Bowen ratio-constrained global dataset of bulk air-sea turbulent heat fluxes from 1993 to 2017

Yizhe Wang[a, b], Ronglin Tang[a, b, *], Meng Liu[c], Lingxiao Huang[a, b], Zhao-Liang Li[a, b, c]

[a] State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

[c] State Key Laboratory of Efficient Utilization of Arable Land in China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

* Authors to whom correspondence should be addressed: tangrl@lreis.ac.cn

Abstract

Air-sea turbulent heat fluxes, including the sensible heat flux (SHF) and latent heat flux (LHF), along with the Bowen ratio (β, ratio of SHF to LHF), are crucial for understanding air-sea interaction and global energy and water budgets. However, the existing products, primarily developed using the semi-empirical bulk aerodynamic methods and data-driven machine learning approaches, are often weak in accuracy and physical rationality, due to the uncertainties in the environmental forcings and inappropriate parameterizations. In this study, we generated a global daily 0.25° product of bulk air-sea turbulent heat fluxes using the Bowen ratio-constrained Neural Network (NN) model (referred to as the BrTHF model) that could coordinately estimate the SHF and LHF, along with the observations from 197 globally distributed buoys and multi-source remote sensing and reanalysis inputs. The spatial ten-fold cross-validation results showed that the BrTHF model, achieving root mean square errors of 6.05 W/m^2, 23.67 W/m^2 and 0.22 and correlation coefficients of 0.93, 0.91 and 0.25 for the SHF, LHF and β, respectively, outperformed the physics-agnostic NN model and seven widely used air-sea turbulent heat flux products (including JOFURO3, IFREMER, SeaFlux, ERA5, MERRA2, OAFlux, and OHF). Furthermore, the inter-comparison of the spatial distribution of multi-year means, as well as intra-annual and inter-annual change patterns showed that the BrTHF product reliably simulated global SHF, LHF and β, in contrast to the machine learning-based OHF product that failed to replicate these patterns. The main advantage of the BrTHF model lies in its improved rationality of β estimates, successfully eliminating the outliers observed in the physics-agnostic NN model and the seven typical products. The improved SHF, LHF, and β estimates can allow for more accurate quantification of the global air-sea energy and water budgets, enhance our understanding of air-sea interaction, and improve projections of climate change under global warming. The 0.25° daily global product from 1993 to 2017 can be freely accessed from the National Tibetan Plateau Data Center (TPDC) [https://doi.org/10.11888/Atmos.tpdc.302578, Tang and Wang (2025)].

Keywords: Air-sea turbulent heat fluxes; Sensible heat flux; Latent heat flux; Bowen ratio

1. Introduction

Air-sea turbulent heat fluxes, comprising the latent heat flux (LHF) and sensible heat flux (SHF), play vital roles in the Earth's climate system by characterizing the exchange of energy and water between the ocean and atmosphere (Wild et al., 2014; Loeb et al., 2021; Fasullo et al., 2014). Accurate estimation of SHF, LHF and their ratio—the Bowen ratio (β = SHF/LHF) is an essential prerequisite for advancing our understanding of atmosphere-sea interaction (Gentemann et al., 2020), improving the quantification of global water and energy budget (Zhang, 2023), and enhancing the predictability of extreme weather events (Yu, 2019).

To estimate global air–sea turbulent heat fluxes, the semi-empirical bulk aerodynamic method was developed based on the Monin–Obukhov similarity theory (Monin and Obukhov, 1954). It establishes scaling relationships between fluxes and near-surface meteorological variables such as wind speed, humidity, and temperature (Yu, 2019). This method, for its ease of application, has been applied to generate tens of widely used products in the past few decades (Shie et al., 2009; Liman et al., 2018; Yu and Weller, 2007; Berry and Kent, 2011; Tomita et al., 2018; Crespo et al., 2019). However, there were huge discrepancies in the global and regional magnitude and patterns of SHF and LHF among these products, which seriously impeded our understanding of the air-sea interaction and the global budget of water and energy (Bentamy et al., 2017; Tang et al., 2024; Yu, 2019). The discrepancies could be partly attributed to the substantial uncertainties in the environmental forcings used to develop these products (Robertson et al., 2020) and the inappropriate parameterizations (Brodeau et al., 2017; Jiang et al., 2024a; Jiang et al., 2024b; Yang et al., 2024). More explicitly, existing parameterizations often rely on simplified assumptions about atmospheric stability and boundary layer dynamics, which may not hold under diverse environmental conditions. For instance, most bulk algorithms are optimized for moderate wind regimes, resulting in degraded performance and increased uncertainty when applied under weak wind regimes (Jiang et al., 2024a; Brunke, 2002). At very high wind speeds, however, observations show that the drag coefficient can decrease due to sea spray and whitecap formation, reducing effective surface roughness and potentially biasing flux estimates (Cai et al., 2025). In addition, simplifications in the treatment of sea surface skin temperature, saturation humidity, and air density in the parameterizations can also introduce substantial uncertainty (Brodeau et al., 2017). Together, these limitations can contribute a lot to the biases in the SHF and LHF estimates which can even lead to the unphysical estimations of β, as Wang et al. (2024) reported. To better describe and comprehend the air-sea interaction and the energy and water budgets, the existing mode to produce global air-sea turbulent heat fluxes needs improvement urgently.

Machine learning techniques have been extensively applied to upscale point-scale in-situ measurements of a single variable (such as soil moisture, roughness, or temperature) into grid-scale global datasets (Wang et al., 2023; Peng et al., 2022; O and Orth, 2021; Nelson et al., 2024; Fu et al., 2023). These efforts highlight the great potential of machine learning for more accurate and consistent multivariate coordinated mapping (Karniadakis et al., 2021; Kashinath et al., 2021; Van Der Westhuizen et al., 2023; Wang et al., 2024). However, the application of machine learning in global mapping of air-sea turbulent heat fluxes remains limited. Among these studies, some have focused on solely improving the accuracy of LHF, while the remaining studies have mostly considered independent modeling of SHF and LHF (Bourras et al., 2007; Cummins et al., 2024; Cummins et al., 2023; Zhou et al., 2024). In both cases, however, most studies have not produced long-term flux products. The only publicly available machine learning-based global air-sea turbulent heat fluxes product, released by the National Oceanic and Atmospheric Administration (NOAA) Ocean heat flux CDR (hereafter dubbed OHF), simultaneously modeled SHF and LHF using a Neural Network (NN) technique (Clayson and Brown, 2016). Although it performed well when validated against the observations from the tropical buoys, it failed to capture the regional characteristics, particularly in areas where air-sea turbulent heat exchange is intense (e.g. oceans with latitudes beyond 45° for SHF and subtropical highs for LHF) (Tang et al., 2024). Additionally, it exhibited different patterns of temporal evolution of global annual mean and opposite inter-annual trends at both regional and global scales to most widely-used physical model-based products, likely due to unreasonable construction of observation datasets [with data before and after 2007 coming from SeaFlux in-situ datasets and ICOADS (International Comprehensive Ocean-Atmosphere Data Set) datasets, respectively]. Furthermore, the product likely suffers from unphysical estimates of the β due to neglecting the interrelations among SHF, LHF and β during the model construction.

To improve the estimation of SHF, LHF, and β in a coordinative framework, we recently proposed an innovative Bowen ratio-informed data-driven model by considering the synergistic changes [on the one hand, ensuring physical consistency (i.e., SHF/LHF = β); on the other hand, achieving high-accuracy estimations of SHF, LHF, and β simultaneously] using a Random Forest (RF) technique (Wang et al., 2024). Validation against hourly eddy covariance (EC) flux measurements from 53 historical cruises demonstrated the model's superior performance, achieving high accuracy in estimating SHF, LHF, and β, with results that are physically consistent. Wang et al. (2024) highlights the feasibility of simultaneously estimating SHF, LHF, and β with high accuracy using machine learning techniques, offering strong potential for global mapping that aligns with physical consistency. However, since EC observations are difficult to obtain at sea due to platform motion and airflow distortion (Bourras et al., 2019; Bourras et al., 2009)—their limited spatio-temporal coverage constrains the application of the model for global mapping. Buoy-based flux observations provide a viable alternative. Buoy data offer globally representative flux samples with adequate volume and acceptable accuracy, which have been widely used to evaluate the performance of global products (Bentamy et al., 2017; Bourras, 2006; Tang et al., 2024;

Weller et al., 2022; Zhou et al., 2020) and support global modeling (Chen et al., 2020a) and analysis (Song et al., 2024; Yan et al., 2024).

The primary objectives of this study are three fold: (1) to develop an innovative Bowen ratio-constrained statistical model for improving the air-sea SHF, LHF and β estimates (referred to as the BrTHF model hereafter) using the machine learning technique and global buoy-based air-sea turbulent heat fluxes observations; (2) to demonstrate the superiority of the statistical model through an inter-comparison with seven widely used global products and the estimates from the physics-free machine learning-based model; (3) to produce a global daily 0.25° dataset based on the BrTHF model over ice-free oceans covering the period from 1993 to 2017. The flux observations from 197 global distributed buoys, along with multi-source satellite-based and reanalysis-based inputs, were collected to construct the models and further produce the global air-sea turbulent heat fluxes dataset. The accuracy and spatio-temporal patterns of the SHF, LHF and β estimates were inter-compared with seven widely used products, including the remote sensing-based JOFURO v3, IFREMER v4.1 and SeaFlux v3, as well as reanalysis-based ERA5 and MERRA2, hybrid-based OAFlux v3 and machine learning-based OHF v2 products.

2. Data and Methods

The following sub-sections provide an overview of the development of the BrTHF product, detailing the construction of air-sea turbulent heat fluxes observation datasets, learning datasets and the BrTHF model, as well as the evaluation strategies used in this study, as indicated in Figure 1.

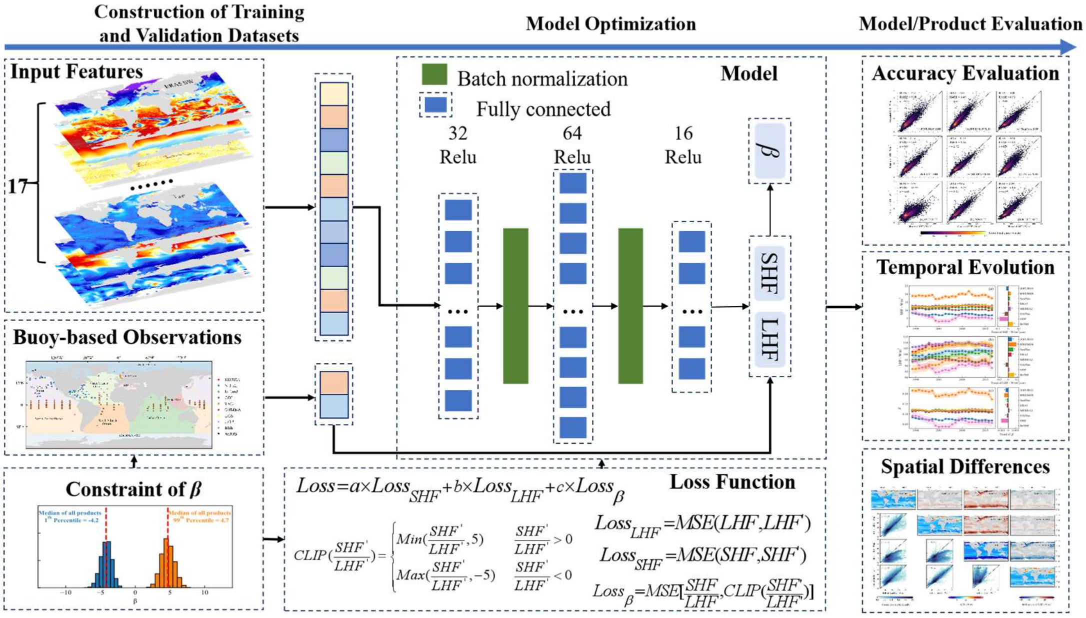


Figure 1. flowchart of the generation of a global product of air-sea SHF, LHF and *β* by the BrTHF model

2.1 Air-sea turbulent heat fluxes observation datasets

To obtain the buoy-derived air-sea turbulent heat fluxes observations, the hourly or sub-hourly oceanic and atmospheric measurements including sea surface temperature (T_s), sea surface air temperature (T_a), sea surface wind speed (*WS*) and relative humidity (*RH*) were firstly collected at 268 buoys covering a variety of ocean basins from 13 organizations or networks. Detailed information about the buoy sources and the number of buoys from each provider is summarized in Table 1.

For certain buoys lacking *RH* measurements [e.g. buoys from NDBC (National Data Buoy Center) provided dew point temperature (*DEW*) rather than *RH*], the *RH* was computed using *DEW* and T_a. To ensure the quality of the measurements, we filtered the records based on the quality control recommendations provided by the data providers. Further refinement was also made by removing the questionable values that exceeded three standard deviations (3σ) for each variable at individual buoys.

Once the data was cleaned, daily mean aggregation was applied to the oceanic and atmospheric measurements. Given the varying temporal resolutions of the measurements (e.g. NDBC provided hourly observations before 2005 and 10-min observations thereafter), we only retained the daily mean data when the fraction of the valid hourly or sub-hourly observations exceeded 80% on a given day.

Table 1 Summary of f buoys used in this study

Network/Organization	Number of Buoys	Network/Organization	Number of Buoys
TAO/TRITON (Tropical Atmosphere Ocean / TRITON)	67	KOREA (Korea Meteorological Administration)	24
PIRATA (Prediction and Research Moored Array)	20	OOI (Ocean Observatories Initiative)	6
RAMA (Research Moored Array for Monsoon Analysis)	23	AOOS (Alaska Ocean Observing System)	2
NDBC (National Data Buoy Center)	73	JKEO (Japan Agency for Marine-Earth Science and Tech.)	1
Copernicus Marine In Situ TAC	19	Irish Weather Buoy Network	6
UOP (Upper Ocean Processes Group)	23	Icelandic Meteorological Office (IMO)	1
OCS (Ocean Climate Stations Project)	3		

After the above mentioned data preprocessing, the daily buoy-derived air-sea turbulent heat fluxes (SHF and LHF) observations were then calculated using the daily oceanic and atmospheric measurements combined with the version 3.5 of Coupled Ocean-Atmosphere Response Experiment (COARE3.5) model (Edson, 2013) (available at https://github.com/NOAA-PSL/COARE-algorithm) with the cool-skin and warm-layer calculation switched off. The configuration follows the practice adopted by Pacific Marine Environmental Laboratory for producing daily air-sea turbulent heat flux products (https://www.pmel.noaa.gov/tao/drupal/flux/documentation-lw.html) at Global Tropical Moored Buoy Array (TAO/TRITON, PIRATA, and RAMA).ss Note that outliers of β present in the observations are likely associated with uncertainties in the model-derived estimates and input data. Considering that such outliers can severely impede model training and evaluation, it was necessary to constrain β within a reasonable range to enable simultaneous high-accuracy estimation of SHF, LHF, and β. Therefore, following the air-sea turbulent heat fluxes computations, we further made a quality control on the derived SHF and LHF observations to exclude the abnormal records, by filtering the observations based on the range of daily β values determined from seven widely-used flux products. Specifically, we calculated the cumulative distribution of daily β for each product and their ensemble (across all products). The medians of the 1st and 99th percentiles, approximately -5 and 5, respectively, were selected as the minimum and maximum of valid daily β, as shown in Figure S1. In total, this study compiled 463,585 observations of valid daily air-sea turbulent heat flux from 197 buoy stations (Figure 2 and Table S1) in the Arctic Ocean, Pacific Ocean, Atlantic Ocean and Indian Ocean.

