Spline retracker

a geometrical retracking algorithm for coastal and open ocean altimetry

Mafi, Saleh; Farzaneh, Saeed; Sharifi, Mohammad Ali; Forootan, Ehsan

Published in:
Marine Geodesy

DOI (link to publication from Publisher):
10.1080/01490419.2023.2291772

Publication date:
2024

Document Version
Accepted author manuscript, peer reviewed version

Spline retracker: a geometrical retracking algorithm for coastal and open ocean altimetry

Saleh Mafi[1], Saeed Farzaneh[1,*], Mohammad Ali Sharifi[1], Ehsan Forootan[2]

[1]*School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.*

[2]*Geodesy Group, Department of Sustainability and Planning, Aalborg University, Rendburggade 14, 9000 Aalborg, Denmark.*

[*]*Corresponding author: Associate Professor, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran; [*]Farzaneh@ut.ac.ir*

Abstract:

Satellite altimetry has enhanced the understanding of ocean dynamics through high-rate sampling and global coverage. However, there are some issues related to coastal areas due to land contamination and bad reflection effects in the altimeter and radiometer footprints. This study presents a geometrical method for retracking Jason-2 and Jason-3 altimetry waveforms in coastal areas. This method follows a geometrical assumption related to the symmetrical reciprocal motion of the radar pulse. Based on this assumption, the altimetry waveform is modeled as a continuous and differentiable third-order spline function, and the symmetry point of this function is considered as the retracking gate. As the spline is made by connecting the polynomial functions, there is no unique symmetry point for this function. This is the main problem of the method which is addressed by a weighted averaging of the polynomials' symmetry point. The spline retracking algorithm is validated against the tide gauges at Onsala, Halmstad, and Muscat stations in Sweden and Omani coasts, and its performance is compared with the OCOG (Offset Center of Gravity), threshold, ALES (Adaptive-Leading-Edge-Sub-waveform), and improved threshold algorithms. The spline results showed a remarkable reduction between 50% to 91% in the unbiased-Root-Mean-Squared-Error (ubRMSE) and an increase of at least 13% in the correlation coefficients when compared with the empirical algorithms in Swedish coastal waters. This algorithm presented almost equivalent results with the threshold and improved threshold retrackings in Muscat station, based on Jason-2 measurements. However, along the Jason-3 pass, our spline method showed a

considerable reduction of 80% in ubRMSE and the minimum increase of 42% in correlation coefficients than the empirical algorithms. This method also outperformed the ALES algorithm in most cases.

keywords: coastal altimetry, retracking, spline function, OCOG algorithm, threshold algorithm

1- Introduction

For more than three decades, satellite altimetry as a space-borne method has been used for collecting global sea level data. The principles of the method are fully-known with the aim of determining the range between the mass center of the satellite and the sea surface (Fu and Cazenave 2001; Vignudelli et al. 2011). At first, altimetry was used to study the Sea Surface Height (SSH) variations of open oceans and large ocean eddies. Nowadays, with the further technical development, SSH determination within a precision of 5 cm has become possible (Tapley et al. 1982; Sailor and Leschack 1987; Vignudelli et al. 2011; Agar et al. 2023). The precise estimations of the SSH motivated researchers to further evaluate the performance of this technique in shallow waters, such as rivers (Villadsen et al. 2016; Gao et al. 2019; Nielsen et al. 2022), lakes (Uebbing, Kusche, and Forootan 2015; Göttl et al. 2016; Deng et al. 2021), and coastal areas (Passaro et al. 2014; Passaro et al. 2018; Wang and Huang 2021; Agar et al. 2023). Moreover, in the new generation of altimetry missions, the SAR (Synthetic Aperture Radar) technique is used in missions, such as CryoSat-2, Sentinel-3A, Sentinel-3B, and Sentinel-6 MF, that takes the advantage of groups of coherent pulses. As a result, the radius of the radar footprint decreases about one order of magnitude along the satellite track, which has a remarkable impact on the altimetry results (Raney 1998; Passaro et al. 2022).

The critical challenge of radar altimetry in coastal areas is the contamination of waveforms with land effects (Cipollini et al. 2009), which yields diverse waveforms, including multi-peak, quasi-specular, Brownian with a Gaussian peak, and etc., see, e.g., (Vignudelli et al.

2011). Extracting SSH relevant parameters using these measurements is rather problematic (Idris and Deng 2012; Halimi 2013). The post-processing methods to obtain the parameters are known as “retracking” (Chelton et al. 2001).

According to the altimetry principles, determining SSH requires corrections, which are applied to the range. The retracking correction is the most important correction in shallow waters (Dinardo 2020; Wang and Huang 2021). So far, various methods have been developed to retrack the altimetry waveforms, which are divided into empirical and physically-based retrackers (Passaro et al. 2022). The empirical algorithms are mostly used in retracking the signals that are affected by the land and ocean ice sheets (Passaro et al. 2022). The implementation of these algorithms is easier than the physical ones, and their optimum parameters are derived through trial and error and empirical experiments (Dinardo 2020). The OCOG (Wingham, Rapley, and Griffiths 1986) and threshold (Bamber 1994) algorithms are typical examples of empirical methods. Previous studies carried out for inland waters (e.g., Roohi et al. 2021), coastal areas (e.g., Khaki, Forootan, and Sharifi 2014), and polar oceans (e.g., Shen et al. 2019) refer to the application of empirical algorithms. The physically-based retrackers, on the other hand, are mathematical models that have been developed to model the interactions of the electromagnetic pulse with the sea surface in the radar footprint (Dinardo 2020; Passaro et al. 2022). The geophysical parameters such as the Significant Wave Height (SWH), backscattering coefficient, and wind velocity are computable by these algorithms (Dinardo 2020). The physical models of open ocean (Brown 1977; Hayne 1980), beta parameter (Martin et al. 1983), and Ice-2 (Legrésy and Rémy 1997) are among physical algorithms, which are used to retrack the radar waveforms.

Previous studies have shown that the non-ocean effects appear on waveforms differently. For example, small features such as ships only influence the first gates before the leading edge

(Tournadre 2007), while the bright targets affect the trailing edge. These effects are very common in the waveforms located at a distance of 10 km from the coastline (Passaro et al. 2014). Other features such as big islands with high backscattering coefficients can make changes to the whole waveform (Tournadre 2007). These effects that cause systematic errors (or systematic effects) on waveforms, are reduced by waveform partitioning techniques. For this, in many cases, the empirical and physical retrackers are applied on a part of the waveform, which is called a sub-waveform. The ALES (Passaro et al. 2014), ALES+ (Passaro et al. 2018), and ALES+SAR (Passaro et al. 2022) are the examples of sub-waveform based retracking algorithms. Sub-waveform based retrackers have been examined in various inland studies (e.g., Uebbing et al. 2015; Roohi et al. 2021; Deng et al. 2021) and coastal area (Khaki et al. 2014; Arabsahebi et al. 2018; Idris et al. 2020; Agar et al. 2023).

In this study, a geometrical retracking method is introduced to reduce non-ocean effects on altimetry waveforms, for which there is no need to partition the waveform. A geometrical algorithm follows a geometrical assumption that is related to the movement of the radar pulse and tries to find one or some points (initial gates) on a waveform that satisfies the geometrical assumption. This is the main difference between our geometrical method and the empirical and physically-based algorithms. The classical waveforms (here, measurements of Jason-2 and Jason-3) are more contaminated by non-ocean effects due to their large footprints (with radius of 8.3 km) than the new SAR altimeters (along-track resolution of about 300 m) (Dinardo 2020). Therefore, these data are more suitable for evaluating the performance of a developed retracking algorithm.

The implementation of our geometrically-based spline retracking algorithm consists of three steps: Step 1, a third-order spline function fits the waveform, which changes it to a

continuous and differentiable function; Step 2, the symmetry point of the spline is computed. As the spline is made by connecting the third-order polynomials, there is no unique symmetry point for this function. Thus, the symmetry points of the constructing polynomials, which are called initial gates, are obtained individually; Step 3, the retracking gate of the waveform is computed by a linear combination of the initial gates. The coefficients of the linear combination are defined in such a way that they remarkably reduce the systematic effects of the waveform on the retracking gate.

An important retracking pre-processing step is the selection of waveforms, which is done to remove the waveforms with higher leading edge noise than a critical value. In other words, the waveforms with straight leading edges are better retracked by the retracking algorithms. Ignoring the waveforms with high thermal noise (Uebbing, Kusche, and Forootan 2015), calculating the dissimilarity value (Arabsahebi, Voosoghi, and Tourian 2018), and visual inspection (Vu et al., 2018) are used in the previous studies. In this study, an Artificial Neural Network (ANN) is applied to select waveforms.

The performance of the spline algorithm is assessed over the coastal waters of Sweden and the Oman Sea, which are two different coastal areas with distinct environments. In what follows in section 2, a brief description of the data and the case studies is provided. In Section 3, we describe the methodology of our spline retracking and the assessment criteria. In section 4, the performance of the proposed algorithm is evaluated in comparison with other algorithms, such as OCOG, threshold, ALES, and improved threshold retrackers. In section 5, the research findings are reviewed and some suggestions for future studies are provided.

2- Data and case studies

The Onsala (Latitude: 57.3920, Longitude: 11.9190) and Halmstad (Latitude: 56.6488, Longitude: 12.8358) stations on the southwestern coast of Sweden, i.e., the Baltic Sea, and the Muscat (Latitude: 23.6280, Longitude: 58.5650) station in the Gulf of Oman are considered for assessing the results. These stations exhibit different tidal regimes. The Baltic Sea is a large and shallow water area and counts as an enclosed basin. The sea is connected with the North Sea through the Danish Straits, and due to its limited connections with open oceans, the oceanic tides penetrate the sea weakly. The harmonic analysis of tide gauges in the Baltic Sea reveals that the amplitude of the diurnal constituents O_1 (period of 25.82 h) and K_1 (period of 23.93 h) is almost equal. The central part of the sea experiences minimum amplitudes of 0.4 cm for the diurnal tides, while maximum amplitudes happen in the Gulf of Finland (~3 cm) and the Gulf of Riga (~1.8 cm). The minimum and the maximum amplitudes of the main semidiurnal harmonic M_2 (12.42 h) are 0.2 cm and 6 cm, which are observed in the Gulf of Bothnia and the Danish Straits, respectively (Medvedev, Rabinovich, and Kulikov 2016).

Oman Sea from the northwest is connected with the Persian Gulf through the Strait of Hormuz, and the eastern part reaches the Arabian Sea. Unlike the Baltic Sea, which has a limited mass of water (Pranzini and Williams 2013), this sea has been connected with the Indian Ocean and is influenced by oceanic tides. The study carried out by Akbari et al. (2016) illustrates the amplitude variations of different diurnal and semidiurnal tidal constituents in the Gulf of Oman and the Arabian Sea. K_1 (amplitude: 35-40 cm), O_1 (15-20 cm), and M_2 (60-70 cm) are the most influential constituents in the tidal regime of this region. The continental slope has rapid changes along the Omani coast (Pous 2004), which makes an important factor in the complexity of the altimetry waveforms close to Muscat. Figure 1 illustrates the bathymetric maps of the Baltic and Oman Seas, which indicates the water depth variations in these areas.

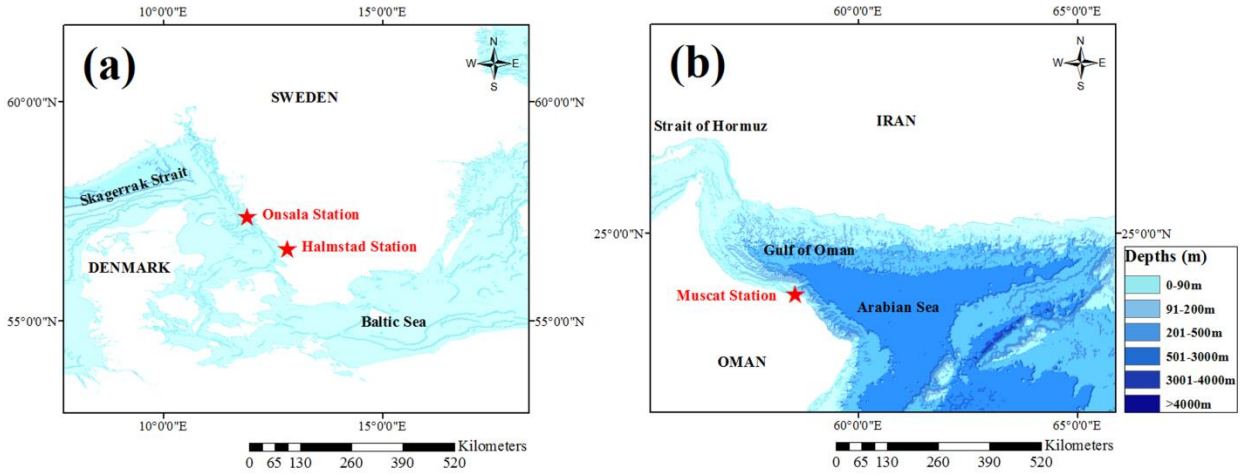


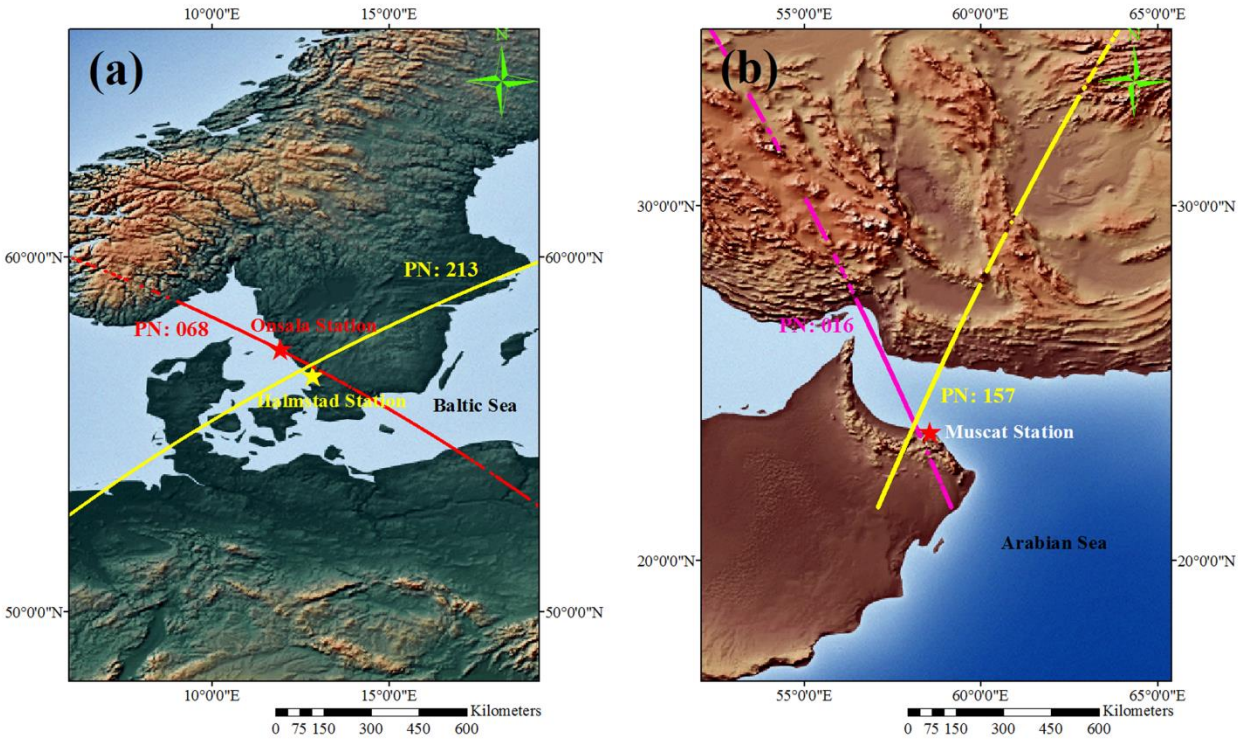
Figure 1.

