Kriging of spatial-temporal water vapor data

Roderik Lindenbergh, Maxim Keshin, Hans van der Marel and Ramon Hanssen

Delft Institute of Earth Observation and Space Systems,
Delft University of Technology,
P.O.Box 5058, 2600 GB,
Delft, The Netherlands,
Tel.: +31 15 27 87 649; Fax: +31 15 27 83 711
(r.c.lindenbergh, m.o.kechine, h.vandermarel, r.f.hanssen)@tudelft.nl

Abstract

Water vapor is the dominant greenhouse gas but is varying strongly both in the spatial and temporal domain. A high spatial resolution of down to 300m is available in MERIS water vapor data but the temporal resolution is only three days. On the other hand water vapor observations from GPS ground stations have temporal resolutions in the order of 1 hour, but nearby stations are mostly tens of kilometers away. A collocated cokriging approach, incorporating the results of a structural spatial and temporal analysis of the different water vapor signals, is used to combine both observation sources in order to obtain a combined water vapor product of high spatial and temporal resolution.

Keywords: spatial-temporal, cokriging, water vapor, accuracy

1 Introduction

Water vapor, the gas phase of water, is reported to account for most of Earth's natural greenhouse effect, (Cess, 2005). Moreover, it is an important contributor to the error budget of remote sensing satellite data. Unlike other atmospheric gases it's distribution is strongly varying with time and location, which makes it difficult to monitor. It is however possible to retrieve water vapor estimations from several satellite systems, including Envisat and GPS, (Elgered et al., 2005; Bennartz et al., 2001). The MERIS spectrometer on the Envisat satellite is able to retrieve integrated water vapor (IWV) at a spatial resolution of up to 300m. The temporal resolution however is restricted to three days. On the other hand it is possible to determine IWV values from Global Positioning System (GPS) ground stations with a higher precision at a temporal resolution of between 15 minutes and 1 hour. The inter-station distances are on the level of 80km.

It will be shown here how to combine MERIS and GPS IWV data. The starting point is one reduced resolution MERIS IWV image from August 13, 2003, 10 am, covering an area in North-West Europe around The Netherlands. At reduced resolution, one IWV pixel is available for every grid cell of 1200m x 1200m. The other data set consists of time series of at least hourly IWV estimates obtained at the same day at 26 different GPS ground-stations, positioned in the area covered by the MERIS image. Elsewhere, (Lindenbergh et al., 2006), a correlation of 0.85 between this MERIS IWV data set and the GPS IWV data near MERIS snapshot time is reported, but only after applying some filters on the data.
For both the GPS and MERIS observations, a spatial experimental variance-covariance analysis is performed, (Goovaerts, 1997). The results are compared but moreover, the dense spatial MERIS data set is used for obtaining optimal parameters for the spatial interpolation by means of Ordinary Kriging and Inverse Distance interpolation of the sparse GPS observations, available near MERIS snapshot time. In this case optimal means that the average absolute difference between the GPS estimates and the MERIS observations is minimized. It turns out that the interpolation range as obtained by this gauging procedure is much longer than shown in the experimental spatial covariance functions. In some detail, the differences in outcome from Ordinary Kriging and from Inverse Distance interpolation will be discussed as well.

The time series of GPS observations are used for determining the temporal continuity of the water vapor signal. In the end the resulting temporal covariance is applied for fixing the weight that a MERIS IWV observation obtains at a certain time moment in the combined interpolation of the GPS and MERIS observations.

As the MERIS observations are spatially exhaustive, a collocated Cokriging approach is chosen for combining the GPS and MERIS observations, (Goovaerts, 1997). This means that at a given location and time, a estimation of the IWV contents is made as a linear combination of the GPS IWV observations, available at that time, and the one MERIS observation, available at that location. Thus, for the GPS observations we need a spatial covariance function and for the one MERIS observation a temporal covariance function. It is assumed that no correlation exists between the spatial and the temporal components. As the temporal resolution of the GPS-IWV observations is high enough, no temporal interpolation of the GPS interpolations is needed. A strong advantage of this setup is that the Kriging system is limited in size and is computational efficient.