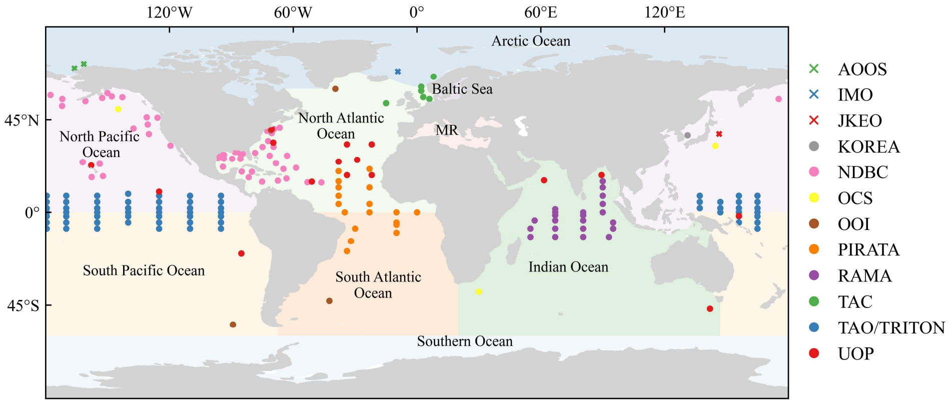


Figure 2. Geographic locations of 197 buoy sites from 12 organizations or networks involved in this analysis including TAO/TRITON, PIRATA, RAMA, NDBC, TAC, UOP, OOI, AOOS, KOREA, OCS, JKEO and IMO. The boundaries of global land and open oceans were sourced from the Natural Earth dataset (https://www.naturalearthdata.com/downloads/) and the Global Oceans and Seas dataset (https://www.marineregions.org/sources.php), respectively. Abbreviations MR refers to the Mediterranean Region. It should be noted that the Caspian Sea was not included within the boundaries of the open oceans and is shown in white.

Finally, the quality-controlled observations were collected to train and validate the BrTHF model. Note that the COARE-based observations at the buoy stations have already widely applied as a benchmark for global air-sea turbulent heat flux product development and evaluation (Bentamy et al., 2017; Chen et al., 2020b; Tang et al., 2024; Weller et al., 2022)

2.2 Learning datasets and state-of-the-art products

2.2.1 Learning datasets for training the neural network

Learning variables were carefully selected based on their potential impacts on the variations of the air-sea turbulent heat fluxes (Grist et al., 2016; Kudryavtsev et al., 2014; Myslenkov et al., 2021; Song, 2020, 2021; Yan et al., 2024) to conduct the feature selection (see section 2.3.1). These variables include T_a, sea surface air specific humidity (Q_a), Mean Sea Level Pressure (*SLP*), Downward Long Wave Radiation Flux (*LW*), Downward Short Wave Radiation Flux (*SW*), T_s, sea surface specific humidity (Q_s), Absolute Dynamic Topography (*ADT*), Sea Level Anomaly (*SLA*), Sea Surface Salinity (*SSS*), Sea Surface Density (*SSD*), Ocean Mixed Layer Current Velocity (*CS*), *WS*, Significant Wave Height (*SWH*), Wave period (T_p), as well as gradient of temperature ($diff_T$) calculated using the T_s and T_a, and gradient of humidity ($diff_Q$) calculated using the Q_s and Q_a.

Datasets of these learning variables were collected from multiple publicly available sources, as summarized in Table 2 and were used as the input features for training the neural network. The Multi Observation Global Ocean Sea Surface Salinity and Sea Surface Density (MOGOSD) dataset and Global Ocean Waves (GOW) Reanalysis dataset were spatially resampled to a 0.25° resolution using mean aggregation, while temporal mean aggregation to daily values was applied to the GOW dataset (originally at 3-hour resolution) and Cross-Calibrated Multi-Platform (CCMP) wind vector analysis dataset (6-hour resolution). Additionally, a daily ERA5 sea-ice mask was applied to the datasets to mitigate the impact of sea ice.

2.2.2 State-of-the-art products for inter-comparison

Seven widely used air-sea turbulent heat flux products, involving remote sensing-based JOFURO3, IFREMER and SeaFlux, as well as reanalysis-based ERA5 and MERRA2, hybrid-based OAFlux and machine learning-based OHF products were selected for inter-comparison.

The remote sensing-based JOFURO3, IFREMER, and SeaFlux products were developed by the Japanese Ocean Flux Data Sets under the Remote Sensing Observations (J-OFURO) research project, the Institute Français de Recherche pour l'Exploitation de la Mer (IFREMER), and the NASA Global Hydrology Resource Center (GHRC) Distributed Active Archive Center (DAAC), respectively. The reanalysis-based ERA5 and MERRA2 products were developed by the ECMWF and NASA Global Modeling and Assimilation Office (GMAO), respectively. The hybrid-based OAFlux and machine learning-based OHF products were developed or published by the Woods Hole Oceanographic Institution (WHOI) and NOAA Ocean Surface Bundle (OSB) Climate Data Record (CDR), respectively. With the exception of the OHF product calculating SHF and LHF simultaneously using a NN model without a constraint, all other products employed bulk aerodynamic methods to estimate SHF and LHF. The JOFURO3, IFREMER, and OAFlux products used the COARE3.0 model, while the SeaFlux used the COARE3.5 model. Differently, the ERA5 adopted the bulk aerodynamic method used by the ECMWF, and the MERRA2 used the method by the GEOS. These products provide SHF and LHF estimates at a 0.25° spatial resolution, except for the MERRA2 (0.5°×0.625°) and OAFlux (1°). Additionally, most products provide daily SHF and LHF estimates, while only the OHF product provides estimates at a 3-hour interval. For further inter-comparison, the daily mean aggregation was applied to the OHF products. More details about the seven products can be found in the review of Tang et al. (2024).

Table 2 Summary of learning datasets used in this study

Dataset source	Resolution	Variables	Urls
ERA5	0.25°/daily	Sea surface air temperature (T_a), sea surface air specific humidity (Q_a), mean sea level pressure (SLP), downward long wave radiation flux (LW) and downward short wave radiation flux (SW)	https://cds.climate.copernicus.eu/datasets/derived-era5-single-levels-daily-statistics?tab=overview
OSCAR	0.25°/daily	Ocean mixed layer current velocity (CS)	https://podaac.jpl.nasa.gov/dataset/OSCAR_L4_OC_FINAL_V2.0
CCMP	0.25°/6-hour	Wind speed (WS)	https://data.remss.com/ccmp/v03.0/daily/
MOGOSD	0.125°/daily	Sea surface salinity (SSS) and sea surface density (SSD),	https://data.marine.copernicus.eu/product/MULTIOBS_GLO_PHY_S_SURFACE_MYNRT_015_013/description
SSH	0.25°/daily	Absolute dynamic topography (ADT) and sea level anomaly (SLA)	https://data.marine.copernicus.eu/product/SEALEVEL_GLO_PHY_CLIMATE_L4_MY_008_057/description
GOW	0.2°/3-hour	Significant wave height (SWH) and wave period (T_p)	https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_WAV_001_032/description
OISST	0.25°/daily	Sea surface temperature (T_s) and sea surface specific humidity (Q_s)	https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/v2.1/access/avhrr/
OISST – ERA5	0.25°/daily	Gradient of temperature ($diff_T$) and gradient of humidity ($diff_Q$)	-

Table 3 Summary of the state-of-the-art air-sea turbulent heat flux products used for inter-comparison in this study

Dataset source	Resolution	Model	Variables	Urls
JOFURO3	0.25°/daily	COARE3.0	Latent heat flux (LHF), sensible heat flux (SHF) and Bowen ratio (β = SHF/LHF)	https://www.j-ofuro.com/en/
IFREMER	0.25°/daily	COARE3.0		ftp://ftp.ifremer.fr/ifremer/cersat/data/heat-flux/ifremer/v4.1/daily
SeaFlux	0.25°/daily	COARE3.5		https://www.earthdata.nasa.gov/data/catalog/ghrc-daac-seaflux-1
MERRA2	0.5°×0.625°/daily	GEOS		https://developers.google.com/earth-engine/datasets/catalog/NASA_GSFC_MERRA_flx_2?hl=zh-cn#bands
ERA5	0.25°/daily	ECMWF		https://cds.climate.copernicus.eu/datasets/derived-era5-single-levels-daily-statistics?tab=overview
OAFlux	1°/daily	COARE3.0		ftp://ftp.whoi.edu/pub/science/oaflux/data_v3
OHF	0.25°/3-hour	Neural Network model		https://www.ncei.noaa.gov/products/climate-data-records/ocean-heat-fluxes

2.3 Construction of the BrTHF model

2.3.1 Feature selection

The study employed a random forest (RF) model to evaluate the importance scores of 17 oceanic and atmospheric learning variables (with datasets collected in Section 2.2) for target variables (SHF and LHF), aiming to filter out less influential variables. As shown in Table S2, the variable importance assessment revealed that $diff_T$ and $diff_Q$ showed the highest importance score (71.56% and 49.93%) for SHF and LHF modelling, respectively; additionally, WS exhibited the second highest importance for both SHF (10.19%) and LHF (27.59%) modelling. Building upon the importance evaluation and through careful screening of highly correlated variables, we ultimately selected 11 key environmental features for subsequent air-sea turbulent heat flux modelling including SLP, LW, SW, SSS, ADT, CS, WS, SWH, T_p, $diff_Q$, and $diff_T$.

2.3.1 Model construction and optimization

We selected the NN technique to build the BrTHF model due to its strong ability to capture the complex nonlinear relationships between the multiple inputs and multiple-target variables with high accuracy (Zhou et al., 2024; Fu et al., 2023; Cummins et al., 2023; Cummins et al., 2024). Additionally, the technique enables the seamless integration of physical constraints, improving the reasonableness of the results (Zhou et al., 2024; Zhao et al., 2019; Shang et al., 2023).

In order to estimate the SHF and LHF with high accuracy in a physics-consistency framework, the β (= SHF/LHF) physical constraint was incorporated into the NN model using the custom multiple-objects (SHF, LHF and β) loss function as follows:

$$Loss = a \times Loss_{SHF} + b \times Loss_{LHF} + c \times Loss_{\beta} \tag{1}$$

$Loss_{SHF}$, $Loss_{LHF}$ and $Loss_{\beta}$ represent the Mean Squared Error (MSE) of SHF, LHF and β, respectively. They were weighted using the factors of a (SHF), b (LHF) and c (β) to balance the different magnitudes of loss during optimization. To prevent potential gradient explosion during model training, predicted β [SHF'/LHF', calculated using the predicted SHF (SHF') and LHF (LHF')] values were clipped within the observed range of β (from -5 to 5) during training:

$$CLIP(\frac{SHF'}{LHF'}) = \begin{cases} Min(\frac{SHF'}{LHF'}, 5) & \frac{SHF'}{LHF'} > 0 \\ Max(\frac{SHF'}{LHF'}, -5) & \frac{SHF'}{LHF'} < 0 \end{cases} \tag{2}$$

$$Loss_{\beta} = MSE(\frac{SHF}{LHF}, CLIP(\frac{SHF'}{LHF'})) \tag{3}$$

Finally, after optimization, the final weights (a, b and c) for SHF, LHF, and β were set to 5, 1, and 250, respectively. The model was constructed consisting of one input layer, three hidden layers, two BatchNormalization layers, and one output layer using the Python TensorFlow library. The numbers of neurons in the three hidden layers were 32, 64, and 16, respectively and the activation function of Leaky Rectified linear unit (ReLU) was used throughout the model.

To illustrate the superiority of the BrTHF model in terms of accuracy and physical consistency, another physics-free NN models, trained without integrating the β constraint, were also constructed to predict SHF and LHF separately for further comparison, where β was calculated to be SHF/LHF.

2.4 Evaluation strategy

A spatial 10-fold cross-validation was employed to assess the performance of the BrTHF model for estimating air-sea SHF, LHF and β. Compared to the traditional 10-fold cross-validation, which randomly split all samples into ten folds and thus may result in overlapping spatial samples between training and validating datasets, the spatial 10-fold cross-validation was conducted in a relatively independent spatial distribution and can provide a more robust and generalizable evaluation.

Specifically, first, all buoy sites were randomly split into ten folds. Then, each fold was in succession selected as the validation dataset and the remaining nine folds was used as the training dataset.

The metrics used to evaluate the performance of the models include: (1) the mean bias error (BIAS); (2) the root mean squared error (RMSE); (3) the correlation coefficient (r).

$$BIAS = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{5}$$

$$r = \frac{\sum_{i=1}^{n}[(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6}$$

where n is the number of samples, $\hat{y}_i$ and y_i are the estimated value and reference truth, $\bar{\hat{y}}$ and $\bar{y}$ are the mean of $\hat{y}_i$ and y_i, respectively. These metrics—BIAS, RMSE, and r—comprehensively evaluate model performance, representing systematic deviation, dispersion between observations and estimates, and the strength and direction of the linear relationship, respectively. Note that RMSE and r can be sensitive to extreme values, particularly for the β, which is a ratio-based variable and may exhibit unrealistically large magnitudes under near-zero flux conditions. To ensure a fair evaluation, model performance is assessed both with all samples retained and with extreme β values excluded in subsequent analyses. This dual evaluation allows us to quantify overall model performance while explicitly accounting for the influence of rare extreme cases.

3. Results and discussion

3.1 Spatial ten-fold cross-validation of the models

3.1.1 Overall accuracy

Figures 3, 4 and 5 present the normalized scatter density plots of the estimated daily SHF, LHF and β from the BrTHF and physics-free NN models, as well as the seven air-sea turbulent heat flux products against the observations obtained from 197 global distributed buoys by the spatial ten-fold cross-validation strategy.

