JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2019MS001876

Machine Learning-Based Prediction of Spatiotemporal Uncertainties in Global Wind Velocity Reanalyses

Christopher Irrgang[1], **Jan Saynisch-Wagner**[1], **and Maik Thomas**[1,2]

[1]Section 1.3: Earth System Modelling, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, [2]Institute of Meteorology, Freie Universität Berlin, Berlin, Germany

Key Points:
- A recurrent neural network is set up to predict spatiotemporal uncertainties in wind velocity reanalyses
- Global uncertainty maps can be derived from only few individual training locations
- This method has benefits for time series prediction, ensemble simulations, and data assimilation

Correspondence to:
C. Irrgang,
irrgang@gfz-potsdam.de

Citation:
Irrgang, C., Saynisch-Wagner, J., & Thomas, M. (2020). Machine learning-based prediction of spatiotemporal uncertainties in global wind velocity reanalyses. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001876. https://doi.org/10.1029/2019MS001876

Received 23 AUG 2019
Accepted 20 APR 2020
Accepted article online 25 APR 2020

Abstract The characterization of uncertainties in geophysical quantities is an important task with widespread applications for time series prediction, numerical modeling, and data assimilation. In this context, machine learning is a powerful tool for estimating complex patterns and their evolution through time. Here, we utilize a supervised machine learning approach to dynamically predict the spatiotemporal uncertainty of near-surface wind velocities over the ocean. A recurrent neural network (RNN) is trained with reanalyzed 10 m wind velocities and corresponding precalculated uncertainty estimates during the 2012–2016 time period. Afterward, the neural network's performance is examined by analyzing its prediction for the subsequent year 2017. Our experiments show that a recurrent neural network can capture the globally prevalent wind regimes without prior knowledge about underlying physics and learn to derive wind velocity uncertainty estimates that are only based on wind velocity trajectories. At single training locations, the RNN-based wind uncertainties closely match with the true reference values, and the corresponding intra-annual variations are reproduced with high accuracy. Moreover, the neural network can predict global lateral distribution of uncertainties with small mismatch values after being trained only at a few isolated locations in different dynamic regimes. The presented approach can be combined with numerical models for a cost-efficient generation of ensemble simulations or with ensemble-based data assimilation to sample and predict dynamically consistent error covariance information of atmospheric boundary forcings.

Plain Language Summary Machine learning is increasingly used for a wide range of applications in geosciences. In this study, we use an artificial neural network in the context of time series prediction. In particular, the goal is to use a neural network for learning spatial and temporal uncertainties that are associated with globally estimated wind velocities. Three well-known wind velocity products are used for the time period 2012–2016 in different training, validation, and prediction scenarios. Our experiments show that a neural network can learn the prevailing global wind regimes and associate these with corresponding uncertainty estimates. Such a trained neural network can be used for different applications, for example, a cost-efficient generation of ensemble simulations or for improving traditional data assimilation schemes.

1. Introduction

The combination of Earth sciences and machine learning is a rapidly growing research area, focusing on novel applications in geosciences (Lary et al., 2016), meteorology and oceanography (Hsieh & Tang, 1998), and climate sciences (Monteleoni et al., 2013). Artificial neural networks (Rosenblatt, 1958) build one branch of machine learning methods, which can be trained and utilized for various purposes, for example, time series prediction, pattern recognition, numerical model emulation, or data assimilation and inversion. In many ways, machine learning has already become a powerful instrument that can augment traditional statistical and numerical approaches for analyzing and interpreting the plethora of available geoscientific data (see also Reichstein et al., 2019). In the research context of our study, neural networks were successfully used to estimate and recover subgrid processes of oceanic, atmospheric, and climate states (Bolton & Zanna, 2019; Brenowitz & Bretherton, 2018; Rasp et al., 2018). Further, neural networks were trained to emulate parametrizations that are applied to climate models (e.g., O'Gorman & Dwyer, 2018). Cintra and Velho (2014) and Cintra et al. (2015) exchanged a data assimilation scheme of an atmospheric general circulation model with a neural network to perform the analysis step, and Wahle et al. (2015) performed a

neural network-based data assimilation in a numerical model of ocean waves. Irrgang et al. (2019) utilized an artificial neural network as a nonlinear inversion scheme to recover global ocean heat content estimates from spaceborne tidal magnetic signals.

The machine learning approach of our study is nested between the topics of time series prediction and recovery of subgrid processes (downscaling). The overall goal is to set up an artificial neural network that is able to estimate and predict spatiotemporal uncertainties of near-surface wind velocities over the open ocean. As a data basis we use currently available state-of-the-art atmospheric reanalysis products. These products result from data assimilation systems, combining numerical models with corresponding measurements from various Earth observation systems (e.g., Hersbach et al., 2018). In all applications of reanalyses products, for example, sensitivity and intercomparison studies, hindcasts and forecasts of geophysical variables, or the usage as boundary forcing for numerical models, the estimation of respective uncertainties is an essential component and subject to broad research activity (e.g., Chaudhuri et al., 2016; Decker et al., 2012; Jakobson et al., 2012; Kim & Alexander, 2013; Schneider & Fogt, 2018). Here, we focus on reanalyzed near-surface wind velocities as they are readily used in the form of wind stresses as boundary forcing for ocean general circulation models (e.g., Thomas et al., 2001) and were found to contain nonnegligible mismatches that can be projected onto the ocean model state (Chaudhuri et al., 2013). Further, knowing and correcting forcing uncertainties often is a crucial component in successful ocean data assimilation systems (e.g., Irrgang et al., 2017; Saynisch et al., 2014).

The goal of our study includes two premises: (1) The neural network should be able to estimate uncertainties in space and time just from a temporal sequence of wind velocities during a time period of interest. (2) The neural network should be able to generalize beyond its training data and, for instance, provide accurate uncertainty estimates in regions that are excluded from the training process. In combination, a trained neural network that fulfills these requirements can be a valuable stand-alone tool for predicting uncertainty characteristics of wind velocity products (or other variables of interest). Moreover, such a tool could be combined with numerical models or data assimilation systems, to enhance their individual capabilities or to reduce computational demand. For instance, utilizing an adaptive online uncertainty estimation would be a major advance for ocean assimilation systems, where forcing uncertainty estimates and error covariances often are assumed static during the time period of interest. In the following, we will demonstrate how an artificial neural network can be set up to achieve these goals and show its performance in a series of consistent experiments. In section 2, we describe the used wind velocity reanalyses and the estimation of associated uncertainties. Further, we outline the machine learning methodology and the neural network training routine. The results of all performed experiments are presented and discussed in section 3. Lastly, a summary of this study's findings and final conclusions are given in section 4.

2. Materials and Methods

2.1. Atmospheric Wind Reanalyses: Data and Uncertainty

For this study, three atmospheric reanalyses are used: ERA5 ((C3S), 2017; Hersbach et al., 2018) from the European Centre for Medium-Range Weather Forecasts (ECMWF), CFSv2 (Saha et al., 2014) from the National Centers for Environmental Prediction (NCEP), and JRA-55 (Kobayashi et al., 2015) from the Japan Meteorological Agency (JMA). From these three data products (hereafter denoted ERA, CFS, and JRA), zonal and meridional 10 m wind velocities are selected for the 6-year time period from 2012 to 2017. The wind velocity data are bilinearly interpolated to a common $1° \times 1°$ grid and selected at 6-hourly (0, 6, 12, and 18 hr) time steps for the purpose of unification and computational demand (see section 2.2). In the following, we will focus on the zonal component u of the wind velocity over the ocean as the quantity of interest, but we will see that the study design and the results can be directly transferred to the meridional component v.

We utilize a commonly chosen multimodel ensemble approach (e.g., Irrgang et al., 2016; Saynisch et al., 2017) to approximate and quantify the range of uncertainty of the used wind velocities. Ensembles utilize the assumption that a multiple representation of a geophysical quantity serves as a more robust quantification of that variable compared to one individual representation alone. Additionally, we assume that all three reanalysis products are equally valid representations of actual wind velocities over the ocean. In this sense, the ensemble spread, taken as the cross-ensemble variation, provides a measure for the uncertainty. Here, we estimate the wind velocity uncertainty from the standard deviation σ of the (ERA, CFS, JRA)

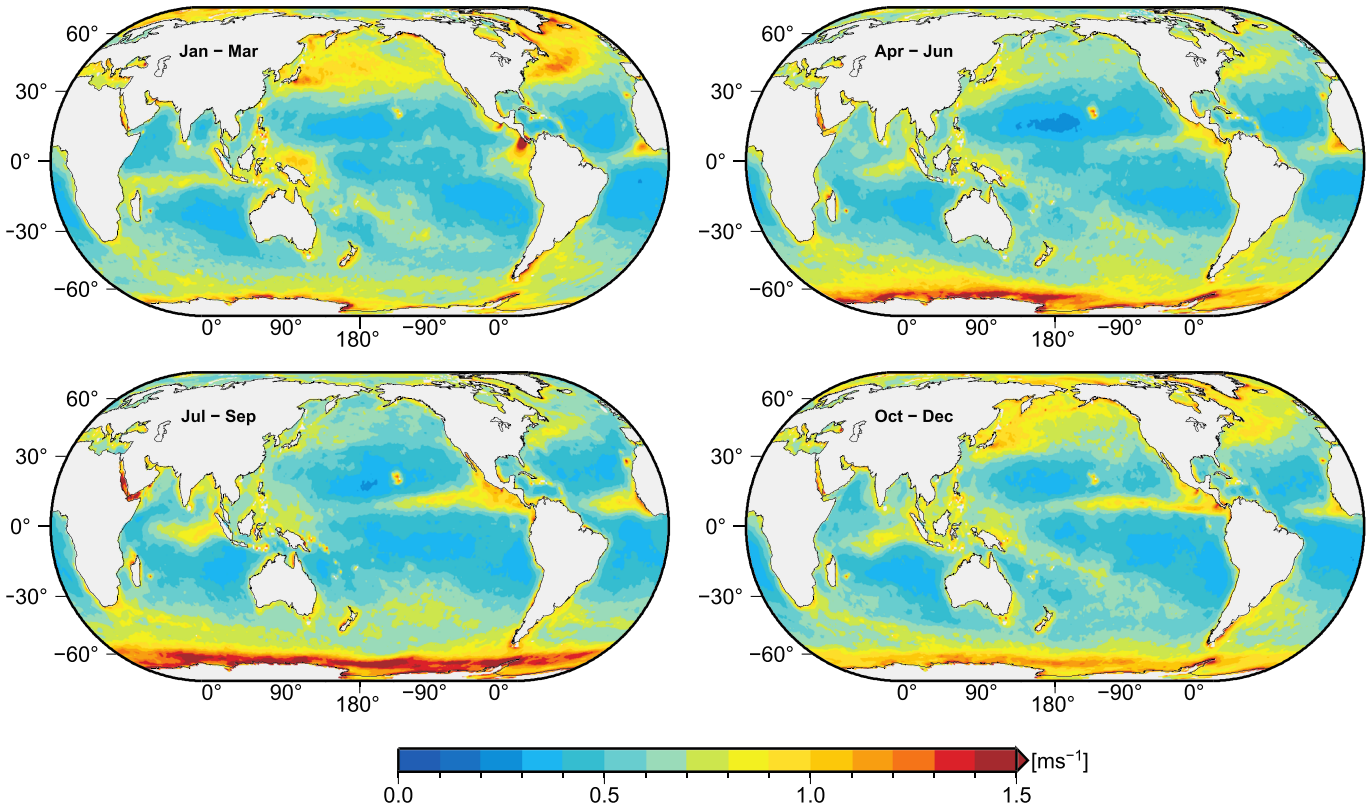

Figure 1. Seasonally averaged uncertainty ε_u, that is, cross-ensemble (ERA, CFS, JRA) standard deviation, of the zonal 10 m wind component during the validation year 2017. The four maps show seasonal uncertainty estimates for January–March (upper left), April–June (upper right), July–September (lower left), and October–December (lower right) of 2017.