A drawback of this approach is that it is assumed that the unique MERIS observation at the estimation location is the best MERIS observation available. As wind acts on the water vapor in between MERIS snapshot time and estimation time, the water vapor field may spatially be moved as a whole. This is the assumption of the so-called Taylor’s frozen flow theory, (Taylor, 1938). If Taylor’s frozen flow hypothesis holds, the wind data should be incorporated in the Kriging system. To some extend Taylor’s frozen flow hypothesis can be tested for by determining the spatial cross-covariance between different epochs of GPS IWV observations: if a certain wind direction and wind speed is strongly dominant this should be revealed by a strong anisotropy in the experimental spatial cross-covariogram. Results that are not presented here show that some dominant wind direction can be distilled from such a spatial cross-covariogram, but the GPS-IWV data is spatially too sparse to allow for useful local wind direction and force estimates.

2 Data description and structural analysis

2.1 Water vapor observations

Water vapor is the gas phase of water. Gaseous water represents a small but environmentally significant constitutive of the atmosphere. Knowledge on water vapor values is not only essential for environmental issues but is also one of the major error constituents for satellite measurements from GPS and (In)SAR: the GPS or SAR signal is delayed by atmospheric water vapor. Most of Earth’s water vapor is contained in the lowest two kilometres of the troposphere. In this article we will consider columnar water vapor values. Such values are
expressed in kg/m² that is as the mass of the water vapor contents in a column of atmosphere above a horizontal square patch of 1m x 1m on the Earth's surface.

2.1.1 GPS-IWV.
The Global Positioning system (GPS) is capable of providing reliable estimates of the Zenith Tropospheric Delay (ZTD), with a high temporal resolution (even down to a few minutes). Such estimates can be used to derive the Integrated Water Vapor (IWV), the vertically integrated quantity of atmospheric water vapor. Numerous validation experiments showed that an accuracy of 1-2 kg/m² IWV is achievable for both post-processed and near real-time GPS IWV estimates using measurements from regional networks of ground-based GPS receivers, (Jarlemark et al., 2002). High-accuracy ZTD estimates with a sufficient temporal and spatial coverage can be used to study distribution and dynamics of water vapor in the atmosphere. The possibility to use data from GPS networks for operational meteorology has been demonstrated in the framework of the COST-716 action, (Elgered et al., 2005), which took place in 2001-2004. In 2003 ten European Analysis Centres were participating in that action, which involved processing a network of more than 350 stations covering the whole of Europe. For this project we use time series of IWV estimates from the 26 GPS COST-716 stations as indicated in Figure 1. From every station a time series of one day of measurements on the acquisition day of the MERIS snapshot is available.

Figure 1. MERIS IWV estimates in kg/m² and locations of GPS ground stations.

Figure 2. Spatial covariograms of the IWV data. Dotted: MERIS covariogram. Dashed: mean of 24 hourly GPS covariograms. Continuous: GPS covariogram at MERIS snapshot time.

Figure 3. Mean of the 26 temporal GPS ground station wise covariograms.

2.1.2 MERIS-IWV.
Since its launch on board the Envisat satellite in March 2002, the Medium Resolution Imaging Spectrometer MERIS gives insight into the properties and dynamics of the Earth system with unprecedented accuracy and resolution by obtaining measurements in 15 spectral bands with a
The maximum spatial resolution of 290 m x 260 m. With MERIS, water vapor and cloud top pressure can be monitored operationally on a global scale with a spatial resolution of 1 km, with the opportunity of zooming into several areas with the full resolution of 300 m, (Bennart et al., 2001). The MERIS instrument is a push broom imaging spectrometer, with 15 spectral bands in the visible and near infrared. The main mission of MERIS is oceanography, observing sea-colour. The secondary mission of MERIS is to observe, amongst others, the water vapor column over land, water or above clouds. Observations are limited to the day-side. Global coverage is obtained after 3 days. MERIS also retrieves cloud type and top height. The theoretical accuracy of the estimated water vapor column is 1.7 kg/m² over land and 2.6 kg/m² over water. For this case study we use a reduced resolution MERIS product, see Figure 1, dating from August 13, 2003, covering a large part of North-West Europe around The Netherlands. At reduced resolution IWV observations are reduced to a 1.2 km² grid. The accuracy of the reduced water vapor product is reported to be within 20 %.

