

Combination of Multi-Mission Altimetry Data Along the Mekong River with Spatio-Temporal Kriging

E. Boergens* S. Buhl† D. Dettmering* C. Klüppelberg† F. Seitz*

December 9, 2016

Abstract

River water-level time series at fixed geographical locations, so-called virtual stations, have been computed from single altimeter crossings for many years. Their temporal resolution is limited by the repeat cycle of the individual altimetry missions. The combination of all altimetry measurements along a river enables computing a water-level time series with improved temporal and spatial resolutions. This study uses the geostatistical method of spatio-temporal ordinary kriging to link multi-mission altimetry data along the Mekong River. The required covariance models reflecting the water flow are estimated based on empirical covariance values between altimetry observations at various locations. In this study, two covariance models are developed and tested in the case of the Mekong River: a stationary and a non-stationary covariance model. The proposed approach predicts water-level time series at different locations along the Mekong River with a temporal resolution of five days. Validation is performed against in situ data from four gauging stations, yielding RMS differences between 0.82 and 1.29 m and squared correlation coefficients between 0.89 and 0.94. Both models produce comparable results when used for combining data from Envisat, Jason-1, and SARAL for the time period between 2002 and 2015. The quality of the predicted time series turns out to be robust against a possibly decreasing availability of altimetry mission data. This demonstrates that our method is able to close the data gap between the end of the Envisat and the launch of the SARAL mission with interpolated time series.

Keywords: Spatio-temporal ordinary kriging; Inland altimetry; Mekong River; Stochastic space-time processes; Covariance models along river; Non-stationary covariance models

*Deutsches Geodätisches Forschungsinstitut, Technische Universität München, Arcisstr. 21, D-80333 Munich, email: eva.boergens@tum.de

†Center for Mathematical Sciences, Technische Universität München, Boltzmannstr. 3, D-85748 Garching, email: sven.buhl@tum.de

1 Introduction

Describing and modelling the global water cycle relies on the knowledge of water levels of inland water bodies such as rivers, lakes and wetlands. To detect long-term changes in surface water storage, a dense (in space and time) network of monitoring stations is required. However, the number of global in situ measurements of gauges has been declining since the 1980s [Global Runoff Data Center (GRDC), 2013].

In the last two decades, many studies have shown the great potential of satellite altimetry for measuring water-level time series of larger inland water bodies worldwide to close the data gap of in situ observations [e.g. Birkett, 1995, 1998, Berry, 2006, De Oliveira Campos et al., 2001, Schwatke et al., 2015b]. Especially in the last few years, the accuracy of inland altimetry observations and their availability for even smaller targets have improved [e.g. Maillard et al., 2015, Boergens et al., 2016]. The accuracy has reached a few centimetres for lakes and a few dozens of centimetres for rivers [Schwatke et al., 2015b]. The onset of SAR altimetry on Cryosat-2 has proven valuable for the observation of small lakes and rivers with high accuracy [Nielsen et al., 2015, Villadsen et al., 2015].

Today, many inland altimetry time series are publicly available through five operational databases: the Database for Hydrological Time Series over Inland Water (DAHITI), developed by the Deutsches Geodaetisches Forschungsinstitut der Technischen Universitaet Muenchen (DGFI-TUM) [Schwatke et al., 2015b]; Hydroweb, developed by the Laboratoire d'Etudes en Geophysique et Oceanographie Spatiales (LEGOS) [Crétaux et al., 2011]; the River and Lake database, provided by the European Space Agency (ESA) [Berry et al., 1997]; the Global Reservoir and Lake Monitor (GRLM), maintained by the Foreign Agricultural Service of the United States Department of Agriculture (USDA) [Birkett et al., 2011] and the Altimetry for Inland Water (AltWater) by the Technical University of Denmark [Nielsen et al., 2015], which incorporates only Cryosat-2 data so far.

Over lakes and reservoirs, it is often possible to link different altimeter missions and passes into one multi-mission multi-pass time series [e.g. Schwatke et al., 2015b, Calmant et al., 2008]. This is possible based on the assumption that the water level of a lake above the geoid remains constant in space. This is true because the water surface forms an equipotential surface if no other effects such as winds or currents are present. By combining different altimetry passes and missions, the temporal resolution of the water-level time series is increased, depending on the number of passes over the water body. Without multi-mission and multi-pass altimetry, the temporal resolution is restricted to 35 days for ERS-1 and 2, Envisat and SARAL measurements, and 10 days for Topex/Poseidon and Jason

missions.

Combining altimetry measurements along rivers cannot follow the same approach as that over lakes, as the water surface is not equipotential along the course of the river. The slope of the river, as well as the water flow along the river, hinders the combination of different observations. Besides, slope and flow velocity might change rapidly over the course of the river. Along rivers, an additional problem is spatial resolution. The locations with available altimetry data do not cover the river consistently due to the meanders of the river; some reaches are measured densely, whereas others are not measured at all. Tourian et al. [2016] attempted to combine such measurements by estimating flow times between different time series along the Po River. For the estimation of flow velocity or flow time, additional information on the river dynamics is required, since flow velocity is not only changing over the course of the river but also over time, depending on the water level. A different approach for data combination is to assimilate altimetry-derived water levels into hydrological river models [Michailovsky et al., 2013].

In this study, we propose another way of linking altimetry measurements along rivers in order to estimate multi-mission time series for any location of the river with a temporal resolution of a few days. The river flow is described by spatial and temporal statistical dependencies using covariances in spatio-temporal ordinary kriging. The key element of the methodology is the modelling of the dependencies in space and time. We test two spatio-temporal covariance models on the Mekong River: a separable stationary spatio-temporal product covariance model and a non-stationary spatio-temporal covariance model. The estimated covariances are then input in an ordinary spatio-temporal kriging algorithm, which combines the different altimetric measurements along the Mekong River. The main advantage of spatio-temporal kriging for combining multi-mission altimetry data is that it can be used to interpolate water-level time series at any given point in space and time. Another advantage is its stability against erroneous measurements.

Ordinary kriging was originally developed in a spatial context by Krige [1952]. It is an interpolation method based on the covariances between the measurement locations and the location to predict. Spatial and spatio-temporal ordinary kriging have been widely applied in recent years [e.g. Cressie and Wikle, 2011, Gräler et al., 2011]. Various authors use spatio-temporal kriging for different sets of observations, for example, to predict soil moisture in Heuvelink et al. [1997] or Snepvangers et al. [2003], as well as top-kriging for river flow [Skøien and Blöschl, 2007]. Yoon et al. [2013] used spatio-temporal kriging to interpolate synthetic SWOT data along the Tennessee River.

In order to apply the kriging method, it is necessary to know the statistical dependencies between

the measurements; the dependencies are modelled with covariances. There are various models for describing spatio-temporal covariances, which have been published with different objectives and applications suitable for different data [e.g. De Iaco et al., 2001, De Cesare et al., 2001, Heuvelink et al., 1997, Snepvangers et al., 2003, Gräler et al., 2011]. Rouhani and Myers [1990] has already investigated the problems of modelling spatio-temporal covariances of geohydrological data. Moreover, Cressie and Wikle [2011] summarised a wide variety of spatio-temporal covariance models, including the product model, which separates space and time.

In all the aforementioned covariance models, the spatial locations are points on a 2D plane and the covariance function depends on the Euclidean distance between two points. However, in order to model covariances along river networks appropriately, it is important to incorporate both the river distance and the Euclidean distance. Due to river meandering, two river locations can be close in terms of their Euclidean distance, but be far apart along the river; thus, the hydrological conditions would differ. Ver Hoef et al. [2006] and Ver Hoef and Peterson [2010] introduced non-stationary covariance models for rivers based on the river distance. However, completely disposing of the Euclidean distance can lead to a loss of information. Asadi et al. [2015] tackled this problem with a mixed covariance model of river and Euclidean distances. Nevertheless, the temporal component was not taken into account in all the above-mentioned studies.

In this study, we demonstrate, through a case study along the Mekong River, the great potential of spatio-temporal kriging for combining multi-mission altimetry data along a river. By linking different altimetry water-level time series, both the spatial and temporal resolutions are enhanced. Each time series itself has a temporal resolution of 10 to 35 days, but all time series together can have a temporal resolution of up to a few days. The most important part of the methodology is modelling the spatio-temporal dependencies of the river flow using covariances. Appropriate covariance models are the essential part of linking multi-mission altimetry data along a river. The method described in this study allows for both high temporal and spatial resolutions of the combined time series, which can be interpolated at any given location along the river at any given time point.

The rest of the paper is structured as follows. In Section 2, the ordinary spatio-temporal kriging method used in this study is presented. Section 3 introduces the study area and Section 4 the altimetry and in situ data applied in this study. Our spatio-temporal covariance models for altimetry measurements along the Mekong River are described in detail in Section 5. Section 6 presents the results of kriging based on the different models and validates them against in situ gauge measurements. The

influence of the different missions on the results is investigated in the same section, and a comparison with previous studies is presented.

2 Spatio-temporal ordinary kriging with uncertain data

Ordinary kriging (hereafter referred to as *kriging*) is a geostatistical interpolation method, which was originally developed for spatial data by Krige [1952]. In this study, altimeter measurements along the Mekong River, scattered in space and time (see more on the data distribution in Subsection 4.1), are used to predict the water level at any given location along the river and at any given point in time.

Kriging produces a statistically unbiased estimator that is optimal with respect to the mean-squared prediction error. The predictor is a weighted average of the observed values, where the weights depend on the spatial or spatio-temporal dependence between the observed locations and the predicted location. The dependency is expressed by the covariance $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$ between the water levels at two space-time points (\mathbf{s}_1, t_1) and (\mathbf{s}_2, t_2) . Spatio-temporal kriging includes one dimension more than spatial kriging. In addition, the spatial and temporal domains have to be treated differently. Time has clear and irreversible ordering, which does not hold for space directly. The difference between space and time is treated in more detail in Section 5. Detailed introductions to spatial and spatio-temporal kriging can be found in Cressie [1993] and Cressie and Wikle [2011], respectively. In the following, we only present the basic formulas.

The kriging predictor $p(\mathbf{s}_0, t_0)$ (i. e. the prediction of the water-level) at some unobserved location (\mathbf{s}_0, t_0) along the river is the weighted average of all available altimetry measurements $\mathcal{Z} := \{Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n)\}$:

$$p(\mathbf{s}_0, t_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i, t_i), \quad (1)$$

where the weights λ_i should add up to 1 to guarantee an unbiased estimation. Note that there might be several altimetry measurements for the same location or the same time point; i.e. $\mathbf{s}_i = \mathbf{s}_j$ or $t_k = t_\ell$ for $i \neq j$ or $k \neq \ell$ is possible. Unlike the classical computation of kriging weights [e.g. Cressie, 1993, Chap. 3.2], where all input data are assigned the same accuracy, we assign different accuracy values for the data according to De Marsily [1986, Chap. 11.4.8]. The weights are then given by

$$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) = (\mathbf{c} + \mathbf{1} \frac{(\mathbf{1} - \mathbf{1}^\top \Sigma^{-1} \mathbf{c})}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}})^\top (\Sigma + \Sigma_{\text{alti}})^{-1}, \quad (2)$$

where $\mathbf{1} = (1, \dots, 1)^\top$, $\mathbf{c} = (C((\mathbf{s}_0, t_0), (\mathbf{s}_1, t_1)), \dots, C((\mathbf{s}_0, t_0), (\mathbf{s}_n, t_n)))^\top$ and $\Sigma = (C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)))_{i,j=1\dots n}$. Σ_{alti} denotes the diagonal matrix of the accuracies σ_i^2 , which are assumed constant for each altimetry mission and are estimated through variance component analysis [Koch, 1999, chap. 3.6]. These variances are not absolute accuracies, but relative accuracies between the observations of the different missions. In particular, the j th diagonal entry of Σ_{alti} is σ_i^2 if the observation (\mathbf{s}_j, t_j) recorded by the i th altimetry mission.

