIGGE ensemble means.
– – ECMWF deterministic forecast.
– · – Our forecast.

Online ensemble learning

Principle: create a linear combination x^* of the members with best performance than any individual member.

$$x^*(t) = \sum_{m=1}^{M} u_m(t) \times x_m(t)\,.$$

x_m	:	value of m-th member.
u_m	:	weight of m-th member.
y	:	observation (or verification) to be forecasted.

The members are sorted in each TIGGE ensemble.

The method is applied at each grid point on the high resolution grid.

Online ensemble learning

$$x^*(t) = \sum_{m=1}^{M} u_m(t) \times x_m(t)\,.$$

x_m : value of m-th member.
u_m : weight of m-th member.
y : observation (or verification) to be forecasted.

	$t-2$	$t-1$	t	$t+1$
	$x_m(t-2)$	$x_m(t-1)$	$x_m(t)$	$x_m(t+1)$
	$u_m(t-2)$	$u_m(t-1)$	$u_m(t)$	$u_m(t+1)$
	$x^*(t-2)$	$x^*(t-1)$	$x^*(t)$	$x^*(t+1)$
	$y(t-2)$	$y(t-1)$	$y(t)$	$y(t+1)$

Online ensemble learning

$$x^*(t) = \sum_{m=1}^{M} u_m(t) \times x_m(t)\,.$$

x_m : value of m-th member.
u_m : weight of m-th member.
y : observation (or verification) to be forecasted.

	$t-2$	$t-1$	t	$t+1$
	$x_m(t-2)$	$x_m(t-1)$	$x_m(t)$	$x_m(t+1)$
	$u_m(t-2)$	$u_m(t-1)$	$u_m(t)$	$u_m(t+1)$
	$x^*(t-2)$	$x^*(t-1)$	$x^*(t)$	$x^*(t+1)$
	$y(t-2)$	$y(t-1)$	$y(t)$	$y(t+1)$

Weights computation

Online regularized regression.

Loss: $\ell_t(\boldsymbol{u}) = (\boldsymbol{u}^\top \boldsymbol{x}_t - y_t)^2$.

Regularization term: $r(\boldsymbol{u}) = \lambda \|\boldsymbol{u}\|^2$.

The weight vector $\boldsymbol{u}_t = [u_1(t), u_2(t), ..., u_M(t)]^\top$ is chosen as:

$$\underset{\boldsymbol{w} \in \mathbb{R}^M}{\text{argmin}} \left[r(\boldsymbol{w}) + \sum_{t'=1}^{t-1} \ell_t(\boldsymbol{w}) \right] .$$

Weights computation

Online regularized regression.

Loss: $\ell_t(\boldsymbol{u}) = (\boldsymbol{u}^\top \boldsymbol{x}_t - y_t)^2$.

Regularization term: $r(\boldsymbol{u}) = \lambda \|\boldsymbol{u}\|^2$.

The weight vector $\boldsymbol{u}_t = [u_1(t), u_2(t), ..., u_M(t)]^\top$ is chosen as:

$$\underset{\boldsymbol{w} \in \mathbb{R}^M}{\operatorname{argmin}} \left[r(\boldsymbol{w}) + \sum_{t'=1}^{t-1} \beta_{t,t'} \, \ell_t(\boldsymbol{w}) \right].$$

Focus on recent data with discount factors: $\beta_{t,t'} = \dfrac{\gamma}{(t-t')^2}$

In practice: low sensitivity around optimal parameters λ, γ.

Theoretical guarantee

Case of online ridge regression (under essentially no assumptions):

- Tends to zero with increasing T.
- Cumulated loss of our forecasts.

$$\frac{1}{T}\left(\sum_{t=1}^{T}\ell_t(\boldsymbol{u}_t) - \min_{\boldsymbol{w}\in\mathbb{R}^M}\sum_{t=1}^{T}\ell_t(\boldsymbol{w})\right) \leq \mathcal{O}\left(\frac{\ln T}{T}\right)$$

