Simulation-based Evaluation of Hydrographic Data Analysis for Dune Tracking on the River Rhine

Felix Lorenz[1] · Thomas Artz[1] · Thomas Brüggemann[1] · Julius Reich[1] · Robert Weiß[1] · Axel Winterscheid[1]

Abstract

Knowledge of the static and morphodynamic components of the river bed is important for the maintenance of waterways. Under the action of a current, parts of the river bed sediments can move in the form of dunes. Recordings of the river bed by multibeam echosounding are used as input data within a morphological analysis in order to compute the bedload transport rate using detected dune shape and migration. Before the morphological analysis, a suitable processing of the measurement data is essential to minimize inherent uncertainties. This paper presents a simulation-based evaluation of suitable data processing concepts for vertical sections of bed forms based on a case study at the river Rhine. For the presented spatial approaches, suitable parameter sets are found, which allow the reproduction of nominal dune parameters in the range of a few centimetres. However, if parameter sets are chosen inadequately, the subsequently derived dune parameters can deviate by several decimetres from the simulated truth. A simulation-based workflow is presented, to find the optimal hydrographic data processing strategy for a given dune geometry.

Keywords Dune tracking · Hydrography · Outlier detection · Parameterized point cloud approximation · Geometric simulation

Zusammenfassung

Simulationsbasierte Bewertung hydrographischer Datenanalyse zur Dünenverfolgung am Rhein. Für die Unterhaltung von Wasserstraßen ist die Kenntnis der statischen und morphodynamischen Komponenten des Flussbetts wichtig. Aufgrund der Strömungswirkung können sich die Sedimente des Flussbetts in Form von Dünen bewegen. Dabei werden Aufnahmen der Flusssohle mittels Fächerecholot als Eingangsdaten für morphologische Analyse verwendet, um anhand der detektierten Dünenformen und deren Migration die Geschiebetransportrate zu berechnen. Vor der morphologischen Analyse ist eine geeignete Aufbereitung der Messdaten notwendig, um inhärente Unsicherheiten zu minimieren. In diesem Beitrag wird anhand einer Fallstudie am Rhein eine simulationsbasierte Auswertung geeigneter Datenverarbeitungskonzepte für vertikale Schnitte von Dünen präsentiert. Für die vorgestellten flächenhaften Ansätze werden geeignete Parametersätze gefunden, die eine Reproduktion der nominalen Dünenparameter im Bereich von wenigen Zentimetern erlauben. Werden die Parametersätze jedoch unzureichend gewählt, können die anschließend abgeleiteten Dünenparameter um mehrere Dezimeter von der simulierten Referenz abweichen. Der Artikel beschreibt einen simulationsbasierten Arbeitsablauf, um die optimale hydrographische Datenverarbeitungsstrategie für eine gegebene Dünengeometrie zu finden.

1 Introduction

The safety and ease of navigation on waterways is an important part of the traffic infrastructure. Thus, in Germany, the German Federal Waterways and Shipping Administration (Wasserstraßen- und Schifffahrtsverwaltung, WSV) is responsible for the implementation of maintenance measures such as traffic safety soundings, maintenance dredging, or bedload input to stabilize the river bed. As the supreme

✉ Felix Lorenz
  lorenz@bafg.de

1 Federal Institute of Hydrology, Am Mainzer Tor 1, 56068 Koblenz, Germany

authority within the Federal Ministry of Transport and Digital Infrastructure (Bundesministerium für Verkehr und digitale Infrastruktur, BMVI), the Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde, BfG) supports the WSV with research and development (R&D) and consulting. One of those R&D projects is the MAhyD project (morphodynamic analyses using hydroacoustic data, Morphodynamische Analysen mittels hydroakustischer Daten – Sohlstrukturen und Geschiebetransport).

Within the scope of maintenance surveys, the knowledge of both the static and morphodynamic component of the river bed is of great importance. Due to sediment transport, in most cases mobile unevenness is formed on the river bed, which are referred to as bed forms (Bechteler 2006) or dunes, leading to short-term changes in depth within the fairway. The ability to accurately predict dune dimensions and hindcast flows is limited by an incomplete understanding of what controls dune size and shape in rivers (Bradley and Venditti 2019). Thus, the available water depth for shipping needs to be monitored regularly by echosounding measurements. A suitable concept for the evaluation of the bed forms with regard to their geometry is necessary in order to estimate the extent of potential shoals in the active shipping area (fairway box). Due to objects on the river bed, measurement uncertainty and errors, such data processing is necessary.

In order to derive a high quality user product from the unevenly spaced 3D bathymetric data, grid-based methods for outlier detection (Debese and Bisquay 1999; Lu et al. 2010) and processing of digital terrain models (Segu 1985; Claussen and Kruse 1988) have been applied. In literature, dune tracking based on 2D bed form profiles (BFP) is often applied in order to estimate bedload transport rates using dune shape and migration (Claude et al. 2012; Gutierrez et al. 2013; van der Mark and Blom 2007). In this work, BFPs have been derived by cutting vertical sections from multibeam echosounder (MBES)-based digital terrain models (DTM), to overcome the deficiencies of single beam measurements. However, the process from high-quality measurement data acquisition and almost traditional surface-based data processing to BFP is only rarely discussed in bed form related publications, although it is crucial for morphological analysis. Thus, a workflow of automatic MBES data processing, including surface-based outlier detection and DTM generation with a comprehensive description of the applied methods is presented in this paper. In this way, we demonstrate the sensitivity of morphological analysis on the hydrographic data processing.