It is obvious from Equation (2) that the setup of the covariance function $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$ is of particular importance. The estimation of an appropriate covariance function for altimetry data along the Mekong River is discussed in Section 5.

3 Study area

This study focuses on the Lower Mekong River in South East Asia, south of the Chinese border, up to the confluence with the Tonle Sap River. North of the Chinese border, the river is not measurable by satellite altimetry due to the small and steep river channel; south of the confluence with the Tonle Sap River, in Phnom Penh, the delta is subject to tidal influence, which needs special handling of the altimetry data. Tributaries are not directly considered in this study, but have an indirect influence on the dependence modelling; the Subsection 5.2. Figure 1 shows the layout of the study area.

The Mekong River is dominated by two hydrological regimes: the precipitation and snow melt on the Tibetan Plateau, the so-called Yunnan Component, and the yearly south-eastern monsoon [Mekong River Commission (MRC), 2005]. The upper part of the Lower Mekong basin is mostly governed by the Yunnan Component, whereas the lower part is under the influence of the monsoon. The change in the water level between the seasons can be as high as 10 m [e.g. Boergens et al., 2016].

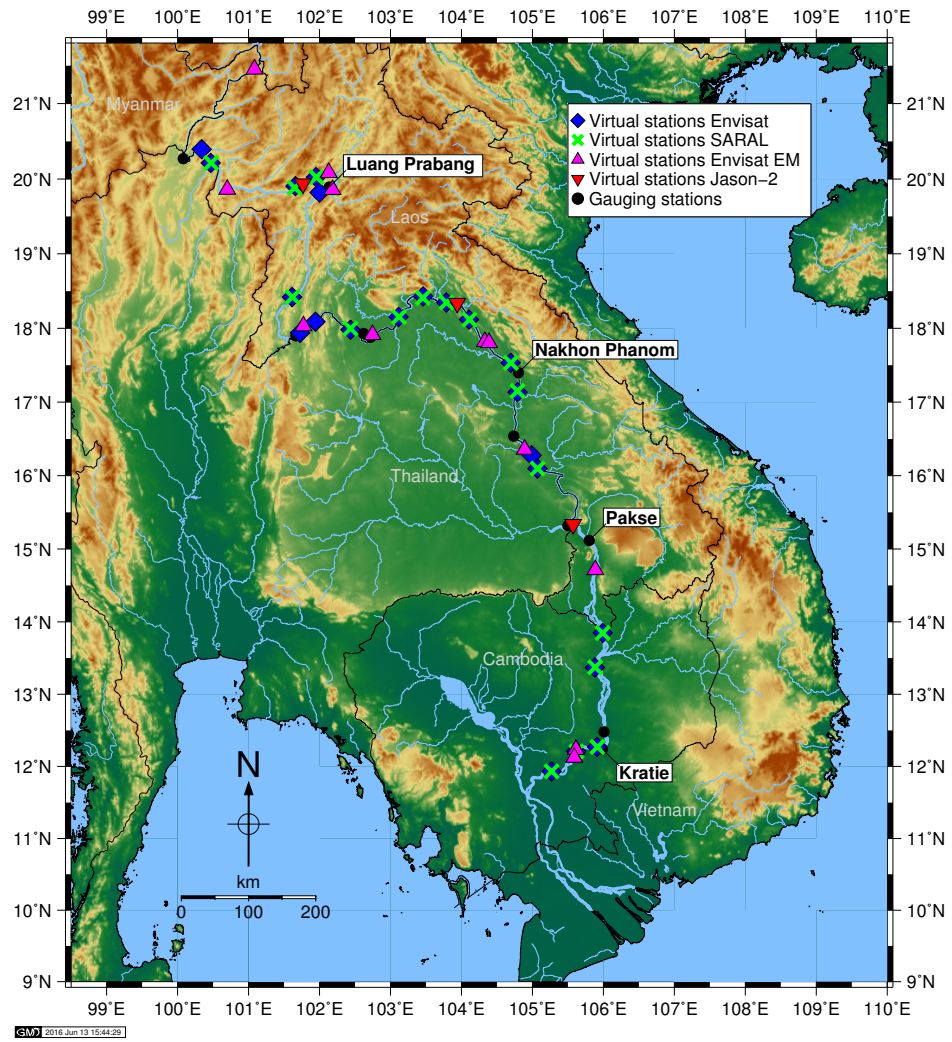


Figure 1: Map of the study area

4 Data

4.1 Water levels from altimetry

In this study, we employ altimetry data recorded by three different satellite missions with repeat orbits: Envisat, SARAL and Jason-2. For Envisat, we use both the repeat orbit phase between 2002 and 2010 and the extended mission from 2010 until 2011 (hereafter Envisat EM). The SARAL mission succeeded the Envisat mission on the same orbit, but commenced in 2013. Jason-2 data are available from 2008 until the present. The repeat time is 35 days for Envisat and SARAL, 30 days for Envisat EM and 10 days for Jason-2. These missions yield data for virtual stations (VSs) at the crossing points with the river, which are processed with the DAHITI database methodology [Schwatke et al., 2015b]. The first step of this method is the pre-processing, i.e. retracking the altimeter measurements and outlier detection. In the second step, a Kalman filter is applied to these heights in order to derive time series of the water-level changes. A correction for the Hooking Effect is included in the pre-processing for the upstream regions [Boergens et al., 2016].

We use 22 VSs of Envisat, 17 of SARAL, 12 of Envisat EM and 3 of Jason-2. We have fewer VSs from SARAL than Envisat because we decide to exclude five VSs that give too few data points in the time series. Missing data points appear mostly in the rainy season and are probably caused by the higher sensitivity of SARAL towards atmospheric water [Schwatke et al., 2015a]. Jason-2 has overall five crossings with the Mekong River in our study area, but in two out of those, it is impossible to derive reliable time series. The locations of all VSs are displayed in Figure 1. The temporal and spatial distributions of the data along the river are shown in Figure 2. The longest time series with the densest spatial pattern is provided by Envisat. The Envisat and SARAL missions share the same orbit; i.e. they have VSs at the same points. However, two years of data are missing in-between the missions. This gap is only sparsely covered by Envisat EM and Jason-2 data. Figures of the time series used in this study can be found in the supplementary data.

4.2 Preprocessing of water level time series

In order to apply kriging to the altimetry data described in Subsection 4.1, a pre-processing step is necessary to make the data roughly satisfy the assumption required for the ordinary kriging method, i.e. a constant mean [Cressie, 1993, p. 120]. In addition, we remove all trends and seasonality in each of the time series according to Brockwell and Davis [2002, Section 1.5.2]. The seasonality is not modelled

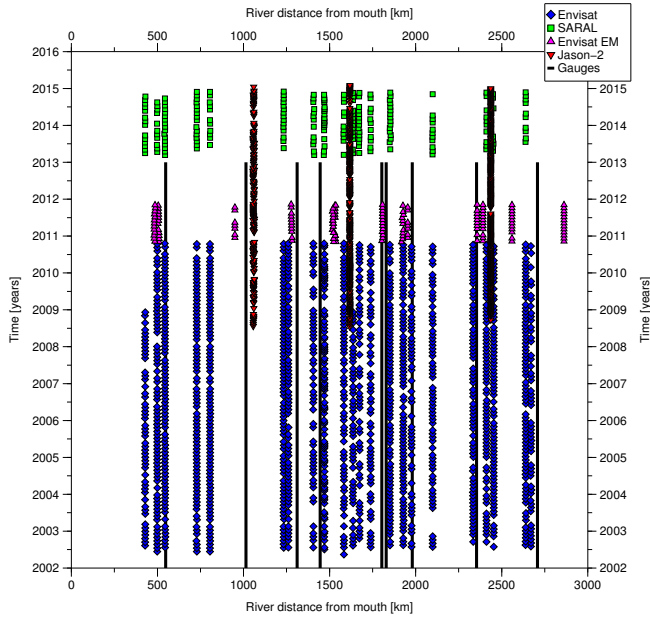


Figure 2: Spatio-temporal data distribution

with a trigonometric function; instead, we compute and remove the monthly seasonal coefficients for each time series. For the Envisat EM data, with only one year of data, we interpolated the seasonal coefficients from the neighbouring time series. The seasonal coefficients method allows reducing the seasonal signals that do not follow a sinusoidal function Figure 3.

We use the transformed observations as residuals and apply kriging to them. The predictions are then back-transformed, so that they yield interpretable information.

4.3 In-situ data

For validation purposes, we use in situ data from gauging stations. The gauging stations are maintained along the main river and tributaries by the Mekong River Commission (MRC, <http://ffw.mrcmekong.org/>). Nine stations are available for this study, but the records reach only until the end of 2012. The gauging stations appear alongside the virtual stations in Figure 1.

4.4 Auxiliary data

During the study, some auxiliary data are required. In Subsection 5.1, we use the width and slope of the Mekong River provided by Gupta and Liew [2007] to estimate the flow velocity of the river. In Subsection 5.2, we use the Global Precipitation Climatology Centre (GPCC) data of monthly mean precipitation placed on a 0.5° grid [Schneider et al., 2011]. A flow direction map from the HydroSHEDS data is used to determine the sub-catchment of each VS in the same section [Lehner et al., 2006]. The

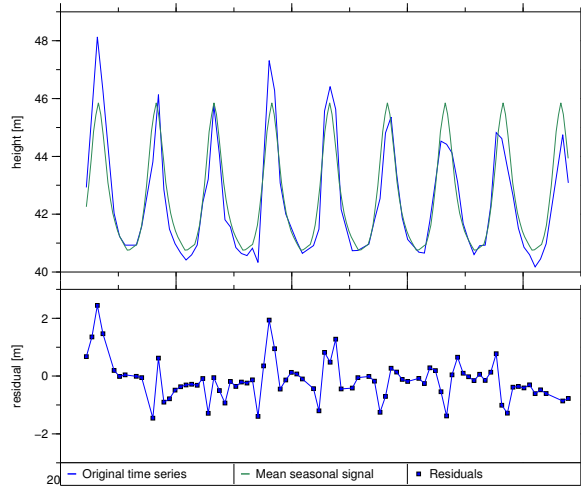


Figure 3: Example of a time series with reduced seasonal signal. The lower panel shows the residuals used for all further computations.

sub-catchments are in turn used to calculate weights and hydrological locations for the non-stationary covariance model in Subsection 5.2.

