Artificial neural network-based storm surge forecast model: practical application to Sakai Minato, Japan

Sooyoul Kim[a*], Shunqi Pan[b] and Hajime Mase[c]

a: Graduate School of Engineering, Tottori University, Koyama-cho Minami, Tottori, 680-850, Japan

b: School of Engineering, Cardiff University, The Parade, Cardiff, CF24 3AA, UK

c: Disaster Prevention Research Institute, Kyoto University, Gokasho, Kyoto, 611-0011, Japan

* Corresponding author: sooyoul.kim@tottori-u.ac.jp

Abstract

The present study describes a novel way of a systematic and objective selection procedure for the development of an Artificial Neural Network-based storm Surge Forecast Model (ANN-SFM) with 5, 12 and 24 h lead times and its application to Sakai Minato area on the Tottori coast, Japan. The selection procedure can guide the determination of the superiority of the best performing model in terms of the appropriate combination of unit number in the hidden layer and parameter in the input layer. In the application of ANN-SFM to Sakai Minato, it is found that the best ANN-SFMs for 5 and 12 h-forecasting are established with the most suitable set of 70 units (the number of hidden neurons) and the input components of surge level, sea level pressure, the depression rate of sea level pressure, longitude, latitude, central atmospheric pressure and highest wind speed. The best ANN-SFM for 24 h-forecasting is determined with 160 units and the input parameters of surge level, sea level pressure, the depression rate of sea level pressure, longitude and latitude. The proposed method of the selection procedure is able to be adaptable to other coastal locations for the development of the artificial neural network-based storm surge forecast model as establishing the superiority of the most relevant set combining unit numbers and input parameters.

Keywords: storm surge forecast; artificial neural network; parameter selection procedure

1. Introduction

Storm surges can be a catastrophic natural hazard arising from typhoons/hurricanes that landfall on coasts or pass through near coasts. They induce flooding in low-lying areas, causing casualty on coastal communities and economic loss of coastal facilities, as well as interruptions to economic activities and transports. The severity of the damages will strongly depend on the timing of peak surge levels being generated. If the time-dependent surge level, especially the peak surge level, is predictable with a specific and sufficient lead time, a relevant government service will be able to issue a suitable warning for the evacuation of the local community and planning for the protection of the coastal facility.

Previously, much attention has been paid to the development of storm surge forecast models to provide reliable surge levels to the coastal community prior to typhoons' landfall. Approaches for the development of storm surge forecast models are mainly to use either physical process-based numerical models (e.g., Luettich and Westerink [19]; Flather [5]; Jelesnianski et al. [8], Kim et al. [10]) or machine learning or data-driven methods (e.g., Lee [16]; Tseng et al. [31]; Kim et al. [13]; Kim et al. [12], Jia et al. [9]). The physical process-based models have the advantage of including physics of storm surge and taking distinct external forces to drive themselves over the machine learning method. However, they are generally time-consuming and cumbersome to be operated. On the other hand, the machine learning method-based forecast models such as an artificial neural network (ANN) are lighter and faster to be operated in comparison to the process-based model (Lee [16], [17], [18]; Kim et al. [13]; Kim et al. [12]; Jia et al. [9]). But, when using these models to study the storm surge phenomena, it is inconvenient and also sometimes laborious to find the appropriate input set between meteorological, hydrodynamic and typhoon-characteristic parameters to efficiently train the ANN-based surge forecast model (ANN-SFM) and accurately get the forecasted surge level (in this way, a surge level is an output parameter). Because of these reasons, efforts have been made to clarify relations between input and output parameters when developing ANN-SFM at specific locations (e.g., Marzenna [22]; Lee [17], [18]; Tseng et al. [31]; Kim et al. [12]). According to the recent studies for time-dependent after-runner storm surge forecasting (e.g., Kim et al. [12]), it was shown that the proper input parameters vary with the area at which the time-dependent surge forecasting is developed through a series of experiments, and potential combination sets of the input parameters are found. The input parameters also differ depending on the given lead time for forecasting. In addition, it was found that an ANN-SFM for a relatively long lead time (for example, 24 hours) is apparently inaccurate in comparison with that for a short lead time (for instance, 5 hours). In other words, these methods were developed with some restrictions indirectly applying to other places as a universal ANN-SFM.

According to Dreyfus [4], the following steps are generally necessary to develop an ANN-based model:

- finding the relevant input parameters;
- determining the appropriate number of hidden neurons (or units);
- selecting the best performing algorithm and its corresponding coefficients;
- determining the proper functions in the hidden and output layers.

A novel classification system has been suggested to improve the accuracy of a neural network. For instance, Zhang et al. [33] suggested fruit-classification system using the principal component analysis and biogeography-based optimization with a feedforward neural network. Wang et al. [32] also proposed a novel computer-vision-based method for automatic detection of the alcohol use disorder based on wavelet Renyi entropy and three-segment encoded Jaya algorithm.

Through the entire procedure mentioned above, the development of ANN-SFM based models is expected to improve the forecast accuracy. However, it will still require a systematic and objective procedure to determine which factor should be first examined or has primary impacts on the accuracy of the model. Therefore, the present study introduces the systematic approach of a selection procedure to seek the best performance combination of the input parameters in the input layer and the corresponding unit number in the hidden layer for the lead times of 5, 12 and 24 hours, and applies the selection procedure of ANN-SFM to Sakai Minato port on the Tottori coast, Japan. The selection procedure suggested in the study is first to explore potential unit numbers of the hidden layer for each possible combination set of input parameters, and then to determine the unit number relevant to each set. Finally, the best set of input parameters can be selected among candidate sets of input parameters. As a result, the best performance model can be selected with the optimized unit number relevant to the best combination set of input parameters. The present study is to demonstrate the systematic selection procedure which can be applied to rather the development of a specified ANN-SFM at a location in any coastal regions than the development of a universal ANN-SFM. Also, in view of the selection procedure, it will show the accuracy of the ANN-SFM being improved, especially, forecasting for a lead time of 24 h, for instance.

This paper is organized as follows: an overview of the artificial neural network is given in Section 2, and a series of numerical experiments is described in Section 3. In Section 4, results and discussion are presented. Finally, conclusions are provided in Section 5.

2. Overview of artificial neural network (ANN)

ANN has been widely applied in many prediction and forecast studies in terms of air quality (e.g., Hanna and Heinold [7]), tides (e.g., Deo and Chaudhari [1]; Lee et al. [14], [15]), sea level rise (e.g., Makarynskyy et al. [20], [21]), maritime structures (e.g., Mase et al.

[23]), waves (e.g., Mase and Kitano [24]; Deo and Naidu [2]; Mase et al. [25]; Peres et al. [29]), wind (e.g., Tagliaferri et al. [30]) and tsunami (Mase et al. [26]).

