Adv. Sci. Res., 6, 1–5, 2011
www.adv-sci-res.net/6/1/2011/
doi:10.5194/asr-6-1-2011

Advances in Science & Research
Open Access Proceedings

10th EMS Annual Meeting and 8th European Conference on Applied Climatology (ECAC) 2010

Geostatistical merging of ground-based and satellite-derived data of surface solar radiation

M. Journée and C. Bertrand

Royal Meteorological Institute of Belgium, Brussels, Belgium

Received: 12 October 2010 – Revised: 13 December 2010 – Accepted: 20 December 2010 – Published: 10 January 2011

Abstract. In this paper, we demonstrate the benefit of using observations from Meteosat Second Generation (MSG) satellites in addition to in-situ measurements to improve the spatial resolution of solar radiation data over Belgium. This objective has been reached thanks to geostatistical methods able to merge heterogeneous data types. Two geostatistical merging methods are evaluated against the interpolation of ground-data only and the single use of satellite-derived information. It results from our analysis that merging both data sources provides the most accurate mapping of surface solar radiation over Belgium.

1 Introduction

Knowledge of the local solar radiation is essential for many applications, including design, planning and operation of solar energy systems, architectural design, crop growth models and evapotranspiration estimates. Traditionally, solar radiation is observed by means of networks of meteorological stations. Costs for installation and maintenance of such networks are very high and national networks comprise only few stations. Consequently the availability of observed solar radiation measurements has proven to be spatially inadequate for many applications. Mapping solar radiation by interpolation/extrapolation of measurements is possible but usually leads to large errors, except for dense networks (Zelenka et al., 1992; Hay, 1981, 1984; Hay and Hanson, 1985; WMO, 1981; Perez et al., 1997).

Because several authors have shown the potentialities of the images of the Earth taken by polar-orbiting and geostationary satellites for mapping the global irradiation impinging on a horizontal surface at the ground level (e.g., Zelenka et al., 1992, 1999; Perez et al., 1997, 2002; Pinker et al., 1995), we evaluate in the present paper the benefit of using space-based observations as an additional information source when interpolating the ground measurements. More specifically, we consider surface incoming global short-wave radiation products derived from Meteosat Second Generation (MSG, Schmetz et al., 2002) in order to improve the spatial resolution of daily surface solar radiation data over Belgium.

Correspondence to: M. Journée
(michel.journee@oma.be)

To reach that objective, we implemented two geostatistical methods able to merge heterogeneous data types (i.e., kriging with external drift and regression kriging) and evaluate these methods against mappings derived from a single source of data (i.e., either in-situ or satellite data).

2 Solar radiation data

2.1 Ground-based solar radiation measurements

The Royal Meteorological Institute of Belgium (RMIB) is currently performing measurements of global solar irradiance (in Wm^{-2}) by means of CNR1 and CM11 pyranometers of Kipp & Zonen at 13 sites well-distributed over Belgium (see Fig. 3). Measurements are made with a 5 s time step and time-integrated on a 10 min basis in the RMIB data warehouse. The 10-min solar irradiation data (in Whm^{-2}) are then subject to a set of semi-automatic quality assessment tests and gaps in the time series are filled by model estimations (Journée and Bertrand, 2011).

2.2 MSG-derived surface solar irradiance

Within the Satellite Application Facility (SAF) network, the down-welling short-wave irradiance at the Earth's surface is operationally retrieved from MSG imageries by three decentralized SAFs: the Ocean and Sea Ice SAF (OSI-SAF, www.osi-saf.org), the Land Surface Analysis SAF (LSA-SAF, land-saf.meteo.pt) and the SAF on Climate Monitoring (CM-SAF, www.cmsaf.eu). To retrieve the same parameter, the different SAFs use their own algorithms and different ancillary input data.

Published by Copernicus Publications.