5 Spatio-temporal covariance models

A prerequisite for the kriging method (see Section 2), which is used in this study to combine the multi-mission altimetry data, is an appropriate covariance modelling. The covariances have to reflect the changing flow of the river along its course, and can be modelled based on knowledge about the river system or by using empirical covariances obtained from the data. This section introduces the covariance models used for the Mekong River and how they are estimated based on empirical covariances.

We assume that the residuals of the altimetry data (see Subsection 4.2) are a realization of the mean-zero random process $\{Z(\mathbf{s}, t) : \mathbf{s} \in \mathcal{S}, t \in \mathcal{T}\}$ for $\mathcal{S} \times \mathcal{T} \subset \mathbb{R}^2 \times [0, \infty)$. The value $Z(\mathbf{s}, t)$ can be interpreted as the deviation from the mean seasonal river level at a space-time location (\mathbf{s}, t) , where \mathbf{s} is a location along the Mekong River and t is some time point (a day between 2002 and 2015).

We follow the ideas presented by De Cesare et al. [2001, Section 2] and define a valid spatio-temporal covariance model C as the mixture of sums and products of valid spatial and temporal models. For the space-time locations (\mathbf{s}_1, t_1) and (\mathbf{s}_2, t_2) , we define a spatio-temporal covariance model $C_{\mathcal{ST}}$ as

$$C_{\mathcal{ST}}((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = k_1 C_S(\mathbf{s}_1, \mathbf{s}_2) C_T(t_1, t_2) + k_2 C_S(\mathbf{s}_1, \mathbf{s}_2) + k_3 C_T(t_1, t_2), \quad (3)$$

where $k_1 > 0$, $k_2, k_3 \geq 0$ and C_S and C_T are valid spatial and temporal covariance models, respectively. We require $Z(\mathbf{s}, t)$ to be (second-order) stationary in time; that is, $C_T(t_1, t_2) = C_T(h_T)$ can be written only in terms of the temporal lag $h_T = |t_1 - t_2|$. Computing empirical variances of the altimetry data and the correspondent confidence intervals of the estimators, we find that the variance does not vary significantly across the different locations, with most confidence intervals overlapping. Thus, it is reasonable to assume that the spatial variance along the Mekong River is constant; that is, $C_S(\mathbf{s}, \mathbf{s}) = C_S(\mathbf{s}', \mathbf{s}')$ for all $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$. Hence, the spatio-temporal variance $C_{ST}((\mathbf{s}, t), (\mathbf{s}, t))$ is also constant over $\mathcal{S} \times \mathcal{T}$. Note that this assumption does not necessarily hold for other river systems.

This assumption is weaker than the one used by De Cesare et al. [2001], which requires stationarity in space as well. Validity, i. e. positive definiteness, of the covariance model (3) follows by validity of models C_S and C_T and the requirement of $k_1 > 0$, $k_2, k_3 \geq 0$, see De Iaco et al. [2011, Section 4].

There are various valid models for spatial or temporal covariance functions (see for instance Cressie [1993, Section 2.3.1]). As outlined by Ver Hoef, Peterson, and Theobald [2006, Section 1], covariance models that are valid when relying on the Euclidean distance need not yield a valid model of a river network if the Euclidean distance is simply replaced by the river distance.

In Subsection 5.1 and Subsection 5.2, we present two models for the spatial covariance function C_S , which are valid for river networks. In Subsection 5.1, we focus on a spatially stationary model applied to observations transformed with respect to flow velocity, whereas Subsection 5.2 deals with a non-stationary approach applied to the original observations.

5.1 Stationary covariance model

In this section, we assume the process $Z(\mathbf{s}, t)$ to be second-order stationary in not only time but also space; i. e. we slightly change the notation and write $C_S(\mathbf{s}_1, \mathbf{s}_2) = C_S(h_S)$, where $h_S = |\mathbf{s}_1 - \mathbf{s}_2|_{\text{riv}}$ is the spatial lag, which is the distance along the river between the two points, expressed as the norm $|\cdot|_{\text{riv}}$. Under this condition, the spatio-temporal covariance model C_{ST} in (3) can be written only in terms of h_S and h_T .

The original altimetry measurements do not fulfil the spatial stationarity assumption in the first place. To make this assumption more feasible, we perform an additional data pre-processing step. We aim for the flow volume between two observed locations on the river within a particular time to be approximately equal whenever the locations are separated by the same river distance. To this end, we transform the observed spatial domain with respect to the estimated flow velocity of the river.

The flow velocity v can be estimated from the slope of the river, its width and the water-level with the Gaukler-Manning-Strickler equation [Jirka, 2007, Chap. 10.3.2]; i. e.

$$v = k_{st} R_h^{2/3} S^{1/2}. \quad (4)$$

In this equation, S is the slope of the river, which is approximately known for the reaches of the river from Gupta and Liew [2007]. By R_h , we denote the so-called hydraulic radius, which is the proportion between the cross-sectional area and the wetted perimeter. For this hydraulic radius, we assume a constant depth of a river with a rectangular cross-section. The width varies in river reaches, as defined in Gupta and Liew [2007], between 500 and 1500 m. The Strickler coefficient k_{st} measures the roughness of the river channel. Due to missing information on this coefficient for the Mekong River, a constant literature value for large rivers is applied ($k_{st} = 35 \frac{m^{1/3}}{s}$) [Jirka, 2007, Chap. 10.3.2, Tab. 10.2]. The transformed distance for each point is then

$$|\mathbf{s} - \mathbf{0}|_{\text{riv}}^{\text{new}} = \int_0^{|\mathbf{s} - \mathbf{0}|_{\text{riv}}} \frac{dx}{v(x)}, \quad (5)$$

where $|\mathbf{s} - \mathbf{0}|_{\text{riv}}$ is the original river distance for the river mouth to the considered point, $|\mathbf{s} - \mathbf{0}|_{\text{riv}}^{\text{new}}$ is the new river distance and $v(x)$ is the flow velocity at the given point s along the river. In our study, the flow velocity is assumed constant over river reaches, and therefore, the above integral can be transformed to a sum of all reaches downstream of the point. Of course, this representation of the flow velocity is only an approximation. For an exact transformation of water levels along the river, a more sophisticated approach would be necessary, taking for instance time variable slopes into account. However, at this point, the transformation is carried out to make the assumption of stationarity in space and the use of a stationary spatial covariance model more plausible.

The spatio-temporal covariance is modelled as the product of a valid spatial model C_S and a valid temporal model C_T ; i. e. we consider the model

$$C_{ST}(h_S, h_T) = C_S(h_S)C_T(h_T) \quad (6)$$

for spatial and temporal lags h_S and h_T . This corresponds to the special case of (3) obtained for $k_1 = 1$ and $k_2 = k_3 = 0$. This model is a separable model. In particular, it implicitly assumes that there is no interaction between the spatial and temporal parts. The temporal dependence structure

is supposed to be location-invariant, and the spatial dependence structure is in turn assumed not to change with time. Spatial invariance of the temporal dependence structure holds approximately for the Mekong River Adamson et al. [2009]. Temporal invariance of the spatial dependence structure would imply that the course of the river does not change with time. This does not hold for the entire river, especially for the delta, but the part of the river investigated in this study is not exposed to major shifting banks or similar effects.

A first step in fitting the spatio-temporal covariance model $C_{\mathcal{S}\mathcal{T}}$ is to obtain an empirical estimate. To this end, we define, for fixed $(h_{\mathcal{S}}, h_{\mathcal{T}})$,

$$N(h_{\mathcal{S}}, h_{\mathcal{T}}) := \{(\mathbf{s}_1, t_1, \mathbf{s}_2, t_2) \in (\mathcal{S} \times \mathcal{T})^2 : Z(\mathbf{s}_1, t_1) \in \mathcal{Z}, Z(\mathbf{s}_2, t_2) \in \mathcal{Z}, |\mathbf{s}_1 - \mathbf{s}_2|_{\text{riv}} = h_{\mathcal{S}}, |t_1 - t_2| = h_{\mathcal{T}}\},$$

where \mathcal{Z} denotes again the set of all observations, which in this study are the residuals of the observed altimetry time series without mean water level and seasonal signal.

Under the stationarity assumption, the covariance function C can be estimated by the empirical covariance, which is given by

$$\widehat{C}(h_{\mathcal{S}}, h_{\mathcal{T}}) = \frac{1}{|N(h_{\mathcal{S}}, h_{\mathcal{T}})|} \sum_{\substack{(\mathbf{s}_1, t_1, \mathbf{s}_2, t_2) \\ \in N(h_{\mathcal{S}}, h_{\mathcal{T}})}} Z(\mathbf{s}_1, t_1)Z(\mathbf{s}_2, t_2), \quad (7)$$

where $|N(h_{\mathcal{S}}, h_{\mathcal{T}})|$ is the cardinality of $N(h_{\mathcal{S}}, h_{\mathcal{T}})$. Since the data are unevenly distributed in space and time, we cluster lags within different tolerance intervals to achieve a reliable estimation of \widehat{C} . The cluster sizes are one month for the temporal and 50 km for the spatial lags. The size of the clusters is a trade-off between the temporal or spatial resolution of the empirical covariance and the reliability of the estimate; the smaller the clusters are, the more variable the estimate becomes. We modify the cluster size only for the estimation of the cluster with lag 0. For the estimation of temporal covariance with lag 0, we only take measurements that are recorded the same day. We take a tolerance of 5 km for lag 0 of the spatial covariance estimation due to the slightly shifting orbit of the altimeter satellite. All covariance estimates are tested for a significant difference from zero with a t-test. Empirical covariances failing the test are set equal to zero.

The separability of model (6) allows us to fit a spatial covariance model $C_{\mathcal{S}}$ and a temporal model

Table 1: Parameter estimates for the stationary product covariance model described in Subsection 5.1 fitted to the altimetry data

a_{exp}	b_{exp}	a_{lin}	b_{lin}
0.4078	5.4942	0.4300	0.0008

$C_{\mathcal{T}}$ separately.

For the temporal covariance function, we choose an exponential model with *nugget effect* n_{exp} ; i. e.

$$C_{\mathcal{T}}(h_{\mathcal{T}}) = \begin{cases} a_{\text{exp}} + n_{\text{exp}} & \text{if } h_{\mathcal{T}} = 0 \\ a_{\text{exp}} \exp(-b_{\text{exp}} h_{\mathcal{T}}) & \text{if } h_{\mathcal{T}} \neq 0 \end{cases} \quad (8)$$

and the parameters that need to be estimated are $a_{\text{exp}} > 0$, $b_{\text{exp}} > 0$ and $n_{\text{exp}} \geq 0$. The nugget effect accounts for measurement errors and arises because the estimation is based on lags $h_{\mathcal{T}}$, which are not arbitrarily close to zero. The nugget occurs as a jump or discontinuity in the fitted covariance model at zero.

The spatial covariance is modelled using a linear tent model [c.f Cressie, 1993] with nugget n_{lin} ; that is,

$$C_{\mathcal{S}}(h_{\mathcal{S}}) = \begin{cases} a_{\text{lin}} + n_{\text{lin}} & \text{if } h_{\mathcal{S}} = 0 \\ \max\{a_{\text{lin}} - b_{\text{lin}} h_{\mathcal{S}}, 0\} & \text{if } h_{\mathcal{S}} \neq 0. \end{cases} \quad (9)$$

and the parameters to be estimated are $a_{\text{lin}} > 0$, $b_{\text{lin}} > 0$ and $n_{\text{lin}} \geq 0$.