2.1 Artificial neural network (ANN)

The present study employs a feedforward network that information flows in the forward direction from the input to the output (e.g., Dreyfus [4]). A structure of ANN used in the present study is commensurate with the previous study (Kim et al. [12]) that consists of three single layers, namely the input, hidden and output layers, with the use of the back-propagation optimization technique for training. The Levenberg–Marquardt algorithm is used that has an advantage in terms of the reduction of computational time among several back-propagation algorithms (the conjugate gradient method, the scaled conjugate gradient method, the Broyden–Fletcher–Goldfarb–Shanno method and the Levenberg–Marquardt method). Table 1 summarizes the parameters of the Levenberg–Marquardt method. Here, the regularization method of early stopping is used to avoid either over- or under-learning of ANNs as described by Mase et al. [26] and Dreyfus [4]: for instance, the training stops when the performance function reaches a pre-defined threshold (for example 10^{-6} m) of the mean squared error between the observed and predicted surge levels, before it completes the pre-set maximum 10,000 iterations. The functions of hyperbolic tangent sigmoid transfer and linear transfer are implemented in the hidden and output layers, respectively. The phases of training and validating are taken for the surge level forecast with the lead times of 5, 12, and 24 hours. The test phase is then carried out to evaluate the ANN-SFM with the set of input parameters and unit number that is determined in the phases of the test and validating. Section 3 will give a detailed description of the phases for training, validation, and test.

2.2 Input and output parameters

This study focuses on hourly measured meteorological and hydrodynamic parameters, as well as three hourly observed typhoon characteristic parameters. The meteorological parameters consist of the wind speed (m/s), wind direction (m/s), sea-level pressure (hPa), and drop of sea-level pressure (hPa) from undisturbed sea-level pressure (1013 hPa) at five hydrographic stations (Hamada, Matsue, Yonago, Ama, and Saigo in Fig. 1). The hydrodynamic parameter is the surge level (m) at the Sakai Minato station. The characteristic parameters of the typhoon are the longitude (degree) and latitude (degree) of the typhoon, central atmospheric pressure (hPa), and highest wind speed (m/s) near typhoon center. These parameters were collected during three typhoon events of Maemi (2003), Songda (2004) and Megi (2004), whose tracks are shown in Fig. 1, which are regarded as the representative typhoon events on the Tottori coast as highlighted in Kim et al. [11] and [12]. Hiyajo et al. [6] indicated that the surge level with the 100-year return period is approximately 0.63 m in this coastal area. This surge level almost corresponds to the maximum surge levels during the Maemi and Megi typhoon events.

As described in Kim et al. [12], the length of each parameter is 168 hours at hourly intervals. In addition, prior to training, all the parameters were normalized to make them dimensionless in the range between -1 and 1 as listed in Table X

Table X ….

Input	**Initial Value**	**Parameter Description**	**Eq.**
$\eta^t = \tilde{\eta}^t$		Water level	(1)
$SLP_p^t = \widetilde{SLP}_p^t$	1013 hPa	sea-level pressure	(2)
$DSLP_p^t = \widetilde{DSLP}_p^t$	100 hPa	drop of sea-level pressure from average sea-level pressure (= 1013 hPa),	(3)

$WS_p^t = \widetilde{WS}_p^t$	100 m/s	wind speed	(4)
$WD_p^t = \widetilde{WD}_p^t$	360 deg	wind direction	(5)
$LG^t = \widetilde{LG}^t$	150°E	longitude of typhoon	(6)
$LT^t = \widetilde{LT}^t$	50°N	latitude of typhoon	(7)
$CAP^t = \widetilde{CAP}^t$	1013 hPa	central atmospheric pressure of typhoon	(8)
$HWS^t = \widetilde{HWS}^t$	100 m/s	highest wind speed near typhoon centre	(9)

Eqs. (1) to (9) represents the raw value of the parameter, *t* the time, and *p* the location, with the symbol tilde (~). All the parameters including the surge level are taken into account as the input data in the input layer, while the surge level with the given lead time is considered as the output data in the output layer.

For each case, three tasks, namely: training, validating, and testing, are carried out to develop the ANN-SFM model with the lead times of 5, 12, and 24 hours. The data length is 460 hourly data. In the process, 75 % of all the input and output parameters is used for the training phase, and the remaining 25 % is used for the validating phase. The testing phase is conducted to evaluate the ANN-SFM with 25 % of all the parameters.

3. A series of numerical experiments

3.1 Data sets for a series of experiments

A series of experiments was carried out with the 12 data sets in order to determine which data set performs best for predicting the surge level with the specific lead time. Table 2 lists those data sets. The data set, $D_{i=1}$, consists of the surge level (SS), the sea level pressure (SLP), and the drop of sea level pressure (DSLP). In the data set, $D_{i=2}$, it is the same, but the wind speed (WS) is added to $D_{i=1}$. In the data set, $D_{i=3}$, the wind direction (WD) is added to $D_{i=2}$. For the data set of $D_{i=4}$, the components of SS, SLP, DSLP, and the longitude (LG) and the latitude (LT) of the typhoon are comprised. In the data set, $D_{i=5}$, the component of WS is added to that of $D_{i=4}$. The data set, $D_{i=6}$, contains all the components in the data set of $D_{i=5}$ as well as WD. The data set, $D_{i=7}$, is designed to include SS, SLP, DSLP, LG, and LT as well as the central atmospheric pressure of the typhoon (CAP) in the data set. For the data set of $D_{i=8}$, the component of WS is added to the data set, $D_{i=7}$. The component of WD is added to the data set, $D_{i=8}$ to construct the data set, $D_{i=9}$. The data set, $D_{i=10}$, consists of SS, SLP, DSLP, LG, LT, CAP, and the highest wind speed of typhoon (HWS). In the data set of $D_{i=11}$, WS is added to the data set, $D_{i=10}$. Finally, WD is added to the data set, $D_{i=11}$, to make the data set, $D_{i=12}$.

For instance, the ANN-SFM that is firstly (k = 1) trained by the data set, $D_{i=1}$, with the unit number, j = 10, can be mathematically expressed by:

$$N_{k=1}^{j=10} D_{i=1}\left(\eta^{t+\Delta t}\right) = (\eta^{t}\ , SLP_p^t, DSLP_p^t) \tag{10}$$

where Δt is the given lead time. For simplification, a symbol in parentheses is ignored.

