

군집분석을 이용한 국지해일모델 지역확장

이다운* · 서장원 · 윤용훈

기상청, 기상연구소 해양기상지진연구실
(2006년 8월 25일 접수; 2006년 10월 18일 승인)

Regional Extension of the Neural Network Model for Storm Surge Prediction Using Cluster Analysis

Da-Un Lee*, Jang-Won Seo and Yong-Hoon Youn

Marine Meteorology & Earthquake Research Laboratory,
Meteorological Research Institute, KMA

(Manuscript received 25 August 2006; in final form 18 October 2006)

Abstract

In the present study, the neural network (NN) model with cluster analysis method was developed to predict storm surge in the whole Korean coastal regions with special focuses on the regional extension. The model used in this study is NN model for each cluster (CL-NN) with the cluster analysis. In order to find the optimal clustering of the stations, agglomerative method among hierarchical clustering methods was used. Various stations were clustered each other according to the centroid-linkage criterion and the cluster analysis should stop when the distances between merged groups exceed any criterion. Finally the CL-NN can be constructed for predicting storm surge in the cluster regions. To validate model results, predicted sea level value from CL-NN model was compared with that of conventional harmonic analysis (HA) and of the NN model in each region. The forecast values from NN and CL-NN models show more accuracy with observed data than that of HA. Especially the statistics analysis such as RMSE and correlation coefficient shows little differences between CL-NN and NN model results. These results show that cluster analysis and CL-NN model can be applied in the regional storm surge prediction and developed forecast system.

Key words: cluster analysis, storm surge prediction, neural network model

1. 서론

폭풍 해일을 예측하기 위해 수치모델이 가장 널리 이용되고 있으나 (문일주 등, 2000; 해양연구소, 1985-1988; Choi *et al.*, 2003) 우리나라와 같은 좁고 복잡한 지형 조건에 있어서 넓은 격자에 대한 수치모델만으로 정확한 예측을 하기에는 무리가 있다. 따라서 지속 (Persistence) 모델, 마르코프 (Markov) 모델, 자기회귀 (Autoregressive) 모델, 신경망 (Neural network)

모델 등과 같은 통계적 방법이 연구에 사용되고 있다. 이전 연구 (이다운 등, 2005)에서도 신경망 모델을 이용하여 폭풍해일을 예측하기 위해서 모델 구조와 입력 변수의 선택에 관한 실험을 하였다. 입력층에 사용되는 입력변수로는 부산, 여수, 완도 등의 각 지점 바람과 기압 자료를 사용하였으며 출력층에서는 목표변수인 폭풍해일 자료를 사용하였다. 여러 번의 실험을 토대로 3개의 은닉마디, 1개의 은닉층을 포함하는 MLP (Multi-Layer Perceptron) 구조를 이용하여 세 지점의 폭풍 해일 예측을 위한 지점별 신경망 모델을 구축하였다.

이러한 최적 모델을 전 지역으로 확장한다면 매우 정확한 지점별 모델을 구축할 수 있지만, 그러한 작업을 위해서는 많은 시간과 여러 번의 반복적인 실험 결과가 토대가 되어야 한다. 또한 그 지점에 대한 데이터

*Corresponding Author: Da-Un Lee, Marine Meteorology & Earthquake Research Lab., Meteorological Research Institute, 460-18, Dongjak-gu, Shindaebang-2-dong, 156-729, Seoul Korea.
Phone : +82-2-847-2495, Fax : +82-2-847-2496
E-mail: ddaung@metri.re.kr

가 충분하지 않을 경우 모델을 구축하는데 있어서 어려움이 발생하게 된다 (Faraway and Chatfield, 1995). 이러한 이유로 지점별 모델을 해안 전 지역으로 확장하기에는 무리가 있으며 본 연구에서는 전 지점을 몇 개의 군집으로 분류하기 위한 객관적 기준을 제공하는 군집 분석을 실시하였으며 군집별 모델을 구축하여 기존의 지점별 모델 결과와 비교하였다. 군집별 모델을 사용함으로써 각각의 지점별 모델을 구축하는데 드는 시간과 비용을 줄일 수 있으며 지점별 데이터가 부족한 경우 대체해서 사용할 수 있다는 장점이 있다. 그러나 군집을 제대로 설정하지 못할 경우 모델링 자체가 제대로 이루어지지 않아 지점별 모델을 사용한 경우보다 정확도 측면에서 크게 나빠질 수 있다.

기상·해양 등의 응용 분야에서 군집 분석 또는 신경망 모델을 이용하여 자료를 분석한 사례를 많이 찾아볼 수 있다. 문영수 (1990)는 군집 분석을 이용하여 한반도 강수 지역을 구분하고 군집별 특성을 분석하였으며 고정용 등 (2005)은 한반도 우기의 강수 특성과 지역 구분을 위해서 인자 분석 후 군집 분석을 사용하여 강수 특성별로 군집을 나누었다. 또한 Tissot *et al.* (2003), Tsai and Lee (1999)는 정확한 해수면 또는 조위 예측을 위해서 입력 변수와 예측 기간을 달리하여 신경망을 구축하고 이 결과를 관측값, 조화분석 (Harmonic Analysis ; HA) 결과와 비교하였다. Cox *et al.* (2002)은 신경망 모델을 이용한 단기간 해수면 예측의 결과를 조화분석, 선형모형 등의 결과와 비교하였으며 Sztobryn (2003)은 여러 구조의 신경망을 이용하여 폭풍 해일이 있을 때의 해수면 높이를 예측하여 그 결과를 기존 수치 모델의 결과와 비교하였다. 군집 분석 후 군집별 모델을 적용한 연구는 많지 않으며 손건태 등 (2005)의 연구에서 통계 모델 훈련과 검증에 사용할 데이터를 나누는 과정에서 군집분석을 사용한 경우를 찾아볼 수 있다.

