Quantifying the Benefits of Nonlinear Methods for Global Statistical Hindcasts of Tropical Cyclones Intensity

S. NEETU,[a] M. LENGAIGNE,[b,c] J. VIALARD,[b] M. MANGEAS,[d] C.E. MENKES,[b,d] I. SURESH,[a] J. LELOUP,[b] J.A. KNAFF[e]

[a] *CSIR-National Institute of Oceanography, Goa, India*
[b] *Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, LOCEAN Laboratory, IPSL, Paris, France*
[c] *Indo-French Cell for Water Sciences, IISc-NIO-IITM–IRD Joint International Laboratory, NIO, Goa, India*
[d] *IRD/UMR ENTROPIE, Nouméa Cedex, New Caledonia*
[e] *NOAA/NESDIS, Center for Satellite Research and Applications, Fort Collins, Colorado*

ABSTRACT

While tropical cyclone (TC) track forecasts have become increasingly accurate over recent decades, intensity forecasts from both numerical models and statistical schemes have been trailing behind. Most operational statistical–dynamical forecasts of TC intensity use linear regression to relate the initial TC characteristics and most relevant large-scale environmental parameters along the TC track to the TC intensification rate. Yet, many physical processes involved in TC intensification are nonlinear, hence potentially hindering the skill of those linear schemes. Here, we develop two nonlinear TC intensity hindcast schemes, for the first time globally. These schemes are based on either support vector machine (SVM) or artificial neural network (ANN) algorithms. Contrary to linear schemes, which perform slightly better when trained individually over each TC basin, nonlinear methods perform best when trained globally. Globally trained nonlinear schemes improve TC intensity hindcasts relative to regionally trained linear schemes in all TC-prone basins, especially the SVM scheme for which this improvement reaches ~10% globally. The SVM scheme, in particular, partially corrects the tendency of the linear scheme to underperform for moderate intensity (category 2 and less on the Saffir–Simpson scale) and decaying TCs. Although the TC intensity hindcast skill improvements described above are an upper limit of what could be achieved operationally (when using forecasted TC tracks and environmental parameters), it is comparable to that achieved by operational forecasts over the last 20 years. This improvement is sufficiently large to motivate more testing of nonlinear methods for statistical TC intensity prediction at operational centers.

1. Introduction

Storm surges associated with tropical cyclones (TCs) are a major contributor to casualties and property loss caused by natural disasters in tropical coastal regions (Needham et al. 2015), especially in the Bay of Bengal and the Gulf of Mexico. For instance, TC Nargis caused 140 000 deaths, 1 million homeless, and $1 billion (U.S. dollars) damages in Myanmar in April–May 2008 (McPhaden et al. 2009). A timely and accurate prediction of the TC track and intensity is therefore of great importance to allow authorities to take preventive actions, such as evacuating areas under threat. Over the last decades, TC track forecasting has improved significantly due to refined numerical weather prediction models, but accurate TC intensity forecasting is still a challenge (Elsberry et al. 2013; Emanuel and Zhang 2016). Track forecast errors have for instance been reduced by ~66% both in the North Pacific and Atlantic over the past decades (Landsea and Cangialosi 2018), whereas intensity forecasts have improved by only one-third to one-half of this rate (Cangialosi and Franklin 2014; DeMaria et al. 2014). The errors of TC intensity forecasts are indeed still large, reaching from ~5–8 kt (1 kt ≈ 0.51 m s^{-1}) at 12-h lead time to 15–25 kt for 120-h lead time (DeMaria et al. 2014; Cangialosi and Franklin 2014; Bushnell and Falvey 2018; Cangialosi 2019).

Operational centers run a hierarchy of operational TC intensity forecast models that range from fully coupled ocean–atmosphere or atmospheric-only numerical

Corresponding author: S. Neetu, neetu@nio.org
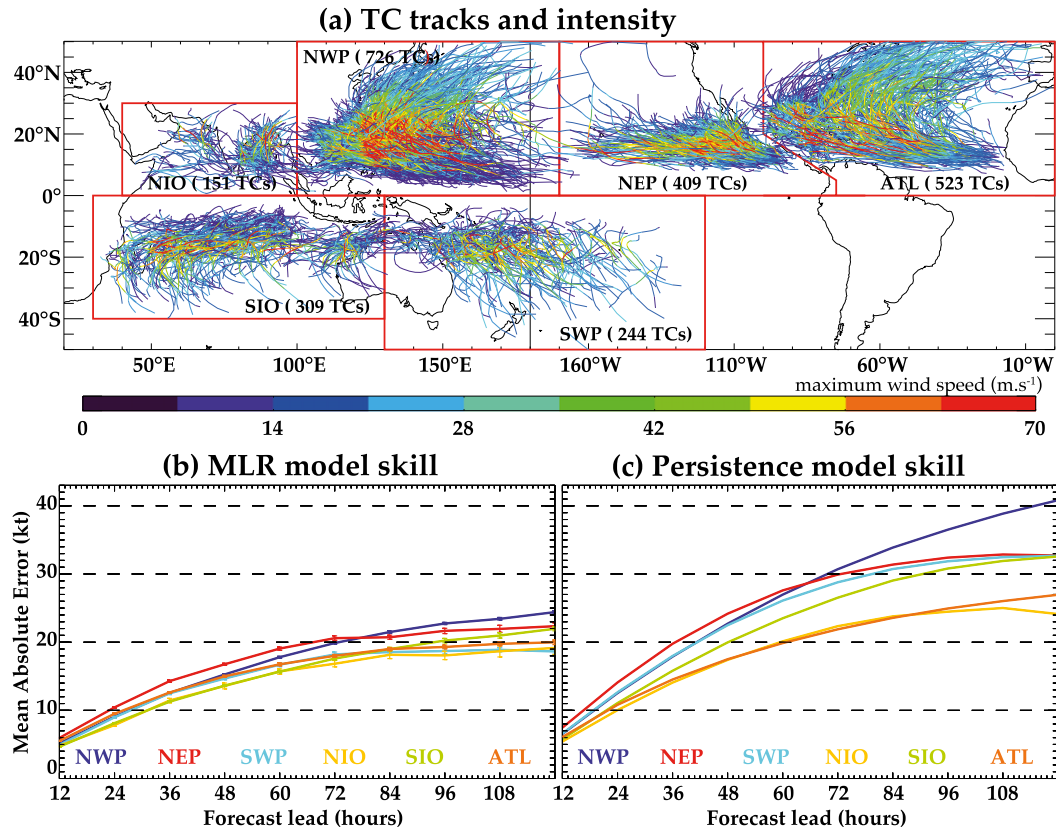
(a) TC tracks and intensity

Fig. 1. (a) Map of TC tracks and maximum wind speed (color) over the 1979–2012 period. The boxes indicate the various TC-prone basins considered in the present study [northwestern Pacific (NWP), northeastern Pacific (NEP), southwestern Pacific (SWP), Atlantic (ATL), southern Indian Ocean (SIO), and northern Indian Ocean (NIO)]. The total number of TCs over the 1979–2012 period is also indicated for each basin. Mean absolute error (MAE) as a function of lead time for (b) the regionally trained MLR model and (c) persistence for each basin. Error bars on (b) provide the 95% confidence interval estimated from a bootstrap method.

models to statistical–dynamical models to simple statistical prediction schemes (DeMaria 2009; DeMaria et al. 2014). Because of the complex physical processes affecting intensity changes, the very high spatial resolution required and the difficulties in initializing the real-time forecasts, statistical–dynamical TC intensity forecast models have remained competitive with dynamical models (Kucas 2010; DeMaria et al. 2007, 2014; Kaplan et al. 2015), except in the Atlantic basin where dynamical models outperform statistical–dynamical forecasts for the most recent hurricane seasons (Cangialosi 2019).

Statistical–dynamical models use statistical techniques to relate the future TC intensity changes to the cyclone initial characteristics (current intensity and its time derivative) and to large-scale environmental parameters encountered by the cyclone along its forecast track. Several large-scale environmental parameters indeed have a well-documented effect on the cyclone intensity. Strong vertical wind shear can for instance inhibit a TC intensification (Gray 1968; DeMaria 1996).

The maximum potential intensity (MPI; e.g., Miller 1958; Emanuel 1995; Holland 1997) increases as a function of the sea surface temperature, and gives an upper bound of the intensity that the cyclone can reach. The midtropospheric relative humidity represents the convection inhibition in dry environments (Emanuel et al. 2004). Other currently used environmental parameters (Emanuel 2007; Knaff et al. 2005; DeMaria et al. 2005) that we also include in the current study are the vorticity at lower levels; air temperature in the upper troposphere and boundary layer equivalent potential temperature.

