

# Direct-Estimation of Sea State Bias in Hy-2 Based on a Merged Dataset

X. Wang

College of Information Science and Engineering  
Ocean University of China  
Qingdao, China

H.L. Miao

College of Information Science and Engineering  
Ocean University of China  
Qingdao, China  
\*Corresponding author

G.Z. Wang

College of Information Science and Engineering  
Ocean University of China  
Qingdao, China

Y.Q. Wang

College of Information Science and Engineering  
Ocean University of China  
Qingdao, China

J. Zhang

First Institute of Oceanography of State Oceanic  
Administration

**Abstract**--In this paper, a merged dataset of collinearly-processed sea surface height (SSH) derived from more than 200 repetition periods of Geophysical Data Record (GDR) data of T/P, Jason-1 and Jason-2 satellites is achieved. A direct-estimation method based on the merged dataset with a resolution of  $0.25 \times 0.25$  is proposed. Compared with the previous non-parametric or parametric models, our direct-estimation method is with a better accuracy owing to the high spatial resolution of the merged dataset from various satellites. By applying the direct-estimation method in HY-2 satellite, we get the sea state bias (SSB) of HY-2 and demonstrate that this method is with a high accuracy, a good data extension and a wide applicability. Hence, the direct-estimation model can be effectively used for the SSB correction in the current and the subsequent in-orbit satellites.

**Keywords**--satellite altimeter; sea state bias; direct-estimation method; significant wave height

## I INTRODUCTION

The Sea State bias (SSB) consisting of the electromagnetic bias, the skewness bias and the tracker bias is one of the fundamental error sources in the measurement of the sea surface height in altimetry [1, 2]. Due to the improved orbital technology in recent years, SSB has replaced the orbit errors and became the largest source of error [3]. The empirical models are previously used to correct the SSB: the parametric model, the nonparametric model and the direct-estimation method [4]. Although the parametric model is simple and straightforward, it fails to effectively estimate the true SSB due to the existed errors in the model [5]. Compared with the parametric model, the nonparametric model is with a better accuracy, however, this approach is very complicated and lacks a good extension [6].

The direct estimation SSB value through the merged dataset can be applicable to a wider range of sea conditions [7] and especially out of modeling using altimeter, the SSB application like in HY-2 or other subsequent altimeters. The

previous direct-estimation method which is based on a single altimeter dataset only can be used for the single altimeter. The future trend of the sea state bias correction is the optimized data set and using more parameters. Guizhong Wang [8] has built a merged dataset for the parametric model. HaoRan Ren [9] has studied on the altimeter-based inversion model of mean wave period. In this paper the direct-estimation is based on the collinear data that in the same location at different time. Statistical analysis was performed by numerous multiple satellite altimeter sea surface height difference to reveal the inherent dependent variable relationship among SSB and significant wave height (SWH), wind speed (U). This paper integrates more than 200 cycles Geophysical Data Record (GDR) data of T/P, Jason-1 and Jason-2 satellites, collinear the sea surface height, and gets the collinear dataset. Then build the direct estimation SSB table by the dataset and apply the table to HY-2. Finally analyze and evaluates the effectiveness of the method.

## II BASIS

The sea surface height after add all of the geophysical correction terms except SSB can be expressed as:

$$SSH' = h_g + \eta + SSB + v \quad (1)$$

The sea surface height measurement that contains the geoid signal ( $h_g$ ), the ocean dynamic topography ( $\eta$ ), the SSB, and some other noise ( $v$ ).

We can get the collinear sea surface height through the collinear processing data of every measurement point, and then average them to obtain the mean sea surface height (MSSH). The difference of SSH' and MSSH is  $\Delta SSH$ .

$$\Delta SSH = SSH' - MSSH = SSB + \varepsilon \quad (2)$$

$\varepsilon$  consisting of other noise ( $v$ ) and the ocean dynamic topography ( $\eta$ ), the geoid signal ( $h_g$ ) as a time invariant signal has been eliminated by the equation (2).