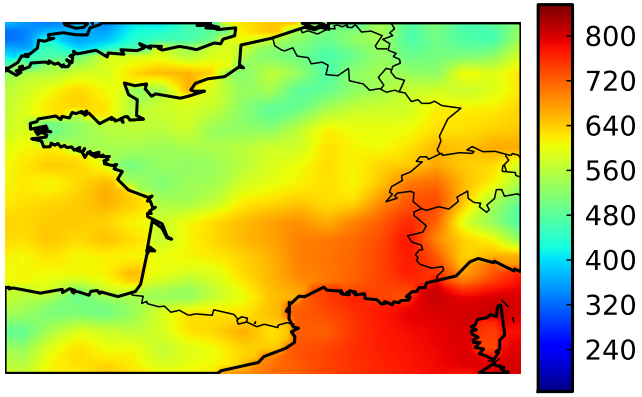
- Cumulated loss of the best combination with constant weights (more skillful than the ensemble mean and each individual member).

Robustness: the bound holds for any time series $y_t, x_{m,t}$.

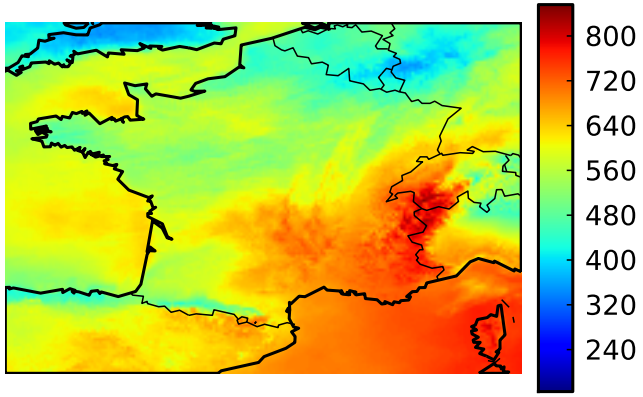
Typical forecast

Localized corrections → improved resolution.

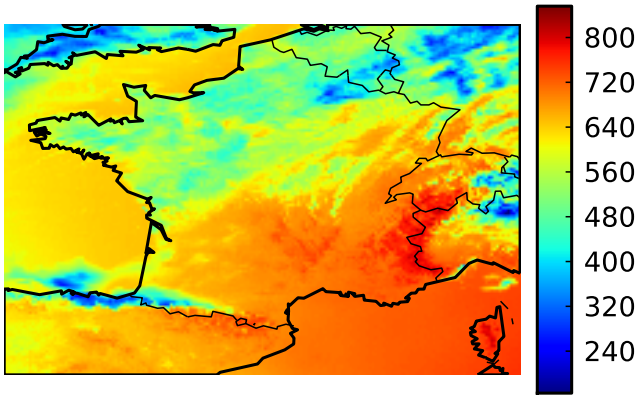
best member



aggregated forecast



verification

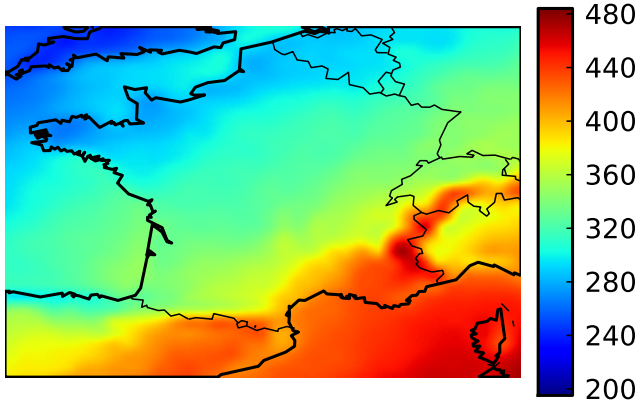


2012/05/16.

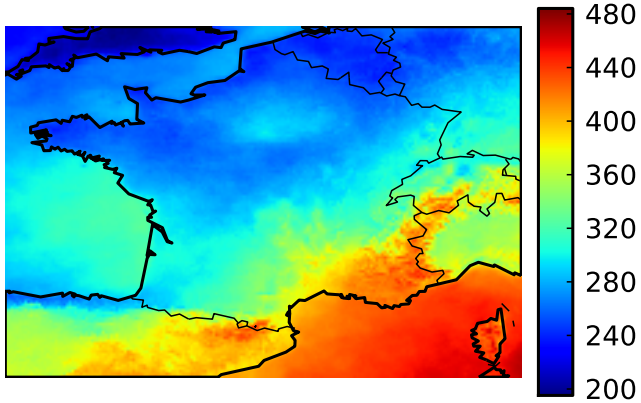
Annual average of the forecasts

Local biases are corrected.

