Assimilation of virtual wide swath altimetry to improve Arctic river modeling

Sylvain Biancamaria, Michael Durand, Kostantinos M. Andreadis, Paul Bates, Aaron Boone, Nelly Mognard, Ernesto Rodriguez, Doug Alsdorf, D. P. Lettenmaier, Elizabeth A. Clark

Assimilation of virtual wide swath altimetry to improve Arctic river modeling

S. Biancamaria [a, b, c, *], M. Durand [d, e], K. M. Andreadis [e], P. D. Bates [f], A. Boone [g], N. M. Mognard [a, b], E. Rodríguez [h], D. E. Alsdorf [d, e, i], D. P. Lettenmaier [j], E. A. Clark [j]

[a] Université de Toulouse; UPS (OMP-PCA); LEGOS; 14 Av. Edouard Belin, F-31400 Toulouse, France

[b] CNES; LEGOS, F-31400 Toulouse, France

[c] CNRS; LEGOS, F-31400 Toulouse, France

[d] Ohio State University; School of Earth Sciences; Columbus OH 43210, USA

[e] Ohio State University; Byrd Polar Research Center; Columbus OH 43210, USA

[f] School of Geographical Sciences; University of Bristol; Bristol BS8 1SS, UK

[g] GAME/CNRM; Météo-France; CNRS; F-31057 Toulouse Cedex, France

[h] Jet Propulsion Laboratory; California Institute of Technology; Pasadena CA 91109, USA

[i] Ohio State University; The Climate, Water & Carbon Program; Columbus OH 43210, USA

[j] Civil and Environmental Engineering; University of Washington; Seattle WA 98195, USA

[*] Corresponding author: sylvain.biancamaria@legos.obs-mip.fr (Postal address: LEGOS, 14 avenue Edouard Belin, 31400 Toulouse, France; Phone: +335 61 33 29 30; Fax: +335 61 25 32 05)

Abstract

Global surface water variations are still difficult to monitor with current satellite measurements. The future Surface Water and Ocean Topography (SWOT) mission is designed to address this issue. Its main payload will be a wide swath altimeter which will provide maps of water surface elevations between 78°S and 78°N over a 120 km swath. This study aims to combine coupled hydrologic/hydraulic modeling of an Arctic river with virtual SWOT observations using a local ensemble Kalman smoother to characterize river water depth variations. We assumed that modeling errors are only due to uncertainties in atmospheric forcing fields (precipitation and air temperature) and different SWOT orbits were tested. First, we tested orbits that all have a three day repeat period but differ in terms of their spatial coverage of the study reach; these orbits correspond to the first three months of the mission, which will be dedicated to calibration and validation experiments. For these orbits, the mean spatial Root Mean Square Error (RMSE) in modeled channel water depth decreased by between 29 % and 79 % compared to the modeled RMSE with no assimilation, depending on the spatial coverage. The corresponding mean temporal RMSE decrease was between 54 % and 91 %. We then tested the nominal orbit with a twenty two day repeat period which will be used during the remaining lifetime of the mission. Unlike the three day repeat orbits, this orbit will observe all continental surfaces (except Antartica and the northern part of Greenland) during one repeat period. The assimilation of SWOT observations computed with this nominal orbit into the hydraulic model leads to a decrease of 59 % and 66 % in the mean spatial and temporal RMSE in modeled channel water depth, respectively. These results show the huge potential of the future SWOT mission for land surface hydrology, especially at high latitudes which will be very well sampled during one orbit repeat period. Still, further work is needed to reduce current modeling uncertainties and to better characterize SWOT measurement errors.

Keywords: SWOT, wide swath altimetry, hydrologic/hydraulic modeling, data assimilation, Kalman filter, Kalman smoother, Arctic, Ob River

1. Introduction

More than 73 % of water used for human activities (for example as drinking water, for irrigation or for energy generation and industrial processes) comes from surface water (Connor *et al.*, 2009). It is therefore crucial to observe and understand the spatial and temporal variations in surface water across the globe. Accordingly, in-situ gage networks have been intensively developed since the second part of the twentieth century. However, these networks are still sparse, especially in remote regions like in the Arctic, and in many areas the coverage is actually now declining. To overcome this issue, hydrologic models and remote sensing data have been used to complement in-situ measurements. However, current remote sensing observations of water surface elevations made by nadir altimeters, which only measure water elevations along the track of the satellite with typical track spacing of ~120 km, miss many of the world's surface water bodies, have relatively large spatial footprints (on the order of 5-10 km) and do not give any information about water extent (Alsdorf *et al.*, 2007).

In order to better characterize surface water and oceanic processes, a wide swath altimeter, the Surface Water and Ocean Topography (SWOT) mission, is currently under study by NASA (National Aeronautics and Space Administration) and CNES (Centre National d'Etudes Spatiales). SWOT will provide maps of water elevation at an unprecedented spatial resolution (on the order of 50-100m) and precision (centimetric accuracy when averaged over areas of 1 km^2; Durand *et al.*, 2010). A small number of recent studies have begun to quantify the benefits of such a mission for land surface hydrology. Biancamaria *et al.* (2010) focus on the benefits of this mission at a global scale for different orbits and show that errors in instantaneous discharge estimated from SWOT measurements using rating curve should be below 25 % for rivers wider than 50m. Errors only due to the SWOT temporal sampling on monthly discharge should be below 20 % for rivers with drainage areas larger than 7000 km^2. Andreadis *et al.* (2007) estimated the benefit of assimilating virtual wide swath measurements, using an Ensemble Kalman filter (EnKF), to reduce modeling errors due to uncertainties on lateral inflows of a mid-latitude river (a segment of the Ohio River). This study compared three different orbits with 8, 16 and 32 days repeat period. They showed that relative errors could be reduced by nearly a factor of two when the filter is used; the best results were obtained for the orbit with the smallest repeat period. Durand *et al.* (2008) assimilated virtual SWOT observations into the Amazon River hydraulic modeling developed by Wilson *et al.* (2007) for estimating bathymetric depths and slopes. They showed that bathymetric slopes can be estimated to within 0.30 $cm.km^{-1}$ and depths to within 56 cm (which was 84 % less than errors without assimilation). They also highlighted that, in their modeling, model errors dominate over measurement errors and therefore estimates of channel bathymetry are relatively insensitive to measurement error characteristics.

The study presented here is a continuation of these works and aims to assess how SWOT could improve the modeling of an Arctic river, where the flow regime is mainly driven by snow melt (contrary to the Ohio and Amazon rivers). In addition, here we use for the first time the actual SWOT orbits, which have recently been selected, a more realistic model boundary condition error computation, and a slightly different assimilation scheme (Local Ensemble Kalman Smoother) compared to previous work. In particular, this paper aims to test the impact of SWOT orbital coverage (depending on orbit parameters), but we do not address the improvement expected from the high spatial resolution of SWOT measurements, as the river modeling used has a 1 km x 1 km spatial resolution (due to the current lack of a high spatial resolution digital elevation model above 60°N).

