

## Ensemble forecast of solar radiation using TIGGE weather forecasts and Helioclim database

Jean Thorey, Vivien Mallet, Christophe Chaussin, Laurent Descamps, Philippe Blanc

► **To cite this version:**

Jean Thorey, Vivien Mallet, Christophe Chaussin, Laurent Descamps, Philippe Blanc. Ensemble forecast of solar radiation using TIGGE weather forecasts and Helioclim database. *Solar Energy*, Elsevier, 2015, 120, pp.232-243. 10.1016/j.solener.2015.06.049 . hal-01184650

**HAL Id: hal-01184650**

**<https://hal-mines-paristech.archives-ouvertes.fr/hal-01184650>**

Submitted on 17 Aug 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Ensemble forecast of solar radiation using TIGGE weather forecasts and HelioClim database

J. Thorey<sup>a,b</sup>, V. Mallet<sup>b</sup>, C. Chaussin<sup>c</sup>, L. Descamps<sup>d</sup>, P. Blanc<sup>e</sup>

<sup>a</sup>CEREA, joint research laboratory ENPC ParisTech – EDF R&D, Université Paris-Est, Marne-La-Vallée, France

<sup>b</sup>INRIA, Rocquencourt, France

<sup>c</sup>EDF R&D, Chatou, France

<sup>d</sup>Météo-France, DIRIC, Saint-Mandé, France

<sup>e</sup>Center for Energy and Processes, MINES ParisTech, Sophia Antipolis, France

---

## Abstract

Medium-range forecasts (one day to two weeks) of solar radiation are commonly assessed with a single forecast at a given location. In this paper, we forecast maps of surface solar irradiance, using ensembles of forecasts from the THORPEX Interactive Grand Global Ensemble (TIGGE) with a 6-h timestep. We compare our forecasts with observations derived from MeteoSat Second Generation (MSG) and provided by the HelioClim-3 database as gridded observations over metropolitan France. First, we study the ensembles from six meteorological centers. Second, we use sequential aggregation to linearly combine all the forecasts with weights that vary in space and time. Sequential aggregation updates the weights before any forecast, using available observations. We use the global numerical weather prediction from the European Center for Medium-range Weather Forecasts (ECMWF) as a reference forecast. The issue of spatial resolution is discussed because the low resolution forecasts from TIGGE are compared to high resolution irradiance estimated from MSG data. We found that the TIGGE ensembles are under-dispersed but rather different from one to another. Aggregation decreases the forecast error by 20%, and produces a more realistic spatial pattern of predicted irradiance.

*Keywords:* solar radiation; ensemble forecast; satellite estimations; TIGGE; HelioClim; sequential aggregation

---

## 1. Introduction

Solar radiation forecasts and especially global horizontal irradiance (GHI) forecasts are needed for the integration of photovoltaic power (PV). The increasing installed capacity of photovoltaic power requires that solar radiation forecasts be always more accurate in terms of spatial and temporal resolutions.

Many meteorological centers provide solar radiation forecasts with two strategies: either a single deterministic forecast, or ensemble forecasts generally with coarser resolution. Deterministic forecasts have been extensively studied for solar and

photovoltaic forecasts. Inman et al. (2013) and Espinar et al. (2010) review numerous modeling techniques to generate solar and PV forecasts from meteorological variables. Deterministic predictions from various meteorological centers are compared by Lorenz et al. (2009) and Perez et al. (2013) in a broad range of sites. In the previously cited papers, some forecasting techniques resort to combinations of forecasts. These combinations are derived from a regression over a moving time-window and are applied to a few members.

Vernay et al. (2013) listed several available maps of solar radiation deriving from cloud products of satellite observations. Even though solar radiation forecasts cover large areas, they

---

\*Corresponding author

Email address: jean.thorey@inria.fr (J. Thorey)

are usually compared to ground measurements (PV or solar) or to satellite observations only at measurement sites (e.g., Gautier et al., 1980, among the firsts). Indeed maps of satellite observations are not broadly used to assess the accuracy of solar radiation forecasts. Morcrette (1991) used satellite observations as reference to assess the performance of numerical weather predictions (NWP), but not predictions of solar irradiance (e.g., short-wave radiation). Perez et al. (1997) studied the interactions between satellite observations and measurement sites with respect to the distances between sites. Due to the variety of error causes, Thelen and Edwards (2013) restricted the comparison between NWP and satellite observations to reflectance for short-wave radiation. Dehghan et al. (2014) give emphasis on the spatial resolution of both NWP and satellite data at ground measurement sites.

Ensemble forecasts are classical in meteorology for any field with large uncertainty and for uncertainty quantification. However, no article using ensemble forecast for solar radiation was found in the literature, despite several conference presentations. Still, Yokohata et al. (2012) studied climatological ensembles of top atmosphere radiation and radiation in cloud-free conditions.

In the framework of sequential aggregation, a single forecast is built as a linear combination of the ensemble members. The weights of the combination may depend on both time and space. The resulting aggregated forecast is hopefully more skillful than the ensemble members. Cesa-Bianchi and Lugosi (2006) detail the strong mathematical background of these methods, which is summarized and tested by Stoltz (2010) and Mallet et al. (2009); Mallet (2010) on forecasts of respectively electricity consumption and ozone concentrations.

We propose here to compare the ensemble forecasts of solar radiation from TIGGE (THORPEX Interactive Grand Global Ensemble (Bougeault et al., 2010)) and to combine them. The satellite observations from HelioClim are used in this article

as reference observations. The use of both ensemble forecasts (from several sources) and satellite observations makes it possible to generate an aggregated forecast with local combinations on the spatial grid. The Integrated Forecast System (IFS) from ECMWF produces our reference forecast.

In Section 2, we describe the TIGGE data sets and we study the performance of the TIGGE ensembles. The HelioClim satellite observations are introduced in Section 2.4, along with a detailed comparison with TIGGE forecasts. Our sequential-aggregation strategy is introduced in Section 3. It is applied in Section 4, where the analysis of the results includes the composition of the ensemble, the spatial patterns in the forecast maps, the forecast time horizon and the sensitivity to the aggregation parameters.

## **2. Analysis of TIGGE solar radiation and HelioClim database**

### *2.1. Description of TIGGE data*

Several meteorological centers provide free-of-charge ensemble forecasts of solar radiation in TIGGE (Table 1). The data sets are available on TIGGE with a 2-day delay after the model ran. Detailed studies of the ensemble forecasts from TIGGE were achieved on geopotential height (Buizza et al., 2005), and 850-hPa and 2-m temperatures (Hagedorn et al., 2012) for example. As far as we know, no such study exists for solar radiation.

The temporal resolution of TIGGE forecasts is 6 h with time horizon up to 15 days. Most meteorological centers provide at least two forecast sets per day. For the sake of clarity and because of the 6-h timestep of TIGGE forecasts, we focus on one forecast set per day for each meteorological center. With these constraints, 6 meteorological centers provide an ensemble: China Meteorological Administration (CMA), European Centre for Medium-Range Weather Forecasts (ECMWF), MetOf-

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

Table 1: Overview of TIGGE ensembles available for solar radiation, with forecasts starting from 1800 or 2400 UTC for the following 108 to 360 hours (horizon). In total, there are 158 ensemble members.

France (UKMO), Korea Meteorological Administration (KMA), Centro de Previsao Tempo e Estudos Climaticos (CPTEC), and Météo-France (M.-F.). We name “whole ensemble” the ensemble including all 158 members without consideration of their origins, as opposed to the 6 TIGGE center ensembles.

While each center ensemble has a native spatial resolution, our TIGGE data sets are obtained on a common regular grid with the spatial resolution of  $0.25^\circ \times 0.25^\circ$ , which is finer or close to the native resolutions. The resolution of the TIGGE data sets is coarser than the  $0.125^\circ \times 0.125^\circ$  resolution of the ECMWF deterministic forecast. Our study focuses on Metropolitan France and the surrounding areas for the 0600–1200 UTC accumulation period of day D with model runs starting at 1800 or 2400 UTC in day D-1. Thus we study the forecasts for either 12 h or 18 h of lead time. We do not study daily radiation but focus on the shortest timestep available in TIGGE. The flux values from TIGGE are averaged over the 6-h timestep so that the values of the forecasts for 1200 UTC are expressed in  $\text{W m}^{-2}$  and correspond to the averaged flux between 0600 and 1200 UTC.

