Hydrol. Earth Syst. Sci. Discuss., 12, 363–404, 2015
www.hydrol-earth-syst-sci-discuss.net/12/363/2015/
doi:10.5194/hessd-12-363-2015

Hydrology and Earth System Sciences
Open Access
Discussions

Water vapor mapping by fusing InSAR and GNSS remote sensing data and atmospheric simulations

F. Alshawaf[1,3], B. Fersch[2], S. Hinz[1], H. Kunstmann[2], M. Mayer[3], and F. J. Meyer[4]

[1]Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany
[2]Institute of Meteorology and Climate Research, Campus Alpin, KIT, 82467 Garmisch-Partenkirchen, Germany
[3]Geodetic Institute, KIT, 76131 Karlsruhe, Germany
[4]Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK 99775 USA

Received: 19 November 2014 – Accepted: 16 December 2014 – Published: 12 January 2015

Correspondence to: F. Alshawaf (fadwa.alshawaf@kit.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Data fusion aims at integrating multiple data sources that can be redundant or complementary to produce complete, accurate information of the parameter of interest. In this work, data fusion of precipitable water vapor (PWV) estimated from remote sensing observations and data from the Weather Research and Forecasting (WRF) modeling system is applied to provide complete, accurate grids of PWV. Our goal is to infer spatially continuous, precise grids of PWV from heterogeneous data sets. This is done by a geostatistical data fusion approach based on the method of fixed-rank kriging. The first data set contains absolute maps of atmospheric water vapor produced by combining observations from Global Navigation Satellite Systems (GNSS) and Interferometric Synthetic Aperture Radar (InSAR). These PWV maps have a high spatial density and an accuracy of submillimeter; however, data are missing in regions of low coherence (e.g., forests and vegetated areas). The PWV maps simulated by the WRF model represent the second data set. The model maps are available for wide areas, but they have a coarse spatial resolution and a yet limited accuracy. The PWV maps inferred by the data fusion at any spatial resolution are more accurate than those inferred from single data sets. In addition, using the fixed-rank kriging method, the computational burden is significantly lower than that for ordinary kriging.

1 Introduction

Water vapor is a vital constituent of the Earth's electrically neutral atmosphere (neutrosphere). Although its average volumetric concentration does not exceed 4 %, the impact of water vapor is important in many respects. Due to the dynamic nature of the neutrosphere and the complex energy exchange with the Earth's surface, the spatio-temporal distribution of water vapor can be highly variable. Accurate information about its content and tendency is the main prerequisite for a precise prediction of clouds and precipitation. Water vapor is important for studies of climate and natural disasters such

HESSD
12, 363–404, 2015

as floods, droughts, deluge or glacier melting. On the other hand, signals transmitted from spaceborne sensors are refracted when traversing the Earth's neutrosphere. Neutrospheric water vapor contributes to less than 10 % of the signal path delay; however, this error source is not easily eliminated. Accurate information about the water vapor concentration along the signal path is required, which is not always obtainable. Although many efforts have been made to produce accurate information about water vapor using ground-based, space-based or numerical methods, the available information is often limited in the temporal resolution, spatial resolution or accuracy (Bevis et al., 1992). In this paper, we present a method to produce water vapor maps of a high spatial resolution by fusing water vapor estimates from space-borne signals and atmospheric prediction models.

The amount of remote sensing data available for monitoring the Earth and its atmosphere is growing in a rapid, continuous way. Interferometric Synthetic Aperture Radar (InSAR) has proved its capability for detecting surface deformation, landslides, tectonic movements (Massonnet et al., 1993; Zebker et al., 1994) and for deriving digital elevation models (Zebker and Goldstein, 1986). The influence of water vapor in the observations can be reduced by averaging a large number of interferograms (Zebker et al., 1997) or by time series analysis that indicates the stable persistent scatterers (Ferretti et al., 2001; Hooper et al., 2007). Besides, InSAR has recently been used to derive the atmospheric phase from the interferograms or by time series analysis (Hanssen, 2001; Meyer et al., 2008; Pichelli et al., 2010; Alshawaf et al., 2012). Global Navigation Satellite Systems (GNSS), however, have been considered since the 1990s as an efficient microwave-based tool for atmospheric sounding (Bevis et al., 1992; Rocken et al., 1995). Since then, numerous methods exploited the GNSS observations to produce estimates of the integrated atmospheric water vapor, and for generating water vapor maps (Luo et al., 2008; Jade and Vijayan, 2008; Karabatić et al., 2011). InSAR and GNSS signals are affected in a similar way by the atmosphere (Onn and Zebker, 2006). Therefore, Alshawaf (2013) presented a new approach to derive absolute, high-

HESSD
12, 363–404, 2015

resolution maps of precipitable water vapor (PWV) by combining data from InSAR and GNSS.

Atmospheric modeling systems are standard approaches to simulate three-dimensional distributions of the neutrospheric water vapor at arbitrary temporal and spatial sampling. Dynamic local area models (LAM) are common tools for scaling down the coarse grids of global circulation models to meso-scale applicability. Several studies employed the Weather Research and Forecasting modeling system (WRF, Skamarock and Klemp, 2008) to compare the LAM simulations of PWV with GNSS point estimates (Mateus et al., 2010; Bender et al., 2008; Cimini et al., 2012) and PWV maps from MERIS (MEdium Resolution Imaging Spectrometer) (Alshawaf et al., 2012). These studies conclude that the medium to long scale water vapor signals are well predicted, whereas short scale fluctuations are often hardly captured in a realistic way. The presence of convective motion or rapid dynamic effects are still a challenge for the model performance. Hence, the model data can be considerably biased with respect to the actual state of the atmosphere. In addition, WRF combines a large number of alternative models for the representation of individual physical compartments. This, in addition to the configurations of the model domains, highly affects the quality of the model simulations (Gong et al., 2010; Fersch and Kunstmann, 2014) as well as the model's intrinsic water balance (Fersch et al., 2012; Fersch and Kunstmann, 2014).

Due to the availability of various data sources that can be complementary or redundant, data fusion has received increasing attention in the Earth observation studies. The focus is put on the combination of multiple sources, which may be spatially, temporally, or spectrally inhomogeneous, to produce a more complete representation of a geophysical process. In this work, we use remote sensing data and numerical atmospheric models through a data fusion approach to provide improved information about the distribution of atmospheric water vapor. This information is important not only for weather forecasting and climate research, but also to better understand how the InSAR interferograms are affected by water vapor, and which is the most appropriate method

HESSD
12, 363–404, 2015

to reduce this noise. In turn, reliable local water vapor maps can support adapting the WRF model configurations and, hence, may improve the model performance.

In the following, we present water vapor maps derived from microwave remote sensing data and numerical atmospheric models. Since the available data have different spatial level of aggregation, it is important to discuss the change of support problem. Then, we present the data fusion approach based on the kriging technique or fixed-rank kriging. We first describe the ordinary kriging and how it can be extended for fusing multiple data sets. Then, we present the reasons behind using the fixed-rank kriging. We use the data fusion approach for predicting maps of the atmospheric PWV from remote sensing data and atmospheric models.

2 Atmospheric water vapor

Several observation systems are used to continuously monitor the vertical and horizontal distributions of water vapor in the atmosphere. These devices are used either from the ground, such as radiosondes and ground-based water vapor radiometers, or from space, such as space-based water vapor radiometers and infrared sensors. In this work, we employ microwave remote sensing systems as well as numerical atmospheric models to provide accurate maps of the atmospheric water vapor at a high spatial resolution.

2.1 Water vapor from remote sensing data

Alshawaf (2013) presented a new approach to derive absolute, high-resolution maps of PWV by combining data from InSAR and GNSS. The data are collected in the region of Upper Rhine Graben in Germany and France over the period 2003–2008. Persistent Scatterer InSAR (PSI) using the Stanford Method for Persistent Scatterers (StaMPS, Hooper et al., 2007) was applied to derive PWV maps from the InSAR interferograms. These maps contain the water vapor signal of short scale spatial variations, while

HESSD
12, 363–404, 2015

the elevation-dependent and long wavelength water vapor components are eliminated when forming the interferograms or phase filtering. Therefore, GNSS-based PWV estimates were used to reconstruct the missing components. Figure 1 shows a map of PWV derived by combining PSI and GNSS data and the corresponding map extracted from MERIS observations. The spatial correlation between the maps is 95 % and the root mean square (RMS) value of the differences is 0.68 mm. We can observe that the persistent scatterers are dense in the urban areas, while they almost disappear in the low coherence regions.

Since PWV data are spatial, their covariance function is exploited by geostatistical techniques to reasonably infer the PWV at regular grids. In order to improve the inferred PWV maps, especially in the areas where the PWV estimates are sparse, we apply data fusion of the remotely-sensed PWV maps with maps produced by the WRF model.

2.2 Water vapor from regional atmospheric models

As depicted in Fig. 2, the WRF model (version 3.1.1, Skamarock et al., 2008) was set up with a parent domain of 27 km × 27 km resolution and two nests with 9 km × 9 km and 3 km × 3 km, respectively. Vertically, the model divides into 42 layers and the model top is defined at 50 hPa. The selection of the physical modules is based on the study of Berg et al. (2013); accordingly, the WRF single moment (WSM) 5-class scheme (Hong et al., 2004) was selected for microphysics. Short and longwave radiation were computed with the community atmospheric model (CAM) scheme (Collins et al., 2004). The processes in the planetary boundary layer were represented by the Yonsai University scheme (Hong et al., 2006). The surface layer was simulated with the Monin–Obukhov scheme, and the Noah land-surface-model (Chen and Dudhia, 2001) was applied for the surface physics. Sub-grid convective processes were included with the Kain-Fritsch parametrization (Kain, 2004). The global dynamic boundary conditions were ingested from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-INTERIM reanalysis at 6 h interval (Uppala et al., 2008).

HESSD
12, 363–404, 2015

The WRF simulations cover the period between July 2004 and September 2005, such that the first 5 months were considered as spin-up. The PWV content was determined at every output time-step (10 min) by a vertical integration of all moisture fields from the land surface to the model top. Two output time slices were compared with the simultaneous MERIS observations. From the compared maps shown in Fig. 3, we observe that the spatial atmospheric patterns are not always correctly resembled by the model. On 27 June 2005, WRF and MERIS PWV maps are strongly correlated with a coefficient of 0.8, whereas the analysis of 5 September 2005 show a lower spatial correlation (0.71). While the patterns in the east of the Upper Rhine valley are reasonably resembled, an unexpected discontinuity exists in the area around 7.7° E, 48.7° N.