2. Study domain and river modeling

This study focuses on the Lower Ob River between the cities of Belogorje and Salekhard; this reach covers the downstream 1120 km of the river before the Ob estuary (Figure 1) and corresponds to a drainage area of 790 000 km^2 (according to the Arctic Rapid Integrated Monitoring System, ArcticRIMS, http://rims.unh.edu). The drainage basin of the entire Ob river covers 2 990 000 km^2 and it is located in Western Siberia, east of the Ural Mountains. In terms of discharge, the Ob is the world's 12th largest river and the 3rd largest in the Arctic (Herschy and Fairbridge, 1998). The Ob is frozen from November to April and its discharge regime is mainly driven by snow melt, with a maximum in May/June during ice breakup (Pavelsky and Smith, 2004). Yang *et al.* (2004) reported that from 1936 to 1990 the monthly mean discharge at the river outlet varied between 500 and 1200 $m^3.s^{-1}$ in the cold season (from November to April), and between 3500 and 9000 $m^3.s^{-1}$ during the summer months. The land cover in this domain is classified as sporadic and discontinuous permafrost (Brown *et al.*, 1998). According to Yang *et al.* (2004), the effects of human activities on the study domain are limited and there are no reservoirs on the lower part of the river. In this study, the modeled time period corresponds to the calendar year 1993.

The river is modeled by the flood inundation model LISFLOOD-FP developed at the University of Bristol, UK (Bates and De Roo, 2000). LISFLOOD-FP is a coupled 1D/2D hydraulic model based on a raster grid. It predicts water depth in each grid cell at each time step and hence can simulate the dynamic propagation of flood waves over fluvial, coastal and estuarine floodplains. Here, the 1D channel flow is based on the kinematic approximation to the 1D St Venant equations. Floodplain flows are similarly described in terms of continuity and momentum equations, discretized over a grid of square cells, which allows the model to represent 2-D dynamic flow fields on the floodplain. There is, however, no exchange of momentum between main channel and floodplain flows, only mass, and ice jam and break up processes are not represented. The kinematic approximation of the channel flow might also be a limitation of the modeling, as the Ob flow regime is likely diffusive, at least in the downstream part of the river. However, according to Trigg et al. (2009), for the Amazon River, this approximation leads to an additional Root Mean Square Error (RMSE) of around 1 m. This error is likely to be lower for the Ob and is much smaller than errors on the floodplain topography and river bathymetry. Finally, backwater effects from inflows are not modeled, but are likely to be minor given the ratio of inflow volume to the mainstem discharge.

The floodplain topography comes from the ACE (Altimeter Corrected Elevation) digital elevation model from De Montfort University, UK, and channel centreline position and width from freely available data sources (CIA World Data Bank II and Landsat imagery). The channel depth is poorly known and is estimated based on a limited number of literature sources and a model sensitivity analysis (see Biancamaria *et al.*, 2009). The Manning coefficients for the river and for the floodplain have been assumed constant in space and time (equal to 0.015 and 0.06, respectively) and the 2D floodplain model is run at 1km resolution. The incoming flow to the study domain from the upstream river and the lateral inflows to the river in the study domain (red arrows in Figure 1) are computed by ISBA (Interactions between the Soil-Biosphere-Atmosphere; Noilhan and Mahfouf, 1996), which is a land surface scheme developed by the CNRM (Centre National de Recherche Meteorologique) in France. Total precipitation (rain and snow) and temperature uncertainties are a main source of errors on the modeled discharge in the coupled ISBA/LISFLOOD-FP scheme (Biancamaria *et al.*, 2009). More details on the Lower Ob modeling can be found in Biancamaria *et al.* (2009).

3. Satellite observations

The main purpose of this work is to estimate the benefits to the accurate estimation of water depths on an Arctic River of combining measurements from the future SWOT mission and hydrologic modeling. This section presents this future satellite mission and how virtual SWOT observations have been generated.

3.1. The SWOT mission

This mission is intended to be launched between 2018 and 2020. SWOT will provide high-resolution images of water surface elevations over oceanic and continental surface water bodies. The core satellite payload is the Ka-band Radar Interferometer (KaRIN), a wide swath radar interferometer. KaRIN has two antennas separated by a 10 m boom which observe two ground swaths of 60 km on each side of the satellite nadir, separated by a 20 km gap. The intrinsic pixel resolution will vary from 60 m (near range) to 10 m (far range) across-track and will be at best around 2 m along-track (however, this value is also dependent upon decorrelation time). The chosen orbits have a 971 km altitude and 78° inclination, in order to observe almost all the continental surfaces (Rodríguez, 2009). The nominal lifetime of the mission is three years.

The first three months of the mission will be a calibration/validation period (called the 'fast sampling period') with a 3 day repeat orbit, allowing a more frequent revisit time but with incomplete spatial coverage. As the satellite has not yet been launched, the orbit phase (i.e. the longitude of the orbit where it first crosses the equator eastward of 0°E) is not known; therefore, three different phases, which observe different parts of the study domain, for this fast sampling orbit have been selected (Figure 2a, b and c). The first orbit (orbit 1, Figure 2a) does not observe the most upstream part of the Lower Ob. In contrast, the second orbit (orbit

2, Figure 2b) does not observe the downstream part of the river. The third orbit (orbit 3, Figure 2c) corresponds to an optimal coverage, as almost all the river inside the study domain is seen. These three orbits thus represent the likely envelope of sampling scenarios possible for a given Arctic river basin during the fast sampling period.

After the initial three months, the remaining time during the mission will be undertaken with an orbit that meets the nominal science requirement to obtain a global coverage of the earth and that has a 22 day repeat. Figure 2d presents the number of observations of the study domain per repeat period (22 days) for this orbit. As the coverage is global, it is not necessary to test different orbit phases.

3.2. Generation of virtual satellite observations

Virtual SWOT observations were generated by first computing the swath coverage over the study domain for both the nominal and fast sampling orbits for each day during the year 1993. The initial modeling of the Lower Ob (see Biancamaria *et al.*, 2009) was taken as the “true” state and was used to compute SWOT measurements. As an example of model outputs, water elevations computed with this modeling for June 28th 1993 is shown in Figure 3. First, modeled water elevations from the nominal LISFLOOD-FP model inside the SWOT swath were selected and then observation errors were added. For the moment, only the instrumental error is taken into account (the errors due to satellite position uncertainties, atmospheric effects such as wet troposphere delays, etc are not considered). Instrument error was modeled by a white noise of 2 cm standard deviation, which corresponds to the expected error of KaRIN (Enjolras *et al.*, 2006; Rodríguez, 2009) at the 1 km^2 resolution of the Lower Ob model.

