공학박사 학위논문

Generation Mechanism and Machine-Learning Forecasting Model of Sudden High Waves in the East Sea

동해 돌연고파의 발생 메커니즘과 기계학습 예측모델

2017년 08월

서울대학교 대학원

건설환경공학부

오 지 희

ABSTRACT OF DISSERTATION

Generation Mechanism and Machine-Learning Forecasting Model of Sudden High Waves in the East Sea

Jihee Oh

Department of Civil and Environmental Engineering

The Graduate School

Seoul National University

Exceptional high waves have occurred repeatedly in the East Sea of Korea. These disastrous waves claimed the losses of life more than 50 people during the ten years between 2005 and 2015 in the east coast of Korea. Several researchers have examined the generation mechanism and characteristics of sudden high waves. However, the definition of the high waves is still vague and insufficient to explain the characteristics of sudden high waves. Also, occurrence of sudden high waves is only roughly forecasted in the daily weather forecast. In this study sudden high waves were defined using a new intensity parameter and the generation mechanism

of the sudden high waves was investigated. Next, significant wave height and period were forecasted in the East Sea of Korea using machine-learning. Finally, sudden high waves were forecasted using the intensity parameter proposed and the forecasted significant waves in the East Sea of Korea.

In this study, the index of sudden high waves was suggested as $\Delta(H^2L)/\Delta t$ and it was calculated using wave data measured in Gangneung and Wangdolcho in 2005–2012. The criteria of sudden high waves was set 80 m^3/hr, which corresponds to the top 20% of cumulative percentage of $\Delta(H^2L)/\Delta t$.

Next, to find the generation mechanism of sudden high waves, the evolution of spatial patterns of wind velocity and sea level pressures was presented during the sudden high wave events by CSEOF analysis and regression analysis. The wave data in Gangneung and Wangdolcho were used and the meteorological data were reanalysis data of National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR). There are two peaks in the modes of all CSLV considered the physical process of sudden high waves. The patterns were categorized two groups. The first pattern was that the first peak was generated by low pressure moving to the north east part of the East Sea and easterly wind blowing for 1 day, whereas the second peak was caused by strong wind. The second pattern was that the first peak was affected by the wind speed in the east coast and the second peak was influenced by wind in the offshore area.

To forecast significant waves in multiple locations simultaneously, an EOFWNN model was developed by combining the EOF analysis and wavelet analysis with the neural network. The wave data used in this research were measured at eight wave observation stations in the East Sea and the meteorological

data were the NCEP/NCAR reanalysis data. The results of the EOFWNN model for significant wave height were compared with those of a wavelet and neural network hybrid (WNN) model in Gangneung, Sakata and Aomori for several lead times. The EOFWNN model is better than the WNN model in that the former shows higher accuracy for longer lead times regardless of the wavelet decomposition level. Significant wave period series were also forecasted using the EOFWNN model. The results of significant wave period also show quite high accuracy. Also, the proposed model was employed to the numerical wave modeling data in the entire area of the East Sea. The results also show relatively high accuracy for one and three hour lead times.

Using the proposed intensity parameter of sudden high waves and the forecasted significant waves by the EOFWNN model, sudden high wave was detected and forecasted. From the forecasted wave data at 24 hour lead time, $\Delta\left(H^2 L\right)/\Delta t$ was calculated in Gangneung and Sakata. Although there is a slight deviation between the results of observed and forecasted wave data, sudden high wave was detected clearly.

Keywords: Artificial neural network; Empirical orthogonal function; Significant wave; Sudden high wave; Wave forecasting; Wavelet.

Student number: 2012-30953

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Comprehensive Explanations

- Unless otherwise stated, all the wave parameters are those of significant waves (e.g. H = significant wave height).

Latin Uppercase

A: approximation

$B_n(x)$: eigenfunction

$B_n(x,t)$, $C_n(\mathrm{x},t), D_n(\mathrm{x},t)$: cyclo-stationary loading vector (CSLV)

D: detail

H: wave height (m)

I_a: index of agreement

L : wave length (m)

NRMSE: normalized root mean square error

R: correlation coefficient

RMSE: root mean square error

T : wave period (s)

$T(x,t)$, $P(x,t)$: spatio-temporal series

$T_n(t)$, $P_n(t)$: principal component (PC) time series

Latin Lowercase

a: scale factor

b: shift factor or ANN bias

d: nested period

slp: sea level pressure series

$t,\ t'$: time

w: ANN weight

wnd: wind velocity series

$x,\ x'$: space

Greek Lowercase

$\alpha_m^{\{n\}}$: regression coefficient

$\varepsilon^{\{n\}}(t)$: regression error time series

λ_n: eigenvalue

ϕ: scaling function

ψ_n: wavelet function

CHAPTER 1. INTRODUCTION

1.1 Background

Recently, exceptional high waves have caused many casualties and serious property damage in the East Sea of Korea. On 21 October 2006, the significant wave height reached 9.69 m and its peak period was 12.8 s near Sokcho Harbor, which was the maximum wave height ever observed on the east coast of Korea (Jeong, Oh, and Lee 2007). Meanwhile, similar events have been repeatedly reported on the west coast of Japan. Many researchers examined the most remarkable event occurred on 24 February 2008 at Toyama Bay, which was highlighted by the significant wave height of 9.92 m (Mase et al. 2008, Lee et al. 2010). These disastrous waves claimed the losses of life more than 50 people during the ten years between 2005 and 2015 in the east coast of Korea because they suddenly occurred under the relatively mild weather in the winter season.

Although storm waves about two times higher than 50 year return period of deep water design waves have been observed in the east coast, sudden high waves have not been considered until now when estimating design waves of coastal structures. The high waves occur repeatedly every year and the risk of sudden high waves will increase due to anomaly climate such as global warming, so it is necessary to take into account such high waves when estimating design waves (Jeong, Oh, and Lee 2007). Since figuring out sudden high waves is the key factor not only in coastal damage and disaster but also in the design of coastal structures, it is critical to forecast sudden high waves accurately.

Sudden high waves in the East Sea have been recognized and studied since

long ago (Kitaide 1952, Isozaki 1971, Isozaki and Yoshio 1972). Recently, as casualties and property damage caused by sudden high waves have been increased, researches have been made actively (Jeong 2009, Lee et al. 2010, Oh et al. 2010, Kashima and Hirayama 2011, Ahn et al. 2013, Oh and Jeong 2013, Lee et al. 2014). The studies are divided into two major parts: analysis of the characteristics and mechanism of sudden high waves and forecasting of sudden high waves using numerical method.

Definition of sudden high waves

Kim and Lee (2008b) analyzed the observed swells in Wangdolcho on 22 – 28 February 2008 using wavelet method. They showed that the peak frequency moves to the lower frequency and the wave energy increases dramatically during the event. Oh et al. (2010) analyzed several events and described the causes of high waves for each case. They defined such high wave as large-height swell-like wave, which often has a larger height than general swell (H > 3 m) and a relatively long period (T > 9 s). Since their research, many researchers have followed their definition of large-height swell-like waves. Also, the many have classified such waves according to only its height and period. However, the definition and criteria of the high waves is vague and insufficient because the high waves have the characteristics of both swell and wind wave and the definition does not reflect the rapid increase of waves in a few hours. In this study, sudden high wave was used as a new terminology to include the characteristics of such high waves.

Generation mechanism of sudden high waves

The mechanism of sudden high waves have been analyzed not only in the field of coastal engineering but also meteorology. According to Kitaide (1952), the waves called “Yorimawari wave” in Japan are generated due to the strong north

and northeasterly winds formed by a quasi-stationary developed atmospheric low-pressure area. Isozaki (1971) and Isozaki and Yoshio (1972) examined the characteristics of abnormal high waves and described the predictability based on meteorological observations of past events. Joung et al. (1984) studied the generation process of a low-pressure system developed on 2 January 1981 by using the results from an adiabatic inviscid quasi-geostrophic model. Jeong, Oh, and Lee (2007) analyzed the characteristics of the high waves observed at 5 stations in the East Sea on 23–24 October 2006 with wind data. According to them, sudden high waves occurs due to superposition of wind waves and swell generated by strong East Sea twister. Oh and Jeong (2013) analyzed the characteristics of sudden high waves and meteorological conditions during the events using the observed wave data at multiple stations along the east coast of Korea. According to them, important is not only the pressure drop during the movement of low pressure system but also other factors such as moving trajectory and staying time of the low pressure system together with arrangement of neighboring atmospheric pressure fields. They also found that the spectral density of the high waves increases two times on 23 February 2008. It was found that the characteristics of waves would be predominantly governed by long-traveled swell in the former case but in contrast, it would be a result of local wind sea development in the latter case. Although many researches related to generation mechanism of sudden high waves have been studied, they are only for a few cases based on casualties or damage of properties. It is necessary to analyze the overall generation mechanism of sudden high waves using the wave data measured long period.

Forecasting of sudden high waves

The forecasting of sudden high waves in the East Sea, as another major part,

have been conducted (Ahn et al. 2013, Hatada and Yamaguchi 1998, Kim and Lee 2008a, Kim et al. 2011, Lee et al. 2008, Lee et al. 2014, Mase et al. 2008). Even though many researches for forecasting of sudden high waves have been done, they have been studied only using numerical methods. There is no research for forecasting of sudden high waves using statistical method such as machine-learning until now.

Machine-learning forecasting of significant wave heights

Despite of considerable advances in computational techniques, the solutions obtained by numerically solving the equations of wave growth are neither exact nor uniformly applicable at all sites and at all times due to the complexity and uncertainty of the wave generation phenomenon (Deo et al. 2001). Prediction of waves is basically uncertain and random process and hence difficult to model by using deterministic equations. Therefore, many researchers have established the forecasting of waves using stochastic models such as auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA) or artificial neural networks (ANN).

Deo et al. (2001) explored the possibility of employing neural networks for weakly mean wave forecasting based on wind speeds. They mentioned that neural network modeling is proper to predict waves since it is primarily aimed at recognition of a random pattern in a given set of input values and does not require knowledge of the physical process as a precondition. However, their results are not very satisfactory, possibly due to the uncertainties of the wind-wave relationship. After their research, many researchers have investigated wave forecasting using artificial neural network by training the observed wave records directly (Tsai, Lin, and Shen 2002, Makarynskyy et al. 2005, Londhe and Panchang 2006). Their

results showed the short-term forecasts (3 and 6 hours) of the wave parameters are more accurate than longer-term forecasts. Meanwhile, some researchers have examined the effects of other meteorological factors on wave height forecasting using ANN (Günaydın 2008, Zamani et al. 2008). However, there is no significant improvement of the forecasting results.

Even though ANN has flexibility, it may not be able to cope with non-stationary data without any preprocessing of the input and output data (Cannas et al. 2006). In the recent years, hybridization of ANN with other techniques has been used in wave height forecasting to provide effective modeling. Özger (2010) proposed the combination of wavelet and fuzzy logic approaches to forecast wave height up to 48 hour lead time. Deka and Prahlada (2012), Prahlada and Deka (2015) used wavelet neural network (WNN) for wave height forecasting up to 48 hour lead time. Dixit and Londhe (2016) also used neuro wavelet technique for extreme wave heights forecasting up to 36 hour lead time. Shahabi, Khanjani, and Kermani (2016) developed genetic programming based wavelet transform to forecast significant wave height up to 48 hour lead time. Hybrid model results showed better prediction performance than single ANN model. Although their results indicates good predictions at lower lead times but slight deviation is observed at higher lead times. Also there are limitations that their models cannot interpret the relationship between spatially distributed meteorological variables and waves and cannot forecast spatially distributed wave height at once. In other words, in their researches modeling is performed for forecasting wave height at different locations separately.

1.2 Research Objectives

The ultimate goal of this study is to forecast sudden high waves in the East Sea of Korea based on the meteorological data and wave data using machine-learning. To forecast the waves, the first objective is to define a new intensity parameter of sudden high waves. Second, the generation mechanism of the waves are investigated by analyzing the relation between waves and meteorological variables. Third, significant wave height and period are forecasted in the East Sea of Korea using machine-learning. Finally, sudden high waves are detected and forecasted from the proposed intensity parameter and the forecasted significant waves in the East Sea of Korea. Fig. 1.1 shows the schematic diagram of this study.

This thesis is organized in the following order. Chapter 2 explains the statistical methods to analyze and forecast sudden high waves. Chapter 3 consists of three parts; (1) description of data (2) definition of sudden high waves and (3) generation mechanism of sudden high waves. Chapter 4 consists of four parts; (1) description of data (2) forecasting of the significant wave height and period using the observed data along the coast of the East Sea, (3) forecasting of the significant wave height and period using the numerical modeling data in the entire area of the East Sea and (4) detection and forecast of sudden high waves. In Chapter 5, the conclusions are given and future work is discussed.

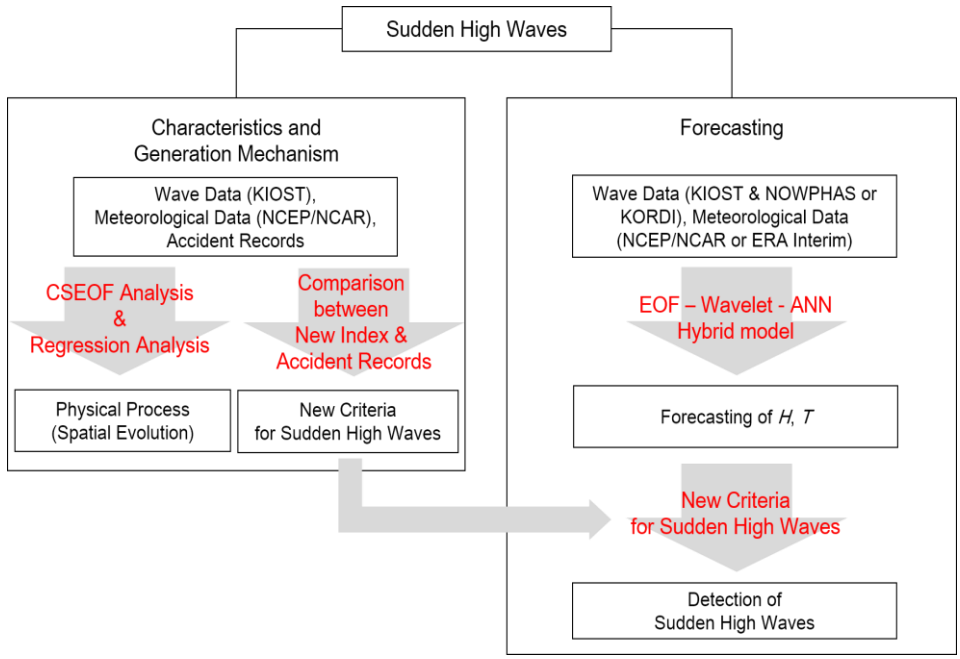


Fig. 1.1 Schematic diagram of the study

CHAPTER 2. THEORETICAL STUDY

2.1 Analysis methods of mechanism of sudden high waves

Some geophysical and climatic variables have periodically time-dependent covariance statistics or non-stationarity. Stationarity assumption is often not appropriate for such geophysical and climate variables even after removing the diurnal cycle or the seasonal cycle (Kim, Hamlington, and Na 2015). A proper recognition of the time-dependent response characteristics is vital in accurately extracting physically meaningful modes and their space-time evolutions from data. Cyclostationary empirical orthogonal function (CSEOF) analysis is an alternative to regular EOF analysis or other eigen-analysis techniques based on the stationarity assumption to extract physical modes. In this study, the CSEOF analysis was used to examine the physical processes of sudden high waves. When a physical process undergoes a stochastic variation for some reasons, two physical variables describing the process evolve in the same fashion (Kim, Hamlington, and Na 2015). If the system is linear, their relationship may be linear. However, due to the complexity of the system, it is difficult to find the relationship between the physical variables directly. Therefore, a proper distinction from stochastic undulation and explanation of time-dependent physical response are vital in accurately determining physically consistent or teleconnection response in different variables. This can be achieved by the regression analysis in CSEOF space (Kim and Chung 2001, Seo and Kim 2003, Hamlington et al. 2011, Kim, Hamlington, and Na 2015).

2.1.1 CSEOF analysis

Kim, North, and Huang (1996) and Kim and North (1997) introduced the concept of CSEOF analysis to capture the time-varying spatial patterns and longer-time-scale fluctuations in geophysical data. The difference between EOF analysis and CSEOF analysis is the ability of CSEOF analysis to extract spatial patterns varying in time and space. This is possible because the CSEOF spatial patterns are time dependent, whereas EOF spatial patterns are only varying spatially (Strassburg et al. 2014).

In a CSEOF analysis, space-time data are decomposed into:

$$T(x,t) = \sum_n B_n(x,t)T_n(t), \tag{2.1}$$

where $B_n(x,t) = B_n(x,t+d)$ are the CSEOF loading vectors (CSLV), which are multiple (d) spatial patterns and repeat themselves in time and $T_n(t)$ are the corresponding principal component (PC) time series (Kim, North, and Huang 1996, Kim and North 1997). In other words, $B_n(x,t)$ are time-dependent physics and $T_n(t)$ are the stochastic undulation of the physical processes. The CSLVs and the corresponding PC time series are obtained by solving:

$$C\left(x,t;x^{'},t^{'}\right)B_n(x^{'},t^{'}) = \lambda_n B_n(x,t), \tag{2.2}$$

with x and $x^{'}$ representing different points in space and time, respectively. The

space-time covariance function is periodic in time with the nested period *d*.

$$C\left(x,t;x^{'},t^{'}\right)=C\left(x,t+d;x^{'},t^{'}+d\right). \tag{2.3}$$

Since the covariance matrix cannot be written as a square matrix, Eq. (2.2) cannot be solved in the same manner as EOF analysis. Instead, it can be solved by taking Fourier transform twice with respect to t and $t^{'}$, making use of the assumption that the covariance matrix is periodic. The PC time series in Fourier space are easy to obtain, and then both the CSLVs and PC time series are transformed back to physical space (Kim, North, and Huang 1996, Hamlington et al. 2011).

While the assumption of periodic statistics may be reasonable for many geophysical variables, it is difficult to prove the periodicity of statistics and identify the period. The periodicity is called "nested period". The nested period is often determined based on a priori physical understanding of the physical process to be investigated (Kim, Hamlington, and Na 2015). However, sometimes the period of a physical process is not obvious mainly because of the lack of understanding of the underlying physical process. It is also difficult to find the nested period because multiple physical processes have different periods. Let us consider a dataset consisting of several physical processes

$$B_n(x,t)=B_n(x,t+d_n) \tag{2.4}$$

where d_n is the period of a physical process $B_n(x,t)$. If it is assumed that the PC time series, $T_n(t)$, are stationary, then the first two moment statistics are:

$$\begin{aligned}\mu(x,t) &= \langle T(x,t)\rangle = \sum_n B_n(x,t)\langle T_n(t)\rangle \\ &= \sum_n B_n(x,t+d)\langle T_n(t+d)\rangle = \langle T(x,t+d)\rangle \\ &= \mu(x,t+d)\end{aligned} \tag{2.5}$$

$$\begin{aligned}C(x,t;x',t') &= \langle T(x,t)T(x',t')\rangle = \sum_n\sum_m B_n(x,t)B_n(x',t')\langle T_n(t)T_m(t')\rangle \\ &= \sum_n\sum_m B_n(x,t+d)B_n(x',t'+d)\langle T_n(t+d)T_m(t'+d)\rangle \\ &= \langle T(x,t+d)T(x',t'+d)\rangle = C(x,t+d;x',t'+d)\end{aligned} \tag{2.6}$$

if *d* is given as the least common multiple of $\{d_n\}$, i.e., $d = LCM(d_n)$. Under the assumption of stationarity of PC time series, the first two moment statistics are functions of time lag $t - t'$. Thus, the period of the first two moment statistics of a given dataset is the least common multiple of all physical periods in the dataset. The consequence of the nested period being the least common multiple of all physical periods is that physical processes with period less than *d* are shown to repeat in CSLVs. Although the nested period can be set to be an integral multiple of least common multiple of all physical periods, the minimum period should be used in order to minimize the contamination of covariance statistics by sampling errors (Kim, Hamlington, and Na 2015).

2.1.2 Regression analysis

First, the CSEOF analysis is conducted on a target variable (wave height and wave period) and predictor variable (meteorological data):

- Target variable:

$$T(x,t) = \sum_n B_n(x,t)T_n(t) \tag{2.7 a}$$

- Predictor variable:

$$P(x,t) = \sum_m C_m(x,t)P_m(t) \tag{2.7 b}$$

where $B_n(x,t)$ and $C_m(x,t)$ are the CSLVs of the target and predictor variables, respectively, and $T_n(t)$ and $P_m(t)$ are the PC time series of the target and predictor time series, respectively. Then, conducting regression analysis between the two PC time series gives:

$$T_n(t) = \sum_{m=1}^{M} a_m^{(n)} P_m(t) + \varepsilon^{(n)}(t), \quad n = 1,2,\ldots \tag{2.8}$$

where $\left\{a_m^{(n)}\right\}$ are the regression coefficients, $\varepsilon^{(n)}(t)$ is the regression error time series, and *M* is the number of predictor time series modes used in the regression. The regression coefficients are determined so that the variance of regression error is minimized. Using the regression coefficient, the evolution of the predictor variable, which is physically consistent with the target evolution, is obtained:

$$D_n(x,t) = \sum_{m=1}^{M} a_m^{(n)} C_m(r,t) \tag{2.9}$$

where $C_m(x,t)$ is the CSLV of predictor variable and $D_n(x,t)$ is the consistent CSLV with the target variable. In this way, the evolution of any variables can be achieved to be physically consistent with that of the target variable. The consistent patterns of two physical variables may not generally have the same physical

response characteristics. However, the physical relationship between physically consistent patterns of two or more physical variables should be dictated by a governing equation describing the particular physical process they represent. It is the stochastic component of undulation that should be identical in the evolution of two physical variables originating from the same physical process (Kim, Hamlington, and Na 2015).

2.2 Wave forecast methods

Prediction of waves is basically uncertain and random process and hence difficult to model by using deterministic equations. Artificial neural network (ANN) is suitable for partially understood underlying physical processes such as wind-wave relationship. In spite of suitable flexibility of ANN, it may not be able to cope with non-stationary data if pre-processing of the input and output data is not performed (Cannas et al. 2006). Also it is difficult to interpret the relationship between spatially distributed meteorological variables and waves. Principal component analysis (PCA), also called the empirical orthogonal function (EOF) analysis is a useful tool to interpret physical processes in the data. The assumption in the EOF analysis is the stationarity of the data. It means that the covariance function of the data does not depend on time. As I mentioned in previous section, the stationarity assumption is often not appropriate for such geophysical and climate variables. In previous section, CSEOF analysis was suggested as an alternative to EOF analysis to extract physical modes. However, to forecast wave series, spatial components and temporal components should be separated completely. Thus, in this section,

EOF analysis is introduced to interpret the relationship between waves and meteorological data. Also, EOF analysis make the proposed model to forecast wave data at multi-stations simultaneously. To overcome stationarity assumption of the EOF analysis, wavelet analysis is combined. Wavelet analysis can handle non-stationary and transient signals as well as fractal-type structure (Murguia and Campos-Cantón 2006). In this study, a hybrid empirical orthogonal function analysis (EOF)–wavelet analysis–ANN (EOFWNN) model is introduced and employed to forecast significant wave height (or period). The decomposed time series are used as inputs to ANN which can handle non-stationarity and non-linearity efficiently.

2.2.1 EOF analysis

EOF analysis, also known as the principal component analysis, or singular value decomposition method, can be utilized effectively to link the spatial and temporal patterns of a data field (Legler 1983). This method partitions the temporal variance of the data into orthogonal spatial patterns called eigenvectors.

Space-time data can be decomposed in terms of EOFs. That is,

$$T(x,t)=\sum_{n}T_n(t)B_n(x) \tag{2.10}$$

where $\{B_n(x)\}$ are called eigenfunctions and $\{T_n(t)\}$ are expansion of coefficients that are functions of time. Eq. (2.10) is called EOF decomposition only

if $\{B_n(x)\}$ are mutually orthogonal to each other and $\{T_n(t)\}$ are mutually uncorrelated. That is,

$$B_n(x) \cdot B_m(x) = \frac{1}{N}\sum_{x=1}^{N} B_n(x) B_m(x) = \delta_{nm} \tag{2.11}$$

$$T_n(t) \cdot T_m(t) = \frac{1}{N}\sum_{t=1}^{N} T_n(t) T_m(t) = \lambda_n \delta_{nm} \tag{2.12}$$

where δ_{nm} is Kronecker delta and λ_n is eigenvalues. Often, $\{B_n(x)\}$ are called loading vectors, $\{T_n(t)\}$ principal component (PC) time series. Spatial covariance function defines covriance between two spatial points *x* and *x'*:

$$C(x,x') = \langle T(x,t) T(x',t) \rangle \tag{2.13}$$

which under the stationarity aussumption can be written as

$$C(x,x') = \frac{1}{N}\sum_{t=1}^{N} T(x,t) T(x',t) \tag{2.14}$$

In terms of EOFs, spatial covariance matrix is

$$\begin{aligned} C(x,x') &= \left\langle \sum_n T_n(t) B_n(x) \sum_m T_m(t) B_m(x') \right\rangle \\ &= \sum_n \sum_m B_n(x) B_m(x') \langle T_n(t) T_m(t) \rangle \\ &= \sum_n \lambda_n B_n(x) B_n(x') \end{aligned} \tag{2.15}$$

Dot product of Eq. (2.15) with $B_m(x')$ yields

$$\begin{aligned} C(x,x')\cdot B_m(x') &= \sum_n \lambda_n B_n(x) B_n(x')\cdot B_m(x') \\ &= \sum_n \lambda_n B_n(x)\delta_{nm} = \lambda_m B_m(x) \end{aligned} \tag{2.16}$$

It is called the Karhunen-Loèvé equation. It is similar to the eigenvalue problem (Kim, Hamlington, and Na 2015).

2.2.2 Wavelet analysis

Wavelet analysis (WA), which provides a perfect filtering characteristic, has been used in various fields of mathematics, science, and engineering. Through the decomposition and reconstruction of a signal, WA can be performed to determine the transient identity in the time-frequency domain (Wang, Lee, and Zhang 2004). Unlike Fourier technique or EOF analysis, the basis function of the wavelet transform has the key property of localization in time (or space) and frequency. This makes the wavelets ideal for handling non-stationary and transient signals, as well as fractal-type structures (Murguia and Campos-Cantón 2006).

Wavelet analysis is used to decompose or reconstruct a signal using the wavelet functions. The wavelets are a family of orthogonal functions of type,

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left[(t-b)/a\right],\ \ a,b\in R,\ a\neq 0 \tag{2.17}$$

generated from a "mother" wavelet function $\psi(t)$ by dilation and translation

operations, which are governed by the scale factor a and shift factor b, respectively.

The continuous wavelet transform (CWT) and its reconstruction version of a signal $f(t) \in L^2(R)$ are defines as,

$$W_f(a,b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t)\psi^*\left[(t-b)/a\right]dt = \left\langle f(t), \Psi_{a,b}(t) \right\rangle \tag{2.18}$$

$$f(t) = C_\Psi^{-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[W_f(a,b)\Psi_{a,b}(t)/a^2\right]dadb \tag{2.19}$$

where, $\psi^*\left[(t-b)/a\right]$ is the complex conjugate of $\psi\left[(t-b)/a\right]$, $C_\Psi = \int_{-\infty}^{\infty} |F_\Psi(\omega)|^2 / |\omega| d\omega$, $F_\Psi(\omega)$ is the Fourier transform of $\Psi(t)$.

To avoid the redundancy in CWT, a so-called dyadic discrete wavelet transform DWT, is usually adopted in engineering practice. Letting $a = 2^{-j}$, $b = 2^{-j}k$ $(j,k \in Z)$, Eq. (3.18) and (3.19) may be rewritten in a dyadic discrete form,

$$W_f(a,b) = W_f\left(2^j, 2^{-j}k\right) = 2^j \int_{-\infty}^{\infty} \psi\left(2^j t - k\right) f(t) dt \tag{2.20}$$

$$f(t) = \sum_j \sum_k W_f(2^{-j}, 2^{-j}k)\Psi(2^j t - k) \tag{2.21}$$

The theory is referred to Wang et al. (2004).

In wavelet analysis, original series are decomposed into approximation and detail. The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components. The filtering process, at its most basic level, the original signal passes through two complementary filters and emerges as two signals. The selection of a suitable level for the hierarchy depends on the signal and experience. Often the level is chosen

based on a desired low-pass cutoff frequency.

In this process, the length of data becomes twice as much data as started. They are the signals *A* and *D*. To fix the extended length into the original length, downsampling is necessary. Through the process of downsampling, two sequences called *cA* and *cD* are achieved. After this process, DWT coefficients are produced. The decomposition can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the wavelet decomposition tree.

The organizing parameter, the scale a, is related to level j by $a = 2^j$. If resolution is defined as $1/a$, then the resolution increases as the scale decreases. The greater the resolution, the smaller and finer are the details that can be accessed. From a technical point of view, the size of the revealed details for any j is proportional to the size of the domain in which the wavelet or analyzing function of the variable x, $\psi(x/a)$ is not too close to 0.

The process of reconstruction assembles the decomposed components back into the original signal without loss of information. The mathematical manipulation that produces this synthesis is called the inverse discrete wavelet transform (IDWT). Where wavelet analysis involves filtering and downsampling, the wavelet reconstruction process consists of upsampling and filtering. Upsampling is the process of lengthening a signal component by inserting zeros between samples. The low- and high-pass decomposition filters (*L* and *H*), together with their associated reconstruction filters (*L*` and *H*`), form a system of what is called quadrature mirror filters.

There are several types of wavelet families such as Daubechies, Biorthogonal, Coiflets, Symlets, Morlet, etc. In this study, Coiflet5 (coif5) was used. This wavelet function ψ has 2*N* moments equal to 0 and the scaling function ϕ has 2*N*-1 moments equal to 0. The two functions have a support of length 6*N*-1. The coifN ψ and ϕ are much more symmetrical than the dbNs (Daubechies). With respect

to the support length, coifN has to be compared to db3N or sym3N (Symlets), which belong to yet another wavelet family. The wavelet analysis was carried out using MATLAB toolbox.

2.2.3 Artificial Neural Network

Artificial neural network (ANN) is suitable for partially understood underlying physical processes such as wind-wave relationship. Fig. 2.1 shows the structure of feed forward back propagation type of ANN, which consists of an input layer, one or more hidden layers, and an output layer. These layers have one or more nodes. The source nodes in the input layer propagate through the network in a forward direction. Each node of a layer connects that of the next layer. The back-propagation learning consists of two passes through the different layers of the network: a forward pass and a backward pass. In the forward pass, an input vector is applied to the sensory nodes of the network and its effect propagates through the network layer by layer. Finally, a set of outputs is produced as the actual response of the network. During the forward pass, all the synoptic weights of the networks are fixed. During the backward pass, on the other hand, all the synoptic weights are adjusted in accordance with an error correction rule (Günaydın 2008).

The outputs of the hidden layer are calculated from

$$H.O_j = f(net_j) = f\left(\sum_{i=1}^{n} w_{ij} IO_i + b_1\right), \quad j = 1, 2, \cdots, n \tag{2.22}$$

where b_1 is the first layer bias, $f(\ )$ is the transfer function between input and

hidden layers. In the present study, a sigmoid transfer function of the type $f\left(net_j\right)=1/\left(1+\exp\left(-net_j\right)\right)$ is used. Outputs of the output layer are determined as follows:

$$O.O_k = f(net_k) = f\left(\sum_{j=0}^{h_m} w_{jk} H.O_j + b_2\right), \quad k=1 \tag{2.23}$$

where b_2 is the second layer bias, $f(\ )$ is the transfer function between hidden and output layers. In this study, a linear transfer function $f(net_k) = net_k$ is used.

To avoid the problem of overfitting that may occur while an ANN is being trained, the number of nodes in the hidden layer, *z*, of the nets employed in the first stages of the study was computed using one of the empirical expressions mentioned by Fletcher and Goss (1993)

$$2n^{1/2} + m \tag{2.24}$$

where *n* is the number of input nodes and *m* is the number of output nodes.

In training schemes the difference or the error between the network-yielded and the target or actual output is minimized using a particular mathematical algorithm. The global (mean sum squared) error (*E*), is defined as follows:

$$e_p = \frac{1}{2}\sum_{k=1}^{l}\left(T_k - O.O_k\right)^2 \tag{2.25}$$

$$E = \frac{1}{P}\sum_{p=1}^{P} e_p \tag{2.26}$$

where T_k is the target (observed) output at the *k*th output node, $O.O_k$ is the predicted output at the *k*th output node, *P* is total number of training patterns, and e_p is the error for the *p*th training pattern. The procedure of back propagation network is to propagate the error at the output layer backward from the output to hidden layer in order to adjust the weights in each layer of the network.

There are several training algorithms such as Resilient back-Propagation (RP), Scaled Conjugate Gradient (SCG), Conjugate Gradient Powell-Beale (CGB), Broyden, Fletcher, Goldfarb (BFG), and Levenberg-Marquardt (LM). The rate of change of error with respect to the connection weights, i.e. the error gradient is used as a path to do so. A general conjugate gradient scheme involves performing a search along the conjugate or orthogonal direction in order to determine the step size to minimize the performance function. Among the algorithms, the LM is the fastest method for training moderate-sized feedforward neural networks. However, it requires the storage of some matrices that can be quite large for certain problems. In this research, the network is very large, so one of the conjugated gradient algorithms is recommended (Demuth and Beale 2000). Kalra et al. (2005) compared the prediction performance of the five algorithms and mentioned that the CGB produced the most accurate results among the algorithms. Therefore, the CGB scheme was used in this study. The ANN model implementation was carried out using MATLAB toolbox.

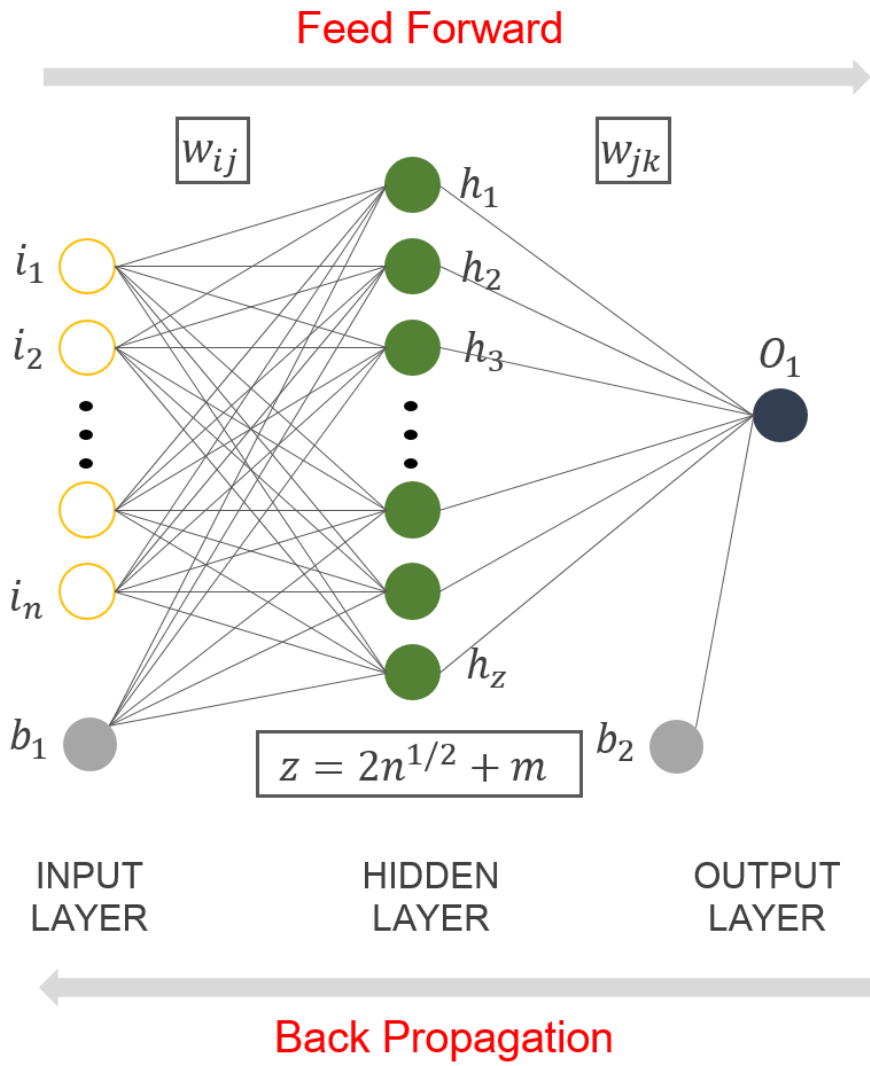


Fig. 2.1 ANN structure

2.2.4 EOF-Wavelet-ANN (EOFWNN) model

In this study, the combination of neural network, EOF analysis and wavelet analysis was employed to forecast significant wave height and period from observed wave data and reanalysis meteorological data. Fig. 2.2 is the schematic diagram of the EOFWNN model, which is the detailed flow chart of the right side of Fig. 1.1. First, EOF analysis was conducted for wave data and meteorological data to separate spatial and temporal components for training period. Second, wavelet analysis was applied to each PC time series of wave data and meteorological data. Next, training was conducted with the decomposed wavelet component time series of wind speed, sea level pressure and wave height (or period) data as input data and each PC time series of wave height (or period) data for various lead times for target. Finally, the forecasted wave height (or period) PC time series and the LVs obtained in the first step were reconstructed to calculate the wave height (or period) time series.

The decomposed wavelet signals of wave data and meteorological data used as the input to the ANN were those of all the decomposed PC time series of significant wave height and the approximations of 90% PC time series of meteorological data. To consider the dominance of persistence in the wave height time series, the significant wave heights at the present time and the previous two time steps were used as predictors. To consider the time lag effect of meteorological data, the meteorological data at the forecast lead time and at one and six time steps ahead of that time were used as predictors. In summary, the scenario formed by predictor configuration to predict $H(t+n\Delta t)$ is $H(t)$,

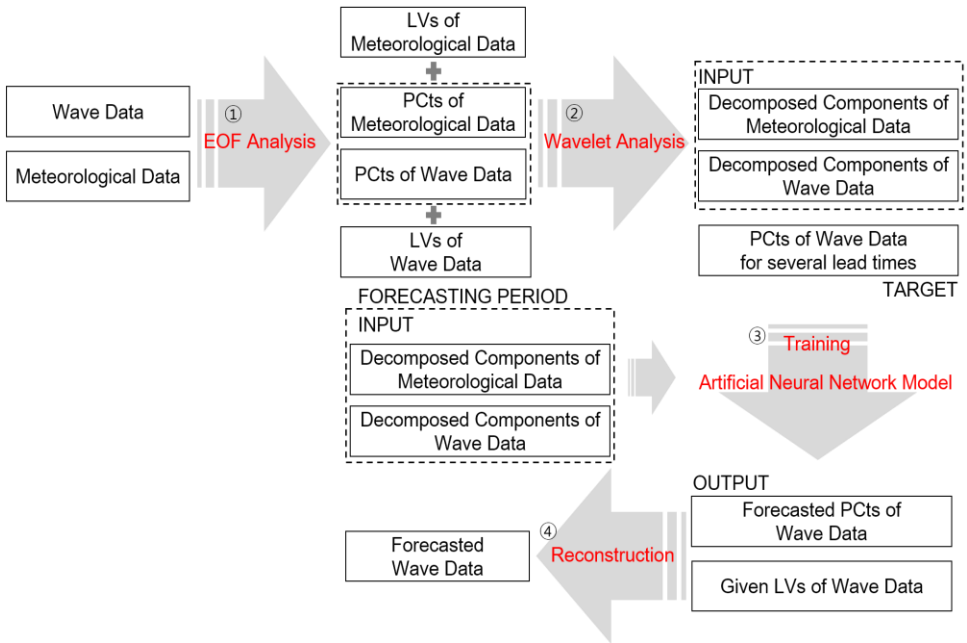


Fig. 2.2 Schematic diagram of the EOFWNN model

$H(t-\Delta t)$, $H(t-2\Delta t)$, $wnd(t+n\Delta t)$, $wnd(t+n\Delta t-\Delta t)$, $wnd(t+n\Delta t-6\Delta t)$, $slp(t+n\Delta t)$, $slp(t+n\Delta t-\Delta t)$, $slp(t+n\Delta t-6\Delta t)$. Where, $H(t)$ is the wavelet component of the present wave height, $H(t-\Delta t)$ and $H(t-2\Delta t)$ are previous time steps, $H(t+n\Delta t)$ is the PC time series of significant wave height at the forecast lead time, $wnd(t+n\Delta t)$ and $slp(t+n\Delta t)$ are the wavelet components of wind speed and sea level pressure at the forecast lead time, and $wnd(t+n\Delta t-\Delta t)$, $wnd(t+n\Delta t-6\Delta t)$, $slp(t+n\Delta t-\Delta t)$, $slp(t+n\Delta t-6\Delta t)$ are those at previous time steps from the forecast lead time, and '$n\Delta t$' denotes the forecast lead time. The lead times were fixed as 1, 3, 12, and 24 hours with the use of $\Delta t = 0.5$ hr. The meteorological data at Δt and $6\Delta t$ ahead of the forecast lead time were arbitrarily chosen by supposing that the wave height at a certain time would be closely related to the meteorological conditions at 30 minutes and 3 hours ahead of the time. Fig. 2.3 shows the predictor configuration of wave forecasting model.