The nature of the solar radiation data sets in TIGGE is mostly net shortwave solar radiation (SSR) as defined for classical meteorological fields. Net shortwave solar radiation is the fraction of the downwards shortwave solar radiation (SSRD) absorbed by the ground on an horizontal plane. Note that KMA data are different from the other TIGGE data sets and are SSRD

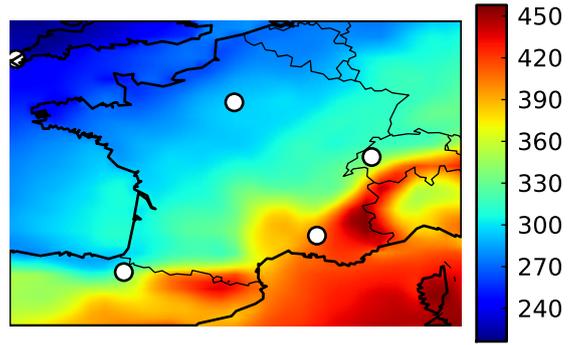


Figure 1: Annual average of 6-hour forecasts at 1200 UTC in  $\text{W m}^{-2}$ , for the KMA ensemble mean in year 2012. The white circles exhibit the locations of ground observation sites for HelioClim evaluation.

data. In the context of photovoltaic production, we are interested in SSRD. The well-known global horizontal irradiance (GHI) refers to SSRD. The albedo coefficient  $\alpha$  is the reflection coefficient of the ground. The albedo defines a relationship between SSR and SSRD, where the incident flux is divided between the absorbed flux and the reflected flux. Thus we deduce that:

$$\text{SSRD} = \frac{\text{SSR}}{1 - \alpha}. \quad (1)$$

Depending on the ground surface, the albedo coefficient can vary in space and time.

## 2.2. Analysis of the TIGGE ensembles of forecasts

Now we analyze and compare the ensembles of forecasts over the area spanning  $41^\circ$  to  $51.50^\circ$  in latitude and  $-5.50^\circ$  to  $10^\circ$  in longitude. Two data sets are used in our study. The main data set consists of 350 consecutive days starting on 2012-01-02. The secondary data set is dated from 2011-06-09 to 2011-09-05 due to data availability and includes only 100 random locations. The secondary data set is only used as learning data set. An example of the annual average of the 6-hour forecasts at 1200 UTC for one center (KMA) is provided in Figure 1, in order to show the large spatial variability of the average forecast.

We propose two sorting procedures so as to consistently

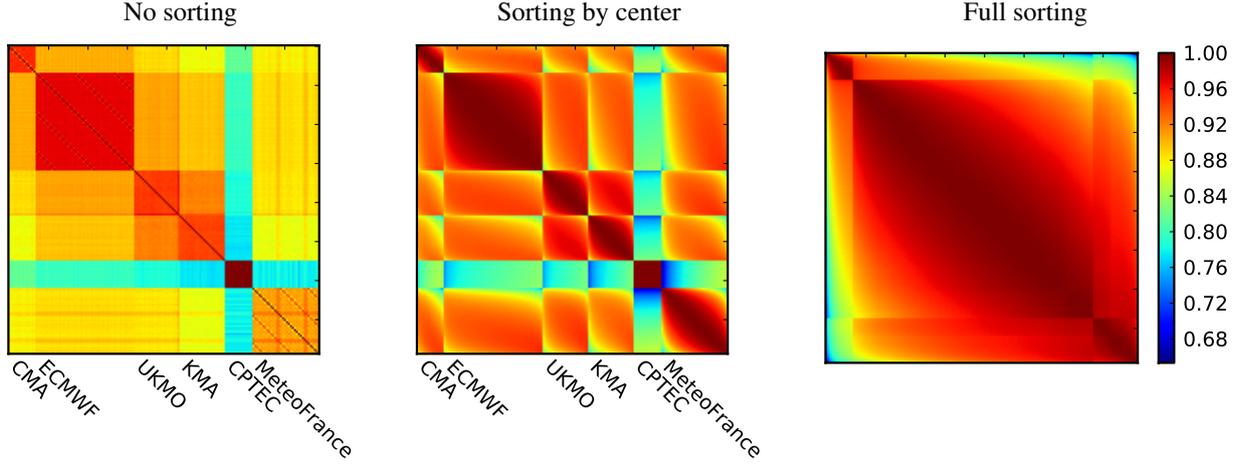


Figure 2: Matrices of correlation between members ( $R^2$  matrices). Each row and column of the correlation matrices is dedicated to one given member. The rows and columns appear in the same order. Each matrix entry is the correlation coefficient between the merged data (all timesteps, 100 random locations) of a pair of members. On the left, the correlations are shown between raw members, as they are retrieved from TIGGE. In the middle, the correlations are computed after sorting within each center ensemble. On the right, the correlations are obtained after a full sorting of the forecasts.

number the forecasts in time. This step is required for the weights in our aggregation to be clearly associated with a given ensemble member and therefore to be meaningful. The two sorting procedures rely on the rank at each grid point. At any time and location, the first member of a sorted ensemble always provides the lowest value, and the last member gives the highest value. The first sorting procedure is applied to the 158 members of the whole ensemble, and the second sorting procedure is applied separately in each center ensemble.

Correlation coefficients between members are computed in order to quantify the similarities between the members (Figure 2). In the correlation matrices, the rectangles and squares materialize some separation between the different ensembles. Indeed the correlation matrices reveal that the members are especially close to one another within the same ensemble. In other words, one member from a given meteorological center ensemble is more correlated to another member of the same ensemble than to another member from a different ensemble. The CPTEC ensemble is very distinguishable due to the extremely high correlations between its own members and also due to the low resemblance between its members and the others.

Sorting has generally two effects on the ensembles. First, the sorted members with close ranks have higher correlations among them than non-sorted members. Second, the pairs of members with the lowest correlation coefficient are found between sorted members of extreme rank.

### 2.3. Reference performance measures

The strengths and weaknesses of the statistical indicators commonly used in solar forecasting are developed in Hoff et al. (2013). The well-known root mean square error (RMSE) and mean absolute error (MAE) are classical performance indicators. The RMSE of the predictions  $\hat{y}$  with respect to the observations  $y$  over the set  $\mathcal{S}$  is given by

$$RMSE = \sqrt{\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} (\hat{y}_s - y_s)^2}, \quad (2)$$

where  $|\mathcal{S}|$  is the number of elements (cardinality) in  $\mathcal{S}$ , and  $s$  indexes space or time or both. In case  $s$  describes all locations at one single timestep, the spatial RMSE is computed. The MAE is calculated in a similar way as

$$MAE = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} |\hat{y}_s - y_s|. \quad (3)$$

It is noticeable that the errors  $\hat{y}_s - y_s$  are computed independently for each  $s$ , hence errors due to geographical or temporal shifts are not detected as such. In case of missing predictions, the missing indices are excluded from the set  $\mathcal{S}$ .

The average observed value is introduced to define relative indicators; e.g. for the relative RMSE:

$$rRMSE = \frac{RMSE}{\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} y_s}, \quad (4)$$

where rRMSE and RMSE are computed over the same set  $\mathcal{S}$ . Except for the scores of Table 2, HelioClim is used as the observation  $y_s$ . Because our main interest is to provide information over land and not over sea, all scores and regression coefficients are given for land locations only.

## 2.4. Comparison with HelioClim

### 2.4.1. Description of HelioClim data

Provided by Transvalor (Armines), HelioClim satellite observations are based on MSG satellite data (Meteosat Second Generation). The operational version HC3-v4 is used here. The HelioSat2 method (Rigollier et al., 2004) has been deployed to generate this HelioClim database (Blanc et al., 2011). The satellite estimations rely on instantaneous reflectance measurements. Data are acquired every 15 min, converted to an estimation of the incident radiation flux, and averaged over the 15-min timestep. The spatial resolution of HelioClim over France is natively between 3 and 5 kilometers; our HelioClim data were retrieved with a spatial resolution of  $1/12^\circ$ , which is already much finer than the resolution of the forecasts.

