



HAL
open science

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

► To cite this version:

Sylvain Biancamaria, Michael Durand, Kostantinos M. Andreadis, Paul Bates, Aaron Boone, et al..
Assimilation of virtual wide swath altimetry to improve Arctic river modeling. *Remote Sensing of
Environment*, 2011, 115 (2), pp.373-381. 10.1016/j.rse.2010.09.008 . hal-00575528

HAL Id: hal-00575528

<https://hal.science/hal-00575528>

Submitted on 10 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Assimilation of virtual wide swath altimetry to improve Arctic river modeling

2

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

5

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

8 ^b CNES; LEGOS, F-31400 Toulouse, France

9 ^c CNRS; LEGOS, F-31400 Toulouse, France

10 ^d Ohio State University; School of Earth Sciences; Columbus OH 43210, USA

11 ^e Ohio State University; Byrd Polar Research Center; Columbus OH 43210, USA

12 ^f School of Geographical Sciences; University of Bristol; Bristol BS8 1SS, UK

13 ^g GAME/CNRM; Météo-France; CNRS; F-31057 Toulouse Cedex, France

14 ^h Jet Propulsion Laboratory; California Institute of Technology; Pasadena CA 91109, USA

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

16 ^j Civil and Environmental Engineering; University of Washington; Seattle WA 98195, USA

17

18 * Corresponding author: sylvain.biancamaria@legos.obs-mip.fr (Postal address: LEGOS, 14
19 avenue Edouard Belin, 31400 Toulouse, France; Phone: +335 61 33 29 30; Fax: +335 61 25
20 32 05)

21 Abstract

22 Global surface water variations are still difficult to monitor with current satellite
23 measurements. The future Surface Water and Ocean Topography (SWOT) mission is
24 designed to address this issue. Its main payload will be a wide swath altimeter which will
25 provide maps of water surface elevations between 78°S and 78°N over a 120 km swath. This
26 study aims to combine coupled hydrologic/hydraulic modeling of an Arctic river with virtual
27 SWOT observations using a local ensemble Kalman smoother to characterize river water
28 depth variations. We assumed that modeling errors are only due to uncertainties in
29 atmospheric forcing fields (precipitation and air temperature) and different SWOT orbits were
30 tested. First, we tested orbits that all have a three day repeat period but differ in terms of their
31 spatial coverage of the study reach; these orbits correspond to the first three months of the
32 mission, which will be dedicated to calibration and validation experiments. For these orbits,
33 the mean spatial Root Mean Square Error (RMSE) in modeled channel water depth decreased
34 by between 29 % and 79 % compared to the modeled RMSE with no assimilation, depending
35 on the spatial coverage. The corresponding mean temporal RMSE decrease was between 54
36 % and 91 %. We then tested the nominal orbit with a twenty two day repeat period which will
37 be used during the remaining lifetime of the mission. Unlike the three day repeat orbits, this
38 orbit will observe all continental surfaces (except Antarctica and the northern part of
39 Greenland) during one repeat period. The assimilation of SWOT observations computed with
40 this nominal orbit into the hydraulic model leads to a decrease of 59 % and 66 % in the mean
41 spatial and temporal RMSE in modeled channel water depth, respectively. These results show
42 the huge potential of the future SWOT mission for land surface hydrology, especially at high
43 latitudes which will be very well sampled during one orbit repeat period. Still, further work is
44 needed to reduce current modeling uncertainties and to better characterize SWOT
45 measurement errors.

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

48

49 1. Introduction

50 More than 73 % of water used for human activities (for example as drinking water, for
51 irrigation or for energy generation and industrial processes) comes from surface water
52 (Connor *et al.*, 2009). It is therefore crucial to observe and understand the spatial and
53 temporal variations in surface water across the globe. Accordingly, in-situ gage networks
54 have been intensively developed since the second part of the twentieth century. However,
55 these networks are still sparse, especially in remote regions like in the Arctic, and in many
56 areas the coverage is actually now declining. To overcome this issue, hydrologic models and
57 remote sensing data have been used to complement in-situ measurements. However, current
58 remote sensing observations of water surface elevations made by nadir altimeters, which only
59 measure water elevations along the track of the satellite with typical track spacing of ~120
60 km, miss many of the world's surface water bodies, have relatively large spatial footprints (on
61 the order of 5-10 km) and do not give any information about water extent (Alsdorf *et al.*,
62 2007).