2장에서는 본 연구에서 사용하게 되는 군집 분석과 신경망 모델에 대해 설명하며 3장에서는 군집 분석 결과와 각 군집별 국지 해일 예측 모델링 결과를 기존에 구축한 지점별 모델 결과, 천문조위와 비교한다. 4장에서는 연구의 결론 및 향후과제에 대해 기술할 것이다.

2. 분석 기법

2.1 군집 분석 (Cluster analysis)

군집 분석은 사전에 어떤 그룹에 속해있는지 알 수

없는 개체들에 대해서 군집 특성과 상관성이 높은 변수들이 주어져 있을 때 변수들을 이용하여 유사성 (similarity)에 대한 거리를 구하고 그 거리로부터 각각의 개체들을 몇 개의 군집으로 집단화하는 분석 기법이다. 즉, 주어진 변수들을 이용하여 각각의 개체들 사이의 유사성에 대한 거리를 구하고 거리가 가까우면 같은 군집에 속하게 되며 일정거리 이상으로 먼 개체들은 다른 군집에 속하게 된다 (Wilks, 1995).

p 개의 변수에 대해서 각각 n 개의 관측치가 존재한다고 했을 때 자료는 식 (1)과 같이 나타낼 수 있다.

$$X = [X_1 \ X_2 \ \cdots \ X_n] = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix} \quad (1)$$

여기서 X_i 는 i 번째 개체에 대해서 p 개의 변수로 이루어진 열벡터를 나타내며 X_{ki} 는 i 번째 개체의 k 번째 변수 값을 나타낸다. 각각의 개체 X_i 와 X_j 사이의 거리를 계산하여 유사성을 측정하게 되는데 가장 많이 사용되는 척도는 유클리디안 (Euclidean) 거리이며 개체 X_i 와 X_j 사이의 유클리디안 거리 d_{ij} 에 대한 계산식은 식 (2)와 같다.

$$\begin{aligned} d_{ij} &= \sqrt{(X_i - X_j) \cdot (X_i - X_j)} \\ &= \left[\sum_{k=1}^p (X_{ki} - X_{kj})^2 \right]^{1/2} \end{aligned} \quad (2)$$

변수들 간의 단위가 매우 다를 때는 모든 변수를 표준화하여 사용하게 되지만 변수들 간에 높은 상관관계가 존재할 때는 거리척도 자체에 상관관계까지 고려한 통계적 거리인 마할라노비스 (Mahalanobis) 거리 등을 사용하기도 하며 그 식은 (3)과 같다.

$$\begin{aligned} d_{ij} &= \sqrt{(X_i - X_j) S^{-1} (X_i - X_j)}, \\ \text{where } S &= \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix} \end{aligned} \quad (3)$$

여기서 S 는 p 개의 변수들간의 표본 분산-공분산 행렬을 나타낸다. 계보적 군집 (Hierarchical clustering) 방법 중 병합적 (agglomerative) 방법에서는 이러한 방법으로

각 개체들간의 거리를 구한 다음 유사성이 높은 개체끼리 군집을 이루어나가게 되는데 그 연결방법에 따라 최단 연결법 (single linkage method), 최장 (complete) 연결법, 평균 (average) 연결법, 중심 (centroid) 연결법 등이 있다. 본 연구에서는 중심 연결법을 사용하였으며 계산 방법은 다음과 같다. 먼저 군집 CL_j 에 속하는 개체의 수가 n_{CL_j} 일 때 CL_j 의 중심 \bar{X}_j 는 식 (4)와 같이 구할 수 있다.

$$\bar{X}_j = \sum_{i \in CL_j} \frac{X_i}{n_{CL_j}} \quad (4)$$

개체의 수가 각각 n_{CL_j} , n_{CL_k} 인 군집 CL_j 와 CL_k 의 중심이 \bar{X}_j , \bar{X}_k 일 때 두 군집 사이의 중심 거리 D_{jk} 는 식 (5)와 같다.

$$D_{jk} = |\bar{X}_j - \bar{X}_k|^2 \quad (5)$$

또한 두 군집이 다시 결합할 때 새로운 군집의 중심 \bar{X} 는 식 (6)과 같은 가중 평균의 형태로 구할 수 있다.

$$\bar{X} = \frac{n_{CL_j}\bar{X}_j + n_{CL_k}\bar{X}_k}{n_{CL_j} + n_{CL_k}} \quad (6)$$

이러한 방법을 반복하면서 군집 사이의 중심 거리를 구한 후 가장 가까운 군집 또는 개체와 새로운 군집을 형성해 나가게 된다 (성용현, 2002).

