

## Improving Nonparametric Estimates of the Sea State Bias in Radar Altimeter Measurements of Sea Level

PHILIPPE GASPAR, SYLVIE LABROUE, AND FRANÇOISE OGOR

*Collecte Localisation Satellites, Space Oceanography Division, Ramonville, France*

GUILLAUME LAFITTE, LAURENCE MARCHAL, AND MAGALI RAFANEL

*Institut National des Sciences Appliquées, Département Génie Mathématique et Modélisation, Toulouse, France*

(Manuscript received 22 October 2001, in final form 6 March 2002)

### ABSTRACT

A fully nonparametric (NP) version of the sea state bias (SSB) estimation problem in radar altimetry was first presented and solved by Gaspar and Florens (GF) using the statistical technique of kernel smoothing. This solution requires solving a large linear system and thus comes with a significant computational burden. In addition, examination of GF SSB estimates reveals a marked bias close to the boundaries of the estimation domain. This paper presents efforts to improve both the skill and the computational efficiency of the GF SSB estimation method. Computational efficiency is rather easily improved by an appropriate kernel choice that transforms the linear system to be solved into a very sparse system for which fast solution algorithms exist. The estimation bias proves to be due to the GF choice of a rudimentary NP estimator for conditional expectations. Use of a more elaborate estimator appears to be possible after a slight adaptation of the method. This solves the bias problem. Further improvement of the estimation skill is obtained by a local tuning of the kernel bandwidth. The refined estimation method is finally used to obtain a new NP estimate of the TOPEX SSB. This estimate yields larger SSB values than most previous estimates, in better agreement with recent in situ observations.

### 1. Introduction

It has long been known that the sea level measured by radar altimeters is lower than the true sea level because the backscattered power per unit area is greater for the wave troughs than for the wave crests (Yaplee et al. 1971). This effect is known as the electromagnetic bias. Other, often smaller, sea state-related biases affect the altimeter range measurements (e.g., Chelton et al. 1989). These are generally combined with the electromagnetic bias to form the sea state bias (SSB). This bias ranges between a few centimeters and a few decimeters. Its physical modeling remains a theoretical challenge (e.g., Rodriguez et al. 1992; Elfouhaily et al. 1999, 2000). Therefore, its estimation still largely relies on empirical models, calibrated on the altimeter data themselves (e.g., Born et al. 1982; Zlotnicki et al. 1989; Fu and Glazman 1991; Ray and Koblinsky 1991; Gaspar et al. 1994; Chelton 1994; Glazman et al. 1994). These

models assume that the SSB can be assigned a parametric form:

$$\text{SSB} = \varphi(\mathbf{x}, \theta), \quad (1)$$

where  $\varphi$  is an a priori specified function of  $\mathbf{x}$ , a vector of pertinent sea state-related variables, and  $\theta$ , a vector of parameters. A simple polynomial form of  $\varphi$  is generally assumed and the components of  $\mathbf{x}$  are always chosen among the few sea state-related parameters that can be directly measured by an altimeter: the significant wave height (SWH), the surface wind speed ( $U$ ), the backscatter coefficient ( $\sigma_0$ ), or any combination thereof. A least squares fitting procedure is then used to estimate the value of  $\theta$  that minimizes the variance of the measured sea surface height (SSH) differences at crossover points or along collinear tracks. However, Gaspar and Florens (1998, hereafter referred to as GF) demonstrated that so-fitted models are not true least squares approximation of the SSB. They differ from it by typically a few millimeters to a few centimeters, depending on sea state. This problem proves to be a consequence of the inevitably imperfect specification of the model's parametric form that corrupts the fitting process performed on SSH differences rather than directly on SSH measurements.

*Corresponding author address:* Dr. Philippe Gaspar, Collecte Localisation Satellites, Space Oceanography Division, 8-10 rue Hermes, Parc Technologique du Canal, Ramonville Cedex 31526, France.  
E-mail: philippe.gaspar@cls.fr

To avoid specification of a parametric formulation of the SSB model, GF reformulated and solved the SSB estimation problem in a totally nonparametric way based on the statistical technique of kernel smoothing. This method was used to compute the first fully nonparametric (NP) estimate of the TOPEX SSB as a function of  $U$  and SWH. This estimate was obtained using (i) a very limited dataset, (ii) an elementary NP estimator for conditional expectations, and (iii) a crude selection of smoothing parameters. Still, it appears to be better than the previously published parametric estimates.

Restriction to a limited dataset was justified on the ground of computational limitations. The solution of the NP SSB estimation problem indeed requires solving a large linear system where size is proportional to the number of observations used for estimation. Therefore, GF only used a small fraction (less than 10%) of the available data to obtain a TOPEX SSB estimate while keeping the computation time within reasonable limits. This demonstrated the pertinence of the NP approach, but further analysis or refinement of the results was not attempted in such a computationally limited context. This limitation is especially frustrating as the SSB is a small signal to be extracted from altimetric measurements that contain other much stronger signals acting as a large noise in the estimation process. Use of the largest possible number of data is thus highly desirable to help reduce the impact of the noise without having to unduly increase the amount of smoothing.

Another weak point of the GF solution is the use of the so-called Nadaraya–Watson (NW) kernel estimator for conditional expectations. Use of this estimator, actually an elementary locally weighted average, eased the derivation and implementation of the GF solution. But the NW estimator is known to yield significantly biased estimates in regions where the data density varies rapidly (e.g., Chu and Marron 1991). This problem likely affects the GF SSB estimation since altimetric observations are far from homogeneously distributed in the  $(U, \text{SWH})$  plane. In particular, large data density gradients are present close to the boundaries of the estimation domain.

This paper describes our efforts to improve the skill of the GF SSB estimation method and to overcome its computational limitations. In section 2, we briefly revisit the GF SSB estimation method and show that it is easily generalized to accommodate a wide range of NP estimators of conditional expectations among which a replacement for the NW estimator can be chosen. We present, in section 3, the numerical setup and the datasets used for testing improvements of the estimation method. Methodological improvements are theoretically investigated and numerically tested in section 4. The best estimator is finally used to obtain a new NP estimation of the TOPEX SSB, which is analyzed in section 5.

## 2. Revisiting nonparametric estimation of the sea state bias

### a. Principle

Statistical estimation of the SSB is classically based on the following observation equation:

$$\text{SSH}'_2 - \text{SSH}'_1 = (\text{SSB}_2 - \text{SSB}_1) + \varepsilon, \quad (2)$$

where indices 1 and 2 denote measurements taken at different times  $t_1$  and  $t_2$  but at the same geographic location;  $\text{SSH}'$  is the sea surface height above the reference ellipsoid, not corrected for the SSB; and  $\varepsilon$  is a zero-mean error term that includes altimetric measurement errors, residual geophysical correction errors, and the dynamic topography variation between  $t_1$  and  $t_2$ . The observations  $\text{SSH}'_2 - \text{SSH}'_1$  can be either crossover or collinear differences.

In this section, we briefly revisit, and slightly generalize, the method devised by GF to nonparametrically estimate the SSB based on series of altimetric observations (2). As a starting point, GF simply assume that the SSB is a nonspecified function  $\varphi$  of  $\mathbf{x}$ , a vector of  $p$  pertinent variables. In practice, GF use  $\mathbf{x} = (U, \text{SWH})$  but the theory holds for any set of sea state-related variables. Setting  $y = \text{SSH}'_2 - \text{SSH}'_1$ , (2) can be re-written:

$$y = \varphi(\mathbf{x}_2) - \varphi(\mathbf{x}_1) + \varepsilon, \quad (3)$$

so that the conditional expectation of  $y$ , given  $\mathbf{x}_2 = \mathbf{x}$ , is

$$E[y | \mathbf{x}_2 = \mathbf{x}] = \varphi(\mathbf{x}) - E[\varphi(\mathbf{x}_1) | \mathbf{x}_2 = \mathbf{x}]. \quad (4)$$

Nonparametric estimation theory provides a variety of estimators of the regression function  $r(\mathbf{x}) = E[\zeta | \mathbf{x}]$ , where  $\zeta$  is any random scalar variable jointly distributed with  $\mathbf{x}$ , (see, e.g., Härdle 1990). Gaspar and Florens directly selected the simple Nadaraya–Watson kernel estimator (Nadaraya 1964; Watson 1964). But one can remain more general and simply assume that the selected estimator can be written under the form

$$\hat{r}(\mathbf{x}) = \sum_{i=1}^n \zeta_i \alpha_n(\mathbf{x}, \mathbf{x}_i), \quad (5)$$

where the weights  $\alpha_n$  have to satisfy

$$\sum_{i=1}^n \alpha_n(\mathbf{x}, \mathbf{x}_i) = 1. \quad (6)$$

Using such an estimator and a set of  $n$  altimetric crossover or collinear measurements  $(y_i, \mathbf{x}_{1i}, \mathbf{x}_{2i})$ , the conditional expectations in (4) can be estimated to yield

$$\varphi(\mathbf{x}) = \sum_{i=1}^n y_i \alpha_n(\mathbf{x}, \mathbf{x}_{2i}) + \sum_{i=1}^n \varphi(\mathbf{x}_{1i}) \alpha_n(\mathbf{x}, \mathbf{x}_{2i}). \quad (7)$$

Indices 1 and 2 still denote measurements taken at different times ( $t_1$  and  $t_2$ ) but at the same location, while index  $i$  simply denotes the  $i$ th measurement in the dataset. Equation (7) provides a practical means of esti-

mating  $\varphi(\mathbf{x})$ , for any  $\mathbf{x}$ , provided that the values of  $\varphi(\mathbf{x}_{1i})$  are known. To estimate these, one rewrites (7) for  $\mathbf{x}$  equal any of the  $\mathbf{x}_{1i}$ :

$$\varphi(\mathbf{x}_{1j}) = \sum_{i=1}^n y_i \alpha_n(\mathbf{x}_{1j}, \mathbf{x}_{2i}) + \sum_{i=1}^n \varphi(\mathbf{x}_{1i}) \alpha_n(\mathbf{x}_{1j}, \mathbf{x}_{2i})$$

$$\forall j = 1, \dots, n. \quad (8)$$

or, in matrix form:

$$(\mathbf{I} - \mathbf{A})\varphi_1 = \mathbf{Ay}, \quad (9)$$

where  $\mathbf{I}$  is the  $(n \times n)$  identity matrix,  $\mathbf{A}$  is the  $(n \times n)$  matrix with elements  $a_{ji} = \alpha_n(\mathbf{x}_{1j}, \mathbf{x}_{2i})$ ,  $\varphi_1^T = [\varphi(\mathbf{x}_{11}), \dots, \varphi(\mathbf{x}_{1n})]$ , and  $\mathbf{y}^T = [y_1, \dots, y_n]$ .