There are 6 major regions of TC development worldwide: the northwestern Pacific (NWP), the northeastern Pacific (NEP), the southwestern Pacific (SWP), the North Atlantic (ATL), the north Indian Ocean (NIO), and the southern Indian Ocean (SIO) (Fig. 1a). Operational centers worldwide have developed separate statistical–dynamical TC intensity forecast models for various TC basins. The Statistical Hurricane Intensity Prediction

Scheme (SHIPS) was initially developed for the Atlantic basin, based on a multiple linear regression (MLR) technique, which relates predictors to the TC intensity change (DeMaria and Kaplan 1994, 1999; DeMaria et al. 2005). Knaff et al. (2005) later adapted this model for the northwestern Pacific [Statistical Typhoon Intensity Prediction Scheme (STIPS)]. Knaff and Sampson (2009) developed the Southern Hemisphere Statistical Typhoon Intensity Prediction Scheme or SH-STIPS jointly for the southwestern Pacific and southern Indian Ocean basins. Finally, Kotal et al. (2008) adapted STIPS to northern Indian Ocean TCs.

Neetu et al. (2017) recently developed a statistical-dynamical hindcast scheme similar to those above (i.e., using an MLR technique) separately for each TC-prone basin, but based on the same set of predictors and datasets for all basins. This set of TC hindcast schemes has mean absolute error (MAE; Fig. 1b) comparable to that of SHIPS, STIPS, and SH-STIPS, with a 20%–40% MAE improvement relative to persistence (Fig. 1c), except in the ATL and NIO (10%–25%; Neetu et al. 2017). A large fraction of this MAE improvement (60%–80%) arises from accounting for the initial TC characteristics (i.e., its intensity at and intensity derivative at and before the beginning of the forecast; Neetu et al. 2017). The environmental parameters that yield the most skill globally are vertical wind shear followed by maximum potential intensity, but with individual contributions that strongly depend on the basin. Neetu et al. (2017) also demonstrated that statistical TC intensity forecasts poorly predict intensity changes of moderate TCs in all basins, with 2–4 times more skillful hindcasts for category 3 and above TCs.

The TC intensity statistical forecasts described above use linear schemes (i.e., they assume a linear relationship between the predictor and the TC intensification rate). There are however nonlinear interactions between TC intensity and environmental parameters, which are not considered in these models (Tang and Emanuel 2012; Lin et al. 2017). Tang and Emanuel (2012) for instance argue that the flux of low-entropy air into the TC center (or midlevel ventilation index) is a major environmental parameter affecting TC intensification. This ventilation index was developed from a theoretical framework, and has a nonlinear dependence on other environmental parameters, as it is formulated as the environmental wind shear multiplied by the nondimensional midlevel entropy deficit divided by maximum potential intensity. Although SHIPS already includes elements of the ventilation index, Tang and Emanuel (2012) suggest that the linear nature of SHIPS does not properly account for the influence of the midlevel ventilation on TC intensification, given that the ventilation index is a nonlinear combination of SHIPS parameters.

Some studies have introduced a nonlinear scaling of some predictors (for instance MPI[2]) in order to introduce a nonlinear dependence of the cyclone intensity evolution to those predictors (DeMaria and Kaplan 1999; Knaff et al. 2005). It would be difficult and time-consuming to test all possible scaling for each parameter in a linear statistical scheme. Our strategy is thus to account for the nonlinear relationships between all selected variables in a more systematic way in the present study. A relevant option is to use statistical schemes that are designed to capture the nonlinear relationships between variables, without applying any ad hoc scaling on the input parameters. DeMaria (2009) for instance introduce a logistical growth equation model, in which the storm growth is linearly related to vertical shear and vertical stability of the atmosphere but saturates to yield a maximum intensity determined by the MPI. This model is used for the Atlantic and northeast Pacific since 2006, with a skill increase of up to 15% relative to SHIPS for a given season, especially for long lead times (DeMaria 2009). Similarly, Lin et al. (2017) used sparse generalized additive models, which allow nonlinear transforms of predictors, to identify and characterize nonlinear effects of environmental parameters on TC intensification. These models only marginally increase the skill of ATL and NWP TC intensity prediction relative to a linear regression approach.

Artificial neural networks (ANN; e.g., Rumelhart et al. 1986; Fine 1999) and support vector machines (SVM; e.g., Cortes and Vapnik 1995) are two popular statistical nonlinear modeling tools for nonparametric predictions used in many fields, and more specifically for oceanographic, meteorological, and climatic and climate impact studies (e.g., Tolman et al. 2005; Lee 2006; Liu et al. 2010 for ANN and e.g., Elbisy 2015; Aguilar-Martinez and Hsieh 2009; Descloux et al. 2012 for SVM). Both schemes have already been used in relation with TC studies. The ANN and SVM schemes have both been used for predicting rain under TCs (e.g., Lin et al. 2009; Wei 2012), with more skill for SVM-based models (Lin et al. 2009). ANN schemes have also been used to predict TC-induced storm surges (Lee 2009), tropical cyclogenesis (Hennon et al. 2005), and TC tracks (Ali et al. 2007; Roy and Kovordányi 2012). Only a couple of studies have more specifically used SVM or ANN for TC intensity forecasts focusing on the NWP region. Lin et al. (2013) showed that the SVM scheme improved NWP TC intensity forecasts by about 6%–11% over the 2002–09 period, relative to official operational forecasts. Using the same input parameters as in STIPS (Knaff et al. 2005),

TABLE 1. Number of cases at each forecast interval for the individual basins and the total dataset used in the present study.

Basin/lead time	12 h	24 h	36 h	48 h	60 h	72 h	84 h	96 h	108 h	120 h
NWP	10 151	9437	8735	8037	7378	6740	6128	5549	5004	4481
NEP	4577	4176	3790	3416	3057	2719	2406	2108	1845	1609
SWP	2746	2530	2317	2111	1919	1739	1564	1400	1246	1108
NIO	1129	992	863	744	637	539	448	370	302	246
SIO	4068	3788	3514	3242	2972	2709	2456	2219	1996	1785
ATL	5329	4860	4420	4005	3626	3279	2971	2688	2417	2164
ALL	28 000	25 783	23 639	21 555	19 589	17 725	15 973	14 334	12 810	11 393

Sharma et al. (2013) showed that their nonlinear ANN scheme yielded a 2%–10% skill improvement relative to STIPS for the 2003–04 TC seasons.

The two papers above have suggested the added value of using nonlinear TC intensity forecasts. These papers however focus on the NWP basin using a rather short dataset ($\sim$7 years). In this paper, we investigate benefits of nonlinear methods for each individual basin, using the same approach and set of input parameters for consistency, as well as an extended dataset for more robust quantitative results. To that end, we will compare global TC intensity forecasts using the ANN and SVM nonlinear methods to those made with a linear MLR approach very similar to that in Neetu et al. (2017). More specifically, we will investigate if nonlinear models should be trained for each basin individually (as the MLR), or globally; the skill improvement they yield and its basin dependency; and whether this skill improvement depends on the cyclone characteristics (intensity, intensifying/decaying phase) as was the case for the MLR.

We use a very similar MLR to that described in Neetu et al. (2017) as a reference in the current paper and investigate whether nonlinear ANN/SVM models built from the same input parameters bring improvements. The MLR model, its input parameters, and our skill assessment method are summarized in section 2. Section 3 describes the ANN and SVM models. Section 4 compares the performance of these nonlinear models to that of the linear model. Section 5 provides a summary and discussion of our results.

2. Methods

a. Datasets

Building a TC intensity hindcast scheme requires the cyclone positions (used for obtaining the surrounding "environmental parameters") and the cyclone intensity, defined from maximum 1-min horizontal winds (which is used for training and verifying the scheme, along with its time derivative). The International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010)

dataset combines the best track data from several operational centers provides into a single database with TC locations and intensities in all basins. The Joint Typhoon Warning Center (JTWC) has the advantage of using the same track and intensity determination methodology for all basins, and was used wherever it was available (i.e., all basins but NEP and ATL). For NEP and ATL basins, we use the National Hurricane Center track data, which uses the same wind-averaging period of 1 min as JTWC. The TC dataset used in this study includes all best track points whatever their intensity (including tropical storm and depression stages) and their location (including extratropical cases). The number of cases at each forecast interval is provided in Table 1, for each TC basin and globally.