4. Methodology

To estimate the benefit of SWOT observations for the study of Arctic rivers, an Observing System Simulation Experiment (OSSE) was implemented. First, virtual SWOT measurements were computed as explained in section 3.2. Then, they were assimilated into the LISFLOOD-FP model of the lower Ob to determine their ability to reduce modeling errors. This section presents the assimilation schemes used (section 4.1) and the estimation of the modeling errors (section 4.2).

4.1. Assimilation schemes

4.1.1. Ensemble Kalman Filter

The assimilation process combines model outputs (called forecast or background) and observations to obtain a better estimate (called analysis) of the river state. A popular assimilation scheme is the Kalman Filter (Kalman, 1960; Kalman and Bucy, 1961). For the Kalman filter the analysis is obtained using the following equations:

$$x^a = x^f + P^f H^T (H P^f H^T + R)^{-1} (y - H x^f) \qquad \text{eq. 1}$$

$$P^a = P^f - P^f H^T (H P^f H^T + R)^{-1} H P^f \qquad \text{eq. 2}$$

Here x^f, x^a and y represent the forecast, analysis and observation, respectively. In this study, the state vector corresponds to water heights along the river. P^f, P^a and R are respectively the error covariances for the forecast, analysis and measurements. H is the measurement operator, which projects the model state into the observation space. The superscript ^{T} corresponds to the matrix transpose.

The analysis computed using the Kalman filter is thus a weighted average of the forecast and the observation. The weight optimally takes into account the error in the forecast and in the observation. The Kalman Filter is a sequential filter, which means that the analysis is computed only at times when an observation is available. It is then used as an updated initial condition for the model, which is subsequently run forward from this updated initial condition until a new observation is available. If the observation operator and model equations are linear and if the forecast and observation errors are zero mean Gaussian random vectors, then the analysis obtained with the Kalman Filter is the best linear unbiased estimate of the model state and its error covariance matrix. In this study, SWOT observations correspond to water height measurements along the river, therefore the observation operator is simply a mask of the swath coverage and is a linear operator. However, the model equations are not linear, as is the case for many fluid flow problems.

The covariances for model forecast, analysis and observations are given by the equations 3, 4 and 5, for which x^t corresponds to the true model state and the overline corresponds to an expectation value.

$$P^f = \overline{(x^f - x^t)\,(x^f - x^t)^T} \qquad \text{eq. 3}$$

$$P^a = \overline{(x^a - x^t)\,(x^a - x^t)^T} \qquad \text{eq. 4}$$

$$R = \overline{(y - H\,x^t)\,(y - H\,x^t)^T} \qquad \text{eq. 5}$$

Since the true model state is never known, the covariance matrices can only be approximated. A widely used Monte Carlo approximation of the covariance matrices was proposed by Evensen (1994), and the resulting filter is commonly called the Ensemble Kalman Filter (EnKF). The EnKF is implemented here. An ensemble of “corrupted” model states is generated, which sample all the possible model errors. Then the error covariance matrices are approximated by the covariance matrices of the ensemble (equations 6 and 7). Of course, increasing the size of the ensemble generally decreases the errors in the Monte Carlo sampling.

$$P^f \approx P_e^f = \overline{\left(x^f - \overline{x^f}\right)\left(x^f - \overline{x^f}\right)^T} \quad \text{eq. 6}$$

$$P^a \approx P_e^a = \overline{\left(x^a - \overline{x^a}\right)\left(x^a - \overline{x^a}\right)^T} \quad \text{eq. 7}$$

The EnKF used in this study has been implemented following the square root analysis algorithm described in Evensen (2004).

4.1.2. Local Ensemble Kalman Filter

In some cases, when the size of the ensemble is small, some spurious long range correlations can appear in the forecast error covariance matrix, leading to large errors in the analysis model state. A solution to this issue consists of limiting the influence of an observation during the analysis step to a localized region near the observation. Hamill *et al.* (2001) suggested replacing equation 1 with equation 8, which includes a correlation matrix, denoted S, representing the region of influence of an observation (the symbol “×” in equation 8 corresponds to the Schur product, i.e. element by element multiplication). From now on, this version of the filter will be referred to as the Local Ensemble Kalman Filter (LEnKF).

$$x^a = x^f + [S \times (P_e^f H^T)]\left\{H[S \times (P_e^f H^T)] + R\right\}^{-1}(y - H\,x^f) \quad \text{eq. 8}$$

Like Hamill *et al.* (2001), in this study a fifth order function of Gaspari and Cohn (1999) has been used to define the correlation matrix S. This correlation function has a shape close to a Gaussian function but it decreases to zero at a finite radius. The length scale of this function has been set to 10 km, which means that the correlation is equal to 0.5 at a distance of 12 km from the observation and is below 0.1 at distance above 22 km. This value was chosen to be one order of magnitude less than the mean distance between two lateral inflows to the river in the Ob modeling, which is equal to 140 km. For the time steps where the EnKF performs well, it is expected that the LEnKF will be slightly less efficient as the influence of observation is reduced. However, this length scale will prevent any spurious long range correlations and thus avoid unrealistic water depth computations.

4.1.3. Local Ensemble Kalman Smoother

As previously stated, the EnKF and LEnKF are sequential filters. For all time steps during which there is no observation, the model is not corrected and thus the spread of the ensemble of model states tends to increase. For the Lower Ob, it takes around 10 days for water to flow from the upstream to the downstream part of the modeled river (Biancamaria *et al.*, 2009). Therefore, after 10 days the benefits from the assimilation are completely lost. Fortunately, for the 22 day SWOT orbit, the study domain is observed every 3 days. Nonetheless, it is important to propagate the benefit of observations to other time steps. This is done by applying the Local Ensemble Kalman Smoother (LEnKS). The LEnKS assumes that differences between observations and model state at a time step *i*, for which an observation is available, are statistically correlated to errors at previous time steps. The equations of the analysis for the time step *j* (*j*<*i*) is the following (Moore, 1973):

$$x_j^a = x_j^f + [S\times(P_{eij}^f\ H^T)]\left\{H[S\times(P_{eii}^f\ H^T)]+R_i\right\}^{-1}(y_i - H\ x_i^f) \quad \text{eq. 9}$$

$$P_{eij}^f = \overline{\left(x_i^f - \overline{x_i^f}\right)\left(x_j^f - \overline{x_j^f}\right)^T} \quad \text{eq. 10}$$

The smoother has been applied over a constant time frame, i.e. for all time steps included in the interval [*i*-timelag; *i*], where timelag is constant for all the analysis steps. Sensitivity of the analysis results to different values of the time lag was explored. It is worth noting that the filter leads to sharp discontinuities in model mass and momentum before and after the update, and that the smoother tends to mitigate these effects.