ensemble, which varies in both space and time. For a given location (φ, ϑ) over the ocean and for a time point t_0, we calculate the uncertainty ε_u for the zonal wind component as

$$\varepsilon_u\left(\varphi, \vartheta, t_0\right) = \sigma\left(u_{\mathrm{ERA}}(\varphi, \vartheta, t_0), u_{\mathrm{CFS}}(\varphi, \vartheta, t_0), u_{\mathrm{JRA}}(\varphi, \vartheta, t_0)\right) \ . \tag{1}$$

The seasonally averaged spatial distribution of ε_u for the Year 2017 is shown in Figure 1. These values will serve as targets that we aim to predict with an artificial neural network that is trained with wind velocity data from the five previous years 2012 to 2016 (see section 2.2).

To quantify the investigation of the machine learning performance by means of different training situations (see section 2.2), we derived an a priori estimation of the wind uncertainties in 2017 without machine learning, which will serve as a baseline for the evaluation of the machine learning performance. We follow the same approach that the true uncertainty of the Validation and Prediction Year 2017 is unknown and needs to be estimated based on the previous and known wind velocity records from the training time period. This canonical estimation approach is often used in data assimilation studies, where model error covariance matrices are calculated from temporal variances of the respective state variables during representative time periods (see, e.g., Evensen, 1994; Irrgang et al., 2017; Saynisch & Thomas, 2012). Thus, the a priori uncertainty estimates are calculated as climatological mean values over the entire training time period from 2012 to 2016. The resulting relative mismatch to the true uncertainties of 2017 (Figure 1) is shown in Figure 2. Positive and negative deviations are visible over the entire ocean area and indicate the inability of the multiyear averages to reproduce the actual seasonal variations of the wind velocity uncertainties in 2017. In particular, a distinctive underestimation of the 2017 uncertainty range is visible in widespread regions with mismatch values of more than 20% (see also Figure A4 in Appendix A). This a priori uncertainty mismatch will serve as a reference to examine whether the proposed machine learning approach can outperform the widely used uncertainty estimation based on climatological averages over the time period of interest. Note that quantitatively similar mismatch values occur (although more balanced between positive and negative values) if just a single training year is used to derive a priori wind velocity uncertainties (not shown).

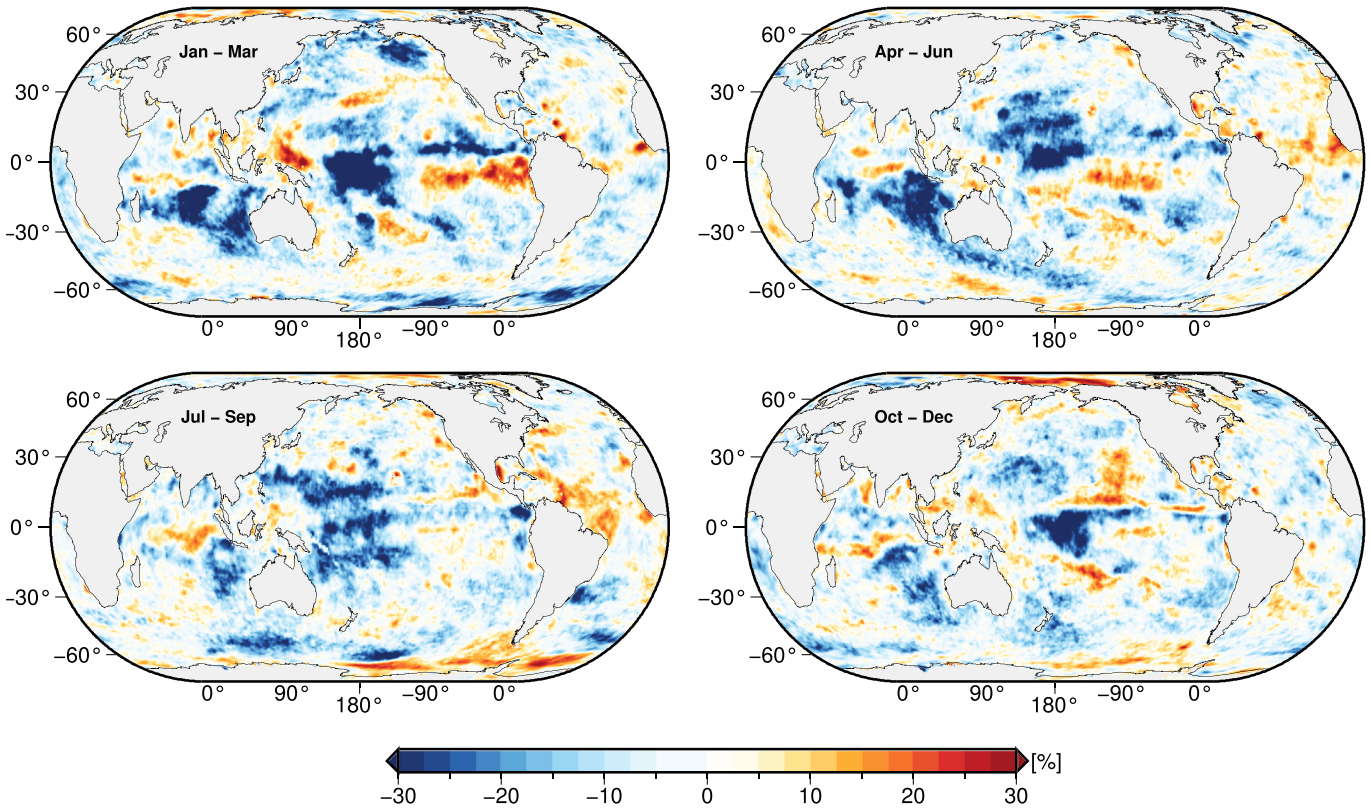

Figure 2. Relative mismatch between an a priori uncertainty estimation of the zonal 10 m wind component in 2017 and the true uncertainty estimates as in Figure 1. The a priori uncertainty values are derived from the 2012–2016 climatological cross-ensemble standard deviation. Negative (blue) values indicate underestimation; positive (red) values indicate overestimation. The four maps show seasonal mismatch values for January–March (upper left), April–June (upper right), July–September (lower left), and October–December (lower right) of 2017.

2.2. Machine Learning: Setup, Training, and Prediction

The machine learning architecture used in this study is realized with the open source framework Tensorflow (Abadi et al., 2016) and the superimposed deep learning library Keras (Chollet, 2015). Using this framework, we set up a recurrent neural network (RNN; e.g., Rumelhart et al., 1986). Like simple feed-forward neural networks (Rosenblatt, 1958), RNNs have the ability to emulate an unknown, and possibly strongly non-linear, functional mapping between input and output values. In addition, RNNs can perform conditional predictions based on temporal sequences of the considered input variables, which renders them as versatile tools for time series applications (Che et al., 2018; Connor et al., 1994). Here, we utilize this feature with the aim that the trained RNN can estimate the zonal wind velocity uncertainty at a given location and time based on a past sequence of the actual zonal wind velocities at that location and also derive predictions for the imminent future.

The RNN contains two hidden layers with 32 artificial neurons (processing nodes) in the first layer and 64 artificial neurons in the second layer. In this setup, we utilize gated recurrent units (GRU, Cho et al., 2014), which are a variant of the long short-term memory (LSTM) architecture (Hochreiter & Schmidhuber, 1997) that solves the vanishing gradient problem inherent in conventional RNNs (Pascanu et al., 2013). LSTM nodes contain a memory cell for storing information and different gates that control how data are passed through the neuron and whether the memory cell is updated with new information. The neurons' ability to remember information renders RNNs especially useful for time series applications. We utilize Keras' default configuration, in which a reset gate is applied in GRU node to control its ability to forget past information (Cho et al., 2014). Additionally, the Rectified Linear Unit (ReLU, Hahnloser et al., 2000) activation function is used, which is computationally efficient and further helps to prevent vanishing gradients in the training process. General and recurrent dropout rates of 0.1 and 0.5, respectively, are applied. In total, the setup of this RNN translates into 22,241 trainable parameters. The final network topology and the hyperparameters were selected from a larger set of different test configurations by optimizing the trade-off between network size,