Sea surface being out of the scope of this study, we focused our investigation on the LSA-SAF and CM-SAF products. The LSA-SAF surface solar radiation product (Geiger et al., 2008) is generated every 30 min and distributed to the users in near real-time at the pixel spatial resolution of the MSG spectral imager (i.e., about 6 km in NS direction and 3.3 km in EW direction over Belgium). The operational CM-SAF surface solar radiation product (Mueller et al., 2009) is an off-line product provided on a 15×15 km sinusoidal grid in daily and monthly average. The monthly mean diurnal cycle is also provided. Because of the relatively coarse spatial and temporal resolution of the CM-SAF operational product, intermediate CM-SAF values (R. Mueller, personal communication, 2010) were considered in the present study, namely instantaneous hourly CM-SAF solar surface irradiance remapped onto 3×3 km, 9×9 km and 15×15 km grids.

3 Geostatistical mapping methods

In this study, we considered three geostatistical mapping methods: ordinary kriging (OK), kriging with external drift (KED) and regression kriging (RK). The aim of these methods is to interpolate a random field, G, (e.g., the spatial distribution of surface solar radiation) from observations, $G(x_i)$, at selected locations $x_i|_{i=1,...,N}$ (e.g., the ground measurements). Based on the knowledge of the spatial dependence of the random field, the kriging estimation $\hat{G}(x_0)$ at an unobserved location x_0 is computed as a linear combination of the observations, $\hat{G}(x_0)=\sum_{i=1}^{N} w_i G(x_i)$, where the weights, w_i, are chosen in such a way that the estimator is unbiased and the error variance is minimized. The spatial correlation of the random field between two locations x_i and x_j is described by means of the variogram $\gamma(x_i,x_j)=\mathbb{E}[(G(x_i)-G(x_j))^2]$, which is often chosen as isotropic, meaning that it is function only of the distance d between x_i and x_j, i.e., $\gamma(x_i,x_j)=\gamma(d)$. When the number of observation locations is sufficient, a model can be fitted to the empirical variogram derived from the observed data. The variogram has otherwise to be assumed.

While in OK the random field G is assumed to have a constant, albeit unknown, mean at all locations, the KED and RK techniques rely on the knowledge of a densely sampled auxiliary variable g (e.g., the SAFs' products) to model G as a non-stationary random field of the form $\mathbb{E}[G(x)] = a_0 + a_1 g(x)$. Although the KED and RK methods are aimed toward a similar objective, they differ in the way to compute the parameters a_i and the weights w_i (Hengl et al., 2003). In KED, a_i and w_i are derived together by forcing an exact interpolation of the auxiliary variable, $g(x_0)=\sum_{i=1}^{N} w_i g(x_i)$. In RK, the parameters a_i are first computed by linear regression from data at the observed locations. The regression residuals are then interpolated by OK. From a computational point of view, all three kriging methods require to solve linear systems of equations. We refer to Wackernagel (1995), Hengl et al. (2003) and references therein for more details on these techniques.

4 Cross validation analysis

This study is focused on the mapping of surface solar radiation data over Belgium on a daily basis. Daily totals of the 10-min RMIB ground data are obtained by simple summation, while the instantaneous satellite data are integrated by trapezoidal integration. The considered interpolation and merging methods are evaluated by leave-one-out cross-validation (CV) on the basis of two years of quality-controlled data (2008 and 2009). In total, we used a set of 491 days for which the ground data and both SAFs' satellite data were available at all stations and over the entire diurnal cycle. The performance of the different methods is assessed by the average on these 491 instances of three indices derived from the bias between the cross-validation prediction $\hat{G}$ and the actual measurement G at the N locations $x_i|_{i=1,...,N}$:

– the cross-validation mean bias error $\mathrm{MBE_{cv}} = \frac{1}{G_{\mathrm{avg}}N}\sum_{i=1}^{N}(\hat{G}(x_i)-G(x_i))$,

– the cross-validation mean absolute error $\mathrm{MAE_{cv}} = \frac{1}{G_{\mathrm{avg}}N}\sum_{i=1}^{N}|\hat{G}(x_i)-G(x_i)|$ and,

– the cross-validation root mean square error $\mathrm{RMSE_{cv}} = \frac{1}{G_{\mathrm{avg}}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{G}(x_i)-G(x_i))^2}$.

where $G_{\mathrm{avg}} = \frac{1}{N}\sum_{i=1}^{N} G(x_i)$ is the average solar radiation over all stations.