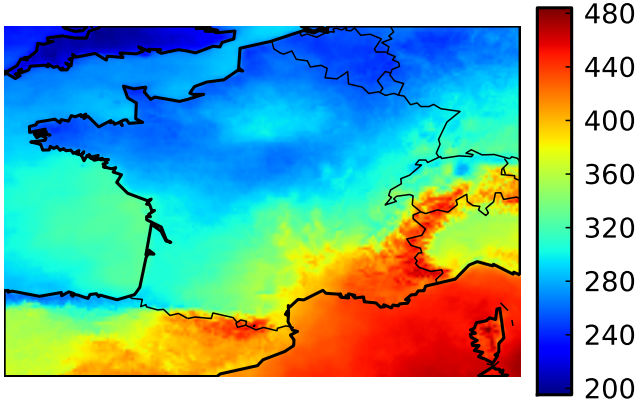
reference forecast



aggregated forecast

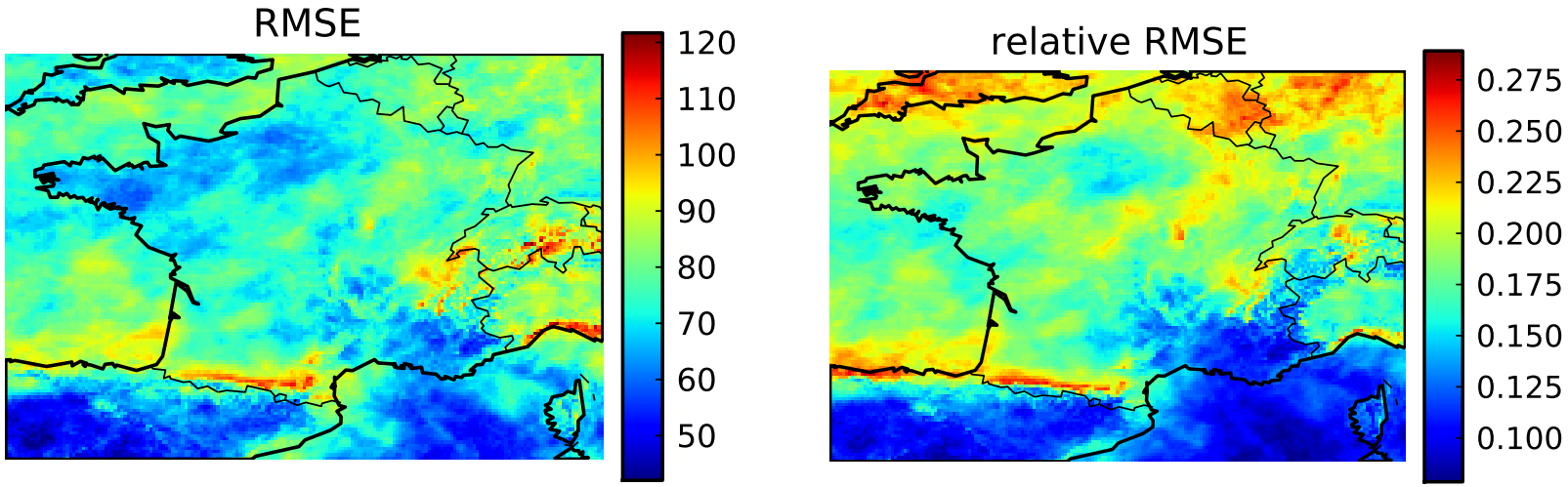


verification



Performance maps

spring summer



$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{u}_t^\top \boldsymbol{x}_t - y_t)^2}\,.$$

