Deciphering the variability in air-sea gas transfer due to sea state and wind history

Mingxi Yang [ID][a,*], David Moffat [ID][a], Yuanxu Dong [ID][b,c] and Jean-Raymond Bidlot[d]

[a]Plymouth Marine Laboratory, Plymouth PL1 3DH, United Kingdom
[b]Marine Biogeochemistry Research Division, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel 24148, Germany
[c]Institute of Environmental Physics, Heidelberg University, 69120 Heidelberg, Germany
[d]European Centre for Medium-Range Weather Forecasts, Reading RG2 9AX, United Kingdom
*To whom correspondence should be addressed: Email: miya@pml.ac.uk
Edited By Vicki Grassian

Abstract

Understanding processes driving air-sea gas transfer and being able to model both its mean and variability are critical for studies of climate and carbon cycle. The air-sea gas transfer velocity (K_{660}) is almost universally parameterized as a function of wind speed in large scale models—an oversimplification that buries the mechanisms controlling K_{660} and neglects much natural variability. Sea state has long been speculated to affect gas transfer, but consistent relationships from in situ observations have been elusive. Here, applying a machine learning technique to an updated compilation of shipboard direct observations of the CO_2 transfer velocity ($K_{CO2,660}$), we show that the inclusion of significant wave height improves the model simulation of $K_{CO2,660}$, while parameters such as wave age, wave steepness, and swell-wind directional difference have little influence on $K_{CO2,660}$. Wind history is found to be important, as in high seas $K_{CO2,660}$ during periods of falling winds exceed periods of rising winds by ∼20% in the mean. This hysteresis in $K_{CO2,660}$ is consistent with the development of waves and increase in whitecap coverage as the seas mature. A similar hysteresis is absent from the transfer of a more soluble gas, confirming that the sea state dependence in $K_{CO2,660}$ is primarily due to bubble-mediated gas transfer upon wave breaking. We propose a new parameterization of $K_{CO2,660}$ as a function of wind stress and significant wave height, which resemble observed $K_{CO2,660}$ both in the mean and on short timescales.

Significance Statement

The ocean is a key sink of CO_2, and the instantaneous rate of ocean CO_2 uptake is proportional to the air-sea gas transfer velocity. Transfer velocity is almost universally parameterized as a function of wind speed only in global models, which is a reasonable approximation in the mean but neglects physical mechanisms and so variability. Here, combining the largest observational dataset to date we demonstrate that there are substantial variations in CO_2 transfer at short timescales due to sea state and wind history. We propose a new parameterization of the gas transfer velocity based on wind and waves data to more accurately predict air-sea CO_2 flux.

Introduction

The ocean has absorbed ∼30% of the atmospheric CO_2 emitted from human activities since the industrial revolution (e.g. (1)). Reducing the uncertainties in air-sea CO_2 flux and improving our understanding in the processes controlling it are thus critical for monitoring ongoing change and projecting the future climate. Air-sea flux of a gas such as CO_2 can be measured directly by the eddy covariance (EC) technique (e.g. (2)), but such monitoring has not been implemented on a large spatial and temporal scale. Instead, flux is generally estimated from a parameterization of the gas transfer velocity, K_{660}, and the gas concentration difference between air and water near the interface, ΔC: Flux = K_{660} $(Sc/660)^{-0.5}$ ΔC. Here, Sc is the Schmidt number of the gas in water. Thus, any error in K_{660} is directly propagated to the flux estimate and is in fact the dominant source of flux uncertainty (e.g. (3)).

Wind blowing over the ocean provides the predominant, but indirect forcing for K_{660}. This is because for sparingly soluble gases like CO_2 and dimethyl sulfide (DMS), their air-sea exchange is ultimately controlled by complex waterside processes (4). K_{660} in global models is almost universally parameterized as a simple function of wind speed (U_{10n}), oversimplifying the underlying physical mechanisms. Such K_{660} parameterizations become especially unsatisfying when gases of different solubility (e.g. CO_2 and DMS) require different wind speed fits (e.g. (5)).

Mechanistic models usually partition K_{660} into two parallel processes ($K_{660} = k_d + k_b$): (i) diffusive transfer through an unbroken surface (k_d) due to viscous wind stress and surface renewal, which is relatively more important at lower wind speeds and (ii) bubble-mediated transfer (k_b) upon wave breaking, which becomes increasingly important at higher wind speeds.

k_d for different waterside-controlled gases should be identical when normalized to a reference Schmidt number, and the transfer velocity of the more soluble DMS ($K_{DMS,660}$) is often thought to approximate k_d (e.g. (6, 7)). In contrast, k_b is relatively more important for the less soluble CO_2 than for DMS (5–9). Reichl and Deike (10) estimated that k_b contributes to 30% of CO_2 exchange globally by combining a mechanistic bubble-breaking model (11) with significant wave height data from the Wavewatch III model. A survey of literature, though, shows an order-of-magnitude discrepancy in the estimation of k_b (11–17), illustrating our poor understanding in bubble-mediated transfer.

Direct air-sea CO_2 flux measurements on ships by the EC method have improved dramatically in quality over the last 1.5 decades. Yang et al. (2) synthesized and reevaluated CO_2 flux observations from 11 cruises, primarily from the North Atlantic and Southern Ocean but also including the Tropical Indian Ocean and the Arctic. Dividing CO_2 flux by the concurrently measured air-sea CO_2 concentration difference and applying a Schmidt number normalization yields $K_{CO2,660}$. This foundational dataset (over 2,000 h) shows that the SD in the mean $K_{CO2,660}$ vs. wind speed relationship is about 20% (range >50%), with typically higher $K_{CO2,660}$ at a given wind speed in regions of larger waves such as the North Atlantic.

The physical interaction between wind and waves is complex and nonlinear. Simplistically, wind blowing over the ocean surface provides the momentum to produce windsea, which after some separation in time and space becomes swell. Breaking waves generate bubbles, which are manifested as whitecaps on the ocean surface. Sea state has long been speculated to affect K_{660} but deciphering consistent relationships from in situ data has been difficult. This is likely due to a combination of factors:

- K_{660} over the open ocean is challenging to measure. Each research cruise typically returns 100–200 valid data points (if measured by EC) or a few data points (if inferred from tracers, e.g. (18)), which do not cover the full range of sea states.
- There is substantial scatter in the K_{660} observations. Until recently, the partitioning in scatter between natural variability and measurement uncertainty was unclear. Dong et al. (19) comprehensively determined the random uncertainty in EC $K_{CO2,660}$ data to be about 30% under typical conditions for hourly measurements (generally decreasing with increasing wind speed and flux magnitude).
- Wind and wave parameters tend to correlate with each other, and which wave parameters are of importance toward K_{660} is not well known.

