
Article

Forecasting the June Ridge Line of the Western Pacific Subtropical High with a Machine Learning Method

Cunyong Sun ¹, Xiangjun Shi ^{1,*}, Huiping Yan ¹, Qixiao Jiang ¹ and Yuxi Zeng ¹

¹ School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China

* Correspondence: shixj@nuist.edu.cn

Abstract: The ridge line of the western Pacific subtropical high (WPSHRL) plays an important role in determining the shift of the summer rain belt in eastern China. In this study, we developed a forecast system for the June WPSHRL index based on the latest autumn and winter sea surface temperature (SST). Considering the adverse condition of the small observed sample size, a very simple neural network (NN) model was selected to extract the non-linear relationship between input predictors (SST) and target predictands (WPSHRL) in the forecast system. In addition, some techniques are used to deal with the adverse condition, enhance the stabilization of forecast skills, and analyze the interpretability of the forecast system. The forecast experiments show that the linear correlation coefficient between the predictions from the forecast system and their corresponding observations is around 0.6, and about three-fifths of the observed abnormal years (the years with an obviously high or low WPSHRL index) are successfully predicted. Furthermore, sensitivity experiments show that the forecast system is relatively stable in terms of forecast skill. The above evaluations suggest that the forecast system is valuable in a real application sense.

Keywords: machine learning; neural network; forecasting system; western Pacific subtropical high

1. Introduction

The ridge line of western Pacific subtropical high (WPSHRL) plays an important role in determining the shift of the summer rain belt in eastern China [1-7]. Over the last few decades, the prediction of the WPSHRL has been widely studied based on traditional statistical methods [8-15] and numerical simulation methods (i.e., dynamical forecast) [16-21]. Unfortunately, there are few studies on the prediction of the WPSHRL more than three months ahead of time, and the forecast skill (evaluated by correlation coefficient) remains unsatisfactory (< 0.4) at lead times of more than one season [21].

In recent years, machine learning methods have been widely used for climate forecast systems and have obviously improved the forecast beyond the seasonal time scale [22-30]. However, these studies mostly focus on the prediction of sea surface temperature (SST) indexes (e.g., El Niño-Southern Oscillation, ENSO). There are few studies about the prediction of general atmospheric circulation indexes more than three months in advance. One major purpose of this study is to investigate whether the machine learning method can provide a skillful prediction of the WPSHRL for lead times longer than one season.

Machine learning methods with non-linear statistical models can extract the non-linear relationship between input predictors and target predictands, which cannot be handled with traditional linear statistical methods (e.g., multivariate linear regression) [31-34]. Because non-linear statistical models (e.g., neural network approaches) are usually much more complex (i.e., too many undetermined parameters), a sufficient number of samples is essential to train the model [35-37]. In terms of climate forecast, the observation period is usually too short to achieve proper training [23,38]. For example, for the observation of the June WPSHRL from 1961 to 2021, the number of samples is only 61. Considering this limitation, we decided to use a simple neural network (NN) model (dozens of

parameters) [31,32]. Furthermore, developing a stable and reliable NN-based forecast system is a multifaceted problem that requires a deep understanding of the technical aspects of the training process and details of NN architecture [39-45]. In this study, some techniques are used to extend the samples of training data, enhance the stabilization of forecast skills, and analyze the interpretability of the forecast system.

The goal of this study is to design a forecast system for the June WPSHRL index based on the latest autumn and winter SST. The paper is organized as follows: the observed data and the NN-based forecast system are described in Section 2, followed by the application introduction and performance analysis in Section 3; finally, the conclusions and discussion are provided in Section 4.

2. Data and Methods

2.1. Data

The global SST data provided by the Physical Sciences Laboratory (PSL) were used in this study. The data set is stored on a $5^{\circ} \times 5^{\circ}$ grid and consists of monthly anomalies from 1856 to present [46]. The monthly WPSHRL indexes (integer degree of latitude) from 1961 to 2016 were downloaded from the National Climate Center of China (NCC). This data set has been widely used for climatic prediction in China [4,6,14,15,20,47,48]. The WPSHRL index (manual statistical data) is defined as the averaged 0 line of $\partial gh/\partial y$ in the area of ($10^{\circ}\text{N}\sim 45^{\circ}\text{N}$, $110^{\circ}\text{E}\sim 150^{\circ}\text{E}$) based on the monthly mean 500 hPa geopotential height (gh) field. More details about the calculating method can be found in the study of Liu [48]. Because the NCC no longer provided this WPSHRL data after 2016, we artificially figured out the monthly WPSHRL index from 2017 to 2021 based on the reanalysis data, following the above definition.

The SST shows an obvious decadal-scale warming trend due to global warming (not shown). This decadal-scale trend is removed by the 21-point (i.e., year) moving average method. The WPSHRL indexes are normalized to the range of -1 to 1. It is noteworthy that only the Pacific and Indian ocean ($60^{\circ}\text{S}\sim 60^{\circ}\text{N}$, $30^{\circ}\text{E}\sim 90^{\circ}\text{W}$) season-average (only autumn and winter) SST are used as input predictors. In order to increase the samples of training data, we artificially produced some samples based on the composite analysis. Figure 1 presents the composites in the low (the lowest $\sim 1/4$) and high (the highest $\sim 1/4$) WPSHRL index years. This composite analysis provides a kind of background knowledge about the linear relationship between predictors (i.e., SST field) and predictands (i.e., WPSHRL indexes). This linear relationship is obvious. The WPSHRL index is apt to be low/high if the pattern of prophase SST anomalies is similar to the low/high composite. Furthermore, the WPSHRL index is assumed to be enhanced/weakened if the composited prophase SST anomalies are enhanced/weakened. Based on the above background knowledge, 40 composite samples (20 low index samples and 20 high index samples) are produced by magnifying or reducing (i.e., multiplying by a random value around 1) the composite data (both SST and WPSHRL).