For computational reasons, we fit the correlation function instead of the correspondent covariance model, so that $a_{\text{exp}} + n_{\text{exp}} = a_{\text{lin}} + n_{\text{lin}} = 1$. This reduces the number of parameters to be estimated in both models by one. Figure 4 compares the respective empirical and fitted correlation functions. The approximately exponential and linear decays of the temporal and spatial dependence are visible. Table 1 shows the correspondent model parameter estimates. The nugget effect already accounts for 60% of the covariance decay. After two months, the temporal covariance drops by another 50%. As to the spatial domain, the covariance reaches the 25% level at a lag of 250 km.

An extension of this model would be to divide the river into hydrological reaches and assume stationarity only within each reach, i. e. relax the spatial stationarity constrain. The application of

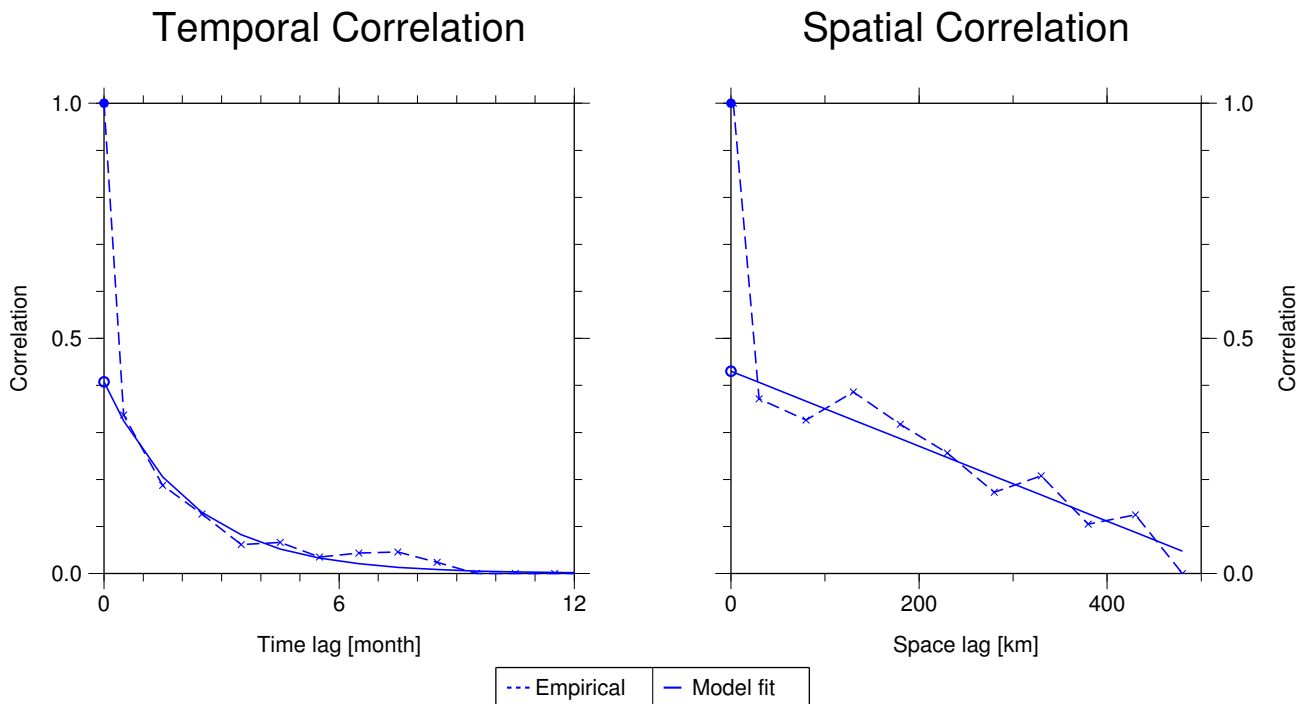


Figure 4: Empirical and fitted temporal and spatial covariance functions

this approach to the Mekong River does not improve the prediction results. When the river is divided into reaches, only data inside each reach can be used for fitting the respective covariance and for the kriging prediction. This reduction of data leads to inferior results, which is why we do not pursue the approach further.

5.2 Non-stationary covariance model

The main drawback of the stationary covariance model described above is the implicit assumption of spatial stationarity. Despite working with data transformed with respect to flow velocity, this might lead to predictions with reduced accuracy. In what follows, we propose a new non-stationary covariance model allowing for a changing spatial dependence structure.

We follow Ver Hoef et al. [2006] with an “upstream” construction on river networks. A covariance model that is valid for river networks can be defined for $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$, where \mathbf{s}_1 is located downstream of \mathbf{s}_2 , by

$$C_{\text{riv}}(\mathbf{s}_1, \mathbf{s}_2) = \begin{cases} 0, & \text{if } \mathbf{s}_1 \leftrightarrow \mathbf{s}_2, \\ \tilde{C}_1(0), & \text{if } \mathbf{s}_1 = \mathbf{s}_2, \\ \prod_{\mathbf{s} \in B_{\mathbf{s}_1, \mathbf{s}_2}} \sqrt{w_{\mathbf{s}}} \tilde{C}_1(|\mathbf{s}_1 - \mathbf{s}_2|_{\text{riv}}), & \text{if } \mathbf{s}_1 \leftrightarrow \mathbf{s}_2 \\ & \text{and } \mathbf{s}_1 \neq \mathbf{s}_2. \end{cases} \quad (10)$$

Here, the symbols \leftrightarrow and \nleftrightarrow stand for flow-connectedness and flow-un-connectedness, respectively. The set $B_{\mathbf{s}_1, \mathbf{s}_2}$ contains the observed measurement locations \mathbf{s} between \mathbf{s}_1 and \mathbf{s}_2 , but excluding \mathbf{s}_1 , and $\tilde{C}_1(\cdot)$ is a stationary covariance model valid in one dimension. We obtain convincing results choosing for $\tilde{C}_1(\cdot)$ again a linear tent model with variance and slope parameters $a_{\text{riv}} > 0$ and $b_{\text{riv}} > 0$, respectively, cf. (9). The factors $w_{\mathbf{s}}$ are weights based on the proportion of flow volume contributed by a measurement location \mathbf{s} to its next location downstream. This proportion is approximated by integrating the mean monthly precipitation over the sub-catchments of both \mathbf{s} and the subsequent location, and then, computing the quotient. The term sub-catchment denotes here the watershed of a location along the river. It consists of all points of the river that drain into this point. Figure 5 shows two examples of sub-catchments in shaded colours corresponding to their VSs (squares). The mean monthly precipitation is obtained using the GPCC data described in Subsection 4.4. Note that incorporating evapotranspiration in our computations is negligible, since this yields nearly identical weights for the Mekong River basin.

The flow volume and the weights themselves in Figure 6 reflect the geomorphology of the river with inflowing tributaries. The flow volume is normalized to the value corresponding to the location most downstream. If a major tributary joins the Mekong River between two stations, the flow volume shows a distinct step between the two stations. The same is even better visible in the weights. A smaller weight at some location indicates that, between this location and the next location downstream, a relatively high inflow occurs. In Figure 6, we add the names of the inflowing tributaries to the distinct data points. The most striking peak is caused by two rivers, Mun and Banghiang. One inflow, Nam Ou, is visible well in the weights chart but worse in the flow volume chart. From Equation (10), it is obvious that a higher inflow between two locations along the river naturally reduces the covariance between them, as they carry less information about each other. Figure 7 shows how the variability in the weights is reflected in the fitted covariance model C_{riv} in comparison with \tilde{C}_1 , in an example of two locations located downstream and upstream of the confluence of tributary Nam Ngum. Since

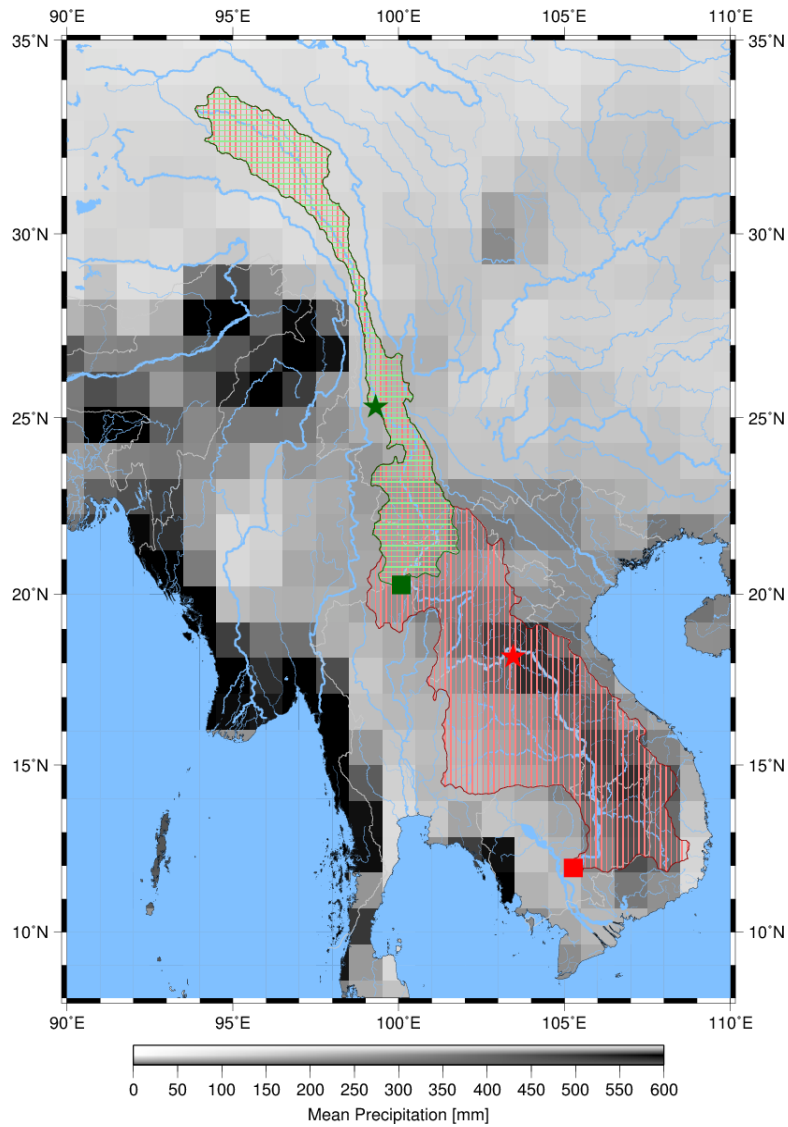


Figure 5: Hydrological locations for two VSs in their sub-catchment. The squares are the locations of the two VSs. The shaded areas indicate the corresponding sub-catchments, and the stars are the hydrological locations. The background is the GPCC precipitation map (c. f. Subsection 4.4) used for the hydrological locations.