Some sea state dependencies in K_{660} have been proposed, but with little consensus. Zhao et al. (20) proposed the wave Reynolds number ($R_{Hw} = u_* \times H_s/v_w$) that includes both wind and wave information as a parameter to describe K_{660}. Here, u_*, H_s, and v_w are the friction velocity (related to surface wind stress), significant wave height, and water viscosity, respectively. This theoretical framework implies that K_{660} should be greater in more developed seas than in developing seas. Brumer et al. (21) showed that the use of R_{Hw} collapses different field observations of K_{660} better than wind speed. Blomquist et al. (5) and Fairall et al. (17) adopted this $R_{Hw}^{0.9}$ scaling for k_b. Deike and Melville (11) developed a mechanistic model of k_b, which can either take on a spectral form or be related to bulk wave parameters: proportional to $u_*^{5/3} H_s^{2/3}$. Using the spectral model, k_b was modeled to be greater during the developing phase of a storm in the Southern Ocean (22), in contrast to

the R_{Hw} formulation. Zavarsky and Marandino (23) further suggested a dependence in $K_{DMS,660}$ on the swell-wind directional difference, which certain directions causing more "shielding" of wind from swell, resulting in suppressed transfer.

In this work, we investigate the sea state dependencies in K_{660} from multiple datasets of $K_{CO2,660}$, $K_{DMS,660}$, waves, and whitecap coverage. We first quantify the variability in the $K_{CO2,660}$ observations that cannot be explained by wind speed and by measurement noise. We then use a machine learning (ML) method to "agnostically" investigate relationships between $K_{CO2,660}$ and various wave parameters. The ML approach elucidates the potential key controlling parameters for $K_{CO2,660}$. Further in depth analysis of these parameters as well as comparison between $K_{CO2,660}$ and $K_{DMS,660}$ enable us to tease out the driver for the sea state dependence in $K_{CO2,660}$. Finally, we propose and evaluate a new parameterization of $K_{CO2,660}$ based on wind and bulk wave data.

Variability in the hourly $K_{CO2,660}$ observations

In this paper, we have added ∼600 hours of $K_{CO2,660}$ data from four recent cruises in the Southern Ocean (24, 25) to the foundational dataset of $K_{CO2,660}$ observations (∼2000 hours) from 11 cruises by Yang et al. (2), forming the largest $K_{CO2,660}$ dataset to date. Please refer to these two papers for the exact location/time of the cruises. A simple power fit as a function of U_{10n} to all the $K_{CO2,660}$ observations returns a R^2 of 0.70 (Fig. 1A). How much of the residual variance is due to noise vs. natural variability? To answer this, we generate a set of synthetic $K_{CO2,660}$ data based on the U_{10n} fit with a Gaussian random noise that averages 30% (19). This synthetic data shows that the maximum possible R^2 between $K_{CO2,660}$ and a dependent variable within this dataset is 0.91 (Fig. 1B). Wind is indeed the predominant driver for gas transfer, but there remains substantial variability not explained by U_{10n} (variance gap of 0.91–0.70 = 0.21). This is particularly true at high wind speeds (e.g. 15 to 20 m s^{-1}), where variance in the synthetic $K_{CO2,660}$ dataset is only about a quarter of the variance in the observations. Clearly, confining our attention to averages of $K_{CO2,660}$ in wind speed bins, as is the norm in past decades of research, neglects much natural variability caused by other factors such as sea state.

Deciphering sea state dependence in $K_{CO2,660}$ using ML

To investigate controlling factors in gas transfer, we developed a ML model (random forest, (26)) for the prediction of hourly $K_{CO2,660}$ based only on the foundational dataset of Yang et al. (2). Model input parameters, consisting primarily of wind (in situ) and bulk wave (ECMWF) parameters, are detailed in the Supplementary Material. Wave parameters from ECMWF were computed from the full modeled wave spectrum, as well as separately for windsea and swell components. In the spectral wave model, the area of the spectrum where the wind input is actively transferring momentum into waves is defined as windsea. The remaining part of the wave spectrum is defined as swell.

The random forest model was developed with 200 trees, a leaf size of one, and the trees were trained to minimize the means squared error. From the cruise data collected, periods of incomplete data were removed, and standard scaling was performed to normalize the rest of the data to zero mean and unit variance. The data were then randomly split into training (65%) and testing (35%) datasets (see Fig. 2D as an example). The 65/35% split was
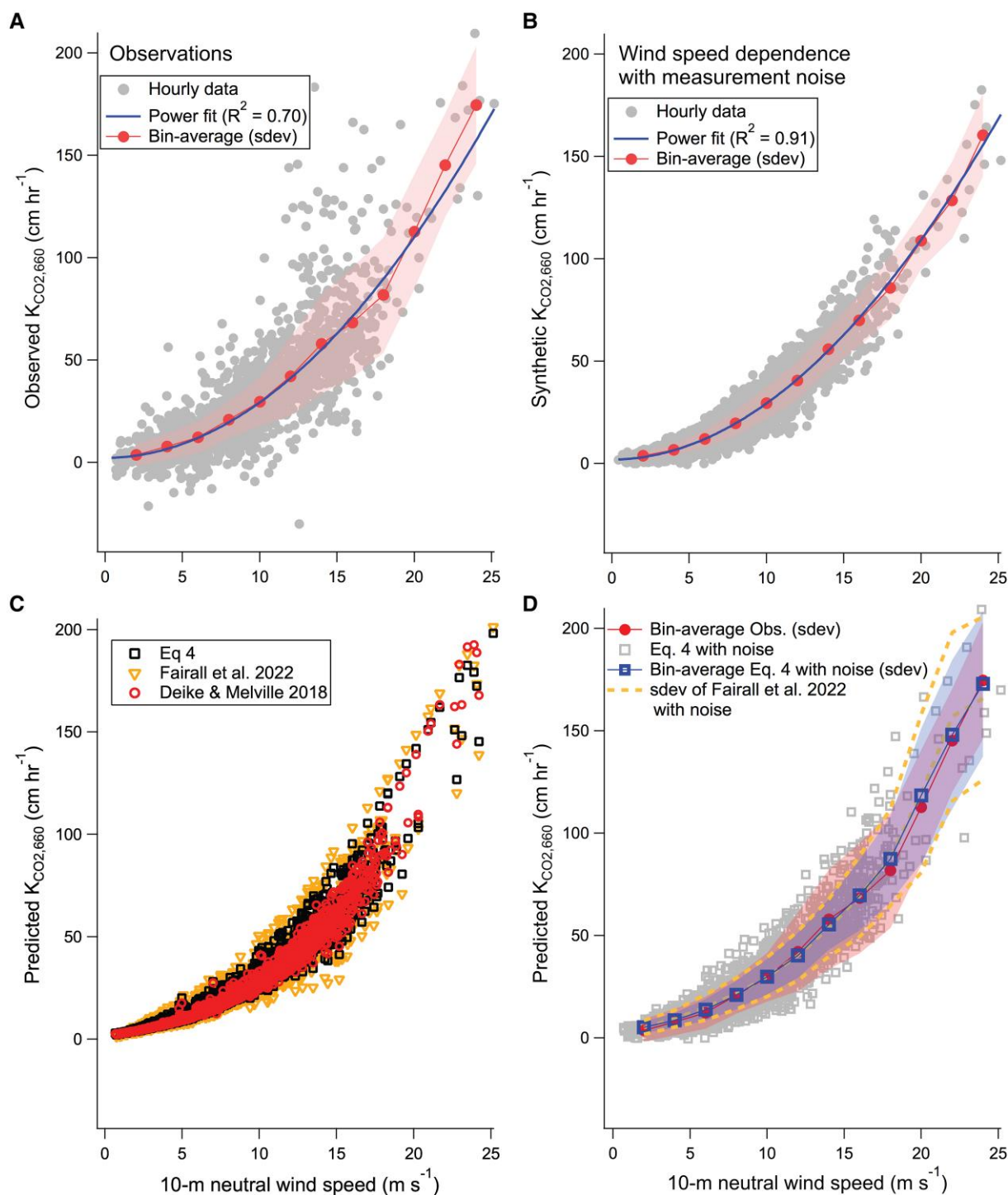
Fig. 1. A) Observed $K_{CO2,660}$ vs. wind speed (hourly) from the entire dataset; the thick solid line shows the mean wind speed dependence (power fit) and the thin line with markers show bin-average with SD. B) Synthetic $K_{CO2,660}$ (hourly) based on the wind speed dependence, accounting for random measurement noise. C) Predicted $K_{CO2,660}$ (hourly) from Eq. 4, Deike and Melville (11), and Fairall et al. (17). D) Predicted $K_{CO2,660}$ (hourly) from Eq. 4 with random measurement noise, which approaches observations in the mean and in variability. Variability predicted by Fairall et al. (17) appears to be too large at high wind speeds relative to observations.