Basis of the outlier detection and DTM generation is the approximation of gridded 3D points by the estimation of polynomial parameters. The derivation of morphological parameters (here: dune parameters) are crucially affected by parameter settings in data processing. As a part of the presented workflow, suitable processing parameters are found using a geometrical simulation in order to conduct nominal-actual comparisons. One major task of the MAhyD project is the generation of suitable data input for dune tracking analysis. The main goals of this paper are to underline the necessity of correctly applied data processing of MBES data for further morphological analysis and to present one possible method on how to handle this issue.

The paper is structured as follows: In Sect. 2 a description of the conducted measurements is given. Section 3 describes the processing methods, including automatic outlier detection (3.1), point cloud approximation (3.2), and dune tracking analysis (3.3).

The methods for outlier detection and point cloud approximation are based on classical geodetic methods. A simulated sampling of a known reference surface, which is presented in Sect. 3.4, enables the conduction of nominal-actual comparisons. These nominal-actual comparisons are used to evaluate suitable parameter sets for the presented processing methods. Section 4 presents the simulation-based results of these nominal-actual comparisons for outlier detection (Sect. 4.1) and point cloud approximation (Sect. 4.2), which are applied to real measurement data in Sect. 4.3. These results are discussed in Sect. 5 and all crucial findings are summarized in Sect. 6.

2 Data Acquisition

The main focus of this paper is on synthetic data. However, real measurement data is necessary in order to generate a simulation using realistic dune shapes. Furthermore, results of the methods applied to synthetic as well as to real data can be compared. Thus, special dune tracking surveys on the river Rhine have been carried out. These echosoundings were collected by a measurement vessel using Kongsberg Maritime multibeam echosounder EM3002. Combined with the positioning by a Trimble antenna SPS185 in PDGPS mode, an estimation of heading by a Seapath 330 system and an inertial measurement unit (MRU5+), measurement points in a consistent reference frame can be retrieved. A typical average depth of 3.5 m below the transducer with a beam divergence of $1.5° \times 1.5°$ yield beam footprints in the range of 0.09–0.27 m, depending on the beam angle. According to the measurement setup the sampling rate of the transducer is $20\,Hz$. Due to the limited measurement frequency a high point cloud density can only be achieved at a low velocity of the vessel. On the river Rhine, a velocity of $1\frac{m}{s}$ over ground can be considered as very low while preserving the manoeuvrability. The point cloud density is about 400 points per m^2.

3 Methods for 3D Data Analysis and Dune Tracking

To demonstrate the importance of a thorough hydrographic data processing, the applied methods are described in this section. Our approaches of both data cleaning and subsequent grid modeling are based on the approximation of polynomials to the raw 3D point cloud. Finally, profiles are cut from the DTM of the river bed to derive an input for the bed form analysis.

3.1 Outlier Detection

MBES data is available in large numbers and is affected by multiple random and systematic deviations (Hare 1995). Furthermore, there may be disturbed measurements or points due to resuspension, wash of propellers, vegetation, et cetera, which are not part of the river bed. Thus, in bathymetry, the percentage of outliers in local areas can be very high. In a spatial outlier detection approach, the erroneous measurements and the points not belonging to the river bed should be detected and marked as invalid and excluded from further processing. In contrast to traffic safety measurements, the focus of the MAhyD project is only on the river bed including the bed forms, not on objects lying on it. The outlier detection should not be too restrained, so points not belonging to the river bed should be marked as invalid. False negatives (i.e. outliers marked as river bed points) should strictly be avoided.

In order to enable automatic spatial outlier detection, the river bed can be described mathematically by a continuous surface in three-dimensional Cartesian space. For this purpose rational polynomials as a function of the easting (E) and northing (N) of the river bed points

$$z = \varphi(E, N) = \sum_{i=0}^{l} \sum_{j=0}^{t} a_i E^i \cdot b_j N^j \qquad (1)$$

are approximated to the measured data with the polynomial parameters a and b (Artz and Willersin 2020), whereby the polynomial order is given by l and t (see Table 1). In order to enable a reliable detection of outliers, the robust parameter estimation in the form of an L1 norm estimator is used. Contrasting the least squares approximation (L2-Norm estimator,

e.g. Koch 1999), the weighted sum of absolute residuals $\mathbf{r}$ is minimized. By applying an iterative weight adjustment, the L1-Norm approximation can be transferred to the classical least squares solution scheme. Thus, as a starting point

$$\left(\mathbf{A}^T \mathbf{W}_k \mathbf{A}\right)\mathbf{x} = \mathbf{A}^T \mathbf{W}_k \mathbf{l}$$
$$\mathbf{A} = \frac{\partial \varphi(E, N)}{\partial \mathbf{x}} \qquad (2)$$

is solved in order to estimate the polynomial parameters $\mathbf{x} = \left[a_0, ..., a_l, b_0, ..., b_t\right]$ from the observations $\mathbf{l}$ (z-component of the 3D points). Subsequently, the main diagonal elements $w_{i,k}$ of the weight matrix $\mathbf{W}_k$ are redefined in each iteration k (Neitzel 2004; Schlossmacher 1973)

$$w_{i,k} = \frac{1}{|r_{i,k-1}| + \epsilon} \qquad (3)$$

by taking the residuals $\mathbf{r}$ into account. In order to avoid division by zero, a small value ϵ is added to the residuals. When all weights are unchanged, the iterative L1-norm approximation has converged. Then, with the help of the studentized residuals