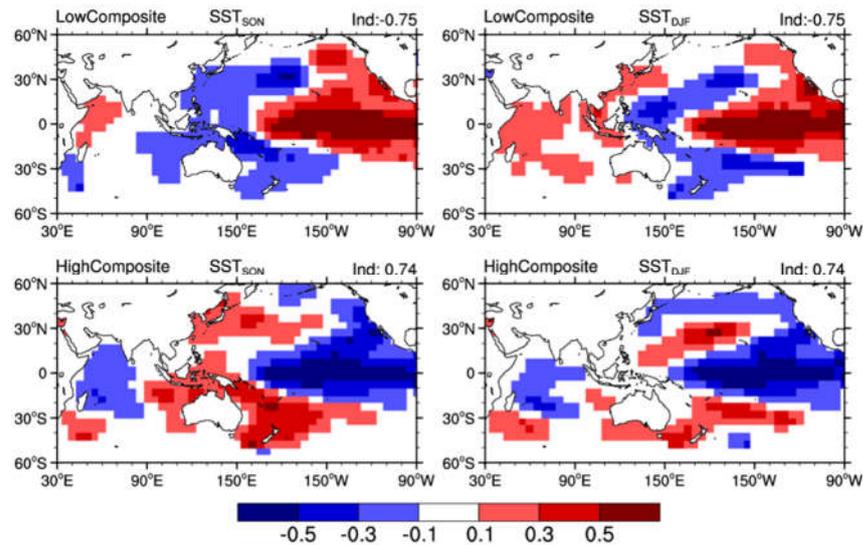


Figure 1. Composites of the prophase autumn (SON, left) and winter (DJF, right) sea surface temperature (SST) anomalies in low (upper panel) and high (lower panel) WPSHRL index years. The corresponding low and high composited WPSHRL indexes are -0.75 and 0.74 , respectively (upper right corner).

2.2. Forecast system

Figure 2 shows the architecture of the NN-based forecast system for the June WPSHRL indexes. The 20 leading empirical orthogonal function (EOF) eigenvectors of observed autumn and winter SST data from 1960 to 2021 are first stored as background knowledge. The N ($N \leq 20$, tunable parameter) leading EOF principal components of the latest SST data are used as predictors (i.e., the input layer of the NN model). There are two hidden layers with tunable neuron numbers (h_1 and h_2). The model architecture is constrained by N , h_1 , and h_2 , which are called the model's hyperparameters [35]. How to obtain the optimal architecture hyperparameters (i.e., N , h_1 , and h_2) is introduced in the next section.

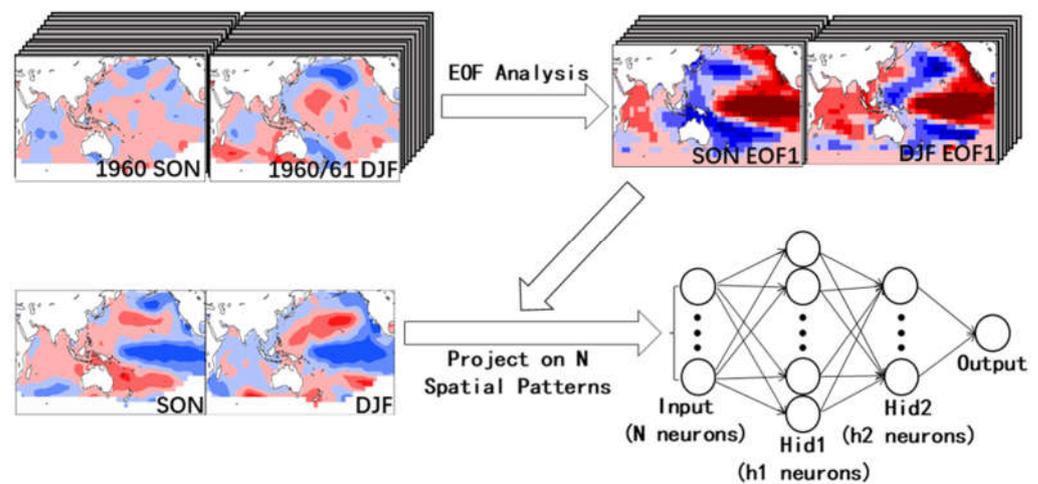


Figure 2. Architecture of the NN-based model used for the June WPSHRL index forecast. The 20 leading EOF eigenvectors of observed SON and DJF SST data from 1960 to 2021 are stored as background knowledge. The N ($N \leq 20$, tunable parameter) leading principal components of input SST (a certain year) obtained by calculating the pattern projection coefficients onto the EOF eigenvectors are used as predictors. The number of hidden layers is two, and the number of neurons at each hidden layer (h_1 , h_2) is also a tunable parameter. The activation function is a hyperbolic tangent function.

The NN model used here is the standard feed-forward multi-layer perceptron NN model. In order to minimize the root-mean-square-error (RMSE) between the predictions (i.e., outputs) and the targets (i.e., the observed WPSHRL indexes), the model's parameters (i.e., the weights and biases associated with each layer) are adjusted through gradient descent [49]. During every update of the model's parameters (i.e., training phase), one-third of the training set (i.e., batch size) is used to calculate the gradients. After every update with the training set (only the observed samples), the artificially produced samples are also used for training the model. The initial biases are random values in the range from -0.01 to 0.01 . The initial weights (i.e., the undetermined parameters of the NN model) are normally distributed, and their variance is regularized by the square root of the layer neuron number [50]. The initial learning rate is set to 0.001 , and it can be adaptively adjusted during model training. As training goes on, the model can be adjusted to best fit the given training set, which causes an overfitting problem. Usually, some samples are reserved as the test set to evaluate the overfitting problem. The training is stopped once the error (i.e., RMSE) from the test set begins to increase. Unfortunately, this common approach is not suited for this study because the total sample size is too small. Here, there is no test set. The NN model would stop training once the training error (i.e., the RMSE from the training set) is less than a threshold value (hereafter stopping threshold, ST). How to set this hyperparameter (i.e., ST) is introduced in the next section.