The altimetry data that are used to produce the SSH time series have been obtained from Jason-2 and Jason-3 missions. Jason's tracks with pass numbers of 68 and 213 were selected for computing SSH time series close to the Onsala and Halmstad stations, respectively. The SSH time series at the Muscat station were obtained from Jason-2 and Jason-3 missions with pass numbers of 16 and 157, respectively. Table 1 shows a summary of the satellite data used in the study.

Table 1. The satellite data used in the selected coastal areas and the locations of the tide gauge stations. The Jason-2 data were not used in the SSH computations of Onsala station, because of the displacement of the satellite ground track.

Coastal zone	Mission/Pass	Tide gauge	Lat	Lon	Distance to station	Data duration
Swedish coast	Jason-3/068	Onsala	57.3920^0	11.9190^0	9 km	05 Sep 2016 09 Sep 2018
	Jason-2/213	Halmstad	56.6488^0	12.8358^0	27 km	03 Jan 2014 16 Mar 2016
	Jason-3/213				27 km	20 Oct 2016 24 Oct 2018
Gulf of Oman	Jason-2/016	Muscat	23.6280^0	58.5650^0	33 km	20 Sep 2014 23 Sep 2016
	Jason-3/157				55 km	13 Feb 2016 16 Feb 2018

١٥٢ We also used data from three permanent tide gauge stations to validate the altimetry time
١٥٣ series. Figure 2 illustrates the locations of Onsala, Halmstad, and Muscat stations in the south of
١٥٤ Sweden and the Omani coast.



١٥٥

١٥٦ **Figure 2.**

١٥٧ The altimetry data used in the study are from the Sensor and Geophysical Data Records
١٥٨ (SGDR) derived from Jason-2 and Jason-3 missions. These datasets include the altimetry
١٥٩ waveforms, the media, and geophysical corrections, which are needed for SSH computations.
١٦٠ Moreover, we also used the SSH time series of altimetry derived for the ALES retracking that
١٦١ have been provided by Open Altimeter Database.

١٦٢ **3- Methodology**

١٦٣ ***3.1- Waveform selection method***

Selecting suitable observations (i.e., altimetry waveforms in this study) is a necessary step to eliminate computational biases (Göttl et al. 2016). In coastal areas, altimetry waveforms are contaminated with non-ocean effects, which in some cases leads to waveforms with noisy leading edges. The goal is to select waveforms with straight leading edges, which was done by implementing an Artificial Neural Network (ANN) algorithm. The designed ANN was a three-layer network with one hidden layer that included six neurons. This architecture, which was implemented in MATLAB software, is considered following Mattes (2019), in which the ANN was used for waveform retracking. The number of neurons in the hidden layer was determined by performing an empirical trial and error (Table 2). The activation function of the neurons was considered to be a sigmoid function (Fausett 1994). A detailed description of the sigmoid function is presented in Appendix A.

The ANN input is selected to be the power from the registered waveforms, and the output is the Boolean code that determines whether the waveform is suitable for retracking or not. Since the sigmoid function is continuous, the ANN produces a float number, which can be rounded to its nearest integer number. Figure 3 illustrates the architecture of the ANN.

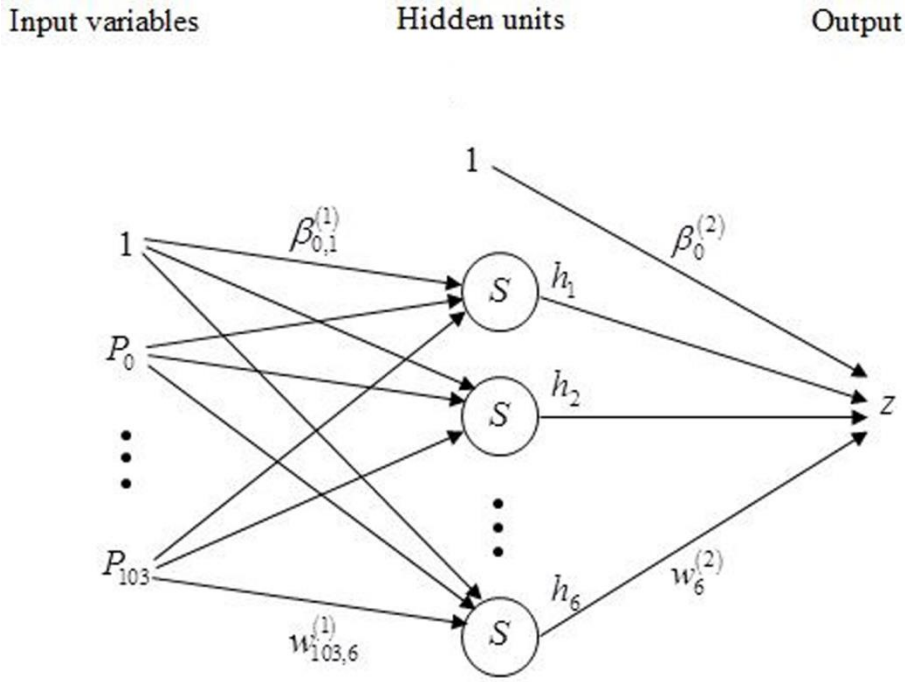


Figure 3.

In our experiment, the ANN parameters (the weights and biases) were set stochastically, and the network training was carried out by the training data (the waveforms' powers with their assigned Boolean codes). The optimum values of the parameters were derived by the back-propagation algorithm. The following equations describe the training process (Goodfellow et al. 2016; Lindholm et al. 2019):

$$\mathbf{b}^{(1)} = \begin{bmatrix} \beta_{0,1}^{(1)} & \cdots & \beta_{0,6}^{(1)} \end{bmatrix};\ \mathbf{W}^{(1)} = \begin{bmatrix} w_{0,1}^{(1)} & \cdots & w_{0,6}^{(1)} \\ \vdots & \cdots & \vdots \\ w_{103,1}^{(1)} & \cdots & w_{103,6}^{(1)} \end{bmatrix};\ \mathbf{b}^{(2)} = \begin{bmatrix} \beta_0^{(2)} \end{bmatrix};\ \mathbf{W}^{(2)} = \begin{bmatrix} w_1^{(2)} \\ \vdots \\ w_6^{(2)} \end{bmatrix};\ \mathbf{P} = \begin{bmatrix} \mathrm{p}_0 \\ \vdots \\ \mathrm{p}_{103} \end{bmatrix} \quad (1)$$

$$\mathbf{h} = \mathrm{S}\left(\mathbf{W}^{(1)^{\mathrm{T}}} \mathbf{P} + \mathbf{b}^{(1)^{\mathrm{T}}} \right) \quad (2)$$

$$z = \mathbf{W}^{(2)^{\mathrm{T}}} \mathbf{h} + \mathbf{b}^{(2)^{\mathrm{T}}} \quad (3)$$

$$Err = y - z \quad (4)$$

where $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$ are the bias vectors, $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ are the weight matrices, $\mathbf{P}$ is the vector of waveform's powers, $\mathbf{h}=\left[h_1 \cdots h_6\right]^{\mathrm{T}}$ is the vector of neurons' responses in the hidden layer, the $\mathrm{S}(x)$ function in Eq. (2) stands for the sigmoid function, and y is the true Boolean code assigned to the waveform. The above equations are called the feed-forward processing of the network. The back-propagation algorithm is a backward process that tries to minimize the cost function of the network (J_{k}) by updating the weight matrices and the bias vectors. The cost function is defined as follows:

$$J_{\mathrm{k}}=\frac{1}{2\mathrm{m}}\left(\mathbf{Err}_k{}^{\mathrm{T}}\mathbf{Err}_k\right)+\frac{\mathrm{q}}{2\mathrm{m}}\left\{\left(\sum_{i=0}^{103}\sum_{j=1}^{6}w_{ij}^{(1),k}\right)^2+\left(\sum_{j=1}^{6}w_j^{(2),k}\right)^2\right\}+\frac{\mathrm{q}}{2\mathrm{m}}\left\{\left(\sum_{j=1}^{6}\beta_{0,j}^{(1),k}\right)^2+\left(\beta_0^{(2),k}\right)^2\right\} \tag{5}$$

The following equations describe the back-propagation algorithm:

$$\Delta=Err\circ\mathrm{S}(z)\circ\left(1-\mathrm{S}(z)\right) \tag{6}$$

$$\mathbf{\Delta}^{(1)}=\mathbf{W}^{(2)^{\mathrm{T}}}\Delta\circ\mathrm{S}(\mathbf{h})\circ\left(1-\mathrm{S}(\mathbf{h})\right) \tag{7}$$

$$\mathbf{dW}^{(2)}=\alpha\Delta\mathbf{h}^{\mathrm{T}};\ \mathbf{db}^{(2)}=\alpha\Delta;\ \mathbf{dW}^{(1)}=\alpha\mathbf{\Delta}^{(1)}\mathbf{P}^{\mathrm{T}};\ \mathbf{db}^{(1)}=\alpha\mathbf{\Delta}^{(1)}; \tag{8}$$

$$\hat{\mathbf{W}}^{(1)}=\mathbf{W}^{(1)}+\mathbf{dW}^{(1)}+\frac{\alpha\mathrm{q}}{\mathrm{m}}\mathbf{W}^{(1)} \tag{9}$$

$$\hat{\mathbf{W}}^{(2)}=\mathbf{W}^{(2)}+\mathbf{dW}^{(2)}+\frac{\alpha\mathrm{q}}{\mathrm{m}}\mathbf{W}^{(2)} \tag{10}$$

$$\hat{\mathbf{b}}^{(1)}=\mathbf{b}^{(1)}+\mathbf{db}^{(1)}+\frac{\alpha\mathrm{q}}{\mathrm{m}}\mathbf{b}^{(1)} \tag{11}$$

$$\hat{\mathbf{b}}^{(2)}=\mathbf{b}^{(2)}+\mathbf{db}^{(2)}+\frac{\alpha\mathrm{q}}{\mathrm{m}}\mathbf{b}^{(2)} \tag{12}$$

where the symbol $\circ$ represents the Hadamard product (Million 2007), α is the learning rate, q is the Lagrange parameter, index k in Eq. (5) refers to the k$^{\mathrm{th}}$ iteration of the training, and m

defines the number of training waveforms. This process was iterated until $\left|\left(J_{\mathrm{k}}-J_{\mathrm{k\text{-}1}}\right)/J_{\mathrm{k}}\right|<\varepsilon$. The ε value was considered to be 10^{-8}, which was defined empirically. If the condition was not satisfied, the training would be terminated after 20000 iterations. $\hat{\mathbf{b}}^{(1)}$, $\hat{\mathbf{b}}^{(2)}$, $\hat{\mathbf{W}}^{(1)}$, and $\hat{\mathbf{W}}^{(2)}$ matrices include the optimum parameters of the ANN.

The visual inspection of the waveforms on the Swedish coastal waters revealed that there are two frequent types of waveforms in the areas. The first type is Brownian waveforms which adapt to Brown's physical model (Brown 1977), see also Figure 4. The second type is the quasi-specular waveforms (Idris and Deng 2012), which are similar to the normal distribution curve (Ghilani and Wolf 2006). These two types of waveforms (Figure 5 (b)) constituted at least 95% of the waveforms within the Swedish coastal waters.

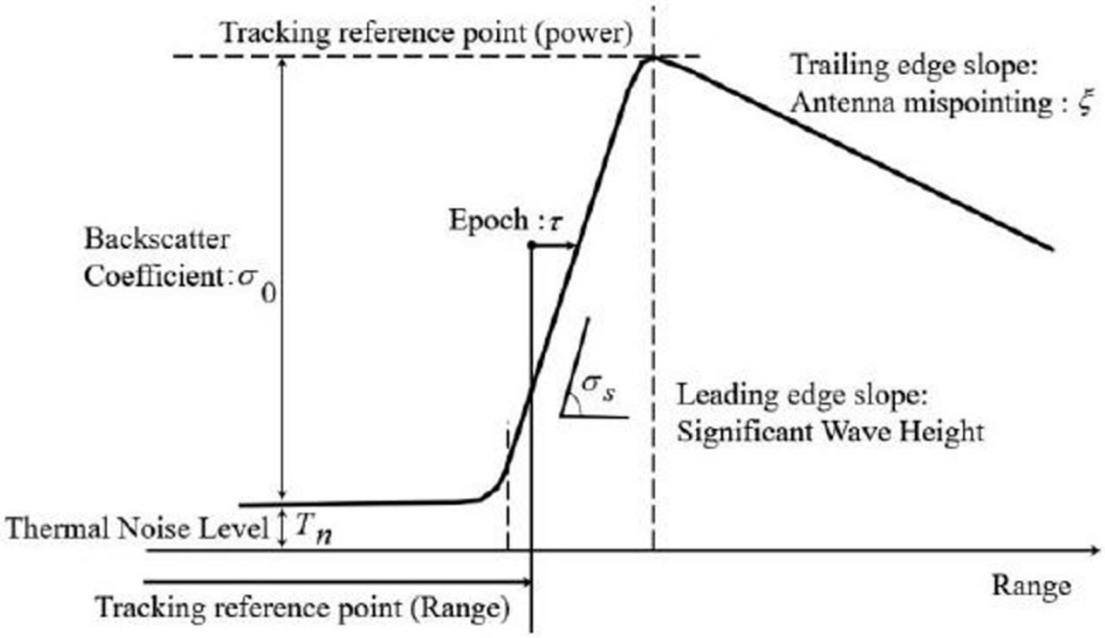


Figure 4.

The physical conditions are different on the Omani coast. At a distance of 0-25 km from the coast, the waveforms close to the coast of Oman were more diverse than the waveforms obtained

from the Swedish coastal waters. The continental slope along the Omani coast seems to be one of the main reasons for the waveform diversity. Using such information was useful in the ANN training process. The land and water waveforms were introduced to the ANN by 0 and 1 digital codes, respectively. The Brownian and quasi-specular waveforms (frequent waveform types in Swedish coastal waters) were tagged with code 1, while the noisy land waveforms were labeled with code 0. In Omani waters, as there were a few quasi-specular waveforms in the cycles, only Brownian waveforms were tagged with code 1. It is essential to note that the training waveforms should have good spatial and temporal distributions. It means that the centers of the footprints should be uniformly distributed along the satellite tracks, and the waveforms must be selected from different cycles so that the ANN can understand the spatial and temporal changes of the waveforms. The ANN training was carried out for each satellite track. Figure 5 illustrates a few of the waveforms that were used for training the ANN.

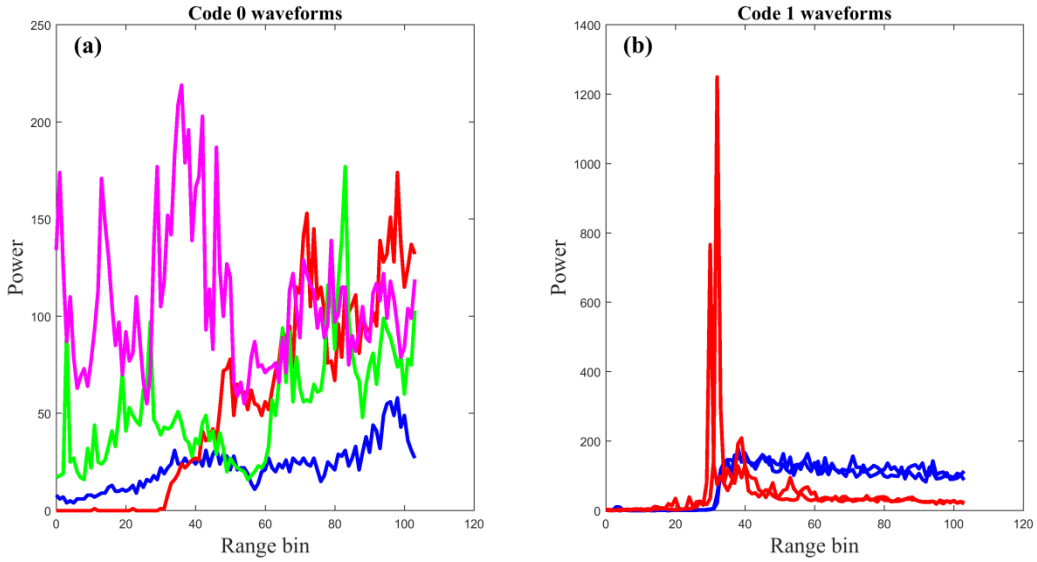


Figure 5.