The performance was repeated 20 times, and then ensemble members were averaged. After removing the two cases of the largest root mean squared error (*RMSE*) between ensemble average and each member, the remaining 18 ensemble members were averaged.

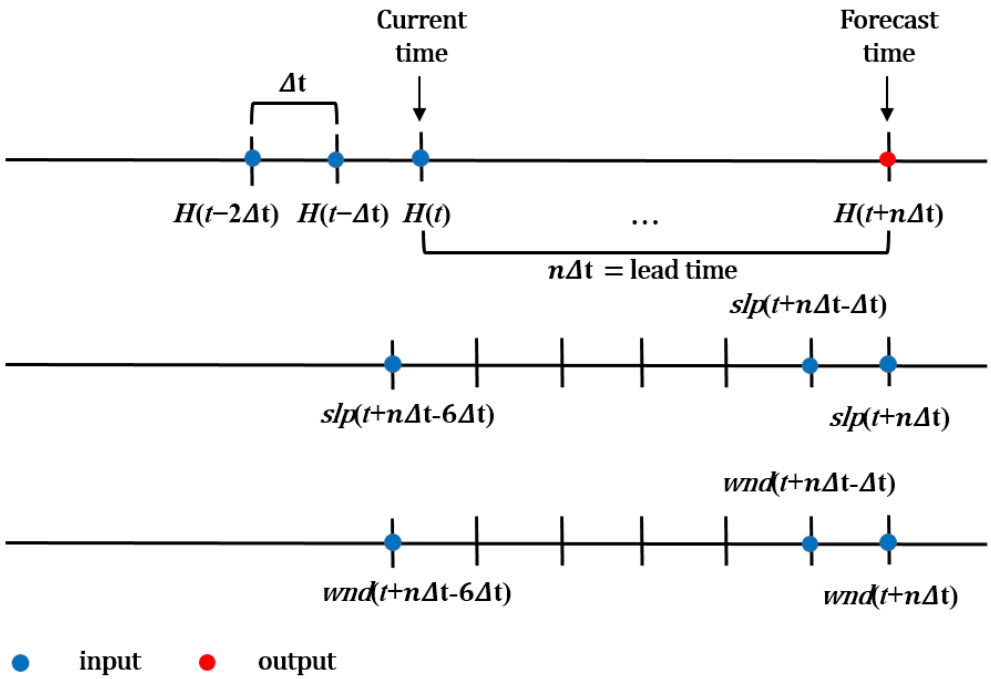


Fig. 2.3 Predictor configuration of wave forecasting model

CHAPTER 3. CHARACTERISTICS OF SUDDEN HIGH WAVES

3.1 Data for analysis of characteristics of sudden high waves

3.1.1 Wave data

To examine the characteristics of sudden high waves that caused marine accidents and property damage, investigations were made into the wave data at Gangneung and Wangdolcho on the east coast of Korea. The wave data were provided by KIOST (Korea Institute of Ocean Science and Technology). Fig. 3.1 shows the locations of wave measurement. The wave data of KIOST was measured every 0.5 s using pressure gauge and sampling rate is $\Delta f \approx 0.008\,Hz$. Wave spectrum was calculated every 30 minutes from the collected wave series. The water depths at the wave measurement stations were 15.0 and 15.3 m, respectively. Experiments were conducted for the winter season from October to February, when sudden high waves occur frequently. The period is 5 years in 2005, 2008, 2010-2013 (Oct. 2005-Feb. 2013).

3.1.2 Meteorological data

The meteorological data used in this study were National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis data provided by National Oceanic and Atmospheric

Administration/Climate Diagnostics Center (NOAA/CDC) (Kalnay et al. 1996). The sea level pressures and 10 m height wind speeds in the *u*-direction (east-west) and *v*-direction (north-south) were used to analyze the relationship between meteorological variables and waves. This dataset can be downloaded from a website (https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html).

The temporal resolution is 30-min interval interpolated from 6 hour interval data, and the spatial resolutions are 2.5 ° × 2.5 ° grid for sea level pressure data and T62 Gaussian grid for wind speed data, which roughly corresponds to 220 km in the *u*-direction and to 280 km in the *v*-direction. Fig. 3.2 shows the region of meteorological data, which is 127°-142° E, 33°-46° N.

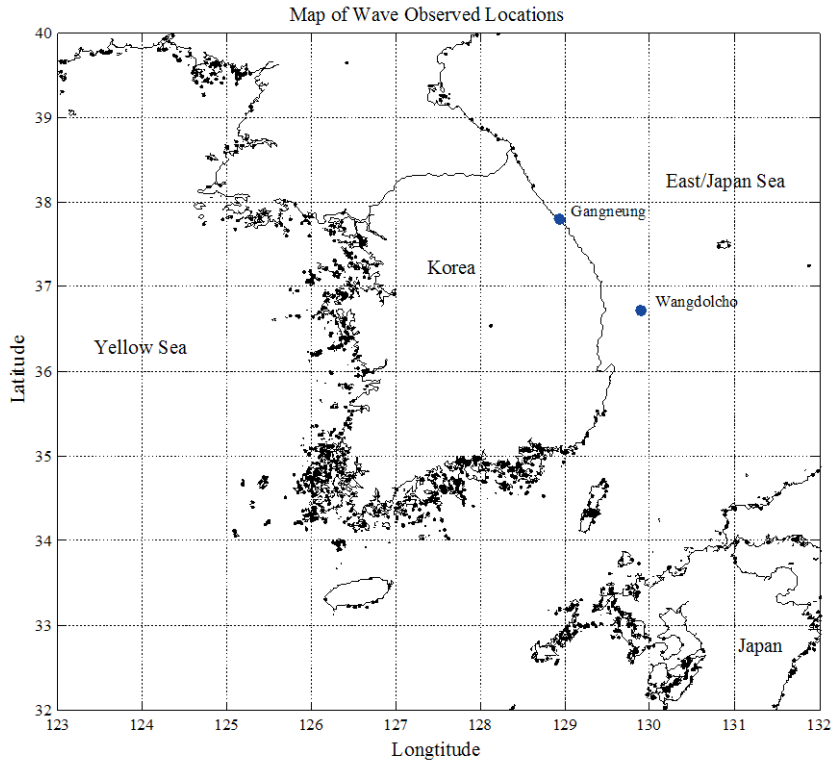


Fig. 3.1 Locations of wave observation by KIOST

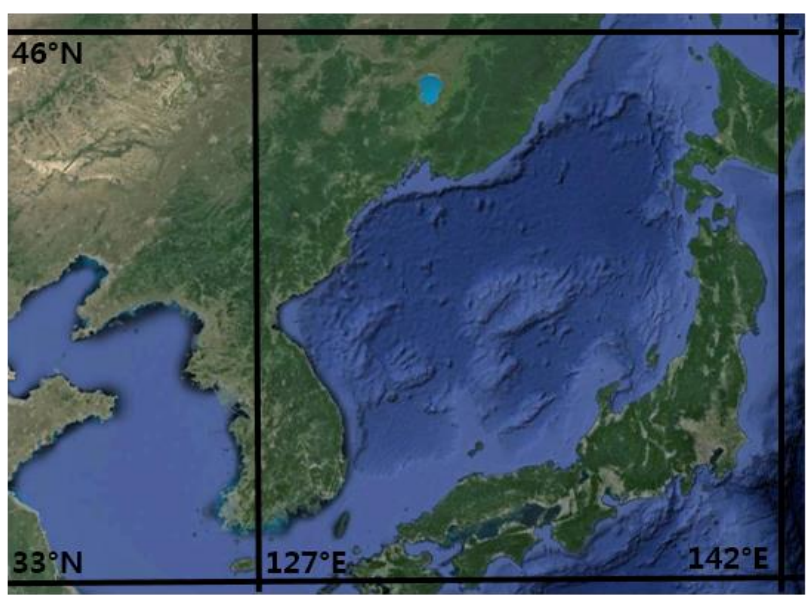


Fig. 3.2 Region of meteorological data

3.2 Definition of sudden high wave

Sudden high waves, which occur frequently in the winter season in the east coast of Korea, have caused serious property damage and many casualties. These high waves are called yorimawari waves in Japan (Nagai et al. 2009, Lee et al. 2010). Nagai et al. (2009) used the terminology of sudden peculiar long period swell. These high waves have a broader frequency band than a swell, and their significant height is frequently greater than 5 m. Also, they have a wave period greater than 9 s, which exceeds that of ordinary wind seas. Therefore, they have characteristics of both swell and wind waves. Oh and Jeong (2013) used the terminology of high swell-like waves to describe the waves having a relatively long period (T > 9 s) and large height (H > 3 m). This terminology, however, does not represent the suddenness of the waves. In this study, therefore, the terminology of sudden high waves was used. Fig.3.3 shows the event of such high waves in October 2005 in Gangneung and Wangdolcho in which both wave height and period suddenly increased from a calm state of sea. The parameter describing sudden high waves should include two information: suddenness and magnitude (i.e. height and period) of the waves. The purpose of this study is to introduce a parameter that includes such information and propose a criterion of sudden high waves that would cause marine accidents and property damage.

Prior to proposing a criterion of sudden high waves, high wave events should be determined. Since in general the wave period changes in proportion to the wave height as shown in Fig. 3.3, the high wave events are determined based on the time series of wave height. Here, a high wave event indicates a time span during which the calm sea surface becomes rough with increasing wave height and then retrieves

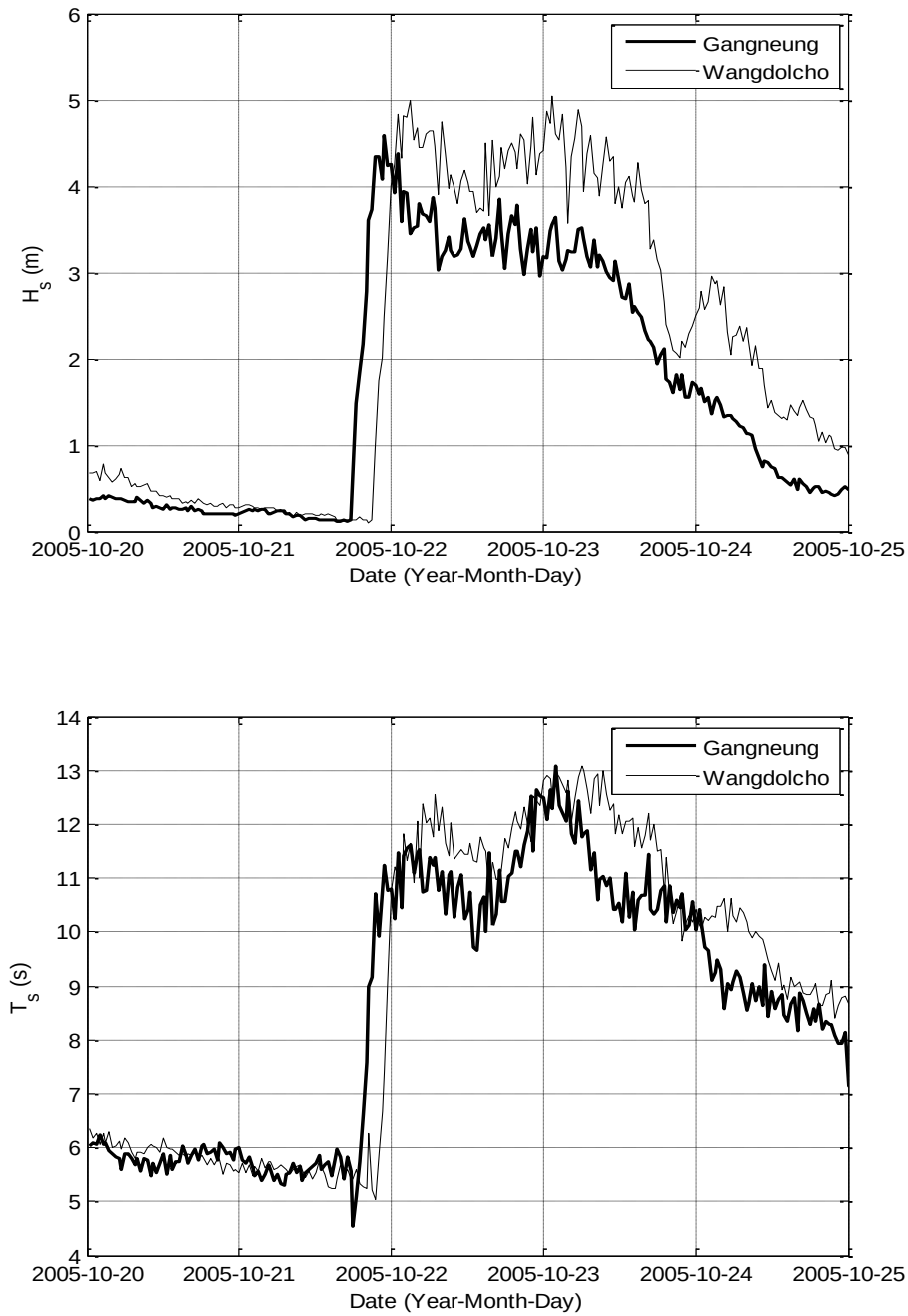


Fig. 3.3 Event of sudden high waves in October 2005 in Gangneung and Wangdolcho

a calm sea state. The slight sea of Degree 3 of Douglas sea scale indicates the sea state of wave heights between 0.5 and 1.25. In this study, the upper bound value, 1.25 m, was used as the criterion of calm sea state.

Referring to Fig. 3.4, the high wave events were determined as follows. The triangles in the figure indicate the start and end of an event. If we only use the criterion that the calm sea state becomes rough and retrieves a calm state, too many high wave events are counted. To resolve this problem, a comparison period of minimum wave height is introduced. Fig. 3.4 is the result of using 60 hours of comparison period. If the wave height at a time satisfies the calm sea state condition and it is the minimum between before and after the comparison period from the time, the time is determined as the starting point of a high wave event. This method enables one to find the minimum value from a graph which is not differentiable due to severe fluctuation as in Fig. 3.4.

The most important variable to represent sudden high waves may be the wave height, which best represents the magnitude of waves and most directly affects the damage due to waves. However, as shown in Fig. 3.3, the wave period also increases rapidly during the occurrence of sudden high waves. Since the wavelength increases with the wave period, H^2L could be a candidate to represent the effects of both height and period of sudden high waves, which also represents the wave energy in one wavelength. Kim and Lee (2008b), using the wavelet analysis, have shown that wave energy increased rapidly during the sudden high wave event in February 23-24, 2008. On the other hand, the most important characteristic of sudden high waves is its suddenness. To represent the suddenness, the time interval Δt has to be considered during which the waves are developed from a calm sea to a rough sea. In this study, therefore, I selected $\Delta H / \Delta t$ and

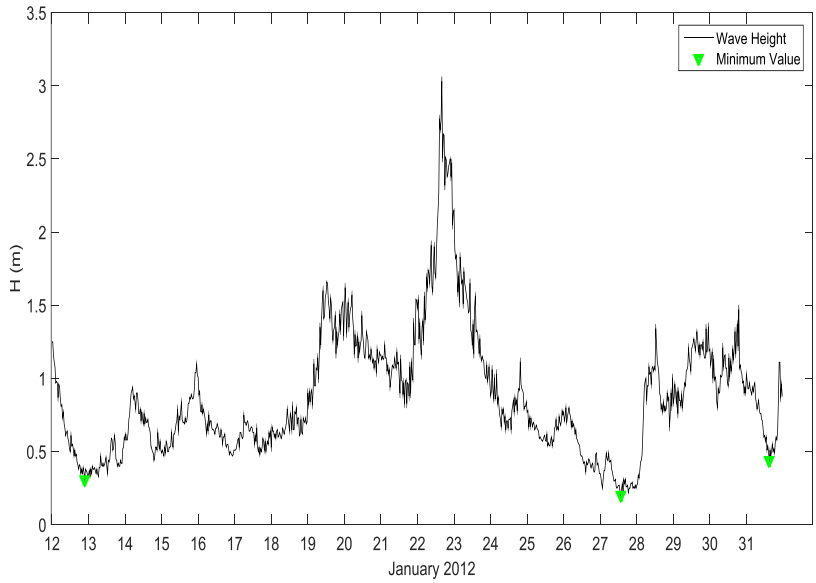


Fig. 3.4 Illustration of high wave events

$\Delta(H^2L)/\Delta t$ as the parameters to represent the intensity of sudden high waves, where ΔH and ΔH^2L are the increments of H and H^2L, respectively, during the time interval Δt.

First, during the period of a high wave event, the point of maximum H or H^2L was taken. In the case of more than two of the same maximum value, the first one was taken. The start point of sudden high waves was then taken as the point where $\Delta H/\Delta t$ or $\Delta(H^2L)/\Delta t$ becomes maximum among the points where H or H^2L is smaller than 1/10 of its maximum value and where the wave height is smaller than 1.25 m, i.e., the upper bound value of slight sea of Degree 3 of Douglas sea scale. Since the rapid increase of H or H^2L is important for the suddenness of high waves, the points where H or H^2L is smaller than 1/10 of its maximum value were considered as the candidates for the starting point of sudden high waves.

Fig. 3.5 shows the time series of H^2L during two high wave events at the Gangneung wave station. The triangle indicates the maximum value, while the diamond indicates the starting point of sudden high waves. In the event of February 2008, shown in Fig. 3.5 (a), the point of maximum $\Delta(H^2L)/\Delta t$ is located in the 24th day, but the point was not selected as the starting point of sudden high waves because it did not satisfy the criteria that the wave height is smaller than 1.25 m and H^2L is smaller than 1/10 of its maximum value. In the other hand, in the event of January 2009, shown in Fig. 3.5 (b), a point at the beginning of the 9th day was selected as the starting point because it satisfied the criteria for the starting point of sudden high waves while showing the maximum $\Delta(H^2L)/\Delta t$. Even

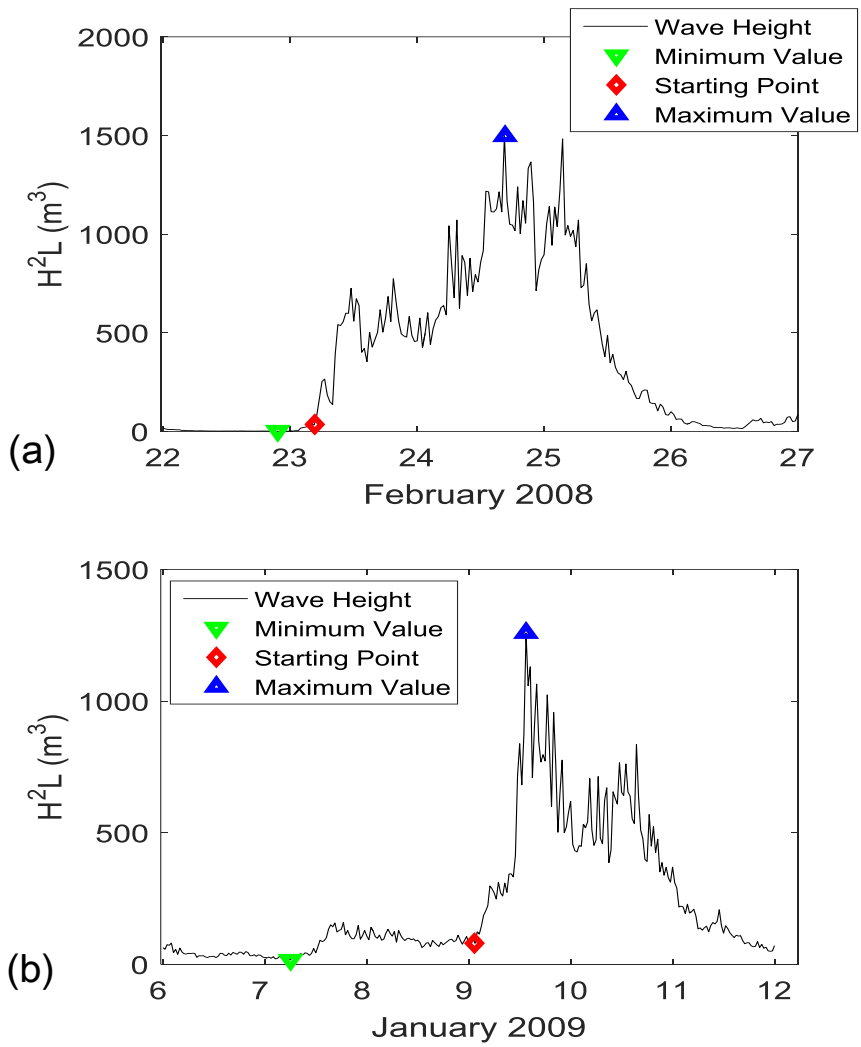


Fig. 3.5 Illustration of calculation of intensity parameter of sudden high waves in Gangnueng: (a) High wave event in February 23-25, 2008; and (b) High wave event in January 9-11, 2009.

though the maximum wave energy in the 2008 event is greater than that in the 2009 event, the values of $\Delta(H^2L)/\Delta t$ are calculated as 40.7 and 98.2 m^3/hr for each event in 2008 and 2009, respectively. This indicates that the suddenness of high waves is well represented by the parameter $\Delta(H^2L)/\Delta t$. The above procedure enables one to distinguish sudden high waves from ordinary high waves.

The wave data at the Gangneung station gave 208 and 206 high wave events for H and H^2L, respectively. The high wave events were then compared with the marine accidents and property damage on the coast of Gangwon-do Province in which the Gangneung station is located. Table 3.1 shows the marine accidents and property damage occurred on the coast of Gangwon-do Province and Gyeongsangbuk-do Province during the winter season between 2005 and 2012. On the other hand, Table 3.2 shows the characteristics of high wave events at the Gangneung station at the times when marine accidents and property damage occurred on the coast of Gangwon-do Province. A comparison of the intensity parameters of sudden high waves shows that the ranks of $\Delta(H^2L)/\Delta t$ are higher than those of $\Delta H/\Delta t$ in most events. It is also observed that not only loss of lives but also severe property damage occurred over a wide area during the 2005 and 2006 events when $\Delta(H^2L)/\Delta t$ showed the highest ranks, indicating that $\Delta(H^2L)/\Delta t$ is proportional to the severity of the events. Therefore, $\Delta(H^2L)/\Delta t$ seems to be more appropriate than $\Delta H/\Delta t$ to measure the intensity of sudden high waves.

On the other hand, the wave data at the Gangneung and Wangdolcho stations gave 206 and 155 high wave events, respectively, for H^2L. For each event,

Table 3.1 Marine accidents and/or property damage along the coast of Gangwon-do Province and Gyeongsangbuk-do Province (from Geosystem Research, 2015).

Date		Area	Specific area	Life damage (Death)	Remarks
2005	10.21	Pohang	Ship at 7 km offshore	14(9)	One ship sank, nine crew members disappeared
	10.22	Pohang	Imgok-ri breakwater	2(2)	Two children killed
	10.22	Donghae	Daejin Port	3(0)	Breakwater damage
	10.22	Donghae	Cheongok Port	3(0)	Breakwater damage
	10.22	Gangneung	Jumunjin Port breakwater	1(1)	One tourist killed
	10.23	Sokcho	Yeongrang-dong	-	Three houses partially destroyed, road damage
	10.23	Sokcho	Jangsa Port	-	Two ships sank
2006	10.08	Pohang	Yangpo Port breakwater	13(2)	Eleven tourists injured, two killed
	10.09	Ulleung	Jeodong Inner Port breakwater	1(1)	One tourist killed
	10.23	East/Japan Sea	Entire area	-	110 ships, 270 houses damaged
	10.23	Sokcho	Ship at 100 km offshore	7(4)	One ship sank
	10.23	Goseong	Bongpo Port breakwater	4(1)	Four tourists swept away, one killed
2007	10.28	Sokcho	Seashore rocks off Yeonggeum-jeong	3(1)	One tourist killed, two injured
2008	2.24	Gangneung	Anmok Port	18(3)	Three tourists killed
2009	1.10	Gangneung	Jumunjin Port breakwater	1(1)	One tourist killed
	1.13	Gangneung	Jumunjin Port north breakwater	5(3)	Three out of five tourists killed
2012	1.1	Pohang	Seashore rocks off Youngam-ri	2(0)	All rescued
	1.20	Samcheok	Seashore rocks at Namaemul, Wondeok-eup	2(1)	One angler killed, one rescued
	11.20	Gangneung	Namhangjin	4(0)	One barge damaged, rescue of human lives

Table 3.2 Characteristics of $\Delta H / \Delta t$ and $\Delta(H^2 L) / \Delta t$ in Gangneung

Date	$\Delta H / \Delta t$		$\Delta(H^2 L) / \Delta t$	
	Value (m/hr)	Rank (Percent)	Value (m^3/hr)	Rank (Percent)
2005/10/21	0.807	1 (0.48)	523.6	1 (0.49)
2006/10/23	0.367	3 (1.44)	353.2	2 (0.98)
2007/10/29	0.128	32 (15.3)	38.6	73 (35.4)
2008/02/24	0.080	65 (31.2)	40.7	67 (32.5)
2009/01/09	0.052	89 (42.8)	98.2	22 (10.7)
2012/01/22	0.012	197 (94.7)	39.4	70 (34.0)
2012/11/20	0.075	71 (34.1)	59.0	47 (22.9)

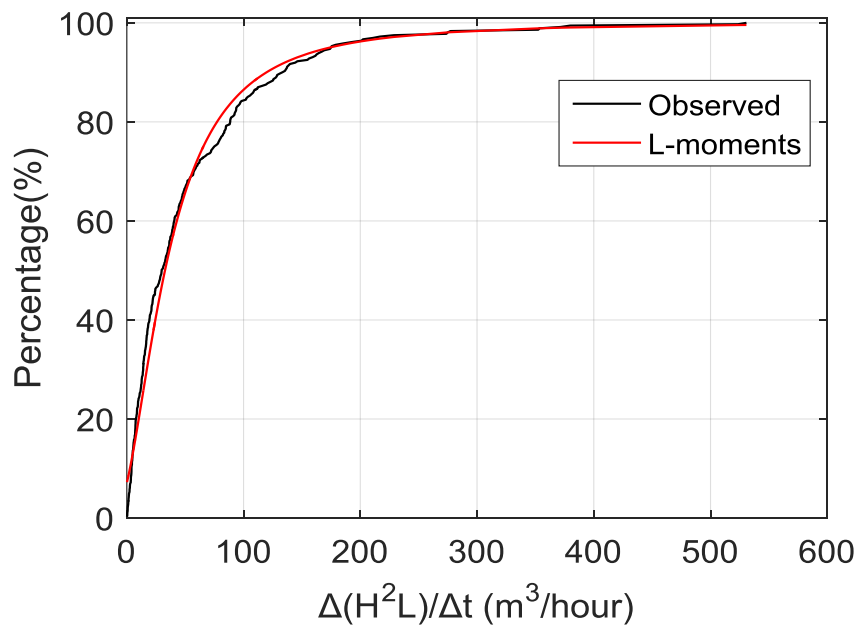


Fig.3.6 Cumulative percentage curve of $\Delta(H^2 L) / \Delta t$

$\Delta(H^2L)/\Delta t$ was also calculated. Fig. 3.6 shows the cumulative percentage curve of the 361 observed values of $\Delta(H^2L)/\Delta t$ along with the generalized extreme value distribution estimated by L-moments, the cumulative distribution function of which is given by

$$F(x)=\exp\left\{-\left[1-\frac{0.39(x-22.02)}{27.15}\right]^{1/0.39}\right\};\quad x=\Delta(H^2L)/\Delta t \tag{3.1}$$

The estimated distribution is adequate at the level of significance of 5% with the Kolmogorov-Smirnov goodness-of-fit test. The rapidly increasing curve for small values of $\Delta(H^2L)/\Delta t$ levels off at about 80% of cumulative percentage. Therefore, the criterion of sudden high waves was determined as $\Delta(H^2L)/\Delta t = 77.3 \approx 80$ m^3/hr, which corresponds to the top 20% on the estimated distribution.

Table 3.3 shows a comparison between sudden high wave events of $\Delta(H^2L)/\Delta t = 77.3 \approx 80$ m^3/hr (at one or both of the two wave stations) and marine accidents and property damage occurred on the coast of Gangwon-do Province and Gyeongsangbuk-do Province. By specifying the period of each event, we examined whether the time of the accident was within the period of the event. The cells shaded grey in the table indicate either the case of $\Delta(H^2L)/\Delta t \geq 80$ m^3/hr or the case that the time of the accident is not included in the period of the event. It can be seen that 13 out of 18 accidents occurred during sudden high wave events of $\Delta(H^2L)/\Delta t \geq 80$ m^3/hr. The other five accidents

Table 3.3 Comparison of sudden high wave events of $\Delta(H^2L)/\Delta t \geq 80\ \mathrm{m}^3/\mathrm{hr}$ and marine accidents

Gangneung		Wangdolcho		Accident	
Period	$\Delta(H^2L)/\Delta t$ (m^3/hr)	Period	$\Delta(H^2L)/\Delta t$ (m^3/hr)	Date	Area
2005/10/21 17:30 – 2005/10/24 06:30	523.6	2005/10/21 21:00 – 2005/10/24 18:30	128.4	2005/10/21	Pohang
				2005/10/22	Pohang
				2005/10/22	Donghae
				2005/10/22	Donghae
				2005/10/22	Gangneung
				2005/10/23	Sokcho
				2005/10/23	Sokcho
2006/10/11 21:30 – 2006/10/12 15:00	80.1	2006/10/12 01:30 – 2006/10/12 20:00	47.7	2006/10/08	Pohang
				2006/10/09	Ulleng
2006/10/22 21:00 – 2006/10/25 18:00	353.2	2006/10/23 03:00 – 2006/10/25 23:30	274.8	2006/10/23	Entire East/Japan Sea
				2006/10/23	Sokcho
				2006/10/23	Goseong
2008/02/23 04:30 – 2008/02/25 20:00	40.7	2008/02/23 06:30 – 2008/02/25 18:00	371.6	2008/02/24	Gangneung
2009/01/09 01:30 – 2009/01/11 08:30	98.2	2009/01/09 07:30 – 2009/01/11 19:00	146.1	2009/01/10	Gangneung
				2009/01/13	Gangenung
2012/01/01 18:00 – 2012/01/03 10:30	112.1	2012/01/01 21:30 – 2012/01/03 15:30	530.8	2012/01/01	Pohang
2012/01/21 21:00 – 2012/01/23 20:30	39.4	2012/01/22 09:30 – 2012/01/24 12:00	108.7	2012/01/20	Samcheok
2012/11/20 05:30 – 2012/11/21 12:30	59.0	2012/11/23 06:30 – 2012/11/24 15:30	84.8	2012/11/20	Gangneung

seem to occur due to ordinary high waves or other causes.

The total number of high wave events at the Gangneung and Wangdolcho stations is 361. Since the criterion of sudden high waves was taken as $\Delta(H^2L)/\Delta t \geq 80\ \mathrm{m}^3/\mathrm{hr}$, which corresponds to the top 24.7% of the observed values, the total number of sudden high wave events is 89. However, as shown in Table 3.3, the number of events that caused accidents while satisfying the criterion of sudden high waves is only 9. This means that no accident occurred during the remaining 80 sudden high wave events. It can be seen that, therefore, in most cases no accident occurs even though sudden high waves occur.

Lastly, to examine the possibility that the accident during a sudden high wave event was simply due to bad weather, the precipitation and maximum wind speed in the area of the accident were investigated as shown in Table 3.4. The data of Korea Meteorological Administration (2005 to 2012) were used. A heavy rain warning is issued when 80 mm or more of rainfall is expected during 24 hours, whereas a high wave warning is issued when 10-minute average wind speed over 14 m/s continues more than 3 hours. The sudden high wave event accompanied by such warnings of bad weather occurred only in October 23, 2006. It seems that bad weather and sudden high waves occurred simultaneously on that day. The remaining accidents occurred due to sudden high waves, not by bad weather. Therefore, all the 13 accidents in Table 3.3 could be considered to occur due to sudden high waves.

Table 3.4 Precipitation and maximum wind speed during marine accident and/or property damage due to sudden high waves

Date		Area	Specific area	Precipitation (mm)	Max wind speed (m/s)
2005	10/21	Pohang	Ship at 7 km offshore	0.4	8.3
	10/22	Pohang	Imgok-ri breakwater	2.0	3.6
	10/22	Donghae	Daejin Port	4.5	10.7
	10/22	Donghae	Cheongok Port	4.5	10.7
	10/22	Gangneung	Jumunjin Port breakwater	1.5	6.9
	10/23	Sokcho	Yeongrang-dong	-	4.8
	10/23	Sokcho	Jangsa Port	-	4.8
2006	10/23	East/Japan Sea	Entire area	232	30.5
	10/23	Sokcho	Ship at 100 km offshore	232	30.5
	10/23	Goseong	Bongpo Port breakwater	232	30.5
2008	2/24	Gangneung	Anmok Port	-	5.2
2009	1/10	Gangneung	Jumunjin Port breakwater	-	8.1
2012	1/1	Pohang	Seashore rocks off Youngam-ri	-	4.9
	11/20	Gangneung	Namhangjin	-	4.6

3.3 Mechanism of sudden high waves

Many studies have been performed for the generation process of high waves in winter season in the east coast of Korea. Oh and Jeong (2013) analyzed the characteristics of high swell-like waves and meteorological conditions during the events using the observed wave data at multiple stations along the east coast of Korea. According to them, important is not only the pressure drop during the movement of low pressure system but also other factors such as moving trajectory and staying time of the low pressure system together with arrangement of neighboring atmospheric pressure fields. In this study, the CSEOF analysis and regression analysis established for wave data and meteorological data are used in order to find the generation mechanism of sudden high waves.

The CSEOF analysis was applied to significant wave height and meteorological data, respectively. Nested period was chosen as 15 day period after checking if the characteristics of sudden high waves were detected or not from the wave series. Significant wave data provided by KIOST at Gangneung and Wangdolcho wave stations were used to analyze the mechanism of sudden high waves. Although the physical patterns of the events in the east coast of Korea and the west coast of Japan are similar, local characteristics could be different. Therefore, the wave data in the east coast of Korea were used to investigate the characteristics of sudden high waves more precisely.

Significant wave height series were decomposed into 40 modes, which corresponded 99% variance of the series. Fig. 3.7 shows the normalized and cumulative eigenvalues of CSEOF modes for the significant wave height. Fig. 3.8-Fig. 3.10 represent the CSEOF mode 1-3 of significant wave height. The first panel

of each figure is the physical pattern during the 15 day period. Blue line is for Gangneung and green line is for Wangdolcho. The second panel is the PC time series and the third panel is the reconstructed significant wave height of the mode. The last panel is the zoom view of the reconstructed significant wave height, which is to check if peak occurs or not during the reported marine accidents date on 21-23 October 2005. The shapes of the second and third CSLVs are similar to sudden high waves, so they were regarded as the physical process of sudden high waves. Also, the second and third modes have a large peak on 21-23 October 2005. Therefore, two modes were considered as the physical process of sudden high waves.

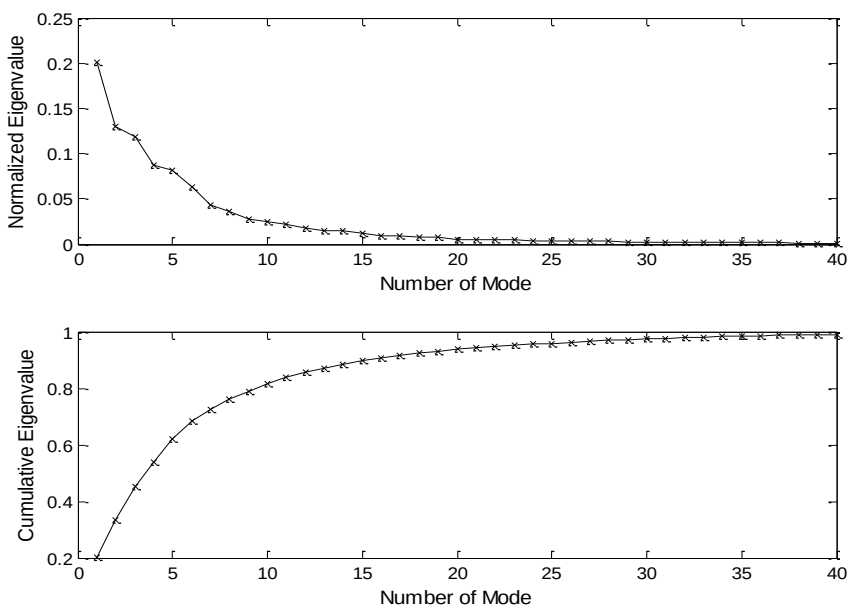


Fig. 3.7 Eigenvalues of CSEOF modes for significant wave height data from KIOST

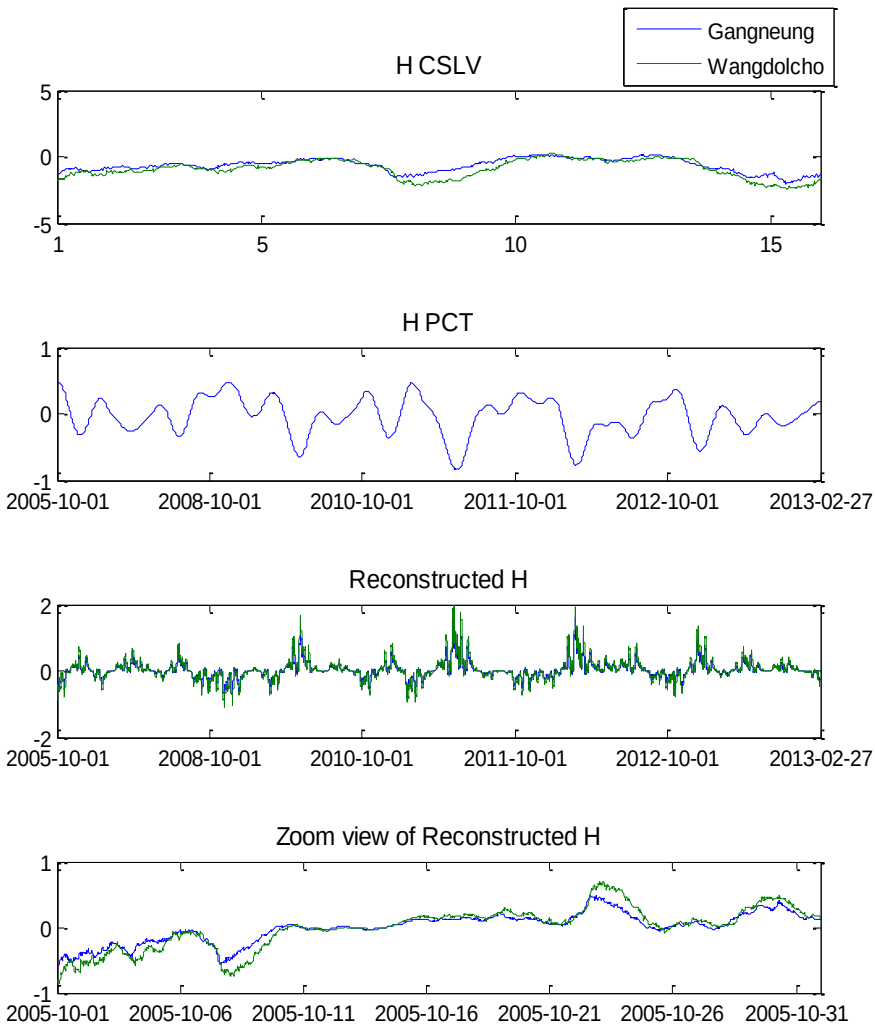


Fig. 3.8 CSEOF mode 1 of significant wave height data from KIOST

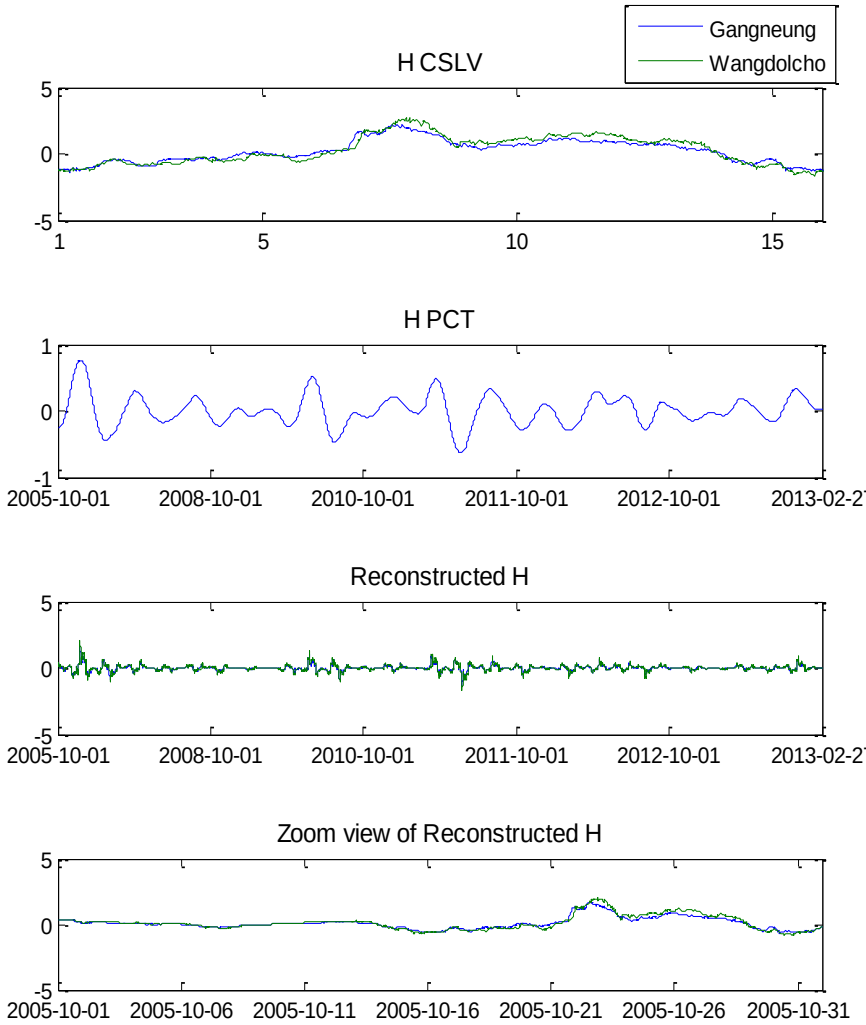


Fig. 3.9 CSEOF mode 2 of significant wave height data from KIOST

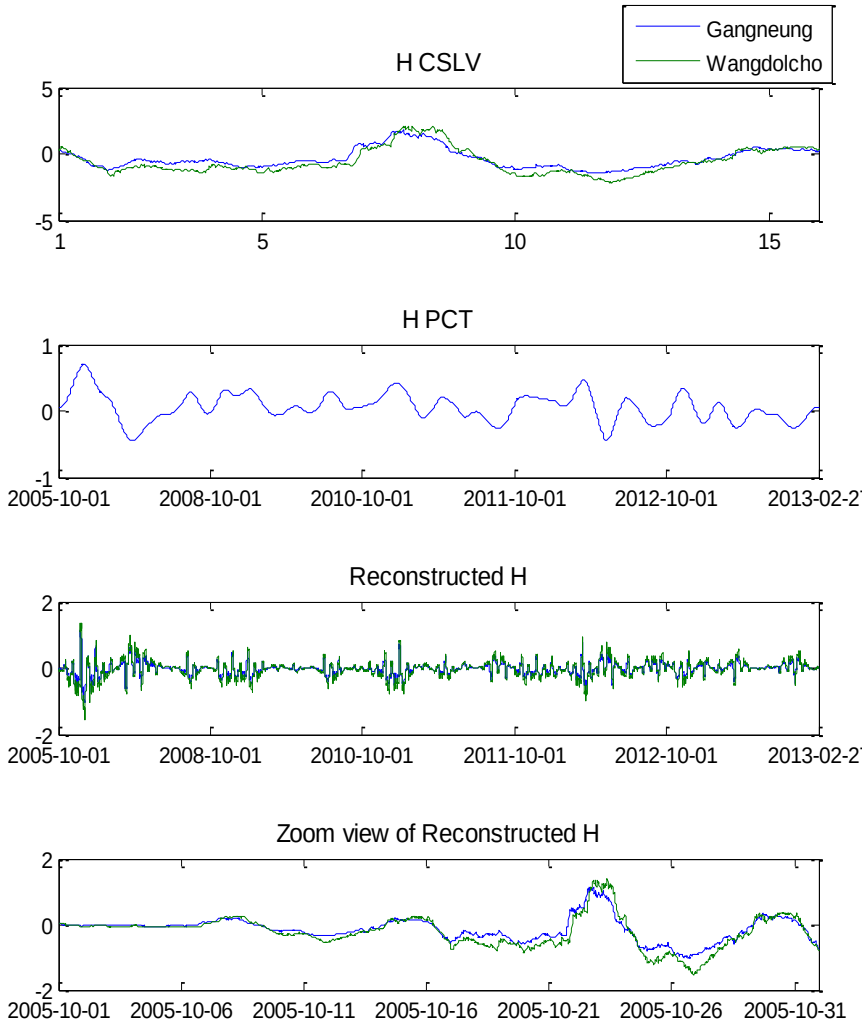


Fig. 3.10 CSEOF mode 3 of significant wave height data from KIOST

To detect sudden high wave modes, the index of sudden high waves, H^2L was analyzed using CSEOF analysis. It is decomposed into 33 modes. Fig. 3.11 presents the normalized and cumulative eigenvalues of CSEOF modes for the index. Fig. 3.12-Fig.3.14 show the first 3 modes for the index. The shapes of the modes are similar to those of significant wave height but they are more distinct. The first 3 modes are similar to sudden high waves. However, comparing the reported events and the reconstructed series, the third mode did not correspond. Therefore, the first two modes were considered as the physical process of sudden high waves.