The NN model is trained by the training set. However, what we actually care about is the performance of the model on new unseen examples (i.e., generalization problem). Usually, some samples are reserved as the validation set to evaluate the model's forecast skill. Considering the small sample size in this study, the leave-one-out cross validation approach was used to quantify the forecast skill of the NN model [35,51]. Just as its name implies, one subset is chosen as the validation set and the rest of the subsets are used as the training set, and so on until all subsets have acted as validation set. For instance, the whole observed data set (e.g., 1961~2021) is divided into 61 subsets, each containing one sample. Firstly, the model is trained based on the 2nd~61st samples and the 1st prediction (i.e., the predicted 1961 WPSHRL index) is calculated with this model using the SST data of the 1st sample. Subsequently, the model is trained again based on the 1st and 3rd~61st samples and the 2nd prediction (i.e., the predicted 1962 WPSHRL index) is calculated with this new model using the SST data of the 2nd sample. This procedure is repeated until the 61st prediction (i.e., the predicted 2021 WPSHRL index) has been calculated. Finally, the forecast skill of the NN model is evaluated by the Pearson linear correlation coefficient (Cor) between the predictions from 1961 to 2021 and their corresponding observations. Furthermore, since more attention should be paid to the obvious lower or higher years (i.e., abnormal years), the number of the abnormal years, which are successfully predicted (Nyears), is also used to evaluate the model forecast skill. It is noteworthy that the Cor and Nyears actually represent the forecast skill of the model framework (i.e., the model's hyperparameters are fixed, but the model's parameters are undetermined) because the predictions from 1961 to 2021 are calculated by different NN models (i.e., the weights and biases of the model are different although the model framework is fixed).

The performance of a machine-learning model might be sensitive to the random seed for weight initialization [52]. Our preliminary analysis shows that some good random seeds can often lead to a relatively higher forecast skill, and the model training often fails to reach the stopping threshold when starting with some bad random seeds. Aside from random seed, there is another random seed, which is used for the order of training data. In contrast to the weight initialization random seed, the impact of the training order random seed is negligible. In other words, the forecast skill of a given model framework (i.e., all hyperparameters have been fixed) is mainly influenced by the weight initialization random seed. Therefore, selecting good seeds becomes an important step of the model training process. At this step, a large number (e.g., 200 in this study) of weight initialization random seeds were carried out, and each seed corresponds to a model training forecast skill estimated by the leave-one-out approach. The reason for the name of "training forecast skill" is because the forecast skill is estimated based on the training set (i.e., during

the training phase). In order to use the leave-one-out approach, one sample of the training set is chosen as the validation set and the other samples are used to train the model, and so on, until all samples from the training set have been acted as the validation set. The best 10 random seeds are selected out based on the training forecast skill (only the Cor). After that, the NN model would be trained again using these 10 good seeds (one seed corresponds to one NN model). At this step, the leave-one-out approach is not needed; thus, all samples from training set are used to train the model. Finally, there are 10 NN models (the model's parameters are fixed) in the forecast system, the actual prediction is the average of the 10 corresponding predictions (i.e., ensemble-mean prediction). The standard deviation calculated from the difference of each prediction for 10 predictions is used to indicate the uncertainty of the ensemble-mean prediction to some extent. The main goal of ensemble prediction is to enhance the robustness of forecast skills. In the rest of this article, the forecast skill usually denotes that which is calculated based on the ensemble-mean prediction. It is necessary to point out that more details about the forecast system can be found in the code, which has been archived in a public repository.

3. Application and Evaluation

3.1. Setting hyperparameters

All hyperparameters must be set at the start of model training. Most of the time, setting hyperparameters requires an expensive guess-and-check process [35]. Similar to selecting good seeds, setting hyperparameters is also based on the model training forecast skill estimated by the leave-one-out approach. At this step, all observed samples are reserved as a training set, and the training forecast skill used for setting hyperparameters is calculated based on the ensemble-mean prediction from the best 10 random seeds.

In this study, there is one hyperparameter for the model training (i.e., ST) and three hyperparameters for the model architecture (i.e., N, h1, and h2). The number of possible combinations is very huge. It is fortunate that the model performance is more sensitive to the hyperparameter of the input layer (i.e., N) as compared to other hyperparameters. Therefore, N should be set first. Table 1 lists the training forecast skills (Cor, Nyears) under different model architectures. It is obvious that the optimal N is 13 at ST= 0.36. We note that, 13 is also the optimal value of N at other STs (e.g., 0.33 and 0.40, not shown). It seems that small N (N< 13) cannot provide enough useful information, and a large N (N> 13) may cause a more serious overfitting problem.

Table 1. Training forecast skills (Cor, Nyears) under different model architectures (N, h1, and h2).

N\h1,h2	(8,5)	(6,5)	(6,4)	(4,4)	(6,3)	(4,3)	(3,3)
8	0.56,14	0.55,16	0.57,16	0.57,16	0.56,17	0.57,16	0.57,17
10	0.61,17	0.60,19	0.62,18	0.59,18	0.62,18	0.59,17	0.61,19
12	0.62,16	0.56,18	0.62,17	0.59,18	0.59,17	0.61,17	0.60,16
13	0.65,18	0.64,18	0.66,18	0.64,19	0.65,17	0.67,18	0.66,18
14	0.64,14	0.62,18	0.63,18	0.61,19	0.62,17	0.63,19	0.64,17
16	0.64,18	0.63,17	0.62,17	0.63,17	0.61,17	0.63,15	0.61,18
18	0.63,15	0.61,16	0.61,13	0.62,18	0.60,16	0.61,12	0.62,18

Here, the ST is set to 0.36, and the observed samples from 1961 to 2021 are used as training set. The number of abnormal years from observation is 30.

After setting N, it becomes relatively simple to set the other hyperparameters. Table 2 lists the forecast skills under different combinations of the other three hyperparameters (i.e., ST, h1, and h2). In contrast to the N, the impacts of the other hyperparameters on the

forecast skill are not remarkable. Generally speaking, the model training phase is insufficient if the ST is too large, and the overfitting problem becomes serious if the ST is too small. Taken overall, the forecast skills at ST= 0.36 are a little better than those at other STs. Considering the generalization problem, a simple model is better than a complex model under similar forecast skills [35]. Based on the above analysis, the Cor= 0.67 and Nyears= 18 at ST= 0.36, h1= 4 and h2= 3 was taken as the best forecast skill although the model framework (ST= 0.36, h1= 4 and h2= 4) produces the highest Nyears (Nyears= 19). In this study, N= 13, h1= 4, h2= 3 and ST= 0.36 were chosen as the optimal hyperparameters when the size of the training set is around 61. If the size of training set is far from 61, the optimal hyperparameters should be reset.

