AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-201703.00283

# An Improved Local Linear Regression Estimator and Its Application to Radar Altimeter Sea State Bias Estimation (Postprint)

**Authors:** Jiang Maofei, Permission, Liu Yalong and Wang Lei

**Date:** 2017-03-10T00:00:00+00:00

## Abstract

In establishing non-parametric models for radar altimeter Sea State Bias (SSB), the Local Linear Regression (LLR) estimator is commonly employed. Traditional local linear regression estimators involve high-dimensional matrix operations; when the volume of modeling data is large, estimating sea state bias requires substantial computational time, thereby rendering non-parametric estimation methods difficult to apply to high-dimensional sea state bias models. This paper proposes an Improved Local Linear Regression (ILLR) estimator that avoids the high-dimensional matrix operations required by traditional LLR estimators. Without affecting sea state bias estimation results, it reduces the time complexity of obtaining the weighting function for the local linear regression estimator from to, thus significantly decreasing the time required for sea state bias estimation and establishing a foundation for real-time computation of high-dimensional non-parametric sea state bias models.

## Full Text

### An Improved Local Linear Regression Estimator and Its Application to Radar Altimeter Sea State Bias Estimation

**Jiang Maofei[123], Xu Ke*[12], Liu Yalong , Wang Lei[12]**

[1](Key Laboratory of Microwave Remote Sensing, Chinese Academy of Sciences, Beijing 100190, China)
[2](National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China)
[3](University of Chinese Academy of Sciences, Beijing 100049, China)
(Yantai Marine Environmental Monitoring Center Station, State Oceanic Administration, Yantai 264000, China)

ChinaRxiv 中X

**Abstract:** When developing nonparametric models for radar altimeter Sea State Bias (SSB), the Local Linear Regression (LLR) estimator is commonly employed. However, conventional LLR estimators involve high-dimensional matrix operations, which require substantial computational time when modeling large datasets, making nonparametric estimation methods impractical for high-dimensional SSB models. This paper proposes an Improved Local Linear Regression (ILLR) estimator that avoids the high-dimensional matrix operations required by traditional LLR estimators. The proposed method reduces the time complexity of obtaining the weight function from $O(N^3)$ to $O(N)$ without affecting SSB estimation accuracy, thereby significantly reducing the time required for SSB estimation and laying the foundation for real-time computation of high-dimensional nonparametric SSB models.

**Keywords:** Radar altimeter; Sea state bias; Nonparametric estimation; LLR estimator

**DOI:** 10.11999/JEIT151280

## 1 Introduction

Global sea level rise has attracted increasing attention, and one important application of radar altimeters is measuring mean sea surface height [1, 2]. However, due to Sea State Bias (SSB), altimeter measurements of mean sea surface height are lower than the actual values, necessitating correction. SSB primarily arises from the non-Gaussian characteristics of the sea surface, including electromagnetic bias and skewness bias [3]. The uncertainty of SSB obtained using current methods can reach 2 cm [4], making it the largest error source for Jason series altimetry satellites [5, 6].

Currently, operational radar altimeter SSB corrections primarily employ nonparametric SSB models that use wind speed and significant wave height from altimeter measurements as inputs [6, 7]. Constructing nonparametric SSB models typically requires the Local Linear Regression (LLR) estimator, which involves extensive matrix operations and consumes substantial computational time, severely limiting the application of nonparametric estimation in high-dimensional SSB models.

This paper proposes an improved LLR estimator that avoids high-dimensional matrix operations and reduces the time complexity of obtaining the weight function from $O(N^3)$ to $O(N)$. Numerical experiments demonstrate that under identical conditions, the improved LLR estimator produces the same SSB estimation results as the conventional LLR estimator while substantially reducing the time required for SSB estimation.

## 2 Local Linear Regression Estimator and Its Application to SSB Nonparametric Estimation

GASPAR et al. [8] first applied nonparametric estimation methods to SSB model construction in 1998. Compared with parametric models [4, 9], nonparametric models offer higher accuracy and have been more widely applied to SSB correction in altimeter sea surface height measurements. The main estimators used for nonparametric SSB models are the Nadaraya-Watson (NW) estimator [8] and the Local Linear Regression (LLR) estimator [10]. The NW estimator exhibits large biases near boundaries, whereas the LLR estimator effectively addresses this issue [11, 12], making the LLR estimator more widely adopted.