63 In order to better characterize surface water and oceanic processes, a wide swath altimeter,
64 the Surface Water and Ocean Topography (SWOT) mission, is currently under study by
65 NASA (National Aeronautics and Space Administration) and CNES (Centre National
66 d'Etudes Spatiales). SWOT will provide maps of water elevation at an unprecedented spatial
67 resolution (on the order of 50-100m) and precision (centimetric accuracy when averaged over
68 areas of 1 km²; Durand *et al.*, 2010). A small number of recent studies have begun to quantify
69 the benefits of such a mission for land surface hydrology. Biancamaria *et al.* (2010) focus on

70 the benefits of this mission at a global scale for different orbits and show that errors in
71 instantaneous discharge estimated from SWOT measurements using rating curve should be
72 below 25 % for rivers wider than 50m. Errors only due to the SWOT temporal sampling on
73 monthly discharge should be below 20 % for rivers with drainage areas larger than 7000 km².
74 Andreadis *et al.* (2007) estimated the benefit of assimilating virtual wide swath
75 measurements, using an Ensemble Kalman filter (EnKF), to reduce modeling errors due to
76 uncertainties on lateral inflows of a mid-latitude river (a segment of the Ohio River). This
77 study compared three different orbits with 8, 16 and 32 days repeat period. They showed that
78 relative errors could be reduced by nearly a factor of two when the filter is used; the best
79 results were obtained for the orbit with the smallest repeat period. Durand *et al.* (2008)
80 assimilated virtual SWOT observations into the Amazon River hydraulic modeling developed
81 by Wilson *et al.* (2007) for estimating bathymetric depths and slopes. They showed that
82 bathymetric slopes can be estimated to within 0.30 cm.km⁻¹ and depths to within 56 cm
83 (which was 84 % less than errors without assimilation). They also highlighted that, in their
84 modeling, model errors dominate over measurement errors and therefore estimates of channel
85 bathymetry are relatively insensitive to measurement error characteristics.

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

94 river modeling used has a 1 km x 1 km spatial resolution (due to the current lack of a high
95 spatial resolution digital elevation model above 60°N).

96

97 2. Study domain and river modeling

98 This study focuses on the Lower Ob River between the cities of Belogorje and
99 Salekhard; this reach covers the downstream 1120 km of the river before the Ob estuary
100 (Figure 1) and corresponds to a drainage area of 790 000 km² (according to the Arctic Rapid
101 Integrated Monitoring System, ArcticRIMS, <http://rims.unh.edu>). The drainage basin of the
102 entire Ob river covers 2 990 000 km² and it is located in Western Siberia, east of the Ural
103 Mountains. In terms of discharge, the Ob is the world's 12th largest river and the 3rd largest in
104 the Arctic (Herschy and Fairbridge, 1998). The Ob is frozen from November to April and its
105 discharge regime is mainly driven by snow melt, with a maximum in May/June during ice
106 breakup (Pavelsky and Smith, 2004). Yang *et al.* (2004) reported that from 1936 to 1990 the
107 monthly mean discharge at the river outlet varied between 500 and 1200 m³.s⁻¹ in the cold
108 season (from November to April), and between 3500 and 9000 m³.s⁻¹ during the summer
109 months. The land cover in this domain is classified as sporadic and discontinuous permafrost
110 (Brown *et al.*, 1998). According to Yang *et al.* (2004), the effects of human activities on the
111 study domain are limited and there are no reservoirs on the lower part of the river. In this
112 study, the modeled time period corresponds to the calendar year 1993.

113 The river is modeled by the flood inundation model LISFLOOD-FP developed at the
114 University of Bristol, UK (Bates and De Roo, 2000). LISFLOOD-FP is a coupled 1D/2D
115 hydraulic model based on a raster grid. It predicts water depth in each grid cell at each time
116 step and hence can simulate the dynamic propagation of flood waves over fluvial, coastal and
117 estuarine floodplains. Here, the 1D channel flow is based on the kinematic approximation to

118 the 1D St Venant equations. Floodplain flows are similarly described in terms of continuity
119 and momentum equations, discretized over a grid of square cells, which allows the model to
120 represent 2-D dynamic flow fields on the floodplain. There is, however, no exchange of
121 momentum between main channel and floodplain flows, only mass, and ice jam and break up
122 processes are not represented. The kinematic approximation of the channel flow might also be
123 a limitation of the modeling, as the Ob flow regime is likely diffusive, at least in the
124 downstream part of the river. However, according to Trigg et al. (2009), for the Amazon
125 River, this approximation leads to an additional Root Mean Square Error (RMSE) of around 1
126 m. This error is likely to be lower for the Ob and is much smaller than errors on the floodplain
127 topography and river bathymetry. Finally, backwater effects from inflows are not modeled,
128 but are likely to be minor given the ratio of inflow volume to the mainstem discharge.

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

143

144 3. Satellite observations

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

149 3.1. The SWOT mission

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

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

167 2, Figure 2b) does not observe the downstream part of the river. The third orbit (orbit 3,
168 Figure 2c) corresponds to an optimal coverage, as almost all the river inside the study domain
169 is seen. These three orbits thus represent the likely envelope of sampling scenarios possible
170 for a given Arctic river basin during the fast sampling period.

