Chapter 4

WHITECAP COVERAGE DATABASE

Sunlight, moonlight,
Day light, night light,
There's no greater sight
Than whitecaps at their height.

Cherokee poem, *WHITECAPS*

A new method has been developed for estimating whitecap coverage, W, on a global scale using the satellite-measured brightness temperature of the ocean surface. Whitecap coverage evaluated with this method incorporates various environmental effects. An extensive database of W and concomitant measurements of environmental variables can be compiled with the new method. This database could be useful for analyzing spatial and temporal characteristics of oceanic whitecaps, and assessing their role for various processes at the air-sea interface. Most importantly, this database allows new dependencies of whitecap coverage to be parameterized. This chapter gives an account of the content, organization, and use of the whitecap coverage database.

4.1 Database content and organization

The new method, described in detail in Chapter 3, is applied to satellite and *in situ*-measured data for T_B, T_s, U_{10}, S, and V for all days of 1998 to obtain daily maps of global whitecap coverage, W.

The number of pixels useful for *W*-estimation after applying all necessary masks to the data (wind, clear sky, rain, and T_s availability), ranges from a minimum of about 6,700 in November to maximum of 27,000 in October. This constitutes 4% to 17% of all pixels representing the oceans on a 0.5°×0.5° global map.

Minimum occurrences of conditions yielding unrealistic negative values for *W* happened in November with about 0.5% of all useful cells. The maximum number of such cases occurred in July, when *W*-values were negative for 7% of all useful cells.

The minimum number of cells discarded due to large relative error is encountered in July, about 2.5% of all useful cells. The maximum number of cases with large errors is registered in May, about 10.5% of all useful cells.

Taking into account the removal of negative and erroneous estimates, 83% to 96% of all cells available for retrieving *W* give acceptable values. On average, the number of credible *W*-estimates is about 14,000 per day—a wealth of values for *W* in comparison to the 500 points *in situ* measurements obtained for years. Calculations for the entire 1998 give more than 5 million reliable entries for the whitecap coverage database.

The whitecap coverage database is organized into two major parts (Table 4.1). One part comprises *W* and the most significant intermediate variables resulting from the new method. An examination and analysis of the data comprising the first part of the whitecap coverage database provides insights for the spatial and temporal characteristics of global oceanic whitecaps. Another part of the database groups *W* values and their errors with simultaneously measured environmental variables. The data stored in the second part of the whitecap coverage database is used for parameterizations of whitecap coverage in terms of various environmental factors.

Table 4.1 Organization of the whitecap coverage database

Part 1: *W* estimates from satellite data			**Part 2: *W* estimates to parameterize**		
Variable	Records	Files	Variable	Records	Files
Daily W Map	Array (720×360) Image	365 HDF 1.1 Mb	W, σ_W, U_{10}, T_s, S, $r.e._W$	Matrix (6×259200)	365 Binary 6 Mb
Monthly mean W Map Num. used pxls	Array (720×360) Image Array (720×360)	365 HDF 10.4 Mb	Only "good" data for W, s_W, U_{10}, T_s, S, $r.e._W$	Matrix (6×≈14000)	365 Text ≈ 1.1 Mb
W, s_W, e, s_e, e_s, s_{es}, e_f, s_{ef}, De_r, s_{er}, e^{ℓ}, e^2, e^{ℓ}_f, e^2_f, composite mask	An array (720×360) for each variable	356 HDF 15.3 Mb	W, s_W, U_{10}, T_s, S, $r.e._W$ for 20 salinity bins	Matrix (6×<100 to ≈2500)	20×365 Text 6 to 130 kb
			W, s_W, U_{10}, T_s, S, $r.e._W$ for 20 salinity bins divided into 12 temperature bins and 4 observational sets (8 days)	Matrix (6×0 to ≈2500)	4×12×20 Text 0 to 400 kb
Total: 1095 HDF files in 12 directories (Jan. to Dec.)			Total: 365 binary and 8625 text files in 12 directories (Jan. to Dec.) each containing 20 directories (S1 to S20)		

All files in the first part of the database are HDF. An array and an image of retrieved W are saved for each day in a separate HDF file. Arrays with monthly averaged W values and the number of pixels processed for each location during the month are saved in another HDF file. Intermediate results for each day are saved in yet another HDF file as multiple records of 720×360 arrays, including the standard deviation, s_W, of the retrieved W, the composite mask, the emissivities e, e_s, Δe_r, and e_f with their corresponding standard deviations, as well as the real and imaginary parts of the specular and foam dielectric constants (e', e'', e_f' and e_f''). All files are organized in 12 directories, one for each month.

To proceed with various parameterizations of W, it is necessary to organize the retrieved W values and corresponding concomitant measurements of wind speed, U_{10}, sea surface temperature, T_s, and salinity, S, into a form convenient for processing by different algorithms and software. Binary and text files store the data in such a form and comprise the second part of the whitecap coverage database (Table 4.1, Part 2). This is implemented in two steps.

First, a set of programs in the *Transform* environment organizes and saves 0.5°×0.5° global maps of six variables into one matrix, namely W, s_W, $r.e._W$, U_{10}, T_s, and S. Daily data for each of these variables are in 360×720 matrices. Each matrix is re-arranged into a single column having 259,200 rows. The re-arrangement proceeds column-wise from left to right. It starts from the leftmost column, i.e., the North-South line at 180° W, and stacks each next North-South line below the previous one. The last 360 rows in this long column are the pixels for the North-South line at 179° E (180° E coincides with 180° W). The six variables produce six such long columns. These are placed in one 6×259,200 matrix. This matrix contains all pixels mapping the world on a 0.5°×0.5° grid, including the pixels representing land and the pixels flagged by the masks. One matrix, saved as a binary file, stores the daily data of all six variables. For the entire 1998, 365 such *complete* matrices are saved in 12 directories, one for each month (Table 4.1, Part 2).

Second, a *Matlab* program further re-organizes these complete daily matrices. Initially, only the pixels useful for retrieving W are extracted, i.e., the pixels representing land and the pixels flagged by the masks are "squeezed out" of the matrices by taking only the rows with useful W. The wind speed values, ranging from 3 to 35 m s^{-1} after applying the wind mask, serve as a reference for the selection of these

pixels. The new *reduced* matrices still contain the negative and erroneous W estimates. Thus, the next step in the re-organization is further reducing these matrices by taking only the pixels with reliably retrieved W. The result of this last step is a matrix for each day containing all "good" data for the 6 variables of interest. These final matrices all have 6 columns, one for each variable, but are of different lengths, which are determined by the number of pixels useful for retrieving W for a given day (on average 14,000). The daily matrices with "good" data are organized in 12 directories, one for each month (Table 4.1, Part 2). These "good" data are subjects to regression analysis (§4.3). Though keeping the unrealistic and erroneous W-values in the complete matrices may seem redundant, their storage could be useful for other analyses and comparisons. For instance, analysis of the environmental variables accompanying the negative W values may help to find a pattern for conditions restricting the estimate of W. These, in turn, can be used to predict locations and events when retrieving whitecap coverage reliably is not possible.

4.2 Global whitecap coverage

4.2.1 Previous and new estimates of global whitecap coverage

Estimates of whitecap coverage, as well as their global distribution and seasonal changes, have been made previously using $W(U_{10})$ relations similar to (2.3) but with different exponents (recall §3.4.2).

Blanchard (1963) first combined a quadratic dependence of W on U_{10} with a map of mean winds over the world ocean to obtain global whitecap distribution for summer and winter. The latitudinal variations of Blanchard's (1963) evaluation show maxima in the whitecapping (4% to 10%) at the mid latitudes (30° to 60°) and 2% to

4 % whitecap coverage for low latitudes (up to 30° North and South of the Equator). Blanchard (1963) estimated annual mean global whitecap coverage of 3.4%. The annual means in the Northern and Southern hemispheres (NH and SH) are 2.9% and 4.0%, respectively. These values are high compared to other evaluations (Spillane et al., 1986; Erickson et al., 1986), and can be considered as an upper limit of global whitecap coverage.

Spillane et al. (1986) produced world whitecap atlas using monthly averaged wind speed and atmospheric stability observations compiled by the National Climatic Data Center (NCDC) at Asheville, North Carolina from ship cruises. They employed a power law for the $W(U_{10})$ dependence pieced for cold, moderate, and warm sea surface temperature. Maps of global whitecap coverage for each month show maximum whitecapping, above 4%, in North Atlantic beyond 40° during winter months. Whitecap coverage in the Southern Ocean does not exceed 2% even during the austral winter. At low and mid latitudes (up to 40° N and S) the whitecap coverage never exceeds 1%. These values are low compared to Blanchard's (1963) estimates, and can be considered as a lower limit of oceanic whitecap coverage.

Erickson et al. (1986) also present seasonal estimates of global oceanic whitecap coverage computed with (2.3) and monthly mean surface wind speeds over the world ocean from the NCDC. Global maps for the four seasons show that the highest whitecap coverage, usually up to 4% and sporadically above 5%, is observed in mid and high latitudes (beyond 30° N and S) during winter and fall in each hemisphere. The low-latitude areas in both hemispheres have whitecap coverage less than 1%. These estimates fall in between the upper and lower limits for W established by Blanchard (1963) and Spillane et al. (1986).

In estimating global whitecap coverage, all authors (Blanchard, 1963, 1983, 1985; Spillane et al., 1986; Erickson et al., 1986) discuss possible underestimation of W values for two reasons. First, ships usually avoid areas with stormy weather leading to under sampling of high-wind-speed conditions. This is the most probable cause for observing low estimates of W in the Southern Ocean in Spillane et al.'s (1986) whitecap atlas. Second, due to the non-linearity of the $W(U_{10})$ dependence, the use of *mean* winds, usually monthly means, instead of *instantaneous* winds, also yields underestimation of the whitecap coverage. Blanchard (1963, 1983, 1985) and Erickson et al. (1986) account for this problem by correcting their W estimates for the standard deviation of the wind speed.

Figure 4.1 demonstrates this underestimation well. The figure plots the difference between monthly averaged values of whitecap coverage, ΔW_U, for March 1998 calculated with (2.3) in two ways. In the first way, the monthly mean W is obtained using the mean wind field for March 1998, ΔW_{Uavr}. In the second way, the monthly mean W is obtained by averaging 31 daily W maps computed from 31 daily wind fields, ΔW_{Uins}. In the figure, red represents cases when $W_{Uins} > W_{Uavr}$, while blue represents cases of $W_{Uins} < W_{Uavr}$. This difference map is mostly reddish, proving that using mean instead of instantaneous winds to estimate W usually leads to underestimation of the global whitecap coverage. The underestimation is well noticeable in the Southern Ocean, North Pacific, and North-East Atlantic.

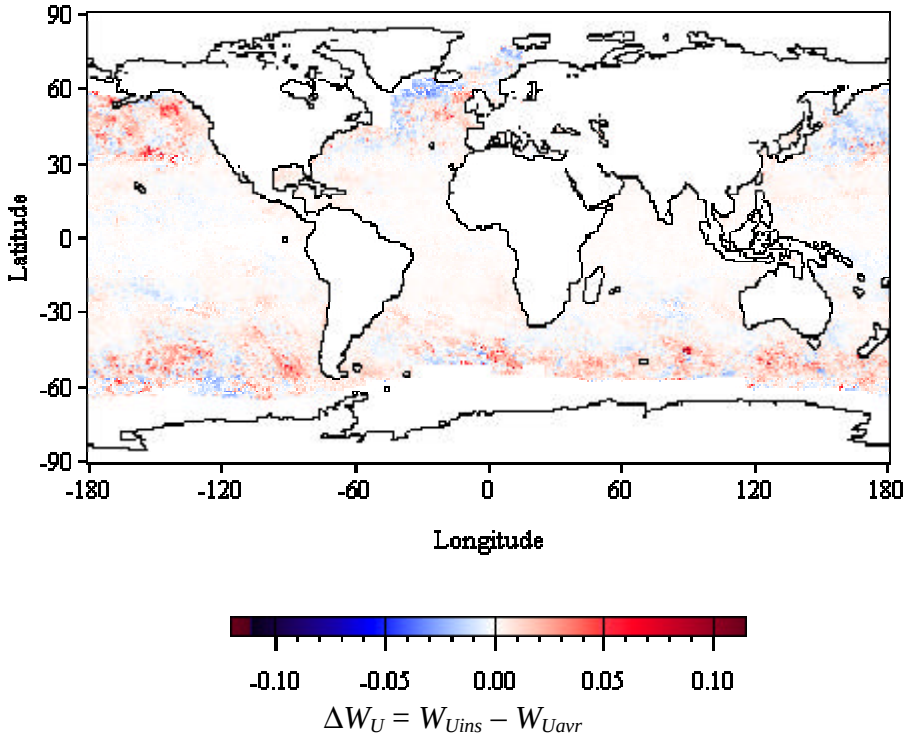


Figure 4.1 Difference in the mean whitecap coverage values, ΔW_U, for March 1998 computed with wind-speed formula (2.3) using monthly mean wind field for March 1998, W_{Uavr}, and 31 daily wind fields, W_{Uins}, for March 1998. Red represents cases of $W_{Uins} > W_{Uavr}$; blue represents cases of $W_{Uins} < W_{Uavr}$.

In this study, whitecap coverage is calculated from daily, in a sense instantaneous, fields of satellite data. The daily values of the whitecap coverage are averaged to obtain monthly, seasonal, or annual means of *W*. These latter are used to investigate the spatial and temporal characteristics of the global whitecap coverage retrieved with the new method.

Global *W* maps for each month of 1998 are archived in the whitecap coverage database. Averaging these monthly-mean maps results in an annually

averaged global map of *W*. From this single map, globally averaged annual *W* obtained with the new method is estimated to be 3.05%.

These *W* values are consistent with previous estimates, especially those of Blanchard (1963). But the validation on more regional scales (§3.4.3) showed that the new-method estimates are usually higher than the *in situ* measurements. No doubt, one possible reason for these higher values is that the satellite retrievals may need tuning and improvement of the computational algorithm. But it is also possible that the new method gives higher estimates of *W* because it measures both stages (A and B, i.e., active and decaying) of the whitecaps and uses daily instead of monthly data.

4.2.2 Spatial characteristics of satellite-measured whitecap coverage

The global spatial distribution of mean oceanic whitecaps for March 1998 obtained with the new method is shown in Figure 4.2. Whitecap coverage over most of the world ocean is up to 4%. The lowest whitecapping, from less than 1% up to 2%, is evident along the equator, on the western edges of the continents, and east of the tip of South America. The highest whitecapping, up to 6%, is observed in the zonal belts of the trade winds (15° to 30° N and S) and the prevailing westerlies (30° to 60° N and S). The average whitecap coverage for March 1998 in the NH is 3.2%, and in the SH it is 2.7%.