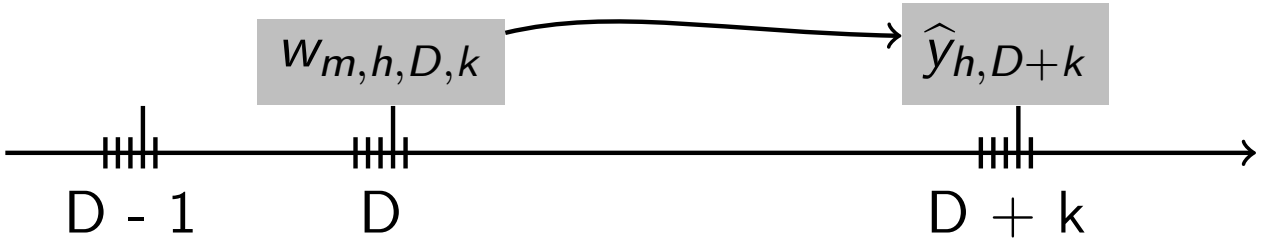
$$\text{relative RMSE} = \text{RMSE} \;/\; \left(\frac{1}{T}\sum_{t=1}^{T} y_t\right).$$

Weights and Lead Time

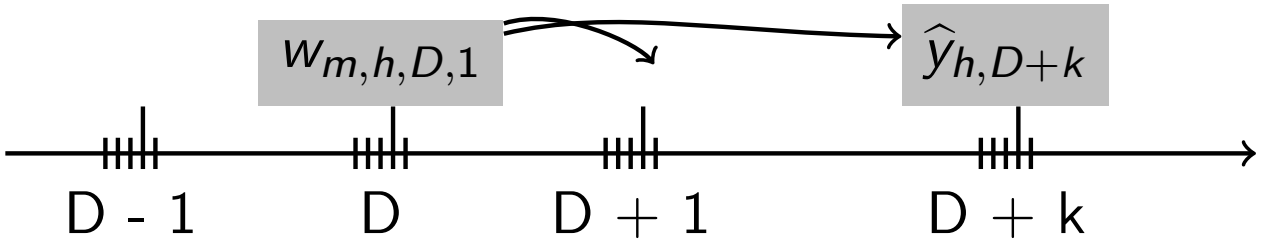
Hourly forecasts: $x_{m,h,D,k}$ (member m, hour h, day D, and lead time $k>1$).

To estimate the observation $y_{h,D+k}$.

- Weights defined per hour and lead time, based on $x_{m,h,D,k}$.