As detailed in section 2b, the TC intensity hindcast is based on documented relations between environmental parameters and the cyclone evolution (e.g., vertical shear in the atmosphere tends to damp cyclones). We use the 1979–2012, 0.75° × 0.75°, 6-hourly European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim; Dee et al. 2011) data to determine the atmospheric environmental parameters and Sea Surface Temperature (which is used to estimate MPI) along the cyclone tracks. Globally 2362 TCs occurred during this 34-yr period.

b. Linear hindcast model (MLR)

The environmental and the storm characteristic parameters that were used for building our statistical TC intensity hindcasts are listed in Table 2. The choice of these parameters closely follows that in previous studies (Knaff et al. 2005; Knaff and Sampson 2009; Sharma et al. 2013; Neetu et al. 2017). The predictors representing the storm characteristics are the TC maximum wind speed at the initial time of forecast (VMAX), its square (VMAX^2) and its change over the period of previous 12 h (PER). The environmental parameters along the cyclone track are the maximum potential intensity (MPI) and its square (MPI^2), vertical wind shear (SHRD), vertical zonal wind shear (USHR), tropospheric relative humidity (RHHI), upper-tropospheric

TABLE 2. List of environmental parameters used as predictors of the TC intensity in the current study. Variables in italics (VMAX2 and MPI2) are used only in the multiple linear regression (MLR) model but not in the artificial neural network (ANN) and support vector machine (SVM) models. Variables marked with a * are estimated from an area average between 200 and 800 km of the cyclone position. Variables marked with a ** are estimated from an area average within 1000 km of the cyclone position. The variables marked with a # are time averaged between the initial and the forecast time. DELV is the predictand.

Serial No.	Predictor	Description
1	VMAX	Initial intensity
2	*VMAX^2*	*Initial intensity squared*
3	PER	Intensity change during previous 12 h
4	MPI#	Maximum potential intensity
5	*MPI^2#*	*Maximum potential intensity squared*
6	SHRD*#	200–850-hPa wind shear magnitude
7	USHR*#	200–850-hPa zonal wind shear magnitude
8	RHHI*#	500–300-hPa average relative humidity
9	T200*#	200-hPa temperature
10	E925*#	925-hPa equivalent potential temperature
11	Z850**#	850-hPa vorticity
12	DELV	Intensity change from the initial forecast time

temperature (T200), boundary layer equivalent potential temperature (E925), and low-level vorticity (Z850). We do not use any predictor such as the ocean heat content (Leipper and Volgenau 1972) to represent the effect of the interactions with the ocean on the cyclone intensity evolution, as Neetu et al. (2017). Hindcasts over a shorter period over which reliable estimates of the oceanic state are available due to the availability of satellite altimetry (1993–2012) and including the ocean heat content as a predictor however give very similar results to those presented in the current paper (not shown).

Following Neetu et al. (2017), MPI is derived empirically for each TC basin, with the same empirical relations generally used in operational models. We use an exponential relation between MPI and sea surface temperature in the NWP, SWP, ATL and SIO basins and a linear relation in the NEP and NIO, in agreement with the previous literature (DeMaria et al. 2005; Knaff et al. 2005; Knaff and Sampson 2009; Kotal et al. 2008; Neetu et al. 2017). To account for the inhibiting effect of strong environmental vertical wind shear on the TC intensity, we use the average 200–850-hPa level zonal and total wind shear (USHR and SHRD), averaged over an annular region between 200 and 800 km away from the TC center. Other environmental parameters that affect the TC intensity include relative humidity (RHHI) at midtropospheric level (averaged between 300 and 500 hPa), upper-level temperature (T200 at 200 hPa), and lower-level equivalent potential temperature

(E925 at 925 hPa). RHHI, T200 and E925 are also averaged over an annular region between 200 and 800 km from the TC center. We also use relative vorticity at lower levels (Z850, 850 hPa), averaged within 1000 km from the TC center, as in Knaff et al. (2005) and Knaff and Sampson (2009). The environmental variables above are also averaged in time, from the beginning of the hindcast to the lead time for which the TC intensity is estimated.

TC intensity changes from the initial forecast time (DELV, see Table 2) at 12-h intervals up to 120-h lead time are the dependent variables (predictands) in our statistical schemes. Our linear statistical MLR model is similar to that of Neetu et al. (2017) and uses a multiple linear regression technique to relate DELV (the predictand) to all the predictors in Table 2. The data are partitioned into two groups, with 80% of the data being used for training the model and the remaining 20% for testing purposes (computing skill scores). Entire TC tracks are attributed to either the training or testing dataset when randomly generating those datasets. This procedure is repeated 50 times to generate 50 different randomly selected training and testing sets. All skill measures (see section 2c) in the current study are obtained by averaging the skill of these 50 models. This also allows 95% confidence intervals on the skill, using Monte Carlo techniques.

Neetu et al. (2017) demonstrated that this MLR was most skillful when trained separately for individual TC basins. In the current study, we either train the MLR (and the two nonlinear schemes) individually on each basin ("*regionally trained*") or globally considering all basins together ("*globally trained*").

c. Skill assessment

The model performance is assessed using MAE as in many similar studies (Knaff et al. 2005; Knaff and Sampson 2009; Sharma et al. 2013; Neetu et al. 2017). The MAE is defined as the average absolute value of the difference between the predicted TC intensity and the actual TC intensity in the IBTrACS database, across all TCs in the testing database. To ease the comparison between various models, we define the skill as the percentage of improvement in MAE of one model relative to a reference model:

$$\text{Skill}(\%) = 100 \times (\text{MAE}_{\text{ref}} - \text{MAE}_{\text{Model}})/\text{MAE}_{\text{ref}}.$$

Such a skill is of course dependent on the reference used. For instance, in this paper, we aim at estimating the added value of nonlinear schemes (ANN and SVM) relative to the MLR model. In such a case, the MLR will be used as the reference model. But we can also estimate

the performance of a globally trained model relative to a regionally trained model. Finally, in order to estimate the additional value of environmental parameters, we have built purely statistical hindcast models (hereafter "baseline" model, as in DeMaria et al. 2007; Neetu et al. 2017) that only use the initial cyclone characteristics as predictors (i.e., VMAX, VMAX^2, and PER for the MLR model).

3. Development of nonlinear TC intensity hindcasts

The previous section introduced the linear model (MLR) that we will take as a reference in the current study. In this section, we describe the nonlinear schemes that we will evaluate: the ANN (section 3a) and the SVM (section 3b). Since nonlinear schemes are supposed to naturally capture nonlinear dependencies between predictors and the predictand, nonlinear input parameters (VMAX^2 and MPI^2) were not included in these nonlinear schemes.

a. ANN scheme

Artificial neural networks are deep-learning algorithms that are used to solve artificial intelligence problems including forecasting and prediction. They can approximate any function (Hornik 1991) and are hence well suited to model nonlinear processes. Typically, the neural network consists of interconnected nodes called neurons, arranged in layers. Each input neuron transmits a linear function (multiplied by a weight with a bias removed) of the input data (predictors) and transmits the signal to next layer (hidden neurons). The weighted sum of inputs is transformed by a nonlinear activation function in the hidden neuron's layer. That process is repeated in each of the following hidden layers and the last neuron layer returns an output (predictand). The "training dataset" is used to determine the values of the weigths and biases that minimizes an optimization function (e.g., mean squared error between the computed output and correct outputs).

Our ANN model uses one input layer, one hidden layer, and one output layer. It is based on a hyperbolic tangent activation function. The input layer contains the same number of neurons as there are inputs parameters (i.e., predictors) while the output layer consists of a single neuron that predicts the TC intensification rate at a given lead time. A back-propagation learning algorithm (Reed and Marks 1999) was used to estimate the biases and weights during the training of the ANN model. Mean squared difference between the output and observations is a nonlinear function of the weights and biases, which has several local minima, where the model solution can be trapped. Another problem of nonlinear modeling is overfitting (i.e., when the model performs almost perfectly on the training dataset but is unable to generalize to new situations). To avoid overfitting on our ANN scheme, we use a technique called early stopping. For this, ANN requires three disjoint datasets for training, validation, and testing. The weights and biases (i.e., the parameters that define the neural network) are obtained during the training step, but the iterative procedure is stopped once errors relative to the validation dataset starts increasing. Finally, the testing dataset allows the computation of a skill score on a completely independent dataset to that used during the training/validation. In the following, the training dataset consistently uses a random selection of 60% of the TCs in the full database, while the remaining 40% are split into 20% for validation and 20% for testing purposes. We follow the same procedure as for the MLR and repeat this procedure 50 times with different randomly selected training, validation and testing realizations. We obtain the ANN parameters (weights and biases) values through an ensemble average of those 50 realizations. This procedure minimizes the problems of being trapped in a local minimum while solving the nonlinear problem.