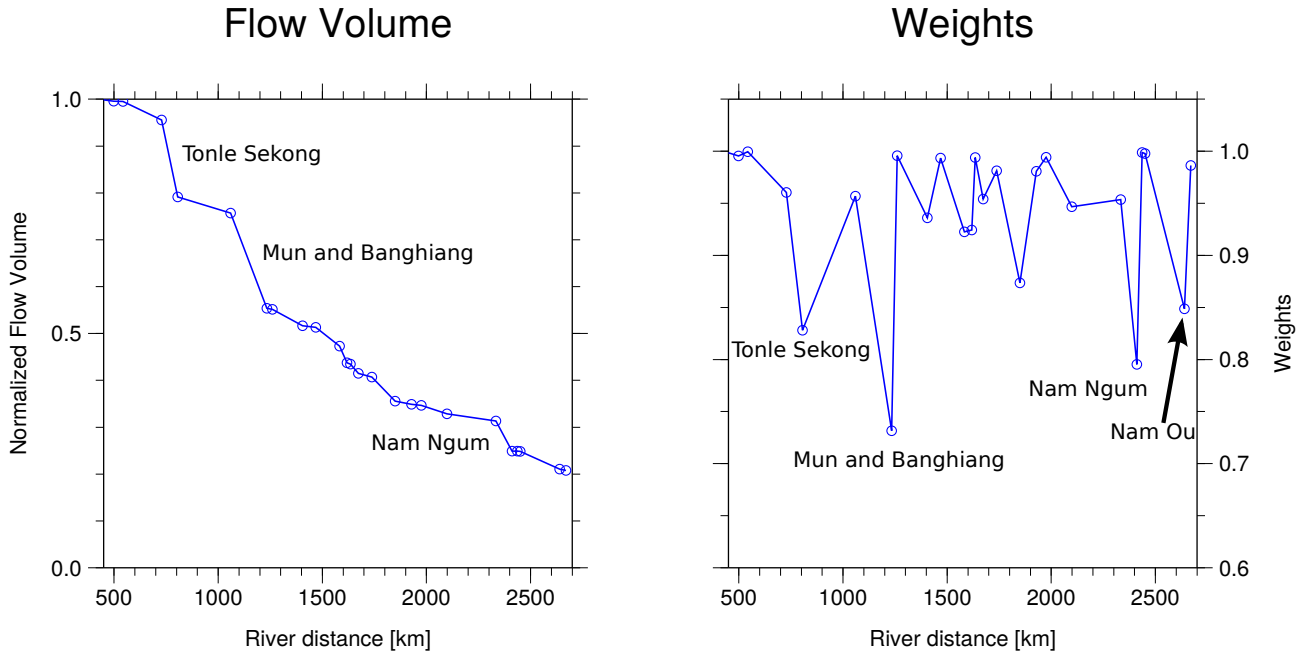


Figure 6: Flow volume and weights used in C_{riv} . The topography of inflowing tributaries is clearly visible. The names of prominent inflowing rivers are indicated.

\tilde{C}_1 is a stationary model, the graphs in the left panel overlap. Obvious in the fitted function C_{riv} is the drop in the dependence at the confluence, which shows non-stationarity of the spatial dependence structure.

The difference between model (10) and the “upstream” construction introduced by Ver Hoef et al. [2006] is that, in our case, the weights w_s are associated with the respective measurement locations s instead of river segments that lie between two locations. This is due to the lack of available observations on all segments of the Mekong River network.

One drawback of model (10) is that it only accounts for isolated river networks. In real data analysis, the Euclidean distance of two measurement locations often plays an important role, since environmental parameters such as precipitation tend to be similar at nearby locations. In the particular case of a river network, it is appropriate not to take into account the Euclidean distance of the measurement locations themselves, but rather the distance of their hydrological locations [Asadi et al., 2015, Section 3.3], which we define as the sub-catchments’ mass centres of the monthly precipitation. In Figure 5, we indicate the position of the hydrological location in the sub-catchment for two VSS with a star.

A covariance model based on Euclidean distances is given by

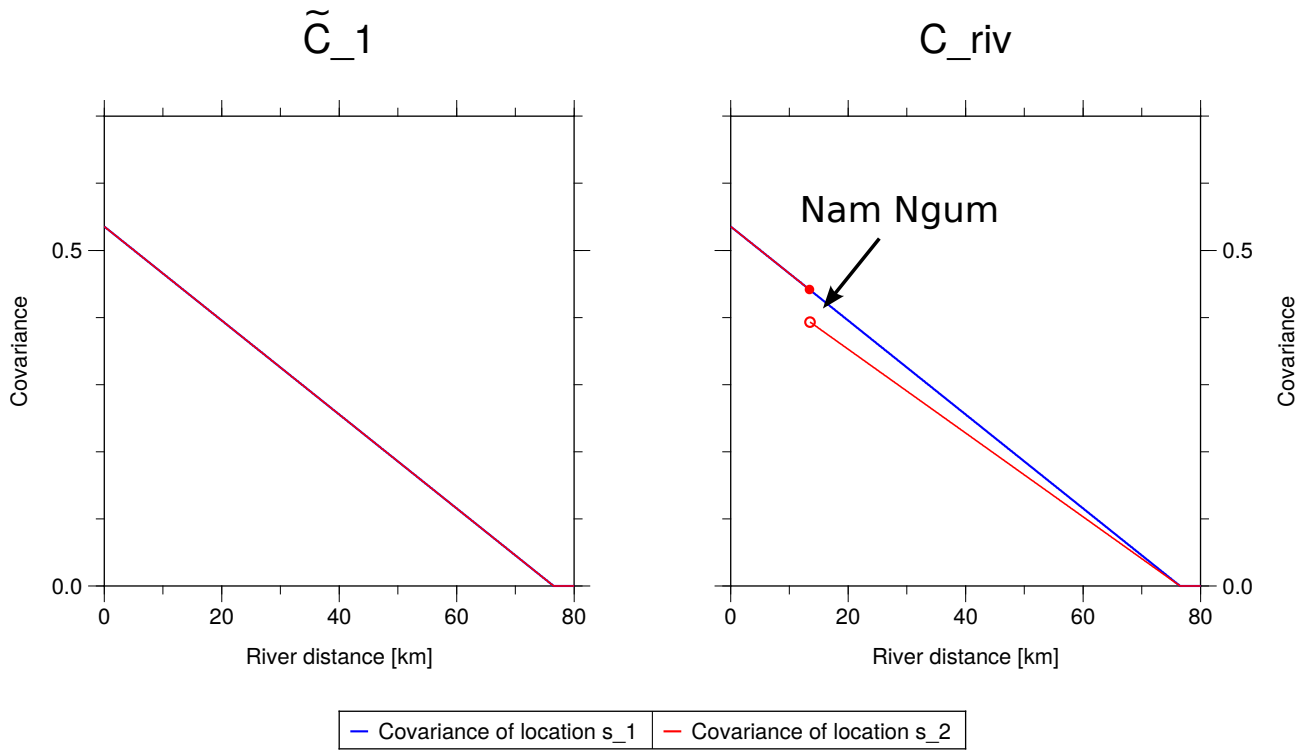


Figure 7: Fitted covariance functions $\tilde{C}_1(\cdot)$, $C_{\text{riv}}(\mathbf{s}_1, \cdot)$ (red) and $\tilde{C}_1(\cdot)$, $C_{\text{riv}}(\mathbf{s}_2, \cdot)$ (blue) for two reference locations \mathbf{s}_1 and \mathbf{s}_2 located downstream (at 2320 km from the river mouth) and upstream (at 2340 km) of the confluence of Nam Ngum. The plots show the respective covariance of the reference locations with locations of up to 80 km upstream.

$$C_{\text{euc}}(\mathbf{s}_1, \mathbf{s}_2) = \tilde{C}_2(\mathbf{A}(\mathbf{L}(\mathbf{s}_1) - \mathbf{L}(\mathbf{s}_2))), \quad (11)$$

where $\mathbf{L}(\mathbf{s})$ is the hydrological location of \mathbf{s} . The function \tilde{C}_2 is a covariance function valid in \mathbb{R}^2 and the matrix

$$\mathbf{A} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ a \sin \alpha & a \cos \alpha \end{pmatrix}, \quad a > 0, \quad \alpha \in \left[0, \frac{\pi}{2}\right],$$

allows for spatially anisotropic effects through rotation and dilation. As a model \tilde{C}_2 valid in two dimensions, we choose the spherical model [c.f. Cressie, 1993, Section 2.3.1], which is given by

$$\tilde{C}_2(\mathbf{h}) = \begin{cases} a_{\text{euc}} \left[1 - \frac{3}{2} b_{\text{euc}} \|\mathbf{h}\| + \frac{1}{2} (b_{\text{euc}} \|\mathbf{h}\|)^3 \right] & \text{if } 0 \leq \|\mathbf{h}\| \leq \frac{1}{b_{\text{euc}}} \\ 0 & \text{if } \frac{1}{b_{\text{euc}}} \leq \|\mathbf{h}\|, \end{cases} \quad (12)$$

where $a_{\text{euc}} > 0$ and $b_{\text{euc}} > 0$ denote the variance and range parameters, respectively. By $\|\cdot\|$, we denote the Euclidean norm.

In the following, we consider an overall spatial covariance model as a weighted sum of (10) and (11); i. e. for $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$,

$$C_{\mathcal{S}}(\mathbf{s}_1, \mathbf{s}_2) = \begin{cases} \pi_{\text{riv}} C_{\text{riv}}(\mathbf{s}_1, \mathbf{s}_1) + \pi_{\text{euc}} C_{\text{euc}}(\mathbf{s}_1, \mathbf{s}_1) + n_{\mathcal{S}}, & \text{if } \mathbf{s}_1 = \mathbf{s}_2, \\ \pi_{\text{riv}} C_{\text{riv}}(\mathbf{s}_1, \mathbf{s}_2) + \pi_{\text{euc}} C_{\text{euc}}(\mathbf{s}_1, \mathbf{s}_2), & \text{otherwise} \end{cases}, \quad (13)$$

with non-negative weights π_{riv} and π_{euc} . The parameter $n_{\mathcal{S}} \geq 0$ accounts for an additional nugget effect at the origin. Model (13) is non-stationary; indeed, even the notion of stationarity is not clear, since the model is not defined on a Euclidean space [see Asadi et al., 2015, Section 3.4]. Note, however, that only one of the parameters π_{riv} and a_{riv} (and likewise π_{euc} and a_{euc}) is statistically identified, since they both appear as a product in (13). Thus, we transform these parameter pairs to single parameters, which we denote again by π_{riv} and π_{euc} . Concerning the temporal covariance model $C_{\mathcal{T}}$, we use, as in Subsection 5.1, a stationary exponential model with variance, scale and nugget parameters $a_{\mathcal{T}} > 0$, $b_{\mathcal{T}} > 0$ and $n_{\mathcal{T}} \geq 0$, see (8).