4.2. Ensemble member generation

The ensemble used in the LEnKF and LEnKS should be representative of all model errors. Possible sources of errors include initial conditions, forcing data, model parameters and model equations used. In this study, only errors from ISBA forcing data (precipitation and temperature) were considered, as these are the primary source of errors in the modeling (Biancamaria *et al.*, 2009). These forcing data come from NCEP-DOE AMIP II reanalysis (National Centers for Environmental Prediction - Department Of Energy, Atmospheric Model Intercomparison Project; Kanamitsu *et al.*, 2002). Errors in reanalysis products are always difficult to estimate, as very few high quality global products exist for comparison, especially at high latitudes. However, Serreze *et al.* (2005) show that for the Ob basin correlations between NCEP monthly total precipitation over the river basin and in-situ measurements vary from 0.60 to 0.86 depending on the month, for the time span 1979/1993. Biancamaria *et al.* (2009) found that downstream Ob discharge modeled using NCEP precipitation has an error of 14 % compared to in-situ discharge time series. In addition, Voisin *et al.* (2008) found that precipitation from another reanalysis (ERA-40 from the European Centre for Medium-Range Weather Forecasts) has errors between 0.7 % and 34.5 % on Eastern Siberian Rivers. Therefore, we have assumed that errors on precipitation are 20 %. The standard deviation between daily air temperature from NCEP and in-situ measurement at Belogorje for 1993 is equal to 0.18. Thus, the error on air temperature has also been set to 20 %.

4.2.1. Methodology

Members of the ensemble correspond to a “corrupted” version of the nominal forcing data (considered to be the truth). The methodology used has been previously developed by Auclair *et al.* (2003) and consists of perturbing the most statistically significant modes of the atmospheric fields. To do so Empirical Orthogonal Functions (EOF) of the atmospheric field temporal anomaly were computed, and the corrupted field ($P^{corrupt}$) was obtained by recombining the first EOF modes which explained 95 % of the variance and the temporal atmospheric field mean, multiplied by white noise (equation 11).

$$P^{corrupt}(l,t) = \overline{P}(l).\varepsilon_m + \sum_{k=1}^{N} \varepsilon_k.\alpha_k(t).\phi_k(l) \qquad \text{eq. 11}$$

In equation 11, *l* is the spatial index, *t* the temporal index, $\overline{P}$ is the temporal mean, *k* is the EOF mode, *N* is the highest EOF mode used, α_k is the temporal component and φ_k is the spatial component of the EOF for the mode k, ε_m is the noise on the mean and ε_k is the noise on the EOF recombination for the mode *k*. ε_m and ε_k are both white noise with a 0.2 standard deviation. It should be noted that ε_m and ε_k are not a function of *l* or *t* (i.e. they are constant in space and time). The last mode N was chosen so that the cumulative explained variance for modes 1 to N is equal to 95 %.

4.2.2. Corrupted precipitation and air temperature

The EOF modes were computed using the algorithm developed by Toumazou and Crétaux (2001), for the total precipitation field (rain rate + snow rate) and air temperature. Table 1 presents the explained variance for the first 10 EOF modes of these two atmospheric forcings. As there is no seasonal cycle in the total precipitation (the mean life time of a depression is roughly around a week, with no strong seasonality), 187 EOF modes are required to explain 95 % of the variance. On the contrary, for air temperature, most of the energy is included in the first 8 modes (which explain 95 % of the variance). The first mode itself (corresponding to the seasonal cycle) explains 84 % of the variance. For these two ISBA inputs, EOFs were computed from August 1991 to July 1995, using the methodology presented in section 4.2.1. For computational reasons, the size of the ensemble was set to 20 (thus 20 corrupted precipitation and temperature fields have been computed).

While LISFLOOD-FP produces a 1km resolution grid of water depths at each modeled time step, for simplicity, we only consider water depths along the channel centre line. Figure 4 presents water depths (in m) along the channel for the truth and all members of the ensemble for a given date, June 28th 1993. Figure 5 shows the time series of water depths along the river channel obtained after running ISBA and LISFLOOD-FP for the "truth" (a.) and the ensemble mean (b.). The ensemble mean water height is greater than the true depth; this is because snow melt occurs earlier in the ISBA ensemble compared to the true ISBA simulation.

In LISFLOOD-FP, the river bathymetry has been set only at the location of lateral inflows. The model does a linear interpolation of the bathymetry between these locations. Therefore, between two lateral inflows the slope is constant, which explains gaps in water depths at lateral inflow locations in Figure 4 and the vertical banding effect in Figure 5. To avoid these effects, future work based on the Ob modeling will use a polynomial interpolation of the bathymetry.

5. Results

The EnKF, LEnKF and the LEnKS were applied to reduce modeling error using virtual SWOT water height observations. The following sections present the results obtained when SWOT observations are computed using the three selected fast sampling orbits (section 5.1) and the nominal orbit (section 5.2). The LEnKS was tested with different time lags (2 days, 3 days, 5 days and 10 days).

5.1. Fast sampling orbits

Table 2 presents the mean spatial and temporal RMSE between the truth and the ensemble mean with and without assimilating SWOT observations for these three orbits. For these orbits, there is an observation every one or two days, depending on the location. Therefore the best results are obtained using a 2 day lag time LEnKS. The percentage of error reduction compared to no assimilation (equation 12) is also indicated in Table 2.

$$\varepsilon = 100 \cdot \frac{RMSE_{no\ assimilation} - RMSE_{assimilation}}{RMSE_{no\ assimilation}} \qquad \text{eq. 12}$$

Model errors after assimilation are highly dependent on the location of the observations and are, therefore, quite different for each orbit phase. The mean spatial RMSE was decreased from between 29 % to 79 % and the mean temporal RMSE was decreased by between 54 % and 91 % for the 2 day time lag LEnKS. In particular, fast sampling orbit 2 (Figure 2b) observes a smaller portion of the river than orbit 1 (Figure 2a); however the mean spatial and temporal RMSE after assimilating SWOT observations generated using orbit 2 are smaller than after assimilating SWOT orbit 1 observations (Table 2). This is due to the location of the ground track: orbit 2 observes the upstream part of the river, near Belogorje, where the incoming flow to the study domain is located. The incoming streamflow is one order of magnitude higher than the lateral inflows to the river from the study domain, as computed by the ISBA model. When orbit 2 is used, the part of the river with the highest error (the upstream) is well observed and thus well corrected; this correction propogates dowstream, even to unobserved river locations. Orbit 1, on the other hand, observes the downstream part of the river; therefore the upstream part of the river, which is not seen, is not corrected. This leads to higher errors than those obtained with orbit 2. This effect is obvious in Figures 5c, 5d and 5e, which show water depths along the river channel versus time for the ensemble mean after assimilating SWOT observations for the three fast sampling orbits using a 2 day time lag LEnKS. Orbit 3 corresponds to the optimum orbit, as almost the entire river is observed every three days. Consequently, the spatial and temporal RMSE for this orbit decrease by 79 % and 91 %, respectively, compared to the RMSE with no assimilation. It is important to note that the main Arctic rivers (Mackenzie, Ob, Yenisey and Lena) are oriented South to North. Therefore, if the SWOT fast sampling orbit is correctly chosen, at least some of these rivers should be very well (if not entirely) observed.