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

176 3.2. Generation of virtual satellite observations

177 Virtual SWOT observations were generated by first computing the swath coverage over the
178 study domain for both the nominal and fast sampling orbits for each day during the year 1993.
179 The initial modeling of the Lower Ob (see Biancamaria *et al.*, 2009) was taken as the “true”
180 state and was used to compute SWOT measurements. As an example of model outputs, water
181 elevations computed with this modeling for June 28th 1993 is shown in Figure 3. First,
182 modeled water elevations from the nominal LISFLOOD-FP model inside the SWOT swath
183 were selected and then observation errors were added. For the moment, only the instrumental
184 error is taken into account (the errors due to satellite position uncertainties, atmospheric
185 effects such as wet troposphere delays, etc are not considered). Instrument error was modeled
186 by a white noise of 2 cm standard deviation, which corresponds to the expected error of
187 KaRIN (Enjolras *et al.*, 2006; Rodríguez, 2009) at the 1 km² resolution of the Lower Ob
188 model.

189

190 4. Methodology

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

197 4.1. Assimilation schemes

198 4.1.1. Ensemble Kalman Filter

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

$$203 \quad x^a = x^f + P^f H^T (H P^f H^T + R)^{-1} (y - H x^f) \quad \text{eq. 1}$$

$$204 \quad P^a = P^f - P^f H^T (H P^f H^T + R)^{-1} H P^f \quad \text{eq. 2}$$

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

210 The analysis computed using the Kalman filter is thus a weighted average of the forecast and
211 the observation. The weight optimally takes into account the error in the forecast and in the
212 observation. The Kalman Filter is a sequential filter, which means that the analysis is
213 computed only at times when an observation is available. It is then used as an updated initial

214 condition for the model, which is subsequently run forward from this updated initial condition
 215 until a new observation is available. If the observation operator and model equations are linear
 216 and if the forecast and observation errors are zero mean Gaussian random vectors, then the
 217 analysis obtained with the Kalman Filter is the best linear unbiased estimate of the model state
 218 and its error covariance matrix. In this study, SWOT observations correspond to water height
 219 measurements along the river, therefore the observation operator is simply a mask of the
 220 swath coverage and is a linear operator. However, the model equations are not linear, as is the
 221 case for many fluid flow problems.

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

$$225 \quad P^f = \overline{(x^f - x^t)(x^f - x^t)^T} \quad \text{eq. 3}$$

$$226 \quad P^a = \overline{(x^a - x^t)(x^a - x^t)^T} \quad \text{eq. 4}$$

$$227 \quad R = \overline{(y - H x^t)(y - H x^t)^T} \quad \text{eq. 5}$$

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

236
$$P^f \approx P_e^f = \overline{(x^f - \overline{x^f})(x^f - \overline{x^f})^T}$$
 eq. 6

237
$$P^a \approx P_e^a = \overline{(x^a - \overline{x^a})(x^a - \overline{x^a})^T}$$
 eq. 7

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

240 4.1.2. Local Ensemble Kalman Filter

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

249
$$x^a = x^f + [S \times (P_e^f H^T)] \{H[S \times (P_e^f H^T)] + R\}^{-1} (y - H x^f)$$
 eq. 8

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

258 observation is reduced. However, this length scale will prevent any spurious long range
 259 correlations and thus avoid unrealistic water depth computations.

260 4.1.3. Local Ensemble Kalman Smoother

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

$$272 \quad x_j^a = x_j^f + [S \times (P_{eij}^f H^T)] \{ H[S \times (P_{eii}^f H^T)] + R_i \}^{-1} (y_i - H x_i^f) \quad \text{eq. 9}$$

$$273 \quad P_{eij}^f = \overline{(x_i^f - \overline{x_i^f})(x_j^f - \overline{x_j^f})^T} \quad \text{eq. 10}$$

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

279 4.2. Ensemble member generation

280 The ensemble used in the LEnKF and LEnKS should be representative of all model errors.
281 Possible sources of errors include initial conditions, forcing data, model parameters and
282 model equations used. In this study, only errors from ISBA forcing data (precipitation and
283 temperature) were considered, as these are the primary source of errors in the modeling
284 (Biancamaria *et al.*, 2009). These forcing data come from NCEP-DOE AMIP II reanalysis
285 (National Centers for Environmental Prediction - Department Of Energy, Atmospheric Model
286 Intercomparison Project; Kanamitsu *et al.*, 2002). Errors in reanalysis products are always
287 difficult to estimate, as very few high quality global products exist for comparison, especially
288 at high latitudes. However, Serreze *et al.* (2005) show that for the Ob basin correlations
289 between NCEP monthly total precipitation over the river basin and in-situ measurements vary
290 from 0.60 to 0.86 depending on the month, for the time span 1979/1993. Biancamaria *et al.*
291 (2009) found that downstream Ob discharge modeled using NCEP precipitation has an error
292 of 14 % compared to in-situ discharge time series. In addition, Voisin *et al.* (2008) found that
293 precipitation from another reanalysis (ERA-40 from the European Centre for Medium-Range
294 Weather Forecasts) has errors between 0.7 % and 34.5 % on Eastern Siberian Rivers.
295 Therefore, we have assumed that errors on precipitation are 20 %. The standard deviation
296 between daily air temperature from NCEP and in-situ measurement at Belogorje for 1993 is
297 equal to 0.18. Thus, the error on air temperature has also been set to 20 %.