In Figure 5, code 0 means that the selected waveform is not suitable for retracking, and those shown by code 1 refer to retrackable waveforms. The training data consisted of 70% of the whole dataset. The other 30% was used for validation of the results. After validation, a general criterion was applied to test the ANN performance. This criterion was defined based on the study conducted by Halimi et al. (2013). They stated that more than a quarter of waveforms did not follow the Brownian model up to 8 km from the coastline. This ratio is smaller at the distance of 15 km (see section 3.2), as given the radius of 8.3 km for the Jasons' footprints (Quartly 1998), almost all of the footprints at the distance of 8-15 km are located at the sea surface. Moreover, in some tracks, we considered the quasi-specular waveforms as code 1. Therefore, we expected that more than 75% of the data should be selected by the ANN. To prove this, we computed a ratio for the echograms that reveals the Number Of (NO) selected waveforms by the ANN to the total number of waveforms in the echogram. This ratio was computed for each echogram. Figure 6 describes the percentage of the selected waveforms in every track and each cycle. Investigation of the ANN performance showed that the network selected an average of 96% of the waveforms for the retracking process.

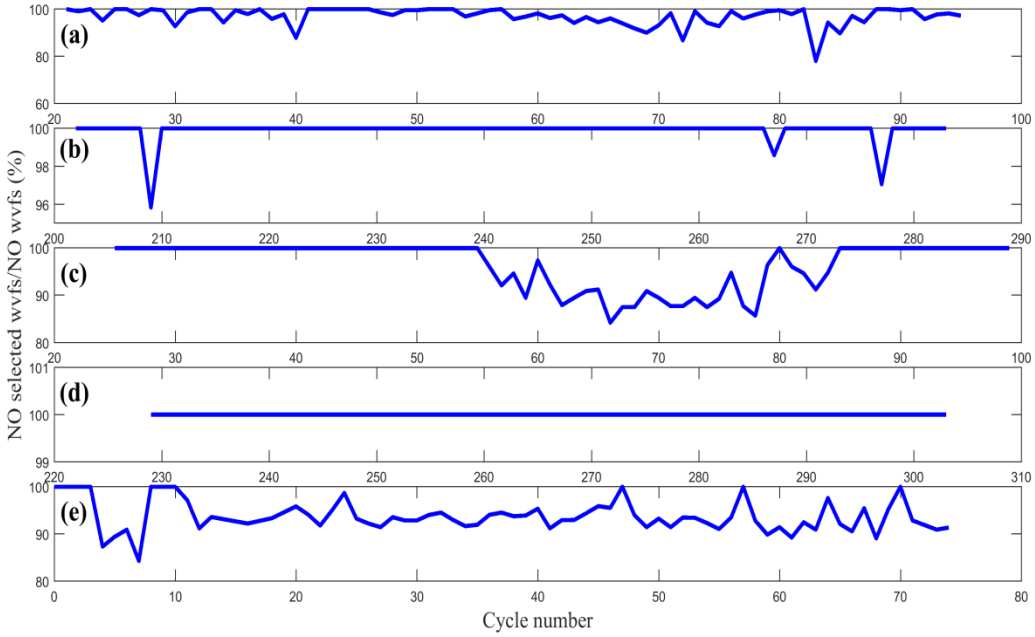


Figure 6.

It is worth mentioning that we also chose the number of neurons in the hidden layer empirically, where the ANN performance for different numbers of neurons is validated for each tide gauge station. For example, Table 2 illustrates the ANN errors in the Onsala station.

Table 2. Comparison of different architectures' performance in prediction of the waveform codes in Onsala station. The training accuracy shows the maximum accuracy in fitting the ANN model to the training data. The validation accuracy also refers to the performance of the trained ANN in the prediction of the validation data codes. Here, 40 waveforms were used for training and 15 for validation.

Number of neurons	Training accuracy (%)	Validation accuracy (%)
1	80	46.67
2	95	86.67
3, 5	97.5	86.67
4, 6-10	100	86.67

The results indicate that the ANN with 4, 6, 7, 8, 9, or 10 neurons in the hidden layer may have an optimized architecture. The same comparison was also carried out for other stations, where we concluded that the ANN with 6 neurons in the hidden layer has the best performance. The values of the learning rate and the Lagrange parameter (Lindholm et al. 2019) were also tested. These values are important for fitting the ANN model to the training data. Testing

different values for these parameters during the training process, α=0.0001 and q=0.002 were found as the most common values for our experiments.

3.2- Spline retracking

In the third-order spline formulation, a dataset is represented by imposing a set of geometrical assumptions, ensuring that the spline function will be continuous and differentiable in data points. A general description about the spline function has been provided in Appendix B.

The spline retracking algorithm relies on the geometrical assumption that the radar pulse follows a symmetrical reciprocal motion from the time of transmission to reception by the satellite antenna. According to this, the symmetry point of the waveform function was considered the retracking point. As the waveform function was made by connecting the third-order polynomials, there was no unique symmetry point for the waveform, which was the main problem of our algorithm. Therefore, the question was how to use the symmetry points to estimate the retracking gate? The symmetry point of a third-order polynomial is the inflection point which is derived from the second-order derivative of the polynomial function (Stewart 2012). Therefore, a third-order spline function which is made by n-1 third-order polynomials, will have n-1 inflection points. The x coordinate of the inflection point (x_I) in the domain $[x_{i-1}, x_i]$ is as follows:

$$x_I = x_{i-1} - \frac{a_{i2}}{3a_{i3}} \tag{13}$$

where a_{i2} and a_{i3} are the coefficients of the third-order polynomial function calculated in the domain of $[x_{i-1}, x_i]$.

It was proved that in cases the inflection point takes place in the domain, the inflection point and the point with Average Radius of Curvature (ARC) are the same. In other words, in this case, the inflection point has the ARC value. Therefore, if the inflection point took place outside the domain, the ARC point was used instead. The ARC was an auxiliary point calculated to make a uniform distribution of initial gates over the waveform. The vectorized equation of the polynomial function (Stewart 2012) was applied to calculate the ARC location:

$$P_i(x) = a_{i0} + a_{i1}(x - x_{i-1}) + a_{i2}(x - x_{i-1})^2 + a_{i3}(x - x_{i-1})^3$$

$$\vec{\mathbf{r}}(x) = \begin{pmatrix} x \\ P_i(x) \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ a_{i0} + a_{i1}(x - x_{i-1}) + a_{i2}(x - x_{i-1})^2 + a_{i3}(x - x_{i-1})^3 \\ 0 \end{pmatrix} \tag{14}$$

The radius of curvature ($R(x)$) obtains by the following equations (Stewart 2012):

$$K(x) = \frac{\left\| \dot{\vec{\mathbf{r}}}(x) \times \ddot{\vec{\mathbf{r}}}(x) \right\|}{\left\| \dot{\vec{\mathbf{r}}}(x) \right\|^3} = \frac{\left| 2a_{i2} + 6a_{i3}(x - x_{i-1}) \right|}{\left(1 + \left(a_{i1} + 2a_{i2}(x - x_{i-1}) + 3a_{i3}(x - x_{i-1})^2 \right)^2 \right)^{\frac{3}{2}}} \tag{15}$$

$$R(x) = \frac{1}{K(x)} = \frac{\left(1 + \left(a_{i1} + 2a_{i2}(x - x_{i-1}) + 3a_{i3}(x - x_{i-1})^2 \right)^2 \right)^{\frac{3}{2}}}{\left| 2a_{i2} + 6a_{i3}(x - x_{i-1}) \right|} \tag{16}$$

where $K(x)$ is the curvature function and $R(x)$ stands for radius of curvature. By using the above equation, it was proved that the inflection point would have the maximum radius of curvature.

Proof:

$$\lim_{x\to x_{\mathrm{I}}} \mathrm{R}(x) = \lim_{x\to x_{i-1}-\frac{a_{i2}}{3a_{i3}}} \frac{\left(1+\left(a_{i1}+2a_{i2}\left(x-x_{i-1}\right)+3a_{i3}\left(x-x_{i-1}\right)^2\right)^2\right)^{\frac{3}{2}}}{\left|2a_{i2}+6a_{i3}\left(x-x_{i-1}\right)\right|} \to \infty$$

Therefore, in cases the inflection point takes place in the domain, the ARC value becomes infinite; and the inflection and the ARC points are the same.

The ARC value in the domain $\left[x_{i-1}, x_i\right]$ derived by applying the mean value theorem (Stewart 2012):

$$\overline{\mathrm{R}} = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} \mathrm{R}(x)\,dx = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} \frac{\left(1+\left(a_{i1}+2a_{i2}\left(x-x_{i-1}\right)+3a_{i3}\left(x-x_{i-1}\right)^2\right)^2\right)^{\frac{3}{2}}}{\left|2a_{i2}+6a_{i3}\left(x-x_{i-1}\right)\right|}\,dx \qquad (17)$$

By defining a new variable $t = x - x_{i-1}$, the integral can be rewritten as:

$$t = x - x_{i-1} \Rightarrow dt = dx$$

$$x_i \to x \Rightarrow t = x_i - x_{i-1} = 1$$

$$\overline{\mathrm{R}} = \int_0^1 \mathrm{f}(t)\,dt = \int_0^1 \frac{\left(1+\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)^2\right)^{\frac{3}{2}}}{\left|6a_{i3}t+2a_{i2}\right|}\,dt \qquad (18)$$

There was no analytical solution for the above integral. However, there were two methods to approximate the integral value: The first method was the numerical integrating (Young and Mohlenkamp 2021), and the second one was to approximate the $\mathrm{f}(t)$ by Mac-Lauren series up to order 2 (good approximation in this case) and then solving the integral. Here, the second method was used. The above integral approximated as follows:

$$\overline{\mathrm{R}} \approx \left| \frac{\left[1+a_{i1}{}^2\right]^{\frac{3}{2}}}{\left|2a_{i2}\right|} + \frac{3a_{i2}a_{i1}\left[1+a_{i1}{}^2\right]^{\frac{1}{2}}\left|2a_{i2}\right| - \frac{3a_{i3}a_{i2}}{\left|a_{i2}\right|}\left[1+a_{i1}{}^2\right]^{\frac{3}{2}}}{\left(2a_{i2}\right)^2} + \frac{c_0}{6} \right| \qquad (19)$$

where $c_0 = f''(0)$. A detailed description about the calculation of $f''(t)$ has been provided in Appendix C.

The x coordinate of the ARC point (x_{ARC}) in the domain $[x_{i-1}, x_i]$ is the solution of the following equation:

$$G(x) = \left(3a_{i3}(x - x_{i-1})^2 + 2a_{i2}(x - x_{i-1}) + a_{i1}\right)^2 - \left(\overline{R}\right)^{\frac{2}{3}}\left(6a_{i3}(x - x_{i-1}) + 2a_{i2}\right)^{\frac{2}{3}} + 1 = 0 \quad (20)$$

The iterative Newton-Raphson method was used to find the solution (Ruszczyński 2006). The center of the domain was considered the initial value of the iteration.

$$x_0 = \frac{x_{i-1} + x_i}{2}$$

$$G'(x) = 2\left(3a_{i3}(x - x_{i-1})^2 + 2a_{i2}(x - x_{i-1}) + a_{i1}\right)\left(6a_{i3}(x - x_{i-1}) + 2a_{i2}\right)... \\ ... - 4a_{i3}\left(\overline{R}\right)^{\frac{2}{3}}\left(6a_{i3}(x - x_{i-1}) + 2a_{i2}\right)^{-\frac{1}{3}}$$

$$x_k = x_{k-1} - \frac{G(x_{k-1})}{G'(x_{k-1})} \quad (21)$$

In order to determine the root of the function using the iterative Newton- Raphson method, a convergence criterion of 10^{-5} was used, indicating that the method stopped when the absolute difference between the current and previous approximations of the root was less than 10^{-5}. At the end of this process, the inflection and the ARC points were obtained for each domain.

If $x_{i-1} < x_I$ and $x_{ARC} < x_i$, the initial gate (g_{init}) was chosen as the closest point to the center of the domain. If x_I was outside the domain and x_{ARC} was inside, the initial gate was chosen as the ARC point, and vice versa. As a result, the first output of the spline retracking algorithm is a list of gates on the waveform. Figure 7 illustrates examples of the initial gates located on two waveforms.

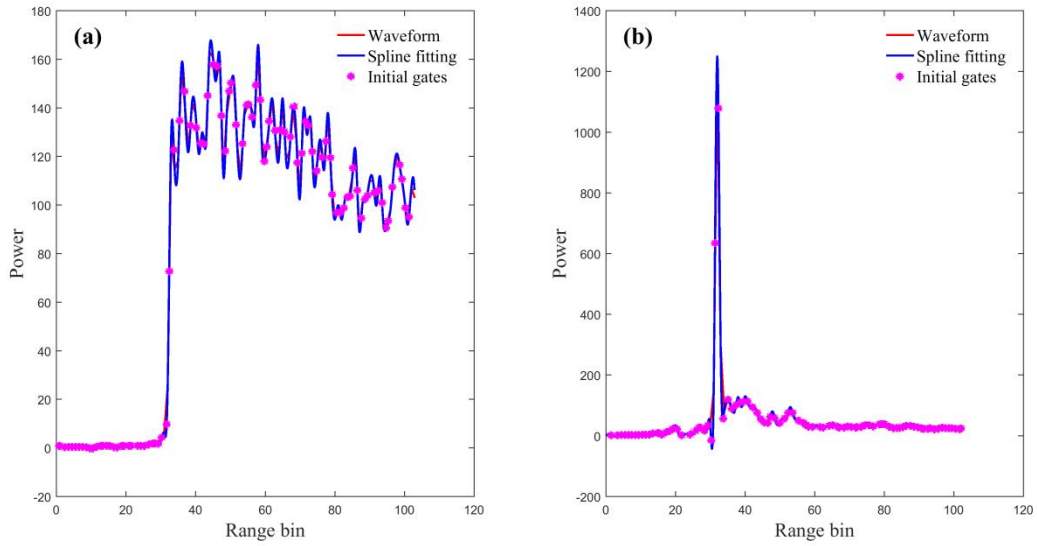


Figure 7.

