



대설피해 취약 등급별 피해예측 회귀 모형 개발

Multiple Regression Models of Snow Damage Prediction According to the Snow Damage Vulnerability Groups

오영록* · 정건희**

Oh, YeoungRok*, and Chung, Gunhui**

Abstract

The frequency of natural disaster has been increased due to the climate change, and the heavy snow has been also increased during the winter. Therefore, many researches have been conducted to estimate snow damage vulnerability. In this study, multiple regression models for snow damage prediction were developed using snow damage vulnerability which was calculated in the previous research. The snow damage vulnerability were categorized into 5 groups. The developed models were even applied in the area where never had a snow damage in the history. Three multiple regression models were developed according to the snow damage vulnerabilities. The input data for the model were snow vulnerability index, snow depth, the exceedance percentage of snow design criteria, daily relative humidity, daily minimum temperature, and daily maximum temperature. As a result, normalized root mean square error (NRMSE) was low enough to apply the models to estimate the snow damage. The developed models could be applied to estimate the snow damage even in the areas where the snow damage have not been occurred.

Key words : Snow Damage, Snow Damage Estimation, Multiple Regression Analysis, Snow Damage Vulnerability Groups

요 지

최근 세계적인 기상이변으로 인한 자연재해가 빈번하게 발생하고 있으며, 그 피해를 예측하기란 더 어려워지고 있는 실정이다. 우리나라의 대표적 겨울철 자연재해인 대설피해의 빈도 또한 기습적인 대설 또는 새로운 기록을 갱신하는 강설량으로 인해 피해가 증가하고 있는 추세이다. 그러므로 대설 피해 취약성 예측을 위한 연구들이 다수 수행되었다. 본 연구에서는 선행연구에서 수행된 대설피해 취약성 분석 결과를 기반으로 각 취약성 등급별 대설 피해 예측을 위한 회귀모형을 개발하였다. 선행연구에서 구분된 5개의 취약성 등급을 바탕으로 과거 대설 피해가 발생한 적이 없는 지역에도 기습적인 폭설이 내릴 경우 발생 가능한 피해액을 대략적으로 예측하는 것을 목적으로 하였다. 이를 위해 5개의 취약성 등급을 3개로 묶어 각 경우별 다중회귀모형을 구축하고 대설피해액을 예측하였다. 다중회귀모형의 독립변수로는 지역별 대설피해 취약성 지수, 최심신적설심, 설계기준 초과율, 상대습도, 최저기온, 최고기온을 이용하였다. 검토 결과 정규화된 표준제곱근 오차인 NRMSE가 낮게 계산되어 모형의 적용성을 입증하였다. 본 연구를 바탕으로 피해사례가 거의 없거나 전혀 없는 지역까지 지역별 특성을 고려한 대략적인 대설피해 예측이 가능할 것으로 예상되며, 향후 고도화 된 대설피해 취약성 결과와, 분석방법을 이용한다면 전국 대설피해 예측 모형의 성능이 향상될 것으로 판단된다.

핵심용어 : 대설피해, 대설피해 추정, 다중회귀분석, 대설피해 취약성 등급

*정회원, 호서대학교 토목공학과 석사과정(E-mail: dudtla224a@nate.com)

Member, Master's Course, Department of Civil Engineering, Hoseo University

**교신저자, 정회원, 호서대학교 토목공학과 교수(Tel: +82-41-540-5794, Fax: +82-41-540-5798, E-mail: gunhuic@gmail.com)

Corresponding Author, Member, Professor, Department of civil Engineering, Hoseo University