The spatial distribution of satellite-derived *W*, as already noticed in §3.4.2, is quite different from that obtained with the wind-speed formula displayed in Figure 4.3a. Let us compare Figure 4.2 and 4.3a. As $W(U_{10})$ relation suggests, all previous evaluations inevitably predict high whitecap coverage in high latitudes, around and beyond 50-60° N and S, where the highest winds blow, whereas in mid and low latitudes the whitecapping is rarely above 1%. The new method derives higher

whitecapping in mid and low latitudes compared to wind-formula predictions. In this sense, the spatial distribution of whitecap coverage from satellite data, $W(U_{10}, \Delta T, T_s, S, f, d, C)$, does not vary as much as the $W(U_{10})$ relation implies.

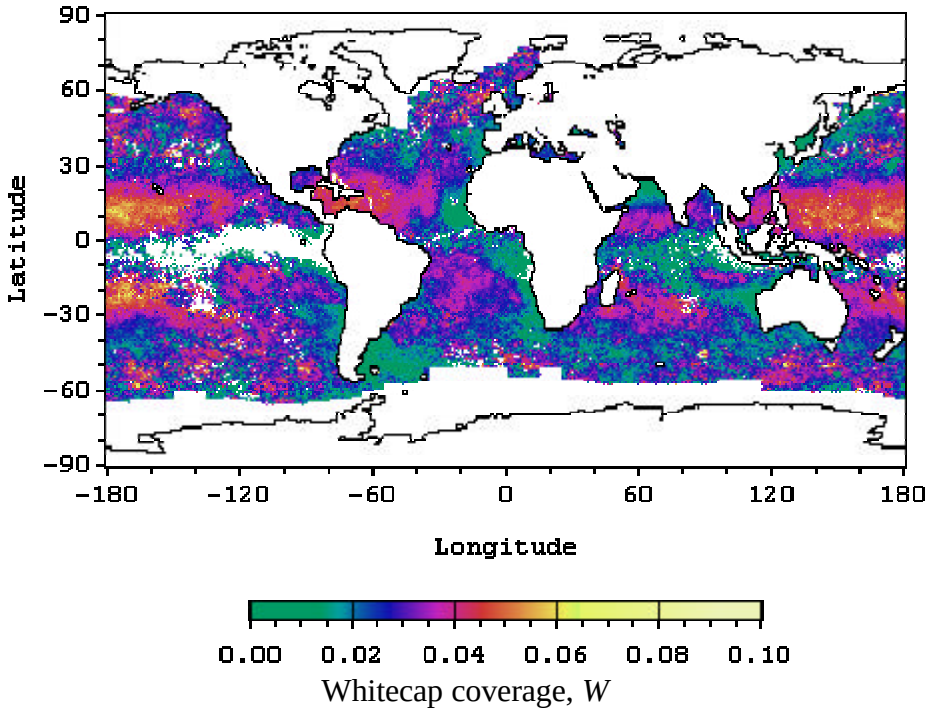


Figure 4.2 Mean whitecap coverage for March 1998 obtained with the new method (average of 31 daily maps of *W*).

The spatial differences between new method results and wind-formula predictions are quantified in Figure 4.3b, which plots $\Delta W = W_{\text{new method}} - W_{\text{wind formula}}$. In the figure, red shows positive ΔW, i.e., $W_{\text{new method}} > W_{\text{wind formula}}$, and blue shows negative ΔW. Systematically, the wind formula gives higher W values in high latitudes, 6% to 10% (blue in the figure), while the new method estimates higher W,

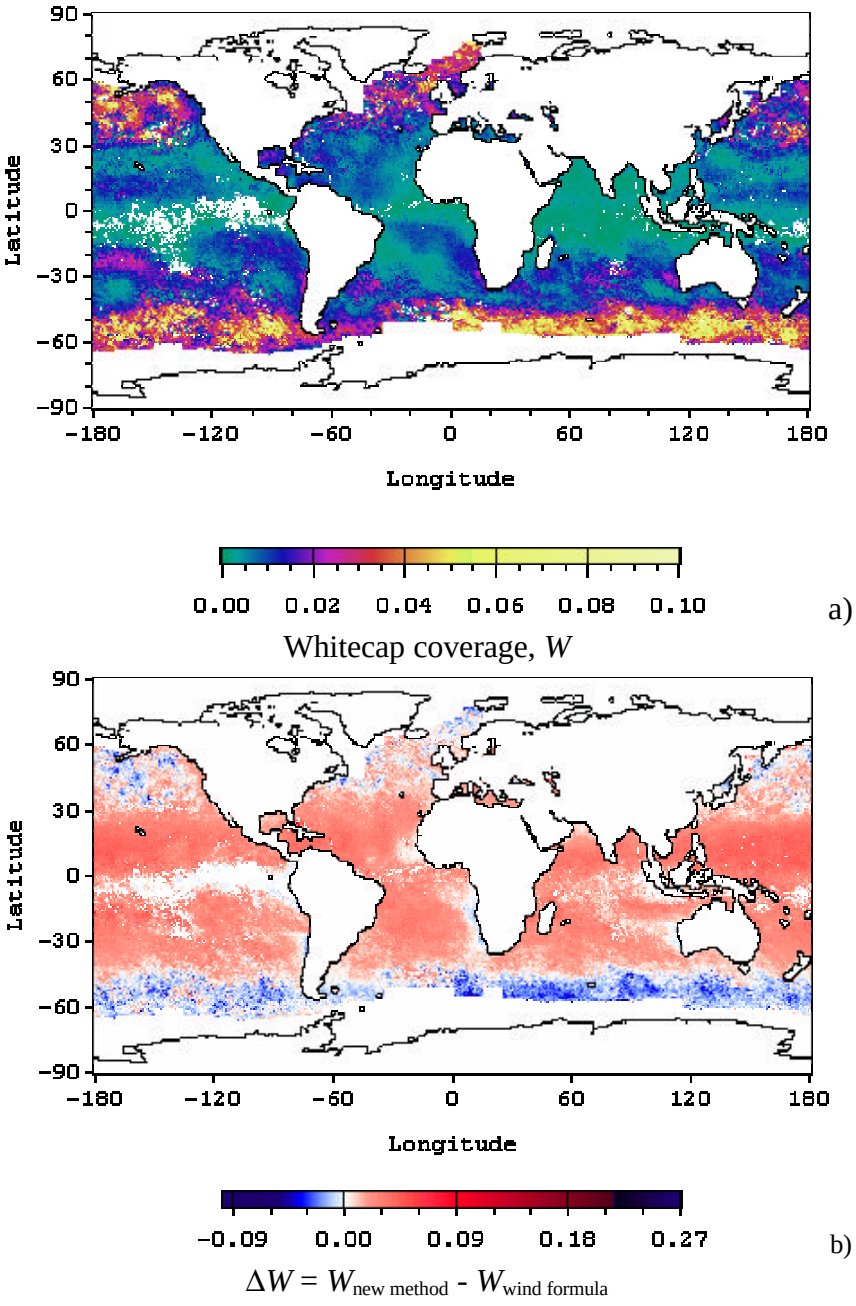


Figure 4.3 **a) Mean whitecap coverage for March 1998 computed with wind-speed formula (2.3) and daily values of wind speed; b) Difference ΔW between the mean W-values for March 1998 obtained with the new method and wind-speed formula: red shows $W_{\text{new method}} > W_{\text{wind formula}}$; blue shows $W_{\text{new method}} < W_{\text{wind formula}}$.**

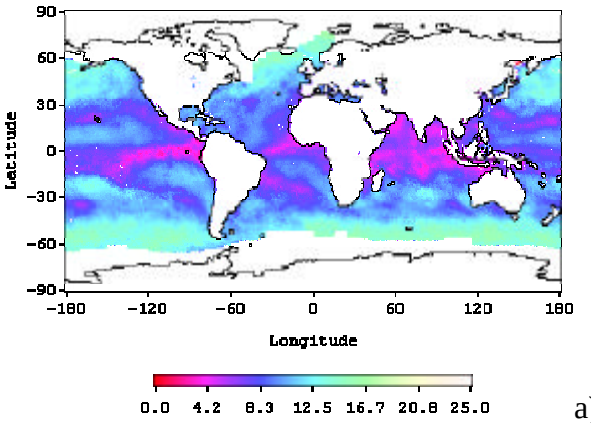
6-8% (red in the figure), in mid and low latitudes. The magnitude of ΔW rarely exceeds ± 5%.

The largest differences between the two computational methods are in a few places. Though realistic as absolute values, these seem inconsistent with the neighboring values. Such high and inconsistent W values are not only present in "difference" maps as the one shown in Figure 4.3 (not seen because of the land contours), but also in the averaged (monthly or annually) maps of W. All positive large differences are along the edges of the continents; the large negative differences are in open ocean. Large ΔW in the open ocean could be due to the interplay between wind speed and other environmental variables accounted differently by the two computations. There are two possible explanations for the large ΔW along the continents. First, they could be due to the flawed performance of SSM/I close to land, which is not noticeable in the daily maps, but "accumulates" in the averaged maps. Second, they could be a real physical feature due to either lower salinity in the coastal zone or more vigorous whitecapping due to topography, both accounted for by the new method and not included in the wind formula. Finally, it is conceivable that they are artifacts left from the problem with re- sampling the salinity maps (§ *WOA98 S data*). Whatever the reason, however, the places with large ΔW are so a few, up to 1% of the useful values, that most probably they would not have any statistical significance when W data are used for different regressions.

Though not large in magnitude, the differences in the W values point toward different physical processes affecting the whitecapping. The most plausible explanation of the differences is that the new method takes into account not only the

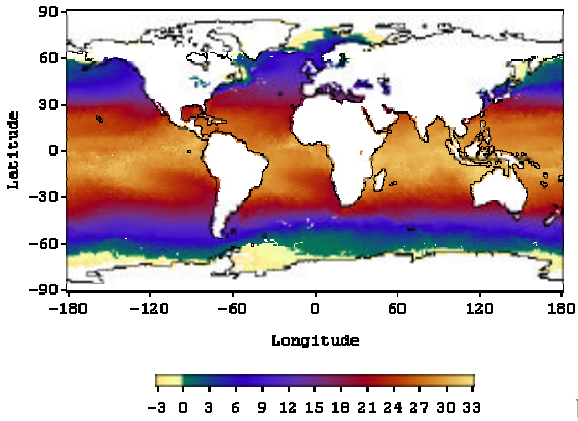
influence of wind speed on W, but also the effects of other environmental variables. Monthly maps of U_{10}, T_s, and S for March 1998 are plotted in Figure 4.4. The spatial features of U_{10}, T_s, and S help to track their impacts on the global W distribution in Figures 4.2 and 4.3a. For instance, areas with high winds in mid and high latitudes (greenish pixels in Figure 4.4a) can be easily recognized as places with high W in Figure 4.3a. But while in Figure 4.3a high whitecapping in high latitudes is a dominant spatial feature, in Figure 4.2 the whitecapping in the high latitudes is suppressed by the lowest sea surface temperatures, from below 0°C up to 3°C (see Figure 4.4b). The Southern Ocean and the northern reaches of the North Atlantic and North Pacific are good examples of this observation. Meanwhile, the effect of moderated winds, up to about 12 m s^{-1}, in mid and low latitudes is enhanced by higher sea surface temperature, so that higher than expected whitecapping is observed. This is especially evident in the tropics in all oceans. The spatial features of the salinity field in Figure 4.4c show that high salinity, e.g., in mid latitudes in Atlantic and Indian Oceans, probably enhances whitecapping.

The influence of T_s on W—suppressing the effect of higher U_{10} in high latitudes and enhancing the effect of lower U_{10} in mid and low latitudes—is expected. As T_s decreases, W decreases due to changes in water viscosity, which, in turn, invokes changes in the effectiveness of turbulent dissipation of the wind energy (Monahan and O'Muircheartaigh, 1986). Water viscosity increases in cold waters, and this lowers the frequency of wave breaking and whitecapping since the turbulence effectively dissipates the excessive wind energy. The opposite is true for warm waters where lower water viscosity makes the turbulent dissipation less effective, which therefore



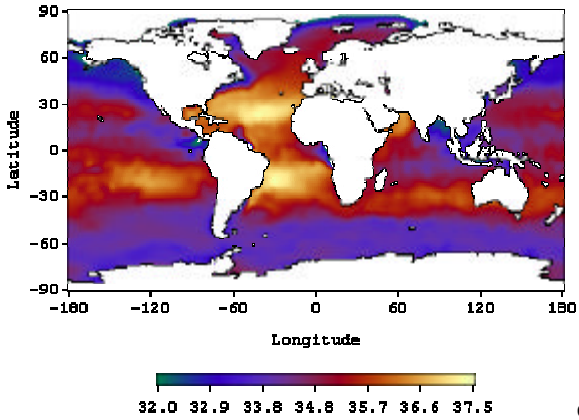
a)

Wind speed, U_{10} (m s^{-1})



b)

Sea surface temperature, T_s (°C)



c)

Salinity, S

Figure 4.4 Monthly mean fields (March 1998) for: a) wind speed; b) sea surface temperature; c) salinity.

requires more wave breaking and whitecap formation in order to dissipate the wind energy.

This effect of T_s on W explains the slowly changing trend of satellite-measured W in Figure 3.20 and is in agreement with previous considerations. Monahan and O'Muircheartaigh (1986) and Spillane et al. (1986) have observed a decrease in the exponent in the $W(U_{10})$ relation for cold water. Bortkovskii (1987) (his fig. 2.5) also shows a strong positive dependence of W on T_s: at the same wind speeds whitecapping in warm waters is higher than in cold water. Accordingly, he proposed a linear dependence of W on U_{10} in cold waters. Thus, the expected behavior of W with changes in T_s seems to be correctly reflected by the new method.

More speculations can be added to explain the high whitecap coverage in mid and low latitudes. Most certainly, the persistent trade winds and the long fetches in the tropics foster whitecapping there. Also, in the central parts of the oceans the water is oligotrophic. A low concentration of organisms leads, presumably, to a low concentration of surface-active materials, thus wave breaking is not suppressed by surfactants and whitecapping is enhanced. In contrast, coastal upwelling west of Africa and South America, as well as high primary production in the Southern Ocean, could lead to more surfactants on the ocean surface, which would hinder wave breaking and diminish whitecapping.

In summary, the new method estimates higher and more evenly distributed whitecapping over the globe compared to the predictions of the $W(U_{10})$ relation.

4.2.3 Temporal characteristics of satellite-measured whitecap coverage

Figure 4.5 shows global maps of whitecap coverage for the four seasons. During the boreal winter (December-January-February), maximum whitecap coverage

is observed in the NH (panel a in Figure 4.5). In spring (March-April-May), whitecapping in the northern parts of the Atlantic and Pacific decreases, while there is an overall increase of *W* southward (panel b). In the boreal summer (austral winter) (June-July-August), maximum whitecap coverage is registered in the SH. Whitecap coverage in trade winds belts in both hemisphere remains about 3-5% (panel c). In the fall (September-October-November), there is 2-3% whitecapping in high latitudes (beyond 30° N and S), and about 3-6% in the tropics (panel d).