298 4.2.1. Methodology

299 Members of the ensemble correspond to a “corrupted” version of the nominal forcing data
300 (considered to be the truth). The methodology used has been previously developed by Auclair
301 *et al.* (2003) and consists of perturbing the most statistically significant modes of the
302 atmospheric fields. To do so Empirical Orthogonal Functions (EOF) of the atmospheric field
303 temporal anomaly were computed, and the corrupted field (P^{corrupt}) was obtained by

304 recombining the first EOF modes which explained 95 % of the variance and the temporal
305 atmospheric field mean, multiplied by white noise (equation 11).

$$306 \quad P^{corrupt}(l,t) = \bar{P}(l) \cdot \epsilon_m + \sum_{k=1}^N \epsilon_k \cdot \alpha_k(t) \cdot \phi_k(l) \quad \text{eq. 11}$$

307 In equation 11, l is the spatial index, t the temporal index, \bar{P} is the temporal mean, k is the
308 EOF mode, N is the highest EOF mode used, α_k is the temporal component and ϕ_k is the
309 spatial component of the EOF for the mode k , ϵ_m is the noise on the mean and ϵ_k is the noise
310 on the EOF recombination for the mode k . ϵ_m and ϵ_k are both white noise with a 0.2 standard
311 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
312 space and time). The last mode N was chosen so that the cumulative explained variance for
313 modes 1 to N is equal to 95 %.

314 4.2.2. Corrupted precipitation and air temperature

315 The EOF modes were computed using the algorithm developed by Toumazou and Crétau
316 (2001), for the total precipitation field (rain rate + snow rate) and air temperature. Table 1
317 presents the explained variance for the first 10 EOF modes of these two atmospheric forcings.
318 As there is no seasonal cycle in the total precipitation (the mean life time of a depression is
319 roughly around a week, with no strong seasonality), 187 EOF modes are required to explain
320 95 % of the variance. On the contrary, for air temperature, most of the energy is included in
321 the first 8 modes (which explain 95 % of the variance). The first mode itself (corresponding to
322 the seasonal cycle) explains 84 % of the variance. For these two ISBA inputs, EOFs were
323 computed from August 1991 to July 1995, using the methodology presented in section 4.2.1.
324 For computational reasons, the size of the ensemble was set to 20 (thus 20 corrupted
325 precipitation and temperature fields have been computed).

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

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

339

340 5. Results

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

346 5.1. Fast sampling orbits

347 Table 2 presents the mean spatial and temporal RMSE between the truth and the ensemble
348 mean with and without assimilating SWOT observations for these three orbits. For these

349 orbits, there is an observation every one or two days, depending on the location. Therefore the
350 best results are obtained using a 2 day lag time LEnKS. The percentage of error reduction
351 compared to no assimilation (equation 12) is also indicated in Table 2.

$$352 \quad \varepsilon = 100 \cdot \frac{RMSE_{no\ assimilation} - RMSE_{assimilation}}{RMSE_{no\ assimilation}} \quad \text{eq. 12}$$

353 Model errors after assimilation are highly dependent on the location of the observations and
354 are, therefore, quite different for each orbit phase. The mean spatial RMSE was decreased
355 from between 29 % to 79 % and the mean temporal RMSE was decreased by between 54 %
356 and 91 % for the 2 day time lag LEnKS. In particular, fast sampling orbit 2 (Figure 2b)
357 observes a smaller portion of the river than orbit 1 (Figure 2a); however the mean spatial and
358 temporal RMSE after assimilating SWOT observations generated using orbit 2 are smaller
359 than after assimilating SWOT orbit 1 observations (Table 2). This is due to the location of the
360 ground track: orbit 2 observes the upstream part of the river, near Belogorje, where the
361 incoming flow to the study domain is located. The incoming streamflow is one order of
362 magnitude higher than the lateral inflows to the river from the study domain, as computed by
363 the ISBA model. When orbit 2 is used, the part of the river with the highest error (the
364 upstream) is well observed and thus well corrected; this correction propogates downstream,
365 even to unobserved river locations. Orbit 1, on the other hand, observes the downstream part
366 of the river; therefore the upstream part of the river, which is not seen, is not corrected. This
367 leads to higher errors than those obtained with orbit 2. This effect is obvious in Figures 5c, 5d
368 and 5e, which show water depths along the river channel versus time for the ensemble mean
369 after assimilating SWOT observations for the three fast sampling orbits using a 2 day time lag
370 LEnKS. Orbit 3 corresponds to the optimum orbit, as almost the entire river is observed every
371 three days. Consequently, the spatial and temporal RMSE for this orbit decrease by 79 % and
372 91 %, respectively, compared to the RMSE with no assimilation. It is important to note that