2.2 신경망 모델(Neural network model)

신경망 모델은 대용량의 데이터로부터 일정한 규칙 또는 패턴을 찾아내는 데이터 마이닝 분류 기법의 하나로서 복잡한 구조를 가진 자료에서의 예측 문제를 해결하기 위해서 사용되는 비선형 모델 (nonlinear models)이다. 이 방법은 입력변수와 목표변수 사이의 관계를 설명하기는 어렵다는 단점이 있으나 예측력이 매우 높다는 장점을 가지고 있다 (김대수, 2001). 신경망 모델은 은닉마디 (hidden units)라고 불리는 구성요소를 포함하고 있으며 이는 인간의 신경세포를 모방화한 것으로서, 각 은닉마디는 입력변수들의 결합을 수신하여 목표변수에 전달한다. 이때 결합에 사용되는 계수들을 연결강도 (synaptic weights)라고 부르며, 활성화수

(activation function)는 각각의 입력 값을 변환하고 이를 다시 입력으로 사용하는 다른 마디로 출력하게 된다 (강현철 등, 1999).

여러 가지 다양한 모델이 있으나 그 중에서도 자료 분석을 위해 널리 사용되며 또한 해일 예측 실험에서 가장 안정적이며 좋은 결과를 보인 모델은 MLP 신경망이다 (Potts, 2000). MLP는 입력층 (input layer), 은닉마디로 구성된 은닉층 (hidden layer), 출력층 (output layer)으로 구성된 전방향 (feed-forward) 신경망이다. 입력층은 각각의 표준화된 입력변수에 대응되는 마디들로 이루어져 있으며 한개 이상의 은닉층은 여러 개의 은닉마디로 구성되어 있는데 각 은닉마디는 입력층으로부터 전달되는 변수 값들의 선형결합을 비선형함수로 처리하여 출력층 또는 다른 은닉층에 전달하는 역할을 한다. 출력층은 목표변수에 대응하는 마디들로 이루어지며 그 구조는 Fig. 1과 같다. p 개의 입력변수로 이루어진 입력마디를 가진 입력층, N 개의 은닉마디를 가진 하나의 은닉층, 하나의 목표변수로 이루어진 출력층으로 구성된 MLP 신경망 구조를 나타내며 이를 수식으로 표현하면 식 (7)과 같다.

$$\begin{aligned} H_1 &= f_1(w_{10} + w_{11}x_1 + \dots + w_{1p}x_p) \\ H_2 &= f_2(w_{20} + w_{21}x_1 + \dots + w_{2p}x_p) \\ &\dots \\ H_N &= f_N(w_{N0} + w_{N1}x_1 + \dots + w_{Np}x_p) \\ E(y) &= g(w_0 + w_1H_1 + w_2H_2 + \dots + w_NH_N) \end{aligned} \quad (7)$$

여기서 w_{ij} , w_k ($i, j, k \neq 0$)는 각 노드에 대한 연결 강도이며 w_{i0}, w_0 는 선형 식의 절편 모수 (bias parameter)를 나타낸다. x_i , $i=1, \dots, p$ 는 입력변수 값이며 H_i , $i=1, \dots, N$ 는 N 개의 은닉마디, $E(y)$ 는 예측된 목표변수 값을 나타낸다. f_1, f_2, \dots, f_N, g 는 활성화 함수를 나타낸다.

결합함수 (combination function)는 입력층 또는 은닉층의 마디들을 결합하는 함수를 말하며 MLP 구조에서는 선형함수를 사용하여 각각의 표준화된 입력변수에 연결강도를 곱한 값들의 선형결합 형태로 사용하였다. 활성화함수는 입력마디 또는 은닉마디의 결합을 변환하는 함수를 의미하며 f_1, f_2, \dots, f_N 은 은닉층에서의 활성화함수, g 는 출력층에서의 활성화함수를 나타내는데 MLP 신경망에서 은닉층에서의 활성화함수로는 식 (8)의 로지스틱 (logistic) 함수, 하이퍼볼릭 탄젠트(hyperbolic

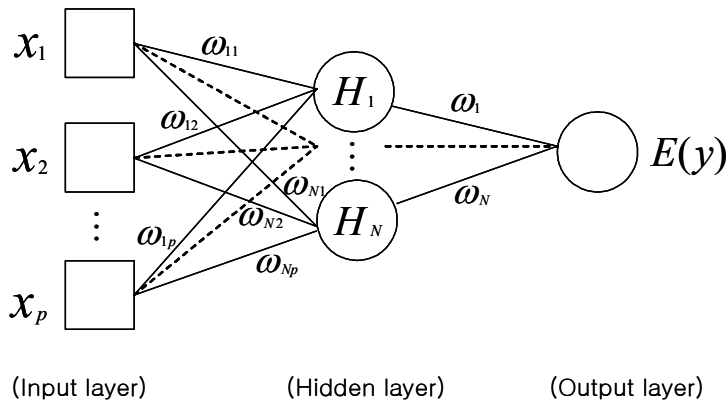


Fig. 1. MLP structure with one hidden layer, N hidden units.

