Use of fuzzy logic clustering analysis to address wave impacts on altimeter sea level measurements

D. Vandemark, H. Feng, T. S. Moore

University of New Hampshire, Ocean Processes Analysis Lab., Durham, New Hampshire

B. Chapron

IFREMER/Centre de Brest, Plouzané, France

N. Tran

CLS/Space Oceanography Division, Ramonville St-Agne, France

Short title: FUZZIFICATION OF THE SEA STATE BIAS

Abstract. Each satellite ocean altimeter measurement must be corrected for surface waves to accurately retrieve a sea level estimate. This wave correction is developed through indirect empirical methods, but because it can not be directly determined using satellite data, it has been difficult to quantify the correction model's uncertainty, much less improve upon it. A new approach for assessing and improving this wave-induced sea state bias algorithm is described here based on fuzzy logic clustering analysis, data from a global ocean wave model, and direct averaging over the altimetric sea surface height anomaly field. These tools combine to provide an objective means to define regimes of nonlinear surface wave variability, to partition wave model and altimeter data into these classes, and subsequently to build class-specific wave correction models in the standard wave height wind speed data domain. Six regimes are objectively determined using input parameters related to the wave height, wave age, and variation in short-scale slope variance. This paper focuses mainly on method demonstration and on detecting and quantifying where the present-day global sea state bias model fails. Several new results emerge. Overall, the methodology yields results anticipated by numerous investigators over the past decades including regional impacts, the effect of swell, and expected ambiguity associated with wave age and steepness. A substantial portion of data samples exhibit range error magnitudes exceeding 0.5% in H_s . As importantly, these errors are not globally or randomly distributed and appear to impose systematic spatial and temporal sea level error in regions as diverse as the equatorial Pacific and Antarctic Circumpolar Current. The six classes combine to show positive skill in sea surface height variance reduction over a single global model.

The analysis is applied equally to TOPEX and Jason-1 data with similar end results. This suggests that the wave physics underpinning the analysis could be divorced from inherent sensor differences, leading to a more universal cross-platform handling of this correction. Further refinement, including fuzzification, should extend this methodology to development of a robust operational solution. Parts I and II provide posters to explain the methodology and the resulting sea state bias error examination.

Introduction

Does one size fit all? Does it have too?

The cm-level range bias attributed to wind wave effects upon ocean altimeter sea level measurements is coined the sea state bias and involves the fact that the nominal gravity wave profile is slightly peaked at its crest and flat in the trough. This leads to asymmetric radar reflection as the microwave signal propagates down upon first the peaks and then the troughs of the many ocean waves illuminated during any given measurement. A range bias results and there has been significant effort expended to empirically correct for this bias using readily-available altimeter wind speed and wave measurements as surrogates for the actual nonlinearity of a given wave field (*Gaspar et al.*, 2002). But numerous studies predict that significant O(cm) errors in sea level remain even after correction using an optimal altimeter-fed empirical sea state bias algorithm (**?**) because the drivers for the sea state error are not directly correlated with the significant wave height and the altimeter backscatter parameters. Further improvement face two hurdles - the first being how to quantify where remaining wave-driven errors occur and the second being how to acquire and utilize new, complementary, and contemporaneous gravity wave information to augment the very useful, but insufficient, altimeter backscatter and significant wave height (H_s) data.

Old text for the rest of the paragraph. The level of this uncertainty is not well quantified or hard to quantify with a single number but the estimate is of the order of 1 % H_s . (i.e. next order problem is that we don't really know where and when the present algo fails - a better validation:assessment tool is needed....) How do we improve on this optimized on-orbit algorithm? Consensus is that the first step is to recognize that the drivers for the sea state error are not directly the significant wave height and the altimeter backscatter. So follow-up questions become - what are the actual controls on the phenomenon and how does one parameterize this? Answers to these queries must recognize the pragmatic matter of how to obtain and enfold relevant new information into an operational approach that corrects each altimeter range estimate.

Sea state bias research attempting to advance beyond the nominal operational two parameter approach generally falls into one of two approaches, either development of an alternate parameterization using theoretical and field experiment efforts (**?***Melville et al.*, 2004; *Gommenginger et al.*, 2003; *Elfouhaily et al.*, 2001), or use of on-orbit satellite data to indirectly suggest where and why the present operational approach fails (*Kumar et al.*, 2003; *Glazman et al.*, 1996; *Minster et al.*, 1992; *Fu and Glazman*, 1991; *Zlotnicki et al.*, 1989). In some sense, these efforts divide out between what we can predict based on field work and theory versus what can be accomplished pragmatically with the space-based data. The approach taken in this paper is to merge aspects of what is known about physical controls on the sea state bias with the operationally-driven constraint that the altimeter H_s and backscatter measurements, while imperfect correlatives, will still be a central part of any correction algorithm. In the process we hypothesize that distinct regimes of wave nonlinearity can be objectively defined and detected within a global data set and that the wave-induced range error characteristics will differ between regimes. The approach taken will essentially be a perturbation analysis where altimeter and gravity wave model data are combined to objectively define differing wave provinces (i.e. regimes or classes) through dynamics in the wave age and wave slope parameters, terms identified with the sea state bias in numerous studies. This objective fuzzy statistical clustering analysis is then used to sort global data into these wave provinces. The effort is, in some sense, a generic exploration of the utility of such a clustering approach to providing a more rigorous evaluation of any globally-derived algorithm's applicability at a given point.

The study objectives are to combine altimeter and wave model data to better quantify where and when perturbations in the global sea state bias correction are occurring and to offer a path forward to an improved operational model. The chosen statistically-based data clustering approach is largely driven by the facts that there are a) no direct means to build an *in* r*situ* versus satellite bias data set and b) we have imperfect altimeter surrogates and wave model information avaialable at the instant and location of each measurement. Methods for the study are described in the following section. To bring needed new data into the effort we have developed a data set where wave model estimates of the directional surface gravity wave spectrum are generated for each individual altimeter estimate. Next we discuss how the data are used to objectively define separate wave nonlinearity provinces using a fuzzy logic clustering analysis, and then to develop empirical two input parameter sea state bias models for these regimes. Results and discussion follow where attention is given to each regime and its characteristics within the context of what is known or predicted regarding the sea state bias. Sensitivity of the results to differences in differing wave model runs, use of Jason-1 versus TOPEX altimeters, and differing analysis periods are addressed. Finally, the study's implications to improved sea state bias modeling and how one might bring these pieces back together in one seamless algorithm are discussed. A longer term goal goal is to use of an objective fuzzy classifier that can provide a continuous, yet partitioned, model to predict wave-induced range errors.

