Surface turbulent fluxes from the MOSAiC campaign predicted by machine learning

Donald P. Cummins[1], Virginie Guemas[2], Christopher J. Cox[3], Michael R Gallagher[4], and Matthew D. Shupe[5]

[1]CNRM, Université de Toulouse, Météo-France, CNRS
[2]Barcelona Supercomputing Center
[3]NOAA-PSL
[4]CIRES/NOAA-ESRL
[5]University of Colorado Boulder

October 27, 2023

Surface turbulent fluxes from the MOSAiC campaign predicted by machine learning

Donald P. Cummins[1], Virginie Guemas[1], Christopher J. Cox[2], Michael R. Gallagher[2,3], Matthew D. Shupe[2,3]

[1]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[2]National Oceanic and Atmospheric Administration, Physical Sciences Laboratory, Boulder, CO, USA
[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

Key Points:

- Neural networks trained on previous Arctic campaigns predict surface turbulent fluxes from MOSAiC more accurately than bulk methods.
- Updated parametrizations using the MOSAiC data have been developed and implemented in Fortran for deployment in weather/climate models.
- Modest performance gains (up to +7% R^2) from recalibration on MOSAiC indicate good generalizability to the pan-Arctic sea ice domain.

Corresponding author: Donald P. Cummins, `donald.cummins@meteo.fr`

Abstract

Reliable boundary-layer turbulence parametrizations for polar conditions are needed to reduce uncertainty in projections of Arctic sea ice melting rate and its potential global repercussions. Surface turbulent fluxes of sensible and latent heat are typically represented in weather/climate models using bulk formulae based on the Monin-Obukhov Similarity Theory (MOST), sometimes finely tuned to high stability conditions and the potential presence of sea ice. In this study, we test the performance of new, machine-learning (ML) flux parametrizations, using an advanced polar-specific bulk algorithm as a baseline. Neural networks, trained on observations from previous Arctic campaigns, are used to predict surface turbulent fluxes measured over sea ice as part of the recent MOSAiC expedition. The ML parametrizations outperform the bulk at the MOSAiC sites, with RMSE reductions of up to 70 percent. We provide a plug-in Fortran implementation of the neural networks for use in models.

Plain Language Summary

Heat can make its way into or out of sea ice via unpredictable air movements, known as turbulence, near the sea surface. In order to predict how quickly Arctic sea ice will melt in the future, we need to know how much heat the turbulence can transport in different weather conditions. Traditionally, turbulence calculations have been performed using sophisticated mathematical formulae from physics. In this study, we test an alternative method for predicting turbulent heat exchange: a computer algorithm known as an artificial neural network. By showing turbulence data, measured in the Arctic during previous scientific expeditions, to the network, it can be "trained" to make predictions in a process known as machine learning. We compare turbulence measurements, taken above sea ice in the recent MOSAiC expedition, with predictions from trained neural networks. We find that the neural networks are better than the traditional physics at predicting what the scientists at MOSAiC observed. The trained neural networks have been made publicly available so that they can be used by scientists for predicting climate change.

1 Introduction

The polar regions, in particular the Arctic, are on the front line of the climate crisis. In recent decades, the rate of surface warming in the Arctic has been two to four times higher than the global mean (Rantanen et al., 2022), a phenomenon known as Arctic amplification (e.g., Serreze & Francis, 2006; Graversen et al., 2008; Serreze & Barry, 2011). Alongside rising temperatures have occurred losses of around 50 percent in both thickness and extent of Arctic sea ice at the end of summer since satellite records began (Gascard et al., 2019). The rate of Arctic sea ice loss in the coming decades remains highly uncertain (Bonan, Lehner, & Holland, 2021; Bonan, Schneider, et al., 2021), however the consequences are expected to be severe: for local ecosystems (Kovacs et al., 2011; Post et al., 2013; Tynan, 2015); for indigenous peoples (Meier et al., 2014); and, potentially, for lower-latitude climate (Cohen et al., 2014; Jung et al., 2015; Cohen et al., 2020; Liu et al., 2022). Heat exchanges between sea ice and the atmosphere are a key driver of the Arctic amplification (e.g., Serreze et al., 2009; Lesins et al., 2012; Previdi et al., 2021) and determine the sea ice melting rate (e.g., Rothrock et al., 1999; Screen & Simmonds, 2010).

Turbulent exchanges of heat and momentum in the planetary boundary layer are not directly simulated in weather/climate models, but are instead represented through parametrizations, typically bulk formulae based on the Monin-Obukhov Similarity Theory (MOST, Monin & Obukhov, 1954; Garratt, 1994). Such parametrizations are semi-empirical: although the MOST provides dimensionless relationships, their final forms cannot be determined without recourse to observational data (e.g., calibration of roughness models and stability functions). The polar boundary layer is influenced by the presence of sea ice and is characterized by high stability and often intermittent turbulence (e.g., Andreas, 1998). Polar-specific stability functions have been proposed (Grachev et al., 2007), as well as formulations of surface roughness (e.g., Andreas, 1987; Andreas, Persson, et al., 2010; Andreas, 2011). More recently, parametrizations have been developed that account for form drag arising from alternating sea ice floes and leads (e.g., Lüpkes et al., 2012; Lüpkes & Gryanik, 2015; Elvidge et al., 2016). Use of polar-specific turbulence parametrizations has been found to reduce biases in atmospheric models (Renfrew et al., 2019; Elvidge et al., 2023), however adoption of these advanced parametrizations in climate models has until recently been limited. The historic scarcity of observations in the Arctic likely goes some way to explaining modelers' caution, yet there are also longstanding unresolved problems with modeling even homogeneous stable boundary layers (e.g., the GABLS experiments, Cuxart et al., 2006; Svensson et al., 2011; Bosveld et al., 2014).

Outside the polar regions, where observations have historically been more readily available, machine learning (ML) has emerged in recent years as an alternative strategy for parametrizing boundary-layer processes (Pal & Sharma, 2021). The basic idea of the ML or *data-driven* approach is that, given sufficient observational data, statistical algorithms can be used to directly infer empirical relationships between quantities of interest, such as surface turbulent fluxes, and mean meteorological variables such as temperature, humidity, etc. Recent studies have found that ML parametrizations, based on artifical neural networks (ANNs), can predict surface turbulent fluxes measured at meteorological towers in extra-polar regions with greater accuracy than bulk algorithms based on the MOST (Leufen & Schädler, 2019; McCandless et al., 2022; Wulfmeyer et al., 2022). These findings were extended to the Arctic by Cummins et al. (2023), hereafter C23, who showed that, even with the relatively small volume of data collected in previous Arctic campaigns, it is nevertheless possible to train ANNs that can outperform a polar-specific bulk algorithm.

The present study is motivated by the recent publication of surface turbulent flux observations collected at the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC, Shupe et al., 2022). The MOSAiC dataset provides a unique opportunity to test the hypothesis, motivated by the encouraging results of C23, that ML parametrizations trained on data with limited spatiotemporal scope in the data-sparse Arctic are broadly applicable to the pan-Arctic sea ice domain. In this paper, we employ the MOSAiC data first to validate the performance of the ANNs of C23, using a MOST-based bulk algorithm as a baseline. We then incorporate the MOSAiC data into the ANN training set to generate an improved set of flux parametrizations for use in polar conditions (see Code Availability Statement). The remainder of this paper is organized as follows. Section 2 briefly recaps the datasets used in C23 and introduces the new MOSAiC data. Section 3 describes the ML and bulk algorithm flux parametrizations used in this study and the statistical methods used to evaluate their performance. Section 4 presents the results. Conclusions and recommendations for modelers are given in Section 5.

