

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

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

24 Abstract

25

26 The present study describes a novel way of a systematic and objective selection procedure for
27 the development of an Artificial Neural Network-based storm Surge Forecast Model (ANN-
28 SFM) with 5, 12 and 24 h lead times and its application to Sakai Minato area on the Tottori
29 coast, Japan. The selection procedure can guide the determination of the superiority of the best
30 performing model in terms of the appropriate combination of unit number in the hidden layer
31 and parameter in the input layer. In the application of ANN-SFM to Sakai Minato, it is found
32 that the best ANN-SFMs for 5 and 12 h-forecasting are established with the most suitable set
33 of 70 units (the number of hidden neurons) and the input components of surge level, sea level
34 pressure, the depression rate of sea level pressure, longitude, latitude, central atmospheric
35 pressure and highest wind speed. The best ANN-SFM for 24 h-forecasting is determined with
36 160 units and the input parameters of surge level, sea level pressure, the depression rate of sea
37 level pressure, longitude and latitude. The proposed method of the selection procedure is able
38 to be adaptable to other coastal locations for the development of the artificial neural network-
39 based storm surge forecast model as establishing the superiority of the most relevant set
40 combining unit numbers and input parameters.

41

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

43

44

45 1. Introduction

46

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

55 Previously, much attention has been paid to the development of storm surge forecast
56 models to provide reliable surge levels to the coastal community prior to typhoons' landfall.
57 Approaches for the development of storm surge forecast models are mainly to use either
58 physical process-based numerical models (e.g., Luettich and Westerink [19]; Flather [5];
59 Jelesnianski et al. [8], Kim et al. [10]) or machine learning or data-driven methods (e.g., Lee
60 [16]; Tseng et al. [31]; Kim et al. [13]; Kim et al. [12], Jia et al. [9]). The physical process-
61 based models have the advantage of including physics of storm surge and taking distinct
62 external forces to drive themselves over the machine learning method. However, they are
63 generally time-consuming and cumbersome to be operated. On the other hand, the machine
64 learning method-based forecast models such as an artificial neural network (ANN) are lighter
65 and faster to be operated in comparison to the process-based model (Lee [16], [17], [18]; Kim
66 et al. [13]; Kim et al. [12]; Jia et al. [9]). But, when using these models to study the storm surge
67 phenomena, it is inconvenient and also sometimes laborious to find the appropriate input set
68 between meteorological, hydrodynamic and typhoon-characteristic parameters to efficiently
69 train the ANN-based surge forecast model (ANN-SFM) and accurately get the forecasted surge

70 level (in this way, a surge level is an output parameter). Because of these reasons, efforts have
71 been made to clarify relations between input and output parameters when developing ANN-
72 SFM at specific locations (e.g., Marzenna [22]; Lee [17], [18]; Tseng et al. [31]; Kim et al.
73 [12]). According to the recent studies for time-dependent after-runner storm surge forecasting
74 (e.g., Kim et al. [12]), it was shown that the proper input parameters vary with the area at which
75 the time-dependent surge forecasting is developed through a series of experiments, and
76 potential combination sets of the input parameters are found. The input parameters also differ
77 depending on the given lead time for forecasting. In addition, it was found that an ANN-SFM
78 for a relatively long lead time (for example, 24 hours) is apparently inaccurate in comparison
79 with that for a short lead time (for instance, 5 hours). In other words, these methods were
80 developed with some restrictions indirectly applying to other places as a universal ANN-SFM.

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

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

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

93 Through the entire procedure mentioned above, the development of ANN-SFM
94 based models is expected to improve the forecast accuracy. However, it will still require a

95 systematic and objective procedure to determine which factor should be first examined or has
96 primary impacts on the accuracy of the model. Therefore, the present study introduces the
97 systematic approach of a selection procedure to seek the best performance combination of the
98 input parameters in the input layer and the corresponding unit number in the hidden layer for
99 the lead times of 5, 12 and 24 hours, and applies the selection procedure of ANN-SFM to Sakai
100 Minato port on the Tottori coast, Japan. The selection procedure suggested in the study is first
101 to explore potential unit numbers of the hidden layer for each possible combination set of input
102 parameters, and then to determine the unit number relevant to each set. Finally, the best set of
103 input parameters can be selected among candidate sets of input parameters. As a result, the best
104 performance model can be selected with the optimized unit number relevant to the best
105 combination set of input parameters. The present study is to demonstrate the systematic
106 selection procedure which can be applied to rather the development of a specified ANN-SFM
107 at a location in any coastal regions than the development of a universal ANN-SFM. Also, in
108 view of the selection procedure, it will show the accuracy of the ANN-SFM being improved,
109 especially, forecasting for a lead time of 24 h, for instance.

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

113

114

115 2. Overview of artificial neural network (ANN)

116

117 ANN has been widely applied in many prediction and forecast studies in terms of
118 air quality (e.g., Hanna and Heinold [7]), tides (e.g., Deo and Chaudhari [1]; Lee et al. [14],
119 [15]), sea level rise (e.g., Makarynskyy et al. [20], [21]), maritime structures (e.g., Mase et al.

120 [23]), waves (e.g., Mase and Kitano [24]; Deo and Naidu [2]; Mase et al. [25]; Peres et al. [29]),
121 wind (e.g., Tagliaferri et al. [30]) and tsunami (Mase et al. [26]).