Previous studies have demonstrated that SSB is the function of SWH and U, therefore, on the base of SWH and U, reconstruction  $\Delta SSH$  data into the bin width of 0.25 m/s in U and 0.25 m in SWH ( $\Delta SWH, \Delta U$ ). The mean value of  $\Delta SSH$ :

$$\overline{\Delta SSH} = \overline{SSB} + \bar{\varepsilon} \quad (3)$$

In each data bin has large amounts of data in the long-term repeated observation. The  $\Delta SSH$  satisfies the normal distribution. Because of the SSB in the relatively small sea range is constant, and the  $\varepsilon$  satisfies the normal distribution [10]. The mean of  $\varepsilon$  therefore is 0. The estimated value of SSB is the mean value of  $\Delta SSH$ :

$$\overline{SSB} = \overline{\Delta SSH} \quad (4)$$

The equation shows that: the estimation of the SSB value is the mean value of sea surface height difference contains all the measuring points in the range of each ( $\Delta SWH, \Delta U$ ) data bin, and that is the basis for the direct estimation methods.

### III DATASET

#### A. Collinear Processing

Extract the data based on the T/P, Jason-1, and Jason-2 satellite altimeter over 200 repeated cycles of GDR and get the collinear processed SSH. First determine the reference orbit and normal points from the same pass in different cycles, using distance weighted average method, according to the SSH of several observation point near the normal point calculated the SSH of the normal point, the MSSH of the normal point is the mean value of the SSH. Process is shown in Figure 1. All the MSSH collinear processing of the whole normal points requires the same treatment on the rest of normal points and on other reference orbits.

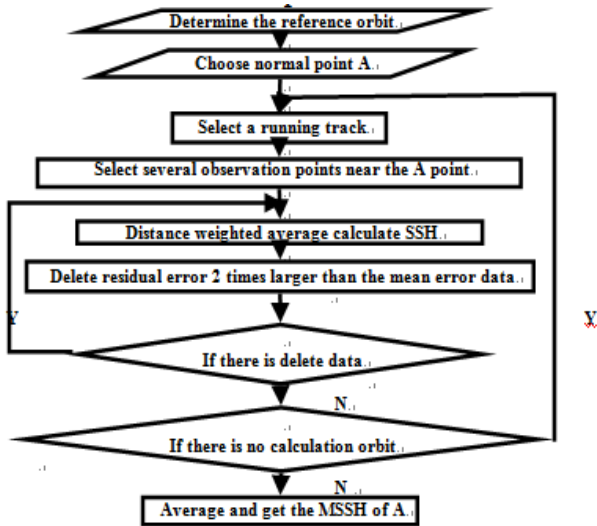


FIGURE 1. SEA SURFACE HEIGHT COLLINEAR PROCESSING FLOW CHART.

#### B. Determine Reference Orbit

Satellite orbit is elliptical, due to the curvature is small, so the orbit can be seen as a circular orbit. Thus we

can calculate from the numerous repeat tracks. Calculate the normal points of the normal track per second interval, and the connection of normal point is the reference orbit. But this calculation method is very complex. In this paper, approximate reference orbit by a relatively simple method. First select the same track in different cycles, and then calculate the number of each track available observation value, select the track with the best quality and the maximum number of points as a reference orbit, every measuring point in the reference orbit as normal point. This method is not only simple, but also its accuracy is almost same as the result from the strict calculation.

#### C. SSH of Normal Point Calculation

After determining the reference orbit and normal point, the values of normal points on reference orbit are calculated by using the observed values on repeat orbit. We use the distance weighted average method.

