Fusing ground measurements and satellite-derived products for the construction of climatological maps in atmosphere optics

Mireille Lefèvre, Jan Remund, Michel Albuission, Thierry Ranchin, Lucien Wald

To cite this version:
Mireille Lefèvre, Jan Remund, Michel Albuission, Thierry Ranchin, Lucien Wald. Fusing ground measurements and satellite-derived products for the construction of climatological maps in atmosphere optics. 23rd EARSeL Annual Symposium ”Remote Sensing in Transition”, Jun 2003, Ghent, Belgium. pp.85-91. hal-00395054

HAL Id: hal-00395054
https://hal.archives-ouvertes.fr/hal-00395054
Submitted on 14 Jun 2009

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.
Fusing ground measurements and satellite-derived products for the construction of climatological maps in atmosphere optics

M. Lefèvre
Groupe Télédétection et Modélisation, Ecole des Mines de Paris, B.P. 207 - 06904 Sophia Antipolis cedex, France

J. Remund
METEOTEST, Fabrikstrasse 14, 3012 Bern, Switzerland

M. Albuisson, T. Ranchin & L. Wald
Groupe Télédétection et Modélisation, Ecole des Mines de Paris, B.P. 207 - 06904 Sophia Antipolis cedex, France

Keywords: data fusion, solar radiation, Linke turbidity factor, remote sensing, resampling

ABSTRACT: Climatological maps (gridded data) of optical parameters of the atmosphere often result from application of numerical models or processing of satellite images. Such maps usually exhibit very large cell sizes, of the order of 1° of arc angle. There is a need to increase their spatial resolution to obtain a cell size closer to the spatial representation obtained by standard meteorological instruments at ground level, i.e. 5' of arc angle. It then permits to fuse ground measurements and gridded data, especially to correct for bias observed in gridded data. Taking advantage of the availability of other data sets of relevance to the parameter under concern, though different, a method is proposed for the synthesis of the initial gridded data at a higher resolution by the means of a fusion process. This paper describes the method with an application for the construction of worldwide maps of the Linke turbidity factor with cells of 5' in size. This factor characterises the atmospheric optical turbidity under clear skies and is a very important parameter in solar radiation studies.

1 INTRODUCTION

Gridded data in environment often result from application of numerical models (e.g., weather models) or processing of satellite images. Examples are climatological maps of geophysical parameters, e.g., air temperature or ozone content. Such maps usually exhibit very large cell sizes, of order of 1° of arc angle. There is a need to increase their spatial resolution for further processing. One example is the need to obtain a cell size close to the spatial representation obtained by standard meteorological instruments at ground level, i.e. 5' of arc angle. It then permits to fuse ground measurements and gridded data, especially to correct for bias observed in gridded data (Beyer et al. 1997; D'Agostino & Zelenka 1992).

The problem is that such climatological maps only exist at low spatial resolution. A resampling of the maps at a finer cell size does not actually increases the spatial resolution of the data since no additional knowledge is brought. Taking advantage of the availability of other data sets of relevance to the parameter under concern, though different, a method is proposed for the synthesis of the initial gridded data at a higher resolution by the means of a fusion process.

2 PROBLEM STATEMENT

Be a set of gridded data $S(0)_{r0}$ available at a spatial resolution $r0$ (taken here as equivalent to the cell size). $S(0)$ may be multi-modality. Let be other sets of gridded data $S(i)_{ri}$, $i=1..N$, exhibiting spatial resolution $ri$. The sets $S(i)$ are ranked by decreasing cell sizes, $rN$ is the smallest cell size (the highest spatial resolution). The sets $S(i)$ can be multi-modality.

It is assumed that $S(i)$ and $S(0)$ are geometrically aligned. It is further assumed that each $S(i)_r$ can be associated to $S(0)_r$, that is a relation exists between these data sets for a given resolution $r$.

The problem is the synthesis of $S(0)_{rN}$.