373 the main Arctic rivers (Mackenzie, Ob, Yenisey and Lena) are oriented South to North.
374 Therefore, if the SWOT fast sampling orbit is correctly chosen, at least some of these rivers
375 should be very well (if not entirely) observed.

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

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

390 5.2. Nominal orbit

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

397 28 %, respectively). They clearly show the LEnKS efficiency decreases when the time lag is
398 much higher than the mean number of observations per repeat period. This Table also shows
399 that, for this specific orbit, the EnKF yields comparable results to the LEnKF; this result is
400 quite different than for the fast sampling phase, where the EnKF led to significant model
401 degradation. The difference is due to the fact that for the nominal orbit, there are fewer
402 observations, leading to less-frequent updates. Thus, the analysis scheme effectively puts less
403 weight on the observations and more weight on the model. This, in turn, decreases the effect
404 of the spurious, long-distance correlations on the EnKF. Thus, in this study, the localization is
405 more important for the fast sampling period than for the nominal orbit. Water depths along the
406 river channel versus time for the corrected ensemble mean obtained after using this
407 assimilation scheme are presented in Figure 5f.

408 As the Ob River is located in the boreal region at high latitudes, there are many observations
409 within the 22 days repeat cycle (Biancamaria *et al.*, 2010); since the whole study domain is
410 observed, both the downstream and upstream part of the river are corrected. Therefore the
411 mean spatial and temporal RMSE are better than those obtained with fast sampling orbits 1
412 and 2. However, it is worth noticing that the downstream part is more frequently observed
413 than the upstream part (Figure 2d). This explains why: 1) the variability in the water depth in
414 June and July near Belogorje after assimilating observations from the nominal orbit (Figure
415 5f) is higher than after assimilating observations from the fast sampling orbits (Figures 5c, 5d
416 and 5e) and 2), the RMSE is higher when using observations from the nominal orbit than ones
417 from the fast sampling orbit 3, as the upstream part of the river is observed less often by the
418 nominal orbit. These results tend to show that, for Arctic rivers, the SWOT nominal orbit has
419 sufficiently good temporal and spatial coverage to significantly decrease modeled water depth
420 errors. Thus, this suggests that, at this basin scale and latitude, spatial coverage is more
421 important for correcting the model than temporal frequency of observations.

422

423 6. Conclusion

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

441 This study has only considered modeling errors due to uncertainties in precipitation and
442 temperature. Even if meteorological forcing is a main source of error, other sources should be
443 considered in future studies. These could include uncertainty in the river bathymetry and
444 errors in the model parameters, such as Manning's roughness coefficient for LISFLOOD-FP,
445 snow on vegetation and drainage parameters for ISBA; see Biancamaria *et al.* (2009) for a

446 discussion of model parameter error. The generation of virtual SWOT observations could also
447 be improved by adding errors that have not yet been considered in this study, such as errors
448 due to satellite shifts (especially uncorrected rolling) and impact of environmental effects,
449 such as delays due to the wet troposphere. Nevertheless, this study shows the potential utility
450 of SWOT observations to improve our understanding of spatial and temporal variations of
451 surface runoff in sparsely gauged Arctic regions.

452

453 Acknowledgements

454 Some authors were funded by the TOSCA SWOT High Resolution Hydrology project from
455 CNES and the MONARCH-A project from the European Union 7th Framework Program.

456 Some participants in this project were funded by NASA's programs in Physical
457 Oceanography and Terrestrial Hydrology. OSU authors were also funded by their Climate,
458 Water, and Carbon Program. One of the authors (S. Biancamaria) was supported by a
459 CNES/Noveltis grant.

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

462

463

464 References

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

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

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

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

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

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

483 Brown, J., Ferrians, O. J. Jr., Heginbottom, J. A., & Melnikov, E. S. (1998, revised February
484 2001). Circum-Arctic map of permafrost and ground-ice conditions. Boulder, CO: National
485 Snow and Ice Data Center/World Data Center for Glaciology, Digital Media.

