DNICast

Direct Normal Irradiance Nowcasting methods for optimized operation of concentrating solar technologies

THEME [ENERGY.2013.2.9.2]

[Methods for the estimation of the Direct Normal Irradiation (DNI)]

Grant agreement no: 608623

Deliverable Nr.: *3.14*	***Deliverable title:*** Report on Combination of nowcasted DNI
Project coordinator: OME	***WP leader:*** *CIEMAT*
Name of the organization: *CIEMAT, DLR, CENER*	***Authors:*** *Lourdes Ramirez (CIEMAT), Natalie Hanrieder (DLR), Stefan Wilbert (DLR), Luis F. Zarzalejo (CIEMAT), J.M. Vindel (CIEMAT), R.X. Valenzuela (CIEMAT), M Gastón (CENER)*
Submission date: *15.8.2017*	***Version nr.:*** *2*

Table of Contents

Abstract

Within the EU FP7 DNICast project (Direct Normal Irradiance Nowcasting methods for the optimized operation of concentrating solar technologies) a set of innovative and improved forecast methods are proposed. One additional task in the project is the proposal and analysis of different options to combine the data provided by forecast sources. This report is focused on this last task of combination, describing the proposed methodologies for the combination. Inputs for the combination come from nowcasting data sets based on different methods already developed in the project. After this task, the data sets are prepared for their validation and their application in CST (Concentrating Solar Technology) models.

After a review of forecasts combination methodologies, the report is focused on showing three different methodologies for combining Direct Normal Irradiance (DNI) nowcast outputs. The first combination methodology weighs the different nowcast data sets according to their uncertainty estimations (UWA). The second combination method uses a time-dependent multi-regressive model (MRA) and the third is based on a weight calculation based on the Euclidean distance between forecasts as functional data (DWA).

The three combination approaches are described in this report. Detailed analysis of the result provided by each methodology is presented in the deliverable 3.13.

Keywords: Nowcasting, DNI, combination, uncertainty

1 General introduction to combining forecasts

Decades ago, it was clearly stated that combining forecast (also called composite forecast) can improve the accuracy of individual forecasts (Clemen, 1989).

The general approach of combining inputs of similar nature for the output improvement was in fact already proposed by Laplace, who claimed that probability law or error will decrease when combining results of two methods (from Armstrong, 2001)

Thus, the forecast combination is related with three main lines of research:

(1) Studies focused in combining estimates.
(2) Studies focused in combining forecasts.
(3) Studies focused in the development of methods for addressing the efficiency of combination models.

These three research topics are in fact in strong relationship to each other. The combining estimates research line was of interest decades ago, but in the late eighties there was an explosion of combining forecasts articles. It was shown in the review articles of Armstrong, 2001 and Clemen, 1989 that the use of a simple combination approach as is the case of equal –weights rule often works reasonably well relative to more complex models.

But this statement depends on the way the results behavior is addressed. Thereby, most of the methods for addressing the efficiency of forecasts uses essentially the composite forecast as benchmark or aggregated differences like: MAD (Mean Absolute Deviation), MAPE (Mean Absolute Percentage Error)…(Chase, 1995) or RMSE (root mean square error).I If the evaluation is based on averaged differences, averaged inputs will provide better results than more complex models.

In parallel to the works demonstrating the advantages of a simple combination, some works tried to look for additional error metrics like the t-statistic and address differences between approaches to model comparison (Cooper and Nelson, 1975)

Some other authors proposed to obtain the combination weighting based on Bayesian inference, where each model's predictions are weighted by their respective probability (Lauret et al., 2012); by the minimum-variance method, where the set of weights which minimizes the variance are sought (Dickinson, 1977); or using regression based methods for the coefficients estimation, what is actually an empirical weights estimation.

More recent models of forecasting combination are using artificial neuronal network or Fussy Inference Systems (FIS) which analyze input values in terms of logical values (Fiordaliso, 1998). Others works are based on a Kalman filter, where estimations are updated with new information (Gupta et al., 2006) and which underlie the Bayesian probability.

In this report we have tested three forecasting combination methods for solar radiation nowcasting combination. Due to the large number of inputs taken into account in this work, only linear methods have been considered, as other techniques are more functional whether the number of variables is reduced.