More hidden layers generally improve the performance of the neural network but increase the risk of artificially overfitting. We have built ANN models with 1–10 hidden neuron layers, in order to determine the optimal number of layers for our problem. Figure 2a, displays the average training and testing MAE as a function of the number of hidden neurons. While the MAE in the training dataset monotonically decreases with the increasing number of neurons, it reaches a plateau or even slightly increases beyond about 5–7 neurons for the testing dataset. The ANN architecture therefore uses 7 neurons in the hidden layer throughout the paper, a number similar to that used in Sharma et al. (2013) study (5 neurons). This ANN architecture hence uses 70 adjustable parameters (i.e., weights) compared to 11 for the MLR. As discussed in Ingrassia and Morlini (2007), the equivalent number of degrees of freedom in the ANN is however much smaller than the number of parameters and does not depend on the number of input variables.

b. SVM scheme

SVM is a supervised learning method that stems from statistical learning theory (Vapnik 2000). SVMs were initially developed for classification purposes in the early 1990s, and later on used for regression problems (Vapnik 1995). The SVM technique requires the choice of a kernel function that maps
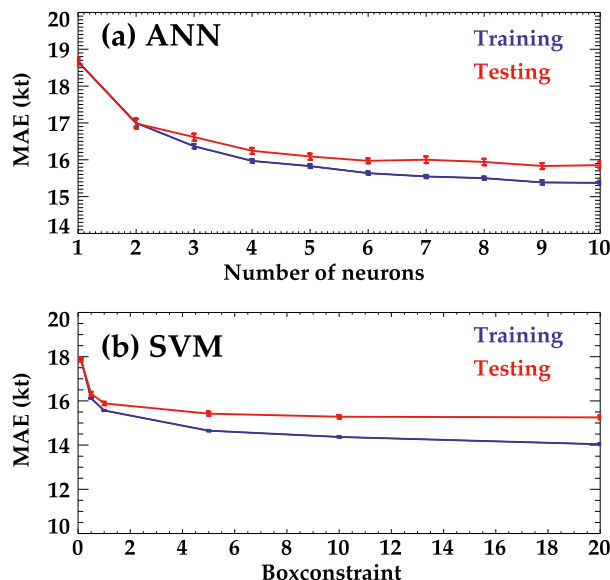
FIG. 2. Globally averaged MAE (kt) for training (blue) and testing (red) datasets at 60-h lead time for (a) the ANN scheme as function of number of neurons and (b) the SVM scheme as a function of the optimization parameter value (see text for details). The MAE was estimated from the average of 50 independent runs (see text for details). Error bars indicate the 95% confidence interval estimated from a bootstrap method. On the basis of this figure, the ANN is built using 7 neurons in the hidden layer and the SVM uses an optimization parameter of 10.

predictors into a higher-dimensional space where they relate linearly to the predictand. The properties of this kernel function are generally dependent on a small number of hyperparameters, that have to be set properly for an optimal performance of the SVM algorithm.

SVM only requires two disjoint sets of data for training and testing. As for the MLR, 80% (20%) of the TCs were used for training (testing). As for the ANN and MLR schemes, SVM models were constructed using 50 randomly selected training/testing datasets and the reported performance is the average over these 50 runs. We chose a radial basis kernel function (Scholkopf et al. 1997) because this widely adopted function (Keerthi and Lin 2003) often outperforms other kernel functions (Ding et al. 2012) and requires a limited number of hyperparameters to be set. The SVM performance depends on the choice of these hyperparameters. To illustrate this aspect, Fig. 2b shows the sensitivity of the SVM predictive skill for different values of the most sensitive hyperparameter (i.e., a smoothing parameter that controls the width of the kernel function and minimizes overfitting). While the MAE monotonically decreases with an increase of this hyperparameter for the training dataset, it only

decreases for values up to 10 for the testing dataset. Further increasing the value of this parameter results in a marginal MAE increase for testing, which deviate from the MAE for training, indicating that there is overfitting in this range. This parameter was therefore fixed to 10. The SVM predictive skills are much less sensitive to the choice of the other hyperparameters (not shown).

4. Results

As described in sections 2 and 3, we have developed three TC intensity hindcast schemes (MLR, ANN, and SVM), including two nonlinear ones (ANN and SVM). These schemes are constructed in a very consistent way, with identical input parameters (Table 2; except that the ANN and SVM scheme do not use the $VMAX^2$ and MPI^2 input parameters). All these schemes are also constructed using 80% of the TC database (with 60% for training and 20% for validation for ANN), and we show the average skill of 50 models trained with random choices of the trainingand testing databases. This allows a fair comparison between the three schemes. (Note that we also tested including the quadratic terms in Table 2 in the linear model as one alternative to using a fully nonlinear scheme, but that the ANN and SVM schemes clearly outperform this approach, not shown).

Neetu et al. (2017) demonstrated that training an MLR scheme individually for each basin performs better than a globally trained model. This may however not be the case for ANN and SVM. Figure 3 illustrates the benefits of training models globally relative to a regional training (see section 2c) for the three schemes. Regionally trained MLR models slightly outperform globally trained MLR models (~1%–5% on average over all basins; Fig. 3a). This superior performance is seen in all Northern Hemisphere TC-prone basins, but not for Southern Hemisphere TCs. It is particularly large for the NIO at long leads (~20%) but this result may not be very reliable given the small amount of long-lived TCs (and TCs in general) in the NIO, and hence the small dataset used to validate the MLR regional model at these extended range (Table 1).

The superior performance of regionally trained models does not hold for nonlinear schemes. Globally trained models behave similarly or slightly outperform the regionally trained ones for both nonlinear schemes. The globally averaged skill improvement for the globally trained ANN model ranges between ~0% and ~3% (Fig. 3b), with strong regional variations. This improvement is larger for the SWP, NIO and SIO basins (~5%–10%), while differences between regional and
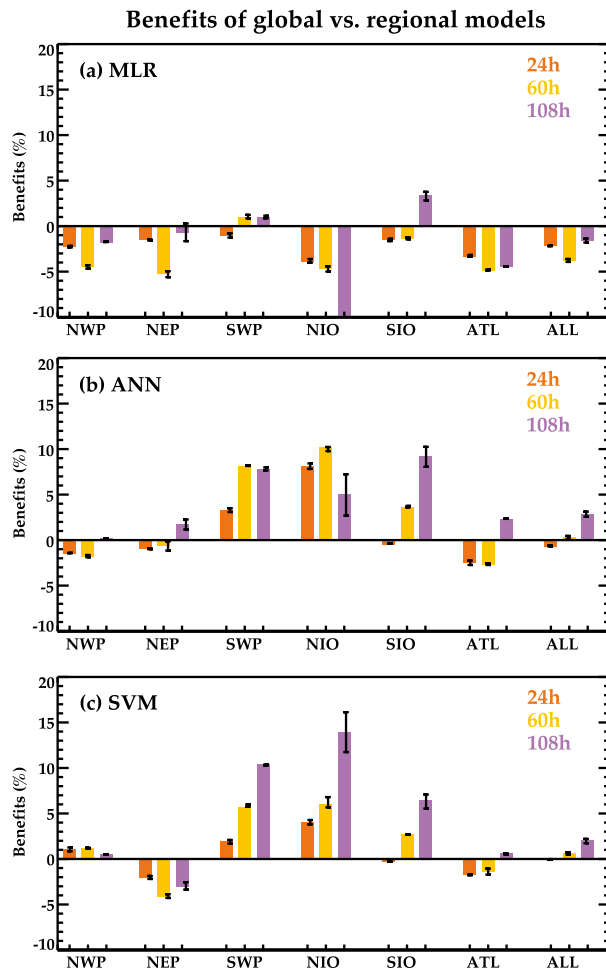
Benefits of global vs. regional models

FIG. 3. Percentage MAE improvement of globally trained over regionally trained models for (a) MLR, (b) ANN, and (c) SVM schemes at 24-, 60-, and 108-h lead times, for each basin and globally averaged. Error bars indicate the 95% confidence interval estimated from a bootstrap method.

globally trained ANN are rather modest for the NWP, NEP, and ATL basins The globally trained SVM model also slightly overcomes the regionally trained model, with an improvement of up to 2% globally (Fig. 3c). The regional dependency of this improvement is similar to that of the ANN, with the largest improvement in the SWP, NIO, and SIO (~2%–13%). In contrast, the skill is reduced for the NEP. Overall, the globally trained SVM and ANN perform better than their regionally trained counterparts, in particular in basins with the least cyclones (SWP, NIO, SIO; Fig. 1a). This suggests that the globally trained nonlinear schemes can use data from other basins to improve forecasts in basins with a small training sample. In contrast, the MLR scheme performs better when trained specifically for each basin. In the following, we will thus compare globally trained nonlinear models (which perform

best) to the best MLR models (i.e., the regionally trained ones).