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

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

494 Durand, M., Fu, L. L., Lettenmaier, D. P., Alsdorf, D. E., Rodríguez, E., & Fernandez, D. E.
495 (2010). The Surface Water and Ocean Topography Mission: Observing Terrestrial Surface
496 Water and Oceanic Submesoscale Eddies. *Proceedings Of the IEEE*, 98(5), 766-779.

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

500 Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model
501 using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research*,
502 99(C5), 10143-10162.

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

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

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

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

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

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

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

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

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

521 Pavelsky, T. M., & Smith, L. C. (2004). Spatial and temporal patterns in Arctic river ice
522 breakup observed with MODIS and AVHRR time series. *Remote Sensing of Environment*,
523 93(3), 328-338.

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

526 Serreze, M., Barrett, A., & Lo, F. (2005). Northern High-Latitude Precipitation as Depicted
527 by Atmospheric Reanalysis and Satellite Retrievals. *Monthly Weather Review*, 133(12), 3407-
528 3430.

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

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

- 533 Voisin, N., Wood, A., & Lettenmaier, D. P. (2008). Evaluation of Precipitation Products for
534 Global Hydrological Prediction. *Journal of Hydrometeorology*, 9(3), 388-407.
- 535 Wilson, M. D., Bates, P. D., Alsdorf, D. E., Forsberg, B., Horritt, M., Melack, J., Frappart, F.,
536 & Famiglietti, J. (2007). Modeling large-scale inundation of Amazonian seasonally flooded
537 wetlands. *Geophysical Research Letters*, 34, L15404,
- 538 Yang, D., Ye, B., & Shiklomanov, A. (2004). Discharge Characteristics and Changes over the
539 Ob River Watershed in Siberia. *Journal of Hydrometeorology*, 5(4), 595-610.

540 Table captions

541

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

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

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

554 Figure captions

555

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

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

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

563 Figure 4. Water depth along the Lower Ob versus channel distance from Belogorje for the
564 truth (blue line) and all members of the ensemble with no assimilation (red lines) for June 28th
565 1993.

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

571 Tables

572

573 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

574

575 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 %)

576

577

578

579 Table 3.

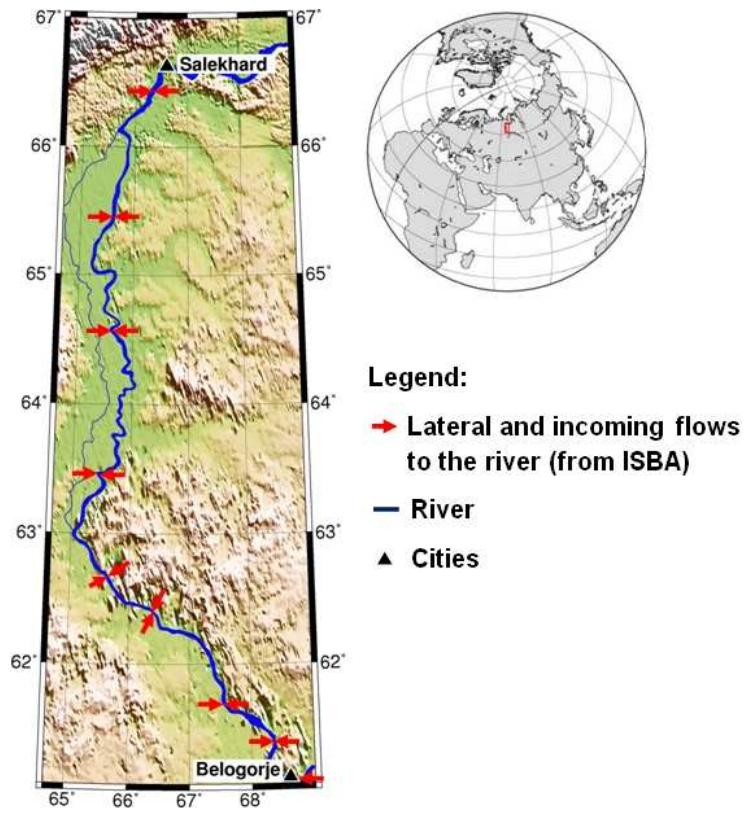
	Mean spatial RMSE (m)	Mean temporal RMSE (m)
No assimilation	0.80	1.11
EnKF	0.39 (51 %)	0.39 (65 %)
LEnKF	0.45 (44 %)	0.55 (51 %)
LEnKS (time lag = 2 days)	0.36 (55 %)	0.42 (62 %)
LEnKS (time lag = 3 days)	0.33 (59 %)	0.38 (66 %)
LEnKS (time lag = 5 days)	0.37 (54 %)	0.45 (60 %)
LEnKS (time lag = 10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