Most models and products showed data distributions closely aligned with the observed SHF, with the majority of samples clustered around the 1:1 line. The BrTHF model slightly overestimated SHF with a BIAS of 0.09 W/m^2, whereas the physics-free NN models, and ERA5 and IFREMER products showed more pronounced overestimations (from 0.42 W/m^2 to 4.05 W/m^2). In contrast, the remaining five products exhibited notable underestimations (from -3.44 W/m^2 to -0.41 W/m^2). As illustrated in Figure 6, the variability of estimated SHF from the BrTHF and the physics-free NN models and ERA5 product closely matched the observed SHF, all with a Standard Deviation (STD) of approximately 16 W/m^2. Notably, the BrTHF model achieved the lowest RMSE (6.05 W/m^2), outperforming both the physics-free NN models (6.29 W/m^2) and the seven air-sea turbulent heat flux products (ranging up to 12.34 W/m^2 for OHF). Additionally, the BrTHF model combined with the physics-free NN models yielded the highest values of r (0.93), surpassing all seven other products. In summary, the BrTHF model showed the best overall performance in estimating SHF among all the models and products.

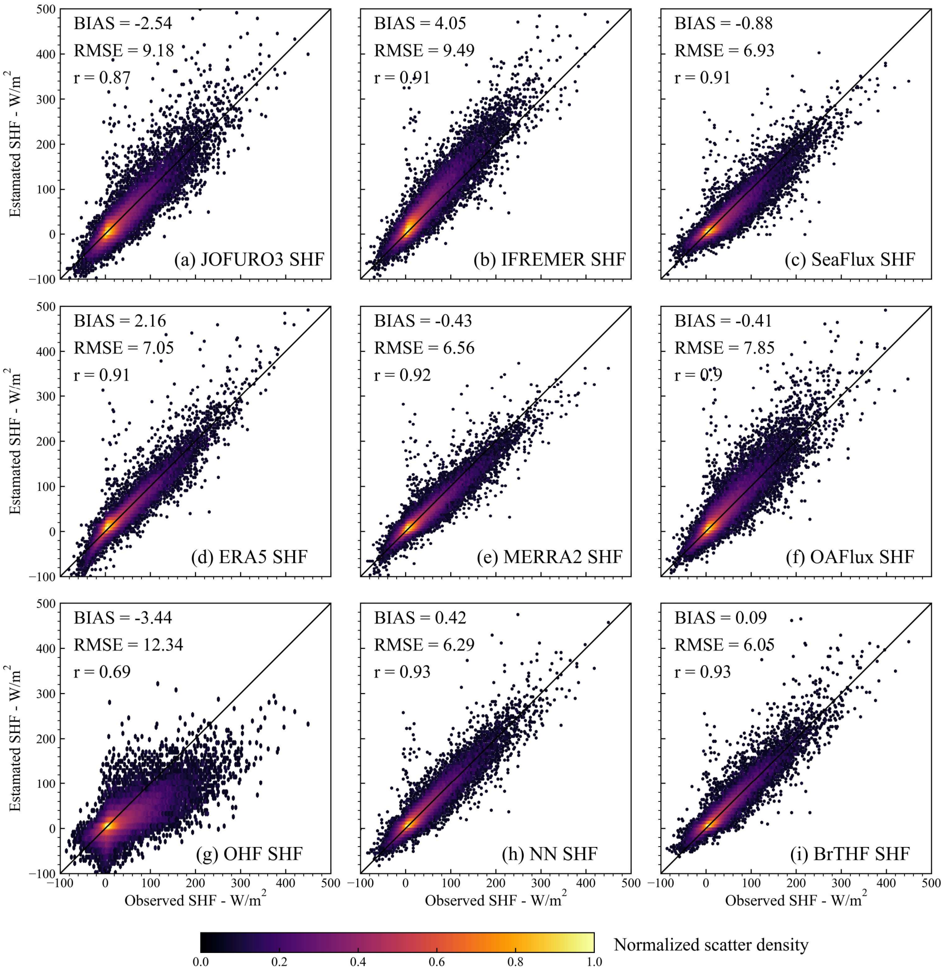


Figure 3. Normalized scatter density plots of estimated SHF from the BrTHF model, the physics-free NN models and seven air-sea turbulent heat flux products against the observed SHF obtained from 197 global distributed buoys.

For LHF, similar to the results for SHF, the BrTHF model demonstrated the best agreement with observations, achieving the lowest RMSE (23.67 W/m^2) and the highest value of r (0.91). Compared to the physics-free NN models and seven products, the BrTHF model reduced RMSE by 2.05 W/m^2 (physics-free NN models) to 12.38 W/m^2 (OHF) and improved r by 0.01 (physics-free NN model) to 0.1 (OHF). Additionally, the BrTHF model showed a slight overestimation of LHF (BIAS = 0.14 W/m^2), lower than that of the SeaFlux, MERRA2, and ERA5 products. In contrast, the remaining products (JOFURO3, IFREMER, OAFlux, and OHF), along with the physics-free NN models, underestimated LHF, with the BIAS values ranging from -10.19 W/m^2 (OHF) to -1.61 W/m^2 (JOFURO3).

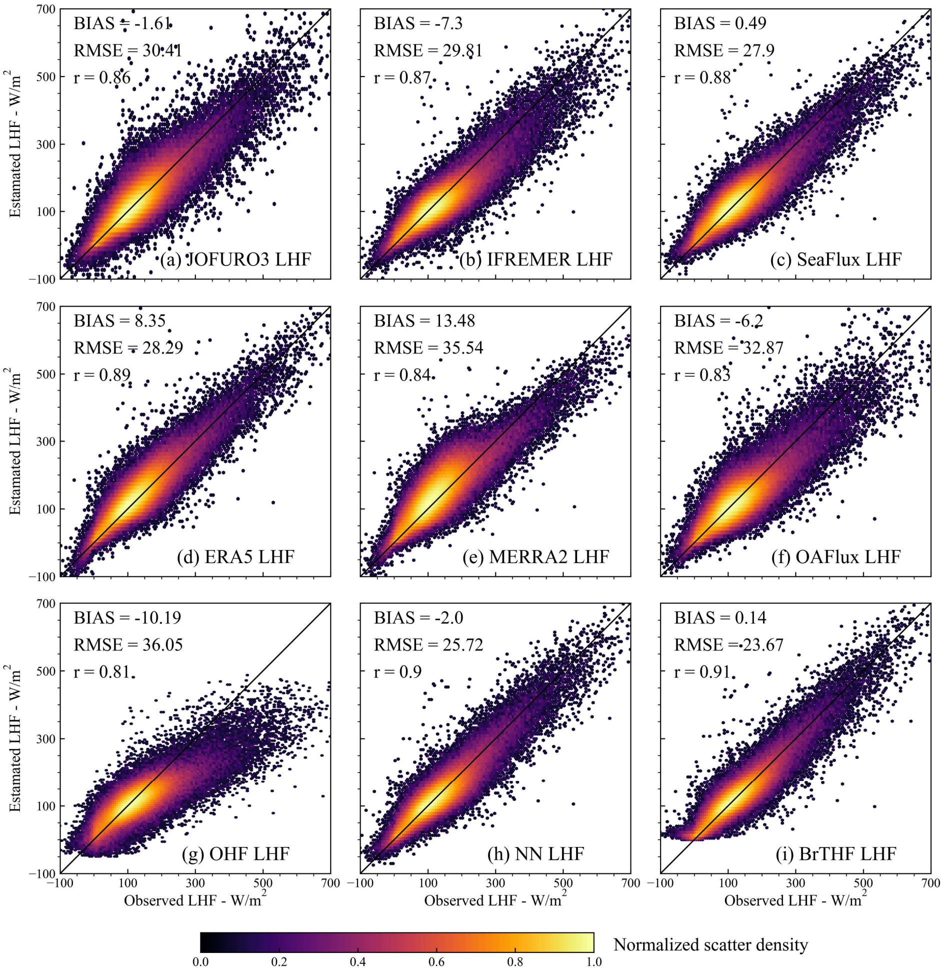


Figure 4. Same as Figure 3 but for LHF.

The BrTHF model exhibited a significantly different distribution of β compared to the physics-free NN models and the seven products, as shown in Figure 5. The β

estimates from the BrTHF model consistently fell within the observed range of -5 to 5, while the physics-free NN model and the seven products occasionally produced estimates outside this range. Specifically, approximately 0.9% of β estimates from both the physics-free NN model and the seven products were out of range. The extreme positive and negative β estimates were found in the OHF (β = 14997) and physics-free NN models (β = -25703) products, respectively. The abnormal β estimates significantly impacted the accuracy of the physics-free NN models and the seven products as Figure 6 indicated. When excluding the abnormal β samples from the physics-free NN models and seven products, the RMSEs ranged from 0.17 (physics-free NN models and SeaFlux) and 0.26 (OHF), with values of r ranging from 0.13 (OHF) to 0.46 (IFREMER), as shown in Figure 6 and Table S3. However, when all estimates were considered, the performances of these model and products deteriorated sharply, with RMSEs rising from 0.87 (SeaFlux) to 39.21 (physics-free NN models), and values of r dropping from 0.06 (SeaFlux) to 0 (JOFURO3, MERRA2 and OHF). In contrast, the BrTHF model maintained robust outperformance, with the lowest RMSEs of 0.22 and 0.15, and higher r values of 0.25 and 0.43, both before and after removing the abnormal β samples from the physics-free NN models and the seven products. Notably, the BIAS values remained stable (ranging from -0.04 to 0.04) for all models and products, regardless of whether the abnormal samples were excluded.

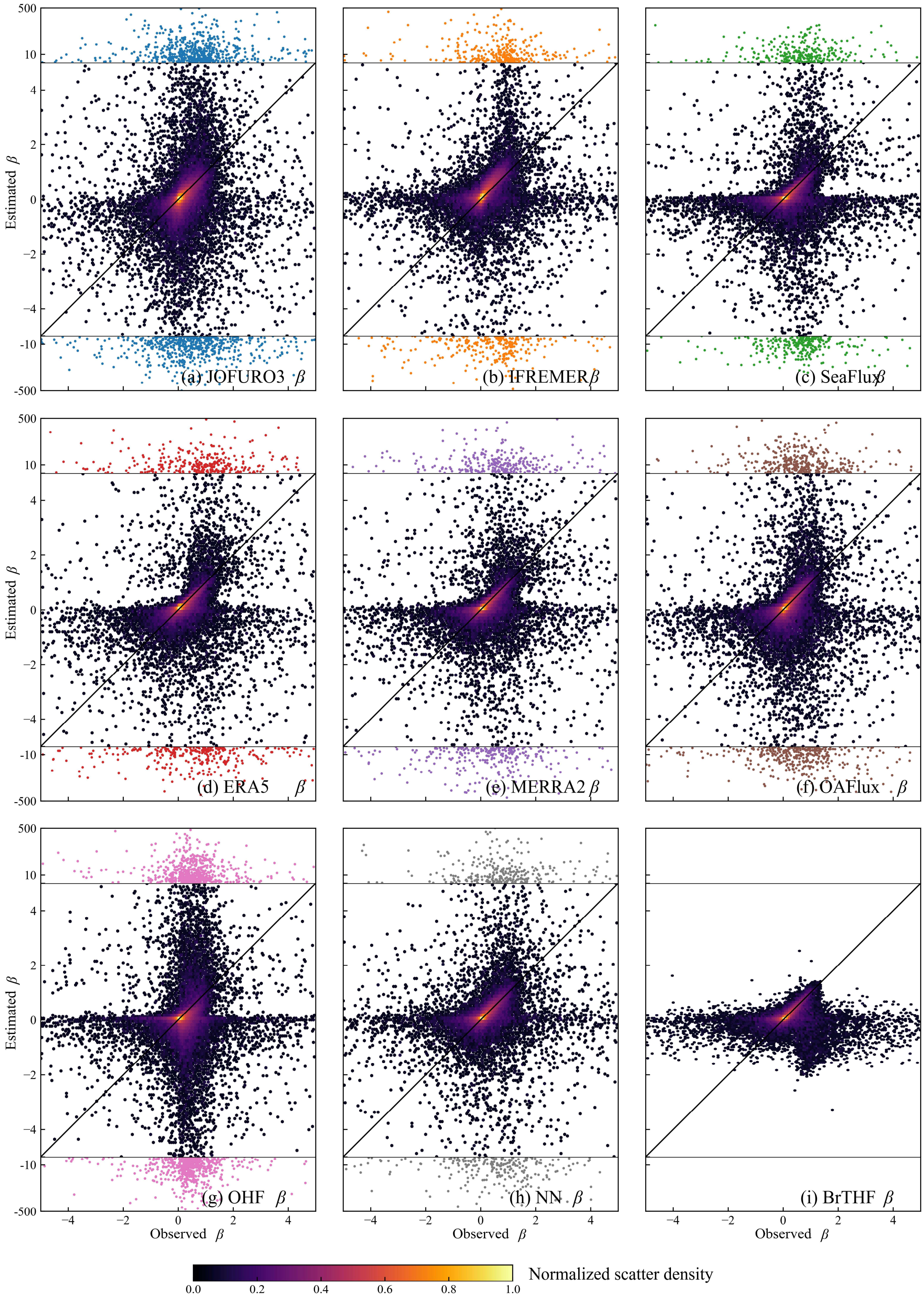


Figure 5. Same as Figure 3 but for *β*. The samples out of the ranges of observed *β* (-5 ≤ *β* ≤ 5) were colored in blue, orange, green, red, purple, brown, pink and gray for JOFURO3, IFREMER, SeaFlux, ERA5, MERRA2, OAFlux, OHF products and the physics-free NN models, respectively. The statistical metrics could be found in Table S3 and Figure 6.