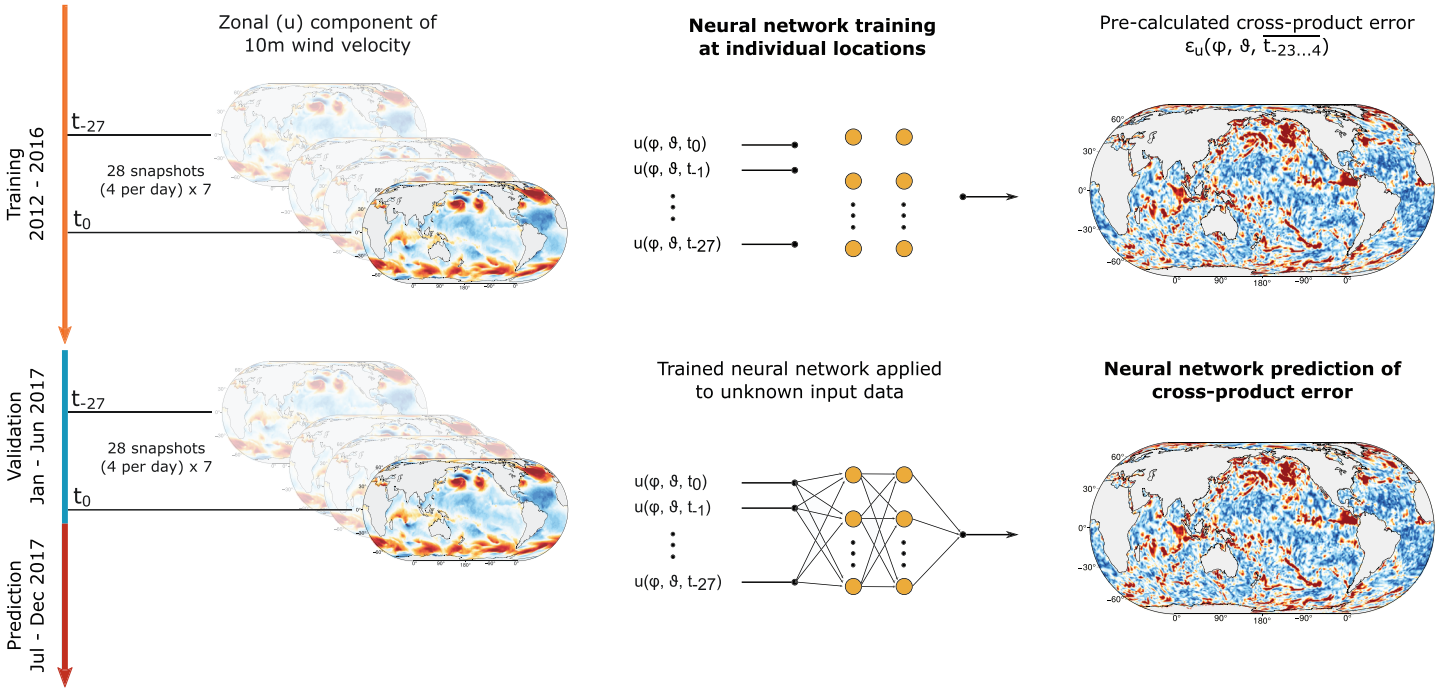

Figure 3. Sketch of the two-phase study design. In the training phase (extending over the time period 2012 to 2016), the recurrent neural network (RNN) is trained with 1-week long samples of 6-hourly zonal wind velocities at a given location as input and the standard deviation across the three reanalyses as output. During the training process, the synaptic weights between the individual neurons are adjusted and the RNN is continuously validated against new data excluded from the training set (January–June 2017; see also Figure 4). In the second phase after the training, the synaptic weights are fixed and the RNN is used to predict new uncertainty estimates for unknown wind velocities that were neither part of the training nor the validation sets (July–December 2017).

network performance, and computational demand. In particular, it is noted that larger networks, that is, networks with more than two hidden layers or with more artificial neurons in the two layers configuration, did not improve the performance. In fact, larger networks exhibited serious overfitting and were discarded.

We follow a supervised learning routine to train the RNN, which is illustrated in Figure 3. The goal of the learning routine is to iteratively change and correct the trainable parameters of the RNN (the weights of the neural connections) in such a way that the RNN emulates a mapping $\mathcal{F}$ between samples of wind velocities at a given location and the corresponding wind velocity uncertainty, that is,

$$\mathcal{F}\left(u_{\{\mathrm{ERA,CFS,JRA}\}}(\varphi, \vartheta, t_{-N,\ldots,0})\right) \mapsto \varepsilon_u(\varphi, \vartheta, \overline{t_{-N+m,\ldots,m}}) = \frac{1}{N+1} \sum_{i=-N+m}^{m} \varepsilon_u(\varphi, \vartheta, t_i)\ . \tag{2}$$

In particular, the RNN is exposed to wind velocity samples at location (φ, ϑ) that reach 1 week into the past from time point t_0 (summing up to $N+1=28$ time steps at a 6-hourly time step). The corresponding target output is given as the weekly averaged cross-product standard deviation according to equation (1). With this training routine, we direct the neural network toward the recovery of long-term variations of the uncertainty. In addition, the averaging window of the weekly averaged output is moved 1 day ($m=4$ time steps) into the future. As a consequence, the RNN is trained (1) to derive realistic wind velocity uncertainties alone from prior knowledge of wind velocities and (2) to predict the change of the wind velocity uncertainty during the next day, which is a commonly requested property in data assimilation.

Five experiments were carried out for this study, all of which include identical (but independent) RNN setups and training routines. The experiments only differ in the number of training locations (see Table 1). As such, the RNN performance resulting from the different training scenarios is examined in a robust way and can be quantitatively compared to the a priori uncertainties (Figures 2 and A4 in Appendix A). In the following, the experiments are denoted E-A (one training location A), E-B (one training location B), E-AB (two training locations A and B), E-10 (10 training locations), and E-19 (19 training locations). In all experiments, the 6 years of reanalyses data described in section 2.1 are split into training, validation, and prediction sets. The training set contains the 5-year long time period of 2012 to 2016. Considering the 6-hourly temporal

Table 1
Summary of the Recurrent Neural Network Training Experiments

Experiment	Number of training locations	Skill score	Generalization skill
A priori	—	42.6	Figure 2
E-A	1	54.8	Figure 6
E-B	1	39.0	Figure 7
E-AB	2	51.2	Figure A1
E-10	10	62.7	Figure A2
E-19	19	**81.7**	Figure 8

Note. For the definition of the skill score see details in section 2 and in Figures A4 and A5. The bold value visually indicates the best performing experiment with the highest skill score value.

resolution of the reanalyses, the 5-year training set comprises 7,308 input-output pairs for each training location (i.e., 138,852 training pairs for E-19). The validation and prediction sets contain the first and second halves of 2017, respectively. We used a training run time of 50 epochs and measured the learning process during the epochs with a mean square error loss function between the RNN predictions and the actual target values from the training and validation sets (e.g., Figure 4 in section 3). After the training, the RNN skill is tested by estimating wind velocity uncertainties for the prediction time period that was excluded from the RNN training routine. Additionally, the generalization skill of the RNN is examined by predicting the wind velocity uncertainty not only at the individual training locations but globally on a $1° \times 1°$ grid.

In the next section, the performance of the trained RNN is analyzed by comparing scatter plots and time series of the predicted versus the true wind velocity uncertainty at training locations. To test the large-scale performance of the RNN, we examine global maps showing absolute and relative mismatches of the RNN during the prediction time period July–December 2017. Additionally, skill scores are derived for the individual experiments that measure the global areal fraction in which the RNN performs best in terms of relative mismatches (see Table 1). As such, the skill scores allow classifying the global machine learning performance in terms of the conducted experiments and with respect to the a priori baseline in a comprehensible way (Figure 2)

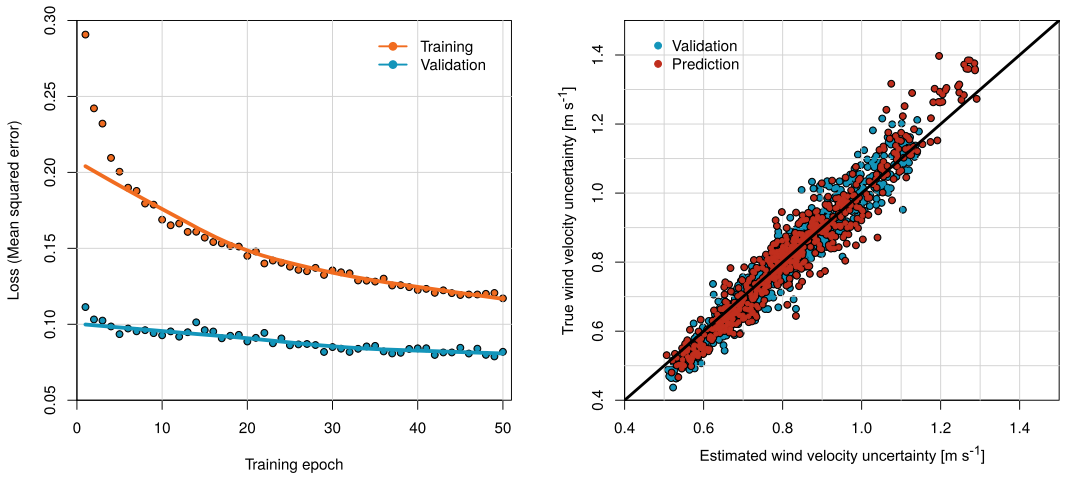

Figure 4. Performance of the recurrent neural network (RNN) in experiment E-A at training location 180°E, 60°S (see black dot in Figure 6). (left panel) Loss improvement, that is, misfit between the RNN prediction and the true values from the training (orange) and validation (blue) set, during the 50 training epochs. Solid lines indicate the overall trend of the training progress (weighted least squares curve). (right panel) Estimated versus true zonal wind velocity uncertainty during the validation (January–June 2017, blue) and prediction (July–December 2017, red) time periods.

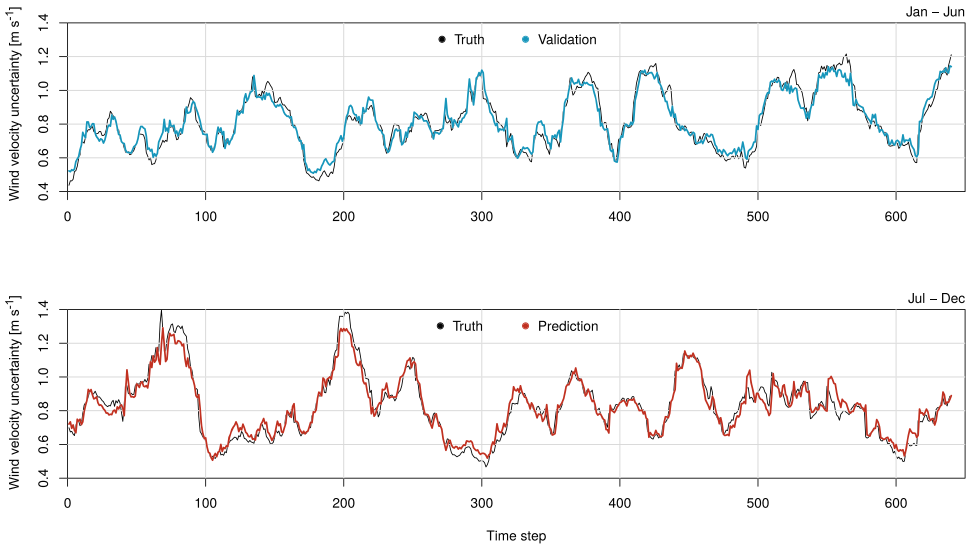

Figure 5. Time series prediction skill of the recurrent neural network (RNN) in experiment E-A at training location 180°E, 60°S (see black dot in Figure 6). (top panel) True versus RNN-based wind velocity uncertainty during the validation time period (January–June 2017). (bottom panel) True versus RNN-based wind velocity uncertainty during the prediction time period (July–December 2017).