The retracking gate was derived from the linear combination of the initial gates:

$$g_{ret} = \sum_{i=1}^{m} \alpha_i g_{init}^{(i)} \tag{22}$$

where α_i s are the coefficients of the initial gates and m refers to the number of initial gates, which is variable for different waveforms as the ARC value was approximated by Eq. (19), and some of the ARC points may not take place inside of the domain. The α_i s were obtained by the following equation:

$$\alpha_i = \frac{\lambda P\left(g_{init}^{(i)}\right)}{\sum_{j=0}^{N_{max}} jP_j} \tag{23}$$

Eq. (22) provides a weighted average of the initial gates where α_i s constitute the weights of averaging. Substitution of Eq. (23) in Eq. (22) results:

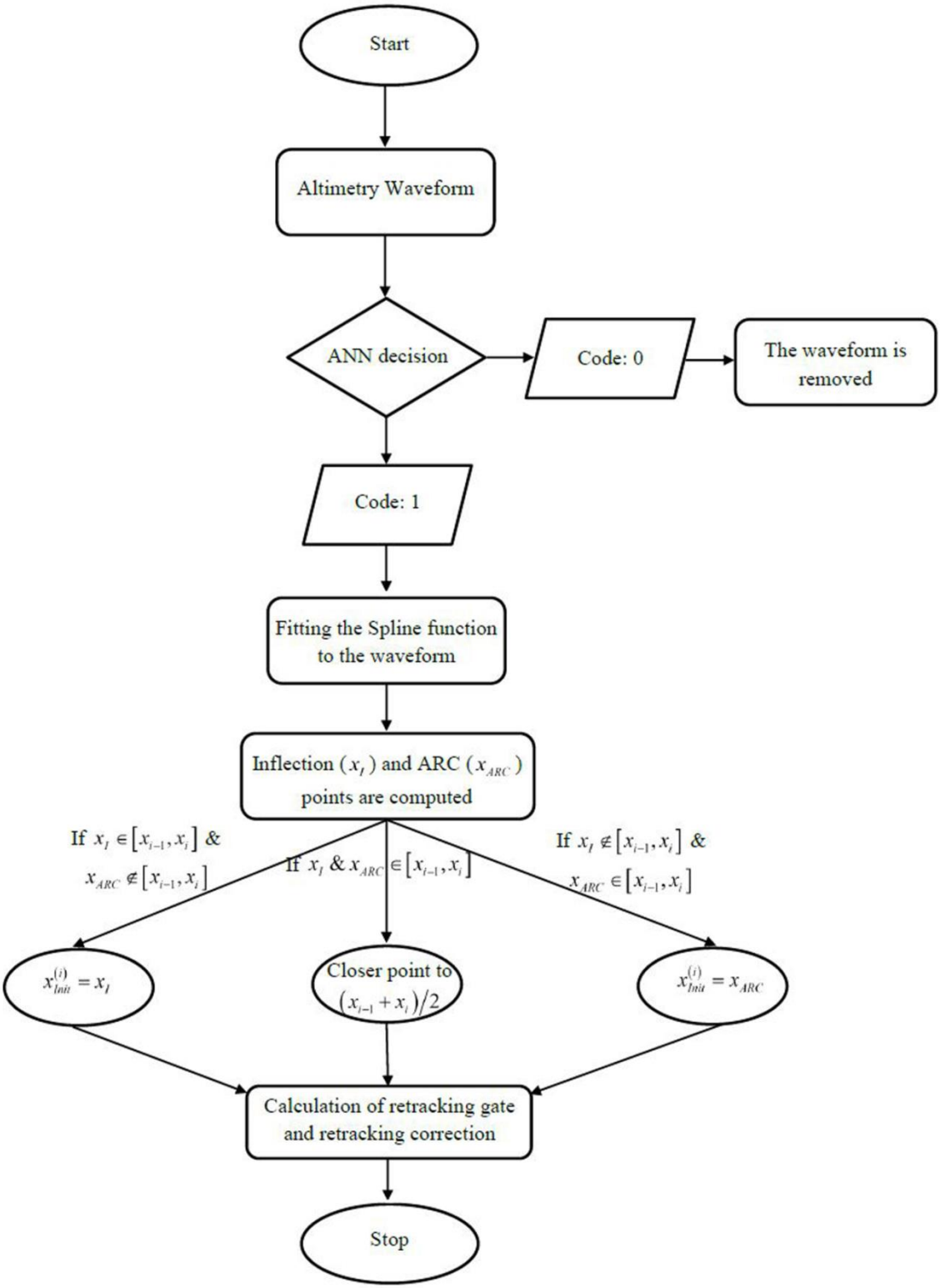
$$g_{ret} = \lambda \frac{\sum_{i=1}^{m} P\left(g_{init}^{(i)}\right) g_{init}^{(i)}}{\sum_{j=0}^{N_{max}} jP_j} \tag{24}$$

where N_{max} is the maximum gate number (103 for Jason-2&3 waveforms), $P(x)$ is the power calculated by the spline interpolation function at the initial gate, and λ is the scale factor that maintains the accuracy of results. It was proved (Appendix D) that if the α_i s were considered as the same as Eq. (23) and the initial gates had an average value of E of systematic errors, then the systematic error of the g_{ret} would be decreased remarkably.

Eq. (24) provides a multiplying scale factor with a ratio that is unique for each waveform and is related to the waveform geometry. The scale factor is a positive and constant real number determined by retracking k pure waveforms (Brownian waveforms with minimum variations on their trailing edge) from different cycles. The retracking gates of the waveforms (g_{ret}) are estimated at approximately. Then, we have:

$$\begin{aligned} \hat{\lambda}_1 &= \frac{\hat{g}_{ret}^1}{r_1} \\ \hat{\lambda}_2 &= \frac{\hat{g}_{ret}^2}{r_2} \\ &\vdots \\ \hat{\lambda}_k &= \frac{\hat{g}_{ret}^k}{r_k} \end{aligned} \Rightarrow \lambda = \frac{\sum_{i=1}^{k} \hat{\lambda}_i}{k} \tag{25}$$

where $\hat{g}_{ret}^1$, $\hat{g}_{ret}^2$,..., $\hat{g}_{ret}^k$ are the estimated retracking gates for the pure waveforms, r_1, r_2 ,... r_k refer to the unique ratios computed for the waveforms, and $\hat{\lambda}_1$, $\hat{\lambda}_2$,... $\hat{\lambda}_k$ describe the approximated scale factors. The best scale factor is derived from averaging the approximated ones. This method requires at least 10 pure waveforms ($k = 10$) distributed uniformly over a satellite track. Figure 8 summarizes the spline retracking algorithm.



٣٦٣

٣٦٤ **Figure 8.**

٣٦٥ The SSH estimation was derived as follows (Vu et al. 2018):

$$SSH = a - \left(R + \sum C_M + \sum C_G + C_{ret} \right) \quad (26)$$

$$\sum C_M = C_I + C_{DT} + C_{WT}$$

$$\sum C_G = C_{PT} + C_{SET} + C_{LT} + C_{SSB}$$

$$C_{ret} = \left(g_{ret} - g_{nom} \right) c \frac{\tau}{2}$$

where a is the satellite altitude and R defines the range between the satellite and the sea surface. $\sum C_M$ is the media correction that derives from the summation of the global Ionospheric model (C_I), and modeled Dry (C_{DT}) and Wet Tropospheric (C_{WT}) corrections. C_{PT}, C_{SET}, C_{LT}, and C_{SSB} are the geophysical effects and refer to Pole Tide, Solid Earth Tide, Load Tide, and Sea State Bias corrections, respectively. C_{ret} is retracking correction which is derived by retracking algorithms. $c = 299792458$ m/sec and $\tau = 3.125$ ns are the light velocity and the gate width, respectively. g_{nom} is the nominal gate that is 32 for Jason-2 and Jason-3 data (Uebbing, Kusche, and Forootan 2015). A usual problem in using Eq. (26) was that sometimes the NaN values occurred in the wet tropospheric delays. To solve the issue, the NaN data were interpolated by the linear interpolation method.

The SSH values are assigned to the centers of the radar footprints located at a distance of 15 km from the coastline. In distances less than 15 km, the difference between the modeled and measured ionospheric corrections is considerable for classical altimetry missions (Vu et al. 2018). Therefore, 15 km is an appropriate distance for assessing the altimetry data. Considering the ANN errors in detecting the suitable waveforms and also the possible biases in the retracking algorithms, gross errors may happen in a few measurements. In this step, the Baarda test removed the erroneous measurements (Baarda 1968). Decision-making about which

measurement should be removed depended on the standard deviation of the measurements and the residual of SSHs. The following criterion was tested to find the biased measurements (Tourian 2012):

$$-K_{\alpha/2} < \frac{e_i}{\sigma} < K_{\alpha/2};\ i=1,2,...,p \tag{27}$$

where e_i is the residual of ith SSH, σ is the standard deviation of the measurements, p is the number of waveforms selected by the ANN, $K_{\alpha/2}$ is the critical value which is 1.96, $\alpha = 0.05$, and $(1-\alpha)$ indicates a confidence level of 95%. Any measurement that exceeded the above criterion would be removed from the list of measurements, and a new standard deviation was calculated based on the new heights. This process iterated until all of the residuals followed the above condition. The altimetry height (H_i^s) was derived by averaging the remained SSHs. It should be noted that outlier detection was applied on the SSH sets of the OCOG, threshold, improved threshold, and spline algorithms. Since the ALES-SSH product was provided in 1 Hz sampling rate, the Baarda test was not used to remove the outliers from it.

3.3- Validation of results

The unbiased RMSE (ubRMSE) (Gao et al. 2019) and the Pearson Correlation Coefficient (PCC) (Evans and Rosenthal 2009) were two statistics used for assessing the altimetry time series. For this, the times of tide gauge measurements were transformed to UTC seconds then the tide gauge heights were interpolated linearly at the times of altimetry measurements to make them harmonize with the altimetry SSHs. Finally, the averages of tide gauge heights and the altimetry heights were computed in each cycle. These statistical measurements are defined as follows:

$$\text{ubRMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(H_i^s - \overline{H_i^s}\right) - \left(H_i^T - \overline{H_i^T}\right)\right]^2}{n}} \tag{28}$$

$$\text{PCC} = \frac{\sum_{i=1}^{n} H_i^s H_i^T - \frac{\sum_{i=1}^{n} H_i^s \sum_{i=1}^{n} H_i^T}{n}}{\sqrt{\left(\sum_{i=1}^{n}\left(H_i^s\right)^2 - \frac{\left(\sum_{i=1}^{n} H_i^s\right)^2}{n}\right)\left(\sum_{i=1}^{n}\left(H_i^T\right)^2 - \frac{\left(\sum_{i=1}^{n} H_i^T\right)^2}{n}\right)}} \tag{29}$$

In the above equations, n is the number of satellite passes (number of cycles) over the studied area, and H_i^s and H_i^T refer to the average of altimetry and the tide gauge heights in each cycle, respectively.

4- Results and discussion

This section is dedicated to investigating the spline retracking algorithm performance. The SSH time series are compared with the OCOG, threshold, ALES, and improved threshold methods. Table 3 shows the results derived from the three retracking algorithms for the Onsala station.

4-1- SSH time series of Onsala station

Table 3. Comparison of retracking algorithms for Onsala station. The 75 cycles were used to provide SSH time series. The waveforms were selected at a distance of 0-15 km from the coastline. The λ parameter is the scale factor that has been computed for this track specifically, in Onsala station.

Mission/Pass	Retracking algorithm	ubRMSE (m)	Correlation (%)
Jason-3/068	OCOG	1.21	40.57
	Threshold-10%	1.20	43.31
	Threshold-20%	1.20	42.51
	Threshold-50%	1.18	40.12
	Spline ($\lambda = 37.00$)	0.10	88.36

Table 3 illustrates the performance of the spline retracking algorithm. The spline results show at least a 91% reduction in ubRMSE and a more than 40% increase in correlation of the SSH time

series than the other algorithms. As mentioned before, the scale factor determines the accuracy of the results. Here, this parameter is identified by the averaging method (Eq. (25)). The threshold level is defined based on the OCOG amplitude (Wingham, Rapley, and Griffiths 1986) of 10%, 20%, and 50% (Jinyum et al. 2006). If the volume scattering is dominant in the physical environment, the threshold level is 10 or 20%. The threshold level of 50% is used in the case of surface scattering (Davis 1997; Vignudelli et al. 2011). The SSH time series of the threshold-10% shows good consistency with the OCOG and spline results. Therefore, the scattering type of the area is volume scattering.

The scatter plots of the algorithms have been depicted in the following:

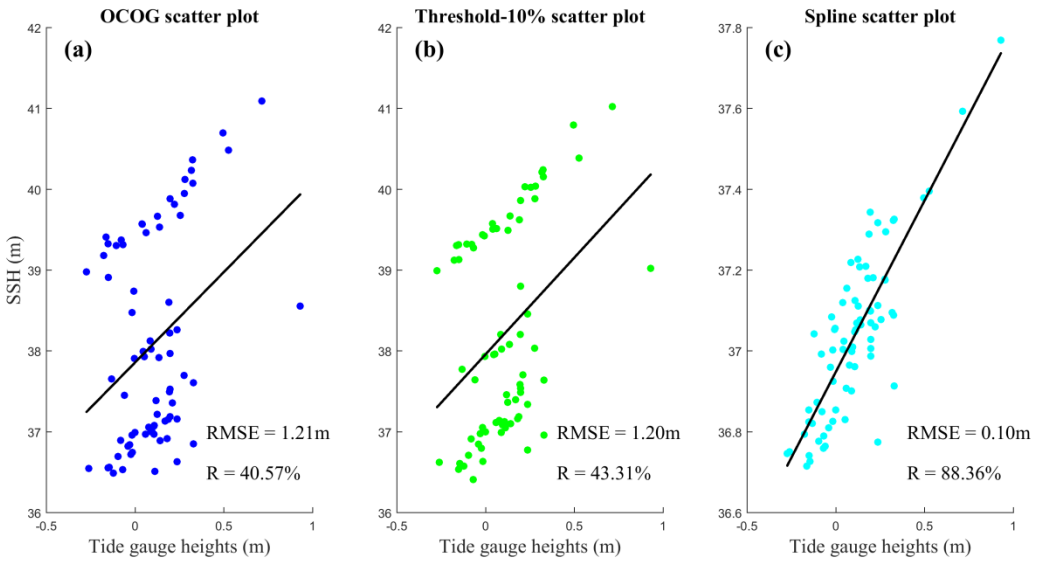


Figure 9.

4-2- SSH time series of Halmstad station

The SSH time series for Halmstad station was validated with Jason-2 and Jason-3 data. Table 4 describes the ubRMSE and correlation of the results.

Table 4. Comparison of retracking algorithms for Halmstad station. The 80 cycles of the Jason-2 and 71 cycles of the Jason-3 were used to provide the SSH time series. The waveforms were selected at a distance of 0-15 km from the coastline. As we see, the λ parameter depends not only on the location of the stations but on the satellite mission.

Mission/Pass	Retracking algorithm	ubRMSE (m)	Correlation (%)
Jason-2/213	OCOG	0.29	58.39
	Threshold-10%	0.30	59.16
	Threshold-20%	0.29	55.99
	Threshold-50%	0.29	46.57
	Spline ($\lambda = 34.5$)	0.14	72.62
Jason-3/213	OCOG	1.44	41.93
	Threshold-10%	0.86	41.63
	Threshold-20%	0.84	38.57
	Threshold-50%	0.84	35.38
	Spline ($\lambda = 33.00$)	0.14	83.53

According to Table 4, the SSH time series of the spline algorithm derived from the Jason-2 data show at least a 50% reduction in ubRMSE and a minimum increase of 13% in the correlation of the time series. This Table also shows at least an 82% reduction in ubRMSE and a more than 40% increase in correlation of the SSH time series derived from Jason-3 data by the spline retracking. According to the threshold level, the scattering type of the area is volume scattering.

4-3- SSH time series of Muscat station

The SSH time series for Muscat station were obtained from Jason-2 and Jason-3 data. Table 5 describes the ubRMSE of the results. As seen in Figure 1, the continental slope varies very sharply along the Omani coast. It is expected that the waveforms around the Muscat station will be more diverse than the waveforms around the Swedish stations. In other words, reaching high-quality results would be possible if the number of waveforms increased. This task carries out by increasing the distances to 25, 30, 35, and 40 km from the coastline.

Table 5. Comparison of retracking algorithms for Muscat station. The 73 cycles of the Jason-2 and 70 cycles of the Jason-3 were used to provide the SSH time series. The waveforms were selected at a distance of 0-25 km from the coastline.