122

123 2.1 Artificial neural network (ANN)

124 The present study employs a feedforward network that information flows in the forward
125 direction from the input to the output (e.g., Dreyfus [4]). A structure of ANN used in the present
126 study is commensurate with the previous study (Kim et al. [12]) that consists of three single
127 layers, namely the input, hidden and output layers, with the use of the back-propagation
128 optimization technique for training. The Levenberg–Marquardt algorithm is used that has an
129 advantage in terms of the reduction of computational time among several back-propagation
130 algorithms (the conjugate gradient method, the scaled conjugate gradient method, the
131 Broyden–Fletcher–Goldfarb–Shanno method and the Levenberg–Marquardt method). Table 1
132 summarizes the parameters of the Levenberg–Marquardt method. Here, the regularization
133 method of early stopping is used to avoid either over- or under-learning of ANNs as described
134 by Mase et al. [26] and Dreyfus [4]: for instance, the training stops when the performance
135 function reaches a pre-defined threshold (for example 10^{-6} m) of the mean squared error
136 between the observed and predicted surge levels, before it completes the pre-set maximum
137 10,000 iterations. The functions of hyperbolic tangent sigmoid transfer and linear transfer are
138 implemented in the hidden and output layers, respectively. The phases of training and
139 validating are taken for the surge level forecast with the lead times of 5, 12, and 24 hours. The
140 test phase is then carried out to evaluate the ANN-SFM with the set of input parameters and
141 unit number that is determined in the phases of the test and validating. Section 3 will give a
142 detailed description of the phases for training, validation, and test.

143

144 2.2 Input and output parameters

145

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

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

163 **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)
$DSL P_p^t = \widetilde{DSL P}_p^t$	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)

164

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

169 For each case, three tasks, namely: training, validating, and testing, are carried out to
170 develop the ANN-SFM model with the lead times of 5, 12, and 24 hours. The data length is
171 460 hourly data. In the process, 75 % of all the input and output parameters is used for the
172 training phase, and the remaining 25 % is used for the validating phase. The testing phase is
173 conducted to evaluate the ANN-SFM with 25 % of all the parameters.

174

175 3. A series of numerical experiments

176

177 3.1 Data sets for a series of experiments

178

179 A series of experiments was carried out with the 12 data sets in order to determine
180 which data set performs best for predicting the surge level with the specific lead time. Table 2
181 lists those data sets. The data set, $D_{i=1}$, consists of the surge level (SS), the sea level pressure
182 (SLP), and the drop of sea level pressure (DSL P). In the data set, $D_{i=2}$, it is the same, but the

183 wind speed (WS) is added to $D_{i=1}$. In the data set, $D_{i=3}$, the wind direction (WD) is added to
184 $D_{i=2}$. For the data set of $D_{i=4}$, the components of SS, SLP, DSLP, and the longitude (LG) and
185 the latitude (LT) of the typhoon are comprised. In the data set, $D_{i=5}$, the component of WS is
186 added to that of $D_{i=4}$. The data set, $D_{i=6}$, contains all the components in the data set of $D_{i=5}$
187 as well as WD. The data set, $D_{i=7}$, is designed to include SS, SLP, DSLP, LG, and LT as well
188 as the central atmospheric pressure of the typhoon (CAP) in the data set. For the data set of
189 $D_{i=8}$, the component of WS is added to the data set, $D_{i=7}$. The component of WD is added to
190 the data set, $D_{i=8}$ to construct the data set, $D_{i=9}$. The data set, $D_{i=10}$, consists of SS, SLP,
191 DSLP, LG, LT, CAP, and the highest wind speed of typhoon (HWS). In the data set of $D_{i=11}$,
192 WS is added to the data set, $D_{i=10}$. Finally, WD is added to the data set, $D_{i=11}$, to make the
193 data set, $D_{i=12}$.

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

$$197 \quad N_{k=1}^{j=10} D_{i=1}(\eta^{t+\Delta t}) = (\eta^t, SLP_p^t, DSLP_p^t) \quad (10)$$

198
199 where Δt is the given lead time. For simplification, a symbol in parentheses is ignored.

201 **3.2 Selection procedure for the best performance forecast model**

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

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

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

226

227 It is found that the ANN-SFMs that are trained and verified by the same unit number

228 and data set predict slightly different surge levels because the initial weight and bias are

229 randomly given and then adjusted through the training phase, which will be shown later. Thus,

230 the present study proposes the selection procedure to judge the best performance ANN-SFM.

231 The selection procedure consists of two parts to decide the best acceptable unit number

232 and data set. First, the data set, $D_{i=1}$, is prepared in Step 1. The initial unit number = 10, $u^{j=10}$,

233 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,
234 and tested with a randomly given weight and bias in the hidden layer in Step 3. In Step 4, the
235 20 individual ANN- SFM, $N_{k=1,\dots,20}^{j=10} D_{i=1}$, with the constant unit number, $u^{j=10}$, are made by
236 initializing the weight and bias every time as repeating Step 3 twenty times. In Step 5, the one
237 good performance ANN-SFM can be then selected among the 20 ones, $N_{k=1,\dots,20}^{j=10} D_{i=1}$, by the
238 indices of Eqs. (11), (12) and (13). Once the one good performance model has been selected,
239 repeat Steps 3 to 5 as increasing the unit number with the interval of 10 units. Steps 3, 4, and
240 5 are iteratively conducted up to the unit number = 200, $u^{j=200}$. Consequently, the 20 good
241 performance models can be chosen for each unit number, as illustrated in Step 7. Among the
242 20 good performance models, the one better performance model with its accompanying unit
243 number can be selected in Step 8. Steps 2 to 8 are, so to speak, the selection procedure for the
244 better model and its accompanying unit number.