Table 2. Similar to Table 1, but for the training forecast skills (Cor, Nyears) under different stopping thresholds (ST) and hidden layer architectures (h1 and h2).

ST\ (h1,h2)	(8,5)	(6,5)	(6,4)	(4,4)	(6,3)	(4,3)	(3,3)
0.28	0.64,15	0.63,14	0.64,14	0.62,17	0.65,17	0.64,16	0.65,16
0.30	0.65,15	0.64,15	0.65,14	0.62,18	0.65,18	0.65,16	0.65,16
0.34	0.66,18	0.64,16	0.66,17	0.64,19	0.65,18	0.65,18	0.66,18
0.36	0.65,18	0.64,18	0.66,18	0.64,19	0.65,17	0.67,18	0.66,18
0.38	0.66,16	0.64,17	0.66,16	0.65,18	0.65,16	0.66,16	0.65,18
0.40	0.66,15	0.64,17	0.66,17	0.65,16	0.66,18	0.65,16	0.64,16
0.42	0.67,16	0.63,18	0.67,16	0.63,18	0.64,16	0.67,15	0.65,17
0.44	0.65,14	0.61,16	0.66,15	0.61,16	0.63,16	0.66,15	0.64,16

Here, the N is set to 13.

The training forecast skills analyzed above do not indicate the actual forecast skill of the forecast system. The training forecast skills are only used for setting hyperparameters and selecting random seeds. After that, the NN model is trained again using the whole training set. To some extent, the training forecast skill might be considered as the maximum potential forecast skill of the forecast system.

3.2. Forecast skill

Firstly, the forecast skill of the forecast system was analyzed based on all observed samples (i.e., 1961~2021). The leave-one-out approach was used twice: once to calculate the training forecast skills during the model training phase (the size of available samples is 60), and again to calculate the actual forecast skill used for estimating the forecast system (the size of available samples is 61). Three forecast experiments, named Exp1, Exp2, and Exp3, were carried out with different selection range of weight initialization random seeds (e.g., 0~199, 400~599, and 1000~1199). The three forecast experiments (Exp1, Exp2, and Exp3) were run twice using the same code, named Round1 and Round 2. The Experimental results show that the forecast system has acceptable predictability in most years (Figure 3). However, the predictability is always bad in a few years. We take the year of 2004 as an example; the observed WPSHRL index is very high, although the spatial pattern of the prophase SST anomalies (not shown) is more similar to the composites of SST anomalies in low index years (Figure 1). In other words, the year of 2004 is contrary to the background knowledge from composite analysis (i.e., Figure 1). All these years are called special years in this paper. It is very difficult to improve the predictability in these special years. Taken overall, the forecast skill of the forecast system is acceptable. The Cor is higher than 0.6 in all experiments. The number of abnormal years from observation is 30. The Nyears value is 17 or 18, which is around three-fifths of the observed abnormal years. This suggests that the forecast system is applicable to both normal years and abnormal

years. Furthermore, the robustness of the forecast system is also analyzed. The prediction uncertainty, which is the standard deviation from the 10 ensemble members, is usually less than 0.2. As compared to the ensemble-mean value (i.e., the final prediction, which is in the range of -1 to 1), the uncertainty range (< 0.2) is usually not significant. The difference in Cor and Nyears among the Exp1, Exp2, and Exp3 experiments is less than 0.02 and 2, respectively. The impact of the range of random seed selection is relatively small and negligible. The forecast skills from these three experiments are almost same between Round1 ((0.603,17), (0.619,17), and (0.604,18)) and Round2 ((0.606,17), (0.619,17), and (0.604,18)). In terms of forecast skill, the forecast system is relatively stable, although the NN model's parameters are different among these forecast experiments.

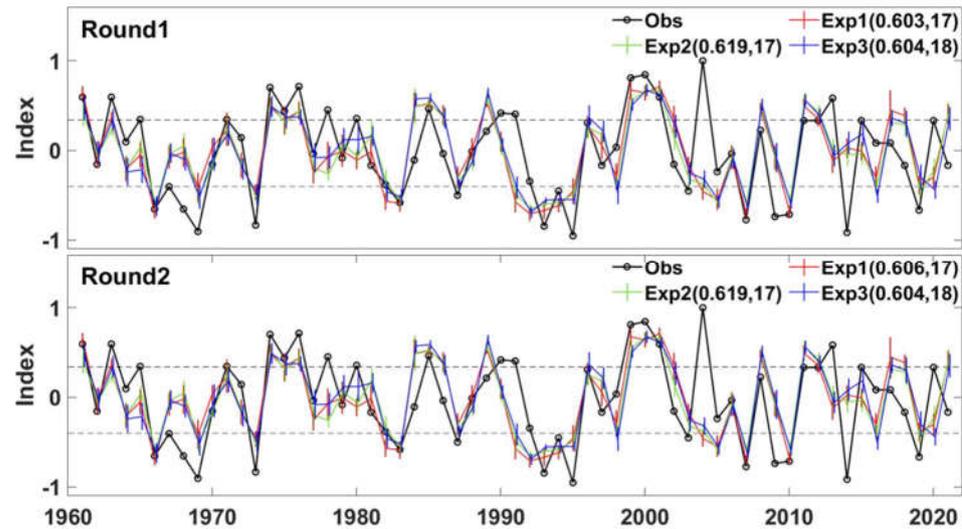


Figure 3. Time series of the WPSHRL index from the observation (black line) and model predictions (colored lines, for different experiments) for the period from 1961 to 2021. The vertical bars indicate the ranges of prediction uncertainty at different years. The dash lines indicate the index thresholds for abnormal years. The three forecast experiments (Exp1, Exp2, and Exp3) were run twice using the same code (upper panel from Round 1, and lower panel from Round 2). Forecast skills (Cor, Nyears) are shown after the corresponding experiment names.