1. 서론

1.1 연구배경 및 목적

최근 세계적인 기상이변의 결과로 국지적인 폭설과 한파가 발생하고 있으며, 이로 인한 피해를 예측하기는 더 어려워지고 있는 추세이다. 최근 폭설피해로는 2016년 1월 폭설과 한파로 인해 제주공항이 폐쇄된 사례가 있었으며, 2017년 1월 일본에 폭설이 발생하여 6명이 사망하고, 항공편 결항 등의 피해가 속출하고 있다. 재해연보에 따르면 우리나라의 자연재해는 크게 호우, 태풍, 대설, 강풍 등 5가지로 구분할 수 있으며, 그중 대설로 인한 피해는 최근 10년간 약 2천억 원으로 집계되었다(MPSS, 2016). 대표적 사례 중 하나인 2014년 2월 경주 마우나 오션 리조트 체육관의 붕괴로 인해 약 10여명이 사망하고 103여명이 부상당하는 사례로 인해 대설피해 예방 및 대책에 관심이 증가하고 있다. 따라서 Jung et al.(2015)은 상대적으로 대설에 취약한 온실의 피해를 저감시키기 위해 적설하중 산정을 위한 노출계수를 제안하였으며, Lee et al.(2015)은 최심신적설의 단위중량에 대해 연구하여 최대가능적설하중을 추정하는 연구를 진행하였다. 또한 Kwon and Chung(2017)은 회귀분석을 이용하여 강원도지역의 대설피해액 추정 모형을 개발하였고, Oh and Chung(2017)은 대설피해가 자주 발생하는 강원도, 전라도, 충청도 지역을 대상으로 대설피해 예측 모형을 개발하고 모형의 적용성을 검토하였다. 그러나 Kwon and Chung(2017)의 연구는 수정된 결정계수가 0.6값 미만을 나타내고, 0.2-0.4 사이의 값이 다수 존재하는 등 적용성이 다소 떨어지는 시·군·구가 있었으며 강원도 지역만을 대상으로 했다는 점에서 전국 적용성이 떨어진다. Oh et al.(2017) 또한 대설피해예측에 적합한 설계기준 적설심 기준을 25 cm 이상으로 제안하며 수정된 결정계수가 0.7 이상이었으나 강원도, 전라도, 충청도 지역만을 대상으로 했다는 점에서 전국에 적용하기에 한계가 있었다. 뿐만 아니라 재해연보의 대설피해에서 가장 크게 집계되는 비닐하우스의 지역별 설계기준을 고려하지 않아 지역별 특성을 제대로 고려하지 못한다는 한계도 존재하였다.

이와 같이 사전에 대설피해를 예측하고자 하는 선행 연구들이 존재하지만 그 효율이 높지 않거나, 지역별 특성을 제대로 고려하지 않고, 특정 시·도만을 대상으로 했다는 점에서 전국 적용성의 한계가 존재하였다. 또한 과거 대설피해 이력이 없거나, 거의 없던 지역은 재해통계 데이터가 부족하여 통계적 방법으로 대설피해 예측 함수 개발이 불가능한 점도 해결해야 할 문제이다. 그러므로 재해통계가 부족한 지역의 특성을 고려한 군집화를 통해 데이터 부족의 문제를 보완하고 지역 특성이 고려된 전국 대설피해 예측함수 개발이 필요한 실정이다. 따라서 본 연구에서는 전국의 대설피해를 예측하기 위해 대설피해 취약성 등급별로 군집을 진행하고 대설피해 취약성 분석 결과를 독립변수로 활용

하여 지역별 특성이 반영된 대설피해 예측함수 개발 가능성을 검토하였다.

1.2 연구범위 및 방법

재해연보의 1994년부터 2016년까지 지난 23년간 대설피해 발생 사례를 살펴보면 Fig. 1과 같이 전라도, 충청도, 강원도 지역이 피해가 빈번하며 최대 19건으로 거의 매년 피해가 발생한 지역까지 존재하였으나 남해지역이나 내륙 등의 지역은 피해가 거의 없을 뿐 아니라 서울, 대구, 부산, 인천 등의 몇몇 지역은 피해 사례가 0건으로 집계되어 전국을 대상으로 대설피해 예측 모형을 개발하기에 어려움이 있다. 그러므로 전국을 대상으로 대설피해 예측함수를 적용하기 위해서는 피해사례가 없는 지역의 지역특성을 고려한 군집화 진행을 통해 데이터 부족의 문제를 해결해야만 한다. Oh et al.(2018)는 전국 229개 행정구역별 사회·기상 및 환경적인 요인 11종류의 데이터를 구축해 취약성 분석을 실시하였으며, 대설피해로부터 상대적으로 안전한 지역인 1등급이 125개 시·군·구, 2등급이 35개 시·군·구, 3등급이 11개 시·군·구, 4등급이 23개 시·군·구, 5등급이 35개 시·군·구로 나타났다. 또한 취약성 분석 결과를 1994년부터 2016년까지의 재해연보의 대설피해 집계양상과 비교하여 타당성을 검증하였다. 따라서 본 연구에서는 대설피해 취약성 분석을 통해 전국을 취약성 5등급으로 구분한 Oh et al.(2018)의 연구 결과를 인용하여 취약성 등급별 대설피해 예측함수를 개발하고 적용성을 검토하였다.