3 Change of support problem

Spatial data, for which close observations correlate more than distant ones, can be collected at points or areal units. The former are called point-level data or simply point data and the latter are areal-level or block data (Gelfand et al., 2001). In geostatistics, this defines the spatial support of the data. When both data types are available, data fusion can be applied to infer the underlying process at any level of support. The change of support problem is concerned with the inference of the underlying process at point- or block-levels different from those at which the data are available. This also includes fusing data at different support levels. Based on the available input data and the desired output grid, there are four prediction possibilities: points to points, points to blocks, blocks to points, or blocks to blocks. These prediction possibilities may be collected under the umbrella of kriging (Cressie, 1993).

For block data that can be expressed as an average of point data as if it is collected within the block, such as rainfall, temperature, surface elevation, and atmospheric water

HESSD
12, 363–404, 2015

vapor, the following model is appropriate

$$Y(B_i) = \frac{1}{|B_i|}\int_{B_i} Y(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \tag{1}$$

where $Y(B_i)$ and $Y(\boldsymbol{s})$ define the block and point data, respectively (Fig. 4). B_i refers to the block over which the data are aggregated and $|B_i|$ is the volume (or cardinality) of the data. The block-level covariance can then be related to the point-level covariance as follows:

$$C(B_i, B_j) = \mathrm{cov}\left(\frac{1}{|B_i|}\int_{B_i} Y(\boldsymbol{u})\mathrm{d}\boldsymbol{u}, \frac{1}{|B_j|}\int_{B_j} Y(\boldsymbol{v})\mathrm{d}\boldsymbol{v}\right) \tag{2}$$

$$= \frac{1}{|B_i||B_j|}\int_{B_i}\int_{B_j} C(\boldsymbol{u},\boldsymbol{v})\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v} \tag{3}$$

where $C(B_i, B_j)$ is the block-to-block or block covariance function and $C(\boldsymbol{u},\boldsymbol{v})$ is the point covariance function.

4 Spatial data fusion using kriging methods

4.1 Ordinary kriging

In geostatistics, a spatial process can be inferred over a continuous spatial domain by exploiting the covariance function as an important source of information. Predictions are obtained based either on single or multiple sets. Kriging is a geostatistical interpolation technique that infers values at new locations by considering spatial correlations (Cressie, 1993). If the considered spatial data set is denoted by Z, the kriging equations are presented by Cressie and Johannesson (2008) as follows. Each element of

HESSD
12, 363–404, 2015

Z is given by:

$$Z(\boldsymbol{s}) = Y(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) \tag{4}$$

where $Y(\boldsymbol{s})$ is the spatial process and $\epsilon(\boldsymbol{s})$ is an independent error term, which is assumed to be a white noise process with a mean of zero and variance σ^2_ϵ. For ordinary kriging (OK), $Y(\cdot)$ is expressed using the following linear model:

$$Y(\boldsymbol{s}) = \mathbf{T}(\boldsymbol{s}) \cdot \boldsymbol{\alpha} + \nu(\boldsymbol{s}) \tag{5}$$

where $\mathbf{T}(\boldsymbol{s})\cdot\boldsymbol{\alpha}$ defines a deterministic linear trend, $\mathbf{T}$ has a size of $N\times 3$ and each row has the following entries: [1 longitude($\boldsymbol{s}$) latitude($\boldsymbol{s}$)]. N is the number of observations and $N = 1$ in Eq. (5), $\boldsymbol{\alpha}$ is a vector of the least squares regression coefficient. $\nu(\boldsymbol{s})$ captures the spatial covariance structure of the process, and it is assumed to have a mean zero and generally a non-stationary covariance function. If the kriging technique is used to infer the signal at a new location, it is required to center the data by estimating and subtracting the linear trend, i.e.,

$$\widetilde{Z} = Z - \mathbf{T}\hat{\boldsymbol{\alpha}} \quad \text{with} \quad \hat{\boldsymbol{\alpha}} = (\mathbf{T}\mathbf{T}')^{-1}\mathbf{T}'Z \tag{6}$$

The kriging estimator $\hat{Y}(\boldsymbol{s}_0)$ at the location $\boldsymbol{s}_0$ is then determined as follows:

$$\hat{Y}(\boldsymbol{s}_0) = \boldsymbol{a}'\widetilde{Z} \tag{7}$$

where the vector $\boldsymbol{a}$ contains the kriging weighting coefficients. The best linear unbiased estimator is found by solving the following constrained minimization problem:

$$\min_{\boldsymbol{a}} \quad \mathrm{E}\left\{(\hat{Y}(\boldsymbol{s}) - Y(\boldsymbol{s}))^2\right\} \quad \text{subject to} \quad \mathrm{E}\{\hat{Y}(\boldsymbol{s})\} = \mathrm{E}\{Y(\boldsymbol{s})\} \tag{8}$$

The constraint is added to guarantee that the estimator is unbiased with respect to the true process. Note that the deterministic signal is calculated from $T(\boldsymbol{s}_0)\hat{\boldsymbol{\alpha}}$ and added to the estimate in Eq. (7) to get the total estimated value of $Y(\boldsymbol{s}_0)$.

HESSD
12, 363–404, 2015

In the next section, a similar strategy is followed to solve for the best unbiased estimator using two data sets as presented in Braverman et al. (2009).

4.2 Spatial statistical data fusion

The spatial statistical data fusion (SSDF) is a method that statistically combines two data sets to optimally infer the quantity of interest and calculate the corresponding uncertainties at any predefined grid (Nguyen, 2009; Braverman et al., 2009). This method extends the kriging technique described above to find the optimal estimator using multiple data sets. Let the underlying process $Y(\boldsymbol{s})$ to be estimated at the location $\boldsymbol{s}$ from the data in Z_1 and Z_2 with the size N_1 and N_2, respectively. The estimator $\hat{Y}(\boldsymbol{s})$ at the location $\boldsymbol{s}$ is obtained from the two data sets as follows:

$$\hat{Y}(\boldsymbol{s}) = \boldsymbol{a}_1' \widetilde{Z}_1 + \boldsymbol{a}_2' \widetilde{Z}_2 \tag{9}$$

where $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are the fusion weighting coefficients, and $\widetilde{Z}_1$ and $\widetilde{Z}_2$ are detrended data sets of Z_1 and Z_2, respectively. Following Eq. (8) and Eq. (9), the Lagrangian function L for the minimization problem under the unbiasedness constraint is

$$\begin{aligned} L = {} & \boldsymbol{a}_1' \boldsymbol{\Sigma}_{11} \boldsymbol{a}_1 + \boldsymbol{a}_2' \boldsymbol{\Sigma}_{22} \boldsymbol{a}_2 + 2\boldsymbol{a}_1' \boldsymbol{\Sigma}_{12} \boldsymbol{a}_2' - 2\boldsymbol{a}_1' \boldsymbol{c}_1 - 2\boldsymbol{a}_2' \boldsymbol{c}_2 \\ & + 2m(\boldsymbol{a}_1' \mathbf{1}_{N_1} + \boldsymbol{a}_2' \mathbf{1}_{N_2} - 1) \end{aligned} \tag{10}$$

where $\boldsymbol{\Sigma}_{ii} = \mathrm{cov}(\widetilde{Z}_i)$, $\boldsymbol{\Sigma}_{ij} = \mathrm{cov}(\widetilde{Z}_i, \widetilde{Z}_j)$, and $\boldsymbol{c}_i = \mathrm{cov}(\widetilde{Z}_i, Y(\boldsymbol{s}))$ are the covariance functions. $\mathbf{1}_{N_i}$ is a vector with all entries one and a length N_i, and m denotes the Lagrange multiplier. The last term of L accounts for the unbiasedness constraint. By differentiating L with respect to $\boldsymbol{a}_1, \boldsymbol{a}_2, m$ and assigning the results to zero, we get in the following system of equations:

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \mathbf{1}_{N_1} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \mathbf{1}_{N_2} \\ \mathbf{1}_{N_1}' & \mathbf{1}_{N_2}' & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ m \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ 1 \end{bmatrix} \tag{11}$$

HESSD
12, 363–404, 2015

and hence

$$\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ m \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \mathbf{1}_{N_1} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \mathbf{1}_{N_2} \\ \mathbf{1}'_{N_1} & \mathbf{1}'_{N_2} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ 1 \end{bmatrix} \tag{12}$$

There are several important discussion points for the solution in Eq. (12). The covariance matrices $\boldsymbol{\Sigma}_{ij}$ are determined without assuming that the underlying process is isotropic or stationary. This is important for atmospheric parameters, particularly the atmospheric water vapor that shows spatial anisotropy as observed from the spatial autocorrelation function in Fig. 5. The covariance function $\boldsymbol{c}_i$ should account for the change of the support between the input and the output data. For massive data sets, the the size of the covariance matrix is huge and the solution in Eq. (12) is not anymore feasible. Also, the covariance matrices should be modeled such that they would allow data prediction to any level of aggregation. The Fixed-rank kriging covariance model suggested by Cressie and Johannesson (2008) provides a comprehensive solution for these problems for single data sets and the generalized model for fusing multiple data sets was presented by Nguyen (2009) and Braverman et al. (2009). In the next section, we describe the Fixed-rank kriging method and the associated covariance model. Then, we describe how the data fusion approach is applied to our data sets.

4.3 Fixed-rank kriging

The Fixed-rank kriging (FRK) approach splits the spatial process into two or three components depending on the roughness of the spatial variations, i.e,

$$Y(\boldsymbol{s}) = \underbrace{\mathbf{T}(\boldsymbol{s}) \cdot \boldsymbol{\alpha}}_{\text{linear trend}} + \underbrace{\mathbf{S}(\boldsymbol{s}) \cdot \boldsymbol{\eta} + \zeta(\boldsymbol{s})}_{\nu(\boldsymbol{s})} \tag{13}$$

The model in Eq. (13) is called the spatial random effects (SRE) model (Cressie and Johannesson, 2008). The first component represents a deterministic linear trend that

HESSD
12, 363–404, 2015

reflects the large scale spatial variations. The second component $\mathbf{S}(\boldsymbol{s})\cdot\boldsymbol{\eta}$ captures the relatively smooth spatial variations, which form the covariance structure of the process. That is, $\mathrm{cov}(\mathbf{S}(\boldsymbol{u})\cdot\boldsymbol{\eta},\mathbf{S}(\boldsymbol{v})\cdot\boldsymbol{\eta}) = \mathbf{S}(\boldsymbol{u})\mathbf{K}\mathbf{S}'(\boldsymbol{v})$ with $\mathbf{K}$ the covariance function of $\boldsymbol{\eta}$. This component is modeled by a linear combination of spatial random effects at multiple spatial scales. The vector $\boldsymbol{\eta}$ contains r hidden spatial random effects, which are estimated from the data at predefined nodes. Therefore, we should be able to estimate $\boldsymbol{\eta}$ regardless of the aggregation level of the input data. When neglecting the last term in Eq. (13), the weighted sum $\sum_{j=1}^{r} S_j(\boldsymbol{s})\eta_j$ should give the detrended value of Y at the location $\boldsymbol{s}$.