Mission/Pass	Retracking algorithm	ubRMSE (m)	Correlation (%)
	OCOG	1.29	51.71
	Threshold-10%	0.64	71.24

Jason-2/016	Threshold-20%	0.67	70.79
	Threshold-50%	0.82	65.27
	Spline (λ = 35.00)	0.69	68.00
Jason-3/157	OCOG	2.10	42.88
	Threshold-10%	1.80	44.57
	Threshold-20%	1.75	47.63
	Threshold-50%	1.77	47.32
	Spline (λ = 38.00)	0.34	90.07

According to the Table, the diversity of waveforms in pass 016 reduces the precision of the spline algorithm. In this case, the spline results are almost equivalent to the threshold. In pass 157, we see the same results as the Swedish waters. In this case, the spline method provides the best-quality time series. The dominant scattering in the area is volume scattering. Table 6 describes the results if the distances increase to 30, 35, and 40 km from the coastline.

Table 6. Comparison of retracking algorithms for Muscat station. The waveforms were selected at distances of 30, 35, and 40 km from the coastline. The scale factor is the same for all of the distances.

Mission/Pass	Retracking algorithm	Distance to coast (km)	ubRMSE (m)	Correlation (%)
Jason-2/016	OCOG	30	1.12	56.67
	Threshold-20%	30	0.56	77.04
	Spline (λ = 35.00)	30	0.55	75.67
	OCOG	35	0.73	65.04
	Threshold-20%	35	0.50	79.85
	Spline (λ = 35.00)	35	0.42	80.94
	OCOG	40	0.34	89.38
	Threshold-20%	40	0.41	86.14
	Spline (λ = 35.00)	40	0.18	95.53
Jason-3/157	OCOG	30	1.35	61.12
	Threshold-20%	30	1.32	59.30
	Spline (λ = 38.00)	30	0.25	94.65
	OCOG	35	1.17	69.46
	Threshold-20%	35	1.22	65.01
	Spline (λ = 38.00)	35	0.23	95.18
	OCOG	40	1.09	71.55
	Threshold-20%	40	1.06	72.47
	Spline (λ = 38.00)	40	0.18	96.90

The above Table explains a new problem in the spline method. In cases where the depth varies dramatically, the geometry of the waveforms becomes more complex. Therefore, we need more data to obtain high-quality SSH time series. This issue will be clarified in the discussion section.

Figure 10 illustrates the SSH time series of different retracking algorithms at Onsala, Halmstad, and Muscat stations.

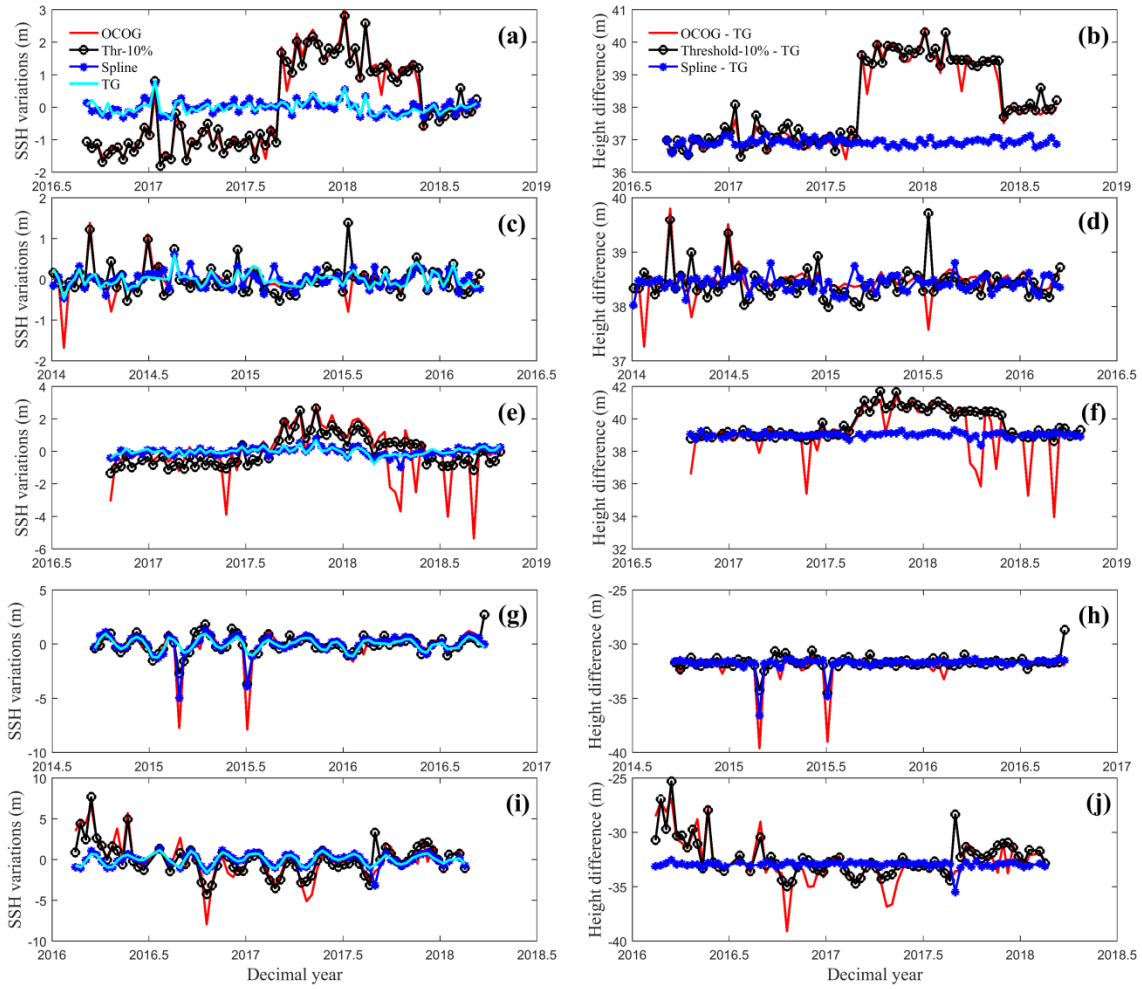


Figure 10.

4-4- Discussion

This section is devoted to the performance analysis of different retracking algorithms in providing SSH time series of the selected stations. The most important question is: "What is the reason behind decreasing the precision of the spline method for regions that are around 25 km of the coastline in Muscat station?" The previous studies revealed that every retracking algorithm

cannot retrack all kinds of waveforms in a region. The spline method is certainly not an exception, and there are limitations in its implementations, such as defining the coefficients of the combination. However, in comparison with the OCOG and threshold methods, it is found to likely provide more high-quality results because of the considerable reduction of systematic errors. In other words, it means that the spline method can retrack more diverse waveforms than the two conventional algorithms.

The classification of coastal waveforms using ANNs which was carried out by Vignudelli et al. (2011) introduced 14 types of waveforms in coastal areas. These waveforms are depicted in Figure 11.

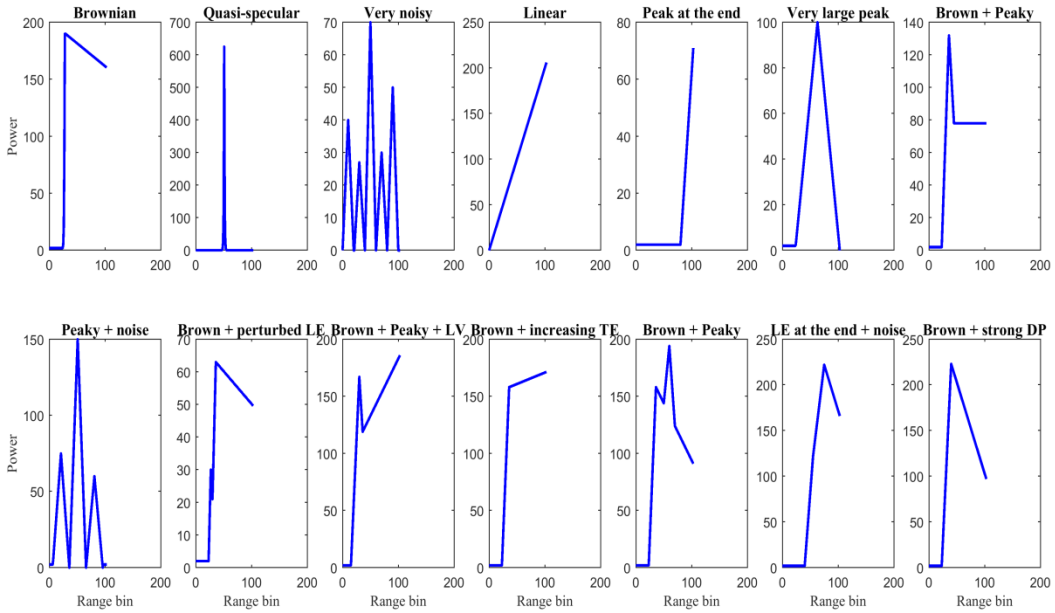


Figure 11.

According to the classification of Figure 11 and using the visual inspection, the waveforms obtained from the Onsala and Halmstad coastal areas are classified into 6 classes, which are: Brownian, quasi-specular, Brownian + peaky, peaky + noise, Brownian + strong decreasing

plateau, and Brownian + leading edge perturbation waveforms. The analysis of the SSH time series on Onsala coastal waters revealed that the spline is a successful algorithm in retracking the above waveform types, as according to Figure 9, there is a high correlation between the tide gauge heights and the spline time series of SSH. The Muscat waveforms were also classified in the same way, and the performance of different retracking algorithms was considered on the samples. In this station, there are 2 special classes of waveforms in addition to the 6 types of waveforms in Swedish waters. These classes are “very large peak” and “Brownian + increasing trailing edge” waveforms. Figure 12 illustrates the performance of different retracking algorithms on the waveforms of Swedish and Muscat coastal waters.

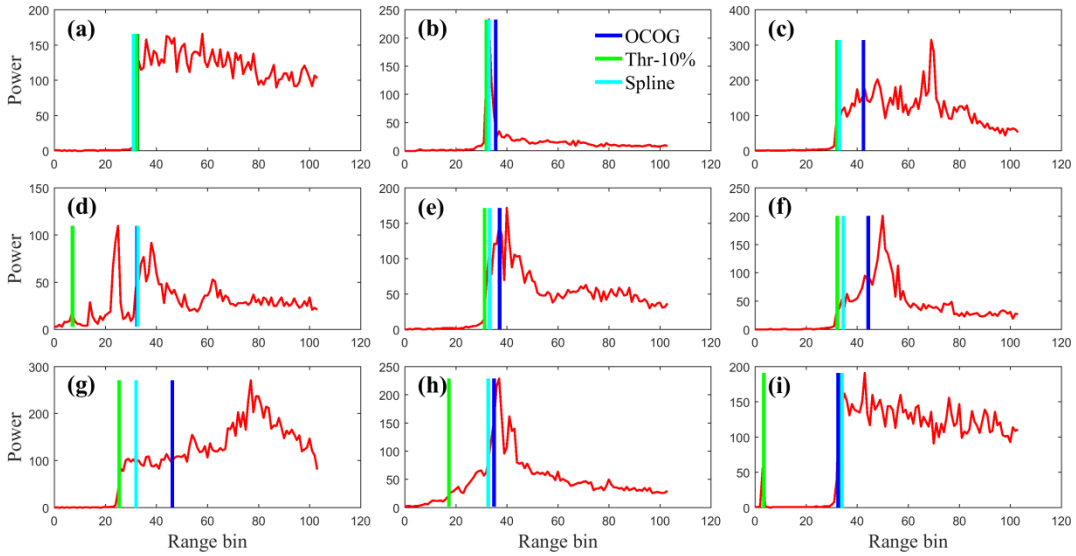


Figure 12.

Considering the Figure above, the following results are obtained:

The retracking algorithms have almost similar performance on Brownian and Brownian with strong decreasing plateau (samples (a) and (e)) waveforms. This result was predictable, as almost all of the algorithms can retrack the oceanic waveforms. In the case of quasi-specular waveform

(sample (b)), the performance depends on the regifted powers around the peak. As these powers decrease, the results become more similar. If the algorithms apply on a waveform with extra peak(s) (samples (d), (g), (h), and (i)), the spline will try to retrack the main sub-waveform, while the empirical algorithms are dependent to the location of the peak(s). Sample (i) represents a special type of Brownian waveform that does not belong to the types introduced in Figure 11. For this, it is shown at the end of the Figure. The peak in the thermal noise level makes a challenge for retracking algorithms, especially threshold retracking. This problem has been investigated before in the study carried out by Lee et al. (2008). However, the spline retracker does not consider the first sub-waveform automatically. Because the number of high-power initial gates in the first sub-waveform is smaller than the ones in the second (main) sub-waveform. Therefore, the weighted average of the initial gates will be a point on the main sub-waveform.

The main difference between the retrackers arises in retracking the waveform with rising trailing edge (sample (g)). Here, the spline represents very erroneous results. The reason for this problem returns to the definition of the coefficients of linear combination (Eq. (23)). As the initial gates on the trailing edge have higher powers, the weights are larger for these gates, and the retracking gate tends to the end gates of the waveform. If the number of rising-trailing-edge waveforms increases in a cycle (such as cycles 245 and 258 in Jason-2 data with pass number 016), it will be very likely to have systematic errors in SSH computations. As the Baarda test can not remove all of the erroneous measurements in the echogram. The more the distance to the coastline increases (Table 6), the more Brownian waveforms appear on the echogram, and the rising-trailing-edge waveforms will be put in a minority position. Therefore, the erroneous measurements are more detectable by the Baarda test.

The only constraint that someone must consider in defining the coefficients is that the remained systematic error of the retracking gate should be smaller than the systematic error of each initial gate (Eq. (D.4)). Therefore, someone can define different coefficients for the linear combination, and evaluate the precision of the algorithm in different areas. The spline results were also compared with two sub-waveform based algorithms. Table 7 shows the spline performance with the renowned ALES (Passaro et al. 2014) and improved threshold (S Roohi 2017) methods that have been developed to retrack the open ocean and coastal waveforms.

Table 7. Comparison of the ALES, improved threshold, and spline performances in the study areas. This Table has been provided separately for a better comparison between the spline and the sub-waveform based retracking algorithms. The improved threshold is assumed to retrack the first sub-waveform.

Station/Mission	Retracking algorithm	Distance to coast (km)	ubRMSE (m)	Correlation (%)
Onsala/Jason-3	ALES	15	0.46	43.21
	Imp threshold-10%	15	1.20	43.09
	Spline ($\lambda = 37.00$)	15	0.10	88.36
	Spline ($\lambda = 37.00$)	20	0.09	89.86
Halmstad/Jason-2	ALES	15	0.20	37.29
	Imp threshold-20%	15	0.30	54.39
	Spline ($\lambda = 34.5$)	15	0.14	72.62
	Spline ($\lambda = 34.5$)	20	0.13	77.37
Halmstad/Jason-3	ALES	15	0.30	50.17
	Imp threshold-10%	15	0.86	41.35
	Spline ($\lambda = 33.00$)	15	0.14	83.53
	Spline ($\lambda = 33.00$)	20	0.11	87.72
Muscat/Jason-2	ALES	25	0.16	96.10
	Imp threshold-20%	25	0.63	69.81
	Spline ($\lambda = 35.00$)	25	0.69	68.00
	Spline ($\lambda = 35.00$)	30	0.55	75.67
Muscat/Jason-3	ALES	25	0.74	56.81
	Imp threshold-20%	25	1.87	41.53
	Spline ($\lambda = 38.00$)	25	0.34	90.07
	Spline ($\lambda = 38.00$)	30	0.25	94.65

Table 7 illustrates the flexibility of the spline retracking algorithm than to the other methods. In Swedish coastal waters, the ubRMSE is reduced by at least 29% and the correlation increases by more than 30% than to the ALES method. In the Muscat station and Jason-2 track, the ALES outperforms the spline method, as the Brownian waveforms with rising trailing edges are better

retracked by the ALES algorithm. Applying the ALES algorithm in this track makes a 77% reduction in the ubRMSE and an increase of 28% in the correlation of the results than to the spline method. In the Jason-3 track, the condition is completely different and the spline shows better performance than the ALES method. The considerable reduction of systematic error that happens on the retracking gate is the reason for this (Appendix D). The improved threshold shows the poorest performance in this comparison. Comparing the ubRMSE and correlation of improved threshold with the threshold algorithm (Tables 3,4, and 5) illustrates that the results are almost the same. It means that applying the threshold algorithm on a full waveform equals to retracking the first sub-waveform which is not always a true assumption, as Table 7 shows. Figure 13 compares the SSH time series of the spline with improved threshold and ALES algorithms.