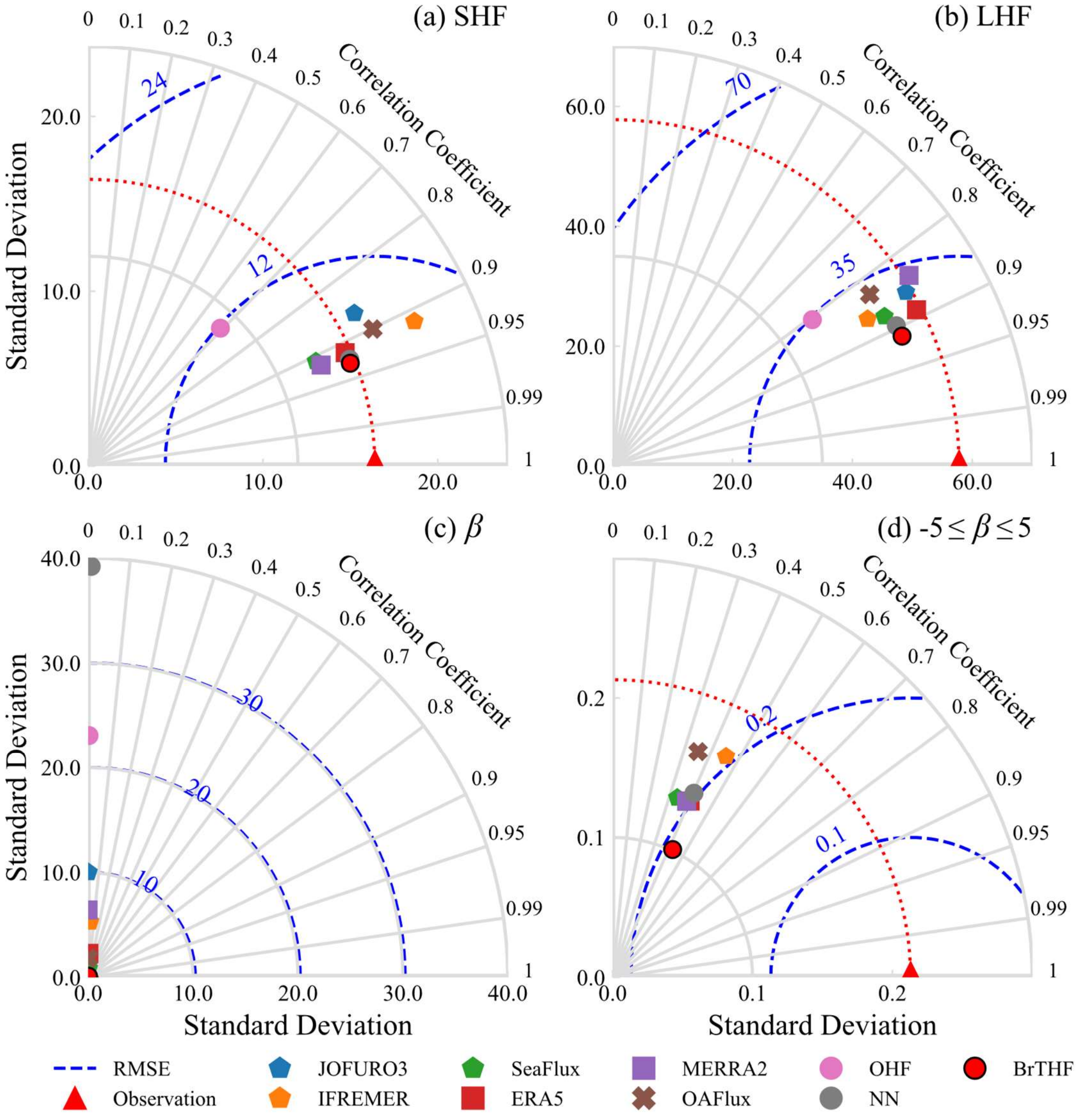


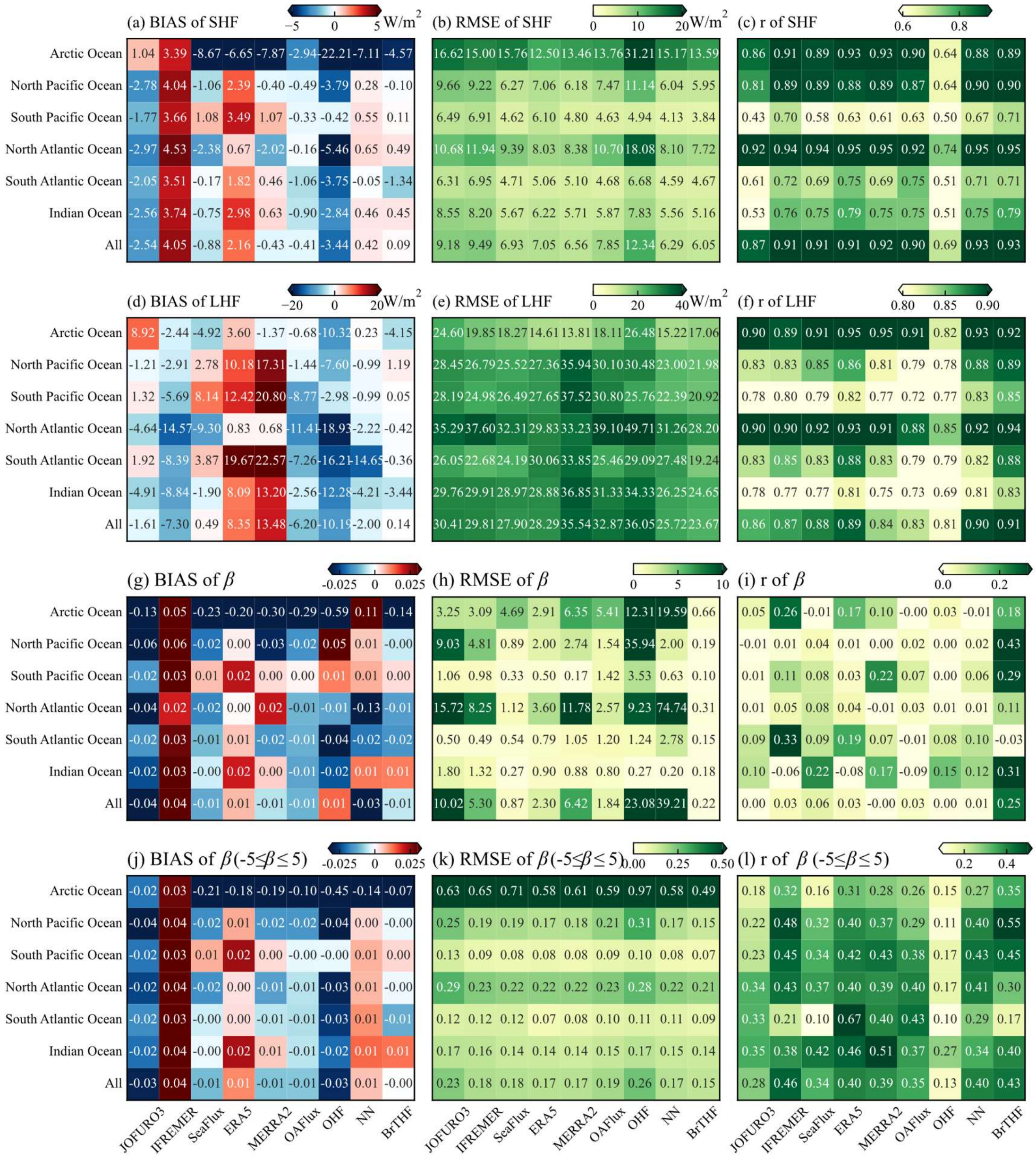
Figure 6. Taylor diagrams of the validation of estimated daily SHF (a), LHF (b), β (c) and β (-5≤β≤5, d) from the BrTHF model, the physics-free NN models and the seven products against the in-situ observations.

3.1.2 Accuracies across oceans

To better understand the accuracy of SHF, LHF and β estimates from the BrTHF and physics-free NN models, as well as the seven products in different oceans, we conducted an additional evaluation by categorizing the observations according to the belonging ocean basins, as shown in Figure 7. The major ocean boundaries, obtained from Marine Regions (https://www.marineregions.org/), were used to define the ocean basins, which include the Arctic Ocean, South Pacific Ocean, North Pacific Ocean, South Atlantic Ocean, North Atlantic Ocean, and Indian Ocean.

For SHF, the BrTHF model exhibited overestimations in the South Pacific Ocean, North Atlantic Ocean, and Indian Ocean, while it underestimated SHF in the remaining three ocean basins. The values of BIAS ranged from -4.57 W/m^2 in the Arctic Ocean to 0.49 W/m^2 in the North Atlantic Ocean. Furthermore, the BrTHF achieved the lowest RMSEs in most ocean basins, ranging from 3.84 W/m^2 in the South Atlantic Ocean to 7.72 W/m^2 in the North Atlantic Ocean, except in the Arctic Ocean where the RMSE of 13.59 W/m^2 was higher than those of the ERA5 (12.5 W/m^2) and MERRA2 (13.46 W/m^2) products, as shown in Figure 7(b). Correlation analysis also demonstrated the robust performance of the BrTHF model in estimating SHF, with values of r exceeding 0.89 in most ocean basins, except for ocean basins in the Southern Hemisphere (ranging from 0.71 to 0.79) where the values of r for all models and products were reduced.

For LHF, the values of BIAS of the BrTHF model ranged from -4.15 W/m^2 in the Arctic Ocean to 1.19 W/m^2 in the North Pacific Ocean. In comparison, the BrTHF model showed more pronounced underestimations in the Arctic Ocean and Indian Ocean. Additionally, the BrTHF model outperformed the physics-free NN models and the seven products across most ocean basins, achieving the lowest RMSEs (ranging from 17.06 W/m^2 in the Arctic Ocean to 28.20 W/m^2 in the North Atlantic Ocean) and the highest values of r (ranging from 0.83 in the Indian Ocean to 0.94 in the North Atlantic Ocean) except for the Arctic Ocean where the value of r was 0.01 less than the physics-free NN models and the RMSE were 2.45 W/m^2, 3.25 W/m^2 and 1.84 W/m^2 higher than the ERA5 and MERRA2 products and the physics-free NN models, respectively. Combining the overall and regional evaluations, we find that, except for the OHF product, the relative performance of SHF and LHF among the seven products is generally consistent with Tang et al. (2024). The OHF product shows notable degradation, particularly in high-latitude oceans, which may be due to limitations in the training datasets (sparse and unevenly distributed in-situ observations) and model training strategy (randomly splitting all observations into training, validation, and test sets without accounting for spatial dependencies), leading to reduced transferability.The BrTHF model consistently performed better in estimating β across most ocean basins, both before and after removing the abnormal β samples that deviated from the observed range ($-5 \leq \beta \leq 5$). In contrast, the physics-free NN models and the seven products did not perform as well. Specifically, the BrTHF model exhibited the lowest RMSEs in almost all ocean basins except in the South Atlantic Ocean after removing β outliers. Moreover, in terms of correlation analysis, the BrTHF model achieved higher values of r in most ocean basins before and after the removal of abnormal β samples, among all models and products.



(a) BIAS of SHF

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	1.04	3.39	-8.67	-6.65	-7.87	-2.94	-22.21	-7.11	-4.57
North Pacific Ocean	-2.78	4.04	-1.06	2.39	-0.40	-0.49	-3.79	0.28	-0.10
South Pacific Ocean	-1.77	3.66	1.08	3.49	1.07	-0.33	-0.42	0.55	0.11
North Atlantic Ocean	-2.97	4.53	-2.38	0.67	-2.02	-0.16	-5.46	0.65	0.49
South Atlantic Ocean	-2.05	3.51	-0.17	1.82	0.46	-1.06	-3.75	-0.05	-1.34
Indian Ocean	-2.56	3.74	-0.75	2.98	0.63	-0.90	-2.84	0.46	0.45
All	-2.54	4.05	-0.88	2.16	-0.43	-0.41	-3.44	0.42	0.09

(b) RMSE of SHF

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	16.62	15.00	15.76	12.50	13.46	13.76	31.21	15.17	13.59
North Pacific Ocean	9.66	9.22	6.27	7.06	6.18	7.47	11.14	6.04	5.95
South Pacific Ocean	6.49	6.91	4.62	6.10	4.80	4.63	4.94	4.13	3.84
North Atlantic Ocean	10.68	11.94	9.39	8.03	8.38	10.70	18.08	8.10	7.72
South Atlantic Ocean	6.31	6.95	4.71	5.06	5.10	4.68	6.68	4.59	4.67
Indian Ocean	8.55	8.20	5.67	6.22	5.71	5.87	7.83	5.56	5.16
All	9.18	9.49	6.93	7.05	6.56	7.85	12.34	6.29	6.05

(c) r of SHF

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	0.86	0.91	0.89	0.93	0.93	0.90	0.64	0.88	0.89
North Pacific Ocean	0.81	0.89	0.89	0.88	0.89	0.87	0.64	0.90	0.90
South Pacific Ocean	0.43	0.70	0.58	0.63	0.61	0.63	0.50	0.67	0.71
North Atlantic Ocean	0.92	0.94	0.94	0.95	0.95	0.92	0.74	0.95	0.95
South Atlantic Ocean	0.61	0.72	0.69	0.75	0.69	0.75	0.51	0.71	0.71
Indian Ocean	0.53	0.76	0.75	0.79	0.75	0.75	0.51	0.75	0.79
All	0.87	0.91	0.91	0.91	0.92	0.90	0.69	0.93	0.93

(d) BIAS of LHF

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	8.92	-2.44	-4.92	3.60	-1.37	-0.68	-10.32	0.23	-4.15
North Pacific Ocean	-1.21	-2.91	2.78	10.18	17.31	-1.44	-7.60	-0.99	1.19
South Pacific Ocean	1.32	-5.69	8.14	12.42	20.80	-8.77	-2.98	-0.99	0.05
North Atlantic Ocean	-4.64	-14.57	-9.30	0.83	0.68	-11.41	-18.93	-2.22	-0.42
South Atlantic Ocean	1.92	-8.39	3.87	19.67	22.57	-7.26	-16.21	-14.65	-0.36
Indian Ocean	-4.91	-8.84	-1.90	8.09	13.20	-2.56	-12.28	-4.21	-3.44
All	-1.61	-7.30	0.49	8.35	13.48	-6.20	-10.19	-2.00	0.14

(e) RMSE of LHF

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	24.60	19.85	18.27	14.61	13.81	18.11	26.48	15.22	17.06
North Pacific Ocean	28.45	26.79	25.52	27.36	35.94	30.10	30.48	23.00	21.98
South Pacific Ocean	28.19	24.98	26.49	27.65	37.52	30.80	25.76	22.39	20.92
North Atlantic Ocean	35.29	37.60	32.31	29.83	33.23	39.10	49.71	31.26	28.20
South Atlantic Ocean	26.05	22.68	24.19	30.06	33.85	25.46	29.09	27.48	19.24
Indian Ocean	29.76	29.91	28.97	28.88	36.85	31.33	34.33	26.25	24.65
All	30.41	29.81	27.90	28.29	35.54	32.87	36.05	25.72	23.67

(f) r of LHF

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	0.90	0.89	0.91	0.95	0.95	0.91	0.82	0.93	0.92
North Pacific Ocean	0.83	0.83	0.85	0.86	0.81	0.79	0.78	0.88	0.89
South Pacific Ocean	0.78	0.80	0.79	0.82	0.77	0.72	0.77	0.83	0.85
North Atlantic Ocean	0.90	0.90	0.92	0.93	0.91	0.88	0.85	0.92	0.94
South Atlantic Ocean	0.83	0.85	0.83	0.88	0.83	0.79	0.79	0.82	0.88
Indian Ocean	0.78	0.77	0.77	0.81	0.75	0.73	0.69	0.81	0.83
All	0.86	0.87	0.88	0.89	0.84	0.83	0.81	0.90	0.91