Methods

There is ample evidence indicating that the present-day two parameter sea state bias model has deficiencies due primarily to the fact that the available input wind speed and wave height terms are only surrogates for the actual wave nonlinearity driving the bias. Examples include results that show likely regional error (*Glazman et al.*, 1994; *Kumar et al.*, 2003; *Gaspar and Florens*, 1998) and the impact of swell systems (*Glazman et al.*, 1996; *Minster et al.*, 1992). But these results have been limited to an implication of error that is difficult to quantify. We desire a methodology where one can examine error from a physical perspective, but still come to a quantitative regional and global evaluation in the end. A more generic problem also exists in this arena, how to use imperfect input data to show how and where an imperfect model is underperforming. A fuzzy logic classification scheme is chosen to address these needs. The approach taken is to cluster data into classes representing somewhat distinct nonlinear wave regimes, e.g. swell-dominated or young seas, and then to create new SSB models for these subsets. These SSB models are developed using the standard two term altimeter inputs [H_s ,U_{10}]. This is done in large part to provide an objective sea state bias assessment that is familiar to the research community. Future work could refine this using alternate forms and input parameters. The resulting class-specific SSB models are then examined alongside the globally-derived model to provide a new view of how much sea surface height error results from use of the global algorithm and where this occurs. Details follow.

Study data set

A critical study component is the archive of collocated wave model and altimeter data created for sea state bias model improvement efforts. This data set merges six hourly global output fields from the Wave Watch III (WW3, (**?**) Ver 2.x model with all available altimeter data from both the TOPEX/Poseidon and Jason-1 altimeter missions spanning the years 2000-2003. This merge involves interpolating the 1 degree grid wave model output in space and time onto the nominal TOPEX altimeter pathfinder groundtrack (**?**). Altimeter data used here include the significant wave height, the 10 m wind speed as derived from the Ku-band cross section (witter and chelton) model, and sea surface height anomaly with corrections per the pathfinder data set version X.X. Wave model statistics used are the significant wave height of the wind sea and the mean wave vertical acceleration variance, both derived from WW3 two-dimensional directional wave spectral output. Remaining details pertaining to these data are found in *Feng et al.* (2006); *Tran et al.* (2006).

Wave model data offer the new opportunity to correlate proposed bias-related wave parameters, for example the root mean square wave slope, to observed sea surface height changes. Recent work (*Kumar et al.*, 2003; *Tran et al.*, 2006) has taken this approach with some success. However, the modeled wave information does not share the accuracy or resolution of the altimeter data it is merged with (*Feng et al.*, 2006). This fact guides both the present approach and our choice of wave parameters. One key lesson drawn from that latter analysis is that the swell field created within the model runs is much less reliable than the wind sea, thus for this study most trust or weight is given to the wind sea related statistics carried in the wave model spectra.

Data classification

An unsupervised fuzzy c-means data clustering algorithm is applied to the data following the methods detailed in *Moore et al.* (2001) where a general summary of the motivation for a fuzzy classification approach can also be found. The algorithm acts to partition a multivariate data set into clusters and also to ascribe a membership value to each individual input sample that is based on the reciprocal distance from a given cluster center. The term fuzzy derives from the fact that each sample can belong to more than 1 cluster or class. The technique's suggested strengths lie in applications where vagueness or ambiguity are present and where an approximate but imperfect solution may suffice. Hard or crisp clustering is a subset of the fuzzy methodology consistent with classical set theory. Calculations herein utilized the FCM algorithm in MATLAB, Math Works Inc. This routine attempts to minimize the objective functionJ_m defined as

$$\mathrm{J}_m = {}_{j=1}^{N}({}_{i=1}^{N_c})(?_{ij}^{m} d^2(X_j, V_i))$$

where N is the total number of observations, j, in the data set; N_c is the number of separate clusters, i; d is the distance between an observation vector X_j and cluster centroid V_i; and μ is the cluster membership value for each sample (value between 0 and 1); and m is a weighting exponent.

The classification objective for this altimeter application is to divine data subsets that identify or illuminate sources of ambiguity residing within the global SSB algorithm. The guiding assumption is that if one can obtain classes related to anomalies in the surface wave field nonlinearity then these classes may also be relevant to altimeter range error explanation. Our choice of a multivariate data vector X for input to the clustering is $[H_s\ ,\ \delta_h\ ,\ \delta_s\ ]$ where δ_h is the ratio $\frac{H_{sea}}{H_s}$ and δ_s is the ratio $\frac{mss_l}{mss_{tot}}$. Here H_{sea} refers that portion of the wave height spectrum attributed to the wind sea; mss_{long} is the slope variance of the long waves down to a cut-off frequency of 0.4 Hz as obtained from the WW3 wave acceleration spectral variance (m_4) using $mss_{long} = \sqrt{m_4 2\pi^4 g^{-2}}$ under the linear gravity wave dispersion assumption; while mss_{tot} is derived from the inverse of the altimeter radar cross section under the nominal quasi-specular relationship (*Barrick*, 1968; *Jackson et al.*, 1992) and using a Fresnel coefficient of 0.45 . Each of these four parameters within the triplet are found within wave nonlinearity and sea state bias literature. H_s is a recognized first-order correlative. δ_h , a ratio in wave elevation, can be interpreted as a pseudo wave age ($\frac{u^2}{H_s} \propto H_{sea}$ (*Fu and Glazman*, 1991; *Glazman et al.*, 1994), or an index for the amount of swell present. A high ratio indicates wind-driven, unimodal, and likely steep nonlinear seas like the canonical inverse wave age term u/c; c being the celerity of a wind-driven wave train and u the wind speed. But the global ocean exhibits a wide range in this ratio due to the common mixture of sea and multiple swell modes. Thus δ_h is a heuristic and somewhat ambiguous, or pseudo, wave age but one found to suit our needs within the present classification. δ_s pertains to a ratio in wave slope variance, or mean square slope (mss), and describes the amount of long (order 10 m) gravity wave steepness relative to the total mss that is typically dominated by the steeper shorter cm-scale wavelets. δ_s is theoretically (*Elfouhaily et al.*, 2001) and observationally (*Vandemark et al.*, 2005) shown to correlate with sea state bias change by its control of the cross-skewness statistic (*Srokosz and Longuet-Higgins*, 1986) and thus the electromagnetic tilt bias term. A simplified conceptual δ_s explanation is that short wave attenuation of the tilt bias term lessens when the long wave slope magnitude exceeds its nominal relationship to the short wave slopes.

Pragmatism was also considered when devising these ratio terms and their inclusion as input to the clustering analysis and eventually sea state bias investigations. It can be seen that δ_h and δ_s are explicitly normalized by H_s and altimeter backscatter σ^o respectively. Knowing *a priori* that the ensuing SSB evaluations will be done in [H_s ,U_{10} (i.e. σ^o)] space, these normalizations are designed to decrease the possibility of self-correlation in the analysis as a whole and to obtain the maximum utility from the wave model parameters.

Multiple realizations of the three-input training vector were submitted to the clustering process to determine the optimal number of classes and class centers. Each realization carried 20,000 globally randomized data points drawn from a one year pool holding more than 1 million samples. The resulting optimized number of clusters chosen from this training is six as determined through ensemble evaluation of the partitioning coefficient, clustering index, and weighting exponent (see *Moore et al.*

(2001)). Next, all data within a given time period can be submitted to the inversion using the known cluster number to ascribe membership values μ to each sample. These values fall between 0 and 1 and each sample can belong to one class. This study uses the distance-based FCM membership function to assign membership and this constrains total membership such that a sample's total must sum to 1.