3 THE ARSIS CONCEPT AND THE GENERAL SOLUTION

The application of the ARSIS concept is one solution to this problem. The acronym ARSIS stands for the French "Amélioration de la Résolution Spatiale par Injection de Structures", which means "spatial resolution enhancement by injection of structures" (Ranchin et al. 2003; Ranchin & Wald 2000; Wald 2002).

Under this concept, it is considered that the missing knowledge permitting to synthesize $S(0)_{rN}$ from $S(0)_{r0}$, is composed of the spatial high-frequencies (formally, wavevectors) of $S(0)$ comprised between $[1/(2 r0), 1/(2 rN)]$. This range of high frequencies
(noted HF) is present in the other data sets, e.g., \( S(i) \) exhibits frequencies in \([1/(2r_0), 1/(2r_{i})]\). An appropriate modeling permits to synthesize the searched \( S(0)HF \) from the known \( S(i)HF \). Then the application of a multi-scale model (MSM) constructs \( S(0)_{rN} \), starting from \( S(0)_{r0} \) and using the estimated HF.

4 ASSESSING \( S(0)HF \)

\( S(0)_{r0} \) does not exhibit frequencies greater than \(1/(2r_0)\):

\[
S(0)HF = 0, \text{ for } r < r_0
\]  

(1)

Let note the multi-scale analysis process MSM. The couple of information \((S(i)_{r0}, S(i)HF_{r0\rightarrow i})\) made of the approximation of \( S(i) \) at resolution \( r_0 \) and of the high frequencies of \( S(i) \) in \([1/(2r_0), 1/(2r_i)]\) is obtained by applying the MSM to the data set \( S(i)_{r0} \):

\[
(S(i)_{r0}, S(i)HF_{r0\rightarrow i}) = \text{MSM}(S(i)_{r0})
\]  

(2)

The MSM process may be the multi-resolution analysis algorithm of Mallat (1989) or Dutilleux (Dutilleux 1989; Starck & Murtagh 1994) or the generalized Laplacian pyramid (Aiazzi et al. 1999). The MSM process is invertible and the synthesis process is noted MSM\(^{-1}\)

\[
S(i)_{ri} = \text{MSM}^{-1}(S(i)_{r0}, S(i)HF_{r0\rightarrow i})
\]  

(3)

Assume that:

\[
S(0)_{r0} = F_0(S(i)_{r0}) + \varepsilon_0
\]  

(4)

where \( F_0 \) and \( \varepsilon_0 \) are known, e.g., by statistical regression. Note \( \text{var}(0) \) the fraction of variance in \( S(0)_{r0} \) explained by \( F_0(S(i)_{r0}) \). If we write

\[
S(0)_{r0} = \text{MSM}^{-1}[S(0)_{r0}, \text{var}(0) F_0(S(i)HF_{r0\rightarrow i})]
\]  

(5)

where \( S(i)HF_{r0\rightarrow i} \) is obtained by applying MSM to \( S(i)_{ri} \), \( S(0)_{r0} \) is an estimation of \( S(0)_{r1} \).

Because the model \( F_0 \) is not exact and was assessed at another resolution \( r0 \) while applied to the interval \([r1, r0]\),

\[
(S(0)_{r0}^*, S(0)HF_{r0\rightarrow i}^*) = \text{MSM}(S(0)_{r1}^*)
\]  

(6)

\( S(0)_{r0}^* \) is not equal to \( S(0)_{r0} \). It should be because the fused product should be identical to the initial product at resolution \( r0 \). One corrects for it by writing:

\[
S(0)_{r1} = \text{MSM}^{-1}(S(0)_{r0}, S(0)HF_{r0\rightarrow i}^*)
\]  

(7)

The process is iterated. At each step and resolution \( ri \), are used the sources of data \( S(i) \) having a better resolution than \( ri \).

At the end of the process, \( S(0)_{rN} \) is obtained that is an approximation of the true but unknown gridded data. In any case, this set better represents the reality than what can be obtained by a resampling of \( S(0)_{r0} \).