### 2.1 Local Linear Regression Estimator

Given $N$ observation pairs $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i$ is a $p$-dimensional vector, $y_i$ and $\mathbf{x}_i$ satisfy:

$$y_i = r(\mathbf{x}_i) + \varepsilon_i$$

where $r(\cdot)$ is the regression function and $\varepsilon_i$ is a zero-mean error term. For the regression function $r(\mathbf{x})$, performing a Taylor expansion in the neighborhood of $\mathbf{x}$ yields:

$$r(\mathbf{z}) \approx r(\mathbf{x}) + \sum_{j=1}^{p} \frac{\partial r(\mathbf{x})}{\partial x_j}(z_j - x_j)$$

For the given $N$ observation pairs $(\mathbf{x}_i, y_i)$, we can find coefficients $\beta_0, \beta_1, \dots, \beta_p$ through weighted least squares to minimize:

$$\sum_{i=1}^{N} \left[ y_i - \beta_0 - \sum_{j=1}^{p} \beta_j (x_{ij} - x_j) \right]^2 K_H(\mathbf{x}_i - \mathbf{x})$$

where $K_H(\cdot)$ is a weighting function, typically chosen as a kernel function. Common kernel functions include the Gaussian kernel and the Epanechnikov kernel. Let $\mathbf{X}$ be the design matrix and $\mathbf{W}$ be the $N \times N$ weighting matrix. Equation (3) can be rewritten as:

$$(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)$$

Minimizing equation (4) yields:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}$$

The LLR estimator for $r(\mathbf{x})$ is $\hat{r}_{LLR}(\mathbf{x}) = \hat{\beta}_0$, while the other components are first-order partial derivatives of the function $r(\mathbf{x})$, i.e., $\hat{\beta}_j = \frac{\partial r(\mathbf{x})}{\partial x_j}$ for $j = 1, \dots, p$.

For the given $N$ observation pairs $(\mathbf{x}_i, y_i)$, according to nonparametric estimation theory, the conditional expectation $E[y_i|\mathbf{x}_i]$ represents the mean of $y_i$ given $\mathbf{x}_i$. Its kernel estimator $\hat{r}(\mathbf{x})$ can be viewed as a weighted average of observations $y_i$:

$$\hat{r}(\mathbf{x}) = \sum_{i=1}^{N} w_i(\mathbf{x}) y_i$$

For the LLR estimator, setting $\mathbf{e}_1^T = [1, 0, \dots, 0]$, the weight function can be obtained as:

$$\mathbf{w}(\mathbf{x}) = \mathbf{e}_1^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}$$

### 2.2 Application of LLR Estimator to SSB Nonparametric Estimation

SSB models are typically constructed using sea surface height discrepancies at crossover points. A satellite’ s ground track includes two arcs per orbital revolution: ascending and descending orbits. The intersection of ascending and descending orbits is called a crossover point, and the crossover discrepancy refers to the difference in sea surface height measured by the ascending and descending orbits at this point [13].

Let $h$ represent the uncorrected sea surface height measurement. Assuming SSB is a function of input variables $\mathbf{x}$, and letting $r(\mathbf{x})$ denote the SSB function, according to [8] and [10]:

$$h_1 - h_2 = r(\mathbf{x}_1) - r(\mathbf{x}_2) + \varepsilon$$

where $\varepsilon$ is a zero-mean error term, and subscripts 1 and 2 correspond to the ascending and descending orbits, respectively. Under the given conditions, the LLR estimator can be applied to estimate the SSB function $r(\mathbf{x})$.

---

## 3 Data and Experiments

### 3.1 Data Source and Processing

This study uses the Jason-2 altimeter Geophysical Data Record (GDR) products. The Jason-2 satellite, launched in June 2008, is the successor to TOPEX/POSEIDON and Jason-1 and is recognized as the most accurate altimetry satellite currently available. For constructing the nonparametric SSB model, data from cycles 51-100 (50 cycles total) were used. For evaluating the impact of the ILLR estimator on SSB estimation results, three years of data were used: 2009 (cycles 19-55), 2010 (cycles 56-92), and 2011 (cycles 93-128).