(g) BIAS of β

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	-0.13	0.05	-0.23	-0.20	-0.30	-0.29	-0.59	0.11	-0.14
North Pacific Ocean	-0.06	0.06	-0.02	0.00	-0.03	-0.02	0.05	0.01	-0.00
South Pacific Ocean	-0.02	0.03	0.01	0.02	0.00	0.00	0.01	0.01	0.00
North Atlantic Ocean	-0.04	0.02	-0.02	0.00	0.02	-0.01	-0.01	-0.13	-0.01
South Atlantic Ocean	-0.02	0.03	-0.01	0.01	-0.02	-0.01	-0.04	-0.02	-0.02
Indian Ocean	-0.02	0.03	-0.00	0.02	0.00	-0.01	-0.02	0.01	0.01
All	-0.04	0.04	-0.01	0.01	-0.01	-0.01	0.01	-0.03	-0.01

(h) RMSE of β

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	3.25	3.09	4.69	2.91	6.35	5.41	12.31	19.59	0.66
North Pacific Ocean	9.03	4.81	0.89	2.00	2.74	1.54	35.94	2.00	0.19
South Pacific Ocean	1.06	0.98	0.33	0.50	0.17	1.42	3.53	0.63	0.10
North Atlantic Ocean	15.72	8.25	1.12	3.60	11.78	2.57	9.23	74.74	0.31
South Atlantic Ocean	0.50	0.49	0.54	0.79	1.05	1.20	1.24	2.78	0.15
Indian Ocean	1.80	1.32	0.27	0.90	0.88	0.80	0.27	0.20	0.18
All	10.02	5.30	0.87	2.30	6.42	1.84	23.08	39.21	0.22

(i) r of β

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	0.05	0.26	-0.01	0.17	0.10	-0.00	0.03	-0.01	0.18
North Pacific Ocean	-0.01	0.01	0.04	0.01	0.00	0.02	0.00	0.02	0.43
South Pacific Ocean	0.01	0.11	0.08	0.03	0.22	0.07	0.00	0.06	0.29
North Atlantic Ocean	0.01	0.05	0.08	0.04	-0.01	0.03	0.01	0.01	0.11
South Atlantic Ocean	0.09	0.33	0.09	0.19	0.07	-0.01	0.08	0.10	-0.03
Indian Ocean	0.10	-0.06	0.22	-0.08	0.17	-0.09	0.15	0.12	0.31
All	0.00	0.03	0.06	0.03	-0.00	0.03	0.00	0.01	0.25

(j) BIAS of $\beta(-5\leq\beta\leq 5)$

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	-0.02	0.03	-0.21	-0.18	-0.19	-0.10	-0.45	-0.14	-0.07
North Pacific Ocean	-0.04	0.04	-0.02	0.01	-0.02	-0.02	-0.04	0.00	-0.00
South Pacific Ocean	-0.02	0.03	0.01	0.02	0.00	-0.00	-0.00	0.01	0.00
North Atlantic Ocean	-0.02	0.04	-0.02	0.00	-0.01	-0.01	-0.03	0.01	-0.00
South Atlantic Ocean	-0.02	0.03	-0.00	0.00	-0.01	-0.01	-0.03	0.01	-0.01
Indian Ocean	-0.02	0.04	-0.00	0.02	0.01	-0.01	-0.02	0.01	0.01
All	-0.03	0.04	-0.01	0.01	-0.01	-0.01	-0.03	0.01	-0.00

(k) RMSE of $\beta(-5\leq\beta\leq 5)$

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	0.63	0.65	0.71	0.58	0.61	0.59	0.97	0.58	0.49
North Pacific Ocean	0.25	0.19	0.19	0.17	0.18	0.21	0.31	0.17	0.15
South Pacific Ocean	0.13	0.09	0.08	0.08	0.08	0.09	0.10	0.08	0.07
North Atlantic Ocean	0.29	0.23	0.22	0.22	0.22	0.23	0.28	0.22	0.21
South Atlantic Ocean	0.12	0.12	0.12	0.07	0.08	0.10	0.11	0.11	0.09
Indian Ocean	0.17	0.16	0.14	0.14	0.14	0.15	0.17	0.15	0.14
All	0.23	0.18	0.18	0.17	0.17	0.19	0.26	0.17	0.15

(l) r of β $(-5\leq\beta\leq 5)$

	JOFURO3	IFREMER	SeaFlux	ERA5	MERRA2	OAFlux	OHF	NN	BrTHF
Arctic Ocean	0.18	0.32	0.16	0.31	0.28	0.26	0.15	0.27	0.35
North Pacific Ocean	0.22	0.48	0.32	0.40	0.37	0.29	0.11	0.40	0.55
South Pacific Ocean	0.23	0.45	0.34	0.42	0.43	0.38	0.17	0.43	0.45
North Atlantic Ocean	0.34	0.43	0.37	0.40	0.39	0.40	0.17	0.41	0.30
South Atlantic Ocean	0.33	0.21	0.10	0.67	0.40	0.43	0.10	0.29	0.17
Indian Ocean	0.35	0.38	0.42	0.46	0.51	0.37	0.27	0.34	0.40
All	0.28	0.46	0.34	0.40	0.39	0.35	0.13	0.40	0.43

Figure 7. Heatmaps of BIAS, RMSE and r metrics for the validation of estimated daily SHF (a - c), LHF (b - e), β (f - i) and β ($-5 \leq \beta \leq 5$, j - l) from the BrTHF model, the physics-free NN models and the seven products against the in-situ observations across different ocean basins. It should be noted that the statistical metrics for each ocean basin were calculated using observations from the available buoys within the corresponding basin.

3.2 Temporal variations in SHF, LHF and β

After spatial ten-fold cross-validation, we produced the daily 0.25° global air-sea turbulent heat flux products from 1993 to 2017 using a combination of the BrTHF model and learning datasets, and further made a comparison of the temporal variation (in this section), spatial distribution (in Section 3.3) and annual trend (in Section 3.4) of SHF, LHF and β estimates from the BrTHF product and those with the seven state-of-the-art global products. The selected period (from 1993 to 2017) was determined by the overlapping availability of input learning datasets.

Figure 8 illustrates the monthly area-weighted global means of SHF, LHF and β from 1993 to 2017 for the BrTHF product and seven state-of-the-art products. The BrTHF product exhibited similar bimodal patterns for SHF, LHF and β as the seven products, with peaks in December-January and May-June-July-August. In addition, the peak in May-June-July-August was less pronounced for SHF and β compared to that for LHF. The monthly area-weighted global means of SHF and β from the BrTHF product were higher than those of most products, except for the MERRA2 product in January, February, March, April, July, August and September, and the IFREMER product in all months. For LHF, the BrTHF showed lower values than the ERA5 and MERRA2 products across all months. Notably, the patterns of SHF and β from the OHF product, with the highest peak occurring in August and smoother intra-annual cycles, differed from those of the corresponding BrTHF product and the other six products developed using the bulk aerodynamic methods.

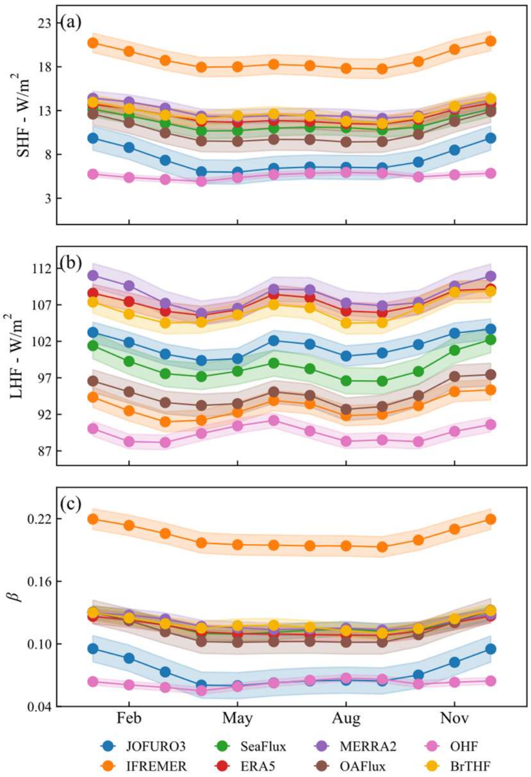


Figure 8. Intra-annual cycles of area-weighted global monthly mean of SHF (a), LHF (b) and *β* (c) from the eight products from 1993 to 2017. The shaded areas indicate ±1 standard deviation around the mean.

Figure 9 presents the temporal evolution of the area-weighted annual global mean of SHF, LHF and *β* from 1993 to 2017 for the eight products over ice-free oceans. The global mean annual SHF of the BrTHF product was 12.7 W/m^2, which was close to those of SeaFlux (11.6 W/m^2), OAFlux (10.6 W/m^2), MERRA2 (13 W/m^2) and ERA5 (12.4 W/m^2), whereas it significantly lower than that of IFREMER (18.8 W/m^2) and higher than those of JOFURO3 (7.5 W/m^2) and OHF (5.6 W/m^2). Meanwhile, the BrTHF product exhibited the largest significant increase of SHF with the trend of 0.04 W/(m^2·year) among all eight products, and showed similar temporal evolution as SeaFlux, MERRA2, ERA5 and OAFlux during the period from 1993 to 2017. As for LHF, the BrTHF exhibited a relatively high global mean annual value of 106.2 W/m^2, which was close to those of the ERA5(107.3 W/m^2) and MERRA2 (108.3 W/m^2), and it was significantly higher than the rest of the products. Moreover, the growth of the BrTHF LHF was significant with a trend of 0.33 W/(m^2·year), which was lower than the IFREMER but higher than the OAFlux, MERRA2, OHF, ERA5, JOFURO3 and SeaFlux, ranging from -0.14 W/(m^2·year) to 0.4 W/(m^2·year). Note that only the OAFlux product showed a negative trend of LHF from 1993 to 2017. For *β*, the BrTHF showed a similar temporal pattern to that of SHF, and most products concentrated within the narrow range of 0.11 to 0.12 for the annual values. The magnitude of annual *β* of the BrTHF was about 0.11, which was close to the OAFlux, SeaFlux, MERRA2 and ERA5, but significantly lower than the IFREMER and higher than the JOFURO3 and OHF. Moreover, in contrast to the significant increasing trends of LHF and SHF, negative trends of *β* were shown for most products. However, the BrTHF product exhibited a weak positive trend, which may be attributed to the relatively smaller differences between the SHF and LHF trends in BrTHF compared to those in other products.

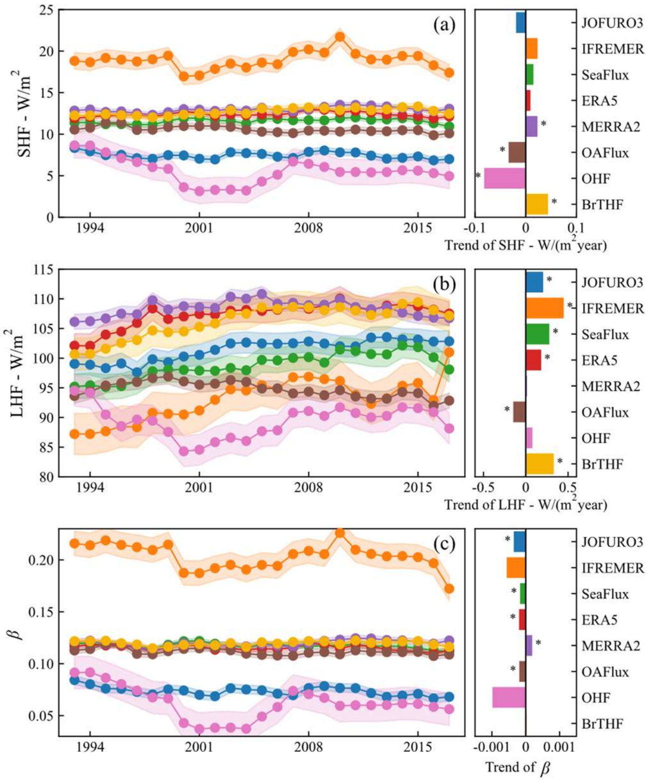


Figure 9. Inter-annual evolution of area-weighted global mean SHF (a - b), LHF (c - d) and *β* (e - f) from 1993 to 2017. The trends were calculated based on the Sen's slope method. The * in the sub-figures (b, d and f) represent the trend passed the Mann-Kendall significant test ($p < 0.05$). The shaded areas indicate ±1 standard deviation around the mean.

3.3 Inter-comparison of the spatial pattern

We selected three representative products including the (reanalysis-based) ERA5, (remote sensing-based) SeaFlux, and (the only publicly available machine learning-based) OHF products to evaluate the BrTHF product's ability in simulating global air-sea turbulent heat fluxes (SHF, LHF, and β) from 1993 to 2017. These products were chosen because they demonstrated relatively high accuracy within their respective categories (as shown in Section 3.1) and shared the same 0.25° spatial resolution with the BrTHF product.

Figure 10 presents the spatial distribution of multi-year mean of SHF from the SeaFlux, ERA5, BrTHF, and OHF products, along with their cross-comparisons. Overall, the BrTHF product exhibited strong consistency with ERA5 and SeaFlux products, with values of r exceeding 0.88, which was significantly higher than the consistency between SeaFlux and OHF (r = 0.33) and between ERA5 and OHF (r = 0.37). Spatially, the BrTHF, SeaFlux and ERA5 products all showed higher SHF (over 50 W/m^2) in the Western Boundary Currents (WBCs, e.g. Kuroshio, Gulf Stream, Brazil Current and Agulhas Current) regions, whereas the OHF product yielded much lower SHF (~25 W/m^2). Additionally, the former three products captured pronounced SHF gradients in the Southern Ocean, features that were absent in the OHF product. SHF differences between BrTHF and SeaFlux/ERA5 remained within ±10 W/m^2 in most oceans. The BrTHF product exhibited slightly higher SHF values than SeaFlux in the Northern Hemisphere, whereas in the Southern Hemisphere—particularly over the Southern Ocean—the BrTHF showed relatively lower SHF. Compared to the ERA5 product, the BrTHF product yielded lower SHF in the equatorial zone, subtropical high-pressure regions and the Southern Ocean, but higher SHF in other areas, particularly in the North Pacific and the southern Indian Ocean.