5 APPLICATION CASE

The Linke turbidity factor (TL, for an air mass equal to 2) is a very convenient approximation to model the atmospheric absorption and scattering of the solar radiation under clear skies. It describes the optical thickness of the atmosphere due to both the absorption by the water vapour and the absorption and scattering by the aerosol particles relative to a dry and clean atmosphere (Kasten 1996; WMO 1981). The larger TL, the larger the attenuation of the radiation by the clear sky atmosphere. It is a convenient parameter to summarize the turbidity of the atmosphere and is often used by engineers and consultants. TL (most usually atmospheric turbidity) is a key input to several models assessing the downwelling irradiance under clear skies that are used several communities in the fields of renewable energies, climatology, agro-meteorology, and atmospheric pollution.

TL can be obtained directly from observations performed during very clear sky periods, but this kind of experimental data is rarely available, thus TL is generally an estimated parameter. Also, time series of radiation data are generally too short to allow estimation on a daily basis. Fortunately, long-term monthly average values, \( TL_m \), are sufficient for most applications. Several methods for estimating \( TL_m \) values can be found in the literature: for example from time series of daily global horizontal irradiation (Aguiar, in ESRA 2000); from monthly Angström sum (Page, 1986); from subjective assessments of the type of atmospheric conditions prevailing in the region of interest (linear regression with latitude and with the Atmospheric Turbidity Index, Dogniaux and Lemoine, 1983).

Most often information for specific locations is presented (Jacovides 1997; Pedros et al. 1999), though at times, larger areas are dealt with. For example, Gueymard (1994) presents data for several Canadian and Northern USA sites and a Web site supplies values for approximately 600 sites in Europe and vicinity (Angles et al. 1998, 1999) derived from the ESRA (2000). The best existing database of surface aerosols is the Aerosol Robotic Network (Aeronet), of which a climatology has been presented by Holben et al. (2000) and data can be viewed online (aeronet.gsfc.nasa.gov). The aerosol values can be converted into TL values using existing approximations.

However, TL is mostly unknown worldwide. There is a need for world climatological maps of the Linke turbidity factor (TL) with a cell size of 5' (approx. 10 km at mid-latitude). The availability of such maps will be an important step for the whole community of the stakeholders in solar radiation. No such maps exist presently, even at regional scales.
Several data sets of the Linke turbidity factor are available. They have been derived from other measurements. They are either values for specific geographical sites or they are under the form of gridded data. The values for specific sites are accurate but have a limited extension in space. The gridded values are covering the whole world. They are given for cells of 280 x 280 km² and are space-averaged.

The principle is to fuse these two different sets of information, in order to obtain the final product on a grid with cells of 5° in size. D’Agostino, Zelenka (1992) or Zelenka et al. (1992) proposed solutions to such a problem. In a case similar to ours since they used the same gridded data, Beyer et al. (1997) proceed as follows. The original gridded data set is resampled by the means of a bicubic spline and regridded with a regular cell size of 5° x 5° in a canonical projection. It is then assumed that the value in a cell should be equal to the mean of those observed for the measuring stations within the cell. For all cells containing at least a measuring station, this field is added to the gridded data, providing an unbiased gridded map.

The methodology that is used in the construction of the maps of the Linke turbidity factor is derived from that of Beyer et al. (1997). The main drawbacks of the previous method lies in the fact that the gridded data that are fused with the sites observations at cells of 5° in size result from an interpolation of the gridded data at cells of 280 km in size. Accordingly, there are no high frequencies in the gridded data at 5°; and more exactly, there is no wavenumbers comprised between 5° of arc angle and 150° (approximately 280 km). The gridded data is very smooth. Some high frequencies are locally injected by fusing with sites observations. For areas with no or few observations, the resulting map is still very smooth. The method was appropriate in the case of Beyer et al. because a large number of sites observations were available. It is not our case, because large areas are unknown.