$$t_i = \frac{r_i}{s_0^2 \rho_i} \qquad (4)$$

for each observation a critical value as a function of the a posteriori variance factor

$$s_0^2 = \frac{\mathbf{r}^T \cdot \mathbf{W} \cdot \mathbf{r}}{n - u}, u = l + t + 2, n = \#\text{observations} \qquad (5)$$

and the partial redundancies ρ_i is calculated. This information is used in order to perform a hypothesis test with a significance level of 95 %, to decide whether an observation represents an outlier or not (Koch and Yang 1998). Controlled by the parameter t_{statmin}, residuals $t_i \leq t_{\text{statmin}}$ are not marked as outliers in order to compensate for ground roughness.

At places with large terrain inclinations, such as a dune's lee side, large residuals can occur, while the perpendicular distance of a point to the model surface is small. As a further optional criterion, it is checked whether this distance also exceeds the specified geometric limit. To determine the distance between point and surface function, an extremum problem is solved. The roots of the first derivative of the function are calculated numerically using the Newtonian method (Gellert et al. 1978).

The outlier detection procedure is applied locally. For this purpose, a gridding of the point cloud with a given grid cell size c_s is performed. The approximation and the statistical test are performed only for the measurement data within a single grid cell, whereby the easting and northing of the data is reduced to the cell centre.

Table 1 Polynomial parameters a_i, b_j of different model types m_t

m_t	l	t	max($l + t$)	Description of model type
2	1	1	1	Plane
3	1	1	2	Hyperbolic paraboloid
4	2	2	2	Elliptical paraboloid
5	3	3	3	Cubic function

In the actual outlier elimination process, only four types of polynomials – with specific order l and t – are used. These are described by the model type m_t and are depicted in Table 1. Only polynomial parameters being significant in a student's t test (Koch 1999) are used in the further computation in order to use the best performing model type.

3.2 DTM Modeling

A transition from point clouds to a surface model is necessary due to the existing measurement noise (even after outlier detection). To obtain a homogeneous DTM for the dune tracking analysis, the measurement data is gridded. Due to the high point cloud density of the dune tracking measurements, comparatively small cell sizes c_s can be used. Decreasing the cell size leads to a smaller grid point distance, which should not be less than the footprint size (see Sect. 2).

For each grid point, a polynomial is approximated to the subset of the 3D point cloud within a given radius according to Eq. 1. In this adjustment approach, there are differences to the outlier detection described in Sect. 3.1. There is no reduction to an L1 norm estimator, but a least squares adjustment (L2 norm estimator) is applied.

Furthermore, the diagonal entries of the weight matrix $\mathbf{W}$ are filled by a distance dependent weighting function

$$a = \frac{r_{A,B}}{w - 1} + \sqrt{\frac{r_{A,B}^2}{(1 - w)^2} - \frac{r_{A,B}^2}{1 - w}}$$
$$W_{i,i} = \frac{a^2}{(d_i + a)^2}. \tag{6}$$

The points within a selected influence radius $r_{A,B}$ are included in the estimation with a weight depending on their horizontal distances to the grid point d_i. A maximum weight of $W_{i,i} = 1$ is given by $d_i = 0$ and $d_i = r_{A,B}$ yields the minimum weight of $W_{i,i} = \frac{1}{w}$. A higher value for w results in a lower weighting of remote points. If the specified redundancy is not achieved with the radius of influence r_A, this radius is incrementally increased to the extended radius of influence r_B. Given the high point densities and high grid resolution, the distant depending weighting might be not necessary. However, to preserve the local consistency at the crest and trough of the dunes, this approach might be considered expedient.

The evaluation of the spatial polynomial at the position of the respective grid point results in the grid point height. Including the height information, the grid represents a DTM. In contrast to an averaging per grid cell, the adjustment of polynomials represent the rather smooth but variable structure of the bed forms in a more realistic way.

By triangular meshing of the DTM, a BFP can be determined with a mathematically clean profile section (i.e. the height of the triangular surfaces along the profile line) for further dune tracking analysis.

3.3 Bedforms ATM and RhenoBT

The applied dune tracking method in this work is based on the software RhenoBT (Frings et al. 2012) and Bedforms ATM (Gutierrez et al. 2018). For the MAhyD project both applications have been combined and automatized for processing large numbers of data sets. It is a 2D profile (vertical section) based method for the determination of geometric parameters and migration of bed forms. In a first step a wavelet analysis is performed to identify dominant wavelengths in a signal (Bedforms ATM). Related to a 2D profile of the river bed, it delivers information about predominant bed form lengths. Thus, it provides an objective decision basis for the definition of the input parameters of the subsequent zero-crossing procedure (van der Mark and Blom 2007), which is performed in RhenoBT. Within that procedure, the wavelet analysis specifies the window size used to compute the moving average of the 2D profile. The window size should approximately correspond to the expected bed form length, which has been determined before by the wavelet analysis. A constructed baseline (i.e. the line connecting two adjacent troughs) separates the bed forms from the immobile river bed. The RhenoBT software yields the geometrical attributes of each individual bed form (e.g. height and length). As depicted in Fig. 1, the dune length λ is the distance between two troughs, while the bed form height h is the height of the triangle formed by a crest and its two adjacent troughs.