Class-specific sea state bias models

Once classes have been defined through the clustering computation it is straightforward to create a sea state bias estimator for a given subset or class (*Vandemark et al.*, 2002). First, pre-corrected altimeter sea surface height anomaly data are obtained by removing the operational sea state bias solution. These data are then conditionally averaged with respect to H_s and U_{10} bins, leading to a data table akin to the nonparametric two parameter solution developed at global scale (**?**). All such models presented in this study are created with a one year data set. The classes used here are built using a hard class selection - each observation is assigned only to the class matching that sample's maximum class membership value. Thus the resulting six class-based range error models together hold all the data and each class-specific model can be compared against a global (total data) solution that is developed using the superset. This straightforward comparison capability against the well-established global benchmark model provides a pragmatic and quantitative means to evaluate the questions of if and when class-specific error models offer improvements.

Results and discussion

Wave regime classification

The clustering analysis developed using TOPEX altimeter and wave model data for the year 2000 leads to six data classes as shown in Fig. 1 and Table 1. Recall that three wave-related parameters are used as input to the process - H_s , δ_h , and δ_s . The multivariate cluster centers are depicted via the three two dimensional panels in Fig. 1 while the cluster mean values and standard deviations are given in the table. The classes are numbered 1-6 according to their increasing value in wave height. Given this identification scheme one observes in Fig.1(a) that the classes are seen to roughly fall into three sea state categories: three low-to-moderate (class 1, 2, 3), two moderate (4, 5), and one extreme (6). Cluster centers show good separation in all three data domains in Fig.1 though the clustering does have a somewhat less monotonic behavior with regard toδ_h and δ_s . Still, amongst classes 1-3 there is an increase observed in both of the latter parameters. According to the hypotheses used in developing these classes, the increase in both ratio parameters should correspond to wave nonlinearity and range error increases between class 1 and class 3 for nearly the same wave height value. A similar observation and prediction is made for classes 4 and 5. Based on the cluster centers in Fig. 1 and the recollection that δ_h refers to pseudo inverse wave age whilst δ_s refers to shorter wave steepness dynamics, we can also provide qualitative labels - the lower sea state classes (1-3) are denoted as swell dominated, mixed, and young seas respectively. The moderate sea state classes (4-5) as mixed and young seas and class 6 as simply high seas.

The contours on Fig.1 represent the population density in each data domain and from this it is apparent that class 6 lies toward the extreme in all three input terms. The well known global H_s probability distribution function (*Callahan et al.*, 1994) is apparent with a mode near 2.3 m. The mode for δ_h lies near 0.4 while the cases of near full development reside near the δ_h maximum 0.8 and strong swell domination (old seas) toward the minimum of 0.2. The mode indicates that at global scale the swell typically carries more than half the total wave field energy, i.e. wind sea dominated wave spectra are unusual. The mode of the slope ratio δ_s is 0.1 and values can extend as high as 0.3. Regarding the total data set then, the class 2 cluster center indicates that this class represents the mode across the three dimensional data space.

Each sample in the global one year data set has been assigned six membership values ranging between 0.0 and 1.0, one value per class. These six values add up to one in this implementation. A membership value probability distribution for each class is shown in Fig. 2. Also shown for each class is the percentage of total samples whose maximum membership value belongs to that class. The percentages indicate that the global data are divided fairly evenly between the defined classes with the extreme wave height class six holding the lowest percentage. Overall, only class one shows a strong tendency for membership values to fall near to 1.0. Otherwise it appears that the distribution mode is typically a value between 0.4 and 0.5. Thus, any single data sample is likely to hold membership is more than one class. Given the prescribed training input choices and the continuous nature of wind waves this vagueness in separation amongst six classes is not unexpected. Is this a discussion section sentence?

Fig. 3 presents global data maps to show where these wave regime subsets are located in an average sense. The image intensity represents the annual average of class membership for a given 2 degree bin in latitude and longitude. For each spatial bin the total of the six classes must again sum to 1.0. Clear spatial patterns appear for each class across the ocean basins. As might be expected for the lower wave height classes (1-3), these subsets are localized towards the equator, below 30-40 degrees in either hemisphere. Classes 4-6 map to higher latitudes. Also of interest are the patterns that emerge amongst classes at a given latitude. The low latitude classes occupy distinct neighboring oceanic features such as the doldrums (class 1), the tradewind belts (class 3), and eastern equatorial boundaries (class 1). Classes 4 and 5 divide across the currents of the Southern Ocean and claim systematic membership in portions of the Indian Ocean. Such spatial details differ substantially from the annual climatology of significant wave height alone (*Young*, 1999) , and this new classified view of the wave regimes is intriguing in and of itself. Is this a discussion section sentence?

Further spatial and temporal information on wave dynamics can be seen if one presents the data seasonally as shown in Figs. 4 and 5. Seasonal variation is presented separately for the low and high latitude classes where now each spatial pixel displays which class dominates in that season. Focus for Fig. 4 should primarily be on the low latitude bands and for Fig. 5 . Seasonal hemispheric migrations are evident as are more subtle shifts; for example note the clear separation and temporal evolution between classes 4 and 5 near the Agulhus and Antarctic Circumpolar currents in the south, and numerous shifts in the equatorial Pacific and Indian Oceans. Systematic low and mid-latitude storm regions also become apparent in the seasonal view. There are many additional ways to view the classified data set, but the objective at hand is sea state bias and thus we proceed to this application.

Error models for the six wave classes

The reference, or benchmark, for evaluating the class-based range correction models will be the globally-derived TOPEX model derived for the year 2000 as seen in Fig. 6. Intensity represents the range error as a percentage of wave height with the mean value falling near 3 % and the range spanning from 1 to 4 %. The observed pattern is consistent with the two parameter operational model used for all TOPEX data (*Gaspar et al.*, 2002). As discussed, Fig. 6 is generated through conditional averaging over the sea surface height anomaly estimates in the data domain of [H_s , U_{10}]. As also discussed, an objective here is to determine remaining error within this correction model. Fig. 6b indicates where the six classified data subsets reside within this two parameter data domain by providing the 90 % contour for their respective population densities. One observes the increasing wave height classes (see Table 1 cluster means) as well as substantial class overlap across both H_s and U_{10} variables. The only class with little overlap is class 6.

Class-specific sea state bias models are readily derived using the same conditional averaging process and the results are shown in Fig.7. A given altimeter plus wave model data point is included in only one class model based on its maximum class membership value as determined from its triplet values in [H_s , δ_h , δ_s]. A sea state bias result is only given in the figure if a given data bin holds more than 50 samples, same as for the global estimator. Collectively the class models span over much of the total space as seen in Fig.6b. Range error values between classes for the same [H_s , U_{10}] value indicate some clear differences. For example, at moderate U_{10} values of 5-10 m/s one sees differences as high as 2 % H_s between 1 and 3, and between 4 and 5. Certainly each realization differs substantially from the global solution (Fig.6a) at some location in the data space. One manner of interpretation for these class-based estimates is that they represent the sea state bias for somewhat distinctive nonlinear wave regimes and this is addressed below.