Accordingly, we designed a new method. The principle of computing differences for specific sites, interpolating them and finally adding the field of differences to the gridded data is kept. The difference between both methods lies in the construction of approximations of the maps with a cell of 5° before the integration of the sites measurements. In the method of Beyer et al., these approximations are resampled gridded data. That means that there is no injection of high frequencies and that the spatial resolution of the gridded data is still 280 km, even if the cell is 5° in size. Only the representation has changed. It follows that we are comparing gridded values and ground measurements that are not comparable because the supports of information are very different. It also follows that the resulting field will be smooth, too. Introduction of the ground measurements will not lead to introduction of high frequencies, except where the network of ground measurements is dense compared to the 2.5 degree distance. In the present method, approximations of the Linke turbidity factor are synthesized at a spatial resolution of 5°, which are close to reality, before correcting them by integrating the site measurements. The general solution presented above is used.

It is made of three steps. In the first step, several gridded data having different spatial resolutions and representing different geophysical parameters are fused in order to construct several approximations of the Linke turbidity factor at increasing spatial resolutions. At the end of the first step, the spatial resolution is 20'. In the second step, it is assumed that the value in a cell of 20' in size should be equal to the mean of those observed for the measuring stations within the cell. For all cells containing at least a measuring station, the difference between each cell, this field is added to the gridded data, providing an unbiased gridded map. In the last step, the final products are constructed by fusing the gridded data at 20' with orography, using a similar approach than in the first step.

In principle, a two-steps approach would suffice. The fusion of sites observations and gridded data can be made at the cell of 5° in size, instead of 20' and the third step is not necessary. However, the size of the data to handle at a 5° resolution is large: the matrix of values is 4320 times 2160 pixels. Running an interpolation process based on a distance weighted by the difference in latitudes and in orography, and calling upon approximately 250 sites observations, request a large amount of computational resources. These resources are beyond the capabilities of Meteotest in the framework of this project. Accordingly, a trade-off was found. The fusion of sites observations and gridded data is performed with a cell of 20' in size and the third step permits to obtain the final product at 5° by a fusion of additional gridded data.

7 THE GRIDDED DATA

Gridded data are available that report monthly clear sky irradiations (Staylor approximation). They are supplied by the NASA-Langley Research Center, in the framework of the Solar Radiation Budget Project SRB (DiPasquale & Whitlock 1992; Whitlock...
From these data, the TL can be computed (Remund et al. 2003) by an iterative application of the ESRA clear-sky model (Rigollier et al. 2000) with modifications (Geiger et al. 2002).

The cell is an equal area cell everywhere of 280 x 280 km². At mid-latitude, it corresponds to approximately 2.5 x 2.5 degrees² of arc angle. The twelve maps (one per month) of TL were re-mapped using the canonical projection. This is performed by the means of the spline bi-cubic operator. A mirror technique is used for the edges. The cells are squared and have a size of 160' of arc angle. This initial set of maps is called the set TL$_{SRB}$ (Fig. 1).

Figure 1. The set TL$_{SRB}$ at resolution 160'. June. The first line is latitude 90 N, the last is 90 S. The first column is longitude 180 W, the last is 180 E. Black: large value (approx. 6); white: low value (approx. 1, or unknown for the Antarctic part).

The atmospheric turbidity is a function of the amount of water vapour in the atmosphere. Within the NVAP (NASA Water Vapor Project), the NASA Langley Research Center Atmospheric Sciences Data Center combined several observations to construct world-wide maps of the content of water vapour integrated over the total column (Randel et al. 1996). The gridded data are available on the Internet (NVAP 2001). The monthly mean value for each cell over the 10 years (1988-97) is computed (Fig. 2).

Figure 2. The map of water vapour amount for June. Black: low value; white: large value.