Our zone of interest includes five BSRN stations (Baseline Surface Radiation Network) (Ohmura et al., 1998) for the evaluation of HelioClim performance. The stations are located in Camborne (United Kingdom), Cener (Spain), Carpentras (South France), Palaiseau (North France), and Payerne

Station	Bias	RMSE	MAE
Camborne	-1.4 % (-3.3)	10.6 % (25.6)	8.3 % (20.0)
Palaiseau	3.4 % (8.5)	10.7 % (26.9)	8.6 % (21.5)
Payerne	-6.9 % (-18.3)	14.4 % (38.3)	10.9 % (29.1)
Carpentras	1.3 % (4.6)	8.3 % (28.9)	6.7 % (23.3)
Cener	2.9 % (9.7)	9.5 % (32.0)	7.6 % (25.6)

Table 2: Evaluation of HelioClim-3 estimation of SSRD compared to in-situ measurements (daily average of daytime data). The relative scores are given, followed by the absolute scores in brackets (in  $\text{W m}^{-2}$ ).

(Switzerland), as exhibited in Figure 1. The evaluation results (Table 2) are computed over several years and show an average relative RMSE of 10.7% and an average relative MAE of 8%.

In Section 4, we wish to produce our own forecasts based on the previously described ensembles and we wish our forecasts to be at the finest available resolution, which is the spatial resolution of our HelioClim data. Consequently our study is carried out at the resolution of HelioClim. All the forecasts are interpolated by bilinear interpolation to reach the same resolution of  $1/12^\circ$ , for a total amount of  $127 \times 187$  grid points. On the high-resolution grid, the forecasts vary spatially slowly compared to the satellite observations.

### 2.4.2. TIGGE ensembles and HelioClim

The difference of nature between the HelioClim incident radiation (SSRD) and the ground-absorbed radiation from TIGGE (SSR) prevents a direct comparison. Therefore we test several conversion methods in Appendix A. The method of linear conversion (lin) based on historical data shows the best RMSE. This method is used until the end of the section.

The monthly RMSEs are impacted by the seasonal variability (Figure 3). Solar radiation forecasting is more difficult between April and July, during the brightest days with large variability. It is worthy of notice that the ranking of the ensemble means is steady over time. The ensemble means of KMA and UKMO (after linear conversion) show better scores than the reference forecast, while their spatial resolution is natively poorer. An improved forecast, called TIGGE-mean, may be built as the mean

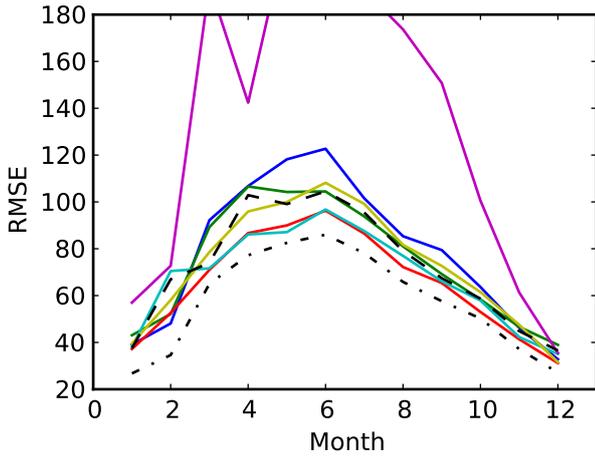


Figure 3: Monthly RMSEs of TIGGE ensemble means in  $W m^{-2}$ , after linear conversion (lin); (blue: CMA), (green: ECMWF), (red: UKMO), (cyan: KMA), (magenta: CPTEC) and (yellow: Météo-France). The dashed line is the score of the reference forecast. The dashed-dotted line is the score of a typical aggregated forecast (see Section 4.2.3).

of the TIGGE ensemble means (CPTEC excluded) with linear conversion. The score of the TIGGE-mean lies at  $66.9 W m^{-2}$ , which is better than any TIGGE ensemble mean.

The same analysis was undertaken with the MAE (not shown) and reveals similar trends:  $54.4 W m^{-2}$  (reference forecast MAE),  $50.2 W m^{-2}$  (UKMO ensemble mean with linear conversion MAE),  $48.0 W m^{-2}$  (TIGGE-mean MAE).

We highlight the fact that no center ensemble is steadily the closest to the satellite observations, whatever the score discrepancy described above. If we track over the consecutive timesteps the origin of the best member at each location, we find that this origin changes from one timestep to the next with a frequency of 75%. Furthermore, the proportion of times and locations where the best member belongs to a given center is reported here: reference ECMWF 3%, CMA 10%, ensemble ECMWF 26%, UKMO 19%, KMA 18%, CPTEC 2%, Météo-France 23%. Considering the fact that the ECMWF reference forecast is one single member compared to 158 members, its frequency of being the closest member to the observation is rather high. On the opposite, the CPTEC ensemble and, to a lesser extent, the CMA ensemble show the lowest frequencies.

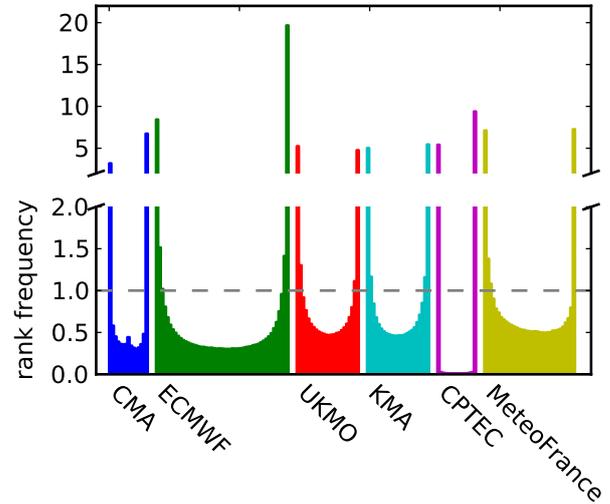


Figure 4: Rank histograms of the center ensembles. In the ideal case of a flat histogram, the height of all the bars would be equal to 1.

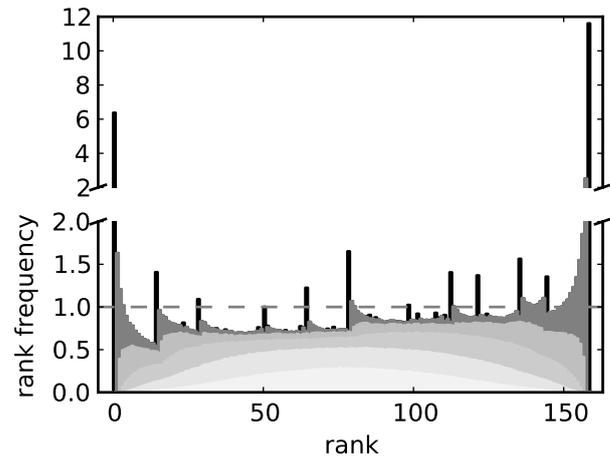


Figure 5: Rank histograms for the whole ensemble. In the ideal case of a flat histogram, the height of all the bars would be equal to 1. The gray scale indexes the number of ensembles whose spread contains the observation, from black (no ensemble) to white (all ensembles, but not observed because of the very low dispersion of CPTEC).

Rank histograms (Anderson, 1996; Talagrand et al., 1999; Hamill and Colucci, 1997) evaluate the quality of the spread of an ensemble. Each observation is given a rank, which corresponds to the number of members with lower values than the considered observation. Then the distribution of rank frequencies can reveal the presence of under-dispersion (U-shaped histogram), over-dispersion, and biases. The rank histograms of the center ensembles and the rank histogram of the whole 158-member ensemble are shown in Figure 4 and 5, for data converted by center (lin). The values of the bars of each histogram are normalized with respect to the total number of observations, so that the height of the bars of the ideally flat histogram is always 1. There are clear outliers on all histograms, which is a clear sign of general under-dispersion (Hamill, 2001). The frequencies with which observations fall within an ensemble envelop are: CMA 34%, ECMWF 45%, UKMO 58%, KMA 56%, CPTEC 1%, Météo-France 59%, whole ensemble 89%.

The whole ensemble is less under-dispersed than each center ensemble and exhibits a rank structure with peaks. The ranks of the major inner peaks (14, 28, 37, 78, 112, 121, 135 and 144) correspond to combinations of ensemble sizes, starting from either extreme peak. For example, the peak at rank 28 corresponds to the combination of the ensemble sizes of CPTEC and CMA (both 14). The peak at rank 135 corresponds to an ensemble of 23 members (UKMO or KMA) starting from the outer peak at rank 158. The peaks are generated by the 4% share of the observations (inner black bars) which do not fall within any of the center ensemble envelops. Indeed these observations are necessarily indexed at ranks related to combinations of ensemble sizes. We conclude that the peaks are due to the variety of the under-dispersed ensembles.