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

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

253

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

254

255

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

256

257

258 where, $\eta_{obs,i}$ is the observed surge level, $\eta_{fore,i}$ is the forecasted surge level, $\bar{\eta}_{obs}$ is the
 259 averaged observation, $\bar{\eta}_{fore}$ is the average value of the forecasted surge level, $\eta_{obs,max}$ is the
 260 observed highest surge level, $\eta_{obs,min}$ is the observed lowest surge level and $i = 1, 2, \dots, n$.

261

262 4. Results and discussion

263

264 4.1 Accuracy of the good performance model

265

266 Due to initially and randomly assigned weights and biases in the training phase, results
 267 obtained from individual ANN-SFMs show differences in accuracy, even though those are
 268 trained by the identical input and output data, and training algorithm. For example, the
 269 independent 957 ANN-SFMs, $N_{k=1,\dots,957}^{j=10} D_{i=1}$, were trained by the data set, $D_{i=1}$, with the 10
 270 units for the 24 h lead time. Then, their correlation coefficients (CCs) were calculated, as shown
 271 in Fig. 3 (a). It was found that CCs are scattered in the range of 0.5411 to 0.9474. When having
 272 a look at $N_{k=1,\dots,20}^{j=10} D_{i=1}$ in the horizontal axis, the highest CC value is found at 0.9046, as seen
 273 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
 274 good performance model among the 20 candidate models, $N_{k=1,\dots,20}^{j=10} D_{i=1}$. Meanwhile, the
 275 model, $N_{k=890}^{j=10} D_{i=1}$, reveals the highest CC of 0.9474 among the 957 models, $N_{k=1,\dots,957}^{j=10} D_{i=1}$.
 276 In other words, the model, $N_{k=890}^{j=10} D_{i=1}$, may be selected for a good performance model among
 277 the 957 ones. Consequently, it was found that the accuracy of ANN-SFM randomly varies with
 278 some ranges. However, the appropriate number ($N_{k=1,\dots,max}$) of ANN-SFM ($N_k^j D_i$), is not well

279 known for determining the good performance model: how many models are sufficient is
 280 uncertain for choosing the most accurate model. Therefore, in the present study, it is assumed
 281 that the maximum = 20 ($N_{k=1,\dots,20}$) is appropriate to decide the good performance model in
 282 order to reduce time-consuming processes. The determination of the optimal number of the
 283 model is beyond the scope of this study.

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

$$292 \quad HH = \sqrt{\frac{\sum_{i=1}^{i=n} (\eta_{fore,j} - \eta_{obs,j})^2}{\sum_{i=1}^{i=n} \eta_{fore,j} \eta_{obs,j}}} \quad (13)$$

294 which is introduced initially by Hanna and Heinold [7], to overcome the above difficulties.
 295

296
 297 4.2 Accuracy of the better performance model

298
 299 First, the good performance model among the twenty models trained with each unit
 300 number in the data set, $D_{i=1}$, consisting of the surge level (SS), the sea-level pressure (SLP),
 301 and the depression rate of the sea-level pressure (DSLPL), is selected in Steps 1 to 4 and
 302 evaluated by the statistical indices of CC, NRMSE and HH. Figure 5 shows the indices of the

303 selected good performance models in each unit number in the data set, $D_{i=1}$. The good
 304 performance ANN-SFMs trained for the lead times of 5 and 12 hours reveal the NRMSE values
 305 of approximately 5 to 7 % with the range of 1 % along the unit number, while those for the
 306 lead time of 24 hours shows the NRMSE values larger than 7 % with the bandwidth of 3 %,
 307 see Fig. 5(a). There is a similar tendency in the value of CC, see Fig. 5 (b): the CC values for
 308 the lead times of 5 and 12 hours vary with the range of 0.005, while those for the 24h lead time
 309 change with the range of 0.06. The change of HH is also apparent as the lead time is longer,
 310 see Fig. 5 (c). Hence, it is clear that the fluctuation width of the model error becomes substantial
 311 when the lead time becomes longer. Figure 5 shows no evidence of a tendency of a significant
 312 improvement in the forecasted surge level, as increasing the unit number in the data set, $D_{i=1}$.
 313 As discussed in the previous section, the values of CC and NRMSE are diverse for the 5 h lead
 314 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
 315 Fig 4 (a). Therefore, from now on the index of HH will be used to evaluate the ANN-SFM
 316 model to remove the uncertainty came from the difference of the evaluations between two
 317 indices of CC and NRMSE. As a result, the better performance ANN-SFMs is achieved based
 318 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
 319 associated unit number (j): 50, 130, and 70 for the 5, 12, and 24 h lead times, respectively.