Assuming that a normal point A, the longitude is  $\lambda$ , the latitude is  $\phi$ . We choose the latitude of A as the reference, and the ground track I of the same pass which in the different cycles. In this pass there are some (assuming the quantity is k) closest points (Q1, Q2.....Qk) to A. Distance weighted average the k points to get the  $SSH(i)$  value of A. We choose k as 8 in this paper, Distance weighted average equation:

$$SSH(i) = \frac{\sum_{j=1}^k (SSH_j \times p_j)}{\sum_{j=1}^k p_j} \quad (5)$$

$SSH_j$  is the observed value of sea surface height corresponding point j. And  $p_j$  is the weights corresponding point j:

$$p_j = 1/s_j \quad (6)$$

$s_j$  is the spherical distance from the calculated points to the normal points, expression is as follows:

$$s_j = \sqrt{(\Delta \phi_j)^2 + (\Delta \lambda_j \cos \phi_j)^2} \quad (7)$$

$\Delta \phi_j$ ,  $\Delta \lambda_j$  are the differences of the longitude and latitude between the calculated points to the normal points.  $\phi_j$  is the latitude of the calculated point j. And the residual error is  $v_j$ :

$$v_j = SSH_j - SSH(i) \quad (8)$$

Root mean square error(RMSE)  $\sigma$  is:

$$\sigma = \sqrt{\frac{\sum_{j=1}^k p_j v_j^2}{k-1}} \quad (9)$$

If there is  $v_j > 2\sigma$  on the point  $j$ , the point will be regarded as not satisfy the standard. Distance weighted average and inspect again after eliminating the point until all the calculated point satisfy the standard. Repeat the previous process after take the place of the  $i$  by  $i+1$ . Calculating the SSH of all the normal point on the same pass of all the cycles. After average the calculated SSH we can get the MSSH of the normal point A.

#### D. The Establishment of a Dataset

Based on the MSSH after average the collinear processing, interpolating the MSSH by the arbitrary

TABLE I. PART OF THE MERGED DATASET (EXAMPLE).

Lon/ $^{\circ}$	Lat/ $^{\circ}$	SWH/m	$U/m \cdot s^{-1}$	SSH'/m	MSSH/m
167.189828	13.424731	1.429564	5.074660	25.688452	25.372418
167.208545	13.473518	1.498096	5.425461	26.138562	25.811412
167.227271	13.522304	1.447417	5.272471	26.634680	26.228908
167.246006	13.571086	1.511679	5.182139	27.087192	26.684649

### IV DIRECT ESTIMATION METHOD

Using the dataset, divide the data bin as  $(\Delta SWH, \Delta U) = (0.25m, 0.25m/s)$ , and  $(SWH_0, U_0) = (0.25m, 0.25m/s)$ . Distribution of the number of each data bin contains data is shown in Figure 2. According to equation (2) make difference between MSSH and its corresponding SSH. The difference is  $\Delta SSH$ , from equation (4) we can know in every data bin the  $\Delta SSH$  is  $\overline{SSB}$ .

Get the SSB by subtracting MSSH from SSH in every effective bin, and averaging the difference the SSB represents the sea conditions of the corresponding data bin.

And set up the SSB query table  $(SSB, SWH, U)$  in the  $(SWH, U)$  plane. For any conditions of  $(SWH, U)$ , we can obtain the SSB corresponding correction value by the bilinear interpolation method. Table 2 is the example of the SSB query table:

TABLE II. PART OF THE SSB QUERY TABLE (EXAMPLE).

$U/m \cdot s^{-1}$ SWH / m	3.875	4.125	4.375	4.625	...
2.875	-0.075924	-0.075804	-0.077766	-0.078488	...
3.125	-0.079474	-0.075945	-0.077152	-0.080252	...
3.375	-0.071285	-0.076249	-0.076915	-0.080462	...
3.625	-0.084077	-0.085750	-0.091718	-0.088348	...

Figure 3 shows the results of the SSB direct estimation correction method. From the figure, most of the estimated SSB values are between -0.35m and 0m. In contrast with Figure 2, we can find that the regions where there are bins contain large number of data, the distributions of estimated SSB are more continuous and the accuracy is higher.

observation point of all cycles, established the dataset together with the point of the information of latitude and longitude, SWH, U, the SSH without SSB correction and MSSH. This merged dataset includes more than 200 cycles GDR of T/P, Jason-1 and Jason-2 multi satellites. After remove the abnormal data, the number of the merged dataset is about 88000000. This dataset gives support to the wide applicability of our method. Table 1 is the example of the merged dataset:

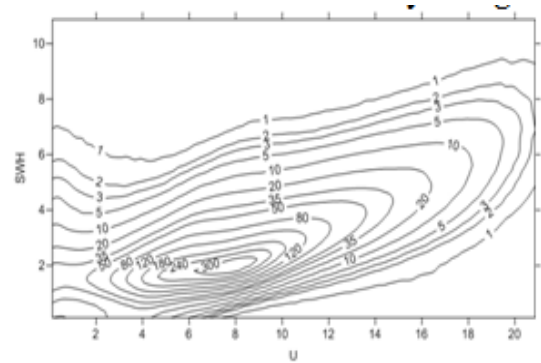


FIGURE II. DISTRIBUTION OF THE AMOUNT DATA BIN CONTAINED DIVIDED (/1000).

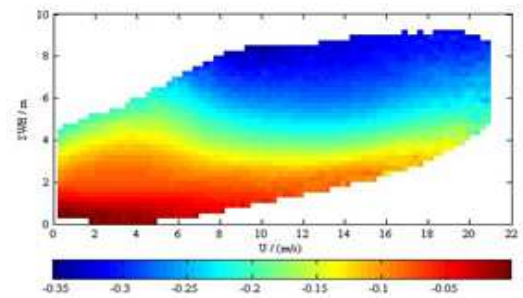


FIGURE III. DISTRIBUTION OF DIRECTLY OF ESTIMATED SSB (UNIT: M).

### V APPLICATION AND EVALUATION

#### A. Application in HY-2

Using the 027 cycle GDR data of HY-2, extract the SWH, U and other information. Bilinear interpolation according to the SSB query table and get the HY-2 SSB direct estimation value. Distribution of directly estimated SSB of HY-2 is shows in Figure 4. We can know that the direct estimation of

query table and the distribution value (Figure 3) have the same trends. The SSB value in most of region is in the range of -0.1m ~ 0m.

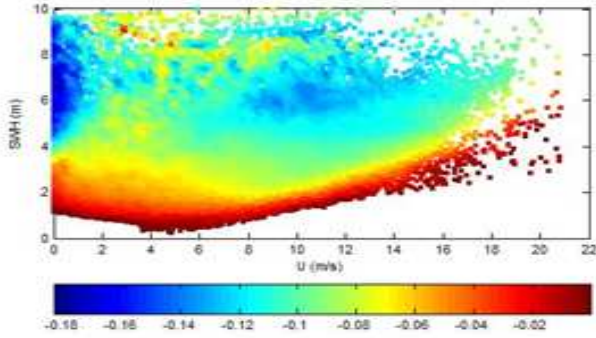


FIGURE IV. DISTRIBUTION OF DIRECTLY ESTIMATED SSB IN HY-2 (UNIT: M).

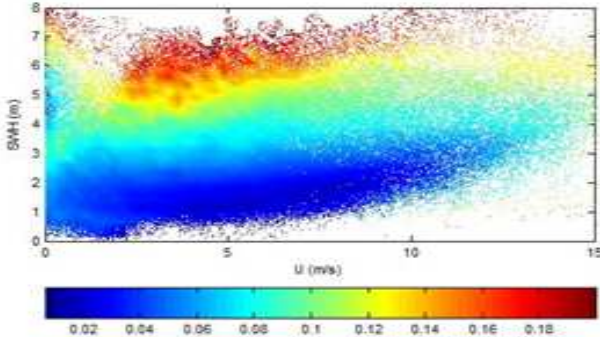


FIGURE V. THE DISTRIBUTION OF THE ABSOLUTE VALUE OF THE DIFFERENCE BETWEEN THE SSB DIRECT ESTIMATION AND THE PARAMETER MODEL IN GDR (UNIT: M).