Constructing SSB models requires wind speed, significant wave height, and uncorrected sea surface height discrepancies at crossover points. All correction terms needed for calculating sea surface height must be interpolated to crossover points. This study employs cubic spline interpolation. To ensure better interpolation quality, only crossover points with at least four consecutive measurement points on each side are considered valid, which guarantees good altimeter data quality.

### 3.2 Kernel Function and Bandwidth Selection

The selection of kernel functions and bandwidth is crucial for nonparametric SSB models. Currently, the most commonly used kernel functions are the Gaussian kernel and the Epanechnikov kernel. Bandwidths for nonparametric SSB models include global bandwidth and local bandwidth. Global bandwidth uses fixed values for wind speed bandwidth $h_U$ and significant wave height bandwidth $h_{SWH}$, whereas local bandwidth automatically adjusts $h_U$ and $h_{SWH}$ according to data density.

Compared with the Gaussian kernel, the Epanechnikov kernel can transform the matrix in linear system equation (14) into a sparse matrix, thereby significantly improving computational efficiency, though its performance in data-sparse regions is relatively poor. Compared with global bandwidth, local bandwidth can adjust according to data density, offering advantages in data-sparse regions. Since the numerical experiments in this paper aim only to evaluate the computational efficiency of the improved local linear estimator and its impact on estimation results, the choice of kernel function and bandwidth is not critical. This study employs two combinations of kernel functions and bandwidths: (1) Gaussian kernel with global bandwidth, corresponding to the model in [10]; and (2) Epanechnikov kernel with local bandwidth, corresponding to the model in [12].

### 3.3.1 Efficiency Comparison Between LLR and ILLR Estimators

We first compare the time required for SSB estimation using LLR and ILLR estimators. Table 1 presents results obtained using the Gaussian kernel with global bandwidth. As the number of crossover points $N$ used for each SSB estimation increases, the matrix dimensions in the LLR estimator grow accordingly, leading to significantly increased computation time. In contrast, the matrix dimensions in the ILLR estimator do not change with $N$, and although its required time also increases with $N$, the rate of increase is relatively small.

Table 2 shows the time required for SSB estimation using the Epanechnikov kernel with local bandwidth. Normally, the Epanechnikov kernel can convert the matrix in linear system equation (14) into a sparse matrix, thereby improving computational efficiency and reducing SSB estimation time. However, due to the use of local bandwidth, determining the appropriate bandwidth for each grid cell corresponding to every input **x** during SSB estimation adds computational overhead, increasing the overall computation time. Similar to Table 1, as $N$ increases, the computation time for the LLR estimator increases substantially, while the increase for the ILLR estimator remains relatively modest.

Both Table 1 and Table 2 demonstrate that the ILLR estimator can significantly reduce SSB estimation time compared with the LLR estimator, with greater improvements observed for larger $N$.

Nonparametric SSB models ultimately produce a lookup table of SSB values as a function of $U$ (wind speed) and $SWH$ (significant wave height). Using only $U$ and $SWH$ provides limited SSB information, and introducing new parameters (such as wave period) into SSB models has become a new research direction [6, 16, 17]. When constructing high-dimensional nonparametric SSB models, the resulting lookup table will have far more grid points than a two-dimensional table. The time required using the LLR estimator would be enormous, and more data would be needed to ensure model validity as dimensionality increases, posing a significant challenge for the LLR estimator. The ILLR estimator effectively solves this problem.

### 3.3.2 Comparison of SSB Estimation Results Between LLR and ILLR Estimators

Figure 1 [Figure 1: see original paper] shows results obtained using the Gaussian kernel with global bandwidth. Based on data density, the data are divided into two regions: the shaded area represents regions with fewer than 100 altimeter measurements, indicating data-sparse regions, while the remaining area represents data-dense regions. The contour lines in Figures 1(a) and 1(b) show the distribution of SSB lookup tables obtained using the LLR and ILLR estimators, respectively, in the $(U, SWH)$ plane, with contour units in cm. Clearly, the SSB variations with $U$ and $SWH$ obtained by both estimators are identical: under the same $U$, SSB magnitude increases with $SWH$; under the same $SWH$, SSB magnitude initially increases with $U$ but gradually decreases after $U$ reaches a certain value. Figure 1(c) shows the distribution of differences between SSB values obtained by the two estimators in the $(U, SWH)$ plane. In data-dense regions, the differences are on the order of $10^{-4}$ cm, which is essentially negligible.