3. Results and Discussion

The E-A results demonstrate the performance of the RNN at one single training location (Figure 4). The left panel shows the training progress over time (training epochs) with continuously decreasing misfit between the RNN prediction and the true target values of the zonal wind velocity uncertainty. During each training epoch, the entire training data set is passed through the neural network and the neural weights are adjusted accordingly to minimize the loss function. After each training epoch, the network is tested against the validation data. In particular, this plot highlights that the RNN training is not overfitting toward the training data, since both the loss curves for the training and the validation data are steadily decreasing during the training. Furthermore, the RNN is able to reproduce the validation data with misfit values below 0.1 m s^{-1}. Note that due to the dropout regularization of the neural network during the training phase, the returned loss of the validation data is lower than the loss of the training data. The right panel of Figure 4 shows the scattered RNN-based (horizontal axis) and true (vertical axis) wind velocity uncertainties during the validation and prediction periods of Year 2017. As desired, most values are gathered closely around the diagonal (perfect match), indicating an accurate prediction skill of the RNN. A high correlation of 0.97 was achieved between the estimated and true uncertainty values during the prediction time period (red dots). A slight skewness is visible, that is, the RNN underestimates and overestimates extreme values close to the upper and lower value boundaries, respectively. The corresponding RNN-predicted time series of the wind velocity uncertainty are shown in Figure 5. In accordance with the results in Figure 4, the RNN-derived time series of the wind velocity uncertainty closely follow their respective true target values and accurately reproduce the monthly variability. Again, it should be noted that the target uncertainty time series (denoted Truth in Figure 5) were not part of the RNN training and, thus, were completely unknown to the RNN. In addition, the previously described skewness of the RNN prediction skill is visible in the time series prediction as small mismatches between the RNN-based and the true wind velocity uncertainty (e.g., around Time Step 200 in the bottom panel of Figure 5). Nevertheless, the RNN is capable of reproducing the majority of local maximal and minimal error values.

So far, we have examined the prediction skill of the RNN at a single location that was also used for the training process. Now, we add a second layer of complexity and test the RNN's ability to transfer its knowledge to different regions that were not part of the training routine. The trained neural network and its parameters remain unchanged and are used together with wind velocities at new locations to derive respective uncertainty estimates. Hereafter, this transfer is denoted generalization skill. Ideally, the RNN should have both a good prediction skill and a good generalization skill. The combination of both skills allows accurate global wind velocity uncertainty estimates to be provided, considering only few localized regions for the training

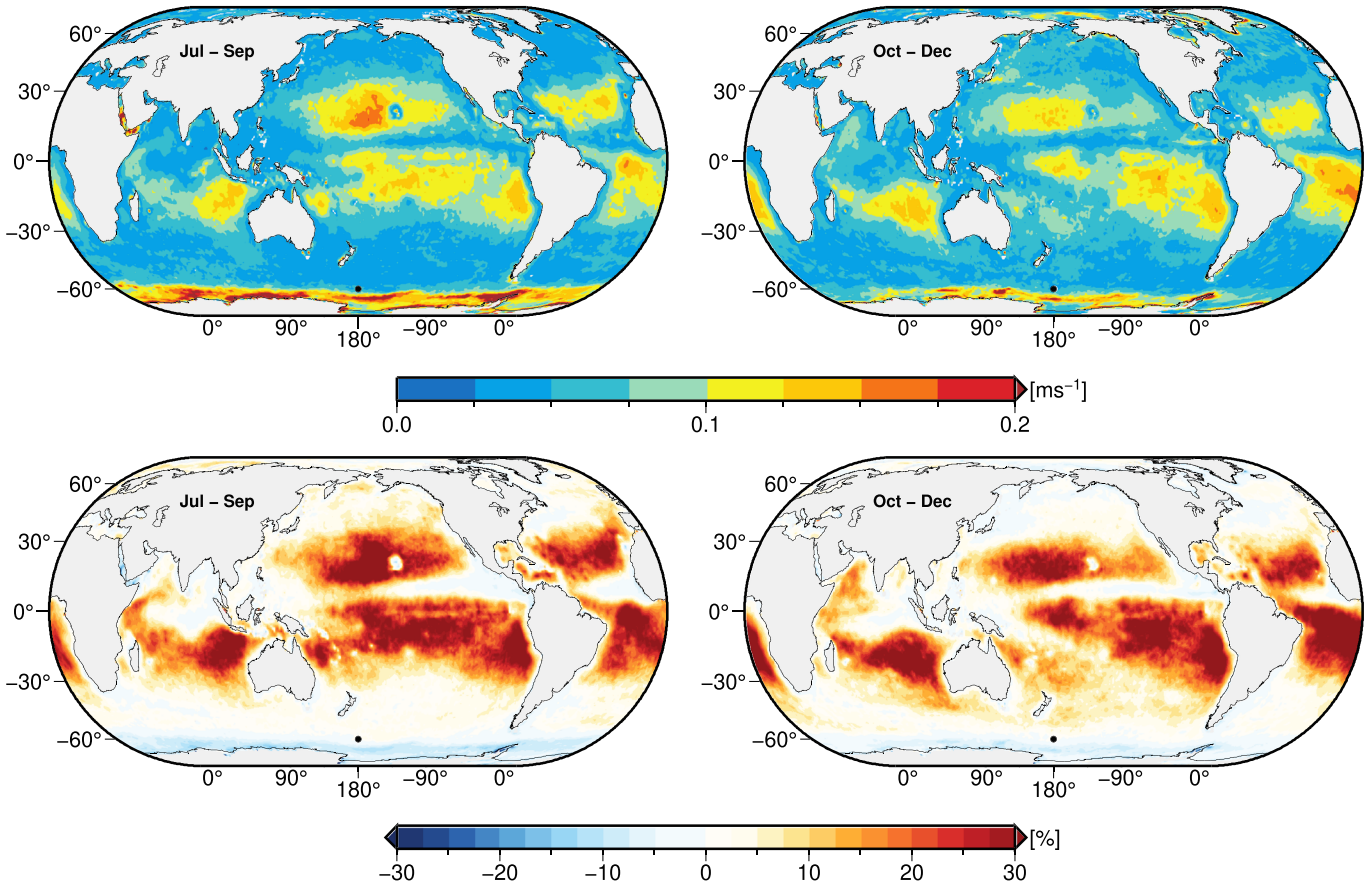

Figure 6. Absolute (RMS, top) and relative (bottom) mismatch of the recurrent neural network (RNN) prediction in experiment E-A with one training location at 180°E, 60°S (left) July–September 2017; (right) October–December 2017. Percentage values indicate the ability to predict the seasonally averaged true zonal wind velocity uncertainty (compare Figure 1). Negative (blue) values indicate underestimation; positive (red) values indicate overestimation.

routine and, thus, minimal computational effort. This skill combination makes the trained RNN especially suitable for applications with sparse or scattered data that might occur due to limited observation availability. In the following, the generalization skill is quantified in terms of root-mean-square (RMS) errors and relative mismatch values between the neural network prediction and the true target values; that is, small mismatch values are associated with a strong generalization skill and vice versa.

The generalization skill of the RNN in experiment E-A is shown in Figure 6. As in Figure 2, the lower maps show the relative mismatch (as seasonal mean values of the of 6-hourly uncertainties) between the RNN-predicted uncertainty values and the true uncertainty values during July–December 2017. The black dot at location 180°E, 60°S indicates the training location used in E-A. The training location is set in the region of the Antarctic Circumpolar Current (ACC), where strong eastward wind velocities and large temporal variations occur (e.g., Lin et al., 2018; Trenberth et al., 1990). Especially during the Southern Hemisphere winter, also high corresponding wind velocity uncertainties are present (see Figure 1 and Chaudhuri et al., 2013). Due to the single training location, the RNN is trained to these extreme conditions in the Southern Ocean. Thus, the RNN generalization skill is particularly high in regions with wind conditions similar to those of the training location (compare Figures 1 and 6). Most prominently, the zonal band between −40°N and −80°N shows consistent low mismatch values around ±5% during both seasons of the prediction time period, which are also lower than in the a priori uncertainty estimation (compare Figures 2 and 6). In this band, absolute errors amount to RMS values up to 0.1 m s^{-1} north of 60°S with regional peak values around 0.3 m s^{-1} below 60°S (upper maps of Figure 6). Equally low mismatches are visible in regions of prevailing westerlies in the Northern Hemisphere (with RMS values below 0.1 m s^{-1}), for example, over the North Atlantic and North Pacific Oceans (Figure 6). The skill score for Experiment E-A amounts to 54.8, which is larger than for "no-knowledge" a priori uncertainty estimation (see Table 1 and compare Figures A4 and A5).

