

The Humidity Composite Product of EUMETSAT's Climate Monitoring SAF: Towards an Optimal Merging of Satellite Data Sets

¹Jörg Schulz, ²Ralf Lindau, and ¹Nathalie Selbach

*¹Satellite Application Facility on Climate Monitoring, Deutscher Wetterdienst,
Offenbach, Germany, e-mail: joerg.schulz@dwd.de*

²Meteorological Institute, University of Bonn, Germany

Abstract

The need for a comprehensive and accurate global water vapour data set as an assisting tool for scientific studies in the atmospheric sciences has been acknowledged during the last 10 years. Such a data set is extremely useful for all aspects of climate science being dependent on accurate water budget data, e.g. general circulation model verification or regional climate studies. A data set is all the more useful if it contains an objective error information. In this study a universally applicable technique to merge two satellite data sets to create daily mean water vapour fields is presented. The technique is based on the commonly known kriging technique which is an optimal interpolation technique that provides not only fully covered fields but also a corresponding map of errors. kriging can be applied to water vapour estimates from simultaneous flying instruments sharing the same retrieval algorithm, e.g., the ATOVS instrument combination on NOAA satellites. The technique also holds the potential to merge estimates from different sources, e.g., SSM/I and AMSU-A if systematic errors between retrievals are understood. Within this study the benefits of kriging are studied for the derivation of daily mean fields of ATOVS derived total precipitable water estimates in one exemplary month.

Introduction

The Humidity Composite Product (HCP) of EUMETSAT's Satellite Application Facility on Climate Monitoring (CM-SAF) will integrate data from several existing and upcoming satellites, e.g., microwave imagers like SSM/I as well as water vapour profiling instruments on the Meteorological Operational polar platform (MetOp) and ATOVS data from NOAA platforms. CM-SAF will provide single sensor estimates as well as some merged estimates, e.g., from SSM/I and AMSU-A. The production of this thematic climate record relies on calibrated and inter-calibrated input data to be provided by the satellite operators.

Two pilot studies on the creation of daily mean fields have been undertaken. The first provided a merged result from total precipitable water (TPW) observations from AMSU-A onboard the NOAA-15 and NOAA-16 satellites and from SSM/I on DMSP-F13, F14, and F15 satellites (Lindau and Schulz, 2004). This second study considers the merging of ATOVS

estimates from NOAA-15 and NOAA-16. In both studies the merging is performed by kriging, an optimal interpolation technique that provides not only fully covered fields but also a corresponding map of errors. The obtained errors reflect mostly the actual sampling situation and should not be mixed up with retrieval errors that have to be determined by external comparisons to other data. The technique has been chosen because of its potential to merge data from several completely different sources if they have no bias errors. It also holds the potential to include spatiotemporal resolved information on the retrieval error itself. Within this paper the kriging technique is presented and an exemplary application to ATOVS data from April 2004 is shown.

Data

In this study observations of the total precipitable water from ATOVS (Advanced TIROS Operational Vertical Sounder) onboard NOAA satellites are used, where TPW is computed from mixing ratio profiles derived at NOAA using the retrieval described in Reale (2003). Data from two satellites, NOAA-15 and NOAA-16 are considered aiming at an optimal merged product. We analysed one month of data (April 2004) over an area spanning from 20°N to 80°N and from 60°W to 60°E.