### 3. Ensemble forecast strategy: sequential aggregation

In this section we detail the principles of sequential aggregation. In particular, we explain the method of discounted ridge regression. The performance and robustness of discounted ridge regression have been tested for the case of air quality (Mallet et al., 2009). While time series of scalar fields were considered above, in this section we only consider scalar time series.

#### 3.1. Notation

Let  $x_{m,t}$  describe the  $m$ -th member of our forecast ensemble at time  $t$ , with  $m \in \{1, \dots, M\}$  indexing the members and  $t \in \{1, \dots, T\}$  indexing the forecast timesteps. The vector  $\mathbf{x}_t$  refers to the ensemble of forecasts  $[x_{1,t}, x_{2,t}, \dots, x_{M,t}]^\top$ . At each timestep, the members should be conveniently combined with the weights  $w_{m,t}$  to generate the forecast

$$\hat{y}_t = \sum_{m=1}^M w_{m,t} x_{m,t} \quad (5)$$

of the observation  $y_t$ .

#### 3.2. Sequential aggregation: method

The aggregation weights  $w_{m,t}$  are updated before the forecast step  $t$ , using only past observations  $y_1, y_2, \dots, y_{t-1}$  and past simulations  $\mathbf{x}_1, \mathbf{x}_2$  and  $\mathbf{x}_{t-1}$ . For the discounted ridge regression with parameters  $(\lambda, \gamma)$ , the weight vector  $\mathbf{w}_t = [w_{1,t}, w_{2,t}, \dots, w_{M,t}]^\top$  is found through minimization of

$$J(\mathbf{u}) = \lambda \|\mathbf{u} - \mathbf{w}_{\text{ref}}\|_2^2 + \sum_{t'=1}^{t-1} \beta_\gamma(t-t') \times (y_{t'} - \mathbf{u} \cdot \mathbf{x}_{t'})^2 \quad (6)$$

with

$$\beta_\gamma(t-t') = 1 + \frac{\gamma}{(t-t')^2} \quad (7)$$

and  $\mathbf{w}_{\text{ref}}$  a reference weight vector chosen by the operator and constant in time. The parameter  $\lambda$  affects the distance between  $\mathbf{w}$  and the reference vector  $\mathbf{w}_{\text{ref}}$  (usually set to zero or to

$[1/M, \dots, 1/M]^\top$ , following the ensemble mean). The function  $\beta_\gamma$  gives higher importance to the most recent timesteps. When both  $\lambda$  and  $\gamma$  are set to zero, a simple recursive least-square regression is achieved.

The classical ridge regression without discount provides the theoretical guarantee that the final score of the aggregated forecast will be close to the final score of the best constant linear combination. Indeed we have

$$\sum_{t=1}^T \frac{1}{T} (y_t - \hat{y}_t)^2 - \min_{\mathbf{u} \in \mathcal{B}_M} \sum_{t=1}^T \frac{1}{T} (y_t - \mathbf{u} \cdot \mathbf{x}_t)^2 \leq \mathcal{O}\left(\frac{\ln T}{T}\right), \quad (8)$$

under the assumption of bounded losses  $(y_t - \hat{y}_t)^2$ . The condition  $\mathbf{u} \in \mathcal{B}_M$ , where  $\mathcal{B}_M$  is a 2-norm ball in  $\mathbb{R}^M$ , means that  $\|\mathbf{u}\|$  is bounded. The best linear combination with constant weights (implicitly defined above by the second term in the left hand side) is named the oracle. The oracle is found by least-square regression over the whole set. By definition, the oracle shows better performance than any member in the ensemble. Consequently the aggregated forecast is more skillful in the long run than the best member. The discounted ridge regression provides asymptotically this guarantee, which is verified for each sequence of discounted regret.

### 3.3. Algorithm

Parameters:  $\lambda, \gamma, \mathbf{w}_{\text{ref}}$ ;

Initialization:  $\mathbf{w}_1$ ;

For each time index  $t = 1, 2, \dots, T$

1. get the predictions  $\mathbf{x}_t$ ,
2. compute  $\hat{y}_t$ , with  $\mathbf{x}_t$  and  $\mathbf{w}_t$ ,
3. get the observation  $y_t$  and compute  $\mathbf{w}_{t+1}$ .

The initial weight vector  $\mathbf{w}_1$  is arbitrarily set, e.g., to  $[1/M, \dots, 1/M]^\top$ .

## 4. Application

### 4.1. Experiment setup

The sequential aggregation with discounted ridge regression as described in 3.2 is applied independently at each location of the  $127 \times 187$  grid. In a similar fashion to the study of Section 2, the forecast variable is the incident radiation flux integrated between 0600 and 1200 UTC. The TIGGE forecasts are available at 1800 or 2400 on day D-1 to forecast the quantity of interest for day D at 1200, also named (D, 1200).

The ensemble data are SSR data from TIGGE without any SSR-SSRD conversion because the aggregation does not depend on any multiplicative coefficient applied to the members. In other words, the weights of the aggregation produce multiplicative corrections and solve the issue of the nature of TIGGE data.

The aggregation parameters  $(\lambda, \gamma)$  are respectively set to  $6 \times 10^6$  and 20 by default (see Section 4.2.6 for the assessment of the parameters). The values of the reference vector  $\mathbf{w}_{\text{ref}}$  are set to  $1/M$  in order to drive the aggregated forecast towards the ensemble mean. Even though the ensemble mean may not be the most appropriate reference vector, the vector  $\mathbf{w}_{\text{ref}}$  mostly impacts the beginning of the aggregation so that the critical parameters are truly  $(\lambda, \gamma)$ .

The aggregation may be achieved in a single step by choosing members from all of the center ensembles. Another approach involves two steps: a first aggregation within each center ensemble and a second aggregation with the resulting forecasts as members. Both procedures have been tested and only the procedure achieved in a single step is presented here since the second procedure did not lead to significantly different results.

In order to study the impact of the number of members  $M$ , the same amount of members are chosen from each meteorological center, but the center ensembles do not have the same size. Therefore, the members are chosen in a way that their

ranks are linearly spaced and centered on the median of each center ensemble. The full sorting procedure is not impacted by this member selection because in this case the members are mixed up before sorting and rank selection. It is possible to realize aggregation with one single member, such as the reference forecast. In this case the weight plays the role of a local correction factor.

Missing data (only CPTEC ensemble) are replaced by the ensemble mean of the available members.

## 4.2. Results

### 4.2.1. Aggregation example with one center ensemble

In this section, the aggregation is first run with the ensemble KMA only (without sorting) and then with an additional deterministic member. In the first case, we indicate that the RMSE of the SSRD ensemble mean is  $75.0 \text{ W m}^{-2}$ , whereas the RMSE of the ensemble mean with linear conversion is  $72.1 \text{ W m}^{-2}$ . The RMSE of the aggregated forecast equals  $70.0 \text{ W m}^{-2}$  and may decrease when data from another source is included in the ensemble. For example, when the ECMWF deterministic forecast is added as a member, the RMSE reaches  $67.5 \text{ W m}^{-2}$ . The scores are to be compared to the  $58.9 \text{ W m}^{-2}$  RMSE of the oracle of the same ensemble and to the  $65.1 \text{ W m}^{-2}$  RMSE of the oracle with only KMA data.

One question is whether the improvement due to the deterministic member originates from its high spatial resolution. We therefore include the ECMWF deterministic forecast at the lower resolution of  $0.25^\circ \times 0.25^\circ$  which is obtained by averaging the fine  $0.125^\circ \times 0.125^\circ$  forecast. In that case, the RMSE is also equal to  $67.5 \text{ W m}^{-2}$ , which means that the resolution of the reference forecast is not the key factor. For comparison, the RMSE of the reference forecast corrected by discounted ridge regression is equal to  $68.0 \text{ W m}^{-2}$ .