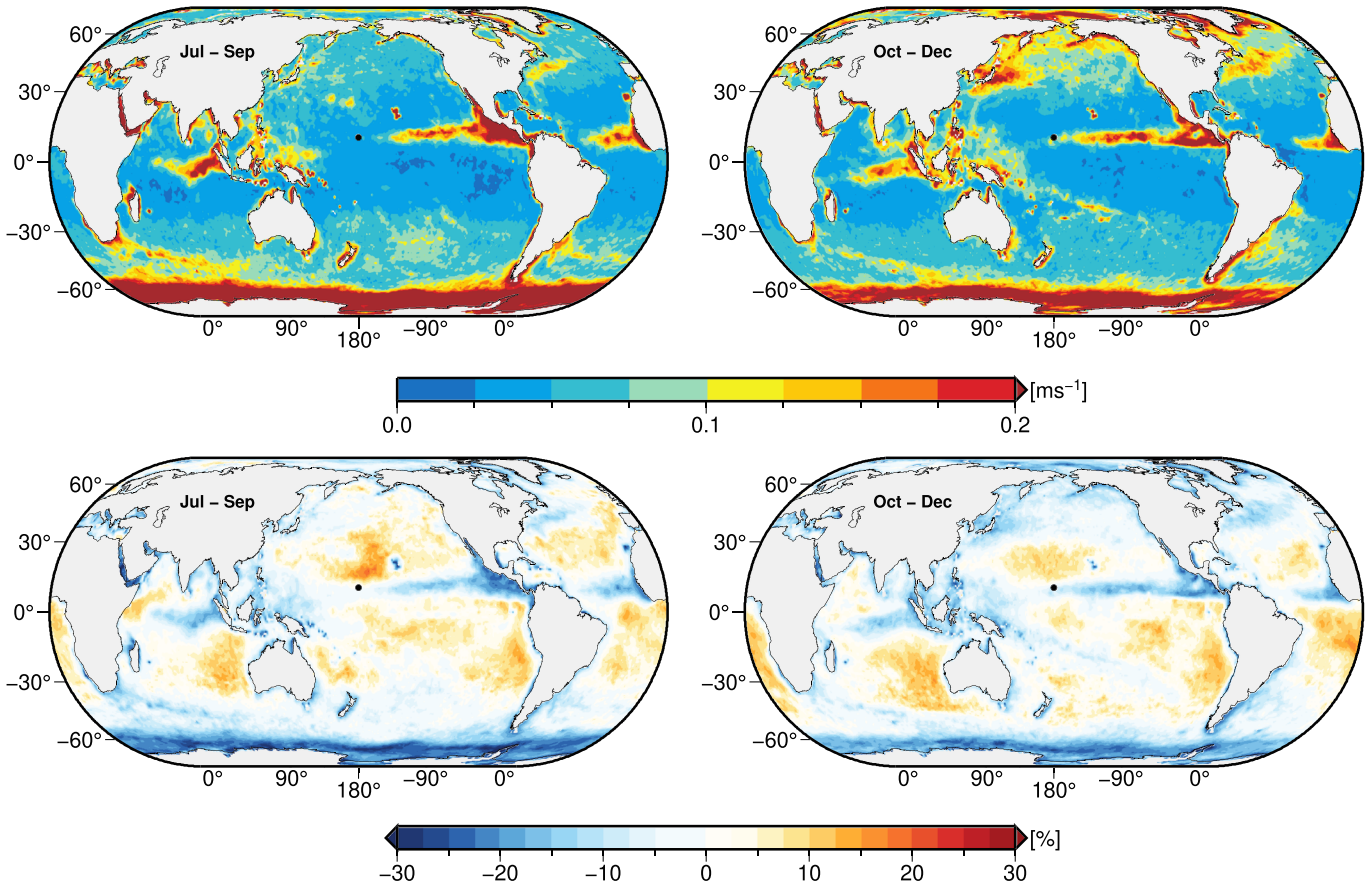

Figure 7. Absolute (RMS, top) and relative (bottom) mismatch of the recurrent neural network (RNN) prediction in experiment E-B with one training location at 180°E, 10°N (see black dot). (left) July–September 2017; (right) October–December 2017. Percentage values indicate the ability to predict the seasonally averaged true zonal wind velocity uncertainty (compare Figure 1). Negative (blue) values indicate underestimation; positive (red) values indicate overestimation.

At the same time, the single training location is leading to a significant bias toward large wind velocity uncertainties during the RNN training. The biased training particularly excludes easterly wind conditions, trade winds, and their respective uncertainties. As a consequence, the RNN is overestimating the target values in almost all midlatitudes and close to the equator with mismatch values of more than +20% (RMS values between 0.1 and 0.2 m s^{-1}). In these regions, the mismatch values are larger than in the a priori uncertainty estimation (compare Figures 2 and 6). In conclusion, the RNN in E-A has an accurate prediction skill within its knowledge horizon but a deficient generalization skill.

The findings from experiment E-A were reassessed by selecting a different single training location in experiment E-B that is located in a northeasterly wind regime at 180°E, 10°N, where much smaller wind velocity uncertainties are present (compare Figures 1 and 7). As expected, the RNN training in E-B is oppositely biased compared to E-A, which results in a general improvement of the RNN mismatch of northeasterly and southeasterly trade wind uncertainties with values below 15% (0.05 m s^{-1}) in widespread regions. Below 60°S, however, the RNN significantly underestimates the true wind velocity uncertainties in both seasons with mismatch values between −20% and −30% (up to 0.5 m s^{-1}) and, thus, also performs worse than the a priori uncertainty estimation (compare Figures 2 and 7). The strongest underestimation is found during the Southern Hemisphere winter season in the ACC region (left panel of Figure 7), where the largest uncertainties of respective eastward wind velocities occur that cannot be reproduced by the RNN in E-B. Additionally, comparing the skill scores of E-A (54.8) and E-B (39.0) further indicates the bias offset between the performance of the respective RNN's generalization capability (see Figure A5).

To improve the RNN generalization skill while maintaining an accurate prediction skill, we successively expanded the training routine and included additional training locations from various regions and wind

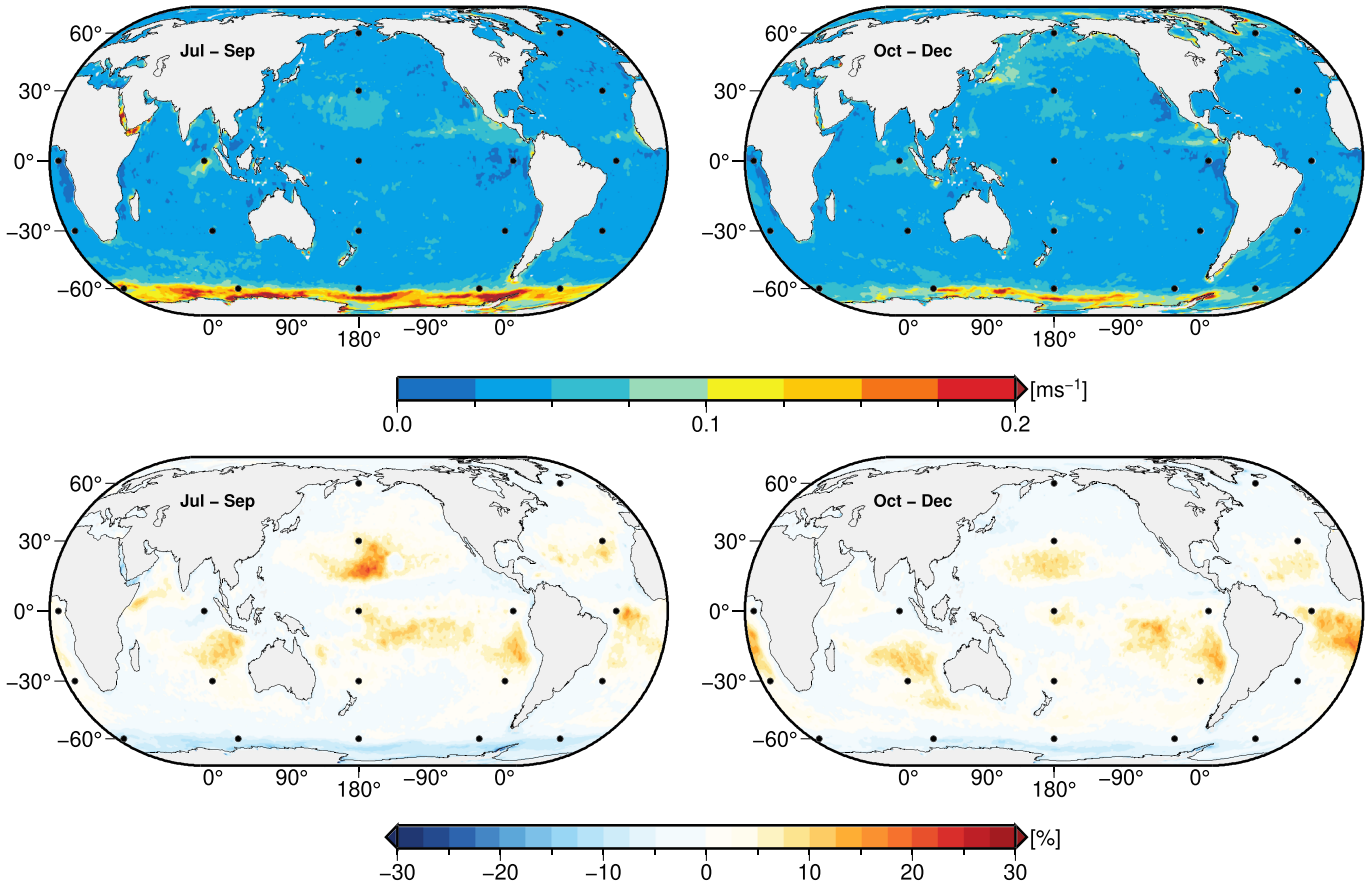

Figure 8. Absolute (RMS, top) and relative (bottom) mismatch of the recurrent neural network (RNN) prediction in experiment E-19 with 19 different training locations (see black dots). (left) July–September 2017; (right) October–December 2017. Percentage values indicate the ability to predict the seasonally averaged true zonal wind velocity uncertainty (compare Figure 1). Negative (blue) values indicate underestimation; positive (red) values indicate overestimation.

regimes on the Earth. In experiment E-AB, this concept is examined by combining both training locations from E-A and E-B into the RNN training (Figure A1 in Appendix A). Although there is still a significant overestimation of midlatitude wind velocity uncertainties with mismatch values of 15% and more, the RNN is able to attenuate the strongest mismatch values visible in E-A and E-B (compare Figure A1 with Figures 6 and 7). RMS errors are diminished and limited accordingly, especially the peak values in the ACC region that remain below 0.2 m s^{-1}. The generalization skill is additionally quantified by a skill score of 51.2 (see Table 1 and Figure A5), which is a similar performance as in E-A (54.6).