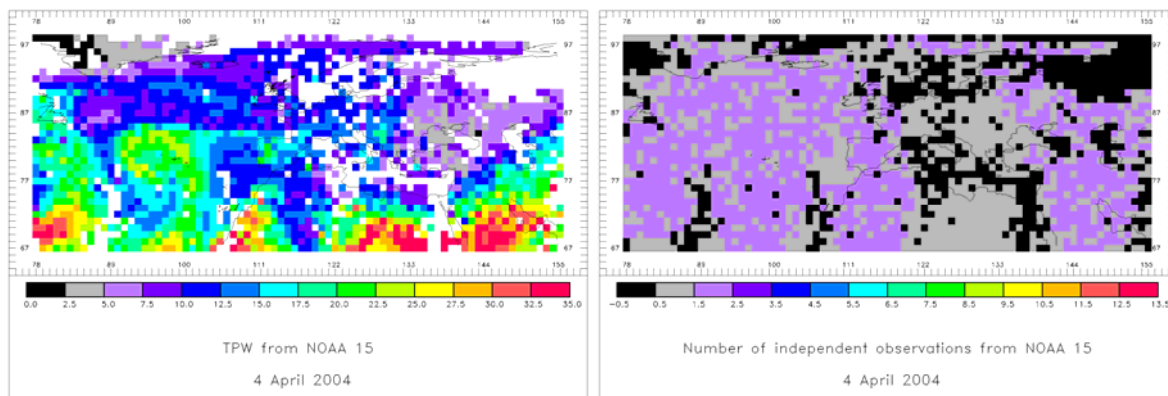


Fig 1: Example of the input data to be merged by the kriging procedure. The map on the left shows the daily mean TPW derived from ATOVS on NOAA-15 on April 4, 2004. TPW is computed from ATOVS derived mixing ratio profile and has been brought to an equal area projection with a nominal grid cell distance of 150 km. The map on the right shows the number of independent observations from NOAA-15. As pixels of the same overpass are considered to be not independent, the maximum number of independent observations per satellite and day is equal to two.

As raw data, swath-oriented observations of the two satellites are used. However, in a first step this data is spatially and temporally averaged because the kriging procedure requires an error estimate that is concluded for each grid box from the internal box variance. However, this approach provides reasonable estimates for the error variance only if the data comprised in each box can be considered as independent. It is obvious, that independency cannot be assumed for neighbouring pixels of the same satellite. Consequently, observations are only considered to be independent, if they come from different satellites or different overpasses.

On the one hand the grid boxes have to be chosen relatively large, because a sufficient number of independent observations should be available in each grid box to allow for statistical calculations. On the other hand, the highest possible resolution should be of course retained. As a compromise the data is spatially averaged over 150 km by 150 km on an equal-area grid. The equal-area grid is used to gather a sufficient amount of data in northern regions. A temporal averaging over one day is simply used to achieve the derivation of a daily mean map. Figures 1 and 2 show the pre-processed data for the 4th April 2004 for NOAA-15 and NOAA-16, respectively.

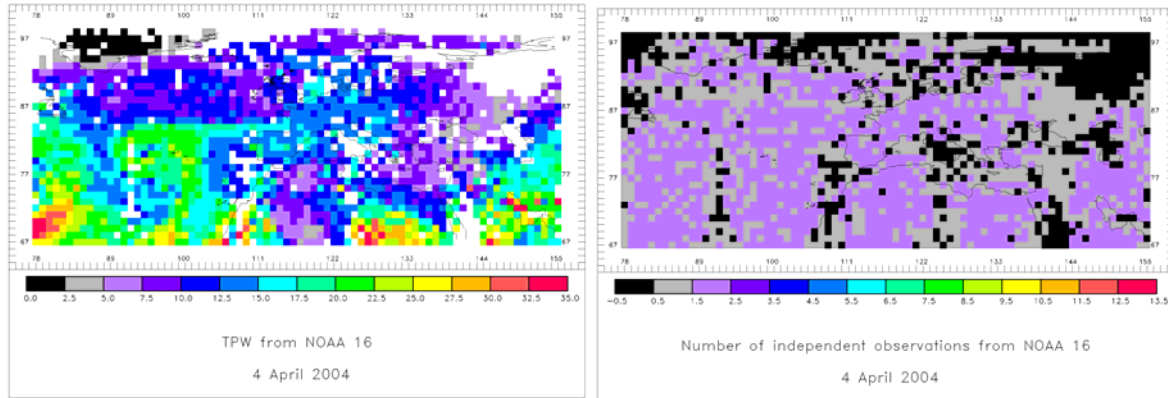


Fig.2: As figure 1, but for NOAA-16

Kriging

Kriging can be regarded as a prediction of a value x_0 at a location P_0 where no measurement is available employing information from measurements at the surrounding locations P_i . A solution is not possible for one single case. However, if a time series of m measurements at each location P_i is available it is reasonable to minimise the expression:

$$\sum_{i=1}^m \left(x_0 - \sum_{i=1}^n \lambda_i (x_i + \Delta x_i) \right)^2 = \min \quad (1)$$