Monthly values of whitecap coverage for both hemispheres are plotted in Figure 4.6a. Whitecap coverage varies from 2.5% to 3.5%, i.e., seasonal changes vary by a factor of 1.4. These variations are less than those estimated by Erickson et al. (1986). This result is consistent with the conclusion of more even whitecap distribution predicted by the new method. The globally-averaged monthly values of whitecap coverage (Figure 4.6b) are less variable, from 2.9% to 3.2%. Yet a seasonal cycle on a global scale is evident. The increase during the boreal summer (austral winter) is mostly due to the high whitecap coverage in the Southern Ocean, emphasizing its importance for air-sea processes involving whitecaps, such as sea-salt aerosol production, gas exchange, and planetary albedo.

4.2.4 Implications of whitecap coverage estimates

Besides being responsible for the production of sea-salt aerosols, oceanic whitecaps are involved in various processes at the air-sea interface. Two of these processes—contribution of whitecaps to changes in planetary albedo and air-sea exchange of CO_2—are considered in this section.

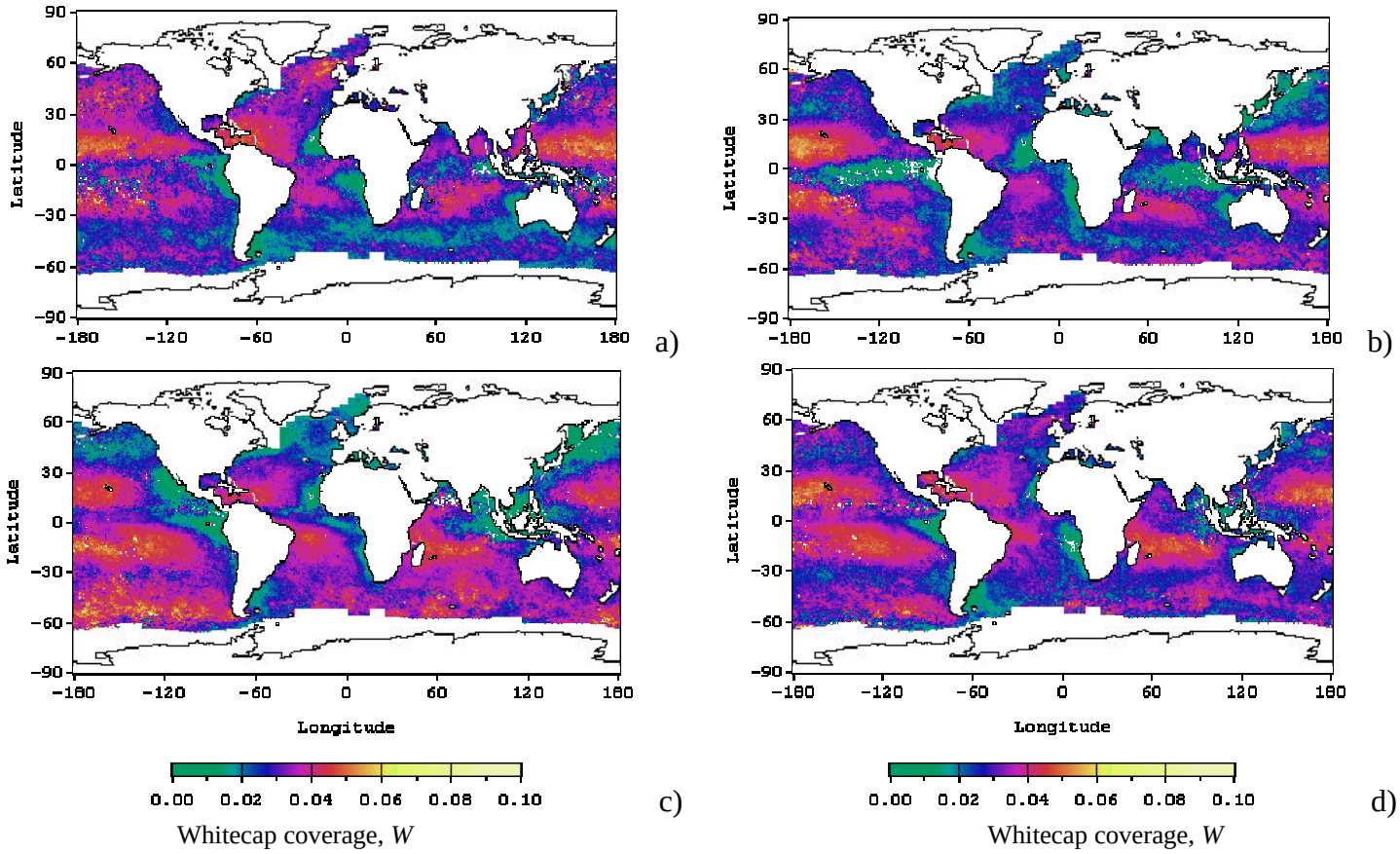


Figure 4.5 Boreal seasonal means of the global whitecap coverage for 1998 (maps 0.5 °×0.5 °): a) winter (Dec-Jan-Feb); b) spring (Mar-Apr-May); c) summer (Jun-Jul-Aug); d) fall (Sep-Oct-Nov).

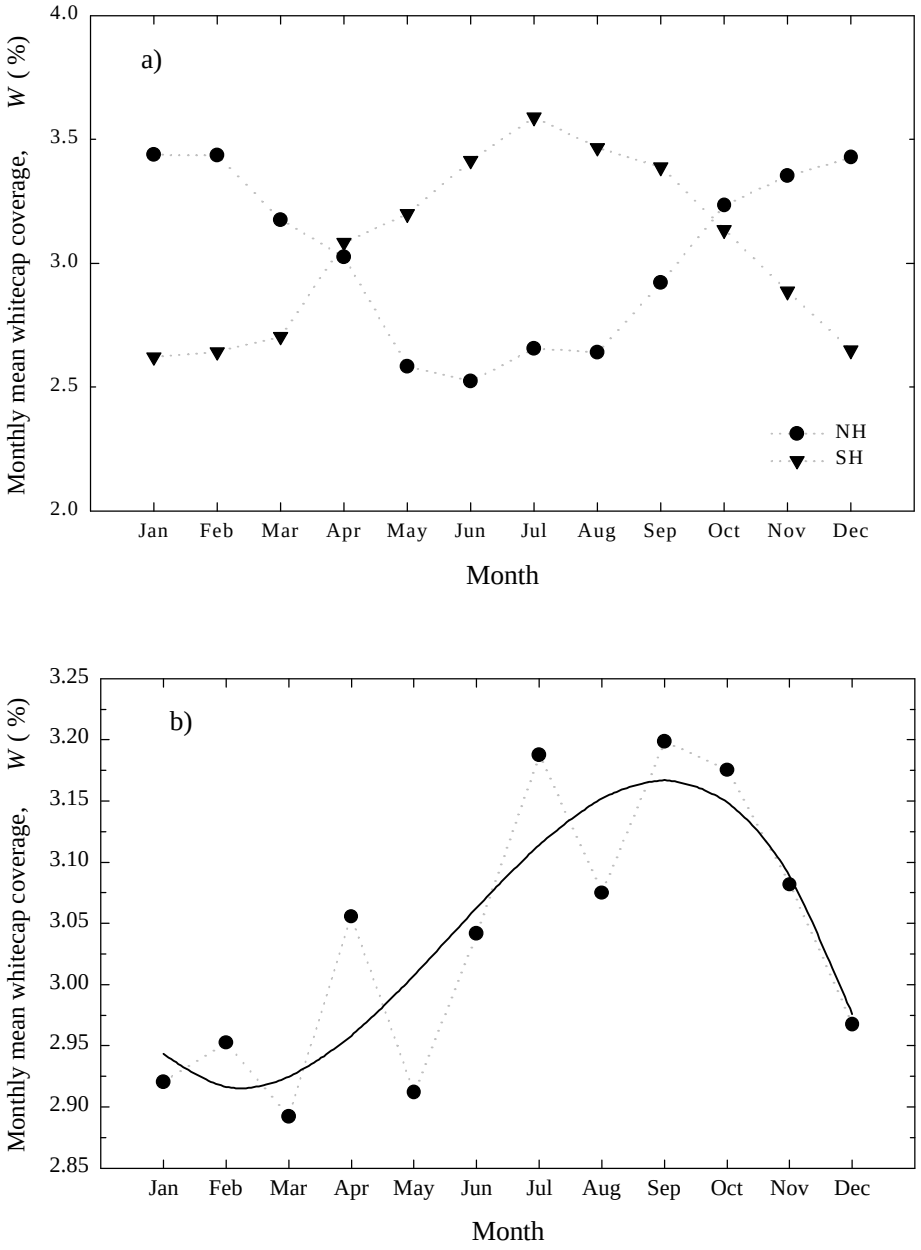


Figure 4.6 Monthly variations of whitecap coverage for 1998: a) in Northern and Southern hemispheres; b) globally.

Whitecaps and ocean surface albedo

We see whitecaps as bright white patches on the ocean surface due to their high reflectivity (§2.5.3). High whitecap reflectivity led to the suggestion that sea foam might influence the shortwave albedo of the ocean-atmosphere system. Gordon and Jacobs (1977) found that an increase in the wind speed from 6 to 14 m s^{-1} would double the surface albedo as highly reflecting sea foam appear. They concluded that newly formed whitecaps could contribute to the radiation budget of the planet.

Frouin et al. (2001) proposed a scheme, convenient for use in climate models, quantifying the effect of oceanic whitecaps on the global radiation budget. They estimated that the globally averaged radiative forcing of whitecaps, defined as changes in the outgoing radiative flux due to changes of surface albedo, lies in the range 0–0.14 W m^{-2} with a probable value of 0.03 W m^{-2}. This forcing could be larger on regional and seasonal scales with radiative forcing values reaching 0.7 W m^{-2} in the Indian Ocean during summer.

Whitecap coverage, *W,* enters the parameterization scheme for the direct radiative forcing of oceanic whitecaps. Frouin et al. (2001) parameterized *W* with wind speed. Because global *W* obtained with the new method exhibits spatial features different from those obtained with wind speed, it is pertinent to re-calculate the radiative forcing of whitecaps.

In quantifying the direct forcing of whitecaps, Frouin et al. (2001) follow the approach used by Charlson et al. (1992) for sulfate aerosols. At any location over the oceans, the change in the outgoing radiative flux due to whitecaps is:

$$\Delta F = F_0 m_0 (1 - A_c) T_a^u T_a^d \Delta R_s$$

where F_0 is the solar irradiance at the top of the atmosphere (solar constant), m_0 is the cosine of the sun zenith angle, A_c is the fraction of the surface covered by clouds, T_a^u

and T_a^d are the clear-sky atmospheric transmissivity for up-welling and down-welling flux, respectively, and R_s is the albedo of the ocean surface. Averaged on a global scale, this equation becomes:

$$\langle \Delta F \rangle \approx 0.5 a F_0 (1 - \langle A_c \rangle) \int_0^1 T_a^u T_a^d \Delta R_s m_0 dm_0 \qquad (4.1).$$

Here, the factor of 0.5 is introduced because any point on the earth's surface is illuminated by sunlight half the time during a year, a is the fraction of annually averaged ice-free oceanic surface, and $\langle\ \rangle$ denotes area averaging over the oceans.

To estimate ΔR_s in (4.1), Frouin et al. (2001) represent the surface albedo, R_s, as a sum of reflections, analogous to (3.1):

$$R_s = (1 - W) R_w + W R_{wc} \qquad (4.2),$$

where R_w is the albedo of foam-free water surface, R_{wc} is the albedo of foam-covered water, and W is the whitecap coverage; when $W = 0$, $R_s = R_w$. The change in surface albedo due to whitecaps then is:

$$\Delta R_s = R_s - R_w = W (R_{wc} - R_w) \qquad (4.3).$$

Plugging (4.3) in (4.1) yields:

$$\langle \Delta F \rangle \approx 0.5 a F_0 (1 - \langle A_c \rangle) \langle W \rangle \int_0^1 T_a^u T_a^d (R_{wc} - R_w) m_0 dm_0 \qquad (4.4).$$

For $\langle W \rangle$, Frouin et al. (2001) use Monahan and O'Muircheartaigh's (1980) relation, $\langle W \rangle = 2.95 \times 10^{-6} \langle U \rangle^{3.52}$, where $\langle U \rangle$ is globally averaged wind speed over a 10-year period. This estimate can be replaced now with a globally and annually averaged estimate of $\langle W \rangle$ for 1998 retrieved with the new method proposed here. As reported in §4.2.2, the global annual value for $\langle W \rangle$ is 0.0305, estimated over the ocean surface representing a fraction $a = 0.46$ of the global surface. The solar constant is

F_0 = 1372 W m^{-2}. For the globally averaged fractional cloud coverage, Frouin et al. (2001) use a climatological value averaged over 7 years (1983 to 1989): $\langle A_c \rangle$ = 0.59. Using a radiation transfer model, Frouin et al. (2001) estimate the value of the integral to be $\approx 2.7\times10^{-2}$. With all these values, the global annual estimate for the direct radiative forcing due to whitecaps amounts to about 0.11 W m^{-2}.

Employing the global annual map of *W* produced with the new method, the spatial changes in the outgoing shortwave flux due to presence of whitecaps are depicted in Figure 4.7. About 99% of the ocean surface provides $\langle \Delta F \rangle$ in the range from 0 to 0.2 W m^{-2} with some values reaching 0.9 W m^{-2}. The largest increase of ocean albedo due to the presence of whitecaps is observed in the trade winds zones of the ocean, yielding an outgoing shortwave flux of 0.16 to 0.20 W m^{-2}. Most of the upwelling zones, east of North America and Asia, and the continental shelf east of the tip of South America have the lowest contribution to reflection of shortwave radiation, up to 0.04 W m^{-2}.

The most probable value for $\langle \Delta F \rangle$ = 0.11 W m^{-2} evaluated with new-method-derived whitecap coverage is a factor of 3.7 higher than that reported by Frouin et al. (2001), 0.03 W m^{-2}. A possible reason for the difference is that the satellite-measured whitecap coverage gives higher *W*-values for mid and low latitudes than the wind formula predicts. When averaged globally, this leads to higher global estimates of $\langle W \rangle$ and $\langle \Delta F \rangle$.

Similarly to the estimate of Frouin et al. (2001), the updated estimate of the direct radiative forcing of the whitecaps is small. It is not negligible, though. Whitecaps, a natural agent for the climate system, provide cooling comparable to that due to anthropogenic agents such as stratospheric ozone (0.18 W m^{-2}), biomass burning

(0.21 W m^{-2}), and land use (0.22 W m^{-2}) (IPCC, 2001). In a warming world with a possible increase of wind speed (Latham and Smith, 1990; IPCC, 2001, p.65), whitecap forcing could increase.