The NVAP maps are re-mapped as above. The cells are squared and have a size of 80' of arc angle. The relation between TL$_{SRB}$ and NVAP (in mm) was assessed over land at resolution 160' by linear regression (Table 1):

\[ TL_{SRB} = a \ln \text{NVAP} + b \]  

Table 1. Parameters of Eq. 8

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>-33.3787</td>
<td>-32.0650</td>
<td>-13.3799</td>
<td>35.4049</td>
<td>82.0606</td>
<td>-32.5927</td>
</tr>
</tbody>
</table>

This data set NVAP is used for resolutions comprised between 80' and 160' over land.

The third data set TL$_{TMAP}$ is composed of maps of the TL, over the oceans only. They are computed from the maps of aerosol optical depth at 630 μm for several years (1985-99) found on the Web server of the Thermal Modeling and Analysis Project (TMAP) at NOAA Pacific Marine Environmental Laboratory (TMAP 2001). The twelve maps were re-mapped as above. The cells are squared and have a size of 80' of arc angle. This data set is used for resolutions 80' to 160' over the oceans.

The fourth data set is a worldwide digital terrain model with a cell size of 5' (TerrainBase 1995). This data set is used from resolution 5' to 80'.

By application of the general solution described above, one obtains twelve maps of the parameter TL with a cell size of 5'. Here, we used the MSM of Starck & Murtagh (1994).

8 MERGING GRIDDED DATA AND GROUND MEASUREMENTS

These gridded data are subsampled down to 20' before merging with ground data. The TL of the stations are corrected to the altitude of the cell. For all cells containing at least a measuring station, the difference between both data sets is computed. Then, these differences are interpolated by the means of a linear unbiased interpolator using the inverse of the squared distance. The distance is that of Lefèvre et al. (2002) which is based on the geodetic distance and takes into account the latitudinal anisotropy of the meteorological phenomena and the difference in elevation.

Once the field of the residuals obtained for each cell, this field is added to the gridded data, providing an unbiased gridded map which is resampled up to 5'. This final map is presented in Fig. 3 for the month of June.

One may note that both maps (Figs. 1 and 3) are similar for large scales. The second one exhibits high frequencies, which are those found in the external data. A close examination reveals some artefacts that originate from the TMAP or NVAP data. The maps look spotty, but most of the spots can be explained. Several regional effects like wood burning are not included in the background map but in the ground information. Because of the low density of ground stations (outside Europe and North America) some of the local effects are mapped as spots. Many tropical stations show high aerosol values just before
rainy season, when the wood is burned (see September/October in the Amazon region). The Sahel zone exhibits high turbidity values during the whole year.

These maps for the twelve months are available on line at the web site (www.soda-is.com) of the SoDa service delivering information on the solar radiation at ground level (Wald et al. 2002).

9 QUALITY OF THE ASSESSMENT

The quality of the retrieval is assessed by performing the above-described operations with all the stations but one. By iteration and changing the removed station at each iteration, one gets a series of discrepancies between the observed TL and the estimates. From this series are computed the bias (MBE) and the root mean square error (RMSE) and reported in Table 2. This is called the effective error. The RMSE is influenced by outliers in the tropics where the distances between sites are very large and the values vary very much in space. About 6 out of 220 sites were eliminated for calculating the effective error.

In order to evaluate the benefit of the fusion with ground measurements, a second test was made and characterises the error for the first step of the procedure, i.e. constructing a gridded map at 20' resolution. The cell values are then compared to ground measurements (columns "first step" in Table 2). Remembering that the TL varies approximately between 1 and 6, the MBE after the first step is quite high as well as the RMSE.

The fusion with ground measurements is very beneficial. The bias is null. The RMSE error lies between 0.6 (Dec.) and 0.9 (March) with an average value of 0.7.