The previous experiments suggest that a well-trained RNN can have a strong generalization skill, provided that the training comprises the majority of the globally occurring wind velocity range and its corresponding uncertainty estimates. Consequently, we selected a coarse global grid of training locations for the next experiment E-19, that covers polar, midlatitudinal, and equatorial regions. The training grid consists of 19 locations, which are set at longitudinal coordinates 5°E, 90°E, 180°E, 270°E, and 330°E, and at latitudinal coordinates 60°S, 30°S, 0°N, 30°N, and 60°N. As such, it is still assumed that the true uncertainty and its spatial distribution is not known. Rather than preselecting specific training locations by considering physical patterns in the different wind regimes, this regular training grid was chosen to prevent a selection bias in the neural network training. Note that this coarse grid training setup particularly renders E-19 a downscaling task in addition to the pure time series prediction. The resulting generalization skill and the underlying training locations are shown in Figure 8. Small mismatch values are visible in almost all regions and in both seasonal plots, indicating that the predicted seasonal wind velocity uncertainties closely resemble their corresponding target values. RMS values mostly remain below 0.05 m s^{-1} and below 0.15 m s^{-1} in the ACC region. The RNN is capable of reproducing the true wind velocity uncertainty with a mismatch of less than 5% in more than 80% of the areal coverage (see also Table 1 and Figure A5). In terms of the skill score (81.7),

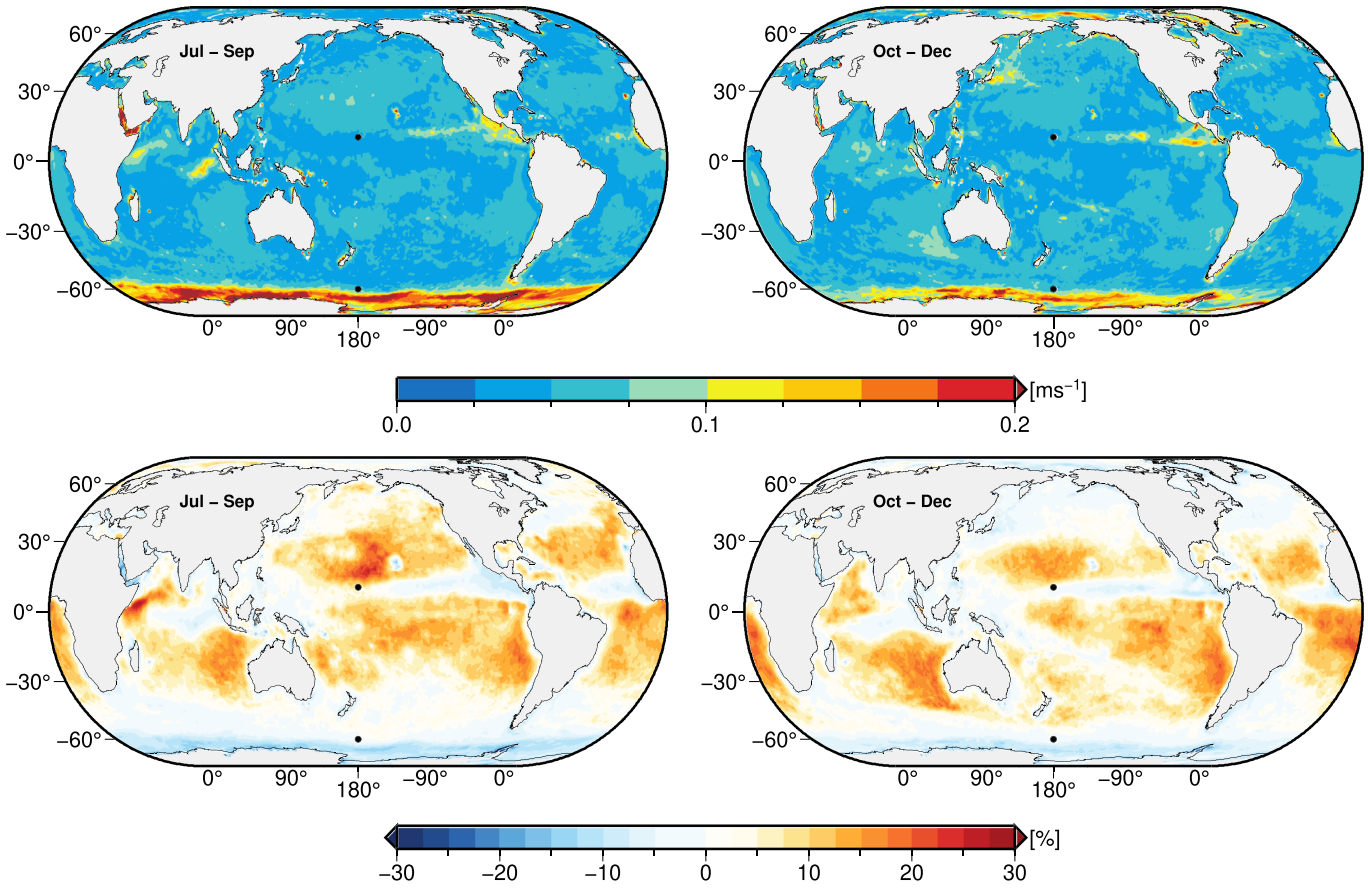

Figure A1. Absolute (RMS, top) and relative (bottom) mismatch of the recurrent neural network (RNN) prediction in experiment E-AB with two different training locations at 180°E, 60°S and 180°E, 10°N (see black dots). (left) July–September 2017; (right) October–December 2017. Percentage values indicate the ability to predict the seasonally averaged true zonal wind velocity uncertainty (compare Figure 1). Negative (blue) values indicate underestimation; positive (red) values indicate overestimation.

the RNN's skill in E-19 almost doubled compared to the a priori tests (see Table 1 and compare Figures A4 and A5). Thus, it can be concluded that the trained neural network in E-19 has captured spatiotemporal characteristics of the wind velocity uncertainty budget and is able to derive accurate predictions from new wind velocity trajectories that were not part of the training process. Most surprisingly, only very few training locations are necessary to teach the RNN the distribution and evolution of the globally prevalent uncertainty dynamics. However, the training setup of E-19 is a minimal requirement to achieve the shown performance of the RNN. Using less training locations, that is, a coarser training grid, leads to an immediate deterioration of the generalization skill (see results of experiment E-10 in Table 1 and Figures A2 and A5 in Appendix A). One reason for the remaining mismatch lies in the intentionally regular training grid that includes locations close to the doldrums in the equator region and in the transition zone between easterly and westerly winds around 30°N and 30°S. Adding further training locations to the E-19 setup, specifically in regions of high remaining mismatch values, could help to further increase the generalization skill. Prominent candidate regions, where the RNN still overestimates the true wind velocity uncertainty, are located close to land, for example, west of Peru and Chile, north of Madagascar, and west of Central Africa. Increasing the number of training locations without prior knowledge only leads to a marginal improvement of the RNN in terms of mismatches and skill score (not shown). Due to the available computational resources, the asymptotic behavior of the RNN prediction and generalization skills with respect to the number of training locations could not be investigated.

Besides the presented results in the context of time series prediction, using machine learning as a tool to estimate uncertainty dynamics of a geophysical quantity provides several further applications. First, an RNN that is trained as described above can be used to sample uncertainty characteristics from just a single data

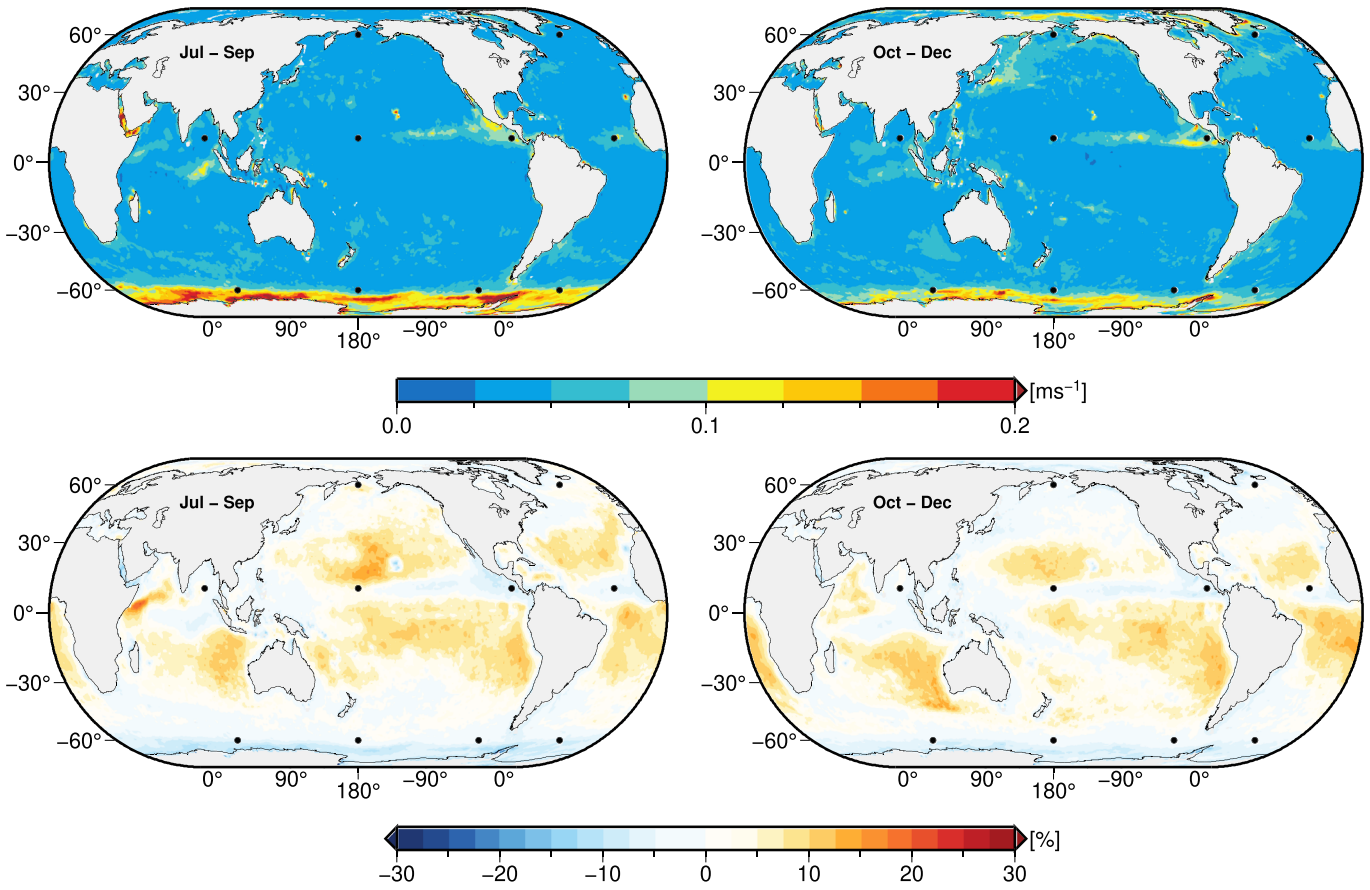

Figure A2. Absolute (RMS, top) and relative (bottom) mismatch of the recurrent neural network (RNN) prediction in experiment E-10 with 10 different training locations (see black dots). (left) July–September 2017; (right) October–December 2017. Percentage values indicate the ability to predict the seasonally averaged true zonal wind velocity uncertainty (compare Figure 1). Negative (blue) values indicate underestimation; positive (red) values indicate overestimation.