Figure 2 [Figure 2: see original paper] presents results obtained using the Epanechnikov kernel with local bandwidth. The contour lines in Figures 2(a) and 2(b) show the SSB lookup tables from the LLR and ILLR estimators, respectively, in the $(U, SWH)$ plane (units in cm). As with Figure 1, the SSB variations with $U$ and $SWH$ are identical for both estimators. Due to the use of the Epanechnikov kernel, the results in data-sparse regions (low wind speed with high sea state and high wind speed with low sea state) are less smooth than those in Figure 1 because fewer data points are used. Figure 2(c) shows the differences between the two estimators' SSB values in the $(U, SWH)$ plane. In data-dense regions, the differences are on the order of $10^{-6}$ cm, indicating that the SSB values obtained by both estimators are essentially identical.

Figure 3 [Figure 3: see original paper] shows scatter plots of SSB estimates obtained using the LLR and ILLR estimators. These results were produced by applying the SSB lookup tables from both estimators to Jason-2 altimeter data from 2011. The SSB estimates from both estimators show excellent consistency. Applying both lookup tables to Jason-2 data from 2009-2011 and performing annual statistical analyses yields the results shown in Table 3 . When using the Gaussian kernel with global bandwidth, the maximum difference between SSB estimates from the two estimators is on the order of $10^{-3}$ cm, with the mean difference on the order of $10^{-4}$ cm. When using the Epanechnikov kernel with local bandwidth, the maximum difference is on the order of $10^{-5}$ cm, with the mean difference on the order of $10^{-6}$ cm. Therefore, regardless of the kernel function and bandwidth combination, the impact of the ILLR estimator on SSB estimation results is negligible.

Currently, explained variance is the primary metric used internationally to evaluate SSB models. Explained variance is defined as the variance of uncorrected sea surface height discrepancies minus the variance after SSB correction. It represents the portion of variance in sea surface height discrepancies that can be explained by SSB, with larger explained variance indicating a more effective SSB model. Using three years of data (2009-2011), we calculated the explained variance for SSB obtained by both estimators for the two kernel function and bandwidth combinations. Table 4 shows the differences in explained variance between the two estimators. When using the Gaussian kernel with global bandwidth, the difference in explained variance is on the order of $10^{-5}$ cm². When using the Epanechnikov kernel with local bandwidth, the difference is on the order of $10^{-9}$ to $10^{-7}$ cm². Thus, regardless of the kernel function and bandwidth combination, the ILLR estimator has negligible impact on SSB model effectiveness compared with the LLR estimator, and consequently does not affect the final altimetry measurement precision.

---

## 4 Conclusion

Nonparametric models have been widely used for radar altimeter SSB correction due to their high accuracy. When estimating SSB using nonparametric methods, the LLR estimator is typically employed to obtain the weight function in the weighting matrix. However, the LLR estimator involves high-dimensional matrix operations that require substantial computation time when dealing with large datasets.

This paper proposes an ILLR estimator that avoids high-dimensional matrix operations and significantly improves computational efficiency. Using two combinations—Gaussian kernel with global bandwidth and Epanechnikov kernel with local bandwidth—the differences between SSB values obtained by the ILLR and LLR estimators are essentially negligible. The ILLR estimator substantially reduces the time required for SSB estimation, with more pronounced improvements for larger datasets. Currently, operational altimeters employ two-dimensional SSB models based on wind speed and significant wave height. Introducing new parameters into SSB models has become a new research direction for empirical SSB models. The ILLR estimator proposed in this paper can significantly reduce SSB estimation time without affecting estimation accuracy, thereby laying the foundation for real-time computation of high-dimensional nonparametric SSB models.

## References

[1] WANG L. Study on the data processing for high precision satellite radar altimeter[D]. [Ph.D. dissertation], National Space Science Center, Chinese Academy of Sciences, 2015.

[2] WANG L, XU K, SHI L W, et al. A new range migration correction algorithm and its simulation for SAR altimeter[J]. Journal of Electronics & Information Technology, 2015, 37(11): 2713-2718. doi: 10.11999/JEIT150282.

[3] LIU Y L. Calibration technology for HY-2 radar altimeter sea surface height[D]. [Ph.D. dissertation], Ocean University of China, 2014.

