

## 전국 해안지역의 오손분포Map 개발에 관한 연구

김준일\*, 김동명\*  
한국전력공사 전력연구원

### The study on contamination criterion map of seashore area in Korea

Kim Joon Eel\*, Kim Dong-myung\*  
KEPCO KEPRI\*

**Abstract** - 삼면이 바다로 둘러 싸여 있는 우리나라의 경우에는 태풍이 바다에서 육지로 이동하게 됨으로써, 염분을 포함한 수분이 배전설비의 표면에 부착되어, 설비의 열화로 인한 정전이 발생할 가능성이 매우 크므로, 염분에 의해 오손된 배전설비의 오손등급을 평가하여 설비운영에 활용함으로써, 정전을 예방하고 정전 관련 민원을 최소화할 필요가 있다. 이를 위해 지형을 고려한 해안지역의 오손분포를 확인할 수 있는 오손 분포 알고리즘을 개발하고 이에 따른 분포 도면을 제작하여, 한국전력공사의 각 지역별 배전설계담당자들이 해당 설계지역의 오손도를 정확히 인지하여 배전설비 설계에 적용하는데 그 목적이 있다.

#### 1. 서 론

본 논문은 국내 해안지역의 오손정도를 판단하는 오손의 분포 맵을 작성하는 알고리즘 및 이에 따른 분포 도면의 제작에 관한 것이다. 과거에 측정된 오손의 정보를 이용하여 지형(DEM을 이용한 고도 및 지형경사의 변화와 토지 피복 현황)과 강우량 및 오손의 축적관계를 데이터마이닝 기법을 이용하여 모델링 한 후, 모델링 된 수식을 해안선 영역의 지형인자와 강우량 인자에 적용하여 오손의 분포도면을 제작하였다.

이를 위하여 기존에 측정된 해안선 부근의 오손 측정치(ESDD)를 근거로 측정 날짜의 강우량 및 지형인자(지형고도, 경사, 토지 이용현황)를 변수로 한 오손의 관계수식을 데이터마이닝 기법 중의 하나인 의사결정나무 알고리즘을 이용하여 설정하였으며 해안선으로부터 20km 이내의 모든 영역에 대하여 1km<sup>2</sup>의 화소로 분류하여 각 화소의 중앙점으로부터 해안까지의 거리, 지형의 고도 및 경사, 토지이용현황을 데이터베이스화 한 이후에 일정 강우량을 설정하여 의사결정나무로 지정된 관계식에 이 인자를 대입하여 각 화소에서의 오손 값을 역산하여 그 분포를 도면화 하였다. 그러므로 오손 분포도면을 제작하기 위해서는 기존의 측정 위치에 대한 데이터를 기반으로 오손과 지형의 관계식을 설정하는 Training 과정과, 이 과정에서 설정된 수식의 신뢰도를 평가하는 Testing 과정 및 설정된 수식을 전국 해안가 전역에 적용하여 예측된 오손값을 계산하기 위한 작업을 수행하였다.

#### 2. 본 론

##### 2.1 오손과 지형인자의 상관관계 적용

오손과 지형의 관계수식이 실제 오손 측정지역의 자료로부터 설정되어야 하므로 <그림 1>에 표기된 조사지역의 지형인자를 추출하여 지형인자와 오손의 관계식을 설정하고 관계식의 신뢰도를 평가하는 과정을 진행하였다. 관계식이 설정되면 해안으로부터 20km<sup>2</sup>이내의 모든 영역에 대하여 설정된 관계식을 대입하여 추정된 오손도면을 제작하여야 하므로 데이터베이스 작업에 대한 기법은 관계식 설정을 위한 자료의 추출과정과 오손 분포도 작성을 위한 기초자료의 추출에 동일하게 적용되었다. 지형의 분석은 해안으로부터 측정까지의 거리, 측정으로부터 해안방향으로 500m 간격으로 20개의 구간을 순차적으로 옮겨가면서 각 지점의 고도와 경사값을 기록하였다. 그러므로 각 위치로부터 해안 방향으로 10km의 구간에 대한 지형기록이 데이터 베이스화 되었다. 지형자료의 구축을 위하여 전국이 234개의 도면으로 구분된 1:50,000 축척의 지형도에서 1초 간격으로 고도를 읽은 DEM(Digital Elevation Model) 자료가 활용되었다. 지형자료는 측정 위치로부터 해안방향으로 10km의 구간에 걸쳐 읽혀졌는데, 해안의 굴곡이 심한 지점에서는 측정으로부터 해안방향을 정의할 방향의 설정이 문제될 수 있으므로, 이를 위하여 <그림 2>와 같이 해안의 방향을 단순화 한 도면을 제작하고, 본 도면에 설정

된 방향을 기준으로 지형의 변화를 데이터베이스화 하였다.