As an overall spatio-temporal covariance model $C_{\mathcal{ST}}$, we use model (3), i. e.

$$C_{\mathcal{ST}}((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = k_1 C_{\mathcal{S}}(\mathbf{s}_1, \mathbf{s}_2) C_{\mathcal{T}}(|t_1 - t_2|) + k_2 C_{\mathcal{S}}(\mathbf{s}_1, \mathbf{s}_2) + k_3 C_{\mathcal{T}}(|t_1 - t_2|),$$

for $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$, $t_1, t_2 \in \mathcal{T}$. Recall that the coefficients k_1, k_2, k_3 need to satisfy $k_1 > 0$ and $k_2, k_3 \geq 0$ for model (3) to be valid.

Fitting the full model $C_{\mathcal{ST}}$ involves estimating the 13 parameters

$$\{\pi_{\text{riv}}, \pi_{\text{euc}}, b_{\text{riv}}, b_{\text{euc}}, \alpha, a, n_{\mathcal{S}}, a_{\mathcal{T}}, b_{\mathcal{T}}, n_{\mathcal{T}}, k_1, k_2, k_3\}.$$

As opposed to Subsection 5.1, we cannot compute the empirical covariance function as done in (7), since, due to non-stationarity, the true covariance function $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$ does not depend only on the spatio-temporal lag ($|\mathbf{s}_1 - \mathbf{s}_2|_{\text{riv}}, |t_1 - t_2|$) but on the space-time locations $(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)$ themselves. However, for each pair of space-time locations (\mathbf{s}_1, t_1) and (\mathbf{s}_2, t_2) , we only have at most one observation to base the estimation on. We, therefore, subdivide the observed temporal domain into K smaller groups, which can be considered approximately independent. Each time group comprises 35 subsequent days $0, \dots, 34$; the first time group contains the first 35 days, the second group the next 35 days, and so on. Choosing 35 days makes the temporal independence assumption more realistic, since 35 days is the return time of the Envisat satellite. Additionally, water needs up to four weeks to flow through our study area according to our flow-velocity estimation (cf. Subsection 5.1). In this way, we obtain for each observed pair of space-time locations $(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)$, where t_1 and t_2 each correspond to one (possibly the same) of the 35 days, several associated realizations of $Z(\mathbf{s}_1, t_1)$ and $Z(\mathbf{s}_2, t_2)$ (roughly one per time group, exceptions are due to missing values). We enumerate the pairs as

$$\{(\mathbf{s}_1, t_1)^{(1)}, (\mathbf{s}_2, t_2)^{(1)}, \dots, (\mathbf{s}_1, t_1)^{(K)}, (\mathbf{s}_2, t_2)^{(K)}\}$$

and compute the empirical covariance as

$$\widehat{C}((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \frac{1}{K} \sum_{j=1}^K Z((\mathbf{s}_1, t_1)^{(j)}) Z((\mathbf{s}_2, t_2)^{(j)}).$$

Model $C_{\mathcal{ST}}$ is then fitted using least squares estimation.

Since the number of parameters in the fully general model is large (13 parameters), we additionally

Table 2: Parameter estimates of the simplified non-stationary covariance model described in Subsection 5.2 fitted to the altimetry data

π_{riv}	π_{euc}	b_{riv}	b_{euc}	$a_{\mathcal{T}}$	$b_{\mathcal{T}}$	$n_{\mathcal{T}}$
0.536	0.471	0.007	1/2145.247	1.047	5.494	1.168

perform the procedure with $k_1 = 1$ and $k_2 = k_3 = 0$, which again yields the simpler separable space-time product model

$$C_{S\mathcal{T}}(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = C_S(\mathbf{s}_1, \mathbf{s}_2)C_{\mathcal{T}}(|t_1 - t_2|).$$

Additionally, we find that incorporating the spatial nugget effect n_S and the anisotropy parameters α and a downgrades the model fit. Therefore, we fix their values to $n_S = 0$, $a = 1$ and $\alpha = 0$. In this case, there are only seven parameters left for estimation, given by

$$\{\pi_{\text{riv}}, \pi_{\text{euc}}, b_{\text{riv}}, b_{\text{euc}}, a_{\mathcal{T}}, b_{\mathcal{T}}, n_{\mathcal{T}}\}.$$

Due the separability property, it is possible to compute the empirical spatial and temporal covariances separately, similarly as in Subsection 5.1. In our subsequent analyses, kriging with the full (13 parameters) and the simplified model (7 parameters) yield qualitatively comparable results. Hence, for computational reasons, we focus on the latter.

Table 2 shows the least squares parameter estimates for the simplified non-stationary spatio-temporal covariance model. Figure 8 visualizes the latter for two reference locations. Animated versions of the two figures can be found in the supplementary material. They provide an improved three dimensional impression of the content. Sometimes, jumps are visible along the y-axis, i.e. the spatial covariance function of the reference location with the other locations. The jumps appear if a location is further apart from the reference location than the previous one in terms of river length, but the hydrological location is closer to that of the reference location in terms of Euclidean distance. This shows the influence of the Euclidean covariance part. Recall that by construction the fit of the spatio-temporal covariance function is based on all empirical values, whereof the black dots only constitute a subset. Hence for some fixed reference locations the fitted function seems to give a better fit than for others.

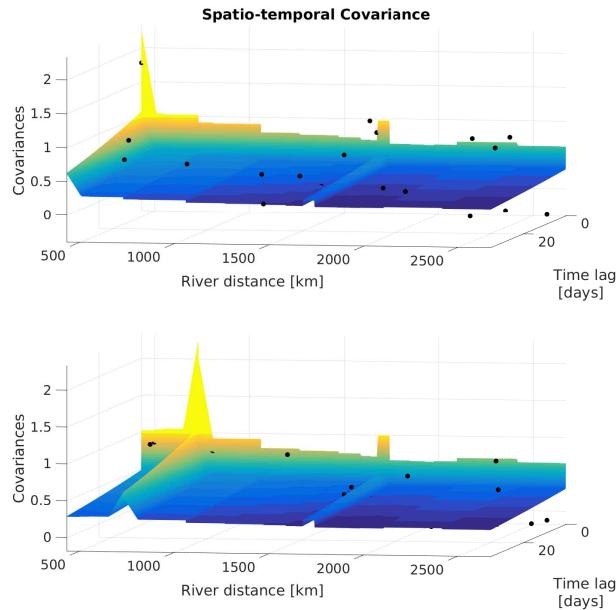


Figure 8: Fitted non-stationary spatio-temporal covariance functions $C_{ST}(\mathbf{s}, \cdot, 0, \cdot)$ (coloured planes) for two reference locations \mathbf{s} in comparison with the correspondent empirical counterparts (black points). The reference locations (visible as green peaks) are 429 km (top) and 729 km (bottom).

6 Results

In this section, we present the results of the spatio-temporal kriging prediction. We use the term prediction to denote interpolation, which is common with kriging. As the first step, the results based on both the stationary product covariance model described in Subsection 5.1 and the (simplified) non-stationary model explained in Subsection 5.2 are presented and compared to in situ gauging data and to each other. Afterwards, the influence of the available altimeter missions on the results is investigated. Finally, the results are compared with a previous study on the combination of altimetry data along river networks.

6.1 Validation of the different covariance models against in situ measurements

For validation purposes and in order to verify the goodness of the estimates obtained in Section 5, we use kriging with the two aforementioned covariance models to predict time series with a temporal resolution of five days between 2002 and 2015 at the locations of four gauging stations: Luang Prabang, Nakhon Phanom, Pakse and Kratie (see Figure 1 for the locations). We then compare the predictions with the available in situ measurements. We choose to predict the time series with a five-day resolution based on data availability. In the central part of our study area, between river distance 1000 and

2300 km (see Figure 2), every two or three days, an altimeter measures the water level at some location. Although, in principle, every temporal resolution of the predicted time series is possible with our approach, any higher resolution than these two or three days is useless. Moreover, we reduce the temporal resolution to five days as a compromise, because it is also appropriate for the more upstream or downstream reaches. To predict water levels at a particular location, for the stationary product covariance model, only observations that are recorded either upstream and earlier in time or downstream and later in time are used, in order to take into account the river flow. The non-stationary model described in Subsection 5.2 does not have this limitation, since it is partly based on the Euclidean distance.

The comparison is performed through the *root-mean-square error* (RMS), the *relative RMS* (rRMS), the *coefficient of determination* R^2 and the *Nash-Sutcliff efficiency coefficient* (NSE). The RMS is a measure for the point-wise absolute difference between the predicted time series and the observed gauge time series. The rRMS relates the RMS to the mean water level variation. If the water level variations are larger, then a higher RMS can be acceptable because the variation is still depicted well. The value of R^2 , which is the squared correlation coefficient between the two time series, is sensitive to the phase shift of the returning periodic signal between the two time series. The NSE measures the power or quality of the kriging predictions. It is given at a location $\mathbf{s} \in \mathcal{S}$ by [Nash and Sutcliffe, 1970]

$$NSE(\mathbf{s}) = 1 - \frac{\sum_{t=1}^T (Z_g(\mathbf{s}, t) - p(\mathbf{s}, t))^2}{\sum_{t=1}^T (Z_g(\mathbf{s}, t) - \overline{Z_g}(\mathbf{s}))^2}. \quad (14)$$

$Z_g(\mathbf{s}, t)$ is the height measured at time t at a gauging location \mathbf{s} with a correspondent mean value $\overline{Z_g}(\mathbf{s})$, and $p(\mathbf{s}, t)$ is the predicted height at the same time and location. The NSE ranges from $-\infty$ to 1, where 1 is the optimal value and allocated to perfect predictions. A value of 0 indicates a prediction that is as good as just taking the observed mean value of the time series; values less than 0 indicate even worse predictions.

The results are summarised in Table 3, and the in situ and predicted time series at Luang Prabang are shown in Figure Figure 9. Plots of the time series corresponding to the remaining three stations are provided together with those shown here in the supplementary material. The hydrological profile of the Mekong River exhibits an almost uniform annual signal. Therefore, the mean annual signal is already a good approximation of the true signal. The prediction quality of the mean annual signal is shown along with the results of the kriging predictions in Table 3. The comparison of each gauge with

Table 3: Comparison of the results using the different covariance models and the time series of gauging stations at four locations. The panel ‘annual signal’ presents the results if a mean annual signal is simply used as a predictor. The panel ‘Closest VS’ shows the results when the values of the closest VSs are used for the prediction; the respective missions are mentioned.