In the first tested approach we will refer to (Meyer et al., 2008), where the weights of the inputs for the combination are estimated by theirs uncertainties. In the second approach we test a regression process for the weights estimation (Turlapaty et al., 2012), but using the previous forecasted events for the same moment (Armstrong, 2001). In addition, a weighting methodology based on the Euclidean distance between previous forecasts and measures has been implemented (Shrestha et al., 2011). From this technique, that could be considered an empirical estimation of the uncertainty weighted methodology, the weights are not previously established but dynamically estimated. On the other hand, variables are not initially fixed, as happens in the multivariate approach.

The efforts have been focused in review the basics of combining forecasts and address the main issues of this methodology when are applied to solar radiation nowcasting.

2 Overview of the input data sets

In WP 3.4 of DNICast a combination of several nowcasts which have been developed in earlier work packages of the project was generated. The combination was performed for four, three-months long forecast periods and two selected stations.

2.1 Characteristics of nowcasting inputs for the combination

Several nowcast from different methods are available:

1. DNI nowcasting method with all sky images.
2. Satellite based cloud and DNI nowcasting methods.
3. Numerical Weather Prediction (NWP) based nowcasting methods.

In Table 1, the main characteristics of the possible nowcast inputs for the combination are summarized.

Table 1: Main characteristics of the available DNICast nowcast inputs for the combination

Data source	Partner	Refresh interval (min)	Time res. (min)	Horizon (min)	Horizon (hour)
SKY IMAGES	ARMINES	1	5	30	0.5
SATELLITE	DLR-DFD	15	1	480	8
SATELLITE	DLR-IPA	15	5	360	6
SATELLITE	Meteotest	15	5	240	4
SATELLITE	SMHI	180	15	540	9
NWPM	SMHI	180	15	540	9

At each specific location the availability of nowcasts suppliers and supplied periods can vary. Thus, the availability of nowcasts suppliers and periods available at Plataforma Solar de Almería (PSA) are shown in Table 2. “(x)” means that the entire month is not available.

Table 2: Available nowcast periods and suppliers at PSA

Data source	Partner	2010 JAN	2010 FEB	2010 MAR	2013 MAR	2013 APR	2013 MAY	2014 JUN	2014 JUL	2014 AUG	2015 SEP	2015 OCT	2015 NOV
SKY IMAGES	ARMINES							(x)	(x)	(x)	(x)	(x)	(x)
SATELLITE	DLR-DFD	X	X	X	X	X	X	X	X	X	X	X	X
SATELLITE	DLR-IPA	X	X	X	X	X	X	X	X	X	X	X	X
SATELLITE	Meteotest	X	X	X	X	X	X	X	X	X	X	X	X
SATELLITE	SMHI					X							
NWPM	SMHI	X	X	X	X	X	X	X	X	X	X	X	X

2.2 Testing sites and time periods

The combination approaches have been tested for two different sites. Figure 1 shows a map including the testing locations: Plataforma Solar de Almería (PSA) placed at the southeast of Spain and Ghardaia (GHA) placed in the north of Algeria.

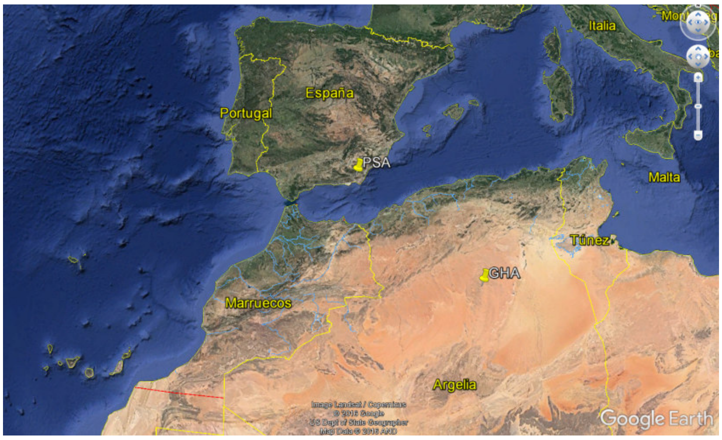

Figure 1: Plataforma Solar de Almería (PSA) and Ghardaia (GHA) location.