3.2 Selection procedure for the best performance forecast model

The present study aims to investigate the effect of the number of unit (or neuron, u^j) in the hidden layer on the accuracy of the ANN-SFM that is relevant to the improvement of the accuracy for the forecasted surge level. In order to determine the best performance ANN-SFM with the desired lead time through the data sets, $D_{i=1}$ to $D_{i=12}$ listed in Table 2, the following selection procedure is introduced, as illustrated in Fig. 2:

1. Set the data set, $D_{i=1}$,

2. Set a unit number = 10, (= $u^{j=10}$) initially where *j* indicates the unit number,
3. Train, validate and test one ANN-SFM ($N_{k=1}^{j=10} D_{i=1}$), for the data set, $D_{i=1}$ with randomly given values of weight and bias in the hidden layer,
4. Repeat above the Step 3 twenty times to make the 20 independent ANN-SFMs (=$N_{k=1,\ldots,20}^{j=10} D_{i=1}$) by initializing the weight and bias every time,
5. Select the one good performance model among the 20 ANN-SFMs (=$N_{k=1,\ldots,20}^{j=10} D_{i=1}$) for the unit number = 10 ($u^{j=10}$) in the data set, $D_{i=1}$, by the indices of Eqs. (11), (12) and (13), which will be shown later,
6. Repeat the Steps 3 to 5 up to the unit number = 200, ($u^{j=200}$), with the increment of the unit number, $u^{j=10}$, in the data set, $D_{i=1}$,
7. Complete the selection of the 20 good performance models for each unit number,
8. Choose the one better performance model with its better acceptable unit number among the 20 good performance models for the data set, $D_{i=1}$,
9. Repeat the Steps of 1 to 8 for the data sets, $D_{i=1}$ to $D_{i=12}$,
10. Select the 12 better performance models and their dependent unit numbers for each data set and,
11. Decide the best performance model with a combination of the relevant unit number and data set among the twelve better performance models chosen in Step 10.

It is found that the ANN-SFMs that are trained and verified by the same unit number and data set predict slightly different surge levels because the initial weight and bias are randomly given and then adjusted through the training phase, which will be shown later. Thus, the present study proposes the selection procedure to judge the best performance ANN-SFM.

The selection procedure consists of two parts to decide the best acceptable unit number and data set. First, the data set, $D_{i=1}$, is prepared in Step 1. The initial unit number = 10, $u^{j=10}$, is set in Step 2. An ANN-SFM, $N_{k=1}^{j=10}D_{i=1}$, with the unit number, $u^{j=10}$, is trained, validated, and tested with a randomly given weight and bias in the hidden layer in Step 3. In Step 4, the 20 individual ANN- SFM, $N_{k=1,\ldots,20}^{j=10}D_{i=1}$, with the constant unit number, $u^{j=10}$, are made by initializing the weight and bias every time as repeating Step 3 twenty times. In Step 5, the one good performance ANN-SFM can be then selected among the 20 ones, $N_{k=1,\ldots,20}^{j=10}D_{i=1}$, by the indices of Eqs. (11), (12) and (13). Once the one good performance model has been selected, repeat Steps 3 to 5 as increasing the unit number with the interval of 10 units. Steps 3, 4, and 5 are iteratively conducted up to the unit number = 200, $u^{j=200}$. Consequently, the 20 good performance models can be chosen for each unit number, as illustrated in Step 7. Among the 20 good performance models, the one better performance model with its accompanying unit number can be selected in Step 8. Steps 2 to 8 are, so to speak, the selection procedure for the better model and its accompanying unit number.

Steps 1 to 9 are repeated for the data sets of $D_{i=1}$ to $D_{i=12}$ since changing the data set as demonstrated in Step 9, may result in obtaining 12 better performance models with their better acceptable unit number in Step 10. In Step 11, the best performance model with its best relevant unit number as well as the best supervising data set can be chosen for the given lead time.

The following indices to judge the performance of the ANN-SFM are implemented: the correlation coefficient (CC) and normalized root mean square error (NRMSE) in percentage as done in Kim et al. [12].

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{i=n}\left(\eta_{obs,i} - \eta_{fore,i}\right)^2}}{\left(\eta_{obs,\max} - \eta_{obs,\min}\right)} \tag{11}$$

$$CC = \frac{\sum_{i=1}^{i=n}\left(\eta_{obs,i}-\bar{\eta}_{obs}\right)\left(\eta_{fore,i}-\bar{\eta}_{fore}\right)}{\sqrt{\sum_{i=1}^{i=n}\left(\eta_{obs,i}-\bar{\eta}_{obs}\right)^2\sum_{i=1}^{i=n}\left(\eta_{fore,i}-\bar{\eta}_{fore}\right)^2}} \quad (12)$$

where, $\eta_{obs,i}$ is the observed surge level, $\eta_{fore,i}$ is the forecasted surge level, $\eta_{obs,i}$ is the averaged observation, $\eta_{fore,i}$ is the average value of the forecasted surge level, $\eta_{obs,max}$ is the observed highest surge level, $\eta_{obs,min}$ is the observed lowest surge level and $i = 1, 2, \ldots, n$.

4. Results and discussion

4.1 Accuracy of the good performance model

Due to initially and randomly assigned weights and biases in the training phase, results obtained from individual ANN-SFMs show differences in accuracy, even though those are trained by the identical input and output data, and training algorism. For example, the independent 957 ANN-SFMs, $N_{k=1,\ldots,957}^{j=10}D_{i=1}$, were trained by the data set, $D_{i=1}$, with the 10 units for the 24 h lead time. Then, their correlation coefficients (CCs) were calculated, as shown in Fig. 3 (a). It was found that CCs are scattered in the range of 0.5411 to 0.9474. When having a look at $N_{k=1,\ldots,20}^{j=10}D_{i=1}$ in the horizontal axis, the highest CC value is found at 0.9046, as seen in Fig. 3 (b). The model, $N_{k=2}^{j=10}D_{i=1}$, indexing the highest CC of 0.9046 can be treated as a good performance model among the 20 candidate models, $N_{k=1,\ldots,20}^{j=10}D_{i=1}$. Meanwhile, the model, $N_{k=890}^{j=10}D_{i=1}$, reveals the highest CC of 0.9474 among the 957 models, $N_{k=1,\ldots,957}^{j=10}D_{i=1}$. In other words, the model, $N_{k=890}^{j=10}D_{i=1}$, may be selected for a good performance model among the 957 ones. Consequently, it was found that the accuracy of ANN-SFM randomly varies with some ranges. However, the appropriate number ($N_{k=1,\ldots,max}$) of ANN-SFM ($N_k^jD_i$), is not well known for determining the good performance model: how many models are sufficient is uncertain for choosing the most accurate model. Therefore, in the present study, it is assumed that the maximum = 20 ($N_{k=1,\ldots,20}$) is appropriate to decide the good performance model in order to reduce time-consuming processes. The determination of the optimal number of the model is beyond the scope of this study.

Furthermore, it was found that the values of CC and NRMSE sometimes indicate different performances of ANN-SFM. For instance, when the data set, $D_{i=1}$, trains the ANN-SFM with the unit number = 60, the values of CC and NRMSE indicated the different models as the good performance models, as seen in Fig. 4: the highest CC value can be obtained from the model, $N_{k=11}^{j=60} D_{i=1}$, while the lowest NRMSE value is acquired from the model, $N_{k=16}^{j=60} D_{i=1}$. These results are in line with results described in Mentaschi et al. [27]. Hence, the use of two indices makes it more difficult in deciding which ANN-SFM is best/better/good. Therefore, the statistical indicator (HH) is finally implemented, as defined by

$$HH = \sqrt{\frac{\sum_{i=1}^{i=n} \left(\eta_{fore\,i} - \eta_{obs\,i}\right)^2}{\sum_{i=1}^{i=n} \eta_{fore\,i}\,\eta_{obs\,i}}} \qquad (13)$$

which is introduced initially by Hanna and Heinold [7], to overcome the above difficulties.