applied to all the cruises within the foundational dataset to ensure suitable temporal and geographical representations within both the training and testing datasets. We reserve the new Southern Ocean data from Dong et al. (24, 25) as an independent dataset for further model validation.

Within the foundational dataset, the ML model captures a R^2 of 0.97 in the training data and 0.81 for the testing data (Figs. S1 and S2). The R^2 for training exceeds the theoretical maximum R^2 of 0.91 (Fig. 1B). This reflects a small degree of "over-fitting" by the ML model, which does not appear to affect how the ML predicts nontraining data (Fig. 2D). The R^2 for testing surpasses the R^2 of a simple wind speed fit by 0.11 but is lower than the maximum possible R^2 of 0.91, likely in part because we have neglected in the ML model some other processes that can also affect $K_{CO2,660}$,
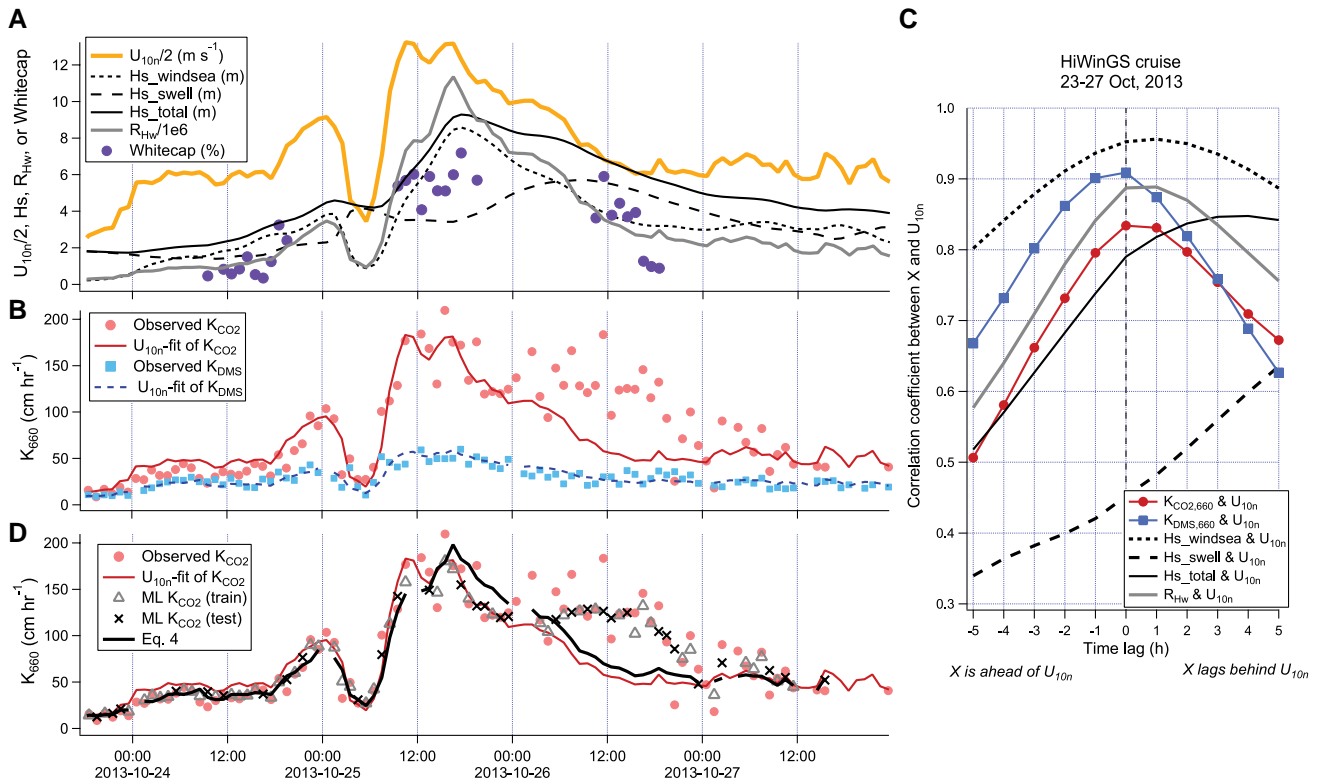
Fig. 2. An example of the impact of waves on gas transfer during the "St. Jude" storm of the HiWinGS cruise. A) Wind speed peaked in excess of 25 m s^{-1} on 25 October, while significant wave heights (total and swell) and whitecap fraction remained elevated for hours longer. B) A simple U_{10n} dependent fit to the HiWinGS $K_{CO2,660}$ observations tends to overestimate during periods of rising winds and developing seas (e.g. 24 October), and substantially underestimate during periods of falling winds and more developed seas (e.g. 26 October). C) Lag correlation analysis of the period during the St. Jude storm shows that $K_{CO2,660}$ is lagged relative to U_{10n}, and this hysteresis is most similar to that of the wave Reynolds number (here the viscosity of water at 20°C was used to compute R_{Hw}). A similar sea state dependent hysteresis is not observed in $K_{DMS,660}$, which is well described by a U_{10n} dependence throughout the storm. D) Predictions from the ML model and Eq. 4 of $K_{CO2,660}$. Overall, the ML model generalizes the mean trend of the observations well. A small degree of over-fitting is apparent as the ML model often tries to match the individual "wiggles" in the training data, which are partially due to measurement noise. Equation 4 from this work outperforms a simple wind speed fit but does not capture as much natural variability as the ML model following this extreme storm event (e.g. 26 October).

such as surfactants (e.g. (27, 28)) and convection (29). A separate random forest ML model run with wind input only yields an R^2 of 0.75 for the testing data. This shows that (i) inclusion of wave parameters improves the ML model of $K_{CO2,660}$ (by R^2 of about $0.81 - 0.75 = 0.06$), and (ii) ML is able to identify wind-dependent patterns in $K_{CO2,660}$ (e.g. wind history) that a simple wind speed fit neglects (by R^2 of about $0.75 - 0.70 = 0.05$).