tangent) 함수 등과 같은 시그모이드 (sigmoid) 활성화 함수를 이용한다.

$$\begin{aligned} \text{logistic}(x) &= \frac{1}{1 + e^{-x}} \\ \text{tanh}(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned} \quad (8)$$

로지스틱 함수는 예측값을 (0, 1) 구간으로, 하이퍼볼릭 탄젠트 함수는 예측값을 (-1, 1) 구간으로 제한하므로 일반적으로 하이퍼볼릭 탄젠트 함수는 시계열 자료 예측에 있어서 출력단계에 적용하지 않는다. 출력층에서의 활성화함수는 목표변수의 특성에 따라 항등 함수 (identity function), 로지스틱 함수 등을 사용한다 (Potts, 2000).

3. 분석 결과 및 검증

3.1 군집분석

군집분석을 위한 변수로는 2000년부터 2004년까지 우리나라를 통과한 태풍에 대한 각각의 최대 폭풍 해일 높이와 각 지점별 위·경도 등의 8개의 변수를 이용하였으며 모든 변수는 표준화하여 사용하였다. 또한 변수들 간의 대부분의 상관계수가 높지 않으므로 유클리디안 거리를 이용하여 각 개체간의 유사성을 계산하였다.

2000년부터 2004년까지의 해수면 높이에 대한 조사원 자료와 기상청 AWS 자료를 이용 가능한 14개 지

점이 군집 분석 대상이 되었으며 군집분석을 위한 지점의 수는 이용 가능한 자료가 증가할수록 늘어나게 될 것이다. 해일 값에 결측치가 존재하는 마산의 경우 가장 가까운 두 지점의 최고 해일에 대한 평균값으로 대체하여 분석하였다.

계보적 군집방법을 사용하였으며 유사성이 밀접한 개체를 먼저 묶어나가는 병합적 방법을 사용하였다. 병합적 방법의 첫 단계에서는 각 개체가 각각의 군집을 형성하며 마지막 단계에서의 군집의 개수는 1이 된다. 또한 각 군집 사이의 연결 방법으로는 이상치의 영향을 가장 받지 않는 중심 연결법을 사용하였다.

그 결과는 Fig. 2의 나무구조도 (Dendrogram)과 같으며 14개 지점에 대한 계보적 군집 형태를 보이고 있다. 각 지점의 위치 정보인 위도, 경도 자료뿐만 아니라 각 태풍에 대한 해일 정보 또한 입력 변수로 사용하였으나 대체적으로 근접한 지점끼리 군집을 이루고 있는 것을 확인할 수 있으며 이는 근접한 지점들의 태풍별 해일고 또한 유사한 패턴으로 나타나고 있음을 반영한 결과라 생각된다.

Fig. 3은 각 군집화 단계 (stage)별 두 개체 또는 두 군집 사이의 중심간 거리를 나타낸다. 병합적 방법에 있어서 군집의 수를 몇 개로 정해야 하는지 정해진 기준은 없으나 Fig. 3을 토대로 하여 군집 사이의 중심간 거리에 대한 차이가 급격하게 증가하는 시점에서 군집화를 끊어주는 방법을 고려해볼 수 있다. 즉 단계 6까지는 중심간 거리가 완만하게 증가하다가 단계 6과 단계 7의 군집 중심간 거리에 있어서 차이가 급격하게 나타나므로 단계 6까지만 군집화를 수행하게 되며 군집 분석 결과는 Fig. 4와 같다.

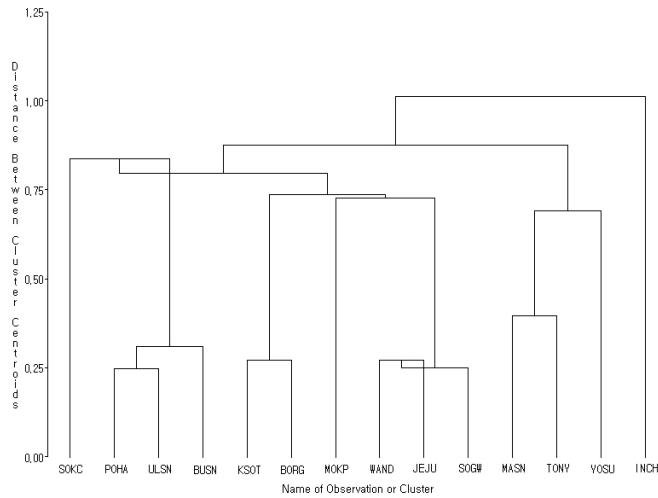


Fig. 2. Dendrogram of station clustering using centroid linkage method.

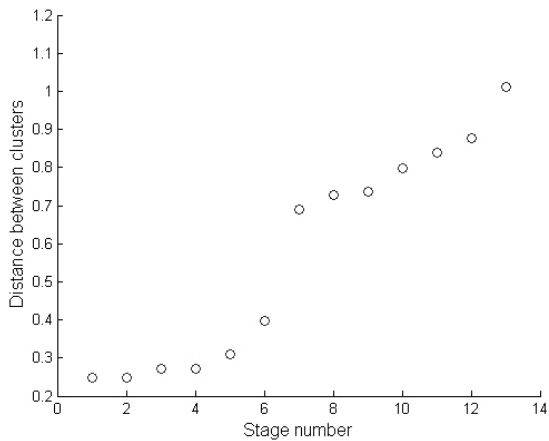


Fig. 3. Distance between cluster centroids for each stage.