where x_i denote the available measurements, Δx_i their errors, and λ_i the weights. Differentiating of Eq.(1) leads to the following expression, where the temporal summation is abbreviated by brackets []. As the errors are considered to be random, error terms occur only on the diagonal of the matrix:

$$\begin{pmatrix} [x_1 x_1] + [\Delta x_1 \Delta x_1] & [x_1 x_2] & \cdots & [x_1 x_n] \\ [x_2 x_1] & [x_2 x_2] + [\Delta x_2 \Delta x_2] & \cdots & [x_2 x_n] \\ \vdots & \vdots & \ddots & \vdots \\ [x_n x_1] & [x_n x_2] & \cdots & [x_n x_n] + [\Delta x_n \Delta x_n] \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} [x_0 x_1] \\ [x_0 x_2] \\ \vdots \\ [x_0 x_n] \end{pmatrix} \quad (2)$$

Thus, a linear set of equations results, containing the covariance matrix between the data points P_i with the error variances on its diagonal, and a vector giving the covariance between

the predicting point P_0 and the locations P_i . If these quantities are known, the solution of the above set of equations leads, via the determination of the weights λ_i , to the optimal prediction at P_0 . Obviously, the minimised expression of Eq.(1) is equal to the error of the predicted value, often called as kriging error. Transformation of that expression leads to:

$$[x_0 x_0] - 2 \sum_{i=1}^n \lambda_i [x_0 x_i] + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j [x_i x_j] + \sum_{i=1}^n \lambda_i \lambda_i [\Delta x_i \Delta x_i] \quad (3)$$

The first term denotes the variance at P_0 and is equal to unity if anomalies are considered. The second term contains the covariance between data points, the so called **information**. The third term contains the so called **redundance** because information from points P_i may not be independent. The last term describes the individual errors at the points P_i .

To determine the weights λ_i information on the spatial covariance and the error of the individual observations is needed. Note that the spatial covariance is spuriously modified by error covariance between the points. Although not explicitly mentioned the average effect of the spatial error covariance is implicitly included within the information and redundance terms.

Spatial Correlation Function

According to Eq.(2), a prerequisite for the application of the discussed kriging technique is the knowledge of the spatial covariance matrix, or since we are dealing with normalised anomalies the knowledge of the correlation matrix. Primarily, the correlation of two time series strongly depends on the distance between the locations of the individual observations. Furthermore, the spatial covariance strongly depends on the considered time scale: The prediction of the daily anomaly at a certain location is indeed possible with measurements from a distant observation, while short-time variability is connected only over short distances. Since we are aiming at the correlation length of daily means on a spatial scale of 150 km, covariances should also be calculated on the base of such time and space averages, which show larger correlation lengths than individual observations representing a quasi instantaneous snap shot.

It should be emphasised that the correlation of the anomalies against the local monthly means have to be considered here because such data will be processed in the kriging procedure instead of raw observations. In the following the covariance of daily means is calculated as a pure function of distance, which means the covariance of different regions is lumped together. Otherwise, regions of high variability would dominate the results. The normalisation allows to take into account also the information of low variable regions adequately. In the concrete accomplishment, an exponential function is fitted to the correlation:

$$r = e^{a_0 + a_1 x} \quad (4)$$

The correlation is calculated for daily averages, which are based on the maximal four independent observations available at each location, two from each platform and two from the morning and evening overpass, respectively. The distances are measured in units of the grid lengths of 150 km. For zero distance, no direct observation is available, but the fit of an exponential function allows a rough estimate of the average error variance, which can be read at the ordinate with 0.969, meaning that a fraction of 0.031 of the total variance can be assigned to the error variance. The correlation length (the distance, where the correlation is decreased by a factor of $1/e$ compared to the correlation found for zero distance) is determined to be 4.3, corresponding to 645 km.