The forecast skill was also analyzed by imitating the actual application of the forecast system (Figure 4). In the forward forecast experiments, all available samples were used as a training set. Assuming it was 1 March 1998, the June 1998 WPSHRL was forecasted using the NN models trained by the observed samples from 1961 to 1997. The backward forecast experiments were carried out under the assumption that the order of years is reversed. In these forecast experiments, the optimal hyperparameters of the NN model need to be reset based on the size of the corresponding training set, and the predicted WPSHRL indexes return back to the integer degree of latitude. The Cor is around 0.66 in the backward experiments from 1990 to 1961. In the forward experiments from 1991 to 2021, Cor is obviously decreased due to a few special years (e.g., 2004, 2014). The Cor from the forward experiments without the years of 2004 and 2014 is around 0.63 (not shown in Figure 4), which is close to that from the backward experiments. The reason that the forecast skill (evaluated by the Cor) for the June WPSHRL is decreased after 2000 might be that the special years occur more frequently. The numbers of observed abnormal years from the backward experiment period and forward experiment period are 12 and 15, respectively. Note that, the year with a WPSHRL higher than 22° N is taken as an abnormal year. In the backward experiment period, there are too many years (7 years) with the WPSHRL at the threshold value (22° N). This is the reason why the number of abnormal years from the backward experiment period is 3 (i.e., $15 - 12$) less than that from the forward experiment period. Nyears is 10 (two-thirds of the observed abnormal years) in the forward experiment period. Taken overall (i.e., 1961~2021), Nyear is 17 or 18, which is around three-fifths

of the observed abnormal years. The forecast skill (both Cor and Nyears) estimated by this imitating approach (Figure 4) is similar to that estimated by the leave-one-out approach (Figure 3).

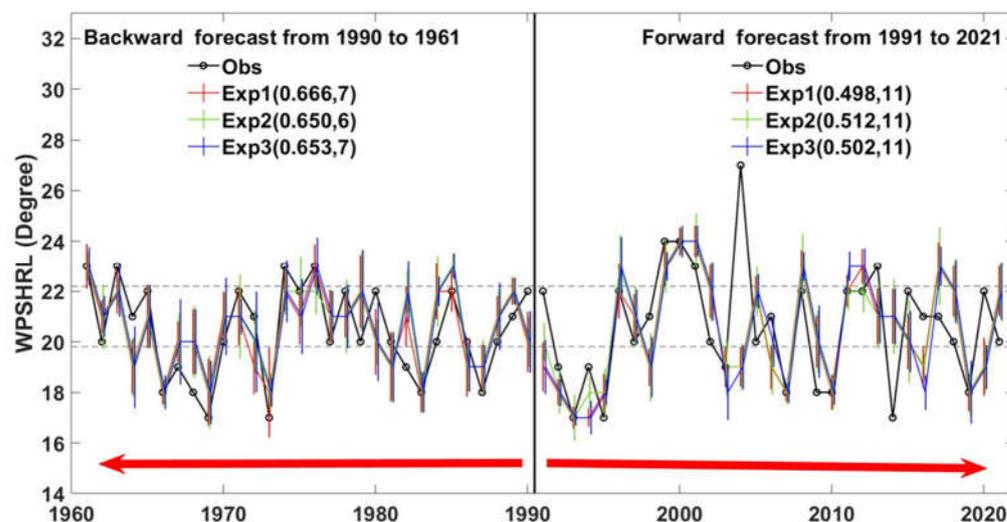


Figure 4. Similar to Figure 3, but for the forward and backward forecast experiments. The number of the observed abnormal years from the above two periods (1961~1990 and 1991~2021) are 12 and 15, respectively.

3.3. Interpretability of the forecast system

It is difficult to illustrate the physical interpretability of an NN model with several hidden layers. However, the Perturbation-Based Method can be used to check the relation between input predictors and target predictands from the NN model [53]. Here, the interpretability of the forecast system was analyzed by the Perturbation-Based Method. Figure 5 shows the sensitivity of the predicted WPSHRL to the prophase SST (i.e., $\partial\text{WPSHRL}/\partial\text{SST}$). Here, the sensitivities were calculated based on the NN models (with given weights and biases) used in Figure 3 (the Exp1 from Round1). The relationship between SST and WPSHRL from the observation can be shown by the linear correlation (Figure 5, upper panel). Unlike the linear correlation from observation, the sensitivity from the non-linear statistical model (i.e., the NN model) at one region (e.g., the North Pacific) is influenced by the SST at another region (e.g., the East Pacific). Therefore, the sensitivity from the NN model is dependent on the SST field of the year. The sensitivities of four years (1962, 1974, 1983 and 2004) are shown in Figure 5. The observed WPSHRL indexes from the years of 1962, 1974, 1983, and 2004 are -0.15 (normal), 0.70 (high), -0.58 (low), and 1.00 (high), respectively. Their corresponding predicted WPSHRL indexes are -0.04 (normal), 0.51 (high), -0.56 (low), and -0.38 (normal), respectively. The years 1962, 1974, and 1983 have good predictions, whereas the year 2004 has a bad prediction. The spatial pattern of these sensitivities from the NN model is generally in agreement with the linear correlation from observation, especially for the years with good predictions. The differences in sensitivity among these good years are not prominent although the input SST fields (not shown) and predicted WPSHRL indexes (normal, high, and low) are obviously different. As compared to the sensitivities from these good years, the sensitivities from the bad year (i.e., 2004, the year with bad prediction) are relatively weak in most regions and the spatial pattern is much more different from the linear correlation. These characteristics can also be found in most bad years (not shown). It seems necessary to check the sensitivity from the NN model, which is calculated based on the input SST data. The weak sensitivity and bad spatial pattern usually indicate that the corresponding prediction is of low reliability.