Besides the determination of the geometric parameters, the profile-related bed form migration can be determined according to Frings et al. (2012) from data sets of different points in time by cross correlation.

3.4 Dune Simulation

In order to evaluate suitable parameter sets for the presented processing methods, a synthetic data set of a geometrically simulated dune field is applied. Generating a known reference enables accurate nominal-actual comparisons. This is a
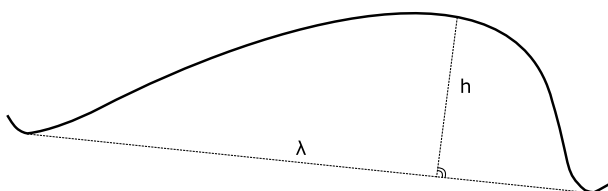

Fig. 1 Principal dune parameters length (λ), height (h) defined by the crest and the line connecting two adjacent troughs

purely geometric simulation and not a physical simulation as applied for example in Goll (2017) or in Nabi et al. (2013).

The longitudinal profile of a dune is mainly described by the parameters dune length λ and dune height h (see Sect. 3.3). Further dune parameters describe the course of the stoss and lee side, the skewness and curvature of the dune, and the height difference from adjacent troughs. Using these parameters various types of dune segments (see Lefebvre et al. 2016) can be generated.

To ensure comparability of simulation and real measurement (and thus to create transferability of findings from simulation to real measurement) it is important to create realistic dune shapes. Data from previous measurement campaigns (see Sect. 2) are used to gather realistic shapes. Large dunes ($\lambda = 6$–8.5 m, $h = 0.3$–0.45 m) and small dunes ($\lambda = 2.6$–3.3 m, $h = 0.17$–0.35 m) are constructed. The main points of a dune (beginning trough, crest, end trough) result from the definition of dune length and height (Sect. 3.3). Further condensing points on the dune profile can be determined by an Akima interpolation. Akima interpolation allows a smooth and natural behaviour of a curve (see Akima 1969). A resampling of the interpolation influences the achievable resolution of the 3D surface in the later course. Several 2D profiles of different dune parameters, compiled and arranged along a main axis, result in a three-dimensional dune field. Finally, a Delaunay triangulation of the resulting 3D point cloud is performed. This nominal surface can now be sampled as a synthetic echosounding data set using the Visualization Toolkit (VTK) (Schroeder et al. 2006). Considering the echosounder sampling rate and a realistic speed over ground, as presented in Sect. 2, every equidistant sampled point in a ping fan is a result of a sequence of transformations starting from the position and orientation
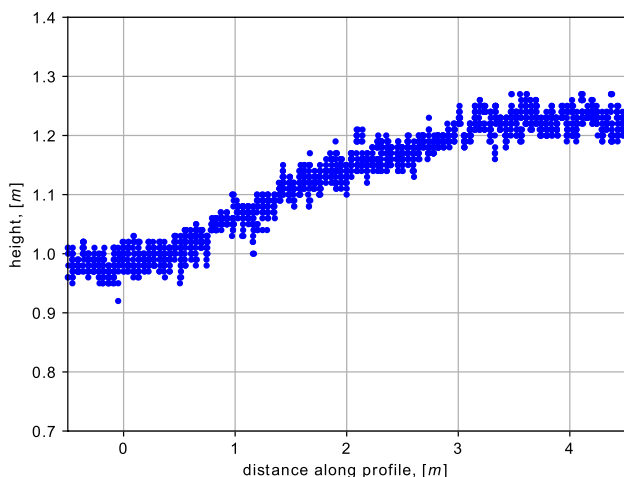
Table 2 Parameter sets and score of outlier detections applied to synthetic data

Set	c_s, [m] (phase 1,2)	False negtives	False positives	mcc, [−]
pp1	5	4675	0	0.04
pp2	1, 0.3	856	1023	0.8
pp3	2	903	6799	0.53

of the vessel. As the simulation is only used to obtain measurement points of a reference river bed with a realistic point density, further influences of the measurement process (e.g. footprint) are not of interest in the scope of this simulation. Thus, a specific sonar measurement principle is not considered. Random noise is added to the z-component of the sampled points in order to consider possible measurement noise of the multibeam echosounder, sediment turbulence, and all other influences that contribute to the noise band of the river bed.

4 Data Analysis Approaches and Results

4.1 Outlier Detection

Table 2 shows the different parameter sets (pp: plausibility parameters) for outlier detection. To evaluate the performance of various outlier detection approaches, the number of false negatives and false positives are compared, whereby false negatives are considered more problematic than false positives. Furthermore, the outlier detection as a binary classifier is evaluated by the Matthew Correlation Coefficient mcc as this evaluation measure is particularly suitable for highly uneven classes (Chicco and Jurman 2020), because all elements of the confusion matrix are considered.

Random noise is added to the sampling of the simulated nominal surface, which is generated according to Sect. 3.4. From the central area to the outer area of the measured strip, the level of added noise is increased from $\sigma = 0.01$ m over $\sigma = 0.015$ m to $\sigma = 0.025$ m. This reflects the increasing uncertainty potential of MBES measurements towards the border of the beam. These standard deviations were not derived from an uncertainty propagation (Hare 1995), but empirically derived from existing dune tracking measurements on the river Rhine. The feasibility of this approach can be seen by comparing real measurements (Fig. 2) with the synthetic data set (Fig. 3) as either noise floor is on the same level.