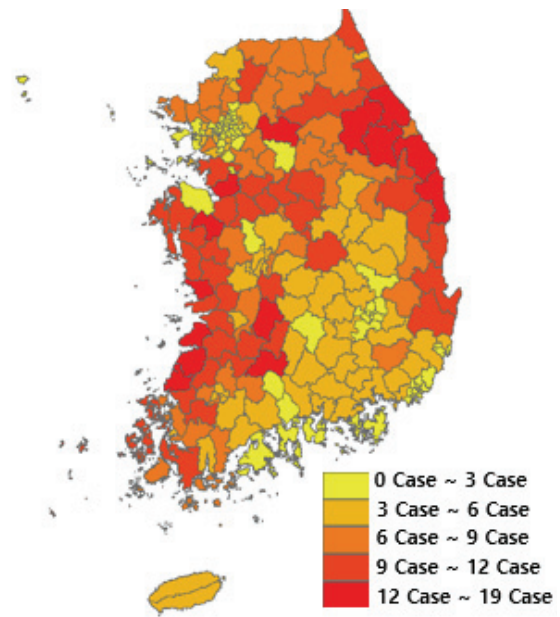


Fig. 1. Frequency of Snow Disasters from 1994~2016

1.2.1 분석방법

재해통계자료 구축을 위해 재해연보를 살펴본 결과 2015년

12월 3일 경기도 이천시에 약 6억 원의 대설피해가 집계되었지만 해당기간 관측된 적설량은 0 cm이었으며, 2011년 2월 11일 경상북도 영천시 또한 약 9천만 원의 대설피해가 집계되었지만 해당 관측소의 적설량 0 cm를 기록하는 등 기상청의 적설심 자료와 재해연보의 대설피해액 간에 보정이 필요한 사례들이 다수 존재하는 것을 확인하였다. 또한 Oh et al.(2017)는 재난통계를 활용한 대설피해 예측 및 대설피해 적설심 기준 방안에서 최심신적설심 25 cm 이하에서 발생한 피해사례는 시설물의 노후화 및 관리소홀 등 적설하중 외적의 요인이 작용했을 것으로 보고, 대설피해 예측모형을 개발할 경우 적설심 25 cm 이상의 피해사례를 이용하도록 제안하였다. 그러므로 본 연구에서는 25 cm 이상의 적설심에서 발생한 대설피해 사례들을 취약성 등급별로 군집하여 데이터를 구축하였다. 대설피해 예측함수 개발에 이용된 모형으로는 기존 대설피해 예측모형으로 자주 이용되는 다중회귀모형을 선택하였으며, 독립변수로는 최심신적설량, 설계기준 초과율, 상대습도, 최저기온, 최고기온, 지역별 대설피해 취약성 지수를 이용하고 종속변수로는 재해연보의 대설피해 집계액을 이용하였다. 취약성 등급별 표본수를 Table 1에 나타냈다.

등급별 적설심 25 cm 이상의 피해사례를 수집한 결과 3등급과 4등급의 표본수가 각각 4건, 18건으로 다소 부족하게 나타났다. 따라서 본 연구에서는 1등급의 피해사례를 Case 1, 2등급~3등급 피해사례를 Case 2, 4등급~5등급 피해사례를 Case 3으로 구분하고 Case 별 다중회귀모형을 구축해 적용성을 검토하였다.