The added value of using nonlinear models is then evaluated in Fig. 4 by computing the skill improvement of the globally trained ANN and SVM models relative to regionally trained MLR models (see section 2c). Using nonlinear models systematically improves the skill relative to the MLR at all lead times and in all basins (Figs. 4a,b), except at 108 h for the NIO in the case of ANN. Comparing light color bars on Figs. 4a and 4b also indicates that this improvement is systematically larger for SVM than for ANN. The globally averaged skill improvement relative to the MLR ranges between 4% and 7% for ANN (light color bars on Fig. 4a) and between 8% and 12% for SVM (light color bars in Fig. 4b). The additional skill yielded by nonlinear schemes is relatively uniform across TC basins (light color bars in Figs. 4a,b). There is thus a clear benefit of using nonlinear models for TC intensity forecasts, in any TC basin (except the NIO at long lead times for ANN).

Neetu et al. (2017) showed that the TC initial characteristics contributed to ~60%–80% of the overall skill in linear models. We will now investigate if this also holds true for our two nonlinear schemes, by comparing the performance of schemes using all predictors to "baseline models" using only the TC initial characteristics (see section 2c). Comparing dark and light colored bars in Fig. 4a reveals that the overall improvement brought by ANN and SVM predominantly arises from a better handling of TC environmental parameters, with a far weaker globally averaged improvement when only accounting for TC initial characteristics compared to the improvement when also accounting for environmental parameters. The overall improvement arising from a better handling of TC initial characteristics is however slightly larger for the SVM than for ANN model.

Figure 5 allows investigating if the respective contribution of individual parameters is similar for our two nonlinear schemes, by comparing the schemes using all predictors to schemes obtained by excluding each of the environmental parameters. In broad agreement with Neetu et al. (2017), SHRD (~3% contribution, globally) and MPI (~2%) are the environmental parameters that contribute most to the MLR skill globally (Fig. 5a), with very variable contributions depending on the basin. No single variable clearly dominates the performance of nonlinear schemes, and environmental parameters generally all contribute more to the performance than in the MLR. Only two parameters (SHRD and MPI) increase the model performance by more than 2% globally in the MLR, while four variables reach or exceed

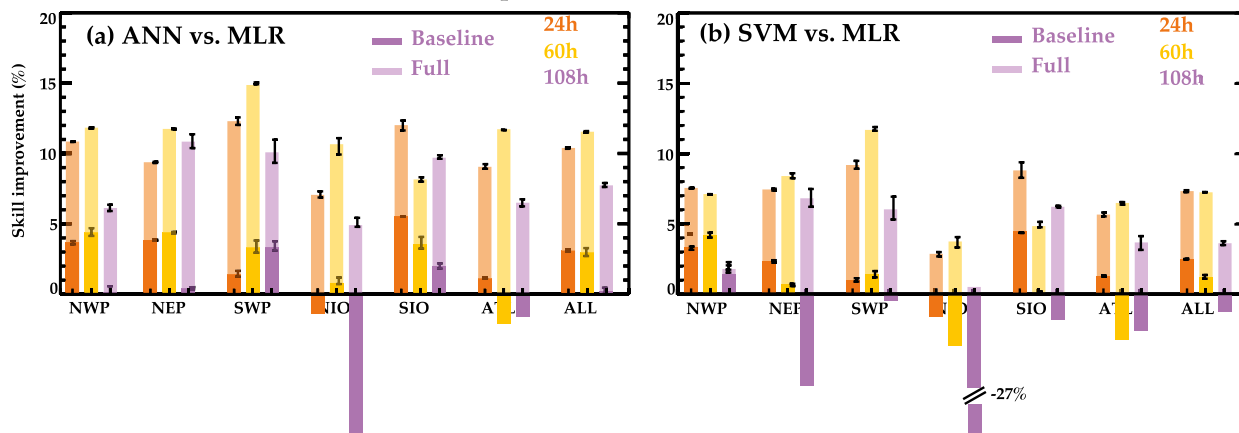## Skill improvement: Nonlinear vs. linear models

FIG. 4. Percentage of skill improvement for (a) globally trained ANN and (b) SVM relative to regionally trained MLR for baseline models (i.e., only TC initial characteristics, dark color bars) and full models (i.e., including also environmental parameters along the TC track, light color bars) at 24-, 60-, and 108-h lead times, for each basin and globally averaged. Error bars indicate the 95% confidence interval estimated from a bootstrap method.

that threshold for the nonlinear schemes (SHRD, MPI, T200, and E925). SHRD remains the most important environmental parameter globally in the ANN scheme (~3.5%), closely followed by MPI, RHHI, T200, and E925 (all around 2%). E925 is the most important environmental parameter globally in the SVM scheme (~5%), followed by SHRD, MPI, and T200 (all around 3%). The contribution of each predictor also varies considerably basin-wise for nonlinear schemes, but more parameters tend to contribute to the model performance than for the MLR. This suggests that nonlinear schemes are able to extract more information from environmental predictors than linear schemes.

Neetu et al. (2017) also demonstrated that an MLR model built considering climatological values for the input environmental parameters performs very similarly to a model constructed from real-time values. While this result implies that the practical implementation of operation TC intensity forecasts can considerably be simplified, it also suggests that the MLR technique only extracts little information from the environmental parameters. To investigate whether nonlinear schemes extract more information, we further compare the performance of experiments using real-time environmental predictors over ones using climatological values (Fig. 6). As in Neetu et al. (2017), Fig. 6a illustrates that the MLR model using real-time predictors does not add much to the model performance (1%–2%), with a slightly larger improvement for Southern Hemisphere TCs (up to 10% for SIO). The use of real-time predictors in nonlinear models results in a larger and more systematic improvement, in all basins and for all lead

times (Figs. 6b,c). The globally averaged added value ranges between 5% and 8% for both ANN and SVM (Figs. 6b,c), with a particularly large improvement for Southern Hemisphere TCs (7%–17%). The basin-wise dependency of the added value brought by real-time environmental parameters is quite similar to that of MLR, but it generally larger (Figs. 6b,c versus Fig. 6a). This suggests that ANN and SVM are able to better capture the influence of the nonseasonal variations of atmospheric parameters on the TC intensification rate.

Neetu et al. (2017) also demonstrated that MLR hindcasts are 3–4 times more skillful for strong than for moderate TCs. We followed the definition of Neetu et al. (2017) and divided the predictands into two subsets based on the TC intensity at hindcast time. Best track points associated with an intensity smaller than 96 kt (category 2 or below on the Saffir–Simpson scale) are referred to as "moderate" (~63% of the dataset), while those exceeding 96 kt (categories 3–5) are considered as "strong" (~37% of the dataset). As discussed in Neetu et al. (2017) for the MLR, nonlinear models predict intensity changes of strong TCs better than those of moderate ones (not shown). Figures 7a and 7b show the SVM and ANN global skill improvement relative to the regionally trained MLR schemes. The ANN scheme has a higher skill than the MLR, but this skill improvement does not particularly depend on whether the cyclone is strong or not. In contrast, the SVM scheme tends to yield a larger skill increase for moderate TCs (21% at 108 h) than for strong TCs (13% at 108 h). The SVM scheme hence partially overcomes the weaker MLR performance for moderate TCs.