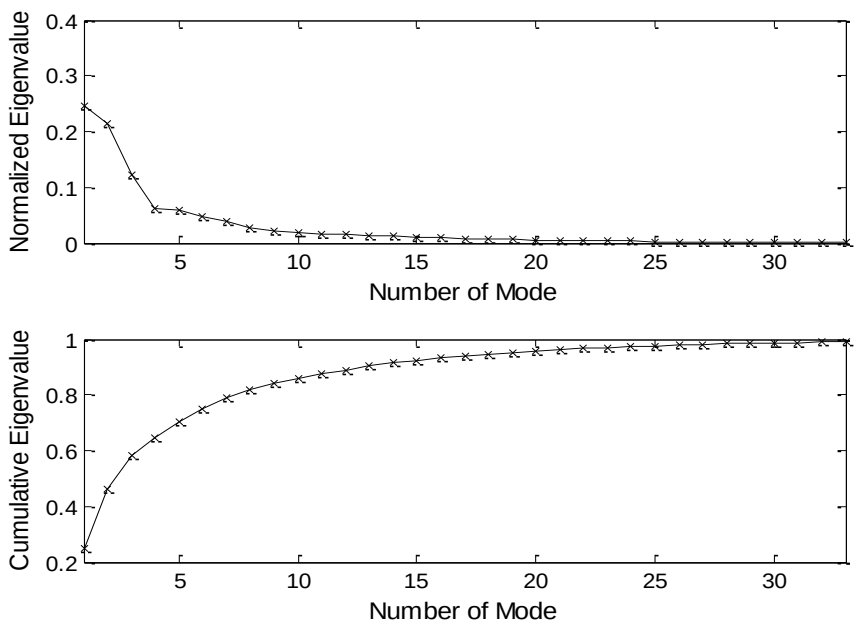


Fig. 3.11 Eigenvalues of CSEOF modes for sudden high wave index

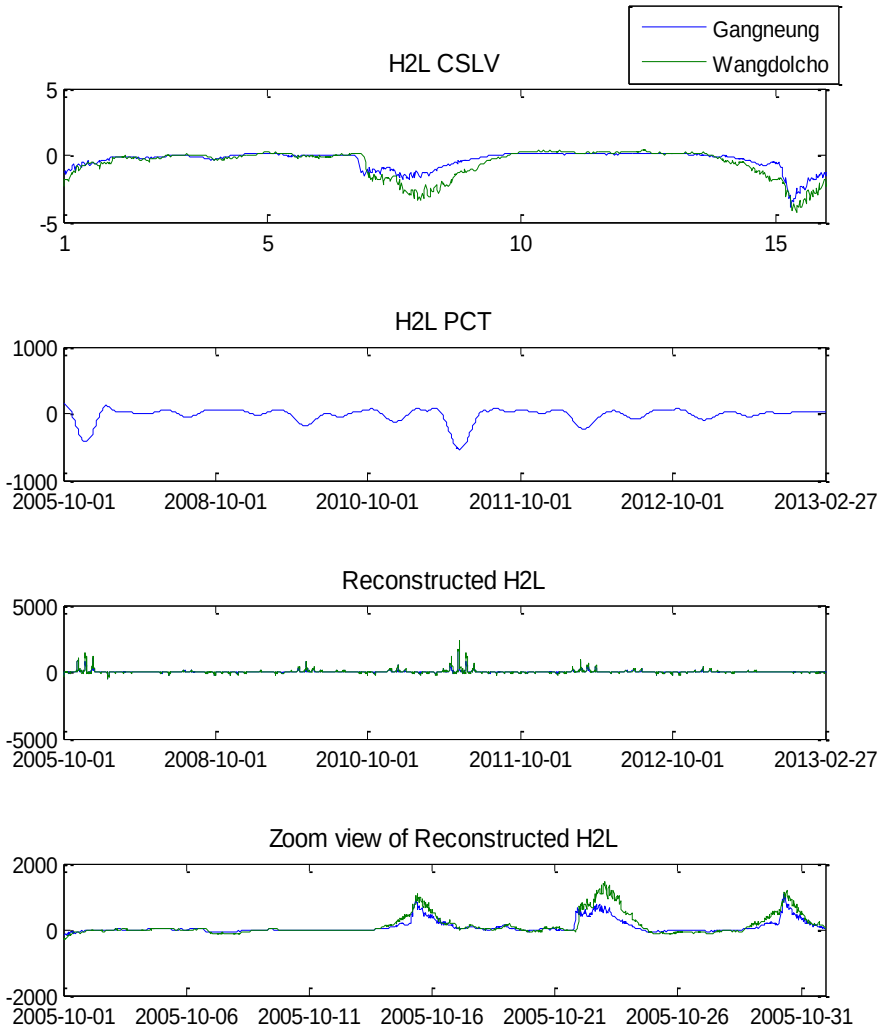


Fig. 3.12 CSEOF mode 1 of sudden high wave index

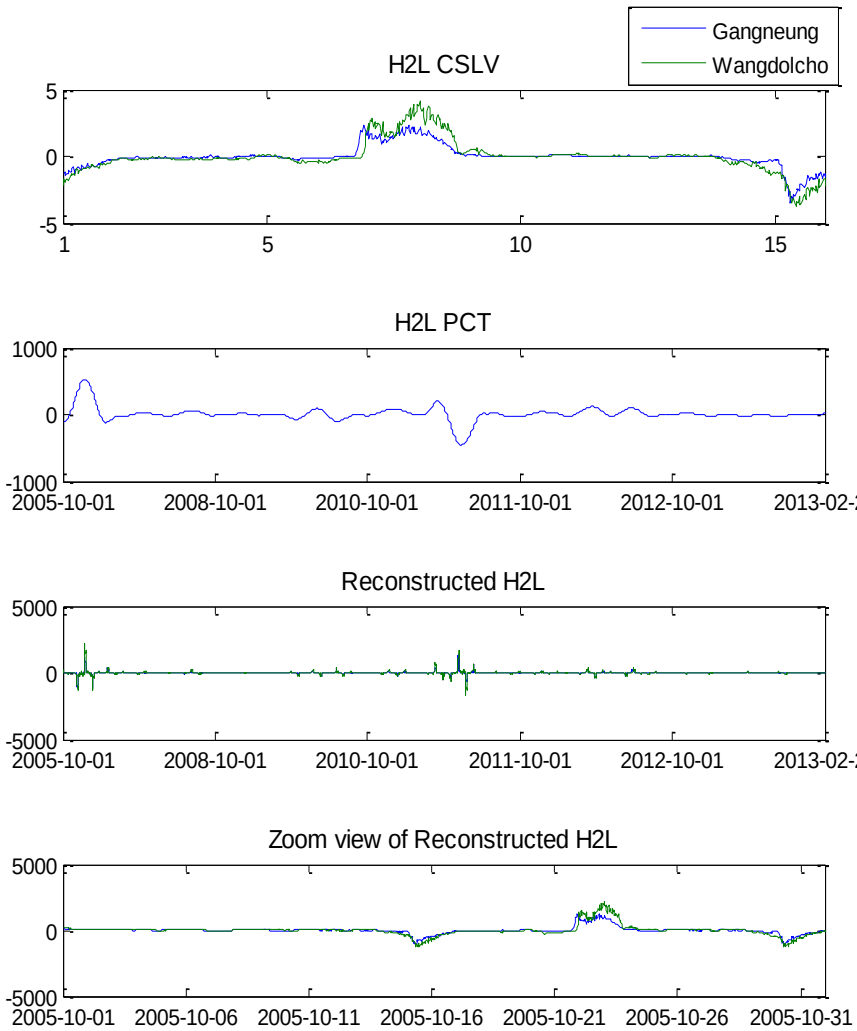


Fig. 3.13 CSEOF mode 2 of sudden high wave index

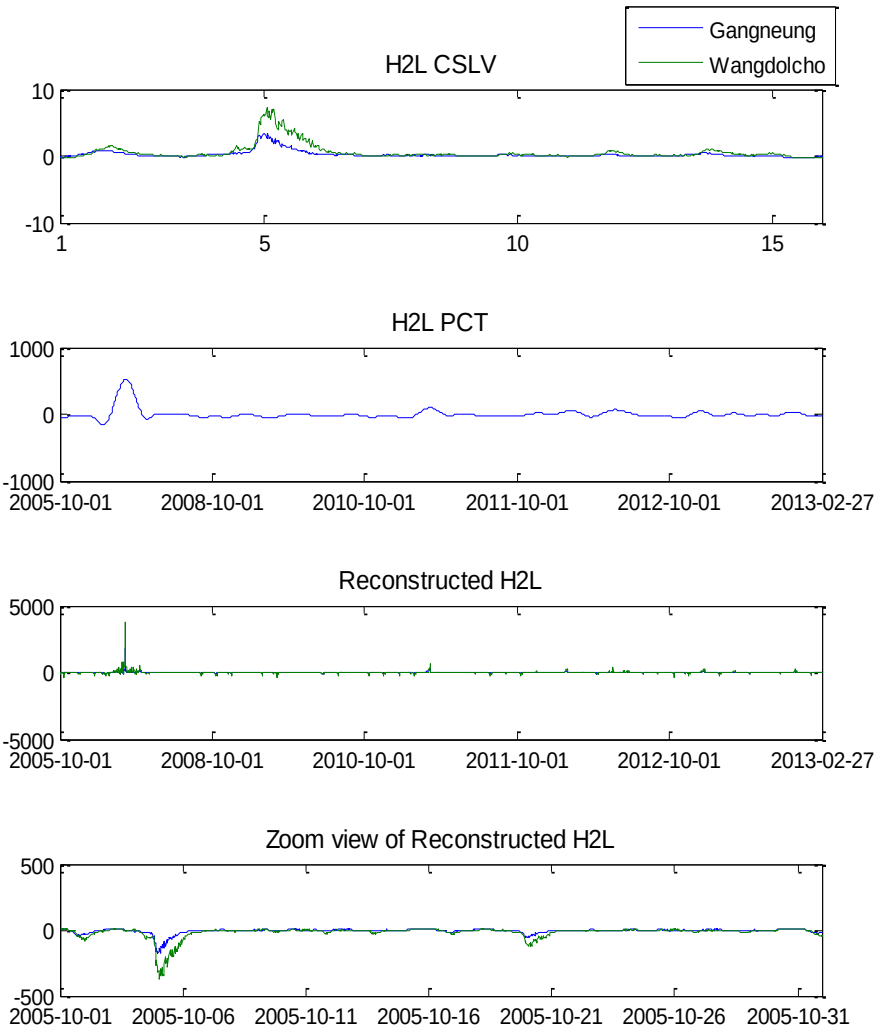


Fig. 3.14 CSEOF mode 3 of sudden high wave index

Next, regression analysis was conducted to find the consistent physical relation between waves and meteorological variables. Fig. 3.15 and Fig. 3.16 show comparisons of the first two modes of the regressed sea level pressure PC time series and wind speed PC time series with the significant wave height PC time series, respectively.

By multiplying the regression coefficients to CSLVs of meteorological variables, the processes of meteorological variables corresponding to the physical processes of significant wave height were produced. Fig. 3.17 shows the evolution of spatial patterns of the second mode of significant wave height and meteorological variables. They are the snapshots of the spatial patterns at the point of the red spot on the graph of the significant wave height CSLV. It shows the physical patterns of the meteorological variables when the second mode of physical pattern of the wave height occurred. Until two days before the rapid growth of the wave height, sea level pressure patterns show high pressure in the west of Korea and low pressure in the east and wind had blown continuously. From two days before the occurrence of sudden high waves, high pressure stayed over the east coast of Korea and wind on the east coast of Korea was very weak, but wind in the northwest of Japan was strong. The weather in the Korean peninsula was mild in those days. As high pressure moved to the east, the weather condition changed rapidly, so wind speed in the East Sea increased. Low pressure passed to the east side before the wave height grew. There are two peaks in the second mode of CSLV for *H*. The first peak seemed to be generated by low pressure and easterly wind blowing for 1 day, whereas the second peak seemed to be caused by strong wind. Fig. 3.18 represents the evolution of spatial patterns of the third mode. Variation of sea level pressure was very small over the period. Wind blew strongly

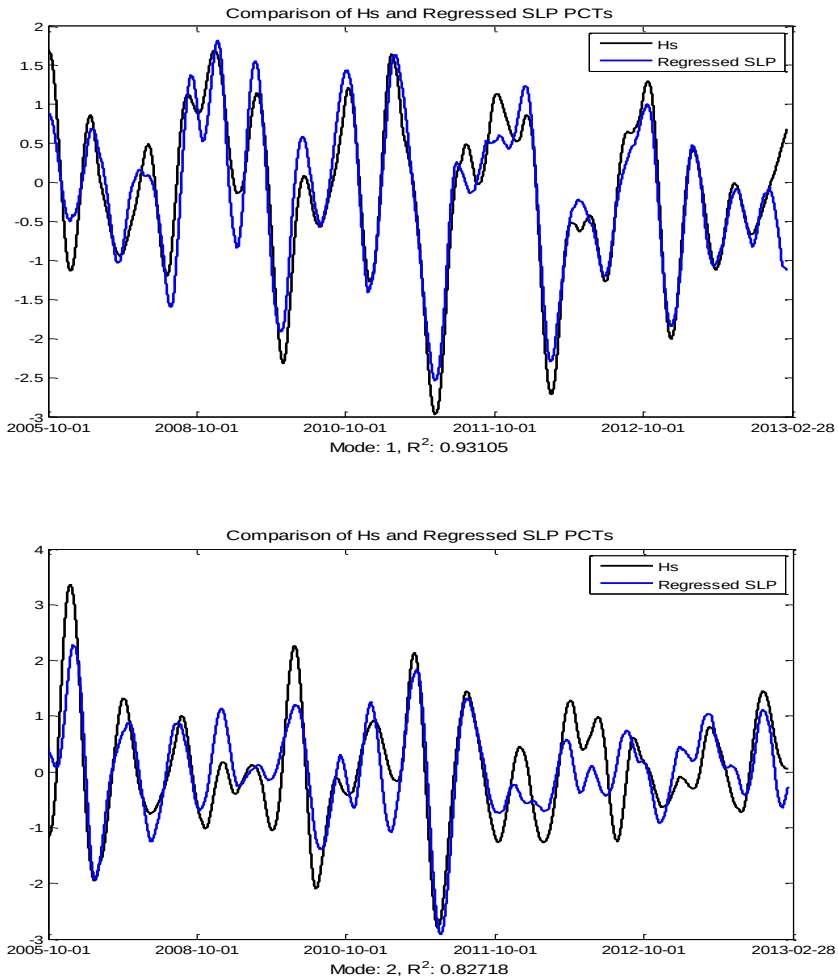
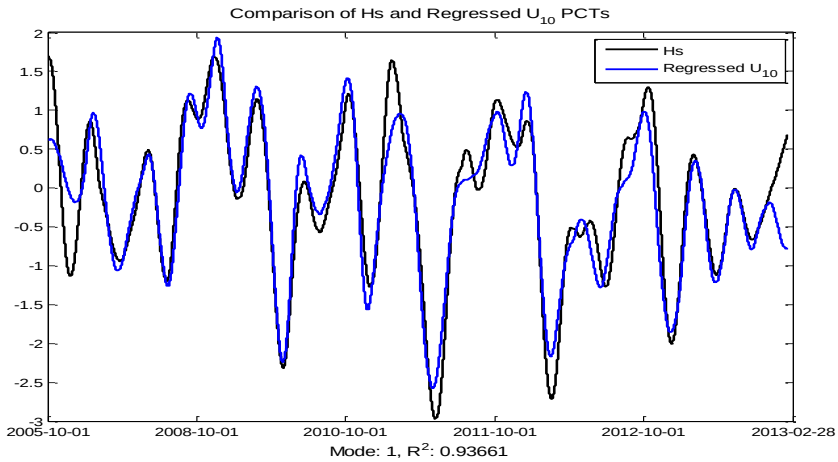


Fig. 3.15 Comparison of the first 2 modes of the regressed sea level pressure PC time series and significant wave height PC time series



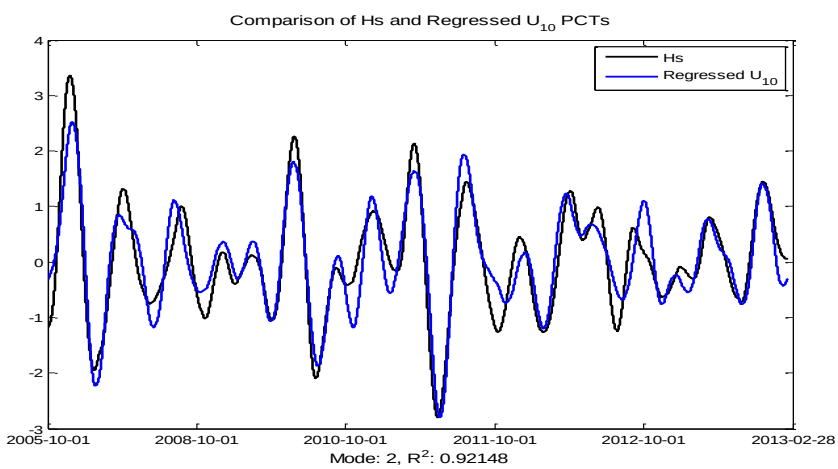


Fig. 3.16 Comparison of the first 2 modes of the regressed wind speed PC time series and significant wave height PC time series

over the period, but there was a spatial variation before the rapid increase of the wave height. Wind on the east coast of Korea was very weak, but wind far from the coast was strong. The first peak of the third mode of wave height seemed to be come from the northern part of the East Sea. After the arrival of the first peak of the wave height, wind blew strongly over the East Sea. The second peak seemed to be affected by strong wind near Gangneung.

Fig. 3.19 and Fig. 3.20 show comparisons of the first two modes of the regressed sea level pressure PC time series and wind speed PC time series with the index of sudden high waves, H^2L, respectively. The physical processes of sudden high waves and the corresponding processes of meteorological variables were compared by multiplying the regression coefficients to the CSLVs of meteorological variables. Fig. 3.21 shows the evolution of the spatial patterns of the first mode of sudden high wave index and meteorological variables. The first mode of sudden high wave index has negative values, so sudden high waves will occur when it combines with the negative PC time series. Until the rapid growth of the index, there was spatial variation in wind speed. Wind near the east coast of Korea was very weak, but wind in the northeast part of the East Sea was strong. As the spatial variation reduces gradually, the wind speed increased near the east coast of Korea and the value of index grew rapidly. After the increase of the magnitude of the index, the wind speed decreased gradually on the east coast, but wind in the offshore area blew continuously. It was inferred that the first peak of the index seemed to be affected by the wind speed on the east coast to increase and the second peak seemed to be influenced by wind in the offshore area. For the second mode of sudden high wave index, there was no distinct change in the spatial patterns of meteorological variables during the period.

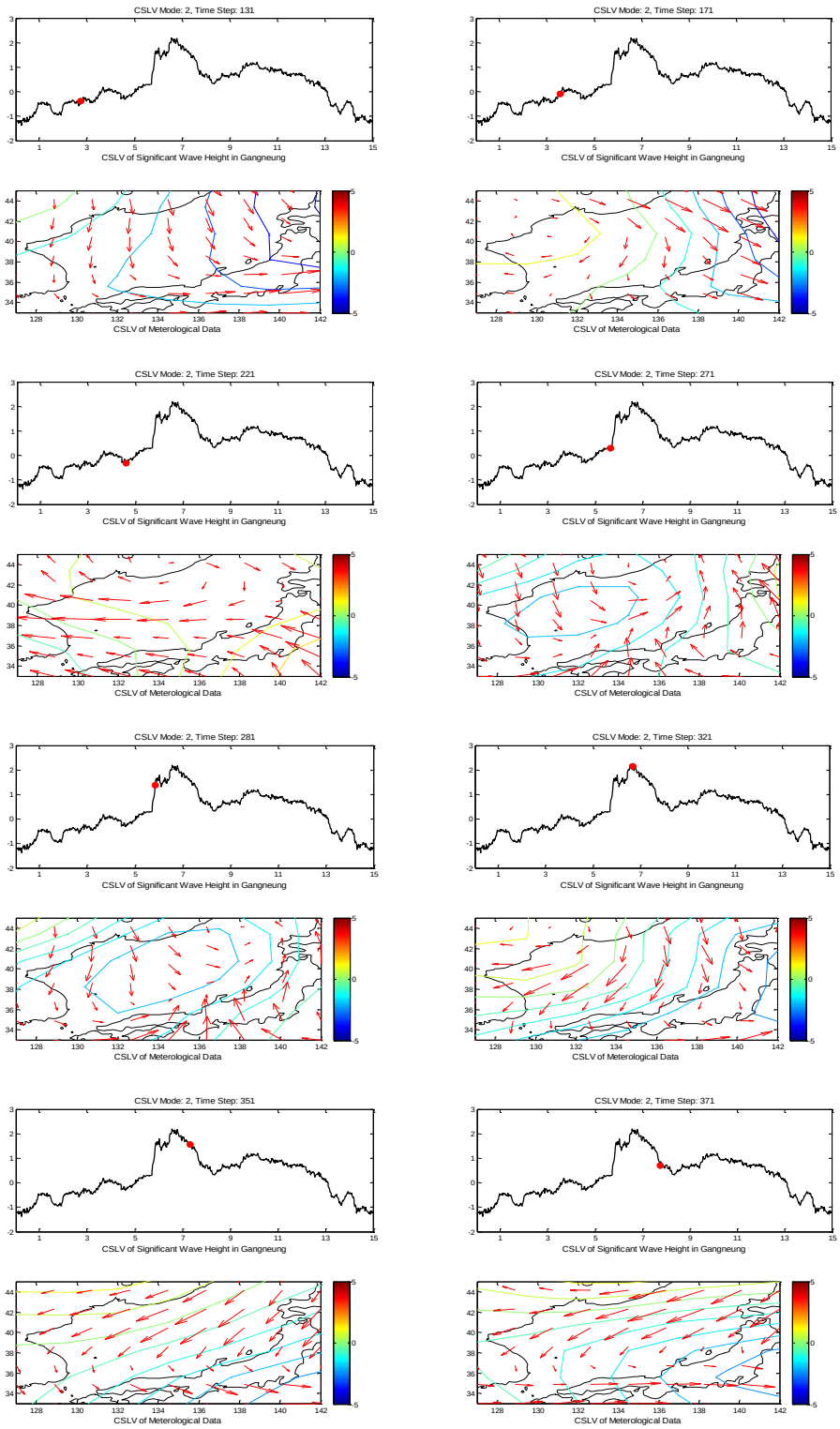


Fig. 3.17 Evolution of spatial patterns for the 2nd mode of significant wave height and meteorological variables

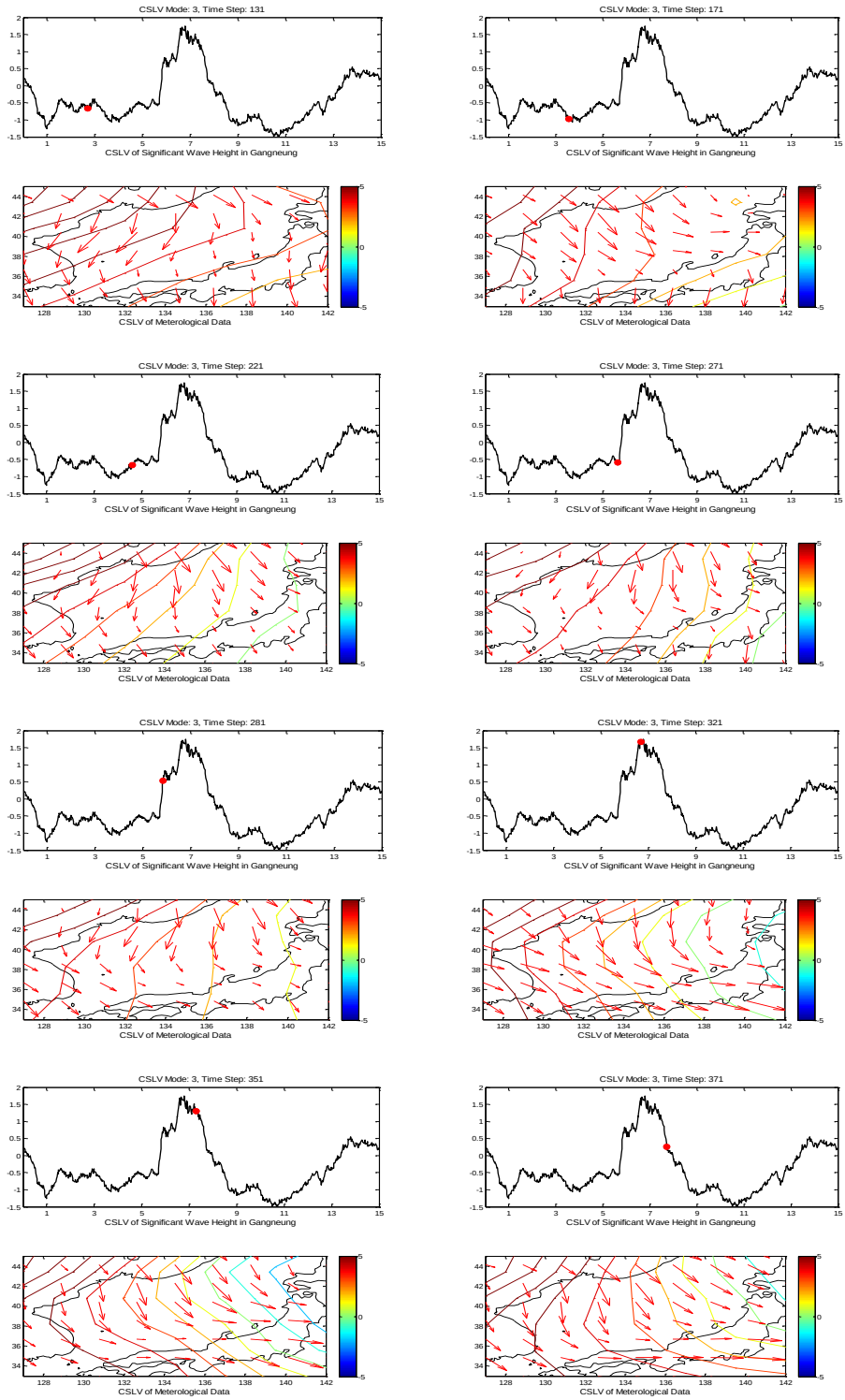
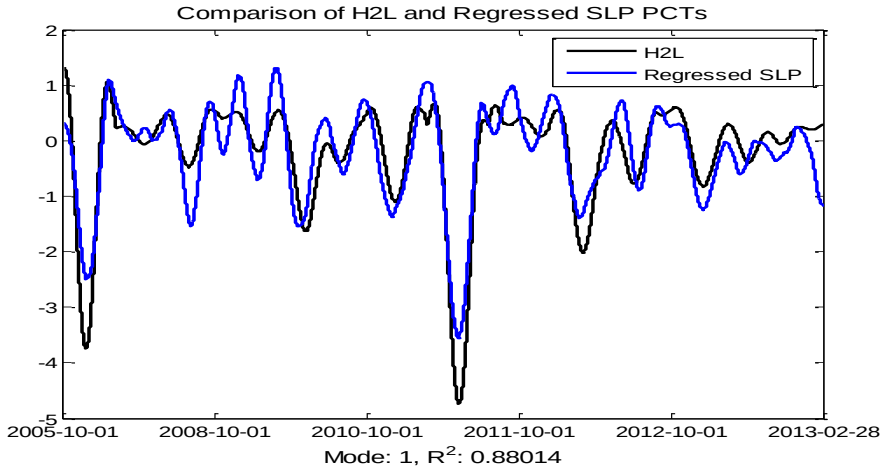


Fig. 3.18 Evolution of spatial patterns for the 3rd mode of significant wave height and meteorological variables



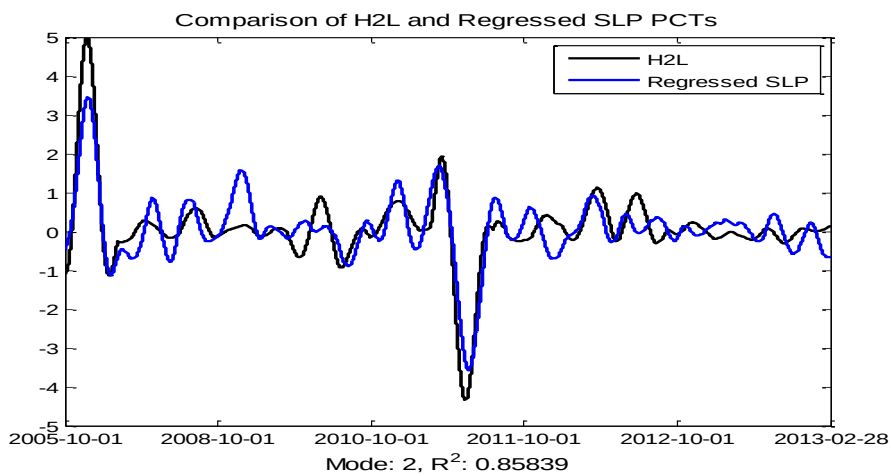
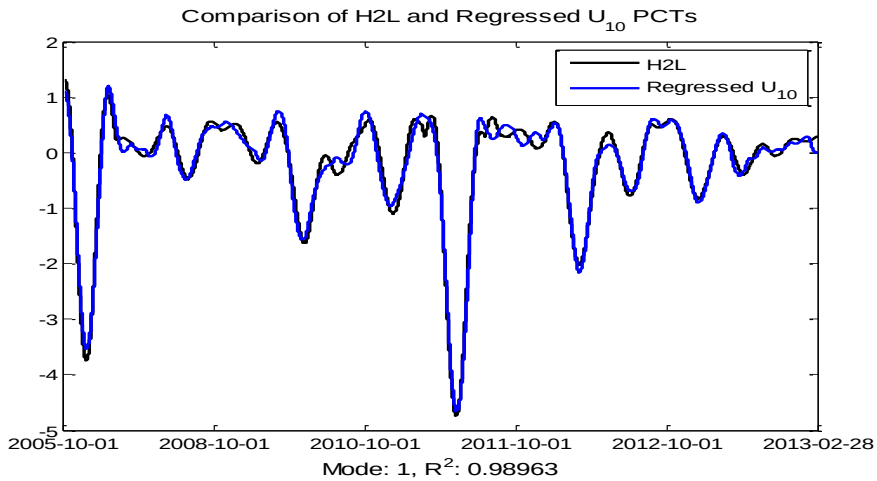


Fig. 3.19 Comparison of the first 2 modes of the regressed sea level pressure PC time series and sudden high wave index PC time series



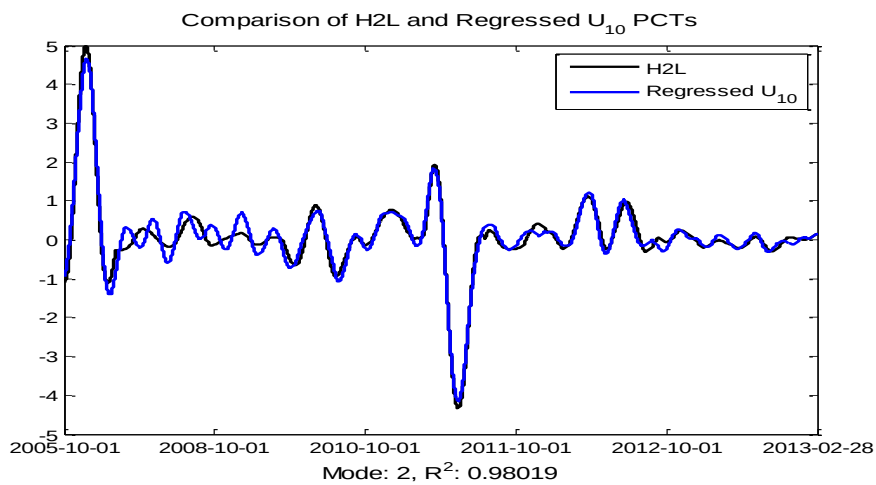


Fig. 3.20 Comparison of the first 2 modes of the regressed wind speed PC time series and sudden high wave index PC time series

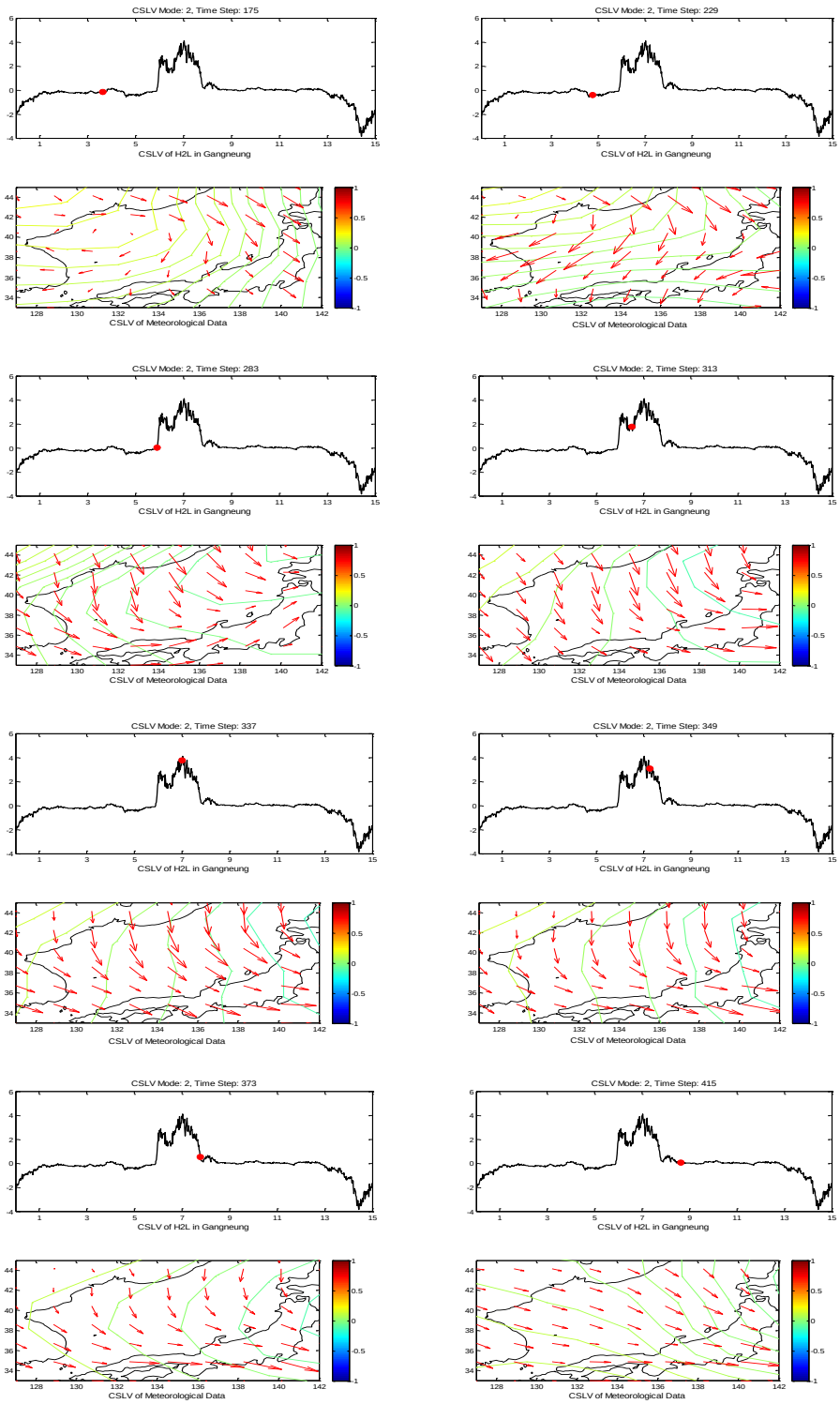


Fig. 3.21 Evolution of spatial patterns for the 1st mode of sudden high wave index and meteorological variables

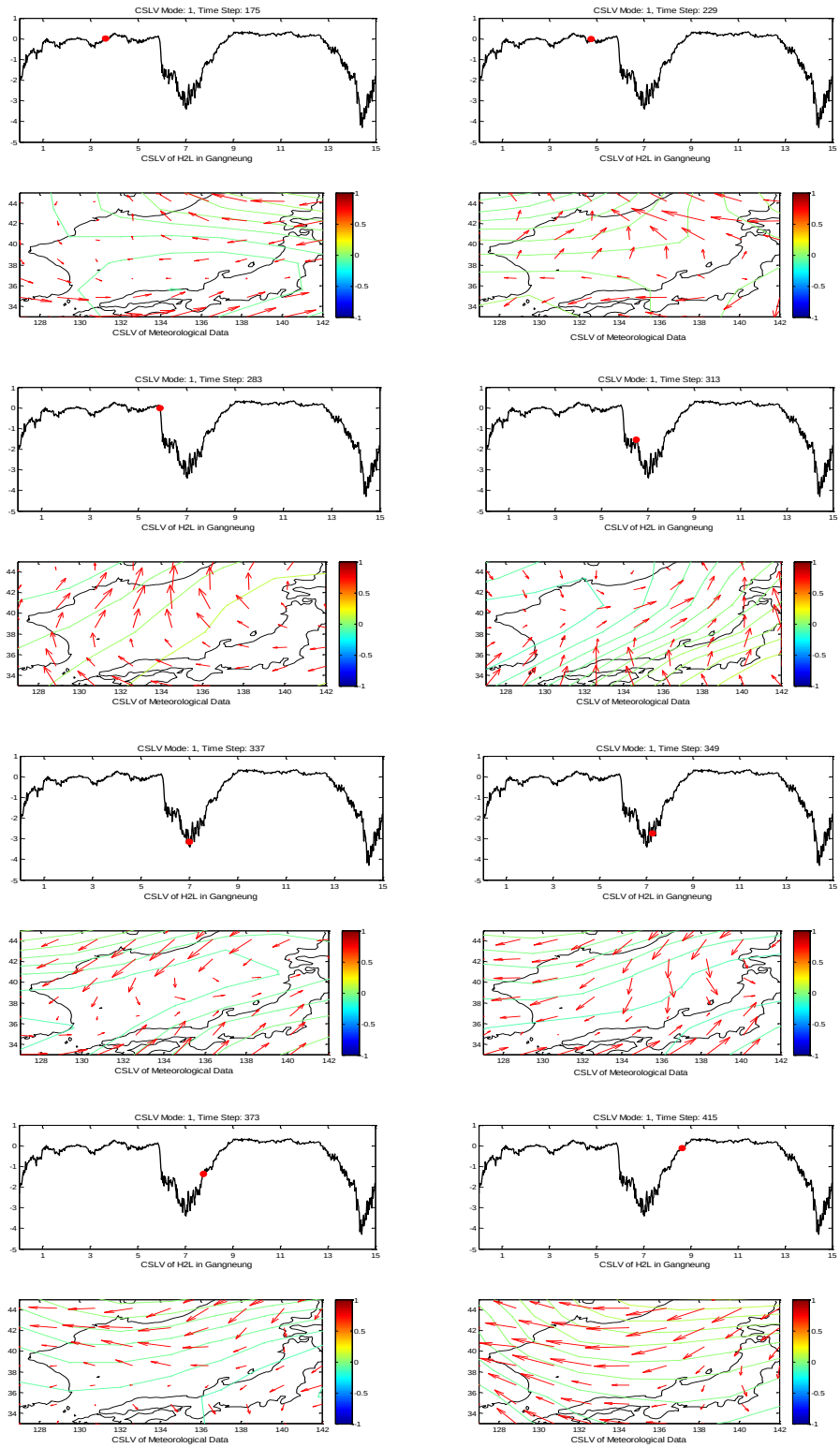


Fig. 3.22 Evolution of spatial patterns for the 2nd mode of sudden high wave index and meteorological variables

CHAPTER 4. FORECASTING OF SUDDEN HIGH WAVES

4.1 Data for forecasting of sudden high waves

4.1.1 Observed wave data

The wave data used in this study were observed at two different locations (Gangneung and Wangdolcho) provided by KIOST (Korea Institute of Ocean Science and Technology) and at six different buoys (Rumoi, Aomori, Sakata, Fukui, Tottori, and Genkainada) by NOWPHAS (Nationwide Ocean Wave information network for Ports and HArbourS). Fig. 4.1 shows the locations of wave measurement and Table 4.1 represents the information of the stations. The wave data of KIOST and NOWPHAS were measured every 0.5 s using pressure gauge and sampling rate is $\Delta f \approx 0.008\,Hz$. Wave spectrum was calculated every 30 minutes from the collected wave series, whereas the wave data of NOWPHAS were calculated every 20 minutes. Therefore, the latter were converted into 30 minutes. The data at every hour were used as they were, but the data at 30 minutes after every hour were calculated by taking the average of the data observed at 20 and 40 minutes after the hour. Experiments were conducted for the winter season from October for February, when sudden high waves occur frequently. The period is 1 years in 2010-2011 (Oct. 2010-Feb. 2011).