	RMS [m]	rRMS [%]	R ²	NSE
stationary product covariance Model				
Luang Prabang	1.02	11	0.91	0.90
Nakhon Phanom	0.82	10	0.94	0.94
Pakse	0.92	11	0.92	0.92
Kratié	1.29	10	0.92	0.92
Non-Stationary Covariance Model				
Luang Prabang	1.11	12	0.89	0.91
Nakhon Phanom	0.89	10	0.93	0.94
Pakse	0.88	10	0.93	0.93
Kratié	1.15	9	0.94	0.94
Annual Signal				
Luang Prabang	1.50	17	0.81	0.83
Nakhon Phanom	1.25	14	0.85	0.86
Pakse	1.20	14	0.86	0.87
Kratié	1.69	13	0.87	0.88
Closest VS				
Luang Prabang (<i>E+S</i>)	1.70	19	0.90	0.87
Nakhon Phanom (<i>E+S</i>)	0.39	4	0.96	0.98
Pakse (<i>J2</i>)	1.18	14	0.88	0.89
Kratié (<i>E+S</i>)	2.04	16	0.83	0.85

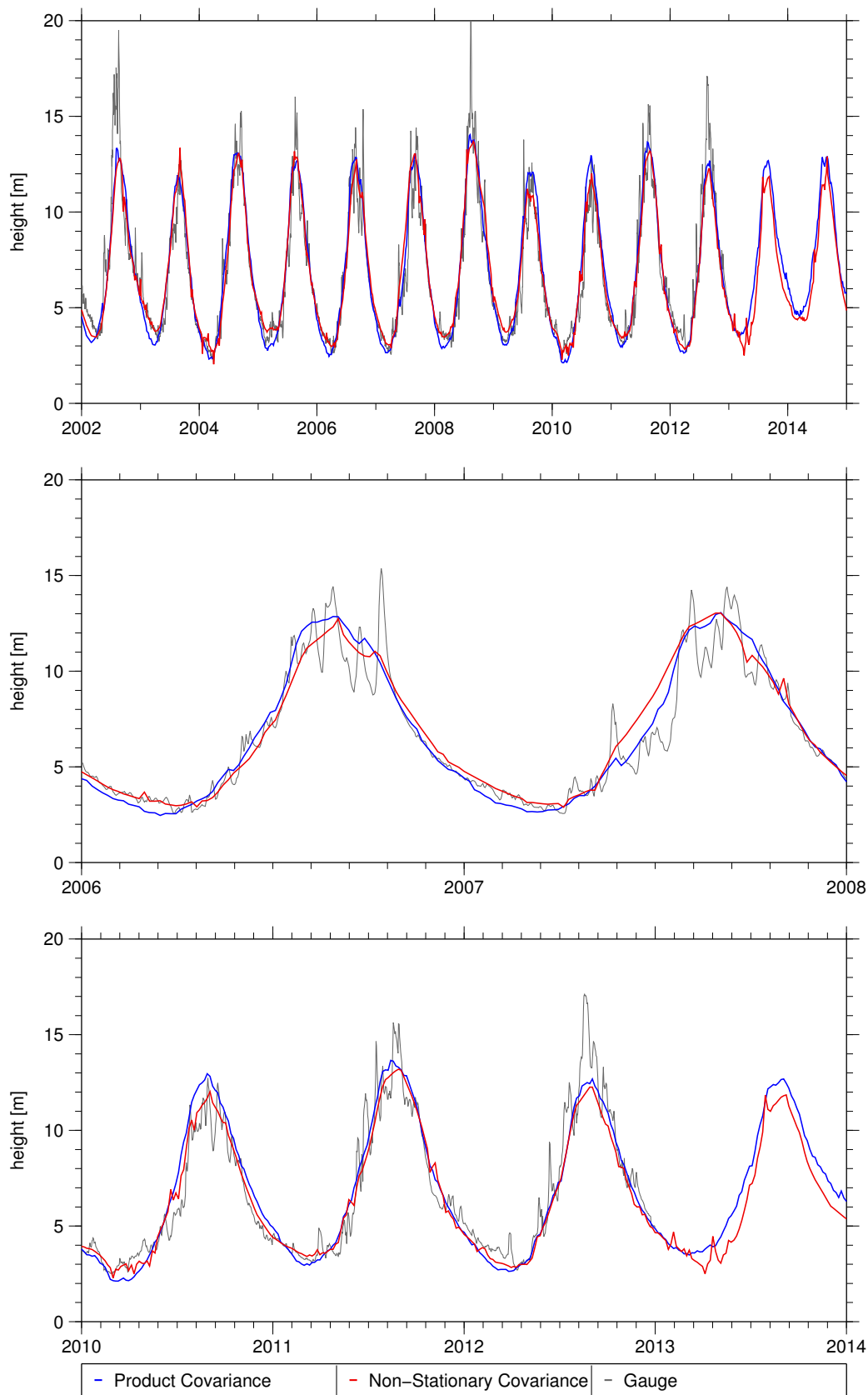


Figure 9: Comparison of the gauge measurements and the predicted five-day time series at the station of Luang Prabang based on the stationary product and the non-stationary covariance model. The two lower panels are enlargements of the upper panels for the years 2006 until 2008 and 2010 until 2013, respectively. The time series is height-referenced to the gauge data.

its nearest VS is shown in the last panel of Table 3.

The stationary product covariance model applied to observations transformed with respect to flow velocity yields, at all four stations, satisfying results with RMS values between 0.82 and 1.29 m. The coefficients of determination are similar for all stations and above 0.90; the same holds for the NSE.

The non-stationary covariance model and the stationary product covariance model provide similar predictions. The RMS lies between 0.88 and 1.15 m. Concerning the coefficient of determination R^2 (0.89–0.94) and NSE (0.91–0.94), the results are not better than those obtained with the stationary covariance model, but equally good. The two downstream stations, Pakse and Kratie, are predicted better with the non-stationary covariance model, whereas the product model performs better for the two upstream stations. In addition, the two stations near the border of our study area, Luang Prabang and Kratie, with less neighbouring data, show inferior results than the other two stations. Nevertheless, we do not have enough results to judge if these differences are random or reflect some difference in the covariance models.

When comparing the kriging results with the measurements at the closest VS, we observe that kriging yields clearly superior results, except for Nakhon Phanom. However, the VS near Nakhon Phanom is one of the best, if not the best, of the VSs used in this study, both in terms of the proximity to a gauge (ca. 15 km distance) and completeness of the time series (no missing data).

The time series of Luang Prabang, as seen in Figure Figure 9, reveals the differences in the two covariance models. For 2007, the rise of the water level is predicted too early with the non-stationary covariance model. On the other hand, the prediction with the non-stationary covariance models shows more small-term variations. It seems that kriging with the non-stationary covariance model results in less smoothing of the result than the stationary product model does. For late 2007, it seems that the correct short-term variation is predicted. However, for 2013, the variations in the low water seem rather unlikely, as in all other years, such variations are not present during low water. Unfortunately, we do not have gauge data for 2013 that could be used to validate this point.

Overall, many short-term variations cannot be detected with satellite altimetry at all. A small peak in water level can occur only locally for a short time frame. If no satellite measures this reach of the river at exactly that time, we cannot gain any information about the variation. Moreover, even if one altimeter measures such a small peak, the kriging tends to even it out, especially with the stationary product covariance model. On the other hand, the kriging method is also robust against outliers in the data and evens them out up to a certain degree.

6.2 Influence of the altimeter data availability

Between 2011 and 2013, there is a gap in the data due to the end of the Envisat mission. In this time span, only three Jason-2 and (at least partly) a few Envisat EM data points are available (see Figure 2). The Envisat observations are the backbone of our model. For this reason, we are particularly interested in how well the kriging method can handle the change in data availability between the end of 2010 and the beginning of 2013. In the time series in Figure Figure 9, no deterioration during these years is visible. For this reason, we examine the RMS values over the time.

Using kriging, we predict time series of water levels at a 5-day resolution at all available gauging stations in the observation area. To compare the predicted with the observed time series and detect possible changes in the quality of the predictions, we compute the RMS in yearly moving windows. The results are shown in Figure Figure 10 (colour coding is according to their location along the river). The RMS value is chosen for this comparison because, among the analysed coefficients, it shows the largest differences in Table 3. To obtain a clear overview, we present this investigation solely for the stationary product covariance model; the findings are similar for the non-stationary covariance model. The mission lifetimes are indicated at the top of Figure Figure 10. We cannot quantify the influence of the SARAL data, as the gauge time series are only available until the end of 2012.

It is not possible to find any correspondence between the data availability and the variation in the RMS over time. In particular, the RMS does not improve with the start of the Jason-2 mission in 2008, and does not deteriorate in 2012 with the end of Envisat EM. The only pattern visible is the overall smaller RMS in the years 2006 and 2007. We do not have any explanation for this pattern. Still gauging stations at closer distance tend to show more similarity in the RMS over time than those further apart.

6.3 Comparison to other published results

To our knowledge, the only other study combining altimetry data over rivers, so far, has been published by Tourian et al. [2016]. They conducted their study along the Po, Mississippi, Congo and Danube rivers with a focus on the Po River in Northern Italy. Data from the satellite altimetry missions of Topex/Poseidon, the Topex/Poseidon extended mission, Jason-2, Envisat, the Envisat extended mission, SARAL and Cryosat-2 were used. Their method was based on estimating the flow time between the different VSs with which the stations could be linked.

Despite the differences between the Mekong and the Po River systems, their results can be com-

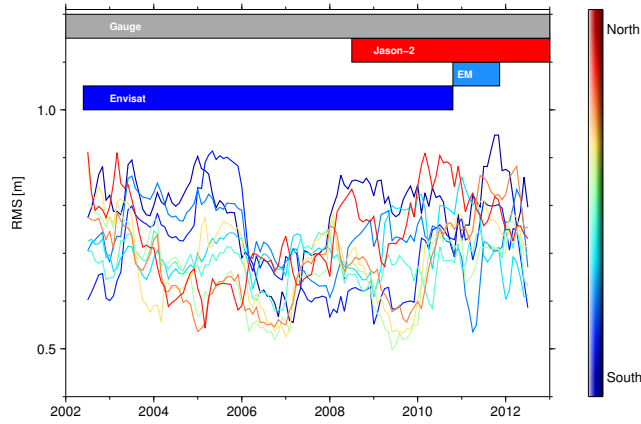


Figure 10: Moving yearly RMS values at all gauging stations along the Mekong River. The colour indicates the location along the river. Thus, VSs close in space have similar colours.

pared to ours under some limitations. The main limitation is that the annual signal of the Po River is not as strong and regular as that of the Mekong River, which suggests that our prediction is more accurate. Comparing the NSE, in our results, the value is never below 0.89 and ranges up to 0.94, whereas, in the Po River study, this value ranged between 0.2 and 0.6. In addition, there was a larger variance between the different gauging stations than that in our study.

In the Po River study, there were three possible merging processes for the altimetry data: all measurements were treated equally, disregarding the distance between the measurement locations and the predicted location; a three-point moving average with weighting according to their errors; and a three-point moving average with weighting according to the distance between measurement and prediction location. In our study, the observations are weighted according to their distance in space and time from the location to be predicted, which might explain the differences in the results.

7 Conclusion

In this study, we demonstrate the potential of spatio-temporal kriging, together with appropriate covariance modelling, to combine measurements of multi-mission altimetry along rivers over space and time. The essential part of the presented approach is the modelling of the spatial and temporal dependencies in the data with appropriate covariance models for predicting the river flow. Two covariance models are tested: a simple stationary spatio-temporal stationary product covariance model based on the river distance between locations, but applied to observations transformed with respect to flow velocity, and a non-stationary model, which is designed for river networks and additionally incorporates the Euclidean distance between the centres of the sub-catchments of the locations. These two

covariance models are used in the kriging approach to obtain multi-mission time series with a temporal resolution of 5 days. These time series are validated against in situ gauge data. Both models perform equally well, with RMS values between 0.82 and 1.29 m and coefficients of determination above 0.89.