The weights stored in the matrix $\mathbf{S}$ for each location $\boldsymbol{s}$ depend on the distance between $\boldsymbol{s}$ and each node. The weighting function $S(\boldsymbol{s})$ has the following form:

$$S(\boldsymbol{s}) = \begin{cases} \left[1 - (\|\boldsymbol{s}-\boldsymbol{m}_i\|/r_i)^2\right]^2, & \text{for } \|\boldsymbol{s}-\boldsymbol{m}_i\| \le r_i, \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$\boldsymbol{m}_i$ is the node location, and r_i is a predefined effective radius. The formula in Eq. (14) represents a bisquare bell-shaped function that has its maximum value at $\boldsymbol{m}_i$ and decreases smoothly until it reaches zero outside the circle. To demonstrate, a schematic diagram for the nodes setup is shown in Fig. 6. Within the domain of the data, four nodes, $\boldsymbol{m}_1,\cdots,\boldsymbol{m}_4$, are defined with a corresponding radius. In Fig. 6, if $\boldsymbol{s}$ is located within the radius of a certain node, it gets a positive weight, otherwise the weight is zero. Hence, $\mathbf{S}(\boldsymbol{s}) = [0, 0, 0, S(\boldsymbol{s})]$.

The last component in Eq. (13) accounts for the roughness of the process that has not been captured so far (Kang and Cressie, 2011). The component ζ is assumed to be an uncorrelated Gaussian process with a mean zero and a variance σ^2_ζ.

Based on the model in Eq. (13), the FRK estimator is found when $\boldsymbol{\eta}$ and ζ are determined, i.e.,

$$\begin{aligned} \hat{Y}(\boldsymbol{s}_o) &= \boldsymbol{S}_p(\boldsymbol{s}_o)\cdot\hat{\boldsymbol{\eta}} + \hat{\zeta}(\boldsymbol{s}_o) \\ &= \boldsymbol{S}_p(\boldsymbol{s}_o)\mathbf{K}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\widetilde{Z} + \sigma^2_\zeta\mathbf{E}(\boldsymbol{s}_o = \boldsymbol{s})\boldsymbol{\Sigma}^{-1}\widetilde{Z} \end{aligned} \tag{15}$$

HESSD
12, 363–404, 2015

where $\boldsymbol{S}_p(\boldsymbol{s}_o)$ is the weighting matrix for the prediction location and $\boldsymbol{\Sigma}$ is the covariance matrix of the input data. The matrix **E** in Eq. (15) has a value of one if $\boldsymbol{s} = \boldsymbol{s}_o$ and zero elsewhere. $\hat{Y}$ represents the detrended estimator. $\hat{\boldsymbol{\eta}}$ and $\hat{\zeta}$ are the optimal a posteriori estimates of $\boldsymbol{\eta}$ and ζ, respectively (Braverman et al., 2011). In order to get the total value of $\hat{Y}_t$, we calculate

$$\hat{Y}_t(\boldsymbol{s}_o) = \boldsymbol{T}(\boldsymbol{s}_o) \cdot \hat{\boldsymbol{\alpha}} + \hat{Y}(\boldsymbol{s}_o) \tag{16}$$

The steps followed to obtain the predictions based on the FRK method are summarized in Fig. 7. The methods to estimate the noise variance σ^2_ϵ, the covariance matrix **K**, and the variance of the fine-scale signal σ^2_ζ are shown in Appendix A.

We classify the spatial variations of the atmospheric water vapor signal into three components: long wavelength, medium to short wavelength, and uncorrelated fine scale. Therefore, we express the water vapor signal using the linear model in Eq. (13) and use the FRK method for prediction.

We applied the OK and FRK to estimate the zenith-directed wet delay derived from remote sensing data. For the FRK, the matrix **S** is constructed using the node setup shown in Fig. 8. The nodes or center locations of 93 basis functions are established at three spatial resolutions, the first resolution is 40 km, the second resolution is 20 km, and the third resolution is 10 km. The results are shown in Fig. 9; the map has a resolution of 3 km × 3 km. Due to the lack of ground truth data that should be used to estimate the bias in the model data, we do not add the long-wavelength component into the figures to enable unbiased comparison. We observe similar results from both ordinary kriging and fixed-rank kriging that agree with the original WRF map. The spatial correlation coefficients with the WRF data are approximately 85 and 83 % for FRK and OK, respectively. For using OK, we assumed the signal spatially isotropic to ease the computations; therefore, the OK prediction map shows results sightly different from the FRK. The most impressive point here is the computational time reported for both algorithms. The FRK algorithm is fast, so that it requires significantly shorter time to produce the predictions. Most of the time is invested in the calculations of the covari-

HESSD
12, 363–404, 2015

ance model parameters and constructing the matrices $\mathbf{S}$ and $\boldsymbol{\Sigma}$. On the other hand, we implemented the OK algorithm such that the predictions are found iteratively. Also, to estimate a value at the location $\boldsymbol{s}$, we do not use the entire data, but only those which exist within a predefined radius around the prediction location. Nevertheless, the OK algorithm requires computational time with an order of magnitude higher than that required by the FRK method using the same machine.

In the next section, we describe the extension of the FRK method for predicting the atmospheric PWV by fusing remote sensing data and the WRF model.

5 Data fusion for water vapor estimation

In this section, we fuse the PWV maps derived from the remote sensing data and WRF model. Since we classify the spatial variations of the atmospheric water vapor signal into long wavelength, medium to short wavelength, and uncorrelated fine scale components, we use the following model setup for prediction.

5.1 Model setup

PWV maps will be derived from the remote sensing data, denoted Z_1, and those from the WRF model, denoted Z_2 with the size N_1 and N_2, respectively. Z_1 contains the point PWV estimates from remote sensing data and Z_2 contains the block WRF data. Following the SME model in Eq. (13), the two data sets can be expressed as

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_1 \\ \overline{\mathbf{T}}_2 \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{S}_1 \\ \overline{\mathbf{S}}_2 \end{bmatrix} \boldsymbol{\eta} + \begin{bmatrix} \boldsymbol{\zeta}_1 \\ 0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{bmatrix} \tag{17}$$

The regression coefficient $\boldsymbol{\alpha}$ should be estimated jointly from both data sets. However, we do not have apriori information about the biases; therefore, we estimate $\boldsymbol{\alpha}$ in this contribution independently for each data set. The matrices $\mathbf{S}_1$ and $\overline{\mathbf{S}}_2$ contain the weights of each location for each data set. To distinguish between point and block

HESSD
12, 363–404, 2015

data, we used the notation $\overline{\mathbf{S}}_2$ for block-level data. The model components for point and block data are given in Table 1. The WRF data are available at a resolution of 3 km × 3 km; therefore, the highly variable signal of water vapor is smoothed. Hence, we do not consider the component $\boldsymbol{\zeta}$ for the model data.

To solve the system in Eq. (12), we determine the covariance structure associated with each SRE model in Eq. (17), i.e.,

$$\boldsymbol{\Sigma}_{11} = \mathrm{var}(\widetilde{Z}_1) = \mathbf{S}_1\mathbf{K}\mathbf{S}_1' + \sigma_\zeta^2\mathbf{V}_\zeta + \sigma_{\epsilon_1}^2\mathbf{V}_{\epsilon_1} \tag{18}$$

$$\boldsymbol{\Sigma}_{22} = \mathrm{var}(\widetilde{Z}_2) = \widetilde{\mathbf{S}}_2\mathbf{K}\widetilde{\mathbf{S}}_2' + \sigma_{\epsilon_2}^2\mathbf{V}_{\epsilon_2} \tag{19}$$

$$\boldsymbol{\Sigma}_{12} = \mathrm{cov}(\widetilde{Z}_1, \widetilde{Z}_2) = \mathbf{S}_1\mathbf{K}\widetilde{\mathbf{S}}_2' = \boldsymbol{\Sigma}_{21}' \tag{20}$$

where $\sigma_\zeta^2\mathbf{V}_\zeta$ and $\sigma_\epsilon^2\mathbf{V}_\epsilon$ are diagonal covariance matrices for ζ and ϵ, respectively. Note that when computing the cross covariance functions $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{21}$, the only part of the signals that is assumed correlated is $\boldsymbol{\eta}$. In order to solve Eq. (12), we need not only to specify the covariance matrices of the input data, but also to find the covariance between the observations and the spatial process at the prediction locations. The covariance terms are obtained from:

$$\begin{aligned}\boldsymbol{c}_1 &= \mathrm{cov}(\widetilde{Z}_1(\boldsymbol{s}), Y(\boldsymbol{s}_o)) \\ &= \boldsymbol{S}_p(\boldsymbol{s}_o)\mathbf{K}\mathbf{S}_1'(\boldsymbol{s}) + \sigma_\zeta^2\mathbf{E}(\boldsymbol{s} = \boldsymbol{s}_o)\end{aligned} \tag{21}$$

$$\boldsymbol{c}_2 = \mathrm{cov}(\widetilde{Z}_2, Y(\boldsymbol{s}_o)) = \boldsymbol{S}_p(\boldsymbol{s}_o)\mathbf{K}\widetilde{\mathbf{S}}_2' \tag{22}$$

The matrix $\mathbf{E}$ in Eq. (21) has a value of one if $\boldsymbol{s} = \boldsymbol{s}_o$ and zero elsewhere. By solving for $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in Eq. (12) and substituting the results in Eq. (9), the estimator $\hat{Y}(\boldsymbol{s}_o)$ becomes

$$\hat{Y}(\boldsymbol{s}_o) = \left(\boldsymbol{S}_p(\boldsymbol{s}_o)\mathbf{K}\begin{bmatrix}\mathbf{S}_1' \\ \widetilde{\mathbf{S}}_2'\end{bmatrix} + \begin{bmatrix}\sigma_\zeta^2\mathbf{E} \\ 0\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}^{-1}\begin{bmatrix}\widetilde{Z}_1 \\ \widetilde{Z}_2\end{bmatrix} \tag{23}$$