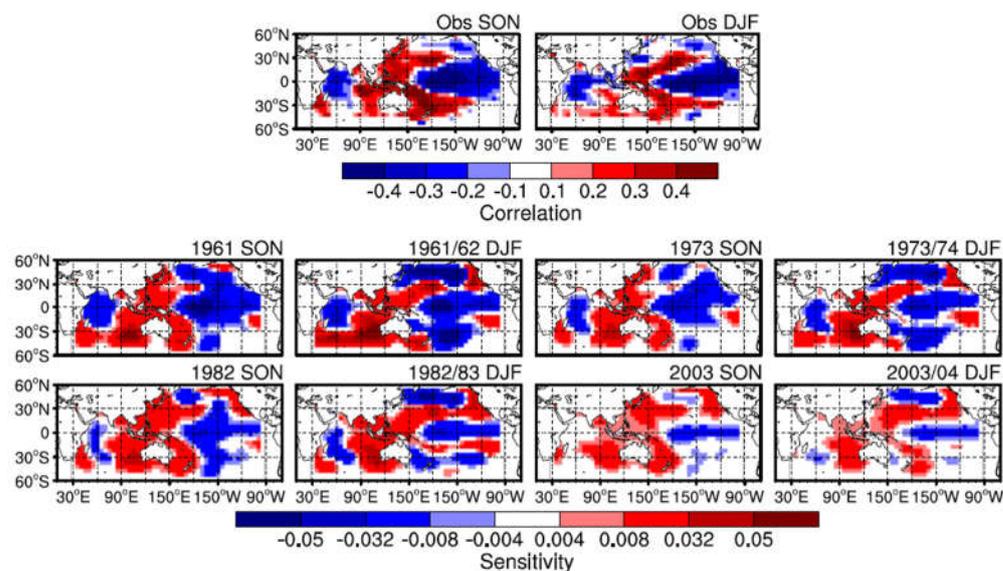


Figure 5. The impact of the prophase SST on the June WPSHRL index from the observation (linear correlation, upper panel) and forecast system (sensitivities, lower panel).

4. Conclusions and Discussion

In this study, we developed a forecast system for the June WPSHRL index based on the latest autumn and winter SST. Considering the limited number of observed samples, a simple NN model was used to extract the non-linear relationship between input predictors and target predictands in the forecast system. The linear correlation coefficient between the predictions from the forecast system and their corresponding observations is around 0.6, and about three-fifths of observed abnormal years are successfully predicted. Furthermore, sensitivity experiments (i.e., Exp1~3, Round1~2) show that the forecast system is relatively stable in terms of forecast skill. Therefore, we think that the forecast system is valuable in a real application sense.

The successful forecast system introduced in this study suggests that the prediction of the general atmospheric circulation beyond one season of lead time might be improved using non-linear statistical methods (e.g., the NN model). However, making full use of the non-linear statistical methods is much more difficult than using the traditional linear statistical methods, especially for climate forecast. Unlike other application areas with huge samples, there are two unique aspects in the area of climate forecast. One is that the size of the observed samples is very small (adverse condition), the other is that previous studies have provided a great deal of useful background knowledge and experience (favorable condition). In this study, to deal with the adverse condition, a very simple NN model is selected, 40 composite samples are produced artificially, the model training is stopped based on a threshold, and the leave-one-out approach is used for calculating the forecast skill. This study also takes advantage of the favorable condition to some extent, such as in the composite analysis in Figure 1, the principal component extraction in Figure 2, and the linear relationship from the observation in Figure 5.

The practicability of the July and August WPSHRL index was also tested using this forecast system. The input predictors are still the latest autumn and winter SST. Although the forecast skill for the July WPSHRL is significantly reduced as compared to that for the June WPSHRL, the forecast skill is still acceptable (Cor is around 0.40). However, the forecast skill for the August WPSHRL is very low (Cor could not reach 0.3). It is well known that the formation mechanism for the WPSHRL index is different in different months. The forecast system introduced in this study is designed only for the prediction of the June WPSHRL at a lead time of three months. For instance, the predictors (i.e., the regions and seasons of SST data) are first selected according to background knowledge (i.e., the strong linear correlation between the prophase SST and the June WPSHRL index). Generally

speaking, the background knowledge and experience from previous studies are useful to improving the performance of the forecast system. In order to make rational use of the background knowledge and experience, the forecast system should be designed according to the month and lead time.

Author Contributions: X.S. designed the study. H.Y. supported the study. C.S developed the forecast system and corresponding code. C.S and X.S. wrote the original draft. All authors contributed to reviewing and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (grant nos. 2020YFA0608000) and the National Natural Science Foundation of China (grant nos. 41775095 and 42075145). The APC was funded by the same funding.

Data Availability Statement: The WPSHRL data from 1961 to 2016 are downloadable from <http://cmdp.ncc-cma.net/160/74.php>. The original SST data are downloadable from https://psl.noaa.gov/data/gridded/data.kaplan_sst.html. The code of the forecast system and related results used in this study have been archived in a public repository <https://doi.org/10.5281/zenodo.6294926>.