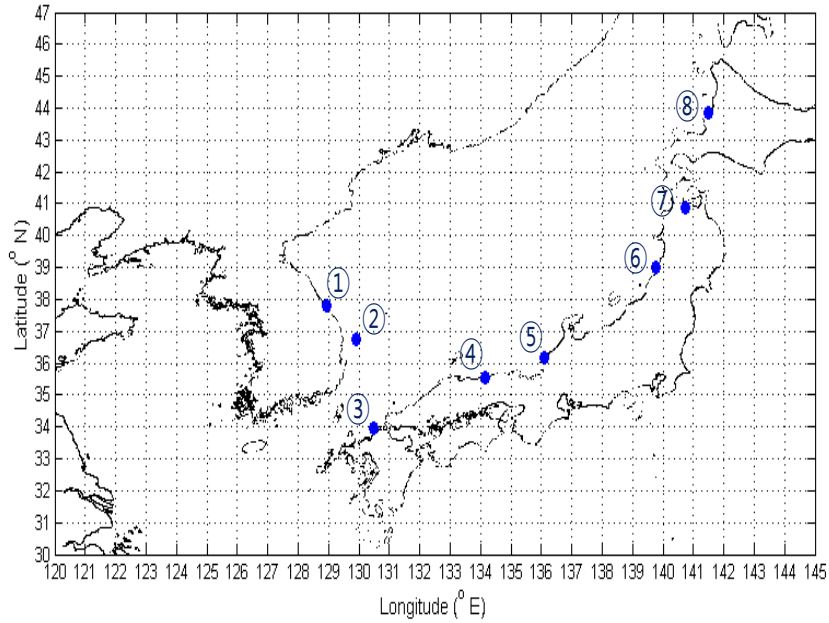


Fig. 4.1 Wave measurement locations

Table 4.1 Information on wave measurement stations and statistical properties of wave data at each station

Station	No.	Location	Waver depth (m)	Min (m)	Max (m)	Mean (m)
Gangneung	1	128° 55′ 43.2″ E, 37° 47′ 50.8″ N	15.0	0.05	5.65	0.79
Wangdolcho	2	129° 43′ 52.9″ E, 36° 43′ 10.3″ N	15.3	0.06	5.89	1.10
Genkainada	3	130° 28′ 05″ E, 33° 56′ 02″ N	39.5	0.16	5.76	1.34
Tottori	4	134° 09′ 41″ E, 35° 33′ 16″ N	30.9	0.13	5.12	1.61
Fukui	5	136° 04′ 30″ E, 36° 09′ 50″ N	36.7	0.13	6.75	1.83
Sakata	6	139° 46′ 45″ E, 39° 00′ 31″ N	45.9	0.14	8.64	1.87
Aomori	7	140° 44′ 21″ E, 40° 51′ 10″ N	24.9	0.09	1.89	0.32
Rumoi	8	141° 28′ 07″ E, 43° 51′ 59″ N	49.8	0.12	5.74	1.46

4.1.2 Numerical wave modeling data

The numerical wave modeling results provided by KORDI (Korea Ocean Research & Development Institute) were used to analyze the spatial patterns of sudden high waves. KORDI (2005) produced long-term wave data in 1979-2003. They simulated wave modeling for typhoon and non-typhoon period separately. The HYybrid PArametrical wave prediction (HYPA) model and a third generation wave model (WAM) model were used for non-typhoon and typhoon period, respectively. The wind field used in the continuous wave simulation is the reanalyzed wind data conducted by European Midrange Weather Forecast (ECMWF) (Lee and Jun 2006). Temporal resolution is 1 hour interval and spatial resolutions are about 18 km (1/6 °) both in longitude and latitude direction. Fig. 4.2 shows the grid map of KORDI (2005). Total number of grid in the region is 91×103, 9373. However, to reduce the number of grids, data were picked with a spacing of 1 ° in the region. Then the number of grid is 16×13, 208 and for the sea area is 96 except the east coast of Japan. Fig. 4.3 shows the grid map of the wave data. Red area is land and blue area is sea.

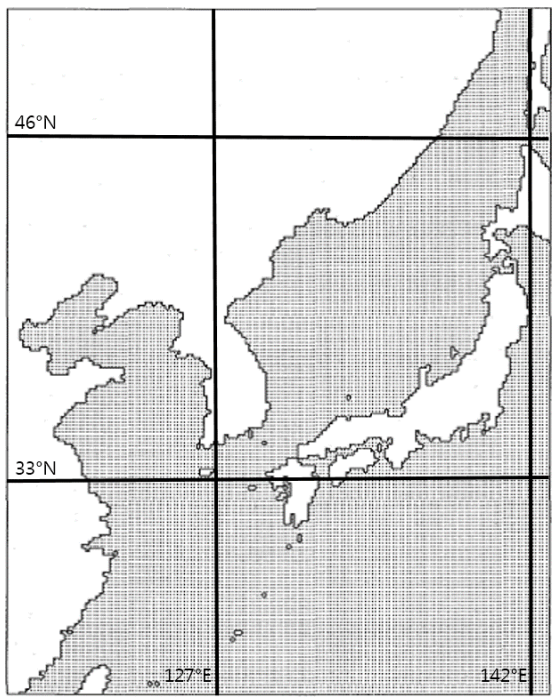


Fig. 4.2 Grid map of KORDI (2005)

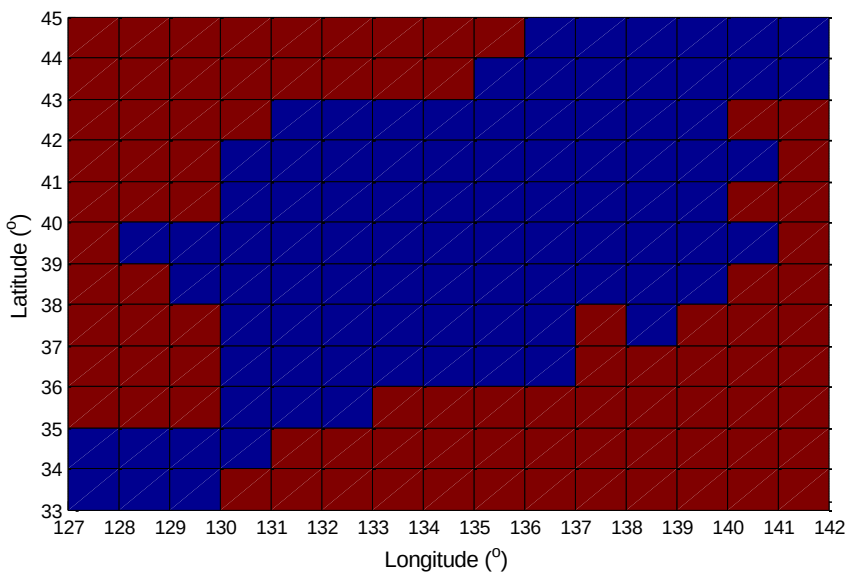


Fig. 4.3 Grid map of numerical wave modeling data

4.1.3 Meteorological data for forecasting

To forecast wave height (or period), the meteorological data used in this study were NCEP/NCAR reanalysis data provided by NOAA/CDC (Kalnay et al. 1996). The sea level pressures and 10 m height wind speeds in the *u*-direction (east-west) and *v*-direction (north-south) were also used. The temporal resolution is 30-min interval interpolated from 6 hour interval data. The region of meteorological data is same as Fig. 3.2, which is 127°-142° E, 33°-46° N.

In order to forecast significant wave series using numerical wave modeling data, the ERA-Interim reanalysis data provided by European Center for Midrange Weather Forecasts (ECMWF) were used (Dee et al. 2011), which were 10 m height wind speeds in the *u*-direction (east-west) and *v*-direction (north-south).This dataset can be downloaded from a website (http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/). The temporal resolution is 1 hour interval interpolated from 6 hour interval data, and the spatial resolutions are N128 reduced Gaussian grid, which are 0.703125 ° × 0.703125 ° grid. The region of the data is same as Fig. 3.2.

4.2 Forecasting of significant wave height and period using the observed data

In recent years, hybridization of ANN with other techniques has been used in wave height forecasting to overcome the limitation of ANN and to provide effective modeling. In this study, an EOFWNN model is developed by combining the EOF analysis and wavelet analysis with the neural network to forecast significant wave heights (or period) in multiple locations simultaneously. The input data of the model are the past wave height (or period) and the past and future meteorological reanalysis data in all the locations. The wave data used in this study were provided by KIOST and NOWPHAS and the meteorological data were the NCEP/NCAR reanalysis data. The model then calculates the wave heights (or periods) in the locations simultaneously for various lead times. The developed model was employed to forecast significant wave heights (or periods) at eight wave observation stations in the coast waters around the East Sea. Experiments were conducted for winter season from October to February. The period used in this study was from October 2010 to February 2011.

The first stage of the EOFWNN model was the EOF analysis of the wave and meteorological data. Through the EOF analysis, the input time series were decomposed into several modes and separated into spatial and temporal components. Fig. 4.4 and Fig. 4.5 show the first four modes of the eigenvectors of wind velocity and sea level pressure, respectively. They represent the physical patterns of the spatial distribution of the meteorological data. Fig. 4.6 and Fig. 4.7 show the corresponding PC time series. Significant wave height series were decomposed into 8 modes, which corresponded 99.9% variance of the series. Fig.

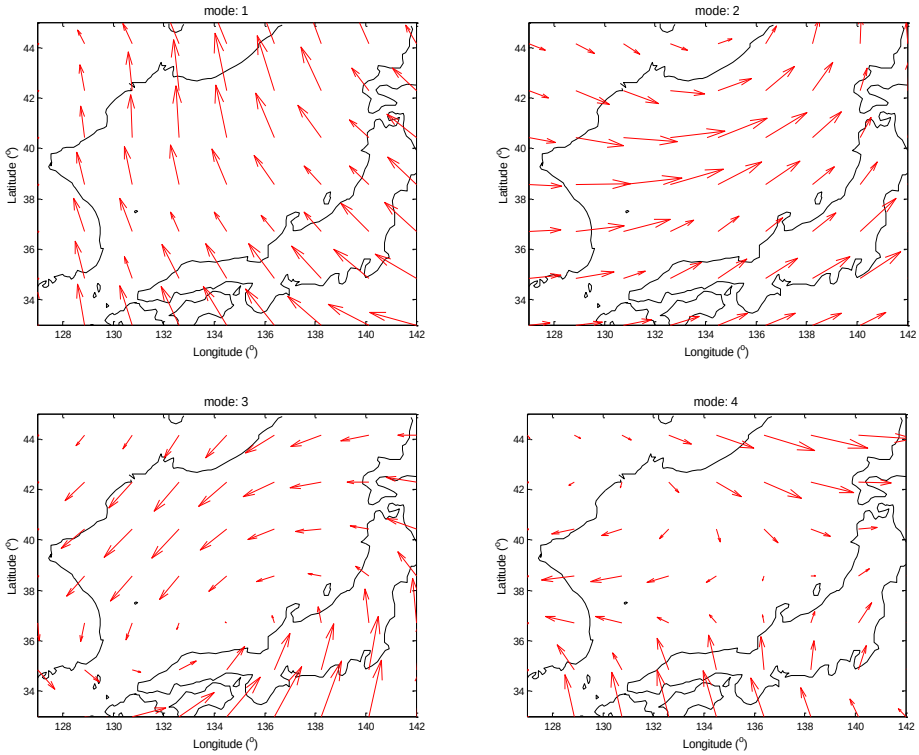


Fig. 4.4 First to fourth mode eigenvectors for wind velocity

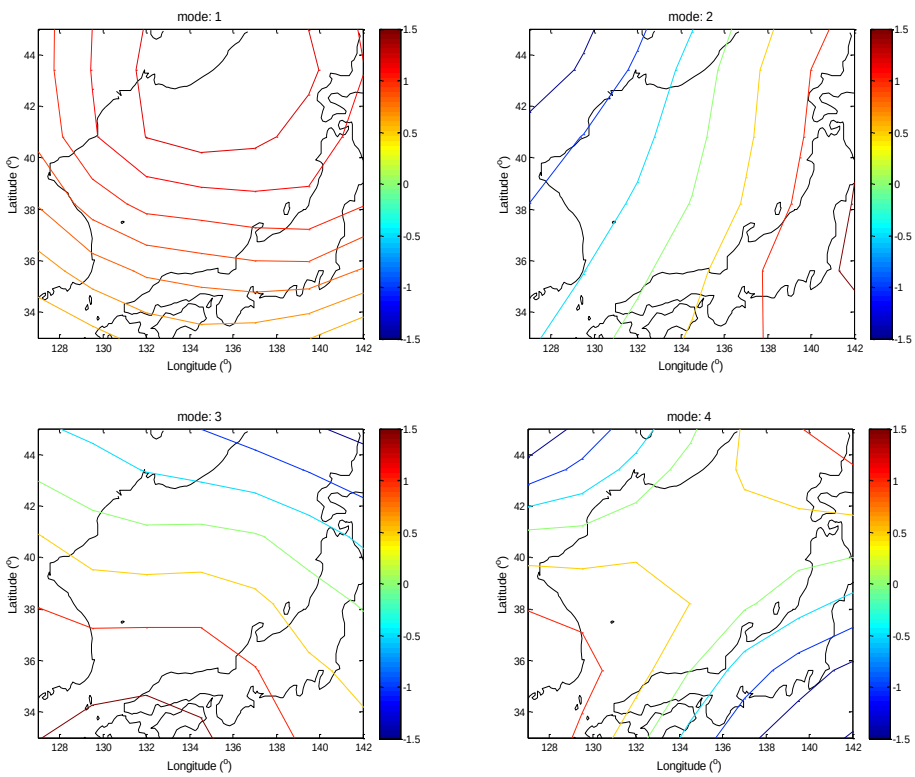


Fig. 4.5 First to fourth mode eigenvectors for sea level pressure

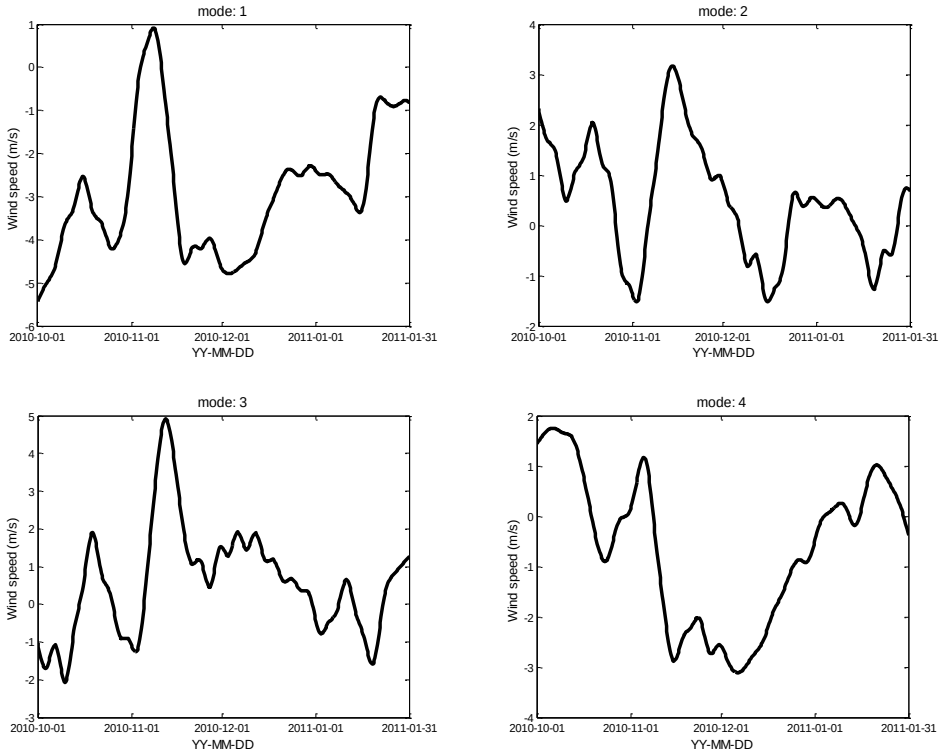


Fig. 4.6 First to fourth mode PC time series for wind velocity

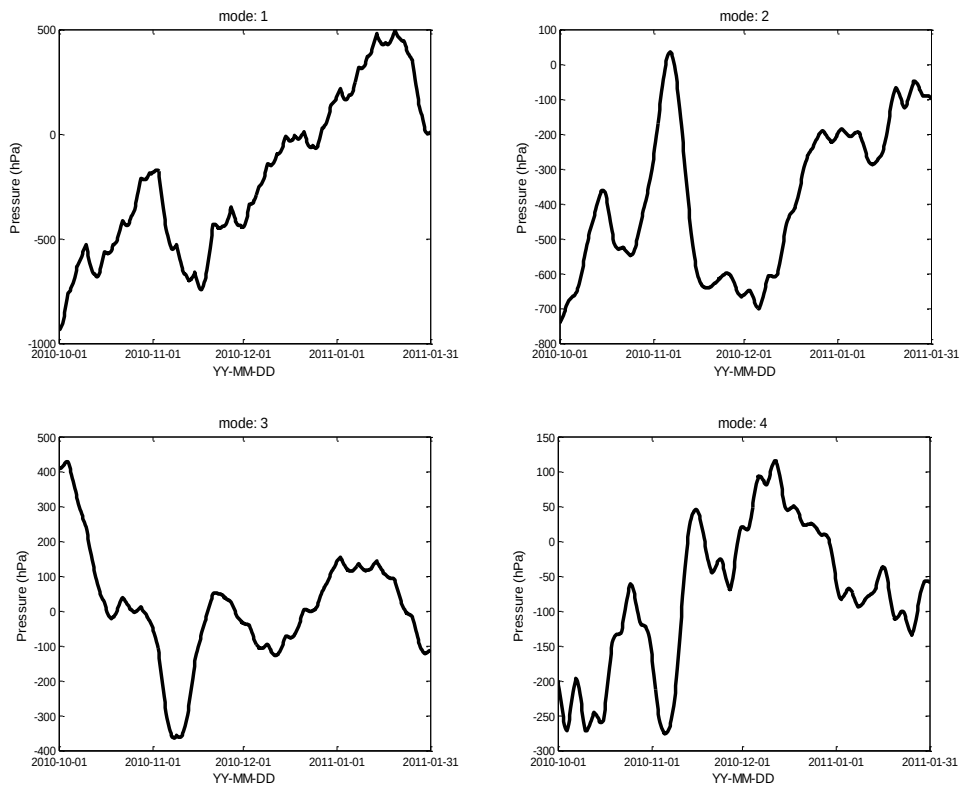


Fig. 4.7 First to fourth mode PC time series for sea level pressure

4.8 shows the normalized and cumulative eigenvalues of EOF modes for the significant wave height. Fig. 4.9 represents the first four modes of the PC time series for significant wave height.

Next, by employing the wavelet analysis for each PC time series for 3, 5, and 7 decomposition levels, several wavelet components were produced. Fig. 4.10 shows the wavelet components for the 3rd decomposition level of the first mode of the significant wave height PC time series, which consist of one approximation and three details. Fig. 4.11 shows the approximations of the 1st wind speed PC time series and sea level pressure PC time series.

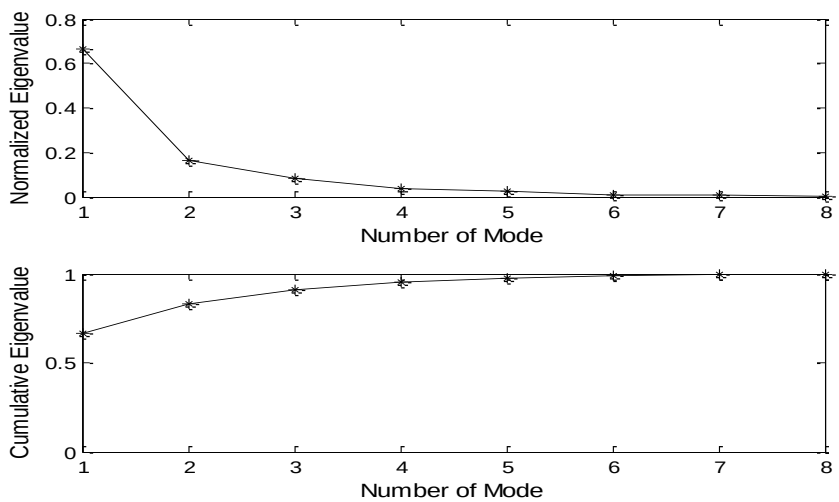


Fig. 4.8 Eigenvalues of EOF modes for significant wave height data

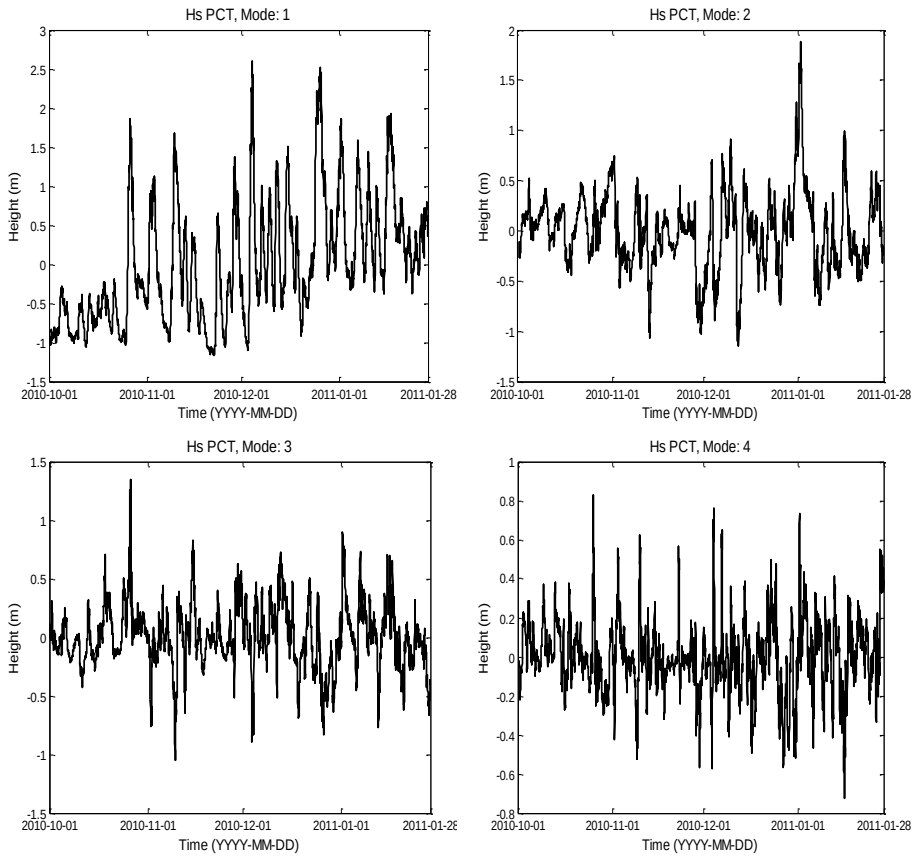


Fig. 4.9 First to fourth mode PC time series for significant wave height

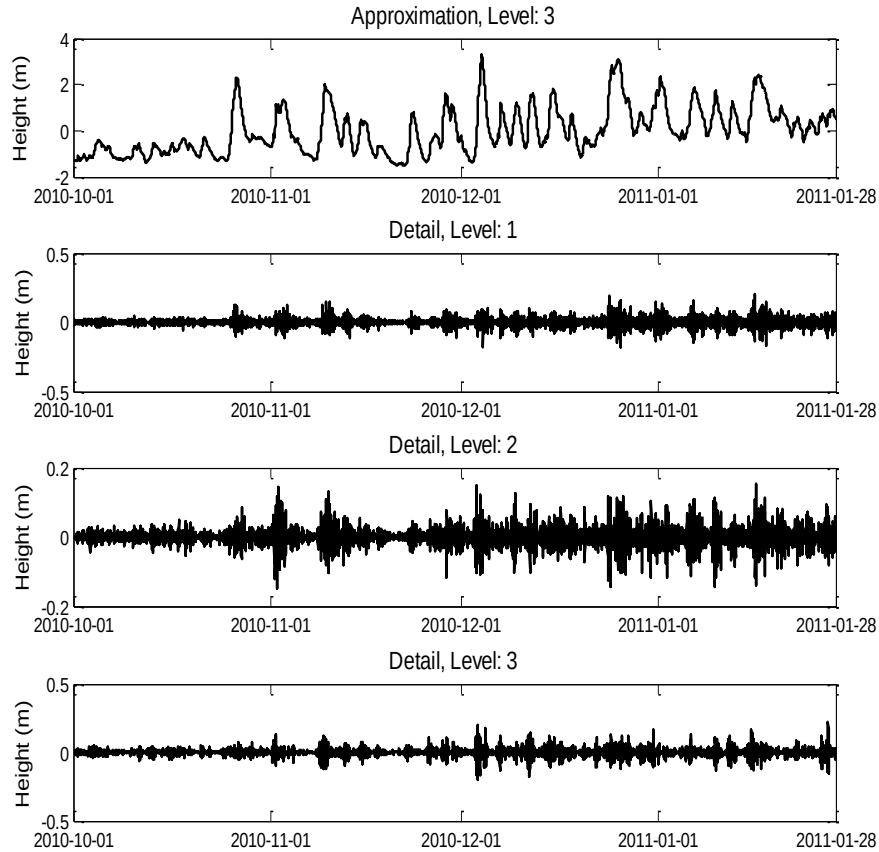


Fig. 4.10 Decomposed wavelet components of the 1st mode of *H* PC time series

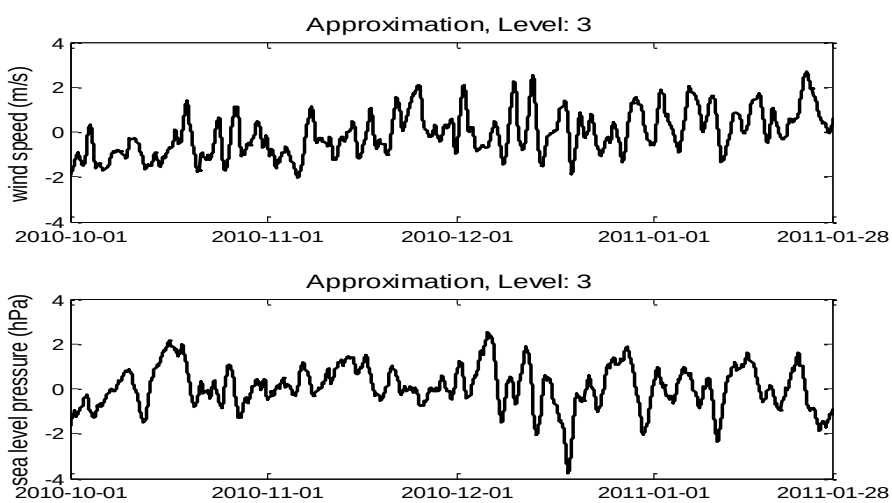


Fig. 4.11 Approximations of the 1st mode of wind speed and sea level pressure PC time series

Third, as mentioned before, the decomposed wavelet signals of wave data and meteorological data were used as the input to the ANN. Experiments were conducted for the winter season from October to February during which extratropical storms frequently occur in the East Sea. The experimental period in this study was from October 2010 to February 2011. Two models were used for forecasting real time significant wave heights. The first model is for one and three hour lead times, and the second is for 12 and 24 hour lead times. In the first model, training was conducted for 120 days (00:00 October 1, 2010 to 23:30 January 28, 2011), and testing was conducted for 7 days (00:00 January 29, 2011 to 23:30 February 4, 2011) using the weights and bias obtained from the training period. In the second model, training was conducted for 120 days before each forecast time, and testing was conducted for the consecutive 3 day period.

Fig. 4.12 shows the results of the 1st mode PC time series of significant wave height for 1 hour lead time forecasting. The upper panel is the result of the largest *R* case among the 20 ensemble members for training period and the lower panel is the ensemble members (black line) and ensemble averaged result (red line) at 00:00:00, Jan. 29, 2011. The value of *R* is close to 1.0 and *RMSE* is close to zero. Fig. 4.13-Fig. 4.15 show the results for 3, 12 and 24 hour lead time forecasting, respectively.

By reconstructing the estimated PC time series and given LV of significant wave height, the wave height series were predicted at the eight stations shown in Fig. 4.16. Figure shows the forecasted significant wave height at 00:00:00, Jan. 29, 2011 at 8 stations for several lead times.

The results of the EOFWNN model are compared with those of a wavelet and neural network hybrid (WNN) model at Gangneung, Sakata and Aomori wave

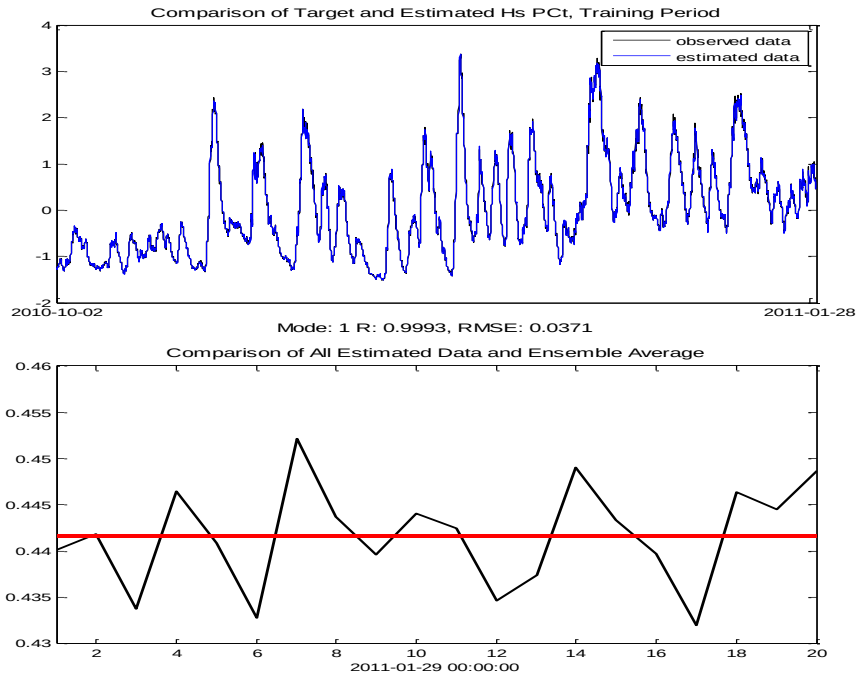


Fig. 4.12 Largest *R* case for training period (Oct. 02, 2010 - Jan. 28, 2011) and ensemble members (black) and ensemble average (red) (00:00:00, Jan. 29, 2011) for the 1st mode *H* PC time series for 1 hour lead time forecasting

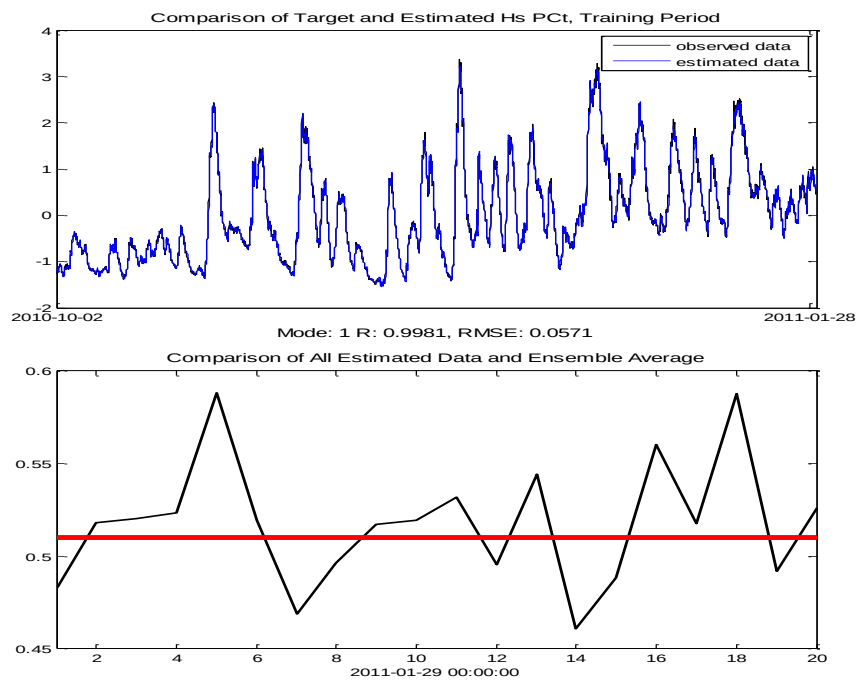


Fig. 4.13 Largest *R* case for training period (Oct. 02, 2010 - Jan. 28, 2011) and ensemble members (black) and ensemble average (red) (00:00:00, Jan. 29, 2011) for the 1st mode *H* PC time series for 3 hour lead time forecasting

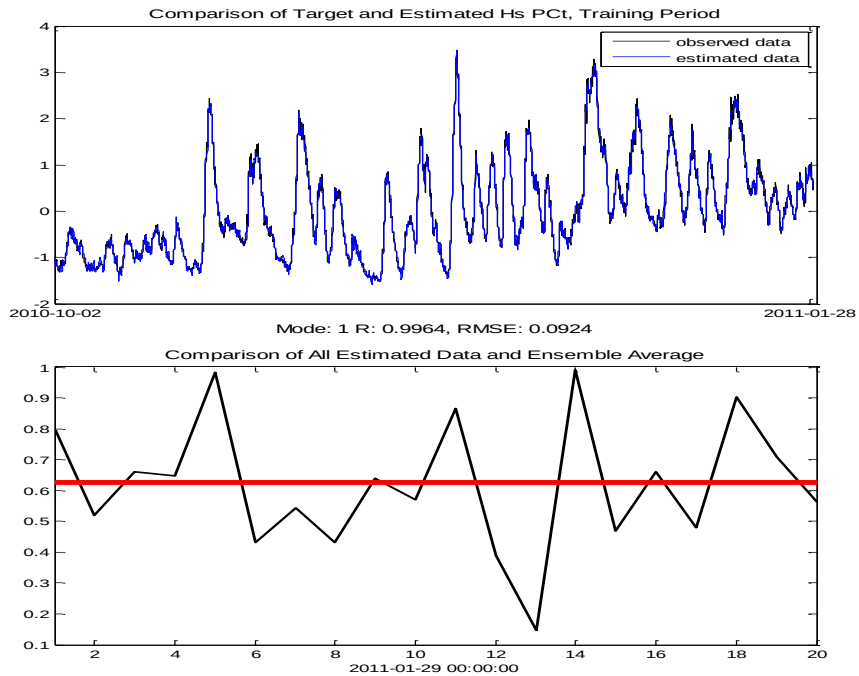


Fig. 4.14 Largest *R* case for training period (Oct. 02, 2010 - Jan. 28, 2011) and ensemble members (black) and ensemble average (red) (00:00:00, Jan. 29, 2011) for the 1st mode *H* PC time series for 12 hour lead time forecasting