Fig.8 provides the global versus class-based difference, or sea state bias anomaly, estimates over this same data space. Distinct patterns emerge for each class and systematic deviations in magnitude exceeding 1 % H_s are evident in classes 1-5. Conversely, each class also holds areas that agree closely with the global solution. Recognizing that globally most data fall between winds of 3-10 m/s and HS of 1-4 m, one can focus attention to these values to see that classes 1-5 essentially split the sea state bias apart. For example, at H_s of 2 m and wind of 6 m/s the class-specific answers for 1-4 would be -1.2, -0.x, 1.x , 1.x with respect to the global solution.

!!!!Fig.**??**re idea!!!! is to map for each pixel within the 90 % global data domain where one has at least 3 classes involved, a weighted average of the cross class anomaly magntitude [x1% data*anom1data+....2,3,4] or maybe not even weighted but just the largest anomaly magnitude for every pixel. Really what we want is an uncertainty measure, a number for each bin, e.g. 3 cm rms error at xy. Perhaps a start would be this worst-case mapping as predicted by our mapping done in A) % H_s and B) absolute.

Has Phillipe or NP ever shown this?

Result sensitivity discussion wrt to Jason/Topex, forcing winds, seasonal, interannual.

Briefly discuss limits of the method to interpretation and accurate perturbation assessment

Discussion

Methodology. We chose a path that puts us through to figure 8. As such it was highly attuned to our perturbation approach. This means that the choice of inputs largely directed the study direction. Another approach could have been taken. Regarding the wave classification, we got what we wanted which was division across [H_s , U_{10}] space. This is evident in Figs. 6b. This provides the potential then to sp

Is our test of Figure 8 independent proof of success - or just an artifact of least squares (Remko)? Is a test of error reduction independent? One answer would be that the classification tool is not created using all the data, only 20000 at a time. A next answer is that the class-specific SSb models are also not created using all the data - only using the data that showed maximum membership for that sample. But I suppose that any bin in 2D ssb space is completely covered by the six classes. Still these 6 are selected somewhat independently of the 2D data via the triplet params and membership.

NOT a big issue for this paper. What about bias in any given class model? Only real chance at proving this is not an issue is the spatial equalization. The training certainly has no ability to bias. Moreover, the global averaging is over the same subset of the class plus all other classes so unless a given class has a spatially systematic, non-zero, dynamic topography shift for a given H_s U_{10} bin with respect to the global then bias should not, in general, be an issue (some pixels off perhaps but not the whole discussion).

Bertrand point. - one key result is that residual is not that huge systematically and that is a good benefit of the tool and also not surprising given that we already have two valuable pieces of information in the coincident wave height and backscatter data.

Biggest discussion should be about Figure 8. What is says physically and what it says pragmatically.

This paper outlines a method for going beyond a one size fits all satellite algorithm. As such, it is generic.

Summary

This paper outlines a method for going beyond a one size fits all satellite algorithm. As such, it is generic. It also helps to illustrate how well the one size algorithm works, where it is physically true and where it is systematically off.

References

Barrick, D. (1968), Rough surface scattering based on the specular point theory, *IEEE Trans. Antennas Propagation*, *AP 16*, 449–454.

Callahan, P. S., C. S. Morris, and S. V. Hsiao (1994), Comparison of topex/poseidon sigma(0) and significant wave height distributions to geosat, *Journal of Geophysical Research-Oceans*, *99*(C12), 25,015–25,024.

Elfouhaily, T., D. R. Thompson, B. Chapron, and D. Vandemark (2001), Improved electromagnetic bias theory: Inclusion of hydrodynamic modulations, *Journal of Geophysical Research-Oceans*, *106*(C3), 4655–4664.

Feng, H., D. Vandemark, Y. Quilfen, B. Chapron, and B. Beckley (2006), Assessment of wind forcing on global wave model output using the topex altimeter, *Ocean Engineering*, *33*, 1431–1461.

Fu, L. L., and R. Glazman (1991), The effect of the degree of wave development on the sea state bias in radar altimetry measurement, *Journal of Geophysical Research-Oceans*, *96*(C1), 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, *Journal of Geophysical Research-Oceans*, *103*(C8), 15,803–15,814.

Gaspar, P., S. Labroue, F. Ogor, G. Lafitte, L. Marchal, and M. Rafanel (2002), Improving nonparametric estimates of the sea state bias in radar altimeter measurements of sea level, *Journal of Atmospheric and Oceanic Technology*, *19*(10), 1690–1707.

Glazman, R., A. Fabrikant, and M. A. Srokosz (1996), Numerical analysis of the sea state bias for satellite altimetry, *Journal of Geophysical Research*, *101*(C2), 3789–3799.

Glazman, R. E., A. Greysukh, and V. Zlotnicki (1994), Evaluating models of sea-state bias in satellite altimetry, *Journal of Geophysical Research-Oceans*, *99*(C6), 12,581–12,591.

Gommenginger, C. P., M. A. Srokosz, J. Wolf, and P. Janssen (2003), An investigation of altimeter sea state bias theories, *Journal of Geophysical Research-Oceans*, *108*(C1), art. no.–3011.

Jackson, F. C., W. T. Walton, D. E. Hines, B. A. Walter, and C. Y. Peng (1992), Sea-surface mean-square slope from Ku-band backscatter data, *J. Geophys. Res.-Oceans*, *97*(C7), 11,411–11,427.

Kumar, R., D. Stammer, W. K. Melville, and P. Janssen (2003), Electromagnetic bias estimates based on topex, buoy, and wave model data, *Journal of Geophysical Research-Oceans*, *108*(C11).

Melville, W. K., F. C. Felizardo, and P. Matusov (2004), Wave slope and wave age effects in measurements of electromagnetic, *Journal of Geophysical Research-Oceans*, *109*(C7).

Minster, J. F., D. Jourdan, C. Boissier, and P. Midolmonnet (1992), Estimation of the sea-state bias in radar altimeter geosat data from examination of frontal systems, *Journal of Atmospheric and Oceanic Technology*, *9*(2), 174–187.

Moore, T. S., J. W. Campbell, and H. Feng (2001), A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms, *Ieee Transactions on Geoscience and Remote Sensing*, *39*(8), 1764–1776.

Srokosz, M. A., and M. S. Longuet-Higgins (1986), On the skewness of sea-surface elevation, *Journal of Fluid Mechanics*, *164*, 487–497.

Tran, N., D. Vandemark, B. Chapron, S. LaBroue, H. Feng, B. Beckley, and V. P. (2006), New models for satellite altimeter sea state bias correction developed using global wave model data, *Journal of Geophysical Research-Oceans*, *111*(C9), art. no.–3406.