2 Data

2.1 Pre-MOSAiC observational campaigns

C23 trained and validated ANN models using surface turbulent flux measurements from four observational campaigns conducted over Arctic sea ice: Surface Heat Budget of the Arctic Ocean (SHEBA, Andreas et al., 1999; Persson et al., 2002; Uttal et al., 2002); Aerosol-Cloud Coupling and Climate Interactions in the Arctic (ACCACIA, Elvidge et al., 2016); Arctic Cloud in Summer Experiment (ACSE, Sotiropoulou et al., 2016; Prytherch et al., 2017); and Arctic Ocean 2016 (AO16, Tjernström & Jakobsson, 2021; Srivastava et al., 2022). These datasets sample a range of seasons and meteorological conditions in the Arctic. The sea ice varies in concentration (between zero and one), as well as in its morphology. For example, the ice surrounding the year-long SHEBA camp was compact and snow-covered in winter (Andreas, Persson, et al., 2010), but littered with deep melt ponds and leads in summer (Andreas, Horst, et al., 2010). It should be noted that C23 omitted from the training set observations in ACCACIA that were collected at heights > 30 m above the surface. Surface turbulent fluxes in climate models are typically calculated much closer to the surface (e.g., $\sim$ 10 m in CNRM-CM6-1, Voldoire et al., 2019; Roehrig et al., 2020). Satellite estimates of sea ice concentration were obtained from the National Snow and Ice Data Center (NSIDC, Meier et al., 2021).

2.2 MOSAiC

For the MOSAiC expedition, the icebreaker *RV Polarstern* was frozen into the Arctic sea ice and drifted with it for most of a year between Oct 2019 and Oct 2020. The original ice floe, on which the MOSAiC camp was established in Oct 2019, exited into the North Atlantic in late July 2020. Polarstern then repositioned near the North Pole at a new ice floe for August and September 2020. Various scientific research sites were established on the ice surrounding the ship, in a fashion similar to SHEBA although on a larger scale. As part of MOSAiC, extensive measurements were taken of the Arctic atmospheric system (Shupe et al., 2022). Surface turbulent fluxes of momentum, sensible heat and latent heat were computed at multiple locations using eddy-covariance techniques together with high-frequency (sampling rates of 10-20 Hz) observations from ultrasonic anemometers. Eddy covariances were computed over 10-minute sampling periods. Turbulence measurements were made at a meteorological tower with sensors at

2, 6 and 10 m above the initial snow/ice surface. Data from all three tower levels were used in this study. Flux measurements were also taken at 3.8 m at the three Atmospheric Surface Flux Stations (ASFS), analogous to the Portable Automated Mesonet (PAM) stations in SHEBA. ASFS 30/40 were deployed at ~ 13 km from the tower and ASFS50 at ~ 23 km. Note that, due to accumulation and ablation of snow, the actual measurement heights varied over time. MOSAiC data used in this study were subject to Level3.4 quality control (see Data Availability Statement). MOSAiC increased the size of the C23 flux database by a factor of five for momentum and sensible heat and four for latent heat.

3 Methods

Surface turbulent fluxes are typically computed in climate models through bulk algorithms using wind, temperature and humidity at the single model level closest to the surface (e.g., in the SURFEX module in CNRM-CM6-1, Voldoire et al., 2019). The MOST, or a simplified version thereof, may then be used to extrapolate the vertical profiles of meteorological variables in the surface layer (Geleyn, 1988). Flux parametrizations in this study have been developed as plug-in replacements for bulk algorithms and therefore expect similar inputs. For the MOSAiC ASFS data, the wind and temperature/humidity measurements were made at different heights above the snow/ice surface (3.86 and 2.13/1.84 m respectively). While this doesn't preclude a direct application of the bulk approach (since the MOST does not require measurements of those variables at the same height), it means that some pre-processing is required before the ANNs of C23 can be used. The wind speed, measured at a single height, was not interpolated. Instead, the temperature/humidity measurements were linearly extrapolated to 3.86 m, using the observed gradient between the surface and the measurement height. Surface specific humidity was computed from temperature and pressure using the *meteolib* Python library (see Code Availability Statement). More sophisticated alternatives include a logarithmic extrapolation, or one based on the full MOST. However, our own numerical tests, conducted using equivalent measurements at 2 and 6 m on the meteorological tower, found the linear extrapolation to outperform the logarithmic in a root-mean-square error (RMSE) sense. Using the MOST approach would naturally introduce a bias in favour of that methodology. Different sensors were also mounted at slightly different heights around the nominal height of each tower level. Taking the heights of different sensors as the measurement height was found to have a small impact on the accuracy of flux predictions ($\pm 10\%$ RMSE). In the final analysis, it was decided to use the nominal heights of the tower levels, corrected for snow thickness, which is consistent with how C23 treated data from the meteorological tower of the SHEBA campaign.

C23 developed ML flux parametrizations based on single-layer, feed-forward ANNs with four nodes in the hidden layer. For a high-level introduction to statistical modeling with neural networks, see Hastie et al. (2009) or Kuhn and Johnson (2013). These models are general-purpose non-linear functions (Hornik et al., 1989), permitting a high degree of variable interaction, and containing 37 tuneable parameters. Each ANN takes seven mean meteorological variables as inputs: the measurement height z; absolute horizontal windspeed $u(z)$; potential temperatures $\theta(z)$, θ_s of the air and at the surface respectively; specific humidities $q(z)$, q_s; and the sea ice concentration C_i, determined over a $25{\times}25$ km^2 domain. The relative importance of the different inputs to the bulk and ANN methods was explored by C23, who found that the ANNs depend less critically on the vertical gradients. The models were trained on the pre-MOSAiC data using the *nnet* library for the statistical programming language R (Venables & Ripley, 2002; R Core Team, 2021). A weight decay of $\lambda = 0.01$ was used for regularization and the networks were fitted in ensembles of 100 models to reduce variability due to random parameter initialization (Ripley, 1996). The fitted ANNs output turbulent fluxes of momentum $u_\star^2$, sensible heat $u_\star\theta_\star$ and latent heat $u_\star q_\star$. Predicted fluxes are returned in kinematic units, i.e. in the same units as the measured eddy covariances, and hence are written here in terms of the MOST scaling parameters $u_\star$, $\theta_\star$, $q_\star$.

The polar-specific bulk algorithm, used in this study as a baseline against which to compare the ANNs, is the same as that described in C23. Over open water, the iterative COARE 3.0 algorithm is used (Fairall et al., 2003; Edson et al., 2013), with stability functions from Grachev et al. (2000) in unstable conditions and from Beljaars and Holtslag (1991) in stable conditions. The COARE 3.0 algorithm has been well tested over the years and is currently in use in large-scale climate models, including CNRM-CM6-1. Bulk transfer coefficients are initialized using a non-iterative estimate of the stability (Grachev & Fairall, 1997). Over sea ice, the stability function from Grachev et al. (2007) is used in stable conditions, as well as the scalar roughness model of Andreas (1987) and the aerodynamic roughness model of Andreas, Persson, et al. (2010). For partial sea ice concentrations, we use the *mosaic* approach (e.g., Vihma, 1995), whereby we take a weighted average of fluxes computed over open water and over sea ice, with the weighting given by the sea ice concentration. An additional form drag contribution is included when computing the momentum flux, to account for the influence of intermittent sea ice coverage (Lüpkes & Gryanik, 2015). Intermittent ice coverage is associated with vertical ice surfaces that tend to increase turbulence. This bulk algorithm is available for download as a Python library (see Code Availability Statement). Compared against estimates from unmodified COARE 3.0, momentum flux estimates from our bulk algorithm have lower RMSE at the MOSAiC sites (up to a 16% reduction). The polar-specific components have less impact on the heat fluxes: there is a 99% correlation between our heat fluxes and those from COARE 3.0. The results of our comparison with ML in Section 4 are robust to the use / non-use of polar-specific components in the bulk algorithm.