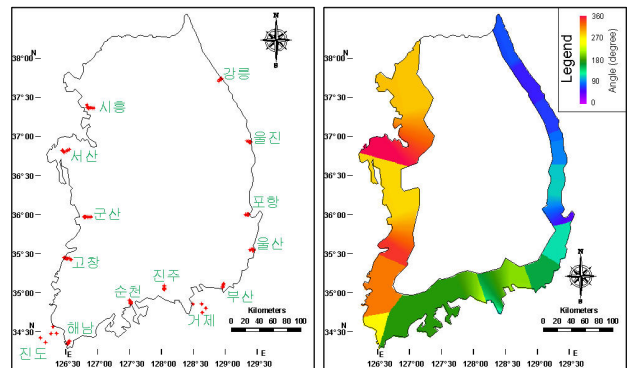


그림 1) 전국 해안지역의 오손측정개소 <그림 2>해안의 방향을 단순화한 도면

해안으로부터의 거리분석은 위에서 구해진 벡터로부터 해안선까지의 거리를 바탕으로 계산하였으며 거리의 계산은 해안선을 향한 벡터와 해안선의 교점과 1 × 1km 크기의 지형격자의 중심까지의 거리를 계산한 결과이다. 계산의 알고리즘은 Map Object의 콤포넌트를 이용한 S/W를 제작하여 수행하였다.

오손 측정위치에서의 지형정보는 해안선을 향한 벡터로 500m씩 이동한 지점에서 10km의 거리까지 지형의 고도변화와 경사값을 읽어서 고도자료 20개의 지형인자를 추출하였다. 지형의 경사는 해당위치로부터 벡터방향으로 각기 좌우로 100m 이동하여 읽혀진 2개의 고도값과 해당위치의 고도값 3개를 최소자승법으로 보간하여 경사값을 수평으로부터 시계방향으로 읽어주는 기법을 활용하였다. 그러므로 반시계방향의 각도는 -ve 값으로 기록된다. 지형의 형태를 500m 간격으로 읽어서 이를 해당점에 대한 지형인자로 모델링한 것은 해당측점으로부터 해안선까지의 지형 변화를 정량화 하고자 하는 시도로서, 오손은 해안으로부터 유입되며 유입되는 경로가 유사한 지형을 가질 경우 40개의 인자가 유사한 배열을 가지는 것을 가정한 결과이다.

##### 2.2 오손과 강우량과의 상관관계 적용

강우량은 기상청에서 제공하는 전국 76개소의 강우자료를 사용하였다. 전국에 분포하는 76개소의 각 관측소에서 2005년 1월부터 2006년 12월 까지 관측된 일별 강우자료를 데이터 베이스화 하였으며, 오손이 측정된 위치의 강우자료는 측정일로부터 1달 전의 누적강우량을 아래와 같은 과정으로 계산하였다.

1. 오손이 측정된 위치로부터 인접한 3개 이상의 관측소를 탐색
2. 각 관측소에서 측정일 1달 이전부터의 누적 강우량을 계산
3. 3개 이상의 관측소에서 측정까지의 거리를 구한 후, 각 관측소의 누적 강우량을 Inverse Distance Weight 알고리즘을 이용해 보간하여 해당 위치의 강우량을 계산

이와 같은 계산을 위하여 강우자료와 공간 DB가 구축되었으며 이를 탐색하고 계산하는 S/W가 제작되었다. 지형고도는 이전 위치와 현 위치의 고도 차이를 읽어주며, 경사는 수평면으로부터 시계방향에 양수(+Ve)이며 반시계방향은 음수(-Ve)이다.

##### 2.3 오손분포 알고리즘의 적용

의사결정나무(Decision tree)는 데이터마이닝 기법의 하나로, 데이터마이닝은 대형 데이터베이스 내부에 존재하는 자료의 구

조(patterns, statistical models, relationships)를 추출하는데 매우 유용한 기법이다. 즉 데이터마이닝은 다량의 자료 내부에 존재하는 규칙을 전산기법으로 정의하고 정의된 규칙을 다른 데이터베이스에 적용하여 그 데이터베이스로부터 유도된 결론을 추정한다. 오손 자료와 같이 지형의 숫자 자료가 대부분이고 토지이용의 문자자료를 포함하는 자료구조에 가장 적합한 방법인 CART 기법을 이용하여 의사결정나무의 제작에 Weka S/W에서 제공하는 M5P라는 알고리즘을 활용하였다.