Table 2. Assessment of the quality

<table>
<thead>
<tr>
<th>Month</th>
<th>Number of sites</th>
<th>First step</th>
<th>Effective error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MBE RMSE</td>
<td>MBE RMSE</td>
</tr>
<tr>
<td>January</td>
<td>222</td>
<td>-0.6 1.2</td>
<td>0.0 0.7</td>
</tr>
<tr>
<td>February</td>
<td>230</td>
<td>-0.7 1.3</td>
<td>0.0 0.7</td>
</tr>
<tr>
<td>March</td>
<td>228</td>
<td>-0.6 1.2</td>
<td>-0.1 0.9</td>
</tr>
<tr>
<td>April</td>
<td>230</td>
<td>-0.4 1.1</td>
<td>0.0 0.7</td>
</tr>
<tr>
<td>May</td>
<td>239</td>
<td>-0.2 1.1</td>
<td>0.0 0.7</td>
</tr>
<tr>
<td>June</td>
<td>240</td>
<td>-0.1 1.0</td>
<td>0.0 0.8</td>
</tr>
<tr>
<td>July</td>
<td>238</td>
<td>0.0 1.0</td>
<td>0.0 0.8</td>
</tr>
<tr>
<td>August</td>
<td>239</td>
<td>-0.2 1.0</td>
<td>0.0 0.8</td>
</tr>
<tr>
<td>September</td>
<td>238</td>
<td>-0.4 1.1</td>
<td>0.0 0.7</td>
</tr>
<tr>
<td>October</td>
<td>229</td>
<td>-0.3 1.0</td>
<td>0.0 0.7</td>
</tr>
<tr>
<td>November</td>
<td>227</td>
<td>-0.4 1.0</td>
<td>0.0 0.7</td>
</tr>
<tr>
<td>December</td>
<td>219</td>
<td>-0.5 0.9</td>
<td>0.1 0.6</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>-0.4 1.1</td>
<td>0.0 0.7</td>
</tr>
</tbody>
</table>

10 CONCLUSION

The ARSIS concept has mostly been used with images at high or very high spatial resolution. The present study demonstrates its potentials in climatology. By the means of a fusion process, it has been possible to synthesize climatological maps of the atmospheric turbidity at a much finer resolution than the initial one.

The standard implementations of the ARSIS concept, and especially the Inter-Modality Models (IMM), apply to the high frequencies (Ranchin & Wald 2000). Here, the IMM is assessed on the approximations themselves; the HRIMM (High Resolution IMM) is applied on the high frequencies.

Another innovation in this implementation is its iterative process: at each resolution, two synthenses are performed. They are separated by an analysis process and a substitution of the analysis results by the fused product obtained at the previous resolution. In current implementations, the HRIMM replicates the model IMM and comprises only one synthesis. This approach was selected here to ensure that i) the fused product is equal to the initial one once resampled to the initial resolution and ii) the intensity of injected high frequencies is reasonable in order to avoid spatial artefacts.

Owing to this fusion method, twelve climatological maps of the Linke turbidity factor were constructed, one per month. Their availability and their quality are an important step for the whole community of the stakeholders in solar radiation. The quality was estimated as offering a RMSE of 0.7. This may be considered as of low quality. It should be stressed that, even in this case, the construction of such maps is a significant step forward the state-of-the-art and that nobody can provide values of the Linke turbidity factor with a much greater accuracy, except at the relatively rare sites where specific observations are made.

ACKNOWLEDGEMENTS

This work was supported by the European Commission (DG-INFSO) under Contract Number IST-1999-12245 (project SoDa). Many thanks goes to H. Gilgen and A. Ohmura of Institute for Atmosphere and Climate ETH (IACETH) for BSRN data and data of Payerne and Reckenholz. Further we thank the principal investigators Brent Holben and Chuck McClain from GSFC, NASA and Didier Tanré from Univ. Lille for Aeronet level 2.0 data. The Australian Regional Instrument Center, Bureau of Meteorology provided a CD-ROM with global and beam irradiance data of several years. Swiss Meteorological Institute (SMI) provided data of ten stations and the Swiss Agency for Environment, Forest and Landscape (SAEFL) data of five NABEL stations.

REFERENCES


Figure 4. The final map TL at resolution 5' for June.