In C23, the ML and bulk algorithm flux parametrizations were tested using a campaign-wise cross-validation scheme. Each campaign (or measurement site in the case of SHEBA) was left out of the training set in turn and the trained models validated on that campaign. Flux predictions from the two methods, together with measured eddy covariances, were used to compute performance metrics, such as RMSE, mean absolute error (MAE) and Pearson correlation. Since the MOSAiC data were not involved in the calibration of either parametrization, they constitute an independent test set and are therefore ideal for model validation and comparison. Mean meteorological variables, measured at each of the MOSAiC sites, were supplied as input variables and predicted fluxes calculated. In addition to these truly out-of-sample predictions, further flux estimates were obtained from ANNs fitted to MOSAiC-augmented training sets: for each site in MOSAiC, an ANN model was fitted to a training set comprising the pre-MOSAiC data plus all MOSAiC data *not* observed at that site. Iterating over the MOSAiC sites then gives a complete set of out-of-sample predictions, which allows us to quantify any gains in predictive power obtained from the MOSAiC data.

4 Results

Performance metrics, computed for the bulk algorithm and ANN parametrizations at each of the MOSAiC sites, are given in Table 1. Figures 1-3 show two-dimensional histograms of predicted fluxes against measured eddy covariances at each site. Note that results at the meteorological tower do not differ qualitatively between tower levels in terms of patterns/biases, however there is a small dependence of predictive accuracy on measurement height. Specifically, both the bulk algorithm and neural network methods have

Table 1. Predictive performance of neural network (nnet/nnet+), and Monin-Obukhov (bulk) flux parametrizations at the MOSAiC sites in kinematic units. The nnet+ columns show results for ANNs trained using MOSAiC-augmented data. Boldface indicates a better score in one of root-mean-square error (RMSE), mean absolute error (MAE) or Pearson correlation. Using bootstrapping, all score differences were found to be statistically significant at the five-percent level (Davison & Hinkley, 1997). Note that direct measurements of $u_\star q_\star$ are not available at ASFS40.

			RMSE			MAE			correlation		
	site	samples	bulk	nnet	nnet+	bulk	nnet	nnet+	bulk	nnet	nnet+
$u_\star^2$	ASFS30	22161	0.039	**0.035**	**0.033**	0.021	**0.018**	**0.019**	0.92	**0.92**	**0.92**
	ASFS40	18498	0.035	**0.034**	**0.029**	0.020	**0.019**	**0.017**	**0.94**	0.93	0.93
	ASFS50	18871	0.031	**0.027**	**0.025**	0.015	**0.013**	**0.014**	0.93	**0.94**	**0.93**
	met. tower	66777	0.049	**0.047**	**0.044**	0.025	**0.023**	**0.023**	**0.90**	0.89	0.89
$u_\star\theta_\star$	ASFS30	22161	0.0091	**0.0057**	**0.0045**	0.0066	**0.0043**	**0.0033**	0.80	**0.81**	**0.85**
	ASFS40	18498	0.0113	**0.0071**	**0.0058**	0.0089	**0.0052**	**0.0041**	0.70	**0.74**	**0.78**
	ASFS50	18871	0.0056	**0.0045**	**0.0049**	0.0041	**0.0033**	**0.0036**	0.77	**0.81**	**0.80**
	met. tower	66777	0.0073	**0.0066**	**0.0062**	0.0046	**0.0040**	**0.0038**	0.69	**0.71**	**0.73**
$u_\star q_\star$	ASFS30	2692	3.20E-06	**8.43E-07**	**1.06E-06**	2.21E-06	**5.83E-07**	**7.85E-07**	**0.75**	0.63	0.63
	ASFS50	3436	9.88E-07	**8.93E-07**	**8.50E-07**	6.55E-07	**4.96E-07**	**4.63E-07**	**0.79**	0.67	0.70
	met. tower	33185	**5.30E-07**	5.94E-07	5.87E-07	**2.97E-07**	3.53E-07	3.43E-07	**0.67**	0.48	0.50

slightly lower RMSE ($\sim 10\%$) when applied at 10 m compared with 2 m. This is as expected: the 10-m differences of the meteorological variables are larger than the corresponding 2 m differences, so if the measurement errors at the different levels are similar in magnitude then the 10-m differences should have lower relative error. Any inaccuracies in the estimated measurement heights should also be proportionally smaller at 10 m. Overall, the results are encouraging, with the ANN parametrizations consistently delivering performance improvements over the bulk algorithm, particularly in the stable conditions which predominate in MOSAiC.

Both methods produce similar estimates of the momentum flux $u_\star^2$ and the two-dimensional histograms in Figure 1 share common features, such as a conservative bias (systematic underprediction of larger fluxes). However, the ANNs achieve a lower RMSE at all the MOSAiC sites: a result which is robust under bootstrap resampling (Davison & Hinkley, 1997). The conservative bias of the ANNs was noted by C23 and is a known property of the models. In short, the ANNs have an inbuilt reluctance to extrapolate when faced with a combination of inputs not seen in training. That the bulk algorithm also underpredicts $u_\star^2$ is unexpected and warrants investigation (see final paragraph of this section). Augmenting the ANN training set with data from MOSAiC reduces the RMSE of the ANNs at all sites and produces a visible attenuation of the conservative bias for larger fluxes. This result indicates that the conditions conducive to large $u_\star^2$ were consistent across the MOSAiC sites.

The ANN parametrization outperforms the bulk algorithm as an estimator of the sensible heat flux $u_\star\theta_\star$, with RMSE 10-40 percent lower across the sites. It can be seen from Figure 2 that the improvements over the bulk are particularly apparent at the ASFS30 and ASFS40 stations. As was the case for $u_\star^2$, the prediction errors of the two $u_\star\theta_\star$ parametrizations share common features, including some clearly non-random deviations from the line $y = x$. In particular, there is a long tail in the panels for the met. tower in Figure 2, comprised of large upwards fluxes (negative eddy covariance) whose magnitude was underestimated by both the bulk and ANN parametrizations. These fluxes occurred in near-neutral conditions, defined by Högström (1988) as $|\zeta| < 0.1$ where ζ is the Obukhov stability parameter. There was little temperature gradient in the surface layer and it is possible that non-local turbulence played a role. Large prediction errors also occurred when there was a strong surface-layer gradient but observed fluxes were small, again in near-neutral conditions. Including MOSAiC data from other sites in the ANN training

