기상데이터 통계를 활용한 일 단위 용설 깊이 예측 가능성 검토

A Study on the Prediction of Daily Snowmelt Depth using Multiple Linear Regression

오영록* · 이규민** · 전경수*** · 선우우연**** · 백승우***** · 정건희*****

Oh, YeoungRok*, Lee, Gyunmin**, Jun, Kyung Soo***, Sunwoo, Wooyeon****, Baek, SeungWoo*****,
and Chung, Gunhui*****

Abstract

In this study, daily snowmelt was predicted using observed meteorological data and multiple regression analysis. Five observation stations (located in Daegwallyeong, Gwangju, Seosan, Mokpo, and Jeonju) were selected to analyze fresh snow depth from 2000 to 2010. The dependent variable used in the multiple regression analysis was daily snowmelt depth, and the independent variables were fresh snow depth, diurnal temperature range, temperature interception, diurnal humidity range, humidity intercept, and solar radiation. Seventy percent of the total observed data was used to develop a multiple regression model and the regression model was verified using the 30% of remaining data. The adjusted R-squared and Root Mean Square Deviation (RMSE) were used to examine the developed regression model. As a result, the adjusted R-squared was higher than 0.769 (except Daegwallyeong); thus the developed model represented well the daily snowmelt depth. Even Jeonju had an adjusted R-squared of 0.869. Also, the RMSE in all of the five stations was lower than 2.5 cm. The lowest value in Seosan was 1.7 cm. From the two types of verification, the developed multiple regression model was judged to be suitable to predict the daily snowmelt depth. However, multicollinearity should be explained, as rapid increases in temperature and sustained high temperature could not be reflected in the model. Therefore, if the limitations were resolved in further research, the model could be used to predict the amount of daily snowmelt depth more reliably.

Key words : Snow Melting Depth, Prediction of Snow Melting, Prediction of Snowdepth, Multiple Regression Analysis

요 지

본 연구에서는 기상자료를 활용한 용설 깊이 예측 가능성에 대해 검토하였다. 분석 방법으로는 다중회귀분석을 이용하였다. 분석 기간은 2000년부터 2020년까지이며, 분석 지점은 대관령, 광주, 서산, 목포, 전주 등 총 5지점을 선정하였다. 분석에 적용된 종속변수는 용설 깊이, 독립변수는 적설량, 기온변화량, 기온 절편, 습도변화량, 습도 절편, 일사량 등 6개를 적용하였다. 수집한 데이터의 약 70%를 다중회귀식 구축에 이용하고, 나머지 30%를 이용해 검증을 실시하였다. 수정된 결정계수와 Root Mean Square Deviation (RMSE)을 검증 방법으로 이용하였다. 검증 결과 대관령 지점에서 수정된 결정계수가 0.589로 다소 가장 낮게 나타났지만, 다른 모든 지점에서 수정된 결정계수가 0.769 이상으로 나타났으며, 전주 지점은 0.869 값을 나타내는 등 다수의 지점에서 1과 가까운 값을 나타냈으므로, 모형이 적합한 것으로 판단된다. RMSE 검증 결과 모든 지점에서 약 2.5 cm 미만의 오차를 기록하였고, 서산 지점에서 약 1.7 cm의 오차를 기록하였다. 두 종류의 검증 결과에서 모형이 적합한 것으로 판단되었으므로, 본 연구에서 제시한 방법은 일일 용설량 깊이를 예측할 수 있는 하나의 접근법으로 활용이 가능할 것으로 판단된다. 그러나 이용된 변수 다수에서 다중회귀식의 성능을 떨어트리는 다중공선성에 대한 문제가 의심되었고, 기온의 급상승이나, 지속되는 고온을 제대로 반영하지 못하는 문제점이 발견되었다. 향후 이러한 문제들을 해소한다면 오차를 크게 줄일 수 있을 것으로 판단된다.

핵심용어 : 용설 깊이, 용설 예측, 적설량 예측, 다중회귀분석

*정회원, 성균관대학교 수자원전문대학원 박사과정(E-mail: dudtla224@skku.edu)

Member, Doctor course, Graduate School of Water Resources, SungKyunKwan University

**정회원, 성균관대학교 건설환경연구소 선임연구원

Member, Research Professor, Construction and Environmental Research Center

***정회원, 성균관대학교 수자원전문대학원 교수

Member, Professor, Graduate School of Water Resources, SungKyunKwan University

****정회원, 성균관대학교 수자원전문대학원 박사과정

Member, Doctor Course, Graduate School of Water Resources, SungKyunKwan University

*****정회원, 기상청 전주기상지청 주무관

Member, Administrative Officer, Jeonju Branch Office of Meteorology, Korea Meteorological Administration