2.2 Experimental covariance functions.
Correlation in time or in space between observations can be detected and modelled by a variogram or covariance analysis, (Goovaerts, 1997). The resulting model is used to determine the variance-covariance matrices of the observations. Using the VC-matrices, a Best Linear Unbiased estimation can be obtained for the IWV content at a given time and location. The underlying assumption used in this framework is that the signal, in our case the IWV, can be considered a random function. This means that every observation is one single outcome of a complete distribution of possible observations at that time and location. Stationarity of a random function means that the expectation of the function value is independent of location or time.

The theoretical covariance function of a stationary random function \( Z(x) \) is an expectation, \( E \), and is defined as \( \text{cov}(s) = E((Z(x) - m)(Z(x + s) - m)) \), where \( m = E(Z(x)) \) denotes the mean of \( Z(x) \) and \( s \) a temporal/spatial distance. Given a set of observations, a discrete experimental covariance function is determined by computing experimental covariances between any two observations and by grouping the outcomes according to some distance interval. By fitting the experimental values into a positive definite model, a continuous covariance function is obtained that is used to fill the VC-matrix for a estimation at an arbitrary location and moment.

2.2.1 GPS spatial correlation
As the number of different GPS station considered is only 26, it is difficult to determine a reliable covariogram for the spatial correlation at a single epoch. Therefore a spatial experimental covariogram is determined from the, if necessary, linearly interpolated GPS measurements at every hour between 0.30 and 23.30. These epoch times are chosen because for most GPS stations and most epochs, IWV estimates are available at exactly the half hours. The covariogram for 10:30 is shown by a continuous line in Figure 2. Only those interpolated GPS IWV values were used for the hourly covariograms for which at least one measurement is available within one hour of the covariogram time. The mean of the experimental covariograms obtained in this way is shown by a dashed line in the same figure. This covariogram displays a range, i.e. the maximal distance at which correlation exists, of about 200 km and this approximately holds for all individual covariograms as well.
2.2.2 MERIS spatial correlation
The experimental covariogram of a subset of about 1100 MERIS IWV observations is represented by the dotted line in Figure 2. This covariogram is fairly comparable to the covariogram of the GPS IWV data of 10:30. Due to the higher spatial resolution, the range of the MERIS covariogram is higher than the GPS range and is equal to almost 500 km.

2.2.3 GPS temporal correlation
The time series of the GPS IWV at the different GPS ground stations display a strong trend. Therefore a linear trend was fitted at each station and removed from the data. The mean of the 26 temporal covariance functions, determined from the detrended data is given by the brown dots in Figure 3. As the IWV values at most stations even show some non-linear trend during the day, the stationarity condition for the random function does not hold very well. This is expressed by the negative covariance values at higher distance.

3 Combining the MERIS and GPS water vapor observations

3.1 Ordinary Kriging’s clustering and screening properties.
Incorporating the variance-covariance structure of the observations into the interpolation method has some important consequences, (Wackernagel, 2003; Chilès et al., 1999). Two interpolation effects occur that do not occur in case of e.g. inverse distance interpolation. The first effect is known as clustering. According to the spatial/temporal continuity assumption, close by observations will be highly correlated. This implies that the additional weight obtained by an extra observation in the direct neighborhood of an existing observation will be limited, as according to the high correlation, there is limited new information in the extra observation. This effect is called clustering because weights can be at first instance being thought of as being divided between the different clusters of observations rather than between individual observations.