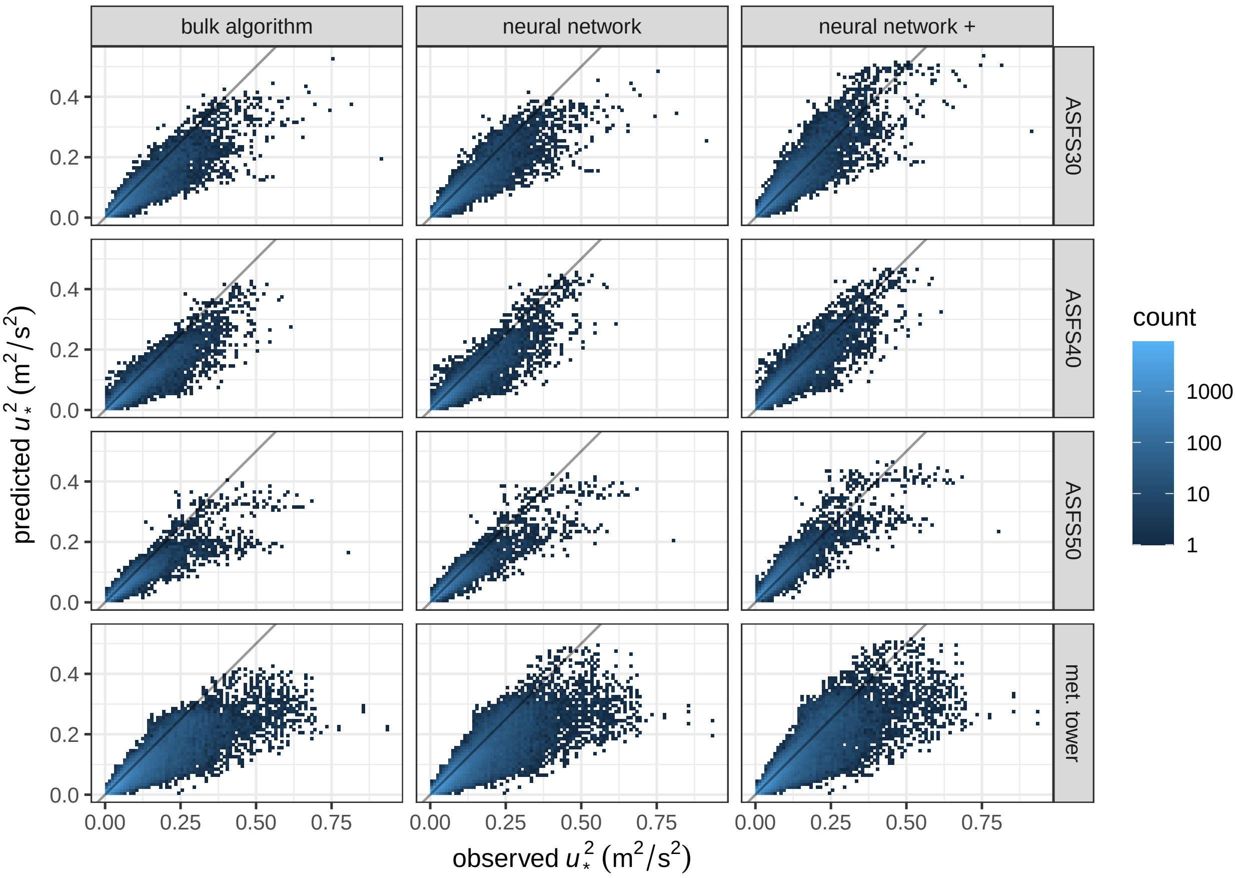

Figure 1. Predicted momentum fluxes $u_\star^2$ at the MOSAiC sites plotted against observed eddy covariances. To the left are estimates obtained from a polar-specific bulk algorithm based on the Monin-Obukhov Similarity Theory; in the centre, estimates from the neural networks of Cummins et al. (2023); to the right, estimates from neural networks trained using MOSAiC-augmented data. The diagonal line $y = x$ would represent a perfect fit.

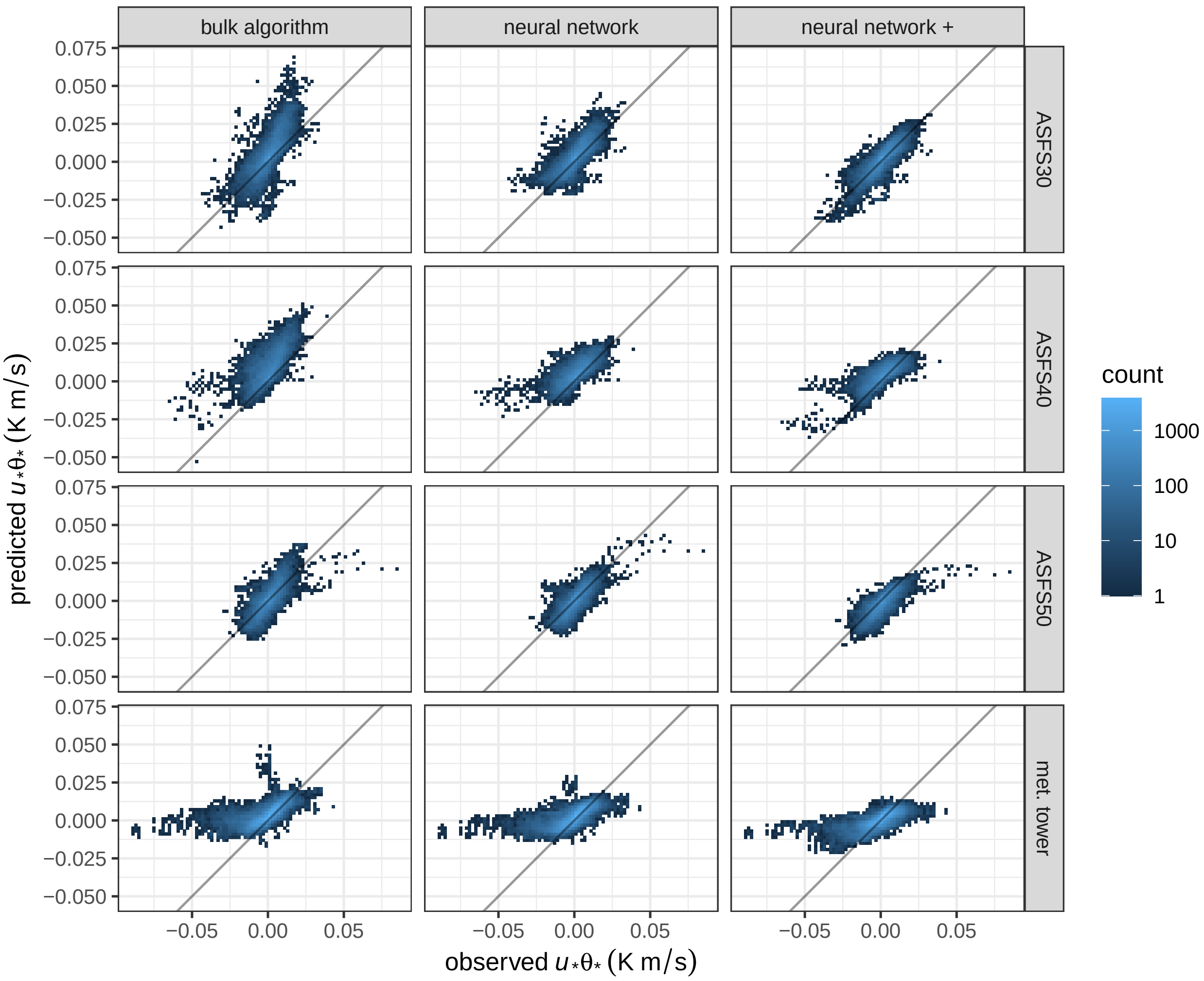

Figure 2. Predicted sensible heat fluxes $u_\star\theta_\star$ at the MOSAiC sites plotted against observed eddy covariances. See Figure 1 caption for details.

set produces clear improvements at three of the four sites, with several systematic features in the residuals disappearing. The predictions at the ASFS50 site, however, became worse. Prediction errors at ASFS50 are the lowest for $u_\star^2$ and $u_\star\theta_\star$, so it doesn't necessarily follow that the site is at fault. There may be locally varying factors, not included in the set of input variables, which affect the fluxes (see final paragraph of this section). Identifying such variables has the potential to deliver further performance gains, especially in near-neutral conditions where performance of both algorithms is worse.

Latent heat flux $u_\star q_\star$ is by far the most difficult of the three fluxes to predict: $u_\star q_\star$ was generally small in magnitude at the MOSAiC measurement sites, suggesting a low signal-to-noise ratio. The ANNs are also disadvantaged here by a small training set from previous campaigns, comprised mainly of very small fluxes (C23). Results for $u_\star q_\star$ are therefore unsurprising: to the extent that the measured fluxes are small in magnitude, the ANNs perform well. For larger fluxes, the ANNs exhibit a strong conservative bias. Conversely, the bulk algorithm tends to overpredict the magnitude of $u_\star q_\star$. It is because of these contrasting biases that the bulk achieves a higher correlation at the ASFS30 and ASFS50 stations, while at the same time the ANNs give RMSE reductions at those sites of about 70 and 10 percent respectively. At the MOSAiC tower, the bulk algorithm performs better, achieving a 10-percent lower RMSE. The ANNs trained on the MOSAiC-augmented data perform better at the ASFS50 site and the tower, but slightly worse at ASFS30. From Figure 3 it can be seen that the underestimation bias, while still present, is improved by training with MOSAiC data.