The estimation problem is now in a closed form, but the linear system (9) cannot be solved for  $\varphi_1$  since  $\mathbf{I} - \mathbf{A}$  is singular. Indeed, the definition of the weights (6) implies that the sum of elements, in each line of this matrix, is zero. The rank of  $\mathbf{I} - \mathbf{A}$  is thus equal to  $n - 1$ . This is the numerical expression of the obvious undetermination problem linked to observation equation (2): only SSB differences are observed so that the SSB can only be determined to within a constant. To eliminate this undetermination we have to impose (rather than estimate) one value of the SSB. In this case, we impose the value of one of the components of  $\varphi_1$ , say,  $\varphi(\mathbf{x}_{11})$  as the dataset can always be rearranged to have the selected component in first place. The constraint is thus

$$\varphi(\mathbf{x}_{11}) = \varphi_0, \quad (10)$$

so that (9) can be rewritten

$$\mathbf{B}_1 \varphi = \mathbf{Ay} - \mathbf{B}_0 \varphi_0, \quad (11)$$

where  $\varphi$  is the vector made of the  $n - 1$  components of  $\varphi_1$  that remain to be determined:  $\varphi^T = [\varphi(\mathbf{x}_{12}), \dots, \varphi(\mathbf{x}_{1n})]$  and  $(\mathbf{B}_0, \mathbf{B}_1)$  is a partition of  $\mathbf{I} - \mathbf{A}$ ,  $\mathbf{B}_0$  being simply the first column of  $\mathbf{I} - \mathbf{A}$ . We now have a system of  $n$  equations with  $n - 1$  unknowns. A linear least squares solution is immediately obtained:

$$\hat{\varphi} = (\mathbf{B}_1^T \mathbf{B}_1)^{-1} \mathbf{B}_1^T (\mathbf{Ay} - \mathbf{B}_0 \varphi_0). \quad (12)$$

This provides the missing estimates of  $\varphi(\mathbf{x}_{1i})$ ,  $i = 2, \dots, n$ . These can then be plugged into (7) to estimate  $\varphi(\mathbf{x})$  for any value of  $\mathbf{x}$ . The NP SSB estimation problem is thus formally solved. This solution is that of GF, with a slight generalization concerning the form (5) of the used NP estimator for the conditional expectations.

### b. Selection of a nonparametric estimator

As previously mentioned, GF used the simple Nadaraya–Watson kernel estimator of the regression function  $r(\mathbf{x}) = E[\zeta | \mathbf{x}]$ . This estimator determines  $\hat{r}(\mathbf{x})$  as a simple weighted average of the measurements  $\zeta_i$ . The weighting scheme is governed by a so-called Kernel function ( $K$ ) that assigns weights as a function of a scaled distance  $\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_i)$  between  $\mathbf{x}_i$  and  $\mathbf{x}$ . In its general multivariate form, the NW estimator reads

$$\hat{r}_{\text{NW}}(\mathbf{x}) = \frac{\sum_{i=1}^n \zeta_i K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}, \quad (13)$$

with

$$K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x}) = \frac{1}{|\mathbf{H}|} K[\mathbf{H}^{-1}(\mathbf{x}_i - \mathbf{x})], \quad (14)$$

where  $|\mathbf{H}|$  is the determinant of  $\mathbf{H}$ , the  $(p \times p)$  non-singular scaling matrix, generally called bandwidth matrix. Individual scaling of each component of  $\mathbf{x}_i - \mathbf{x}$  is usually applied so that  $\mathbf{H}$  is, in most practical cases, a diagonal matrix. The kernel function  $K$  is a nonnegative scalar function normalized to satisfy  $\int K(\mathbf{x}) d\mathbf{x} = 1$ . It is usually chosen to be a symmetric probability density function.

The complete specification of the NW estimator, or any other kernel estimator, is thus based on three elements: the formulation of the estimator [e.g., Eq. (13)], the choice of a Kernel function  $K$ , and the determination of the bandwidth matrix  $\mathbf{H}$ . As will be shown in section 4, careful selection of each of these three elements can help significantly improve the skill and the computational efficiency of the GF SSB estimation method.

## 3. Numerical setup

### a. Datasets

#### 1) TOPEX DATA

Improvements of the NP SSB estimation technique are conveniently checked using the TOPEX crossover dataset previously gathered by GF. This dataset of over 600 000 valid crossover differences, includes 100 cycles of TOPEX side A altimetric measurements obtained between 4 April 1993 and 15 April 1996. Data quality is homogeneous throughout this period. In particular, the dataset does not include the very first repeat cycles nor the end of the TOPEX side A operation period, during which a progressive degradation of the measurements was observed.

The crossover differences ( $y = \text{SSH}_2' - \text{SSH}_1'$ ) correspond to the descending arc measurement minus the ascending arc measurement, inside the same repeat cycle. Therefore, in the rest of this paper, index 1 (index 2) denotes measurements taken on the ascending (descending) arc. Each crossover difference is associated with the corresponding altimeter-derived wind speed and wave height measurements ( $U_1, \text{SWH}_1, U_2, \text{SWH}_2$ ), where  $U$  is computed using the modified Chelton–Wentz algorithm (Witter and Chelton 1991a). More details about this data set (hereafter denoted as TP dataset) and its quality control can be found in GF.

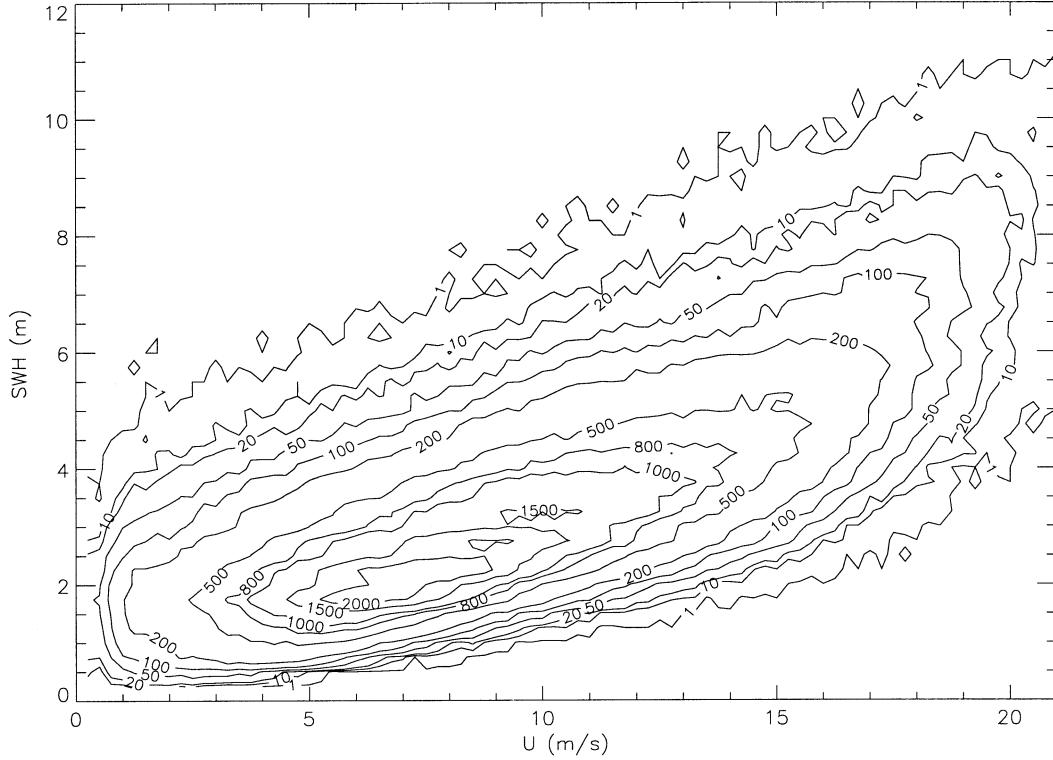


FIG. 1. Distribution of altimetric measurements in the  $(U, \text{SWH})$  plane. Measurements come from the TP dataset described in section 3a and are binned into  $(U, \text{SWH})$  boxes of width  $(0.25 \text{ m s}^{-1}, 0.25 \text{ m})$ . Isolines of the number of data per box are plotted.

## 2) SIMULATED DATA

The skill of SSB estimators cannot be precisely evaluated using the TP dataset alone. Indeed, classical diagnostics such as analysis of estimation residuals or of the explained variance can only be applied to crossover residuals, that is, to crossover differences corrected for the estimated SSB. As explained by GF, an SSB estimate that minimizes the variance of crossover residuals is not necessarily a true least squares approximation of the SSB. In addition, diagnostics on crossover residuals can only be related to wind speed and wave height differences ( $\Delta U, \Delta \text{SWH}$ ) at the crossover points, but not to one specific value of  $(U, \text{SWH})$ . Any problem detected in the residuals is thus difficult to trace back to the corresponding estimation problem in a well-identified zone of the  $(U, \text{SWH})$  plane.

To allow easier and more detailed evaluation of our new SSB estimates, we created a set of simulated crossover measurements containing only a known SSB signal plus some noise. SSB estimates obtained using this simulated (SIM) dataset can be directly compared to the known “true” SSB and estimation errors are directly diagnosed in the  $(U, \text{SWH})$  plane.

The SIM dataset is simply obtained by replacing, in the TP dataset, the actual crossover differences ( $\text{SSH}_2' - \text{SSH}_1'$ ) by simulated values  $z_2 - z_1$ :

$$z_2 - z_1 = B(U_2, \text{SWH}_2) - B(U_1, \text{SWH}_1) + w, \quad (15)$$

where  $w$  is a simulated noise, and  $B$  is an analytical expression of the SSB as a function of  $(U, \text{SWH})$ . For this purpose, we use the BM4 model of Gaspar et al. (1994) fitted over the TP dataset; that is,

$$\begin{aligned} B(U, \text{SWH}) &= \text{SWH}[-0.021 - 0.0035U + 0.00014U^2 \\ &\quad + 0.0027\text{SWH}]. \end{aligned} \quad (16)$$

For simulation realism, the standard deviation of the random noise  $w$  varies regionally and is taken to be the actual standard deviation of the TP crossover residuals, after correction for the BM4 SSB. This standard deviation varies from 3 cm in the quietest ocean regions up to about 20 cm in the most energetic areas.

### b. Numerical solution procedure

To solve (12), we first have to choose the measurement  $\mathbf{x}_{11}$ , where the constraint (10) is imposed. For the TP dataset, the only obvious choice is to impose that the SSB is zero over a flat sea with no wind, that is  $\varphi(0, 0) = 0$ . However, the data density close to  $(0, 0)$  is so weak (Fig. 1) that imposing the constraint in that neighborhood can create numerical problems. As GF suggested, a numerically effective procedure to satisfy this constraint is the following.