For all three fast sampling orbits in Table 2, the EnKF updates degraded the LISFLOOD-FP model run, e.g., the river bed became dry in certain parts of the study area. These degradations are apparently due to spurious, long-distance correlations. In the fast sampling phase, the updates occur very frequently, such that the model does not have adequate time to self-correct after a spurious update. These results highlight the importance of suppressing long-distance correlations when working with modest ensemble sizes, especially when working with frequent updates.

When the time lag is above the mean time between two observations, some parts of the river channel can be updated twice. However, for the second update the hypothesis that the correction computed during the observation time can be used for previous time steps no longer holds because the error has already been decreased. Thus an unrealistic update is performed, and error increases; this is similar to the issue raised in section 4.1.2 about the need for a local filter. For this reason, errors for the LEnKS in Table 2 have a tendency to increase for time lags above the mean time between two observations.

5.2. Nominal orbit

Table 3 presents the mean spatial and temporal RMSE between the truth and the ensemble mean after assimilating SWOT observations for the nominal orbit. For this orbit, on average, the study domain is observed every three days. For this reason, the best results are obtained with a 3 day time lag LEnKS (in this case the mean spatial and temporal RMSE are reduced by 59 % and 66 %, respectively). The results obtained using a 10-day time lag LEnKS are also indicated in Table 3 (the mean spatial and temporal RMSE are only reduced by 34 % and

28 %, respectively). They clearly show the LEnKS efficiency decreases when the time lag is much higher than the mean number of observations per repeat period. This Table also shows that, for this specific orbit, the EnKF yields comparable results to the LEnKF; this result is quite different than for the fast sampling phase, where the EnKF led to significant model degradation. The difference is due to the fact that for the nominal orbit, there are fewer observations, leading to less-frequent updates. Thus, the analysis scheme effectively puts less weight on the observations and more weight on the model. This, in turn, decreases the effect of the spurious, long-distance correlations on the EnKF. Thus, in this study, the localization is more important for the fast sampling period than for the nominal orbit. Water depths along the river channel versus time for the corrected ensemble mean obtained after using this assimilation scheme are presented in Figure 5f.

As the Ob River is located in the boreal region at high latitudes, there are many observations within the 22 days repeat cycle (Biancamaria *et al.*, 2010); since the whole study domain is observed, both the downstream and upstream part of the river are corrected. Therefore the mean spatial and temporal RMSE are better than those obtained with fast sampling orbits 1 and 2. However, it is worth noticing that the downstream part is more frequently observed than the upstream part (Figure 2d). This explains why: 1) the variability in the water depth in June and July near Belogorje after assimilating observations from the nominal orbit (Figure 5f) is higher than after assimilating observations from the fast sampling orbits (Figures 5c, 5d and 5e) and 2), the RMSE is higher when using observations from the nominal orbit than ones from the fast sampling orbit 3, as the upstream part of the river is observed less often by the nominal orbit. These results tend to show that, for Arctic rivers, the SWOT nominal orbit has sufficiently good temporal and spatial coverage to significantly decrease modeled water depth errors. Thus, this suggests that, at this basin scale and latitude, spatial coverage is more important for correcting the model than temporal frequency of observations.

6. Conclusion

In this study, we investigated the potential of future wide swath altimetry data to decrease errors in water depths in a coupled 1D/2D hydraulic model. In particular, virtual observations of the future SWOT mission were computed and assimilated in an Arctic river hydrodynamic model using a local Ensemble Kalman Smoother. The results are very promising. For the fast sampling phase (first three months) of the mission the virtual SWOT observations decrease the mean spatial RMSE on modeled channel water depth by between 29 % and 79 % and the mean temporal RMSE by between 54 % and 91 % depending on the orbit phase compared to the RMSE with no assimilation. For the nominal phase of the mission, the mean spatial and temporal RMSE in modeled channel water depth are reduced by 59 % and 66 %, respectively. These results depend highly on the temporal and spatial coverage and thus are expected to be different at lower latitudes, where there will be fewer observations per repeat cycle for the nominal orbit. For example, low latitudes rivers like the Amazon, Brahmaputra and Ganges rivers, which flow more perpendicularly to the orbit, will only be seen two or three times per repeat period. Therefore, lower error reduction after the assimilation process is expected. Of course, huge rivers, like the Amazon, will be less impacted than smaller watersheds because the temporal persistence of corrections should last longer due to lower sensitivity to small-scale meteorological events.

This study has only considered modeling errors due to uncertainties in precipitation and temperature. Even if meteorological forcing is a main source of error, other sources should be considered in future studies. These could include uncertainty in the river bathymetry and errors in the model parameters, such as Manning's roughness coefficient for LISFLOOD-FP, snow on vegetation and drainage parameters for ISBA; see Biancamaria *et al.* (2009) for a discussion of model parameter error. The generation of virtual SWOT observations could also be improved by adding errors that have not yet been considered in this study, such as errors due to satellite shifts (especially uncorrected rolling) and impact of environmental effects, such as delays due to the wet troposphere. Nevertheless, this study shows the potential utility of SWOT observations to improve our understanding of spatial and temporal variations of surface runoff in sparsely gauged Arctic regions.

Acknowledgements

Some authors were funded by the TOSCA SWOT High Resolution Hydrology project from CNES and the MONARCH-A project from the European Union 7th Framework Program. Some participants in this project were funded by NASA’s programs in Physical Oceanography and Terrestrial Hydrology. OSU authors were also funded by their Climate, Water, and Carbon Program. One of the authors (S. Biancamaria) was supported by a CNES/Noveltis grant.

We gratefully acknowledge the staff and computing resources provided by the Ohio Supercomputer Center (OSC; www.osc.edu).

References

Alsdorf, D. E., Rodríguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, 45(2), RG2002 doi:10.1029/2006RG000197.

Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., & Alsdorf, D. E. (2007). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, 34, L10403, doi:10.1029/2007GL029721.

Auclair, F., Marsaleix, P., & De Mey, P. (2003). Space-time structure and dynamics of the forecast error in a coastal circulation model of the Gulf of Lions. *Dynamics of Atmospheres and Oceans*, 36, 309-346.

Bates, P. D., & De Roo, A. (2000). A simple raster-based model for floodplain inundation. *Journal of Hydrology*, 236(1-2), 54-77.

Biancamaria, S., Bates, P. D., Boone, A., & Mognard, N. M. (2009). Large-scale coupled hydrologic and hydraulic modelling of an arctic river: the Ob river in Siberia. *Journal of Hydrology*, 379(1-2), 136-150.

Biancamaria, S., Andreadis, K. M., Durand, M., Clark, E. A., Rodríguez, E., Mognard, N. M., Alsdorf, D. E., Lettenmaier, D. P., & Oudin, Y. (2010). Preliminary characterization of SWOT hydrology error budget and global capabilities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(1), doi:10.1109/JSTARS.2009.2034614.