580

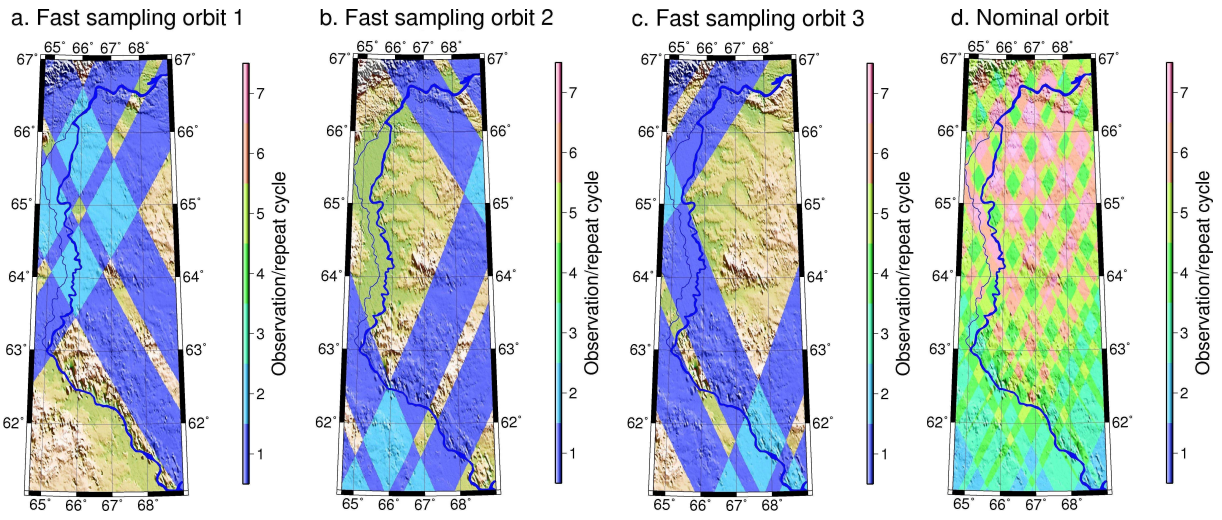
581 Figures

582

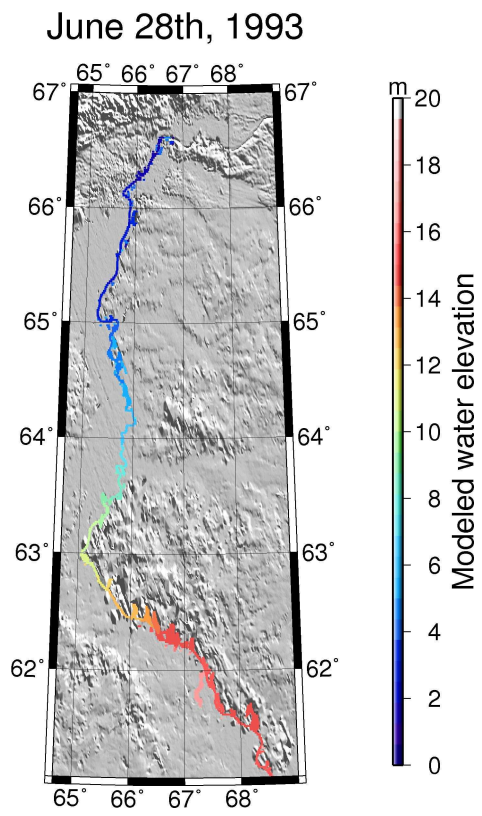
583 Figure 1