Table 1. Number of Data by Criteria

Criteria	No. of city	No. of Data
1	125	59
2	35	20
3	11	4
4	23	18
5	35	23
All	229	124

2. 회귀분석

2.1 회귀분석 및 모형 진단 방법

다중회귀모형은 예측하고자 하는 종속변수를 설명하기 위해 두 개 이상의 설명변수가 사용되는 선형회귀모형을 말하는 것으로, 단순회귀모형의 확장형이라 할 수 있다. 다중회귀모형의 종속변수 Y 를 n 개의 설명변수 x_1, x_2, \dots, x_n 의 선형식으로 설명하려는 모형으로 Eq. (1)과 같이 나타낼 수 있다.

$$Y = B_0 + B_1x_1 + B_2x_2 + \dots + B_nx_n \quad (1)$$

여기서 B_0, B_1, \dots, B_n 는 모집단의 회귀계수이며 Y 는 대설피해 집계금액이다. 본 연구에서 선택한 독립변수들은 대설피해에 영향을 끼치는 중요한 요인이기 때문에 동시입력방식을 이용하였으며, 변수 간 상관관계 분석과 분산팽창계수(VIF)를 검토하여 다중공선성 진단에 유무의를 판단하였다. 또한 Kim et al.(2017)은 다중회귀분석을 통한 호우피해예측함수를 개발하여 모형의 성능을 검토하기 위해 평균제곱근오차(Root Mean Square Error, RMSE)를 표준화한 NRMSE(Normalized Root Mean Squared Error, NRMSE)를 이용하였다. Oh et al.(2017) 또한 다중회귀분석을 통해 모의 피해액과 집계피해액간의 NRMSE를 이용하여 모형의 성능을 나타낸 바 있으므로, 같은 분석방법을 이용한 본 연구에서도 NRMSE를 이용하여 모형의 성능을 검토하였다. NRMSE는 모의피해액과 집계피해액 사이의 평균제곱근오차(RMSE)를 정규화하여 비교할 수 있는 정규화한평균제곱근오차(NRMSE)를 나타냄으로서, 분석 Case 별 성능 비교를 가능하게 하기 위함이다. 또한 NRMSE는 0에 가까울수록 모의 피해액과 집계 피해액의 차이가 없음을 나타내며 높은 성능을 뜻한다. RMSE와, NRMSE를 도출하기 위한 식은 Eqs. (2)와 (3)에 나타냈다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (2)$$

$$NRMSE = \frac{RMSE}{(Max(y_i) - \min(y_i))} \quad (3)$$

여기서 y_i 는 i 번째 실제피해액을 의미하며 \hat{y}_i 는 i 번째 모의된 피해액을 의미한다.

2.2 다중공선성 진단 결과

다중회귀모형을 동시입력방식으로 구축할 경우, 독립변수들 간의 상관관계로 인해 발생할 수 있는 다중공선성의 진단을 반드시 거쳐야 한다. 다중공선성은 개발된 모형의 독립변수들 간의 상관관계 문제로 인해 모형의 정확성을 떨어뜨리는 것으로 분산팽창계수(VIF)를 통해 검토할 수 있다. 분산팽창계수란 설명하는 변수 간에 상관관계가 있을 경우 추정된 회귀 계수의 분산이 증가하는 정도를 측정하는 것으로서, 일반적으로 분산팽창계수가 1이면 다중공선성이 존재하지 않음을 의미하며, 분산팽창계수가 10 이상인 경우에는 다중공선성에 문제가 있다고 판단한다. Case별 다중공선성 진단 결과를 Table 2에 나타냈다.