또한 각 단계별 지역 군집을 표로 나타내면 Table 1과 같다. 인천 (INCH), 보령 (BORG), 군산외항 (KSOT), 목포 (MOKP), 완도 (WAND), 제주 (JEJU), 서귀포 (SOGW), 여수 (YOSU), 통영 (TONG), 마산 (MASN), 부산 (BUSN), 울산 (ULSN), 포항 (POHA), 속초 (SOKC) 등의 지역에 대해 각 단계별로 그룹화 되는 지점들을 표시하였으며 전체 지점 중에서 각 단계에서 군집화 되어 있는 지점들은 괄호로 나타내었다. 또한 그 단계에서 군집화 된 지점들은 음영으로 표시하였다. 첫 번째 단계에서는 개체 간 중심 거리가 가장 짧은 개체들 (포항 · 울산)이 그룹을 형성하게 되며 다음 단계로 갈수록 중심 거리가 길어진다.

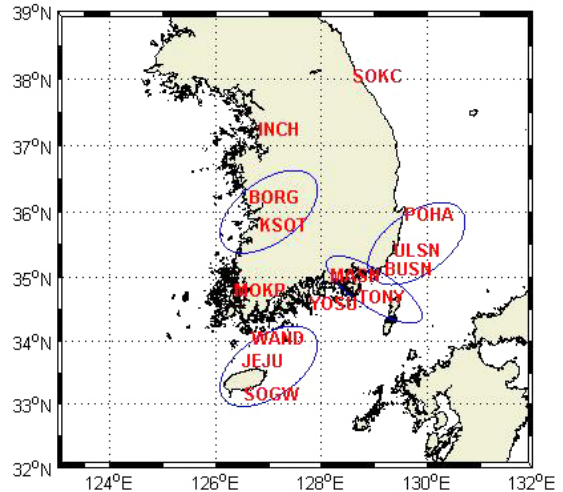


Fig. 4. Stations used for cluster analysis and Clusters (Boryeong · Kunsan,Wando · Jeju · Seogwipo,Masan · Tongyoung, Busan · Ulsan · Pohang) from results of the cluster analysis.

3.2 군집별 신경망 모델

위의 군집 분석 결과를 바탕으로 각각의 지역 군집 별 모델링을 실시하였으며 특히 부산, 울산, 포항으로 이루어진 군집 (군집1)과 완도, 제주, 서귀포 등의 지역 으로 이루어진 군집 (군집2)별로 해일 예측을 위한 신경망 모델링을 하여 그 결과를 군집1과 군집2에 포함된 부산, 제주의 지점별 모델, 예측 천문조와 비교하였다.

Table 1. Station clustering for each stage.

Stage	Stations clustered in each stage	Clusters in each stage
1	POHA, ULSN	SOKC, (POHA, ULSN), BUSN, KSOT, BORG, MOKP, WAND, JEJU, SOGW, MASN, TONG, YOSU, INCH
2	KSOT, BORG	SOKC, (POHA, ULSN), BUSN, (KSOT, BORG), MOKP, WAND, JEJU, SOGW, MASN, TONG, YOSU, INCH
3	WAND, JEJU	SOKC, (POHA, ULSN), BUSN, (KSOT, BORG), MOKP, (WAND, JEJU), SOGW, MASN, TONG, YOSU, INCH
4	WAND, JEJU, SOGW	SOKC, (POHA, ULSN), BUSN, (KSOT, BORG), MOKP, ((WAND, JEJU, SOGW)), MASN, TONG, YOSU, INCH
5	POHA, ULSN, BUSN	SOKC, ((POHA, ULSN), BUSN), (KSOT, BORG), MOKP, ((WAND, JEJU, SOGW)), MASN, TONG, YOSU, INCH
6	MASN, TONG	SOKC, ((POHA, ULSN), BUSN), (KSOT, BORG), MOKP, ((WAND, JEJU, SOGW)), (MASN, TONG), YOSU, INCH
7	MASN, TONG, YEOSU	SOKC, ((POHA, ULSN), BUSN), (KSOT, BORG), MOKP, ((WAND, JEJU, SOGW)), ((MASN, TONG), YOSU), INCH
∴	∴	∴

먼저 군집1과 군집2의 각각 세 지역에 대한 태풍 프라피룬 (Prapiroon), 사오마이 (Saomai), 루사 (Rusa), 매미(Maemi), 메기 (Megi) 통과 시점 전후 7일 동안의 매 시간 자료를 사용하였으며 입력변수로는 MM5 (Mesoscale Model version 5)로부터의 예측 기압과 바람 응력의 동-서, 남-북 방향 성분을 이용하였다.