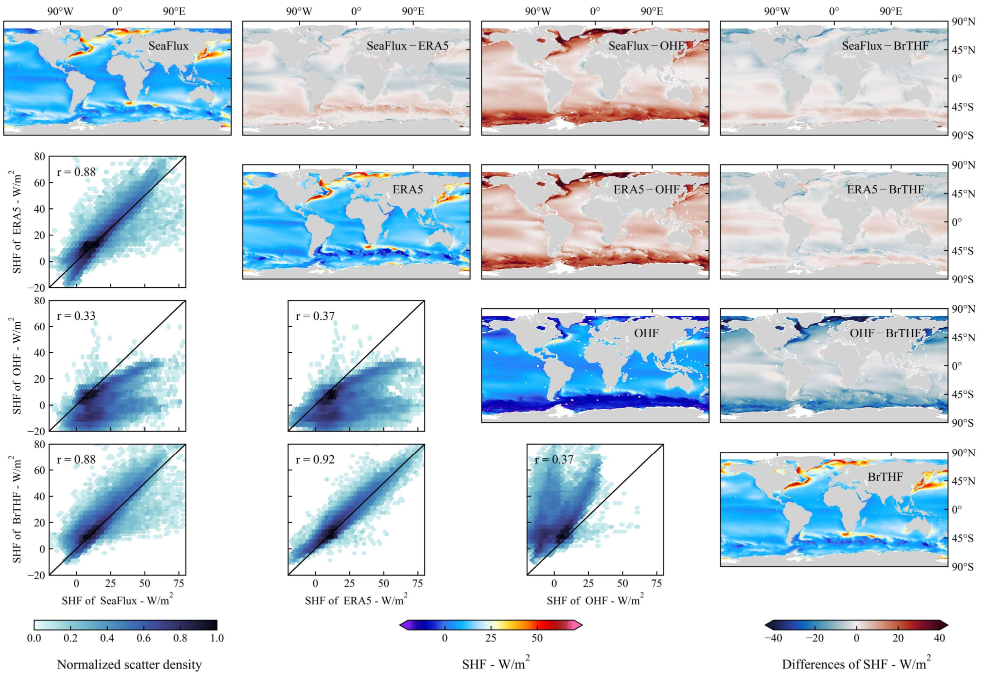


Figure 10. Inter-comparison of the spatial distributions of multi-year means of SHF among the SeaFlux, ERA5, OHF and BrTHF products from 1993 to 2017.

For LHF, the BrTHF and the three reference products exhibited more similar spatial distribution patterns, with the values of r exceeding 0.98, compared to the results for the SHF, as shown in Figure 11. The higher LHF (over 150 W/m^2) primarily occurred around the regions of WBCs and the sub-tropic highs, while lower LHF (below 50 W/m^2) appeared in the Eastern Equatorial Pacific and Atlantic Warm Tongue and the oceans with latitudes higher than 45°. The spatial distribution of LHF in the BrTHF product generally agreed better with that of the ERA5 product, though the BrTHF showed significantly lower LHF in sub-tropic highs. Additionally, the BrTHF exhibited relatively lower LHF than the ERA5 over the Southern Ocean and the central North Atlantic. Compared to the SeaFlux, the BrTHF yielded slightly higher LHF in most oceans except the Southern Ocean and equatorial zones.

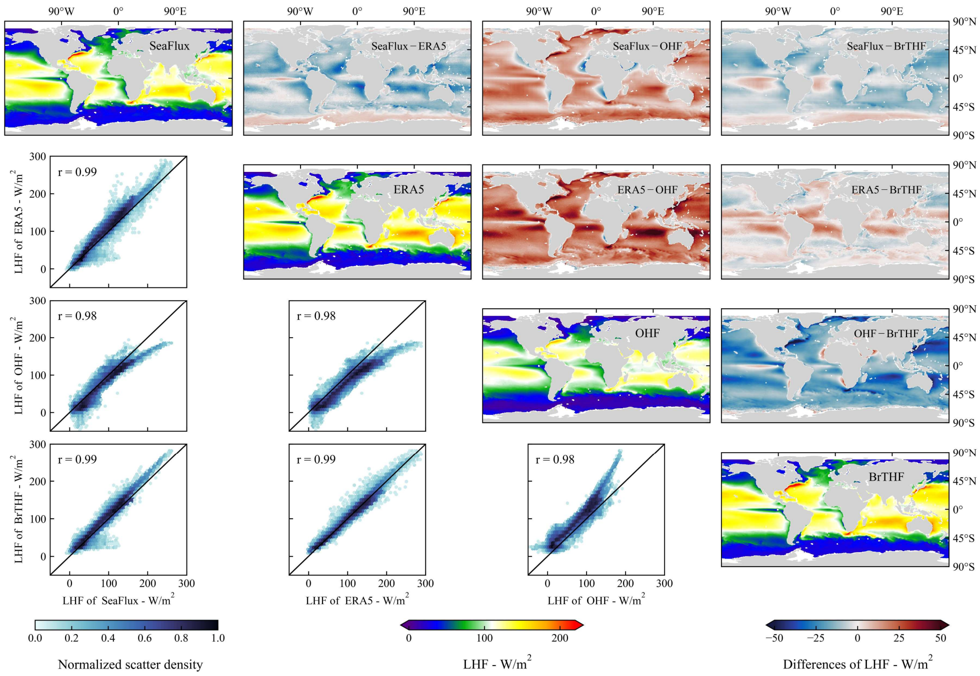


Figure 11. Same as Figure 10 but for LHF.

For β, the BrTHF product demonstrated strong spatial correlation with the ERA5 and SeaFlux in multi-year mean distributions, with values of r exceeding 0.81. In contrast, the OHF showed a markedly different spatial pattern of β, exhibiting negative correlations when compared to the other three products. Spatially, the BrTHF product's β distribution aligned more closely with the SeaFlux, both displaying higher β (up to 1) in high-latitude oceans particularly in the Northern Hemisphere and the similar wavelike textures of β over the Southern Ocean's Antarctic Circumpolar Current zone. The differences between the BrTHF and OHF products were more evident. Specifically, the BrTHF product showed overall overestimation of β in the oceans where latitudes were larger than 45° compared to the OHF product.

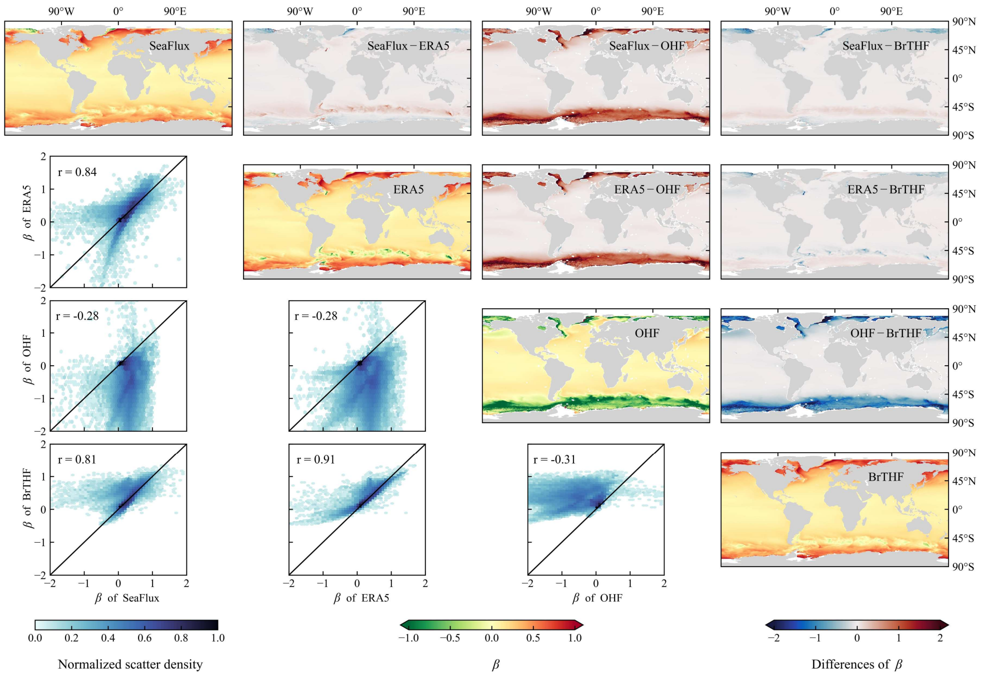


Figure 12. Same as Figure 10 but for β.

3.4 Spatial pattern of trends in SHF, LHF and β from the BrTHF product

Figure 13 illustrates the spatial distribution of inter-annual trends of SHF, LHF and β in the BrTHF product from 1993 to 2017. The SHF showed increasing trends across 71.4% of the oceans, with statistically significant increases in 26.2% of regions. In contrast, decreasing trends were observed in 28.6% of the oceans, with only 3% showing significant reductions. Overall, the trends of zonal annual averages of SHF remained stable between 60°N to 45°S, with significant increases occurring southward and decreases northward. Specifically, moderate increases (~0.2 W/(m^2 year)) dominated between 45°N and 45°S, while more pronounced increases (>0.8 W/(m^2 year)) were observed in mid- and high-latitude oceans, including the Kara Sea, Gulf Stream, Baffin Bay, Brazil Current, Sea of Okhotsk, and Sea of Japan. Notable decreases (< -0.8 W/(m^2 year)) were concentrated in the Barents Sea and the central North Atlantic.

The LHF exhibited markedly different characteristics of the spatial distribution, with 87.5% of oceans showing increasing trends (38.9% were significant), versus 12.5% decreasing (0.7% were significant). In contrast to those of the SHF, the trends of zonal annual averages for LHF weakened poleward from the oceans of Equator. The substantial increases (>0.6 W/(m^2 year)) occurred in the oceans between 45°N to 45°S, particularly in the Gulf Stream, Brazil Current, and Agulhas Current systems, while notable decreases (lower than -0.3 W/(m^2 year)) were observed in the central North Atlantic and Kuroshio extension regions.

For β, approximately 53% of the oceans showed increasing trends, with 12.2% of these being statistically significant. Conversely, about 47% of the oceans showed decreasing trends, with 9.8% being significant. Most oceans between 45°N to 45°S exhibited near-zero trends, while significant trends were concentrated in the high-latitude oceans. Notable increases were found in Baffin Bay, Kara Sea, and the Southern Ocean, while decreases were observed in the Barents Sea and the Southern Ocean near South America.

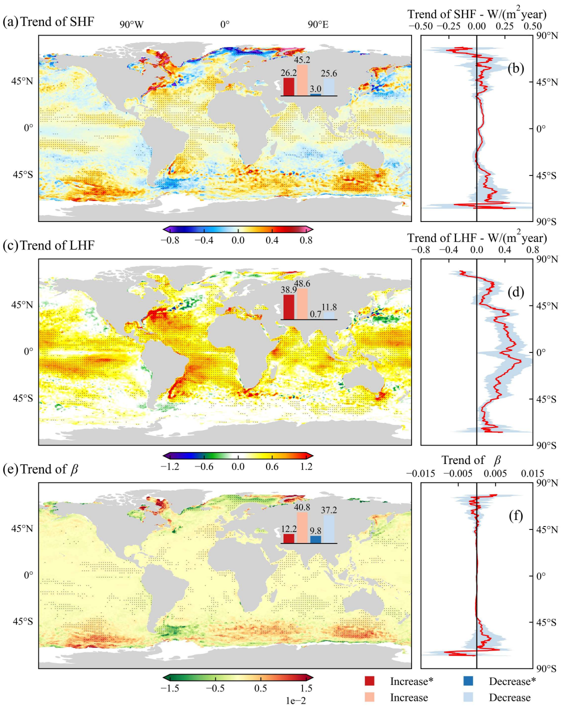


Figure 13. Spatial maps of inter-annual trends for SHF (a), LHF (c), and β (e) from the BrTHF product for the period 1993 to 2017. The trends were calculated using the Sen's slope method. Dotted areas indicate oceans where the p-value of the Mann-Kendall significance test is less than 0.05. Panels (b), (d) and (f) represent the inter-annual trends of zonal annual averages for SHF, LHF and β, respectively.

3.5 Discussion

Advancing our understanding of the air-sea interaction and achieving the global closure of the ocean surface energy budget require accurate global-scale simulations of air-sea turbulent heat fluxes (Yu, 2019). Existing global air-sea turbulent heat flux products, primarily generated using the semi-empirical bulk aerodynamic methods and data-driven machine learning approach, are often weak in accuracy and physical rationality, arising from uncertainties in environmental forcings and inappropriate parameterizations (Brodeau et al., 2017; Jiang et al., 2024a; Wang et al., 2024). To improve the simulation of the global air-sea turbulent heat fluxes, this study presents the BrTHF product, generated using a Bowen ratio-constrained NN technique with a custom multiple-objective loss function, as well as observations from 197 globally distributed buoys along with multi-source remote sensing and reanalysis inputs.

3.5.1 Advantages

The primary advantage of the BrTHF model lies in its accurate estimation of β, which shows the most pronounced improvement among all flux components. As a key indicator of surface energy partitioning, β is widely used within the surface energy balance framework to ensure physically consistent and reliable estimates of SHF and LHF (Yang et al., 2025). In addition, β serves as an effective diagnostic variable in the studies of large-scale climate variability (e.g., ENSO) (Jo, 2002) and in investigations of how surface energy constraints regulate the hydrological cycle (e.g., precipitation) (Wang et al., 2021). With its enhanced representation of β, the BrTHF product is expected to provide more reliable support for such applications.

To achieve these improvements in β and flux, the BrTHF model was designed differently from our previous study (Wang et al., 2024), which simultaneously predicted SHF, LHF and β in the constructed RF model. Specifically, this study employed an NN model constrained by the Bowen ratio to jointly estimate SHF and LHF. The new approach avoided the issue of selection of β derived from either the calculated β [β_{cal} equals predicted SHF ($\mathrm{SHF}_{\mathrm{pre}}$) divided by predicted LHF ($\mathrm{LHF}_{\mathrm{pre}}$)] or the predicted β

(β_{pre}), as reported by Wang et al. (2024). Furthermore, the custom loss function in our new approach provides a flexible approach to adjust the weights of SHF, LHF, and β, allowing the model to balance attention among these variables. As a result, the accuracy of SHF, LHF, and β from our newly developed BrTHF model outperformed that of the mainstream air-sea turbulent heat flux products and the physics-free NN models on both global and regional scales. In contrast, the accuracy of SHF and LHF in the model constructed by Wang et al. (2024) was somewhat marginally lower than that of the physics-free RF model.