선택된 변수들의 다중공선성을 진단한 결과 Table 2와 같이 대부분의 분산팽창계수가 1~3 사이의 값을 나타냈으며, 가장 큰 분산팽창계수가 8.3을 나타냈다. 모든 변수의

Table 2. Result of Multi-Collinearity

Independent variable	Result of Multi-Collinearity by Criteria (VIF Value)		
	Case 1	Case 2	Case 3
Snow depth	1.739	3.109	2.172
Exceedance percentage of snow design criteria	2.352	3.071	1.806
Daily relative humidity	1.878	8.308	2.445
Daily minimum temperature	1.454	1.313	1.506
Daily maximum temperature	2.183	7.946	2.367
Snow vulnerability index	1.275	1.903	1.200

분산팽창계수가 10 미만을 나타냈으므로 모든 독립변수들을 동시에 입력하는 방식의 다중회귀모형을 구축하여도 문제가 없는 것으로 판단되었다.

3. 회귀분석 결과

본 연구에서는 다중회귀모형을 이용하여 전국의 대설피해 예측모형을 개발하기 위해 Oh et al. (2017)연구의 대설피해 취약성 분석 결과를 인용하였다. 취약성 분석 결과를 바탕으로 1등급을 Case 1, 2등급-3등급을 Case 2, 4등급-5등급을 Case 3으로 구분하여 다중회귀모형을 구축하였으며, 독립변수로는 최심신적설량, 설계기준 초과율, 상대습도, 최저기온, 최고기온, 취약성 지수를 이용하였다. 구축한 Case 별 표본수와 결과를 Table 3과 Fig. 2에 나타냈다. 분석 결과 Case 1의 경우 RMSE가 약 35억 원으로 나타났고, Case 2는 약 52억 원, Case 3의 경우 119억 원으로 나타났다.

Case 1에서 Case 3으로 갈수록 RMSE가 커지는 이유는 Case 3으로 갈수록 취약성 등급이 높아지고, 해당 지역에 대한 대설피해 집계 금액 또한 증가하기 때문에 모의피해액과, 실제 집계피해액간의 차이가 커지는 것으로 나타난다. 그러나 Case 3의 최대 집계 피해액이 970억 원인 점을 감안한다면 모형의 성능이 나쁘지 않은 것으로 판단되며 보다 정확한 Case 별 비교를 위해서는 RMSE를 정규화한 NRMSE를 이용한 비교가 필요하다.

Table 3. Result of Multiple Regression Analysis

Case	No. of Data	RMSE(Unit : Million Won)	NRMSE
1	59	3,559	11.5%
2	24	5,249	16.6%
3	41	11,937	12.1%

NRMSE를 비교한 결과 Case 1의 경우 11.5%로 나타났으며, Case 2가 16.6%, Case 3의 NRMSE가 약 12.1%로 나타났다. 0에 가까울수록 모형의 성능이 최고치를 나타내는 NRMSE가 모든 Case에서 20% 미만으로 도출되었으므로 모형의 예측력이 우수하다고 판단된다. 또한 본 회귀모형의 독립변수에 지역별 대설피해 취약성 지수를 이용하였으므로 지역별 특성인자가 고려되었다 할 수 있다. 이는 기존의 재해통계 기반 대설피해 예측 함수 개발에서 대설피해가 발생하지 않았던 지역의 표본 개수 부족이라는 고질적인 문제를 취약성 분석 결과물을 이용한 군집을 방안을 도입한다면 해결 할 수 있는 것으로 판단된다.

4. 결론

본 연구에서는 대설피해에 대해 선제적 대응 또는 사전 대비에 도움이 되는 대설피해 예측 모형을 개발하고자 하였으며, 표본 개수 부족으로 인해 대설피해 예측 함수 개발이