4.2 Accuracy of the better performance model

First, the good performance model among the twenty models trained with each unit number in the data set, $D_{i=1}$, consisting of the surge level (SS), the sea-level pressure (SLP), and the depression rate of the sea-level pressure (DSLP), is selected in Steps 1 to 4 and evaluated by the statistical indices of CC, NRMSE and HH. Figure 5 shows the indices of the selected good performance models in each unit number in the data set, $D_{i=1}$. The good performance ANN-SFMs trained for the lead times of 5 and 12 hours reveal the NRMSE values of approximately 5 to 7 % with the range of 1 % along the unit number, while those for the lead time of 24 hours shows the NRMSE values larger than 7 % with the bandwidth of 3 %, see Fig. 5(a). There is a similar tendency in the value of CC, see Fig. 5 (b): the CC values for the lead times of 5 and 12 hours vary with the range of 0.005, while those for the 24h lead time change with the range of 0.06. The change of HH is also apparent as the lead time is longer, see Fig. 5 (c). Hence, it is clear that the fluctuation width of the model error becomes substantial when the lead time becomes longer. Figure 5 shows no evidence of a tendency of a significant improvement in the forecasted surge level, as increasing the unit number in the data set, $D_{i=1}$. As discussed in the previous section, the values of CC and NRMSE are diverse for the 5 h lead time: $N_{k=2}^{j=50}D_{i=1}$ is better for CC, while $N_{k=14}^{j=30}D_{i=1}$ is better for NRMSE, respectively, see Fig 4 (a). Therefore, from now on the index of HH will be used to evaluate the ANN-SFM model to remove the uncertainty came from the difference of the evaluations between two indices of CC and NRMSE. As a result, the better performance ANN-SFMs is achieved based on three indices in the data set, $D_{i=1}$: $N_{k=6}^{j=50}D_{i=1}$, $N_{k=2}^{j=130}D_{i=1}$, and $N_{k=17}^{j=70}D_{i=1}$ with the associated unit number (j): 50, 130, and 70 for the 5, 12, and 24 h lead times, respectively.

Figures 6, 7, and 8 show the HH values of the good performance models evaluated in each unit number in each data set for the 5, 12, and 24 h lead times, respectively. For the data set, $D_{i=2}$, combining SS, SLP, DSLP, and WS, the ANN-SFMs for the 5 and 12 h lead times reveal the HH values in the range of approximately 0.17 to 0.23 against the unit number, while the ANN-SFMs for the 24 h lead time show the highest HH in the range of 0.27 to 0.33. Consequently, it can be found that a fluctuation of the HH values becomes larger, when the lead time is longer, as found in the data set $D_{i=1}$. The statistical indicators of HH show that accuracies of the ANN-SFMs for the 5, 12, and 24 h lead times are insignificantly improved, as increasing the unit number. Nevertheless, the better performance models with the associated unit numbers of 40, 190, and 110 are selected for each lead time: $N_{k=17}^{j=40}D_{i=2}$, $N_{k=20}^{j=190}D_{i=2}$, and $N_{k=5}^{j=110}D_{i=2}$, respectively.

Next, the HH values of the good performance models selected in each unit number trained by the data set, $D_{i=3}$, are plotted, when adding the wind direction (WD) to the data set, $D_{i=2}$. Overall, a behavior of the HH values against the unit number in the data set, $D_{i=3}$, is very similar to that in the data set, $D_{i=2}$. The values of HH for the 5 and 12 h lead times are similar each other in the range of 0.15 to 0.22, while those for 24 h lead time are slightly larger in the range of 0.25 to 0.32. The results show that the better performance models with the associated unit number of 70, 40, and 170 were obtained for the 5, 12, and 24 h lead times: $N_{k=3}^{j=70}D_{i=3}$, $N_{k=18}^{j=40}D_{i=3}$, and $N_{k=20}^{j=170}D_{i=3}$, respectively.

The results of the data set, $D_{i=4}$, which combines SS, SLP, DSLP, and the typhoon position (both the longitude (LG) and latitude (LT)), show that the HH values of the good performance models for all the lead times are in the range of 0.07 to 0.14. The HH values in the data set, $D_{i=4}$, are significantly lower than those in the data sets, $D_{i=1,2,3}$. Especially, it was found that the 24 h lead time forecast model trained by the data set, $D_{i=4}$, obtains the significantly lower HH values in comparison with those acquired by the 24 h lead time forecast models trained by the data sets, $D_{i=1,2,3}$. In addition, the ranges of the HH values in the data set, $D_{i=4}$, are apparently narrower than those in the data sets, $D_{i=1,2,3}$. As a result, the use of the unit numbers, $u^{j=170}$, $u^{j=120}$, and $u^{j=160}$, gives the better performance for all the lead times: $N_{k=10}^{j=170}D_{i=4}$, $N_{k=10}^{j=120}D_{i=4}$, and $N_{k=13}^{j=160}D_{i=4}$, respectively.

Contrary to the results in the previous data sets, $D_{i=1,2,3,4}$, it is found in the data set, $D_{i=5}$, that accuracies of the models tend to diverse and become lower when the good performance models are trained by SS, SLP, DSLP, LG, LT, and WS in the data set, $D_{i=5}$. For example, the HH values for the 24 h lead time seem to increase from 0.15 to 0.24 as increasing the unit number. Based on the HH values, the 24 h-forecast ANN-SFM, $N_{k=18}^{j=130}D_{i=5}$, could be selected as the better performance model among the 20 good performance models. For the 5 and 12 h-forecast ANN-SFMs, the models, $N_{k=8}^{j=40}D_{i=5}$ and $N_{k=4}^{j=70}D_{i=5}$, appear to be the better-performed models, respectively.

In the data set, $D_{i=6}$, including SS, SLP, DSLP, LG, LT, WS and WD, the HH values tend to be larger as increasing the unit number, which is similar to the tendency of the data set, $D_{i=5}$. This tendency is much significant in the HH evaluated for the 24 h-forecast model. From the results of the data set, $D_{i=6}$, the use of the unit number, $u^{j=10}$, is the best rather than the use of others. As a result, the ANN-SFMs, $N_{k=3}^{j=10}D_{i=6}$, $N_{k=12}^{j=10}D_{i=6}$, and $N_{k=6}^{j=10}D_{i=6}$, are chosen for the better performance models for all the lead times, respectively.