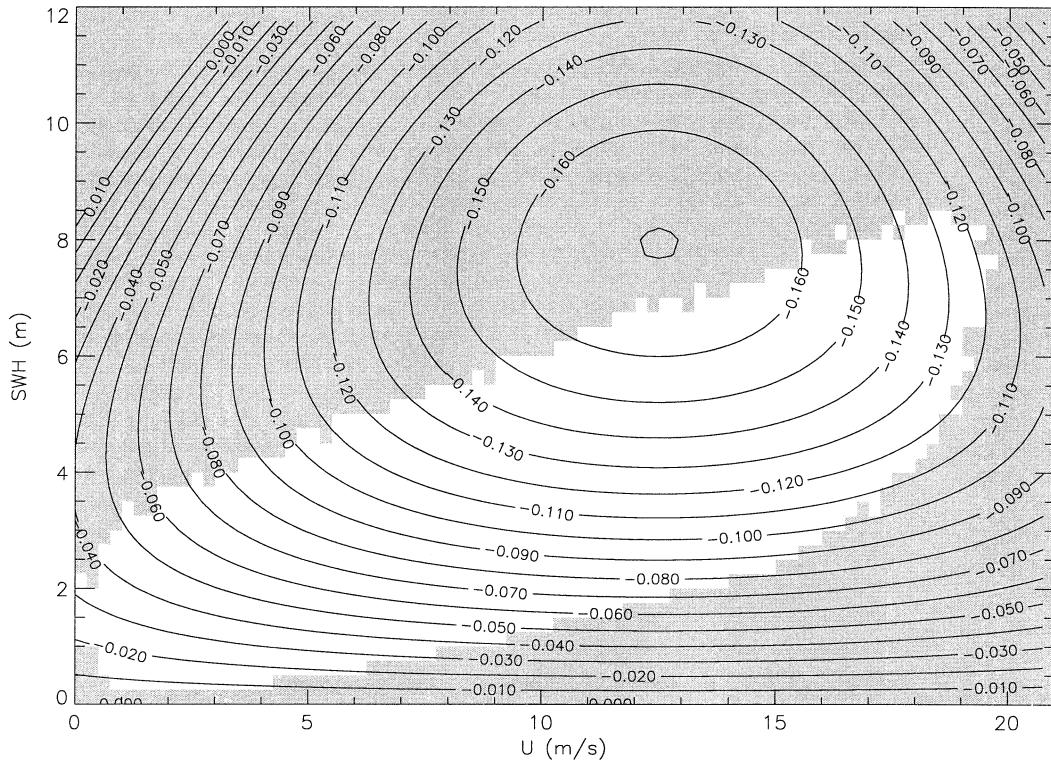


FIG. 2. BM4 SSB (in m) in the  $(U, \text{SWH})$  plane. The shaded area corresponds to data boxes containing less than 30 measurements of the TP dataset.

- 1) Select the ascending arc measurement  $\mathbf{x}_{11}$  closest to  $(U_m, \text{SWH}_m)$ , the mean of the  $(U, \text{SWH})$  distribution. In the TP dataset, this mean is  $8 \text{ m s}^{-1}, 2.7 \text{ m}$ .
- 2) Impose  $\varphi(\mathbf{x}_{11}) = \varphi_0$ , with any reasonable value of  $\varphi_0$ , for example,  $\varphi_0 = -5 \text{ cm}$ , and compute  $\hat{\varphi}$  using (12).
- 3) Then, using (7), compute  $\varphi(\mathbf{x})$  over a regular grid of  $(U, \text{SWH})$  values, including  $(0, 0)$ . We typically use a grid size of  $0.25 \text{ m s}^{-1}$  in  $U$  and  $0.25 \text{ m}$  in  $\text{SWH}$ .
- 4) Finally, subtract  $\varphi(0, 0)$  from  $\varphi(\mathbf{x})$  at all grid points. The so-shifted solution obviously satisfies  $\text{SSB}(0, 0) = 0$ .

For the SIM dataset, the problem is somewhat different as the SSB is known everywhere. Thus, the exact value of  $\varphi(\mathbf{x}_{11})$ , rather than a guess, can be imposed in step 2 of the above procedure. By doing so and skipping step 4, one obtains SSB estimates that are nearly exact close to the mean of the data distribution  $(U_m, \text{SWH}_m)$  but are not necessarily equal to zero at  $(0, 0)$ . Use of such nonshifted-to-zero solutions proves to be interesting for error analysis, as shown in section 3c.

#### c. Computing the final SSB estimate and its uncertainty

Even if we intend to significantly improve the numerical implementation of the NP SSB estimation meth-

od, we cannot reasonably envision solving (12) at once for the whole TP or SIM dataset, as it means solving a linear system with over 600 000 unknowns. We will thus follow the GF approach and compute SSB estimates, repeat cycle per repeat cycle. Unlike GF though, complete cycles will be used, without data subsampling. This requires solving a linear system with typically 6000 unknowns for each cycle. Then the final SSB estimate is obtained as the average of the individual cycle estimates computed at each grid point  $\mathbf{x}$ . In the usual case where the solution is shifted to zero, the final solution is thus

$$\text{SSB}(\mathbf{x}) = \bar{\varphi}(\mathbf{x}) - \bar{\varphi}(0, 0), \quad (17)$$

where the overbar denotes an average over individual cycle estimates. Assuming independence of the  $m$  individual cycle estimates, the variance of the final SSB estimate is given by

$$\begin{aligned} \sigma_{\text{SSB}}^2(\mathbf{x}) &= m^{-1} \{ \sigma_\varphi^2(\mathbf{x}) + \sigma_\varphi^2(0, 0) \\ &\quad - 2 \text{ cov}[\varphi(\mathbf{x}), \varphi(0, 0)] \}, \end{aligned} \quad (18)$$

where  $\sigma_\varphi^2(\mathbf{x})$  is the variance of the nonshifted cycle estimates  $\varphi(\mathbf{x})$ . This expression is specially interesting as it clearly shows the impact of the solution-shifting procedure on the uncertainty of the final SSB estimate. This procedure actually adds the variance of the shift estimate  $\bar{\varphi}(0, 0)$  to the variance of the nonshifted solution. This

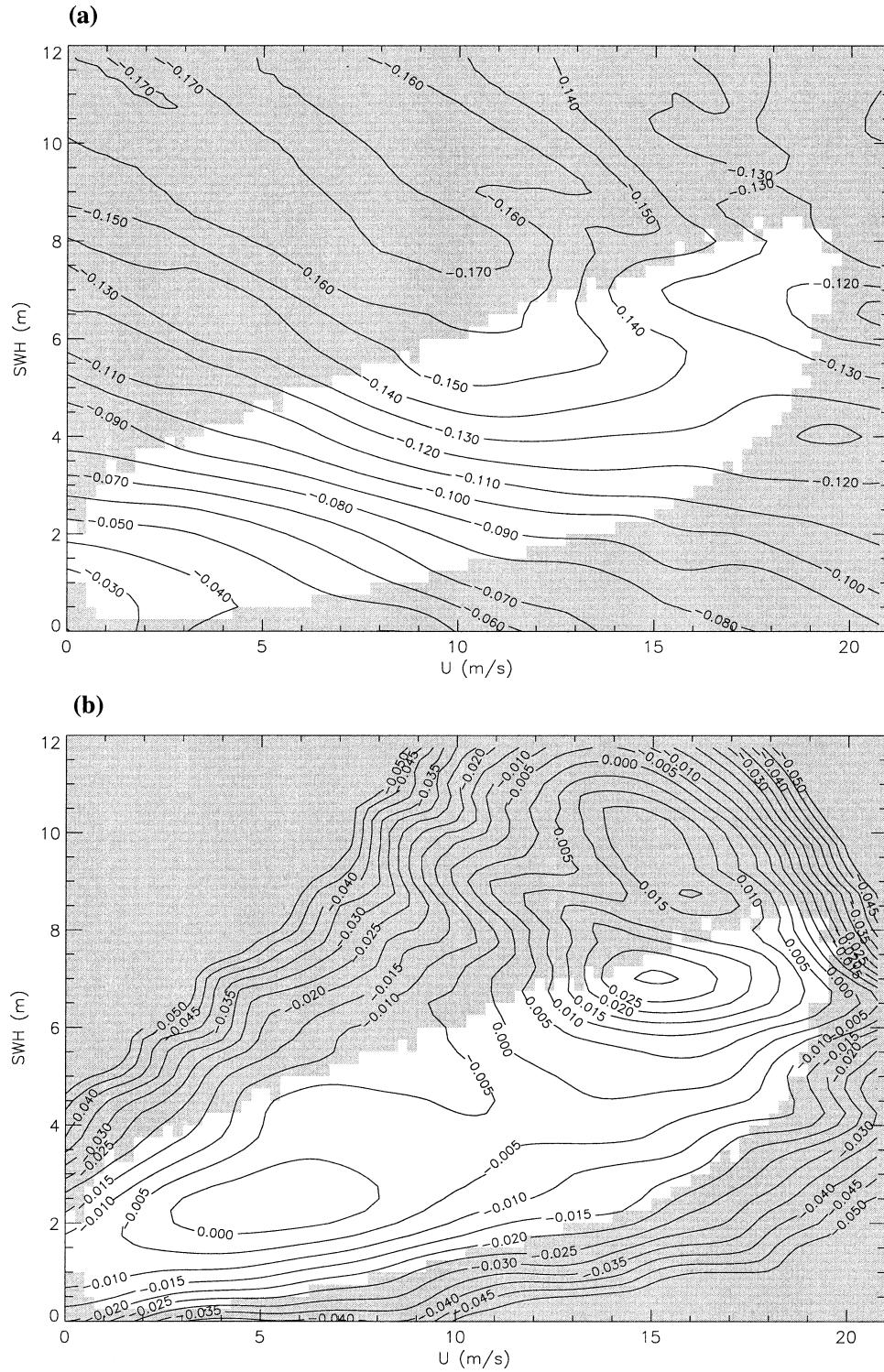


FIG. 3. (a) SSB estimate (in m) obtained in EXNW and (b) the corresponding estimation error, i.e., the difference between this estimate and the exact BM4 solution (16).

TABLE 1. Technical characteristics of the different estimation experiments.

Expt	Dataset	Solution shift	Estimator	Kernel	Bandwidth
EXNW	Subsampled SIM	No	NW	Gaussian	$h_U = 1 \text{ m s}^{-1}$ $h_{\text{SWH}} = 0.4 \text{ m}$
EXLLR1	Subsampled SIM	No	LLR	Gaussian	$h_U = 1 \text{ m s}^{-1}$ $h_{\text{SWH}} = 0.4 \text{ m}$
EXLLR2	Subsampled SIM	No	LLR	Epanechnikov	$h_U = 2.2 \text{ m s}^{-1}$ $h_{\text{SWH}} = 0.9 \text{ m}$
EXLLR3	Complete SIM	No	LLR	Epanechnikov	$h_U = 2 \text{ m s}^{-1}$ $h_{\text{SWH}} = 0.9 \text{ m}$
EXLLR4	Complete SIM	No	LLR	Epanechnikov	Local Eq. (33)
EXTP	Complete TP	Yes	LLR	Epanechnikov	Local Eq. (33)

added variance is constant, but unfortunately large as the estimation uncertainty is large in the data-poor region near  $(0, 0)$ . An additional covariance term is also present but generally small compared to  $\sigma_\varphi^2(0, 0)$ , except for values of  $\mathbf{x}$  close to  $(0, 0)$ .

In practice, as the variance of the final shifted solution (18) is a rather convoluted diagnosis including both estimation and shifting effects, it is useful to separately analyze the variance of the nonshifted estimate:  $m^{-1}\sigma_\varphi^2(\mathbf{x})$ . This variance is a direct measure of the estimator skill unaffected by shifting effects, even if its value at  $(0, 0)$  is the constant variance added by the shifting process:  $m^{-1}\sigma_\varphi^2(0, 0)$ .

#### d. Validity of the results

Any statistical estimate of the SSB as a function of  $(U, \text{SWH})$  is meaningful only in the zone of  $(U, \text{SWH})$  plane where a sufficient number of observations exists. To help readers visualize this zone, the data boxes (as defined in Fig. 1) containing less than 30 measurements will be shaded in all figures presenting results in the  $(U, \text{SWH})$  plane. As an example, Fig. 2 shows such a display of the BM4 SSB (16). Note that the shaded area contains slightly over 1% of the TP (and SIM) data.