이전 실험으로부터 선택된 MLP 구조를 사용하였으며 3개의 은닉노드를 포함하는 1개의 은닉층을 가지는 모델 구조를 선택하였다. 모델 검증은 5개의 태풍에 대해서 교차 검증 (cross-validation) 형태로, 검증하려는 태풍을 제외한 나머지 태풍에 대한 자료를 분석용 (training or learning) 자료로 사용하였으며 검증하려는 태풍을 평가용 (test or validation) 자료로 하여 그 비율을 80:20으로 사용하였다. 평가용 자료에 대한 RMSE (Root Mean Square Error)와 상관 계수 (correlation coefficient) 값을 비교하였으며 군집별 신경망 모델 (CL-NN) 결과를 조화 분석으로 예측한 천문조 (Tide), 각 지점별 자료만을 이용한 지점 신경망 모델 (NN)과 비교하였다.

그 결과는 Table 2와 같으며 부산과 제주 지역의 각 태풍별 RMSE 와 상관 계수를 보여주고 있다. 부산 지역의 경우 최고 해일고가 20 cm 정도로 비교적 낮게 나타났던 2000년 태풍 프라피룬의 경우 두 신경망 모델 모두 조화분석에 의한 천문조보다도 오히려 높은 RMSE와 낮은 상관계수를 보였다. 해일고가 60 cm 이상으로 높게 나타났던 2000년 사오마이와 2002년 루

사의 경우에는 신경망 모델의 결과가 다른 태풍 사례보다도 낮은 RMSE를 보이고 있었다. 특히 부산, 울산, 포항 등의 군집 자료를 이용하여 모델링한 군집 신경망 (CL-NN)의 결과가 사오마이의 경우 RMSE 8.82 cm, 상관계수 0.99, 루사의 경우 RMSE 8.25 cm, 상관계수 0.97로 부산 지점만의 자료를 이용한 지점 신경망 모델 (NN) (사오마이의 경우 RMSE 9.32 cm, 상관계수 0.98, 루사의 경우 RMSE 8.80 cm, 상관계수 0.96)보다도 좋게 나타나고 있음을 확인할 수 있다.

2003년 태풍 매미와 2004년 메기의 경우도 해일고가 높게 나타났으며 이 경우는 군집별 모델보다는 지점별 모델이 좀더 낮은 RMSE를 보이고 있다. 제주의 경우를 살펴보면 태풍 프라피룬, 매미, 메기의 경우에 있어서는 완도, 제주, 서귀포 등의 군집에 대한 신경망 모델의 결과가 가장 좋게 나타나고 있으며 사오마이와 루사의 경우는 제주 지점만의 자료를 이용한 신경망 모델이 가장 낮은 RMSE를 보이고 있다.

또한 Table 3은 각각의 태풍 기간 중 가장 높은 관측 해수면 높이와 그때의 천문조 높이, 신경망으로 예측된 해일과 천문조의 합인 예측 해수면 높이를 나타내고 있다. 부산 지역 사오마이와 루사, 매미의 경우 2, 11, 14 cm 정도의 오차로 군집 신경망 모델의 결과가 좋았으며 부산 지역의 경우 대부분 군집 모델이 실제 관측값이나 지점 모델의 결과보다 과대추정하는 경향을 보이고 있다. 제주 지역의 경우 부산 지역과는 반대로 군집 모델이 과소추정하는 경향을 보이고 있으며

Table 2. Comparison of harmonic analysis and neural network(NN, CL-NN) model performance for Busan and Jeju.

			Tide	NN	CL-NN
Busan	Prapiroon	RMSE	7.03	11.15	16.46
		CORR	0.9917	0.9788	0.9754
	Saomai	RMSE	24.45	9.32	8.82
		CORR	0.9622	0.9839	0.9859
	Rusa	RMSE	21.96	8.80	8.25
		CORR	0.8254	0.9615	0.9690
	Maemi	RMSE	14.00	10.55	12.08
		CORR	0.9517	0.9748	0.9705
	Megi	RMSE	18.21	12.39	15.13
		CORR	0.9532	0.9670	0.9587
Jeju	Prapiroon	RMSE	24.89	14.45	13.70
		CORR	0.9889	0.9854	0.9863
	Saomai	RMSE	34.43	14.44	17.00
		CORR	0.9755	0.9838	0.9812
	Rusa	RMSE	34.47	16.10	19.32
		CORR	0.9174	0.9662	0.9579
	Maemi	RMSE	19.16	18.50	18.08
		CORR	0.9788	0.9832	0.9785
	Megi	RMSE	28.77	35.53	17.32
		CORR	0.9868	0.8902	0.9659

Table 3. Maximum observed sea level, predicted sea level and tide at Busan and Jeju during Typhoons.

station	Typhoon	Max. sea level (cm)			
		OBS.	TIDE	NN	CL-NN
Busan	Prapiroon	152	146	144	177
	Saomai	167	134	163	165
	Rusa	159	99	145	170
	Maemi	211	132	185	197
	Megi	147	97	142	157
Jeju	Prapiroon	314	293	323	321
	Saomai	295	262	287	281
	Rusa	260	193	231	226
	Maemi	268	241	275	272
	Megi	298	269	303	310

이는 각 근집 모델 훈련에 사용한 여러 지점의 특성이 복합적으로 반영된 결과라 생각된다. 부산과 제주 지점 모두 전반적으로 해일고가 높게 나타난 태풍 사례의 경우에 있어서 지점 모델보다 근집 모델의 결과가 좋으며 해일고가 낮은 사례에서는 좋지 못한 결과를 보이고 있다.