The second effect is known as screening and can be thought of as happening within one cluster. If one observation is behind another observation with respect to the estimation location, the observation behind contains limited new information. Therefore it will obtain a low weight compared to the observation in front, that is preferred because it is closer to the estimation location. It is even possible that the observation behind gets a negative weight. The stress on the weights can be relieved by increasing the so-called nugget value or white noise. This value encodes the short time variability or the size of the measurements errors. The clustering and screening effect will have relevant influence on the result, especially for the GPS-IWV observations where the observations are spatially sparse and not regularly spaced.

3.2 Optimizing the GPS IWV estimation.
When producing IWV maps from the GPS IWV observations, the additional MERIS observations can be used in two different ways. First of all, the MERIS observations are directly incorporated in producing the maps by means of the Cokriging approach. For this purpose the individual VC-matrices of both the GPS and the MERIS observations are needed together with the cross-correlation VC-matrix containing the cross-covariances between the MERIS and GPS observations. Moreover, the MERIS IWV observations can be used to gauge and control the covariance parameters of the GPS IWV observations. For this purpose the MERIS observations are gridded to a 0.25 x 0.1 degree longitude-latitude grid. This implies that one grid cell has an approximate size of 17 x 11 km. The cell size is chosen such that local
variations in the MERIS signal are still representable, while the size is not so small that computations become too time consuming. The gridding is done by identifying the nearest neighbor in each of the eight octants around a grid-point and combining these eight points to an interpolated value by means of a quadratic inverse distance scheme. Figure 4, left. Note that in comparison to Figure 1 all smaller gaps have been filled. Grid points for which no observation is available in all eight octants within 100 km did not get a value.

![Figure 4. Grided MERIS (far-left) and OK-interpolated GPS IWV observations at MERIS snapshot time. Both data sets are grided/interpolated to the same grid. On the right the difference between the two maps is given.](image)

In the following, this gridded MERIS data set is used as reference data for determining optimal parameters for the interpolation of the GPS-IWV data near MERIS time. As a reference, the mean absolute difference between the gridded MERIS data and the mean of the GPS-IWV observations equals 4.565 kg/m². Two methods were used for interpolation of the GPS-IWV data. For both methods optimal parameter values were determined by minimizing the mean absolute difference at the 0.25 x 0.1 degree longitude-latitude grid between the interpolated MERIS-IWV values and the interpolated GPS-IWV values. Firstly, the GPS-IWV data were interpolated using plain inverse distance, that is every observation gets a weight

$$w_i(p) = \frac{1}{d_i^p \sum_{j=1}^{n} \frac{1}{d_j^p}}$$

(1)

where $n$ denotes the number of observations and $d_i$ the spheroidal distance between the interpolation location and the observation location. Here the only parameter to determine is the power $p$ of the interpolation. The best result, for $p$ in {2,3, ..., 6} is obtained for $p=3$, when the mean absolute difference equals 4.059.

The mean absolute difference was determined for different parameters within the ordinary Kriging framework. Here a minimum absolute difference of 3.989 was obtained with a nugget of zero, a sill of 10, and a long range of 3000 km, while using the exponential model. The result of this interpolation is shown in Figure 4, middle left. while, at the right, the difference between the interpolated MERIS data and this GPS map is given. It is remarkable that the experimental covariance functions as derived above indicate a much shorter range of 200 km from just the GPS-IWV data and still only 500 km from the MERIS-IWV data than found by the optimization procedure of above. Indeed, a range of 500 km, together with a sill of 50 and a filtered nugget of 1, while still using the exponential model, results in a mean absolute difference of 4.097. The need for using a long range is caused by the often large distances...
between the interpolation locations and the observations. When using a short range, the Ordinary Kriging system will not attach higher weights to the nearest observations, as according to the range there is no correlation between the estimation location and the observations. As a result, the estimation value will be close to the Kriged mean, that is the mean which incorporates the correlation/clustering of the observations. Using a long range forces the Kriging system to attend more weight to the nearest observation and this leads to better results in comparison to the MERIS data.