3.5.2 Generalizability

Based on Figure 2 and Table S8, we observe that the spatial coverage of observations varies across different ocean regions: the Northern Hemisphere generally has higher coverage than the Southern Hemisphere, with the Northern Pacific Ocean exhibiting the highest coverage, while the Arctic Ocean shows the lowest. Comparing spatial coverage with accuracy metrics reveals a more complex relationship between model performance and data coverage. Specifically, the values of r tend to be lower in regions with lower coverage — a pattern consistent across SHF, LHF, and β. However, RMSE does not follow this trend. For SHF and β, RMSEs in the Northern Hemisphere are generally higher than those in the Southern Hemisphere. Similarly, for LHF, RMSEs are higher in the Northern Hemisphere except in the Indian Ocean, where the pattern differs.

We applied a spatial 10-fold cross-validation, which provides a more generalized assessment than traditional random cross-validation, to evaluate the BrTHF model. However, it is important to acknowledge that the spatial distribution of the training dataset is inherently imbalanced, with a heavy concentration of observations in the Tropics and the Northern Hemisphere. In contrast, the Southern Hemisphere—particularly the Southern Ocean—suffers from sparse or even missing observational coverage. Given that the environmental conditions in these underrepresented or data-sparse regions may differ significantly from those captured in the training dataset, the selected input variables for the observations may lead to large uncertainty in the model's performance in these areas. To further assess the model’s ability to extrapolate to such regions, we conducted an additional targeted cross-validation. Specifically, we excluded stations from the Southern Ocean [i.e., Southern Ocean Flux Station (SOFS) and Global Southern Ocean Station (GSOS)] from the training dataset and used them solely for validation. Results presented in Tables S4 and S5 show that the BrTHF model achieved the best performance in terms of LHF and β at the SOFS with lower RMSE of 15.6 W/m^2 and 0.73 and higher values of r of 0.96 and 0.34, respectively, while its SHF was slightly outperformed by ERA5 and the physics-free NN model. At the GSOS, BrTHF yielded more accurate estimates for SHF and β with RMSEs of 6.38 W/m^2 and 0.74 and values of r of 0.95 and 0.16, respectively, compared to other products, while its LHF was marginally less accurate than that of SeaFlux and the physics-free NN model. Moreover, under both spatially-informed cross-validation and targeted cross-validation, the model demonstrates comparable accuracy at the two sites, as shown in Figures S4–S7. These findings suggest that BrTHF retains competitive accuracy of SHF, LHF and β even in regions entirely excluded from training, reflecting promising generalization. While these results are encouraging, it is important to note that the validation remains limited to a small number of sites with available observations. Therefore, the reported r values and RMSE reflect model performance in these specific locations and do not necessarily guarantee similar accuracy in broader, unobserved ocean regions. Consequently, the BrTHF product should be viewed as being primarily optimized within the geographical coverage of existing buoy networks. In remote regions far from the observation-rich regions, such as the high-latitude Southern Ocean, the lack of direct ground-truth constraints may result in certain uncertainties. Users should therefore exercise caution when interpreting the global-scale performance, particularly in data-sparse basins where spatial sampling limitations are most pronounced.

The generalization capability of the model can also affect the accuracy of simulated long-term trends. In Figure 13, we present the spatial distributions of long-term trends for SHF, LHF, and β simulated by the BrTHF product. Considering the scarcity of training data in high-latitude oceans, the simulated long-term trends in these regions may be associated with larger uncertainties. However, due to the lack of long-term observations in high-latitude oceans, we cannot validate the simulated trends using observational records as has been done in previous studies for mid- and low-latitude regions (Weller et al., 2022; Tang et al., 2024). To address this, we examined the spatial distribution of long-term trends from the other seven widely used products. Specifically, in these high-latitude regions, the trends simulated by the BrTHF are largely consistent with those of most other products—for example, SHF exhibits a pronounced increase in the Kara Sea, Gulf Stream, Baffin Bay, Brazil Current, Sea of Okhotsk, and Sea of Japan, with differences mainly in magnitude.

3.5.3 Sensitivity and uncertainty

Given the key role of the β constraint in the BrTHF model, it is also important to assess the sensitivity of the estimated SHF and LHF to the imposed β range. A series of sensitivity experiments with progressively relaxed β constraints indicate that the BrTHF model exhibits weak sensitivity to the specific choice of the β range (see Table S9). The resulting variations in RMSE are small relative to their mean values, while values of r and BIAS remain largely unchanged across different β configurations. These results suggest that the improved performance of BrTHF is not driven by a particular predefined range of β, but instead reflects the robustness of the Bowen ratio–constrained machine learning framework.

Beyond the sensitivity to the imposed β range, we further examined the robustness of the BrTHF model to uncertainties in its environmental inputs, using one-at-a-time perturbation experiments (Table S10). The results indicate that the model is generally insensitive to perturbations (from ±5 to ±20%) in most auxiliary inputs, with changes in SHF and LHF RMSE typically remaining within 1%. Noticeable sensitivities are mainly associated with WS, $diff_Q$, and $diff_T$, which are physically expected given their direct roles in controlling air–sea turbulent heat exchanges. Even under extreme perturbations, the resulting variations in SHF and LHF remain within physically reasonable ranges, suggesting that the BrTHF model does not rely excessively on precise tuning of individual inputs. For β, most perturbations lead to minor changes; however, perturbations of $diff_Q$ can induce large variability due to the ratio-based nature of β, further highlighting the necessity of the β constraint.

To further address potential methodological inconsistencies between buoy-derived fluxes and flux products, we conducted an additional baseline experiment. Specifically, the COARE3.5 model was forced with the same subset of daily meteorological variables used to drive the BrTHF model, thereby providing a method-consistent reference under identical forcing conditions. The results (Table S12) show that the COARE3.5-driven estimates exhibit substantially larger errors for SHF and LHF, with RMSEs of approximately 10.3 W/m^2 and 34.4 W/m^2, respectively, accompanied by systematic underestimations relative to buoy observations. More importantly, physically unrealistic values of the β emerge in the COARE3.5 estimates, leading to a degradation in β accuracy (RMSE = 6.35). These findings suggest that, even when driven by the same meteorological inputs, traditional bulk formulations remain highly sensitive to forcing uncertainties. In contrast, the BrTHF model demonstrates improved robustness by explicitly incorporating physical constraints within the machine-learning framework.

Finally, we examined the potential impact of temporal aggregation, as the buoy-derived daily fluxes used for model training and evaluation were calculated from daily-averaged meteorological variables via COARE3.5, which may introduce biases due to the nonlinearity of bulk flux formulations. By recalculating fluxes from high-frequency buoy observations and aggregating to daily means, we found that although absolute errors differ to some extent, the relative performance among different flux products and models remains unchanged, with the BrTHF model still achieving the best overall accuracy (Table S11). This confirms that our conclusions are robust to the choice of daily flux calculation strategy.

3.5.4 Limitaions and recommendations

Although the results demonstrated significant improvements in the accuracy and physical consistency of SHF, LHF, and β estimates from the BrTHF model compared to those from the physics-free NN models and the seven products, the BrTHF product also has some limitations. First, extreme values of β may arise either from physically plausible but rare ocean conditions or from numerical instability or poorly constrained estimates associated with uncertainties in model-derived fluxes and input variables. In practice, these two sources are difficult to distinguish. To ensure robust model training and stable performance across the vast majority of β conditions (approximately the 1-99% of the β distribution), we applied a conservative β constraint (-5 to 5) during training. This compresses the standard deviation and constrains the extreme tails of the predicted β, but substantially improves model stability and accuracy for the majority of ocean conditions. Although this choice leads to a narrower dynamic range compared to some other products, it ensures that the 1-99% of the β distribution is well-represented for most practical applications. It should be noted that this strategy represents an interim solution rather than a final one. With future improvements in the quality and spatiotemporal representativeness of observational datasets, physically plausible extremes (e.g., 0-1% and 99-100% of the β distribution) could be better constrained and incorporated into model training, allowing expansion of the dynamic range of β. Secondly, the estimated SHF and LHF values exhibited a narrower distribution compared to the observations. This issue possibly stems from the uncertainty of the BrTHF model that was constructed from the uneven distribution of SHF and LHF in the observation datasets, which contain a low proportion of extreme samples, especially negative LHF values. Moreover, due to the insufficient observations, validation and modelling in high-latitude oceans, especially in the Southern Ocean, was limited. To address these problems, more experiments are highly recommended to collect observations covering more regions of oceans with better spatial and temporal representativeness, which could further enhance the product.

The BrTHF model demonstrated the feasibility and potential of jointly estimating multiple interrelated air-sea variables through a machine learning model that incorporates appropriate physical constraints to account for their interrelations. In the future, the predicted variables in the BrTHF model could be expanded to include surface radiation, heat storage, and precipitation over the ocean, by integrating the physical mechanisms of energy and water exchange. This would allow for the collaborative optimization of estimates across all components of the air-sea energy and water budgets, potentially contributing to achieving global closure of the air-sea energy and water budgets.

4. Data and code availability

The daily 0.25° BrTHF product, consisting of SHF and LHF estimates from 1993 to 2017, can be freely accessed from the National Tibetan Plateau Data Center (TPDC) [https://doi.org/10.11888/Atmos.tpdc.302578, Tang and Wang (2025)]. The code for developing the product can be found on the GitHub platform (https://github.com/zhezhe1996/BrTHF).

5. Summary and Conclusion

In this study, we generated a daily 0.25° air-sea turbulent heat fluxes product for the period from 1993 to 2017 using our developed BrTHF model and multi-source remote sensing and reanalysis data. A comprehensive validation was performed against observations from 197 buoys and inter-comparisons were made with seven representative gridded products. The key findings are as follows:

The BrTHF model demonstrated the most significant improvement in estimating the β, while its performance in estimating SHF and LHF was generally comparable to or slightly better than that of the physics-free NN models and the seven widely used air-sea turbulent heat products (including the JOFURO3, IFREMER, SeaFlux, ERA5, MERRA2, OAFlux and OHF products). Through the spatial ten-fold cross-validation against the observations from the 197 buoys, the BrTHF model achieved RMSEs of 6.05 W/m^2 for SHF, 23.67 W/m^2 for LHF and 0.22 for β, and showed values of r of 0.93, 0.91, and 0.25 for SHF, LHF, and β, respectively. Additionally, the BrTHF model performed better in evaluations across six major ocean basins, with lower RMSEs and higher values of r, in comparison to the physics-free NN models and the seven products. Notably, the BrTHF model significantly improved the rationality of β estimates, successfully eliminating the outliers observed in both the physics-free NN models and the seven products. Furthermore, the global distributions for SHF, LHF, and β from the BrTHF product closely matched those of the physically-based ERA5 and SeaFlux products. The global mean annual estimates of SHF, LHF, and β from the BrTHF product from 1993 to 2017 were 12.7 W/m^2, 106.2 W/m^2 and 0.11, respectively, all within the ranges of the seven products. The BrTHF product exhibited similar intra-annual cycles for SHF, LHF and β, with bimodal patterns featuring lower and higher peaks in May-June-July-August and December-January, respectively, which was consistent with the results of the seven state-of-the-art products. Additionally, the BrTHF product exhibited significant increasing trends for global SHF and LHF, with rates of 0.04 W/(m^2 year) and 0.33 W/(m^2 year), respectively, which were consistent to most of the seven products. In contrast, the BrTHF product displayed weak growth in β, with a trend approaching 0, which was opposite to the results of the seven products except for the MERRA2 product. The increasing (significant increasing) trends dominated the oceans, with areas of 71.4% (26.2%) for SHF, 87.5% (38.9%) for LHF, 53% (12.2%) for β in the BrTHF product.

The BrTHF product shows significant advantages in the accuracy and rationality of estimates for key parameters (SHF, LHF, and β) related to air-sea interaction and the global energy and water budgets compared to the existing products. It holds great potential for quantifying the global air-sea energy and water budgets, enhancing our understanding of the air-sea interaction, and projecting climate change under global warming.

Author contribution

YW and RT conceived the study and designed the experimental framework. YW performed the experiment and prepared the initial manuscript draft. All authors contributed to manuscript revision, and approved the final version of the manuscript.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgement

We thank the flux datasets and learning datasets provided by the J-OFURO project, IFREMER, ECMWF, NASA, WHOI, NOAA, CEMES and RSS. Moreover, we thank the observations provided by the TAO/TRITON, PIRATA, RAMA, NDBC, TAC, UOP, OOI, AOOS, KOREA, OCS, JKEO and IMO networks or organizations. This work is supported by the National Natural Science Foundation of China [42271378,42071332], and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB0740202).

References

Bentamy, A., Piollé, J. F., Grouazel, A., Danielson, R., Gulev, S., Paul, F., Azelmat, H., Mathieu, P. P., von Schuckmann, K., Sathyendranath, S., Evers-King, H., Esau, I., Johannessen, J. A., Clayson, C. A., Pinker, R. T., Grodsky, S. A., Bourassa, M., Smith, S. R., Haines, K., Valdivieso, M., Merchant, C. J., Chapron, B., Anderson, A., Hollmann, R., and Josey, S. A.: Review and assessment of latent and sensible heat flux accuracy over the global oceans, Remote Sensing of Environment, 201, 196-218, 10.1016/j.rse.2017.08.016, 2017.

Berry, D. I. and Kent, E. C.: Air-Sea fluxes from ICOADS: the construction of a new gridded dataset with uncertainty estimates, International Journal of Climatology, 31, 987-1001, 10.1002/joc.2059, 2011.

Bourras, D.: Comparison of five satellite-derived latent heat flux products to moored buoy data, Journal of Climate, 19, 6291-6313, 2006.

Bourras, D., Reverdin, G., Caniaux, G., and Belamari, S.: A Nonlinear Statistical Model of Turbulent Air–Sea Fluxes, Monthly Weather Review, 135, 1077-1089, 10.1175/mwr3335.1, 2007.

Bourras, D., Cambra, R., Marié, L., Bouin, M. N., Baggio, L., Branger, H., Beghoura, H., Reverdin, G., Dewitte, B., Paulmier, A., Maes, C., Ardhuin, F., Pairaud, I., Fraunié, P., Luneau, C., and Hauser, D.: Air-Sea Turbulent Fluxes From a Wave-Following Platform During Six Experiments at Sea, Journal of Geophysical Research: Oceans, 124, 4290-4321, 10.1029/2018jc014803, 2019.

Bourras, D., Weill, A., Caniaux, G., Eymard, L., Bourlès, B., Letourneur, S., Legain, D., Key, E., Baudin, F., Piguet, B., Traullé, O., Bouhours, G., Sinardet, B., Barrié, J., Vinson, J. P., Boutet, F., Berthod, C., and Clémençon, A.: Turbulent air-sea fluxes in the Gulf of Guinea during the AMMA experiment, Journal of Geophysical Research: Oceans, 114, 10.1029/2008jc004951, 2009.