### 4.2.2. Oracles with orthogonal members

We now want to quantify the potential improvements brought by sequential aggregation with all members, using a relevant oracle. The problem of overfitting can arise because of the large number of members (158) compared to the length  $T = 350$  of the time sequence. In that case, the score of the oracle is artificially good, and the competition against the oracle does not carry any meaning. We present here oracles computed with the help of principal component analysis (PCA), so as to avoid overfitted oracles. The PCA generates orthogonal modes, and consequently orthogonal members. The 158 members are sorted within each center ensemble and then orthogonalized by PCA, independently at each grid point. We compute the oracle of the ensemble of size  $M'$ , by selecting  $M'$  orthogonal members based on the  $M'$  modes explaining the largest variance share of the ensemble. In Figure 6 the RMSE of the oracle and the total amount of variance explained by its members are plotted against the number of orthogonal members. An indication of possible overfitting is shown in Figure 6, because the RMSE still decreases with the number of PCA modes while the share of unexplained variance is small. Indeed the 32 first orthogonal members explain 95% of the variance and generate an oracle with an RMSE of  $54.4 \text{ W m}^{-2}$ , whereas the RMSE of the oracle with all PCA modes (100% of explained variance) equals  $33.9 \text{ W m}^{-2}$ . Consequently, the 32-member oracle is considered as the relevant oracle; its score is a relevant evaluation of the best score that may be achieved by linear combination without overfitting.

### 4.2.3. Typical aggregation

In this paper, the ‘‘typical aggregation’’ refers to the aggregation with 30 members (5 per sorted center ensemble), with default parameters  $\lambda = 6 \times 10^6$  and  $\gamma = 20$ . The resulting forecast is simply referred to as typical aggregated forecast and is used below to illustrate spatial and temporal features resulting

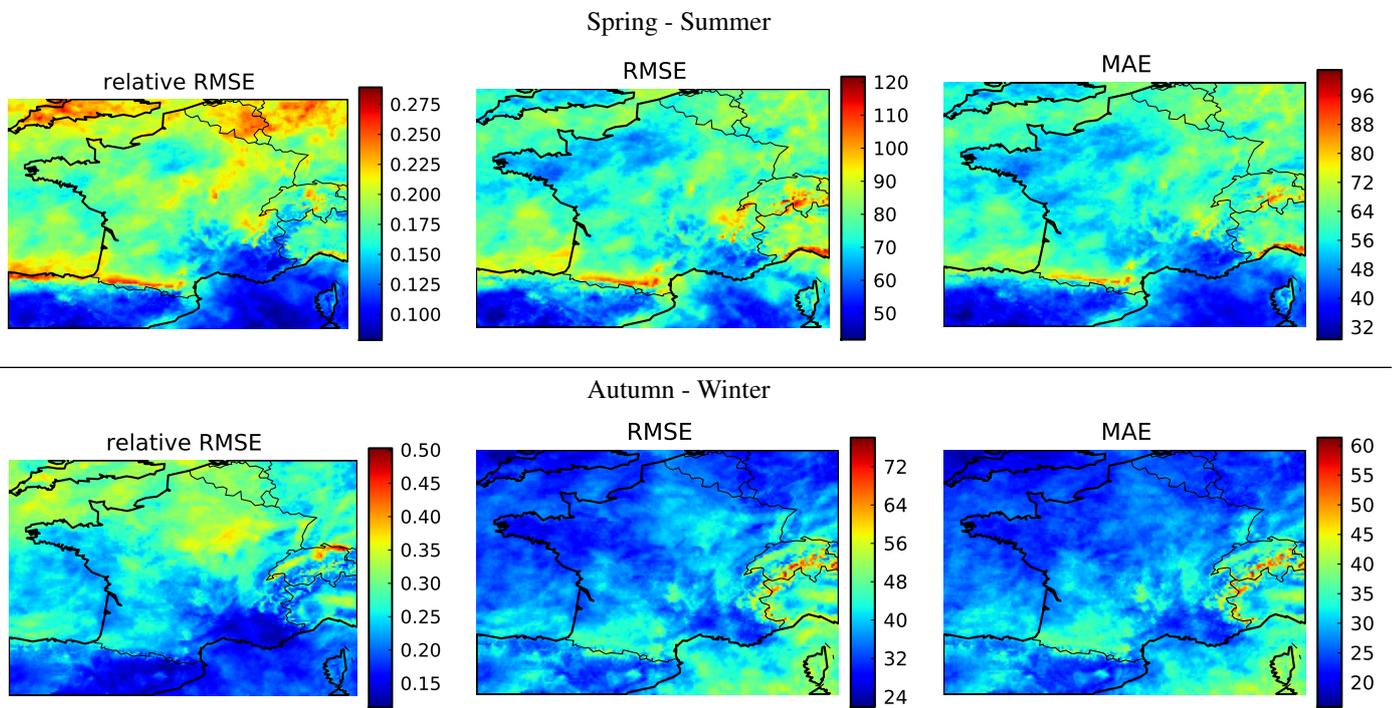


Figure 7: Maps of statistical scores for the typical aggregated forecast (unitless for rRMSE, in  $\text{W m}^{-2}$  for MAE and RMSE).

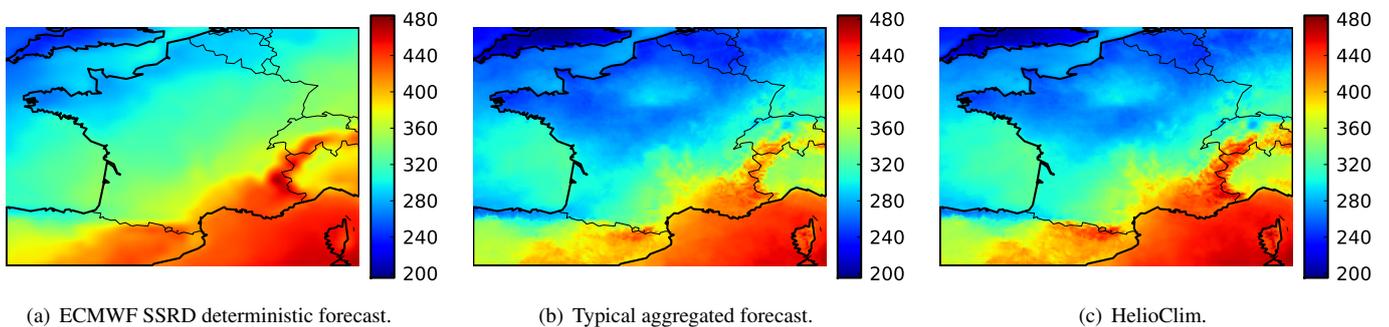


Figure 8: Annual averaged estimation for 1200 UTC in  $\text{W m}^{-2}$ . Although it is based on individual members with low resolution, the aggregated forecast shows fine structures that are comparable to those of HelioClim.

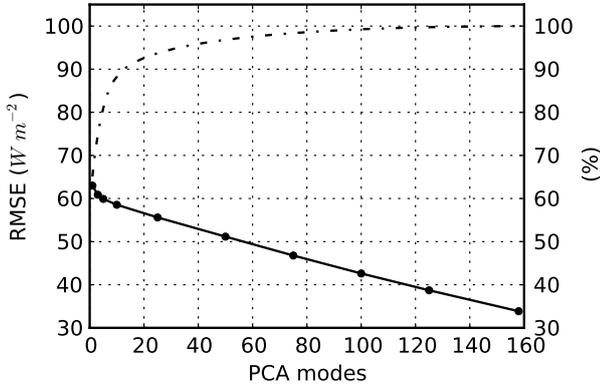


Figure 6: RMSE (solid line with disks, left axis) and explained variance (dashed-dotted line, right axis) of oracles with orthogonal members plotted against number of orthogonal members.

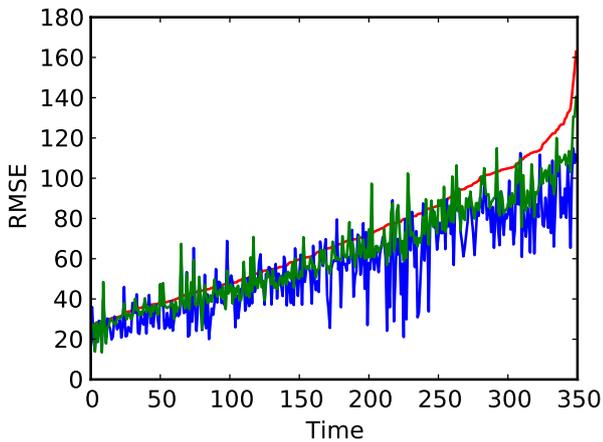


Figure 9: Sorted spatial RMSE in  $W m^{-2}$  along 350 days; ECMWF deterministic forecast (red and used to sort the days), typical aggregated forecast (blue), ECMWF deterministic forecast corrected by discounted ridge regression (green).

from sequential aggregation. Note that the typical aggregation is compared to all members in Figure 13 and to the ensemble means for each month in Figure 3. We highlight the fact that the typical aggregated forecast performs better than any of the individual members, steadily over time.