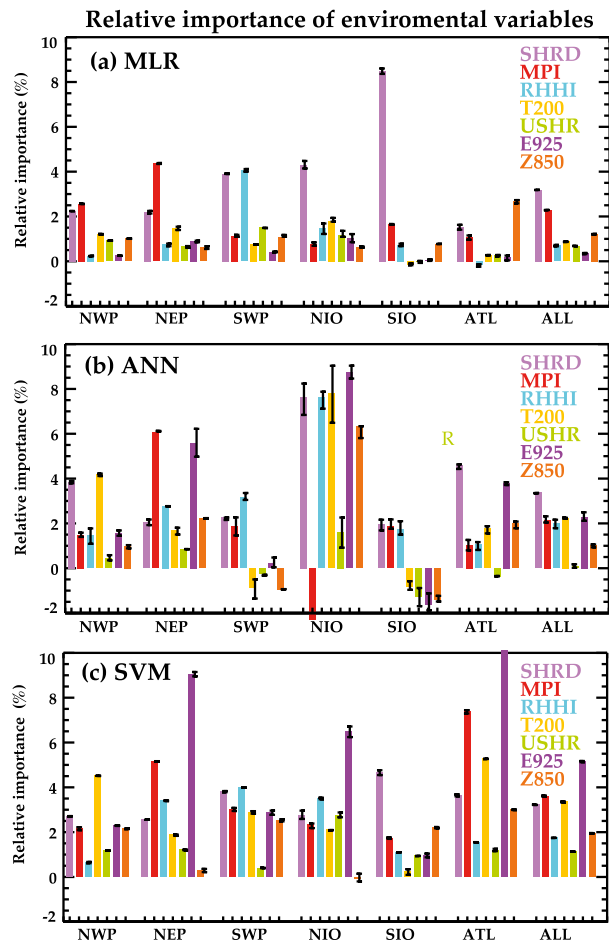### Relative importance of enviromental variables

FIG. 5. Respective contributions of each predictor to the overall skill for (a) regionally trained MLR, (b) globally trained ANN, and (c) globally trained SVM for each basin and globally. These contributions are estimated as the percentage of skill reduction at 60-h lead time when excluding all the predictors one by one within each model framework. Error bars indicate the 95% confidence interval estimated from a bootstrap technique. In the MLR scheme, the $VMAX^2$ (MPI^2) predictor is also removed when testing the impact of VMAX (MPI).

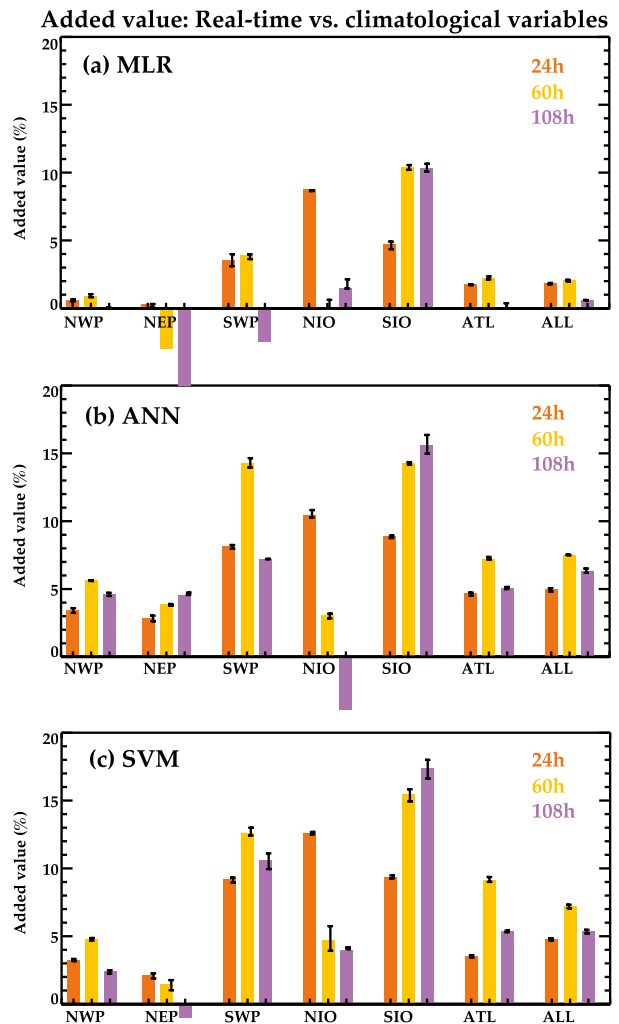### Added value: Real-time vs. climatological variables

FIG. 6. Added value of using real-time rather than climatological environmental atmospheric parameters for (a) regionally trained MLR, (b) globally trained ANN, and (c) globally trained SVM models, at 24-, 60-, and 108-h lead times, for each basin and globally averaged. This added value is measured as the percentage of skill improvement when using real-time environmental atmospheric parameters against climatological ones in each model. Error bars indicate the 95% confidence interval estimated from a bootstrap technique.

Lee et al. (2016) did show that linear models tend to have a lower skill for predicting decaying TCs than for intensifying TCs. We hence tested the skill sensitivity to whether the TC is intensifying or decaying. To do so, we divided the predictands into two subsets based on the sign of the TC intensification rate at the beginning of the hindcast. Linear models tend to have a higher skill for intensifying (~40% skill improvement relative to persistence at 24 h) than for decaying TCs (~30%; not shown). Figures 7c and 7d shows the SVM and ANN global skill improvement relative to the regionally trained MLR schemes: nonlinear schemes more specifically outperform linear schemes for decaying TCs. This is particularly striking for the SVM scheme, with a skill improvement of ~7%–10% relative to the MLR for the intensifying phase and 16%–22% for the decaying phase. The SVM in fact yields a similar skill improvement for intensifying or decaying TC relative to performance (46% at 24 h; not shown) (i.e., it corrects the tendency of the MLR to perform less well for decaying TCs).

5. Conclusions and discussion

Our results indicate that nonlinear schemes (ANN and SVM) systematically improve the skill of TC hindcasts

Skill improvement: TCs characterisitcs
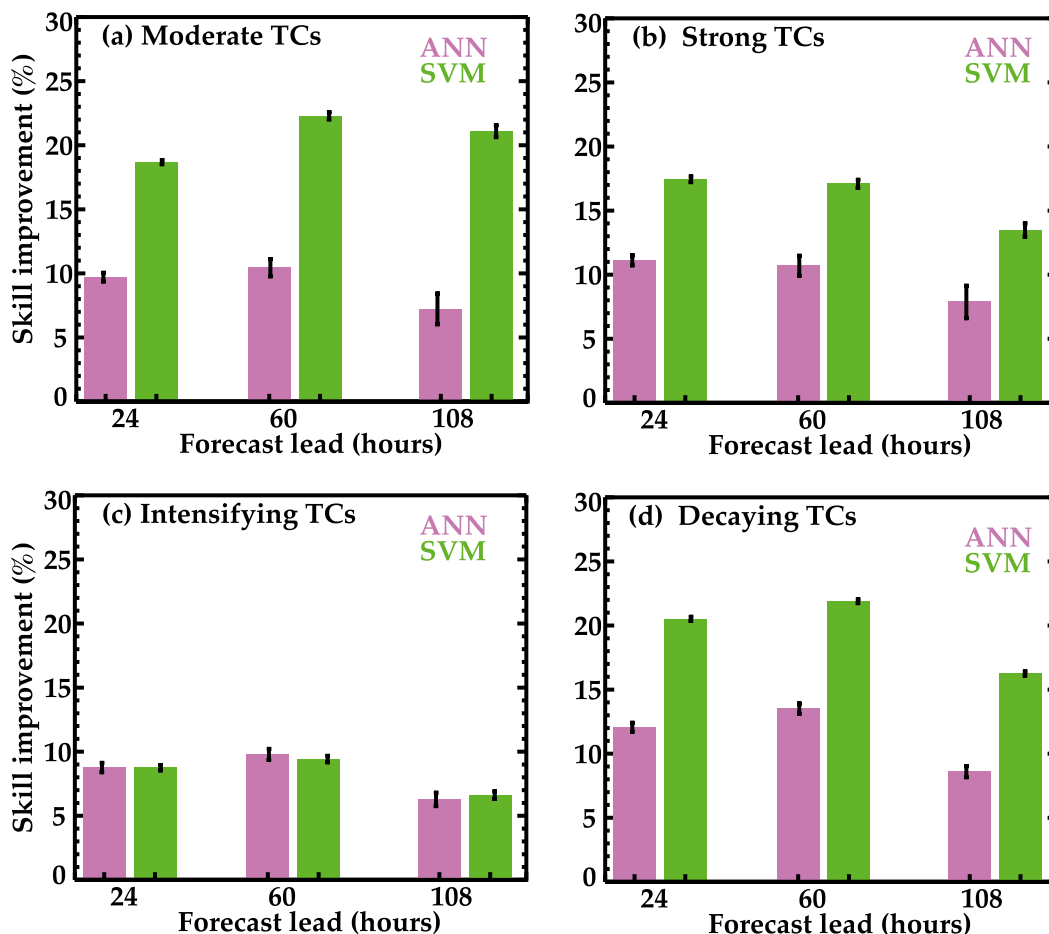

FIG. 7. Globally averaged percentage of skill improvement relative to regionally trained MLR for (a) moderate (i.e., <96 kt: category 2 and below) and strong TCs (i.e., >96 kt: category 3 or more) and for (b) intensifying (DELV > 0) and decaying TCs (DELV < 0) for globally trained ANN (blue) and SVM (green) schemes at 24-, 60-, and 108-h lead times. Error bars provide the 95% confidence interval estimated from a bootstrap method.