It should be noted that the biases, visible in the MOSAiC-augmented ANN predictions in Figures 1-3, should be further reduced by the next step, which is to incorporate all the MOSAiC data in the ANN training set. As more observations become available, we would expect the as-yet-unsampled regions of the input space to diminish, along with the associated biases. That is not to say that all biases can be resolved through more training data. As mentioned above, omission of important predictor variables has the potential to induce biases that will persist regardless of the volume of training data. For example, the upwards and downwards radiation terms are known to contribute signficant explanatory power (e.g., Wulfmeyer et al., 2022). These radiation terms are nevertheless unsuitable for use as parametrization inputs, because the radiative fluxes in GCMs are themselves the output of complex parametrizations with their own errors and uncertainties. The surface characteristics may also be an important missing variable. In

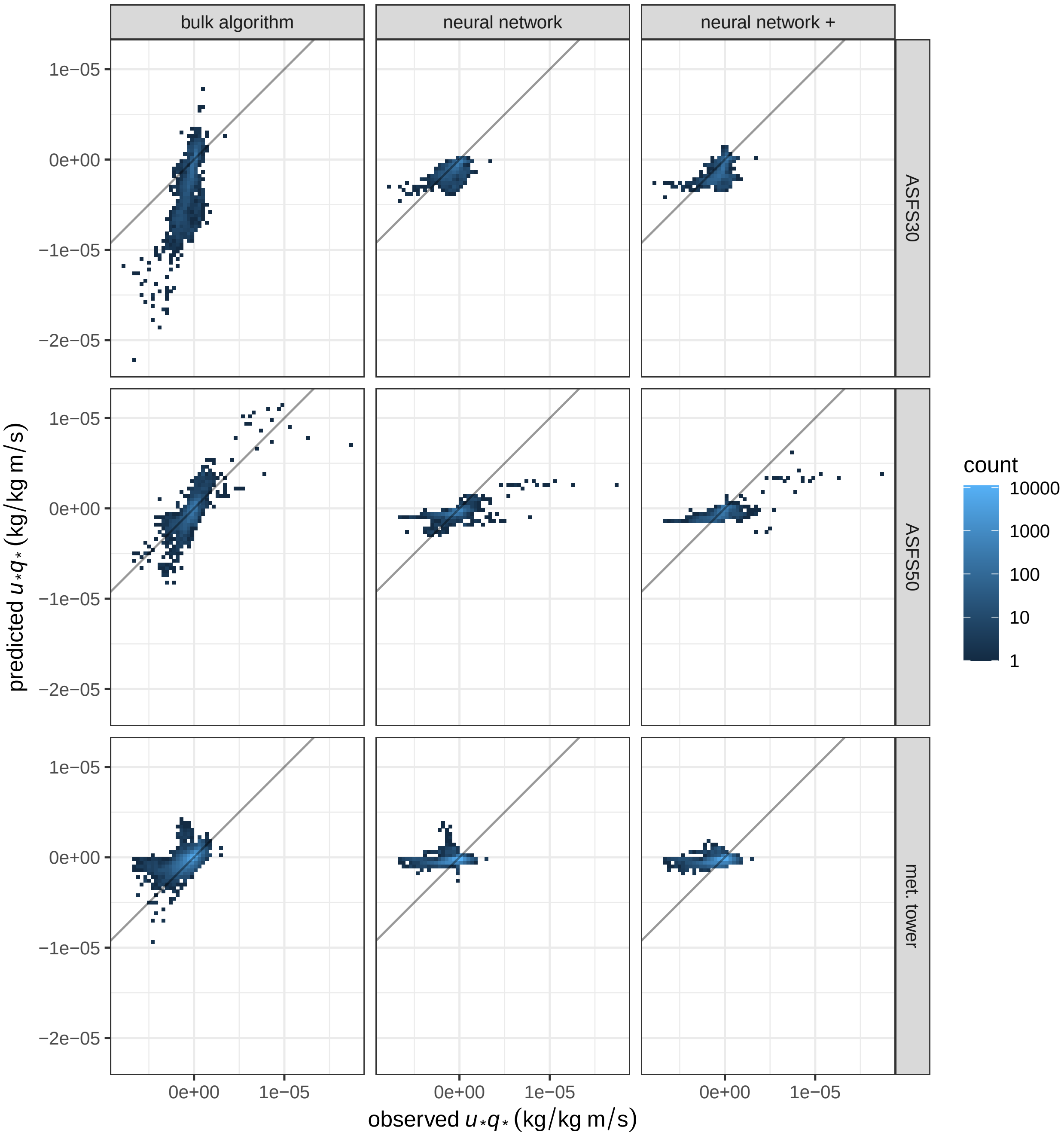

Figure 3. Predicted latent heat fluxes $u_\star q_\star$ at the MOSAiC sites plotted against observed eddy covariances. See Figure 1 caption for details.

MOSAiC, the winter sea ice may have been generally aerodynamically rougher than that seen in the SHEBA campaign. This could potentially explain the underestimation of $u_\star^2$ by both the ANN and bulk algorithm parametrizations (see Figure 1). Finally, turbulent heat fluxes over sea ice are known to be influenced by lead width (e.g., Marcq & Weiss, 2012), although the exact nature of the dependency is a topic of ongoing research (Gryschka et al., 2023). Including the lead-width distribution as an input to the ANNs could shed additional light on this question.

5 Conclusions

Accurate representation in climate models of turbulent heat exchanges between the surface and atmosphere in polar regions is essential for constraining predictions of future climate change, locally and potentially globally. Surface turbulent fluxes in the polar boundary layer are currently parametrized using the traditional MOST, although alternative ML parametrizations based on ANNs have recently been proposed (C23). The wealth of new flux observations collected in the Arctic during the MOSAiC campaign has provided an excellent opportunity to validate and calibrate these alternative parametrizations.

In this study, the MOSAiC data have been used to validate ANN parametrizations of momentum, sensible heat and latent heat fluxes, that were originally trained on data from previous Arctic campaigns. The ANNs have been found to generalize well to the new data, particularly for momentum and sensible heat, yielding substantial reductions in error metrics such as RMSE when compared against a polar-specific bulk algorithm based on the MOST. Although the ANNs performed well at predicting small latent heat fluxes, limitations of the training data resulted in systematic underprediction of larger fluxes.

The ANN parametrizations, developed in C23 and validated in this study, have been recalibrated on an augmented training dataset that incorporates the observations from MOSAiC. The largest increase in variance explained (R^2) after recalibration was only seven percent, despite the training set growing by a factor of 4-5, indicating that a high level of generalizability has already been achieved. These updated parametrizations have been implemented as a Fortran subroutine, suitable for deployment in weather/climate models as a plug-in replacement for bulk algorithms (see Code Availability Statement).

An important next step will be to perform sensitivity studies with these new parametrizations in a climate model. In this way, the implications for the polar atmosphere and melting of Arctic sea ice can be assessed.

Open Research

Code Availability Statement

The following publicly available software tools can be used to reproduce the results presented in this study:

- The Python library *CDlib* (Guemas, 2023a) provides functions to compute transfer coefficients and related variables (zeta, stability functions, aerodynamic and scalar roughness etc.), as well as to apply the bulk algorithm parametrizations used in this study.
- The Python library *meteolib* (Guemas, 2023b) provides functions to estimate meteorological parameters (humidity, latent heat as a function of temperature etc.).
- The Fortran subroutine *PolarFlux* (Cummins, 2023) implements the neural network flux parametrizations developed in this study. The networks have been trained on all available datasets, including MOSAiC.

Neural network ensembles were fitted using the R package *caret* (Kuhn & Johnson, 2013), which itself depends on the R package *nnet* (Venables & Ripley, 2002) to fit the underlying models. Bootstrapping of model performance metrics was performed using the R package *boot* (Canty & Ripley, 2022).

Data Availability Statement

MOSAiC campaign sites

The MOSAiC data used in this study are available from the National Science Foundation Arctic Data Center: met. tower (Cox, Gallagher, Shupe, Persson, Blomquist, et al., 2023); ASFS30 (Cox, Gallagher, Shupe, Persson, Grachev, et al., 2023a); ASFS40 (Cox, Gallagher, Shupe, Persson, Grachev, et al., 2023b), ASFS50 (Cox, Gallagher, Shupe, Persson, Grachev, et al., 2023c).