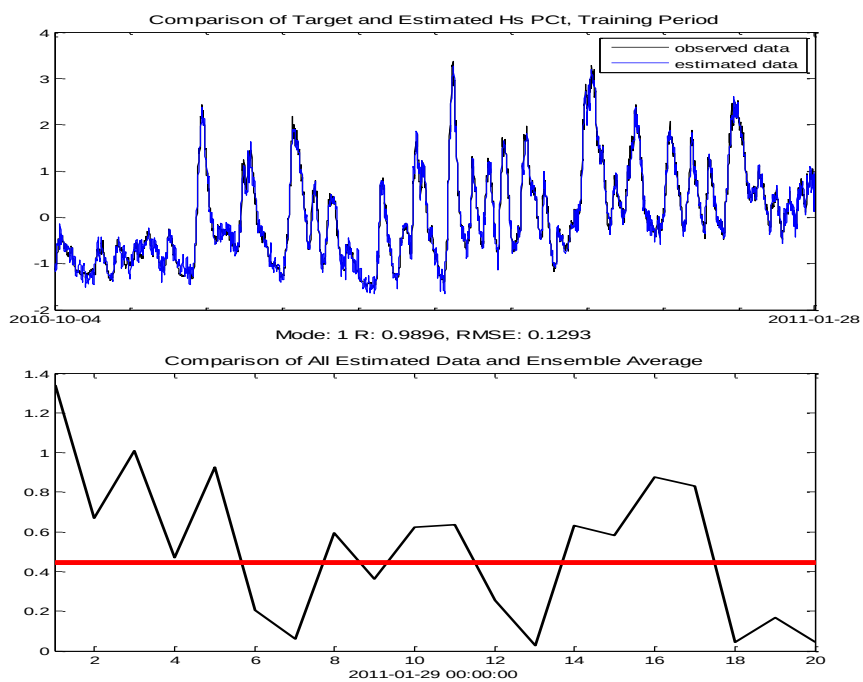


Fig. 4.15 Largest *R* case for training period (Oct. 03, 2010 - Jan. 27, 2011) and ensemble members (black) and ensemble average (red) (00:00:00, Jan. 29, 2011) for the 1st mode *H* PC time series for 24 hour lead time forecasting

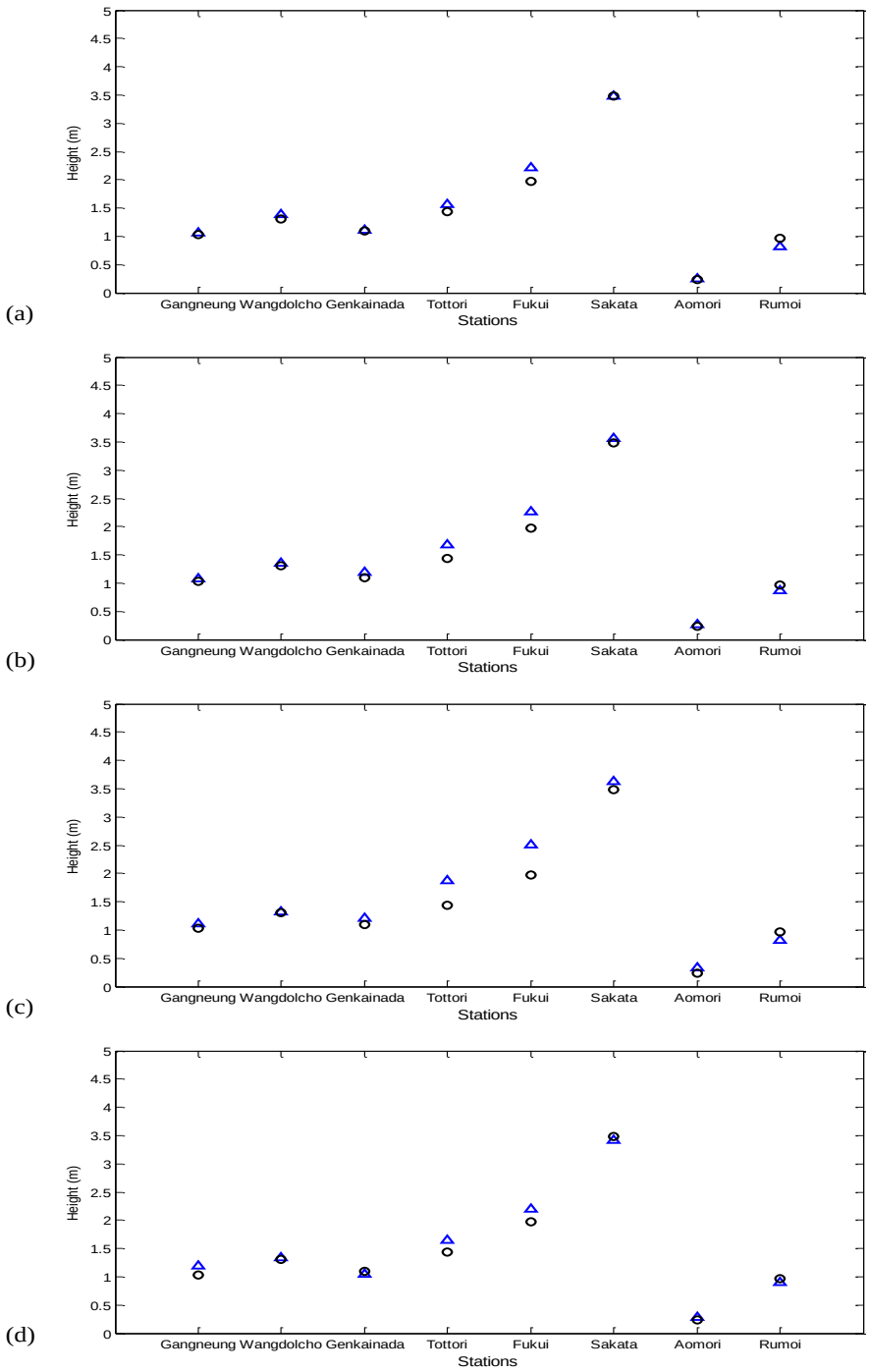


Fig. 4.16 Comparison of observed (black circle) and estimated (blue triangle) wave height at 8 stations at 00:00:00, Jan. 29, 2011 (a) for 1 hour lead time, (b) for 3 hour lead time, (c) for 12 hour lead time, (d) for 24 hour lead time forecasting

stations, which is developed and run for each location separately. The first two stations are located at the open coasts on the Korean side and Japanese side, respectively, with similar latitudes, whereas the Aomori station is located inside a bay where the wave heights are relatively small. Only decomposed wavelet components of significant wave heights were used as the input data to the WNN model, which were 3, 5 and 7 decomposed levels. To evaluate the performance of the models, three different performance indices are employed, which are correlation coefficient (*R*), index of agreement (I_a) and normalized root mean square error (*NRMSE*) calculated by Eqs. (4.1) to (4.3).

$$R = \sqrt{1 - \frac{\sum_i \left(x_i - y_i\right)^2}{\sum_i \left(x_i - \bar{x}\right)^2}} \tag{4.1}$$

$$I_a = 1 - \frac{\sum_i \left(x_i - y_i\right)^2}{\sum_i \left(\left|y_i - \overline{y_i}\right| + \left|x_i - \overline{x_i}\right|\right)^2} \tag{4.2}$$

$$NRMSE = \sqrt{\frac{\sum_i \left(x_i - y_i\right)^2}{n}} / \bar{x} \tag{4.3}$$

where x_i, y_i, $\bar{x}$, $\bar{y}$ and n are the observed wave height, forecasted wave height, mean of observed wave height, mean of forecasted wave height, and the number of observations, respectively. Table 4.2 shows the results of the WNN model at Gangneung, Sakata and Aomori for several lead times. For 1 and 3 hour lead times, the values of *R* and I_a were close to one regardless of the decomposition level, and the *NRMSE* values were distributed between 0.073and 0.164. For longer lead times, the values of *R* and I_a decreased drastically and the *NRMSE* values increased. The

Table 4.2 Test results of WNN model for *H* in Gangneung, Sakata and Aomori for several lead times

Lead time	Station	Decomposition Level	R	NRMSE	Index of Agreement
1 hr	Gangneung	3	0.991	0.101	0.995
		5	0.991	0.099	0.995
		7	0.991	0.097	0.996
	Sakata	3	0.991	0.073	0.995
		5	0.990	0.074	0.995
		7	0.991	0.074	0.995
	Aomori	3	0.971	0.128	0.985
		5	0.971	0.127	0.985
		7	0.972	0.125	0.986
3 hr	Gangneung	3	0.982	0.145	0.990
		5	0.987	0.121	0.993
		7	0.987	0.117	0.994
	Sakata	3	0.987	0.088	0.993
		5	0.989	0.079	0.994
		7	0.990	0.077	0.995
	Aomori	3	0.952	0.164	0.974
		5	0.957	0.153	0.978
		7	0.962	0.146	0.981
12 hr	Gangneung	3	0.448	0.305	0.667
		5	0.851	0.177	0.910
		7	0.882	0.161	0.934
	Sakata	3	0.374	0.333	0.622
		5	0.748	0.208	0.849
		7	0.868	0.141	0.927
	Aomori	3	-0.256	0.446	0.086
		5	0.293	0.340	0.538
		7	0.522	0.238	0.708
24 hr	Gangneung	3	-0.463	0.391	0.134
		5	-0.113	0.361	0.280
		7	0.809	0.208	0.860
	Sakata	3	0.008	0.391	0.431
		5	0.212	0.357	0.559
		7	0.870	0.140	0.927
	Aomori	3	0.074	0.367	0.365
		5	0.736	0.375	0.661
		7	0.461	0.289	0.672

model prediction accuracy is reduced with increasing lead time, whereas it is enhanced with higher decomposition level. It is worthy of note that the model accuracy at Aomori is lower than those at other stations because the wave heights are relatively small at Aomori which is located inside a bay. The results obtained from the EOFWNN model at the eight stations are summarized in Tables 4.3 to 4.6 for different lead times. The performance of the EOFWNN model for short lead times (i.e. 1 and 3 hours) was similar to those of the WNN model. The values of *R* and I_a were close to one regardless of the decomposition level, and the *NRMSE* values were distributed between 0.074 and 0.265. The EOFWNN model showed high accuracy even for longer lead times (i.e. 12 and 24 hours). Even at the 3rd decomposition level, the values of *NRMSE* were under 0.226. Figs. 4.17 to 4.19 show the observed and forecasted significant wave height time series by the WNN and EOFWNN models with the 7th decomposition level at 24 hour lead time in Gangneung, Sakata and Aomori, respectively. Both models forecasted the overall behavior of the observed data, but the EOFWNN model proposed in this study showed more accurate results than the WNN model. Figures for other lead times are attached in appendix. Fig. 4.20 compares the index of agreement of the EOFWNN and WNN models corresponding to different decomposition levels for different lead times at Gangneung, Sakata and Aomori. Both models show lower accuracy with increasing lead time and higher accuracy with increasing decomposition level. The accuracy of the WNN model with low decomposition level drastically declined for longer lead times. However, the accuracy of the EOFWNN model was not significantly affected by the decomposition level. The accuracy of the EOFWNN model was much better than that of the WNN model for longer lead times.

Table 4.3 Test results of EOFWNN model for *H* at 1 hr lead time depending on decomposition level

Station	Decomposition Level	R	NRMSE	Index of Agreement
Gangneung	3	0.992	0.096	0.996
	5	0.991	0.097	0.996
	7	0.992	0.098	0.995
Wangdolcho	3	0.988	0.103	0.994
	5	0.988	0.104	0.994
	7	0.988	0.106	0.993
Genkainada	3	0.989	0.080	0.995
	5	0.990	0.078	0.995
	7	0.989	0.080	0.995
Tottori	3	0.990	0.080	0.995
	5	0.990	0.079	0.995
	7	0.990	0.081	0.995
Fukui	3	0.991	0.080	0.995
	5	0.991	0.081	0.995
	7	0.990	0.084	0.995
Sakata	3	0.990	0.074	0.995
	5	0.990	0.075	0.995
	7	0.990	0.080	0.994
Aomori	3	0.967	0.135	0.983
	5	0.967	0.136	0.983
	7	0.967	0.136	0.983
Rumoi	3	0.956	0.080	0.978
	5	0.958	0.078	0.979
	7	0.956	0.079	0.978

Table 4.4 Test results of EOFWNN model for *H* at 3 hr lead time depending on decomposition level

Station	Decomposition Level	R	NRMSE	Index of Agreement
Gangneung	3	0.979	0.164	0.988
	5	0.987	0.125	0.993
	7	0.987	0.125	0.992
Wangdolcho	3	0.981	0.139	0.989
	5	0.988	0.105	0.994
	7	0.986	0.129	0.990
Genkainada	3	0.980	0.124	0.986
	5	0.988	0.088	0.993
	7	0.987	0.091	0.993
Tottori	3	0.988	0.091	0.994
	5	0.988	0.087	0.994
	7	0.988	0.089	0.994
Fukui	3	0.989	0.091	0.994
	5	0.990	0.086	0.995
	7	0.989	0.093	0.993
Sakata	3	0.987	0.094	0.992
	5	0.989	0.081	0.994
	7	0.990	0.081	0.994
Aomori	3	0.891	0.265	0.923
	5	0.953	0.160	0.975
	7	0.955	0.169	0.975
Rumoi	3	0.923	0.108	0.957
	5	0.945	0.093	0.971
	7	0.942	0.092	0.969

Table 4.5 Test results of EOFWNN model for *H* at 12 hr lead time depending on decomposition level

Station	Decomposition Level	R	NRMSE	Index of Agreement
Gangneung	3	0.938	0.126	0.964
	5	0.909	0.141	0.953
	7	0.923	0.131	0.960
Wangdolcho	3	0.893	0.121	0.944
	5	0.907	0.112	0.951
	7	0.924	0.101	0.961
Genkainada	3	0.909	0.095	0.952
	5	0.930	0.085	0.964
	7	0.945	0.077	0.971
Tottori	3	0.912	0.090	0.951
	5	0.905	0.093	0.950
	7	0.928	0.081	0.963
Fukui	3	0.935	0.097	0.962
	5	0.947	0.088	0.973
	7	0.955	0.080	0.977
Sakata	3	0.898	0.124	0.943
	5	0.959	0.080	0.979
	7	0.962	0.078	0.981
Aomori	3	0.687	0.193	0.813
	5	0.538	0.228	0.714
	7	0.696	0.201	0.826
Rumoi	3	0.863	0.143	0.921
	5	0.922	0.111	0.959
	7	0.921	0.114	0.958

Table 4.6 Test results of EOFWNN model for *H* at 24 hr lead time depending on decomposition level

Station	Decomposition Level	R	NRMSE	Index of Agreement
Gangneung	3	0.904	0.164	0.941
	5	0.924	0.138	0.957
	7	0.926	0.130	0.961
Wangdolcho	3	0.891	0.126	0.937
	5	0.861	0.141	0.924
	7	0.902	0.117	0.949
Genkainada	3	0.923	0.088	0.956
	5	0.936	0.084	0.964
	7	0.943	0.076	0.971
Tottori	3	0.893	0.103	0.937
	5	0.893	0.099	0.940
	7	0.911	0.094	0.953
Fukui	3	0.929	0.103	0.959
	5	0.938	0.094	0.967
	7	0.960	0.078	0.979
Sakata	3	0.918	0.112	0.954
	5	0.946	0.093	0.972
	7	0.958	0.083	0.978
Aomori	3	0.625	0.226	0.765
	5	0.653	0.209	0.798
	7	0.725	0.196	0.841
Rumoi	3	0.906	0.122	0.946
	5	0.900	0.123	0.945
	7	0.925	0.107	0.961

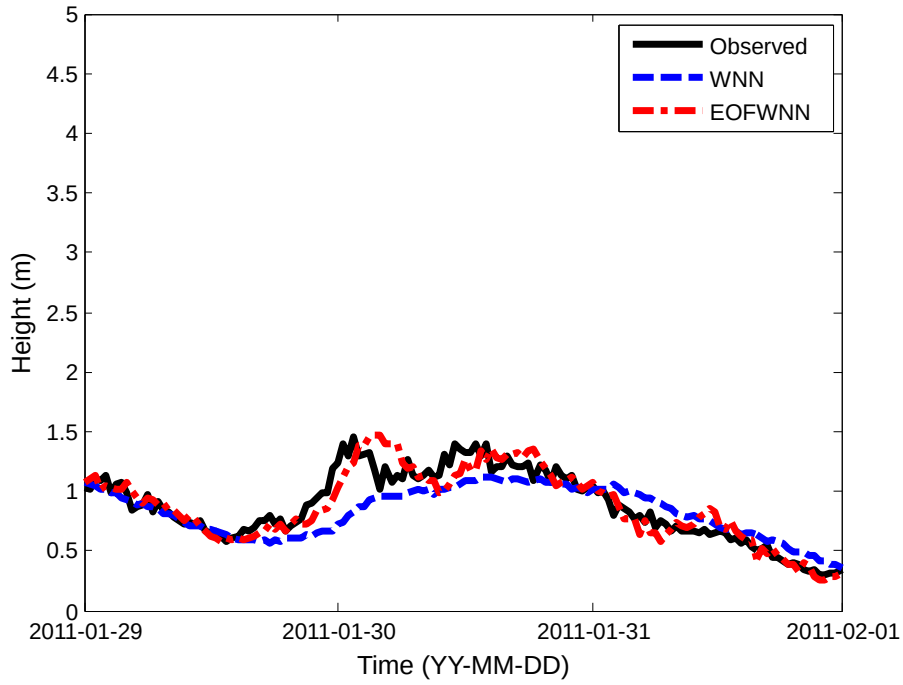


Fig. 4.17 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 24 hour lead time in Gangneung

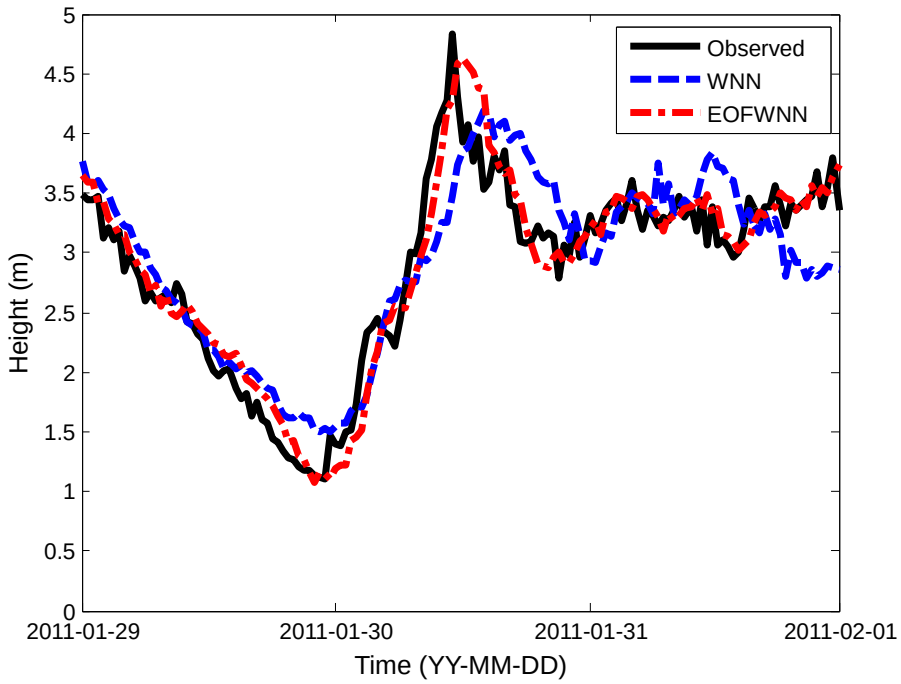


Fig. 4.18 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 24 hour lead time in Sakata

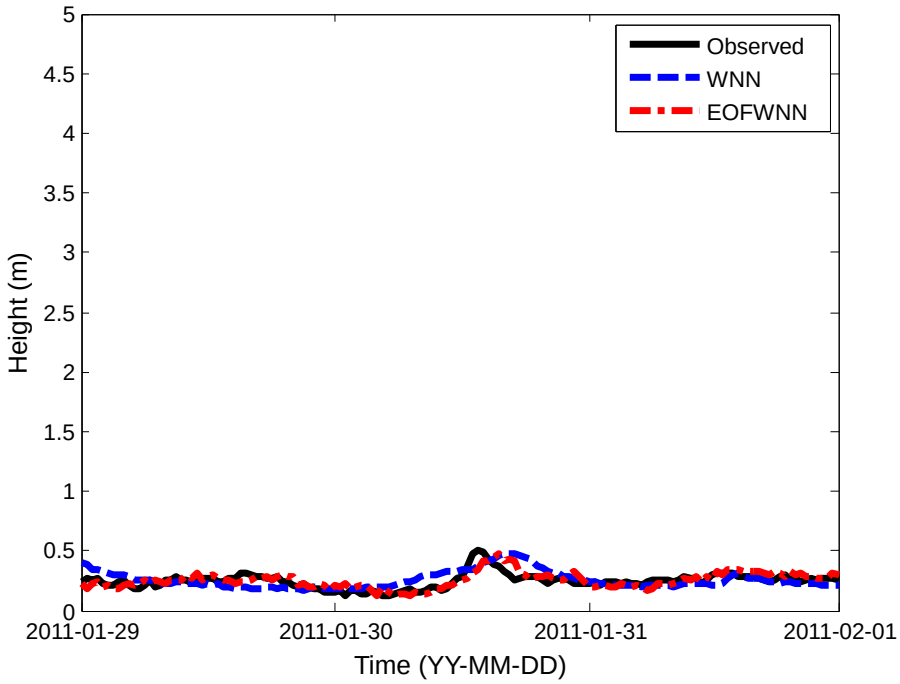


Fig. 4.19 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 24 hour lead time in Aomori

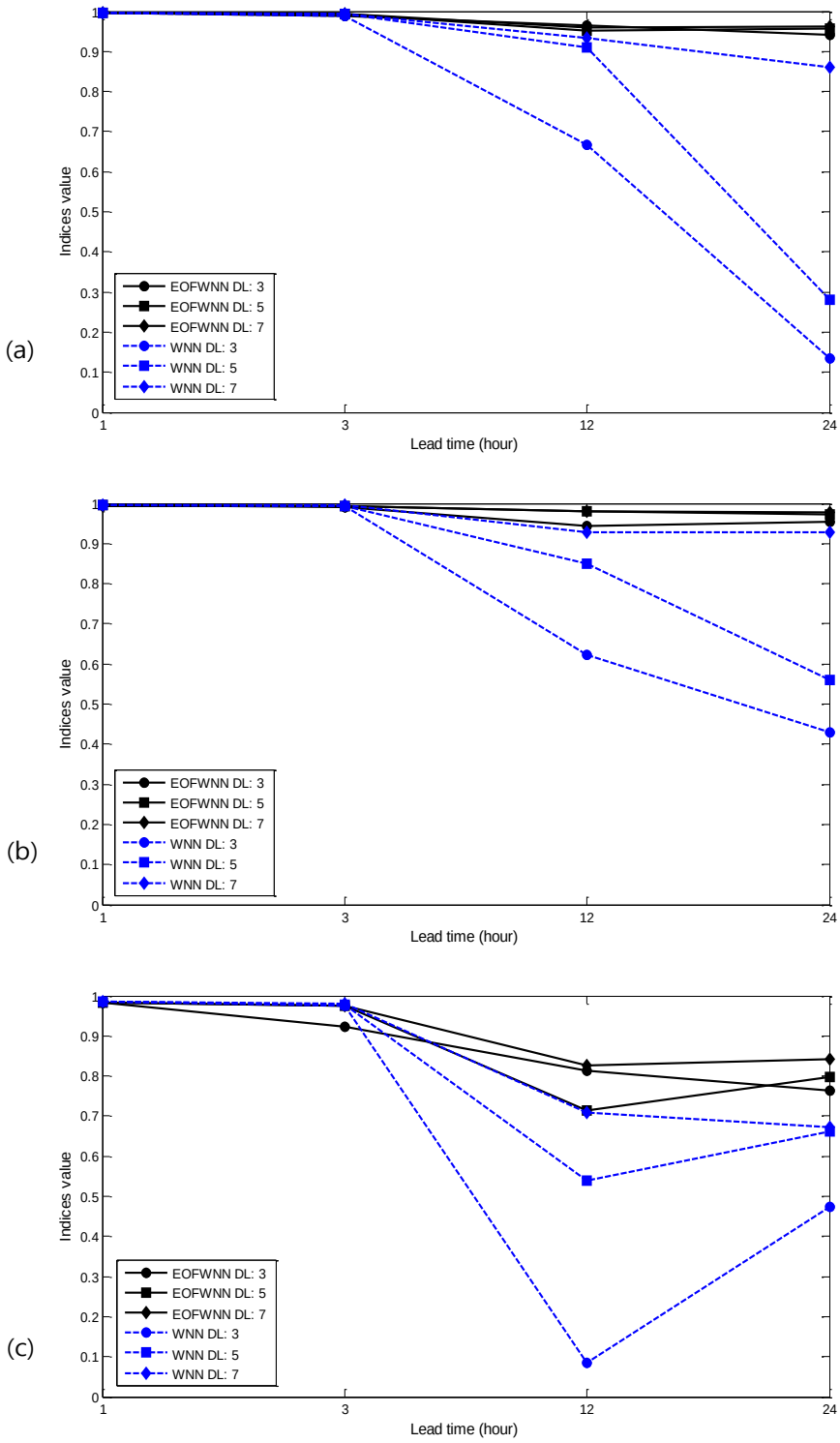


Fig. 4.20 Comparison of index of agreement between EOFWNN and WNN model with 3, 5, and 7 decomposition level in several lead times in (a) Gangneung, (b) Sakata, (c) Aomori

Forecasting of significant wave period series was also conducted using EOFWNN model. The procedure is same as that of significant wave height. Fig. 4.21 represents the first four modes of the PC time series for significant wave period. Since the effect of wavelet decomposition level was shown before, the wavelet decomposition level was fixed as 7^{th} level for prediction of significant wave period. The results obtained from the EOFWNN model at the eight stations are summarized in Tables 4.7 for different lead times. The model shows lower accuracy with increasing lead time. The model showed high accuracy for short lead times. The values of R and I_a were close to one except Aomori, and the *NRMSE* values were distributed between 0.034 and 0.087. The model showed relatively high accuracy even for longer lead times. Figs. 4.22 to 4.24 show the observed and forecasted significant wave period time series by the EOFWNN models with the 7th decomposition level at 24 hour lead time in Gangneung, Sakata and Aomori, respectively. The forecasted time series follow the trend of the observed one, but it does not show the small peaks clearly. Figures for other lead times are attached in appendix.

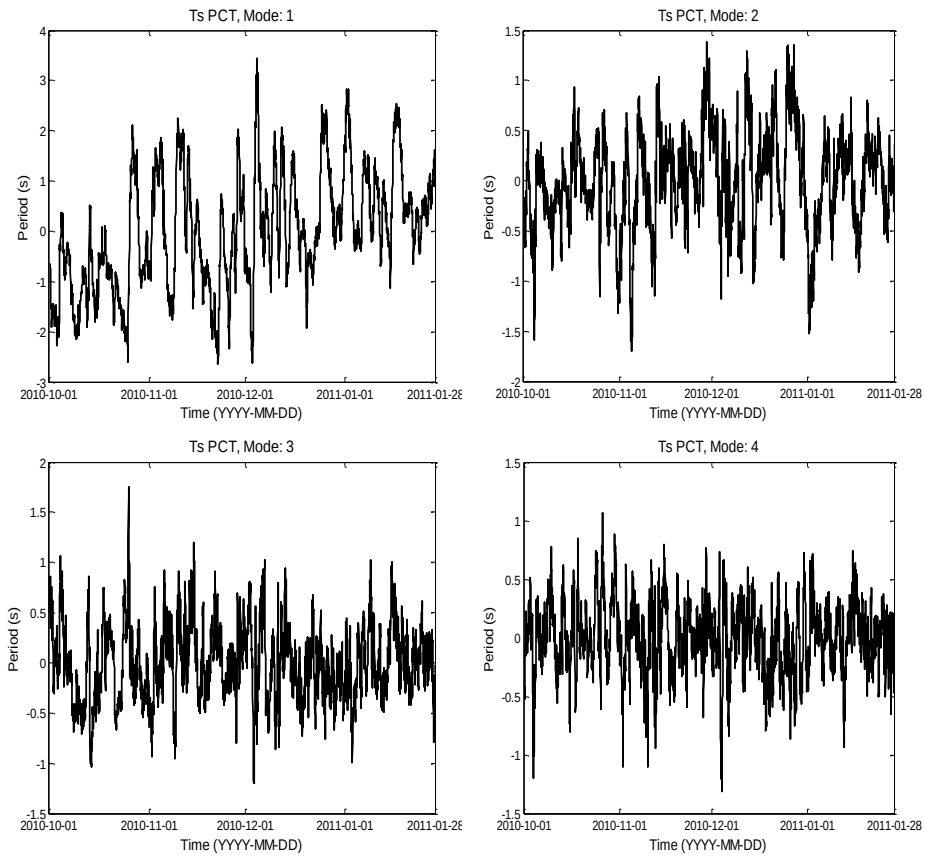


Fig. 4.21 First to fourth mode PC time series for significant wave period

Table 4.7 Test results of EOFWNN model for *T* at 8 stations for several lead times with decomposition level 7

Lead time	Station	R	NRMSE	Index of Agreement
1 hr	Gangneung	0.931	0.038	0.964
	Wangdolcho	0.966	0.038	0.982
	Genkainada	0.979	0.042	0.989
	Tottori	0.977	0.037	0.988
	Fukui	0.985	0.034	0.992
	Sakata	0.983	0.035	0.991
	Aomori	0.739	0.083	0.857
	Rumoi	0.932	0.034	0.965
3 hr	Gangneung	0.929	0.039	0.963
	Wangdolcho	0.963	0.040	0.980
	Genkainada	0.978	0.044	0.988
	Tottori	0.972	0.042	0.983
	Fukui	0.982	0.038	0.990
	Sakata	0.980	0.038	0.989
	Aomori	0.706	0.087	0.839
	Rumoi	0.924	0.036	0.961
12 hr	Gangneung	0.874	0.039	0.934
	Wangdolcho	0.890	0.042	0.943
	Genkainada	0.786	0.048	0.883
	Tottori	0.884	0.037	0.937
	Fukui	0.896	0.043	0.944
	Sakata	0.882	0.045	0.938
	Aomori	0.607	0.077	0.767
	Rumoi	0.770	0.046	0.876
24 hr	Gangneung	0.858	0.043	0.934
	Wangdolcho	0.900	0.040	0.948
	Genkainada	0.756	0.052	0.866
	Tottori	0.879	0.038	0.936
	Fukui	0.898	0.044	0.947
	Sakata	0.911	0.039	0.954
	Aomori	0.709	0.069	0.835
	Rumoi	0.769	0.046	0.875

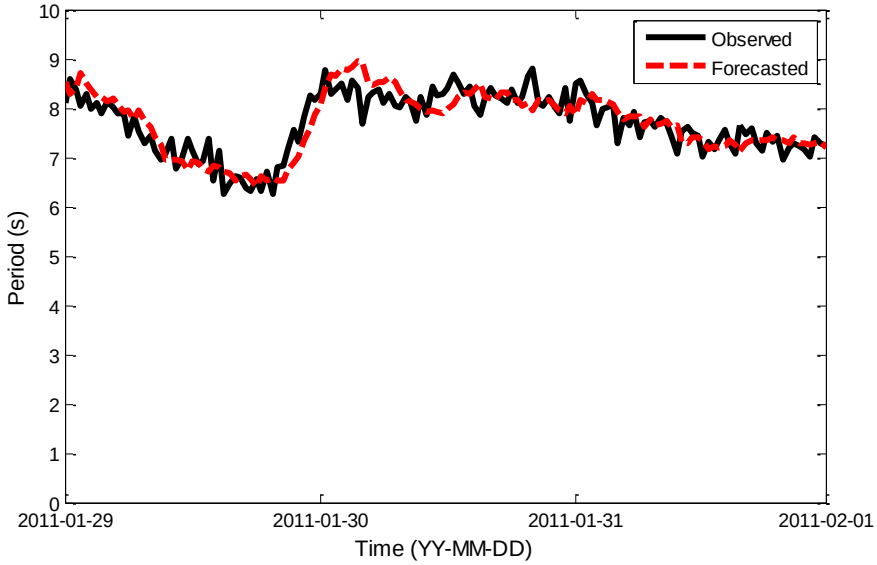


Fig. 4.22 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 24 hour lead time in Gangneung

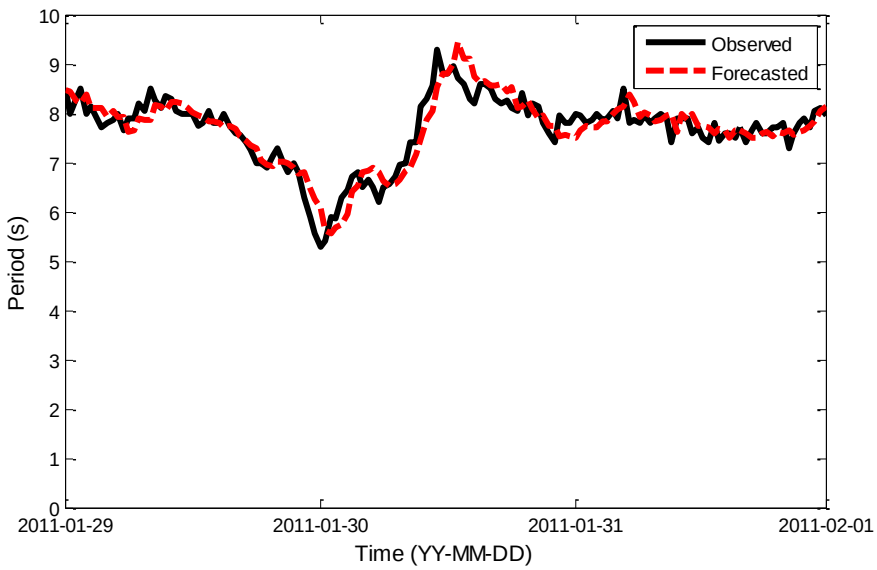


Fig. 4.23 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 24 hour lead time in Sakata

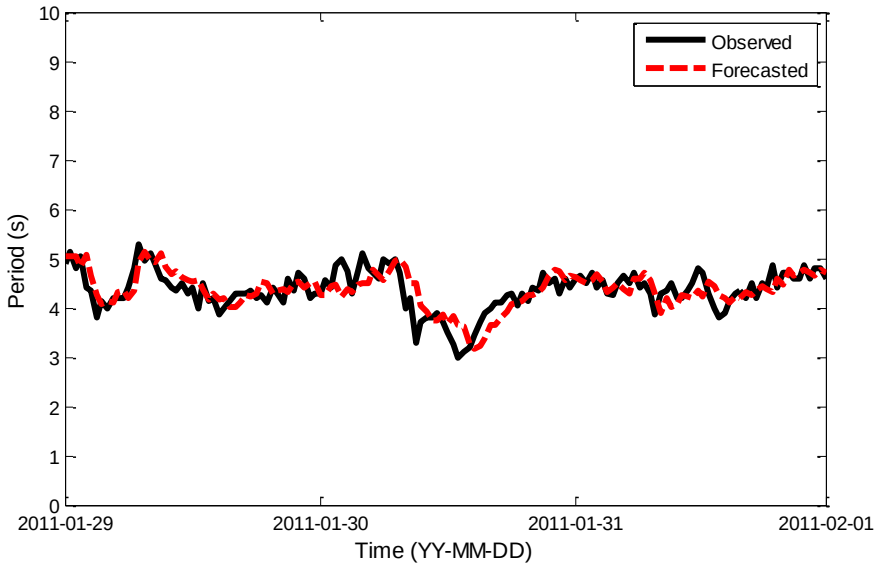


Fig. 4.24 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 24 hour lead time in Aomori

4.3 Forecasting of significant wave height and peak period using the numerical modeling data

The numerical wave modeling data were used to employ the EOFWNN model to the entire area of the East Sea. In the case of forecasting of numerical wave modeling data, wind velocity and wave heights (or periods) were used for input variables. The lead times were fixed as 1 and 3 hours with the use of $\Delta t = 1\,\mathrm{hr}$. The experimental period in this study was from October 1999 to February 2000. Training was conducted for 120 days (00:00 October 1, 1999 to 23:00 January 28, 2000), and testing was conducted for 7 days (00:00 January 29, 2000 to 23:00 February 4, 2000) using the weights and bias obtained from the training period.

Through the EOF analysis, wind velocity and wave height time series were decomposed into several modes and separated into spatial and temporal components. Fig. 4.25 shows the first four modes of the eigenvectors of wind velocity and Fig. 4.26 represents corresponding PC time series. Fig. 4.27 shows the first mode of the eigenvector and corresponding PC time series of significant wave height. The wavelet decomposition level was fixed as 7^{th} level for prediction of significant wave height.

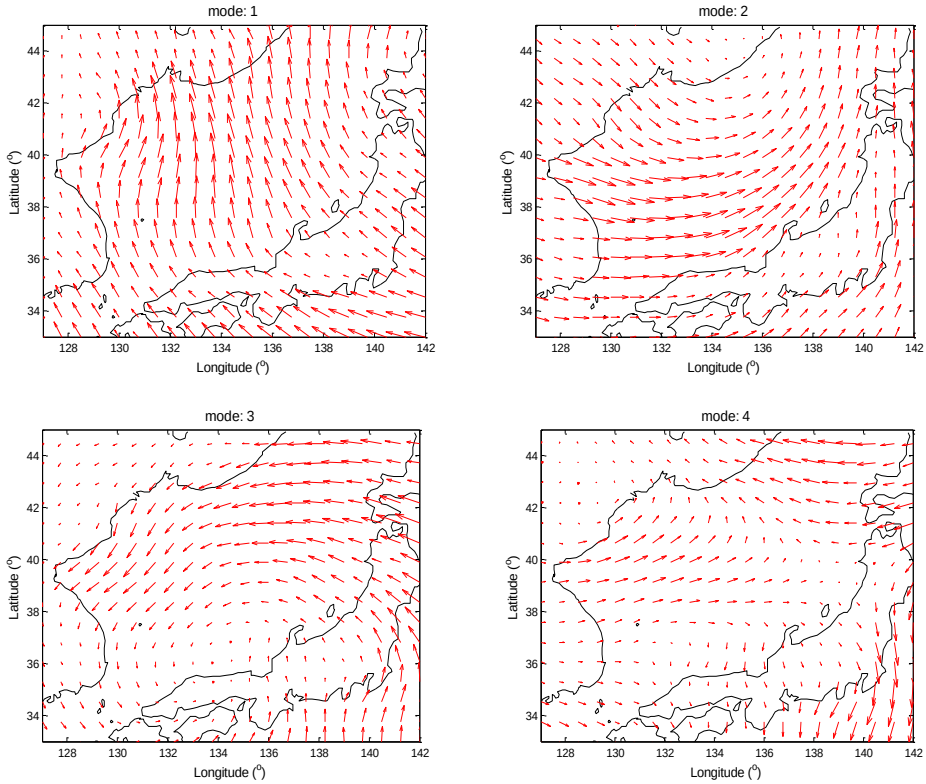


Fig. 4.25 First four modes of eigenvectors for wind velocity

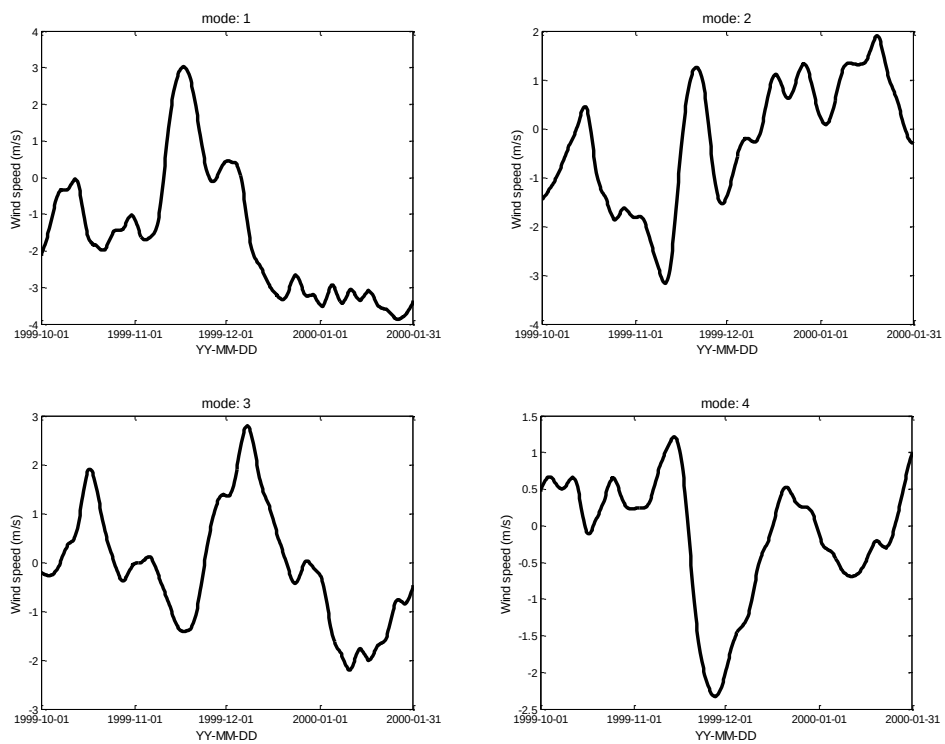


Fig. 4.26 First four modes of PC time series for wind velocity

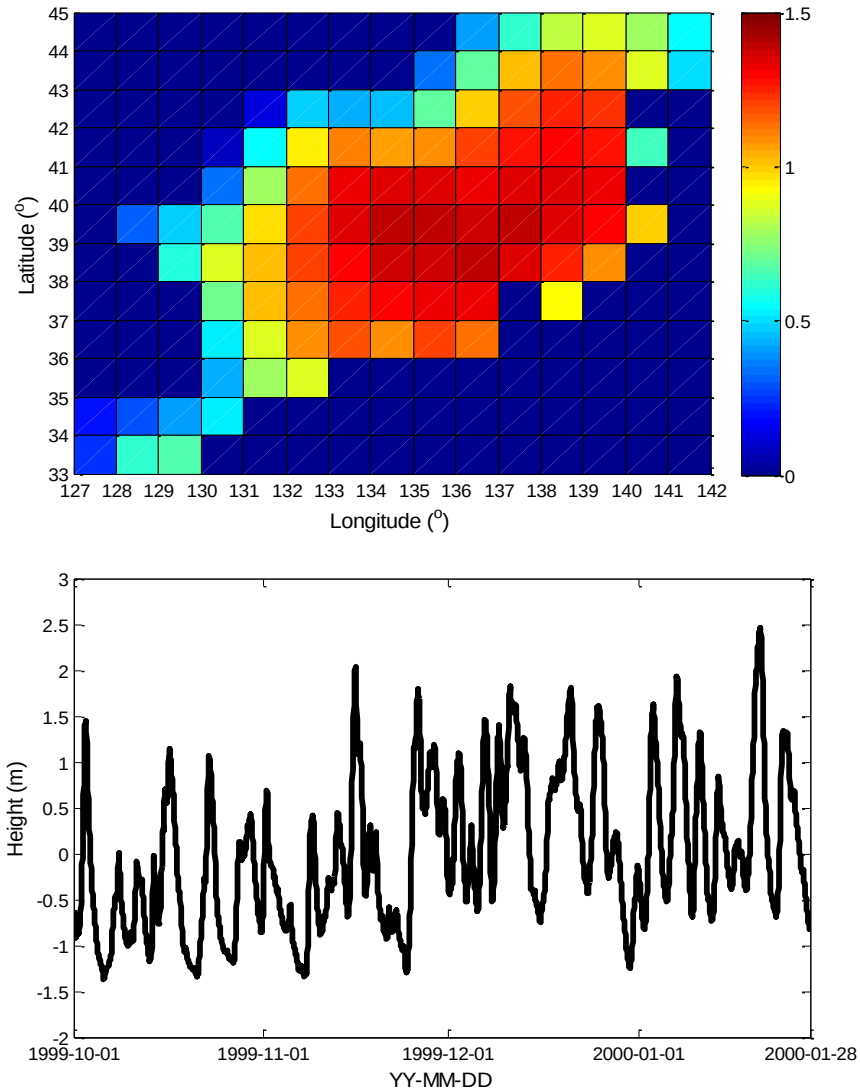


Fig. 4.27 First mode of the eigenvector and corresponding PC time series of significant wave height for numerical wave modeling data

To evaluate the performance of the models, correlation coefficient (*R*), index of agreement (I_a) and normalized root mean square error (*NRMSE*) are also employed. The figures of the results of the EOFWNN model for numerical results of significant wave heights at 1 hour lead time are attached in Appendix. The values of *R* were distributed between 0.872 and 0.995 and for IOA were between 0.925 and 0.997. The range of *NRMSE* values were from 0.044 to 0.228. Fig. 4.28- Fig. 4.30 show the performance of the EOFWNN model for 3 hour lead time. The results were similar to those of 1 hour lead time. The values of *R* were distributed between 0.863 and 0.994 and for IOA were between 0.919 and 0.997. The range of *NRMSE* values were from 0.048 to 0.259. Fig. 4.31 shows the highest performance case for IOA and Fig. 4.32 shows the lowest performance case. Since the wave heights are relatively small for the lowest performance case, the accuracy is lower than those at other locations.