## 4. Toward an improved nonparametric estimation of the SSB

#### a. Characterizing errors of the GF SSB estimates

To precisely characterize errors likely to be present in the GF estimate of the TOPEX SSB, we first estimated the SSB of the SIM dataset in exactly the technical conditions as GF estimated the TOPEX SSB. We thus subsampled SIM to retain only 500 measurements per repeat cycle and used the NW estimator with the same Gaussian kernel and bandwidth as GF.

$$K_H(\mathbf{x} - \mathbf{x}_i)$$

$$= \frac{1}{2\pi h_U h_{\text{SWH}}} \exp \left[ \frac{-(U - U_i)^2}{2h_U^2} - \frac{(\text{SWH} - \text{SWH}_i)^2}{2h_{\text{SWH}}^2} \right], \quad (19)$$

with  $h_U = 1 \text{ m s}^{-1}$  and  $h_{\text{SWH}} = 0.4 \text{ m}$ . This first experiment is noted EXNW. Its numerical characteristics are summarized in Table 1, like the characteristics of all subsequent experiments. The only difference between EXNW and the GF solution procedure is that we did not shift the solution to zero but rather imposed the exact value of BM4 at the ascending arc measurement closest to the mean of the  $(U, \text{SWH})$  distribution. As previously explained, this allows a clear identification of the errors associated with the estimation procedure with no superimposed effect due to the shifting procedure.

The EXNW solution and its error (difference from the simulated BM4) are shown in Figs. 3a and 3b, respectively. The error is disappointingly large, in the range  $[-3 \text{ cm}, 3 \text{ cm}]$ . More worrisome is the fact that this error is far from random: it exhibits well-organized variations especially close to the boundaries of the nonshaded region. This appears to be a manifestation of the known bias of the NW estimator in regions of the estimation domain where the data density varies rapidly.

To understand the origin of this bias one must realize that, from a function approximation point of view,  $\hat{r}_{\text{NW}}(\mathbf{x})$  is the value of the constant that locally best fits  $r(\mathbf{x})$ , in a weighted least squares sense. Indeed, basic calculus shows that  $\hat{r}_{\text{NW}}(\mathbf{x})$  as defined by (13) is the value of the constant  $r_0$  that minimizes

$$\sum_{i=1}^n (\zeta_i - r_0)^2 K_H(\mathbf{x}_i - \mathbf{x}). \quad (20)$$

The NW estimator is therefore called a local constant estimator. A useful measure of the error associated to this, or any other, regression estimator  $\hat{r}(\mathbf{x})$  is the pointwise mean-square error (MSE):

$$\text{MSE}(\mathbf{x}) = E\{[\hat{r}(\mathbf{x}) - r(\mathbf{x})]^2 | X\}, \quad (21)$$

where  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ . The MSE is easily decomposed under the form

$$\text{MSE}(\mathbf{x}) = \{E[\hat{r}(\mathbf{x}) | X] - r(\mathbf{x})\}^2 + \text{var}[\hat{r}(\mathbf{x}) | X]. \quad (22)$$

The first component of the MSE is often referred to as the square of the bias and the second as the variance of

the regression estimator. In the one-dimensional case (i.e., for a scalar  $x$ ), the leading terms in the asymptotic

expansion of the bias and variance of the NW estimator are (e.g., Fan 1992)

$$\text{bias}[\hat{r}_{\text{NW}}(x)] = h^2 \left( \frac{1}{2} r''(x) + \frac{r'(x)f'(x)}{f(x)} \right) \int_{-\infty}^{+\infty} u^2 K(u) du \quad (23)$$

$$\text{variance}[\hat{r}_{\text{NW}}(x)] = \frac{\sigma^2(x)}{f(x)nh} \int_{-\infty}^{+\infty} K^2(u) du, \quad (24)$$

where  $f$  is the design density, that is, the density of the random variable  $X$ ;  $h$  is the scalar bandwidth and  $\sigma^2(x)$  is the conditional variance of  $\zeta$ ; that is,  $\text{var}(\zeta | X = x)$ . Equivalent results exist in the multivariate case (Ruppert and Wand 1994), but the one-dimensional version suffices here to highlight the basic problem of the NW estimator. Indeed, the bias equation (23) shows that the NW estimator contains a natural smoothing error proportional to  $r''$  (the well-known peak and trough flattening effect of any smoother) plus a spurious error term  $(r'f'/f)$  due to the local constant fit. This term simply reflects the fact that, in the presence of a slope in  $r$  and in the data density, a constant fit will be biased toward the  $r$  values found in the most densely populated zone of the local estimation window. Use of a local linear approximation of  $r$  instead of a local constant approximation is a known, simple and elegant, solution to this problem (e.g., Chu and Marron 1991).

### b. Improving the estimator formulation

Let us thus assume a local linear model for the regression function  $r$ :

$$r(\mathbf{z}) = \beta_0 + \sum_{j=1}^p \beta_j(z_j - x_j), \quad (25)$$

with  $\mathbf{z}$  in the neighborhood of  $\mathbf{x}$ . The value of  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$  that minimizes

$$\sum_{i=1}^n \left[ \zeta_i - \beta_0 - \sum_{j=1}^p \beta_j(x_{ij} - x_j) \right]^2 K_H(\mathbf{x}_i - \mathbf{x}) \quad (26)$$

is the well-known solution of the weighted least squares regression problem,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D)^{-1} \mathbf{X}_D^T \mathbf{W} \zeta, \quad (27)$$

where  $\mathbf{W} = \text{diag}\{K_H(\mathbf{x}_i - \mathbf{x})\}$  is the  $(n \times n)$  matrix of the weights, and

$$\mathbf{X}_D = \begin{pmatrix} 1 & x_{11} - x_1 & \cdots & x_{1p} - x_p \\ 1 & x_{21} - x_1 & \cdots & x_{2p} - x_p \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} - x_1 & \cdots & x_{np} - x_p \end{pmatrix}.$$

The so-called local linear regression (LLR) estimator of

$r(\mathbf{x})$  is  $\hat{\beta}_0$ , while the other components of  $\hat{\boldsymbol{\beta}}$  are estimators of the first-order partial derivatives of  $r$  (Fan 1992). One can thus write

$$\hat{r}_{\text{LLR}}(\mathbf{x}) = \hat{\beta}_0 = \mathbf{e}_1^T (\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D)^{-1} \mathbf{X}_D^T \mathbf{W} \zeta, \quad (28)$$

where  $\mathbf{e}_1$  is the  $(p \times 1)$  unit vector:  $\mathbf{e}_1^T = (1, 0, \dots, 0)$ . Obviously,  $\hat{r}_{\text{LLR}}$  conforms to (5) and, with a little algebra, it can be shown to satisfy (6). This estimator is thus directly usable in the SSB estimation procedure as a replacement for  $\hat{r}_{\text{NW}}$ . The asymptotic bias and variance of  $\hat{r}_{\text{LLR}}$  are (Fan 1992)

$$\text{bias}[\hat{r}_{\text{LLR}}(x)] = \frac{h^2}{2} r''(x) \int_{-\infty}^{+\infty} u^2 K(u) du \quad (29)$$

$$\text{variance}[\hat{r}_{\text{LLR}}(x)] = \frac{\sigma^2(x)}{f(x)nh} \int_{-\infty}^{+\infty} K^2(u) du. \quad (30)$$

This shows that the spurious  $r'f'/f$  term has been eliminated from the bias expression with strictly no impact on the asymptotic variance, an excellent result that also holds in the multivariate case (Ruppert and Wand 1994). This improved estimator comes with a rather elaborate weighting scheme (28) involving tedious matrix multiplications and the inversion of  $\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D$ , a  $[(p+1) \times (p+1)]$  matrix. In our case, however,  $p = 2$  so that the additional time needed to compute weights only represents a small fraction of the total computation time. The bulk of the computational effort remains on the least squares solution of the large linear system (11).

To test  $\hat{r}_{\text{LLR}}$  we performed a new numerical experiment (EXLLR1) in the same conditions as EXNW except that  $\hat{r}_{\text{NW}}$  was replaced by  $\hat{r}_{\text{LLR}}$ . The resulting SSB estimate and its error are shown in Fig. 4. The estimation error is considerably smaller than in EXNW, generally below 1 cm in the nonshaded area, except in the vicinity of  $(0, 0)$ . This error no longer exhibits well-organized variations, as in EXNW. This confirms that use of  $\hat{r}_{\text{LLR}}$  largely solves the bias problem, leaving essentially random errors which magnitude increases as the local data density decreases. Thus, either more data or more smoothing is requested to further reduce the estimation error.

### c. Improving the kernel choice

More data are at hand. But, so far, the number of data actually used has been limited for computational rea-

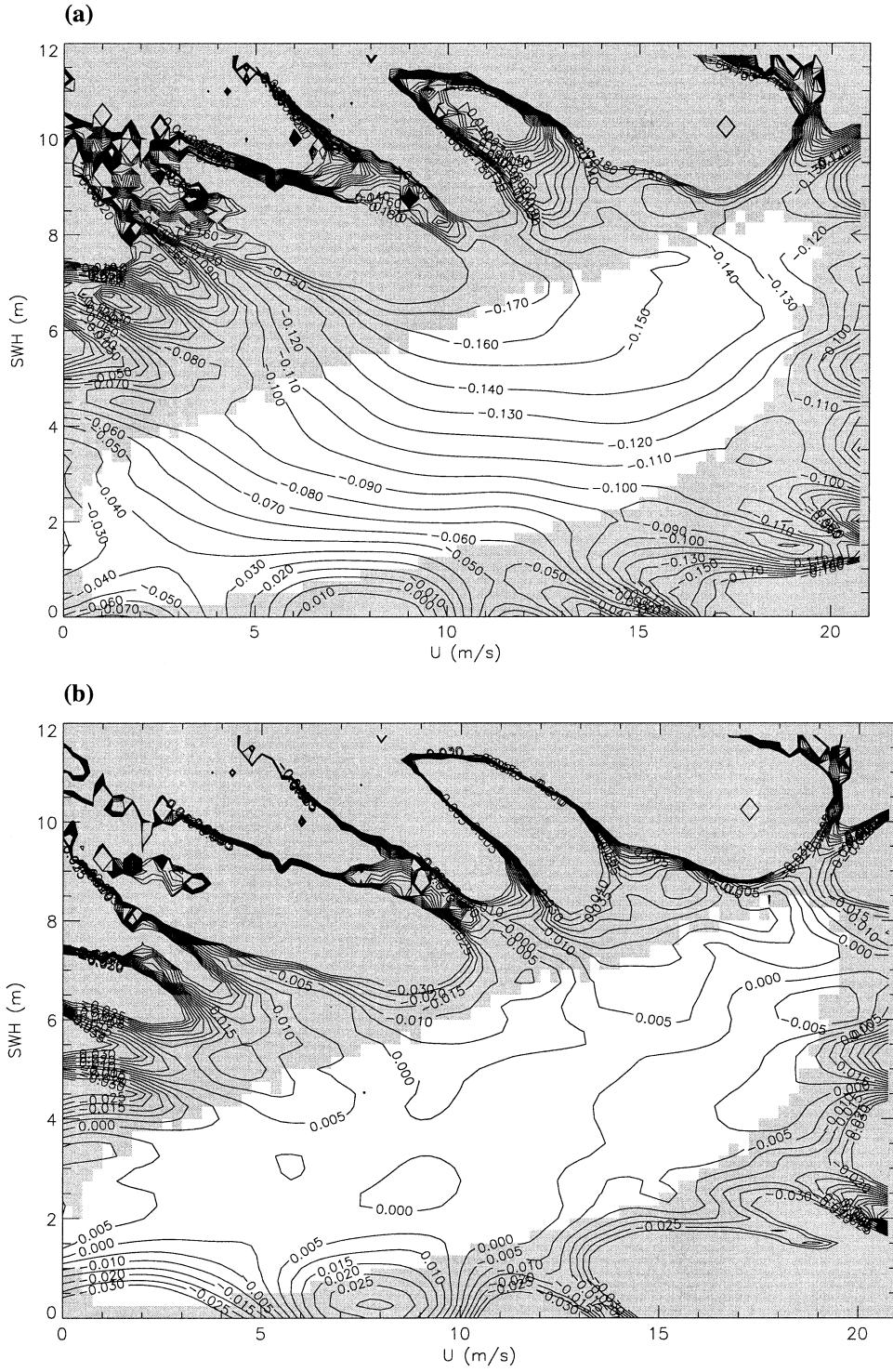


FIG. 4. (a) SSB estimate (in m) obtained in EXLLR1 and (b) the corresponding estimation error.