Pre-MOSAiC observational campaigns

The SHEBA data are available from the NCAR Earth Observing Laboratory: met. tower (Andreas et al., 2007); PAM stations (Andreas et al., 2012). The ACCACIA flight data are available from the CEDA archive: MASIN (British Antarctic Survey (BAS), 2014); FAAM (Facility for Airborne Atmospheric Measurements et al., 2015). The ACSE cruise data are available from the CEDA archive (Brooks et al., 2022a). The AO16 cruise data are available from the CEDA archive (Brooks et al., 2022b). The NSIDC sea ice concentration data are available from the NSIDC archive (Meier et al., 2021).

Acknowledgments

This work was supported by a national funding by the Agence Nationale de la Recherche within the framework of the Investissement d'Avenir program under the ANR-17-MPGA-0003 reference. This article has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003826 via project CRiceS (Climate Relevant interactions and feedbacks: the key role of sea ice and Snow in the polar and global climate system).

Data used in this manuscript were produced as part of the international Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) with the tag MOSAiC20192020. We thank all persons involved in the expedition of the *Research Vessel Polarstern* during MOSAiC in 2019-2020 (AWI_PS122_00) as listed in Nixdorf et al. (2021). The MOSAiC observations received support from the US National Science Foundation Office of Polar Programs (OPP-1724551); the NOAA Physical Sciences Laboratory; and NOAA's Global Ocean Monitoring and Observing Program (FundRef `https://doi.org/10.13039/100018302`).

This is a contribution to the Year of Polar Prediction (YOPP), a flagship activity of the Polar Prediction Project (PPP), initiated by the World Weather Research Programme (WWRP) of the World Meteorological Organisation (WMO). We acknowledge the WMO WWRP for its role in coordinating this international research activity.

We thank the two anonymous reviewers for their careful and constructive comments, which helped us to improve the manuscript.

References

Andreas, E. L. (1987, January). A theory for the scalar roughness and the scalar transfer coefficients over snow and sea ice. *Boundary-Layer Meteorology*, *38*(1), 159–184. doi: 10.1007/BF00121562

Andreas, E. L. (1998). The Atmospheric Boundary Layer Over Polar Marine Surfaces. In M. Leppäranta (Ed.), *Physics of Ice-Covered Seas* (Vol. 2, pp. 715–773). Helsinki: Helsinki University Press.

Andreas, E. L. (2011). A relationship between the aerodynamic and physical roughness of winter sea ice. *Quarterly Journal of the Royal Meteorological Society*, *137*(659), 1581–1588. doi: 10.1002/qj.842

Andreas, E. L., Fairall, C., Guest, P., & Persson, O. (2007). *Tower, 5-level hourly measurements plus radiometer and surface data at Met City (ASFG). Version 1.0* [Dataset]. UCAR/NCAR - Earth Observing Laboratory. doi: 10.5065/D65H7DNS

Andreas, E. L., Fairall, C., Guest, P., & Persson, O. (2012). *Ice Camp Surface Mesonet NCAR PAM-III 1 hour (FINAL). Version 1.0* [Dataset]. UCAR/NCAR - Earth Observing Laboratory. doi: 10.5065/D6ZC8170

Andreas, E. L., Fairall, C. W., Guest, P. S., & Persson, P. O. G. (1999). An overview of the SHEBA atmospheric surface flux program. In *13th symposium on boundary layers and turbulence. Dallas, TX, Amer. Meteorol. Soc., Proc.* (pp. 550–555).

Andreas, E. L., Horst, T. W., Grachev, A. A., Persson, P. O. G., Fairall, C. W., Guest, P. S., & Jordan, R. E. (2010). Parametrizing turbulent exchange over summer sea ice and the marginal ice zone. *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 927–943. doi: 10.1002/qj.618

Andreas, E. L., Persson, P. O. G., Grachev, A. A., Jordan, R. E., Horst, T. W., Guest, P. S., & Fairall, C. W. (2010, February). Parameterizing Turbulent Exchange over Sea Ice in Winter. *Journal of Hydrometeorology*, *11*(1), 87–104. doi: 10.1175/2009JHM1102.1

Beljaars, A. C. M., & Holtslag, A. a. M. (1991, March). Flux Parameterization over Land Surfaces for Atmospheric Models. *Journal of Applied Meteorology and Climatology*, *30*(3), 327–341. doi: 10.1175/1520-0450(1991)030⟨0327:FPOLSF⟩2.0.CO;2

Bonan, D. B., Lehner, F., & Holland, M. M. (2021, March). Partitioning uncertainty in projections of Arctic sea ice. *Environmental Research Letters*, *16*(4), 044002. doi: 10.1088/1748-9326/abe0ec

Bonan, D. B., Schneider, T., Eisenman, I., & Wills, R. C. J. (2021). Constraining the Date of a Seasonally Ice-Free Arctic Using a Simple Model. *Geophysical Research Letters*, *48*(18), e2021GL094309. doi: 10.1029/2021GL094309

Bosveld, F. C., Baas, P., Steeneveld, G.-J., Holtslag, A. A. M., Angevine, W. M., Bazile, E., ... Svensson, G. (2014, August). The Third GABLS Intercomparison Case for Evaluation Studies of Boundary-Layer Models. Part B: Results and Process Understanding. *Boundary-Layer Meteorology*, *152*(2), 157–187. doi: 10.1007/s10546-014-9919-1

British Antarctic Survey (BAS). (2014). *British Antarctic Survey Twin Otter aircraft Meteorological Airborne Science INstrumentation (MASIN) core data for the Aerosol Cloud Coupling and Climate Interactions in the Arctic (ACCACIA) project.* [Dataset]. NCAS British Atmospheric Data Centre (NCAS BADC). doi: 10.5285/0844186DB1BA9E20319A2560F8D61651

Brooks, I. M., Prytherch, J., & Srivastava, P. (2022a). *CANDIFLOS : Surface fluxes from ACSE measurement campaign on icebreaker Oden, 2014* [Dataset]. NERC EDS Centre for Environmental Data Analysis. doi: 10.5285/C6F1B1FF16F8407386E2D643BC5B916A

Brooks, I. M., Prytherch, J., & Srivastava, P. (2022b). *CANDIFLOS : Surface fluxes from AO2016 measurement campaign on icebreaker Oden, 2016* [Dataset]. NERC EDS Centre for Environmental Data Analysis. doi: 10.5285/614752D35DC147A598D5421443FB50E8

Canty, A., & Ripley, B. D. (2022, November). *Boot: Bootstrap Functions (Originally by Angelo Canty for S).*

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., ... Jones, J. (2014, September). Recent Arctic amplification and extreme midlatitude weather. *Nature Geoscience*, *7*(9), 627–637. doi: 10.1038/ngeo2234

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., ... Yoon, J. (2020, January). Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nature Climate Change*, *10*(1), 20–29. doi: 10.1038/s41558-019-0662-y

Cox, C., Gallagher, M., Shupe, M., Persson, O., Blomquist, B., Grachev, A., ... Uttal, T. (2023). *Met City meteorological and surface flux measurements (Level 3 Final), Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC), central Arctic, October 2019 - September 2020.* [Dataset]. NSF Arctic Data Center. doi: 10.18739/A2PV6B83F

Cox, C., Gallagher, M., Shupe, M., Persson, O., Grachev, A., Solomon, A., ... Uttal, T. (2023a). *Atmospheric Surface Flux Station #30 measurements (Level 3 Final), Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC), central Arctic, October 2019 - September 2020.* [Dataset]. NSF Arctic Data Center. doi: 10.18739/A2FF3M18K

Cox, C., Gallagher, M., Shupe, M., Persson, O., Grachev, A., Solomon, A., ... Uttal, T. (2023b). *Atmospheric Surface Flux Station #40 measurements (Level 3 Final), Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC), central Arctic, October 2019 - September 2020.* [Dataset]. NSF Arctic Data Center. doi: 10.18739/A25X25F0P