When training a model with the data set, $D_{i=7}$, which combines SS, SLP, DSLP, LG, LT, and the central atmospheric pressure of typhoon (CAP), the HH value of good performance models shows a much similar change to that of the previous models trained by the data set, $D_{i=4}$, which consists of SS, SLP, DSLP, LG, and LT. As discussed in the data set, $D_{i=4}$, the HH values for all three lead times scatter below 0.12 along with the unit number. The scattering pattern of $D_{i=7}$ is very similar to that of the data set, $D_{i=4}$. On the other hand, a variation of HH in the data set, $D_{i=7}$, is significantly narrower and more stable than that in the data set, $D_{i=4}$. For instance, the HH value of the 24 h lead time in the data set, $D_{i=7}$, shows the smaller range of 0.0276 between 0.0784 to 0.106 in comparison with the corresponding HH value in the data set, $D_{i=4}$, which is in the range of 0.0644 between 0.0774 to 0.1418. As done in the previous data sets, the better performance ANN-SFMs, $N_{k=1}^{j=100}D_{i=7}$, $N_{k=6}^{j=90}D_{i=7}$, and $N_{k=3}^{j=40}D_{i=7}$, with the unit numbers, $u^{j=100}$, $u^{j=90}$, and $u^{j=40}$, could be chosen for the 5, 12, and 24 h lead times, respectively.

When the ANN-SFM is trained by the data set, $D_{i=8}$, which combines SS, SLP, DSLP, LG, LT, CAP, and WS, the HH value in $D_{i=8}$ is generally larger than that in $D_{i=7}$. In this data set, the HH value has a slightly increasing tendency, as increasing the unit number. Among the twenty good performance models, the better performance ANN-SFMs, $N_{k=14}^{j=120}D_{i=8}$, $N_{k=8}^{j=80}D_{i=8}$, and $N_{k=7}^{j=70}D_{i=8}$, with the unit numbers, $u^{j=10}$, $u^{j=80}$, and $u^{j=70}$, could be selected for the lead times of` 5, 12, and 24 hours, respectively.

In the data set, $D_{i=9}$, consisting of SS, SLP, DSLP, LG, LT, CAP, WS and WD, the HH value shows a similar trend to that obtained from the data set, $D_{i=8}$: an accuracy becomes lower, as increasing the unit number. Among the twenty good performance models, the better performance models are found with the associated unit numbers of 70, 190, and 10 for the lead times of 5, 12, and 24 hours: $N_{k=6}^{j=70}D_{i=9}$, $N_{k=16}^{j=80}D_{i=9}$, and $N_{k=3}^{j=10}D_{i=9}$ respectively.

When training the ANN-SFM with the data set, $D_{i=10}$ that combines SS, SLP, DSLP, LG, LT, CAP, and HWS, the HH value particularly for the 24 h-forecast ANN-SFM increases in the range of 0.13 to 0.20, as increasing the unit number. On the other hand, the HH values are relatively stable in the 5 and 12 h-forecast models, fluctuating in the range of 0.0574 to 0.1018. Among the selected twenty good performance models, the better performance models with the associated unit numbers of 70, 160, and 30 were chosen for the lead times of 5, 12, and 24 hours: $N_{k=14}^{j=70}D_{i=10}$, $N_{k=6}^{j=160}D_{i=10}$, and $N_{k=12}^{j=30}D_{i=10}$, respectively.

When using the data set, $D_{i=11}$, of SS, SLP, DSLP, LG, LT, CAP, HWS, and WS, the behavior of the HH values for all the forecasts is similar to that of the data set, $D_{i=10}$: it becomes in general larger, as increasing the unit number. In the case of the data set 11, $D_{i=11}$, the better performance models with the unit numbers of 30, 20, and 20 were selected for the lead times of 5, 12, and 24 hours: $N_{k=14}^{j=30}D_{i=11}$, $N_{k=8}^{j=20}D_{i=11}$, and $N_{k=15}^{j=20}D_{i=11}$, respectively.

Finally, in the data set, $D_{i=12}$, when enlarging the unit number, the HH value significantly increases in all the good performance forecast models. The HH values apparently vary from 0.095 to 0.20. Similar to the models in the data set, $D_{i=11}$, the better performance models appear at the smallest/small unit number and then, the relevant unit numbers were 10, 10, and 30 for the lead times of 5, 12, and 24 hours: $N_{k=18}^{j=10}D_{i=12}$, $N_{k=17}^{j=10}D_{i=12}$, and $N_{k=12}^{j=30}D_{i=12}$, respectively.

Until now, we have assessed the performance of the 5, 12, and 24 h-forecast ANN-SFMs trained, validated, and tested by the data sets, $D_{i=1,\ldots,12}$, which combine the meteorological, hydrodynamic and typhoon-characteristic parameters, to find out the appropriate set of the input parameter and the unit number in the hidden layer. Through the selection procedure, it was found that as increasing the unit number, the HH value for the 5 h-forecast ANN-AFM constantly fluctuates within the specific ranges in the data sets of $D_{i=1,2,3,4,7,8,10}$. On the other hand, the HH value becomes gradually larger in the data sets of $D_{i=5,6,9,11,12}$. For the 12 h-forecast ANN-AFM, the HH value is constant in the data sets of $D_{i=1,2,3,4,7,10}$ when the unit number increases, but it increases in the data sets of $D_{i=5,6,8,9,11,12}$. For the 24 h-forecast ANN-AFM, the HH value fluctuates with the constant range in the data sets of $D_{i=1,3,4,7}$, while it gradually increases in other the data sets. Also, it was found for the 24 h-forecast model trained by the data sets of $D_{i=5,6,8,9,10,11,12}$ that the better performance ANN-AFM is obtained by using the smallest unit number as shown in other studies. It becomes apparent that the accuracy of the ANN-SFM may not always accompany with the increase of the unit number. As in consequence, the performance of the ANN-SFM shows the unique feature depending on not only the data set but also the unit number. Therefore, it can be said that the selection process proposed in the present study is able to determine the better performance model with the appropriate set of the input parameter and the unit number.

4.3 Accuracy of the best performance model

The best performance model accompanied by the most appropriate set of the input parameters and the unit number can be determined among the twelve better performance models previously chosen in the twelve data sets, as shown in Fig. 9. When the data sets, $D_{i=1,2,3}$, which combine the parameters of SS, SLP, DSLP, WS, and WD, train ANN-SFMs, their accuracies are generally lower than the others, for instance, trained by the data sets, $D_{i=4,\dots,12}$. When training with the data set, $D_{i=10}$, which consists of SS, SLP, DSLP, LG, LT, CAP, and HWS, the highest accuracies of the 5 and 12 h-forecast ANN-SFMs are obtained with the most relevant unit numbers of 70 and 160, respectively (also see Fig. 10). For the 24 h-forecast ANN-AFM, the higher accuracy could be acquired when training with the data sets, $D_{i=4}$ (the parameters: SS, SLP, DSLP, LG, and LT) and $D_{i=7}$ (SS, SLP, DSLP, LG, LT, and CAP). Thus, the best performance 5 and 12 h-forecast ANN-SFMs ($N_{k=14}^{j=70} D_{i=10}$ and $N_{k=6}^{j=160} D_{i=10}$) are chosen with the unit numbers, $u^{j=70}$ and $u^{j=160}$, that are trained by the data set, $D_{i=10}$ (SS, SLP, DSLP, LG, LT, CAP, and HWS). The best performance 24 h-forecast ANN-SFM, $N_{k=13}^{j=160} D_{i=4}$, is determined with the most appropriate combination of the unit number, $u^{j=160}$ and the data set, $D_{i=4}$, consisting of SS, SLP, DSLP, LG, and LT (also see Fig. 11).