sons. Indeed, to solve the GF SSB estimation method requires, in the least squares sense, a linear system (11) in which size is proportional to the number of data used. Gaspar and Florens implemented a classical solution

based on the Cholesky decomposition of the  $\mathbf{B}_1^T \mathbf{B}_1$  matrix. Other numerical solution techniques might prove to be slightly more efficient, but the real breakthrough comes from recognizing that  $\mathbf{B}_1$  is almost a sparse ma-

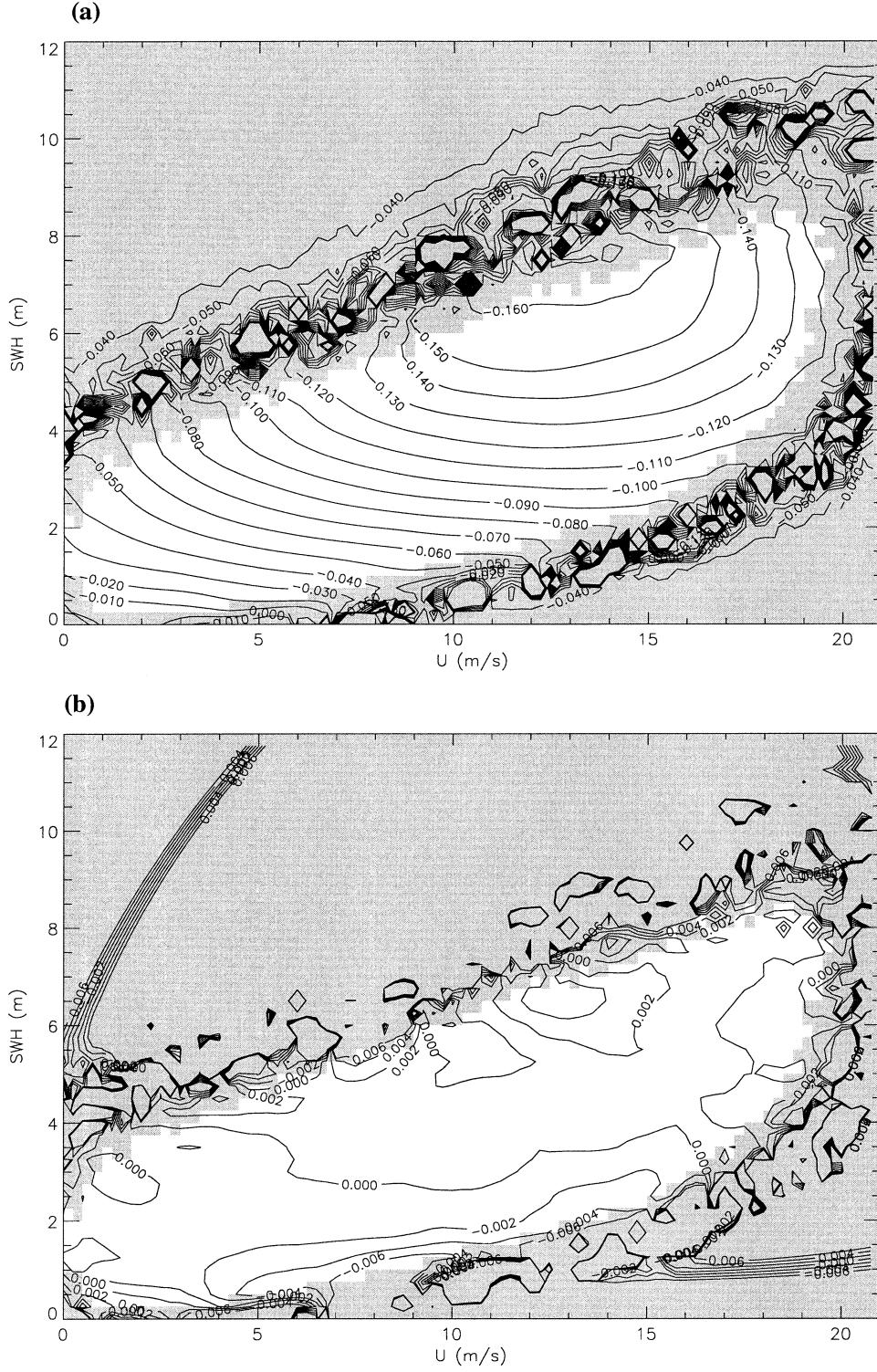


FIG. 5. (a) SSB estimate (in m) obtained in EXLLR3 and (b) the corresponding estimation error.

trix. Indeed, the matrix coefficients are the weights  $\alpha_n(\mathbf{x}_{1j}, \mathbf{x}_{2i})$ . Many of them concern far apart measurements  $\mathbf{x}_{1j}$  and  $\mathbf{x}_{2i}$  and are thus very small, but never exactly zero in the case of a Gaussian kernel. However,

other popular kernels have a compact support and thus yield truly zero weights when the distance between  $\mathbf{x}_{1j}$  and  $\mathbf{x}_{2i}$  exceeds the bandwidth. Since the few commonly used kernels have very similar statistical properties

(e.g., Simonoff 1996), replacing one by another is known to have little effect on the skill of the estimator provided that the bandwidth is correspondingly adapted (see next section). We can thus readily replace the Gaussian kernel by a kernel with a compact support. The obvious choice is the spherical Epanechnikov kernel, which, under certain conditions, proves to minimize the asymptotic MSE of the LLR estimator (Fan and Gijbels 1996). In our two-dimensional case, this kernel reads

$$\begin{aligned} K_H(\mathbf{x} - \mathbf{x}_i) \\ = \max \left\{ 0, \frac{2}{\pi h_u h_{SWH}} \right. \\ \times \left[ 1 - \left( \frac{U - U_i}{h_u} \right)^2 - \left( \frac{SWH - SWH_i}{h_{SWH}} \right)^2 \right] \left. \right\}. \end{aligned} \quad (31)$$

Using this kernel, (11) becomes a truly sparse linear system, for which very efficient iterative solution methods exist. We tested a number of these and finally selected the LSQR algorithm of Paige and Saunders (1982), which was specifically developed for large sparse linear systems to be solved in the least squares sense. This algorithm was implemented together with the Modified Sparse Row (MSR) storage scheme that allows compact storage of the  $\mathbf{B}_1$  matrix (for a description, see Saad 1996). The new version of the estimation software was then tested in a simple experiment (EXLLR2) analogous to EXLLR1 but with an Epanechnikov instead of a Gaussian kernel, and with adapted bandwidths:  $h_u = 2.2 \text{ m s}^{-1}$  and  $h_{SWH} = 0.9 \text{ m}$ . The reason for multiplying the bandwidth by a factor of 2.2 when switching from a Gaussian to an Epanechnikov kernel will be explained in the next section. The SSB estimate obtained in EXLLR2 (not shown here) is very close to that of EXLLR1. Differences between the two estimates are random, ranging from nearly zero close to the peak of the data distribution up to 4–5 mm at the boundaries of the nonshaded region where the estimation uncertainty is large due to data scarcity. From an estimation point of view, switching from a Gaussian to an Epanechnikov kernel is thus essentially a blank operation. From a computational point of view, however, the effect is quite impressive: on the same workstation EXLLR1 required over 2 h of CPU time, while EXLLR2 ran in less than 15 min. The storage gain is also significant as over 85% of the elements of  $\mathbf{B}_1$  are zero in the EXLLR2 experiment. Thanks to the achieved reduction in both computer time and storage requirements, the new software version allows us to reasonably consider solving the NP SSB estimation problem without subsampling repeat cycles, that is, using complete cycles, each containing 6 to  $7 \times 10^3$  valid crossovers. This is achieved in the next section, together with the bandwidth tuning.

#### d. Improving the bandwidth selection

Bandwidth selection in the context of multivariate NP regression is a difficult and still very active field of research (for an excellent overview, see the textbook of Fan and Gijbels 1996). Before trying to implement complex bandwidth determination procedures, we performed a few simple numerical experiments in which a first guess of the bandwidths was progressively refined to reduce the difference between the estimated SSB and its known exact value (BM4). These experiments were performed using the complete SIM dataset, the LLR estimator and the Epanechnikov kernel. After a few trials we obtained a very satisfactory estimate using  $h_u = 2 \text{ m s}^{-1}$  and  $h_{SWH} = 0.9 \text{ m}$ . This estimate (EXLLR3) and its error are shown in Fig. 5. The estimation error is close to zero in most of the nonshaded zone. It reaches values close to, or slightly above, 2 mm near the boundaries of that zone but then increases very rapidly when entering the shaded zone. The estimation uncertainty at  $(0, 0)$  is close to 1 cm. This rapid increase of the estimation error as one enters the shaded zone is essentially due to the compact support of the Epanechnikov kernel. Indeed, this kernel provides markedly less smoothing than a Gaussian kernel in data-poor regions as only the very few measurements inside the kernel support area (i.e., inside the bandwidth) have a nonzero weight and hence contribute to smoothing.