각 태풍별 관측, 예측 해수면 높이 및 천문조위에 대

한 태풍 통과시점 전후 5일의 매 시간 시계열 자료를 Fig. 5에 나타냈다. Fig. 5(a)는 태풍 사오마이 경우의 예측 결과를 나타내며 저조시에는 실제 관측값 보다 약간 낮게 예측하고 있으나 고조의 경우 30 cm 이상의 해일을 몇 cm의 오차로 잘 예측하고 있으며 Fig. 5(b) 루사의 경우 고조에서 높은 해일이 지속적으로 나타났으며 최고 60 cm 정도의 해일을 보였다. 지점, 근집 신

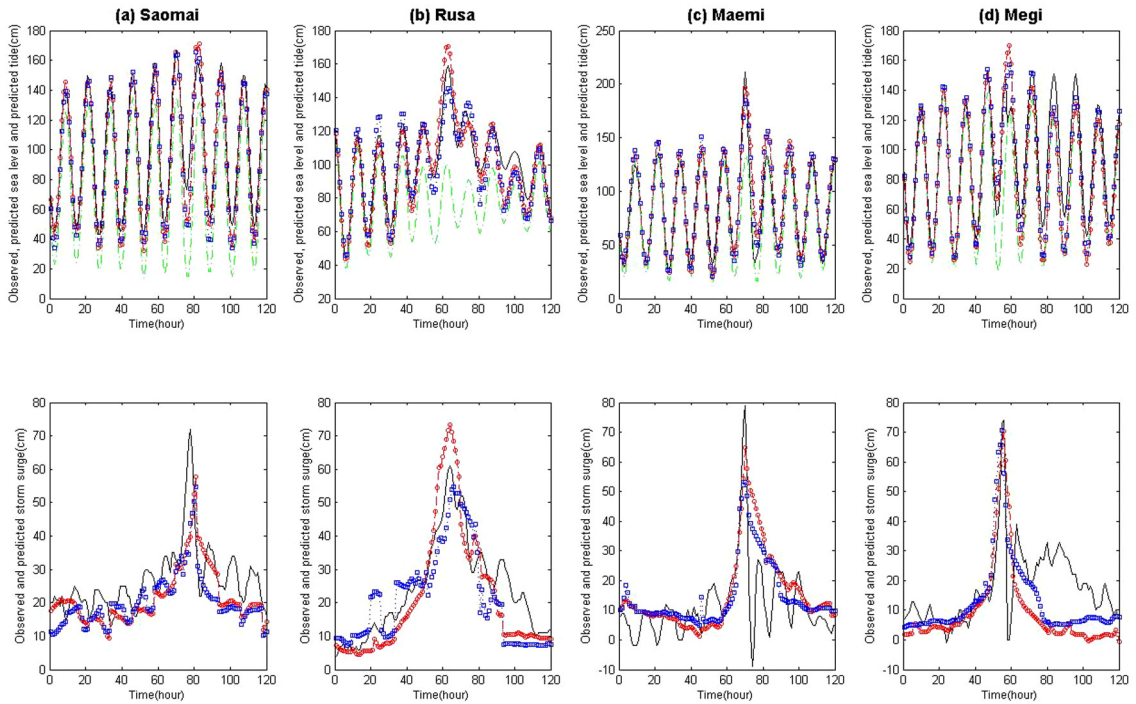


Fig. 5. Observed (solid/black), predicted (NN:dotted/squares/blue, CL-NN:dashed/circles/red) sea level and predicted tide (dash-dotted/green) (upper); Observed (solid/black) and predicted (NN:dotted/squares/blue, CL-NN:dashed/circles/red) storm surge (lower) at Busan during Typhoon Saomai, Rusa, Maemi and Megi.

경망 모두 10 cm 정도의 오차만을 보였다. Fig. 5(c)는 태풍 매미의 경우이며 고조시 최고 80 cm 정도의 해일을 보였으며 군집 모델의 경우 14 cm, 지점 모델의 경우 25 cm 정도의 오차를 보였다. Fig. 5(d)는 태풍 메기 사례이며 이 경우는 저조시에 74 cm 정도의 최고 해일을 보였으며 군집 모델, 지점 모델 모두 70 cm 정도로 해일을 예측하고 있다.

4. 결론 및 향후계획

본 논문에서는 세 지점에 대한 신경망 국지 해일 예측 모델의 확장을 위한 군집 분석의 결과와 군집별 모델링 결과를 나타냈다. 먼저 전 지점으로 모델을 확장하기 위한 지역 군집화의 객관적인 기준을 설정하기 위해 지점 정보와 태풍별 해일고 변수 등을 고려하여 군집 분석을 실시하였다. 군집 사이의 유사성을 나타내는 척도를 토대로 가장 가까운 개체끼리 군집을 형성해 나가며 각 군집 사이의 거리의 차이가 급격하게 증가하는 시점에서 군집화를 끊어주었다. 그 결과 지

점들을 몇 개의 군집으로 그룹화 할 수 있었다.