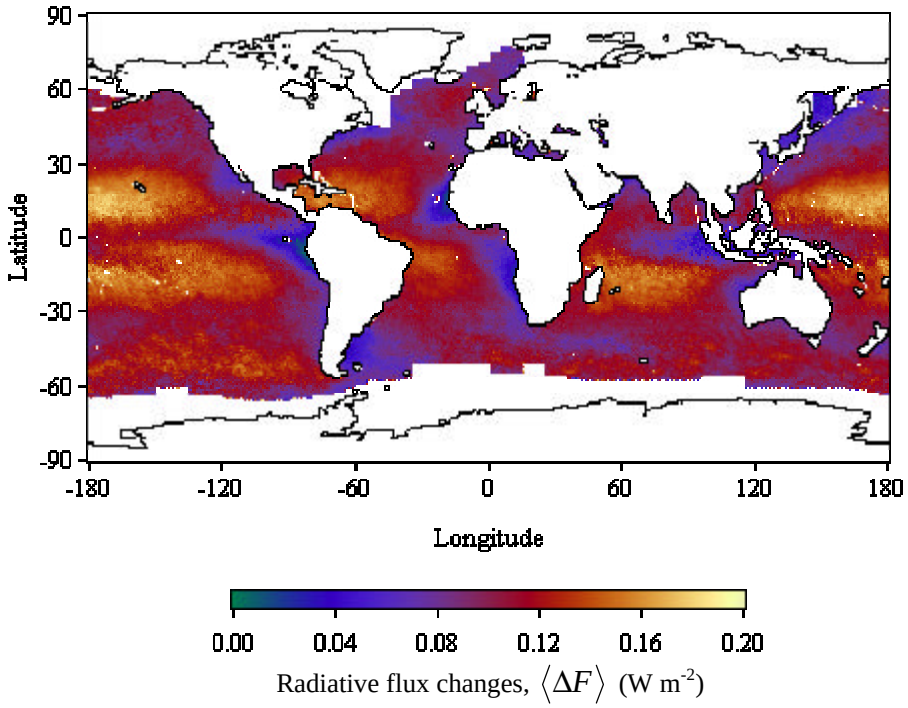


Figure 4.7 Global map (0.5 °×0.5 °) for 1998 of annually averaged changes in outgoing radiative flux $\langle \Delta F \rangle$ due to presence of whitecaps on the ocean surface.

Whitecaps and transfer velocity of CO_2

Accurate estimates of air-sea gas fluxes are essential for understanding the global cycles of CO_2, CH_4, NO_x, DMS and other trace gases that affect the earth's radiation budget (Frew, 1997). The flux, *F* (mass per unit time per unit area), of any

gas across the air-sea interface is proportional to its air-sea concentration difference, ΔC, giving $F = k_T \Delta C$ (Liss, 1983). The coefficient of proportionality, k_T, has the dimension of velocity and is called the transfer (or piston) velocity. Vigorous research is currently conducted to understand all mechanisms possibly contributing to the gas transfer across the air-sea interface, and embody them as accurately as possible in a parameterization of k_T.

Wind, waves, and the formation of bubble clouds, associated with whitecaps, enhance the direct diffusive flux of any gas, including CO_2. The transfer velocity is thus frequently parameterized in terms of wind speed U_{10} (Liss and Merlivat, 1986; Tans et al., 1990; Wanninkhof, 1992). However, the relation $k_T(U_{10})$ is not well constrained by field data (Wanninkhof, 1992). Other factors, besides wind speed, contribute to the efficiency of the gas transfer, presumably sea surface temperature, atmospheric stability, wind speed variability, wind fetch, and surface-active materials. The effects of some of these factors were investigated to explain the scatter in experimental data. Wanninkhof (1992) concluded that variability of wind speed and wind fetch, nonlinearity of the $k_T(U_{10})$ relation, and chemical effects for reactive species such as CO_2 do not fully explain the variance in laboratory and field data.

Monahan and Spillane (1984) parameterized k_T for CO_2 in terms of whitecap coverage, W. Using this parameterization, Erickson (1993) evaluated the effect of atmospheric stability, ΔT, on k_T by representing W as a function of wind speed and the stability-dependent drag coefficient. He found that gas exchange is sensitive to variations in ΔT and suggested that ΔT forcing could account for the scatter in the experimental data. Because global W obtained with the new method incorporates the effects of various environmental factors, an evaluation of $k_T(W)$ may give more realistic

k_{CO2} values and serve as a measure for what to expect when these factors act in concert. Computations outlined by Erickson (1993) are used for this purpose.

Analogously to (3.1) and (4.2), the transfer velocity can be written as a sum of contributions by: i) molecular diffusion over most of the ocean area with no whitecaps; and, ii) turbulent processes within the whitecap-covered area (Monahan and Spillane, 1984):

$$k_T = (1-W)k_w + Wk_{wc} \qquad (4.5),$$

where k_w is the transfer velocity through the foam-free water surface, k_{wc} is the transfer velocity through foam-covered water, and W is the whitecap coverage; when $W = 0$, $k_T = k_w$. Erickson (1993) used $k_w = 5.0$ cm h^{-1} and $k_{wc} = 1300$ cm h^{-1}, determined via several model runs and comparison with data. Since the initial set of k_w and k_{wc} used for this adjusting process was obtained from data for radon (Rn), k_T computed with (4.5) is for Rn, i.e., k_{Rn}. It is necessary, therefore, to convert k_{Rn} to k_{CO2} with the following expression (Erickson, 1993):

$$k_{CO2} = k_{Rn}\left[\frac{Sc_{CO2}(T_s)}{Sc_{Rn}(T_s)}\right]^n \qquad (4.6)$$

with an exponent, n, having values

$$n = -1/2 \qquad U_{10} \geq 3.6 \text{ m s}^{-1}$$
$$n = -2/3 \qquad U_{10} < 3.6 \text{ m s}^{-1} \quad .$$

Here $Sc(T_s)$ denotes sea surface temperature dependence of Schmidt numbers for CO_2 and Rn. $Sc(T_s)$ is derived as a cubic polynomial regression fitted to data:

$$Sc = a + bx + cx^2 + dx^4 \qquad (4.7)$$

with $x = T_s$ in °C and regression coefficients listed in Table A.5 (Appendix A).

For W, Erickson (1993) used the expression proposed by Wu (1975), $W = 0.2u_*^3$, where the friction velocity can be computed as $u_* = C_D^{0.5} U_{10}$. The effect of atmospheric stability, ΔT, on k_{CO2} is introduced via the drag coefficient, C_D, which can be determined from the bulk aerodynamic method. All environmental variables, such as wind speed, humidity, and air and sea surface temperatures, used in the calculations are 5-year model simulations. Erickson (1993) presents the results for k_{CO2} in annual and monthly (January and July) maps of k_{CO2}.

This study estimates annual k_{CO2} for 1998 in three steps: i) k_{Rn} is calculated with (4.5) using the chosen values for k_w and k_{wc} and new-method-derived global annually (or monthly) averaged maps of W; ii) the Schmidt numbers for CO_2 and Rn are estimated with (4.7) employing annual (or monthly) mean fields of T_s; iii) k_{CO2} is evaluated with (4.6) using the results from steps i) and ii). Due to the linearity of (4.5), the use of annually (or monthly) averaged instead of daily estimates of W would not impact the k_{CO2} results, and would be suitable for comparison with Erickson's (1993) computations.

Figure 4.8 shows global distribution of annually averaged transfer velocity for CO_2 in cm h^{-1}. On a global scale, k_{CO2} ranges from 5 to 150 cm h^{-1}, with most probable values around 50.0 cm h^{-1}. High values, up to about 400 cm h^{-1}, are sporadically encountered in so a few places (0.3% of all estimates) that they do not change the overall result at all. For the range up to 150 cm h^{-1}, high k_{CO2} values (above 60 cm h^{-1}) are observed in the trade winds zones in all oceans (5° N to 30° N and 5° S to 30° S). The values in the Southern Ocean (40° S to 60° S) are in the 50-70 cm h^{-1} range. The lowest k_{CO2} values, from 5 to 40 cm h^{-1}, are along the equator, in upwelling zones, and east of North America and Asia coasts. The global average value of CO_2

transfer velocity is k_{CO2} = 56.8 cm h^{-1}. These results differ in k_{CO2} absolute values and resemble in k_{CO2} spatial distribution those reported by Erickson (1993).

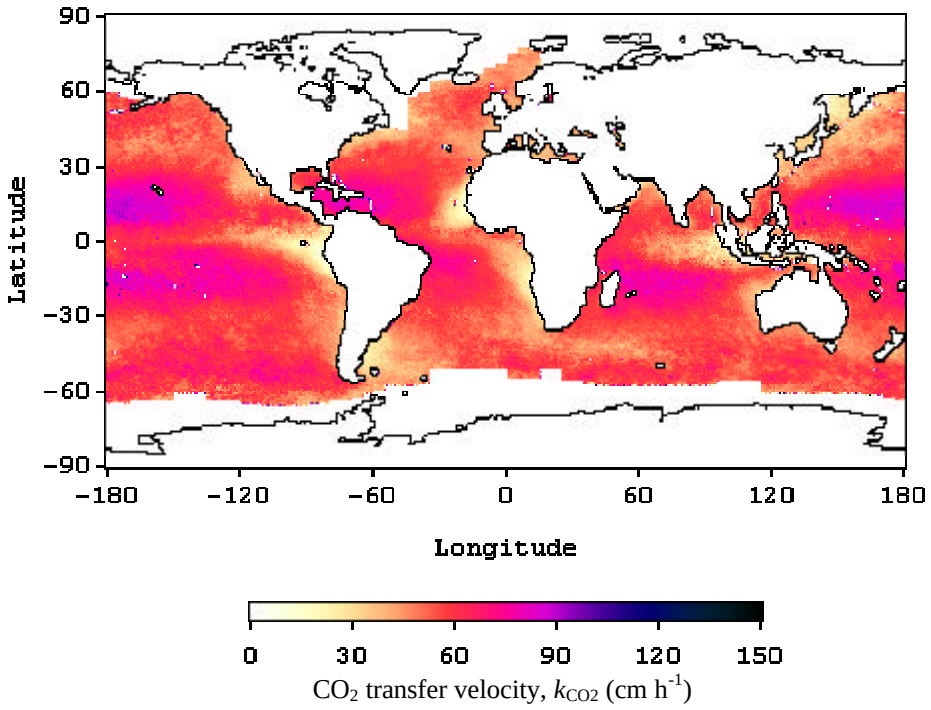


Figure 4.8 Global map (0.5 °×0.5 °) for 1998 of annually averaged CO_2 transfer velocity estimated from annual whitecap coverage obtained with the new method.

Erickson (1993) has listed a global mean k_{CO2} value of 20.9 cm h^{-1}. On one hand, such a value compares well with global mean k_{CO2} data measured via bomb and natural ^{14}C methods (Broecker et al., 1986), about 21.5 cm h^{-1}. The global mean estimate of k_{CO2} in this study is a factor of 2.7 higher than Erickson's (1993). It is lower, however, than the controversial data obtained with the eddy correlation method, usually an order of magnitude higher than those from carbon method measurements

(Broecker et al., 1986; Frew, 1997). On the other hand, Erickson's (1993) estimates differ from those obtained with alternative parameterizations. His data are consistently higher (by 20 to 80%) than those obtained with Liss and Merlivat (1986) scheme, which produces k_{CO2} values up to 20 cm h^{-1}. Erickson data are usually, though not always, lower (by 20 to 40%) than those published by Tans et al. (1990), whose values are up to 80 cm h^{-1} on a seasonal scale. Thus, the present study estimates are consistent with values resulting from some of the parameterizations proposed. In view of the uncertainties in the measuring techniques (Phillips, 1997) and the absence of adequate parameterization of k_{CO2} (Frew, 1997), the values estimated in this study should be given attention since they intrinsically account for the effects of all environmental variables acting together and are, therefore, closer to the real ones.

As expected, the spatial distribution of k_{CO2} estimated in this study mimics that of the whitecap coverage (though Figure 4.2 shows a monthly averaged W map, it can still serve as a guide for comparison). Such a spatial distribution, while different, is not far from the gross features of the k_{CO2} maps shown by Erickson (1993). If Erickson (1993) had used the relation $W(U_{10})$ alone, the spatial distribution of his k_{CO2} maps and those in Figure 4.8 would have been much more different, as different as are the spatial features in Figures 4.2 and 4.3a. Because Erickson (1993) used $W(U_{10}, \Delta T)$ instead, the spatial features in his k_{CO2} maps are closer to reality and hence resemble those evaluated with the new-method estimates of W.

The major differences between the k_{CO2} spatial distributions in Erickson's and the present study maps are in the equatorial regions. In Erickson's maps, low k_{CO2} values are spread wide along both sides of the equator—regions generally thought of as strong sources of CO_2 (Volk and Liu, 1998). The most probable reason for this

deficiency of Erickson's model is the difficulty in accurately estimating C_D under low wind conditions usual for the equatorial area. Meanwhile, the present study shows high k_{CO2} values in these regions, fostering the well-accepted contention that sea surface temperature affects the gas transfer. High k_{CO2} values computed in this study coupled with high negative concentration gradients, ΔC (Asher, 1997), reveal the Southern Ocean as a major sink of CO_2, confirming another well-accepted opinion in the gas-exchange community.

4.3 Parameterization of whitecap coverage

The new method gives estimates of whitecap coverage, W, containing the effects of environmental parameters such as wind speed, U_{10}, atmospheric stability, ΔT, sea surface temperature, T_s, salinity, S, wind fetch, f, wind duration, d, and surfactant concentration, C. To model various processes coupled with W, it is necessary to parameterize W in terms of these variables. Thus, the next goal of this study is extending the existing wind speed relation (2.3) from a $W(U_{10})$ form to a form including the effects of additional factors, $W(U_{10}, \Delta T, T_s, S, f, d, C)$. This section documents the search for approaches and implementation procedures achieving this goal, and reports initial results.

This study does not pursue parameterization of the entire suit of relations in $W(U_{10}, \Delta T, T_s, S, f, d, C)$ since data for f, d, and C are not readily available. Monahan and O'Muircheartaigh (1986) have proposed an expression for the $W(U_{10}, \Delta T)$ relation and the effect of ΔT is not further considered here. The whitecap coverage database, built with the new method, contains information suitable for extracting the empirical relations involved in $W(U_{10}, T_s, S)$.

4.3.1 Regression analysis

Regression analysis is a traditional and reliable tool for analyzing observational data and deriving empirical relationships between observed variables. Regression procedures find an equation and a line, straight or curved, that best fits the behavior of a dependent variable Y when an independent variable X is varied. Coefficients entering the equations and connecting the independent and dependent variables, called regression coefficients, could be estimated from a set of n observations (values) of these variables, (X_k, Y_k), where $k = 1, 2, \ldots, n$. Regression analysis defines the linearity and non-linearity of a regression model by considering its regression coefficients (Ratkowsky, 1990). A regression model, in which all coefficients appear linearly, is a linear regression model. A cubic polynomial, $Y_k = y_0 + aX_k + bX_k^2 + cX_k^3$, is an example of a linear regression model because the regression coefficients y_0, a, b, and c appear linearly. A regression model, in which at least one of its coefficients appears nonlinearly, is a nonlinear regression model; a power law, $Y_k = aX_k^b$, is a nonlinear regression model.