When we validate the ML model with wave parameters against the new $K_{CO2,660}$ observations that were not used for the training (24, 25), the model slightly over predicts (R^2 of 0.66, Fig. S2), perhaps due to the aforementioned other controlling factors for gas transfer that have not been considered. The Pearson correlations between observations and predictions remains high though ($r^2 = 0.74$), implying that the ML model does capture most of the variations in $K_{CO2,660}$ in the validation dataset. The random forest ML algorithm is publicly available at GITHUB (https://github.com/djmoffat/air-sea-gas-prediction). We also explored the use of a small Artificial Neural Network, but this underperformed compared with the random forest model, likely due to the still relatively limited observational dataset.

To further assess which input predictor has the largest impact on the predictions from the ML model, we turn to SHAP (SHapley Additive exPlanations; (30)) analysis. The SHAP value (Fig. S3) shows that the most important parameters in predicting $K_{CO2,660}$ are U_{10n} and significant wave heights (windsea, swell, and total),

while parameters such as wave age, swell-wind direction difference, swell period, and wave steepness are generally not important.

Example of sea state dependence from the HiWinGS cruise

Clear sea state dependence in $K_{CO2,660}$ is apparent in the data from the HiWinGS cruise (2013 in the North Atlantic; originally published by (5) and recomputed in the case of CO_2 by (2)), during which the ship stationed through multiple exceptionally large storms (Fig. 2). The highest wind speed (>25 m s^{-1}) was observed on 25th October 2013 in what eventually was named the "St Jude" storm upon landfall in the United Kingdom. $K_{CO2,660}$ was expectedly very high at the peak in wind speed, but interestingly the high $K_{CO2,660}$ values persisted for many hours following the wind peak. A simple U_{10n} fit to all the HiWinGS $K_{CO2,660}$ observations clearly underestimates the observed $K_{CO2,660}$ for most of 26th October, when the seas declined following the peak of the storm. At the same time, the U_{10n} fit overestimates $K_{CO2,660}$ for most of 24th October, when the seas were building up. In contrast, we do not see any difference in $K_{DMS,660}$ (a proxy of diffusive transfer) between rising and falling winds/seas during the HiWinGS cruise.

Fairly long periods of continuous measurements and the large range in wind conditions during HiWinGS allow us to assess the time lag between gas transfer and its drivers (Fig. 2C). Consistent
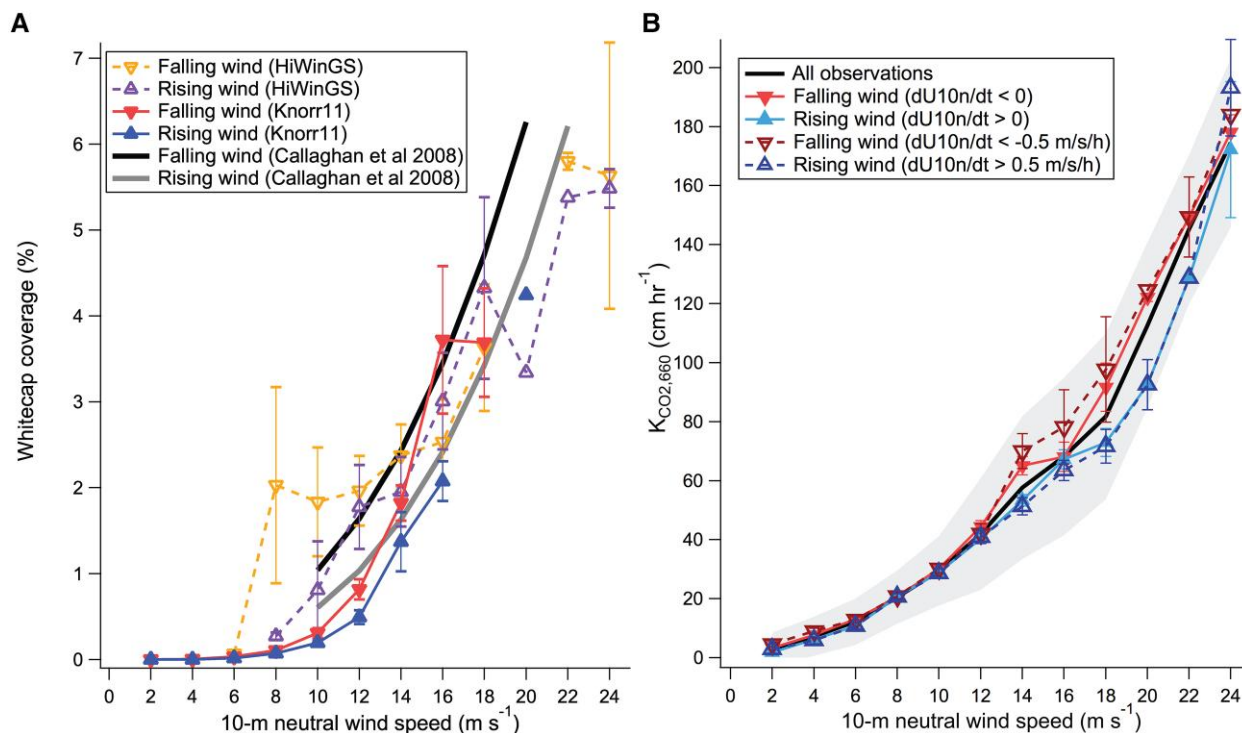
Fig. 3. Wind history dependence in whitecap coverage from three different cruises (A) and in CO_2 transfer from 15 cruises (B). At high wind speeds, whitecap coverage tends to be greater during falling winds than during rising winds (by about 50% for the Knorr11 cruise and (33)), while $K_{CO2,660}$ is on average about 20% greater during falling winds than during rising winds. The error bars on rise/fall categories indicate SE, while the gray shading on the overall bin-average in B (black line) indicates SD.

with our understanding of the development of the seas, windsea lagged behind wind by 1–2 h, while swell lagged behind wind by much longer. There is a clear asymmetry in the $K_{CO2,660}$:U_{10n} lag correlation. This hysteresis implies that $K_{CO2,660}$ has some lag relative to U_{10n}, similar to the behavior of the wave Reynolds number computed from the total significant wave height. In contrast to $K_{CO2,660}$, $K_{DMS,660}$ correlates symmetrically with respect to U_{10n} and its peak correlation with U_{10n} is higher than that of $K_{CO2,660}$, suggesting minimal sea state effect.

The measured whitecap coverage (W_f) was nearly three times higher for most of 26th October than on 24th October, despite very similar wind speeds at around 12 m s^{-1}. Measurements of bubbles during the HiWinGS cruise also showed higher bubble void fractions during falling winds than rising winds (31). On aggregate, the higher whitecap coverage during falling winds and more developed seas agrees well with previous observations from the Knorr11 cruise (32) and with Callaghan et al. (33), as shown in Fig. 3A and Table S1. Here rising (falling) wind is defined as dU_{10n}/dt of the preceding three hours (including hour of interest) being >0 (<0), as in Hanson and Phillips (34) and Callaghan et al. (33). An analogous relationship with wind history was observed for sea spray flux from a coastal site (35). The similar behaviors in $K_{CO2,660}$ and W_f toward wind history and the lack of wind history dependence in $K_{DMS,660}$ suggest that the hysteresis in $K_{CO2,660}$ is mostly due to bubble-mediated transfer upon wave breaking.