Brodeau, L., Barnier, B., Gulev, S. K., and Woods, C.: Climatologically Significant Effects of Some Approximations in the Bulk Parameterizations of Turbulent Air–Sea Fluxes, Journal of Physical Oceanography, 47, 5-28, 10.1175/jpo-d-16-0169.1, 2017.

Brunke, M. A.: Uncertainties in sea surface turbulent flux algorithms and data sets, Journal of Geophysical Research, 107, 10.1029/2001jc000992, 2002.

Cai, L., Wang, B., Wang, W., and Feng, X.: The Impact of Air–Sea Flux Parameterization Methods on Simulating Storm Surges and Ocean Surface Currents, Journal of Marine Science and Engineering, 13, 10.3390/jmse13030541, 2025.

Chen, X., Yao, Y., Zhao, S., Li, Y., Jia, K., Zhang, X., Shang, K., Xu, J., and Bei, X.: Estimation of High-Resolution Global Monthly Ocean Latent Heat Flux from MODIS SST Product and AMSR-E Data, Advances in Meteorology, 2020, 1-19, 10.1155/2020/8857618, 2020a.

Chen, X., Yao, Y., Li, Y., Zhang, Y., Jia, K., Zhang, X., Shang, K., Yang, J., Bei, X., and Guo, X.: ANN-Based Estimation of Low-Latitude Monthly Ocean Latent Heat Flux by Ensemble Satellite and Reanalysis Products, Sensors (Basel), 20, 10.3390/s20174773, 2020b.

Clayson, C. and Brown, J.: NOAA Climate Data Record Ocean Surface Bundle (OSB) Climate Data Record (CDR) of Ocean Heat Fluxes, Version 2, Clim. Algorithm Theor. Basis Doc. C-ATBD Asheville NC NOAA Natl. Cent. Environ. Inf. Doi, 10, V59K4885, 2016.

Crespo, J., Posselt, D., and Asharaf, S.: CYGNSS Surface Heat Flux Product Development, Remote Sensing, 11, 10.3390/rs11192294, 2019.

Cummins, D. P., Guemas, V., Cox, C. J., Gallagher, M. R., and Shupe, M. D.: Surface Turbulent Fluxes From the MOSAiC Campaign Predicted by Machine Learning, Geophysical Research Letters, 50, 10.1029/2023gl105698, 2023.

Cummins, D. P., Guemas, V., Blein, S., Brooks, I. M., Renfrew, I. A., Elvidge, A. D., and Prytherch, J.: Reducing Parametrization Errors for Polar Surface Turbulent Fluxes Using Machine Learning, Boundary-Layer Meteorology, 190, 10.1007/s10546-023-00852-8, 2024.

Edson, J. B. a. J., Venkata and Weller, Robert A and Bigorre, Sebastien P and Plueddemann, Albert J and Fairall, Christopher W and Miller, Scott D and Mahrt, Larry and Vickers, Dean and Hersbach, Hans: On the Exchange of Momentum over the Open Ocean, Journal of Physical Oceanography, 43, 1589-1610, 10.1175/jpo-d-12-0173.1, 2013.

Fasullo, J. T., Trenberth, K. E., and Balmaseda, M. A.: Earth's Energy Imbalance, Journal of Climate, 27, 3129-3144, 10.1175/jcli-d-13-00294.1, 2014.

Fu, S., Huang, W., Luo, J., Yang, Z., Fu, H., Luo, Y., and Wang, B.: Deep Learning-Based Sea Surface Roughness Parameterization Scheme Improves Sea Surface Wind Forecast, Geophysical Research Letters, 50, 10.1029/2023gl106580, 2023.

Gentemann, C. L., Clayson, C. A., Brown, S., Lee, T., Parfitt, R., Farrar, J. T., Bourassa, M., Minnett, P. J., Seo, H., Gille, S. T., and Zlotnicki, V.: FluxSat: Measuring the Ocean–Atmosphere Turbulent Exchange of Heat and Moisture from Space, Remote Sensing, 12, 10.3390/rs12111796, 2020.

Grist, J. P., Josey, S. A., Zika, J. D., Evans, D. G., and Skliris, N.: Assessing recent air-sea freshwater flux changes using a surface temperature-salinity space framework, Journal of Geophysical Research: Oceans, 121, 8787-8806, 10.1002/2016jc012091, 2016.

Jiang, Y., Li, Y., Lu, Y., Wu, T., and Gao, Z.: Evaluating modifications to air–sea momentum flux parameterizations under light wind conditions in CAM6, Climate Dynamics, 62, 9687-9701, 10.1007/s00382-024-07415-8, 2024a.

Jiang, Y., Li, Y., Lu, Y., Wu, T., Zhang, J., and Gao, Z.: Evaluating nine different air-sea flux algorithms coupled with CAM6, Atmospheric Research, 10.1016/j.atmosres.2024.107486, 2024b.

Jo, Y.-H.: Calculation of the Bowen ratio in the tropical Pacific using sea surface temperature data, Journal of Geophysical Research, 107, 10.1029/2001jc001150, 2002.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L.: Physics-informed machine learning, Nature Reviews Physics, 3, 422-440, 2021.

Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., and Singh, A.: Physics-informed machine learning: case studies for weather and climate modelling, Philosophical Transactions of the Royal Society A, 379, 20200093, 2021.

Kudryavtsev, V., Chapron, B., and Makin, V.: Impact of wind waves on the air-sea fluxes: A coupled model, Journal of Geophysical Research: Oceans, 119, 1217-1236, 10.1002/2013jc009412, 2014.

Liman, J., Schröder, M., Fennig, K., Andersson, A., and Hollmann, R.: Uncertainty characterization of HOAPS 3.3 latent heat-flux-related parameters, Atmospheric Measurement Techniques, 11, 1793-1815, 10.5194/amt-11-1793-2018, 2018.

Loeb, N. G., Johnson, G. C., Thorsen, T. J., Lyman, J. M., Rose, F. G., and Kato, S.: Satellite and Ocean Data Reveal Marked Increase in Earth s Heating Rate, Geophysical Research Letters, 48, 10.1029/2021GL093047, 2021.

Monin, A. S. and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere, Contrib. Geophys. Inst. Acad. Sci. USSR, 151, e187, 1954.

Myslenkov, S., Shestakova, A., and Chechin, D.: The impact of sea waves on turbulent heat fluxes in the Barents Sea according to numerical modeling, Atmospheric Chemistry and Physics, 21, 5575-5595, 10.5194/acp-21-5575-2021, 2021.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Zhang, W., Hamdi, Z. M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op de Beeck, M., Billesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J. R., Desai, A., Dong, G., El-Madany, T. S., Euskirchen, E. S., Feigenwinter, I., Galvagno, M., Gerosa, G. A., Gielen, B., Goded, I., Goslee, S., Gough, C. M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M. A., Klosterhalfen, A., Knox, S., Kobayashi, H., Kohonen, K.-M., Korkiakoski, M., Mammarella, I., Gharun, M., Marzuoli, R., Matamala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J.-M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C. R., Shekhar, A., Silberstein, R., Silveira, M. L., Spano, D., Tagesson, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E. R., Wang, Y., Woodgate, W., Yepez, E. A., Zhang, J., Zona, D., and Jung, M.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X, Biogeosciences, 21, 5079-5115, 10.5194/bg-21-5079-2024, 2024.

O, S. and Orth, R.: Global soil moisture data derived through machine learning trained with in-situ measurements, Sci Data, 8, 170, 10.1038/s41597-021-00964-1, 2021.

Peng, Z., Tang, R., Jiang, Y., Liu, M., and Li, Z.-L.: Global estimates of 500 m daily aerodynamic roughness length from MODIS data, ISPRS Journal of Photogrammetry and Remote Sensing, 183, 336-351, 10.1016/j.isprsjprs.2021.11.015, 2022.

Robertson, F. R., Roberts, J. B., Bosilovich, M. G., Bentamy, A., Clayson, C. A., Fennig, K., Schröder, M., Tomita, H., Compo, G. P., Gutenstein, M., Hersbach, H., Kobayashi, C., Ricciardulli, L., Sardeshmukh, P., and Slivinski, L. C.: Uncertainties in Ocean Latent Heat Flux Variations over Recent Decades in Satellite-Based Estimates and Reduced Observation Reanalyses, Journal of Climate, 33, 8415-8437, 10.1175/jcli-d-19-0954.1, 2020.

Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J. B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., and Zhang, L.: Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the Tibetan Plateau, Remote Sensing of Environment, 289, 10.1016/j.rse.2023.113519, 2023.

Shie, C.-L., Chiu, L. S., Adler, R., Nelkin, E., Lin, I. I., Xie, P., Wang, F.-C., Chokngamwong, R., Olson, W., and Chu, D. A.: A note on reviving the Goddard Satellite-based Surface Turbulent Fluxes (GSSTF) dataset, Advances in

Atmospheric Sciences, 26, 1071-1080, 10.1007/s00376-009-8138-z, 2009.

Song, X.: The Importance of Relative Wind Speed in Estimating Air–Sea Turbulent Heat Fluxes in Bulk Formulas: Examples in the Bohai Sea, Journal of Atmospheric and Oceanic Technology, 37, 589-603, 10.1175/jtech-d-19-0091.1, 2020.

Song, X.: The Importance of Including Sea Surface Current when Estimating Air–Sea Turbulent Heat Fluxes and Wind Stress in the Gulf Stream Region, Journal of Atmospheric and Oceanic Technology, 38, 119-138, 10.1175/jtech-d-20-0094.1, 2021.

Song, X., Xie, X., Yan, Y., and Xie, S.-P.: Observed sub-daily variations in air–sea turbulent heat fluxes under different marine atmospheric boundary layer stability conditions in the Gulf Stream, Monthly Weather Review, 10.1175/mwr-d-24-0003.1, 2024.

TANG, R. and WANG, Y.: Global dataset of air-sea turbulent heat fluxes (sensible heat flux and latent heat flux) (1993–2017), National Tibetan Plateau Data Center [dataset], https://dx.doi.org/10.11888/Atmos.tpdc.302578, 2025.

Tang, R., Wang, Y., Jiang, Y., Liu, M., Peng, Z., Hu, Y., Huang, L., and Li, Z.-L.: A review of global products of air-sea turbulent heat flux: accuracy, mean, variability, and trend, Earth-Science Reviews, 249, 10.1016/j.earscirev.2023.104662, 2024.

Tomita, H., Hihara, T., Kako, S. i., Kubota, M., and Kutsuwada, K.: An introduction to J-OFURO3, a third-generation Japanese ocean flux data set using remote-sensing observations, Journal of Oceanography, 75, 171-194, 10.1007/s10872-018-0493-x, 2018.

van der Westhuizen, S., Heuvelink, G. B., and Hofmeyr, D. P.: Multivariate random forest for digital soil mapping, Geoderma, 431, 116365, 2023.

Wang, J., Tang, R., Jiang, Y., Liu, M., and Li, Z.-L.: A practical method for angular normalization of global MODIS land surface temperature over vegetated surfaces, ISPRS Journal of Photogrammetry and Remote Sensing, 199, 289-304, 10.1016/j.isprsjprs.2023.04.015, 2023.

Wang, W., Chakraborty, T. C., Xiao, W., and Lee, X.: Ocean surface energy balance allows a constraint on the sensitivity of precipitation to global warming, Nat Commun, 12, 2115, 10.1038/s41467-021-22406-7, 2021.

Wang, Y., Tang, R., Huang, L., Liu, M., Jiang, Y., and Li, Z.-L.: A Bowen ratio-informed method for coordinating the estimates of air–sea turbulent heat fluxes, Environmental Research Letters, 19, 10.1088/1748-9326/ad9341, 2024.

Weller, R. A., Lukas, R., Potemra, J., Plueddemann, A. J., Fairall, C., and Bigorre, S.: Ocean Reference Stations: Long-Term, Open-Ocean Observations of Surface Meteorology and Air–Sea Fluxes Are Essential Benchmarks, Bulletin of the American Meteorological Society, 103, E1968-E1990, 10.1175/bams-d-21-0084.1, 2022.

Wild, M., Folini, D., Hakuba, M. Z., Schär, C., Seneviratne, S. I., Kato, S., Rutan, D., Ammann, C., Wood, E. F., and König-Langlo, G.: The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models, Climate Dynamics, 44, 3393-3429, 10.1007/s00382-014-2430-z, 2014.

Yan, Y., Song, X., Wang, G., and Li, X.: Tropical Cool-Skin and Warm-Layer Effects and Their Impact on Surface Heat Fluxes, Journal of Physical Oceanography, 54, 45-62, 10.1175/jpo-d-23-0103.1, 2024.

Yang, Y., Sun, H., Wang, J., Zhang, W., Zhao, G., Wang, W., Cheng, L., Chen, L., Qin, H., and Cai, Z.: Global ocean surface heat fluxes revisited: A new dataset from maximum entropy production framework with heat storage and Bowen ratio optimizations, Earth System Science Data Discussions, 2024, 1-44, 2024.

Yang, Y., Sun, H., Wang, J., Zhang, W., Zhao, G., Wang, W., Cheng, L., Chen, L., Qin, H., and Cai, Z.: Global ocean surface heat fluxes derived from the maximum entropy production framework accounting for ocean heat storage and Bowen ratio adjustments, Earth System Science Data, 17, 1191-1216, 10.5194/essd-17-1191-2025, 2025.

Yu, L.: Global Air–Sea Fluxes of Heat, Fresh Water, and Momentum: Energy Budget Closure and Unanswered Questions, Annual Review of Marine Science, 11, 227-248, 10.1146/annurev-marine-010816-060704, 2019.

Yu, L. and Weller, R. A.: Objectively Analyzed Air–Sea Heat Fluxes for the Global Ice-Free Oceans (1981–2005), Bulletin of the American Meteorological Society, 88, 527-540, 10.1175/bams-88-4-527, 2007.

Zhang, R., Guo, Weihao, Wang, Xin, and Wang, Chunzai: Ambiguous Variations in Tropical Latent Heat Flux since the Years around 1998, Journal of Climate, 36, 3403–3415, 2023.

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., and Qiu, G. Y.: Physics-Constrained Machine Learning of Evapotranspiration, Geophysical Research Letters, 46, 14496-14507, 10.1029/2019gl085291, 2019.

Zhou, S., Shi, R., Yu, H., Zhang, X., Dai, J., Huang, X., and Xu, F.: A Physical-Informed Neural Network for Improving Air-Sea Turbulent Heat Flux Parameterization, Journal of Geophysical Research: Atmospheres, 129, 10.1029/2023jd040603, 2024.

Zhou, X., Ray, P., Barrett, B. S., and Hsu, P.-C.: Understanding the bias in surface latent and sensible heat fluxes in contemporary AGCMs over tropical oceans, Climate Dynamics, 55, 2957-2978, 10.1007/s00382-020-05431-y, 2020.