Each site has different nowcast availability. For instance, in Table 3, available nowcasts periods for the two selected locations from SATELLITE DLR-DFD are shown.

Table 3: Available periods at the selected locations from SATELITE DLR-DFD

Station	Lat. (°N)	Lon. (°E)	Altitude (m)	Full Name	Time period
PSA	37.0909	-2.3581	500	DLR_PSA	2010-Jan, Feb, Mar
					2013-Mar, Apr, May
					2014-Jun, Jul, Aug
					2015-Sep, Oct, Nov
GHA	32.386	3.78	463	Enermena_Ghardaia	2013-Mar, Apr, May
					2014-Jun, Jul, Aug

3 Uncertainty weighted approach (UWA) by DLR-IFS

3.1 Methodology

To calculate the optimal combination of the different provided nowcasts, the method of Meyer et al. (2008) is applied.

In a first step, nowcasted DNI in Wm^{-2} is extracted for each available data set:

DLR-DFD one-minute accumulated nowcasted irradiation given in Whm^{-2} is transformed first in Whm^{-2} and then translated in averaged Wm^{-2} for each instant time step *i* with:

$$DNI_{nowcast,i} = \frac{DNI_{nowcast,i} + DNI_{nowcast,i+1}}{2}$$

The 15-minutes accumulated nowcast of SMHI is given in kJm^{-2}. It is first converted in Wm^{-2} and then transformed with the same method so that an averaged nowcast for the time step *i* is available.

For each set of refresh time and forecast time for the combined nowcast data set, it is verified if a nowcasted DNI from the different data sets is available.

The combined DNI is calculated from all available nowcasted DNI_i from the data set *i* for each set of refresh time and forecast time:

$$DNI_{combined} = \left(\frac{1}{\sum_{i=1}^{n} \frac{1}{\Delta i}}\right) \cdot \left(\sum_{i=1}^{n} \frac{DNI_i}{\Delta i}\right)$$

Where $DNI_{combined}$ is the resulting combined DNI and DNI_i is the nowcasted DNI of the data set *i*, both given in Wm^{-2} for one time stamp. Δi is the absolute error for each data set *i* and this time stamp. Δi is provided by WP3.1-3.3 and WP4 as a function of the forecast horizon and the meteorological conditions.

The estimated absolute uncertainty of combined DNI $\Delta_{DNIcombined}$ is calculated with:

$$\Delta_{DNI_{combined}} = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial DNI_{combined}}{\partial DNI_i} \cdot \Delta i\right)^2}$$

given in Wm^{-2}.

Recycled SMHI nowcasts

The nowcasts of SMHI are only available every three hours in comparison to the other combined nowcasts which are refreshed every 15 minutes. It could be seen in the validation report presented in the DNICast Deliverable D.4.3 (Validation of nowcasted DNI methods) as well as in an internal validation analysis, that the MAE as well as RMSE are rather constant for lead times from 0 up to 240 minutes. Therefore, older nowcasts of SMHI for a certain nowcasted time stamp are considered as well for the DNI combination in between the 3 hours refresh time stamps of the SMHI data set.

Table 4: Combined data sets by DLR-IFS

Work package	Project partner	Method	Data sets	Refresh rate [min]	Forecast resolution [min]	Forecast horizon [h]
3.2	DLR-IPA	Satellite based	instant value	15	5	6
3.2	Meteotest	Satellite based	clim post-processed, instant value climdirect post-processed, instant value maccdirect post-processed, instant value	15	5	4
3.2	DLR-DFD	Satellite based	accumulated	15	1	8
3.3	SMHI	NWP based	- 15-min accumulated single grid point - 15-min accumulated average of 5x5 grid points around the station	180	15	9
-	DLR-ISF	Persistence	instant value	15	1	6

3.2 Uncertainty treatment by DLR-IFS

The data sets of DLR-IPA, Meteotest and SMHI contain uncertainty estimates. For DLR-DFD and the persistence nowcast, absolute uncertainties have been derived as described below. If no absolute uncertainty is given for one set of refresh time and forecast time and one data set, a default uncertainty is estimated dependent on the data set. All absolute uncertainties are displayed in Figure 2 and **¡Error! No se**

encuentra el origen de la referencia.. Absolute uncertainties within a 10 Wm^{-2} range are displayed here in one bar.