HESSD
12, 363–404, 2015

The mean squared prediction error (MSPE) corresponding to $\hat{Y}$ can be obtained from

$$\text{MSPE} = \boldsymbol{a}_1'\boldsymbol{\Sigma}_{11}\boldsymbol{a}_1 + \boldsymbol{a}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{a}_2 + 2\boldsymbol{a}_1'\boldsymbol{\Sigma}_{12}\boldsymbol{a}_2 - 2\boldsymbol{a}_1'\boldsymbol{c}_1 - 2\boldsymbol{a}_2'\boldsymbol{c}_2 \tag{24}$$

Using the FRK covariance model in Eq. (20) makes the matrix inversion of Eq. (23) scalable. That is, the matrix inversion can be achieved by applying a recursive block-wise inversion as follows:

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{O}_1 & \mathbf{O}_2\\ \mathbf{O}_3 & \mathbf{O}_4\end{bmatrix} \tag{25}$$

where

$$\mathbf{O}_1 = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$$
$$\mathbf{O}_2 = -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$$
$$\mathbf{O}_3 = -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$$
$$\mathbf{O}_4 = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$$

and $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are matrices of any size, and $\mathbf{A}, \mathbf{D}$ must be square. The inversion of individual matrices in Eq. (25) is achieved by applying the formula of Sherman–Morrison–Woodbury, which is made possible due to the FRK covariance structure,

$$\begin{aligned}\boldsymbol{\Sigma}_{ii}^{-1} &= (\mathbf{D}_i + \mathbf{S}_i\mathbf{K}_i\mathbf{S}_i')^{-1}\\ &= \mathbf{D}_i^{-1} - \mathbf{D}_i^{-1}\mathbf{S}_i(\mathbf{K}^{-1} + \mathbf{S}_i'\mathbf{D}_i^{-1}\mathbf{S}_i)^{-1}\mathbf{S}_i'\mathbf{D}_i^{-1}\end{aligned} \tag{26}$$

The computations require the inversion of the matrices $\mathbf{K}$ and $(\mathbf{K}^{-1} + \mathbf{S}_i'\mathbf{D}_i^{-1}\mathbf{S}_i)$, where each of them has the size $r \times r$ with r significantly smaller than the data size. Note that $\mathbf{D}_i$ is a diagonal matrix, for which the inversion is achieved by inverting the diagonal elements. Using the FRK covariance model makes the computational burden for the matrix inversion linear with the data size (Cressie and Johannesson, 2008).

HESSD
12, 363–404, 2015

5.2 Results and discussion

In this section, we build PWV maps by fusing remote sensing and model data using the spatial statistical data fusion method. The first PWV map, derived by combining GNSS and PSI, has 169 688 points. The WRF model provides a block-level map of 1296 cells of the size 3 km × 3 km. The data to be fused have different qualities, huge size, different spatial support, and gaps in the remote sensing data. The output maps are compared with the MERIS PWV maps.

Following the work flow in Fig. 7, we first estimate the long wavelength trends and remove them from the data using Eq. (6). By comparing the PWV from the WRF model and remote sensing data, we found it is most likely that the model data have a bias. Due to the lack of apriori information about the bias and accurate ground truth data, we estimated $\boldsymbol{\alpha}$ independently for each data set. The centered maps are shown in Fig. 10.

The matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ are constructed for the first data set (remote sensing data) and the second data set (model data). The node setup is shown in Fig. 8. The number of nodes must be the same for both data sets and they are selected such that $\mathbf{S}$ does not contain columns of zeros, otherwise the corresponding node has to be removed. When PWV is available at point-level, a weighting value is calculated for each point with respect to all nodes. However, the WRF model simulates block-level data, hence we superimpose the model grid with a lattice of regular points such that each cell in the WRF grid contains 9 points. A weighting value is calculated for each point and the values are averaged to get a weighing value for each WRF cell to form the matrix $\overline{\mathbf{S}}_2$. Building the matrix $\boldsymbol{S}_p$ for the prediction locations is done in a similar way, either at point-level or block-level, depending on the output grid.

Next, the covariance parameters $(\mathbf{K}, \sigma_\zeta^2, \sigma_\epsilon^2)$ are estimated from the centered data $\widetilde{Z}_1$ and $\widetilde{Z}_2$. The error variances for both data sets, $\mathbf{K}$ and σ_ζ^2 are estimated as described in Appendix A. Note that when the two data sets are combined to infer a single process, i.e., PWV, one $\mathbf{K}$ is estimated for all data sets.

HESSD
12, 363–404, 2015

The output grid is defined at 3 km × 3 km (block-level support). So far, all components required in Eq. (23) are obtained. In Fig. 11, we show the prediction maps obtained by applying FRK to individual data sets as well as the map obtained by data fusion. The figure also shows the MSPE maps associated with each prediction map. We compare the interpolations obtained by applying FRK to single data sets with those obtained by SSDF and we compare both with the MERIS data. The results show that the map obtained by data fusion correlates more consistently with the map predicted only from PSI + GNSS (Table 2). In the PWV map generated by WRF, shown in Fig. 10, the area in the lower left corner shows artifacts that do not reflect the correct values of PWV as observed from the MERIS PWV map, Fig. 3c and d. Applying FRK to the WRF data does not remove these artifacts from the prediction map. However, in the map obtained by the fusion of both data sets, the artifacts in the lower corner disappeared, but the corresponding MSPE values are large for this region. The MSPE values corresponding to the SSDF predictions are generally smaller, and we should note that in the regions of sparse observations, the corresponding MSPE values tend to increase. For regions of sparse observations in the PWV map (Fig. 10), i.e., the areas in the west of the Rhine valley or in the lower right corner, the map from the WRF model contributes to improve the estimation of the PWV values in the prediction map. The region in the lower right corner has a higher topography and the wet delay values are expected to decrease as we observe from the map of WRF. In the prediction map obtained by applying FRK to PWV from PSI and GNSS, the predicted values tend to increase since the data in this area are sparse and partially biased. By applying the SSDF approach, the data available from WRF influence the predictions such that the PWV values in this area are more reasonable and they decrease by moving to the lower right corner. In a similar way, the data from WRF improve the predictions in the region around (7.8° E, 49.25° N), where only sparse PWV data exist. The data from the model, however, affects the prediction in the lower left corner such that they are smaller than those observed in the MERIS map.

HESSD
12, 363–404, 2015

In addition, we show the PWV profiles over the line drawn horizontally at the latitude 49.37° N in Fig. 11 (lower right). It is observed from the plots that the predictions made by data fusion are affected more by the data from WRF in region A, where the remote sensing data are sparse. However, in region B, the WRF data are significantly overestimated. In the prediction map made by data fusion, these data have a lower effect in than those received from the remote sensing data. The map received by applying the data fusion shows the best spatial correlation with the data from MERIS and the smallest RMS value (see Table 2).

6 Conclusions and outlook

We presented a method to obtain the atmospheric PWV over any aggregation level by the fusion of remote sensing data and atmospheric models. The PWV maps derived by combining data from PSI and GNSS are available at discrete points that are absent in regions of low coherence. On the other hand, the WRF model provides simulations of PWV in the atmosphere on regular grids at a coarse spatial resolution. Both the quality of the model data, and the model skills for representing meso-scale atmospheric structures should be improved. For that purpose, the spatial statistical data fusion method, first presented in (Nguyen, 2009), was employed. This method is based on the FRK approach and attempts to solve the problems of computational complexity of huge data sets, change of support, and bias. We inferred PWV data on a grid of 3 km × 3 km and compared the results with maps from MERIS inferred on the same grid. The results show a strong correlation between data fusion maps and those maps from MERIS. The difference between both maps has an RMS value of 0.82 mm, which is lower than that obtained from inferring data based on single sets.

To further improve the results, we suggest the following. So far, the matrix $\mathbf{S}_i$ is constructed for each data source by defining a set of spatial nodes. The number of the nodes is empirically adjusted such that the covariance function computed for the data set based on the estimated matrix $\mathbf{K}$ approximates the empirical covariance. In future

HESSD
12, 363–404, 2015

work, the size and the locations of nodes have to be optimized by minimizing the difference between the empirical and the estimated covariance functions. We should also estimate the biases for each data set (if exist), so that they can be accounted for in the fusion approach. The data fusion approach can be extended such that more than two data sets are used, for example, by including the MERIS maps in the fusion. With the increasing number of satellite missions and improved atmospheric models, we are able to produce complete, accurate information about the Earth's atmosphere based on data fusion approaches. Moreover, the improved PWV maps can be iteratively assimilated to the local area atmospheric model to generate more accurate 3-dimensional water vapor fields. Also, testing other combinations of physical schemes within the WRF model can further improve the resulting water vapor maps.

Appendix A: Estimation covariance parameters

Predicting the stochastic component of the atmospheric signal using kriging requires obtaining the covariance function $\mathbf{\Sigma}$ and fitting a covariance model. Using the FRK covariance model, we need to estimate the matrix $\mathbf{K}$, the noise variance σ^2_ϵ, and the variance of the fine-scale signal σ^2_ζ. The first method proposed to estimate $\mathbf{K}$ is called binned method-of-moments (MM) (Cressie and Johannesson, 2008; Nguyen, 2009). This approach derives the empirical estimator for $\mathbf{\Sigma}$ and obtains $\mathbf{K}$ such that $\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F$ is minimum, where $\|\cdot\|_F$ refers to the Frobenius norm. This method is also described in details in Alshawaf (2013).

Another approach proposed by Katzfuss and Cressie (2009) targets to determine the covariance parameters using the algorithm of maximum likelihood estimation (MLE). Furthermore, they estimated the covariance paramters using the expectation-maximization (E-M) algorithm (Dempster et al., 1977) to reduce the computational burden. This algorithm provides estimates not only of $\mathbf{K}$ but also of σ^2_ζ, where the solution for the MLEs is found iteratively. Within each iteration the algorithm performs two steps, the expectation and maximization. In the following, we present a description of how to

HESSD
12, 363–404, 2015

obtain the maximum likelihood estimates of the covariance model parameters via the E-M algorithm.