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

Corresponding Author, Member, Associate Professor, Department of Civil Engineering, Hoseo University

1. 서 론

1.1 연구배경 및 목적

우리나라는 사계절이 뚜렷한 나라로서 겨울철에는 눈으로 인한 교통의 혼잡이나 고립, 차량 사고와 같은 피해가 해마다 발생하고 있다. 눈은 지표면에 쌓인 이후 시간 경과에 따라 융해 및 승화되는 특성이 있으므로, 기존에 내린 눈이 남아있는 상태에서 추가로 강설이 발생할 경우 그 피해는 더욱 가중될 수 있다. 그렇기 때문에 남아있는 눈에 대한 예측은 대설피해 예측 또는 대비 측면에서 중요한 인자로 작용할 수 있다. 적설량과 관련된 연구를 살펴보면 융설로 인한 유출량 모의에 대한 연구나, 적설하중에 의한 피해 분석에 대한 연구가 대부분이며 융설 깊이에 대한 예측은 거의 찾아보기 어렵다. Kim et al. (2011)은 Terra MODIS 영상을 기반으로, 적설분포도 구축에 대한 연구를 진행하였으며, Yu et al. (2014)은 100년 빈도 적설하중을 이용하고 ArcGIS 크리깅(Kriging) 기법을 통해 우리나라 설계 적설하중 지도를 작성하였다. Oh and Chung (2018)은 기상학적 요소와 사회·경제적인 요소를 활용하여 대설피해 예측 방법을 제안하였다. Shin et al. (2006)은 DEM, 토지피복도, NDVI, 수문기상자료 등을 이용하여, 융설매개변수를 포함한 유출량 산출 방안에 대해 연구하였다. 이처럼 대부분의 연구가 융설 깊이에 대해서는 상대적으로 관심이 낮다. Guan et al. (2013)은 수문학적 모델링을 활용하여 수자원 관리를 개선하기 위해 Sierra Nevada 지점을 대상으로 Snow Water Equivalent (SWE)를 추정해보았으나, 대상 지점이 평균 해발고도 1,719 m인 네바다 설원으로 지중해성 기후를 띄고 있으며, 연평균 강설량이 10.4 m를 넘는 지역이므로, 우리나라에 적용하기에는 다소 어려움이 존재하였다. 그러므로 우리나라에 적용이 가능한 융설 깊이 예측 방법 개발에 대한 연구가 필요하다.

1.2 연구범위 및 방법

1.2.1 연구 범위

일반적으로 자연적인 적설량의 감소는 크게 융해나 승화 또는 자체 하중에 의한 압축 등 세 가지로 나눌 수 있다. Roebber et al. (2003)에 따르면 눈은 자체하중으로 인해

압축될 수 있다고 하였으며, Tarboton (1994)은 기온과 습도, 풍속 및 일사량 등을 이용하여 눈의 녹는 양과 승화되는 양을 모델링하고자 하였다. 기존의 연구들은 융설로 인한 유출량에 관심을 두고, 감소한 적설 깊이에서 승화와 압축으로 감소한 양을 제외한 나머지를 통해 유출량을 계산하고자 하였다. 그러나 본 연구에서는 통계적 방법을 이용하여 남아있는 적설량 예측 가능 여부에 관심을 두었다.

본 연구에서는 2000년 이후부터 2020년 까지 운영 중인 우리나라 종관기상관측소(Automated Synoptic Observing System, ASOS) 75개 지점 중, 일사량을 관측하면서 5 cm 이상의 적설 사상이 가장 많이 관측된 5개 지점을 대표로 선정하였다. 이때 한 개의 사상이란 기존의 잔여 적설량이 모두 사라진 이후 12시간 이내에 재강설이 발생하지 않은 경우를 가정하였다. 대표지점으로 선정된 관측소의 정보와 위치를 Table 1과 Fig. 1에 나타냈다. 일사량을 기록하는 관측소 중 5 cm 이상의 강설량이 발생한 건수는 대관령이 100건, 광주가 60건, 서산이 42건, 목포가 41건, 전주가 40건 순으로 가장 많이 발생하였다. 따라서 본 연구에서는 대관령, 광주, 서산, 목포, 전주 지점을 대상으로 선정하였으며, Table 2에 각 지점별 강설 사상을 나타내고 선정된 지점을 검게 나타내었다.