There are various methods for estimating the regression coefficients. The most popular one is the method of least squares, which seeks the values of the regression coefficients that minimize the sum of the squared differences between the observed and predicted values of the dependent variable, Y. The least-squared estimates are optimal when the analyzed data fulfill two assumptions: normality and constant variance. The normality assumption requires the observed data (X_k, Y_k) to be normally distributed around the regression line. The constant variance assumption requires the variances of the dependent variable values, σ_{Yk}, to be constant regardless of the values of the independent variable, X_k. The least-squared regression procedure must test if the set of observations (X_k, Y_k) fulfills these two assumptions. Failure of the

normality test indicates the presence of outlying influential points among the set of observations or an inappropriate regression model. If the constant-variance test fails, a different regression model should be considered or the values of the independent variable, X_k, must be transformed to obtain stable variances, σ_{Yk}.

The choice of a regression model is not an easy task. If previous studies or underlying theory point toward an appropriate mathematical expression, this expression should be used as a regression model. If there is not a suggested mathematical model, and there is a single independent variable (X), various equations should be plotted against the data and the one that matches the data trend best should be chosen. If several independent variables (X_1, X_2 and so on) are candidates for explaining the behavior of Y, a multiple linear regression model should be tried.

A useful guideline when choosing a regression model is the famous dictum, known as Occam's razor, that "Plurality should not be assumed without necessity" (Ratkowsky, 1990). This principle expresses the belief, not only in this particular case of regression analysis, but also in any other scientific endeavor, that simplicity is to be preferred to complexity. Often attempts to fit a complicated regression model where a simple one suffices fail. The estimates of an over-parameterized model containing unnecessarily numerous terms and coefficients are usually severely biased compared to the estimates of a simpler model with good statistical properties.

Good statistical properties show that a chosen regression model fits the set of observations well. Numerous statistical properties helping the assessment of model performance have been established in the specialized literature (Ratkowsky, 1990; Birkes and Dodge, 1993). Among these, the following three statistical properties are the most informative. The squared correlation coefficient, R^2, is a reliable measure for

the performance of a linear regression model: the closer R^2 value to 1, the better the model equation describes the relation between the independent and dependent variables. The standard error of the estimate, $\sigma_{Y|X}$, is a measure of the variability of the observations (X_k, Y_k) about the regression line. Generally, the observed values should fall within about two standard errors around the line of the regression model. A magnitude of standardized residuals is a gauge for the performance of a nonlinear regression model; R^2 is not useful at all in judging a nonlinear model. The standardized residual is defined as the difference between the predicted and observed values of the dependent variable Y divided by the standard error of the estimate, $\sigma_{Y|X}$. The model is performing well when the standardized residuals are in the interval from –2 to +2. Residual values outside this interval indicate outlying points. In addition, the residuals should not exhibit a consistent trend; they should be randomly spread around zero.

4.3.2 Implementation

Regression analysis is used in this study for extracting the empirical relations between whitecap coverage, W (a dependent variable, i.e., $Y = W$), and wind speed, U_{10}, sea surface temperature, T_s, and salinity, S (independent variables, i.e., $X = U_{10}$ or T_s or S). The least square method is employed for estimating the regression coefficients. The regression analysis is applied to the W estimates with relative errors, $r.e._W = \sigma_W / W$, below 100% and the environmental variables U_{10}, T_s, and S corresponding to them. These are the "good" data in Part 2 of the whitecap coverage database (Table 4.1).

Parameterization approach

Two approaches for parameterizing $W(U_{10}, T_s, S)$ are possible.

The first approach is multiple regression analysis. Applying multiple regression analysis to W estimates, which contain the effects of all environmental variables, would derive a mathematical formula expressing the combined effect of U_{10}, T_s, and S. However, one should be cautious choosing this approach as a first step in parameterizing whitecap coverage. As the review in §2.5.1 shows, the effects of the environmental variables are only qualitatively discussed in the literature and quantitative expressions are not offered. These variables influence the whitecapping differently and with opposing results. Most certainly, a tractable mathematical function, e.g., a simple combination of linear terms each representing one variable, would not suffice to describe the overall behavior of $W(U_{10}, T_s, S)$. In addition, there are not theoretical suggestions helping a construction of a more complex mathematical expression. The Occam razor principle further supports this caution.

Thus, the second approach, a single-variable regression analysis, seems appropriate for a first attempt to parameterize $W(U_{10}, T_s, S)$. Performing a single-variable regression analysis, it is possible to extract the whitecap coverage dependence on each environmental variable separately, i.e., $W(U_{10})$, $W(T_s)$, and $W(S)$. Then, these can be combined to derive an expression for their effect in concert:

$$W(U_{10}, T_s, S) = W(U_{10}) \cdot W(T_s) \cdot W(S) \tag{4.8}$$

Derivation of one relation at a time, $W(U_{10})$, $W(T_s)$, or $W(S)$, with this approach would also give the possibility to investigate the various effects on the whitecapping separately and evaluate their relative contribution to whitecap coverage.

Previous considerations (Monahan and O'Muircheartaigh, 1986) have shown, and Figure 3.20 confirms, that most probably $W(T_s)$ and $W(S)$ are not strong dependencies by themselves, but they may affect significantly the coefficients in the

power law of the wind dependence, $W(U_{10}) = aU_{10}^{b}$, especially the exponent b. Accepting this suggestion, the parameterization of $W(U_{10}, T_s, S)$ may proceed as a revision of the $W(U_{10})$ relation aiming the inclusion of the T_s and S effects in its coefficients. For instance, using as an example the widely recognized power law for $W(U_{10})$, the parameterization can be represented as:

$$W(U_{10}, T_s, S) = a(T_s, S) U_{10}^{b(T_s, S)} \tag{4.9a}$$

The same approach can be used with any other mathematical function that may express $W(U_{10})$ plausibly, for instance, an exponential law:

$$W(U_{10}, T_s, S) = a(T_s, S)\ \exp\left[b(T_s, S) \cdot U_{10}\right] \tag{4.9b}$$

The dependence of coefficients a and b on T_s and S can be expressed as:

$$\begin{aligned} a(T_s, S) &= a(T_s) \cdot a(S) \\ b(T_s, S) &= b(T_s) \cdot b(S) \end{aligned} \tag{4.10}$$

where each of the dependencies $a(T_s)$, $a(S)$, $b(T_s)$, and $b(S)$ can also be obtained with a single-variable regression analysis. Note that the regression coefficients could be more than a and b if $W(U_{10})$ is represented by a mathematical law other than 4.9.

Single-variable regression analysis and the use of (4.9)–(4.10) comprise the approach chosen in this study to parameterize $W(U_{10}, T_s, S)$.

Data binning

To apply single-variable regression to a dependent variable affected by several independent variables, any of the analyzed dependent-independent pairs should be processed at all remaining independent variables held constant. For instance, to revise the $W(U_{10})$ relation, single-variable regression analysis should be applied to a set of $(W_k, U_{10,k})$ values at constant T_s and S. Thus, to prepare the W estimates and their

corresponding U_{10}, T_s, and S values for such analysis, the data must be allocated into bins within which S and T_s remain approximately constant.

Data are not binned by the values of f, d, and C. Their effects on the data are excluded by narrowing the range of the considered S-values. The full range of observed S-values (used unabridged to derive W estimates with the new method in Chapter 3) is from 3.42 to 42.22 psu with most S-values clustered in the range of 20-40 psu. Higher S-values, around 35 psu, are usual for open ocean conditions. Salinity values below 20 psu are characteristic of coastal zones, where, compared to open-ocean conditions, the wind fetch, f, is shorter, the wind duration, d, is more variable due to the land proximity, and surface active material, C, is more abundant due to a higher amount of nutrients delivered by river run off. Bottom topography additionally affects wave breaking and W in coastal zones. Thus, narrowing the range of S-values excludes the coastal zone influences on W.

The range of S-values used for regression analysis is from 33 to 37 psu. Only about 5% of the pixels with useful W estimates are lost as a result of this restriction. The salinity range from 33 to 37 psu is divided in 20 salinity bins each with 0.2 psu width. This number of salinity bins results from applying the rule of thumb (Panofsky and Brier, 1958): the number of necessary bins is 5 log m, m being the number of observations available for distribution in bins; m is around 14,000.

Using the S-values falling within a given S-bin as a reference, all W estimates for a particular S-bin, together with their corresponding σ_W, $r.e._W$, U_{10}, T_s, and S values, are extracted from the daily files with "good" data (Table 4.1, Part 2) and stored in a new text file. This results, in total, in about 7,300 text files: 365 daily text files containing W, σ_W, $r.e._W$, U_{10}, T_s, and S for each of the 20 salinity bins. These files

are organized in 12 (monthly) directories each containing 20 directories—one for each salinity bin.

Next, the data within each *S*-bin are divided in 12 temperature bins each with a width of 3 °C covering the entire range of sea surface temperature from –1.8 to 33 °C. Table 4.2 lists the values of all salinity and sea surface temperature bins. Similarly to the salinity binning, the T_s-values falling within a given T_s-bin are used as a reference to extract the necessary *W* estimates and their corresponding additional variables for a particular T_s-bin.

Sets of observations

How many observations, *n*, should a set prepared for regression analysis, i.e., binned by *S* and T_s, have? An attempt was first made to proceed with regression analysis using observations (W_k, U_{10k}), $k = 1, 2, \ldots, n$, for one day. The number of observations, *n*, turned out to be insufficient to reveal the $W(U_{10})$ relation. Too often only 1 or 2 points were available for a particular combination of *S*- and T_s-bins. The next attempt was to run regressions using observations (W_k, U_{10k}) for all days in a month. In this case, very often the points were too many, above 7,000, and any trend, if present, was lost in the variability of data.

A decision was then made to construct sets of observations including (W_k, U_{10k}) values for 8 consecutive days. In this way, the number of observations in a set formed for given T_s and *S*-bins usually ranges from 10 to 2,000. There are exceptions: for the *S*-bins around 35 psu, some T_s-bins may still have more than 3,000 points, and some T_s-bins in the lower (close to 33 psu) and higher (close to 37 psu) salinity ends may not include any observations. Such cases, however, are more rarely

encountered compared to the formation of sets with 1-day or 30-day observations. Grouping the observations for 8 days allows 4 observational sets for each T_s-bin within

Table 4.2 Values for salinity, sea surface temperature, and wind speed bins.

Salinity bins			Sea surface temperature bins			Wind speed bins		
Bin name	Bin center psu	Bin range psu	Bin name	Bin center °C	Bin range °C	Bin name	Bin center m s^{-1}	Bin range m s^{-1}
S1	33.1	33.0-33.2	T1	-0.9	-1.8-0.0	U1	4.0	3.0-5.0
S2	33.3	33.2-33.4	T2	1.5	0.0-3.0	U2	6.0	5.0-7.0
S3	33.5	33.4-33.6	T3	4.5	3.0-6.0	U3	8.0	7.0-9.0
S4	33.7	33.6-33.8	T4	7.5	6.0-9.0	U4	10.0	9.0-11.0
S5	33.9	33.8-34.0	T5	10.5	9.0-12.0	U5	12.0	11.0-13.0
S6	34.1	34.0-34.2	T6	13.5	12.0-15.0	U6	14.0	13.0-15.0
S7	34.3	34.2-34.4	T7	16.5	15.0-18.0	U7	16.0	15.0-17.0
S8	34.5	34.4-34.6	T8	19.5	18.0-21.0	U8	18.0	17.0-19.0
S9	34.7	34.6-34.8	T9	22.5	21.0-24.0	U9	20.0	19.0-21.0
S10	34.9	34.8-35.0	T10	25.5	24.0-27.0	U10	22.0	21.0-23.0
S11	35.1	35.0-35.2	T11	28.5	27.0-30.0	U11	24.0	23.0-25.0
S12	35.3	35.2-35.4	T12	31.5	30.0-33.0	U12	26.0	25.0-27.0
S13	35.5	35.4-35.6				U13	28.0	27.0-29.0
S14	35.7	35.6-35.8				U14	30.0	29.0-31.0
S15	35.9	35.8-36.0				U15	32.0	31.0-33.0
S16	36.1	36.0-36.2				U16	34.0	33.0-35.0
S17	36.3	36.2-36.4						
S18	36.5	36.4-36.6						
S19	36.7	36.6-36.8						
S20	36.9	36.8-37.0						

a given *S*-bin for each month. Each observational set is saved in a new text file. There are 960 (= 4(sets)×12(T_s-bins)×20(*S*-bins)) such files for a month saved in the directories allotted for the *S*-bins (Table 4.1, Part 2). From the perspective of the entire 1998, there are 48 sets (= 4(sets)×12(months)) for each combination of T_s-*S* bins.

Figure 4.9 plots the $W(U_{10})$ observations comprising set #1 for March 1998 (Julian days 60 to 67) for all 12 T_s-bins in salinity bin S5 (Table 4.2). In the figure, W estimates are given with their standard deviations only for temperature bin T2; for the other plots, the error bars are omitted to avoid cluttering the graphs.

Note in Figure 4.9 that though the number of observations in a set (observations for 8 days) reveals a trend in the data well, some variability in W estimates remains. When least-square regression is applied to the data, this variability often leads to the failure of either normality or the constant variance test, precluding an estimation of regression coefficients for some T_s-bin. This, in turn, diminishes the reliability of obtaining $a(T_s)$ and $b(T_s)$ relations.

Smoothing the data within each T_s-bin circumvents this problem. The smoothing is achieved by dividing the wind speed range from 3 to 35 m s^{-1} into 16 bins, each 2 m s^{-1} wide (Table 4.2), and averaging the W values coupled with the U_{10} values in each U_{10}-bin. In this way, the variability is effectively removed and a smooth curve represents the $W(U_{10})$ trend for each T_s-bin. These smoothed curves are shown in Figure 4.9 with gray lines.