[4] GASPAR P, OGOR F, LETRAON P Y, et al. Estimating the sea-state bias of the topex and poseidon altimeters from crossover differences[J]. Journal of Geophysical Research-Oceans, 1994, 99(C12): 24981-24994. doi: 10.1029/94JC01430.

[5] PRANDI P, PHILIPPS S, PIGNOT V, et al. SARAL/AltiKa global statistical assessment and cross-calibration with Jason-2[J]. Marine Geodesy, 2015, 38(10): 297-312. doi: 10.1080/01490419.2014.995840.

[6] TRAN N, VANDEMARK D, CHAPRON B, et al. New models for satellite altimeter sea state bias correction developed using global wave model data[J]. Journal of Geophysical Research-Oceans, 2006, 111(C9): 141-152. doi: 10.1029/2005JC003406.

[7] BONNEFOND P, EXERTIER P, LAURAIN O, et al. SARAL/AltiKa absolute calibration from the multi-mission corsica facilities[J]. Marine Geodesy, 2015, 38(10): 171-192. doi: 10.1080/01490419.2015.1029656.

[8] GASPAR P and FLOREANS J P. Estimation of the sea state bias in radar altimeter measurements of sea level: Results from a new nonparametric method[J]. Journal of Geophysical Research-Oceans, 1998, 103(C08): 15803-15814. doi: 10.1029/98JC01194.

[9] CHELTON D B. The sea-state bias in altimeter estimates of sea-level from

collinear analysis of topex Data[J]. Journal of Geophysical Research-Oceans, 1994, 99(C12): 24995-25008. doi: 10.1029/94JC02113.

[10] GASPAR P, LABROUE S, OGOR F, et al. Improving nonparametric estimates of the sea state bias in radar altimeter measurements of sea level [J]. Journal of Atmospheric and Oceanic Technology, 2002, 19(10): 1690-1707. doi: 10.1175/1520-0426(2002)019<1690:INEOTS>2.0.CO;2.

[11] WANG J X and XIAO Q X. Local polynomial estimation of time-dependent diffusion parameter for discretely observed SDE models[J]. Filomat, 2014, 28(4): 871-878. doi: 10.2298/FIL1404871W.

[12] SU L and ULLAH A. Local polynomial estimation of nonparametric simultaneous equations models[J]. Journal of Econometrics, 2008, 144(1): 193-218. doi:10.1016/j.jeconom.2008.01.002.

[13] DETTMERING D, SCHWATKE C, and BOSCH W. Global calibration of SARAL/AltiKa using multi-mission sea surface height crossovers[J]. Marine Geodesy, 2015, 38(10): 206-218. doi: 10.1080/01490419.2014.988832.

[14] PAIGE C C and SAUNDERS M A. LSQR: An algorithm for sparse linear-equations and sparse least-squares[J]. Acm Transactions on Mathematical Software, 1982, 8(1): 43-71. doi: 10.1145/355984.355989.

[15] FONG D C L and SAUNDERS M. LSMR: An iterative algorithm for sparse least-squares problems[J]. Siam Journal on Scientific Computing, 2010, 33(5):2950-2971. doi: 10.1137/10079687X.

[16] TRAN N, VANDEMARK D, LABROUE S, et al. Sea state bias in altimeter sea level estimates determined by combining wave model and satellite data[J]. Journal of Geophysical Research-Oceans, 2010, 115(C03020):1-7. doi: 10.1029/2009JC005534.

[17] FENG H, YAO S, LI L Y, et al. Spline-based nonparametric estimation of the altimeter sea-state bias correction[J]. IEEE Geoscience and Remote Sensing Letters, 2010, 7(3): 577-581. doi: 10.1109/LGRS.2010.2041894.

**Jiang Maofei:** Born in 1989, Ph.D. candidate. Research interests include radar altimeter data processing.

**Xu Ke:** Born in 1963, Ph.D., Professor, Ph.D. supervisor. Main research interests include spaceborne radar altimeter system technology, synthetic aperture radar altimeter system technology, and signal processing.

**Liu Yalong:** Born in 1985, Ph.D. Research interests include radar altimeter data processing.

**Wang Lei:** Born in 1986, Ph.D. Research interests include radar altimeter signal processing.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv —Machine translation. Verify with original.*