320 Figures 6, 7, and 8 show the HH values of the good performance models evaluated in
 321 each unit number in each data set for the 5, 12, and 24 h lead times, respectively. For the data
 322 set, $D_{i=2}$, combining SS, SLP, DSLP, and WS, the ANN-SFMs for the 5 and 12 h lead times
 323 reveal the HH values in the range of approximately 0.17 to 0.23 against the unit number, while
 324 the ANN-SFMs for the 24 h lead time show the highest HH in the range of 0.27 to 0.33.
 325 Consequently, it can be found that a fluctuation of the HH values becomes larger, when the
 326 lead time is longer, as found in the data set $D_{i=1}$. The statistical indicators of HH show that
 327 accuracies of the ANN-SFMs for the 5, 12, and 24 h lead times are insignificantly improved,

328 as increasing the unit number. Nevertheless, the better performance models with the associated
329 unit numbers of 40, 190, and 110 are selected for each lead time: $N_{k=17}^{j=40} D_{i=2}$, $N_{k=20}^{j=190} D_{i=2}$,
330 and $N_{k=5}^{j=110} D_{i=2}$, respectively.

331 Next, the HH values of the good performance models selected in each unit number
332 trained by the data set, $D_{i=3}$, are plotted, when adding the wind direction (WD) to the data set,
333 $D_{i=2}$. Overall, a behavior of the HH values against the unit number in the data set, $D_{i=3}$, is
334 very similar to that in the data set, $D_{i=2}$. The values of HH for the 5 and 12 h lead times are
335 similar each other in the range of 0.15 to 0.22, while those for 24 h lead time are slightly larger
336 in the range of 0.25 to 0.32. The results show that the better performance models with the
337 associated unit number of 70, 40, and 170 were obtained for the 5, 12, and 24 h lead times:
338 $N_{k=3}^{j=70} D_{i=3}$, $N_{k=18}^{j=40} D_{i=3}$, and $N_{k=20}^{j=170} D_{i=3}$, respectively.

339 The results of the data set, $D_{i=4}$, which combines SS, SLP, DSLP, and the typhoon
340 position (both the longitude (LG) and latitude (LT)), show that the HH values of the good
341 performance models for all the lead times are in the range of 0.07 to 0.14. The HH values in
342 the data set, $D_{i=4}$, are significantly lower than those in the data sets, $D_{i=1,2,3}$. Especially, it
343 was found that the 24 h lead time forecast model trained by the data set, $D_{i=4}$, obtains the
344 significantly lower HH values in comparison with those acquired by the 24 h lead time forecast
345 models trained by the data sets, $D_{i=1,2,3}$. In addition, the ranges of the HH values in the data
346 set, $D_{i=4}$, are apparently narrower than those in the data sets, $D_{i=1,2,3}$. As a result, the use of
347 the unit numbers, $u^{j=170}$, $u^{j=120}$, and $u^{j=160}$, gives the better performance for all the lead
348 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.

349 Contrary to the results in the previous data sets, $D_{i=1,2,3,4}$, it is found in the data set,
350 $D_{i=5}$, that accuracies of the models tend to diverse and become lower when the good
351 performance models are trained by SS, SLP, DSLP, LG, LT, and WS in the data set, $D_{i=5}$. For

352 example, the HH values for the 24 h lead time seem to increase from 0.15 to 0.24 as increasing
 353 the unit number. Based on the HH values, the 24 h-forecast ANN-SFM, $N_{k=18}^{j=130} D_{i=5}$, could be
 354 selected as the better performance model among the 20 good performance models. For the 5
 355 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
 356 better-performed models, respectively.

357 In the data set, $D_{i=6}$, including SS, SLP, DSLP, LG, LT, WS and WD, the HH values
 358 tend to be larger as increasing the unit number, which is similar to the tendency of the data set,
 359 $D_{i=5}$. This tendency is much significant in the HH evaluated for the 24 h-forecast model. From
 360 the results of the data set, $D_{i=6}$, the use of the unit number, $u^{j=10}$, is the best rather than the
 361 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
 362 chosen for the better performance models for all the lead times, respectively.

363 When training a model with the data set, $D_{i=7}$, which combines SS, SLP, DSLP, LG,
 364 LT, and the central atmospheric pressure of typhoon (CAP), the HH value of good performance
 365 models shows a much similar change to that of the previous models trained by the data set,
 366 $D_{i=4}$, which consists of SS, SLP, DSLP, LG, and LT. As discussed in the data set, $D_{i=4}$, the HH
 367 values for all three lead times scatter below 0.12 along with the unit number. The scattering
 368 pattern of $D_{i=7}$ is very similar to that of the data set, $D_{i=4}$. On the other hand, a variation of
 369 HH in the data set, $D_{i=7}$, is significantly narrower and more stable than that in the data set,
 370 $D_{i=4}$. For instance, the HH value of the 24 h lead time in the data set, $D_{i=7}$, shows the smaller
 371 range of 0.0276 between 0.0784 to 0.106 in comparison with the corresponding HH value in
 372 the data set, $D_{i=4}$, which is in the range of 0.0644 between 0.0774 to 0.1418. As done in the
 373 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
 374 $N_{k=3}^{j=40} D_{i=7}$, with the unit numbers, $u^{j=100}$, $u^{j=90}$, and $u^{j=40}$, could be chosen for the 5, 12,
 375 and 24 h lead times, respectively.