Regression models are applied to these smoothed curves. Initially, a group of regression coefficients (2 or more coefficients, depending on the mathematical function chosen for a regression) was determined for each of the 4 observational sets available for a month. Then these 4 groups of coefficients were averaged to obtain one group of coefficients for each T_s-bin. Though this is the most rigorous way to estimate the regression coefficients, this procedure shows two deficiencies: i) high variability of the coefficient values from one set to another; and ii) inability to construct a complete temperature dependence of the coefficients, e.g., $a(T_s)$ and $b(T_s)$. The reason

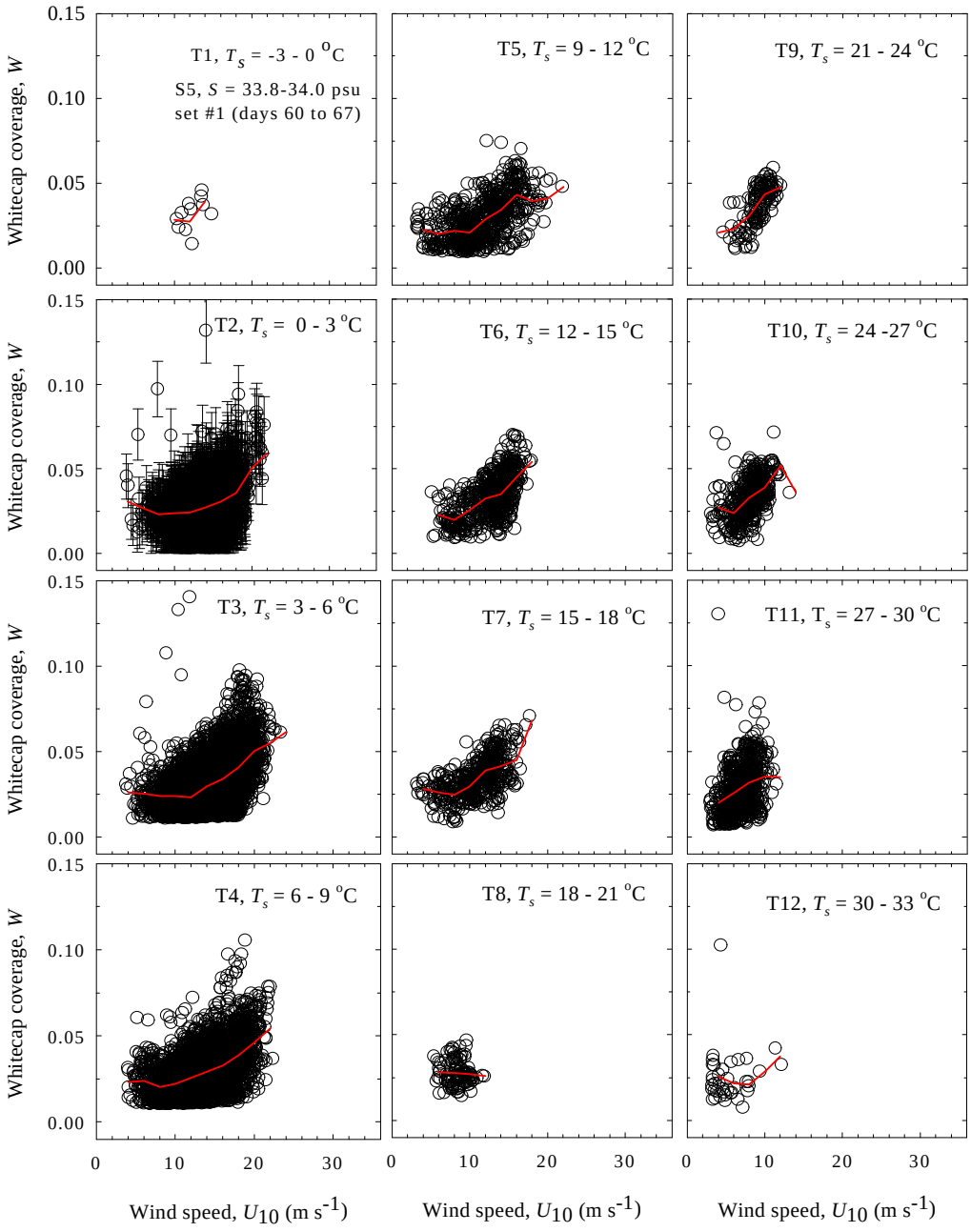


Figure 4.9 The relation $W(U_{10})$ for March 1998, observational set #1 (Julian days 60 to 67) in 12 T_s-bins at salinity bin $S5$ (S = 33.8 - 34.00 psu).

for these deficiencies is the fact that $W(U_{10})$ relation is not fully presented in all T_s-bins. Figure 4.9 illustrates the problem: relation $W(U_{10})$ is well expressed covering a relatively wide range of wind speeds (up to 24 m s^{-1}) only in some temperature bins, e.g., T2 to T7. For other temperature bins the range of available U_{10} values does not exceed 12-14 m s^{-1}, e.g., T1 and T8 to T12. In these temperature bins, only a short section of the real $W(U_{10})$ relation is available and cannot be used reliably to extract an empirical expression.

How can we ensure observational sets long enough to extract a trustworthy $W(U_{10})$ relation in all T_s-bins? Two considerations help to construct such observational sets.

The first consideration is related to the salinity value. A survey of data for various S-bins shows that as salinity changes, the range of U_{10} values available for regressions may improve in the T_s-bins understated in Figure 4.9. The improvement in representing the $W(U_{10})$ relation in the understated T_s-bins, however, is often accompanied with a deterioration of the $W(U_{10})$ relation in the T_s-bins well represented in Figure 4.9. Moreover, looking for long enough observational sets in various salinity bins would compromise the requirement of holding S constant, which is necessary for applying a single-variable regression analysis. A final result of this survey is the choice of the salinity bin to be held constant: S5 (S = 33.8 to 34.0 psu). This salinity bin not only encompasses the most likely salinity values encountered in open ocean, but also offers the largest number of T_s-bins with relatively long (W_k, U_{10k}) sets.

The second consideration in constructing long observational sets for $W(U_{10})$ in all T_s-bins at a chosen S-bin is related to the use of the observational sets.

Instead of using them separately, the smoothed curves for all 4 observational sets available for a month are averaged. This yields one long averaged observational set. Indeed, some of the points in the averaged observational set may happen to be the average of 2 or 3 points, instead of 4. It may even happen that only one point is available for some wind speed values. But this approach certainly provides the longest possible observational sets (W_k, U_{10k}) in *all* T_s-bins.

The procedure of constructing one long averaged observational set for each T_s-bin would be valid only if the 4 observational sets available for a month are stationary and do not exhibit a systematic trend in the course of the month. This stationarity is verified by the similarity of the smoothed curves for all 4 sets (Figure 4.10) and the lack of variability in the averaged W estimates for each set (Figure 4.11). Figure 4.12 piles the averaged observational sets for all T_s-bins at S5.

In a similar manner, averaged observational sets are constructed for 12 salinity bins (S1, S3, S5, S7, S9, S11, S12, S13, S15, S17, S19, S20) at T8 (Figure 4.13). The choice of the sea surface temperature bin to be held constant (T8 with T_s from 18.0 to 21.0 °C) offers the largest number of S-bins with relatively long (W_k, U_{10k}) sets.

Regression models

Three regression models are chosen for the regression analysis of $W(U_{10})$.

The choice was done among numerous possible functions, some coming of interest from previous studies, others because they apparently fit the trend of the data well.

Various regression equations, usually modified to meet the specifics of this study, and diagnostic tests available from *Sigma Plot* software were used in this

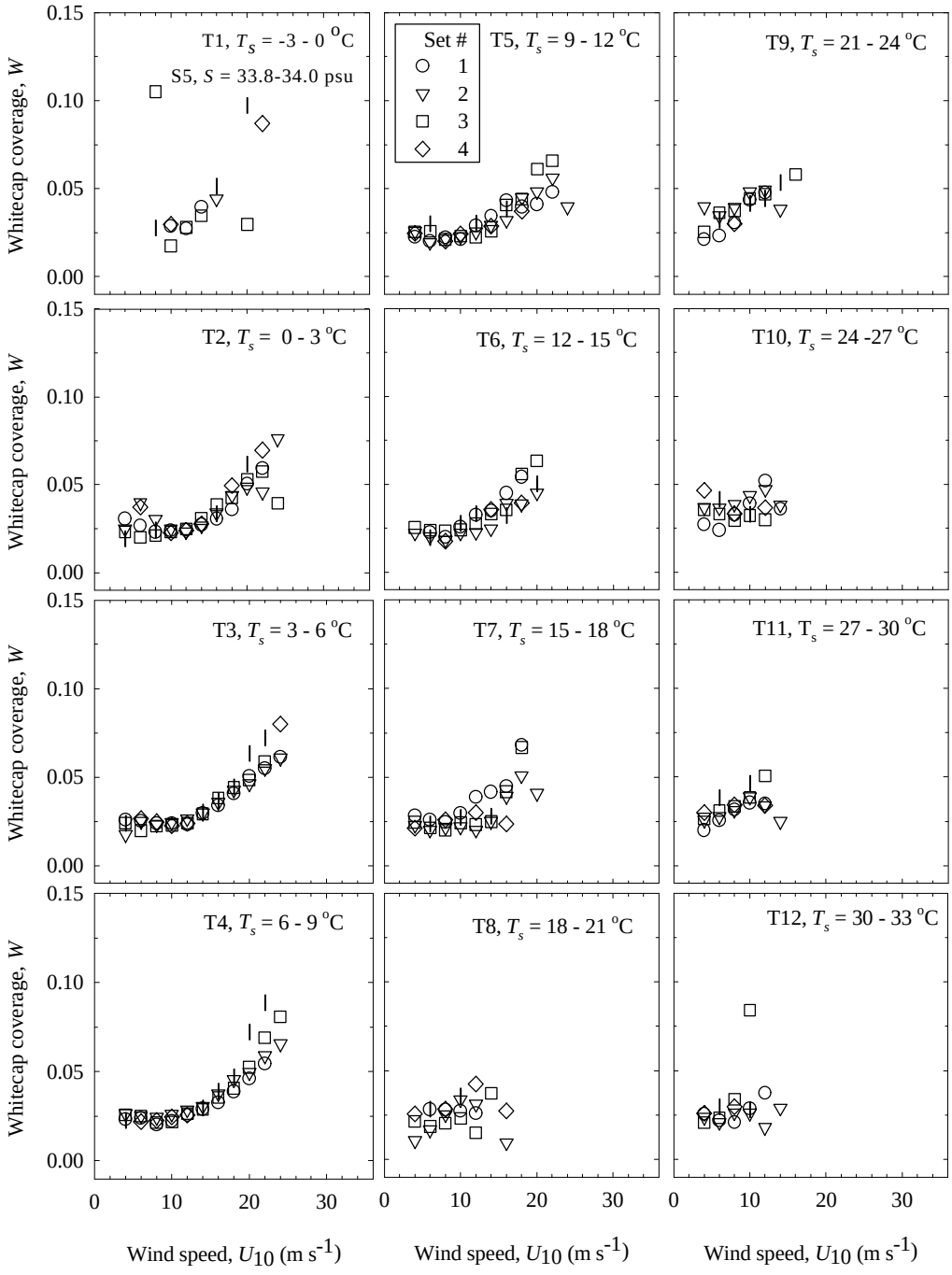


Figure 4.10 All 4 observational sets for the $W(U_{10})$ relation for March 1998.

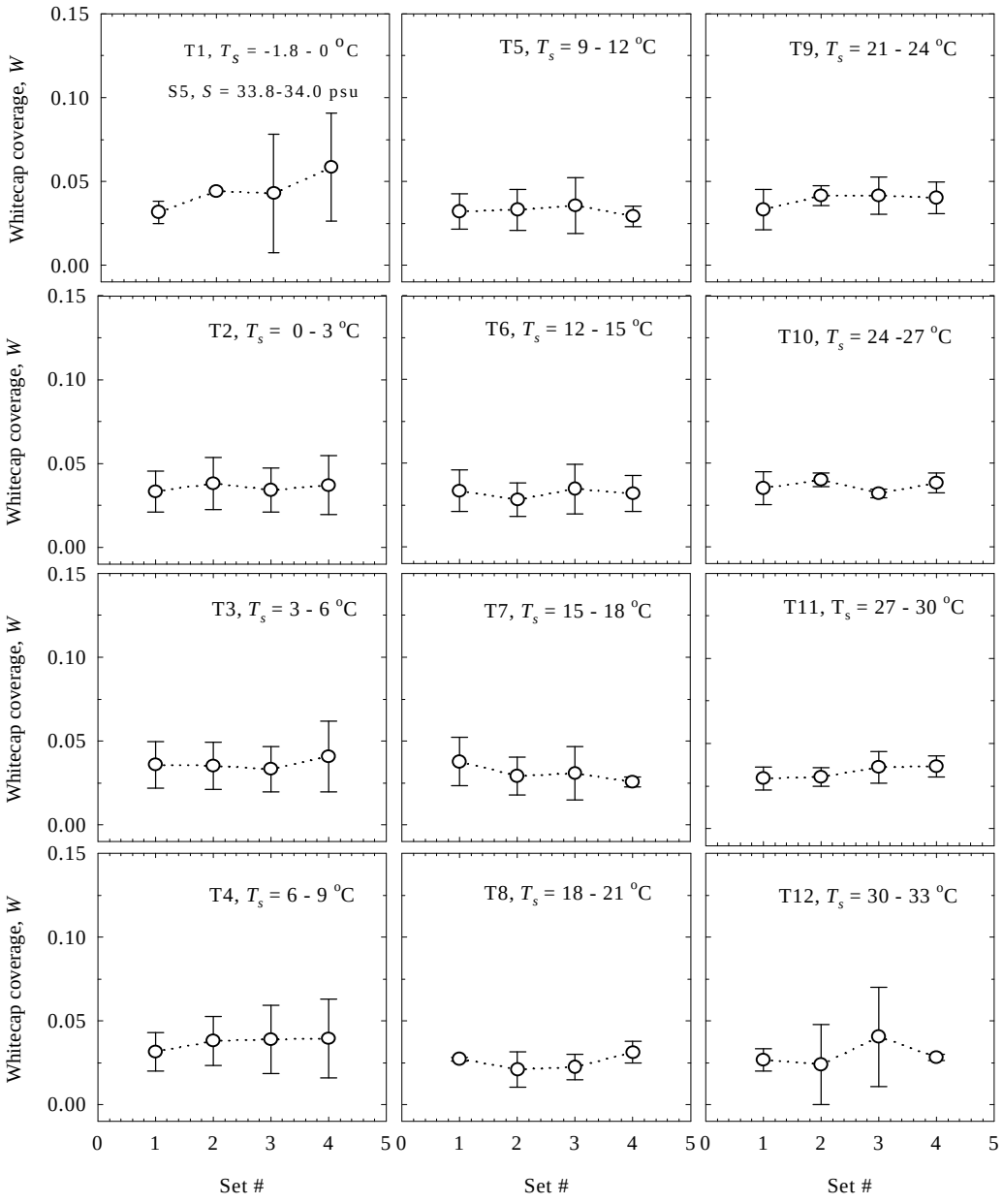


Figure 4.11 Average whitecap coverage for each observational set in all T_s-bins at salinity bin S5.