Acknowledgments: The authors would like to thank Dong Si for guidance in the calculating method of the WPSHRL index. This study was conducted in the High-Performance Computing Center of Nanjing University of Information Science & Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tao, S.; Wei, J. The Westward, Northward advance of the Subtropical High over the West Pacific in Summer. *J Appl. Meteor. Sci. (in Chinese)* 2006, 17, 513-525. [<https://doi.org/10.3969/j.issn.1001-7313.2006.05.001> (in Chinese)]
2. Liu, Y.; Hong, J.; Liu, C.; Zhang, P. Meiyu flooding of Huaihe River valley and anomaly of seasonal variation of subtropical anticyclone over the Western Pacific. *Chin. J Atmos. Sci. (in Chinese)* 2013, 37, 439-450. [<http://dx.doi.org/10.3878/j.issn.1006-9895.2012.12313> (in Chinese)]
3. Jiang, W.; Zhang, Z.; Liu, Y. Relationship between the dry-season precipitation in Southwest China and decadal changes of the Western Pacific Subtropical High since the 21st century. *Meteor. Mon.* 2016, 42, 1335-1341. [<https://d.wanfangdata.com.cn/periodical/qx201611005> (in Chinese)]
4. Gao, H.; Yuan, Y.; Hong, J.; Wang, D. Overview of climate prediction of the Summer 2016 and the precursory signals. *Meteor. Mon.* 2017, 43, 486-494. [<https://d.wanfangdata.com.cn/periodical/qx201704011> (in Chinese)]
5. Yuan, Y.; Gao, H.; Liu, Y. Analysis of the characteristics and causes of precipitation anomalies over Eastern China in the Summer of 2016. *Meteor. Mon.* 2017, 43, 115-121. [<https://d.wanfangdata.com.cn/periodical/qx201704011> (in Chinese)]
6. Wu, S.; Guo, D. Effects of East Asian summer monsoon and Western Pacific Subtropical High on summer precipitation in China. *Science and Technology Innovation Herald. (in Chinese)* 2019, 16, 112-119. [<https://doi.org/10.16660/j.cnki.1674-098X.2019.16.112> (in Chinese)]
7. Shi, W.; Li, X.; Zeng, M.; Zhang, B.; Wang, H.; Zhu, K.; Zhu Ge X. Multi-model comparison and high-resolution regional model forecast analysis for the "7-20" Zhengzhou severe heavy rain. *Trans. Atmos. Sci. (in Chinese)* 2021, 44, 688-702. [<https://doi.org/10.13878/j.cnki.dqkxxb.20210823001> (in Chinese)]
8. Zhang, E.; Mao, Y.; Han, Z.; Tang, B. The multi-level recursion long-range forecasting models of the Summer characteristic quantities for the Western Pacific Subtropical High. *Plateau. Meteor. (in Chinese)* 1984, 3, 51-56. [<http://www.gyqx.ac.cn/article/id/5882> (in Chinese)]
9. Zhang, E.; Tang, B.; Han, Z. A long-range forecasting model for the Subtropical High using the integral multi-level recursion. *J Appl. Meteor. Sci. (in Chinese)* 1989, 1, 69-74. [<http://qikan.camsma.cn/article/id/19890110> (in Chinese)]
10. Ai, Y.; Chen, X. Analysis of the correlation between the Subtropical High over Western Pacific in Summer and SST. *J Trop. Meteor. (in Chinese)* 2000, 16, 1-8. [<https://doi.org/10.3969/j.issn.1004-4965.2000.01.001> (in Chinese)]
11. Dong, Z.; Zhang R. A prediction of the Western Pacific Subtropical High based on wavelet decomposition and ANFIS model. *J Trop. Meteor. (in Chinese)* 2004, 20, 419-425. [<https://doi.org/10.3969/j.issn.1004-4965.2004.04.010> (in Chinese)]
12. Zhang, R.; Wang, H.; Liu, K.; Hong, M.; Yu, D. Dynamic randomness and complexity of Subtropical High index based on phase space reconstruction. *J Nanjing Inst. Meteor. (in Chinese)* 2007, 30, 723-729. [<https://doi.org/10.3969/j.issn.1674-7097.2007.06.001> (in Chinese)]
13. Wang, Y.; Teng, J.; Zhang, R.; Wan, Q.; Dong, Z. Predicting the Subtropical High index by coupling self-organizing feature map And generalized regression neural network. *J Trop. Meteor. (in Chinese)* 2008, 24, 475-482. [<https://doi.org/10.3969/j.issn.1004-4965.2008.05.005> (in Chinese)]

14. Jia, Y.; Hu, Y.; Zhong, Z.; Zhu, Y. Statistical forecast model of Western Pacific Subtropical High indices in Summer. *Plateau Meteor.* (in Chinese) 2015, 34, 1369-1378. [<http://www.gyqx.ac.cn/article/doi/10.7522/j.issn.1000-0534.2014.00079> (in Chinese)]
15. Qian, D.; Guan, Z.; Xu, J. Prediction models for summertime Western Pacific Subtropical High based on the leading SSTA modes in the tropical Indo-Pacific sector. *Trans. Atmos. Sci.* (in Chinese) 2021, 44, 405-417. [<https://doi.org/10.13878/j.cnki.dqkxxb.20200331010> (in Chinese)]
16. Ma, D.; Mao, H.; Bao, Y.; Lin, Y. The Subtropical High forecast verification and interpretation of the NWP in Summer. *Meteor. Mon.* 1999, (3), 49-53. [<https://doi.org/10.3969/j.issn.1000-0526.1999.03.011> (in Chinese)]
17. Ren, H.; Zhang, P.; Guo, B.; Chou, J. Dynamical model of Subtropical High ridge-line section and numerical simulations with its simplified scheme. *Chin. J Atmos. Sci.* (in Chinese) 2005, 29, 71-78. [<http://dx.doi.org/10.3878/j.issn.1006-9895.2005.01.09> (in Chinese)]
18. Liu, Y.; Li, W.; Zuo, J.; Hu, Z. Simulations and projections of the Western Pacific Subtropical High in CMIP5 models. *Acta Meteor. Sinica.* (in Chinese) 2014, (2), 277-290. [<http://dx.doi.org/10.11676/qxxb2014.025> (in Chinese)]
19. Fu, B.; Wang, T.; Wei, C.; Deng, Z. Testing and assessment of capabilities of day-to-day predicting of summertime West Pacific Subtropical High based on CFSv2. *Guangdong Meteor.* (in Chinese) 2016, 38, 15-19. [<https://doi.org/10.3969/j.issn.1007-6190.2016.02.004> (in Chinese)]
20. Duan, C.; Xu, M.; Cheng, Z.; Luo, L. Evaluation on monthly prediction of Western Pacific Subtropical High by DERF2.0 model. *Meteor. Mon.* 2017, 43, 1267-1277. [<https://d.wanfangdata.com.cn/periodical/qx201710011> (in Chinese)]
21. Zhou, F.; Ren, H.; Hu, Z.; Liu, M.; Wu, J.; Liu, C. Seasonal predictability of primary East Asian Summer circulation patterns by three operational climate prediction models. *Quart. J Roy. Meteor. Soc.* 2020, 146, 629-646. [<https://doi.org/10.1002/qj.3697>]
22. Huntingford, C.; Jeffers, E.S.; Bonsall, M.B.; Christensen, H.M.; Lees, T.; Yang, H. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environ. Res. Lett.* 2019, 14, 124007. [<https://doi.org/10.1088/1748-9326/ab4e55>]
23. Ham, Y.G.; Kim, J.H.; Luo, J.J.; Deep learning for multi-year ENSO forecasts. *Nature* 2019, 573, 568-572. [<https://doi.org/10.1038/s41586-019-1559-7>]
24. Hartigan, J.; Macnamara, S.; Leslie, L.M. Application of machine learning to attribution and prediction of seasonal precipitation and temperature trends in Canberra, Australia. *Climate* 2020, 8, 76. [<http://dx.doi.org/10.3390/cli8060076>]
25. Kuang, Z.; Song, Z.; Dong, C. Study on the future projection of global sea surface temperature over 21st century using a biases correction model based on machine learning. *Climate Change Res. Lett.* (in Chinese) 2020, 9, 270-284. [<https://doi.org/10.12677/ccrl.2020.94031> (in Chinese)]
26. Toms, B.A.; Barnes, E.A.; Ebert-Uphoff, I. Physically interpretable neural networks for the geosciences: applications to earth system variability. *arXiv* 2020, arXiv:1912.01752.
27. Geng, H.; Wang, T. Spatiotemporal model based on deep learning for ENSO forecasts. *Atmosphere* 2021, 12, 810. [<https://doi.org/10.3390/atmos12070810>]
28. Hussein, E.A.; Ghaziasgar, M.; Thron, C.; Vaccari, M.; Bagula, A. Basic statistical estimation outperforms machine learning in monthly prediction of seasonal climatic parameters. *Atmosphere* 2021, 12, 539. [<https://doi.org/10.3390/atmos12050539>]
29. He, S.; Wang, H.; Li, H.; Zhao, J. Machine learning and its potential application to climate prediction. *Trans. Atmos. Sci.* (in Chinese) 2021, 11, 2680-2693. [<https://doi.org/10.13878/j.cnki.dqkxxb.20201125001> (in Chinese)]
30. Liu, J.; Tang, Y.; Wu, Y.; Li, T.; Wang, Q.; Chen, D. Forecasting the Indian Ocean Dipole with deep learning techniques. *Geophys. Res. Lett.* 2021, 48, e2021GL094407. [<https://doi.org/10.1029/2021GL094407>]
31. Tangang, F.T.; Hsieh W.W.; Tang, B. Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Climate Dynamics* 1997, 13, 135-147. [<https://doi.org/10.1007/s003820050156>]
32. Tangang, F.T.; Hsieh W.W.; Tang, B. Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *J Geophys. Res.* 1998, 103, 7511-7522. [<https://doi.org/10.1029/97JC03414>]
33. Wu, A.; Hsieh W.W.; Tang, B. Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Networks* 2006, 19, 145-154. [<https://doi.org/10.1016/j.neunet.2006.01.004>]
34. Abdullah, S.; Ismail, M.; Ahmed, A.N.; Abdullah, A.M. Forecasting particulate matter concentration using linear and non-Linear approaches for air quality decision support. *Atmosphere* 2019, 10, 667. [<http://dx.doi.org/10.3390/atmos10110667>]
35. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, USA, 2016; pp. 99-208. [<http://www.deeplearningbook.org>]
36. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*; O'Reilly Media Inc: USA, 2017; pp. 29-35.
37. Tran, T.; Lee, T.; Kim, J.S. Increasing Neurons or Deepening Layers in Forecasting Maximum Temperature Time Series? *Atmosphere* 2020, 11, 1072. [<http://dx.doi.org/10.3390/atmos11101072>]
38. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* 2019, 566, 195-204. [<http://doi.org/10.1038/s41586-019-0912-1>]
39. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv* 2016, arXiv:1602.04938.
40. Fong, R.C.; Vedaldi, A. Interpretable explanations of black boxes by meaningful perturbation. *arXiv* 2017, arXiv:1704.03296.
41. Kindermans, P.J.; Schütt, K.T.; Alber, M.; Müller, K.R.; Erhan, D.; Kim, B.; Dähne, S. Learning how to explain neural networks: PatternNet and PatternAttribution. *arXiv* 2017, arXiv:1705.05598.