Mean sea state and wind history dependencies in $K_{CO2,660}$

Conditions during the HiWinGS cruise were rather extreme but a similar sea state dependence is apparent at lower wind speeds. We now turn our attention to the entire $K_{CO2,660}$ dataset. As with

W_f, for each hour we compute wind history as dU_{10n}/dt over the preceding three hours. The $K_{CO2,660}$ data are then stratified into categories of "rising wind" (dU_{10n}/dt >0, N = 1074; or >0.5 m s^{-1} h^{-1}, N = 465) and "falling wind" (dU_{10n}/dt <0, N = 1197; or < −0.5 m s^{-1} h^{-1}, N = 487). At the same wind speed, on average $K_{CO2,660}$ during falling winds exceeds $K_{CO2,660}$ during rising winds by on the order of 20%, with the steeper threshold of |0.5 m s^{-1} h^{-1}| usually leading to a greater enhancement than a threshold of 0 (Fig. 3B; Table S1). The enhancement appears to be most pronounced at wind speeds above 12 m s^{-1}, though above 20 m s^{-1} the number of observations becomes very limited.

Periods of falling wind are typically associated with more developed seas (and relatively high wave age), while periods of increasing wind are typically associated with less developed seas (and relatively low wave age). The ML model does not find wave age (and inverse wave age) to be an important controller factor of $K_{CO2,660}$. This may be because the parameter wave age (phase speed of wave divided by wind speed) is highly dependent on wind speed, with high wind speed conditions usually corresponding to less developed seas (e.g. (36)). Including it thus provides limited additional information over what is already provided by wind speed. In comparison, wind history is not very dependent on wind speed on the whole.

So what causes the wind history dependence in the $K_{CO2,660}$ observations? Here, we separate the hourly $K_{CO2,660}$ data in 1 m s^{-1} wind speed bins (to remove the U_{10n} dependence) and then compute the correlations between $K_{CO2,660}$ and significant wave heights as well as wind history within the bins (Table S2). The key findings from this analysis are:

- Correlations between $K_{CO2,660}$ and significant wave heights are almost always positive, and the correlations are usually statistically significant (95% confidence) at U_{10n} above 7 m s^{-1}, consistent with substantial bubble-mediated transfer for CO_2.

- Total significant wave height correlates most strongly with $K_{CO2,660}$, with both windsea and swell components contributing toward the correlation, qualitatively consistent with ML SHAP analysis (Fig. S3).
- Correlations between $K_{CO2,660}$ and wind history are generally negative (consistent with Fig. 3B), but the correlations are weaker than with significant wave heights.

The above findings support the approach of parametrizing k_b (and not K_{660} directly) as a function of total H_s, as in Blomquist et al. (5), Deike and Melville (11), and Fairall et al. (17). The wind history dependence in $K_{CO2,660}$ appears to reflect the H_s dependence to a large extent—that $K_{CO2,660}$ is higher during falling winds than rising winds at the same wind speed is partly because waves (and whitecap coverage) tend to be greater during falling winds than rising winds (Fig. S4). This understanding forms the foundation for our new parameterization of $K_{CO2,660}$ based on bulk wind and sea state parameters.

Building a new wind-wave parameterization of $K_{CO2,660}$

Deike and Melville (11) proposed a parameterization of K_{660} based on bulk wind and wave parameters:

$$K_{660} = k_d + k_b = Au_* + Bu_*{}^{5/3}(gH_s)^{2/3} \qquad (1)$$

Here, we have for simplicity removed from their formula the dependence on Ostwald solubility (which for CO_2 is close to unity). A and B are tunable coefficients, while g is gravity. Zhou et al. (37) adopted the same formulation as Eq. 1 but with different A and B coefficients.

Fairall et al. (17), following Blomquist et al. (5), proposed an alternative representation of K_{660} based on the COARE gas transfer model, which (ignoring buoyance effects at low wind speeds) can be simply represented as:

$$K_{660} = k_d + k_b = Au_{v*} + BW_f = Au_{v*} + B(R_{Hw})^{0.9}. \qquad (2)$$

Here, u_{v*} is the viscous part of u_* — a parameter described by Mueller and Veron (38) that cannot be readily measured in the field. W_f is the whitecap coverage, which is parameterized as a function of the wave Reynolds number ($R_{Hw} = u_* H_s/v_w$ at a mean HiWinGS temperature of 10°C) according to the works from Brumer et al. (39). A and B again are tunable coefficients, which are constrained by limited observations of both $K_{CO2,660}$ and $K_{DMS,660}$. Compared with Deike and Melville (11), k_b in Fairall et al. (17) is ~70% greater while k_d is 30%~smaller.

In both equations above, k_b depends on wind stress as well as wave heights. Yang et al. (2) showed that at low to moderate wind speeds, $K_{CO2,660}$ has an essentially linear dependence on u_*, consistent with the idea that diffusive transfer is driven by wind stress. Here, we opt to parameterize k_d as a function of u_* instead of u_{v*}, since the former is more readily measurable in the field and already available in large datasets (e.g. ECMWF reanalysis). Since W_f appears to have a near linear dependence on R_{Hw} (39), for our parameterization we start with the basic form below:

$$K_{CO2,660} = k_d + k_b = Au_* + BW_f = Au_* + Bu_*H_s. \qquad (3)$$

Here, we have neglected the dependence on v_w, which is convenient for simplifying units in R_{Hw}, but does not help to constrain variability in $K_{CO2,660}$ (2). The bulk u_* is used, which is derived from in situ meteorological and underway seawater observations using the COARE3.5 model (with a wind speed dependent Charnock relation). We note that during windsea-dominated

conditions, H_s approximately scales with u_*^2. Then, the k_b term in Eq. 3 (as well as in Eq. 1) scales with u_*^3, which is in line with historical W_f observations and the concept that W_f is strongly related to the energy flux from the wind (40). Compared with the Eq. 1, the k_b term in Eq. 3 is more weighted toward wave height and less toward wind stress in the presence of swell.

Since $K_{CO2,660}$ in Eqs. 1 and 3 are both dependent on u_*, we can divide observed $K_{CO2,660}$ by u_* to evaluate the functional form for k_b and assess the coefficients A and B. For this analysis, we neglect periods when $u_* < 0.1$ m s^{-1}, as the relative measurement uncertainty is substantially larger and other effects such as buoyancy may become important under these calm conditions. For the Deike and Melville (11) approach, $K_{CO2,660}/u_*$ appears to have a nonlinear relationship with $(u_* g H_s)^{2/3}$ (Fig. 4A). In contrast, $K_{CO2,660}/u_*$ and H_s has essentially a linear relationship (Fig. 4B). From this it seems that the functional form of Eq. 3 is reasonable and seems more consistent with $K_{CO2,660}$ observations than Eq. 1 (11) over the full range of sea state.

If we fit the observed $K_{CO2,660}$ as a function of u_* and H_s simultaneously following the functional form of Eq. 3, we arrive at the following parameterizations of $K_{CO2,660}$ (in units of cm h^{-1}):

$$K_{CO2,660} = 360,000(1.52e-4u_* + 2.90e-5u_*H_s). \qquad (4)$$

Interestingly, fitting $K_{CO2,660}$ with $H_{s\_windsea}$ instead of total H_s does not improve the fit, similar to findings from Blomquist et al. (5) and Brumer et al. (21). Swell appears to be important for CO_2 transfer, as also implied from the SHAP analysis (Fig. S3) as well as from correlations between $K_{CO2,660}$ and the different H_s components (Table S2). It is worth cautioning though that the spectral separation between windsea and swell in the ECMWF wave model is fairly simplistic. In cases of rapidly changing wind fields, the part of the spectrum that is defined as swell might still contain steep breaking waves that contribute to gas exchange.