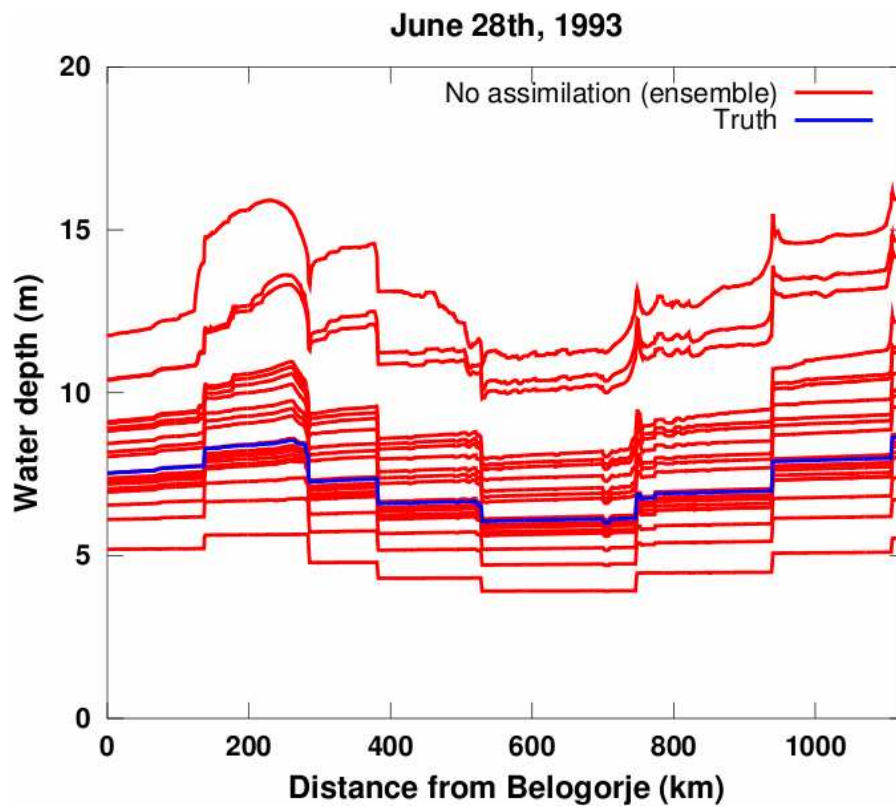
584



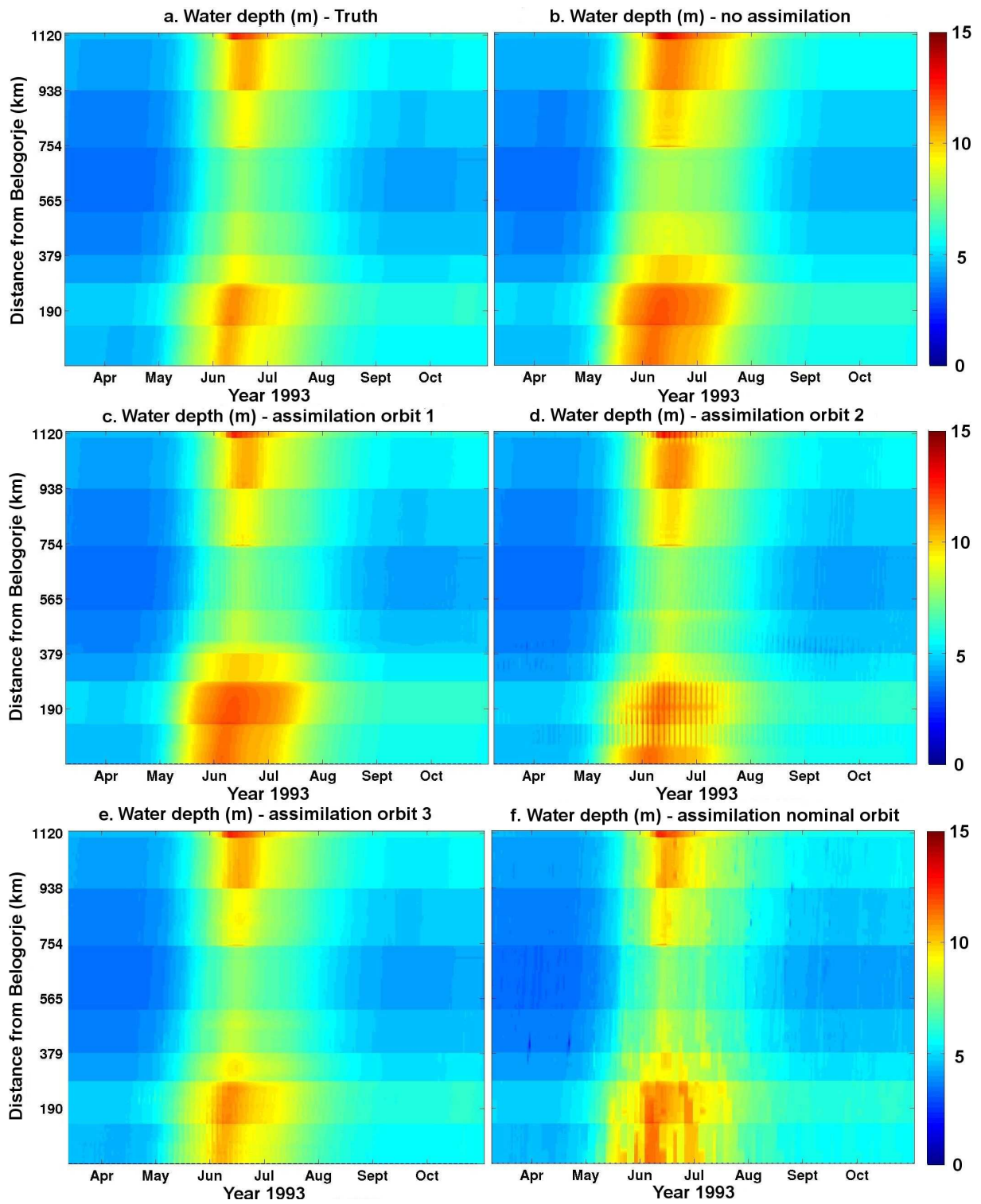
586



589 Figure 4



590



592

593

594