Even if these results are satisfying over most of the data distribution, one is tempted to find a way to reduce the rapid error growth in the data-sparse regions, especially close to  $(0, 0)$ . Additional smoothing would certainly help, but the use of unnecessarily large bandwidths should be avoided in data-rich regions as the estimator bias increases with the bandwidth (29). We are thus naturally led to consider use of locally variable bandwidth instead of a single global bandwidth, a quite common approach (Ruppert et al. 1995). One theoretical option for determining a locally optimal bandwidth is to look for the bandwidth that minimizes the asymptotic MSE. Using (22), (29), and (30), it is easily demonstrated that, in the one-dimensional case, the value of  $h$  that minimizes the asymptotic MSE of  $\hat{r}_{LLR}$  is

$$h_{opt}(x) = \left[ C(K) \frac{\sigma^2(x)}{nf(x)\{r''(x)\}^2} \right]^{1/5}, \quad (32)$$

where  $C(K)$  is a kernel-dependent constant:

$$C(K) = \int_{-\infty}^{+\infty} K^2(u) du \left\{ \int_{-\infty}^{+\infty} u^2 K(u) du \right\}^{-2}.$$

The interpretation of (32) is simple. A larger local variance  $\sigma^2(x)$  or a smaller local data density  $nf(x)$  requires more smoothing and hence a larger bandwidth. A larger local curvature  $r''(x)$  requires a smaller bandwidth to minimize the peak and trough flattening effect of smoothing. For a Gaussian and an Epanechnikov

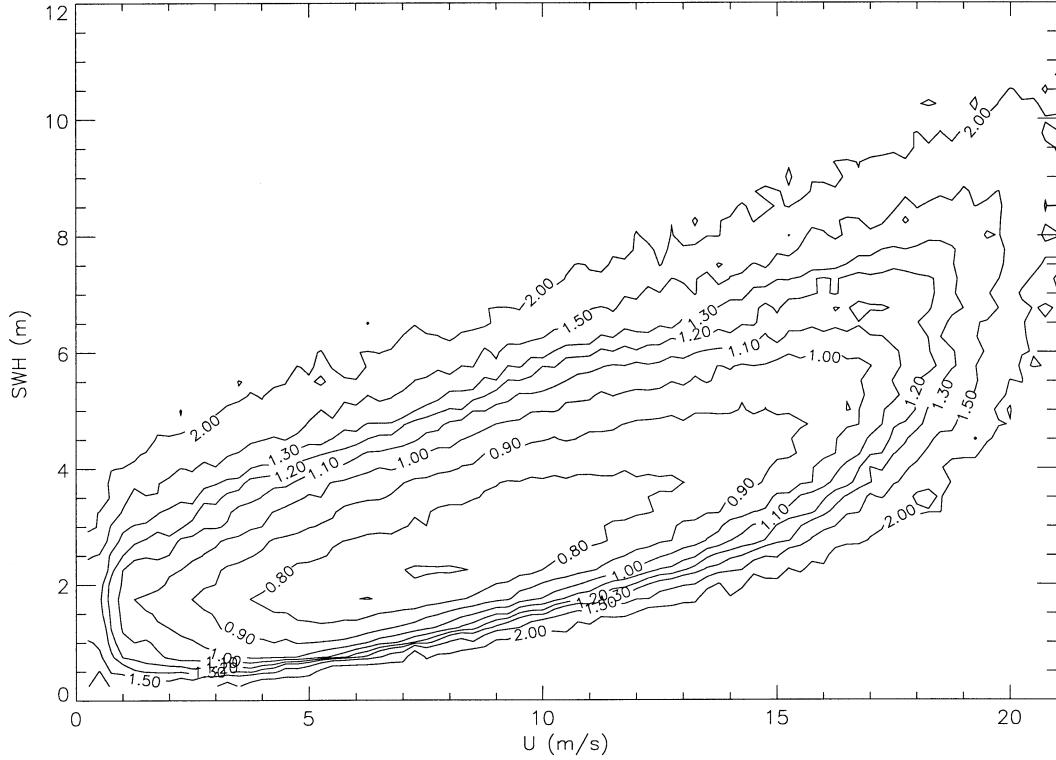


FIG. 6. Modulation factor for the local bandwidth as a function of data density  $[n(x)/\bar{n}]^{-1/6}$ .

kernel, the value of the constant  $C^{1/5}(K)$  is 0.776 and 1.719, respectively. This explains why we multiplied the bandwidth by a factor of 2.2 when switching from a Gaussian to an Epanechnikov kernel in EXLLR2. Beyond such simple proportionality rules, practical use of (32) requires considerable work as  $\sigma^2(x)$ ,  $f(x)$ , and  $r''(x)$  are, a priori, unknown. One approach is to replace these unknowns by some estimates, leading to the so-called plug-in bandwidth selection technique. This is not a simple technique: the estimation of  $r''(x)$  only is already more difficult than the original problem of estimating  $r$ . One can, however, take advantage of expressions like (32) to design simple local bandwidth schemes that have, at least, some properties of the optimal choice. In our case, we wish to modulate the bandwidth as a function of the local data density. Equation (32) indicates that, in the one-dimensional case, the optimal modulation factor is  $f(x)^{-1/5}$ . The generalization to the multivariate case can be found in Fan and Gijbels (1996). It shows that the optimal bandwidth is actually proportional to  $f(x)^{-1/(p+4)}$ . A simple proxy for the local density  $f(x)$  is the number of measurements per  $(U, \text{SWH})$  box, as shown in Fig. 1. This suggests the following simple local bandwidth rule in our two-dimensional problem:

$$(h_U, h_{\text{SWH}})(\mathbf{x}) = (h_U, h_{\text{SWH}})_0 [n(\mathbf{x})/\bar{n}]^{-1/6}, \quad (33)$$

where  $(h_U, h_{\text{SWH}})_0$  is a reference bandwidth,  $n(\mathbf{x})$  is the number of altimetric measurements in the box

where  $\mathbf{x}$  is found, and  $\bar{n}$  is the mean value of  $n(\mathbf{x})$ , the mean being taken on all boxes that contain at least one measurement. As in Fig. 1, we use  $(U, \text{SWH})$  boxes of width  $(0.25 \text{ m s}^{-1}, 0.25 \text{ m})$ . The mean  $\bar{n}$  is a normalization factor used to obtain  $(h_U, h_{\text{SWH}})(\mathbf{x}) = (h_U, h_{\text{SWH}})_0$  in boxes with average data density. The resulting local modulation of the bandwidth is shown in Fig. 6. It varies from about 0.7 near the peak of the data distribution up to 2.0 at the boundary of the nonshaded region.

A new SSB estimation (EXLLR4) was obtained using the local bandwidth rule (33). As in EXLLR3, we used the complete SIM dataset, the LLR estimator and the Epanechnikov kernel. The global bandwidth of EXLLR3 is used as the reference; that is,  $(h_U, h_{\text{SWH}})_0 = (2 \text{ m s}^{-1}, 0.9 \text{ m})$ . Results are shown in Fig. 7. In data-dense regions, the estimation error remains very small, as was the case in EXLLR3. Actually, the EXLLR4 estimate appears to be clearly better than EXLLR3 only in regions with very scarce data coverage, that is, with 10 to 30 measurements per box, which is just outside of the nonshaded zone. In that region, the estimation error now remains well below the centimeter level. Figure 8 confirms that, while the standard deviation of this (nonshifted) estimate is typically between 1 and 2 mm in the nonshaded region, it also keeps low values (3 to 4 mm) immediately outside this region. Note that the magnitude of the standard deviation is very consistent with the magnitude of the observed errors shown in Fig. 7b.

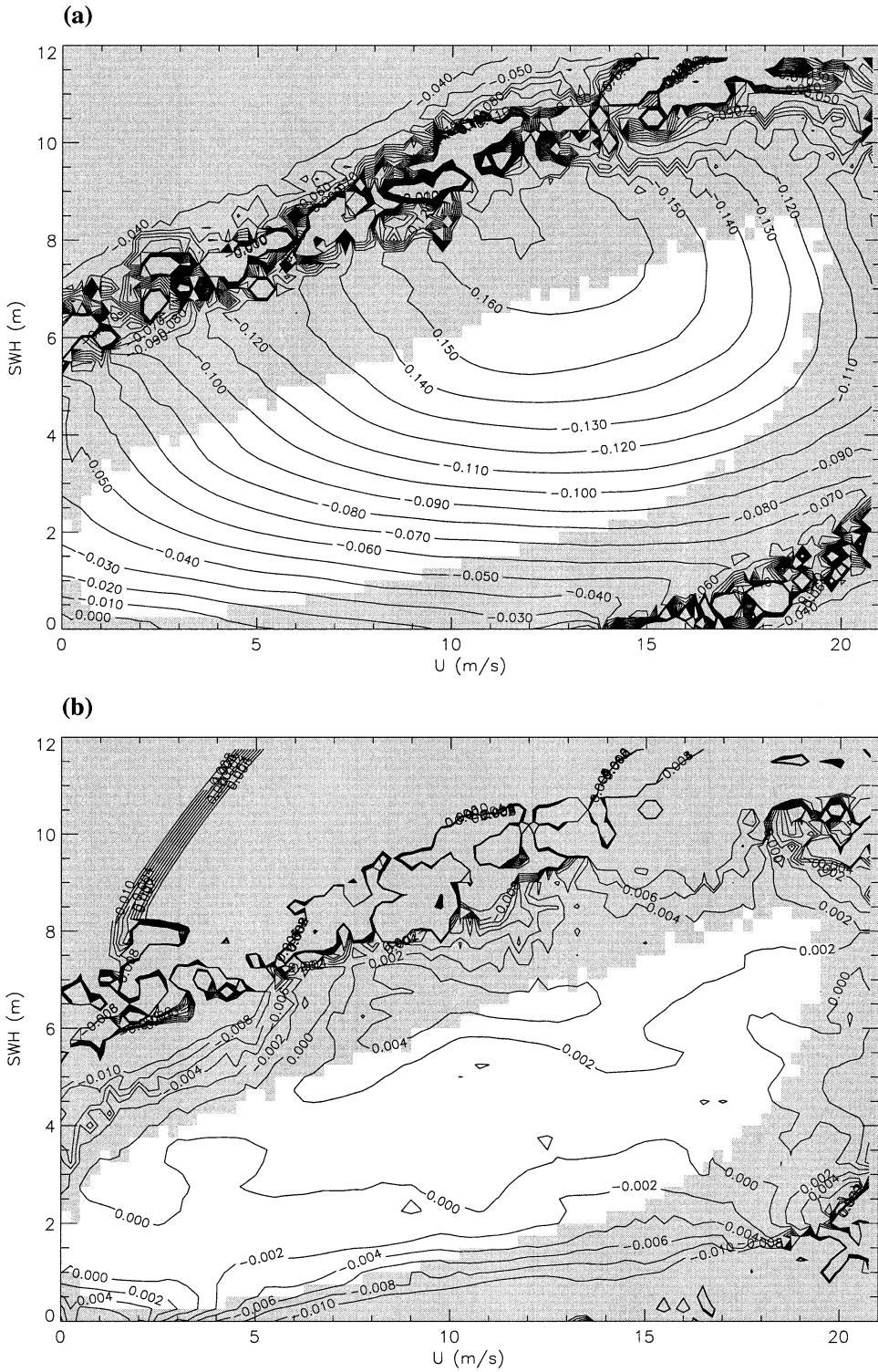


FIG. 7. (a) SSB estimate (in m) obtained in EXLLR4 and (b) the corresponding estimation error.