Tuned to overlapping EC datasets, recent parameterizations based on wind and waves all predict similar $K_{CO2,660}$ in the mean. However, these parameterizations differ in the partitioning between k_d and k_b. At a wind speed of 15 m s^{-1}, k_b accounts for on average 39% of total $K_{CO2,660}$ in Deike and Melville (11), 66% in Fairall et al. (17), 43% in Zhou et al. (37), and 42% according to Eq. 4. At this wind speed, the relative difference in observed $K_{CO2,660}$ between rising/falling wind is on the order of 20%, while the relative difference in W_f between rising/falling wind is on the order of 50% (Fig. 3; Table S1). If k_b scales proportionally with W_f, we expect k_b to contribute roughly 40% of the total $K_{CO2,660}$ (20%/0.5), which is more consistent with Eq. 4 (as well as (11, 37)), and less than the prediction by Fairall et al. (17).

Out of these parameterizations, Eq. 4 generally matches observations the best when averaged in bins of wind speed, wind history, and swell impact. This is demonstrated in Fig. 5 as the ratio between observed and parameterized $K_{CO2,660}$, with the most optimal parameterization giving a ratio of unity during all conditions and showing no trend. Parametrizations from Deike and Melville (11) and Zhou et al. (37) underpredict $K_{CO2,660}$ at moderate to high wind speeds (8 to 16 m s^{-1}) and slightly overpredict $K_{CO2,660}$ at even higher wind speeds (Fig. 5A), probably related to the nonlinearity in their formula as shown in Fig. 4A. In contrast, Eq. 4 and Fairall et al. (17) both reproduce the mean wind speed dependence in $K_{CO2,660}$ reasonably well.

Figure 5B, like Fig. 3B, illustrates the wind history dependence in $K_{CO2,660}$ at wind speeds over 7 m s^{-1} (i.e. observation/wind speed fit is negatively correlated with wind history). This wind history dependence is well accounted for by Eq. 4. Formulations from
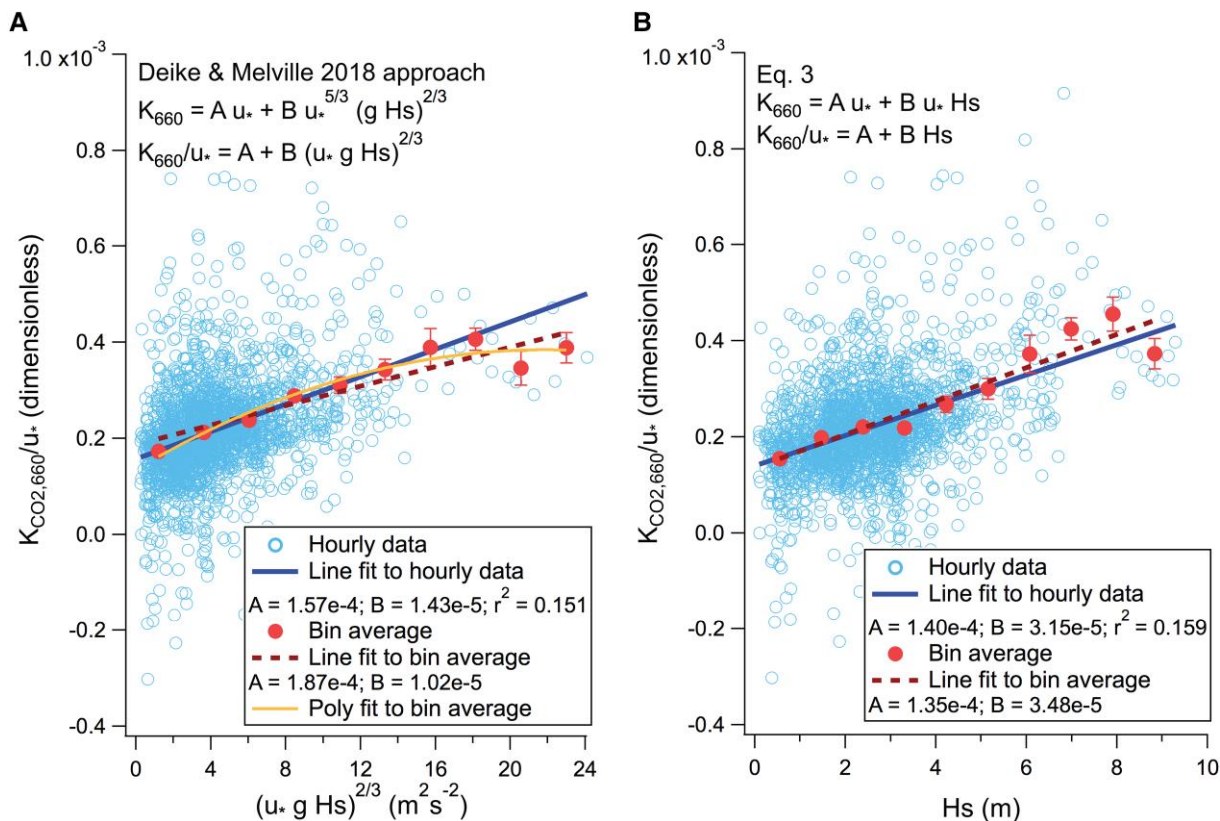
Fig. 4. A) $K_{CO2,660}/u_*$ vs. $(u_* g H_s)^{2/3}$, which clearly shows a nonlinear relationship. Here, the intercept and the slope correspond to tuning coefficients A for diffusive transfer (dimensionless) and B for bubble-mediated transfer (s^2 m^{-2}); B) $K_{CO2,660}/u_*$ vs. H_s, which has a more linear relationship and higher r^2. A polynomial fit to (B) looks nearly identical to the linear fit. Here, the intercept and the slope correspond to tuning coefficients A for diffusive transfer (dimensionless) and B for bubble-mediated transfer (m^{-1}).
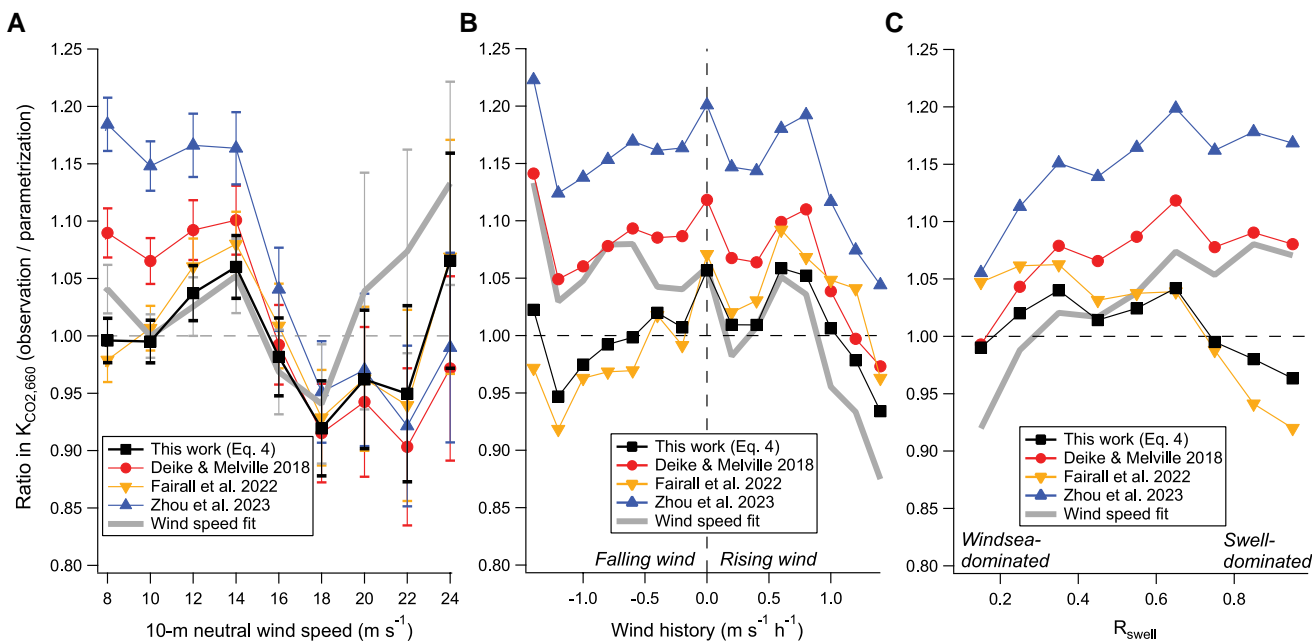