Assuming that the observations in $\widetilde{Z}$ follow a multivariate Gaussian distribution, that is $\widetilde{Z} \sim N(0, \mathbf{\Sigma})$. Let the parameters of interest $\mathbf{K}$ and σ_ζ^2 be summarized in the vector $\mathbf{\Theta}$, then the likelihood function $L(\mathbf{\Theta})$ (Katzfuss and Cressie, 2009)

$$\begin{aligned} -2\log L(\mathbf{\Theta}) &= -2f(\widetilde{Z};\mathbf{\Theta}) \\ &= \log\ \det \mathbf{\Sigma} + \widetilde{Z}'\mathbf{\Sigma}^{-1}\widetilde{Z} + c \\ &= \log\ \det \mathbf{\Sigma} + \ \mathrm{tr}\ (\Sigma^{-1}\widetilde{Z}\widetilde{Z}') + c \end{aligned} \tag{A1}$$

where $c = (N/2)\log 2\pi$ is a constant independent of $\mathbf{\Theta}$ and hence it cancels out in the maximization step. tr(·) denotes the trace operator of a square matrix, with $\mathrm{tr}\ (\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$.

In the expectation step of the algorithm, we calculate

$$Q(\mathbf{\Theta};\mathbf{\Theta}^{[t]}) = \mathrm{E}_{\mathbf{\Theta}^{[t]}}\{-2\log L(\boldsymbol{\eta},\boldsymbol{\zeta};\mathbf{\Theta})|\widetilde{Z}\} \tag{A2}$$

given that:

$$\begin{aligned} -2\log L(\boldsymbol{\eta},\boldsymbol{\zeta};\mathbf{\Theta}) = &\log\ \det \mathbf{K} + \ \mathrm{tr}\ (\mathbf{K}^{-1}\boldsymbol{\eta}\boldsymbol{\eta}') + N\log\sigma_\zeta^2 \\ &+ \sigma_\zeta^{-2}\ \mathrm{tr}\ (\boldsymbol{\zeta}\boldsymbol{\zeta}') + N\log\sigma_\epsilon^2 + \sigma_\epsilon^{-2}\ \mathrm{tr}\ (\boldsymbol{\epsilon}\boldsymbol{\epsilon}') \end{aligned}$$

Then Eq. (A2) becomes

$$\begin{aligned} Q(\mathbf{\Theta};\mathbf{\Theta}^{[t]}) = -\frac{1}{2}[&\log\ \det \mathbf{K} + \ \mathrm{tr}\ (\mathbf{K}^{-1}\mathrm{E}_{\mathbf{\Theta}^{[t]}}\{\boldsymbol{\eta}\boldsymbol{\eta}'|\widetilde{Z}\}) \\ &+ N\log\sigma_\zeta^2 + \sigma_\zeta^{-2}\ \mathrm{tr}\ (\mathbf{V}_\zeta^{-1}\mathrm{E}_{\mathbf{\Theta}^{[t]}}\{\boldsymbol{\zeta}\boldsymbol{\zeta}'|\widetilde{Z}\}) \\ &+ N\log\sigma_\epsilon^2 + \sigma_\epsilon^{-2}\ \mathrm{tr}\ (\mathbf{V}_\epsilon^{-1}\mathrm{E}_{\mathbf{\Theta}^{[t]}}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\widetilde{Z}\})] \end{aligned} \tag{A3}$$

HESSD
12, 363–404, 2015

We should remind the reader that the parameters to be estimated here are $\mathbf{K}$ and σ_ζ^2, while σ_ϵ^2 is estimated from the robust semivariogram, as described later. To proceed with the solution, it is required to quantify the conditional expectations in Eq. (A3). Using the standard formula required for calculating conditional expectations for multivariate normal distribution, the expectations will have the following form (Katzfuss and Cressie, 2009)

$$\mathrm{E}_{\boldsymbol{\Theta}^{[t]}}\{\boldsymbol{\eta}\boldsymbol{\eta}'|\widetilde{Z}\} = \boldsymbol{\Sigma}_\eta^{[t]} + \boldsymbol{\mu}_\eta^{[t]}\boldsymbol{\mu}_\eta^{\prime[t]}$$
$$\mathrm{E}_{\boldsymbol{\Theta}^{[t]}}\{\boldsymbol{\zeta}\boldsymbol{\zeta}'|\widetilde{Z}\} = \boldsymbol{\Sigma}_\zeta^{[t]} + \boldsymbol{\mu}_\zeta^{[t]}\boldsymbol{\mu}_\zeta^{\prime[t]}$$

with

$$\boldsymbol{\mu}_\eta{}^{[t]} = \mathrm{E}_{\boldsymbol{\Theta}^{[t]}}\{\boldsymbol{\eta}|\widetilde{Z}\} = \mathbf{K}^{[t]}\mathbf{S}'{\boldsymbol{\Sigma}^t}^{-1}\widetilde{Z}$$
$$\boldsymbol{\mu}_\zeta{}^{[t]} = \mathrm{E}_{\boldsymbol{\Theta}^{[t]}}\{\boldsymbol{\zeta}|\widetilde{Z}\} = \sigma_\zeta^{2[t]}\mathbf{V}_\zeta{\boldsymbol{\Sigma}^{[t]}}^{-1}\widetilde{Z}$$
$$\boldsymbol{\Sigma}_\eta^{[t]} = \mathrm{cov}_{\boldsymbol{\Theta}^{[t]}}(\boldsymbol{\eta}|\widetilde{Z}) = \mathbf{K}^{[t]} - \mathbf{K}^{[t]}\mathbf{S}'{\boldsymbol{\Sigma}^{[t]}}^{-1}\mathbf{S}\mathbf{K}^{[t]}$$
$$\boldsymbol{\Sigma}_\zeta^{[t]} = \mathrm{cov}_{\boldsymbol{\Theta}^{[t]}}(\boldsymbol{\zeta}|\widetilde{Z}) = \sigma_\zeta^{2[t]}\mathbf{V}_\zeta - \sigma_\zeta^{2[t]}\mathbf{V}_\zeta{\boldsymbol{\Sigma}^{[t]}}^{-1}\sigma_\zeta^{2[t]}\mathbf{V}_\zeta$$

After the expectation step, we perform a maximization step. The parameters $\mathbf{K}$ and σ_ζ^2 in Eq. (A3) should be selected such that $Q(\cdot)$ is maximized. The partial derivative is taken with respect to both parameters and the result is assigned to zero. Finding the derivative here is rather simple since $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ do not show dependency on each other, as observed from Eq. (A3). The updating scheme of the E-M algorithm in each iteration is

$$\mathbf{K}^{[t+1]} = \mathbf{K}^{[t]} + \mathbf{K}^{[t]}\left(\mathbf{S}'{\boldsymbol{\Sigma}^{[t]}}^{-1}\left(\widetilde{Z}\widetilde{Z}'{\boldsymbol{\Sigma}^{[t]}}^{-1} - \mathbf{I}_N\right)\mathbf{S}\right)\mathbf{K}^{[t]} \tag{A4}$$

$$\sigma_\zeta^{2[t+1]} = \sigma_\zeta^{2[t]} + \sigma_\zeta^{2[t]}\,\mathrm{tr}\left(\frac{1}{N}{\boldsymbol{\Sigma}^{[t]}}^{-1}\left(\widetilde{Z}\widetilde{Z}'{\boldsymbol{\Sigma}^{[t]}}^{-1} - \mathbf{I}_N\right)\mathbf{V}_\zeta\right)\sigma_\zeta^{2[t]} \tag{A5}$$

HESSD
12, 363–404, 2015

We keep updating the solution until the algorithm converges. One criterion to monitor convergence is to calculate the norm of the difference between the current and last update of the vector $\boldsymbol{\Theta}$ (which is of size r^2+1). That means $\|\boldsymbol{\Theta}^{[t+1]} - \boldsymbol{\Theta}^{[t]}\| < b$ should hold for small enough and positive value of b. Following Katzfuss and Cressie (2009), b is assigned a value of $10^{-6}r^2$. The starting choice of $\mathbf{K}$ and σ_ζ^2 should be valid; strictly speaking, $\mathbf{K}^{[0]}$ must be symmetric and positive-definite and $\sigma_\zeta^{2[0]}$ must be positive, i.e., $\mathbf{K}^{[0]} = 0.9 \cdot \mathrm{var}(\widetilde{Z})\mathbf{I}_r$ and $\sigma_\zeta^{2[0]} = 0.1 \cdot \mathrm{var}(\widetilde{Z})$.

The measurement error variance σ_ϵ^2 is estimated separately from the empirical semivariogram of the data. Estimating both σ_ϵ^2 and σ_ζ^2 from the data is not a trivial task. That is because the nugget effect in the semivariogram reflects not only the error variance but may be affected by the fine-scale variance. Therefore, having information about the error distribution and variance is worthwhile. In our case we estimate σ_ϵ^2 using the method of robust semivariogram (Cressie, 1993),

$$2\gamma(h) = \frac{\left(\frac{1}{|N(h)|}\sum_{N(h)}\left|Z(\boldsymbol{u}_i) - Z(\boldsymbol{u}_j)\right|^{1/2}\right)^4}{\left(0.457 + \frac{0.494}{|N(h)|}\right)} \tag{A6}$$

where h is separation distance, assuming the signal is spatially isotropic. To obtain an estimate of σ_ϵ^2, a strait line is fitted to the estimated semivariogram at short h. Since the slope of the structure function (variogram) describing atmospheric turbulence is expected to vary with h, we made the line fitting based on the estimates of the first 3 km (empirically defined). Let the line fit be $\hat{\gamma}(h) = \hat{\gamma}(0+) + bh$, then the estimate of σ_ϵ^2 is

$$\hat{\sigma}_\epsilon^2 = \hat{\gamma}(0+) \tag{A7}$$

Should $\hat{\gamma}(0+)$ have a negative value, $\hat{\sigma}_\epsilon^2$ is set to zero.

HESSD
12, 363–404, 2015

The estimate of **K** using the detrended PWV maps estimated from the PSI + GNSS and model data on 5 September 2005 is shown in Fig. 12. The corresponding covariance function is also shown. The matrix **S** is constructed as described in Sect. 4.3 using the nodes setup in Fig. 8. The $\mathbf{K}_{EM}$ has a maximum value for the element (29,29), which is equivalent to estimate at the node in the lower right corner at the location (8.524° E, 48.69° N), see Fig. 8. This can be explained by the sparseness of PWV estimates close to this node and the PWV values from PSI and GNSS are significantly higher than those from the model. The covariance matrix is computed for the observations binned into 7 km × 7 km blocks to demonstrate covariance structure. We observe from the covariance matrices that the variances, on the main diagonal, increase in areas of sparse observations. The reader should note that the observations do not exist on a regular grid (due to the spatial distribution of PS points); hence, the covariance values in the off-diagonal cells can be negative and then again positive.