For the 5 and 12 h lead times, the forecast model starts to operate from a time when the typhoon is more closely approaching to the coast of Sakai Minato where the typhoon's wind and pressure field affects the sea level rise. Because of this reason, the inclusion of the typhoon characteristics as well as the local sea level pressure in the input data set is highly imperative, for achieving the best performance. As the typhoon either approaches to or gets away from Sakai Minato, it can be expected that the relation between the storm surge and the typhoon characteristics might be strongly correlated with each other even though there is a short lag.

Therefore, for the 24 h lead time, the forecast model starts from a time when the typhoon might not impact weather conditions in Sakai Minato. It might not generate the sea level disturbance on the coast of Sakai Minato, because the typhoon is in offshore sea toward the southwest of Kyushu Island, where is in a distance of approximately 700 km. As the typhoon moves to or away from Sakai Minato, one would expect a weaker correlation between the storm surge and the typhoon characteristics except for the typhoon position. Therefore, the critical parameters may be clarified into two groups: one is the local sea level pressure and the typhoon position; another is the typhoon characteristics, especially, for the 24-lead time at Sakai Minato.

4.4 Discussion

In the present study, the best performance ANN-SFM is chosen through the selection procedure to improve the accuracy of the storm surge forecast at the Sakai Minato on the Tottori coast, as focusing on the determination of the input parameter in the input layer and the unit number in the hidden one. As a result, the accuracy of the ANN-SFM is improved. The lowest HH values (HH_{best}) are 0.057, 0.065, and 0.1185 evaluated in the best performance 5, 12, and 24 forecast models ($N_{k=14}^{j=70}D_{i=10}$, $N_{k=6}^{j=160}D_{i=10}$, and $N_{k=13}^{j=160}D_{i=4}$), respectively, on the other hand, the highest HHs (HH_{worst}) are 0.420, 0.5398, and 0.8869 in the worst performance 5, 12, and 24 forecast models ($N_{k=14}^{j=150}D_{i=2}$, $N_{k=5}^{j=100}D_{i=2}$, and $N_{k=12}^{j=180}D_{i=2}$)). Their differences between the highest and lowest HHs are 0.363, 0.4748, and 0.7684 for the 5, 12, and 24 h-forecast models, respectively: the difference can be reduced up to 86.4 % to 138.6 %, which is calculated by:

$$((HH_{worst} - HH_{best})/HH_{worst}) \times 100 \quad (14)$$

Through the systematic and objective selection procedure, the indefiniteness comes from the following questions might be at least removed: how many units in the hidden layer are necessary; what input parameters consist of in the data set; what combination of the unit and the parameter is most appropriate.

Furthermore, the following question will remain from the physical perspectives: why the data sets, $D_{i=10}$ (storm surge (SS), sea level pressure (SLP), depression of sea level pressure (DSLP), longitude (LG), latitude (LT), central atmospheric pressure (CAP), and highest wind speed (HWS)) and $D_{i=4}$ (SS, SLP, DSLP, LG, and LT) are suitable for the 5 and 12, and 24 h lead times, respectively. Based on the current results, the parameters, SS, SLP, DSLP, LG, LT, CAP, and HWS, significantly affected, and are most relevant to the accuracy of the forecast models with the 5 and 12 h lead times. In addition, it can be said that the data sets, $D_{i=4}$, (SS, SLP, DSLP, LG, and LT) and, $D_{i=7}$, (SS, SLP, DSLP, LG, LT, and CAP) are also impactive for the 5 and 12 h-forecast models. In other words, when the storm surge is shortly forecasted with the 5 or 12 h lead times at Sakai Minato, the storm surge level is absolutely governed by the typhoon characteristics that are the longitude, the latitude, the central atmospheric pressure (CAP), and the highest wind speed (HWS) near the typhoon center. However, when the storm surge is predicted with the 24 h lead time at Sakai Minato, the accuracy of the 24 h-forecast model is apparently influenced by the parameters of the local atmospheric components observed near Sakai Minato, the typhoon position, and its central pressure.

Also, the further question can be directed to the regional specification for the best combination of the appropriate unit number and input parameter set. While the best performing data set of the input parameters is highly dependent on a specific area, the most suitable unit number in the hidden layer relevant to the best data set seems to be less accurate for the area.

In other words, the clarification of the input parameters should have a higher priority over the determination of the unit number in a region. The first approach to the development of the 5 and 12 h-forecast models might be the use of the combination of the local sea level pressures and typhoon characteristics, while that of the 24 h-forecast one might be the use of the local sea level pressures and typhoon position. Then, other combinations of the potential input parameters are able to suggest the candidates to be applied to the present novel selection procedure. Once the appropriate data set is determined, and the unit number can be then found out by using the selection procedure proposed in the study. Unlike Sakai Minato, the best performing set may vary with a particular location, probably depending on characteristics of a storm surge and a typhoon. Nevertheless, the novel systematic and objective selection procedure can be applied to other sites when developing an artificial neural network-based storm surge forecast model for a relatively long lead time.

The best performed-storm surge forecast models show the highly promising accuracy, as shown in Figs. 18 and 19. In the study, we have dealt with the data 460 hours long collected from three typhoon events in order to propose the novel selection procedure. In the meanwhile, one might be curious about the actual forecast capacity of the best performing model. The present study has been aiming at the introduction to the novel method, which one could determine the best performing model and its relevant unit number among the potentials, the capacity in the real practice has been beyond the present study so far. For this reason, the current best performing model chosen in the study could not be sustained in the practical application. One thing was not considered was the training data size that is relatively short to directly adopt the forecast model chosen here to the field site. Therefore, for guaranteeing the feasibility of the storm surge forecast model, which is trained and then decided, against the independent typhoon event or future one, the training with the massive data size collected from the typhoons and accompanied input components is essential. In the practical application, if the storm surge and relevant atmospheric components measured are sufficiently long, the feasibility could promise. However, the real site might not acquire such the data length for ensuring the viability. One of its alternatives to make a sufficiently lengthy database is the reproduction of the historical typhoon and storm surge event using an atmospheric general circulation model and ocean circulation one. Inclusion of the future event, extracted from, for instance, a Database for Policy Decision making for Future climate change (d4PDF) (Mizuta et al., [28]), in the database might be one of alternative ways.