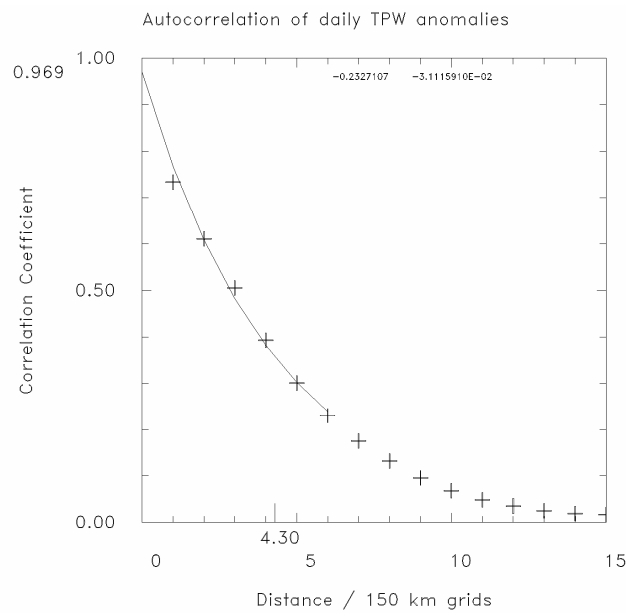


Fig.3: Correlation function of daily TPW anomalies for April 2004

Error Determination

To solve the kriging equation (2) the individual error variance of the used data points has to be known. The determination of the error variance is accomplished by a decomposition of the total variance into four components (von Storch and Zwiers 1999; Lindau 2003). These are i) the error of the monthly mean, ii) the seeming extra daily variance, iii) the mean error of daily means, and iv) the true intra daily variance.

Fig. 4 shows the error variance averaged over the entire month, as calculated from the actually used daily errors. Note the extreme differences between the eastern Atlantic with errors in the order of 1 mm^2 and the southern Sahara with errors of 50 mm^2 . This emphasises the need of determining the errors individually instead of using an average error estimate, which would be easily available from the correlation function.

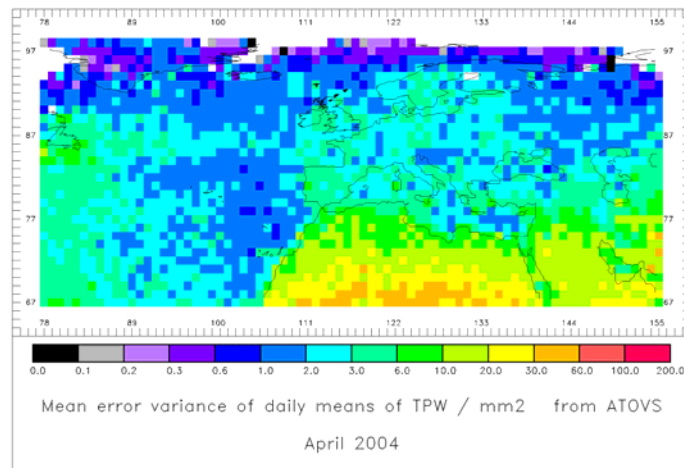


Fig.4: Spatial distribution of the errors of daily TPW means from ATOVS. To give an overview, the map shows the daily error variance averaged over one month.

Products

In this section the kriging procedure is outlined and the results for April 2004 are discussed. In a first step, the actual TPW measurements were averaged within an equal-area grid of 150 km resolution for each day. Four values for each temporal and spatial grid point resulted, derived from the two overpasses of each of the two satellites. Within the next step the TPW values are normalised by subtracting the monthly mean and dividing by the extra daily standard deviation at that specific location. This normalisation has the advantage that no constraint concerning the sum of kriging weights is necessary. In contrary, this sum is a free parameter within our technique and can be interpreted as explained variance. Then the error variance at each individual grid point and day is estimated and accordingly normalised with the extra daily variance. Using the spatial correlation function for the actual month yields an optimal interpolated field for each day. The benefit of kriging is that each interpolated daily field is obtained together with a daily error map.

The standard kriging output comprises of the daily mean fields and the corresponding error maps (Fig. 5). The two maps show the normalised TPW anomaly (left) from ATOVS on April 4, 2004. The values are normalised by subtracting the monthly mean and dividing by the extra daily standard deviation. The right-hand side map shows the corresponding normalised error of the TPW anomaly. Due to the normalisation the maximum error is equal to 1, corresponding to the extra daily standard deviation, which would be obtained if all daily values were set to the monthly mean.