Fig. 5. A) Bin-averaged ratio between observed and parameterized $K_{CO2,660}$ at wind speeds above 7 m s^{-1} where wave breaking becomes important, with error bars indicate SE. B) Same as A), but in bins of wind history. Error bars not shown to avoid visual clutter. C) Same as B), but in bins of R_{swell}.

Deike and Melville (11) and Zhou et al. (37) also account for the relative wind history trend, but on average underestimates $K_{CO2,660}$. The Fairall et al. (17) formula, which has a ~60% larger k_b than Eq. 4, appears to inadvertently over-correct for this wind history dependence.

Observed $K_{CO2,660}$ at a given wind speed appears to be greater during swell-dominated conditions (here indicated by $R_{swell} = (Hs\_swell/H_{s\_total})^2$) than windsea-dominated conditions (Fig. 5C). This could be another sign of the wind history dependence, and/or be related to the idea that breaking in aged seas is

less energetic but more conducive to producing longer-lasting bubbles (41). Parametrizations from Deike and Melville (11) and Zhou et al. (37) do not reduce this swell dependence, while Fairall et al. (17) appears to over-correct for it. Equation 4 is again the most optimal. We note that re-tuning Eqs. 1 and 2 with the latest observations only very marginally improves their performance and does not alter the general trends above.

Turning our attention over to short term variability, Eq. 4 gives a R^2 of 0.753 when applied to the hourly testing dataset for the ML model, which is significantly better than wind speed (0.693 for the same dataset) but falls short compared with the ML model (0.813), probably because there are other subtle wind/wave dependent effects that Eq. 4 does not consider. Nevertheless, we can estimate the sea state driven $K_{CO2,660}$ variability by evaluating the different gas transfer parameterizations at the ambient u_* and H_s (Fig. 1C). At U_{10n} of 15 m s^{-1}, the standard deviation (σ) of the predicted $K_{CO2,660}$ from Eq. 4 is about 10 cm h^{-1}. This sea state effect can account for over a third of the variance (σ^2) in observed $K_{CO2,660}$ at these high wind speeds once measurement noise is considered (Fig. 1A and D). In Eq. 2, the use of u_{v*} instead of u_* to fit k_d necessitates a ∼60% larger k_b compared with Eq. 4. The sea state driven variability in $K_{CO2,660}$ from Fairall et al. (17), even without considering measurement noise, exceeds the variability in observed $K_{CO2,660}$ at high wind speeds (Fig. 1D), again hinting that their k_b may be too large.

Concluding remarks

Parameterizing K_{660} as a simple function of wind speed, given sufficient observations across a wide range of sea states, can yield a reasonable "climatological fit" in the mean. This study reveals significant sea state dependent variability in observed hourly $K_{CO2,660}$ that is due to bubble-mediated transfer—a process that differs between gases of different solubility. We provide a new method of parameterizing $K_{CO2,660}$ using wind/wave data (Eqs. 3 and 4) that explains more observed variability than a wind speed fit, and better accounts for the effects of wind history and swell than other recently proposed wind-wave parameterizations.

It is worth noting that Eq. 4, developed here for CO$_2$, should not be used "as is" to predict transfer of other gases (e.g. DMS). A universal framework for modeling gas transfer requires the specification of the solubility dependence in k_b (e.g. (8, 42)), which is not very well understood. Quantification of K_{660} of another waterside-controlled gas (in addition to CO$_2$ and DMS) should help to further constrain the impact of bubble-mediated transfer.

This work has made use of modeled wave data to decipher patterns in field $K_{CO2,660}$ measurements. There have been very few field campaigns with concurrent wave, whitecap/bubble, and K_{660} measurements. Further improvements in mechanistic understanding require more of these concurrent measurements under a wide range of conditions. We have used ML techniques here to mostly estimate the total variability that is explainable by wind and wave data, as well as identify wave parameters that influence $K_{CO2,660}$. The total number of direct $K_{CO2,660}$ observations made by the international community to date is still rather limited (a few thousand hours). We anticipate that more $K_{CO2,660}$ observations will likely lead to further improvements in the ML-based predictions of gas transfer. Recent developments in buoy-based flux measurements (43) appear to offer a highly promising approach to drastically increase the number of $K_{CO2,660}$ observations.

Acknowledgments

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

Data Availability

This article mostly makes use of already published data: Yang et al. (2) synthesis of $K_{CO2,660}$ and wave data; HiWinGS $K_{DMS,660}$ data from Blomquist et al. (5); whitecap data from Scanlon and Ward (32), Callaghan et al. (33), and Brumer et al. (39). New $K_{CO2,660}$ observations from Dong et al. (24, 25) along with wave data are included in the Supplementary Material of this article.

References

1 Khatiwala S, *et al*. 2013. Global ocean storage of anthropogenic carbon. *Biogeosciences*. 10(4):2169–2191.

2 Yang M, *et al*. 2022. Global synthesis of air-sea CO2 transfer velocity estimates from ship-based eddy covariance measurements. *Front Mar Sci*. 9:826421.

3 Woolf DK, *et al*. 2019. Key uncertainties in the recent air-sea flux of CO2. *Global Biogeochem Cycles*. 33(12):1548–1563.

4 Liss PS, Slater PG. 1974. Flux of gases across the air-sea interface. *Nature*. 247(5438):181–184.

5 Blomquist BW, *et al*. 2017. Wind speed and sea state dependencies of air-sea gas transfer: results from the high wind speed gas exchange study (HiWinGS). *J Geophys Res*. 122:8034–8062.