A main advantage of the kriging approach for combining multi-mission altimetry over rivers is its flexibility in both temporal and spatial resolutions. The method can theoretically predict a time series of any given temporal resolution at any point along the river. However, a meaningful result cannot be expected if the temporal resolution of the prediction is higher than that of the altimetry data. Considering measurements and the possibly larger distance between two altimeter measurements closest in time, the temporal resolution should be even lower than the resolution of the altimetry data.

Although the non-stationary model does not yield better results, in this case study along the Mekong River, its advantage is its flexibility to approach more complex river systems and to include measurements along tributaries in the river system. However, this requires further investigation for the Mekong River system. The downside of the non-stationary model is its complexity and the need for additional data such as sub-catchment information. Therefore, the simple stationary product covariance model is sufficient for the rather non-complex main stream of the Mekong River if a spatial transformation according to the flow velocity is used. This transformation is not needed for the non-stationary model, which is again one of its advantages.

We show in this study that using the kriging method, it is not only possible to combine data from different altimeter measurements, but we are also able to bridge data gaps where only a few measurements are available. Particularly, the time between the end of the Envisat mission and the start of the SARAL mission with only three Jason-2 time series was bridged without any apparent decline in the quality of the predictions.

The next steps of this study involve incorporating the data of the long-repeat mission Cryosat-2 (launched 2010), which will densify the observations, especially in the gap between the Envisat and SARAL missions, as well as improve the temporal resolution of the available data. Additionally, the data of the newly launched missions Jason-3 and Sentinel-3 will be incorporated. Furthermore, data along the tributaries of the Mekong River will be included. We will use real-time processed altimetry data with the kriging method to do short-term predictions into the future. The transferability of the multi-mission approach introduced in this study to other river networks will be tested in a future work as well.

Acknowledgements

The altimeter observations and geophysical corrections are taken from OpenADB (<http://openadb.dgfi.tum.de>). The altimeter missions are operated and maintained by ESA (Envisat), ISRO/CNES (SARAL), and NASA, CNES, EUMETSAT and NOAA (Jason-2). The original data sets are disseminated by AVISO, ESA, and PODAAC. We thank Christian Schwatke for the help by the processing of the altimeter time series in DAHITI. SB thanks Anthony Davison and his group, especially Sebastian Engelke and Peiman Asadi, for interesting discussions and advice during his research stay at EPFL Lausanne. This work was supported by the German Research Foundation (DFG) through fund SE 1916/4-1 and by the TUM International Graduate School of Science and Engineering (IGSSE).

References

- P. T. Adamson, I. D. Rutherford, M. C. Peel, and I. A. Conlan. The Hydrology of the Mekong River. In *The Mekong*, pages 53–76. Mekong River Commission, 2009. ISBN 9780123740267. doi: 10.1016/B978-0-12-374026-7.00004-8.
- P. Asadi, A. C. Davison, and S. Engelke. Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023–2050, 2015.
- P. A. M. Berry. Two decades of inland water monitoring using satellite radar altimetry. *European Space Agency (Special Publication). 15 Years of Progress in Radar Altimetry. ESA SP 614, No 8*, 2006.
- P. A. M. Berry, H. Bracke, and A. Jasper. Retracking ERS-1 altimeter waveforms over land for topographic height determination: an expert systems approach. *European Space Agency (Special Publication). 3. ERS Symposium. ESA SP 414*, pages 403–408, 1997.
- C. M. Birkett. The contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes. *Journal of Geophysical Research: Oceans (1978–2012)*, 100(C12):25179–25204, 1995. doi: 10.1029/95JC02125.
- C. M. Birkett. Contribution of the TOPEX NASA radar altimeter to the global monitoring of large rivers and wetlands. *Water Resources Research*, 34(5):1223–1239, 1998. doi: 10.1029/98WR00124.

- C. M. Birkett, C. Reynolds, B. Beckley, and B. Doorn. From research to operations: the USDA global reservoir and lake monitor. In *Coastal Altimetry*, pages 19–50. Springer, 2011. doi: 10.1007/978-3-642-12796-0.
- E. Boergens, D. Dettmering, C. Schwatke, and F. Seitz. Treating the Hooking Effect in satellite altimetry data: A case study along the Mekong River and its tributaries. *Remote Sensing*, 8(2): 1–22, 2016. doi: 10.3390/rs8020091.
- P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, New York, 2nd edition, 2002. ISBN 0040-1706.
- S. Calmant, F. Seyler, and J. F. Cretaux. Monitoring continental surface waters by satellite altimetry. *Surveys in Geophysics*, 29(4-5):247–269, 2008. doi: 10.1007/s10712-008-9051-1.
- N. A. C. Cressie. *Statistics for Spatial Data*. New York : John Wiley & Sons, Inc., 1993. ISBN 9780471843368.
- N. A. C. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. New York : John Wiley & Sons, Inc., 2011. ISBN 9780471692744.
- J.-F. Crétaux, W. Jelinski, S. Calmant, A. Kouraev, V. Vuglinski, M. Bergé-Nguyen, M.-C. Gennero, F. Nino, R. A. Del Rio, and A. Cazenave. SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data. *Advances in Space Research*, 47(9): 1497–1507, 2011. doi: 10.1016/j.asr.2011.01.004.
- L. De Cesare, D. E. Myers, and D. Posa. Estimating and modeling space-time correlation structures. *Statistics and Probability Letters*, 51(1):9–14, 2001. ISSN 01677152. doi: 10.1016/S0167-7152(00)00131-0.
- S. De Iaco, D. E. Myers, and D. Posa. Space-time analysis using a general product-sum model. *Statistics and Probability Letters*, 52(1):21–28, 2001. ISSN 01677152. doi: 10.1016/S0167-7152(00)00200-5.
- S. De Iaco, D. E. Myers, and D. Posa. On strict positive definiteness of product and product-sum covariance models. *Journal of Statistical Planning and Inference*, 141(3):1132–1140, 2011. ISSN 03783758. doi: 10.1016/j.jspi.2010.09.014.

- G. De Marsily. Quantitative hydrogeology. Technical report, Paris School of Mines, Fontainebleau, 1986.
- I. De Oliveira Campos, F. Mercier, C. Maheu, G. Cochonneau, P. Kosuth, D. Blitzkow, and A. Cazenave. Temporal variations of river basin waters from Topex/[Poseidon satellite altimetry. Application to the Amazon basin. *Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science*, 333(10):633–643, 2001. doi: 10.1016/S1251-8050(01)01688-3.
- Global Runoff Data Center (GRDC). Long-term mean monthly discharges and annual characteristics of GRDC stations. Technical report, Federal Institute of Hydrology, Global Runoff Data Center, 2013.
- B. Gräler, L. Gerharz, and E. J. Pebesma. Spatio-temporal analysis and interpolation of PM10 measurements in Europe. *ETC/ACM Technical Paper 2011/10*, 2011.
- A. Gupta and S. C. Liew. The Mekong from satellite imagery: a quick look at a large river. *Geomorphology*, 85(3):259–274, 2007.
- G. B. M. Heuvelink, P. Musters, and E. J. Pebesma. Spatio-temporal kriging of soil water content. *Geostatistics Wollongong*, 96:1020–1030, 1997.
- G. H. Jirka. *Einführung in die Hydromechanik*. KIT Scientific Publishing, 2007.
- K. Koch. *Parameter estimation and hypothesis testing in linear models*. Springer Science & Business Media, 1999.
- D. G. Krige. A statistical analysis of some of the borehole values in the Orange Free State goldfield. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 53:47–70, 1952.
- B. Lehner, K. Verdin, and A. Jarvis. HydroSHEDS technical documentation, version 1.0. *World Wildlife Fund US, Washington, DC*, pages 1–27, 2006.
- P. Maillard, N. Bercher, and S. Calmant. New processing approaches on the retrieval of water levels in ENVISAT and SARAL radar altimetry over rivers: A case study of the São Francisco River, Brazil. *Remote Sensing of Environment*, 156:226–241, 2015. doi: 10.1016/j.rse.2014.09.027.
- Mekong River Commission (MRC). *Overview of the Hydrology of the Mekong Basin*. Vientiane, 73pp, 2005.

- C. I. Michailovsky, C. Milzow, and P. Bauer-Gottwein. Assimilation of radar altimetry to a routing model of the Brahmaputra River. *Water Resources Research*, 49(8):4807–4816, 2013. ISSN 00431397. doi: 10.1002/wrcr.20345.
- J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970. ISSN 00221694. doi: 10.1016/0022-1694(70)90255-6.
- K. Nielsen, L. Stenseng, O. B. Andersen, and H. Villadsen. Validation of CryoSat-2 SAR mode based lake levels. *Remote Sensing of Environment*, 171:162–170, 2015. ISSN 00344257. doi: 10.1016/j.rse.2015.10.023.
- S. Rouhani and D. E. Myers. Problems in space-time kriging of geohydrological data. *Mathematical Geology*, 22(5):611–623, 1990. ISSN 08828121. doi: 10.1007/BF00890508.
- U. Schneider, A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese. GPCP full data reanalysis version 6.0 at 0.5: monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. doi: 10.5676/DWD_GPCP. *FD_M_V6_050*, 2011.
- C. Schwatke, D. Dettmering, E. Börgens, and W. Bosch. Potential of SARAL/AltiKa for Inland Water Applications. *Marine Geodesy*, 38, 2015a. doi: 10.1080/01490419.2015.1008710.
- C. Schwatke, D. Dettmering, W. Bosch, and F. Seitz. DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, 19:4345–4364, 2015b.
- J. O. Skøien and G. Blöschl. Spatiotemporal topological kriging of runoff time series. *Water Resources Research*, 43(9), 2007. ISSN 00431397. doi: 10.1029/2006WR005760.
- J. J. J. C. Snelvangers, G. B. M. Heuvelink, and J. A. Huisman. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 112(3-4):253–271, 2003. ISSN 00167061. doi: 10.1016/S0016-7061(02)00310-5.
- M. J. Tourian, A. Tarpanelli, O. Elmi, T. Qin, L. Brocca, T. Moramarco, and N. Sneeuw. Spatiotemporal densification of river water level time series by multimission satellite altimetry. *Water Resources Research*, 2016.

- J. M. Ver Hoef and E. E. Peterson. A Moving Average Approach for Spatial Statistical Models of Stream Networks. *Journal of the American Statistical Association*, 105(489):6–18, 2010. ISSN 0162-1459. doi: 10.1198/jasa.2009.ap08248.
- J. M. Ver Hoef, E. Peterson, and D. Theobald. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, 13(4):449–464, 2006. ISSN 13528505. doi: 10.1007/s10651-006-0022-8.
- H. Villadsen, O. B. Andersen, L. Stenseng, K. Nielsen, and P. Knudsen. CryoSat-2 altimetry for river level monitoring — Evaluation in the Ganges–Brahmaputra River basin. *Remote Sensing of Environment*, 168:80–89, 2015. ISSN 00344257. doi: 10.1016/j.rse.2015.05.025.
- Y. Yoon, M. Durand, C. J. Merry, and E. Rodriguez. Improving temporal coverage of the SWOT mission using spatiotemporal kriging. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3):1719–1729, 2013. ISSN 19391404. doi: 10.1109/JSTARS.2013.2257697.