In addition to the noise, several pings were also distorted in order to generate erroneous observations. According to the grain size of coarse gravel, which is to be preserved, a point is considered an outlier if the difference compared to

Fig. 2 Profile section of real MBES data reveals the noise level of 3D measurement points

the nominal height exceeds 0.05 m. The measurement set contains 4682 outliers and 366,544 points in total.

In order to demonstrate the necessity of the outlier detection, some outliers in the range of –0.21–0.3 m have been added at several positions of the sampled dunes. A DTM has been derived from the 3D point cloud according to Sect. 3.2 to evaluate the impact of remaining outliers in the data set. Anticipating the findings in Sect. 5.2, the most suitable modeling parameter set mp3.1 (see Table 4) has been applied. The impact is evaluated by the average of the absolute values of the nominal-actual dune parameter differences $\overline{|\Delta\lambda|}$ and $\overline{|\Delta h|}$ in Table 3.

4.2 Modeling

Depending on the choice of the parameter set for modeling, the DTM (and thus finally any BFP derived from it) represents the river bed in more or less detail. Unfavourably chosen modeling parameters can result in incorrectly determined dune parameters. In the dune simulation, a noise band with a standard deviation of $\sigma = 0.025$ m, which has been chosen to match test surveys on the river Rhine, is applied. The choice of suitable modeling parameter sets should not be influenced by the outlier detection. However, in contrast to Sect. 4.1, no artificial outliers have been added. Several DTMs have been derived from the 3D point cloud by approximating polynomials (see Sect. 3.2) to evaluate the impact of various modeling parameters on the computation of dune parameters, which are depicted in Table 4. The dune parameters have been derived according to Sect. 3.3. The impact is evaluated by $\overline{|\Delta\lambda|}$ and $\overline{|\Delta h|}$ in Table 4 and by differences of the nominal and actual point cloud height in Fig. 4.

Table 3 Effect of remaining outliers on the derivation of dune parameters based on synthetic data

Outlier description	$\overline{|\Delta\lambda|}$, [m]	$\overline{|\Delta h|}$, [m]
Outliers remaining at crest, trough	0.38	0.02
Outliers eliminated at crest, trough	0.1	0.005
Outliers remaining before crest, at slip face	0.04	0.01

Table 4 Effect of modeling parameter sets—applied to synthetic measurement data—on the computation of dune parameters, grid cell average (avg) and median (med) compared to nominal parameters

| Set	c_s [m]	r_A [m]	r_B [m]	m_t [−]	w [−]	$\overline{|\Delta\lambda|}$ [m]	$\overline{|\Delta h|}$ [m]
mp0.0	0.5	1.0	5.0	3	100	0.788	0.08
mp0.1	0.1	1.0	5.0	3	100	0.728	0.072
mp1.1	0.1	0.16	0.26	3	50	0.099	0.008
mp1.2	0.1	0.16	0.26	3	25	0.089	0.008
mp2.0	0.1	0.4	0.8	3	0	0.094	0.018
mp2.1	0.1	0.4	0.8	3	2	0.087	0.017
mp2.2	0.1	0.4	0.8	3	25	0.072	0.012
mp2.3	0.1	0.4	0.8	3	50	0.075	0.01
mp2.4	0.1	0.4	0.8	3	75	0.083	0.009
mp2.5	0.1	0.4	0.8	3	100	0.084	0.009
mp3.0	0.1	0.4	0.8	4	0	0.09	0.006
mp3.1	0.1	0.4	0.8	4	2	0.064	0.003
mp3.2	0.1	0.4	0.8	4	25	0.07	0.005
mp3.3	0.1	0.4	0.8	4	50	0.092	0.006
mp3.4	0.1	0.4	0.8	4	75	0.091	0.007
mp3.5	0.1	0.4	0.8	4	100	0.093	0.007
mp4.1	0.25	0.4	0.8	4	2	0.105	0.004
mp4.2	0.25	0.4	0.8	4	25	0.102	0.003
avg1	0.1				Avg	0.126	0.017
avg2	0.5				Avg	0.213	0.018
med1	0.1				Med	0.143	0.019
med2	0.5				Med	0.203	0.016
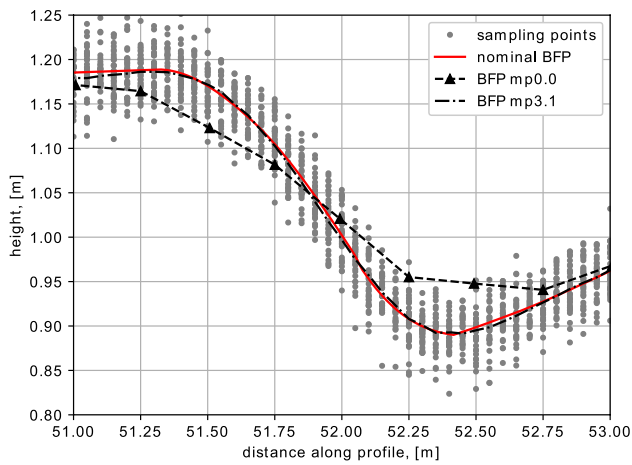
Fig. 3 BFP of set mp0.0, mp3.1 and nominal BFP based on synthetic measurement data