Since surface solar radiation is measured at only 13 stations, variograms can hardly be estimated from the ground data. Hence, we chose a fixed variogram model, e.g., an exponential variogram $\gamma(d)=1-\exp(-d/\bar{d})$. The evolution of $\mathrm{RMSE_{cv}}$ as a function of the range parameter $\bar{d}$ indicates that the best performance is reached when $\bar{d}=500$ km for OK and $\bar{d}=50$ km for KED and RK (see Fig. 1). This difference in optimal values of $\bar{d}$ results from the high spatial resolution of the satellite data used by KED and RK. Even if the variogram is expected to vary with the sky conditions, we used these values of $\bar{d}$ for all the 491 days.

Because of the possible non-uniformity of the surface global radiation within the MSG pixel, the SAFs' satellite products have been used at various spatial resolutions as auxiliary information for the KED and RK methods (i.e., from 1 pixel to 3×3 pixels aggregates for the LSA-SAF product and from 3×3 km to 15×15 km areas for the CM-SAF product). The geostatistical interpolation of ground data and the geostatistical merging of ground data with satellite information are compared against the SAFs' estimations in Table 1. Since the $\mathrm{MBE_{cv}}$ error is in overall very small for all methods, performance comparison relies essentially on the $\mathrm{MAE_{cv}}$ and $\mathrm{RMSE_{cv}}$ indices. First, the largest $\mathrm{RMSE_{cv}}$ and $\mathrm{MAE_{cv}}$ are found for OK of the ground measurements and

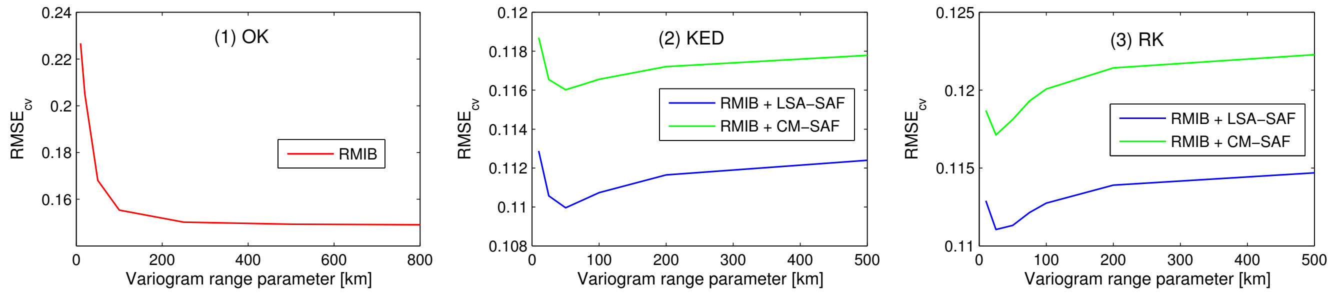

Figure 1. Distribution of the cross-validation root mean square error (RMSE$_{cv}$) as a function of the variogram range parameter $\bar{d}$ for the three kriging methods (left panel: OK; center panel: KED; right panel: RK). The SAFs' products are used at the finest spatial resolution as auxiliary information for KED and RK (i.e., MSG pixel resolution for the LSA-SAF and 3×3 km resolution for the CM-SAF).

Table 1. Cross validation mean bias error (MBE$_{cv}$), mean absolute error (MAE$_{cv}$) and root mean square error (RMSE$_{cv}$) for the geostatistical interpolation by OK of the RMIB ground data, the SAFs' products used at various spatial resolutions, and the geostatistical merging by KED and RK of the RMIB ground data with the SAFs' products.