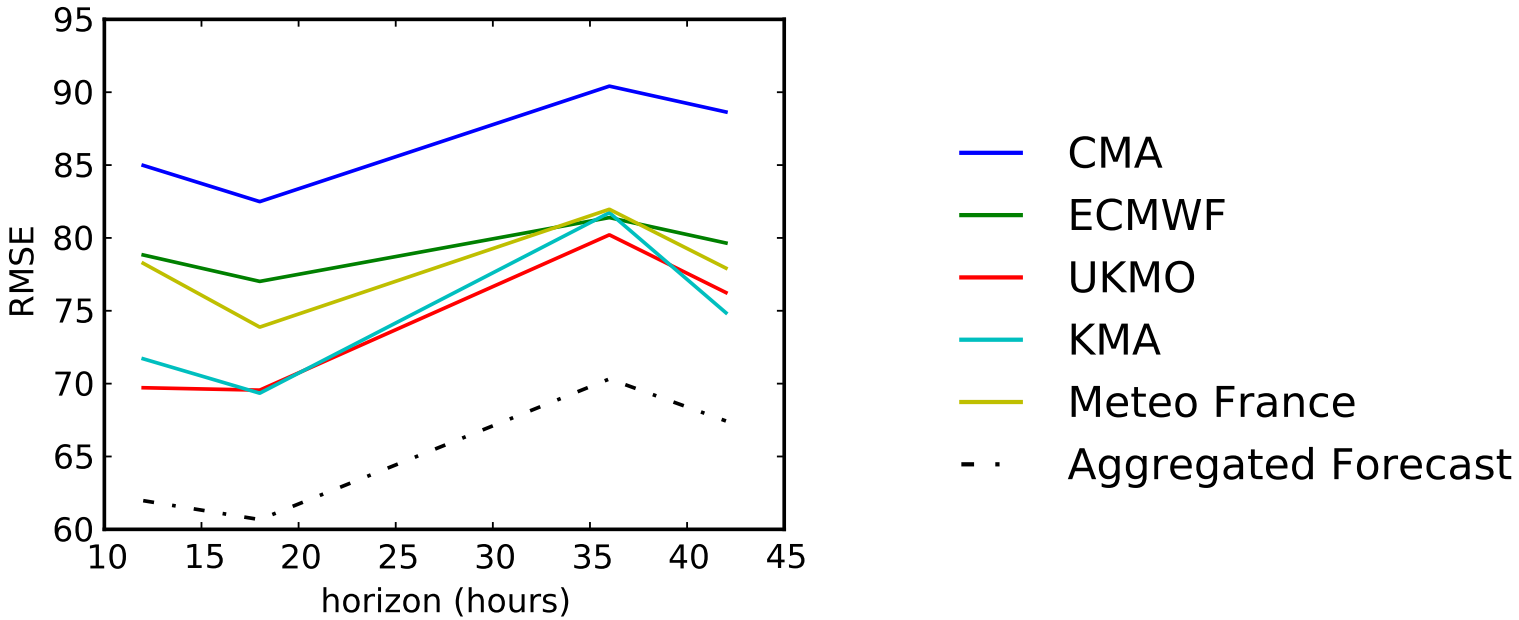


- Weights defined per hour, based on $x_{m,h,D,1}$ (for the shortest lead time), by lagging the weights: $w_{m,h,D,1} = w_{m,h,D,2} = ... = w_{m,h,D,k}$.



Weights and Lead time

With lagged weights computed for first next forecast of 1200 UTC.



Weights can be kept for longer lead times.

Conclusion

- Ensemble forecasting of solar radiation.
- From low resolution forecasts to high resolution observations.
- Online ensemble learning brings theoretical guarantee.
- Robust and simple method, compatible with operationnal forecasts.

Perspectives

- Combine forecasts with finer resolution in time and space.
- Use of spatial information, aggregation of spatial structures.

References

- Cesa-Bianchi, N., Lugosi, G., 2006. Prediction, learning, and games. Cambridge University Press.
- Mallet, V., Stoltz, G., Mauricette, B., 2009. Ozone ensemble forecast with machine learning algorithms. jgr 114 (D05307).
- Mallet, V., 2010. Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation. Journal of Geophysical Research 115 (D24303).
- Stoltz, G., 2010. Agrégation séquentielle de prédicteurs : méthodologie générale et applications à la prévision de la qualité de l'air et à celle de la consommation électrique. Journal de la Société Française de Statistique 151 (2), 66–106.
- Thorey, J., Mallet, V., Chaussin, C., Descamps, L., Blanc, P., 2015. Ensemble forecast of solar radiation using TIGGE weather forecasts and HelioClim database. (under revision).

Thank you for your attention.

Assumptions for the theoretical guarantee

$$\frac{1}{T}\left(\sum_{t=1}^{T}\ell_t(\boldsymbol{u}_t) - \min_{\boldsymbol{w}\in\mathbb{R}^M}\sum_{t=1}^{T}\ell_t(\boldsymbol{w})\right) \leq \mathcal{O}\left(\frac{\ln T}{T}\right)$$

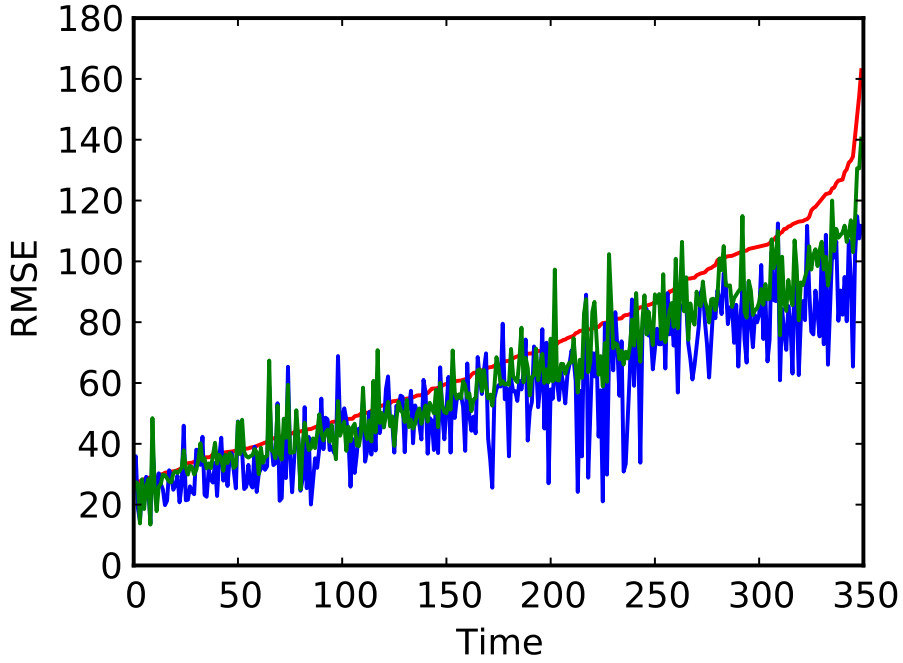
- Our losses are bounded: $\ell_t(\boldsymbol{u}_t) < C_1$.
- The oracle weights are bounded for the 2-norm: $\boldsymbol{w}^\top\boldsymbol{w} < C_2$.