376 When the ANN-SFM is trained by the data set, $D_{i=8}$, which combines SS, SLP, DSLP,
 377 LG, LT, CAP, and WS, the HH value in $D_{i=8}$ is generally larger than that in $D_{i=7}$. In this data
 378 set, the HH value has a slightly increasing tendency, as increasing the unit number. Among the
 379 twenty good performance models, the better performance ANN-SFMs, $N_{k=14}^{j=120}D_{i=8}$,
 380 $N_{k=8}^{j=80}D_{i=8}$, 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
 381 selected for the lead times of 5, 12, and 24 hours, respectively.

382 In the data set, $D_{i=9}$, consisting of SS, SLP, DSLP, LG, LT, CAP, WS and WD, the HH
 383 value shows a similar trend to that obtained from the data set, $D_{i=8}$: an accuracy becomes
 384 lower, as increasing the unit number. Among the twenty good performance models, the better
 385 performance models are found with the associated unit numbers of 70, 190, and 10 for the lead
 386 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.

387 When training the ANN-SFM with the data set, $D_{i=10}$ that combines SS, SLP, DSLP,
 388 LG, LT, CAP, and HWS, the HH value particularly for the 24 h-forecast ANN-SFM increases
 389 in the range of 0.13 to 0.20, as increasing the unit number. On the other hand, the HH values
 390 are relatively stable in the 5 and 12 h-forecast models, fluctuating in the range of 0.0574 to
 391 0.1018. Among the selected twenty good performance models, the better performance models
 392 with the associated unit numbers of 70, 160, and 30 were chosen for the lead times of 5, 12,
 393 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.

394 When using the data set, $D_{i=11}$, of SS, SLP, DSLP, LG, LT, CAP, HWS, and WS, the
 395 behavior of the HH values for all the forecasts is similar to that of the data set, $D_{i=10}$: it
 396 becomes in general larger, as increasing the unit number. In the case of the data set 11, $D_{i=11}$,
 397 the better performance models with the unit numbers of 30, 20, and 20 were selected for the
 398 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.

399 Finally, in the data set, $D_{i=12}$, when enlarging the unit number, the HH value
 400 significantly increases in all the good performance forecast models. The HH values apparently
 401 vary from 0.095 to 0.20. Similar to the models in the data set, $D_{i=11}$, the better performance
 402 models appear at the smallest/small unit number and then, the relevant unit numbers were 10,
 403 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
 404 $N_{k=12}^{j=30}D_{i=12}$, respectively.

405 Until now, we have assessed the performance of the 5, 12, and 24 h-forecast ANN-
 406 SFMs trained, validated, and tested by the data sets, $D_{i=1,\dots,12}$, which combine the
 407 meteorological, hydrodynamic and typhoon-characteristic parameters, to find out the
 408 appropriate set of the input parameter and the unit number in the hidden layer. Through the
 409 selection procedure, it was found that as increasing the unit number, the HH value for the 5 h-
 410 forecast ANN-AFM constantly fluctuates within the specific ranges in the data sets of
 411 $D_{i=1,2,3,4,7,8,10}$. On the other hand, the HH value becomes gradually larger in the data sets of
 412 $D_{i=5,6,9,11,12}$. For the 12 h-forecast ANN-AFM, the HH value is constant in the data sets of
 413 $D_{i=1,2,3,4,7,10}$ when the unit number increases, but it increases in the data sets of $D_{i=5,6,8,9,11,12}$.
 414 For the 24 h-forecast ANN-AFM, the HH value fluctuates with the constant range in the data
 415 sets of $D_{i=1,3,4,7}$, while it gradually increases in other the data sets. Also, it was found for the
 416 24 h-forecast model trained by the data sets of $D_{i=5,6,8,9,10,11,12}$ that the better performance
 417 ANN-AFM is obtained by using the smallest unit number as shown in other studies. It becomes
 418 apparent that the accuracy of the ANN-SFM may not always accompany with the increase of
 419 the unit number. As in consequence, the performance of the ANN-SFM shows the unique
 420 feature depending on not only the data set but also the unit number. Therefore, it can be said
 421 that the selection process proposed in the present study is able to determine the better
 422 performance model with the appropriate set of the input parameter and the unit number.

423

4.3 Accuracy of the best performance model

425

426 The best performance model accompanied by the most appropriate set of the input
427 parameters and the unit number can be determined among the twelve better performance
428 models previously chosen in the twelve data sets, as shown in Fig. 9. When the data sets,
429 $D_{i=1,2,3}$, which combine the parameters of SS, SLP, DSLP, WS, and WD, train ANN-SFMs,
430 their accuracies are generally lower than the others, for instance, trained by the data sets,
431 $D_{i=4,\dots,12}$. When training with the data set, $D_{i=10}$, which consists of SS, SLP, DSLP, LG, LT,
432 CAP, and HWS, the highest accuracies of the 5 and 12 h-forecast ANN-SFMs are obtained
433 with the most relevant unit numbers of 70 and 160, respectively (also see Fig. 10). For the 24
434 h-forecast ANN-AFM, the higher accuracy could be acquired when training with the data sets,
435 $D_{i=4}$ (the parameters: SS, SLP, DSLP, LG, and LT) and $D_{i=7}$ (SS, SLP, DSLP, LG, LT, and
436 CAP). Thus, the best performance 5 and 12 h-forecast ANN-SFMs ($N_{k=14}^{j=70} D_{i=10}$ and
437 $N_{k=6}^{j=160} D_{i=10}$) are chosen with the unit numbers, $u^{j=70}$ and $u^{j=160}$, that are trained by the
438 data set, $D_{i=10}$ (SS, SLP, DSLP, LG, LT, CAP, and HWS). The best performance 24 h-forecast
439 ANN-SFM, $N_{k=13}^{j=160} D_{i=4}$, is determined with the most appropriate combination of the unit
440 number, $u^{j=160}$ and the data set, $D_{i=4}$, consisting of SS, SLP, DSLP, LG, and LT (also see
441 Fig. 11).