Acknowledgements. The authors would like to thank the GNSS data providers: RENAG, RGP, Teria, and Orpheon (France), SA*POS*®-Baden-Württemberg and Rheinland-Pfalz (Germany), European Permanent Network, and IGS. We thank the ESA for the ASAR and MERIS data. We also acknowledge the ECWMF for providing the ERA-INTERIM reanalysis data.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

References

Alshawaf, F.: Constructing Water Vapor Maps by Fusing InSAR, GNSS, and WRF Data, PhD thesis, Karlsruhe Institute of Technology, 2013. 365, 367, 382

Alshawaf, F., Fersch, B., Hinz, S., Kunstmann, H., Mayer, M., Thiele, A., Westerhaus, M., and Meyer, F.: Analysis of atmospheric signals in spaceborne InSAR – toward water vapor mapping based on multiple sources, in: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, 1960–1963, 2012. 365, 366

HESSD
12, 363–404, 2015

Bender, M., Dick, G., Wickert, J., Schmidt, T., Song, S., Gendt, G., Ge, M., and Rothacher, M.: Validation of GPS slant delays using water vapour radiometers and weather models, Meteorol. Z., 17, 807–812, 2008. 366
Berg, P., Wagner, S., Kunstmann, H., and Schädler, G.: High resolution regional climate model simulations for Germany: part I – validation, Clim. Dynam., 40, 401–414, 2013. 368
Bevis, M., Businger, S., Herring, T. A., Rocken, C., Anthes, R. A., and Ware, R. H.: GPS meteorology: remote sensing of atmospheric water vapor using the global positioning system, J. Geophys. Res.-Atmos., 97, 15787–15801, 1992. 365
Braverman, A., Nguyen, H., and Olsen, E.: Space-Time Data Fusion, Earth Science, Earth Science Technology Forum, Pasadena, California, 2011. 375
Braverman, A. J., Cressie, N., Katzfuss, M., Michalak, A. M., Miller, C. E., Nguyen, H., Olsen, E. T., and Wang, R.: Geostatistical data fusion for remote sensing applications, AGU Fall Meeting Abstracts, p. C1014, 2009. 372, 373
Chen, F. and Dudhia, J.: Coupling an advanced land surface-hydrology model with the Penn State–NCAR MM5 Modeling System. Part I: Model implementation and sensitivity, Mon. Weather Rev., 129, 569–585, 2001. 368
Cimini, D., Pierdicca, N., Pichelli, E., Ferretti, R., Mattioli, V., Bonafoni, S., Montopoli, M., and Perissin, D.: On the accuracy of integrated water vapor observations and the potential for mitigating electromagnetic path delay error in InSAR, Atmos. Meas. Tech., 5, 1015–1030, doi:10.5194/amt-5-1015-2012, 2012. 366
Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Kiehl, J. T., Briegleb, B., Bitz, C., Lin, S., Zhang, M., and Dai, Y.: Description of the NCAR Community Atmosphere Model (CAM 3.0), Tech. rep., Climate And Global Dynamics Division National Center For Atmospheric Research Boulder, Colorado, 2004. 368
Cressie, N.: Statistics for Spatial Data, Wiley, New York, 1993. 369, 370, 385
Cressie, N. and Johannesson, G.: Fixed rank kriging for very large spatial data sets, J. Roy. Stat. Soc. B Met., 70, 209–226, 2008. 370, 373, 378, 382
Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Stat. Soc. B, 39, 1–38, 1977. 382
Ferretti, A., Prati, C., and Rocca, F.: Permanent scatterers in SAR interferometry, IEEE T. Geosci. Remote, 39, 8–20, 2001. 365

HESSD
12, 363–404, 2015

Fersch, B. and Kunstmann, H.: Atmospheric and terrestrial water budgets: sensitivity and performance of configurations and global driving data for long term continental scale WRF simulations, Clim. Dynam., 42, 2367–2396, 2014. 366

Fersch, B., Kunstmann, H., Devaraju, B., and Sneeuw, N.: Large scale water storage variations from regional atmospheric water budgets and comparison to the GRACE spaceborne gravimetry, J. Hydrometeorol., 13, 1589–1603, 2012. 366

Gelfand, A. E., Zhu, L., and Carlin, B. P.: On the change of support problem for spatio-temporal data, Biostatistics, 2, 31–45, 2001. 369

Gong, W., Meyer, F., Webley, P., Morton, D., and Liu, S.: Performance analysis of atmospheric correction in InSAR data based on the Weather Research and Forecasting Model (WRF), in: Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 25–30 July 2010, Honolulu, Hawaii, USA, 2900–2903, 2010. 366

Hanssen, R.: Radar Interferometry: Data Interpretation and Error Analysis, Kluwer Academic Publishers, Dordrecht, the Netherlands, 2001. 365

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, Mon. Weather Rev., 132, 103–120, 2004. 368

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, 2006. 368

Hooper, A., Segall, P., and Zebker, H.: Persistent scatterer interferometric synthetic aperture radar for crustal deformation analysis, with application to Volcán Alcedo, Galápagos, J. Geophys. Res.-Solid Earth, 112, B07407, doi:10.1029/2006JB004763, 2007. 365, 367

Jade, S. and Vijayan, M.: GPS-based atmospheric precipitable water vapor estimation using meteorological parameters interpolated from NCEP global reanalysis data, J. Geophys. Res.-Atmos., 113, D03106, doi:10.1029/2007JD008758, 2008. 365

Kain, J. S.: The Kain–Fritsch convective parameterization: an update, J. Appl. Meteorol., 43, 170–181, 2004. 368

Kang, E. L. and Cressie, N.: Bayesian inference for the spatial random effects model, J. Am. Stat. Assoc., 106, 972–983, 2011. 374

Karabatić, A., Weber, R., and Haiden, T.: Near real-time estimation of tropospheric water vapour content from ground based GNSS data and its potential contribution to weather now-casting in Austria, Adv. Space Res., 47, 1691–1703, 2011. 365

HESSD
12, 363–404, 2015

Katzfuss, M. and Cressie, N.: Maximum likelihood estimation of covariance parameters in the spatial-random-effects model, in: Proceedings of Joint Statistical Meetings Alexandria, American Statistical Association, 1–6 August 2009, 3378–3390, 2009. 382, 383, 384, 385

Luo, X., Mayer, M., and Heck, B.: Extended neutrospheric modelling for the GNSS-based determination of high-resolution atmospheric water vapor fields, Bol. Cien. Geod., 14, 149–170, 2008. 365

Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., and Rabaute, T.: The displacement field of the Landers earthquake mapped by radar interferometry, Nature, 364, 138–142, 1993. 365

Mateus, P., Nico, G., Tomé, R., Catalão, J., and Miranda, P.: Comparison of precipitable water vapor (PWV) maps derived by GPS, SAR interferometry, and numerical forecasting models, SPIE, 2008, 782714–782714–7, 2010. 366

Meyer, F., Bamler, R., Leinweber, R., and Fischer, J.: A comparative analysis of tropospheric water vapor measurements from MERIS and SAR, in: Geoscience and Remote Sensing Symposium, IGARSS 2008, IEEE International, 6–11 July, 2008, Boston, Massachusetts, Vol. 4, IV–228–IV–231, 2008. 365

Nguyen, H.: Spatial Statistical Data Fusion for Remote Sensing Applications, PhD thesis, University of California, Los Angeles, USA, 2009. 372, 373, 381, 382

Onn, F. and Zebker, H. A.: Correction for interferometric synthetic aperture radar atmospheric phase artifacts using time series of zenith wet delay observations from a GPS network, J. Geophys. Res.-Solid Earth, 111, B09102, doi:10.1029/2005JB004012, 2006. 365

Pichelli, E., Ferretti, R., Cimini, D., Perissin, D., Montopoli, M., Marzano, F. S., and Pierdicca, N.: Water vapour distribution at urban scale using high-resolution numerical weather model and spaceborne SAR interferometric data, Nat. Hazards Earth Syst. Sci., 10, 121–132, doi:10.5194/nhess-10-121-2010, 2010. 365

Rocken, C., Hove, T. V., Johnson, J., Solheim, F., Ware, R., Bevis, M., Chiswell, S., and Businger, S.: GPS/STORM-GPS sensing of atmospheric water vapor for meteorology, J. Atmos. Ocean. Tech., 12, 468–478, 1995. 365

Skamarock, W. C. and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comput. Phys., 227, 3465–3485, doi:10.1016/j.jcp.2007.01.037, 2008. 366

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A Description of the Advance Research WRF Version 3, NCAR Technical Note

HESSD
12, 363–404, 2015

NCAR/TN-468+STR, National Center for Atmospheric Research, Boulder, Colorado, USA, 2008. 368

Uppala, S., Dee, D., Kobayashi, S., Berrisford, P., and Simmons, A.: Towards a climate data assimilation system: status update of ERA-Interim, ECMWF Newsl., 115, 12–18, 2008. 368

Zebker, H. A. and Goldstein, R. M.: Topographic mapping from interferometric synthetic aperture radar observations, J. Geophys. Res.-Solid Earth, 91, 4993–4999, 1986. 365

Zebker, H. A., Rosen, P. A., Goldstein, R. M., Gabriel, A., and Werner, C. L.: On the derivation of coseismic displacement fields using differential radar interferometry: the Landers earthquake, J. Geophys. Res.-Solid Earth, 99, 19617–19634, 1994. 365

Zebker, H. A., Rosen, P. A., and Hensley, S.: Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps, J. Geophys. Res.-Solid Earth, 102, 7547–7563, 1997. 365

HESSD
12, 363–404, 2015

Table 1. Model components from point-level and areal-level data.

	Point data	Block data
True process	$Y(\boldsymbol{s})$	$Y(B_i) = \frac{1}{\lvert B_i \rvert} \sum_{\boldsymbol{s} \subset B_i} Y(\boldsymbol{s})$
Trend	$T(\boldsymbol{s})\boldsymbol{\alpha}$	$\left(\frac{1}{\lvert B_i \rvert} \sum_{\boldsymbol{s} \subset B_i} T(\boldsymbol{s}) \right) \boldsymbol{\alpha}$
Weighting matrix	$\mathbf{S}(\boldsymbol{s})$	$\widetilde{\mathbf{S}}(B_i) = \frac{1}{\lvert B_i \rvert} \sum_{\boldsymbol{s} \subset B_i} \mathbf{S}(\boldsymbol{s})$
Medium-scale signal	$\mathbf{S}(\boldsymbol{s})\boldsymbol{\eta}$	$\widetilde{\mathbf{S}}(B_i)\boldsymbol{\eta}$
Fine-scale signal	$\zeta(\boldsymbol{s})$	$\zeta(B_i) = \frac{1}{\lvert B_i \rvert} \sum_{\boldsymbol{s} \subset B_i} \zeta(\boldsymbol{s})$
Error	$\epsilon(\boldsymbol{s})$	$\epsilon(B_i)$

HESSD
12, 363–404, 2015

Table 2. Spatial correlation coefficients (CC) and RMS values when comparing the prediction maps with MERIS PWV maps.