Vandemark, D., N. Tran, B. D. Beckley, B. Chapron, and P. Gaspar (2002), Direct estimation of sea state impacts on radar altimeter sea level measurements, *Geophysical Research Letters*, *29*(24), art. no.–2148.

Vandemark, D., B. Chapron, T. Elfouhaily, and J. W. Campbell (2005), Impact of high frequency waves on the ocean altimeter range bias, *Journal of Geophysical Research-Oceans*, *110*, art. no.–2929.

Young, I. R. (1999), Seasonal variability of the global ocean wind and wave climate, *International Journal of Climatology*, *19*(9), 931–950.

Zlotnicki, V., L. L. Fu, and W. Patzert (1989), Seasonal variability in global sea-level observed with geosat altimetry, *Journal of Geophysical Research-Oceans*, *94*(C12), 17,959–17,969.

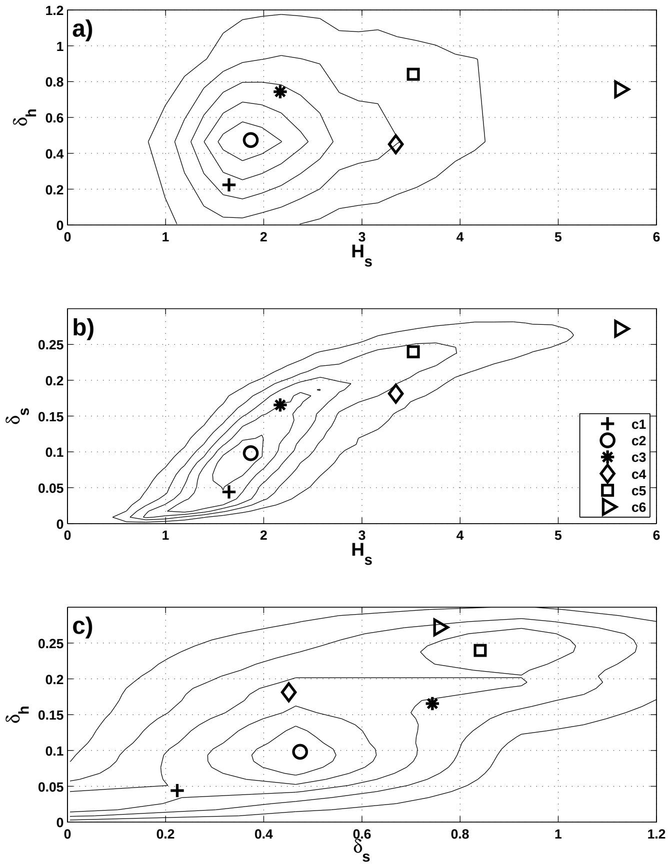

Figure 1. Centroid values for the cluster centers obtained from the nonlinear wave data classification scheme. Six class centers are shown in the respective two dimensional data spaces representing the input training variables [H_s , δ_h , δ_s]. Contours represent the two dimensional data probability distribution with innermost being the most populated. Our prescribed order for the class definition follows ascending values in significant wave height. Data used come from global TOPEX altimeter and WaveWatch 3 model datasets and extend over the entire year 2000. Same data are used for the following 8 figures.

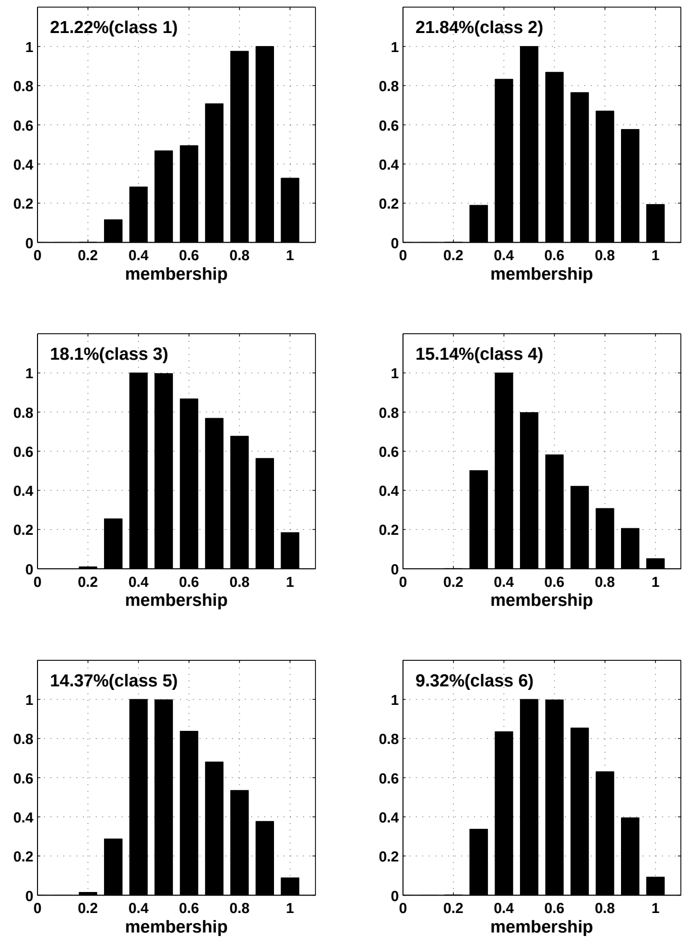

Figure 2. Membership probability for each class where a given input data sample may have membership within multiple classes but the total membership value must add up to one. The percentage of the total data set where input samples have their highest membership value in that class is noted for each class.