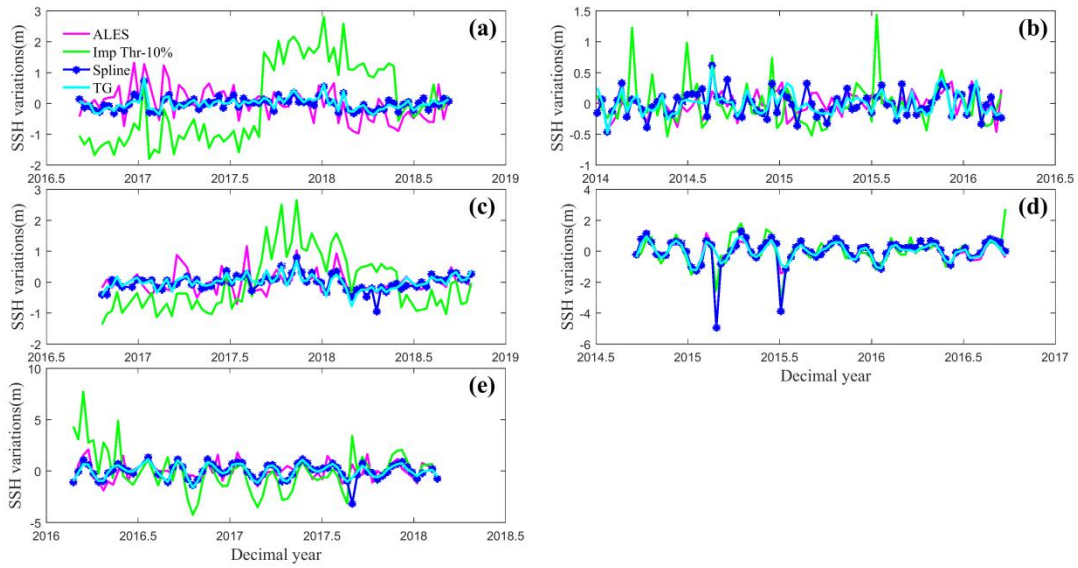


Figure 13.

As we have a considerable number of multi-peak waveforms in the satellite data, it is likely to obtain non-precise SSH time series by the empirical algorithms (As depicted in Figures 10 (a)

and 13 (a)). As a result, they need to retrack different sub-waveforms to reach a good precision. Therefore, it will be a time consuming process to get suitable results by these methods. However, the spline retracker almost considers the best sub-waveform each time.

5- Conclusions

This study represents a novel idea for retracking classical altimetry waveforms. Spline retracking was developed based on a geometrical assumption, i.e., the symmetrical reciprocal motion of the radar pulse, and the whole waveform was considered for retracking, such as OCOG and threshold algorithms. To reduce the systematic errors of the waveform (non-ocean effects) on the retracking gate, the weighted average of the initial gates was considered, and the spline results proved the error reduction. The SSHs derived from the spline algorithm around Swedish tide gauge stations showed good consistency with the tide gauge data. The ubRMSE of spline results was reduced by at least 50%, and the correlation was also increased between 13% to 40% more than the empirical algorithms for the Onsala and Halmstad stations, respectively. The results were different for Muscat station as the continental slope varies sharply in the Gulf of Oman (Pous 2004), and we had more diverse waveforms in some of the cycles. The most typical property of the spline retracker was retracking the main sub-waveform, which was concluded from the investigation of Figure 12. However, its performance was degraded in retracking the Brownian waveforms with rising trailing edges. The dominant scattering type in all three areas is volume scattering.

By comparing the performance of the spline algorithm with the ALES method, it is clear that the spline has provided better results than the ALES (Table 7), which is the most important achievement of the study. As the waveforms become more complex by approaching the coastline, the mathematical model can manage the diversity of the waveforms, which is a benefit

for the proposed algorithm. Therefore, we do not face problems related to fitting the mathematical models to the waveforms, which mostly occur in physical retrackers. Moreover, the spline retracker outperformed the improved threshold algorithm which was applied to retrack the first sub-waveform. The results revealed that retracking the first sub-waveform, which mostly happens in empirical algorithms, is not always a true assumption in altimetry processes, and in some cases (e.g., track 068 in Onsala station) makes systematic errors in the SSH computations.

In conclusion, to enhance the performance of the spline algorithm in future studies, we suggest the following ideas: As mentioned before, the scale factor is a constant parameter defined to manage the diversity of waveforms in the study area. This parameter may have very complex behavior for some missions. Therefore, it is better to apply this method to sub-waveforms. In this case, the retracking gate will be the weighted average of initial gates on the sub-waveform. Defining new weights (the coefficients of the linear combination in Eq. (22)) can be an appealing idea to evaluate the precision of the retracking algorithm. Moreover, the evaluation of the algorithm may be further investigated if the geometrical assumption of the method is altered.

Data availability statement

The SGDR files are accessible by the following link:

https://www.ncei.noaa.gov/products/jason-satellite-products

Tide gauge data are also accessible by the following links:

https://www.smhi.se/data/oceanografi/ladda-ner-oceanografiska-observationer

http://uhslc.soest.hawaii.edu/network/

The temporal resolutions of Onsala, Halmstad, and Muscat tide gauge data are 1 minute, 10 minutes, and 60 minutes, respectively.

The SSH time series of the ALES method are accessible by the following link:

https://openadb.dgfi.tum.de/en/data_access/#

6- References

Agar, Parisa, Shirzad Roohi, Behzad Voosoghi, Arash Amini, and Davod Poreh. 2023. “Sea Surface Height Estimation from Improved Modified, and Decontaminated Sub-Waveform Retracking Methods over Coastal Areas” 15 (804). https://doi.org/10.3390/rs15030804.

Akbari, P, M Sadrinasab, V Chegini, and M Siadatmousavi. 2016. “Tidal Constituents in the Persian Gulf, Gulf of Oman and Arabian Sea: A Numerical Study” 45 (August).

Amarouche, L., P. Thibaut, O. Z. Zanife, J.-P. Dumont, P. Vincent, and N. Steunou. 2004. “Improving the Jason-1 Ground Retracking to Better Account for Attitude Effects.” *Marine Geodesy* 27 (1–2): 171–97. https://doi.org/10.1080/01490410490465210.

Arabsahebi, Reza, Behzad Voosoghi, and Mohammad J. Tourian. 2018. “The Inflection-Point Retracking Algorithm: Improved Jason-2 Sea Surface Heights in the Strait of Hormuz.” *Marine Geodesy* 41 (4): 331–52. https://doi.org/10.1080/01490419.2018.1448029.

Baarda, W. 1968. *A Testing Procedure for Use in Geodetic Networks*. Netherlands Geodetic Commission. Publications on Geodesy. Netherlands Geodetic Commission. https://books.google.com/books?id=4Ho-SwAACAAJ.

Bamber, J. L. 1994. “Ice Sheet Altimeter Processing Scheme.” *International Journal of Remote Sensing* 15 (4): 925–38. https://doi.org/10.1080/01431169408954125.

Brown, G. 1977. “The Average Impulse Response of a Rough Surface and Its Applications.” *IEEE Transactions on Antennas and Propagation* 25 (1): 67–74. https://doi.org/10.1109/TAP.1977.1141536.

Chelton, D.B., J.C. Ries, B.J. Haines, L.-L Fu, and P.S. Callahan. 2001. “Satellite Altimetry.” *Satellite Altimetry and Earth Sciences* 69 (January): 1–131.

Cipollini, Paolo, Christine Gommenginger, Henrique Coelho, J. Fernandes, Jesus Gomez-Enri, Cristina Martin-Puig, S. Vignudelli, P. Woodworth, Salvatore Dinardo, and J. Benveniste. 2009. “Progress in Coastal Altimetry: The Experience of the COASTALT Project,” April.

Davis, C.H. 1997. “A Robust Threshold Retracking Algorithm for Measuring Ice-Sheet Surface Elevation Change from Satellite Radar Altimeters.” *IEEE Transactions on Geoscience and Remote Sensing* 35 (4): 974–79. https://doi.org/10.1109/36.602540.

Deng, Xiaoli, Ren-Bin Wang, Fukai Peng, Yong Yang, and Nan-Ming Mo. 2021. “Retracking Cryosat-2 Data in SARIn and LRM Modes for Plateau Lakes: A Case Study for Tibetan and Dianchi Lakes.” *Remote Sensing* 13 (6): 1078. https://doi.org/10.3390/rs13061078.

Dinardo, S. 2020. “Techniques and Applications for Satellite SAR Altimetry over Water, Land and Ice.” Ph.D Thesis, Technische Universität Darmstadt.

Evans, M.J., and J.S. Rosenthal. 2009. *Probability and Statistics: The Science of Uncertainty*. W. H. Freeman. https://books.google.com/books?id=plokAAAAQBAJ.

Fausett, Laurene V. 1994. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice-Hall.

Fu, Lee-Leung, and Anny Cazenave, eds. 2001. *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*. International Geophysics Series, v. 69. San Diego: Academic Press.

Gao, Qi, Eduard Makhoul, Maria Jose Escorihuela, Mehrez Zribi, Pere Quintana Segui, Pablo Garcia, and Monica Roca. 2019. “Analysis of Retrackers’ Performances and Water Level Retrieval over the Ebro River Basin Using Sentinel-3” 11 (6). https://doi.org/10.3390/rs11060718.

Ghilani, Charles D., and Paul R. Wolf. 2006. *Adjustment Computations: Spatial Data Analysis*. 4th ed. Hoboken, N.J: John Wiley & Sons.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Adaptive Computation and Machine Learning. Cambridge, Mass: The MIT press.

Göttl, Franziska, Denise Dettmering, Felix L. Müller, and Christian Schwatke. 2016. “Lake Level Estimation Based on CryoSat-2 SAR Altimetry and Multi-Looked Waveform Classification” 8 (885). https://doi.org/10.3390/rs8110885.

Halimi, Abderrahim. 2013. “From Conventional to Delay/Doppler Altimetry.” PhD Thesis, University of Toulouse.

Halimi, Abderrahim, Corinne Mailhes, Jean-Yves Tourneret, Pierre Thibaut, and Francois Boy. 2013. “Parameter Estimation for Peaky Altimetric Waveforms.” *IEEE Transactions on Geoscience and Remote Sensing* 51 (3): 1568–77. https://doi.org/10.1109/TGRS.2012.2205697.

Hayne, G. 1980. “Radar Altimeter Mean Return Waveforms from Near-Normal-Incidence Ocean Surface Scattering.” *IEEE Transactions on Antennas and Propagation* 28 (5): 687–92. https://doi.org/10.1109/TAP.1980.1142398.

Hell, Benjamin, Barry Broman, Lars Jakobsson, Martin Jakobsson, Åke Magnusson, and Pat Wiberg. 2012. “The Use of Bathymetric Data in Society and Science: A Review from the Baltic Sea.” *AMBIO* 41: 138–50.

Idris, N. H., and X. Deng. 2012. “The Retracking Technique on Multi-Peak and Quasi-Specular Waveforms for Jason-1 and Jason-2 Missions near the Coast.” *Marine Geodesy* 35 (sup1): 217–37. https://doi.org/10.1080/01490419.2012.718679.

Idris, Nurul Hazrina, Stefano Vignudelli, and Xiaoli Deng. 2020. “Analysis of Sentinel-3A Synthetic Aperture Radar (SAR) Altimetry Waveforms Over the Southeast Asia Region.” In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 5843–46. Waikoloa, HI, USA: IEEE. https://doi.org/10.1109/IGARSS39084.2020.9324276.

Jinyum, Guo, Hwang Cheiway, Chang Xiaotao, and Liu Yuting. 2006. “Improved Threshold Retracker for Satellite Altimeter Waveform Retracking over Coastal Sea*.” *Progress in Natural Science* 16 (7): 732–38. https://doi.org/10.1080/10020070612330061.

Khaki, M., E. Forootan, and M.A. Sharifi. 2014. “Satellite Radar Altimetry Waveform Retracking over the Caspian Sea.” *International Journal of Remote Sensing* 35 (17): 6329–56. https://doi.org/10.1080/01431161.2014.951741.

Lee, Hyongki, C.K. Shum, Yuchan Yi, Alexander Braun, and Chung-Yen Kuo. 2008. “Laurentia Crustal Motion Observed Using TOPEX/POSEIDON Radar Altimetry over Land.” *Journal of Geodynamics* 46 (3–5): 182–93. https://doi.org/10.1016/j.jog.2008.05.001.

Legrésy, Benoît, and Frédérique Rémy. 1997. “Altimetric Observations of Surface Characteristics of the Antarctic Ice Sheet.” *Journal of Glaciology* 43 (144): 265–75. https://doi.org/10.3189/S002214300000321X.

Lindholm, Andreas, Niklas Wahlström, Fredrik Lindsten, and Thomas B Schön. 2019. *Supervised Machine Learning (Lecture Notes for the Statistical Machine Learning Course)*. Uppsala University.

Martin, Thomas V., H. Jay Zwally, Anita C. Brenner, and Robert A. Bindschadler. 1983. "Analysis and Retracking of Continental Ice Sheet Radar Altimeter Waveforms." *Journal of Geophysical Research* 88 (C3): 1608. https://doi.org/10.1029/JC088iC03p01608.

Mattes, Dennis Frederic. 2019. "Analysis of Waveforms in the Satellite Altimetry by Using Neural Networks." https://doi.org/10.18419/OPUS-10323.

Medvedev, Igor P., Alexander B. Rabinovich, and Evgueni A. Kulikov. 2016. "Tides in Three Enclosed Basins: The Baltic, Black, and Caspian Seas." *Frontiers in Marine Science* 3. https://doi.org/10.3389/fmars.2016.00046.

Mikhail, Edward M., and Friedrich E. Ackermann. 1982. *Observations and Least Squares*. Washington, D.C: University Press of America.

Million, Elizabeth. 2007. "The Hadamard Product."

Nielsen, Karina, Elena Zakharova, Angelica Tarpanelli, Ole B. Andersen, and Jérôme Benveniste. 2022. "River Levels from Multi Mission Altimetry, a Statistical Approach." *Remote Sensing of Environment* 270 (March): 112876. https://doi.org/10.1016/j.rse.2021.112876.

Passaro, Marcello, Paolo Cipollini, Stefano Vignudelli, Graham D. Quartly, and Helen M. Snaith. 2014. "ALES: A Multi-Mission Adaptive Subwaveform Retracker for Coastal and Open Ocean Altimetry." *Remote Sensing of Environment* 145 (April): 173–89. https://doi.org/10.1016/j.rse.2014.02.008.

Passaro, Marcello, Laura Rautiainen, Denise Dettmering, Marco Restano, Michael G. Hart-Davis, Florian Schlembach, Jani Särkkä, Felix L. Müller, Christian Schwatke, and Jérôme Benveniste. 2022. "Validation of an Empirical Subwaveform Retracking Strategy for SAR Altimetry." *Remote Sensing* 14 (16): 4122. https://doi.org/10.3390/rs14164122.