Fig. 1. Location of Observation Point

Table 1. Observation Point Information

Point No.	Name	Latitude	Longitude	Level (m)	Source	Data Period
100	Daegwallyeong	37.6771	128.7183	772.43	Korea Meteorological Administration	2000 ~ 2020
156	Gwangju	35.1729	126.8916	72.38		
129	Seosan	36.7766	126.4939	25.25		
165	Mokpo	34.8137	126.3815	44.70		
146	Jeonju	35.8408	127.119	61.40		

Table 2. Number of Snow Fall Occurrences from 2000~2020

Point No.	Area	Total SnowFall	over than 5 cm	Solar Radiation Quantity (O/X)	Point No.	Area	Total SnowFall	over than 5 cm	Solar Radiation Quantity (O/X)
90	Sokcho	126	59	X	189	Seogwipo	67	12	X
95	CheorWon	186	37	X	192	Jinju	109	32	O
98	Dongducheon	147	32	X	201	Ganghwa	67	12	X
100	Daegwallyeong	260	100	O	202	Yangpyeong	39	5	X
101	ChunCheon	164	39	O	203	Icheon	73	13	X
102	Baengnyeongdo	165	38	X	211	Inje	94	16	X
105	Gangneung	51	23	O	212	Hongcheon	150	34	X
106	Donghae	81	39	X	216	Taebaek	95	22	X
108	Seoul	153	30	O	221	Jecheon	106	22	X
112	Incheon	152	25	O	226	Boeun	99	30	X
114	Wonju	213	33	O	232	Cheonan	115	27	X
115	Ulleungdo	232	114	X	235	Boryeong	113	24	X
119	Suwon	161	26	O	236	Buyeo	183	46	X
121	Yeongwol	186	44	X	238	Geumsan	168	44	X
127	Chungju	171	24	X	243	Buan	81	17	X
129	Seosan	181	42	O	244	Imsil	71	16	X
130	Uljin	76	26	X	245	Jeongeup	73	23	X
131	Cheongju	199	39	O	247	Namwon	99	35	X
133	Daejeon	178	29	O	248	Jangsu	162	84	X
135	Chupungryeong	164	30	O	256	Juam	156	44	X
136	Andong	126	18	O	260	Jangheung	104	29	X
138	Pohang	36	9	O	261	Haenam	102	18	X
140	Gunsan	163	45	X	262	Goheung	52	8	X
143	Daegu	59	7	O	265	Seongsanpo	66	9	X
146	Jeonju	163	40	O	271	Bonghwa	30	2	X
152	Ulsan	28	8	X	272	Yeongju	27	6	X
155	Changwon	31	3	X	273	Mungyeong	84	22	X
156	Gwangju	155	60	O	277	Yeongdeok	88	10	X
159	Busan	17	4	O	278	Uiseong	83	20	X
162	Tongyeong	12	3	X	279	Gumi	20	4	X
164	Muan	6	0	X	281	Yeongcheon	56	8	X
165	Mokpo	171	41	O	284	Geochang	89	16	X
168	Yeosu	26	1	O	285	Hapcheon	36	6	X
169	Heuksando	114	19	O	288	Milyang	130	25	X
170	Wando	105	12	X	289	Sancheong	32	5	X
184	Jeju	61	13	O	294	Geoje	15	3	X
185	Gosan	42	2	O	295	Namhae	42	8	X
188	Seongsan	109	32	X					

1.2.2 연구 방법

적설량과 SWE에 대한 연구 사례들을 살펴보면, Elder et al. (1998)과 Molotch et al. (2005)은 적설깊이와 SWE 간에 비선형 관계가 존재함을 밝힌바 있으며, Winstral et al. (2002)은 비선형 관계를 모델링하기 위해 이진 회귀 트리 모델을 이용하였다. 그러나 본 연구에서는 통계적 방법을 이용하기 위해 남아있는 적설량에 대한 예측을 위한 것이므로, 일사량, 기온, 습도 등 여러 변수들과의 거동을 살펴보고 이를 기반으로 다중회귀분석 방법을 채택하였다. 다중회귀 분석은 통계학적으로 설명하고자 하는 변수와 여러 가지의 종속변수들 사이의 상관관계를 나타내는 선형 관계식을 구하는 기법이며, 이렇게 얻어낸 회귀식의 적합도를 측정하는 분석 방법으로 해석된다(S.H. Park, 2007).

2. 다중회귀분석

다중회귀분석은 두 개 이상의 독립변수를 이용해 하나의 종속변수를 설명하는 분석 방법으로서, 단순회귀분석의 확장형이라 볼 수 있다. 다중회귀분석의 종속변수 Y 를 n 개의 독립변수 $x_1, x_2, \dots, x_n$ 의 선형 식으로 설명하려는 방법으로 Eq. (1)과 같이 나타낼 수 있다(Oh and Chung, 2018).