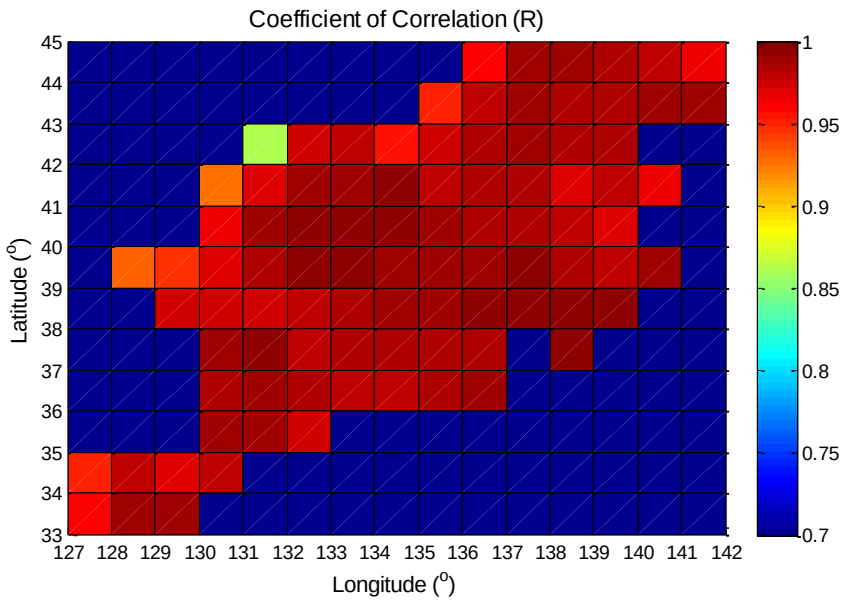


Fig. 4.28 Coefficient of correlation of EOFWNN model for numerical results of significant wave height with 7th decomposition level at 3 hr lead time

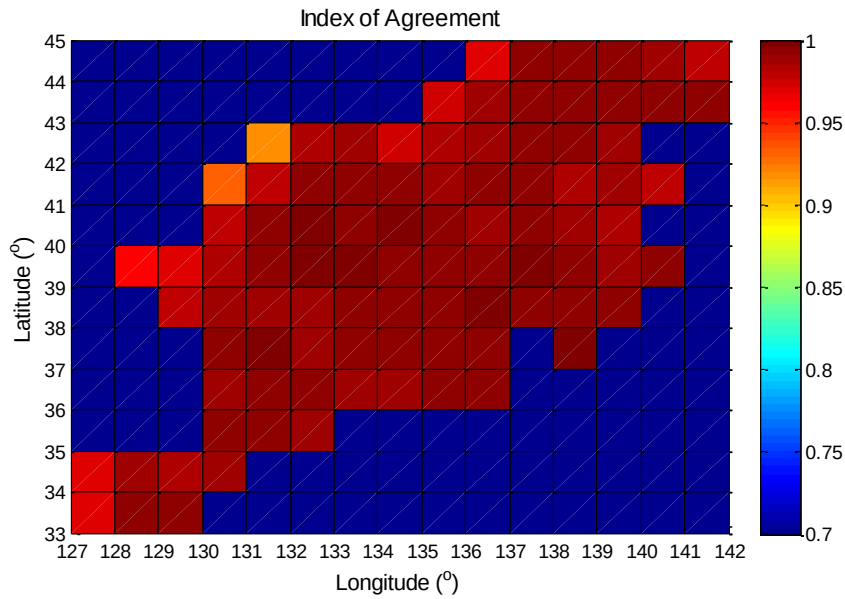


Fig. 4.29 Index of agreement of EOFWNN model for numerical results of significant wave height with 7th decomposition level at 3 hr lead time

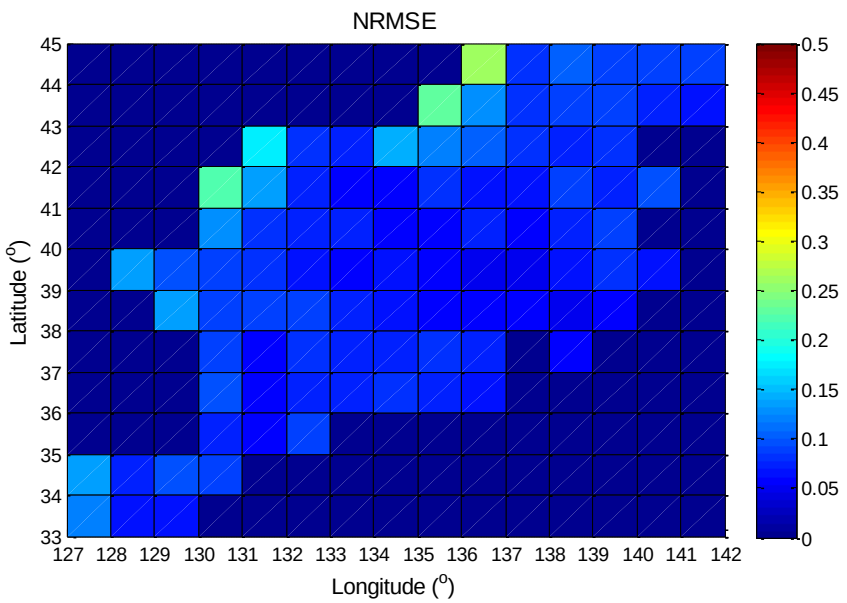


Fig. 4.30 *NRMSE* of EOFWNN model for numerical results of significant wave height with 7th decomposition level at 3 hr lead time

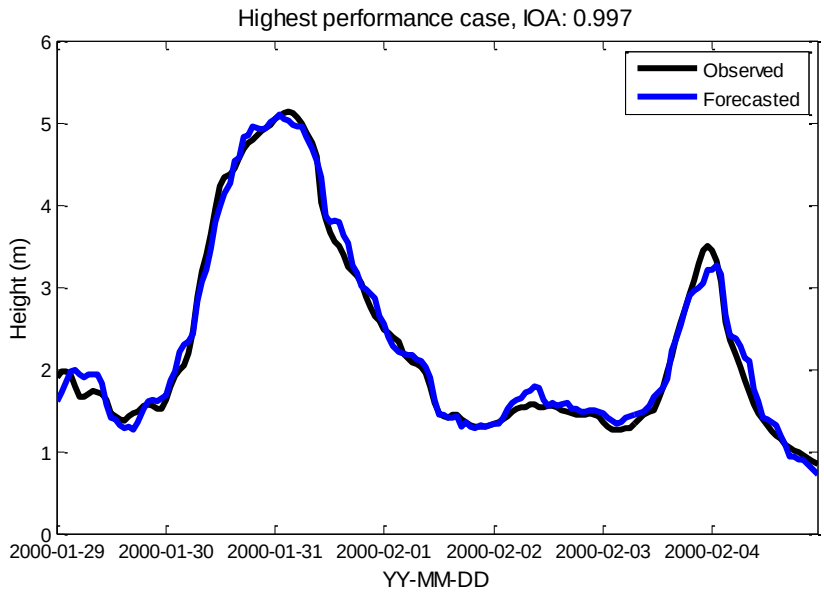


Fig. 4.31 Highest performance case of EOFWNN model for numerical results of significant wave height with 7th decomposition level at 3 hr lead time

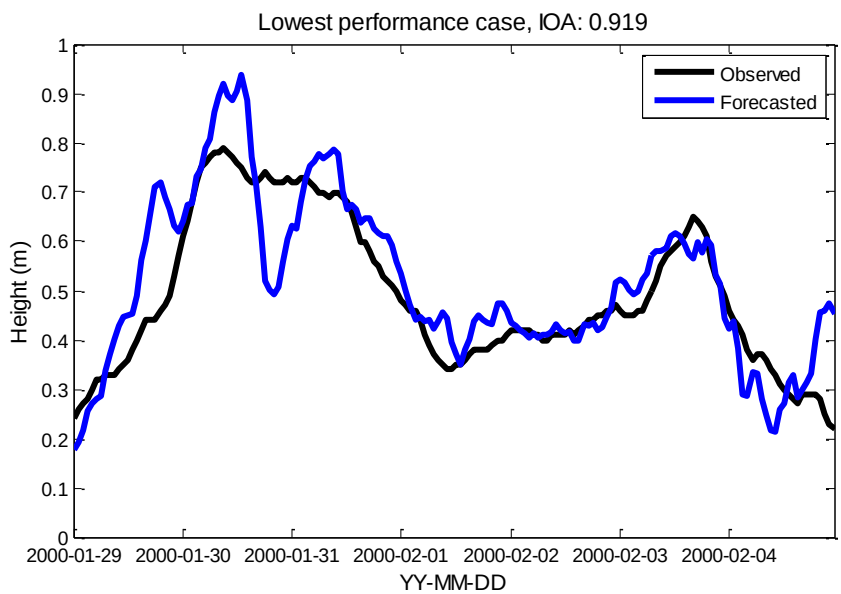


Fig. 4.32 Lowest performance case of EOFWNN model for numerical results of significant wave height with 7th decomposition level at 3 hr lead time

Forecasting of numerical results of peak period series was also conducted using EOFWNN model for 1 and 3 hour lead times. The procedure of forecasting is same as that of numerical results of significant wave height. Fig. 4.33 shows the first mode of the eigenvector and corresponding PC time series of significant wave period. The figures of the results of the EOFWNN model for numerical results of peak periods at 1 hour lead time are attached in Appendix. The values of *R* were distributed between 0.722 and 0.990 and for IOA were between 0.850 and 0.995. The range of *NRMSE* values were from 0.027 to 0.249. The range of *R* and IOA for peak period was wider than those of significant wave height, but the values of *NRMSE* was similar to those of significant wave height. Fig. 4.34-Fig. 4.36 show the performance of the EOFWNN model for 3 hour lead time. The results were similar to those of 1 hour lead time. The values of *R* were distributed between 0.676 and 0.974 and for IOA were between 0.819 and 0.987. The range of *NRMSE* values were from 0.042 to 0.247. Fig. 4.37 shows the highest performance case for IOA and Fig. 4.38 shows the lowest performance case. Even the lowest performance case, the peaks are shown clearly. In future work the EOFWNN model will be employed for longer lead times.

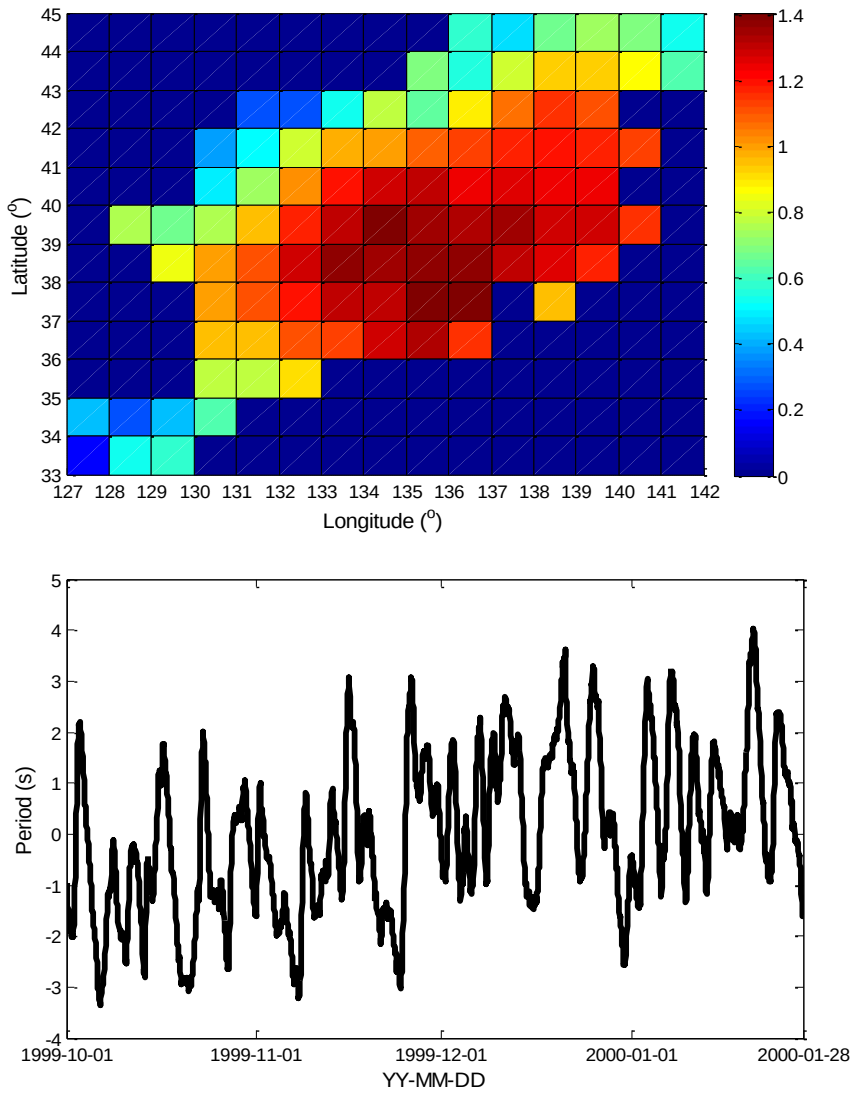


Fig. 4.33 First mode of the eigenvector and corresponding PC time series of peak period for numerical wave modeling data

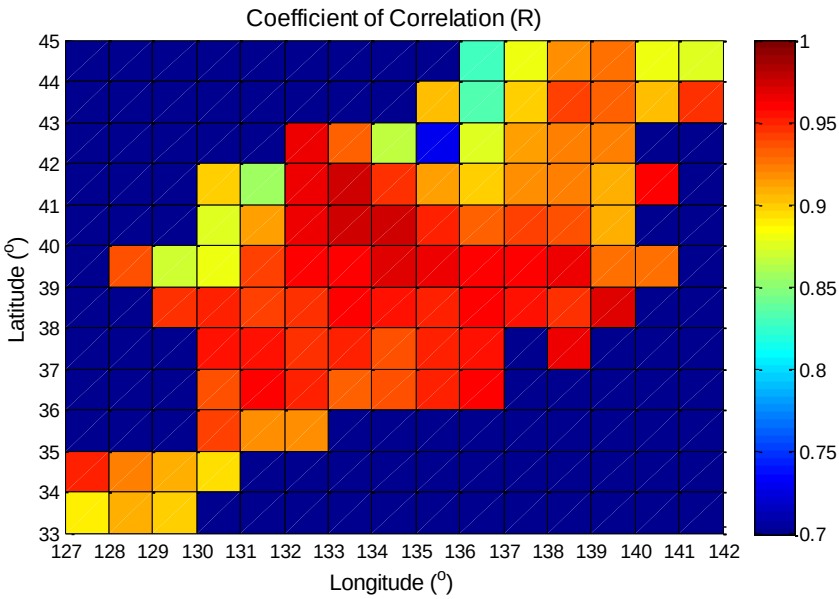


Fig. 4.34 Coefficient of correlation of EOFWNN model for numerical results of peak period with 7th decomposition level at 3 hr lead time

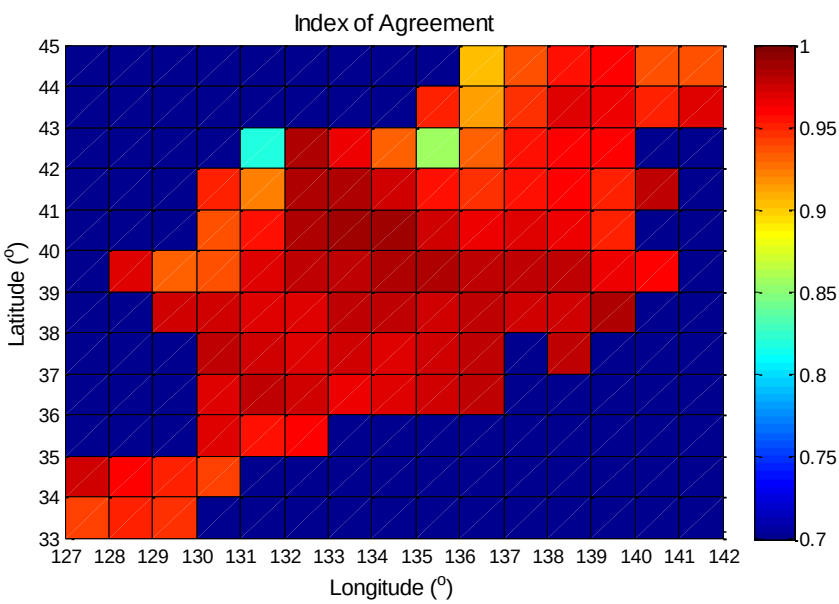


Fig. 4.35 Index of agreement of EOFWNN model for numerical results of peak period with 7th decomposition level at 3 hr lead time

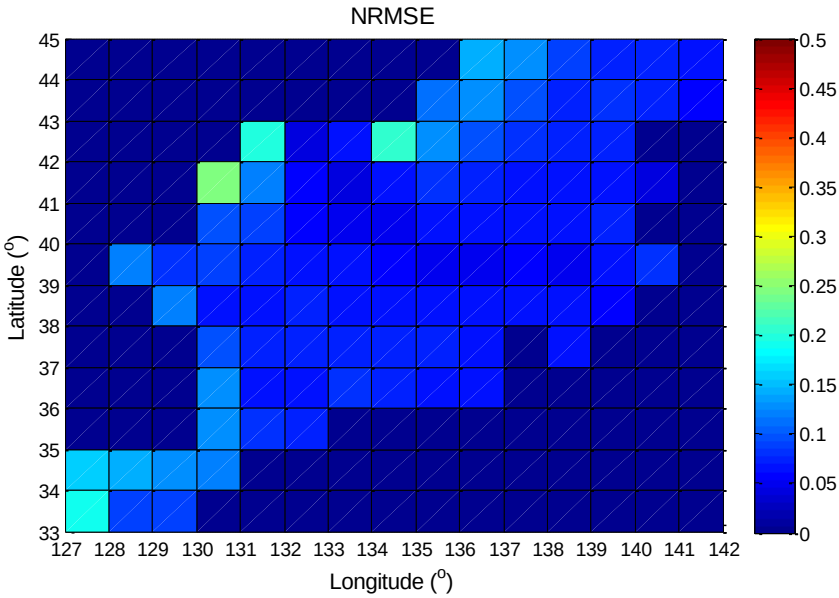


Fig. 4.36 *NRMSE* of EOFWNN model for numerical results of peak period with 7th decomposition level at 3 hr lead time

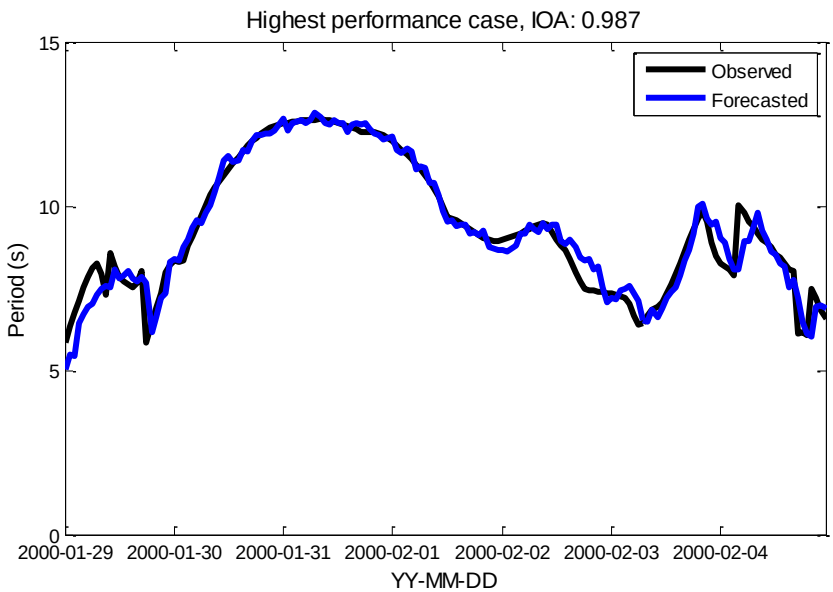


Fig. 4.37 Highest performance case of EOFWNN model for numerical results of peak period with 7th decomposition level at 3 hr lead time

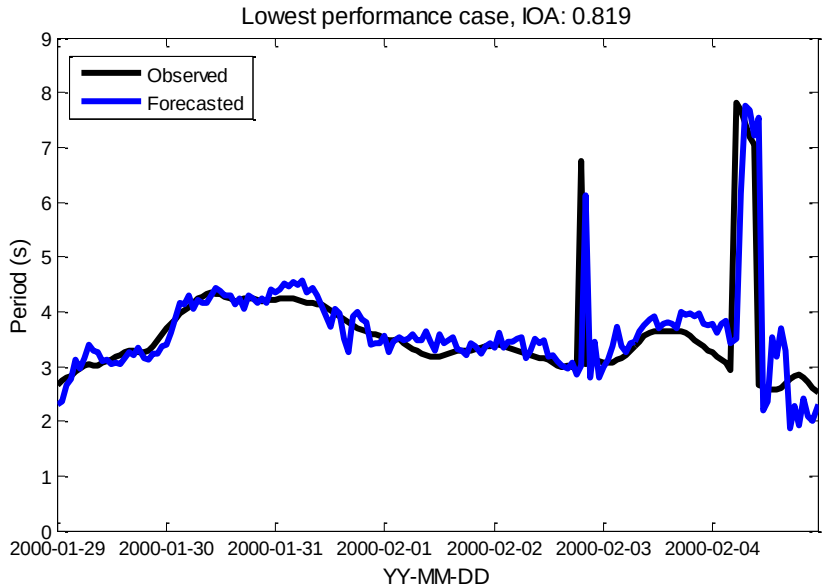


Fig. 4.38 Lowest performance case of EOFWNN model for numerical results of peak period with 7th decomposition level at 3 hr lead time

4.4 Detecting and forecasting of sudden high waves

In the previous chapter 3.2, a criterion of sudden high waves based on wave measurement data and the record of marine accidents and property damage was proposed. It also could be used for forecasting of sudden high waves based on wave forecast data. To examine this possibility, the spatio-temporal variation of $\Delta(H^2L)/\Delta t$ was calculated in the East Sea using the wave hindcast data provided by the KORDI. Fig. 4.39 shows several snapshots taken from the spatio-temporal variation of $\Delta(H^2L)/\Delta t$ calculated during the period of February 2-3, 1987. In order to forecast sudden high waves at each time step, $\Delta(H^2L)/\Delta t$ was calculated at each time step using the maximum value detected from the minimum point of the preceding data to each time step. Note that the maximum value of H^2L during a high wave event can be obtained at each time step and the value of H^2L during non-event time is zero. As shown in Fig. 4.39, the high values of $\Delta(H^2L)/\Delta t$ formed a string, which traveled southeastward across the sea, indicating that a high wave event could suddenly occur from a calm sea. The string arrived near the northwest coast of Honshu, Japan, around 2pm February 3, 1987, as shown in Fig. 4.39 (d).

In order to validate the calculated results in Fig. 4.39, the wave data measured along the coast of Japan by the NOWPHAS system was used, which were measured at three locations at the northwest coast of Honshu, Japan, i.e., Hamada, Tottori, and Kanazawa, as shown in Fig. 4.40. Fig. 4.41 shows the temporal variation of H^2L measured at these stations in February 1-5, 1987. The values of H^2L were small until near the noon February 3rd, after which they increased

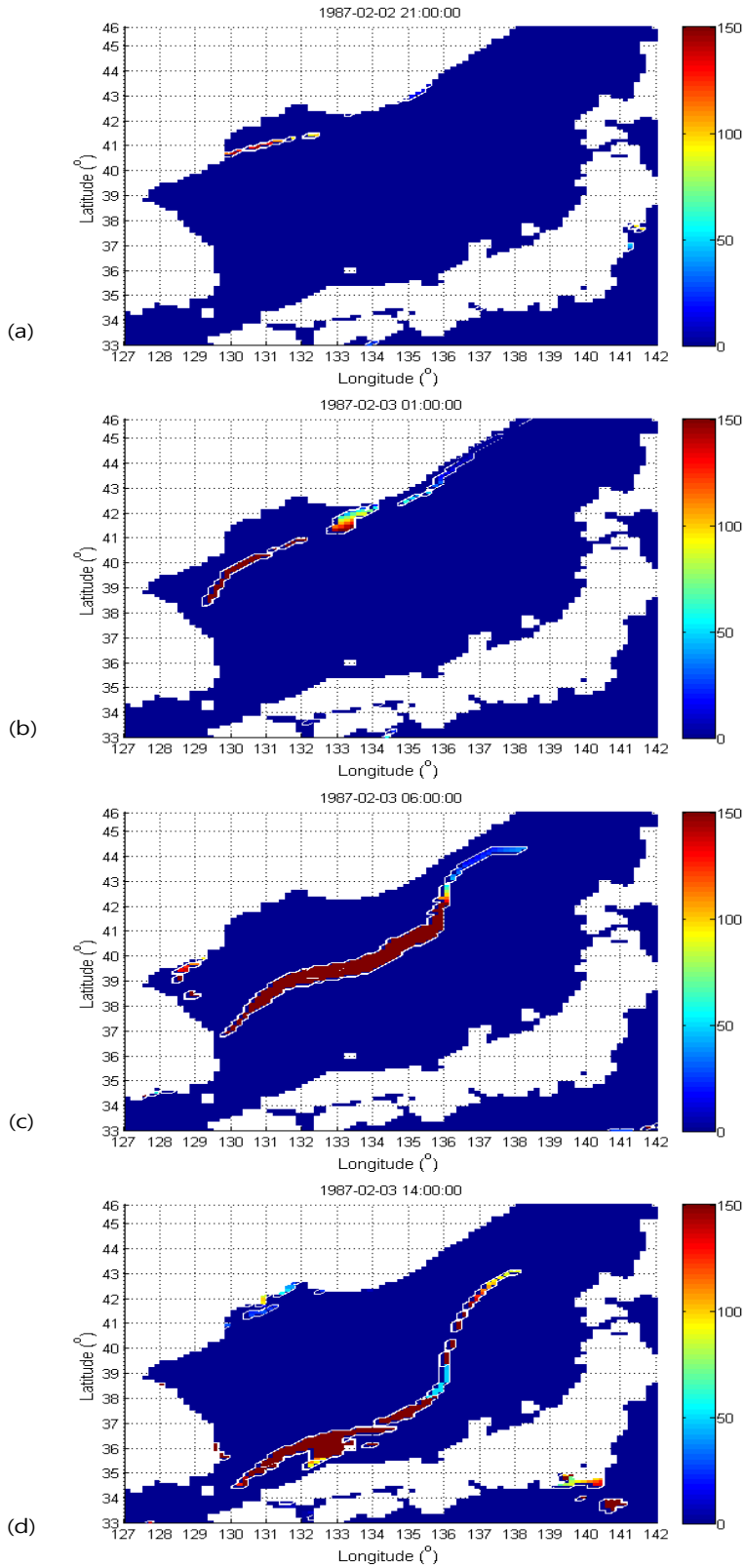


Fig. 4.39 Snapshots of $\Delta(H^2L)/\Delta t$ calculated during the period of February 2-3, 1987: (a) 9pm Feb.2nd (b) 1am Feb. 3rd (c) 6am Feb. 3rd (d) 2pm Feb. 3rd

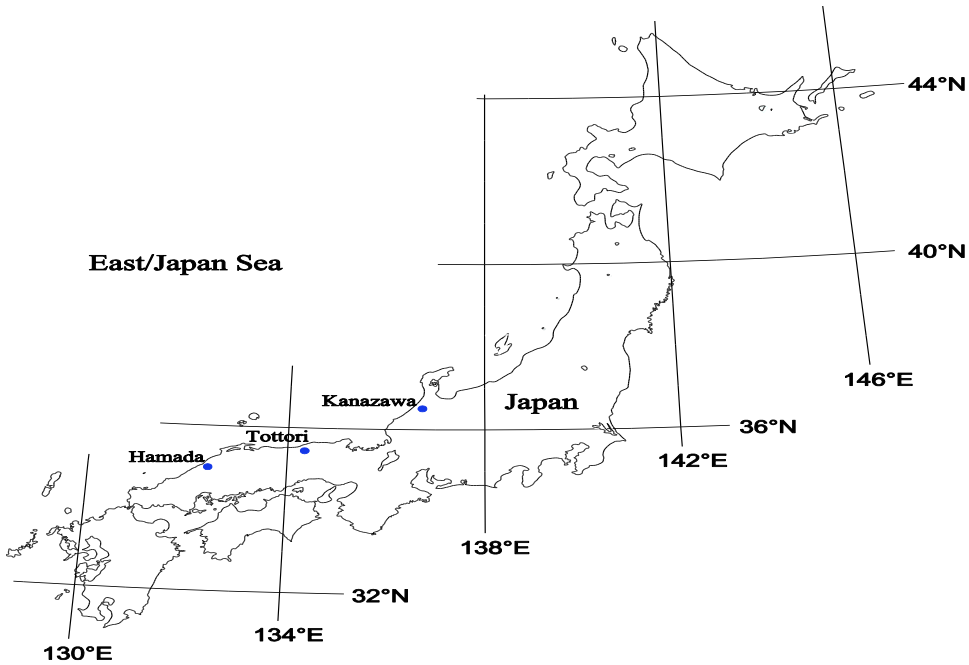


Fig. 4.40 Locations of wave observation stations of NOWPHAS system

Table 4.8 Disaster damage occurred in February 3-4, 1987, in Shimane Prefecture, Japan (unit = 1,000 Japanese Yen)

Classification (Date)	Civil facilities	Fishery port facilities	Agricultural products	Fishery products and facilities	Total
Winter wind waves (February 3-4)	139,249	300,340	18,665	400,689	858,943

rapidly to reach their maxima near the midnight. According to Fig. 4.39, it is expected that the sudden high wave events would occur during the late afternoon or early evening February 3rd, which corresponds well with the results from the measurement shown in Fig. 4.41. The values of $\Delta H^2 L / \Delta t$ for the events shown in Fig. 4.41 were 138.4, 221.9, and 35.8 m^3/hr at Hamada, Tottori, and Kanazawa, respectively. According to the criterion proposed in the present study, a sudden high wave event would have occurred at Hamada and Tottori, but not at Kanazawa. It may be interesting to investigate if any marine accident or property damage had occurred in these areas during this period. Table 4.8 shows a part of the record of the disaster damage occurred in 1987 in Shimane Prefecture, Japan, in which Hamada is located (Shimane Prefecture, 1987). In February 3-4, 1987, there was no human casualty or injury, but property damage of 858,943,000 Japanese Yen occurred by winter wind waves mostly for civil and fishery port facilities and marine products facilities. This indicates that the proposed intensity parameter is a good measure for sudden high waves and it could be used to forecast sudden high waves.

To evaluate forecast of sudden high waves, sudden high waves were forecasted using the wave forecast data in the previous chapter. $\Delta H^2 L / \Delta t$ was calculated using the wave data measured at eight stations by KIOST and NOWPHAS. The present time was fixed at 12:00 January 29, 2011. Then, the each forecast time was at 13:00 January 29, 2011 for 1 hour lead time, at 15:00 January 29, 2011 for 3 hour lead time, at 00:00 January 30, 2011 for 12 hour lead time and at 12:00 January 30, 2011 for 24 hour lead time. Figs. 4.42 – 4.43 show the comparison of temporal variation of $\Delta H^2 L / \Delta t$ between the observed and the forecasted wave

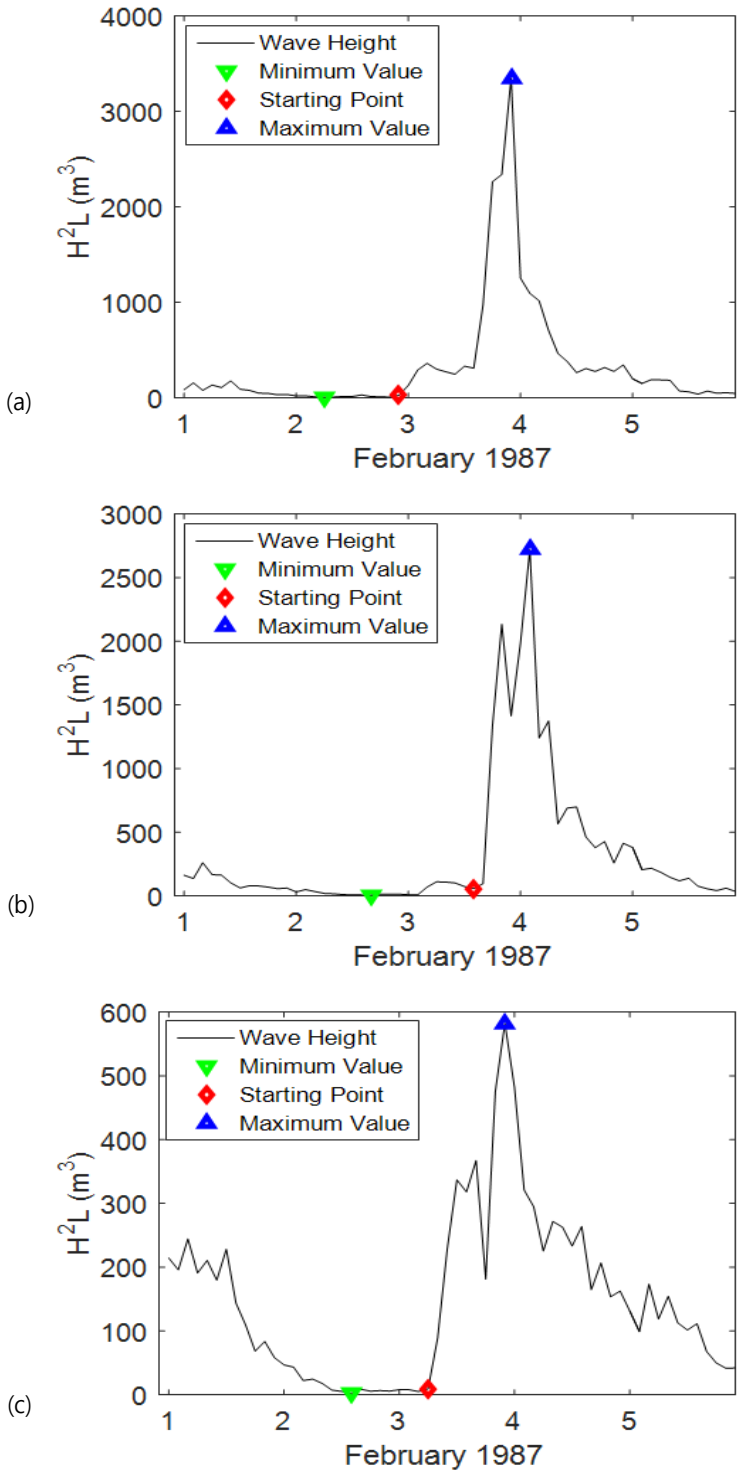


Fig. 4.41 Temporal variation of H^2L measured at three wave stations of NOWPHAS system in February 1-5, 1987: (a) Hamada; (b) Tottori; and (c) Kanazawa

data at 24 lead times in Gangneung and Sakata, respectively. The red line is the present time, 12:00 January 29, 2011. The value of $\Delta H^2 L / \Delta t$ at 00:00 January 25, 2011 in Gangneung was $14.02\ m^3/hr$, and the value during the forecast period was $20.04\ m^3/hr$ for observed data and $17.74\ m^3/hr$ for forecasted data. During this forecast period, there was no sudden high waves in Gangneumg, but sudden high wave was detected clearly in Sakata. The value of $\Delta H^2 L / \Delta t$ in the afternoon January 28, 2011 in Sakata was $108.93\ m^3/hr$, and the value during the forecast period was $814.65\ m^3/hr$ for observed data and $631.19\ m^3/hr$ for forecasted data. Although there was a slight deviation between the results of observed and forecasted wave data, the maximum peak was forecasted quite accurately.