From a statistical point of view, this improvement of the SSB estimation in data-poor regions, typically concerns 1% of the altimetric measurements and may thus seem of little interest. Still, for studies of extreme

events, it can be important to have reasonable SSB estimates even in very unusual sea surface conditions. In addition, use of a local bandwidth reduces the estimation uncertainty at (0, 0) and thus the uncertainty of the

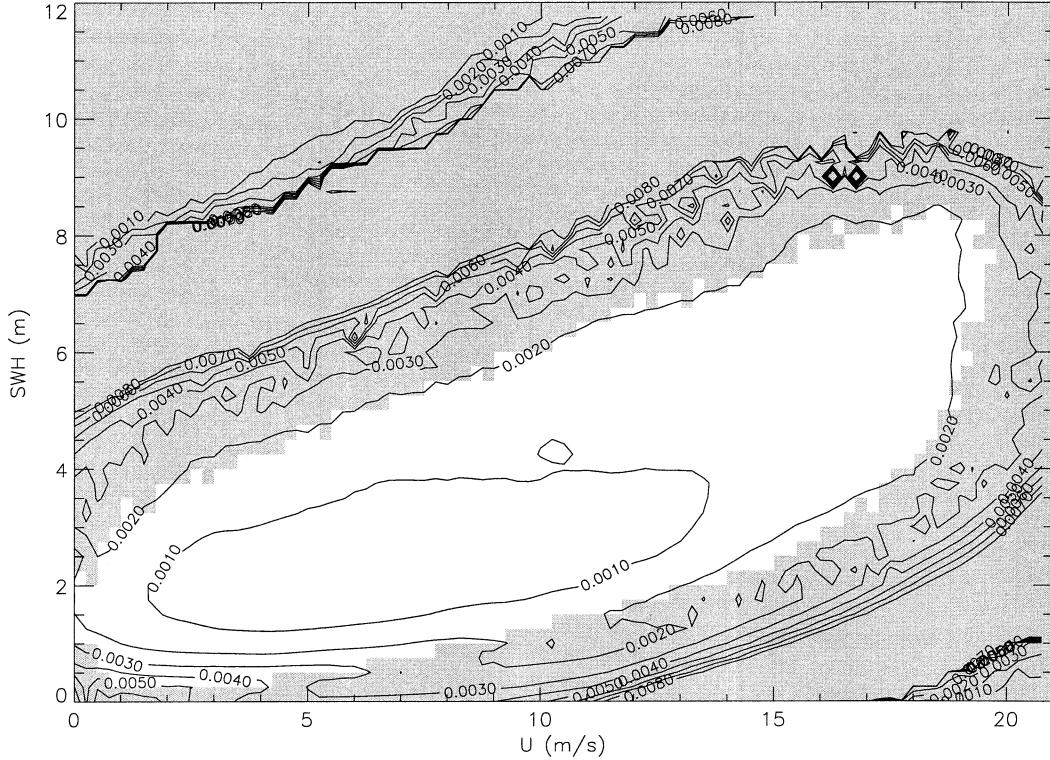


FIG. 8. Standard deviation (in m) of the final SSB estimate obtained in EXLLR4.

shifting process. We will thus keep the simple local bandwidth determination rule (33) for further work with the TOPEX data.

### 5. New TOPEX SSB estimate

The series of experiments performed with the SIM dataset show that the revised NP SSB estimation method based on the LLR estimator, the spherical Epanechnikov kernel, and a local bandwidth (33) leads to significantly improved results compared to those previously published by GF. The last step is thus to apply this revised method to obtain a new NP estimate of the TOPEX SSB. This is readily obtained by performing an estimation experiment (EXTP) in the same technical conditions as EXLLR4 but with the TP dataset instead of SIM. Note, however, that, in this case, the final solution has to be shifted to satisfy the constraint  $\text{SSB}(0, 0) = 0$ .

Let us first examine the standard deviation of the EXTP solution before it is shifted:  $m^{-1/2}\sigma_{\varphi}(\mathbf{x})$  (Fig. 9a), and the standard deviation of the final shifted solution  $\sigma_{\text{SSB}}(\mathbf{x})$  (Fig. 9b). As expected, Fig. 9a closely resemble Fig. 8. For both the SIM and TP datasets, the estimation uncertainty remains close to 1 mm near the peak of the data distribution and increases to 2 or 3 mm as one approaches the boundaries of the nonshaded area. Unfortunately, the estimation uncertainty is still rather large in the vicinity of (0, 0) where it reaches a value of 9

mm. This value is the standard deviation of the shift estimate. Figure 9b accordingly shows that, in most of the nonshaded area, the uncertainty of the final SSB estimate is close to 1.1 cm, roughly corresponding to a constant 9-mm uncertainty due to the solution shift and a 1–3-mm uncertainty due to the estimation itself. The covariance between the shift and estimation effects has little impact, except close to (0, 0) where it naturally causes the uncertainty to become zero, as all individual SSB estimates are constrained to be zero.

The new TOPEX SSB estimate is shown in Fig. 10a with the corresponding relative bias ( $\text{SSB}/\text{SWH}$ ) in Fig. 10b. The wind- and wave-related variability of this new estimate is consistent with the known properties of the Ku-band SSB (see Chelton et al. 2001 for a review). In particular, for any given SWH, the new estimate confirms that the SSB, and relative bias, increase with  $U$  for wind speeds up to 11 or 12  $\text{m s}^{-1}$  and then decrease at higher wind speeds. Also, the relative bias decreases with SWH as first noted by Witter and Chelton (1991b).

The estimated magnitude of the SSB is slightly more surprising. Most previous empirical estimates of the TOPEX relative bias are in the range  $-1\%$  to  $-3\%$  of SWH, with an exception for the BM4 estimate (Gaspar et al. 1994) that reaches somewhat larger values, close to  $-4\%$ , for underdeveloped seas. The newly estimated relative bias is typically in the range  $-3\%$  to  $-4\%$  of

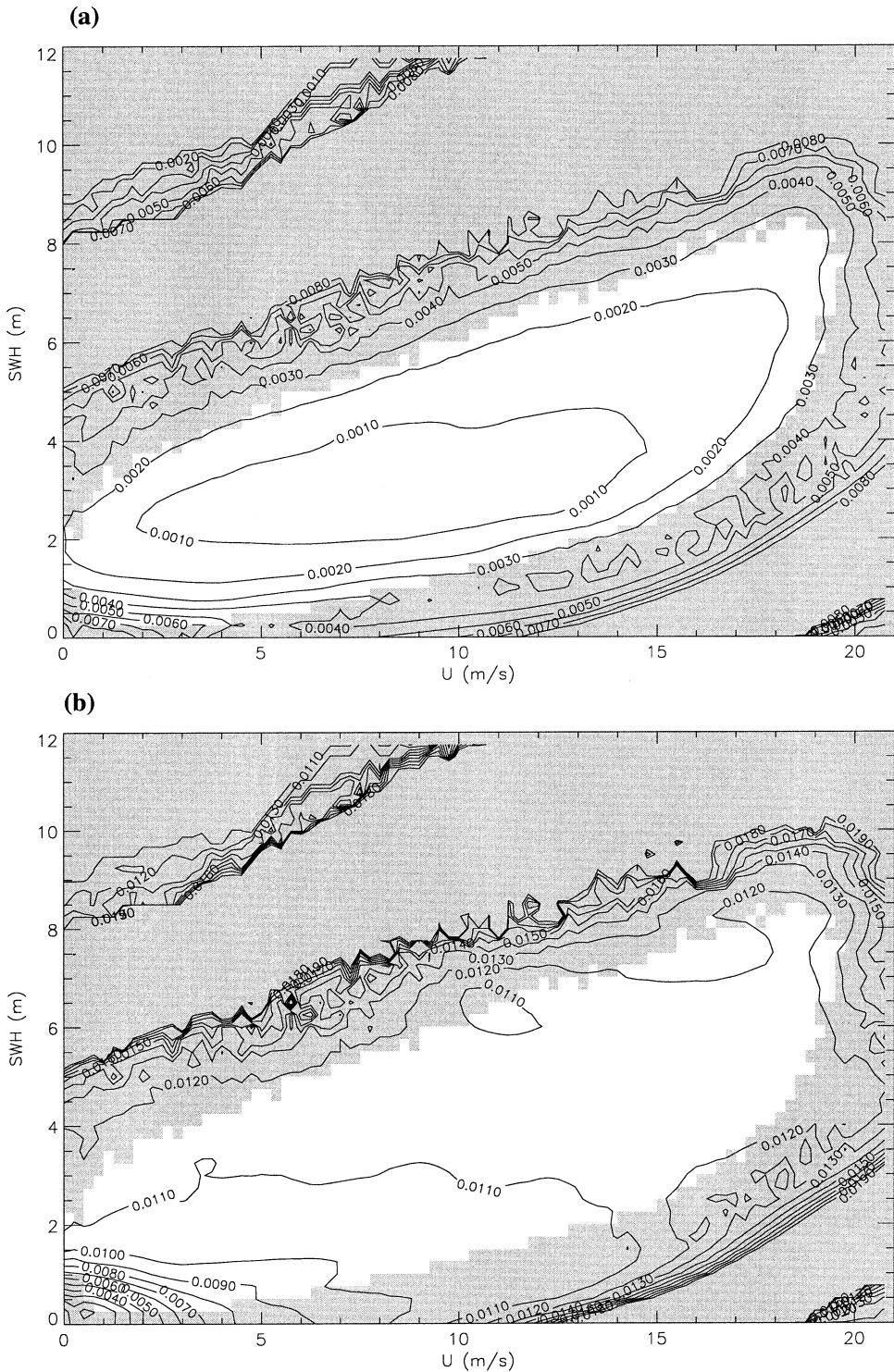


FIG. 9. (a) Standard deviation (in m) of the nonshifted TOPEX SSB estimate obtained in EXTP and (b) standard deviation of the corresponding final shifted solution.