Method & Data		MBE$_{cv}$	MAE$_{cv}$	RMSE$_{cv}$
OK	RMIB	−0.0032	0.116	0.149
–	LSA-SAF (1×1 px)	−0.0004	0.120	0.144
–	LSA-SAF (2×2 px)	0.0008	0.120	0.144
–	LSA-SAF (3×3 px)	0.0006	0.120	0.140
–	CM-SAF (3×3 km)	−0.0078	0.112	0.136
–	CM-SAF (9×9 km)	−0.0310	0.112	0.135
–	CM-SAF (15×15 km)	−0.0321	0.110	0.133
KED	RMIB+LSA-SAF (1×1 px)	0.0006	0.087	0.110
KED	RMIB+LSA-SAF (2×2 px)	0.0001	0.087	0.111
KED	RMIB+LSA-SAF (3×3 px)	0.0006	0.087	0.105
KED	RMIB+CM-SAF (3×3 km)	−0.0005	0.092	0.116
KED	RMIB+CM-SAF (9×9 km)	−0.0001	0.089	0.112
KED	RMIB+CM-SAF (15×15 km)	0.0003	0.088	0.111
RK	RMIB+LSA-SAF (1×1 px)	0.0007	0.087	0.111
RK	RMIB+LSA-SAF (2×2 px)	0.0002	0.087	0.112
RK	RMIB+LSA-SAF (3×3 px)	0.0007	0.087	0.106
RK	RMIB+CM-SAF (3×3 km)	−0.0003	0.094	0.118
RK	RMIB+CM-SAF (9×9 km)	0.0002	0.091	0.114
RK	RMIB+CM-SAF (15×15 km)	0.0003	0.088	0.111

the unmerged LSA-SAF mappings, respectively, while the best performance is observed once ground and satellite data are merged together. The KED and RK merging methods exhibit virtually identical results. Using both ground-based and satellite-derived information provides a significant improvement with respect to mappings based solely on one of these data sources (e.g., in case of the LSA-SAF product, the RMSE$_{cv}$ is reduced by about 25% and 30% with respect to the satellite and the OK mappings, respectively). Regarding the satellite information used as auxiliary information for KED and RK, the both SAF products provide comparable results although the best scores are obtained by KED with the LSA-SAF data. Finally, the spatial resolution of the satellite products appears to have little impact on the resulting performance, albeit that slight improvements are obtained with larger resolutions.

In Fig. 2, we investigate the impact of sky conditions on the mapping performance. Sky-type classification is made upon both the mean and standard deviation over all stations of the daily clearness index (i.e., the ratio of the daily totals of surface and top-of-the-atmosphere incoming solar radiation). First, as far as the average sky condition over Belgium is concerned (see Fig. 2, left panel), the OK interpolation of ground data outperforms the single use of SAFs' data for overcast and very clear skies. In overcast conditions, it is well-known that the SAF's data overestimate the surface incoming solar radiation, while the exact mechanism that causes this overestimation is still unclear (Ineichen et al., 2009; Journée and Bertrand, 2010). The geostatistical merging of ground and satellite data exhibits the best performance for all types of sky. The improvement is however less pronounced for very clear skies. Second, regarding the influence of the spatial variability in sky conditions, the benefit of using the SAF's products is the largest for sky conditions that are highly inhomogeneous over the country (see Fig. 2, right panel).

5 Maps of surface solar radiation

Figure 3 compares the spatial distribution over Belgium of the average daily clearness index as computed by the OK interpolation of ground data, by the single use of LSA-SAF data, and by the KED merging of ground and LSA-SAF data. The daily clearness index at a specific location is inferred by means of the geostatistical interpolation and merging methods or directly from the satellite data for each of the 491 days selected in this study. Averages on these 491 instances are then computed.

All maps clearly highlight the global south-east to north-west positive gradient in clearness index. Satellite-based information is however needed to capture more regional features. The values derived from the LSA-SAF data only