### B. Effectiveness Analysis

The Effectiveness of model can be described by explained variance (D). Explained variance means the subtraction of the SSH differences with and without the SSB, it also can be understood as the part of the error which can be explained by SSB. Additionally, the model is more effective when the explained variance is larger. Its formula is as following:

$$D = \frac{1}{n-1} \sum_{i=1}^n (\Delta SSH'_i - \overline{\Delta SSH'})^2 - \frac{1}{n-1} \sum_{i=1}^n ((\Delta SSH'_i - SSB_i) - (\overline{\Delta SSH'} - \overline{SSB}))^2 \quad (10)$$

$\Delta SSH'_i$  is the SSH differences without SSB correction of the point  $i$ .  $\overline{\Delta SSH'}$  is the mean value of  $\Delta SSH'_i$ .  $\Delta SSH'_i - SSB_i$  is the SSH differences with SSB correction of the point  $i$ , similarly  $\overline{\Delta SSH'} - \overline{SSB}$  is the mean value of  $\Delta SSH'_i - SSB_i$ .

Using the direct estimation of SSB on sea surface height correction, the explained variance is 39cm<sup>2</sup>, and using the SSB from the parameter model in GDR on sea surface height correction, he explained variance is 43cm<sup>2</sup>. The values of

two different explained variances are very close, it indicates that the direct estimation method is effective.

### C. Applicability Analysis

Further test of the applicability of the model, we contrast the direct estimation method results with the selected parameter model SSB of cycle 027 data of the HY-2 GDR. The result is shown in Figure 5: there is a good agreement between them when the SWH under 3.5m and the U under 15m/s. Their difference is within 0~2cm in that area. The differences are bigger when the data distribution is smaller. It indicated that there is a correlation between the accuracy and applicability of direct estimation method with the amount of data within the each data bin. This is the reason why we use the merged dataset which set up by the multiple satellite data to increase the amount of data and improve the applicability of the direct estimation method. Though the direct estimation SSB query table is made by the dataset without HY-2 data, it also can be used in HY-2, further proves that this method has extensive and good applicability.

### D. Skewness Inspection

The Figure 2 shows that: the amount of data contained within each data bin is not balanced. Ranging from 244~303197. As the range of every data bin is small, the SSB value of every bin changes very little and can be viewed as a constant. From the equation (3) and (4), we can know that if the  $\overline{SSB} + \varepsilon$  satisfies the normal distribution, the mean of  $\varepsilon$  is approximate 0. The estimated value of SSB is the mean value of  $\Delta SSH$ . It satisfies the normal distribution better, the SSB estimates are more accurate.

The skewness (skewness) is a digital content of the degree of asymmetry distribution of statistical data. The calculation formula of skewness in the data bin for:

$$SK = \frac{n \sum (\Delta SSH_i - \overline{\Delta SSH_i})^3}{(n-1)(n-2)S_d^3} \quad (11)$$

In this equation n is the amount of data for the bin,  $S_d$  is the standard deviation of the SSB in each bin. Calculating the skewness distribution value of each data bin, we can make a quantitative evaluation of the distribution state. From the skewness calculated results and the comparative analysis, we know that with the increase of the amount of data in the bin the skewness decreased. It indicates that with the increase of the amount of data in the bin, the data density distribution satisfies the normal distribution better. Extraction calculation 4 bins results are shown in Table 3, the data density distribution results are shown in Figure 6.

TABLE III. SKEWNESS OF FOUR BINS.

The middle value of data bin (SWH,U)/(m, m/s)	The amount of data (n)	Skewness(s k)
(4.875,2.875)	1705	0.35922
(3.625,3.875)	11501	0.2991
(2.125,4.875)	156050	0.21624
(1.625,5.875)	257165	0.17625

In the Figure 6, (a),(b),(c),(d) represent the  $\Delta$ SSH data density distribution of results of each data bin whose the middle value of data bin are  $(4.875 \pm 0.125\text{m}, 2.875 \pm 0.125\text{m/s})$ ,  $(3.625 \pm 0.125\text{m}, 3.875 \pm 0.125\text{m/s})$ ,  $(2.125 \pm 0.125\text{m}, 4.875 \pm 0.125\text{m/s})$ ,  $(1.625 \pm 0.125\text{m}, 5.875 \pm 0.125\text{m/s})$ .