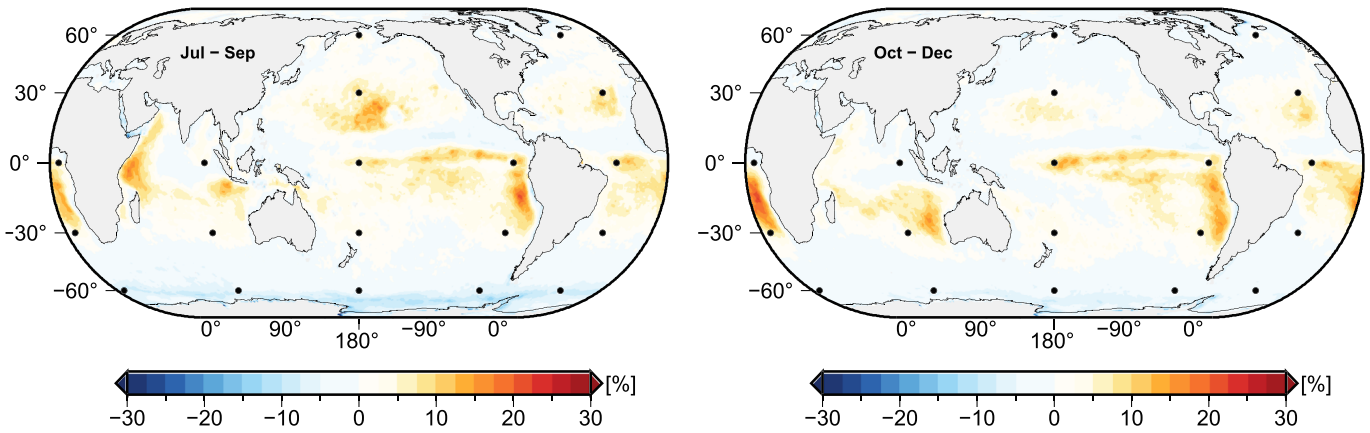

Figure A3. Relative mismatch of the recurrent neural network (RNN) prediction in experiment E-19 with 19 different training locations (see black dots). (left) July–September 2017; (right) October–December 2017. In contrast to Figure 8, the percentage values indicate the ability to predict the seasonally averaged true *meridional* wind velocity uncertainty. Negative (blue) values indicate underestimation; positive (red) values indicate overestimation.

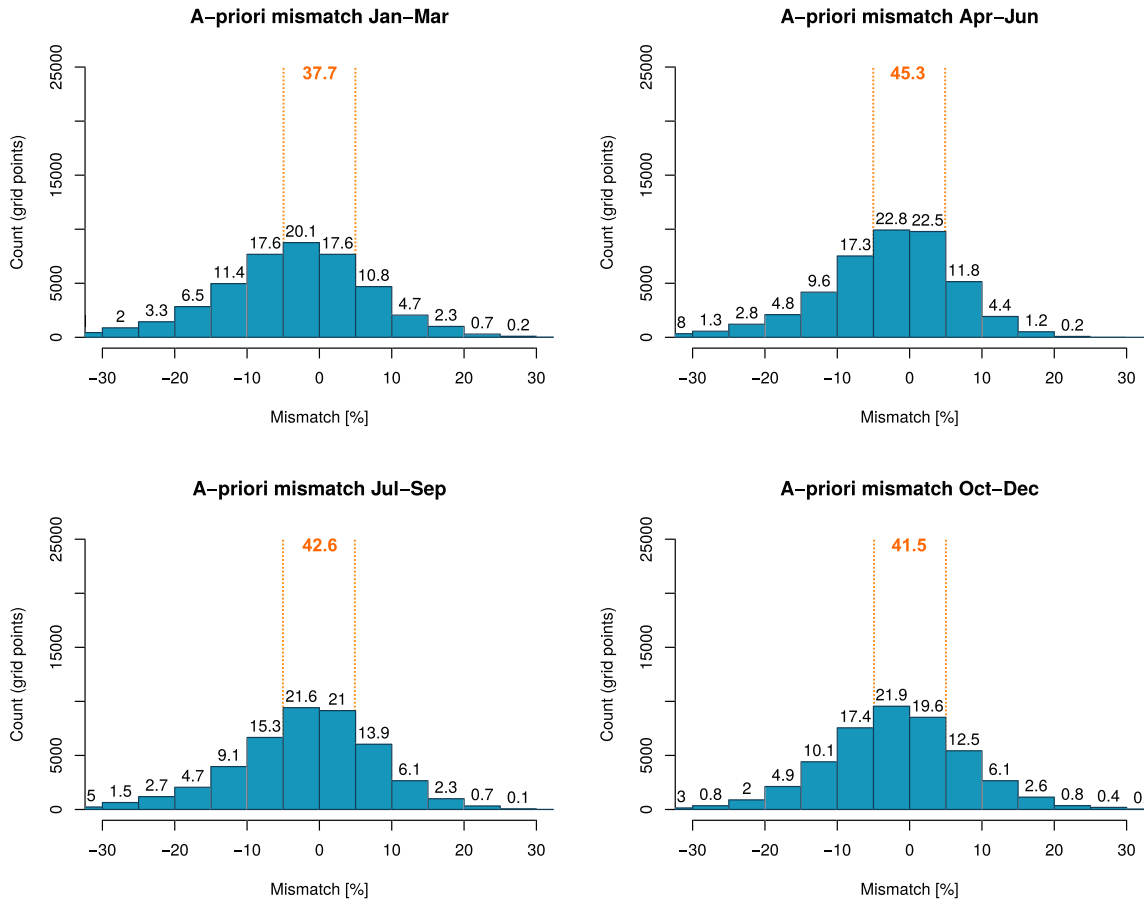

Figure A4. A priori mismatch histograms for the year 2017. The blue bins aggregate the respective grid points visible in Figure 2. Numbers on top of the bins are percentage values with respect to the total number of ocean grid points. Orange numbers indicate the areal proportion of the globe, where mismatch values are within the ±5% interval (the respective largest seasonal value is labeled as skill score in Table 1).

set or reanalysis. After the successive training with different reanalyses data, the RNN has learned not only the connection between input and target output but also information about the range, variability, and cross relation between the different input data, which are stored in the neural weights. This information can be resampled with new or modified input trajectories, which can significantly diminish the need to employ an entire ensemble of different realizations of a geophysical quantity. This application is especially useful for time periods where no ensemble data are available or when computational resources are limited. Similarly, the trained RNN could be coupled to a numerical model to generate a spatiotemporally highly resolved ensemble spread for single-model trajectories. Second, the trained RNN can be incorporated into ensemble-based data assimilation routines and be used for the dynamic evolution of uncertainty through time. Especially in ocean data assimilation, effective corrections of the model state depend on corrections of the atmospheric boundary forcing (Irrgang et al., 2017; Saynisch et al., 2014). The temporal prediction skill of the neural network is useful for this application, as boundary forcing uncertainties could be adaptively preprocessed in advance of the model step and be used to correct the model ensemble span. This combination would be particularly useful for the boundary forcing of a numerical model that is not propagated through time by the model and, thus, needs an external description of associated uncertainties. Third, unlike a numerical model, it is a virtue of artificial neural networks to detach the physics of a geophysical quantity from its actual spatiotemporal patterns. As such, a trained RNN can be directly transferred and used for variables with comparable amplitudes and variations in space and time, given the assumption that the uncertainty trajectories contain similar statistical properties as in the training data. For instance, the RNN trained to estimate the uncertainty of the zonal wind velocity component can be used without further training to predict the respective values for the meridional component of the wind velocity, while maintaining the previously seen high prediction and generalization skills (Figure A3).

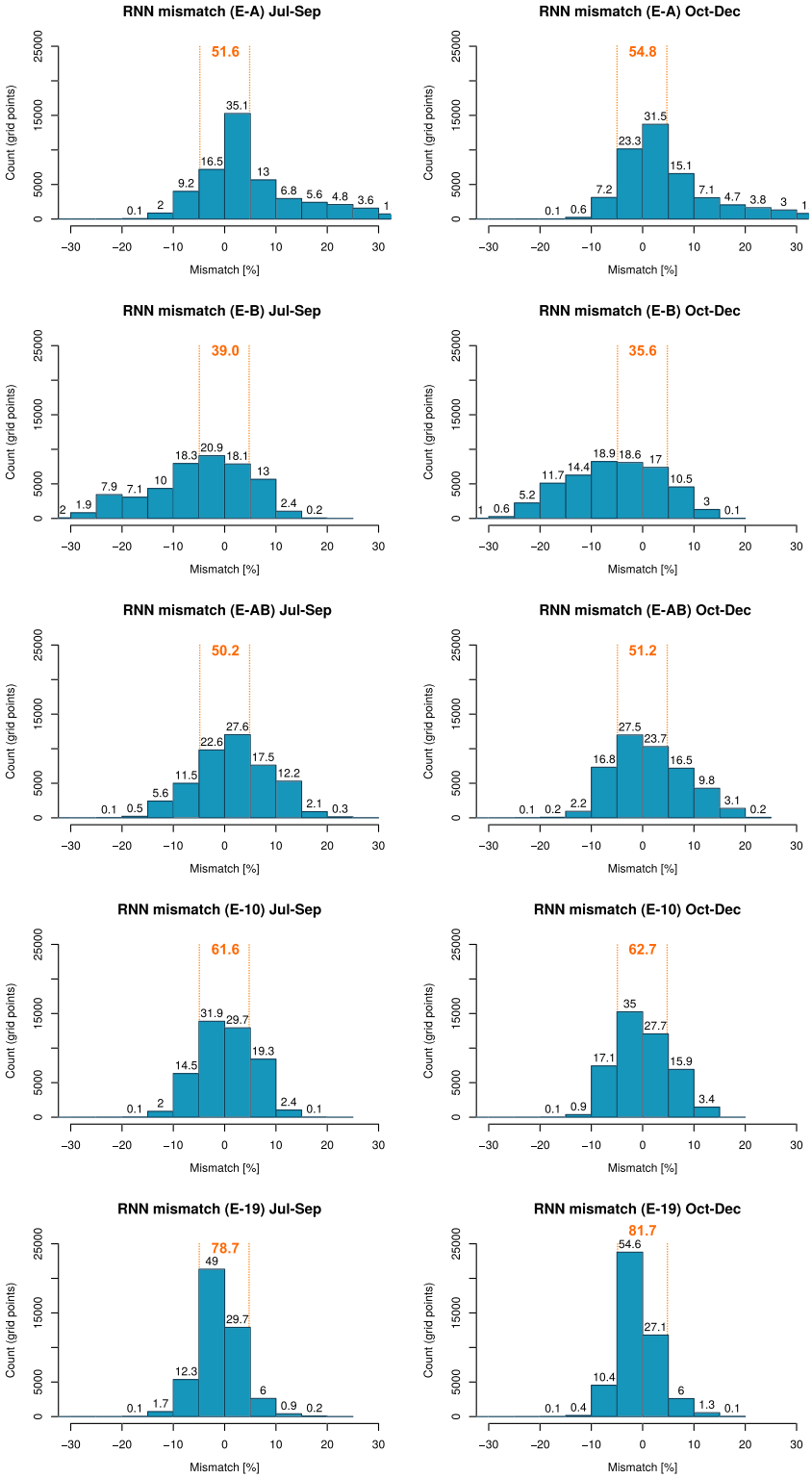

Figure A5. Recurrent neural network (RNN) mismatch histograms for the prediction time period July–December 2017 of all experiments. The blue bins aggregate the respective grid points visible in Figures 6–8 and A1–A3. Numbers on top of the bins are percentage values with respect to the total number of ocean grid points. Orange numbers indicate the areal proportion of the globe, where mismatch values are within the ±5% interval (the respective largest seasonal value is labeled as skill score in Table 1).