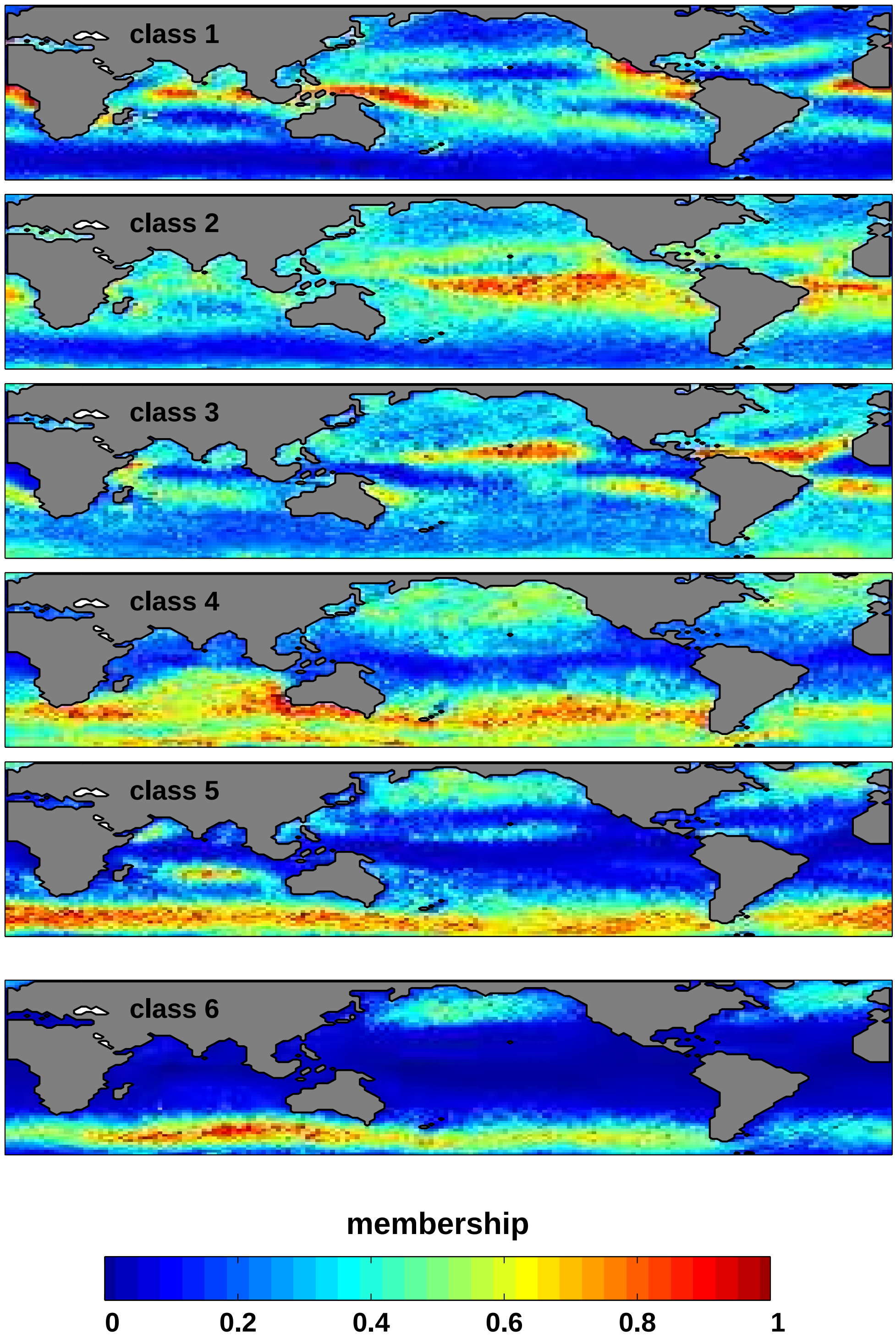

Figure 3. An annual average of the membership value for each class at global scale with the grid being x degrees in latitude and longtitude. Highest membership indicates strong likelihood for this regime to be present at that location.

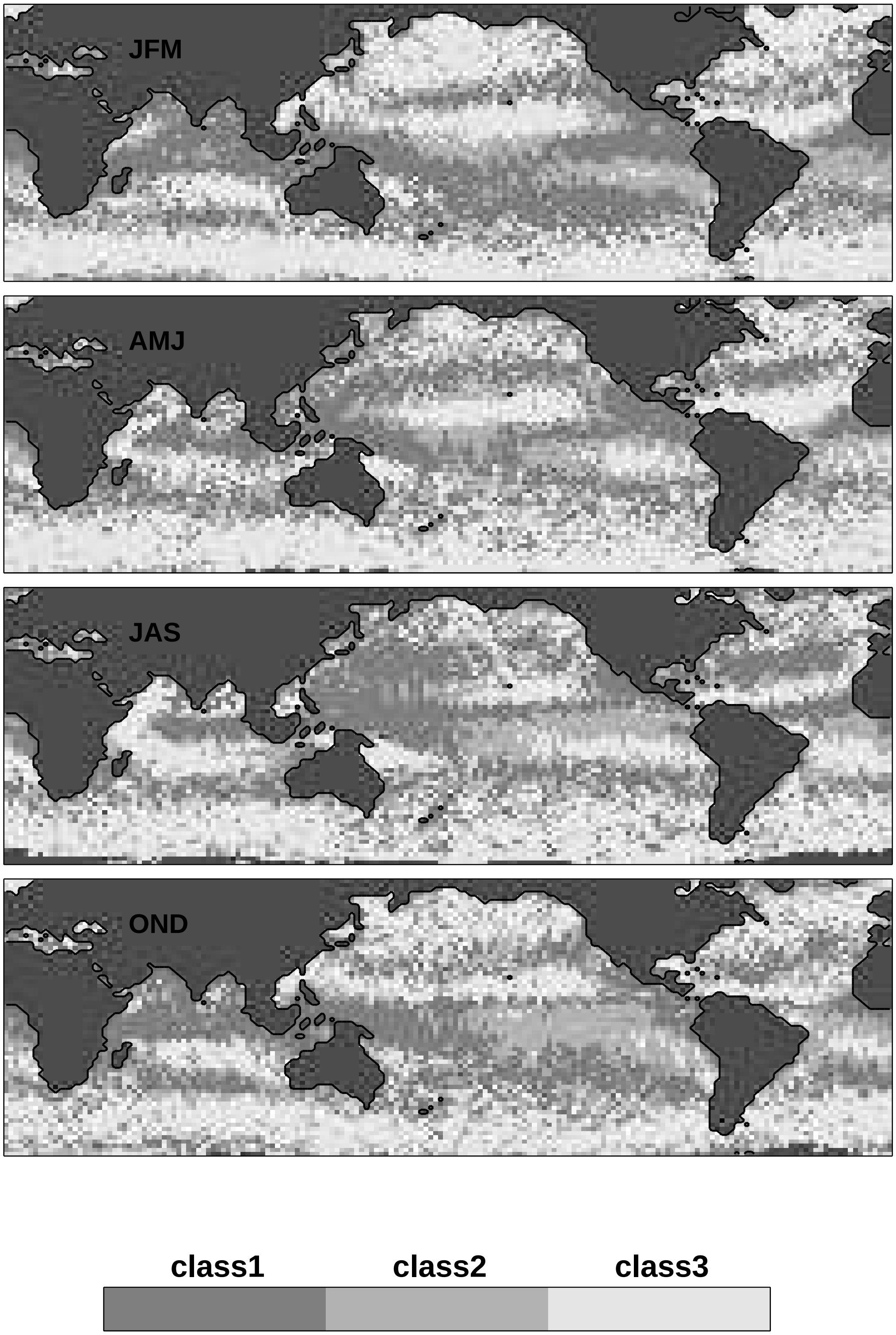

Figure 4. Seasonal mapping for year 2000 that combines classes 1-3 membership information to show spatial and temporal variation. Each color depicts the dominant class for that location and season as determined by a simple maximum total value criterion.