4.3 Modeling of Measured Data

In this section, the presented method for surface modeling is applied to real measurements, whereby the crucial question is whether the lessons learned from the simulation (Sect. 4.2) can be found in real measurements. The differences in dune parameters from BFPs consisting of 13 dunes are compared, which are based on different parameter sets for modeling (see Table 5). According to previous results,
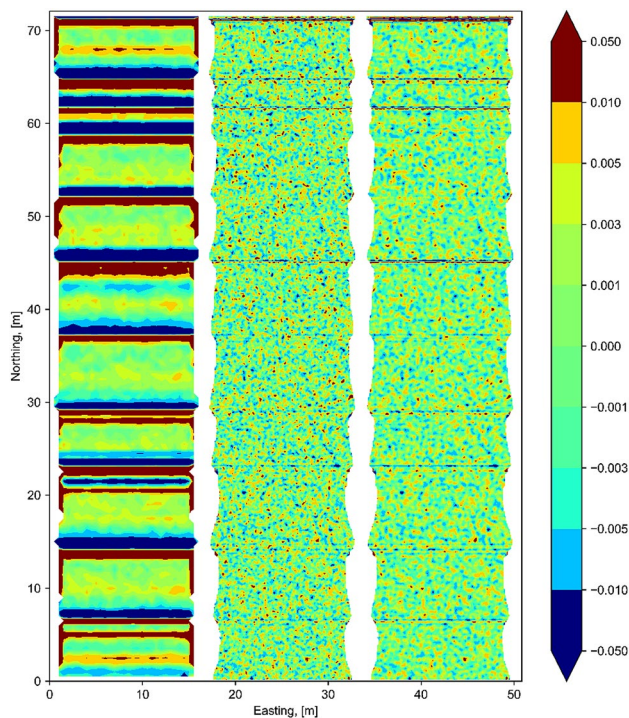
Fig. 4 Height differences in [m] to nominal height (left side: mp0.0, centre: mp3.2, right side: mp3.1) of the synthetic measurement data

the BFP with set pp2.2 (outlier detection) and set mp3.1 (modeling) serves as reference. Differences in model height due to different modeling parameters are presented in Fig. 5.

5 Discussion

5.1 Evaluation of Outlier Detection

A suitable outlier detection set in terms of a traffic safety survey is given by pp1 (see Table 2). In order to preserve objects lying on the river bed, pp1 uses $t_{statmin} = 0.30$ m. Using this parameter set, the vast majority of outliers remains undetected, the value of the mcc score is 0.04. For bed form analysis this approach is obviously not feasible as – in contrast to traffic safety surveys – the outlier detection should classify rather too many points as outliers, since the river bed and not the objects on it are of interest within dune tracking analysis. Applying Parameter set pp3 yields a high number of false positives. Since pp3 is too sensitive due to the large cell size of $c_s = 2$ m (mcc = 0.53), the cell size has to be reduced. Yielding the best outlier detection performance of $mcc = 0.8$, the most suitable set is given by pp2, an approach consisting of two phases of outlier detection with decreasing cell size. The remaining outliers (false positives) are close to the nominal height and therefore do not affect the further data processing. As a single outlier detection phase can be either too restrained or too sensitive, more phases have to be added successively. Here, these two phases are sufficient. In order to remove outliers close to the true surface,

Table 5 Effect of different modeling sets—applied to real measurement data—on the derivation of dune parameters in relation to the dune parameters derived by mp3.1. The outlier detection set is pp2.2

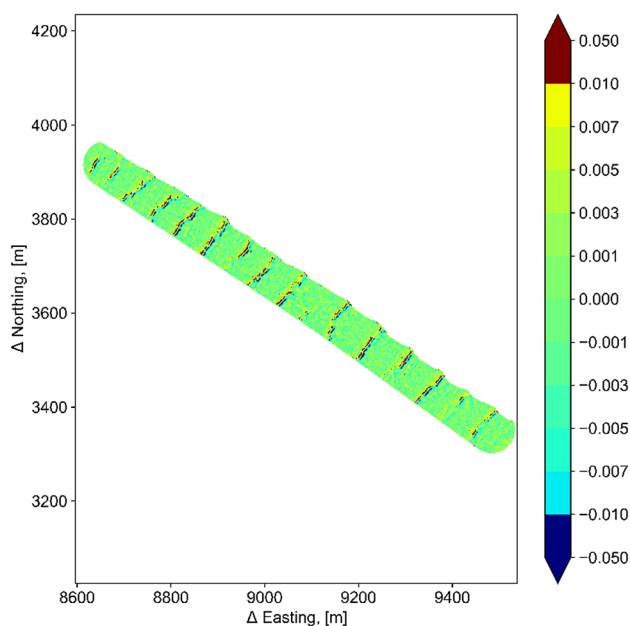| Set	$\overline{|\Delta\lambda|}$, [m]	$\overline{|\Delta h|}$, [m]
mp0.0	0.212	0.042
mp0.1	0.194	0.036
mp1.1	0.126	0.007
mp1.2	0.104	0.006
mp2.0	0.146	0.014
mp2.1	0.145	0.014
mp2.2	0.112	0.011
mp2.3	0.104	0.009
mp2.4	0.102	0.008
mp2.5	0.093	0.007
mp3.0	0.059	0.002
mp3.2	0.077	0.003
mp3.3	0.046	0.004
mp3.4	0.069	0.005
mp3.5	0.099	0.005
mp4.1	0.103	0.005
mp4.2	0.189	0.004

Fig. 5 Height differences in [m] of a DTM using mp2.4 to a DTM using mp3.1 based on real measured data

the outlier detection parameters have to be adapted to the geometric conditions of the specific river bed. Thus, a rule of thumb regarding the best possible parameter set for outlier detection cannot be pointed out. Hence, experience and a stepwise approach using the presented simulation are useful in order to apply a suitable outlier detection.