42. Riccardo, G.; Anna, M.; Salvatore, R.; Franco, T.; Fosca, G.; Dino, P. A survey of methods for explaining black box models. *ACM Comput. Surv.* 2018, 51, 1-42. [<https://doi.org/10.1145/3236009>]
43. Fan, F.; Xiong, J.; Wang, G. On Interpretability of Artificial Neural Networks. *arXiv* 2020, arXiv:2001.02522.
44. Belochitski, A.; Krasnopolsky, V. Robustness of neural network emulations of radiative transfer parameterizations in a state-of-the-art general circulation model. *Geosci. Model Dev.* 2021, 14, 7425–7437. [<http://doi.org/10.5194/gmd-14-7425-2021>]
45. Yuan, H.; Yu, H.; Gui, S.; Ji, S. Explainability in Graph Neural Networks: A Taxonomic Survey. *arXiv* 2021, arXiv:2012.15445.
46. Kaplan, A.; Cane, M.A.; Kushnir, Y.; Clement, A.C. Analyses of global sea surface temperature 1856-1991. *J Geophys. Res.* 1998, 103, 18567-18590. [<http://dx.doi.org/10.1029/98JC01736>]
47. Yao, Y.; Yan, H. Relationship between preceding pacific sea surface temperature and Subtropical High indexes of main raining seasons. *J Trop. Meteor.* (in Chinese) 2008, 24, 483-489. [<https://doi.org/10.3969/j.issn.1004-4965.2008.05.006> (in Chinese)]
48. Liu, Y.; Li, W.; Ai, W.; Li, Q. Reconstruction and application of the monthly Western Pacific Subtropical High indices. *J Appl. Meteor. Sci.* (in Chinese) 2012, 23, 414-423. [<http://qikan.camsma.cn/article/id/20120404> (in Chinese)]
49. Hornik, K.; Stinchcombe, M.; White, H. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks.* 1989, 2, 359-366. [[https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)]
50. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research.* 2010, 9, 249-256. [<http://proceedings.mlr.press/v9/glorot10a.html>]
51. Opper, M.; Winther, O. Gaussian processes and SVM: Mean field results and leave-one-out. *Advances in Large Margin Classifiers*, 8rd ed.; Smola, A.J., Bartlett, P.L., Schölkopf, B., Schuurmans, D., Eds.; MIT Press: Cambridge, USA, 2000; pp. 311-326. [<https://www.researchgate.net/publication/40498234>]
52. Dodge, J.; Ilharco, G.; Schwartz, R.; Farhadi, A.; Hajishirzi, H.; Smith, N. Fine-tuning pretrained language models: weight initializations, data orders, and early stopping. *arXiv* 2020, arXiv:2002.06305.
53. Yuan, H.; Yu, H.; Gui, S.; Ji, S. Explainability in Graph Neural Networks: A Taxonomic Survey. *arXiv* 2020, arXiv:2012.15445.