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

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

457

458

459 4.4 Discussion

460

461 In the present study, the best performance ANN-SFM is chosen through the selection
 462 procedure to improve the accuracy of the storm surge forecast at the Sakai Minato on the Tottori
 463 coast, as focusing on the determination of the input parameter in the input layer and the unit
 464 number in the hidden one. As a result, the accuracy of the ANN-SFM is improved. The lowest
 465 HH values (HH_{best}) are 0.057, 0.065, and 0.1185 evaluated in the best performance 5, 12, and
 466 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
 467 hand, the highest HHs (HH_{worst}) are 0.420, 0.5398, and 0.8869 in the worst performance 5,
 468 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
 469 between the highest and lowest HHs are 0.363, 0.4748, and 0.7684 for the 5, 12, and 24 h-
 470 forecast models, respectively: the difference can be reduced up to 86.4 % to 138.6 %, which is
 471 calculated by:

472

$$473 \quad ((HH_{worst} - HH_{best})/HH_{worst}) \times 100 \quad (14)$$

474

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

479 Furthermore, the following question will remain from the physical perspectives: why
480 the data sets, $D_{i=10}$ (storm surge (SS), sea level pressure (SLP), depression of sea level
481 pressure (DSLPP), longitude (LG), latitude (LT), central atmospheric pressure (CAP), and
482 highest wind speed (HWS)) and $D_{i=4}$ (SS, SLP, DSLPP, LG, and LT) are suitable for the 5 and
483 12, and 24 h lead times, respectively. Based on the current results, the parameters, SS, SLP,
484 DSLPP, LG, LT, CAP, and HWS, significantly affected, and are most relevant to the accuracy of
485 the forecast models with the 5 and 12 h lead times. In addition, it can be said that the data sets,
486 $D_{i=4}$, (SS, SLP, DSLPP, LG, and LT) and, $D_{i=7}$, (SS, SLP, DSLPP, LG, LT, and CAP) are also
487 impactive for the 5 and 12 h-forecast models. In other words, when the storm surge is shortly
488 forecasted with the 5 or 12 h lead times at Sakai Minato, the storm surge level is absolutely
489 governed by the typhoon characteristics that are the longitude, the latitude, the central
490 atmospheric pressure (CAP), and the highest wind speed (HWS) near the typhoon center.
491 However, when the storm surge is predicted with the 24 h lead time at Sakai Minato, the
492 accuracy of the 24 h-forecast model is apparently influenced by the parameters of the local
493 atmospheric components observed near Sakai Minato, the typhoon position, and its central
494 pressure.

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

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

511 The best performed-storm surge forecast models show the highly promising accuracy,
512 as shown in Figs. 18 and 19. In the study, we have dealt with the data 460 hours long collected
513 from three typhoon events in order to propose the novel selection procedure. In the meanwhile,
514 one might be curious about the actual forecast capacity of the best performing model. The
515 present study has been aiming at the introduction to the novel method, which one could
516 determine the best performing model and its relevant unit number among the potentials, the
517 capacity in the real practice has been beyond the present study so far. For this reason, the current
518 best performing model chosen in the study could not be sustained in the practical application.
519 One thing was not considered was the training data size that is relatively short to directly adopt
520 the forecast model chosen here to the field site. Therefore, for guaranteeing the feasibility of
521 the storm surge forecast model, which is trained and then decided, against the independent
522 typhoon event or future one, the training with the massive data size collected from the typhoons
523 and accompanied input components is essential. In the practical application, if the storm surge

524 and relevant atmospheric components measured are sufficiently long, the feasibility could
525 promise. However, the real site might not acquire such the data length for ensuring the viability.
526 One of its alternatives to make a sufficiently lengthy database is the reproduction of the
527 historical typhoon and storm surge event using an atmospheric general circulation model and
528 ocean circulation one. Inclusion of the future event, extracted from, for instance, a Database
529 for Policy Decision making for Future climate change (d4PDF) (Mizuta et al., [28]), in the
530 database might be one of alternative ways.

531

532

533 5. Conclusions

534

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

549 the central atmospheric pressure of the typhoon, and the highest wind speed near the typhoon
550 center). Hence, the twelve better performance ANN-SFMs are selected in each data set. Among
551 the twelve better performance models, the best performance ANN-SFM is finally determined
552 with the associated set of the unit number and the input parameters.

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

570 As discussed in Section 4.4, the characteristics of the best storm surge forecast model
571 systematically determined in this study may not be universal in terms of the appropriate set of
572 the unit number and the input parameters. Otherwise, the systematic selection procedure

573 proposed in the present study is applicable to develop the ANN-based storm surge forecast
574 models on a coast.

575

576 Acknowledgements

577 This work was partially conducted under JSPS KAKENHI, Japan.

578

579 References

580 [1] Deo, M. C. and G. Chaudhari (1998) Tide prediction using neural networks, *Computer-*
581 *Aided Civil and Infrastructure Engineering*, 13, p113-120.