### 2.3.1 M5P 알고리즘을 적용한 의사결정나무 제작

M5P 알고리즘은 의사결정나무의 잎(Leaves)이 다변량의 선형 다항식 (multi-variate linear) 으로 제작된다. 의사결정나무를 제작할 때 가장 중요한 과정은 마디에서 가치를 결정하는 것이다. 즉 어떤 속성을 선택하여 어떠한 분류기준으로 그 속성을 분류하여 가치를 만드느냐 하는 것이다. 이들 중 어떤 속성을 먼저 마디로 지정하여 분류를 시작할 것인가가 중요하다. 본 논문에서 사용하는 M5P 알고리즘은 마디에 지정될 속성의 선택을 위해서 SDR(Standard Deviation Reduction) 이론을 적용하였다.

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} \times sd(T_i)$$

위에서  $sd(T)$ 는 전체 자료의 ESDD값의 표준편차이며  $sd(T_i)$ 는 선택된 속성의 영역을  $i$  개의 그룹으로 분류한 자료군을 의미한다. 예를 들어 해안선으로부터의 거리 속성이 30m에서 800m까지의 자료군으로 형성되어 있다면 30~200m, 201~400m, 401~800m의 3개 자료군으로 분류하여 각 자료군에 속한 ESDD 자료의 평균이  $|T_i|/|T|$ 가 되는 것이다. 이와 같이 각 속성의 SDR을 구한 후 SDR의 값이 최소가 되는 속성을 마디로 지정하게 되는데, 처음 마디가 결정되면 다음 단계에서 분류된 자료군을 상대로 다시 위의 계산을 수행하여 다음 마디를 결정해 내려가면서 전체적인 의사결정나무를 제작하게 된다. 각 단계에서 전체 자료의 표준편차와 분류된 자료의 표준편차를 비교하여 분류된 자료의 표준편차가 전체자료의 표준편차의 5% 이내가 되거나 자료의 개수가 4개 이하가 되면 그 분류를 잎(leaf)으로 지정하면서 분류를 종료하게 된다. 가지의 형태가 결정되면 최종 결정나무의 잎에 할당된 자료를 갖고 아래와 같은 다항식을 작성하게 된다.

$$w_0 + w_1 a_1 + w_2 a_2 + w_3 a_3 + \dots + w_k a_{3k}$$

위의 다항식에서 계산된 값은 ESDD값이며,  $a_k$  항은 속성변수이며  $w_k$  항은 각 속성의 가중치가 된다. 각 속성의 중요도는 SDR로 계산되지만 이를 계산하기 위해서는 속성자료를 적절한 그룹으로 분할하여야 하는데 분할하는 과정은 아래의 수식으로 분할의 경계를 지정하게 된다.

$$\Phi(s/t) = 2P_L P_R \sum_{j=1}^m [P(C_j | t_L) - P(C_j | t_R)]$$

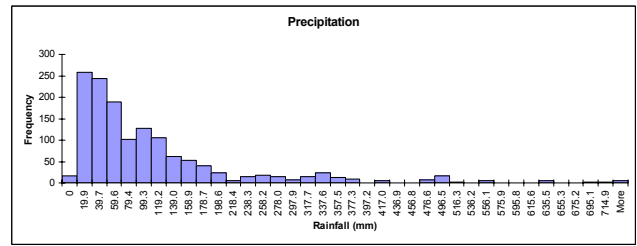
위의 수식에서  $P_L$ 과  $P_R$ 은 각기 구분된 자료가 분류기준에서 좌측과 우측에 존재할 확률이며,  $P(C_j | t_L)$ 와  $P(C_j | t_R)$ 는 각기 좌측과 우측 자료군 내에서 분할된 속성들(예, 거리속성의 분할을 30~200m, 201~400m, 401~800m 등으로 함)의 확률을 의미한다.

오손분포도를 제작하기 위해 고려된 대부분의 속성자료는 숫자 자료이나 토지이용도의 자료는 문자자료이다. 문자자료의 경우는 속성내의 분류그룹을 2개의 조합그룹으로 묶어서 조합의 좌측과 우측의 확률을 이용하여 SDR을 계산하여 가치를 분류하는 알고리즘이 이용된다. 위의 과정으로 제작된 초기 의사결정나무는 학습자료를 너무 세분한 나머지 정보자료를 지나치게 많은 갖는 가지가 있는 "overfitting" 특성을 갖는다. 이러한 overfitting 문제는 속성의 자료값이 넓은 영역을 갖고있을 때 작성된 초기 의사결정나무에서는 대부분 발생한다(Witten I.H. and Frank E. 1999). 이러한 문제를 보완하기 위해서는 의사결정나무의 가지치기(prune)를 수행하여야 한다. 가지치기는 작성된 다항식에 의해 계산된 ESDD의 값과 실제의 ESDD 값을 비교하여 그 오차가 큰 부분을 소거하고 결정나무의 가치를 통합하는 과정으로 진행된다. 위의 이론을 적용하여 Weka의 M5P 알고리즘을 적용하여 오손의 측정치를 이용해 의사결정나무를 작성하였다.