The maps of scores of RMSE, MAE and relative RMSE are shown in Figure 7. The MAE and RMSE maps show very similar patterns. The relative MAE is not shown for it is similar to the relative RMSE. The maps indicate that the topographic relief of mountains severely degrades the forecast accuracy. Indeed the absolute largest errors are found in the region of the Alps and in the region of the Pyrenees. The coarse resolution of the members may explain the poor performance in the regions of complex terrain. Also, in the mountainous regions, the albedo shows higher temporal variations, which is difficult to catch for the aggregation. On the opposite, the best scores are steadily achieved in the inner lands of Spain and in the south-east of France. High relative errors are numerous in the northern area especially during the spring-summer period, because of the low values of the observations combined with large errors. For example in spring and summer period, the British area shows a relative RMSE higher than 20%, whereas the south-east of France reaches a relative RMSE below 12.5%, even though the two areas are associated with similar RMSEs over the same period. We recall here that the relative RMSE of the satellite observations compared to BSRN stations is worth 10.7% on average.

Most of the members are computed with a low spatial resolution and do not show fine long-term spatial structures. On the contrary, the aggregated forecast is built independently at each location, which allows the procedure to adapt locally and to finally show fine structures, that are resolved by HelioClim. These structures are finer than any structures found in the ensemble members (due to their low resolutions), even in the

ECMWF deterministic forecasts (Figure 8).

In order to compare predictions performance at each date, the spatial RMSE of various predictions are temporally sorted according to the performance of the ECMWF reference forecast (Figure 9). The overall trend is the same for the three forecasts. The aggregated forecast often shows the best performance, but it does not always perform better than the aggregation applied to the reference forecast only. One benefit of aggregation is that large improvements are achieved at the most difficult timesteps, even with one single member in the ensemble.

#### 4.2.4. Members selection and sorting

The performance of the aggregated forecast is impacted by the number of members up to a certain extent (Figure 10). Beyond a total amount of roughly 60 members (10 members from each center ensemble), the scores reach a plateau. The full sorting procedure is the first to reach its own plateau, which may be caused by the merger process of the members with sorting. Besides, the sorting procedures impact the scores with limited consequences (less than  $1 \text{ W m}^{-2}$ ) for large enough ensembles. According to the observed performance, the sorting procedure by center should be preferred if more than 5 members per center are chosen. It is noticeable that the reference forecast brings a score improvement whatever the number of aggregated members. The best score achieved with default  $(\lambda, \gamma)$  equals  $61.5 \text{ W m}^{-2}$  for the RMSE and  $43.6 \text{ W m}^{-2}$  for the MAE. Compared to the scores of reference forecast, the improvements brought by the aggregation equal 21.2% for the RMSE and 19.8% for the MAE. Besides, compared to the scores of TIGGE-mean, these improvements are worth 8.1% for the RMSE and 9.1% for the MAE.

The aggregation ensemble may also be built with only 5 out of 6 center ensembles. We compare the aggregated forecast of 30 members, so that 6 members are chosen in each of 5 center ensembles. In this case, the reference aggregated forecast is the

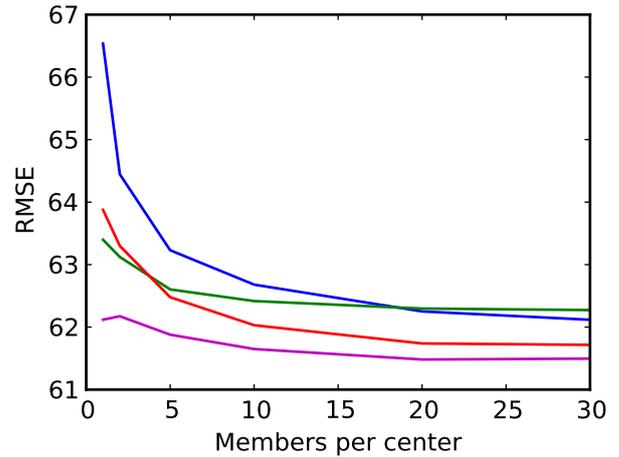


Figure 10: RMSE in  $\text{W m}^{-2}$  against the maximal number of members chosen from each center ensemble; (blue: no sorting), (green: sorting full ensemble), (red: sorting by center), (magenta: sorting by center and reference forecast included in the ensemble).

typical aggregated forecast with 5 members chosen in each of the 6 center ensembles. Either way, the members are sorted per ensemble. We found that the aggregated forecast with 5 centers performs better than the typical aggregated forecast, when the omitted members are from CMA or CPTEC. The largest improvement occurs when the CPTEC is excluded with a benefit of only  $0.3 \text{ W m}^{-2}$ . On the opposite, the score is the most severely degraded when the members from KMA are left out, generating the RMSE of  $64.1 \text{ W m}^{-2}$ , while the RMSE of the typical aggregated forecast equals  $62.5 \text{ W m}^{-2}$ . To conclude, not using all center ensembles to build the aggregation ensemble can generate both benefits and loss in terms of score. However, these benefits are lower than the loss caused by leaving out skillful forecasts. In practice, one may arguably keep all center ensembles.

#### 4.2.5. Time horizon

We consider an ensemble forecast delivered at time  $t$  for the  $k$  next steps. The weights of the sequential aggregation are commonly used for the first step ahead (see the algorithm, in Section 3.3), but can also be used for the following timesteps  $t + k$ , where  $k > 1$ . In fact, the weights can even be used for timesteps

that are not included in the sequential aggregation, such as the forecast for 1800 UTC. This new framework allows us to generate the aggregated forecast for (D, 1800), (D+1, 1200) and (D+1, 1800) with the weights initially computed for the prediction (D, 1200). In terms of time horizon, the prediction for (D, 1200) corresponds to the 12 h horizon, (D, 1800) to 18 h, (D+1, 1200) to 36 h and (D+1, 1800) to 42 h. The time horizons of Météo-France forecasts are actually 6 h longer, because its predictions start at 1800 and not 2400. The night time predictions steps are skipped because of their very low values.

The RMSE of the ensemble means and the scores the typical aggregated forecast are shown in Figure 11, depending on the time horizon. The scores of CPTEC ensemble mean are not shown, because of their high values. The six presented forecasts share the same trends. The scores of predictions for 1800 beat the scores for predictions 1200 of the same day, by less than  $4 \text{ W m}^{-2}$  on average, which is consistent with the fact that the forecasts for 1200 have higher values than the forecasts for 1800. Furthermore, the predictions for day D+1 show an average degradation of more than  $5 \text{ W m}^{-2}$  compared to the prediction for day D. Additionally, we notice that the ensemble mean from ECMWF shows the steadiest performance over time and that the benefits of the aggregated forecast are retained for longer time horizons.

#### 4.2.6. Influence of the aggregation parameters

We searched the best set of parameters  $(\lambda, \gamma)$  that provides the lowest RMSE for the data set of year 2011, already presented in Section 2.2. We found that at most  $3 \text{ W m}^{-2}$  were to be gained in terms of RMSE on a wide range of parameters (Figure 12(a)). Our default parameters  $(6.10^6, 20)$  produce the score of  $67.1 \text{ W m}^{-2}$ , which is close to the best score achieved on the grid of Figure 12(a) (e. g.  $66.7 \text{ W m}^{-2}$  with  $\lambda = 6.10^6$  and  $\gamma = 5$ ). Furthermore, we found that the best set of parameters varies in terms of space and time. Indeed, if we exclude