Passaro, Marcello, Stine Kildegaard Rose, Ole B. Andersen, Eva Boergens, Francisco M. Calafat, Denise Dettmering, and Jérôme Benveniste. 2018. "ALES+: Adapting a Homogenous Ocean Retracker for Satellite Altimetry to Sea Ice Leads, Coastal and Inland Waters." *Remote Sensing of Environment* 211 (June): 456–71. https://doi.org/10.1016/j.rse.2018.02.074.

Pous, S. P. 2004. "Hydrology and Circulation in the Strait of Hormuz and the Gulf of Oman—Results from the GOGP99 Experiment: 2. Gulf of Oman." *Journal of Geophysical Research* 109 (C12): C12038. https://doi.org/10.1029/2003JC002146.

Pranzini, Enzo, and A. T. Williams, eds. 2013. *Coastal Erosion and Protection in Europe*. 1st ed. London ; New York, NY: Routledge.

Quartly, Graham D. 1998. "Determination of Oceanic Rain Rate and Rain Cell Structure from Altimeter Waveform Data. Part I: Theory." *Journal of Atmospheric and Oceanic Technology* 15 (6): 1361–78. https://doi.org/10.1175/1520-0426(1998)015<1361:DOORRA>2.0.CO;2.

Raney, R.K. 1998. "The Delay/Doppler Radar Altimeter." *IEEE Transactions on Geoscience and Remote Sensing* 36 (5): 1578–88. https://doi.org/10.1109/36.718861.

Roohi, S. 2017. "Performance Evaluation of Different Satellite Radar Altimetry Missions for Monitoring Inland Water Bodies." Ph.D Thesis, Universität Stuttgart.

Roohi, Sh., N. Sneeuw, J. Benveniste, S. Dinardo, E.A. Issawy, and G. Zhang. 2021. "Evaluation of CryoSat-2 Water Level Derived from Different Retracking Scenarios over Selected

Inland Water Bodies." *Advances in Space Research* 68 (2): 947–62. https://doi.org/10.1016/j.asr.2019.06.024.
Ruszczyński, Andrzej P. 2006. *Nonlinear Optimization*. Princeton, N.J: Princeton University Press.
Sailor, Richard V., and A. Richard Leschack. 1987. "Preliminary Determination of the Geosat Radar Altimeter Noise Spectrum." *Johns Hopkins APL Technical Digest* 8 (June): 182.
Shen, Xiaoyi, Markku Similä, Wolfgang Dierking, Xi Zhang, Changqing Ke, Meijie Liu, and Manman Wang. 2019. "A New Retracking Algorithm for Retrieving Sea Ice Freeboard from CryoSat-2 Radar Altimeter Data during Winter–Spring Transition." *Remote Sensing* 11 (10): 1194. https://doi.org/10.3390/rs11101194.
Stewart, James. 2012. *Calculus: Early Transcendentals*. 7th ed. Belmont, CA: Brooks/Cole, Cengage Learning.
Tapley, BD, GH Born, and ME Parke. 1982. "The SEASAT Altimeter Data and Its Accuracy Assessment" 87: 3179–88.
Tourian, M.J. 2012. *Controls on Satellite Altimetry Over Inland Water Surfaces for Hydrological Purposes*. Universitätsbibliothek der Universität Stuttgart. https://books.google.com/books?id=SsJyuwEACAAJ.
Tournadre, Jean. 2007. "Signature of Lighthouses, Ships, and Small Islands in Altimeter Waveforms." *JOURNAL OF ATMOSPHERIC AND OCEANIC TECHNOLOGY* 24 (June): 1143–49. https://doi.org/10.1175/JTECH2030.1.
Uebbing, Bernd, Jurgen Kusche, and Ehsan Forootan. 2015. "Waveform Retracking for Improving Level Estimations From TOPEX/Poseidon, Jason-1, and Jason-2 Altimetry Observations Over African Lakes." *IEEE Transactions on Geoscience and Remote Sensing* 53 (4): 2211–24. https://doi.org/10.1109/TGRS.2014.2357893.
Vignudelli, Stefano, Andrey G. Kostianoy, Paolo Cipollini, and Jérôme Benveniste, eds. 2011. *Coastal Altimetry*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12796-0.
Villadsen, Heidi, Xiaoli Deng, Ole B. Andersen, Lars Stenseng, Karina Nielsen, and Per Knudsen. 2016. "Improved Inland Water Levels from SAR Altimetry Using Novel Empirical and Physical Retrackers." *Journal of Hydrology* 537 (June): 234–47. https://doi.org/10.1016/j.jhydrol.2016.03.051.
Vu, Phuong, Frédéric Frappart, José Darrozes, Vincent Marieu, Fabien Blarel, Guillaume Ramillien, Pascal Bonnefond, and Florence Birol. 2018. "Multi-Satellite Altimeter Validation along the French Atlantic Coast in the Southern Bay of Biscay from ERS-2 to SARAL." *Remote Sensing* 10 (2): 93. https://doi.org/10.3390/rs10010093.
Wang, Haihong, and Zhengkai Huang. 2021. "Waveform Decontamination for Improving Satellite Radar Altimeter Data Over Nearshore Area: Upgraded Algorithm and Validation." *Frontiers in Earth Science* 9 (September): 748401. https://doi.org/10.3389/feart.2021.748401.
Wingham, DJ, CG Rapley, and H Griffiths. 1986. "New Techniques in Satellite Tracking System." In , 1339–44.
Young, Todd, and Martin J Mohlenkamp. 2021. *Introduction to Numerical Methods and MATLAB Programming for Engineers*. Ohio University.

Appendix A

Sigmoid function

Sigmoid (Eq. (A.1)) is a continuous and differentiable function that stretches between 0 and 1 (Fausett 1994; Mattes 2019) and is suitable for producing Boolean codes.

$$S(x)=\frac{1}{1+e^{-x}} \tag{A.1}$$

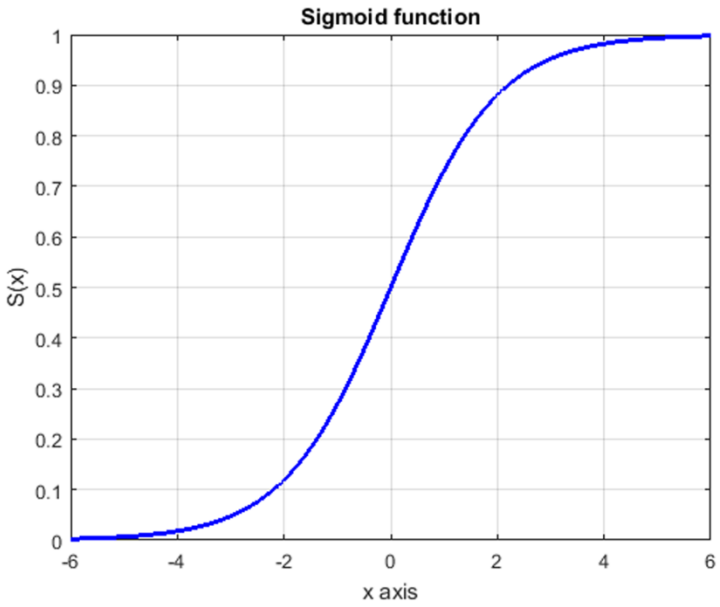


Figure A.1. The sigmoid activation function. This function approaches to 1 for large positive x and 0 for large negative values of x.

The first derivative of the sigmoid function is calculated as follows:

$$S'(x)=\frac{1}{1+e^{-x}}\left(1-\frac{1}{1+e^{-x}}\right)=S(x)\left(1-S(x)\right) \tag{A.2}$$

We used the $S'(x)$ in back-propagation algorithm in Eq. (6) and (7).

Appendix B

Spline function

The general form of a third-order spline function is (Young and Mohlenkamp 2021):

$$S_{3,n}(x)=\begin{cases} P_1(x)=a_{10}+a_{11}(x-x_0)+a_{12}(x-x_0)^2+a_{13}(x-x_0)^3,\ x\in[x_0,x_1] \\ P_2(x)=a_{20}+a_{21}(x-x_1)+a_{22}(x-x_1)^2+a_{23}(x-x_1)^3,\ x\in[x_1,x_2] \\ \vdots \\ P_n(x)=a_{n0}+a_{n1}(x-x_{n-1})+a_{n2}(x-x_{n-1})^2+a_{n3}(x-x_{n-1})^3,\ x\in[x_{n-1},x_n] \end{cases} \quad \text{(B.1)}$$

In Eq. (B.1), n points are given for each waveform. Therefore, one needs to build n-1 third-order polynomial functions, where each has four (a_{i0}, a_{i1}, a_{i2}, and a_{i3}) unknown coefficients, that in total 4*(n-1) unknown coefficients to be estimated, which is more than the number of known points. The following conditions are considered to obtain a unique third-order spline function:

$$\begin{cases} P_i(x_i)=f_i,\ i=0,1,...,n \\ P_i(x_i)=P_{i+1}(x_i),\ i=1,2,...,n\text{-}1 \\ P_i'(x_i)=P_{i+1}'(x_i),\ i=1,2,...,n\text{-}1 \\ P_i''(x_i)=P_{i+1}''(x_i),\ i=1,2,...,n\text{-}1 \\ S_{3,n}'(x_0)=f_0' \\ S_{3,n}'(x_n)=f_n' \end{cases} \quad \text{(B.2)}$$

In the above equations set, the first constraint refers to the fact that the spline function should pass all the data points. The second constraint keeps the continuity of the function in the data points. The third and fourth equations ensure the first and second derivatives in the data points. The last two equations illustrate that the slope of the function in the first and last data points should be equal to the slopes defined by the discrete data.

Appendix C

Calculation of Average Radius of Curvature (ARC)

As mentioned in the article, the ARC point is an auxiliary point calculated to make a uniform distribution of initial gates over the waveform. This coordinate derives from the vectorized equation of the polynomial function:

$$P_i(x) = a_{i0} + a_{i1}(x - x_{i-1}) + a_{i2}(x - x_{i-1})^2 + a_{i3}(x - x_{i-1})^3 \quad \text{(C.1)}$$

$$\vec{\mathbf{r}}(x) = \begin{pmatrix} x \\ P_i(x) \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ a_{i0} + a_{i1}(x - x_{i-1}) + a_{i2}(x - x_{i-1})^2 + a_{i3}(x - x_{i-1})^3 \\ 0 \end{pmatrix} \quad \text{(C.2)}$$

The first and second derivatives of $\vec{\mathbf{r}}(x)$ are:

$$\dot{\vec{\mathbf{r}}}(x) = \begin{pmatrix} 1 \\ a_{i1} + 2a_{i2}(x - x_{i-1}) + 3a_{i3}(x - x_{i-1})^2 \\ 0 \end{pmatrix} \quad \text{(C.3)}$$

$$\ddot{\vec{\mathbf{r}}}(x) = \begin{pmatrix} 0 \\ 2a_{i2} + 6a_{i3}(x - x_{i-1}) \\ 0 \end{pmatrix} \quad \text{(C.4)}$$

The radius of curvature obtains by the following equation (Stewart 2012):

$$K(x) = \frac{\left\| \dot{\vec{\mathbf{r}}}(x) \times \ddot{\vec{\mathbf{r}}}(x) \right\|}{\left\| \dot{\vec{\mathbf{r}}}(x) \right\|^3} = \frac{\left| 2a_{i2} + 6a_{i3}(x - x_{i-1}) \right|}{\left(1 + \left(a_{i1} + 2a_{i2}(x - x_{i-1}) + 3a_{i3}(x - x_{i-1})^2 \right)^2 \right)^{\frac{3}{2}}} \quad \text{(C.5)}$$

$$R(x) = \frac{1}{K(x)} = \frac{\left(1 + \left(a_{i1} + 2a_{i2}(x - x_{i-1}) + 3a_{i3}(x - x_{i-1})^2 \right)^2 \right)^{\frac{3}{2}}}{\left| 2a_{i2} + 6a_{i3}(x - x_{i-1}) \right|} \quad \text{(C.6)}$$

The ARC value in the domain $[x_{i-1}, x_i]$ derives by applying the mean value theorem (Stewart 2012):

$$\overline{R} = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} R(x)\,dx = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} \frac{\left(1 + \left(a_{i1} + 2a_{i2}(x - x_{i-1}) + 3a_{i3}(x - x_{i-1})^2 \right)^2 \right)^{\frac{3}{2}}}{\left| 2a_{i2} + 6a_{i3}(x - x_{i-1}) \right|} dx \quad \text{(C.7)}$$

By defining a new variable, $t = x - x_{i-1}$, the integral rewritten as:

$$t = x - x_{i-1} \Rightarrow dt = dx$$
$$x_i \to x \Rightarrow t = x_i - x_{i-1} = 1 \tag{C.8}$$

$$\overline{\mathrm{R}} = \int_0^1 \mathrm{f}(t)\,dt = \int_0^1 \frac{\left(1+\left(3\mathrm{a}_{i3}t^2 + 2\mathrm{a}_{i2}t + \mathrm{a}_{i1}\right)^2\right)^{\frac{3}{2}}}{\left|6\mathrm{a}_{i3}t + 2\mathrm{a}_{i2}\right|}\,dt \tag{C.9}$$

There is no analytical solution for the above integral. Therefore, the integral is approximated by the Mac-Lauren approximation of $\mathrm{f}(t)$ function. The Mac-Lauren series of the function up to order 2 obtains as follows:

$$\mathrm{f}(t) \approx \mathrm{f}(0) + \frac{\mathrm{f}'(0)}{1!}t + \frac{\mathrm{f}''(0)}{2!}t^2 \tag{C.10}$$

$$\mathrm{f}(t) = \frac{\left(1+\left(3\mathrm{a}_{i3}t^2 + 2\mathrm{a}_{i2}t + \mathrm{a}_{i1}\right)^2\right)^{\frac{3}{2}}}{\left|6\mathrm{a}_{i3}t + 2\mathrm{a}_{i2}\right|} \Rightarrow \mathrm{f}(0) = \frac{\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}}}{\left|2\mathrm{a}_{i2}\right|} \tag{C.11}$$

To calculate the first derivative of $\mathrm{f}(t)$, the function is considered as a division of two $\mathrm{u}(t)$ and $\mathrm{v}(t)$ functions.