The TPW field of that day is characterised by anomalous high moisture over the Mid Atlantic and opposed deviations in the Turkey and the Black Sea. The deviations attain a normalised value of 3, which means that they are three times higher than the extra daily variance for that location and month. However, these are extremes. Broad regions exhibit near zero anomalies denoting that the TPW on 4th of April do not differ substantially from the monthly mean. Important additional information is provided by the error map also given in Fig. 5. The highest theoretical value of the normalised kriging error is 1, which means it is as high as the extra daily variability. Predicting for each day of the month just the monthly mean

would yield that error. Only at the northern edge of the considered area the normalised error attains 1 together with a TPW anomaly near 0 (see Fig. 5). This illuminates the general behaviour of kriging: to suggest the monthly mean average, if no information at all is available, but to warn the potential user by a maximum error of 1.

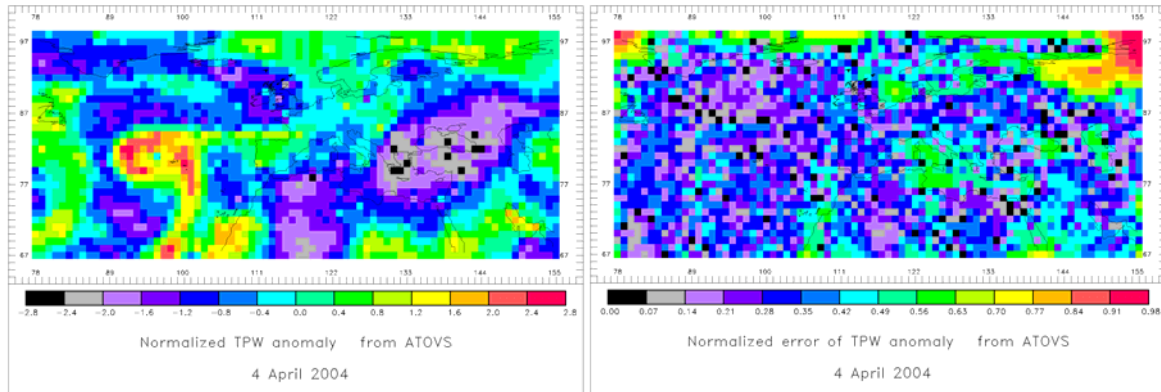


Fig. 5: Example of the standard kriging output. The left map shows the normalised TPW anomaly from ATOVS for April 4, 2004. The map on the right gives the corresponding error.

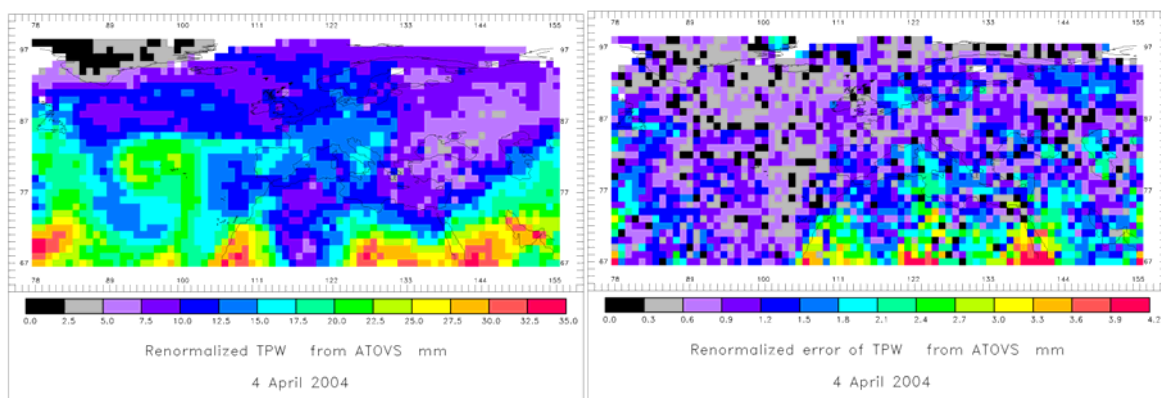


Fig.6: Resulting TPW given in mm from ATOVS for the 4 April 2004 (left). The field is obtained by reconvertng the normalised field shown in Fig.5 with the monthly mean and the extra daily standard deviation. The map on the right-hand side shows the corresponding error (also in mm), as obtained by renormalisation with the extra daily standard deviation.