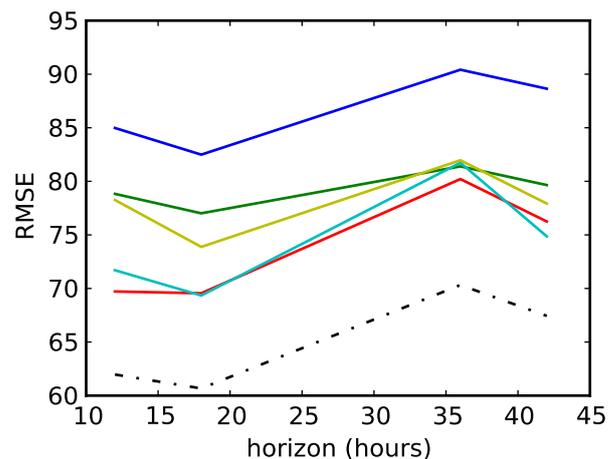
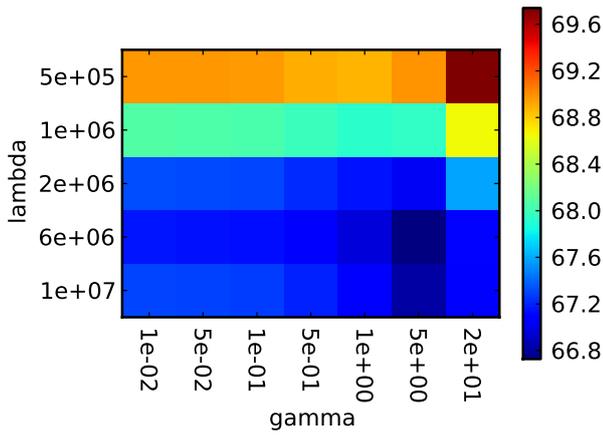


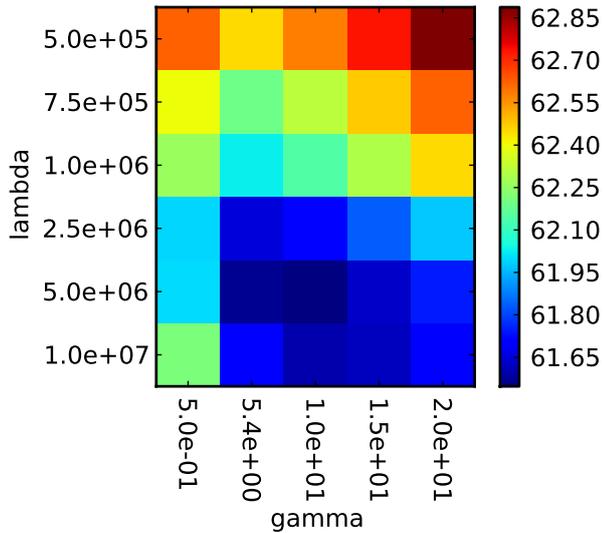
Figure 11: Impact of the horizon up to 42h on the RMSE of the aggregated forecast and ensemble mean (lin); (dash dot: aggregated with 60 members), (blue: CMA), (green: ECMWF), (red: UKMO), (cyan: KMA), and (yellow: Météo-France).

the 50 first timesteps in the RMSE, then the most appropriate parameters are  $\lambda = 10^7$  and  $\gamma = 20$ . If we choose a posteriori the optimal  $(\lambda, \gamma)$  for each grid point, only  $1 \text{ W m}^{-2}$  is to be gained from default parameters. The minor variations of performance guarantee that we may choose our default parameters within one order of magnitude, and test them on another data set without significant loss of performance.

We produce the same analysis for the data set of year 2012 in Figure 12(b). In this case, the best parameters are  $\lambda = 5.10^6$  and  $\gamma = 10$ , with the score of  $61.5 \text{ W m}^{-2}$  that is very similar to the score of  $61.7 \text{ W m}^{-2}$  obtained with default parameters. Therefore, the order of magnitude of the parameters for year 2012 can be deduced from the data set of year 2011. Besides we also found that if we choose the best  $(\lambda, \gamma)$  for each grid point, the gain does not exceed  $1 \text{ W m}^{-2}$ . As stated with the data set of year 2011, once the relevant order of magnitude for the parameters  $(\lambda, \gamma)$  is known, only little improvement is possible by adjusting them.



(a) Data set of year 2011 (RMSE for 100 random locations).



(b) Data set of year 2012 (RMSE for all locations).

Figure 12: RMSE grid ( $W m^{-2}$ ) depending on the aggregation parameters, aggregating a 134-member ensemble using at most 30 sorted members from each center ensemble.

The ensemble forecasts from TIGGE provide a wide range of meteorological fields including net short-wave solar radiation, with the large timestep of 6 h. After conversion based on a constant albedo, the resulting ensembles are under-dispersed, even when grouped together. The performance of the ensemble means is assessed with RMSE and MAE and compared to the deterministic forecast from ECMWF. The reference forecast has higher resolution than the best TIGGE ensemble means, but produces similar scores. Sequential aggregation brings improvements to the TIGGE ensembles on several features, including RMSE and MAE, with theoretical guarantees. In this study, the aggregated forecast performs better than any member of the ensembles and any ensemble mean. On average, the aggregated forecast retrieves all spatial patterns, even at a much finer resolution than any of the members. Besides, the members combination proves to be consistent with the time horizon. Finally sequential aggregation is easy-to-use for its parameters do not need accurate values.

Practical applications of the aggregation algorithm should investigate higher temporal resolutions, especially the hourly timestep. Next developments may focus on the study of uncertainty with sequential aggregation, which is possible using filtering (Mallet et al., 2013), but without the same theoretical robustness as in this paper. The introduction of multiple model runs per day with temporal interpolations may result in a robust framework for intraday forecasting. Furthermore, new procedures are to be developed to explore sequential aggregation with spatial dependencies between grid points, with the objective of better forecasting daily spatial patterns. There is also a need for quantifying the possible resolution improvements brought by the aggregation.

## Acknowledgements

The EDF company (Électricité de France) and ANRT (Association nationale de la recherche et de la technologie) are sincerely acknowledged for providing thesis fellowship and supporting this work. We also express our warm thanks to Luc Musson-Genon (EDF), Stéphanie Dubost (EDF) and Laurent Dubus (EDF) for their remarks and involvement.

## Appendix A Conversion from SSR to SSRD and reference forecast

### A.1 Methods

Several conversion methods allowing the estimation of the albedo are tested. Even though the KMA data set is already SSRD and does not need data conversion, the below methods are also tested on the KMA data set, for the sake of completeness.

First, we infer three conversions from our reference forecast. The SSR and SSRD forecasts from ECMWF are used to provide three SSR-SSRD empirical conversions (Table 3): with a constant 1.18 coefficient (glob), with local multiplicative coefficients (mult), and with local additive coefficients (add). The local multiplicative conversion is in fact the local estimation (in space and time) of the factor  $1/(1 - \alpha)$  based on ECMWF continuous forecast.

Although it may lead to local large errors due to seasonal changes in snow cover and vegetation, we model a constant albedo in space and time (glob). The linear relationship (Eq. (1)) between SSR and SSRD forecasts from ECMWF is found by linear regression on data from 100 random locations and for 350 days in year 2012. The resulting slope value (supposed to be equal to  $1/(1 - \alpha)$ ) is 1.18 and the intercept value equals  $13 \text{ W m}^{-2}$ . The squared correlation coefficient  $R^2$  of 0.968 shows that SSR and SSRD forecasts are strongly correlated in practice. According to the ratio between ECMWF

label	conversion formula
glob	$x_{tigge} \times 1.18$
mult	$x_{tigge} \times \frac{x_{ecmwf}^{ssrd}}{x_{ecmwf}^{ssr}}$
add	$x_{tigge} - (x_{ecmwf}^{ssr} - x_{ecmwf}^{ssrd})$
lin	$x_{tigge} \times a_{center} + b_{center}$

Table 3: Empirical conversion formula of TIGGE data  $x_{tigge}$  from SSR to SSRD. The deterministic forecasts from ECMWF are named  $x_{ecmwf}$ .

SSRD and SSR forecasts, the relative standard deviation (standard deviation divided by mean) of the coefficient  $1/(1 - \alpha)$  is equal to 8.8% on average over all the grid points. Consequently the local variations of the ratio  $1/(1 - \alpha)$  are rather small compared to its local mean value. Besides, more than 90% of the values of the coefficient  $1/(1 - \alpha)$  are comprised between 1.13 and 1.48.

Second, a constant albedo is computed for each center ensemble. This method is referred to as “lin” in Table 3 and is also referred to as the linear conversion below, because the method is based on linear regressions. Taking HelioClim observations as SSRD data and TIGGE ensemble means as SSR data, linear regressions are carried out on past data sets of year 2011 described in Section 2.2. The results of the regressions (Table 4) show two slopes lower than one (KMA and CPTEC), and intercepts ranging from  $-21 \text{ W m}^{-2}$  to  $41 \text{ W m}^{-2}$ . Remind that KMA data is already SSRD data so that its slope is not related to the albedo. The diversity of slopes and intercepts values suggest that the albedo coefficient should be evaluated independently for each center ensemble.