Method	Spatial CC	RMS [mm]
WRF data	0.70	1.33
Remote sensing data	0.87	0.90
Data fusion	0.91	0.82

HESSD
12, 363–404, 2015

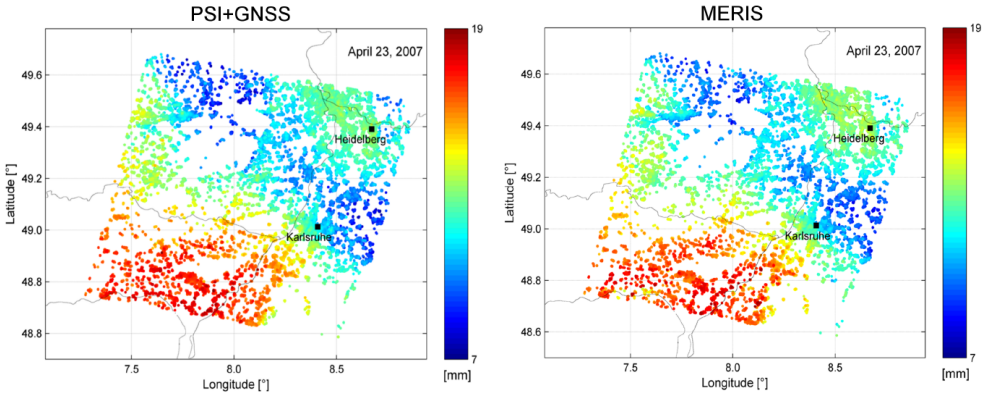

Figure 1. Maps of the absolute atmospheric PWV derived by combining PSI and GNSS data and the corresponding map from MERIS. The spatial correlation coefficient is 95 % and the RMS value of the differences is 0.68 mm.

HESSD
12, 363–404, 2015

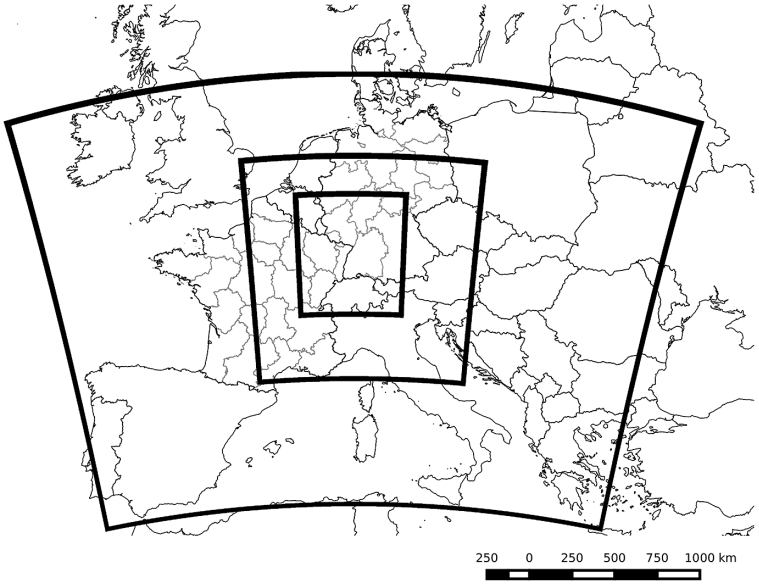

Figure 2. WRF model set up with a parent domain of resolution 27 km × 27 km and two nests of 9 km × 9 km and 3 km × 3 km, respectively.

HESSD
12, 363–404, 2015

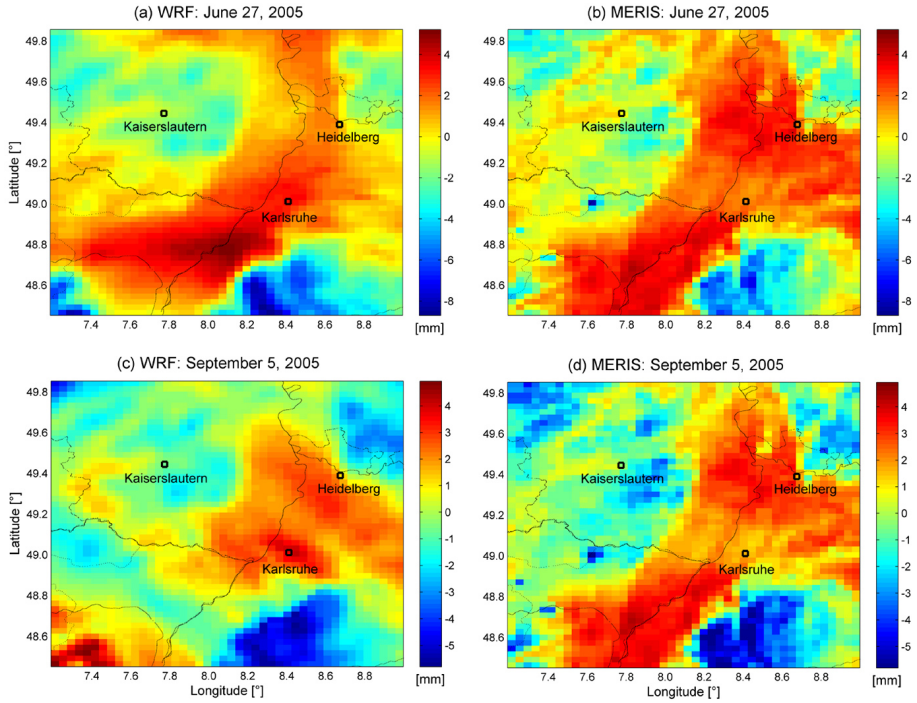

Figure 3. Maps of PWV content as received from MERIS and WRF, where a linear trend is subtracted from each map. The upper data are received on 27 June 2005 (SAR overpass), while the lower data on 5 September 2005. Gaussian averaging is applied to scale the MERIS data at WRF resolution, 3 km × 3 km. The spatial correlation coefficient between the upper maps is 0.8 and 0.71 for the lower.

HESSD
12, 363–404, 2015

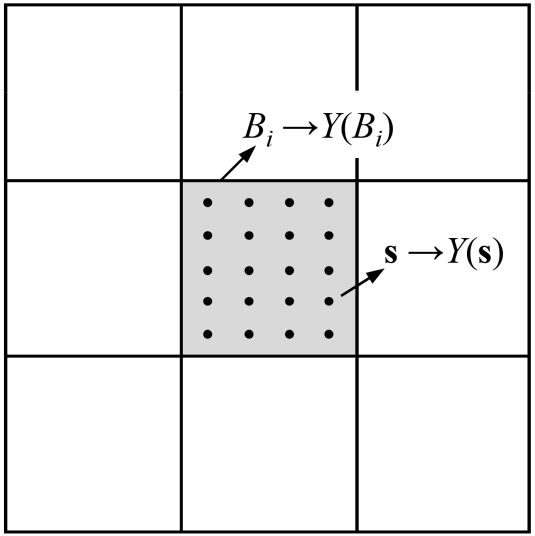

Figure 4. Point and block data, such that for spatial data, $Y(B_i)$ represents the average of the point data within the block.

HESSD
12, 363–404, 2015

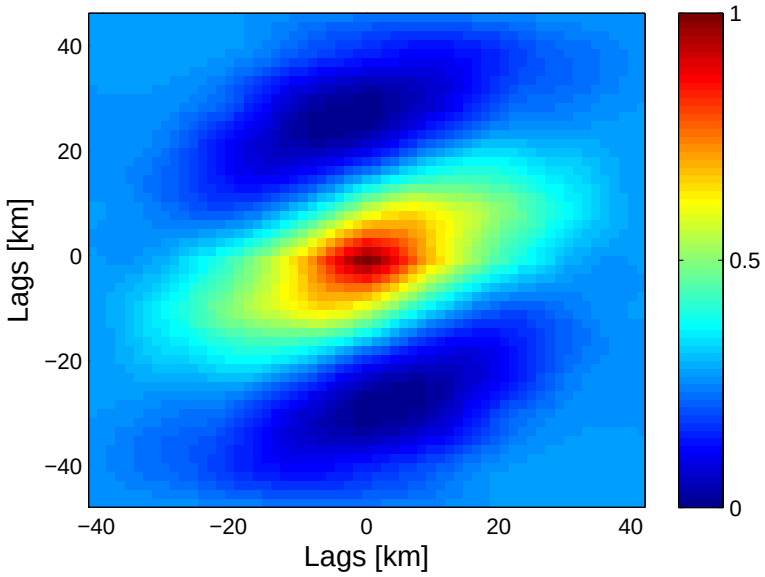

Figure 5. Spatial autocorrelation function for a PWV map, with the long-wavelength component removed, computed from remote sensing data acquired on 5 September 2005, 10:51 UTC.

HESSD
12, 363–404, 2015

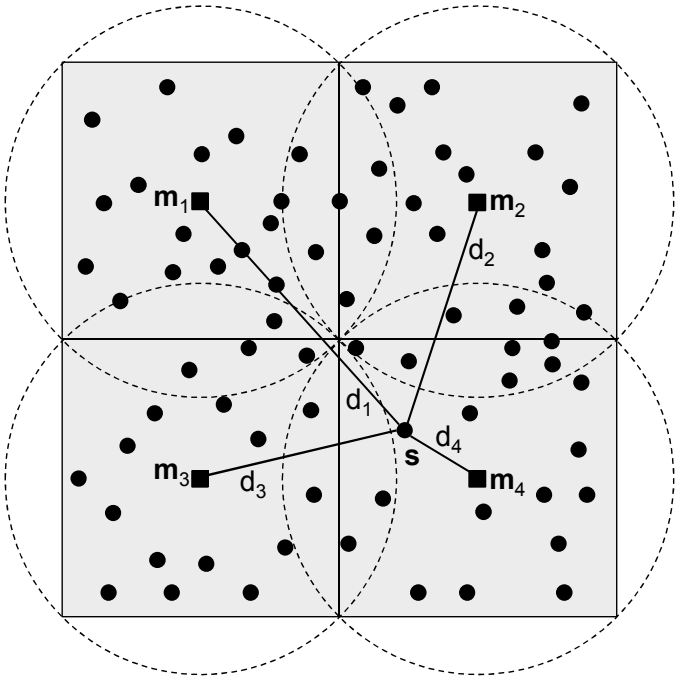

Figure 6. The observation domain with the black dots define the locations at which the data are available. The black little squares indicate the nodes. The weights for each location ***s*** are related to the distances d_i. The dashed circles define the radius for each node.