For many applications only the absolute values are significant for the interpretation of the fields, so that a renormalisation is required. It is performed by a reversed application of monthly mean and the extra daily standard deviation. After renormalisation, the fields shown in Fig. 6 are obtained, which are examples for the final kriging result.

Instead of depicting the kriging result spatially for one arbitrary day (Figs 5 and 6), the obtained data can also be discussed by using time series of the entire month for one arbitrary grid point. This alternative depiction of the results is performed for two locations, one in the Mid Atlantic (Fig. 7a) and one in the Baltic Sea (Fig. 7b).

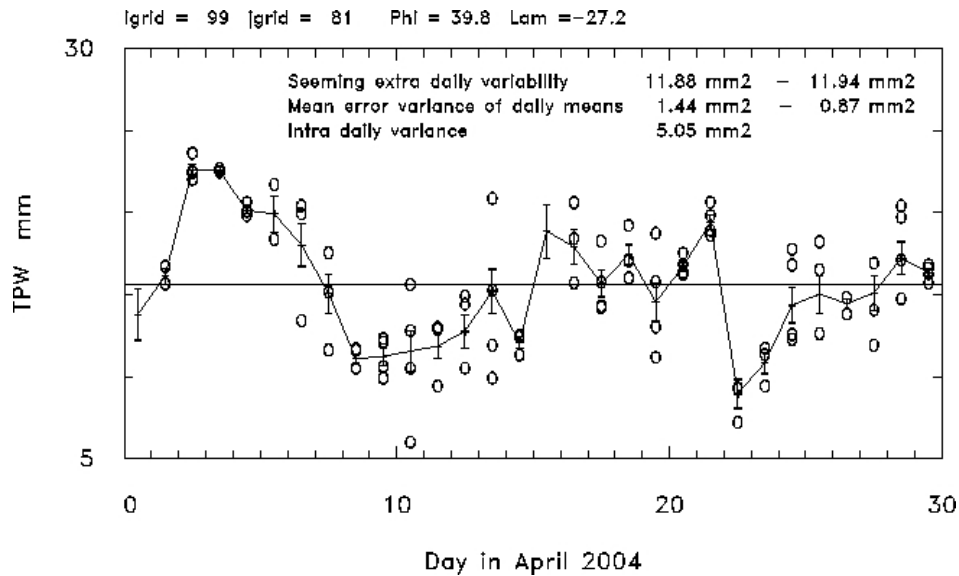


Fig. 7a: Time series showing both the independent satellite observations used as kriging input (circles) and the obtained kriging result, given by the solid line, here for a location the mid Atlantic (27.8°W, 39.8°N). The bars show the kriging error for each day.

For the grid point in the mid Atlantic (Fig. 7a) the following results are found. The extra daily variability as it was estimated from the input observations that is the difference between the seeming extra daily variability and the mean error variance for the daily means ($11.88 \text{ mm}^2 - 1.44 \text{ mm}^2 = 10.44 \text{ mm}^2$) is actually underestimated compared to the kriging procedure (11.94 mm^2). Daily means directly derived from satellite pixels have a mean error variance of 1.44 mm^2 , which is almost halved to 0.87 mm^2 by the kriging procedure. On 4 April, three independent satellite observations are available at this location (only the morning overpass of NOAA-16 is missing), with almost identical TPW observations (22.45, 22.50, 22.65 mm). Consequently, the error of this individual input value is estimated to be very small, so that the kriging procedure provides a rather similar TPW value (22.53 mm) together with a small error estimate (0.06 mm). On 14 April, the satellite input varies considerably between 9.90 mm and 20.85 mm. However, the kriging error remains moderate with 1.33 mm as information from neighbouring grid boxes is included. The most striking benefit from kriging certainly is the fact that values can be attached to days with no observations at all, as e.g., for 1 April, where the error (1.55 mm) is not essentially higher than normal.