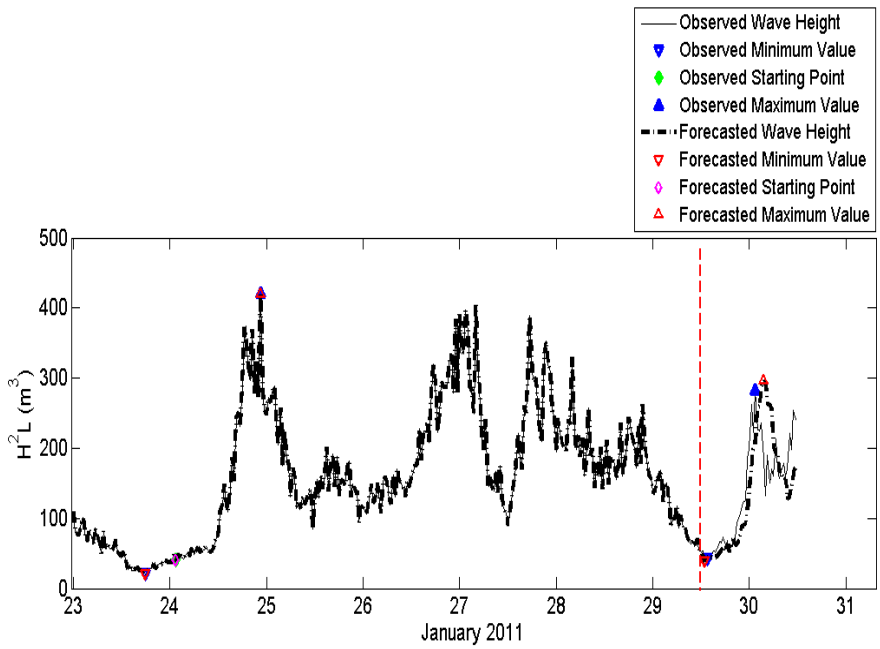


Fig. 4.42 Comparison of temporal variation of H^2L between observed and forecasted wave data at 24 hour lead time in Gangneung

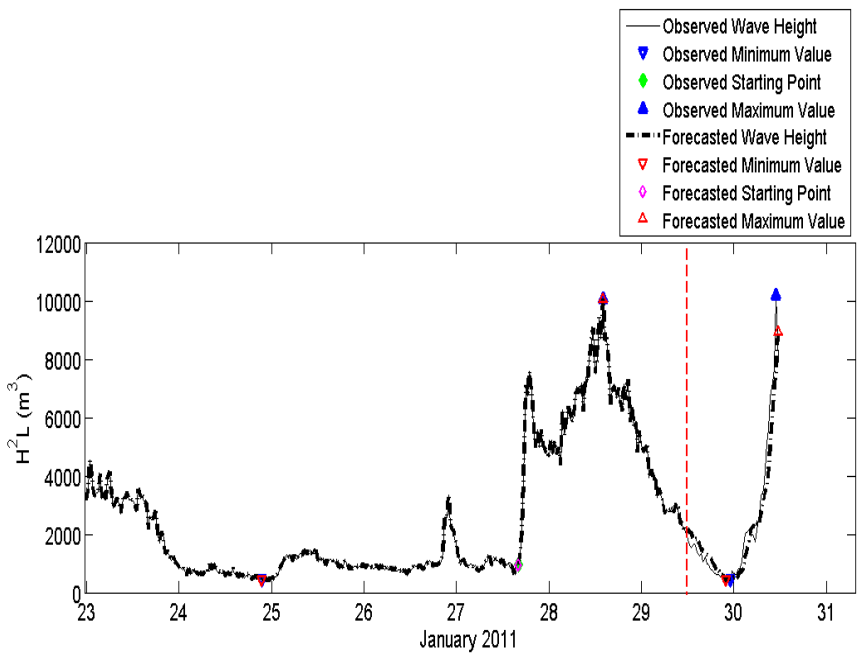


Fig. 4.43 Comparison of temporal variation of H^2L between observed and forecasted wave data at 24 hour lead time in Sakata

CHAPTER 5. CONCLUSIONS

5.1 Summary and conclusions

The main purposes of this study are to analyze the generation mechanism of sudden high waves and forecast sudden high waves in the East Sea of Korea using statistical methods. The objectives are as follows: 1) to define sudden high waves 2) to investigate the generation mechanism of the waves by analyzing the relation between waves and meteorological variables 3) to forecast significant wave height and period in the East Sea of Korea 4) to detect and forecast sudden high waves from the proposed intensity parameter and wave forecast data in the East Sea of Korea.

In this study, a more clear intensity parameter and criteria of sudden high waves were suggested. Wave data measured in Gangneung and Wangdolcho in 2005- 2012 provided by KIOST were used and the occurrences and damages of high wave events on the east coast of Korea during the period were referenced. Since the wave energy increases dramatically during the event, the index of sudden high waves set as $\Delta(H^2L)/\Delta t$ and the values of $\Delta(H^2L)/\Delta t$ were calculated. From the results it was found that $\Delta(H^2L)/\Delta t$ was proportional to the severity of the events and the rapidly increasing curve of $\Delta(H^2L)/\Delta t$ levels off at about 80% of cumulative percentage. Therefore, the criteria of sudden high waves was set 80 m^3/hr, which corresponds to the top 20%.

Next, the generation mechanism of overall sudden high waves was analyzed unlike previous researches focusing on the specific events. Experiments were

conducted for the winter season from October to February, when sudden high waves occur frequently. The wave data provided by KIOST at Gangneung and Wangdolcho were used and the meteorological data were NCEP/NCAR reanalysis data. To detect the relationship between sudden high wave modes and meteorological variables, CSEOF analysis was used, which is a useful tool to extract the physical modes from geophysical variables. The significant wave height and the index of sudden high waves, H^2L was analyzed using CSEOF analysis. From the shapes of the modes and comparison between the reported events and the reconstructed series, the second and third modes of significant wave height and the first two modes of the index of sudden high waves were considered as the physical process of sudden high waves. By regression analysis, the evolution of spatial patterns of wind velocity and sea level pressures was analyzed during the sudden high wave events. The patterns were categorized two groups. There are two peaks in the modes of all CSLV considered as the physical process of sudden high waves. The common patterns were until the rapid growth of the wave height, there was spatial variation in sea level pressure and wind speed, and the weather in the Korean peninsula was mild in those days. The first pattern was that the first peak seemed to be generated by low pressure moving to the north east part of the East Sea and easterly wind blowing for 1 day, whereas the second peak seemed to be caused by strong wind. The second patterns was that the first seemed to be affected by the strong wind in the east coast and the second peak seemed to be influenced by wind in the offshore area. The result was consistent with the previous researches.

To forecast significant waves at multiple stations simultaneously, an EOFWNN model was developed by combining the EOF analysis and wavelet analysis with the neural network. The wave data used in this research were

measured at eight wave observation stations in the East Sea by KIOST and NOWPHAS and the meteorological data were the NCEP/NCAR reanalysis data. The period was from October 2010 to February 2011. The results of the EOFWNN model for significant wave height were compared with those of a wavelet and neural network hybrid (WNN) model in Gangneung, Sakata and Aomori for several lead times. The EOFWNN model which used the data at multiple locations together showed better performance than the WNN model which used the data at each location separately. The high accuracy of the EOFWNN model was attributed to considering the effect of spatially distributed meteorological variables by the EOF analysis. The EOFWNN model is better than the WNN model in that the former shows higher accuracy for longer lead times regardless of the decomposition level and that it forecasts the wave heights at multiple locations together. The results of significant wave period also showed quite high accuracy, although they were less accurate than those of significant wave height. Also, the EOFWNN model was employed to the entire area of the East Sea, which were numerical wave modeling data provided by KORDI. ECMWF reanalysis data were used for wind velocity. The lead time was fixed as 1 and 3 hours. Even though the number of grid for sea area is 96, the results showed relatively high accuracy. By combining wavelet analysis to neural networks, non-stationarity and non-linearity problem could be overcome, but WNN model has still low accuracy at higher lead times. While the EOFWNN model shows the high performance at higher lead times. It came from the reflection of the relationship between spatially distributed meteorological variables and waves by combining EOF analysis. Also, EOF analysis helped for EOFWNN model to forecast wave series at multi-stations simultaneously. Therefore, the EOFWNN model can be a promising tool for forecasting the

significant wave heights or periods at multiple locations for a relatively long lead time with high accuracy.

Using the proposed intensity parameter and the forecasted significant waves by the EOFWNN model, sudden high wave was detected and forecasted. From the forecasted wave data at 24 hour lead time, $\Delta(H^2L)/\Delta t$ was calculated in Gangneung and Sakata. During the forecast period, there was no sudden high waves in Gangneung, but the value of $\Delta(H^2L)/\Delta t$ in Sakata was over the criteria of sudden high waves, 80 m^3/hr. Although there was a slight deviation between the results of observed and forecasted wave data, sudden high wave was detected clearly. Until now, sudden high waves have been forecasted only using numerical methods. From this study, we saw the potential and reality to forecast sudden high waves using machine-learning.

5.2 Future study

Fig. 5.1 shows the autocorrelation graph of residuals of observed and forecasted significant wave height series at 24 hour lead time in Gangneung. Until 1 and half hours the values are quite correlated. Also, in Fig. 4.17 – 4.19 there are phase shifts about 1 – 1.5 hours. In this research, to consider the time lag effect of the meteorological data, the meteorological data at 30 minutes and 3 hours ahead of the forecast lead time as input data of EOFWNN model were arbitrarily chosen. It seems that the arbitrarily chosen time lags of meteorological data affect the forecasting results for significant waves. In the future research, the effect of different time lags of meteorological data will be compared. In this study,

forecasting lead times of the EOFWNN model were 1, 3, 12 and 24 hours. However, to deal with sudden high waves, more than 24 hour lead time forecasting is required. Thus, the EOFWNN model will be employed at longer lead times more than 24 hours in the future study. Lastly, it should be mentioned that the model showed lower accuracy at the Aomori station, which is located inside a bay and thus in which the wave height is small compared with other stations. It may be desirable to use the model for the wave stations which are located in open coasts so that the magnitudes of wave height are similar to one another.

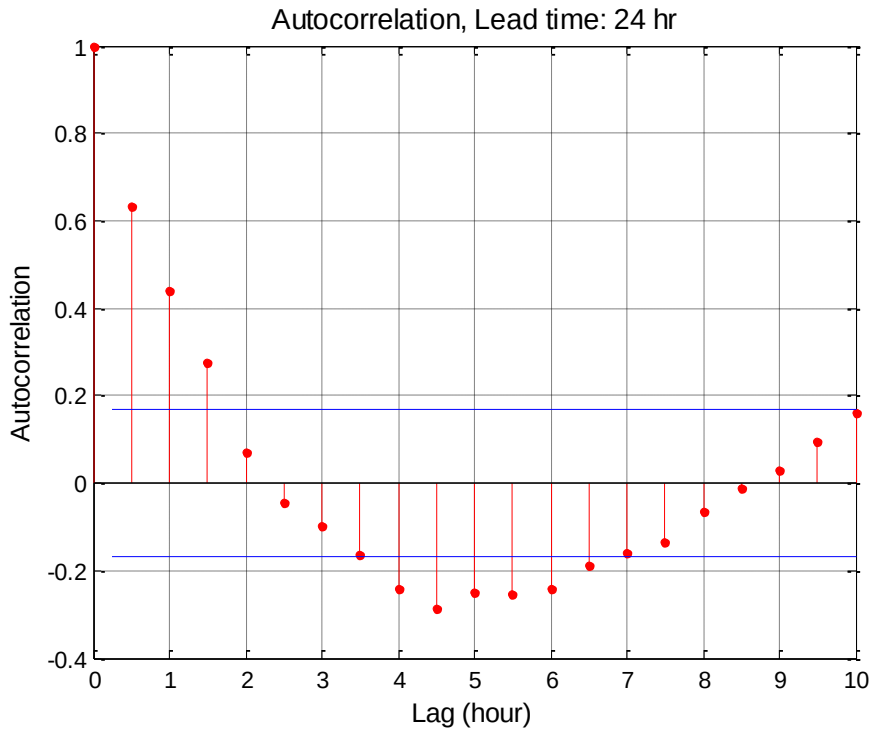


Fig. 5.1 Autocorrelation of residuals of observed and forecasted results for significant wave height at 24 hour lead time in Gangneung

REFERENCES

Ahn, Suk Jin, Byeong Wook Lee, Seok Jae Kwon, and Changhoon Lee. 2013. "Algorithm of Predicting Swell-like Significant Waves in the East Coast of Korea." *Journal of The Korean Society of Civil Engineers* no. 33 (6):2329-2341.

Cannas, Barbara, Alessandra Fanni, Linda See, and Giuliana Sias. 2006. "Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning." *Physics and Chemistry of the Earth, Parts A/B/C* no. 31 (18):1164-1171.

Dee, DP, SM Uppala, AJ Simmons, Paul Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, and P Bauer. 2011. "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system." *Quarterly Journal of the royal meteorological society* no. 137 (656):553-597.

Deka, Paresh Chandra, and R Prahlada. 2012. "Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time." *Ocean Engineering* no. 43:32-42.

Demuth, Howard, and Mark Beale. 2000. "Neural network toolbox user's guide."

Deo, Mc C, A Jha, AS Chaphekar, and K Ravikant. 2001. "Neural networks for wave forecasting." *Ocean Engineering* no. 28 (7):889-898.

Dixit, Pradnya, and Shreenivas Londhe. 2016. "Prediction of extreme wave heights using neuro wavelet technique." *Applied Ocean Research* no. 58:241-252.

Fletcher, Desmond, and Ernie Goss. 1993. "Forecasting with neural networks: an application using bankruptcy data." *Information & Management* no. 24 (3):159-167.

Günaydın, Kemal. 2008. "The estimation of monthly mean significant wave heights by using artificial neural network and regression methods." *Ocean Engineering* no. 35 (14):1406-1415.

Hamlington, BD, RR Leben, RS Nerem, W Han, and K-Y Kim. 2011. "Reconstructing sea level using cyclostationary empirical orthogonal functions." *Journal of Geophysical Research: Oceans (1978–2012)* no. 116 (C12).

Hatada, Yoshio, and Masataka Yamaguchi. 1998. "Preliminary study on a prediction method of large swell-like waves in Toyama Bay "Yorimawari-Nami"." *MEMOIRS-FACULTY OF ENGINEERING EHIME UNIVERSITY* no. 17:261-272.

Isozaki, I. 1971. "On the characteristics of coastal waves on the coast of Toyama bay (Report I)." *Notes of Cooperative Research for Disaster Prevention* no. 25: 3–15.

Isozaki, I. , and O. Yoshio. 1972. "On the characteristics of coastal waves on

the coast of Toyama bay (Report II)." *Notes of Cooperative Research for Disaster Prevention* no. 28:3–17.

Jeong, W.-M., Oh, S.-H., Ryu, K.-H., Lee, D. Y., and Chae, J. W. 2009. "Multi-station observation of the high swell along the east coast of Korea." *Proc. 5th Int. Conf. Asian Pac. Coasts, Singapore, CD-ROM*.

Jeong, Weon-Mu, Sang-Ho Oh, and Dong-Young Lee. 2007. "Abnormally high waves on the East Coast." *J. Korean Soc. Coast. Ocean Engrs* no. 19:295-302.

Joung, C.-H., S. S. Kim, S.-U. Park, K. D. Min, and H.-S. An. 1984. "A case study on the extratropical cyclone development on the East Sea." *J. Korean Met. Soc.* no. 20:1-21.

Kalnay, Eugenia, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, and John Woollen. 1996. "The NCEP/NCAR 40-year reanalysis project." *Bulletin of the American meteorological Society* no. 77 (3):437-471.

Kalra, Ruchi, MC Deo, Raj Kumar, and Vijay K Agarwal. 2005. "Artificial neural network to translate offshore satellite wave data to coastal locations." *Ocean Engineering* no. 32 (16):1917-1932.

Kashima, Hiroaki, and Katsuya Hirayama. 2011. Effects of surf beat caused by long period swell on wave overtopping rate on complex bathymetry. Paper read at th 5 SCACR 2011 International Short Conference on Applied Coastal Research.

Kim, KO, HS Lee, BI Min, and BH Choi. 2011. "Hindcasting of high swell waves off the eastern Korean coast on 24 February 2008 using structured and unstructured wave models." *Journal of Coastal Research* (64):1073.

Kim, Kwang-Y, and Chul Chung. 2001. "On the evolution of the annual cycle in the tropical Pacific." *Journal of climate* no. 14 (5):991-994.

Kim, Kwang-Y, and Gerald R North. 1997. "EOFs of harmonizable cyclostationary processes." *Journal of the atmospheric sciences* no. 54 (19):2416-2427.

Kim, Kwang-Y, Gerald R North, and Jianping Huang. 1996. "EOFs of one-dimensional cyclostationary time series: Computations, examples, and stochastic modeling." *Journal of the atmospheric sciences* no. 53 (7):1007-1017.

Kim, Kwang-Yul, Benjamin Hamlington, and Hanna Na. 2015. "Theoretical foundation of cyclostationary EOF analysis for geophysical and climatic variables: Concepts and examples." *Earth-Science Reviews* no. 150:201-218.

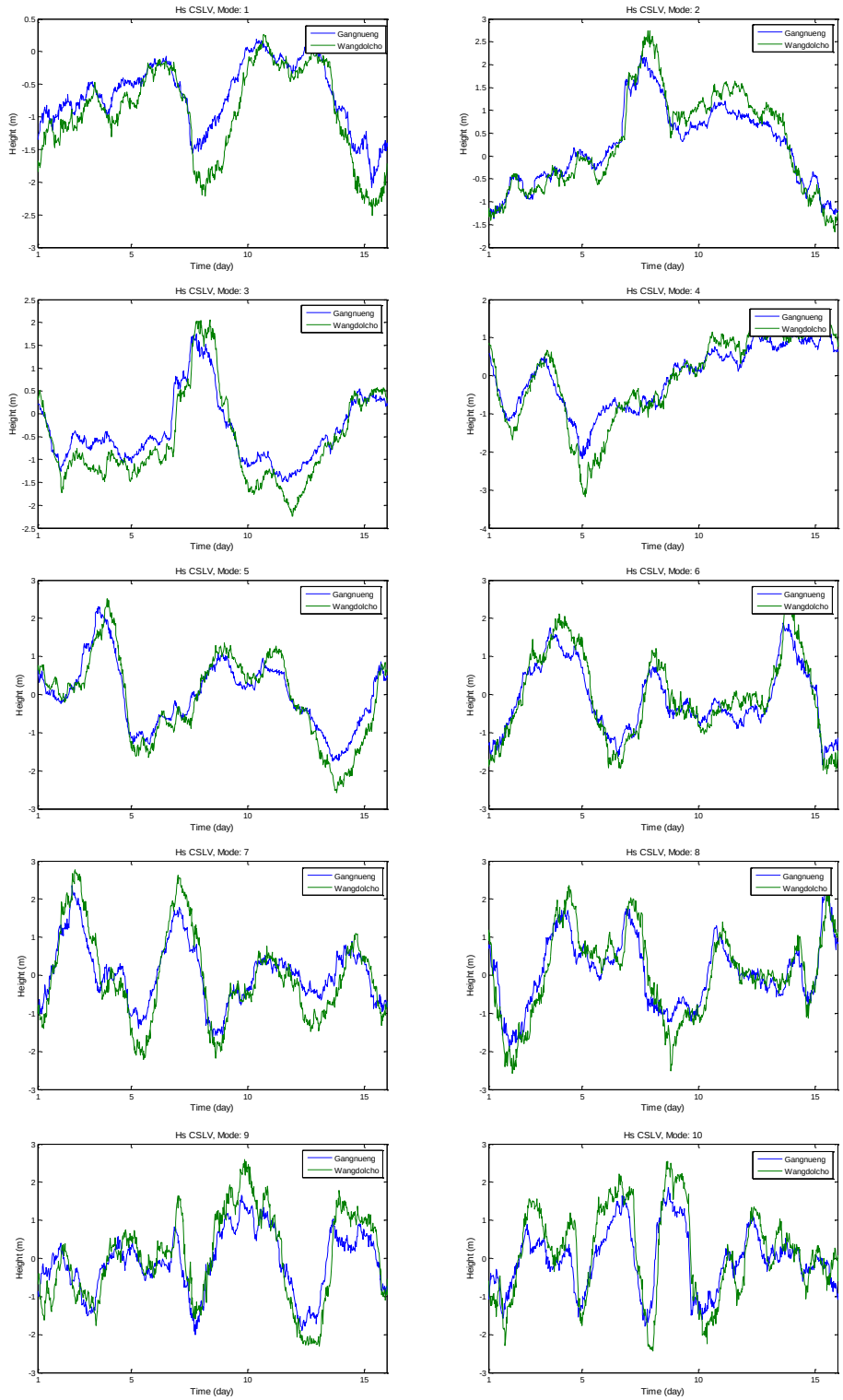
Kim, Tae-Rim, and Kang-Ho Lee. 2008a. "Examinations on the wave hindcasting of the abnormal swells in the East Coast." *Journal of Ocean Engineering and Technology* no. 22 (6):13-19.

Kim, Taerim, and Dong-Young Lee. 2008b. "Wavelet analysis of swells in the East Sea." *Journal of Korean Society of Coastal and Ocean*

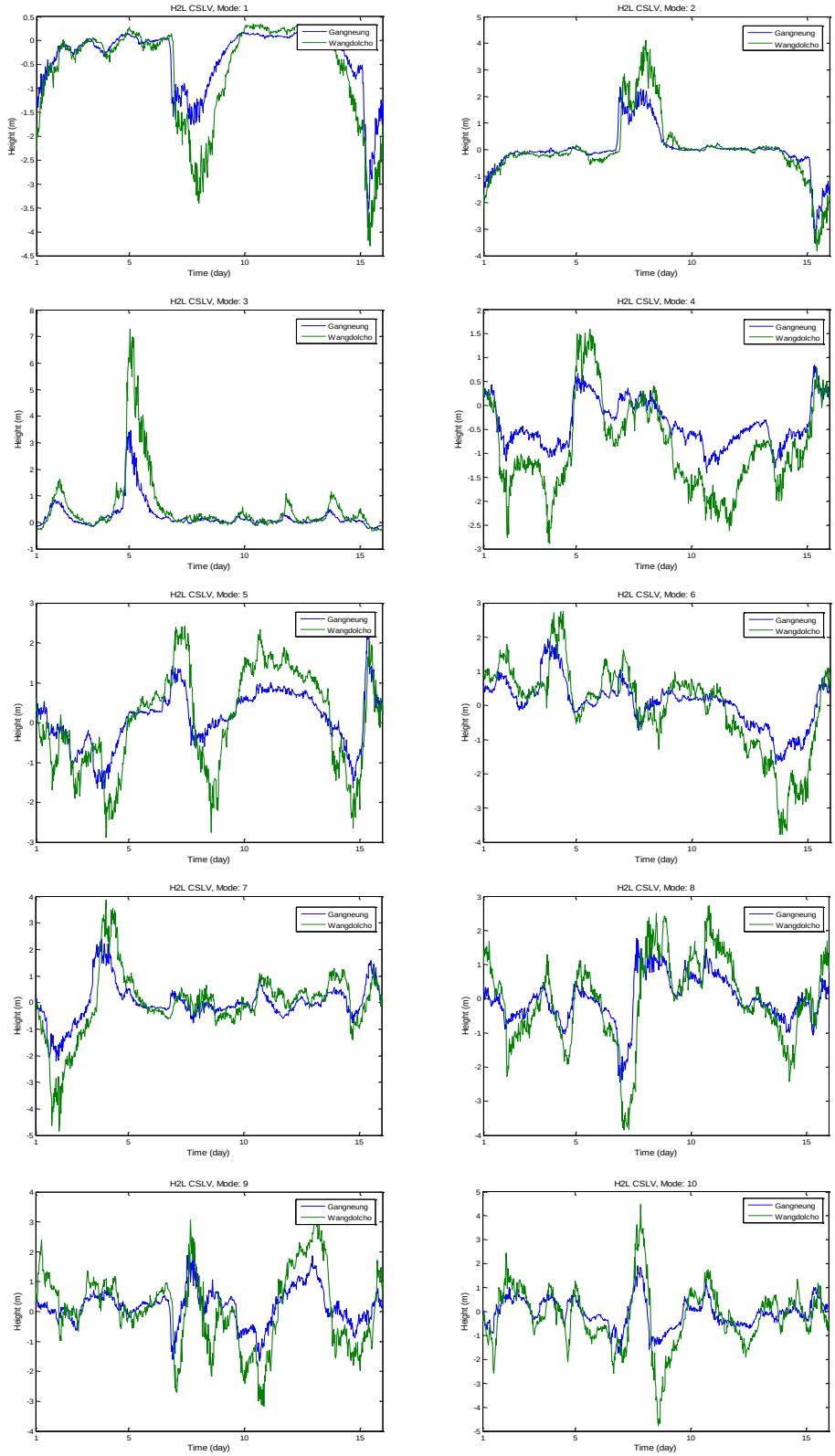
Engineers no. 20 (6):583-588.
Kitaide, M. 1952. "On the mechanism and forecasting of the so-called "Yorimawari" wave." *Marine Report of the Central Meteorological Observatory* no. 2:125–151.
KORDI. 2005. Report on the Deep Sea Design Wave Estimation for Seas Adjacent to Korea. KORDI.
Lee, Changhoon, Suk Jin Ahn, Byeong Wook Lee, Shin Woong Kim, and Seok Jae Kwon. 2014. "Prediction of swell-like high waves using observed data on the east coast of Korea." *Journal of Korean Society of Coastal and Ocean Engineers* no. 26 (3):149-159.
Lee, Dong -Young, and Ki -Cheon Jun. 2006. "Estimation of design wave height for the waters around the Korean Peninsula." *Ocean Science Journal* no. 41 (4):245-254.
Lee, HS, KO Kim, T Yamashita, T Komaguchi, and T Mishima. 2010. "Abnormal storm waves in the winter East/Japan Sea: generation process and hindcasting using an atmosphere-wind wave modelling system." *Nat. Hazards Earth Syst. Sci* no. 10 (4):773-792.
Lee, HS, T Yamashita, T Komaguchi, and T Mishima. 2008. "Numerical simulation of so-called 'Yorimawari-Waves' caused by the low pressures on Feb. 2008 using a meso-scale meteorological model and the wave model." *Annual Journal of Coastal Engineering (JSCE)* no. 55:161-165.
Legler, David M. 1983. "Empirical orthogonal function analysis of wind vectors over the tropical Pacific region." *Bulletin of the American Meteorological Society* no. 64 (3):234-241.
Londhe, SN, and Vijay Panchang. 2006. "One-day wave forecasts based on artificial neural networks." *Journal of Atmospheric and Oceanic Technology* no. 23 (11):1593-1603.
Makarynskyy, Oleg, AA Pires-Silva, Dina Makarynska, and C Ventura-Soares. 2005. "Artificial neural networks in wave predictions at the west coast of Portugal." *Computers & Geosciences* no. 31 (4):415-424.
Mase, H, T Yasuda, TH Tracey, and D Tsujio. 2008. "Forecast and hindcast simulations of wind waves which caused disasters along Toyama Coasts on February 2008." *Annual Journal of Coastal Engineering (JSCE)* no. 55:156-160.
Murguia, JS, and E Campos-Cantón. 2006. "Wavelet analysis of chaotic time series." *Revista mexicana de física* no. 52 (2):155-162.
Nagai, Toshihiko, Tetsuya Hiraishi, Hiroyasu Kwawai, Koji Kawaguchi, Akira Nihei, and Tatsuo Ohkama. 2009. Characteristics of Yorimawari-nami, peculiar Japan-Sea low frequency swell, observed by NOWPHAS seabed wave gauge network. Paper read at The Nineteenth International Offshore and Polar Engineering Conference.
Oh, S-H, and W-M Jeong. 2013. "Characteristics of high waves observed at multiple stations along the east coast of Korea." *Natural Hazards*

and Earth System Sciences Discussions no. 1 (4):3373-3412.
Oh, Sang-Ho, Weon-Mu Jeong, Dong-Young Lee, and Sang-Ik Kim. 2010. "Analysis of the reason for occurrence of large-height swell-like waves in the east coast of Korea." *Journal of Korean Society of Coastal and Ocean Engineers* no. 22 (2):101-111.
Özger, Mehmet. 2010. "Significant wave height forecasting using wavelet fuzzy logic approach." *Ocean Engineering* no. 37 (16):1443-1451.
Prahlada, R, and Paresh Chandra Deka. 2015. "Forecasting of Time Series Significant Wave Height Using Wavelet Decomposed Neural Network." *Aquatic Procedia* no. 4:540-547.
Seo, Kyong-Hwan, and Kwang-Yul Kim. 2003. "Propagation and initiation mechanisms of the Madden-Julian oscillation." *Journal of Geophysical Research: Atmospheres (1984–2012)* no. 108 (D13).
Shahabi, Sajad, Mohammad-Javad Khanjani, and Masoud-Reza Hessami Kermani. 2016. "Significant wave height modelling using a hybrid wavelet-genetic programming approach." *KSCE Journal of Civil Engineering*:1-10.
Strassburg, MW, BD Hamlington, RR Leben, and K-Y Kim. 2014. "A comparative study of sea level reconstruction techniques using 20 years of satellite altimetry data." *Journal of Geophysical Research: Oceans* no. 119 (7):4068-4082.
Tsai, Ching-Piao, Chang Lin, and Jia-N Shen. 2002. "Neural network for wave forecasting among multi-stations." *Ocean Engineering* no. 29 (13):1683-1695.
Wang, YS, CM Lee, and LJ Zhang. 2004. "Wavelet analysis of vehicle nonstationary vibration under correlated four-wheel random excitation." *International journal of automotive technology* no. 5 (4):257-268.
Zamani, Ahmadreza, Dimitri Solomatine, Ahmadreza Azimian, and Arnold Heemink. 2008. "Learning from data for wind–wave forecasting." *Ocean engineering* no. 35 (10):953-962.

APPENDIX. A

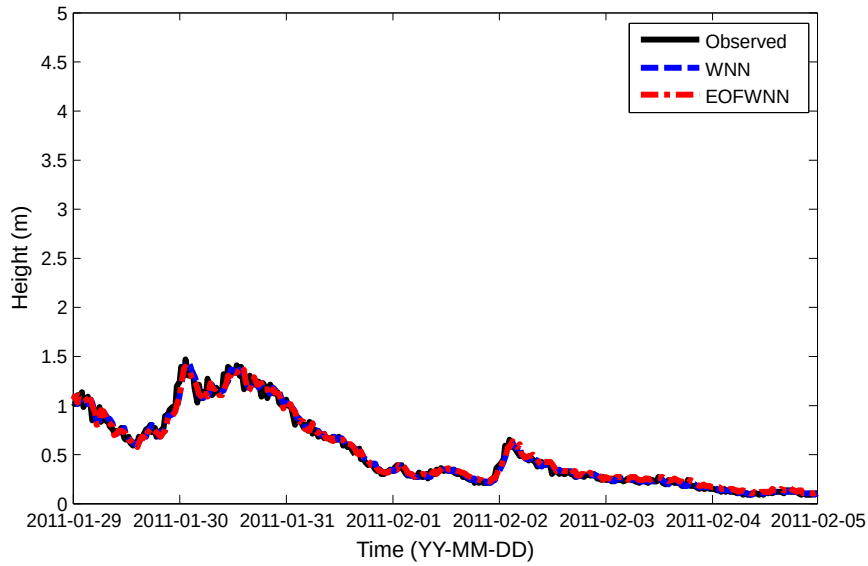


App. A1 CSLV mode 1 - 10 of significant wave height from KIOST

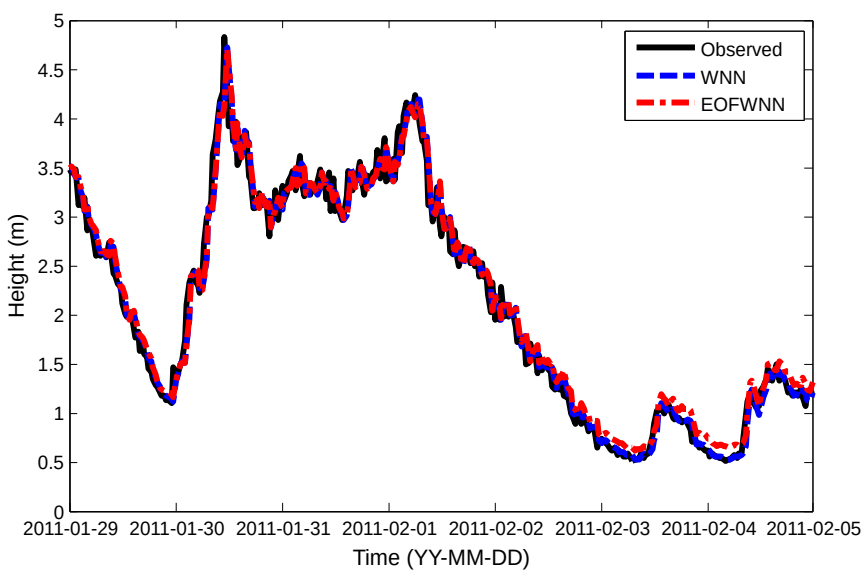


App. A2 CSLV mode 1 - 10 of sudden high wave index

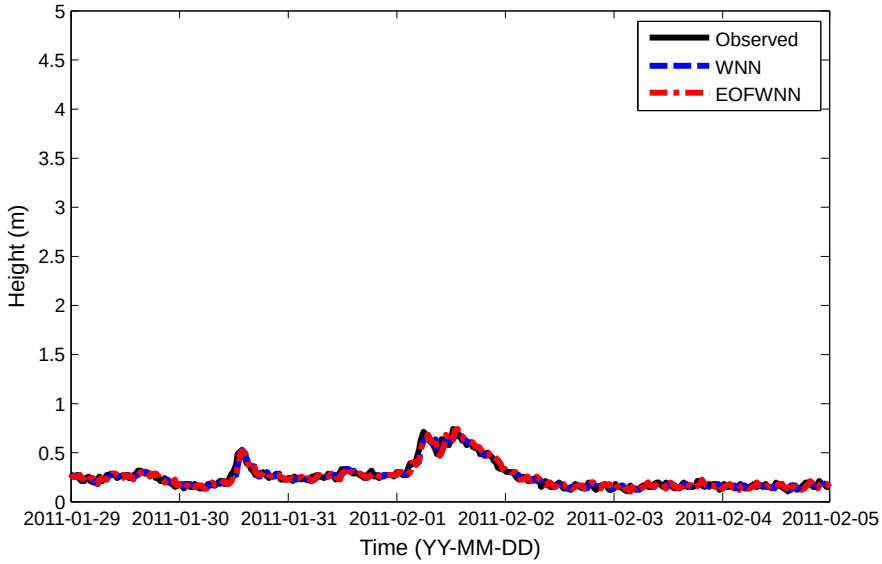
APPENDIX. B



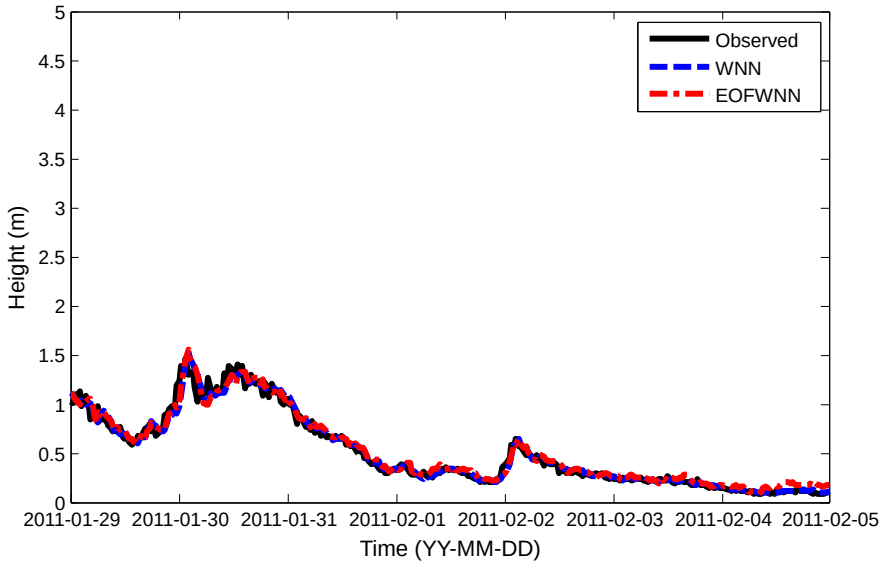
App. B1 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 1 hour lead time in Gangneung



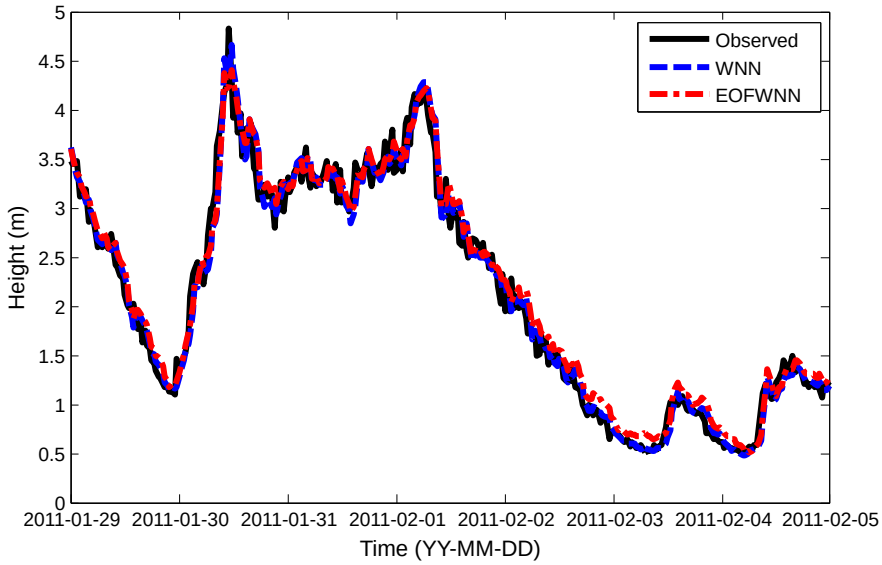
App.B2 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 1 hour lead time in Sakata



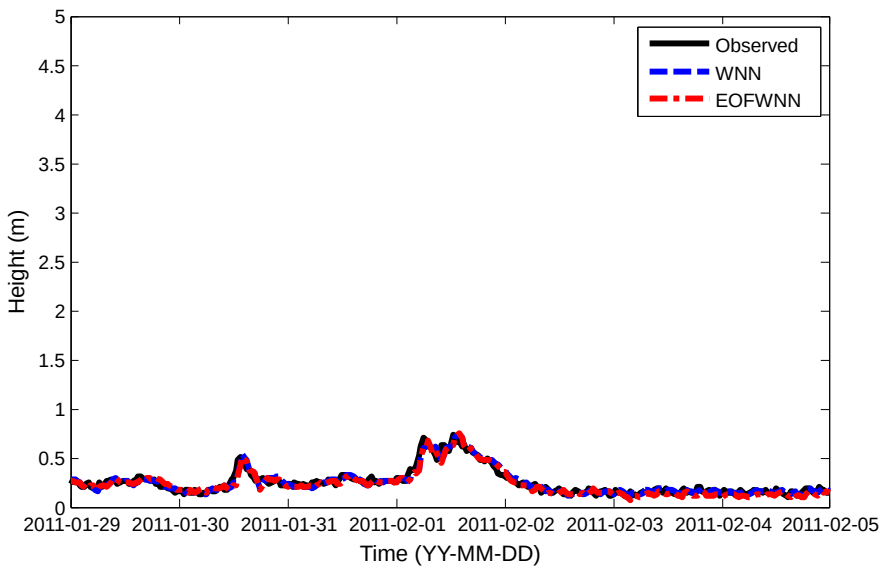
App.B3 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 1 hour lead time in Aomori



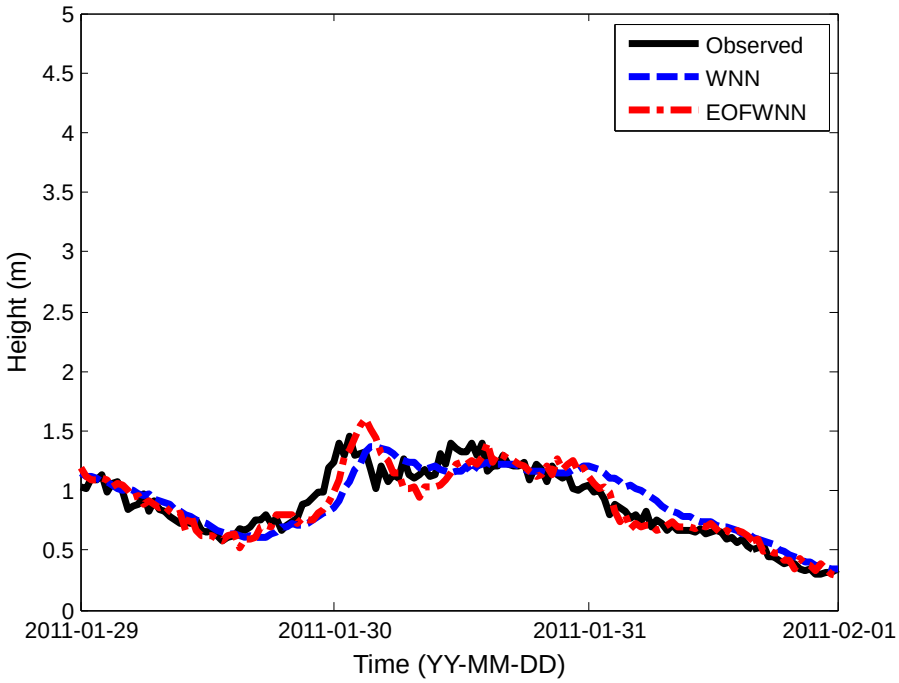
App.B4 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 3 hour lead time in Gangneung