relative to the widely used linear models (MLR) by 5%–15%, in all TC-prone basins, for all lead times and all TC categories (except for the NIO at long lead times in the ANN model). These results hence demonstrate that accounting for the nonlinear relationship between input parameters and the TC intensity greatly benefits the hindcast skill of statistical–dynamical forecasts. In addition, the comparison of the ANN and SVM schemes reveals a better performance of SVM, with 10%–15% skill improvement over the MLR globally versus 6%–11% for ANN). The SVM scheme also corrects the tendency for a lower linear scheme skill in predicting moderate (category 2 or below) and decaying TCs intensity evolution. This improved performance of SVM likely results from two desirable mathematical properties of this approach. Contrary to

ANN, SVM transforms the nonlinear optimization problem into a linear one, hence avoiding the multiple minima issue and providing a more robust estimate of the global solution. It also uses a more robust error norm, with error rates and model complexity minimized simultaneously, while only the error rate is minimized once the model architecture has been designed for ANN. This usually results in SVM performing better than ANN when moving from the training to the testing dataset (Bisgin et al. 2018).

Our results echo those of the few other studies that compared the performance of linear and nonlinear schemes (DeMaria 2009; Lin et al. 2013; Sharma et al. 2013). For the 2006/07 cyclonic seasons, the nonlinear logistic equation growth model of DeMaria (2009) yields a skill increase from 0% to 15% in the northeast

Pacific and from 0% to 10% in the Atlantic relative to SHIPS, especially at long lead times, with a simpler set of input parameters. Lin et al. (2013) found an improvement of 5%–10% when comparing their nonlinear SVM scheme to operational MLR forecasts in the northwest Pacific over the 2002–09 period. Our results for the NWP indicate an even larger improvement (∼8%–16%; see Fig. 4b). However, Lin et al. (2013) did not use the same set of input parameters as the operational forecast, as we do, which makes a quantitative comparison difficult. Sharma et al. (2013) on the other hand used the same parameters as the linear STIPS model (Knaff et al. 2005) and found a 3%–10% skill improvement for their nonlinear ANN scheme at all lead times for the 2003–04 TC seasons in the NWP. We find a similar improvement (5%–11%, Fig. 4a) over the longer 1979–2012 period. By applying nonlinear TC intensity forecasting methods to all TC basins globally, the present study extends previous results obtained over a limited set of basins (NWP and ATL) and limited period (∼7 years vs to 34 years here), and demonstrates the added value of using nonlinear schemes for TC intensity prediction for all basins, lead time, and TC intensity. This improvement arises from the nonlinear relationship between the environmental parameters and intensity of the cyclone, which are more easily captured by nonlinear schemes.

Sensitivity experiments using environmental predictors calculated from climatological data (rather than real-time values) indicate that nonlinear schemes make a far better use of the nonseasonal variations of these environmental predictors. Accounting for these real-time variations indeed yield a far larger improvement (5%–8% for ANN and SVM, globally) than the linear scheme (∼1%–3%).

Finally, our results also demonstrate that the SVM scheme is particularly efficient at improving the hindcast skill of moderate TCs (category 2 and below), which were particularly poor for linear models (Neetu et al. 2017). Similarly, the SVM scheme corrects the tendency of linear models to have a lower skill for decaying than for intensifying TCs (Lee et al. 2016). This larger improvement for moderate TCs (which represent 63% of TCs worldwide) and decaying TCs is a strong incentive for using SVM TC intensity prediction schemes operationally.

Our results indicate that nonlinear schemes have a strong potential to improve statistical TC intensity forecasts, especially SVM schemes. The best TC intensity forecasts have improved by ∼8%–15% over the last decade, both in dynamical and linear statistical-dynamical models (DeMaria et al. 2014; Emanuel and Zhang 2016). The skill gains from using nonlinear schemes

are thus comparable to the TC intensity forecast skill gain over a decade. While we expect a skill degradation for operational application, due to use of forecasted track and environment, rather than reanalyzed environmental parameters along the observed track, the TC intensity skill improvement expected from the present results are sufficiently large to motivate more trials in operational mode.

Finally, the good performance of global training for nonlinear schemes also has interesting applications. The first is that a single system can be built and applied to various basins, in contrast with current statistical–dynamical schemes, which are generally built for a single or a couple of TC basins. This technique also offers great hope in basins such as the northern Indian Ocean, where the limited number of TCs prevents an efficient training of MLRs, yielding poor skill. The ability of the nonlinear schemes to be trained globally takes advantages of the numerous cyclones in other basins, yielding a skill improvement of up to 10% relative to the MLR at 108-h lead time in the Northern Indian Ocean for the SVM scheme (Fig. 4b). This offers great promise in this basin that just represents 5% of the TCs globally, but almost 80% of the casualties they cause (Needham et al. 2015).

REFERENCES

Aguilar-Martinez, S., and W. W. Hsieh, 2009: Forecasts of tropical Pacific sea surface temperatures by neural networks and support vector regression. *Int. J. Oceanogr.*, **2009**, 167239, https://doi.org/10.1155/2009/167239.

Ali, M. M., C. M. Kishtawal, and J. Sarika, 2007: Predicting cyclone tracks in the north Indian Ocean: An artificial neural network approach. *Geophys. Res. Lett.*, **34**, L04603, https://doi.org/10.1029/2006GL028353.

Bisgin, H., and Coauthors, 2018: Comparing SVM and ANN based machine learning methods for species identification of food contaminating beetles. *Sci. Rep.*, **8**, 6532, https://doi.org/10.1038/s41598-018-24926-7.

Bushnell, J. M., and R. J. Falvey, 2018: Annual tropical cyclone report 2017. Joint Typhoon Warning Center Report, 133 pp., https://www.metoc.navy.mil/jtwc/products/atcr/2017atcr.pdf.

Cangialosi, J. P., 2019: National Hurricane Center forecast verification report: 2018 hurricane season. NOAA/NHC, 73 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2018.pdf.

——, and J. L. Franklin, 2014: 2013 National Hurricane Center Forecast Verification Report. NOAA/NWS/NCEP/National Hurricane Center, 84 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2013.pdf.

Cortes, C., and V. N. Vapnik, 1995: Support-vector networks. *Mach. Learn.*, **20**, 273–297, https://doi.org/10.1007/BF00994018.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

DeMaria, M., 1996: The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.*, **53**, 2076–2088, https://doi.org/10.1175/1520-0469(1996)053<2076:TEOVSO>2.0.CO;2.

——, 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.*, **137**, 68–82, https://doi.org/10.1175/2008MWR2513.1.

——, and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220, https://doi.org/10.1175/1520-0434(1994)009%3C0209:ASHIPS%3E2.0.CO;2.

——, and ——, 1999: An updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337, https://doi.org/10.1175/1520-0434(1999)014<0326:AUSHIP>2.0.CO;2.

——, M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvements in the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543, https://doi.org/10.1175/WAF862.1.

——, J. A. Knaff, and C. R. Sampson, 2007: Evaluation of long-term trend in tropical cyclone intensity forecasts. *Meteor. Atmos. Phys.*, **97**, 19–28, https://doi.org/10.1007/s00703-006-0241-4.

——, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, https://doi.org/10.1175/BAMS-D-12-00240.1.

Descloux, E., and Coauthors, 2012: Climate-based models for understanding and forecasting dengue epidemics. *PLOS Negl. Trop. Dis.*, **6**, e1470, https://doi.org/10.1371/journal.pntd.0001470.

Ding, C., L.-F. Yuan, S.-H. Guo, H. Lin, and W. Chen, 2012: Identification of mycobacterial membrane proteins and their types using over-represented tripeptide compositions. *J. Proteomics*, **77**, 321–328, https://doi.org/10.1016/j.jprot.2012.09.006.

Elbisy, M. S., 2015: Sea wave parameters prediction by support vector machine using a genetic algorithm. *J. Coastal Res.*, **314**, 892–899, https://doi.org/10.2112/JCOASTRES-D-13-00087.1.

Elsberry, R. L., L. S. Chen, J. Davidson, R. Rogers, Y. Wang, and L. Wu, 2013: Advances in understanding and forecasting rapidly changing phenomena in tropical cyclones. *Trop. Cyclone Res. Rev.*, **2**, 13–24, https://doi.org/10.6057/2013TCRR01.02.