군집별 신경망 모델의 정확도를 측정하기 위해서 부산, 울산, 포항 등으로 이루어진 군집1과 완도, 제주, 서귀포 등으로 이루어진 군집2에 대해서 군집 모델링을 실시하여 이를 부산과 제주의 지점 모델과 비교하였다. 과거 태풍 사례에 대해서 지점별 모델과 비슷한 RMSE 또는 상관 계수를 보였으며 해일이 높게 나타난 사오마이, 루사 사례의 경우 오히려 더 낮은 RMSE, 상관계수 등을 보이는 경우도 존재했다. 이는 여러 지점의 태풍 자료를 사용하여 더 많은 정보로 모델을 훈련시킨 결과라 하겠다. 즉, 신경망 모델링 결과 지점·군집 신경망 모델이 해일을 잘 예측하고 있으며 특히 낮은 해일보다 높은 해일 예측에서 좋은 결과를 보였다. 따라서 국지 해일 예보 시스템의 지역 확장을 위해서 지점 군집화 및 군집별 모델링은 의미가 있을 것으로 사료된다. 그러나 태풍 프라피룬 사례에 대해서 부산 지점의 경우 두 신경망 모델의 결과가 조화분석에 의한 천문조보다도 좋지 못했으며 이는 프라피룬의 경우 입력변수로 사용된 예측 바람응력이 실제 관측

바람응력보다 매우 강하게 나타났으며 그에 비해 낮게 나타난 해일고, 태풍기간의 높은 해일고 자료만을 이용한 모델 훈련 등에 기인한 결과라 생각된다.

본 연구의 결과로부터 군집별 신경망 모델을 국지 해일 예보시스템에 활용하여 전 지점에서의 확대가 가능할 것이며, 앞으로 신경망 국지 해일 모델의 전 해안으로의 안정적인 확장을 위해서 더 많은 지점에 대한 태풍 정보가 요구된다. 또한, 모델 정확도를 높이기 위해서 모델 입력 자료로 사용되는 예측장의 보정, 실시간으로 태풍 정보를 모델에 포함시킬 수 있는 동적 모델링 시스템 구축 등이 필요하다.

감사의 글

본 연구는 기상청 기상연구소 기본 사업 “해양기상 관측 및 조사” 과제의 일환으로 수행된 것입니다.

참고문헌

- 강현철, 한상태, 최종후, 김차용, 김은석, 김미경, 1999: *데이타마이닝 방법론 및 활용*. 자유아카데미, 233-258.
- 고정웅, 백희정, 권원태, 2005: 한반도 우기의 강수 특성과 지역 구분. *한국기상학회지*, **41**, 101-114.
- 김대수, 2001: *신경망 이론과 응용 I*. 진한엠앤비, 91-117.
- 문영수, 1990: 클러스터분석에 의한 한국의 강수지역 구분. *한국기상학회지*, **26**, 203-215.
- 문일주, 오임상, 윤용훈, 2000: 한반도 서해안 실시간 해수면 예측 수치실험 연구. *한국기상학회지*, **36**, 203-218.
- 성용현, 2002: *응용 다변량분석 이론·방법론·SAS 활용*. 탐진, 321-338.
- 손건태, 이정형, 이순환, 류찬수, 2005: 호남지역 집중호우 예측을 위한 통계 모형 개발. *한국기상학회지*, **41**, 897-907.
- 이다운, 이호만, 서장원, 유승협, 윤용훈, 2005: 신경망을 이용한 폭풍 해일 예측 연구. *한국기상학회지*, **41**, 661-670.
- 해양연구소, 1985-1988: 해안, 재해 예보 시스템연구(I-IV) BSPG. 해상대응 예보 및 해일예보, 197-227.
- Choi, B. H., H. M. Eum, and S. B. Woo, 2003: A synchronously coupled tide-wave-surge model of the Yellow Sea. *Coastal Eng.*, **47**, 381-398.
- Cox, D. T., P. E. Tissot, and P. R. Michaud, 2002: Water level observations and short-term predictions including meteorological events for entrance of Galveston Bay, Texas. *J. Waterw., Port, Coastal, Ocean Eng.*, **128**, 21-29.
- Faraway, J., and C. Chatfield, 1995: Time series forecasting with neural networks: A case study. Research Report 95-06 of the Statistics Group, University of Bath.
- Potts, W. J. E., 2000: *Neural network modeling course notes*. SAS Institutes Inc, 240-259.
- Sztobryn, M., 2003: Forecasting of storm surge by means of artificial neural network. *J. Sea Res.*, **49**, 317-322.
- Tissot, P. E., P. R. Michaud, and D. T. Cox, 2003: Optimization and performance of a neural network model forecasting water levels for the Corpus Christi, Texas, Estuary. *3rd Conf. on Artificial Intelligence Applications to the Environmental Science*.
- Tsai, C. P., and T. L. Lee, 1999: Back-propagation neural network in tidal-level forecasting. *J. Waterw., Port, Coastal, Ocean Eng.*, **125**, 195-202.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Science*. Academic Press, 419-427.