As shown in Table 3, remaining outliers in the data set have an impact on the derivation of dune parameters, especially if these outliers are located at critical positions of the dunes, which are the crest and trough. The detection and elimination of these outliers yield a better estimation of the dune parameters. Remaining outliers located before the crest and at the slip face have a significant smaller impact. However, these effects on the derivation of the dune parameters are in the range of the variations induced by employing different modeling parameters.

5.2 Evaluation of DTM Modeling

The objective of DTM modeling is to reduce the inherent measurement noise and to represent the dune geometry with a smooth model. Especially at the dunes crest and trough the actual geometry has to be represented adequately. Several parameters can be used to control the DTM modeling.

The effect of parameter variation on the derivation of dune parameters can be found in Table 4. Using set mp0.0, which is suitable for traffic safety sounding surveys, causes a loss of dune geometry due to the relatively large cell size.

A smaller cell size does not necessarily yield a more detailed approximation of the dune structure as the large influence radius in mp0.1 causes a strong smoothing effect.

Applying a too small influence radius (mp1.x) the DTM is rough, because the measurement noise can not be eliminated adequately.

A comparison of the parameter set groups mp2.x with mp3.x clarifies the benefit of a higher model type for the representation of the actual dune geometry. This effect is particularly evident in the estimation of dune height (see Table 4). In mp2.x a suitable influence radius is chosen, but the dune geometry cannot be approximated in detail by using model type $m_t = 3$. Especially lower model types require a high weighting factor in order to represent the structure of the dunes. Using a higher weighting factor results in a rough model due to the measurement noise. The considered dune geometry can only be suitably represented by using the model type $m_t = 4$. To reduce model roughness a small weighting factor should be applied. However, using equal weights is not an optimal approach as relatively large influence radii in combination with the variable dune geometry require a slight down-weighting of more distant measurement points. Increasing the weighting factor decreases the reduction of measurement noise (Fig. 4 centre and right side). All of these discussed effects of the various modeling parameters on the DTM can be found in Fig. 4. The results of applying parameter set group mp4.x demonstrate the necessity of a small cell size.

Systematic effects, visible as stripes in Fig. 4, are a result of the smoothing effect of the modeling and of the limited ground sampling, which is 0.05 m along track. These systematic effects could be reduced by applying a smaller cell size, but due to the echosounding footprint (see Sect. 2), a cell size less than $c_s = 0.1$ m is not feasible. Compared to the DTM derived by using mp3.2 (Fig. 4 centre), the systematic effects are more visible due to the smoother surface, retrieved by applying mp3.1. The reason for this phenomenon is that the model by applying mp3.1 is more consistent in transverse direction. Using parameter set mp3.2 yields a model, where systematic deviations are superimposed by random deviations.

Both the computation of average (avg1,2) and median (med1,2) of comparably small cells of 10 cm or 50 cm, respectively, lead to better results than set mp0.0 without any choice of further parameters. However, these approaches are not relevant for dune tracking, since the river bed is inaccurately represented due to the neglection of topographic parameters like slope or curvature around the point.

Based on the visual and numerical quality assessment, set mp3.1 is considered the most suitable one for the given dune geometry and the simulated survey characteristics.

5.3 Evaluation of DTM Modeling of Measured Data

The lessons learned from the simulation can be found again in the processing of real measurement data. As shown in Sect. 4.2, parameter group mp3.x yields comparable results (Table 5). Figure. 5 again reveals the importance of an appropriately chosen model type. By applying a too small chosen model type (parameter set mp2.4) the dune structure cannot be sufficiently represented. This leads to larger systematic model differences at the dunes crest and slope. Furthermore, small (< 1 cm) systematic differences appear, which are part of further investigations.

Thus, the findings of the simulation have been confirmed by the analysis of the real measurements. Finally, to find the optimal analysis procedure for a new dune tracking survey, the following procedure should be applied: (1) a rough estimation of the dune geometry should be done; (2) a simulation of the dunes in the surveyed area should be performed; (3) various analysis options should be tested by nominal-actual comparisons; (4) the actual measurements should be analysed with the best setting of step (3).

6 Conclusion

This paper presented spatial methods with suitable parameter sets for preparing MBES measurements to compute clean BFPs for further dune tracking analysis. The choice of parameters for the outlier detection and point cloud approximation has an impact on the derivation of dune parameters and finally on the estimation of bed load transport. Using the presented simulation, the effect of parameter variation is demonstrated.

Suitable parameter sets can be determined by nominal-actual comparisons, which are possible by simulation-based generation of known reference bed form data. Outliers have to be properly removed from the data set in order to estimate correct dune parameters. The outlier detection parameters have to be adapted to the specific geometrical conditions of the analysed river bed. Here, two phases of outlier detection are necessary as a single phase is either too restrained or too sensitive. For the river Rhine, a small cell size of $c_s = 0.1\,\mathrm{m}$ with a wider influence radius $r_A = 0.4\,\mathrm{m}$ and model type $m_t = 4$ is a preferred modeling parameter set.

The conducted simulation-based nominal-actual comparisons demonstrate reproducibility of the nominal dune parameters in the range of a few centimetres. In general, the height of dunes is easier to determine than the length.