582 [2] Deo, M. C. and C. Sridhar Naidu (1999) Real time wave forecasting using neural networks,
583 *Ocean Engineering*, 26, 191-203.

584 [3] Deo, M. C., A. Jha, A. S. Chaphekar, K. Ravikant (2001) Neural networks for wave
585 forecasting, *Ocean Engineering*, 28, 889-898.

586 [4] Dreyfus, G. (2002) *Neural Networks: Methodology and Applications*, Springer, p497.

587 [5] Flather, R. A. (1994) A Storm Surge Prediction Model for the Northern Bay of Bengal with
588 Application to the Cyclone Disaster in April 1991, *J. Phys. Oceanogr.*, 224, 172–190.

589 [6] Hiyajo, H., Okubo, S., Takasa, S., Kobashigawa, Y., Nishimura, F, Toomine, T., Daimon,
590 H., Itagaki, S., Fukuda, M., Sakaji, T., Taguchi, H., Egami, H., Suzuki, H. and Nozaki, F.
591 (2011) Digitizing the Historical Sea Level Data and Re-Analysis of Storm Surges with the
592 Data, *Meteorology Bulletin by Japan Meteorological Agency*, 78, S1-S32 (In Japanese)
593 (<http://www.jma.go.jp/jma/kishou/books/sokkou-kaiyou/78/vol78s001.pdf>)

594 [7] Hanna, S. and Heinold, D. (1985) Development and application of a simple method for
595 evaluating air quality. In: API Pub. No. 4409, Washington, DC, Washington, USA.

- 596 [8] Jelesnianski, C. P., J. Chen, and W. A. Shaffer (1992) SLOSH - Sea, Lake, and Overland
597 Surges from Hurricanes. U.S. Department of Commerce, NOAA Tech. Report NWS 48,
598 71pp.
- 599 [9] Jia, G., Taflanidis, A.A., Nadal-Caraballo, N.C., Melby J.A., Kennedy A.B., Smith J.M.
600 (2016) Surrogate modeling for peak or time-dependent storm surge prediction over an
601 extended coastal region using an existing database of synthetic storms, *Nat Hazards*, 81,
602 909. doi:10.1007/s11069-015-2111-1
- 603 [10] Kim, S.Y., Yasuda, T., Mase, H., (2008) Numerical analysis of effects of tidal variations
604 on 1065 storm surges and waves. *Appl. Ocean Res.* 30, 311–322.
- 605 [11] Kim, S., Matsumi, Y., Yasuda, T., Mase, H. (2014) Storm surges along the Tottori coasts
606 following a typhoon, *Ocean Engineering*, 91, 133-145.
- 607 [12] Kim, S., Y., Matsumi, S., Pan, H., Mase (2016) A real-time forecast model using artificial
608 neural network for after- runner storm surges on the Tottori coast, Japan, *Ocean*
609 *Engineering* 122, 44–53.
- 610 [13] Kim, S., Melby, J.A., Nadal-Caraballo, N.C., Ratcliff j. (2015) A time-dependent
611 surrogate model for storm surge prediction based on an artificial neural network using
612 high-fidelity synthetic hurricane modeling, *Nat Hazards*, 76, 565. doi:10.1007/s11069-
613 014-1508-6
- 614 [14] Lee, T. L., C. P. Tsai, D. S. Jeng and R. J. Shieh (2002) Neural network for the prediction
615 and supplement of tidal record in Taichung Harbor, Taiwan, *Advances in engineering*
616 *software*, 33, 329-338.
- 617 [15] Lee, T. L. (2004) Back-propagation neural network for long-term tidal predictions, *Ocean*
618 *Engineering*, 31, 225-238.

- 619 [16] Lee, T. L. (2006) Neural network prediction of a storm surge, *Ocean Engineering*, 33,
620 483-494.
- 621 [17] Lee, T. L. (2008) Back-propagation neural network for the prediction of the short-term
622 storm surge in Taichung harbor, Taiwan, *Eng. Appl. Of Artificial Intelligenece*, 21, 63-72.
- 623 [18] Lee, T. L. (2009) Predictions of typhoon storm surge in Taiwan using artificial neural
624 networks, *Advances in Engineering Software*, 40, 1200-1206.
- 625 [19] Luettich, R.A. and J.J. Westerink (1991) A solution for the vertical variation of stress,
626 rather than velocity, in a three-dimensional circulation model, *International Journal for*
627 *Numerical Methods in Fluids*, 12:911-928
- 628 [20] Makarynskyy, O., D. Makarynska, M. Kuhn, W.E. Featherstone (2004) Predicting sea
629 level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia,
630 *Estuarine, Coastal and Shelf Science*, 61, 351–360.
- 631 [21] Makarynska, D., O. Makarynskyy (2008) Predicting sea-level variations at the Cocos
632 (Keeling) Islands with artificial neural networks, *Computers & Geosciences*, 34, 1910-
633 1917.
- 634 [22] Marzenna S. (2003) Forecast storm surge by means of artificial neural network, *Journal*
635 *of Sea Research*, 49, 317-322.
- 636 [23] Mase, H., Sakamoto, M. and Sakai, T. (1995) Neural network for stability analysis of
637 rubble-mound breakwaters, *Jour. Waterway, Port, Coastal, and Ocean Eng.*, ASCE,
638 Vol.121, No.6, pp.294-299
- 639 [24] Mase, H. and Kitano, T. (1999) Prediction model for occurrence of impact wave force,
640 *Ocean Eng.*, Vol.26, No.10, pp.949-961.
- 641 [25] Mase, H., Reis, M.T., Nagahashi, S., Saitoh, T., T.S. Hedges, T.S. (2007) Effects of zero-
642 overtopping data in artificial neural network predictions, *Proc. 5th Coastal Structures Int.*
643 *Conf.*, pp.1542-1551.