### A.2 Numerical results

The deterministic forecasts from ECMWF provide several scores. We build another deterministic forecast, called SSRD-lin, based on deterministic SSR and the linear conversion method. Compared to HelioClim, the RMSE of the

Center	Slope $a_{\text{center}}$	Intercept $b_{\text{center}}$ ( $\text{W m}^{-2}$ )	$R^2$
CMA	1.11	18	0.72
ECMWF	1.18	-21	0.78
UKMO	1.10	19	0.81
KMA	0.91	18	0.81
CPTEC	0.82	41	0.48
Météo-France	1.07	37	0.76

Table 4: Linear regression SSR-SSRD, based on data from year 2011.

ECMWF SSRD deterministic forecast equals  $78.0 \text{ W m}^{-2}$  while the RMSE of SSRD-lin equals  $76.9 \text{ W m}^{-2}$ . When the largest error percentiles of SSRD and SSRD-lin forecasts are compared, we see that the largest errors belong to SSRD-lin only beyond the 99<sup>th</sup> percentile. On the one hand, the physical approach with local albedos (as opposed to SSRD-lin) does not generate the largest errors. On the other hand, local albedos do not provide the best score.

Furthermore, the resolution of the ECMWF deterministic forecast does not impact its performance. Indeed the ECMWF deterministic forecast at degraded resolution of  $0.25^\circ \times 0.25^\circ$ , by interpolation, shows the RMSE of  $77.7 \text{ W m}^{-2}$  when compared to HelioClim, which is below the RMSE of the reference forecast. The two predictions at different resolutions have similar performance, since the distance between them, measured with a root mean square discrepancy (in time and space), equals  $3.0 \text{ W m}^{-2}$ .

The RMSEs are shown in Figure 13 for all members without sorting procedures. The RMSEs between individual members and satellite observations are at least twice as big as the RMSEs between satellite observations and ground measurements (Table 2). It is noteworthy that members of a same center ensemble perform similarly. The scores strongly depend on the conversion methods and on the origin of each forecast. We highlight again the fact that the nature of KMA data is already SSRD.

The linear conversion (lin) is the best conversion method

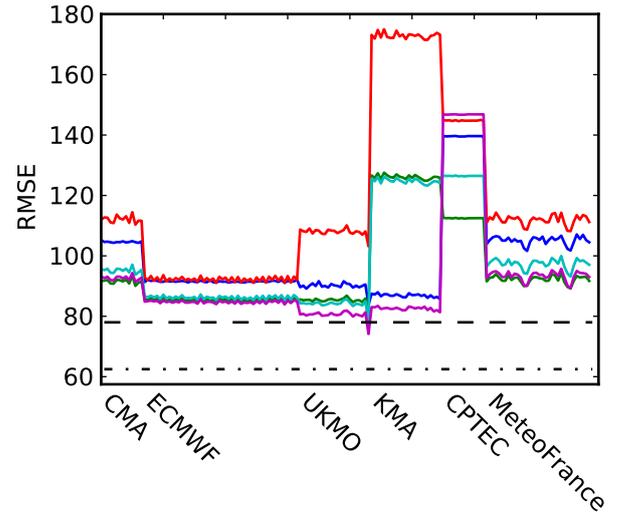


Figure 13: Annual RMSEs of all members in  $\text{W m}^{-2}$  according to SSR-SSRD conversion method; (blue: no conversion), (green: add), (red: mult), (cyan: glob) and (magenta: lin). The abscissa is the member label and the members are grouped by origin. The dashed line is the score of the reference forecast. The dashed-dotted line is the score of a typical aggregated forecast (see Section 4.2.3).

tested here except for the CPTEC ensemble. The multiplicative conversion method was supposed to provide accurate forecasts. However, here is confirmed the need for the conversion method to adapt to each center ensemble.

## References

- Anderson, J. L., 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate* 9 (7), 1518–1530.
- Blanc, P., Gschwind, B., Lefevre, M., Wald, L., 2011. The HelioClim Project: Surface Solar Irradiance Data for Climate Applications. *Remote Sensing* 3 (2), 343–361.
- Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., et al., 2010. The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society* 91 (8), 1059–1072.
- Buizza, R., Houtekamer, P., Pellerin, G., Toth, Z., Zhu, Y., Wei, M., 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review* 133 (5), 1076–1097.
- Cesa-Bianchi, N., Lugosi, G., 2006. *Prediction, learning, and games*. Cambridge University Press.
- Dehghan, A., Prasad, A. A., Sherwood, S. C., Kay, M., 2014. Evaluation and

- improvement of TAPM in estimating solar irradiance in Eastern Australia. *Solar Energy* 107 (0), 668 – 680.
- Espinar, B., Aznarte, J.-L., Girard, R., Moussa, A. M., Kariniotakis, G., et al., 2010. Photovoltaic Forecasting: A state of the art. In: *Proceedings 5th European PV-Hybrid and Mini-Grid Conference*.
- Gautier, C., Diak, G., Masse, S., Aug. 1980. A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. *Journal of Applied Meteorology* 19, 1005–1012.
- Hagedorn, R., Buizza, R., Hamill, T. M., Leutbecher, M., Palmer, T. N., 2012. Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society* 138 (668), 1814–1827.
- Hamill, T. M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review* 129 (3), 550–560.
- Hamill, T. M., Colucci, S. J., 1997. Verification of Eta/RSM short-range ensemble forecasts. *Monthly Weather Review* 125, 1312–1327.
- Hoff, T. E., Perez, R., Kleissl, J., Renne, D., Stein, J., 2013. Reporting of irradiance modeling relative prediction errors. *Progress in Photovoltaics: Research and Applications* 21 (7), 1514–1519.
- Inman, R. H., Pedro, H. T., Coimbra, C. F., 2013. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science* 39 (6), 535–576.
- Lorenz, E., Remund, J., Müller, S. C., Traunmüller, W., Steinmaurer, G., Pozo, D., Antonio, J., Ruiz-Arias, V. L. F., Ramirez, L., Romeo, M. G., et al., 2009. Benchmarking of different approaches to forecast solar irradiance. In: *Proceedings of the 24th European Photovoltaic Solar Energy Conference*. pp. 4199–4208.
- Mallet, V., 2010. Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation. *Journal of Geophysical Research* 115 (D24303).
- Mallet, V., Nakonechny, A., Zhuk, S., 2013. Minimax filtering for sequential aggregation: Application to ensemble forecast of ozone analyses. *Journal of Geophysical Research* 118 (11), 11,294–11,303.
- Mallet, V., Stoltz, G., Mauricette, B., 2009. Ozone ensemble forecast with machine learning algorithms. *JGR* 114 (D05307).
- Morcrette, J.-J., 1991. Radiation and cloud radiative properties in the European centre for medium range weather forecasts forecasting system. *Journal of Geophysical Research: Atmospheres* 96 (D5), 9121–9132.
- Ohmura, A., Gilgen, H., Hegner, H., Müller, G., Wild, M., Dutton, E.G., Forgan, B., Fröhlich, C., Philipona, R., Heimo, A., et al., 1998. Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bulletin of the American Meteorological Society* 79, 2115–2136.
- Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., Hemker Jr, K., Heinemann, D., Remund, J., Müller, S. C., Traunmüller, W., et al., 2013. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Solar Energy* 94, 305–326.
- Perez, R., Seals, R., Zelenka, A., 1997. Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data. *Solar Energy* 60 (2), 89–96.
- Rigollier, C., Lefevre, M., Wald, L., 2004. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. *Solar Energy* 77 (2), 159 – 169.
- 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.
- Talagrand, O., Vautard, R., Strauss, B., 1999. Evaluation of probabilistic prediction system. *Proceedings of the ECMWF Workshop on Predictability*.
- Thelen, J.-C., Edwards, J. M., 2013. Short-wave radiances: comparison between SEVIRI and the Unified Model. *Quarterly Journal of the Royal Meteorological Society* 139 (675), 1665–1679.
- Vernay, C., Pitaval, S., Blanc, P., et al., 2013. Review of satellite-based surface solar irradiation databases for the engineering, the financing and the operating of photovoltaic systems. In: *ISES Solar World Congress*.
- Yokohata, T., Annan, J. D., Collins, M., Jackson, C. S., Tobis, M., Webb, M. J., Hargreaves, J. C., 2012. Reliability of multi-model and structurally different single-model ensembles. *Climate Dynamics* 39 (3-4), 599–616.