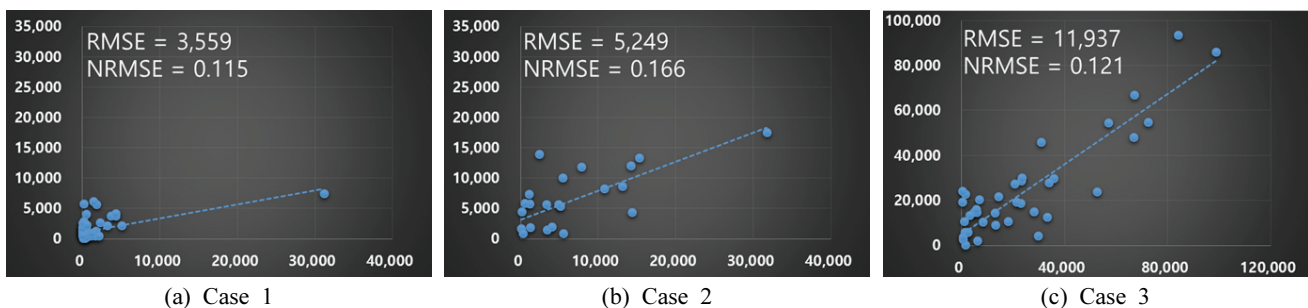


Fig. 2. Comparison of Observed (X-axis) and Calculated(Y-axis) Snow Damage from Multiple Regression Analysis (Unit : Million Won)

불가능한 지역의 문제를 취약성 결과를 통해 해결하고자 하였다. 전국을 5등급으로 구분한 취약성 분석 결과를 바탕으로 등급별 대설피해 예측함수를 개발하였으며, 1등급은 Case 1, 2등급~3등급은 Case 2, 3등급~4등급은 Case 3으로 구분하였다. 모형 구축 결과 각 Case 별로 NRMSE 15.5%에서 16.6% 사이의 결과를 나타냈으며, 전반적으로 NRMSE가 20% 미만으로 도출되어 모형의 예측력이 우수하다고 판단되었다. 따라서 본 연구의 결과물을 바탕으로 대설피해 사례가 거의 없던 지역까지 포함한 전국의 초기 대설피해액 예측이 어느 정도 가능 할 것으로 판단되며, 취약성 분석 결과를 이용한 군집으로 표본 개수 부족에 대한 고질적인 문제를 해결할 수 있는 것으로 판단된다. 또한 향후 기후변화 또는 삼한사온 등의 우리나라 기후 특성을 고려한다면 대설 피해 예측 모형의 성능이 향상 될 것으로 판단된다.

감사의 글

본 연구는 정부(행정안전부)의 재원으로 재난안전기술개발 사업단의 지원을 받아 수행된 연구임 [MOIS-재난-2015-05].

References

Jung, S.H., Yoon, J.S., Lee, J.W., and Lee, H.W. (2015) Comparison and Decision of Exposure Coefficient for Calculation of Snow Load on Greenhouse Structure. *Protected Horticulture and Plant Factory*, The Korean Society for Bio-Environment Control, Vol. 24, No. 3, pp. 226-234.

Kim, J.S., Choi, C.H., Lee, J.S., and Kim, H.S. (2017) Damage

Prediction Using Heavy Rain Risk Assessment: (2) Development of Heavy Rain Damage Prediction Function. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, pp. 371-379.

Kwon, S.H., and Chung, G.H. (2017) Estimation of Snow Damages using Multiple Regression Model: The Case of Gangwon Province. *Journal of the Korean Society of Civil Engineers*, Vol. 37, No. 1, pp. 61-72.

Lee, Y.K., Lee, C.J., and Ahn, S.I. (2015) Estimation of Freshly Fallen Snow Unit Weight and Maximum Probable Snow Load. *J. Korean Soc. Hazard Mitig.*, Vol. 15, No. 1, pp. 47-55.

Ministry of Public Safety and Security (MPSS) (2016) *The 2016 Annual Natural Disaster Report*.

Oh, Y.R., and Chung, G.H. (2017) Estimation of Snow Damage and Proposal of Snow Damage Threshold based on Historical Disaster Data. *Journal of the Korean Society of Civil Engineers*, Vol. 37, No. 2, pp. 325-331.

Oh, Y.R., and Chung, G.H. (2018) Snow Damage Vulnerability Analysis for Evaluation of Snow Damage in South Korea. *J. Korean Soc. Hazard Mitig.*, Vol. 18, No. 2, pp. 347-353.

<i>Received</i>	December 26, 2017
<i>Revised</i>	December 27, 2017
<i>Accepted</i>	January 2, 2018