- 644 [26] Mase, H., T. Yasuda, N. Mori (2011) Real-time prediction of tsunami magnitudes in Osaka
645 Bay, Japan, using an artificial neural network, *J. Waterway, port, coastal, and ocean eng.*,
646 DOI: 10.1061/(ASCE)WW.1943-5460.0000092
- 647 [27] Mentaschi, L., Besio, G., Cassola, F., Mazzino, A. (2013) Problems in RMSE-based wave
648 model validations, *Ocean Modelling*, 72, 53-58.
- 649 [28] Mizuta, R., A. Murata, M. Ishii, H. Shiogama, K. Hibino, N. Mori, O. Arakawa, Y. Imada,
650 K. Yoshida, T. Aoyagi, H. Kawase, M. Mori, Y. Okada, T. Shimura, T. Nagatomo, M.
651 Ikeda, H. Endo, M. Nosaka, M. Arai, C. Takahashi, K. Tanaka, T. Takemi, Y. Tachikawa,
652 K. Temur, Y. Kamae, M. Watanabe, H. Sasaki, A. Kitoh, I. Takayabu, E. Nakakita, and
653 M. Kimoto, 2016: Over 5000 years of ensemble future climate simulations by 60 km global
654 and 20 km regional atmospheric models. *Bull. Amer. Meteor. Soc.* doi:10.1175/BAMS-
655 D-16-0099.1
- 656
- 657 [29] Peres, D. J., C. Iuppa, L. Cavallaro, A. Cancelliere, E. Foti (2015) Significant wave height
658 record extension by neural networks and reanalysis wind data, *Ocean Modelling*, 94, 128-
659 140.
- 660 [30] Tagliaferri, F., I.M. Viola, R.G.J. Flay (2015) Wind direction forecasting with artificial
661 neural networks and support vector machines, *Ocean Eng.*, 97, 65–73.
- 662 [31] Tseng, C. M., C. D. Jan, J. S. Wang and C. M. Wang (2007) Application of artificial neural
663 networks in typhoon surge forecasting, *Ocean Eng.*, 34, 1757-1768.
- 664 [32] Wang, S. H., K., Muhammad, Y. Lv, Y. Sui, L. Han, and Y. D., Zhang (2018)
665 Identification of alcoholism based on wavelet renyi entropy and three-segment encoded
666 jaya algorithm, , *Complexity*, 2018, vol. 2018, Article ID: 3198184

667 [33] Zhang, Y., P., Phillips, S., Wang, G., Ji, J., Yang and J., Wu (2016) Fruit Classification
668 by Biogeography-based Optimization and Feedforward Neural Network, Expert
669 Systems, 2016, 33(3): 239-253
670

671 Captions

672

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

675

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

680

681

682 Fig. 1 The Tottori coast with typhoon tracks. (a) Typhoons with symbols (☪ is typhoon and
683 ○ is cyclone or tropical depression); (b) the stations for the meteorological and

684 hydrodynamic parameters (●: the meteorological station and ✦: the hydrodynamic
685 station)

686

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

689

690 Fig. 3 Correlation coefficients of the ANN-SFMs with the 24 h lead time after testing using the
691 data set 1: (a) is the correlation coefficients obtained by the 957 ANN-SFMs, $N_{k=1,\dots,957}^{u^{j=10}} D_{i=1}$.

692 (b) same but the 1st ANN-SFM to 20th one ($= N_{k=1,\dots,20}^{u^{j=10}} D_{i=1}$).

693

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

696 $N_{k=1,\dots,20}^{u^{j=10}} D_{i=1}$) in each unit number after training, verifying and testing them by the data set,
697 $D_{i=1}$, for the 5 h lead time.

698

699 Fig. 5 Normalized root mean square errors (NRMSEs, %) in (a), Correlation coefficients (CCs)
700 in (b) and statistical indicators (HHs) in (c) of the good performance ANN-SFMs in each unit
701 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
702 hours.

703

704 Fig. 6 Statistical indicators (HHs) of the good performance ANN-SFMs in each unit number
705 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.

706

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

708

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

710

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

712

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

714

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

716

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

718

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

720

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

722

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

724

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

726

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

729

730 Fig. 18 Comparisons of observation and forecasts from the 5 and 12 h lead time ANN-SFMs

731 ($N_{k=14}^{j=70} D_{i=10}$ and $N_{k=6}^{j=160} D_{i=10}$) after training, validating and testing with the relevant set of

732 the 70 and 160 unit numbers and the input parameters of SS, SLP, DSLP, LG, LT, CAP and

733 HWS.

734

735 Fig. 19 Comparisons of observation and forecasts from the 24 h lead time ANN-SFMs

736 ($N_{k=13}^{j=160} D_{i=4}$) after training, validating and testing with the relevant set of the 160 unit

737 number and the input parameters of SS, SLP, DSLP, LG and LT.

738