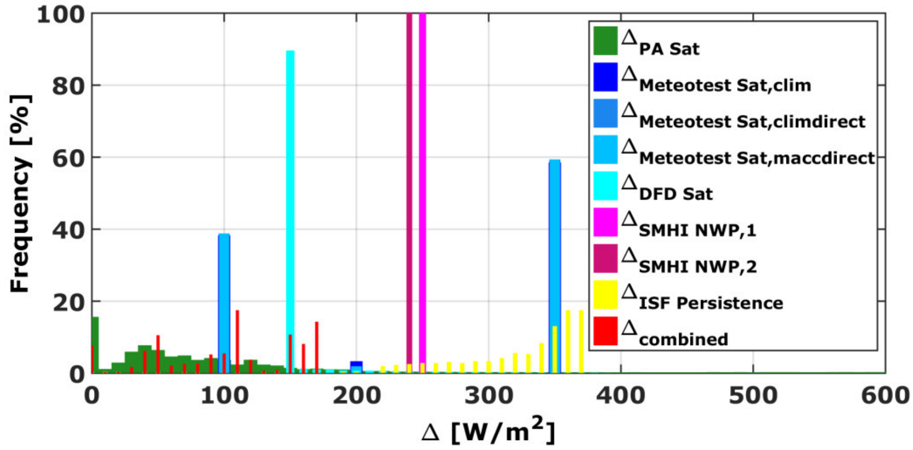

Figure 2: Frequency distribution of absolute uncertainties for all combined data sets and PSA and all available data points of each individual data set.

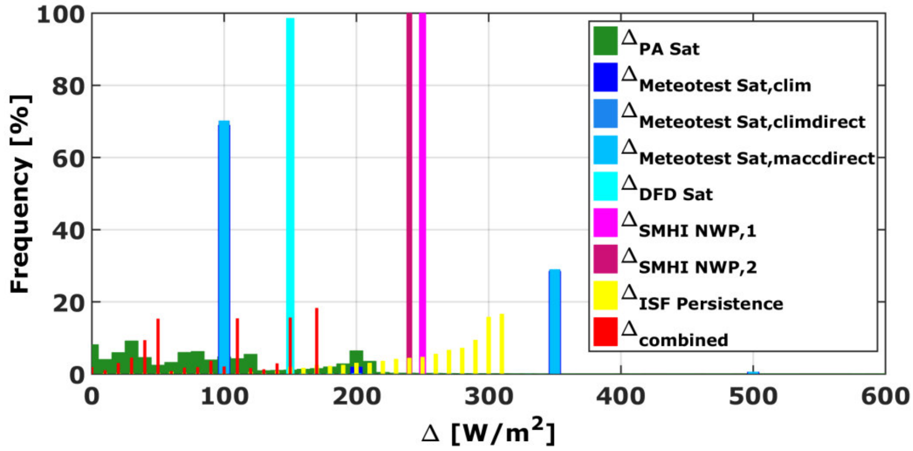

Figure 3: Frequency distribution of absolute uncertainties for all combined data sets and GHA and all available data points of each individual data set.

3.2.1 DLR-PA:

The total uncertainty of the DLR-PA nowcast is the sum of the uncertainty of the optical flow method and the calculation of the DNI.

To derive the uncertainty of the optical flow method which is used to derive the cloud motion vectors, the DNI values of the surrounding pixels of the station are considered. The uncertainty of the calculation of DNI is determined by using daily minimum and maximum AOD values (aerosol optical depth) in case of the PSA and standard deviations of the monthly mean AOD for the other stations.

The absolute uncertainty is given in Wm^{-2}. The frequency for the analyzed time periods for PSA and GHA can be seen in Figure 2 and **¡Error! No se encuentra el origen de la referencia.**.