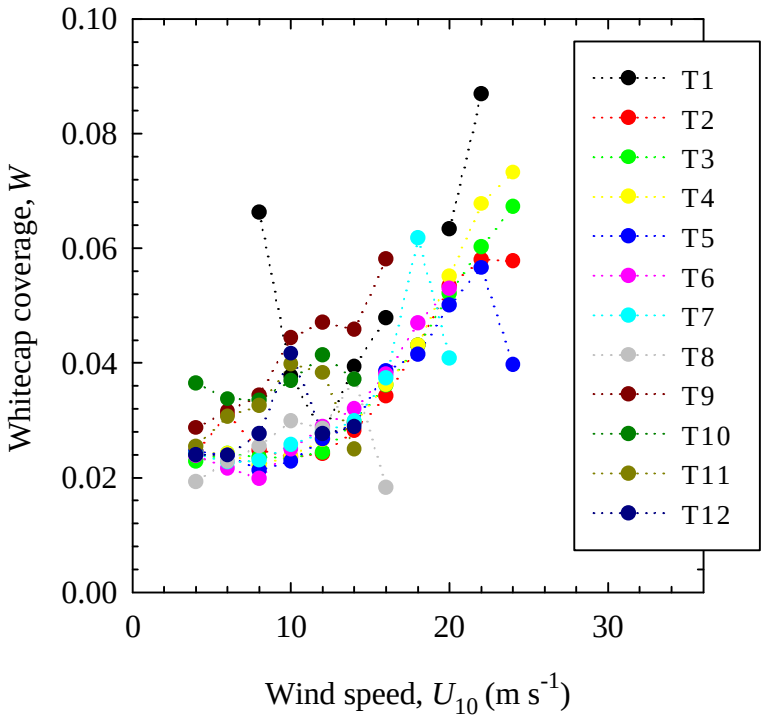


Figure 4.12 Averaged observational sets of the $W(U_{10})$ relation for March 1998 in all T_s-bins at salinity S5.

investigation. The performance of the applied regression models was assessed by examining: i) the passing or failing of normality and constant-variance tests; ii) the squared correlation coefficient, R^2; and iii) the magnitude and distribution pattern of standardized residuals. The three chosen regression models are the ones that have shown the best statistical properties: low standard error of the estimate, $\sigma_{Y|X}$, high squared correlation coefficient, R^2, and magnitude of standardized residuals in the ±2 range.

The chosen regression models, with $x \equiv U_{10}$ and $y \equiv W$, are:

1) Cubic polynomial: $y = ax + bx^2 + cx^3$—linear model;

2) Power law: $y = ax^b$—nonlinear model;

3) Exponential law: $y = ae^{bx}$—nonlinear model;

These regression models are interesting not only for their good statistical properties, but also because they are suggested by previous theory. Model 1 (cubic polynomial) combines linear, quadratic, and cubic dependencies on U_{10}, all suggested

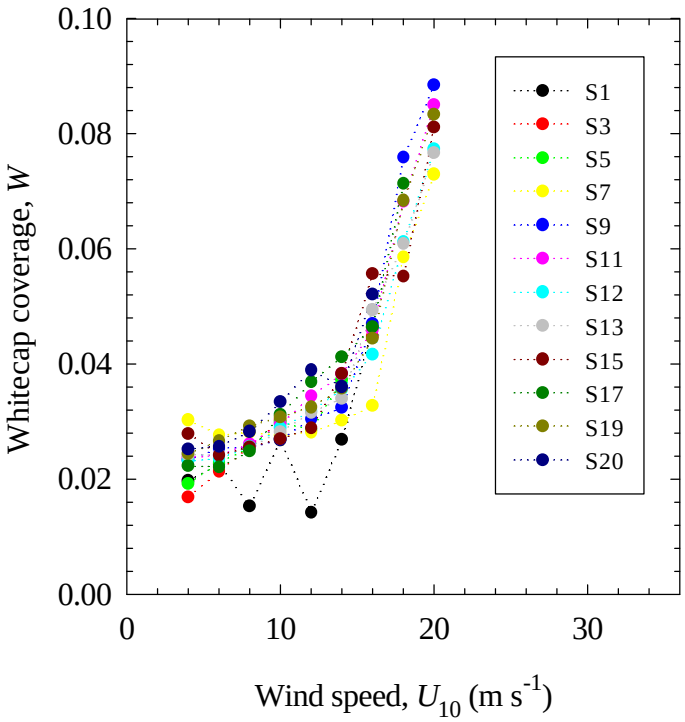


Figure 4.13 Averaged observational sets of the $W(U_{10})$ relation for March 1998 in all S-bins at sea surface temperature T8.

in previous studies (Blanchard, 1963; Wu, 1975; Bortkovskii, 1983). Besides, it is a linear model, the simplest of all chosen models, and, as Occam Razor suggests, should be tried first. Model 2 (power law) represents the currently used expression (2.3). Applied to the satellite-derived estimates of W, however, this model does not perform

better than any other. Model 3 (exponential law) is often suggested and used to obtain aerosol concentration in the atmosphere from wind speed (Lovett, 1978, Erickson et al., 1986). Applied to the satellite W estimates, Model 3 often performs better than Model 2.

These three models are applied to all averaged observational sets for March 1998 shown in Figures 4.12 and 4.13. As a result, the relation $W(U_{10})$ is represented by a cubic polynomial, a power, and an exponential laws. Regression coefficients are estimated for each of these models, once for all T_s-bins at S5 and then for all S-bins at T8. The coefficient values for each T_s-bin provide the sea surface temperature dependence of the regression coefficients, e.g., $a(T_s)$. Analogously, the coefficient values for each S-bin provide the salinity dependence of the regression coefficients, e.g., $a(S)$.

The sets of values for $a(T_s)$, $a(S)$, and all other regression coefficients derived for $W(U_{10})$ relation are now subjects of a second regression. In this second regression the independent variables are the temperature and salinity, i.e., $X \equiv T_s$, or S, and the dependent variables are the regression coefficients from the first regression to $W(U_{10})$, i.e., $Y \equiv a$, b, or c.

The steps of searching an appropriate regression model among various mathematical functions, determining a regression model with the best statistical properties, and applying this model to $a(T_s)$ and $a(S)$ dependencies are repeated. The newly estimated regression coefficients are the values sought. Once specific mathematical expressions for $a(T_s)$ and $a(S)$ are determined, they are used to obtain $a(T_s, S)$ with (4.10). The same is done for the regression coefficient b and c.

The concrete expressions for the regression coefficients $a(T_s, S)$, $b(T_s, S)$, and $c(T_s, S)$ enter the corresponding regression model parameterizing $W(U_{10})$: cubic polynomial, power (4.9a), or exponential (4.9b) law. The result is a $W(U_{10})$ relation revised to include parameterization of the sea surface temperature and salinity effects on whitecap coverage, i.e., a parameterization of $W(U_{10}, T_s, S)$.

4.3.3 Results

Sea surface temperature dependence

Following the implementation steps outlined in the previous section, the sea surface temperature dependence is parameterized. Figure 4.12 reveals that the sea surface temperature dependence is not a clear-cut case. As expected, the effect of temperature is not strong by itself (the modest spread of the $W(U_{10})$ curves at different T_s confirms that), yet it brings noticeable changes. These changes do not follow a monotonic function, e.g., a clear increase or a decrease with T_s. Rather, there are subtle changes in the slopes or slight "jumps" in the magnitudes of the $W(U_{10})$ curves at higher T_s. Compare, for instance, the curves for bins T8 (gray circles), T9 (brown), T10 (dark green) and T11 (dark yellow) to the curves for the rest of the T_s-bins. A clear decrease (T5, blue) or flattening (T2, red) of the $W(U_{10})$ curves at high winds, not predicted by (2.3), is evident. The distinct $W(U_{10})$ behavior at the lowest sea surface temperatures (T1, black) is most probably a consequence of the scarcity of data at these temperatures.

The cubic polynomial, power, and exponential laws are the three regression models applied to each of the $W(U_{10})$ curves in Figure 4.12. A first batch of regression coefficients is produced estimating regression coefficients for each regression model at

each T_s-bin. The values of regression coefficients for each model are plotted as a function of sea surface temperature. Figure 4.14 illustrates this for the regression coefficients *a*, *b*, and *c* of the cubic polynomial model.

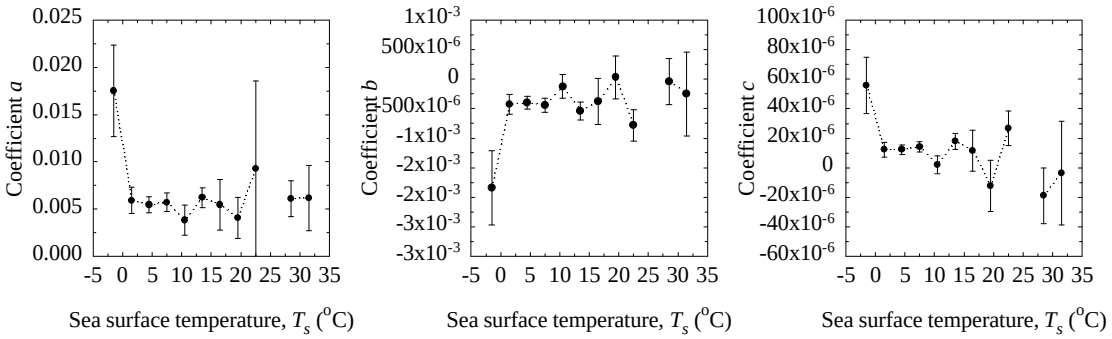


Figure 4.14 Regression coefficients *a*, *b*, and *c* for cubic polynomial applied to $W(U_{10})$ as a function of sea surface temperature.

A second regression is now performed on each of the regression coefficients in this first batch. For each regression coefficient in the batch, a regression model with the best statistical characteristics is determined and its regression coefficients are estimated. Figure 4.15 demonstrates this step for the temperature dependence of the regression coefficients, $a(T_s)$ and $b(T_s)$, of the power-law model applied to $W(U_{10})$. Table 4.3 summarizes the estimated values of the regression coefficients for the power and exponential regression models applied to the $W(U_{10})$ curves. These are the values determining the sea surface temperature dependence.

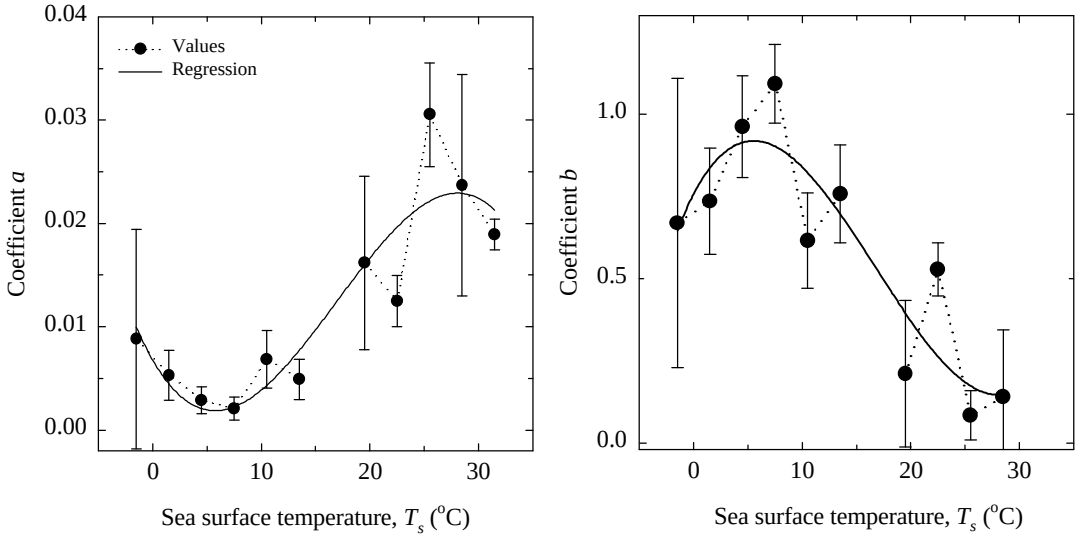


Figure 4.15 Values and regression curves for the sea surface temperature dependence of the regression coefficients *a* and *b* resulting from the power-law model applied to $W(U_{10})$.

Salinity dependence

Figure 4.13 shows that over the salinity range from 33 to 37 psu, the salinity invokes little change in the behavior of $W(U_{10})$: the slopes of the curves do not change and the magnitudes are similar. Once again, the widest changes are observed for the salinity bins with least available data. In spite of the anticipation of weak, if any, salinity dependence, the procedure of parameterizing the salinity effect was carried out.

The power-law regression model only was applied to all the $W(U_{10})$ curves in Figure 4.13, $W = aU_{10}^{\,b}$. A pair of values for the regression coefficients *a* and *b* is estimated for each salinity bin. These are plotted as a function of salinity in Figure 4.16 (circles). A second regression with a model having acceptable statistical properties is applied to $a(S)$ and $b(S)$ (solid lines in Figure 4.16) and a

Table 4.3 Regression coefficients parameterizing the sea surface temperature dependence in the power-law and exponential-law models applied to $W(U_{10})$.

Model to $W(U_{10})$	Model to Coef.(T_s)	Parameterization	R^2	$s_{\text{Coef}\|Ts}$
Power $W = aU_{10}^{\ b}$	Cubic polynomial $a(T_s) = a_0 + a_1T_s + a_2T_s^2 + a_2T_s^3$	$a_0 = 6.779\times10^{-3}$ $a_1 = -1.83\times10^{-3}$ $a_2 = 1.917\times10^{-4}$ $a_3 = -3.778\times10^{-6}$	0.83	0.0045
	Cubic polynomial $b(T_s) = b_0 + b_1T_s + b_2T_s^2 + b_2T_s^3$	$b_0 = 0.7566$ $b_1 = 6.096\times10^{-2}$ $b_2 = -6.547\times10^{-3}$ $b_3 = 1.276\times10^{-4}$	0.83	0.1726
Exponential $W = a\exp(bU_{10})$	Cubic polynomial $a(T_s) = a_0 + a_1T_s + a_2T_s^2 + a_2T_s^3$	$a_0 = 0.0194$ $a_1 = -3.449\times10^{-3}$ $a_2 = 3.413\times10^{-4}$ $a_3 = -7.633\times10^{-6}$	0.63	0.0062
	Cubic polynomial $b(T_s) = b_0 + b_1T_s + b_2T_s^2 + b_2T_s^3$	$b_0 = 0.0561$ $b_1 = 3.655\times10^{-3}$ $b_2 = -3.478\times10^{-4}$ $b_3 = 6.016\times10^{-6}$	0.65	0.0166

parameterization of $a(S)$ and $b(S)$ is obtained. Including the parameterized $a(S)$ and $b(S)$ in the $W(U_{10})$ parameterization, however, renders erroneous W values. The weak dependence of W on S yields very low values for the regression coefficients, especially $a(S)$. Obviously, the range of salinity considered, 33 to 37 psu, is too narrow and extending the parameterization outside this range does not reproduce the real changes correctly.

Wind-speed dependence—revised

Since the parameterization of $a(S)$ and $b(S)$ fails, the wind-speed dependence is revised to include the parameterization of sea surface temperature effect only, changing (4.9) to:

$$W(U_{10},T_s) = a(T_s)U_{10}^{b(T_s)}$$
$$W(U_{10},T_s) = a(T_s)\exp\left[b(T_s)U_{10}\right] \qquad (4.11)$$

where $a(T_s)$ and $b(T_s)$ are listed in Table 4.3 for the power-law and exponential-law models.