4. Summary and Conclusion

In this study, we describe a machine learning-based approach for estimating and predicting the uncertainty of near-surface wind velocities over the ocean. A RNN is set up and is trained with 10 m wind velocity data from an ensemble of three different global atmospheric 1° × 1° reanalyses with a 6-hourly temporal resolution. The RNN training comprises the years 2012 to 2016, while the year 2017 is used for validation and prediction scenarios. In particular, the RNN is trained to derive spatially and temporally varying uncertainty estimates only from temporal samples of the respective wind velocity time series.

We carried out a set of experiments that differ in terms of the RNN training routine, ranging from single training locations to a coarse global grid of training locations. This setup allowed us to examine the performance and prediction skill of the RNN in a consistent way, as well as to test the generalization skill of the RNN for regions where no specific training data were considered. Our results demonstrate that the RNN is able to capture the prevailing wind regimes over the ocean during the 2012–2016 time period and can reproduce the corresponding wind velocity uncertainty through space and time. Additionally, accurate predictions were derived by the neural network for the year 2017, which was unknown to the network, that is, excluded from the training routine. In particular, the RNN outperforms common-sense a priori uncertainty estimates for 2017, which were derived from climatological mean values during 2012–2016, by a factor of $\sim$ 2. Furthermore, we highlight that only very few training locations are necessary for deriving globally reliable uncertainty estimates, which makes this method a computationally cheap tool with versatile applications. Besides the pure application to estimate the uncertainty of a geophysical quantity, an RNN trained as described above could be nested and used in combination with numerical models or data assimilation techniques. In these applications, a trained RNN can assume different tasks, for example, generating dynamic ensembles of a boundary forcing of a numerical model and evolving respective error covariance information through time.

Appendix A: Additional Results of the Recurrent Neural Network Generalization Skill

We provide additional results, that is, absolute and relative mismatches, for the machine learning experiments E-AB (Figure A1), E-10 (Figure A2), and E-19 (Figure A3). Furthermore, we show mismatch histograms for the a priori wind velocity uncertainty estimation (Figure A4) and for the machine learning experiments E-A, E-B,E-AB, E-10, and E-19 (Figure A5).

Acknowledgments
This study was funded by the Helmholtz Association and by the Initiative and Networking Fund of the Helmholtz Association through the project "Advanced Earth System Modelling Capacity (ESM)." We thank the ECMWF, NCEP, and JMA for providing their atmospheric forcing reanalyses ERA5, CFSv2, and JRA-55. The ERA-5 data can be obtained from the Copernicus Climate Change Service (C3S) at the Climate Data Store (via https://cds.climate.copernicus.eu). The CFSv2 and JRA-55 data were provided by the Research Data Archive (RDA) of the Computational and Information Systems Laboratory at the National Center for Atmospheric Research (via https://rda.ucar.edu/). NCAR is supported by grants from the National Science Foundation. Tensorflow and Keras libraries and codes can be downloaded online (from https://www.tensorflow.org/ and https://keras.io/). The generated data within this study will be made available in a public repository of the German Research Centre for Geosciences (GFZ).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., & others (2016). Tensorflow: A system for large-scale machine learning. In *12th Symposium on Operating Systems Design and Implementation*, pp. 265–283.

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, *11*, 376–399. https://doi.org/10.1029/2018MS001472

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*, 6289–6298. https://doi.org/10.1029/2018GL078510

Chaudhuri, A. H., Ponte, R. M., & Forget, G. (2016). Impact of uncertainties in atmospheric boundary conditions on ocean model solutions. *Ocean Modelling*, *100*, 96–108.

Chaudhuri, A. H., Ponte, R. M., Forget, G., & Heimbach, P. (2013). A comparison of atmospheric reanalysis surface products over the ocean and implications for uncertainties in air-sea boundary forcing. *Journal Climate*, *26*(1), 153–170.

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Science Report*, *8*(1), 1–12.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Doha, Qatar: Association for Computational Linguistics. https://www.aclweb.org/anthology/D14-1179

Chollet, F. (2015). Keras. https://github.com/fchollet/keras, GitHub.

Cintra, R. S., & Velho, H. F. D. C. (2014). Data assimilation by artificial neural networks for an atmospheric general circulation model: Conventional observation. *Bulletin of the American Meteorological Society*, *77*(3), 437–471.

Cintra, R. S., Velho, H. F. D. C., Anochi, J., & Cocke, S. (2015). Data assimilation by artificial neural networks for the global FSU atmospheric model: Surface pressure. In *2015 Latin America Congress on Computational Intelligence (LA-CCI)* (pp. 1–6). IEEE. https://ieeexplore.ieee.org/document/7435937/

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, *5*(2), 240–254.

(C3S) (2017). Copernicus Climate Change Service: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS). https://cds.climate.copernicus.eu/cdsapp#!/home

Decker, M., Brunke, M. A., Wang, Z., Sakaguchi, K., Zeng, X., & Bosilovich, M. G. (2012). Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. *Journal Climate*, *25*(6), 1916–1944.

Evensen, G. (1994). Inverse methods and data assimilation in nonlinear ocean models. *Physica D*, *77*(1-3), 108–129.

Hahnloser, R. H. R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, *405*(6789), 947–951.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtolg, P., Berrisford, P., Bidlot, J.-R., de Boissson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A. J., Haiden, T., Hlm, E., Haimberger, L., Hogan, R., Hornyi, A., Janiskova, M., Laloyaux, P., Lopez, P., Muoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thpaut, J.-N., Vitart, F., Yang, X., Zsoter, E., & Zuo, H. (2018). Operational global reanalysis: Progress, future directions and synergies with NWP. (27). https://www.ecmwf.int/node/18765

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hsieh, W. W., & Tang, B. (1998). Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin America Meteorology Society*, *79*(9), 1855–1870.

Irrgang, C., Saynisch, J., & Thomas, M. (2016). Ensemble simulations of the magnetic field induced by global ocean circulation: Estimating the uncertainty. *Journal of Geophysical Research: Oceans*, *121*, 1866–1880. https://doi.org/10.1002/2016JC011633

Irrgang, C., Saynisch, J., & Thomas, M. (2017). Utilizing oceanic electromagnetic induction to constrain an ocean general circulation model: A data assimilation twin experiment. *Journal Advances in Modeling Earth Systems*, *9*, 1703–1720. https://doi.org/10.1002/2017MS000951

Irrgang, C., Saynisch, J., & Thomas, M. (2019). Estimating global ocean heat content from tidal magnetic satellite observations. *Science Report*, *9*, 7893.

Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., & Jaagus, J. (2012). Validation of atmospheric reanalyses over the central Arctic Ocean. *Geophysical Research Letters*, *39*, L10802. https://doi.org/10.1029/2012GL051591

Kim, J. E., & Alexander, M. J. (2013). Tropical precipitation variability and convectively coupled equatorial waves on submonthly time scales in reanalyses and TRMM. *Journal Climate*, *26*(10), 3013–3030.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., & Takahashi, K. (2015). The JRA-55 Reanalysis: General specifications and basic characteristics. *Jornal Meteorology Society Japan Series II*, *93*(1), 5–48.

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3–10.

Lin, X., Zhai, X., Wang, Z., & Munday, D. R. (2018). Mean, variability, and trend of Southern Ocean wind stress: Role of wind fluctuations. *Journal Climate*, *31*(9), 3557–3573.

Monteleoni, C., Schmidt, G. A., & McQuade, S. (2013). Climate informatics: Accelerating discovering in climate science with machine learning. *Computing in Science & Engineering*, *15*(5), 32–40.

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*, 2548–2563. https://doi.org/10.1029/2018MS001351

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International Conference on Machine Learning*, pp. 1310–1318.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychology Review*, *65*(6), 386–408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., & Becker, E. (2014). The NCEP Climate Forecast System Version 2. *Journal Climate*, *27*(6), 2185–2208.

Saynisch, J., Bergmann-Wolf, I., & Thomas, M. (2014). Assimilation of GRACE-derived oceanic mass distributions with a global ocean circulation model. *Journal of Geodesy*, *89*(2), 121–139.

Saynisch, J., Petereit, J., Irrgang, C., & Thomas, M. (2017). Impact of oceanic warming on electromagnetic oceanic tidal signals: A CMIP5 climate model-based sensitivity study. *Geophysical Research Letters*, *44*, 4994–5000. https://doi.org/10.1002/2017GL073683

Saynisch, J., & Thomas, M. (2012). Ensemble Kalman-filtering of Earth rotation observations with a global ocean model. *Journal of Geodesy*, *62*, 24–29.

Schneider, D. P., & Fogt, R. L. (2018). Artifacts in century-length atmospheric and coupled reanalyses over Antarctica due to historical data availability. *Geophysical Research Letters*, *45*, 964–973. https://doi.org/10.1002/2017GL076226

Thomas, M., Sündermann, J., & Maier-Reimer, E. (2001). Consideration of ocean tides in an OGCM and impacts on subseasonal to decadal polar motion excitation. *Geophysical Research Letters*, *28*(12), 2457–2460.

Trenberth, K. E., Large, W. G., & Olson, J. G. (1990). The mean annual cycle in global ocean wind stress. *Journal of Physical Oceanography*, *20*(11), 1742–1760.

Wahle, K., Staneva, J., & Guenther, H. (2015). Data assimilation of ocean wind waves using neural networks. A case study for the German Bight. *Ocean Modelling*, *96*, 117–125.