200 Wm^{-2} is used as a default absolute uncertainty for time steps for which no uncertainty value is available in the input data set.

3.2.2 Meteotest:

The given absolute uncertainty for the Meteotest data sets is a standard deviation which is based on an expert guess (100, 200, 350 and 500 Wm^{-2}) based on the clearness index in the surrounding area (see Figure 2 and **¡Error! No se encuentra el origen de la referencia.**).

A default uncertainty of 200 Wm^{-2} is set for time steps for which no uncertainty value is given in the data set.

3.2.3 SMHI:

The SMHI data sets include a constant RMSE (root mean square error) for the two data sets DNI_1 (15-min accumulated single grid point DNI) and DNI_2 (15-min accumulated average of 5x5 grid points around the station). The RMSE values have been derived from the analysis of the site of PSA and January 2010. For DNI_1, a constant RMSE of 251.11 Wm^{-2} is assumed and for DNI_2 an RMSE of 242.22 Wm^{-2} is given. It has to be kept in mind that the analysis of SMHI showed that the RMSE are in fact larger in the mornings and smaller in the afternoons.

3.2.4 DLR-DFD:

DLR-DFD provides, besides the nowcasted DNI values (the P50 value), also the P10 and P90 percentiles as well as the minimum (DNI_{min}) and maximum values (DNI_{max}) of the nowcasted DNI.

For cloud-free situations, all given values are the same, this additional information can therefore only be used for possibly cloudy situations.

In the investigated time period, cloud-free situations have been nowcasted in about 46% of all examined time steps for PSA. In about 28%, no DNI could be nowcasted due to low solar elevation angles. For the remaining time steps (~26%), the minimum and maximum nowcasted DNI has been used to derive the additional absolute uncertainty due to the nowcasting of cloudy situations. The complete absolute uncertainty of the DLR-DFD data sets is then calculated according to:

$$\Delta_{DLR-DFD} = \sqrt{(150\mathrm{Wm}^{-2})^2 + \left(\frac{DNI_{min} - DNI_{max}}{2}\right)^2}$$

As a baseline absolute uncertainty 150 Wm^{-2} was chosen.

For cloud-free situations, a default absolute uncertainty is consequently 150 Wm^{-2}. It can be seen in Figure 2 and Figure 3 that the added contribution to the absolute uncertainty for nowcasted cloudy situations has only a small impact on the overall distribution of the DLR-DFD uncertainty used for the combination. As more cloudy situations are nowcasted for PSA in comparison to GHA, this method has a larger effect for PSA.

3.2.5 Persistence:

The absolute uncertainty of the persistence nowcast has been derived with the help of the standard deviation of 12 months between 2010 and 2015 for PSA (9 whole months of 2013-2015 for GHA, respectively) in comparison to on-site DNI measurements of 1 minute resolution (10 minutes for GHA, respectively). The measurements of GHA have been linearly interpolated first to 1 minute resolution and the persistence nowcast has been performed using these data with a refresh time of 15 minutes and a temporal resolution of 1 minute.

The absolute uncertainty is estimated dependent on the lead time and is displayed in Figure 4.

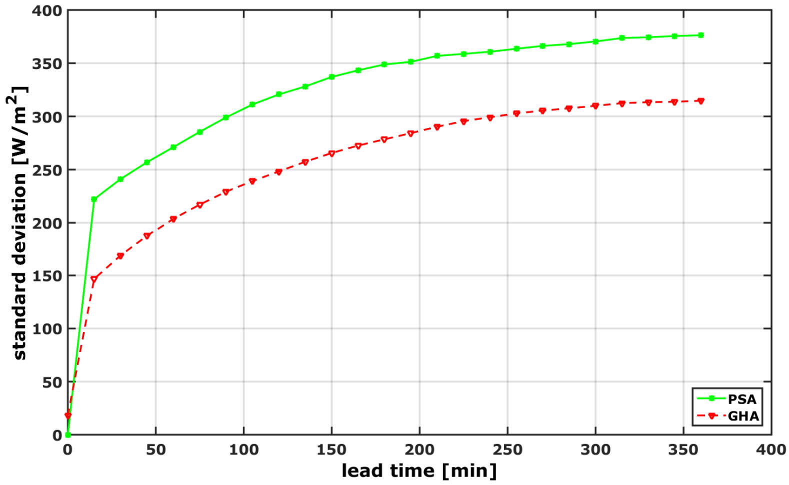

Figure 4: Absolute uncertainty for the persistence nowcasts for PSA and GHA.