Brown, J., Ferrians, O. J. Jr., Heginbottom, J. A., & Melnikov, E. S. (1998, revised February 2001). Circum-Arctic map of permafrost and ground-ice conditions. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology, Digital Media.

Connor, R., Faurès, J.-M., Kuylenstierna, J., Margat J., Steduto, P., Vallée, D., & van der Hoek, W. (2009). Chapter 7: Evolution of water use. In World Water Assesment Programme, *The United Nations World Water Development Report 3: water in a changing world* (pp. 100). Paris: UNESCO Publishing, and London: Earthscan.

Durand, M., Andreadis, K. M., Alsdorf, D. E., Lettenmaier, D. P., Moller, D., & Wilson, M. D. (2008). Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophysical Research Letters*, 35, L20401, doi:10.1029/2008GL034150.

Durand, M., Fu, L. L., Lettenmaier, D. P., Alsdorf, D. E., Rodríguez, E., & Fernandez, D. E. (2010). The Surface Water and Ocean Topography Mission: Observing Terrestrial Surface Water and Oceanic Submesoscale Eddies. *Proceedings Of the IEEE*, 98(5), 766-779.

Enjolras, V., Vincent, P., Souyris, J.-C., Rodríguez, E., Phalippou, L., & Cazenave, A. (2006). Performances study of interferometric radar altimeters: from the instrument to the global mission definition. *Sensors*, 6, 164-192.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143-10162.

Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54(6), 539-560.

Gaspari, G., & Cohn, S.E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554)**,** 723-757.

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, 129(11), 2776-2790.

Herschy, R. W., & Fairbridge, R. W. (1998). *Encyclopedia of Hydrology and Water Resources*. (1st ed., pp. 572-583). Dordrecht: Kluwer Academic Publishers.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82(Series D), 35-45.

Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Transactions of the ASME - Journal of Basic Engineering*, 83(Series D), 95-107.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., & Yang, S.-K. (2002). NCEP/DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*, 83(11), 1631-1643.

Moore, J. B. (1973). Discrete-time fixed-lag smoothing algorithms. *Automatica*, 9, 163-173.

Noilhan, J., & Mahfouf, J.-F. (1996). The ISBA land surface parametrisation scheme. *Global and Planetary Change*, 13(1-4), 145-159.

Pavelsky, T. M., & Smith, L. C. (2004). Spatial and temporal patterns in Arctic river ice breakup observed with MODIS and AVHRR time series. *Remote Sensing of Environment*, 93(3), 328-338.

Rodríguez, E. (2009). SWOT science requirements Document. JPL document, Initial release, (http://swot.jpl.nasa.gov/files/SWOT_science_reqs_final.pdf).

Serreze, M., Barrett, A., & Lo, F. (2005). Northern High-Latitude Precipitation as Depicted by Atmospheric Reanalysis and Satellite Retrievals. *Monthly Weather Review*, 133(12), 3407-3430.

Toumazou, V., & Crétaux, J.-F. (2001). Using a Lanczos Eigensolver in the Computation of Empirical Orthogonal Functions. *Monthly Weather Review*, 129(5), 1243-1250.

Trigg, M. A., Wilson, M. D., Bates, P. D., Horritt, M. S., Alsdorf, D. E., Forsberg, B. R., & Vega, M. C. (2009). Amazon flood wave hydraulics. *Journal of Hydrology*, 374(1-2), 92-105.

Voisin, N., Wood, A., & Lettenmaier, D. P. (2008). Evaluation of Precipitation Products for Global Hydrological Prediction. *Journal of Hydrometeorology*, 9(3), 388-407.

Wilson, M. D., Bates, P. D., Alsdorf, D. E., Forsberg, B., Horritt, M., Melack, J., Frappart, F., & Famiglietti, J. (2007). Modeling large-scale inundation of Amazonian seasonally flooded wetlands. *Geophysical Research Letters*, 34, L15404,

Yang, D., Ye, B., & Shiklomanov, A. (2004). Discharge Characteristics and Changes over the Ob River Watershed in Siberia. *Journal of Hydrometeorology*, 5(4), 595-610.

Table captions

Table 1. Explained variance for the first ten EOF modes of the total precipitation (rain + snow) and air temperature

Table 2. Mean spatial and temporal RMSE in channel water depth between the truth and the ensemble mean with and without assimilating SWOT observations for the fast sampling orbits. When there is assimilation, the percentage of error reduction compared to no assimilation (see equation 12) is indicated in parentheses. Dashes in the table represent assimilation runs when the updates have so much degraded the LISFLOOD-FP model that the model was forced to stop running (e.g., if the river bed became dry).

Table 3. Mean spatial and temporal RMSE in channel water depth between the truth and the ensemble mean with and without assimilating SWOT observations for the nominal orbit. When there is assimilation, the percentage of error reduction compared to no assimilation (see equation 12) is indicated in parentheses.

Figure captions

Figure 1. Study domain (Lower Ob). Red arrows represent the boundary conditions (lateral inflows and incoming flow).

Figure 2. Number of observations for the three selected fast sampling orbits (a., b. and c.) and for the nominal orbit (d.) during one repeat cycle (3 days for the fast sampling orbits and 22 days for the nominal orbits).

Figure 3. Lower Ob water elevations (above OSU91A geoid) for June 28th 1993 from the initial modeling (the “truth”).

Figure 4. Water depth along the Lower Ob versus channel distance from Belogorje for the truth (blue line) and all members of the ensemble with no assimilation (red lines) for June 28th 1993.

Figure 5. Water height (in m) along the river channel (y-axis) versus time (x-axis) for the truth (a.), the ensemble mean with no assimilation (b.), the ensemble mean after assimilation using the LEnKS, with a 2 day time lag, for the SWOT fast sampling orbits number 1 (c.), number 2 (d.) and number 3 (e.), and the ensemble mean after assimilation using the LenKS, with a 3 day time lag, for the SWOT nominal orbit (f.).

Tables

Table 1.

EOF Modes	1	2	3	4	5	6	7	8	9	10
Total precipitation explained variance (%)	8.9	4.8	4.7	3.9	3.2	2.9	2.6	2.4	2.2	2.1
Air temperature explained variance (%)	84.1	3.6	2.5	1.5	1.3	1.2	0.6	0.5	0.4	0.4

Table 2.