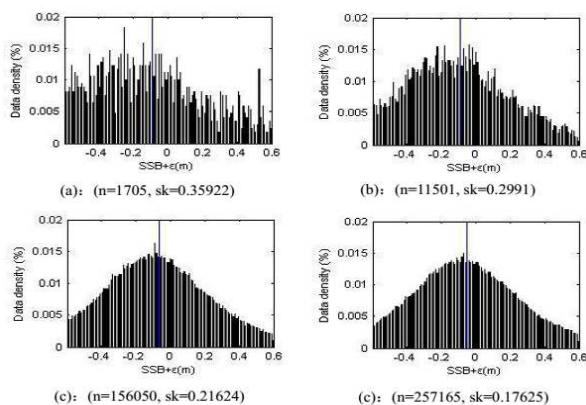


FIGURE VI.  $\Delta$ SSH DATA DENSITY DISTRIBUTION OF RESULTS OF FOUR DATA BIN.

## VI CONCLUSION

Establish a merged dataset with T/P, Jason-1 and Jason-2 multiple altimeter data. Put to use this dataset and select the MSSH obtained by collinear processing as a reference. Statistics the relationship among the SSH discrepancy and SWH, U point by point. Obtain the SSB query table based on SWH, U. It can be used widely and give a convenient method to examine if the SSB corrected in the GDR is right.

Application in HY-2 and contrast the direct estimation method results with the parameter model SSB in the HY-2 GDR. The two explained variance are very close (39cm<sup>2</sup>, 43cm<sup>2</sup>), and the difference of SSB between the two method is within 0~2cm when the SWH under 3.5m and the U under 15m/s. It indicated that the direct estimation method has extensive and good applicability and high precision. Through the skewness inspection, it shows that there is a close relationship between the amount of data in the bin and the accuracy of the direct estimation method. In most of the data range the direct estimation SSB query table achieves a good accuracy.

## ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of China (NSFC, 41176157 and 41406197).

## REFERENCE

- [1] Yapple BS, Shapiro A, Hammond DL et, al. Nanosecond Radar Observation of the Ocean Surface from a Surface from a Stable Platform[J], Geoscience Electronics, IEEE Transactions on, 1971, 9(3): 170-174
- [2] Fu Lee-Lueng, Anny Cazenave. Satellite Altimetry and Earth Sciences[M]. San Diego: ACADEMIC Press, 2001. 57-65
- [3] Zhai GuoJun, Huang MoTao, Xie XiJun. Theory and methods of satellite altimetry data processing[M]. Surveying and Mapping Press. 2000.

- [4] Gaspar P, Labroue S, Ogor F. Improving nonparametric estimates of the sea state bias in radar altimeter measurements of sea level[J]. Journal of atmospheric and oceanic technology, 2002, 19: 1690-1707
- [5] Gaspar P, Ogor F, Traon P et al. Estimating the sea state bias of the TOPEX and POSEIDON altimeter from crossover differences [J]. Journal of geophysical research, 1994, 99(C12): 24 981-24 994
- [6] Gaspar P, Florens JP. Estimation of the sea state bias in radar altimeter easurements of sea level: Results from a new non parametric method[J]. Journal of geophysical research, 1998, 103(C8):803-814
- [7] Tran N, Labroue S, Philipps S. et al. Overview and Update of the Sea State Bias Corrections for the Jason-2, Jason-1 and Topex Missions[J]. Marine Geodesy, 2010, 33: 348-362.
- [8] Wang GuiZhong, Miao HongLi, Wang Xin, et al. Study on Parametric Model of Sea State Bias in Altimeter based on Fusion Dataset of Collinear and Crossover[J]. Remote sensing technology and Application, 2014, 29(1):176-180.
- [9] Miao HongLi, Ren HaoRan, Zhou XiaoGuang, Wang GuiZhong, Zhang Jie. "Study on altimeter-based inversion model of mean wave period", [J], J. Appl. Remote Sens, 2012,6(1), 063591 (Nov 07, 2012).<http://dx.doi.org/10.1117/1.JRS.6.063591>
- [10] Vandemark D, Tran N, Beckley. et al. Direct estimation of sea state impacts on radar altimeter sea level measurements[J]. Geophysical Research Letters, 2002, 29(24): 48-52.