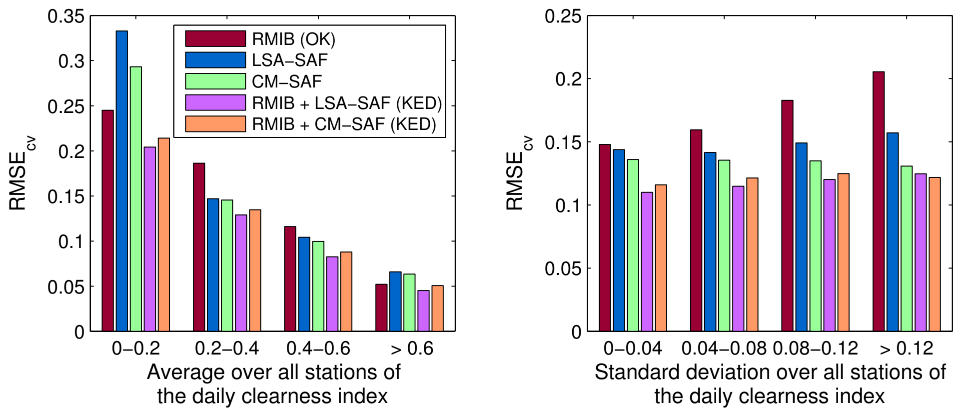

Figure 2. Distribution of the cross-validation root mean square error RMSE$_{cv}$ as a function of sky conditions. Sky-type classification is made upon the average (left-hand plot) as well as the standard deviation (right-hand plot) over all stations of the daily clearness index. The SAFs' products are used at the finest spatial resolution (i.e., MSG pixel resolution for the LSA-SAF and 3×3 km resolution for the CM-SAF). The results obtained with RK are not represented since they are virtually identical to those of KED.

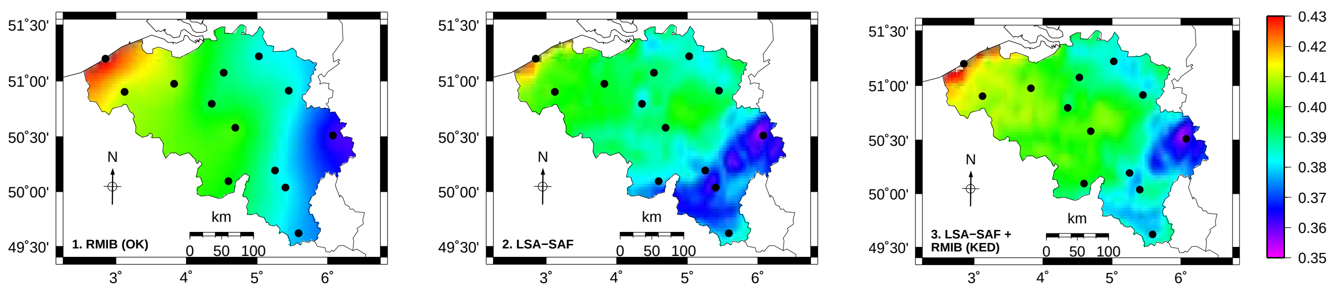

Figure 3. Spatial distribution over Belgium of the average daily clearness index as computed **(1)** by the OK interpolation of ground data, **(2)** by the single use of satellite-based estimations, and **(3)** by the KED merging of ground and satellite data. The displayed values of the daily clearness index are averages on the 491 days selected in this study. Dots indicate the ground stations' locations.

(Fig. 3, middle panel) are at most stations slightly below those derived from ground measurements only (Fig. 3, left panel). Hence, merging the two data sources (Fig. 3, right panel) enables to take advantage of both the accuracy of ground measurements and the global spatial coverage of satellite observations.

6 Conclusions

In this paper, we demonstrated the benefit of using MSG satellites imageries in addition to in-situ measurements to improve the spatial resolution of solar radiation data over Belgium. Regarding the MSG-derived information, we considered two products delivered by the LSA-SAF and the CM-SAF, respectively. The variograms used in the implemented geostatistical interpolation/merging methods were estimated to provide the best cross-validation scores over 13 sites in Belgium. Differences by a factor of 10 were observed for the optimal variogram range parameter, depending on whether the high spatial resolution satellite data are involved or not in the interpolation process.

The best performance has been observed once ground observations are merged with the LSA-SAF product. No significant difference has been noted between the two implemented geostatistical merging methods (kriging with external drift and regression kriging). This reflects the fact that both methods are conceptually very close. Concerning the impact of sky conditions, mappings inferred by merging ground and SAFs' data were systematically more accurate than when using each data source separately, whatever the sky-type, while the benefit of the method is less apparent in case of very clear skies. Finally, the benefit of using satellite information appeared to be the largest in case of highly variable sky conditions over the studied area. In overall, merging ground and satellite data enables to take advantage of both the high accuracy of ground data and the global spatial coverage of satellite information.