In Figure 5, a difference plot is given between the optimal map as obtained by Ordinary Kriging (OK) and by Inverse Distance. In the areas filled with vertical lines Inverse Distance gives a higher estimation, Ordinary Kriging is higher in the areas filled with horizontal lines. The differences are mostly explainable by the clustering and screening properties of OK as discussed above. In the middle of the picture for example OK gives a higher prediction value. This is because the lower observation values in the west of The Netherlands and in Belgium are screened of by the higher values at EUSK and WSRA. In the difference picture, Figure 4 right, green indicates under estimation by the GPS interpolation and red over estimation. This figure shows that even the (higher) OK interpolated GPS-IWV values underestimate the MERIS observations, which shows that in this case the OK estimates are closer to the MERIS values as the Inverse Distance estimates. We conclude that for interpolation of the GPS IWV stations, OK is giving the best result.

3.3 Spatial-temporal interpolation.

The MERIS IWV observations have a high spatial resolution, but only one epoch of observations is available. On the other hand, the GPS IWV observations are spatially sparse but are available at, say, hourly epochs. Therefore it is decided to estimate an IWV value \( I(t_0, (x,y)) \) at epoch time \( t_0 \) and arbitrary location \( p=(x,y) \) as a linear combination

\[
I^*(t_0,p) = (w_1 \cdot I_{GPS}(p_1,t_0) + \ldots + w_n \cdot I_{GPS}(p_n,t_0)) + v \cdot I_M(p,t_M)
\]  

(2)
of GPS-IWV observations $I_{\text{GPS}}(p_1, t_h), \ldots, I_{\text{GPS}}(p_n, t_h)$ at $n$ different GPS stations $p_1, \ldots, p_n$ at epoch time $t_h$ and of the (only) one MERIS-IWV observation $I_{\text{MERIS}}(p, t_h)$ at location $p$, obtained at MERIS-IWV observation time $t_h$. The weights $w_1, \ldots, w_n$ for the $n$ GPS-IWV observations and the single weight $v$ for the one MERIS-IWV observation are obtained by solving the following Ordinary Kriging like system.

$$
\begin{pmatrix}
1 & c_{S_{i2}} & \cdots & c_{S_{in}} & 0 & 1 \\
 c_{S_{i2}} & 1 & \ddots & \vdots & 0 & 1 \\
 \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\
 c_{S_{in}} & \cdots & c_{S_{(n-1)n}} & 1 & 0 & 1 \\
 0 & 0 & \cdots & 0 & 1 & 1 \\
 1 & 1 & \cdots & 1 & 1 & 0 \\
\end{pmatrix}
\begin{pmatrix}
w_1 \\
w_2 \\
 \vdots \\
w_n \\
v \\
 \zeta \\
\lambda \\
\end{pmatrix}
= 
\begin{pmatrix}
c_{S_{p1}} \\
c_{S_{p2}} \\
 \vdots \\
c_{S_{pn}} \\
 c_{t_{\text{MERIS}}} \\
1 \\
\end{pmatrix}
$$

(3)

The VC-matrix on the left in Equation (3) contains the spatial correlations $c_{S_{ij}}$ between the IWV-GPS observations in the $n \times n$ top left part. The $(n+1)$-th row and column express that no correlation is assumed between the spatial and temporal contributions to the system, while the last row and column are added for obtaining an unbiased estimation. In the proximity vector on the right, the first $n$ entries contain the spatial correlations $c_{S_{ip}}$ between the estimation location, $p$, and the locations of the GPS-IWV stations. The $(n+1)$-th entry $c_{t_{\text{MERIS}}}$ gives the temporal correlation between the estimation time, $t_h$, and the MERIS-IWV observation time, $t_h$, while the last entry again is obtained from the unbiasedness condition $w_1 + \ldots + w_n + v = 1$. Solving System (3) gives a unique solution for the weights, because of the positive definiteness of the VC-matrix.