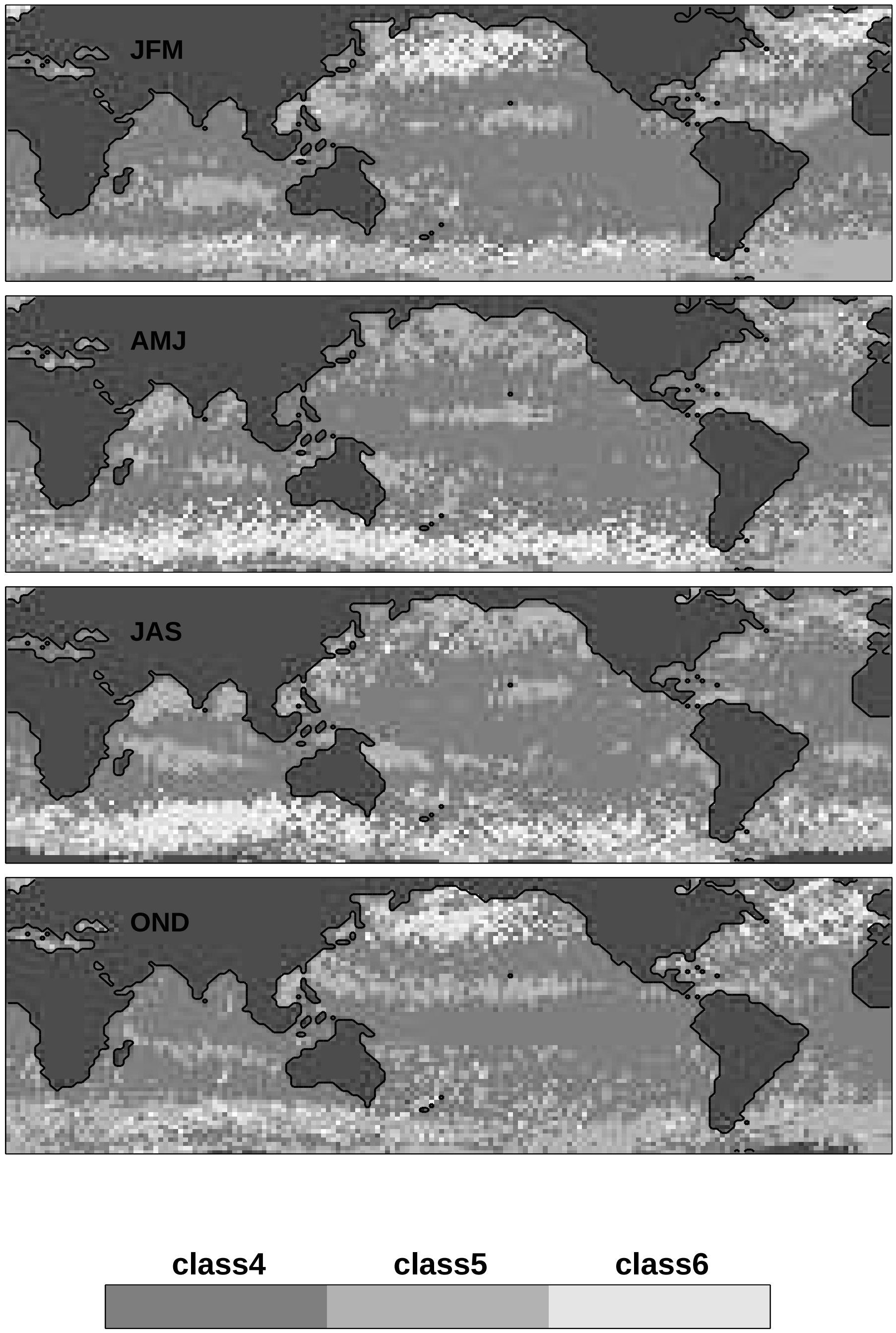

Figure 5. Seasonal mapping for year 2000 that combines classes 4-6 membership information to show spatial and temporal variation. Each color depicts the dominant class for that location and season as determined by a simple maximum total value criterion.

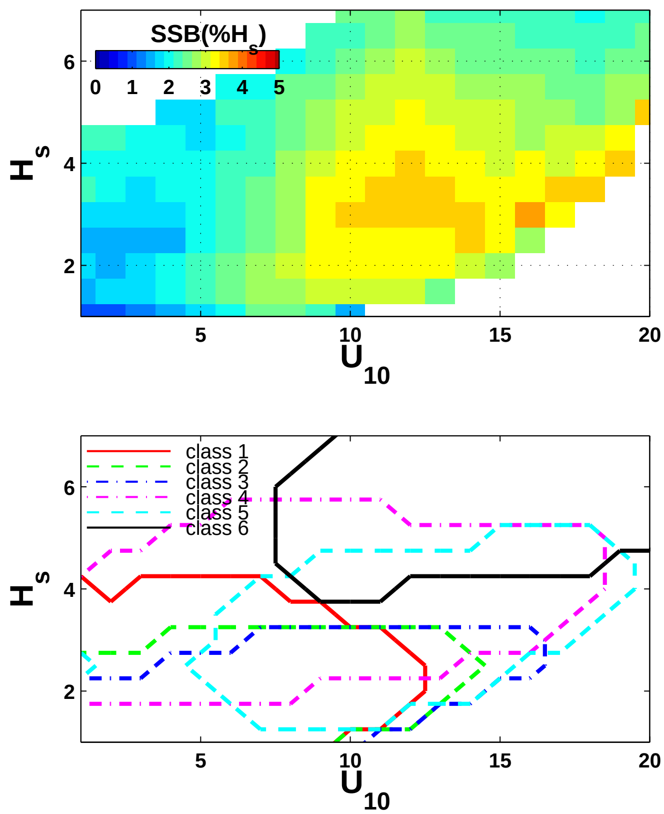

Figure 6. First panel is the global and annual average result for the value of the altimeter sea state bias given as a percentage of H_s versus the two standard altimeter correlatives H_s and wind speed. Second panel shows the 90 % isopleth wherein the majority of data samples for each of the respective six subsets from the data classification lie in this two dimensional data domain.