5. Conclusions

The present study proposed a systematic and objective selection procedure for establishing a suitable Artificial Neural Network-based storm Surge Forecast Model (ANN-SFM) and described the application of ANN-SFM to the Sakai Minato port on the Tottori coast, Japan to predict the storm surge level with the lead times of 5, 12, and 24 hours. In the selection procedure, twenty ANN-SFMs are individually trained by the data set with 10 unit numbers, and the best performed model is then selected from the statistical assessment of HH given in Eq. (3). The procedure is repeated by varying the unit number from 20 to 200 with the increment of 10 units. As a result, the twenty good performance ANN-SFMs are chosen in each unit number. Among the twenty good performance ANN-SFMs, the better performance model is selected with the relevant set of the unit number and the input parameter. These steps are iteratively conducted against the twelve data sets, consisting of the meteorological and hydrodynamic parameters observed at the local stations on the Tottori coast (the surge level, the sea level pressure, the drop rate of the sea level pressure, the wind speed, and the wind direction) and the typhoon characteristic parameters (the longitude and latitude of the typhoon, the central atmospheric pressure of the typhoon, and the highest wind speed near the typhoon center). Hence, the twelve better performance ANN-SFMs are selected in each data set. Among the twelve better performance models, the best performance ANN-SFM is finally determined with the associated set of the unit number and the input parameters.

From the results of applications at Sakai Minato, it was found that the accuracy of the ANN-SFM fluctuated with some ranges and was not necessarily improved with solely increasing the unit number. For instance, when training the ANN-SFM for the 24 h lead time with the data set of surge level, sea level pressure and the drop rate of sea level pressure, the accuracy of the model was constant even though the unit number was increased. But, when training it with the data set of the surge level, the sea level pressure, and the drop rate of the sea level pressure, the longitude and latitude of the typhoon, and the wind speed, the accuracy is improved when increasing the unit number. In this manner, the better performance model can be chosen among the twenty good performance models as varying the unit number. Of twelve better performance models with the associated unit numbers in each data set, the best performance model can be established with the relevant set of the unit number and the input parameters. At Sakai Minato on the Tottori coast, the best performance 5 h (12 h)-forecast models are made when paring 70 units (160 units) and the input parameters of surge level, sea level pressure, the depression rate of sea level pressure, longitude and latitude, central atmospheric pressure and highest wind speed. The best performed 24 h-forecast ANN-SFM was determined as combining the 160 units and the input parameter of the surge level, the sea level pressure, the depression rate of the sea level pressure, and the longitude and latitude.

As discussed in Section 4.4, the characteristics of the best storm surge forecast model systematically determined in this study may not be universal in terms of the appropriate set of the unit number and the input parameters. Otherwise, the systematic selection procedure proposed in the present study is applicable to develop the ANN-based storm surge forecast models on a coast.

Acknowledgements

This work was partially conducted under JSPS KAKENHI, Japan.

References

[1] Deo, M. C. and G. Chaudhari (1998) Tide prediction using neural networks, *Computer-Aided Civil and Infrastructure Engineering*, 13, p113-120.

[2] Deo, M. C. and C. Sridhar Naidu (1999) Real time wave forecasting using neural networks, *Ocean Engineering,* 26, 191-203.

[3] Deo, M. C., A. Jha, A. S. Chaphekar, K. Ravikant (2001) Neural networks for wave forecasting, *Ocean Engineering*, 28, 889-898.

[4] Dreyfus, G. (2002) Neural Networks: Methodology and Applications, Springer, p497.

[5] Flather, R. A. (1994) A Storm Surge Prediction Model for the Northern Bay of Bengal with Application to the Cyclone Disaster in April 1991, J. Phys. Oceanogr., 224, 172–190.

[6] Hiyajo, H., Okubo, S., Takasa, S., Kobashigawa, Y., Nishimura, F, Toomine, T., Daimon, H., Itagaki, S., Fukuda, M., Sakaji, T., Taguchi, H., Egami, H., Suzuki, H. and Nozaki, F. (2011) Digitizing the Historical Sea Level Data and Re-Analysis of Storm Surges with the Data, *Meteorology Bulletin by Japan Meteorological Agency*, 78, S1-S32 (In Japanese) (http://www.jma.go.jp/jma/kishou/books/sokkou-kaiyou/78/vol78s001.pdf)

[7] Hanna, S. and Heinold, D. (1985) Development and application of a simple method for evaluating air quality. In: API Pub. No. 4409, Washington, DC, Washington, USA.

[8] Jelesnianski, C. P., J. Chen, and W. A. Shaffer (1992) SLOSH - Sea, Lake, and Overland Surges from Hurricanes. U.S. Department of Commerce, NOAA Tech. Report NWS 48, 71pp.

[9] Jia, G., Taflanidis, A.A., Nadal-Caraballo, N.C., Melby J.A., Kennedy A.B., Smith J.M. (2016) Surrogate modeling for peak or time-dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms, Nat Hazards, 81, 909. doi:10.1007/s11069-015-2111-1

[10] Kim, S.Y., Yasuda, T., Mase, H., (2008) Numerical analysis of effects of tidal variations on 1065 storm surges and waves. Appl. Ocean Res. 30, 311–322.

[11] Kim, S., Matsumi, Y., Yasuda, T., Mase, H. (2014) Storm surges along the Tottori coasts following a typhoon, Ocean Engineering, 91, 133-145.

[12] Kim, S., Y., Matsumi, S., Pan, H., Mase (2016) A real-time forecast model using artificial neural network for after- runner storm surges on the Tottori coast, Japan, Ocean Engineering 122, 44–53.

[13] Kim, S., Melby, J.A., Nadal-Caraballo, N.C., Ratcliff j. (2015) A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling, Nat Hazards, 76, 565. doi:10.1007/s11069-014-1508-6

[14] Lee, T. L., C. P. Tsai, D. S. Jeng and R. J. Shieh (2002) Neural network for the prediction and supplement of tidal record in Taichung Harbor, Taiwan, *Advances in engineering software*, 33, 329-338.

[15] Lee, T. L. (2004) Back-propagation neural network for long-term tidal predictions, *Ocean Engineering*, 31, 225-238.

[16] Lee, T. L. (2006) Neural network prediction of a storm surge, *Ocean Engineering*, 33, 483-494.

[17] Lee, T. L. (2008) Back-propagation neural network for the prediction of the short-term storm surge in Taichung harbor, Taiwan, *Eng. Appl. Of Artificial Intellignece*, 21, 63-72.

[18] Lee, T. L. (2009) Predictions of typhoon storm surge in Taiwan using artificial neural networks, *Advances in Engineering Soft*ware, 40, 1200-1206.

[19] Luettich, R.A. and J.J. Westerink (1991) A solution for the vertical variation of stress, rather than velocity, in a three-dimensional circulation model, *International Journal for Numerical Methods in Fluids*, 12:911-928

[20] Makarynskyy, O., D. Makarynska, M. Kuhn, W.E. Featherstone (2004) Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia, *Estuarine, Coastal and Shelf Science*, 61, 351–360.

[21] Makarynska, D., O. Makarynskyy (2008) Predicting sea-level variations at the Cocos (Keeling) Islands with artificial neural networks, *Computers & Geosciences*, 34, 1910-1917.