Except for a IWV estimation $I^*(t_h, p)$ at time $t_h$ and at location $p$ we obtain a quality description in the form of a (scaled, $S$) residual variance $\text{var}_S [r(t_h, p)]$, with $r(t_h, p) = I^*(t_h, p) - I(t_h, p)$ the difference between the estimation $I^*(t_h, p)$ and the (unknown) real value $I_0(t_h, p)$ at the estimation space-time point $(t_h, p)$, compare (Goovaerts, 1997; Wackernagel, 2003). The residual variance is given by

$$\text{var}_S [r(t_h, p)] = 1 - \sum_{i=1}^n w_i \cdot c_{S_{0i}} - v \cdot \zeta_{\text{MERIS}} - \lambda$$

(4)

Basically, the residual variance expresses the proximity, both in space and time, to the observations. It is locally minimal at the GPS stations and at MERIS observation time while it increases together with the decrease in correlation between the estimation space-time point and the observations.

This approach is known as the collocated approach. It is fast as the Kriging System (3) is small. The size depends only on the number of GPS ground stations. Moreover the VC-matrix is fixed during the interpolation, which implies that it only needs to be inverted once. The values in the proximity vector however are different for each estimation space-time point. Disadvantage of this method is that it will not fill any spatial gaps in the MERIS grid. This could be resolved for by including a spatial interpolation component for the MERIS observations as well, but then the computational efficiency will be lost.
3.4 Spatial-temporal combination.

In Figure 6 some results of the spatial-temporal interpolation procedure are shown. The left most figure gives the estimation at 10:30, close to MERIS-IWV observation time. Here a lot of small-scale detail is visible, in the same pattern as in Figure 1, that showed the MERIS-IWV observation alone. The combination with the GPS-IWV observations reduces the size of local deviations that are only present in the MERIS-IWV observations however. The middle figure gives an estimation for 14:30 and here the GPS-IWV observations are dominating, but still some smaller scale details are visible, for example off the Norway coast, that were only present in the MERIS-IWV data set. At 18:30 the estimation map is very smooth, compare Figure 4, middle left, because it is by now completely dominated by the GPS-IWV observations as there is no temporal correlation anymore with the MERIS acquisition time. The same kind of change in pattern is observed in the estimation map produced for time points before MERIS-IWV acquisition time.

What remains is to give a good quality description of the IWV estimations obtained in this way. As discussed above, the Kriging error variance gives an indication for the proximity of the observation points. A step that still has to be added however, is to incorporate the drift of the water vapor field due to wind power into the spatial-temporal interpolation procedure. This step, combined with a validation step using e.g. a one month timeseries of both GPS-IWV and MERIS-IWV data sets can lead to an adequate quality description of a combined MERIS-GPS water vapor product.

4 Conclusions and outlook

By determining experimental covariance functions, insight can be gained into the spatial and temporal correlation of the IWV observations. The resulting correlation description can be encoded in a spatial-temporal interpolation procedure for combining MERIS-IWV and GPS-IWV observations, similar to Ordinary Kriging.

The spatially dense MERIS-IWV observations can be used as reference data when finding optimal parameter values for the spatial interpolation of the GPS-IWV observations. In this way a much longer correlation range is found however than by the experimental covariance function analysis.
The results of the spatial-temporal combination of the GPS and MERIS IWV observations match intuition. A next step is to incorporate local wind data in the spatial-temporal combination, as, according to Taylor's frozen flow assumption, wind transports water vapor fields as a whole, to some extent. An experimental cross-covariance analysis of GPS-IWV observations in different epoch already reveals some anisotropy, probably due to wind, but the windfield is too inhomogeneous at larger spatial scales to allow direct estimation from the data set at hand.

These first results were obtained with observations of only one day. Testing of larger data sets will allow for better validation that will result in an adequate quality description of a final combined MERIS-GPS water vapor product. Currently a database of recent IWV MERIS and GPS observations over North-West Europe is gathered, that will be analyzed in the next phase of the project.

Acknowledgements
Sybren de Haan from the Dutch Royal Meteorological Society (KNMI) is thanked for providing the authors with the GPS IWV data and for his useful comments. The MERIS data is distributed free of charge by the European Space Agency, ESA. This project is funded under number EO-085 by the Netherlands Institute for Space Research (SRON).

References