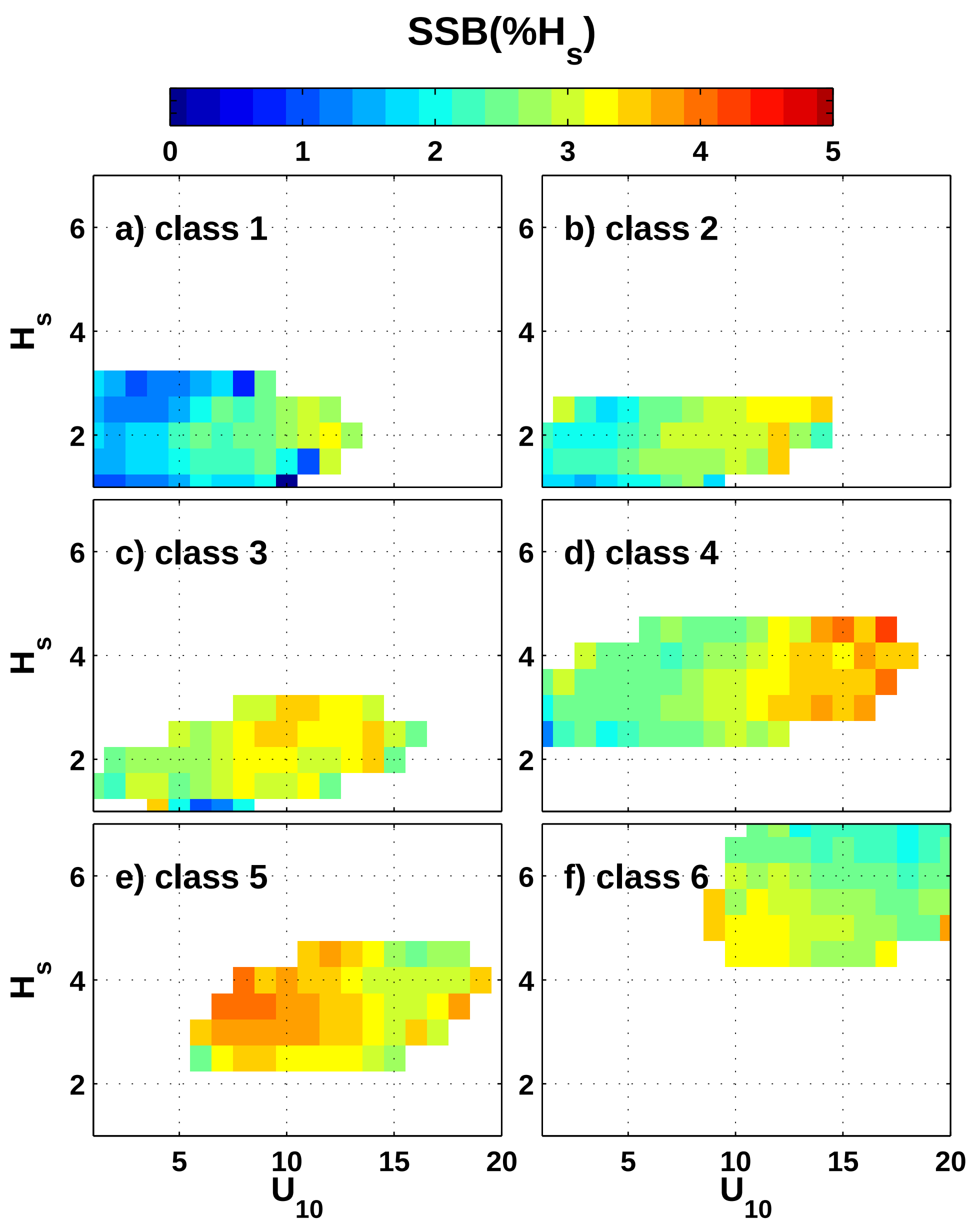

Figure 7. Mean sea state bias values as a percentage of H_s as for Fig. 6 but now computed only with the input samples having highest membership in the respective classes, i.e. these are hard threshold sea state bias class-based models. Results for classes 1-6 are shown. Values are only provided for bins in the data space having more than 100 samples.

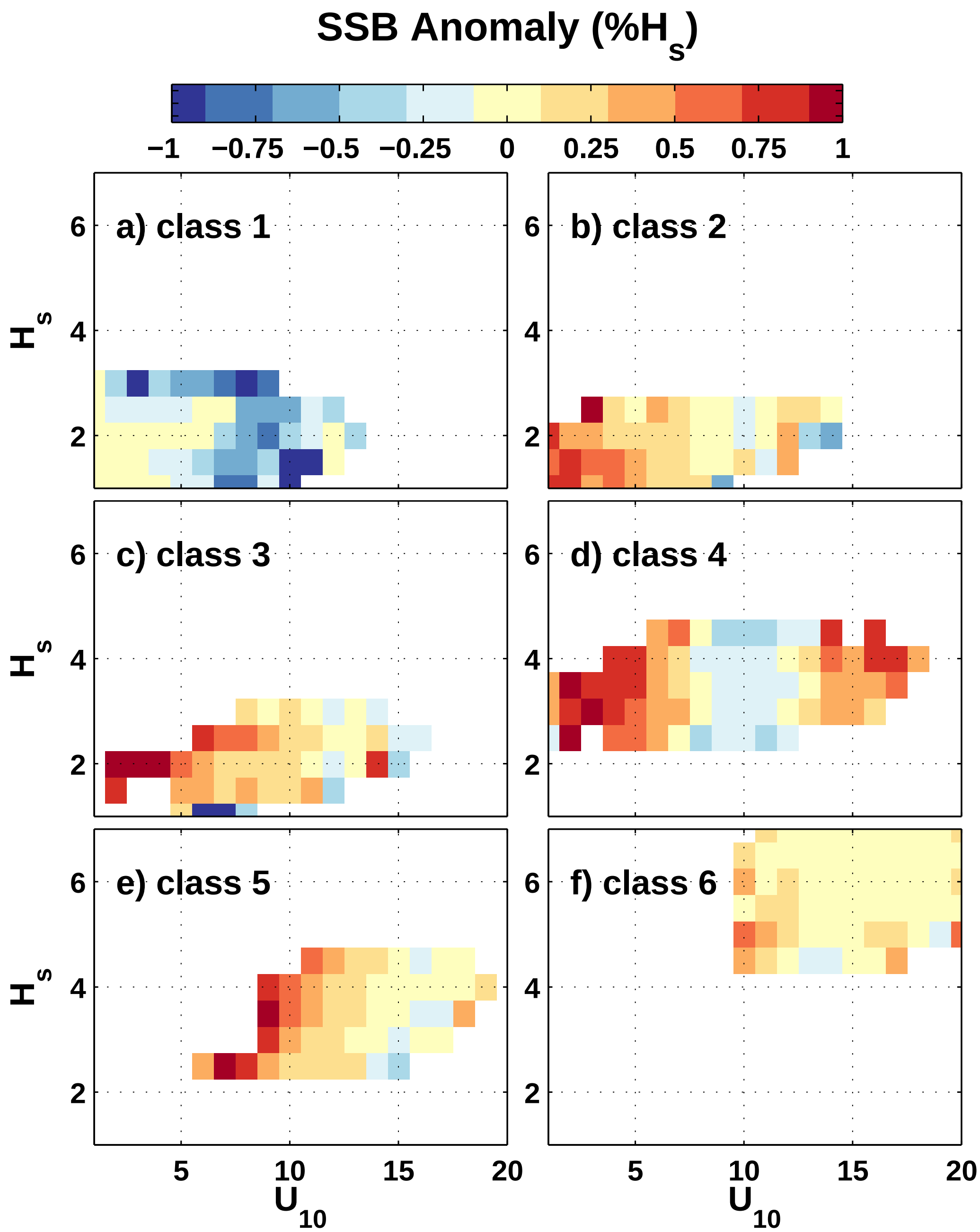

Figure 8. Differences between the global and class-based sea state bias models for each of the six regimes. The intensity represents departure or an anomaly from the global average SSB model of Fig. 6 for the given value of wave height and altimeter wind speed.