The benefit of kriging becomes even more evident at locations with lower data coverage. At a grid box located in the Baltic Sea between Germany and Sweden (Fig. 7b), satellite data are available for only 6 days within April 2004 due to extensive cloud coverage. Nevertheless, a reconstruction of daily TPW time series is possible with errors of about 1 mm.

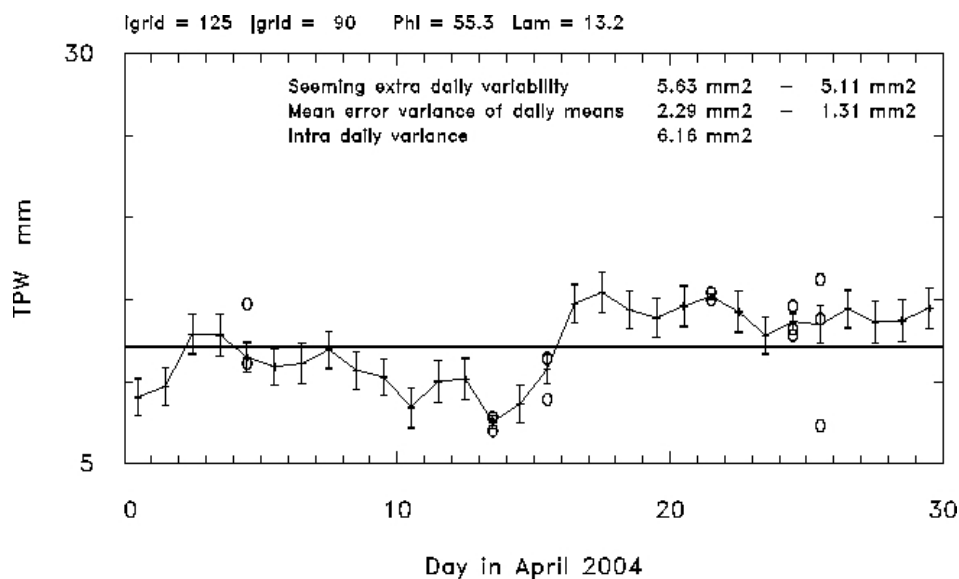


Fig. 7b: As Fig. 7a, but for a location in the Baltic Sea (13.2°E, 55.3°N).

Summary

In this study a universally applicable technique to merge two data sets is presented. The method is based on the commonly known kriging technique applied to daily anomalies of total precipitable water that are normalised by the inter-daily variance. In contrast to ordinary block kriging it is not necessary to prescribe a fixed radius in which observations are considered. The technique presented here manages the aggregation of additional information step by step, deciding at each stage which observation could contribute a maximum of new, not redundant information. This depends on the three characteristics of the potentially added observation: its distance to the predicting point, its individual error and its redundancy with the already aggregated observations.

The procedure is used to combine the total precipitable water fields derived from ATOVS instruments on the NOAA-15 and NOAA-16 satellites. Altogether, information from the two satellites has been merged successfully. It is an inherent benefit of kriging to provide not only optimal interpolations, but also an error field for each map as gratis bonus. It is this field that makes the derived data set valuable for variability analysis. This error fields mainly show high errors due to bad sampling in regions with high variability as retrieval errors have to be measured through comparisons to other external reference data. However, retrieval errors indirectly show up in the error maps as for instance over desert surfaces where ATOVS temperature and mixing ratio retrievals are hindered by the not well known surface emissivity and the poor knowledge of the actual surface temperature and its correlation to the air temperature in the lower atmosphere. Additionally, the method has the potential to include retrieval errors in its processing if those are for instance known as function of the TPW or the spatial distribution of errors is given.

References

- Lindau, R. and J. Schulz, 2004: Gridding/merging techniques for the humidity composite product of the CM-SAF. Proceedings of the 2004 EUMETSAT meteorological Satellite Conference, Prague, Czech Republic, EUM P41, 519-526.
- Lindau, R., 2003. Errors of Atlantic air-sea fluxes derived from ship observations. *J. Climate*, **16**, 783-788.
- Reale, Anthony L., 2003: NOAA Operational Sounding Products for Advanced-TOVS, ITSC XIII Proceedings, Sainte Adèle, Canada, 29 October 2003 – 4 November 2003.
- Von Storch, H. and F. W. Zwiers, 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.