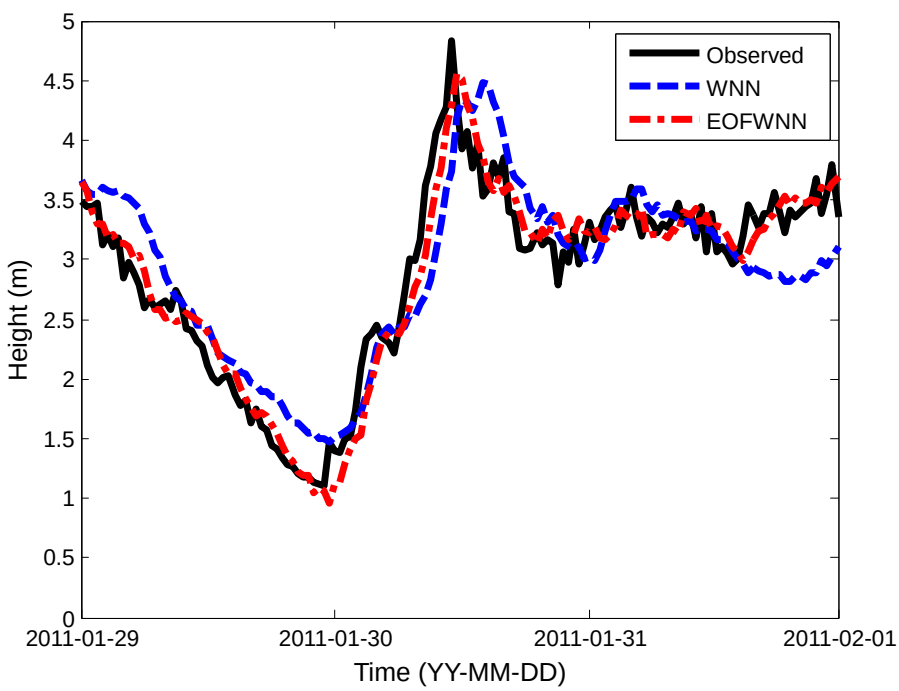
App.B5 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 3 hour lead time in Sakata



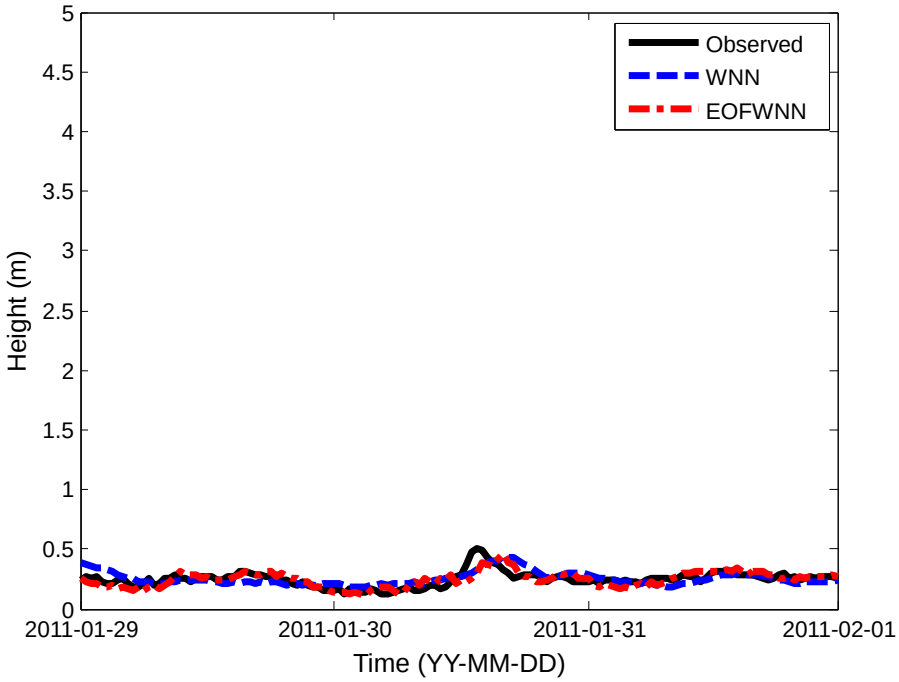
App.B6 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 3 hour lead time in Aomori



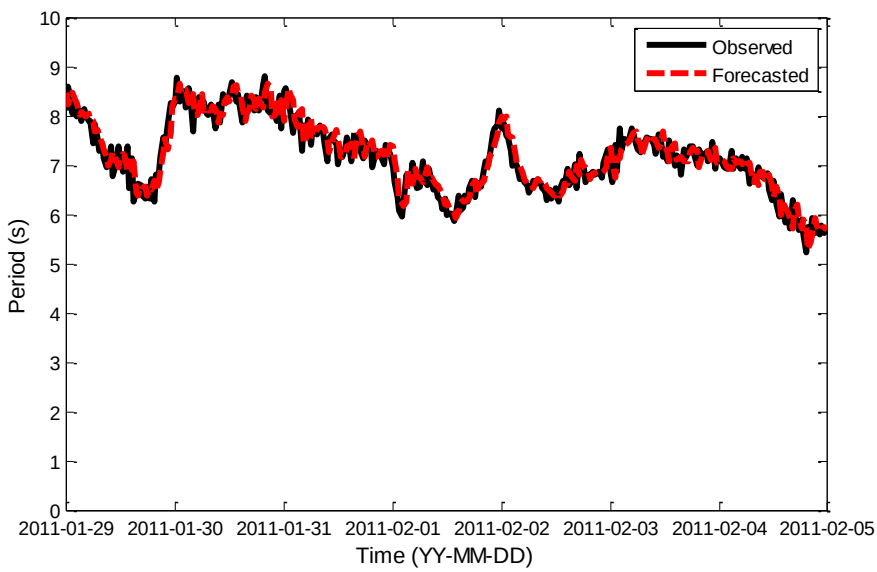
App. B7 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 12 hour lead time in Gangneung



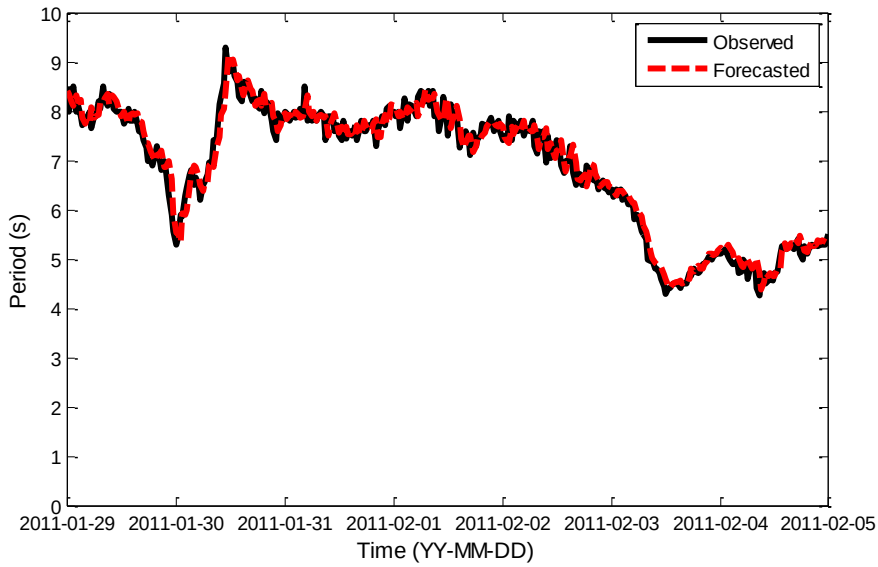
App. B8 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 12 hour lead time in Sakata



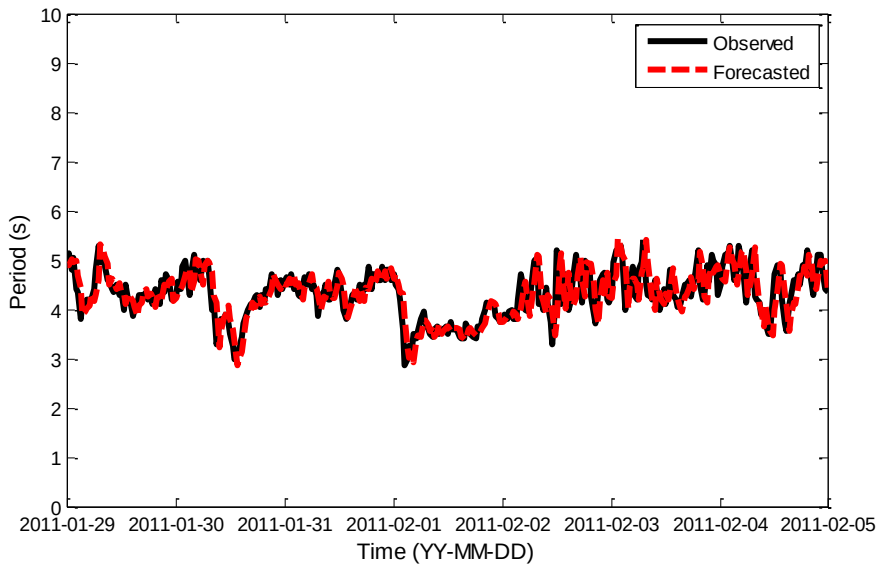
App. B9 Observed and forecasted significant wave heights by WNN and EOFWNN models with 7th decomposition level at 12 hour lead time in Aomori



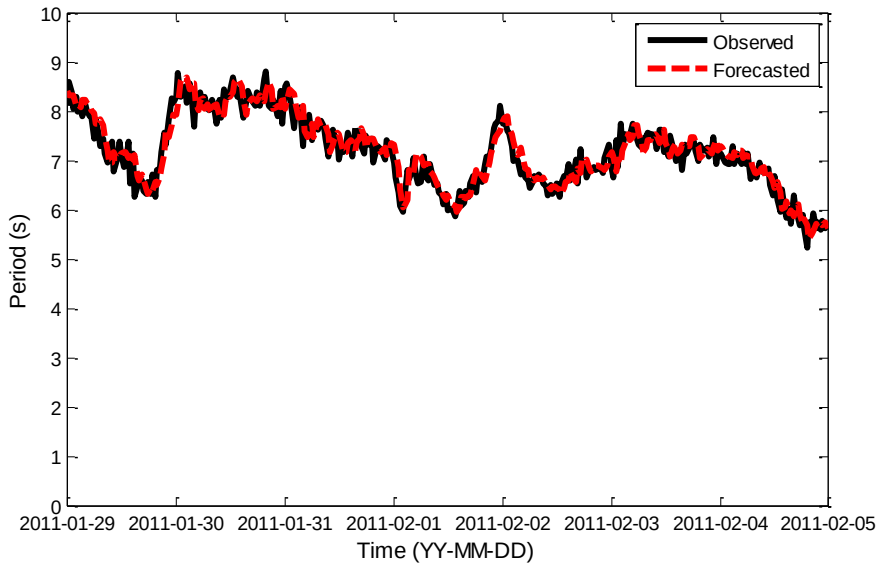
App. B10 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 1 hour lead time in Gangneung



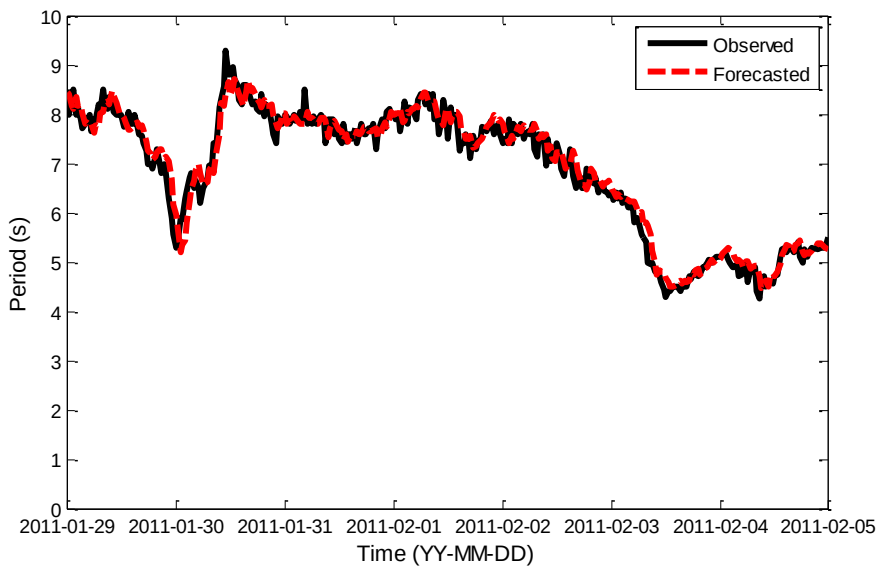
App. B11 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 1 hour lead time in Sakata



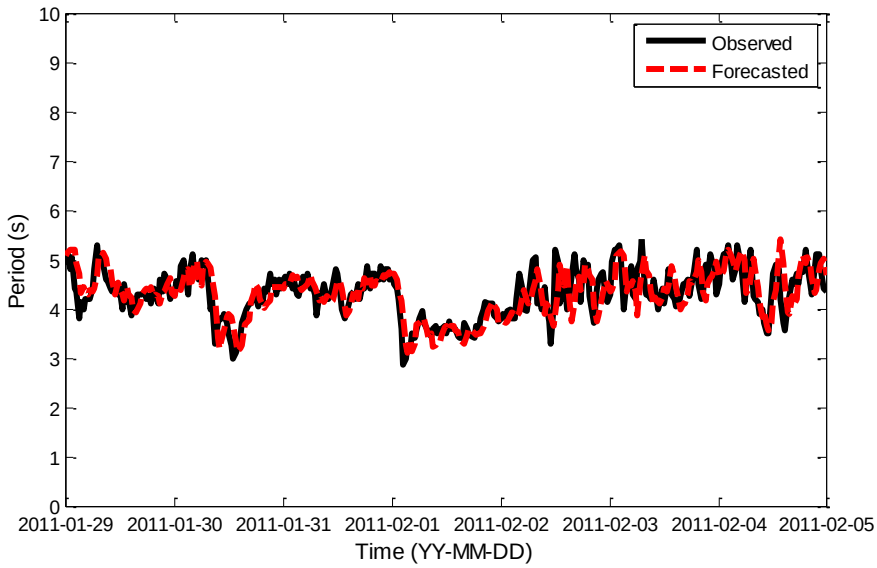
App. B12 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 1 hour lead time in Aomori



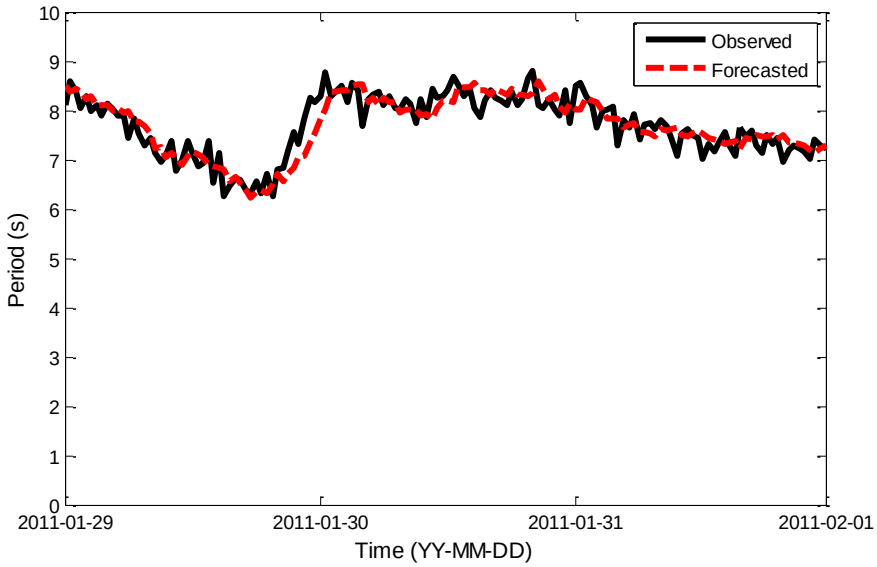
App. B13 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 3 hour lead time in Gangneung



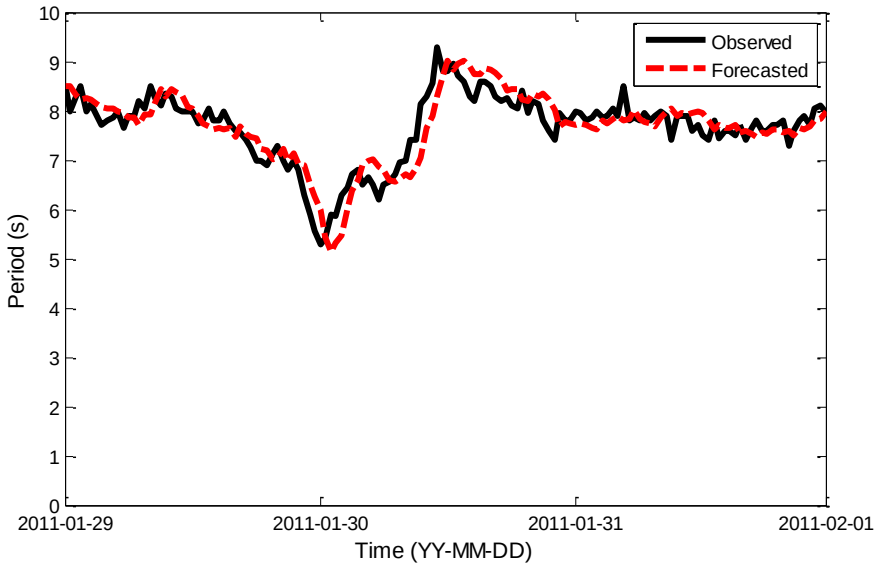
App. B14 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 3 hour lead time in Sakata



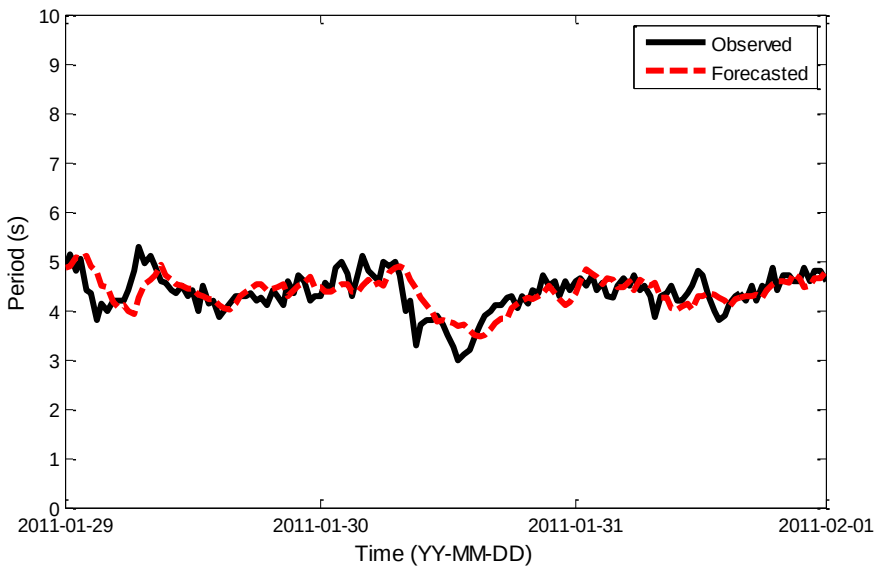
App. B15 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 3 hour lead time in Aomori



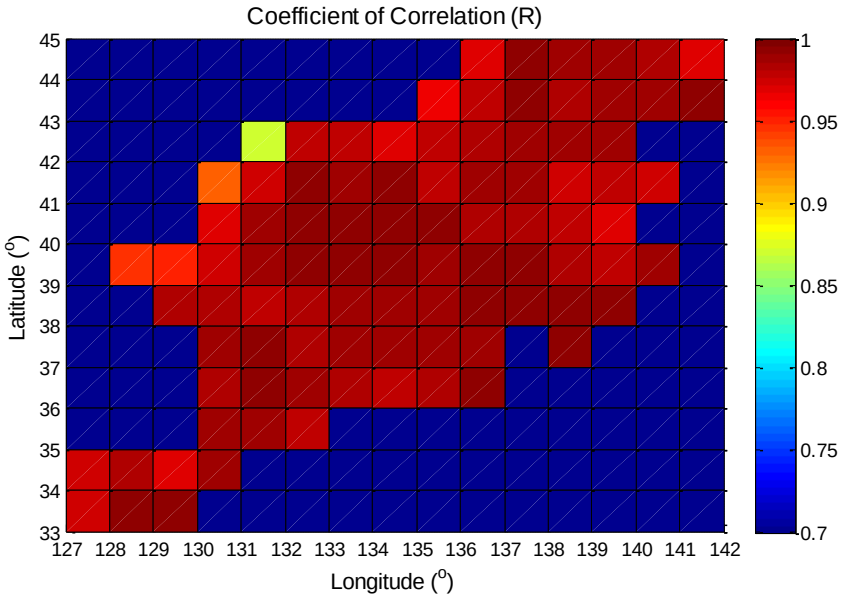
App. B16 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 12 hour lead time in Gangneung



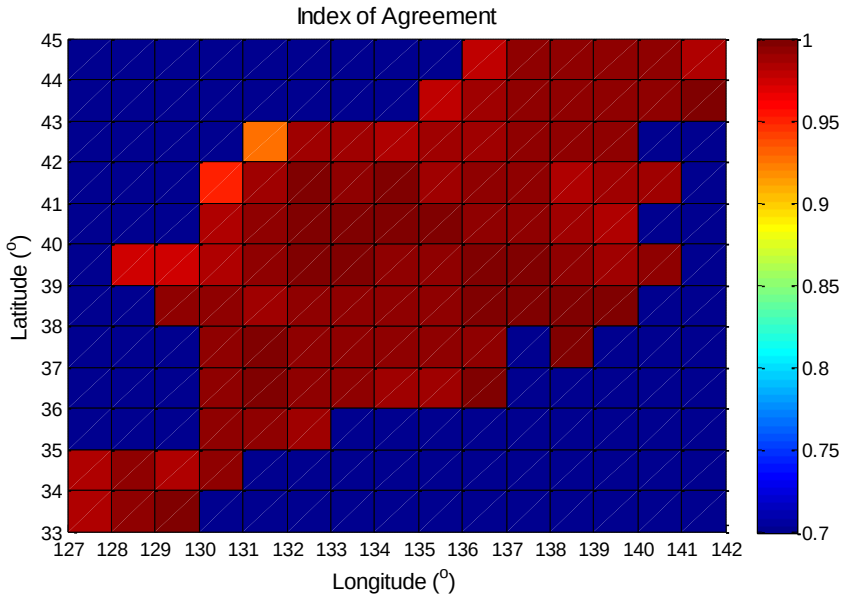
App. B17 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 12 hour lead time in Sakata



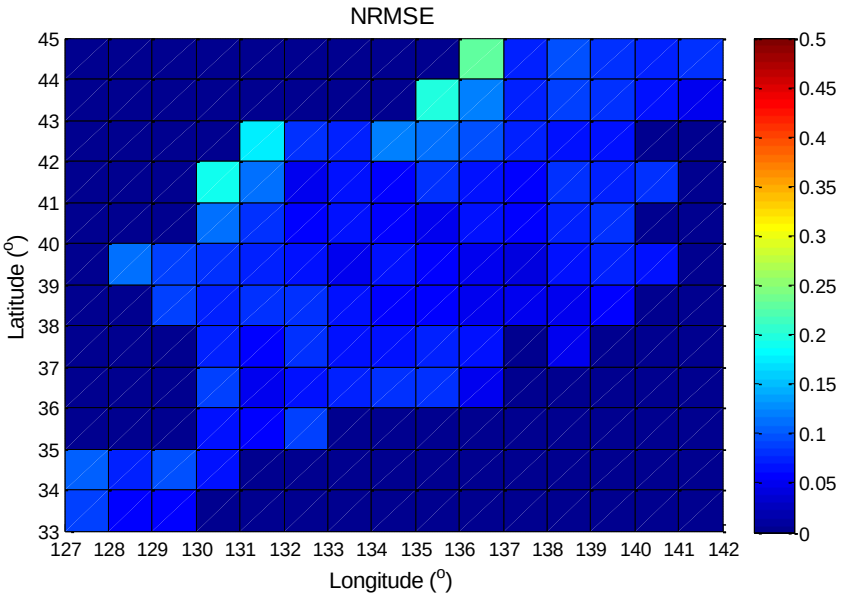
App. B18 Observed and forecasted significant wave period by EOFWNN model with 7th decomposition level at 12 hour lead time in Aomori



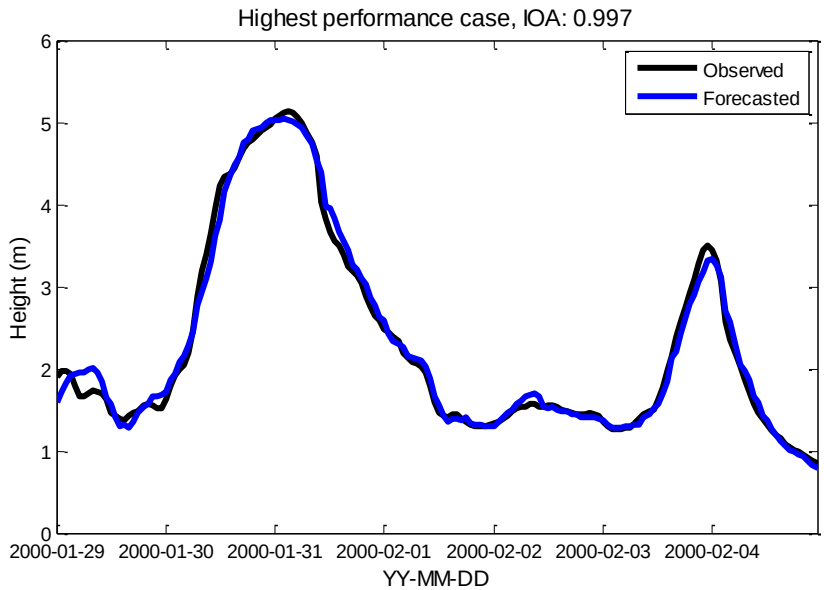
App. B19 Coefficient of correlation of EOFWNN model for numerical significant wave height with 7th decomposition level at 1 hr lead time



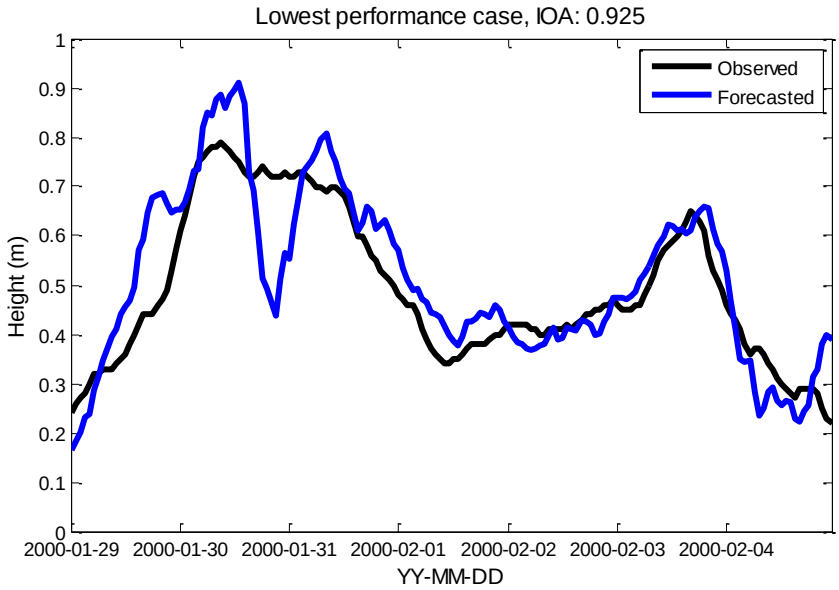
App. B20 Index of agreement of EOFWNN model for numerical significant wave height with 7th decomposition level at 1 hr lead time



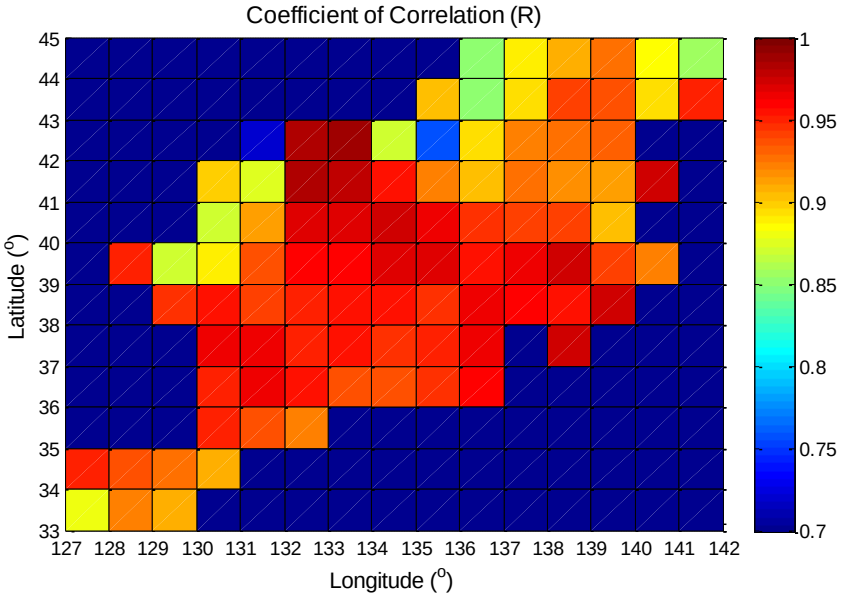
App. B21 *NRMSE* of EOFWNN model for numerical significant wave height with 7th decomposition level at 1 hr lead time



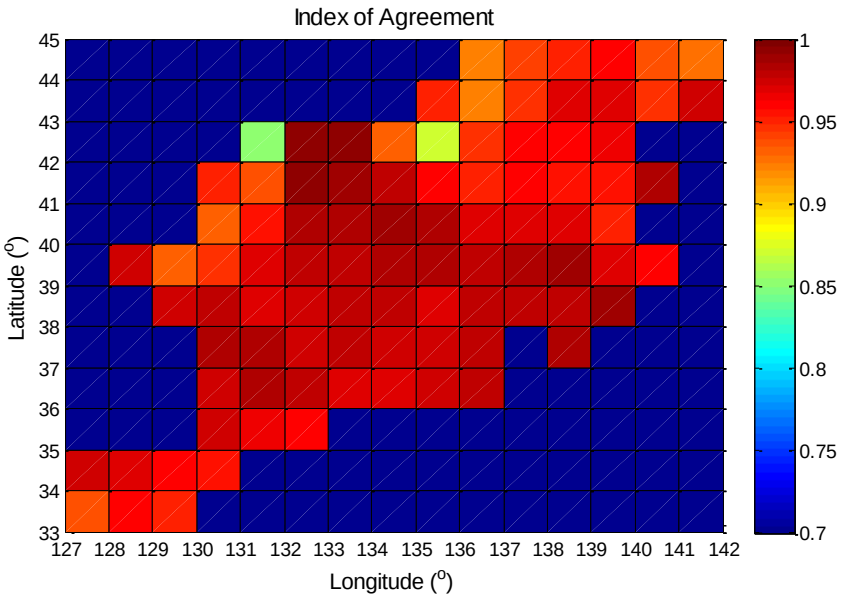
App. B22 Highest performance case of EOFWNN model for numerical significant wave height with 7th decomposition level at 1 hr lead time



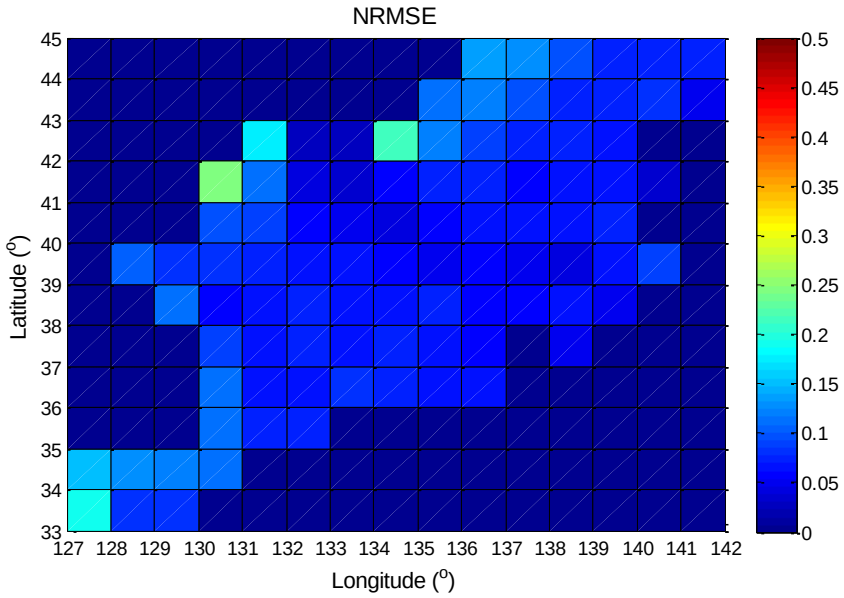
App. B23 Lowest performance case of EOFWNN model for numerical significant wave height with 7th decomposition level at 1 hr lead time



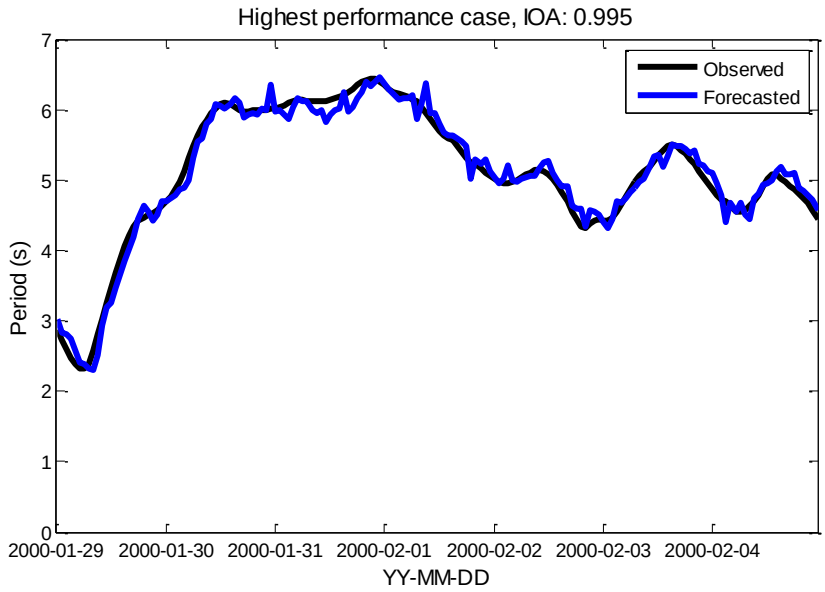
App. B24 Coefficient of correlation of EOFWNN model for numerical peak period with 7th decomposition level at 1 hr lead time



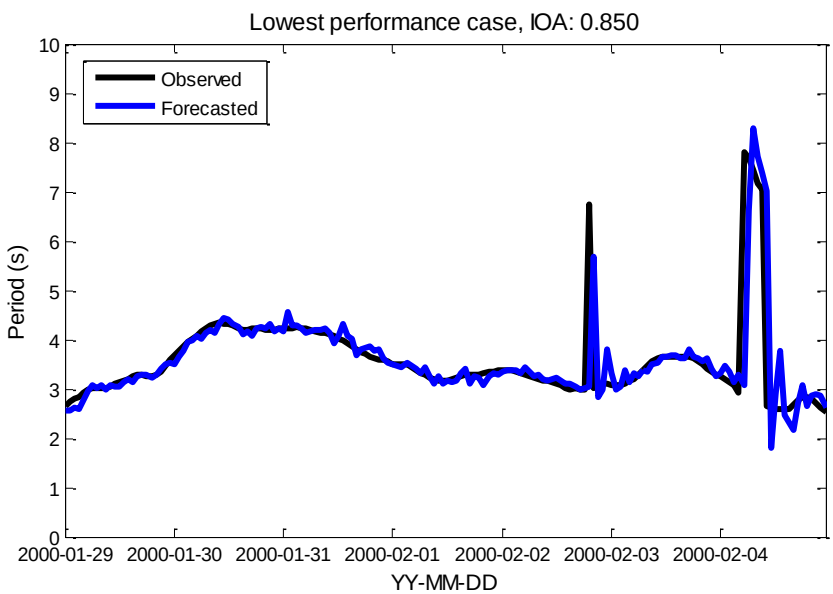
App. B25 Index of agreement of EOFWNN model for numerical peak period with 7th decomposition level at 1 hr lead time



App. B26 *NRMSE* of EOFWNN model for numerical peak period with 7th decomposition level at 1 hr lead time



App. B27 Highest performance case of EOFWNN model for numerical peak period with 7th decomposition level at 1 hr lead time



App. B28 Lowest performance case of EOFWNN model for numerical peak period with 7th decomposition level at 1 hr lead time

국문 초록

동해 돌연고파의 발생 메커니즘과 기계학습 예측모델

서울대학교 대학원

건설환경공학부

오 지 희

동해에서는 예외적으로 큰 파랑이 반복적으로 발생하고 있다. 이 처참한 파랑은 지난 십 년 간 동해에서 50 명 이상의 인명피해를 야기시켰다. 몇몇 연구가들은 돌연고파의 발생 메커니즘과 특성에 대하여 연구해오고 있으나 그 정의는 여전히 모호하고 돌연고파의 특성을 설명하기에는 불충분하다. 또한 돌연고파의 발생은 일일 기상 예보에서 대략적으로 예측되고 있다. 본 연구에서는 새로운 강도 변수를 사용하여 돌연고파를 정의하고 발생 메커니즘을 조사하였다. 다음으로 동해에서의 유의파고와 유의파주기를 예측하였고 마지막으로 앞서 제안된 강도 변수와 동해에서 예측된 유의파고와 유의파주기를 이용하여 돌연고파를 예측하였다.

본 연구에서는 돌연고파 인덱스를 $\Delta H^2 L / \Delta t$ 로 제안하였고, 2005년부터 2012년까지의 강릉과 왕돌초의 파랑 자료를 이용하여 그 값을 계산하였다. 계산된 값에 대하여 분포를 추정한 결과, 누적 분포의 상위 20% 에 해당하는 80 m^3/hr를 돌연고파의 기준으로 삼았다.

다음으로, 돌연고파의 발생 메커니즘을 분석하기 위하여 CSEOF 분석과 회귀분석을 이용하여 돌연고파가 발생하는 동안의 풍속과 해면기압의 공간적 패턴의 시간에 따른 변화를 살펴보았다. 파랑 자료는 강릉과 왕돌초의 자료를 사용하였고, 기상 자료는 NCEP/NCAR 재분석 자료를 사용하였다. 돌연고파의 물리적 기작과 관련있는 모든 CSLV의 성분에는 두 개의 최고점이 있었다. 패턴은 두 그룹으로 분류할 수 있었고, 첫 번째 패턴은 동해의 북동쪽으로 이동하는 저기압과 하루 동안 지속적으로 부는 동풍에 의해 파고의 첫 번째 최고점이 생성되었고, 반면 두 번째 최고점은 강한 풍속에 의한 것으로 보인다. 두 번째 패턴은 동해안에서의 풍속에 의해 첫 번째 파고의 최고점이 생성되고 두 번째 최고점은 원해에서의 풍속에 영향을 받은 것으로 보인다.

여러 지점에서의 유의파고 및 유의파주기를 동시에 예측하기 위하여, EOF 분석과 웨이블릿 분석과 인공신경망 모형을 결합한 EOFWNN 모델은 개발하였다. 본 연구에서 사용된 파랑 자료는 동해의 여덟 지점에서 관측된 자료를 사용하였고 기상 자료는 NCEP/NCAR 재분석 자료를 사용하였다. EOFWNN 모델을 이용하여 유의파고를 예측한 결과는 강릉, 사카다, 아오모리에서의 여러 선행기간에 대하여 웨이블릿과 인공신경망 모형을 결합한 WNN 모델과 비교하였다. EOFWNN 모델은 WNN 모델

보다 웨이블릿 분해 단계와 상관없이 더 긴 선행 기간에 대하여 더 높은 정확도를 보여주었다. EOFWNN 모델을 이용하여 유의파주기도 예측하였다. 유의파주기의 결과 역시 상당히 높은 정확도를 보여주었다. 또한 제안된 모델의 더 넓은 지역의 자료에 대한 적용 가능성을 검증하기 위하여 파랑 수치해석 모델 자료에 적용하였다. 1 시간과 3 시간의 선행기간에 대한 결과 역시 상당히 높은 정확도를 보여주었다.

제안된 돌연파고의 강도 변수와 EOFWNN 모델에 의하여 예측된 유의파고 및 유의파주기를 이용하여 돌연고파가 감지되고 예측되었다. 24 시간 선행기간에 대해 예측된 파랑 자료로부터 강릉과 사카타에서의 $\Delta H^2 L / \Delta t$ 값이 계산되었다. 비록 관측된 자료에 의한 결과와 예측된 자료에 의한 결과 간에 약간의 차이는 있지만, 돌연고파는 정확하게 감지되었다.

주요어: 인공신경망; 경험직교함수; 유의파랑; 돌연고파; 파랑 예측; 웨이블릿.

학번: 2012-30953