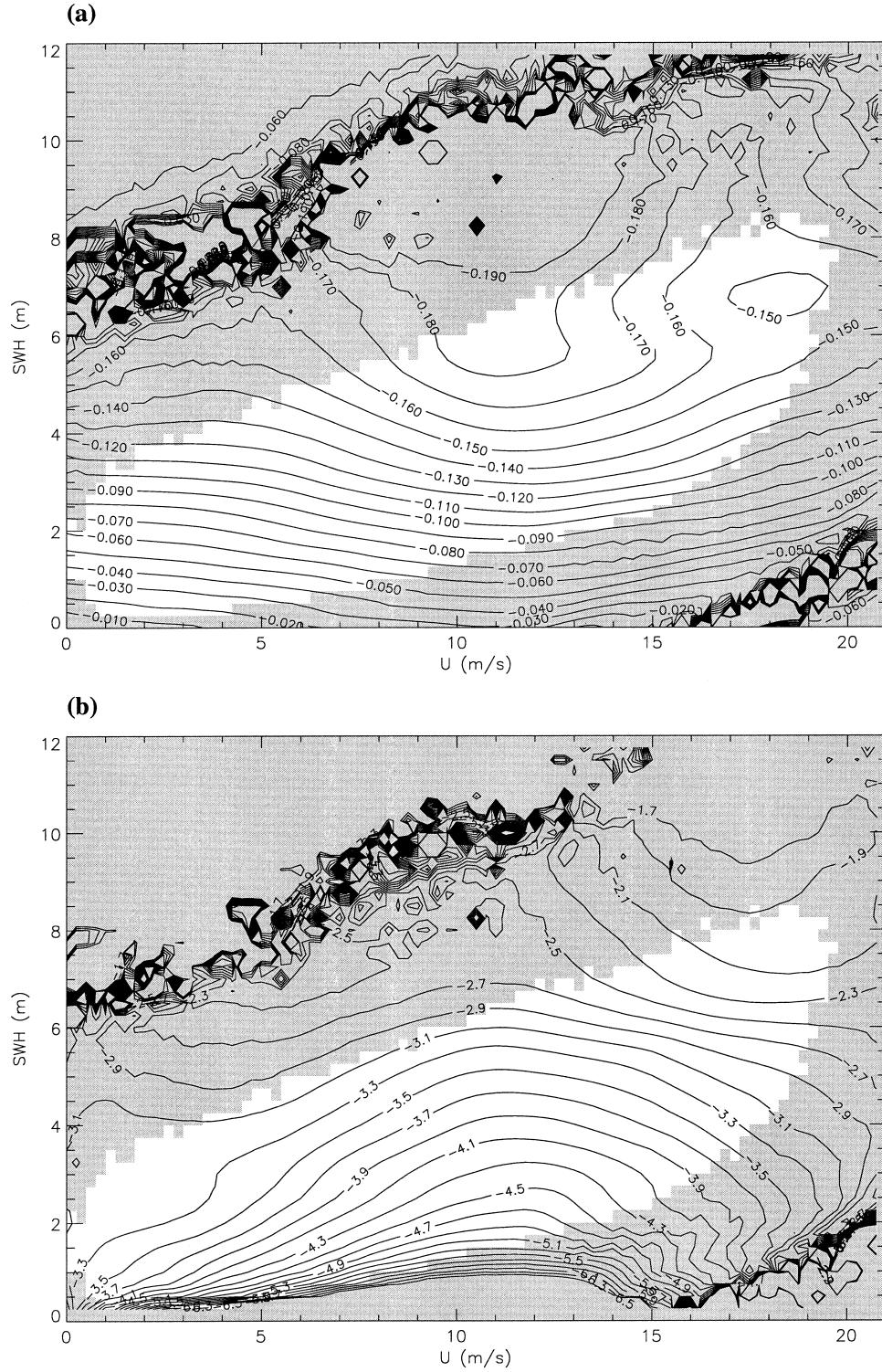


FIG. 10. (a) SSB estimate (in m) obtained in EXTP and (b) the corresponding relative bias (% of SWH).

SWH, with values as high as  $-5\%$  for markedly underdeveloped seas and as low as  $-2\%$  for the highest observed waves. This range of value better matches in situ measurements of the Ku-band SSB made by Mel-

ville et al. (1991) during the SAXON experiment off the Chesapeake Bay, by Arnold et al. (1995) in the Gulf of Mexico, and by W. K. Melville (1999, personal communication) in the Bass strait. Millet et al. (2001a) ac-

tually demonstrated that the differences between our new SSB estimate and in situ measurements in the Gulf of Mexico and Bass Strait are the order of 1 cm. This difference is well within the error bars of both the in situ measurements and our SSB estimate.

This new SSB estimate is directly available from the authors through email ([gaspar@cls.fr](mailto:gaspar@cls.fr)) as a lookup table.

## 6. Summary and conclusions

The refined NP SSB estimation method presented in this paper significantly improves upon the first version of the method devised by GF. The main changes are as follows.

- 1) The refined method uses an LLR estimator for conditional expectations instead of the NW estimator used by GF. This eliminates the bias affecting GF SSB estimates in areas with large data density variations, such as the boundaries of the estimation domain.
- 2) The refined method uses an Epanechnikov kernel instead of a Gaussian kernel. This has little impact on the estimation itself but transforms the large linear system to be solved into a sparse system for which the very efficient LSQR solution method (Paige and Saunders 1982) is implemented together with a compact matrix storage scheme. The resulting gain in computational efficiency allows us to obtain SSB estimates based on complete repeat cycles.
- 3) The refined method uses a local bandwidth determination rule instead of a global one. This allows adaptation of the amount of smoothing to the local data density. This is specially beneficial in the data-poor regions of the  $(U, \text{SWH})$  plane.

All these technical refinements are validated through numerical experiments with simulated data containing an exactly known SSB plus a realistic geographically varying noise. A new TOPEX SSB estimate is finally obtained using the refined estimation technique. This estimate exhibits a wind- and wave-related variability in agreement with the latest findings on this topic. The magnitude of the corresponding relative bias is in the range  $-2\%$  to  $-5\%$  of SWH, larger than most previous empirical estimates based on TOPEX data but in good agreement with different sets of in situ measurements (Melville et al. 1991; Arnold et al. 1995; W. K. Melville 1999, personal communication).

The uncertainty of the new TOPEX SSB estimate is close to 1.1 cm. It combines the estimation uncertainty itself with the (constant) uncertainty of the solution shift needed to impose  $\text{SSB}(0, 0) = 0$ . The uncertainty of the solution shift is, by far, the largest (9 mm), while the estimation uncertainty is in the range 1 to 3 mm. The only way to significantly reduce the uncertainty of the solution shift is to obtain and use more good-quality measurements close to  $(0, 0)$  in order to better constrain the estimation of  $\text{SSB}(0, 0)$ .

Finally, it should be clear that the uncertainties mentioned here only concern the fraction of the SSB possibly modeled as a function of  $(U, \text{SWH})$ . Other factors influencing the SSB such as the fetch or the wave slope are only partly related to  $U$  and SWH. Their impact on the SSB will therefore only be partly reproduced in our estimate. Significant work is now under way to identify the most pertinent missing factors needed to reproduce the largest possible fraction of the SSB variability. The long-wave slope is certainly one of them (Millet et al. 2001b; Chapron et al. 2001). If we become able to accurately estimate this, or another, pertinent parameter at the global scale, the NP SSB estimation technique is now mature enough to help reliably quantify its impact on the SSB.

**Acknowledgments.** The authors wish to thank Jean-Pierre Florens for useful discussions on nonparametric estimation methods and Alain Huard for his expert recommendations on numerical solution techniques for large linear systems. This study is part of the preparation activities for the JASON and ENVISAT missions. It was supported by the Centre National d'Etudes Spatiales (CNES) under Contract 731/CNES/00/8251/00 and by the European Space Agency (ESA) under Contract 12132/96/NL/GS.

## REFERENCES

- Arnold, D. V., W. K. Melville, R. H. Stewart, J. A. Kong, W. C. Keller, and E. Lamarre, 1995: Measurements of electromagnetic bias at Ku and C bands. *J. Geophys. Res.*, **100**, 969–980.
- Born, G. H., M. A. Richards, and G. W. Rosborough, 1982: An empirical determination of the effects of sea state bias on SEASAT altimetry. *J. Geophys. Res.*, **87**, 3221–3226.
- Chapron, B., D. Vandemark, T. Elfouhaily, D. R. Thompson, P. Gaspar, and S. Labroue, 2001: Altimeter sea state bias: A new look at global range error estimates. *Geophys. Res. Lett.*, **28**, 3947–3950.
- Chelton, D. B., 1994: The sea state bias in altimeter estimates of sea level from collinear analysis of TOPEX data. *J. Geophys. Res.*, **99**, 24 995–25 008.
- , E. J. Walsh, and J. L. MacArthur, 1989: Pulse compression and sea level tracking in satellite altimetry. *J. Atmos. Oceanic Technol.*, **6**, 407–438.
- , J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, 2001: Satellite altimetry. *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*, L.-L. Fu and A. Cazenave, Eds., Academic Press, 1–122.
- Chu, C. K., and J. S. Marron, 1991: Choosing a kernel regression estimator (with discussions). *Stat. Sci.*, **6**, 404–436.
- Elfouhaily, T., D. R. Thompson, D. Vandemark, and B. Chapron, 1999: Weakly nonlinear theory and sea state bias estimations. *J. Geophys. Res.*, **104**, 7641–7647.
- , —, B. Chapron, and D. Vandemark, 2000: Improved electromagnetic bias theory. *J. Geophys. Res.*, **105**, 1299–1310.
- Fan, J., 1992: Design-adaptive nonparametric regression. *J. Amer. Stat. Assoc.*, **87**, 998–1004.
- , and I. Gijbels, 1996: *Local Polynomial Modelling and its Applications*. Chapman and Hall, 341 pp.
- Fu, L.-L., and R. Glazman, 1991: The effect of the degree of wave development on the sea state bias in radar altimetry measurements. *J. Geophys. Res.*, **96**, 829–834.
- Gaspar, P., and J. P. Florens, 1998: Estimation of the sea state bias

- in radar altimeter measurements of sea level: Results from a new nonparametric method. *J. Geophys. Res.*, **103**, 15 803–15 814.
- , F. Ogor, P.-Y. Le Traon, and O. Z. Zanife, 1994: Estimating the sea state bias of the TOPEX and POSEIDON altimeters from crossover differences. *J. Geophys. Res.*, **99**, 24 981–24 994.
- Glazman, R. E., A. Greysuckh, and V. Zlotnicki, 1994: Evaluating models of the sea state bias in satellite altimetry. *J. Geophys. Res.*, **99**, 12 581–12 591.
- Härdle, W., 1990: *Applied Non-Parametric Regression*. Cambridge University Press, 349 pp.
- Melville, W. K., R. H. Stewart, W. C. Koeller, J. A. Kong, D. V. Arnold, A. T. Jessup, M. R. Loewen, and A. M. Slinn, 1991: Measurements of electromagnetic bias in radar altimetry. *J. Geophys. Res.*, **96**, 4915–4924.
- Nadaraya, E. A., 1964: On estimating regression. *Theory Probab. Its Appl.*, **19**, 133–141.
- Paige, C. C., and A. M. Saunders, 1982: LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, **8**, 43–71.
- Ray, R. D., and C. J. Koblinsky, 1991: On the sea state bias of the GEOSAT altimeter. *J. Atmos. Oceanic Technol.*, **8**, 397–408.
- Rodriguez, E., Y. Kim, and J. M. Martin, 1992: The effect of small-wave modulation on the electromagnetic bias. *J. Geophys. Res.*, **97**, 2379–2389.
- Ruppert, D., and M. P. Wand, 1994: Multivariate weighted least squares regression. *Ann. Stat.*, **22**, 1346–1370.
- , S. J. Sheather, and M. P. Wand, 1995: An effective bandwidth selector for local least squares regression. *J. Amer. Stat. Assoc.*, **90**, 1257–1270.
- Saad, Y., 1996: *Iterative methods for sparse linear systems*, PWS, 447 pp.
- Simonoff, J. S., 1996: *Smoothing Methods in Statistics*. Springer-Verlag, 338 pp.
- Watson, G. S., 1964: Smooth regression analysis. *Sankhya A*, **26**, 359–372.
- Witter, D. L., and D. B. Chelton, 1991a: A GEOSAT altimeter wind speed algorithm and a method for altimeter wind speed algorithm development. *J. Geophys. Res.*, **96**, 8853–8860.
- , and —, 1991b: An apparent wave height dependence in the sea state bias in GEOSAT altimeter range measurements. *J. Geophys. Res.*, **96**, 8861–8867.
- Yaplee, B. S., A. Shapiro, D. L. Hammond, B. D. Au, and E. A. Uliana, 1971: Nanosecond radar observation of the ocean surface from a stable platform. *IEEE Trans. Geosci. Electron.*, **GE-9**, 170–174.
- Zlotnicki, V., L.-L. Fu, and W. Patzert, 1989: Seasonal variability in global sea level observed with Geosat altimetry. *J. Geophys. Res.*, **94**, 17 959–17 969.