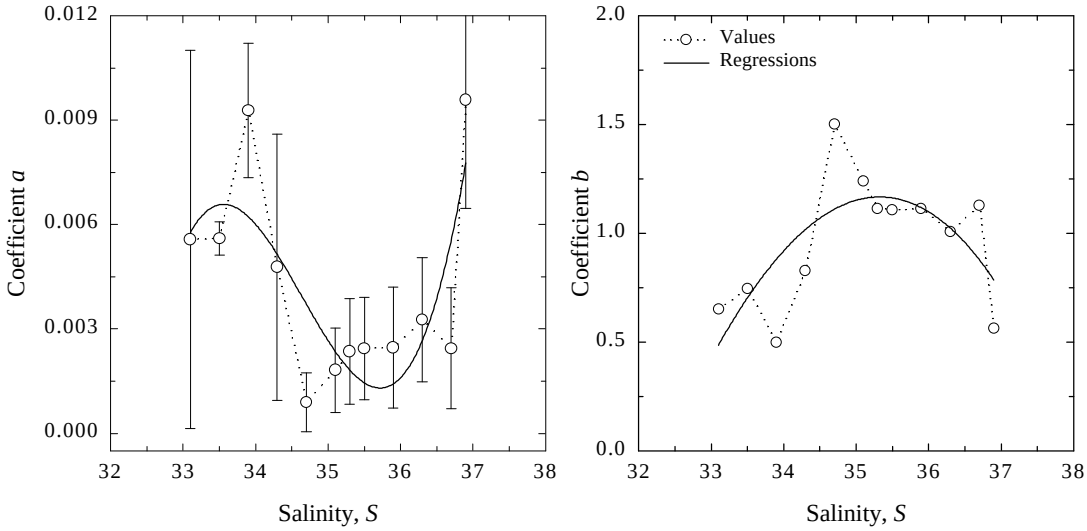


Figure 4.16 Values and regression curves for the salinity dependence of the regression coefficients *a* and *b* resulting from the power-law model applied to $W(U_{10})$.

Employing the $a(T_s)$ and $b(T_s)$ parameterizations in Table 4.3 and (4.11), whitecap coverage, W, is estimated. Monthly (March 1998) values for U_{10} and T_s along a North-South line (31.8° N to 60.45° S) in the Pacific Ocean (133.4° W) are used for the calculations.

Results for the power-law parameterization of $W(U_{10}, T_s)$ are displayed in Figure 4.17. Panel a in the figure compares parameterized W estimates (open circles)

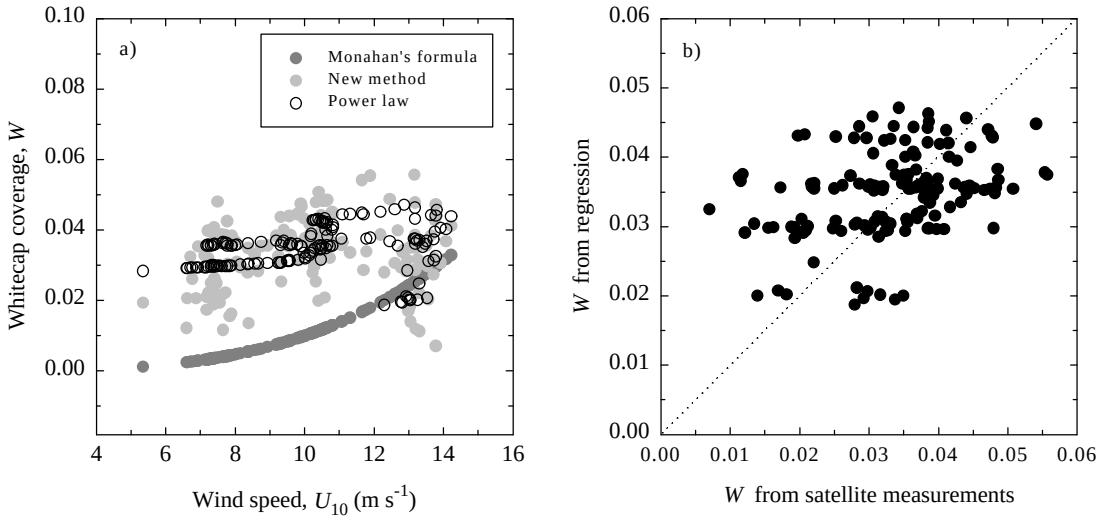


Figure 4.17 Performance of the power-law parameterization: a) comparison between satellite-measured W (light gray circles), $W(U_{10},T_s)$ parameterization (open circles), and $W(U_{10})$ parameterization (dark gray circles); b) comparison between satellite-measured W and $W(U_{10},T_s)$ parameterization.

with the satellite-derived W (light gray circles) and the $W(U_{10})$ estimates with (2.3) along the same North-South line. Generally, the parameterization succeeds in producing W-values comparable in magnitude with those from satellite data, and exhibiting higher variability than (2.3) estimates. The parameterization fails to reproduce the full range of variations in W displayed by the satellite-derived estimates. Panel b in Figure 4.17 clearly shows that the power-law parameterization of $W(U_{10}, T_s)$ largely overestimates W and does not predict well the lower W-values.

Results for the exponential-law parameterization of $W(U_{10}, T_s)$ are displayed in Figure 4.18 in a fashion similar to that in Figure 4.17. Panel a in the figure shows that the exponential parameterization model compares better with the satellite-derived W, especially in the 2% to 4% range of W, than the power-law

parameterization. This is confirmed in Figure 4.18b from the clustering of the *W* estimates in the range from 2% to 4 % close to the 1-to-1 line. But it is also evident that the exponential-law parameterization does not predict well the low (< 2%) and the high (> 4%) *W*-values. Similarly to the power-law parameterization, the exponential model still lacks the ability to predict the full range of possible *W* estimates.

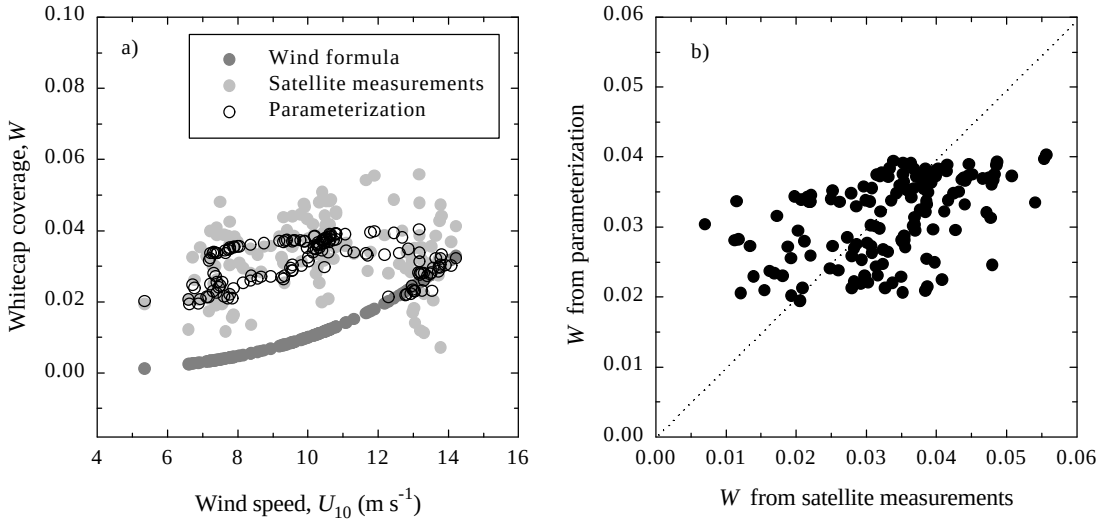


Figure 4.18 Performance of exponential law parameterization: a) comparison between satellite-measured *W* (light gray circles), $W(U_{10},T_s)$ parameterization (open circles), and $W(U_{10})$ parameterization (dark gray circles); b) comparison between satellite-measured *W* and $W(U_{10},T_s)$ parameterization.

Figure 4.19 shows the exponential-law performance for different temperature ranges. The model predicts the *W* variations at high temperatures (T_s from 20 to 30 °C) best, and misses the variability at moderate temperature (T_s from 10 to 20 °C) most severely.

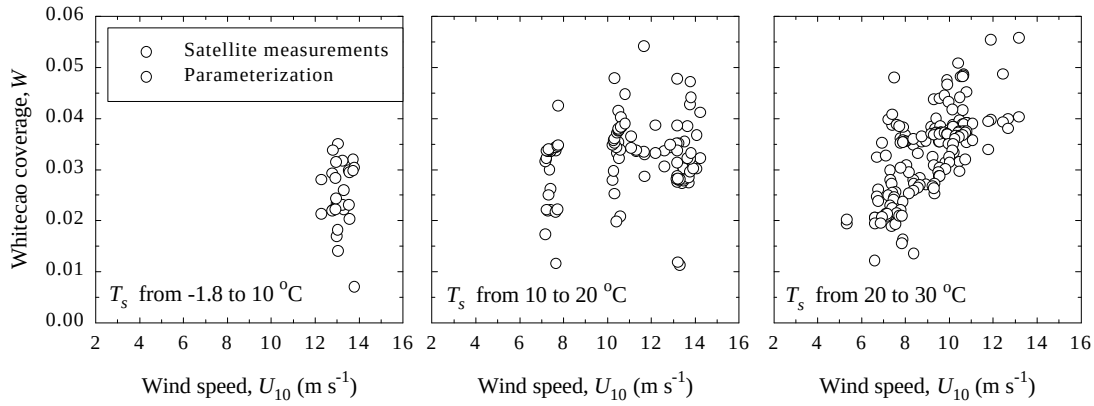


Figure 4.19 Comparison of satellite-measured and exponential-law parameterized *W* at different sea surface temperatures T_s.

Figure 4.20 shows a monthly map (March 1998) of the whitecap coverage estimated using the exponential-law parameterization of $W(U_{10}, T_s)$ and monthly maps of U_{10} and T_s. The strengths and weaknesses of the parameterization discussed in Figure 4.18 are also evident here. The range of *W*-values depicted in Figure 4.20 is restricted: *W* does not exceed 4%. The spatial distribution of whitecap coverage in Figure 4.20 is more similar to that displayed by the *W* estimates from satellite data (Figure 4.2) than to that estimated with (2.3) (Figure 4.3a). Yet, the parameterization shows less of the *W*-variability illustrated in Figure 4.2 and even misplaces some of the main features. For instance, the low *W* east of the tip of South America in Figure 4.2 is missing in Figure 4.20; in Figure 4.20, *W* is overestimated in East-South Pacific and Southern Atlantic; and, the variability at mid latitudes in North Atlantic, according to Figure 4.2, is not present in Figure 4.20.

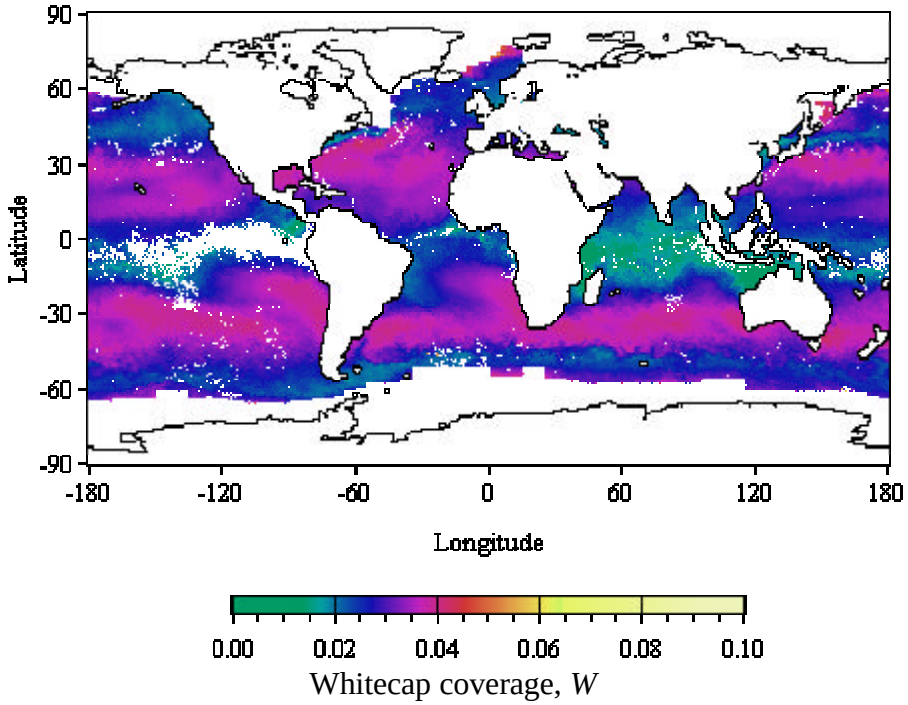


Figure 4.20 Monthly map (March 1998) of parameterized whitecap coverage $W(U_{10}, T_s) = a(U_{10}, T_s)\cdot\exp[b(U_{10}, T_s)^{x}U_{10}]$.

4.3.4 Summary and possible improvements

The initial results of parameterizing $W(U_{10}, T_s, S)$ show that the regression analysis is an appropriate tool for deriving a model accounting for the effects of various environmental parameters on the whitecap coverage. Problems and difficulties in implementing such a regression analysis on satellite-derived estimates of W are established and procedures for running the analysis are proposed. Several pertinent regression models are determined and applied to the data. The inclusion of only the sea surface temperature effect in the W parameterization, in addition to that of wind speed, improved the prediction of the W variability.

The analysis described here is only a first attempt. Undoubtedly, the inability of the proposed parameterizations to model the full range of *W*-variability is caused by the smoothing and then averaging of the observational data. The smoothing and the averaging are used in this first attempt to circumvent problems with the high variability of the data and the inability of the least square-method to deal with it (frequent failure of the normality and constant-variance tests). There are numerous possibilities deserving in-depth investigation, whose employment would make the smoothing and averaging a nuisance and would improve the implementing procedures. Some of these possibilities are listed.

Providing data for wind fetch and surfactant concentrations will help the data binning and most probably diminish the high data variability.

Regression functions suitable for data with high variability should be identified. Or, different regression methods, alternative to the least-square method, dealing better with highly variable data, should be scrutinized.

Different data binning and organization may provide a better population of the temperature and salinity bins with longer observational sets.

The full potential of the least-square method should be investigated by considering different combinations of regression models applied to the temperature and salinity dependencies of the regression coefficients. For instance, the proposed parameterization uses cubic polynomials for both coefficients, *a* and *b*, in the power-law and exponential-law models (Table 4.3). The performance of other functions modeling $a(T_s)$ and $b(T_s)$ should be tried, e.g., cubic polynomial and exponential law, cubic polynomial and power law, and so on.

A parameterization of the effect of salinity over a wider salinity range, from 20 to 40 psu, would most probably yield positive results.

Though more complex than single-variable regression analysis and advised against by Occam's razor, multiple regression analysis is worth investigation.