4 Multi-regressive approach (MRA) by CIEMAT

This CIEMAT's combination uses a time-dependent multi-regressive model. It is inspired in the forecasting optimization of precipitation (Turlapaty et al., 2012). For the merging purpose, an adaptive linear merging model is presented. In our scheme, the explanatory variables are those DNI values predicted in previous forecast events. Thus, each DNICast nowcasting output provides a number of variables depending on the forecasted horizon, the refresh time and the time step. The combined model result N is calculated with the following equation using the total number of input variables:

$$N = \sum_i n_i$$

$$n_i = \sum_i FO_i \cdot d_i$$

$$d_i = \frac{Hm_i}{Rm_i}$$

Table 5 shows the definition of the used symbols and details to evaluate the number of inputs variables from each supplier (n_i).

Table 5: Nowcast input variables at CIEMAT methodology

Method	Partner	Refresh interval (min)	Horizon (min)	Data by time	Forecasted outputs	Inputs for comb.
		$\boldsymbol{Rm_i}$	$\boldsymbol{Hm_i}$	$\boldsymbol{d_i}$	$\boldsymbol{FO_i}$	$\boldsymbol{n_i}$
CAMARAS	ARMINES	1	30	30	1	30
SATELLITE	DLR-DFD	15	480	32	1	32
SATELLITE	DLR-IPA	15	360	24	1	24
SATELLITE	Meteotest	15	240	16	3	48
SATELLITE	SMHI	180	540	3	2	6
NWPM	SMHI	180	540	3	2	6

In order to determine the significant variables among the total inputs for the combination, a stepwise methodology is applied. Stepwise regression is a systematic method for adding and removing terms from a multilinear model based on their statistical significance in a regression.

Some tests have been done training general models for long periods such as the three months available by each year and it has been detected that the significant variables change when the training period changes. This is justified because stepwise models are locally optimal, but may not be globally optimal. Better results are obtained when using dynamic fitting. In addition, dynamic fitting aims to include new input when an additional supplier is available. For the dynamic fitting a moving window of two days has been established as this is the time needed for an impact of the input changes in the model.

In addition to the combination, a post process treatment is necessary. Physical limits are applied to the combined output. Due to the time-dependent multi-regressive model methodology increases the forecasted DNI variability, the treatment attempts to adjust this variability.

5 Approach based on Euclidean distance (DWA) by CIEMAT

This methodology make use of the previous measures for a local adaptation like in (Anadranistakis et al., 2002). In our model, we have designed a two steps methodology.

In the first step empirical weights are calculated. Two inputs are taken into account:

- the Euclidean distance between each forecasted series and measurements of the previous day at the same lead time,
- the coefficient of variation difference between those series.

$$Ed_n = \sqrt{\sum_{t=1}^{lead\ time} (DNI_n^t - DNI_m^t)^2}$$

$$CoVd_n = \frac{\overline{DNI_n}}{\sigma_n} - \frac{\overline{DNI_m}}{\sigma_m}$$

Ed_n being the Euclidean distance and $CoVd_n$ the distance between the coefficients of variation of each nowcast (n) and the corresponding measures (m) for each forecasted series relatively to the previous day. Weights for each nowcast are calculated as follows, where w'^{ed}_n are the normalized weights to the sum of $w_n^{Ed}$and w_n^{Cov}:

$$w_n^{Ed} = \frac{1/_{Ed_n^2}}{\sum_n \left(1/_{Ed_n^2}\right)}$$

$$w_n^{Cov} = \frac{1/_{CoV_n^2}}{\sum_n \left(1/_{CoV_n^2}\right)}$$

$$DNI_c^t = \sum_{n=1}^{n} w'^{ed}_n DNI_n^t + w'^{Cov}_n DNI_n^t$$

In a second step, combining forecasts for clear sky is fitted to the persistence TL derived to previous measurements.