### 2.3.2 누적강우의 공간적 분포예측을 통한 오손분석

위와 같이 제작된 의사결정나무를 이용하여 해안선 인접부의 오손 분포도를 작성하기 위해서는 강우의 공간적 분포예측을 고려하여야 한다. 즉, 강우량은 시간과 위치에 따라서 변하는 자료이므로 전국의 해안선 인접지역의 강우량 강도를 공간적으로 예측해야 하는 문제가 있다. 강우량의 공간적 분포는 시간적 변화를 보이는 강우자료를 어떻게 처리하느냐에 따라 달라지는 문제

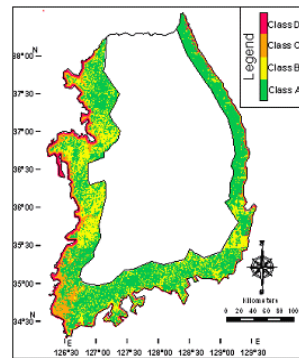
로서 단순 공간분포도를 작성하는 것은 불가능하다. 그러므로 강우량이 많을 경우 오손이 빗물에 씻겨서 상대적으로 적게 누적될 것 이라는 가정 하에 최소 강우량을 예측에 적용함으로써 최대 오손 누적의 상황을 예측하여 이에 대응하고자 하는 소극적 설계 기법을 적용하였다. 적절한 최소 강우량을 구하기 위하여 오손의 측정일에 적용된 모든 누적강우량의 분포를 분석하였다.



〈그림 3〉 누적강우분포 히스토그램

〈그림 3〉은 의사결정나무 제작에 사용된 모든 누적강우량 1,391개의 분포를 히스토그램으로 표시한 것이다. 누적강우량은 평균 99.5mm에 표준편차가 117.26mm의 분포를 보이고 있다. 그러므로 자료의 약 67%가 포함된 표준편차의 경계는 40.87mm(99.5-(117.26/2))에서 시작된다. 이를 근거로 40mm의 누적강우량을 전 지역에 적용하였다.

해안선으로부터 20km 이내의 지역을 1km 화소 단위로 구분한 도면은 37,328개의 화소로 구분되므로 각 화소의 중심점으로부터 전기한 지형의 인자추출 방법을 이용하여 500m 간격의 지형경사와 고도차, 토지피복분류, 지형고도를 추출하여 데이터베이스를 작성하고, 40mm의 누적강우량을 제작된 의사결정나무에 적용하여 각 위치의 오손 누적량을 계산하여 오손의 분포도를 〈그림 5〉와 같이 제작하였다.



〈그림 5〉 전국 해안지역의 오손분포도

## 3. 결 론

국내 해안지역의 오손정도를 판단하는 오손의 분포도를 제작하기 위해 M5P 알고리즘을 이용한 의사결정나무를 작성하였으며 누적강우의 공간적 분포예측을 통한 오손 분석을 시행하였다. 한국전력공사의 각 배전사업소 담당자들이 해당지역의 배전설비를 설계할 경우 해당지역의 오손정도를 인지하여 이를 배전설비에 적용하면 배전선로 운전 여건에 적합한 배전설비를 구축할 수 있는 효과가 있으리라 생각한다. 본 오손 분포도의 제작을 위하여 2년간의 오손 측정데이터를 이용하여 이를 재현주기를 설정해 보정하였으나, 보다 신뢰도 있는 데이터를 구축하기 위해서는 지역별 담당자들이 지역별 오손측정을 주기적으로 시행하여 이를 오손 분포도에 데이터 베이스로 구축한다면 향후 제작되는 오손 분포도는 현장 적용성 측면에서 신뢰도가 좀 더 높아지리라 예상된다.

### [참 고 문 헌]

- [1] 심규일 외, "다중회귀를 이용한 기후와 오손도 간의 상관관계 분석" 한국조명·전기설비학회 학술대회 논문집, pp 199-202, 2001.
- [2] 김찬영 외, "해안거리에 따른 ESDD의 통계적 분석", 대한전기학회 2007년도 하계학술대회 논문집, pp 1474-1475, 2007.
- [3] 심승보 외, "염진해 오손정도 및 기준정립에 관한 연구(최종보고서)", 2002.
- [4] 최남호 외, "통계적 처리방법을 이용한 동해안 염해 오손물의 분포 특성", 대한전기학회 논문지, 제50권 3호, pp130-136, 2000.