Cox, C., Gallagher, M., Shupe, M., Persson, O., Grachev, A., Solomon, A., ... Uttal, T. (2023c). *Atmospheric Surface Flux Station #50 measurements (Level 3 Final), Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC), central Arctic, October 2019 - September 2020.* [Dataset]. NSF Arctic Data Center. doi: 10.18739/A2XD0R00S

Cummins, D. P. (2023, August). *donaldcummins/PolarFlux: v0.1 (v0.1)* [Software]. Zenodo. doi: 10.5281/ZENODO.8207288

Cummins, D. P., Guemas, V., Blein, S., Brooks, I. M., Renfrew, I. A., Elvidge, A. D., & Prytherch, J. (2023, March). *Reducing parametrization errors for polar surface turbulent fluxes using machine learning* (Preprint). In Review. doi: 10.21203/rs.3.rs-2592358/v1

Cuxart, J., Holtslag, A. A. M., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., ... Xu, K.-M. (2006, February). Single-Column Model Intercomparison for a Stably Stratified Atmospheric Boundary Layer. *Boundary-Layer Meteorology*, *118*(2), 273–303. doi: 10.1007/s10546-005-3780-1

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511802843

Edson, J. B., Jampana, V., Weller, R. A., Bigorre, S. P., Plueddemann, A. J., Fairall, C. W., … Hersbach, H. (2013, August). On the Exchange of Momentum over the Open Ocean. *Journal of Physical Oceanography*, *43*(8), 1589–1610. doi: 10.1175/JPO-D-12-0173.1

Elvidge, A. D., Renfrew, I. A., Edwards, J. M., Brooks, I. M., Srivastava, P., & Weiss, A. I. (2023). Improved Simulation of the Polar Atmospheric Boundary Layer by Accounting for Aerodynamic Roughness in the Parameterization of Surface Scalar Exchange Over Sea Ice. *Journal of Advances in Modeling Earth Systems*, *15*(3), e2022MS003305. doi: 10.1029/2022MS003305

Elvidge, A. D., Renfrew, I. A., Weiss, A. I., Brooks, I. M., Lachlan-Cope, T. A., & King, J. C. (2016, February). Observations of surface momentum exchange over the marginal ice zone and recommendations for its parametrisation. *Atmospheric Chemistry and Physics*, *16*(3), 1545–1563. doi: 10.5194/acp-16-1545-2016

Facility for Airborne Atmospheric Measurements, Natural Environment Research Council, & Met Office. (2015). *FAAM B769 ACCACIA Transit flight: Airborne atmospheric measurements from core instrument suite on board the BAE-146 aircraft* [Dataset]. NCAS British Atmospheric Data Centre (NCAS BADC). Retrieved from `https://catalogue.ceda.ac.uk/uuid/c064b0c150274a1cbd18c563573f392e`

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., & Edson, J. B. (2003, February). Bulk Parameterization of Air–Sea Fluxes: Updates and Verification for the COARE Algorithm. *Journal of Climate*, *16*(4), 571–591. doi: 10.1175/1520-0442(2003)016⟨0571:BPOASF⟩2.0.CO;2

Garratt, J. R. (1994). *The Atmospheric Boundary Layer.* Cambridge, UK: Cambridge University Press.

Gascard, J.-C., Zhang, J., & Rafizadeh, M. (2019, January). Rapid decline of Arctic sea ice volume: Causes and consequences. *The Cryosphere Discussions*, 1–29. doi: 10.5194/tc-2019-2

Geleyn, J.-F. (1988). Interpolation of wind, temperature and humidity values from model levels to the height of measurement. *Tellus A*, *40A*(4), 347–351. doi: 10.1111/j.1600-0870.1988.tb00352.x

Grachev, A. A., Andreas, E. L., Fairall, C. W., Guest, P. S., & Persson, P. O. G. (2007, September). SHEBA flux–profile relationships in the stable atmospheric boundary layer. *Boundary-Layer Meteorology*, *124*(3), 315–333. doi: 10.1007/s10546-007-9177-6

Grachev, A. A., & Fairall, C. W. (1997, April). Dependence of the Monin–Obukhov Stability Parameter on the Bulk Richardson Number over the Ocean. *Journal of Applied Meteorology and Climatology*, *36*(4), 406–414. doi: 10.1175/1520-0450(1997)036⟨0406:DOTMOS⟩2.0.CO;2

Grachev, A. A., Fairall, C. W., & Bradley, E. F. (2000, March). Convective Profile Constants Revisited. *Boundary-Layer Meteorology*, *94*(3), 495–515. doi: 10.1023/A:1002452529672

Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E., & Svensson, G. (2008, January). Vertical structure of recent Arctic warming. *Nature*, *451*(7174), 53–56. doi: 10.1038/nature06502

Gryschka, M., Gryanik, V. M., Lüpkes, C., Mostafa, Z., Sühring, M., Witha, B., & Raasch, S. (2023). Turbulent Heat Exchange Over Polar Leads Revisited: A Large Eddy Simulation Study. *Journal of Geophysical Research: Atmospheres*, *128*(12), e2022JD038236. doi: 10.1029/2022JD038236

Guemas, V. (2023a, August). *donaldcummins/CDlib: v0.1 (v0.1)* [Software]. Zenodo. doi: 10.5281/ZENODO.8207293

Guemas, V. (2023b, August). *donaldcummins/meteolib: v0.1 (v0.1)* [Software]. Zenodo. doi: 10.5281/ZENODO.8207302

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Neural Networks. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 389–416). New York, NY: Springer. doi: 10.1007/978-0-387-84858-7_11

Högström, U. (1988, Jan 01). Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation. *Boundary-Layer Meteorology*, *42*(1), 55-78. Retrieved from `https://doi.org/10.1007/BF00119875` doi: 10.1007/BF00119875

Hornik, K., Stinchcombe, M., & White, H. (1989, January). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. doi: 10.1016/0893-6080(89)90020-8

Jung, T., Doblas-Reyes, F., Goessling, H., Guemas, V., Bitz, C., Buontempo, C., ... Yang, S. (2015, November). Polar Lower-Latitude Linkages and Their Role in Weather and Climate Prediction. *Bulletin of the American Meteorological Society*, *96*(11), ES197-ES200. doi: 10.1175/BAMS-D-15-00121.1

Kovacs, K. M., Lydersen, C., Overland, J. E., & Moore, S. E. (2011, March). Impacts of changing sea-ice conditions on Arctic marine mammals. *Marine Biodiversity*, *41*(1), 181–194. doi: 10.1007/s12526-010-0061-0

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling.* New York, NY: Springer. doi: 10.1007/978-1-4614-6849-3

Lesins, G., Duck, T. J., & Drummond, J. R. (2012, December). Surface Energy Balance Framework for Arctic Amplification of Climate Change. *Journal of Climate*, *25*(23), 8277–8288. doi: 10.1175/JCLI-D-11-00711.1

Leufen, L. H., & Schädler, G. (2019, May). Calculating the turbulent fluxes in the atmospheric surface layer with neural networks. *Geoscientific Model Development*, *12*(5), 2033–2047. doi: 10.5194/gmd-12-2033-2019

Liu, J., Song, M., Zhu, Z., Horton, R. M., Hu, Y., & Xie, S.-P. (2022, August). Arctic sea-ice loss is projected to lead to more frequent strong El Niño events. *Nature Communications*, *13*(1), 4952. doi: 10.1038/s41467-022-32705-2

Lüpkes, C., & Gryanik, V. M. (2015). A stability-dependent parametrization of transfer coefficients for momentum and heat over polar sea ice to be used in climate models. *Journal of Geophysical Research: Atmospheres*, *120*(2), 552–581. doi: 10.1002/2014JD022418

Lüpkes, C., Gryanik, V. M., Hartmann, J., & Andreas, E. L. (2012). A parametrization, based on sea ice morphology, of the neutral atmospheric drag coefficients for weather prediction and climate models. *Journal of Geophysical Research: Atmospheres*, *117*(D13). doi: 10.1029/2012JD017630

Marcq, S., & Weiss, J. (2012, February). Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere. *The Cryosphere*, *6*(1), 143–156. doi: 10.5194/tc-6-143-2012

McCandless, T., Gagne, D. J., Kosović, B., Haupt, S. E., Yang, B., Becker, C., & Schreck, J. (2022, November). Machine Learning for Improving Surface-Layer-Flux Estimates. *Boundary-Layer Meteorology*, *185*(2), 199–228. doi: 10.1007/s10546-022-00727-4

Meier, W. N., Fetterer, F., Windnagel, A., & Stewart, S. (2021). *NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4*

[Dataset]. NSIDC. doi: 10.7265/EFMZ-2T65

Meier, W. N., Hovelsrud, G. K., van Oort, B. E., Key, J. R., Kovacs, K. M., Michel, C., ... Reist, J. D. (2014). Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. *Reviews of Geophysics*, *52*(3), 185–217. doi: 10.1002/2013RG000431

Monin, A. S., & Obukhov, A. M. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere. *Tr. Akad. Nauk SSSR Geophiz. Inst.*, *24*(151), 163–187.