The oracle can be built on selected orthogonalized members to avoid overfitting.

Alternatives

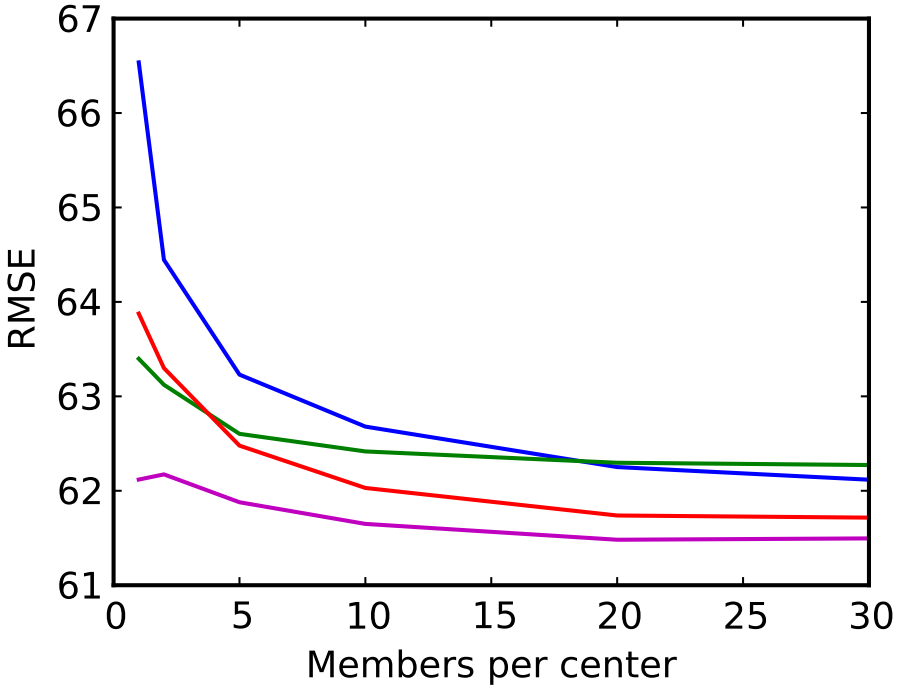
- Pre-process time series: clear sky normalization.
- Chose few members in each sub-ensemble.
- Two-step aggregation (first within each sub-ensemble).
- Identify the members per rank or ?
- Introduce new members:
  - lagged ensemble,
  - analogues,
  - near grid points forecasts.

Correction of the largest errors



- ECMWF deterministic forecast.
- ECMWF deterministic forecast with local corrections.
- Aggregated forecast.