$$\mathrm{f}(t) = \frac{\mathrm{u}(t)}{\mathrm{v}(t)} \Rightarrow \mathrm{f}'(t) = \frac{\mathrm{u}'(t)\mathrm{v}(t) - \mathrm{v}'(t)\mathrm{u}(t)}{\mathrm{v}^2(t)} \tag{C.12}$$

$$\mathrm{u}(t) = \left[1+\left(3\mathrm{a}_{i3}t^2 + 2\mathrm{a}_{i2}t + \mathrm{a}_{i1}\right)^2\right]^{\frac{3}{2}} \Rightarrow \mathrm{u}(0) = \left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}} \tag{C.13}$$

$$\mathrm{u}'(t) = \frac{3}{2}\left[1+\left(3\mathrm{a}_{i3}t^2 + 2\mathrm{a}_{i2}t + \mathrm{a}_{i1}\right)^2\right]^{\frac{1}{2}} \times 2\left(3\mathrm{a}_{i3}t^2 + 2\mathrm{a}_{i2}t + \mathrm{a}_{i1}\right) \times \left(6\mathrm{a}_{i3}t + 2\mathrm{a}_{i2}\right) \Rightarrow$$
$$\Rightarrow \mathrm{u}'(0) = 6\mathrm{a}_{i1}\mathrm{a}_{i2}\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{1}{2}} \tag{C.14}$$

$$\mathrm{v}(t) = \left|6\mathrm{a}_{i3}t + 2\mathrm{a}_{i2}\right| = \sqrt{\left(6\mathrm{a}_{i3}t + 2\mathrm{a}_{i2}\right)^2} \Rightarrow \mathrm{v}(0) = \left|2\mathrm{a}_{i2}\right| \tag{C.15}$$

$$v'(t)=\frac{2\left(6a_{i3}t+2a_{i2}\right)\times 6a_{i3}}{2\sqrt{\left(6a_{i3}t+2a_{i2}\right)^2}}=\frac{6a_{i3}\left(6a_{i3}t+2a_{i2}\right)}{\left|6a_{i3}t+2a_{i2}\right|}\Rightarrow v'(0)=\frac{6a_{i2}a_{i3}}{\left|a_{i2}\right|} \tag{C.16}$$

$$f'(t)=\frac{3\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)\left[1+\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)^2\right]^{\frac{1}{2}}\sqrt{\left(6a_{i3}t+2a_{i2}\right)^2}}{\left(6a_{i3}t+2a_{i2}\right)}-\ldots$$
$$-\frac{6a_{i3}\left[1+\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)^2\right]^{\frac{3}{2}}}{\left(6a_{i3}t+2a_{i2}\right)\sqrt{\left(6a_{i3}t+2a_{i2}\right)^2}} \tag{C.17}$$

$$f'(0)=\frac{u'(0)v(0)-v'(0)u(0)}{v^2(0)}=\frac{6a_{i1}a_{i2}\left[1+a_{i1}{}^2\right]^{\frac{1}{2}}\left|2a_{i2}\right|-\frac{6a_{i2}a_{i3}}{\left|a_{i2}\right|}\left[1+a_{i1}{}^2\right]^{\frac{3}{2}}}{\left(2a_{i2}\right)^2} \tag{C.18}$$

Here, the first derivative is represented by $U(t)$, $V(t)$, $W(t)$, $Z(t)$ and $K(t)$ functions, respectively.

$$f'(t)=\frac{3UVW}{Z}(t)-6a_{i3}\frac{K}{ZW}(t) \tag{C.19}$$

$$f''(t)=3\left[\frac{(UVW)'Z-Z'(UVW)}{Z^2}(t)\right]-6a_{i3}\left[\frac{K'ZW-(ZW)'K}{(ZW)^2}(t)\right]$$

$$\Rightarrow f''(t)=3\left[\frac{U'VWZ+UV'WZ+UVW'Z-UVWZ'}{Z^2}(t)\right]\ldots$$
$$\ldots-6a_{i3}\left[\frac{WZK'-Z'WK-ZW'K}{(ZW)^2}(t)\right] \tag{C.20}$$

$$U(t)=3a_{i3}t^2+2a_{i2}t+a_{i1} \tag{C.21}$$

$$U'(t)=6a_{i3}t+2a_{i2} \tag{C.22}$$

$$U(0)=a_{i1},\ U'(0)=2a_{i2} \tag{C.23}$$

$$V(t)=\left[1+\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)^2\right]^{\frac{1}{2}} \tag{C.24}$$

$$V'(t)=\frac{1}{2}\left[1+\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)^2\right]^{-\frac{1}{2}}\times 2\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)\times\left(6a_{i3}t+2a_{i2}\right) \tag{C.25}$$

$$V(0)=\left[1+a_{i1}{}^2\right]^{\frac{1}{2}},\ V'(0)=\frac{2a_{i1}a_{i2}}{\sqrt{1+a_{i1}{}^2}} \tag{C.26}$$

$$W(t)=\sqrt{\left(6a_{i3}t+2a_{i2}\right)^2} \tag{C.27}$$

$$W'(t)=\frac{6a_{i3}\left(6a_{i3}t+2a_{i2}\right)}{\left|6a_{i3}t+2a_{i2}\right|} \tag{C.28}$$

$$W(0)=\left|2a_{i2}\right|,\ W'(0)=\frac{6a_{i2}a_{i3}}{\left|a_{i2}\right|} \tag{C.29}$$

$$Z(t)=6a_{i3}t+2a_{i2} \tag{C.30}$$

$$Z'(t)=6a_{i3} \tag{C.31}$$

$$Z(0)=2a_{i2},\ Z'(0)=6a_{i3} \tag{C.32}$$

$$K(t)=\left[1+\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)^2\right]^{\frac{3}{2}} \tag{C.33}$$

$$K'(t)=\frac{3}{2}\left[1+\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)^2\right]^{\frac{1}{2}}\times 2\left(3a_{i3}t^2+2a_{i2}t+a_{i1}\right)\times\left(6a_{i3}t+2a_{i2}\right) \tag{C.34}$$

$$K(0)=\left[1+a_{i1}{}^2\right]^{\frac{3}{2}},\ K'(0)=6a_{i1}a_{i2}\left[1+a_{i1}{}^2\right]^{\frac{1}{2}} \tag{C.35}$$

Therefore, the second-order approximation of the $f(t)$ is as follows:

$$\mathrm{f}(t) \approx \mathrm{f}(0) + \frac{\mathrm{f}'(0)}{1!}t + \frac{\mathrm{f}''(0)}{2!}t^2 =$$

$$= \frac{\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}}}{\left|2\mathrm{a}_{i2}\right|} + \frac{6\mathrm{a}_{i1}\mathrm{a}_{i2}\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{1}{2}}\left|2\mathrm{a}_{i2}\right| - \frac{6\mathrm{a}_{i2}\mathrm{a}_{i3}}{\left|\mathrm{a}_{i2}\right|}\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}}}{\left(2\mathrm{a}_{i2}\right)^2}t + \frac{\mathrm{C}_0}{2}t^2 \qquad \text{(C.36)}$$

The Average Radius of Curvature (ARC) equals:

$$\overline{\mathrm{R}} = \int_0^1 \mathrm{f}(t)\,dt \approx \frac{\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}}}{\left|2\mathrm{a}_{i2}\right|}t\Big|_{t=0}^{t=1} + \ldots$$

$$+ \frac{6\mathrm{a}_{i1}\mathrm{a}_{i2}\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{1}{2}}\left|2\mathrm{a}_{i2}\right| - \frac{6\mathrm{a}_{i2}\mathrm{a}_{i3}}{\left|\mathrm{a}_{i2}\right|}\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}}}{\left(2\mathrm{a}_{i2}\right)^2} \times \frac{1}{2}t^2\Big|_{t=0}^{t=1} + \frac{\mathrm{C}_0}{2} \times \frac{1}{3}t^3\Big|_{t=0}^{t=1} \qquad \text{(C.37)}$$

$$\Rightarrow \overline{\mathrm{R}} \approx \left| \frac{\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}}}{\left|2\mathrm{a}_{i2}\right|} + \frac{3\mathrm{a}_{i1}\mathrm{a}_{i2}\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{1}{2}}\left|2\mathrm{a}_{i2}\right| - \frac{3\mathrm{a}_{i2}\mathrm{a}_{i3}}{\left|\mathrm{a}_{i2}\right|}\left[1+\mathrm{a}_{i1}{}^2\right]^{\frac{3}{2}}}{\left(2\mathrm{a}_{i2}\right)^2} + \frac{\mathrm{C}_0}{6} \right| \qquad \text{(C.38)}$$

where $\mathrm{C}_0 = \mathrm{f}''(0)$. The absolute value of the above integral is considered for the ARC, as the radius of curvature always has a positive value.

Appendix D

Estimation of systematic error on the retracking gate

According to the law of propagation of systematic errors (Mikhail and Ackermann 1982), if $y = \mathrm{F}(x_1, x_2, ..., x_n)$ and x_1, x_2, ..., x_n have the systematic errors of Δx_1, Δx_2, ..., Δx_n, then the systematic error of y is derived by the following equation:

$$\Delta y = \frac{\partial \mathrm{F}}{\partial x_1}\Delta x_1 + \frac{\partial \mathrm{F}}{\partial x_2}\Delta x_2 + ... + \frac{\partial \mathrm{F}}{\partial x_n}\Delta x_n \qquad \text{(D.1)}$$

Eq. (22) describes a functional relationship between $\mathrm{g_{ret}}$ and the initial gates of the waveform. Therefore, by considering the systematic error of the initial gates (which is initiated from the non-ocean effects) the systematic error of $\mathrm{g_{ret}}$ is:

$$\Delta \mathrm{g_{ret}} = \frac{\partial \mathrm{g_{ret}}}{\partial \mathrm{g}_{\mathrm{init}}^{(1)}}\Delta \mathrm{g}_{\mathrm{init}}^{(1)} + \frac{\partial \mathrm{g_{ret}}}{\partial \mathrm{g}_{\mathrm{init}}^{(2)}}\Delta \mathrm{g}_{\mathrm{init}}^{(2)} + ... + \frac{\partial \mathrm{g_{ret}}}{\partial \mathrm{g}_{\mathrm{init}}^{(\mathrm{m})}}\Delta \mathrm{g}_{\mathrm{init}}^{(\mathrm{m})} \qquad \text{(D.2)}$$

$$\Rightarrow \Delta \mathrm{g_{ret}} = \alpha_1 \Delta \mathrm{g}_{\mathrm{init}}^{(1)} + \alpha_2 \Delta \mathrm{g}_{\mathrm{init}}^{(2)} + ... + \alpha_{\mathrm{m}} \Delta \mathrm{g}_{\mathrm{init}}^{(\mathrm{m})} \qquad \text{(D.3)}$$

Let's suppose $\Delta \mathrm{g}_{\mathrm{init}}^{(1)} = \Delta \mathrm{g}_{\mathrm{init}}^{(2)} = ... = \Delta \mathrm{g}_{\mathrm{init}}^{(\mathrm{m})} = \mathrm{E}$ (which has a value between 0 and 1), then

$$\Delta \mathrm{g_{ret}} = \alpha_1 \mathrm{E} + \alpha_2 \mathrm{E} + ... + \alpha_{\mathrm{m}} \mathrm{E} = \mathrm{E}\sum_{i=1}^{\mathrm{m}} \alpha_i$$

By substituting the α_{i}s (Eq. (23)) in the above equation, we have:

$$\Delta \mathrm{g_{ret}} = \mathrm{E}\sum_{i=1}^{m} \frac{\lambda \mathrm{P}\left(\mathrm{g}_{\mathrm{init}}^{(i)}\right)}{\sum_{j=0}^{\mathrm{N_{max}}} j\mathrm{P}_j} = \frac{\lambda \mathrm{E}\sum_{i=1}^{\mathrm{m}} \mathrm{P}\left(\mathrm{g}_{\mathrm{init}}^{(i)}\right)}{\sum_{j=0}^{\mathrm{N_{max}}} j\mathrm{P}_j} \qquad \text{(D.4)}$$

According to Eq. (D.4), the denominator of the fraction is significantly larger than the nominator as the waveform powers multiply with the gates. Therefore, the retracking gate will have relatively small systematic errors compared with the initial gates.

Figure captions

Figure 1. (a): The bathymetric map has been prepared for the Baltic Sea and Skagerrak by combining different depth data. The maximum depth in the Baltic Sea is about 70m (Hell et al. 2012). (b): The bathymetric map has been provided for the Gulf of Oman and the northern Arabian Sea. As we see in the Figure, there is a sharp change in the bathymetric data around the Muscat station, which influences the diversity of waveforms in this area. The maximum

depth in the northern Arabian Sea reaches to 3000m (Pous 2004). The bathymetric data are accessible by the following link: (https://www.opendem.info/download_bathymetry.html).

Figure 2. (a): Locations of Onsala and Halmstad stations on the Swedish coast. Jason-3 satellite with pass number 68 passes over the Onsala, and Jason-2 and Jason-3 satellites with pass number 213 pass over the Halmstad station. (b): Location of Muscat station on the Omani coast. This station is also monitored by Jason-2 satellite, pass number 16, and Jason-3 satellite, pass number 157. The coloured lines depict the satellite tracks passing from the vicinity of the stations and the stations illustrated by stars. The topography data are also provided by the natural earth website (https://www.naturalearthdata.com/).

Figure 3. The ANN architecture used for selecting the waveforms. There are 6 neurons in the hidden layer of the network, which have been selected by trial and error. $\beta_{i,j}^{(k)}$ and $w_{i,j}^{(k)}$ are the biases and the weights of the network.

Figure 4. Illustration of a Brownian (ideal) waveform with its geophysical parameters (Amarouche et al. 2004).

Figure 5. (a): Some of the waveforms with noisy leading edges, selected from land, entered to the ANN with code 0. These waveforms are not suitable for retracking. (b): The waveforms with straight leading edges, selected from the ocean surface, entered to the ANN with code 1. The blue waveforms are Brownian, and the red ones are quasi-specular waveforms. These waveforms have been selected from Jason-3 data with pass number 68.

Figure 6. Variations of the ratio in different tracks and cycles. As we see in the Figure, all of the ratios are larger than 75%. Diagrams (a), (b), (c), (d), and (e) refer to the Onsala (Jason-3, 068), Halmstad (Jason-2, 213), Halmstad (Jason-3, 213), Muscat (Jason-2, 016), and Muscat (Jason-3,157) ratios of the selected data, respectively.

Figure 7. Sample (a) is a Brownian, and (b) refers to a quasi-specular waveform. These waveforms were selected from Jason-3 data, cycle number 21, and pass number 68. As we see in the Figure, the initial gates are distributed uniformly over the waveform.

Figure 8. The flowchart of the spline method.

Figure 9. The scatter plots of SSH estimations from different retrackers against tide gauge measurements over Onsala station. (a) OCOG, (b) Threshold-10%, and (c) Spline retrackers. The black line is the regression line of water level estimates against tide gauge data.

Figure 10. plots (a), (c), (e), (g), and (i) refer to height variations of water derived from Jason-3 data for Onsala station, Jason-2 data for Halmstad station, Jason-3 data for Halmstad station, Jason-2 data for Muscat station, and Jason-3 data for Muscat station. Plots (b), (d), (f), (h), and (j) are also the height differences between the altimetry heights and tide gauge data. The random behavior of the “Spline - TG” diagram refers to the fact that there is no systematic effect in the time series. The SSH time series of the Swedish coastal waters have been computed at the distance of 0-15km, and the Muscat SSHs were obtained at the distance of 0-25km from the coastline.

Figure 11. Classification of waveform shapes for Jason-2 mission using ANNs (Vignudelli et al. 2011). LE, LV, TE, and DP abbreviations in the titles of subplots refer to Leading Edge, Linear Variation, Trailing Edge, and Decreasing Plateau, respectively. Some of these classes can be found in the Swedish and Omani coastal areas, e.g., Brownian, quasi-specular, and peaky + noisy waveforms.

Figure 12. Comparison of the performance of three retracking algorithms on nine waveform types. Waveforms (a), (b), ..., (f), and (i) have been selected from the Jason-3 track, pass number 68, and Jason-2 track, pass number 213. Waveforms (g) and (h) also belong to the Jason-2 track with pass number 16. Sample (a) is a Brownian waveform, sample (b) is a quasi-specular waveform, sample (c) is a Brownian + Peaky waveform, sample (d) is a Peaky + Noise waveform, sample (e) is a Brownian + strong decreasing plateau waveform, sample (f) is a Brownian +

leading edge perturbation waveform, sample (g) is a waveform with rising trailing edge, and sample (h) represents the waveform with very large peak. Sample (i) is a Brownian waveform with a peak in thermal noise level.

Figure 13. Comparison of SSH time series derived from the ALES, improved threshold, and the spline retracking algorithms. The SSH time series of the Jason-3 track in Onsala (a), Jason-2 in Halmstad (b), Jason-3 in Halmstad (c), Jason-2 in Muscat (d), and Jason-3 in Muscat (e) have been provided here. As we see in the Figure (and Table 7) the spline retracking outperforms the ALES in cases (a), (b), (c), and (e), and its performance is degraded in case (d).