Acknowledgements. The authors are grateful to Richard Mueller (German Meteorological Service, Germany) for providing us with 2 years of hourly CM-SAF solar surface irradiance remapped onto regular latitude-longitude grids at various spatial resolutions (3×3 km, 9×9 km and 15×15 km). This study was supported by the Belgian Science Policy under the research project: Contribution de l'IRM au développement de l'énergie renouvelable en Belgique.

Edited by: M. Dolinar
Reviewed by: J. Walawender and another anonymous referee

sc | nat The publication of this article is sponsored by the Swiss Academy of Sciences.

References

Geiger, B., Meurey, C., Lajas, D., Franchistéguy, L., Carrer, D., and Roujean, J.-L.: Near real-time provision of downwelling shortwave radiation estimates derived from satellite observations, Meteorol. Appl., 15, 411–420, 2008.

Hay, J. E.: The mesoscale distribution of solar radiation at the Earth's surface and the ability of satellites to resolve it. In Proceedings of the First Workshop on Terrestrial Solar resource Forecasting and on Use of Satellites for Terrestrial Solar Resource Assessment, Washington D.C., 2–5 February, 1981.

Hay, J. E.: An assessment of the mesoscale variability of solar radiation at the Earth's surface, Sol. Energy, 32, 425–434, 1984.

Hay, J. E. and Hanson, K. J.: Evaluating the solar resource: a review of problems resulting from temporal, spatial and angular variations, Sol. Energy, 34, 151–161, 1985.

Hengl, T., Geuvelink, G. B. M., and Stein, A.: Comparison of kriging with external drift and regression-kriging, Technical note, ITC, Enschede, Netherlands, 2003.

Ineichen, P., Barroso, C. S., Geiger, B., Hollmann, R., Marsouine, A., and Mueller, R.: Satellite Application Facilities irradiance products: hourly time step comparison and validation over Europe, Int. J. Remote Sens., 30, 5549–5571, 2009.

Journée, M. and Bertrand, C.: Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measurements, Remote Sens. Environ., 114, 2692–2704, 2010.

Journée, M. and Bertrand, C.: Quality control of solar radiation data within the RMIB solar measurements network, Sol. Energy, 85, 72–86, 2011.

Mueller, R. W., Matsoukas, C., Gratzki, A., Behr, H. D., and Hollmann, R.: The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance – A LUT based eigenvector hybrid approach, Remote Sens. Environ., 113, 1012–1024, 2009.

Perez, R., Seals, R., and Zelenka, A.: Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data, Sol. Energy, 60, 89–96, 1997.

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., and Vignola, F.: A new operational model for satellite-derived irradiances: Description and Validation, Sol. Energy, 73(5), 307–317, 2002.

Pinker, R. T., Frouin, R., and Li, Z.: A review of satellite methods to derive surface shortwave irradiance, Remote Sens. Environ., 51, 105–124, 1995.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: An introduction to Meteosat Second Generation (MSG), B. Am. Meteorol. Soc., 83, 977–992, 2002.

Wackernagel, H.: Multivariate geostatistics: an introduction with applications, Springer-Verlag, Berlin, 1995.

World Meteorological Organization (WMO): Meteorological aspects of the utilization of solar radiation as an energy source, Annex: World maps of relative global radiation, Technical Note No. 172, WMO-No. 557, Geneva, Switzerland, 298 pp., 1981.

Zelenka, A., Czeplak, G., d'Agostino, V., Josefson, W., Maxwell, E., and Perez, R.: Techniques for supplementing solar radiation network data. Technical Report, International Energy Agency, IEA-SHCP-9D-1, Swiss Meteorological Institute, Krahbuhlstrasse, 58, CH-8044 Zürich, Switzerland, 1992.

Zelenka, A., Perez, R., Seals, R., and Renné, D.: Effective accuracy of satellite-derived hourly irradiances. Theor. Appl. Climatol., 62, 199-207, 1999.