6 Summary of the combination approaches

Three combination approaches have been applied. The first approach called UWA (uncertainty weighted approach) makes use of the uncertainty of each input nowcasting data set. To calculate the optimal combination of the different provided nowcasts, the method of Meyer et al. (2008) is applied to derive a combined product, considering the individual uncertainties of the nowcast products. For each set of refresh time and forecast time of the combined nowcast data set, the available different data sets are included in the combination.

The second combination method is a multi-regressive approach (MRA) that uses a time-dependent multi-regressive model. It is inspired in the forecasting optimization of precipitation (Turlapaty et al., 2012). For the merging purpose, an adaptive linear merging model is presented. In this scheme, the explanatory variables are those DNI values predicted in previous forecast events. Thus, each DNICast nowcasting output provides a number of variables depending on the forecasted horizon, the refresh time and the time step.

In the third approach, an empirical combination inspired by Anadranistakis et al. (2002) is applied. The so called distance weighted based approach (DWA) makes use of the distances to the previous measurements to the derivation of combining weights. This methodology is similar to the UWA but weights are calculated empirically. The combined forecast is calculated in two steps. In the first step, weights are calculated using two inputs: the Euclidean distance but also the comparisons of the coefficient of variation (CoV) with the same series relative to the previous day. This use of the inputs variability is an innovative view trying to use the input variability in the combined nowcast. In a second step, combining forecasts for clear sky is fitted to the persistence TL derived from previous measurements.

Results of these three combination forecasts are discussed in deliverable 3.13.

7 References

Anadranistakis, M., Lagouvardos, K., Kotroni, V., Skouras, K., 2002. Combination of Kalman filter and an empirical method for the correction of near-surface temperature forecasts: Application over Greece. Geophys. Res. Lett. 29. doi:Artn 1776\nDoi 10.1029/2002gl014773

Armstrong, J.S., 2001. Combining forecasts. Retrieved from http//repository.upenn.edu/marketing_papers/34 Postprint.

Chase, C.W.J., 1995. Measuring forecast accuracy. J. Bus. Forecast. Methods Syst. 14, 9.

Clemen, R.T., 1989. Combining forecasts: A review and annotated. Int. J. Forecast. 5, 559–583. doi:https://doi.org/10.1016/0169-2070(89)90012-5

Cooper, J.P., Nelson, C.R., 1975. The Ex Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models and Some Results on Composite Predictors. J. Money. Credit. Bank. 7, 1–32.

Dickinson, P., 1977. Some Comments on the Combination of Forecasts. Oper. Res. Q. 26, 205–206.

Fiordaliso, A., 1998. A nonlinear forecasts combination method based on Takagi–Sugeno fuzzy systems. Int. J. Forecast. 14, 367–379. doi:10.1016/S0169-2070(98)00010-7

Gupta, R., Venugopal, V., Foufoula-Georgiou, E., 2006. A methodology for merging multisensor precipitation estimates based on expectation-maximization and scale-recursive estimation. J. Geophys. Res. Atmos. 111, 1–14. doi:10.1029/2004JD005568

Lauret, P., Rodler, A., Muselli, M., David, M., Diagne, H.M., Voyant, C., Id, H., 2012. A Bayesian model committee approach to forecasting global solar radiation, in: World Renewable Energy Forum. Denver, Colorado, p. 7.

Meyer, R., Geuder, N., Lorenz, E., Hammer, A., Ossietzky, C. Von, Oldenburg, U., Beyer, H.G., 2008. Combining solar irradiance measurements and various satellite-derived products to a site-specific best estimate, in: SolarPACES Conference. pp. 1–8.

Shrestha, R., Houser, P.R., Anantharaj, V.G., 2011. An optimal merging technique for high-resolution precipitation products. J. Adv. Model. Earth Syst. 3, M12003. doi:10.1029/2011MS000062

Turlapaty, A.C., Younan, N.H., Anantharaj, V.G., 2012. A linear merging methodology for high-resolution precipitation products using spatiotemporal regression. Int. J. Remote Sens. 33, 7844–7867. doi:10.1080/01431161.2012.703345