		Mean spatial RMSE (m)	Mean temporal RMSE (m)
No assimilation		0.80	1.11
EnKF	Orbit 1	-	-
	Orbit 2	-	-
	Orbit 3	-	-
LEnKF	Orbit 1	0.61 (24 %)	0.62 (44 %)
	Orbit 2	0.43 (46 %)	0.50 (55 %)
	Orbit 3	0.24 (70 %)	0.21 (81 %)
LEnKS (time lag = 2 days)	Orbit 1	0.57 (29 %)	0.51 (54 %)
	Orbit 2	0.40 (50 %)	0.44 (60 %)
	Orbit 3	0.17 (79 %)	0.10 (91 %)
LEnKS (time lag = 3days)	Orbit 1	0.59 (26 %)	0.57 (49 %)
	Orbit 2	0.43 (46 %)	0.49 (56 %)
	Orbit 3	0.19 (76 %)	0.15 (87 %)
LEnKS (time lag = 5 days)	Orbit 1	0.58 (28 %)	0.55 (51 %)
	Orbit 2	0.44 (45 %)	0.51 (54 %)
	Orbit 3	0.21 (74 %)	0.18 (84 %)

Table 3.

	Mean spatial RMSE (m)	Mean temporal RMSE (m)
No assimilation	0.80	1.11
EnKF	0.39 (51 %)	0.39 (65 %)
LEnKF	0.45 (44 %)	0.55 (51 %)
LEnKS (time lag = 2 days)	0.36 (55 %)	0.42 (62 %)
LEnKS (time lag = 3 days)	0.33 (59 %)	0.38 (66 %)
LEnKS (time lag = 5 days)	0.37 (54 %)	0.45 (60 %)
LEnKS (time lag = 10 days)	0.53 (34 %)	0.80 (28 %)

Figures

Figure 1

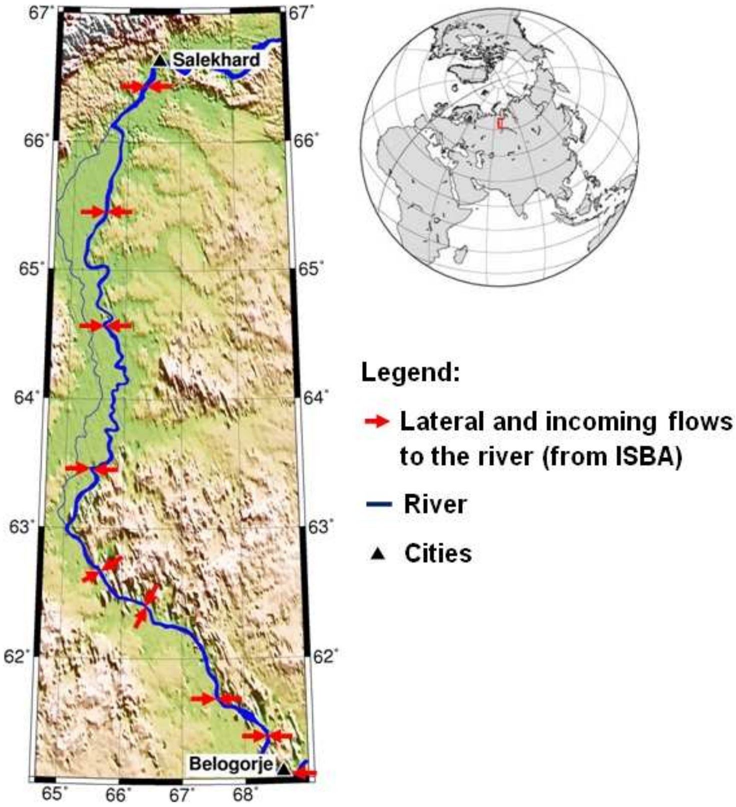
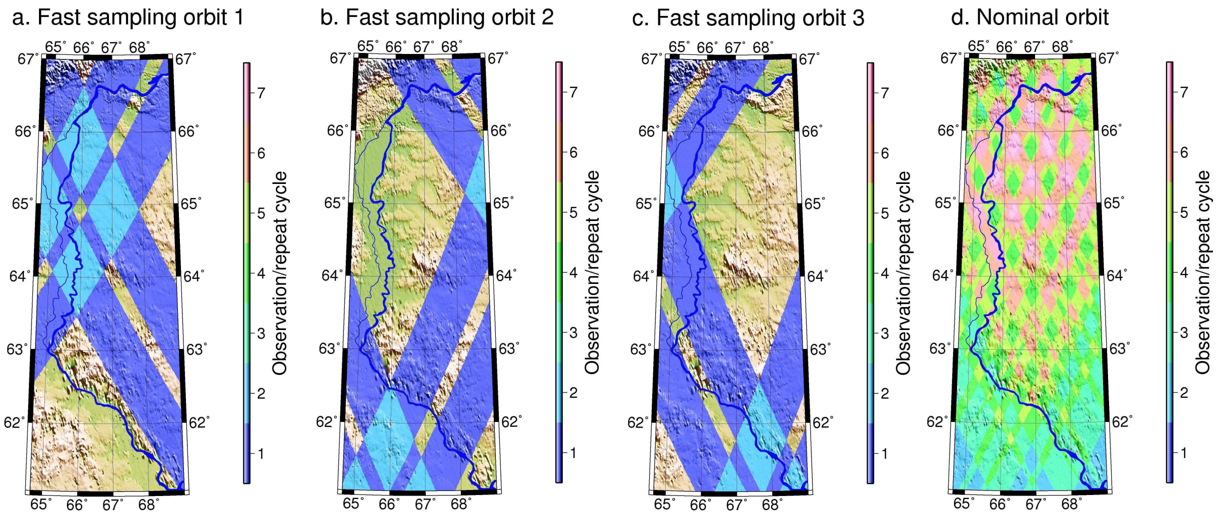
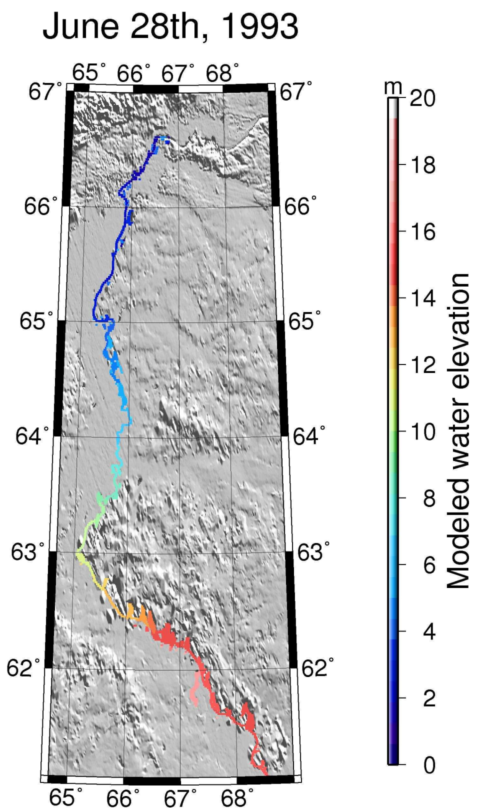


Figure 2


a. Fast sampling orbit 1
b. Fast sampling orbit 2
c. Fast sampling orbit 3
d. Nominal orbit
65° 66° 67° 68°
67° 66° 65° 64° 63° 62°
Observation/repeat cycle
1 2 3 4 5 6 7