Nixdorf, U., Dethloff, K., Rex, M., Shupe, M., Sommerfeld, A., Perovich, D. K., ... Boetius, A. (2021, September). MOSAiC Extended Acknowledgement. doi: 10.5281/ZENODO.5179738

Pal, S., & Sharma, P. (2021, March). A Review of Machine Learning Applications in Land Surface Modeling. *Earth*, *2*(1), 174–190. doi: 10.3390/earth2010011

Persson, P. O. G., Fairall, C. W., Andreas, E. L., Guest, P. S., & Perovich, D. K. (2002). Measurements near the Atmospheric Surface Flux Group tower at SHEBA: Near-surface conditions and surface energy budget. *Journal of Geophysical Research: Oceans*, *107*(C10), SHE 21-1-SHE 21-35. doi: 10.1029/2000JC000705

Post, E., Bhatt, U. S., Bitz, C. M., Brodie, J. F., Fulton, T. L., Hebblewhite, M., ... Walker, D. A. (2013, August). Ecological Consequences of Sea-Ice Decline. *Science*, *341*(6145), 519–524. doi: 10.1126/science.1235225

Previdi, M., Smith, K. L., & Polvani, L. M. (2021, September). Arctic amplification of climate change: A review of underlying mechanisms. *Environmental Research Letters*, *16*(9), 093003. doi: 10.1088/1748-9326/ac1c29

Prytherch, J., Brooks, I. M., Crill, P. M., Thornton, B. F., Salisbury, D. J., Tjernström, M., ... Humborg, C. (2017). Direct determination of the air-sea CO2 gas transfer velocity in Arctic sea ice regions. *Geophysical Research Letters*, *44*(8), 3770–3778. doi: 10.1002/2017GL073593

R Core Team. (2021). *R: A language and environment for statistical computing* [Manual]. Vienna, Austria.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., ... Laaksonen, A. (2022, August). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth &*

Environment, *3*(1), 1–10. doi: 10.1038/s43247-022-00498-3

Renfrew, I. A., Elvidge, A. D., & Edwards, J. M. (2019). Atmospheric sensitivity to marginal-ice-zone drag: Local and global responses. *Quarterly Journal of the Royal Meteorological Society*, *145*(720), 1165–1179. doi: 10.1002/qj.3486

Ripley, B. D. (1996). Feed-forward Neural Networks. In *Pattern Recognition and Neural Networks* (pp. 143–180). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511812651.006

Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Guérémy, J.-F., ... Sénési, S. (2020). The CNRM Global Atmosphere Model ARPEGE-Climat 6.3: Description and Evaluation. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2020MS002075. doi: 10.1029/2020MS002075

Rothrock, D. A., Yu, Y., & Maykut, G. A. (1999). Thinning of the Arctic sea-ice cover. *Geophysical Research Letters*, *26*(23), 3469–3472. doi: 10.1029/1999GL010863

Screen, J. A., & Simmonds, I. (2010, April). The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, *464*(7293), 1334–1337. doi: 10.1038/nature09051

Serreze, M. C., Barrett, A. P., Stroeve, J. C., Kindig, D. N., & Holland, M. M. (2009, February). The emergence of surface-based Arctic amplification. *The Cryosphere*, *3*(1), 11–19. doi: 10.5194/tc-3-11-2009

Serreze, M. C., & Barry, R. G. (2011, May). Processes and impacts of Arctic amplification: A research synthesis. *Global and Planetary Change*, *77*(1), 85–96. doi: 10.1016/j.gloplacha.2011.03.004

Serreze, M. C., & Francis, J. A. (2006, June). The Arctic Amplification Debate. *Climatic Change*, *76*(3), 241–264. doi: 10.1007/s10584-005-9017-y

Shupe, M. D., Rex, M., Blomquist, B., Persson, P. O. G., Schmale, J., Uttal, T., ... Yue, F. (2022, February). Overview of the MOSAiC expedition: Atmosphere. *Elementa: Science of the Anthropocene*, *10*(1), 00060. doi: 10.1525/elementa.2021.00060

Sotiropoulou, G., Tjernström, M., Sedlar, J., Achtert, P., Brooks, B. J., Brooks, I. M., ... Wolfe, D. (2016, December). Atmospheric Conditions during the Arctic Clouds in Summer Experiment (ACSE): Contrasting Open Water and Sea Ice Surfaces during Melt and Freeze-Up Seasons. *Journal of Climate*, *29*(24), 8721–8744. doi: 10.1175/JCLI-D-16-0211.1

Srivastava, P., Brooks, I. M., Prytherch, J., Salisbury, D. J., Elvidge, A. D., Renfrew, I. A., & Yelland, M. J. (2022, April). Ship-based estimates of momentum transfer coefficient over sea ice and recommendations for its parameterization. *Atmospheric Chemistry and Physics*, *22*(7), 4763–4778. doi: 10.5194/acp-22-4763-2022

Svensson, G., Holtslag, A. A. M., Kumar, V., Mauritsen, T., Steeneveld, G. J., Angevine, W. M., … Zampieri, M. (2011, August). Evaluation of the Diurnal Cycle in the Atmospheric Boundary Layer Over Land as Represented by a Variety of Single-Column Models: The Second GABLS Experiment. *Boundary-Layer Meteorology*, *140*(2), 177–206. doi: 10.1007/s10546-011-9611-7

Tjernström, M., & Jakobsson, M. (2021). *Data from expedition Arctic Ocean, 2016.* Bolin Centre Database. doi: 10.17043/ODEN-AO-2016-EXPEDITION-1

Tynan, E. (2015, July). Effects of sea-ice loss. *Nature Climate Change*, *5*(7), 621–621. doi: 10.1038/nclimate2708

Uttal, T., Curry, J. A., McPhee, M. G., Perovich, D. K., Moritz, R. E., Maslanik, J. A., … Grenfeld, T. C. (2002, February). Surface Heat Budget of the Arctic Ocean. *Bulletin of the American Meteorological Society*, *83*(2), 255–276. doi: 10.1175/1520-0477(2002)083⟨0255:SHBOTA⟩2.3.CO;2

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (J. Chambers, W. Eddy, W. Härdle, S. Sheather, & L. Tierney, Eds.). New York, NY: Springer. doi: 10.1007/978-0-387-21706-2

Vihma, T. (1995). Subgrid parameterization of surface heat and momentum fluxes over polar oceans. *Journal of Geophysical Research: Oceans*, *100*(C11), 22625–22646. doi: 10.1029/95JC02498

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., … Waldman, R. (2019). Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2177–2213. doi: 10.1029/2019MS001683

Wulfmeyer, V., Pineda, J. M. V., Otte, S., Karlbauer, M., Butz, M. V., Lee, T. R., & Rajtschan, V. (2022, November). Estimation of the Surface Fluxes for Heat and Momentum in Unstable Conditions with Machine Learning and Similarity Approaches for the LAFE Data Set. *Boundary-Layer Meteorology*. doi:

10.1007/s10546-022-00761-2