Comparable impacts on the derivation of dune parameters can be found based on simulated as well as on real data. All investigations refer to simulated and real measurements of the river Rhine. In order to estimate parameter sets that are

suitable for other territories, a simulation adapted to that territory has to be carried out.

Assuming the approximate dune shapes of a conducted survey are known, optimized parameter sets for data processing can be determined using dune simulation. It is advisable, to analyse the impact of the processing options on the morphological analysis by simulation-based nominal-actual comparisons. In this way, the optimal processing strategy for the final analysis of the real measurements can be found. Furthermore, the simulation-based approach enables an evaluation of a suitable measurement concept to provide the best possible data input for dune tracking analyses.

Declarations

References

Akima H (1969) A method of smooth curve fitting. Tech. Rep, ESSA Research Laboratories

Artz T, Willersin D (2020) Hydrographische maßnahmen für eine smart-bwastr. Interdisziplinärer Wasserbau im digitalen Wandel 63:225–234

Bechteler W (2006) Sedimentquellen und Transportprozesse, Sustainable Sediment Management in Alpine Reservoirs considering ecological and economical aspects, vol 2. Universität der Bundeswehr München, Institut für Wasserwesen, Munchen

Bradley RW, Venditti JG (2019) Transport scaling of dune dimensions in shallow flows. J Geophys Res Earth Surf 124(2):526–547. https://doi.org/10.1029/2018JF004832

Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (mcc) over f1 score and accuracy in binary

classification evaluation. BMC Genom. https://doi.org/10.1186/s12864-019-6413-7

Claude N, Rodrigues S, Bustillo V, Bréhéret JG, Macaire JJ, Jugé P (2012) Estimating bedload transport in a large sand-gravel bed river from direct sampling, dune tracking and empirical formulas. Geomorphology 179:40–57. https://doi.org/10.1016/j.geomorph.2012.07.030

Claussen H, Kruse I (1988) Application of the DTM-program tash for bathymetric mapping. Int Hydrogr Rev 65(2)

Debese N, Bisquay H (1999) Automatic detection of punctual errors in multibeam data using a robust estimator. Int Hydrogr Rev

Frings R, Gehres N, de Jong K, Beckhausen C, Schüttrumpf H, Vollmer S (2012) Rheno bt, user manual. Institute of Hydraulic Engineering and Water Resources Management, RWTH Aachen University, Tech. Rep

Gellert W, Kästner H, Neuber S (1978) Auswertung geodätischer Überwachungsmessungen. Verlag Harri Deutsch, Fachlexikon ABC Mathematik, Germany

Goll A (2017) 3D Numerical Modelling of Dune Formation and Dynamics in Inland Waterways. PhD thesis, Universite Paris-Est

Gutierrez RR, Abad JD, Parsons DR, Best JL (2013) Discrimination of bed form scales using robust spline filters and wavelet transforms: methods and application to synthetic signals and bed forms of the río paraná, argentina. J Geophys Res Earth Surf 118(3):1400–1418. https://doi.org/10.1002/jgrf.20102

Gutierrez R, Mallma J, Núñez-González F, Link O, Abad J (2018) Bedforms-ATM, an open source software to analyze the scale-based hierarchies and dimensionality of natural bed forms. SoftwareX 7:184–189. https://doi.org/10.1016/j.softx.2018.06.001

Hare R (1995) Depth and position error budgets for mulitbeam echosounding. Int Hydrogr Rev

Koch KR (1999) Parameter estimation and hypothesis testing in linear models, 2nd edn. Springer, https://doi.org/10.1007/978-3-662-03976-2

Koch R, Yang Y (1998) Konfidenzbereiche und Hypothesentests für robuste Parameterschätzungen. ZfV Heft 1/1998

Lefebvre A, Paarlberg A, Winter C (2016) Characterising natural bedform morphology and its influence on flow. Geo Mar Lett 2016:379–393. https://doi.org/10.1007/s00367-016-0455-5

Lu D, Li H, Wei Y, Zhou T (2010) Automatic outlier detection in multibeam bathymetric data using robust its estimation. In: 2010 3rd International Congress on Image and Signal Processing 9:4032–4036

Nabi M, de Vriend HJ, Mosselman E, Sloff CJ, Shimizu Y (2013) Detailed simulation of morphodynamics: 3 ripples and dunes. Water Resour Res 49(9):5930–5943. https://doi.org/10.1002/wrcr.20457

Neitzel F (2004) Identifizierung konsistenter Datengruppen am Beispiel der Kongruenzuntersuchung geodätischer Netze. PhD thesis, Fakultät Bauingenieurwesen und Angewandte Geowissenschaften, Technische Universität Berlin

Schlossmacher EJ (1973) An iterative technique for absolute deviations curve fitting. J Am Stat Assoc 68(344):857–859

Schroeder WJ, Lorensen B, Martin K (2006) The visualization toolkit, 4th edn. Kitware Inc, Tech. Rep

Segu WP (1985) Terrain approximation by fixed grid polynomial. Photogramm Rec 11(65):581–591. https://doi.org/10.1111/j.1477-9730.1985.tb00525.x

van der Mark C, Blom A (2007) A new and widely applicable tool for determining the geometric properties of bedforms. No. 1568-4652 in Civil Eng. and Man. Res. Reports 2007R-003/WEM-002, Water Engineering and Management (WEM)