Emanuel, K. A., 1995: Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *J. Atmos. Sci.*, **52**, 3969–3976, https://doi.org/10.1175/1520-0469(1995)052<3969:SOTCTS>2.0.CO;2.

——, 2007: Environmental factors affecting tropical cyclone power dissipation. *J. Climate*, **20**, 5497–5509, https://doi.org/10.1175/2007JCLI1571.1.

——, and F. Zhang, 2016: On the predictability and error sources of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **73**, 3739–3747, https://doi.org/10.1175/JAS-D-16-0100.1.

——, C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858, https://doi.org/10.1175/1520-0469(2004)061<0843:ECOTCI>2.0.CO;2.

Fine, T. L., 1999: *Feedforward Neural Network Methodology*. Springer-Verlag, 356 pp.

Gray, W. M., 1968: A global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669–700, https://doi.org/10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2.

Hennon, C. C., C. Marzban, and J. S. Hobgood, 2005: Improving tropical cyclogenesis statistical model forecasts through the application of a neural network classifier. *Wea. Forecasting*, **20**, 1073–1083, https://doi.org/10.1175/WAF890.1.

Holland, G. J., 1997: The maximum potential intensity of tropical cyclones. *J. Atmos. Sci.*, **54**, 2519–2541, https://doi.org/10.1175/1520-0469(1997)054<2519:TMPIOT>2.0.CO;2.

Hornik, K., 1991: Approximation capabilities of multilayer feedforward networks. *Neural Networks*, **4**, 251–257, https://doi.org/10.1016/0893-6080(91)90009-T.

Ingrassia, S., and I. Morlini, 2007: Equivalent number of degrees of freedom for neural networks. *Advances in Data Analysis*, R. Decker and H.-J. Lenz, Eds., Springer, 229–236.

Kaplan, J., and Coauthors, 2015: Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Wea. Forecasting*, **30**, 1374–1396, https://doi.org/10.1175/WAF-D-15-0032.1.

Keerthi, S. S., and C.-J. Lin, 2003: Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput.*, **15**, 1667–1689, https://doi.org/10.1162/089976603321891855.

Knaff, J. A., and C. R. Sampson, 2009: Southern Hemisphere tropical cyclone intensity forecast methods used at the Joint Typhoon Warning Center. Part II: Statistical-dynamical forecasts. *Aust. Meteor. Oceanogr. J.*, **58**, 9–18, https://doi.org/10.22499/2.5801.002.

——, ——, and M. DeMaria, 2005: An operational statistical typhoon intensity prediction scheme for the western North Pacific. *Wea. Forecasting*, **20**, 688–699, https://doi.org/10.1175/WAF863.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

Kotal, S. D., S. K. Roy Bhowmik, P. K. Kundu, and A. K. Das, 2008: A Statistical Cyclone Intensity Prediction (SCIP) model for the Bay of Bengal. *J. Earth Syst. Sci.*, **117**, 157–168, https://doi.org/10.1007/s12040-008-0006-1.

Kucas, M. E., 2010: Challenges of forecasting tropical cyclone intensity change at the Joint Typhoon Warning Center. *29th Conf. on Hurricanes and Tropical Meteorology*, Phoenix, AZ, Amer. Meteor. Soc., 9C.6, https://ams.confex.com/ams/pdfpapers/168732.pdf.

Landsea, C. W., and J. P. Cangialosi, 2018: Have we reached the limits of predictability for tropical cyclone track forecasting? *Bull. Amer. Meteor. Soc.*, **99**, 2237–2243, https://doi.org/10.1175/BAMS-D-17-0136.1.

Lee, C., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2016: Autoregressive modeling for tropical cyclone intensity climatology. *J. Climate*, **29**, 7815–7830, https://doi.org/10.1175/JCLI-D-15-0909.1.

Lee, T.-L., 2006: Neural network prediction of a storm surge. *Ocean Eng.*, **33**, 483–494, https://doi.org/10.1016/J.OCEANENG.2005.04.012.

——, 2009: Predictions of typhoon storm surge in Taiwan using artificial neural networks. *Adv. Eng. Software*, **40**, 1200–1206, https://doi.org/10.1016/j.advengsoft.2007.06.005.

Leipper, D. F., and D. Volgenau, 1972: Hurricane heat potential of the Gulf of Mexico. *J. Phys. Oceanogr.*, **2**, 218–224, https://doi.org/10.1175/1520-0485(1972)002<0218:HHPOTG>2.0.CO;2.

Lin, G. F., G. R. Chen, M. C. Wu, and Y. C. Chou, 2009: Effective forecasting of hourly typhoon rainfall using support vector machines. *Water Resour. Res.*, **45**, W08440, https://doi.org/10.1029/2009WR007911.

——, P. K. Huang, and H. Y. Lin, 2013: Forecasting tropical cyclone intensity change in the western North Pacific. *J. Hydroinform.*, **15**, 952–966, https://doi.org/10.2166/hydro.2013.155.

Lin, N., R. Jing, Y. Wang, E. Yonekura, J. Fan, and L. Xue, 2017: A statistical investigation of the dependence of tropical cyclone intensity change on the surrounding environment. *Mon. Wea. Rev.*, **145**, 2813–2831, https://doi.org/10.1175/MWR-D-16-0368.1.

Liu, Z. L., C. H. Peng, W. H. Xiang, D. L. Tian, X. W. Deng, and M. F. Zhao, 2010: Application of artificial neural networks in global climate change and ecological research: An overview. *Chin. Sci. Bull.*, **55**, 3853–3863, https://doi.org/10.1007/s11434-010-4183-3.

McPhaden, M. J., and Coauthors, 2009: Ocean-atmosphere interactions during Cyclone Nargis. *Eos, Trans. Amer. Geophys. Union*, **90**, 53–54, https://doi.org/10.1029/2009EO070001.

Miller, B. I., 1958: On the maximum intensity of hurricanes. *J. Meteor.*, **15**, 184–195, https://doi.org/10.1175/1520-0469(1958)015<0184:OTMIOH>2.0.CO;2.

Needham, H. F., B. D. Keim, and D. Sathiaraj, 2015: A review of tropical cyclone-generated storm surges: Global data sources, observations, and impacts. *Rev. Geophys.*, **53**, 545–591, https://doi.org/10.1002/2014RG000477.

Neetu, S., and Coauthors, 2017: Global assessment of tropical cyclone intensity statistical-dynamical hindcasts. *Quart. J. Roy. Meteor. Soc.*, **143**, 2143–2156, https://doi.org/10.1002/qj.3073.

Reed, R. D., and R. J. Marks II, 1999: *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. The MIT Press, 346 pp.

Roy, C., and R. Kovordányi, 2012: Tropical cyclone track forecasting techniques—A review. *Atmos. Res.*, **104–105**, 40–69, https://doi.org/10.1016/j.atmosres.2011.09.012.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning internal representations by error propagation. *Foundations*, D. E. Rumelhart and J. L. Mcclelland, Eds., Vol. 1, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, 318–362.

Scholkopf, B., K. K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, 1997: Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans. Sig. Proc.*, **45**, 2758–2765, https://doi.org/10.1109/78.650102.

Sharma, N., M. M. Ali, J. A. Knaff, and P. Chand, 2013: A soft-computing cyclone intensity prediction scheme for the Western North Pacific Ocean. *Atmos. Sci. Lett.*, **14**, 187–192, https://doi.org/10.1002/asl2.438.

Tang, B., and K. E. Emanuel, 2012: A ventilation index for tropical cyclones. *Bull. Amer. Meteor. Soc.*, **93**, 1901–1912, https://doi.org/10.1175/BAMS-D-11-00165.1.

Tolman, H. L., V. M. Krasnopolsky, and D. V. Chalikov, 2005: Neural network approximations for nonlinear interactions in wind wave spectra: Direct mapping for wind seas in deep water. *Ocean Modell.*, **8**, 253–278, https://doi.org/10.1016/j.ocemod.2003.12.008.

Vapnik, V. N., 1995: *The Nature of Statistical Learning Theory*. 1st ed. Springer, 188 pp.

——, 2000: *The Nature of Statistical Learning Theory*. 2nd ed. Springer-Verlag, 314 pp.

Wei, C.-C., 2012: Wavelet support vector machines for forecasting precipitation in tropical cyclones: Comparisons with GSVM, regression, and MM5. *Wea. Forecasting*, **27**, 438–450, https://doi.org/10.1175/WAF-D-11-00004.1.