587 Figure 3



588

Figure 4

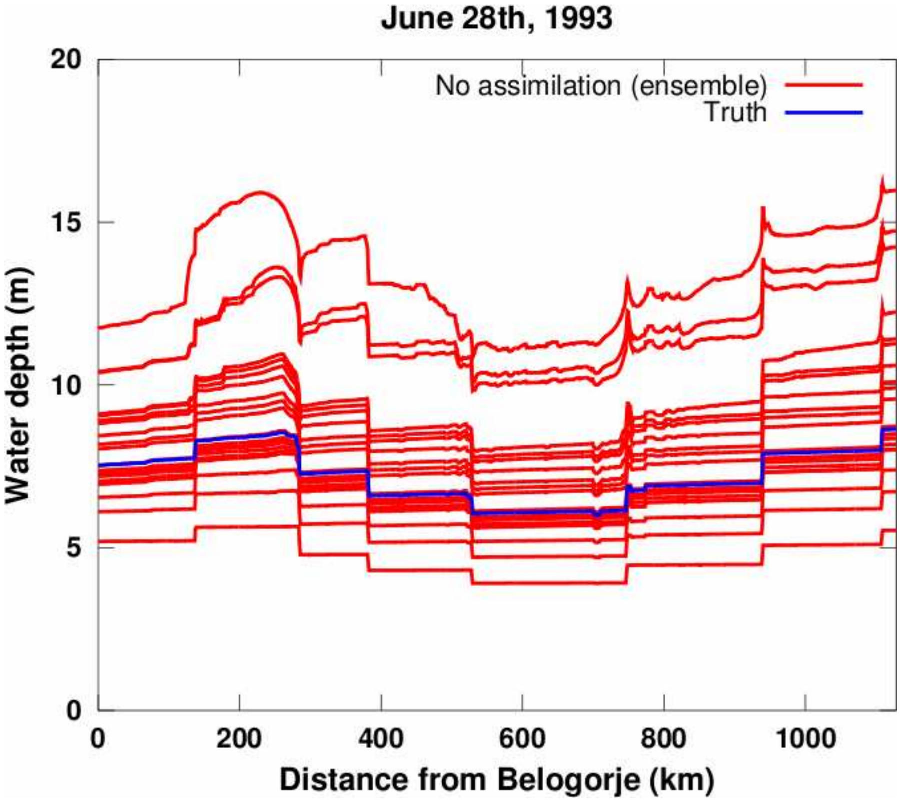
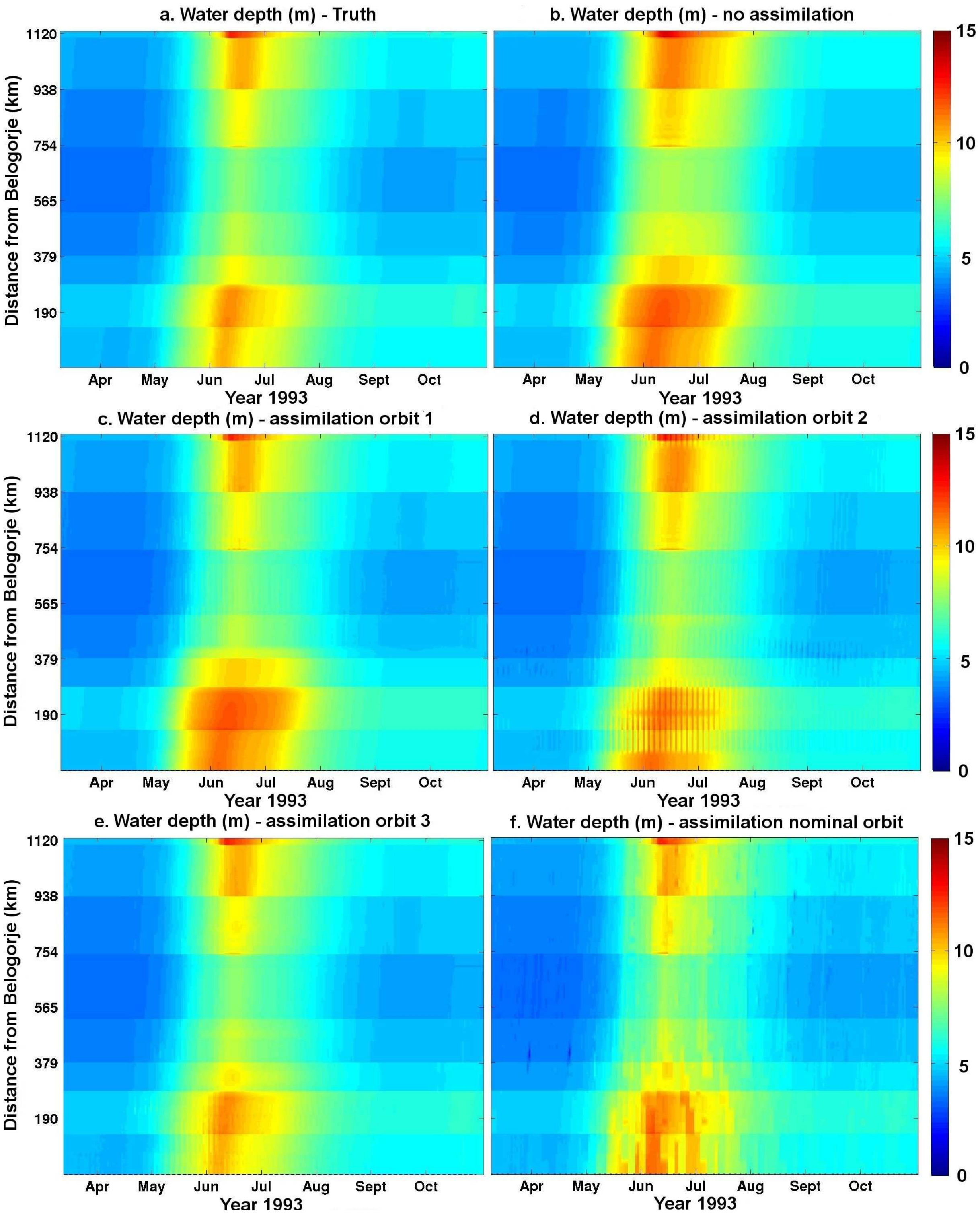

June 28th, 1993
No assimilation (ensemble)
Truth
Water depth (m)
Distance from Belogorje (km)
0
5
10
15
20
0
200
400
600
800
1000

Figure 5



a. Water depth (m) - Truth
b. Water depth (m) - no assimilation
c. Water depth (m) - assimilation orbit 1
d. Water depth (m) - assimilation orbit 2
e. Water depth (m) - assimilation orbit 3
f. Water depth (m) - assimilation nominal orbit
Distance from Belogorje (km)
1120
938
754
565
379
190
Apr
May
Jun
Jul
Aug
Sept
Oct
Year 1993
15
10
5
0