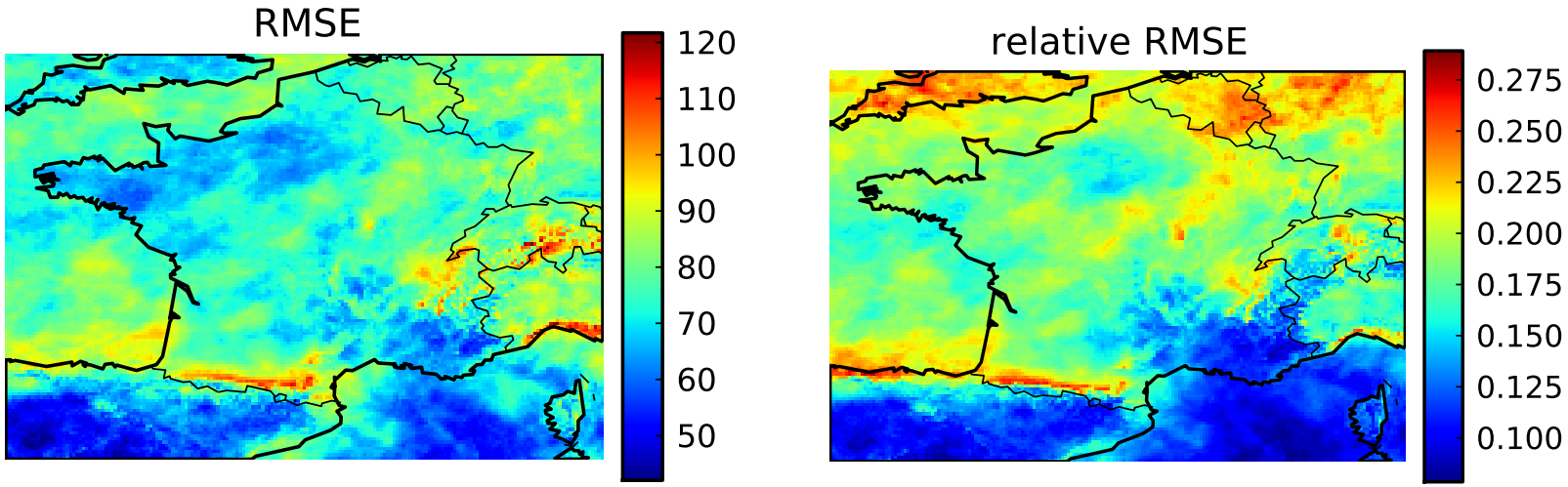
Size effect for the aggregation



- No sorting.
- Sorting full ensemble.
- Sorting for each TIGGE center.
- Sorting for each TIGGE center + ECMWF deterministic forecast.