[22] Marzenna S. (2003) Forecast storm surge by means of artificial neural network, *Journal of Sea Research*, 49, 317-322.

[23] Mase, H., Sakamoto, M. and Sakai, T. (1995) Neural network for stability analysis of rubble-mound breakwaters, Jour. Waterway, Port, Coastal, and Ocean Eng., ASCE, Vol.121, No.6, pp.294-299

[24] Mase, H. and Kitano, T. (1999) Prediction model for occurrence of impact wave force, Ocean Eng., Vol.26, No.10, pp.949-961.

[25] Mase, H., Reis, M.T., Nagahashi, S., Saitoh, T., T.S. Hedges, T.S. (2007) Effects of zero-overtopping data in artificial neural network predictions, Proc. 5th Coastal Structures Int. Conf., pp.1542-1551.

[26] Mase, H., T. Yasuda, N. Mori (2011) Real-time prediction of tsunami magnitudes in Osaka Bay, Japan, using an artificial neural network, *J. Waterway, port, coastal, and ocean eng.*, DOI: 10.1061/(ASCE)WW.1943-5460.0000092

[27] Mentaschi, L., Besio, G., Cassola, F., Mazzino, A. (2013) Problems in RMSE-based wave model validations, *Ocean Modelling*, 72, 53-58.

[28] Mizuta, R., A. Murata, M. Ishii, H. Shiogama, K. Hibino, N. Mori, O. Arakawa, Y. Imada, K. Yoshida, T. Aoyagi, H. Kawase, M. Mori, Y. Okada, T. Shimura, T. Nagatomo, M. Ikeda, H. Endo, M. Nosaka, M. Arai, C. Takahashi, K. Tanaka, T. Takemi, Y. Tachikawa, K. Temur, Y. Kamae, M. Watanabe, H. Sasaki, A. Kitoh, I. Takayabu, E. Nakakita, and M. Kimoto, 2016: Over 5000 years of ensemble future climate simulations by 60 km global and 20 km regional atmospheric models. Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-16-0099.1

[29] Peres, D. J., C. Iuppa, L. Cavallaro, A. Cancelliere, E. Foti (2015) Significant wave height record extension by neural networks and reanalysis wind data, *Ocean Modelling*, 94, 128-140.

[30] Tagliaferri, F., I.M. Viola, R.G.J. Flay (2015) Wind direction forecasting with artificial neural networks and support vector machines, *Ocean Eng.*, 97, 65–73.

[31] Tseng, C. M., C. D. Jan, J. S. Wang and C. M. Wang (2007) Application of artificial neural networks in typhoon surge forecasting, *Ocean Eng.*, 34, 1757-1768.

[32] Wang, S. H., K., Muhammad, Y. Lv, Y. Sui, L. Han, and Y. D., Zhang (2018) Identification of alcoholism based on wavelet renyi entropy and three-segment encoded jaya algorithm, , Complexity, 2018, vol. 2018, Article ID: 3198184

[33] Zhang, Y., P., Phillips, S., Wang, G., Ji, J., Yang and J., Wu (2016) Fruit Classification by Biogeography-based Optimization and Feedforward Neural Network, Expert Systems, 2016, 33(3): 239-253

Captions

Table 1. Parameters of the Levenberg-Marquardt back-propagation method in the training and validation phases.

Table 2. List of data sets: wind speed (WS), wind direction (WD), sea-level pressure (SLP), drop of sea-level pressure (DSLP), longitude (LG) and latitude (LT) of typhoon, central atmospheric pressure of typhoon (CAP), highest wind speed near typhoon center (HWS), and surge level (SS: the difference between the observed and predicted sea surface levels).

Fig. 1 The Tottori coast with typhoon tracks. (a) Typhoons with symbols (is typhoon and is cyclone or tropical depression); (b) the stations for the meteorological and hydrodynamic parameters (●: the meteorological station and : the hydrodynamic station)

Fig. 2 Schematic flow of the selection procedure for selecting the best performance model with the pair of relevant unit number and data set for a given lead time.

Fig. 3 Correlation coefficients of the ANN-SFMs with the 24 h lead time after testing using the data set 1: (a) is the correlation coefficients obtained by the 957 ANN-SFMs, $N_{k=1,\ldots,957}^{u^{j=10}} D_{i=1}$. (b) same but the 1st ANN-SFM to 20th one (= $N_{k=1,\ldots,20}^{u^{j=10}} D_{i=1}$).

Fig. 4 Correlation coefficients (CCs), normalized root mean square errors (NRMSEs, %) and statistical indicators (HHs) against the unit number, evaluated among the 20 ANN-SFMs (=

$N_{k=1,\dots,20}^{u^{j=10}} D_{i=1}$) in each unit number after training, verifying and testing them by the data set, $D_{i=1}$, for the 5 h lead time.

Fig. 5 Normalized root mean square errors (NRMSEs, %) in (a), Correlation coefficients (CCs) in (b) and statistical indicators (HHs) in (c) of the good performance ANN-SFMs in each unit number of $u^{j=10}$ to $u^{j=200}$trained by the data set, $D_{i=1}$ for the lead times of 5, 12 and 24 hours.

Fig. 6 Statistical indicators (HHs) of the good performance ANN-SFMs in each unit number of $u^{j=10}$ to $u^{j=200}$trained by the data set, $D_{i=2}$ for the lead times of 5, 12 and 24 hours.

Fig. 7 Same as Fig. 6 but using the data set, $D_{i=3}$.

Fig. 8 Same as Fig. 7 but using the data set, $D_{i=4}$.

Fig. 9 Same as Fig. 8 but using the data set, $D_{i=5}$.

Fig. 10 Same as Fig. 9 but using the data set, $D_{i=6}$.

Fig. 11 Same as Fig. 10 but using the data set, $D_{i=7}$.

Fig. 12 Same as Fig. 11 but using the data set, $D_{i=8}$.

Fig. 13 Same as Fig. 12 but using the data set, $D_{i=9}$.

Fig. 14 Same as Fig. 13 but using the data set, $D_{i=10}$.

Fig. 15 Same as Fig. 14 but using the data set, $D_{i=11}$.

Fig. 16 Same as Fig. 15 but using the data set, $D_{i=12}$.

Fig. 17 Statistical indicators (HHs) against the data sets for the best performance 5, 12 and 24 h-forecast ANN-AFMs among the twelve better ANN-SFMs.

Fig. 18 Comparisons of observation and forecasts from the 5 and 12 h lead time ANN-SFMs ($N_{k=14}^{j=70}D_{i=10}$ and $N_{k=6}^{j=160}D_{i=10}$) after training, validating and testing with the relevant set of the 70 and 160 unit numbers and the input parameters of SS, SLP, DSLP, LG, LT, CAP and HWS.

Fig. 19 Comparisons of observation and forecasts from the 24 h lead time ANN-SFMs ($N_{k=13}^{j=160}D_{i=4}$) after training, validating and testing with the relevant set of the 160 unit number and the input parameters of SS, SLP, DSLP, LG and LT.