6 Blomquist BW, Fairall CW, Huebert BJ, Kieber DJ, Westby GR. 2006. DMS sea-air transfer velocity: direct measurements by eddy covariance and parameterization based on the NOAA/COARE gas transfer model. *Geophys Res Lett*. 33(7):L07601.

7 Bell TG, *et al*. 2017. Estimation of bubble-mediated air–sea gas exchange from concurrent DMS and CO2 transfer velocities at intermediate–high wind speeds. *Atmos Chem Phys*. 17(14): 9019–9033.

8 Woolf DK. 1997. Bubbles and their role in gas exchange. In: Liss PS, Duce RA, editors. *The sea surface and global change*. Cambridge University Press. p. 173–206.

9 Yang M, Blomquist BW, Fairall CW, Archer SD, Huebert BJ. 2011. Air-sea exchange of dimethylsulfide in the Southern Ocean: measurements from SO GasEx compared to temperate and tropical regions. *J Geophys Res Oceans*. 116(C4):C00F05.

10 Reichl BG, Deike L. 2020. Contribution of sea-state dependent bubbles to air-sea carbon dioxide fluxes. *Geophys Res Lett*. 47(9): e2020GL087267.

11 Deike L, Melville WK. 2018. Gas transfer by breaking waves. *Geophys Res Lett*. 45:10482–10492.

12 Asher WE, *et al.* 1996. The influence of bubble plumes on air-seawater gas transfer velocities. *J Geophys Res.* 101:12027–12041.

13 Stanley RHR, Jenkins WJ, Lott DE III, Doney SC. 2009. Noble gas constraints on air-sea gas exchange and bubble fluxes. *J Geophys Res.* 114:C11020.

14 Liang J-H, *et al.* 2013. Parameterizing bubble-mediated air-sea gas exchange and its effect on ocean ventilation. *Global Biogeochem Cycles.* 27:894–905.

15 Goddijn-Murphy L, Woolf DK, Callaghan AH, Nightingale PD, Shutler JD. 2016. A reconciliation of empirical and mechanistic models of the air-sea gas transfer velocity. *J Geophys Res.* 121: 818–835.

16 Krall KE, Smith AW, Takagaki N, Jähne B. 2019. Air–sea gas exchange at wind speeds up to 85 m s^{-1}. *Ocean Sci.* 15(6):1783–1799.

17 Fairall CW, *et al.* 2022. Air-sea trace gas fluxes: direct and indirect measurements. *Front Mar Sci.* 9:826606.

18 Ho DT, *et al.* 2006. Measurements of air-sea gas exchange at high wind speeds in the Southern Ocean: implications for global parameterizations. *Geophys Res Lett.* 33:L16611.

19 Dong Y, Yang M, Bakker DC, Kitidis V, Bell TG. 2021. Uncertainties in eddy covariance air–sea CO_2 flux measurements and implications for gas transfer velocity parameterisations. *Atmos Chem Phys.* 21(10):8089–8110.

20 Zhao D, Toba Y, Suzuki Y, Komori S. 2003. Effect of wind waves on air'sea gas exchange: proposal of an overall CO2 transfer velocity formula as a function of breaking-wave parameter. *Tellus B Chem Phys Meteorol.* 55(2):478–487.

21 Brumer SE, *et al.* 2017. Wave-related Reynolds number parameterizations of CO2 and DMS transfer velocities. *Geophys Res Lett.* 44(19):9865–9875.

22 Deike L. 2022. Mass transfer at the ocean–atmosphere interface: the role of wave breaking, droplets, and bubbles. *Annu Rev Fluid Mech.* 54:191–224.

23 Zavarsky A, Marandino CA. 2019. The influence of transformed Reynolds number suppression on gas transfer parameterizations and global DMS and CO_2 fluxes. *Atmos Chem Phys.* 19(3):1819–1834.

24 Dong Y, *et al.* 2024. Data from: direct observational evidence of strong CO_2 uptake in the Southern Ocean [dataset]. *Dryad* [online]. [accessed 2024 Sep 12]. https://doi.org/10.5061/dryad.b2rbnzspm.

25 Dong Y, *et al.* 2024. Direct observational evidence of strong CO_2 uptake in the Southern Ocean. *Sci Adv.* 10(30):eadn5781.

26 Breiman L. 2001. Random forests. *Mach Learn.* 45:5–32.

27 Salter ME, *et al.* 2011. Impact of an artificial surfactant release on air-sea gas fluxes during deep ocean gas exchange experiment II. *J Geophys Res Oceans.* 116(C11):C11016.

28 Yang M, *et al.* 2021. Natural variability in air–sea gas transfer efficiency of CO_2. *Sci Rep.* 11(1):1–9.

29 McGillis WR, *et al.* 2004. Air-sea CO_2 exchange in the equatorial pacific. *J Geophys Res Ocean.* 109(C8):C08S02.

30 Lundberg SM, Lee SI. 2017. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 30.

31 Czerski H, *et al.* 2022. Ocean bubbles under high wind conditions–Part 1: Bubble distribution and development. *Ocean Sci.* 18(3): 565–586.

32 Scanlon B, Ward B. 2016. The influence of environmental parameters on active and maturing oceanic whitecaps. *J Geophys Res Oceans.* 121(5):3325–3336.

33 Callaghan A, de Leeuw G, Cohen L, O'Dowd CD. 2008. Relationship of oceanic whitecap coverage to wind speed and wind history. *Geophys Res Lett.* 35(23):L23609. https://doi.org/10.1029/2008GL036165 [dataset].

34 Hanson JL, Phillips OM. 1999. Wind sea growth and dissipation in the open ocean. *J Phys Oceanogr.* 29(8):1633–1648.

35 Yang M, Norris SJ, Bell TG, Brooks IM. 2019. Sea spray fluxes from the southwest coast of the United Kingdom-dependence on wind speed and wave height. *Atmos Chem Phys.* 19(24):15271–15284.

36 Edson JB, *et al.* 2014. On the exchange of momentum over the open ocean. *J Phys Oceanogr.* 44(9):1589.

37 Zhou X, Reichl BG, Romero L, Deike L. 2023. A sea state dependent gas transfer velocity for CO_2 unifying theory, model, and field data. *Earth Space Sci.* 10(11):e2023EA003237.

38 Mueller JA, Veron F. 2009. Nonlinear formulation of the bulk surface stress over breaking waves: feedback mechanisms from airflow separation. *Boundary Layer Meteorol.* 130:117–134.

39 Brumer SE, *et al.* 2017. Whitecap coverage dependence on wind and wave statistics as observed during SO GasEx and HiWinGS. *J Phys Oceanogr.* 47(9):2211–2235.

40 Phillips OM. 1985. Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *J Fluid Mech.* 156: 505–531.

41 Hwang PA, *et al.* 2019. Effects of short scale roughness and wave breaking efficiency on sea spray aerosol production: multisensor field observations. arXiv, arXiv:1906.11201, preprint: not peer reviewed.

42 Woolf DK, *et al.* 2007. Modelling of bubble-mediated gas transfer: fundamental principles and a laboratory test. *J Mar Syst.* 66(1-4): 71–91.

43 Miller SD, *et al.* 2024. Field evaluation of an autonomous, low-power eddy covariance CO_2 flux system for the marine environment. *J Atmos Ocean Technol.* 41(3):279–293.