Performance maps

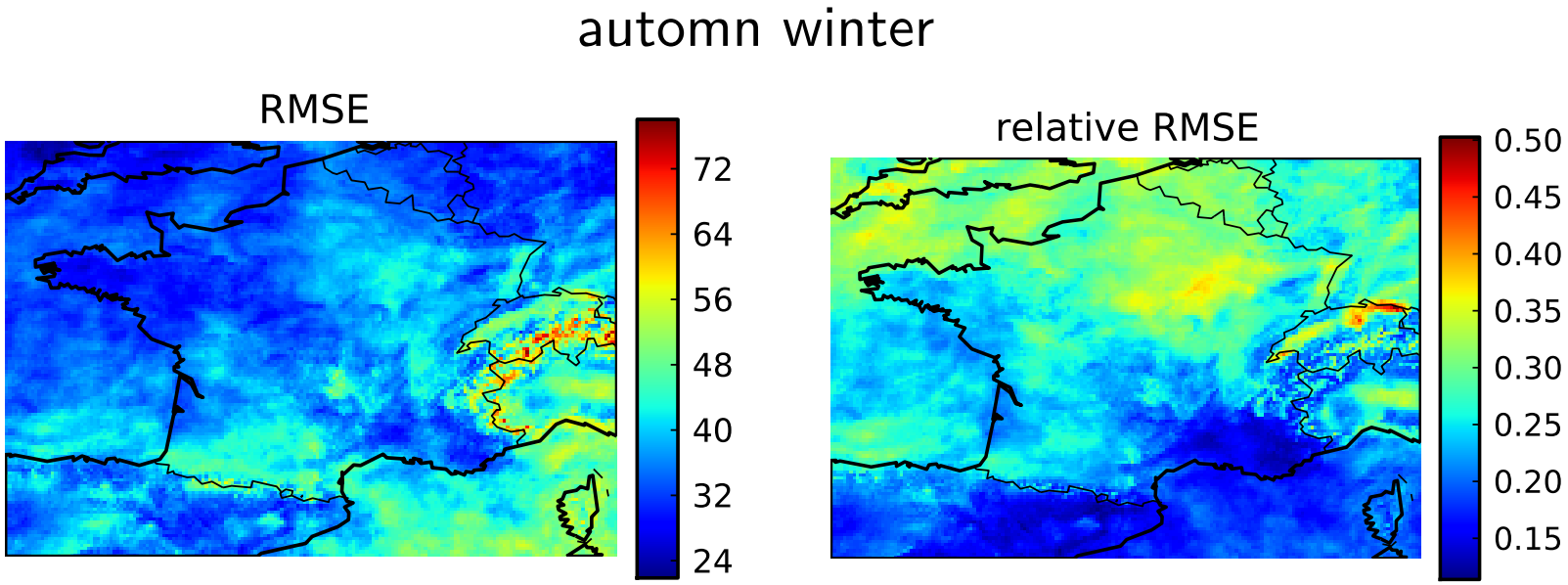
spring summer



$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{u}_t^\top \boldsymbol{x}_t - y_t)^2}\,.$$

$$\text{relative RMSE} = \text{RMSE} \;/\; \left(\frac{1}{T}\sum_{t=1}^{T} y_t\right).$$

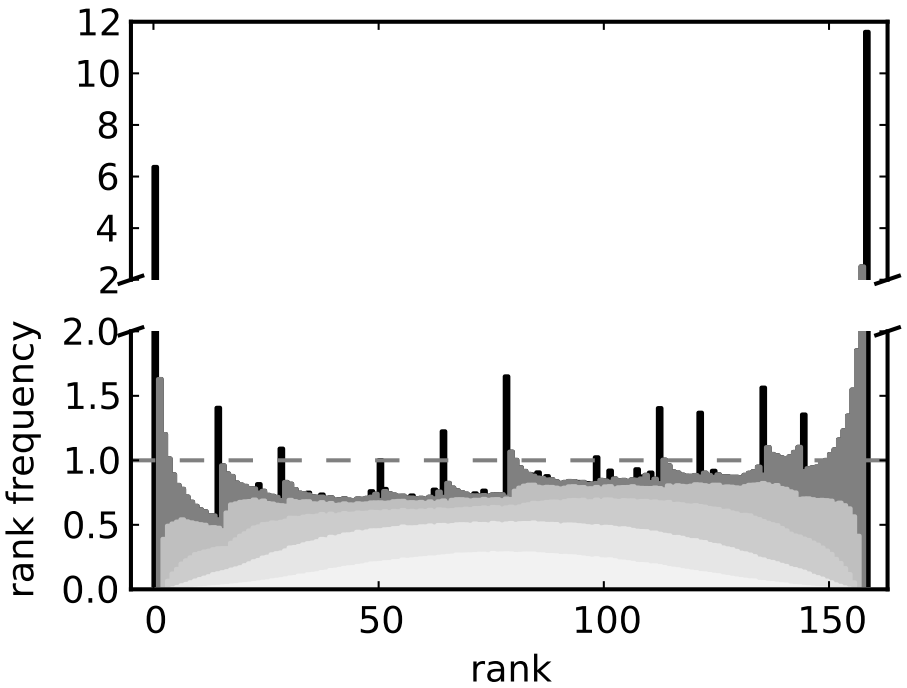
Performance maps



$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{u}_t^\top \boldsymbol{x}_t - y_t)^2}\,.$$

$$\text{relative RMSE} = \text{RMSE} \;/\; \left(\frac{1}{T}\sum_{t=1}^{T} y_t\right).$$

Ensemble from TIGGE



Rank diagram of the whole ensemble (158 members).

The gray scale indexes the number of ensembles whose spread contains the observation.

Conversion methods

label	conversion formula
glob	$x_{tigge} \times 1.18$
mult	$x_{tigge} \times \frac{x^{ssrd}_{ecmwf}}{x^{ssr}_{ecmwf}}$
add	$x_{tigge} - \left(x^{ssr}_{ecmwf} - x^{ssrd}_{ecmwf}\right)$
lin	$x_{tigge} \times \mathrm{a}_{\mathrm{center}} + \mathrm{b}_{\mathrm{center}}$

no conversion