HESSD
12, 363–404, 2015

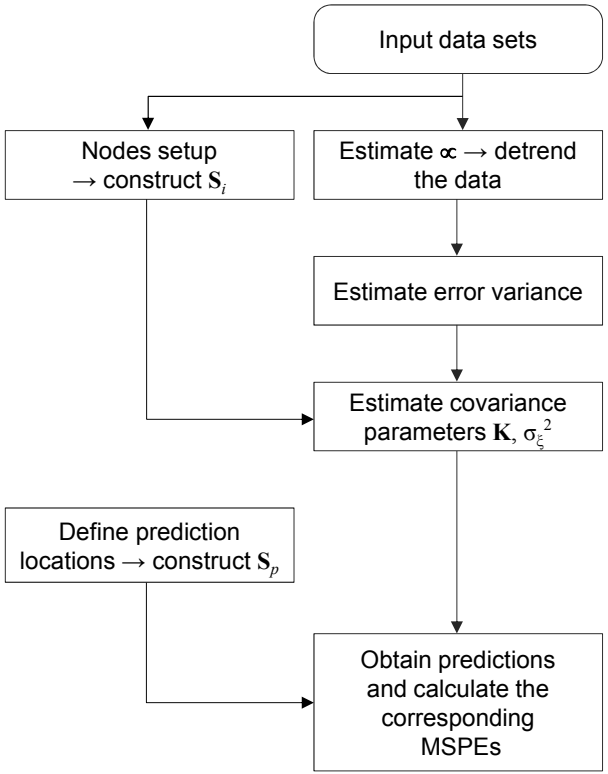

Figure 7. Obtaining predictions via the FRK method.

HESSD
12, 363–404, 2015

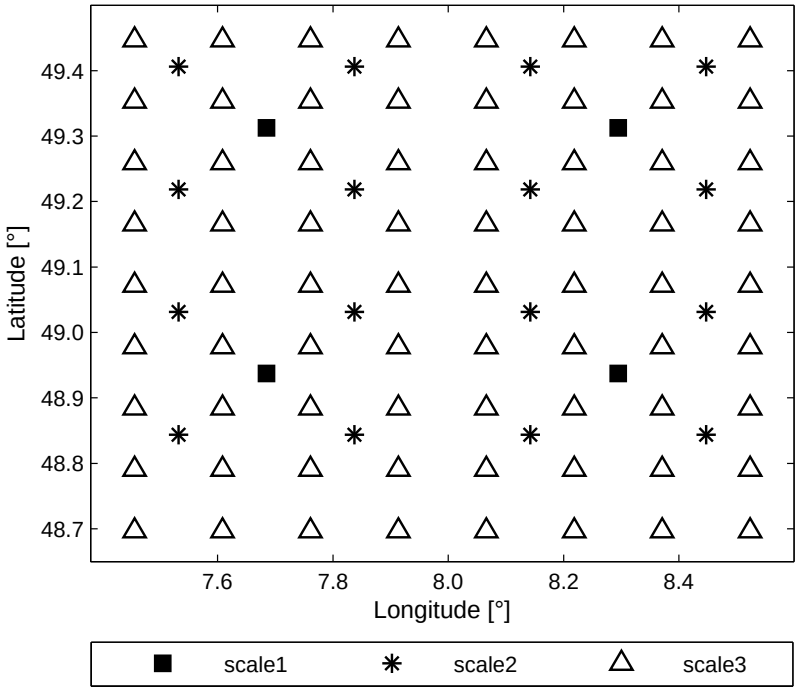

Figure 8. FRK nodes or center locations of 93 basis functions at three spatial resolutions. The first resolution is 40 km, the second resolution is 20 km, and the third resolution is 10 km.

HESSD
12, 363–404, 2015

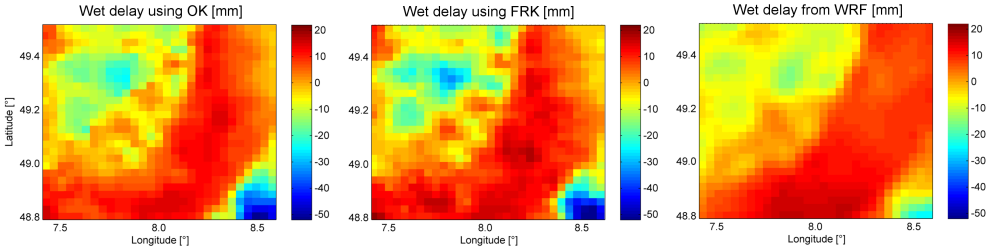

Figure 9. Wet delay prediction map using the block OK and FRK. The resolution of the grid is 3 km × 3 km. A point-level wet delay map, on 23 May 2005 at 09:51 UTC, is used as input to the algorithms.

HESSD
12, 363–404, 2015

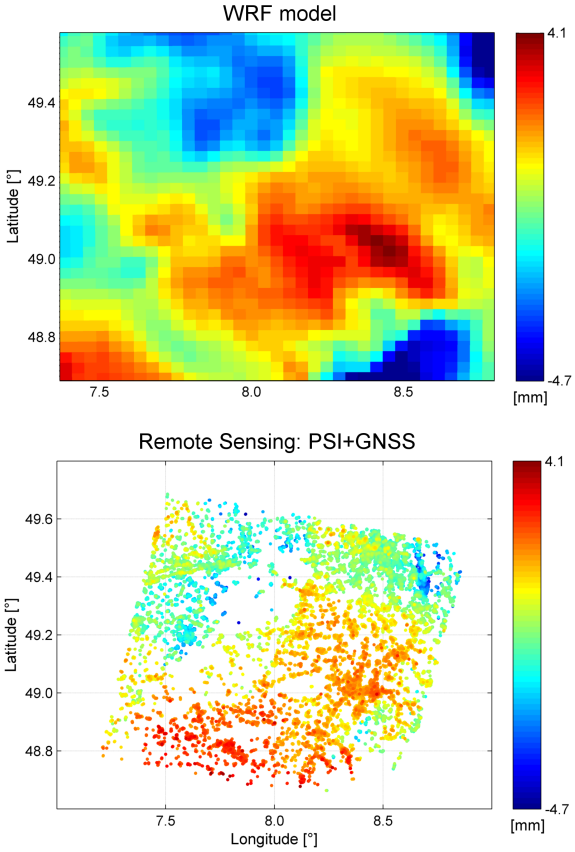

Figure 10. PWV maps from PSI + GNSS combination and WRF on 5 September 2005, with a linear trend subtracted from each map. PSI + GNSS provide point-level observations, while WRF generates block data with a block size of 3 km × 3 km.

HESSD
12, 363–404, 2015

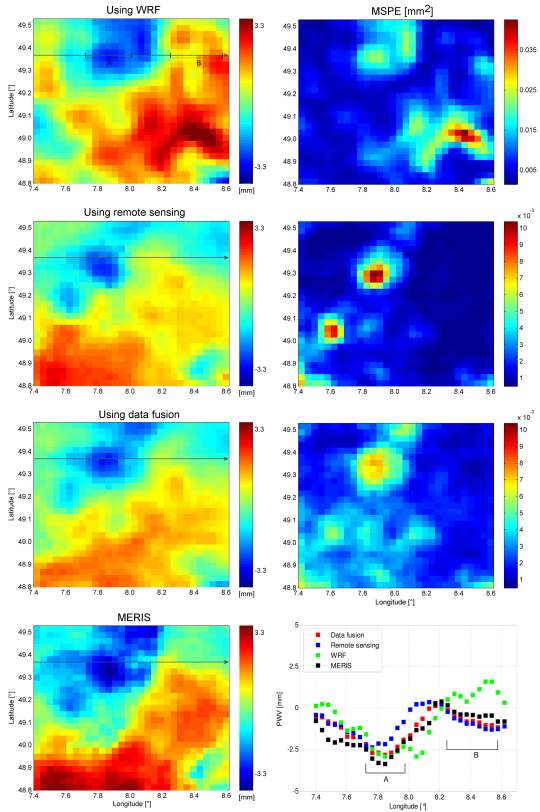

Figure 11. PWV prediction and MSPE maps obtained by data fusion of PWV estimates from PSI and GNSS and maps from WRF as well as predictions obtained by applying FRK to individual data sets. The data are available on 5 September 2005 at SAR overpass time. The output grid has a block size of 3 km × 3 km.

HESSD
12, 363–404, 2015

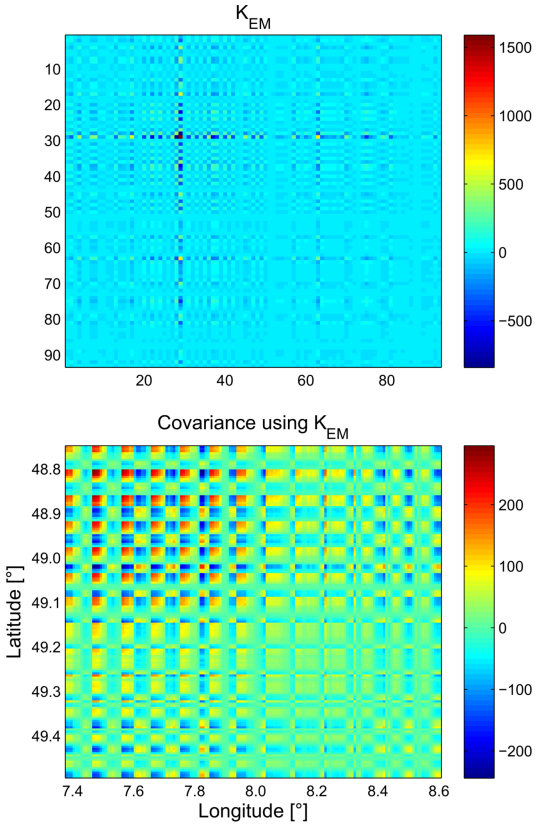

Figure 12. Estimate of the covariance matrix **K** using the E-M algorithm and the corresponding covariance matrix for the Wet delay map from PSI + GNSS. The wet delay observations are aggregated into maps of 7×7 km^2 cells before their covariance matrices are computed.

HESSD
12, 363–404, 2015