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

여기에서 $B_0, B_1, \dots, B_n$ 는 모집단의 회귀계수를 나타낸다. 다중회귀식을 추정하는 방식으로는 동시입력방식과 단계입력방식 등이 있으며, 동시입력방식은 사용자가 고려하는 모든 설명변수들을 동시에 분석하는 방법으로, 다른 설명변수들이 통제된 상태에서 특정 설명변수의 영향력을 분석하여 모든 설명변수들이 얼마나 종속변수에 영향을 끼치는지 정도를 알 수 있다. 본 연구에서는 모든 독립변수들이 종속변수인 융설 깊이에 어떠한 영향을 끼치는지 알아보려고 하였기에 동시입력방식을 이용하였다.

다중회귀분석의 적용성 검토는 일반적으로 수정결정계수를 사용하며, 본 연구에서는 수정된 결정계수와 평균 제곱근 편차(Root Mean Square Deviation, RMSE)를 이용하였다. 수정결정계수는 단순회귀분석의 결정계수에서 독립변수가 많아질수록 결정계수 값이 커지는 점을 보완한 것으로 결정계수를 보정하는 역할을 한다.

수정결정계수(R_a^2)를 산정하는 공식은 Eq. (2)와 같다.

$$R_a^2 = 1 - \left(\frac{n-1}{n-k-1} \right) \frac{SSE}{SST} \quad (2)$$

여기서 n 은 자료의 수이며, k 은 독립변수의 수, SSE 는 오차 제곱합(Sum of Squares error, SSE)과 SST 는 총제곱합(Sum of Squares Total, SST)를 나타내며, Eqs. (3), (4)로

나타낼 수 있다.

$$SSE = \sum (y - \hat{y})^2 \quad (3)$$

$$SST = \sum (y - \bar{y})^2 \quad (4)$$

여기서 y 는 관측된 적설 변화량, $\hat{y}$ 는 산출한 적설 변화량, $\bar{y}$ 는 관측된 적설 변화량의 평균을 나타낸다.

평균 제곱근 편차(Root Mean Square Deviation, RMSE)는 관측치와 추정값 사이의 오차를 숫자로 나타내 한눈에 파악하기 쉬운 장점이 있으며 Eq. (5)와 같은 방법으로 구할 수 있다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_i - O_i)^2}{n}} \quad (5)$$

여기서 S_i 는 다중회귀식을 통해 산출한 남아 있는 적설량의 추정치를 나타내며, O_i 는 관측치를 나타낸다.

3. 변수 구성

3.1 기상데이터 수집

우리나라 기상청(Korea Meteorological Administration, KMA)에서는 ‘기상청 날씨마루’ 서비스를 통해 당일부터 다음날까지, 총 48시간 동안 매 시간별로 일사량 예측값을 제공하고 있다. 또한 ‘동네예보’ 서비스를 통해 당일부터 모레까지 3시간 단위의 기온과 습도, 풍향 등의 예측값을 제공하고 있다. 미래의 융설 깊이 예측 함수를 적용하기 위해서는 현재 관측되고 있거나 예보되는 독립변수를 이용해야 하므로, 현재 기상청에서 단기예보 중이면서 융설에 영향을 미칠 것으로 예상되는 일사량, 기온, 습도에 대한 데이터들을 수집하였다. 현재 기상관측소에서 관측·기록하는 적설량은 목측관측으로서 각 지점별로 매 시간 관측하는 곳과, 3시간 단위로 관측하는 곳으로 나뉘기 때문에 모든 자료를 매 시간별 단위로 이용하기에 무리가 있다. 또한 ‘동네예보’에서 3시간 단위의 예보 서비스를 제공하므로, 수집한 데이터의 시간단위는 매 시간 단위로 구축된 자료를 3시간 단위로 끊어서 이용하였다.

3.1.1 융설 영향 인자 분석

다중회귀분석에 이용될 독립변수를 구축함에 앞서, 융설률에 영향을 미치는 요인이 무엇인지 파악하여야 한다. Kim et al. (2014)에서 SWAT 모형을 이용하여 융설이 충주댐 유입량에 미치는 영향을 평가하였다. 이때 융설은 임계온도에 의한 선형함수로 나타내었으며, Eq. (6)에 나타났다.

$$SNO_{mlt} = b_{mlt} \times sno_{cov} \times \left[\frac{T_{snow} + T_{max}}{2} - T_{mlt} \right] \quad (6)$$

여기서, b_{mlt} 는 융설률(mm/day/°C), sno_{cov} 는 유역의 눈으로 덮인 면적비, T_{snow} 는 적설온도(°C), T_{max} 는 일 최고기온(°C)이다. 이때 융설률(b_{mlt})는 다음 Eq. (7)과 같이 나타낼 수 있다.

$$b_{mlt} = \frac{(b_{mlt6} + b_{mlt12})}{2} + \frac{(b_{mlt} - b_{mlt12})}{2} \times \sin\left(\frac{2\pi}{365}(d_n - 81)\right) \quad (7)$$

여기서, d_n 는 1년의 일(day number), b_{mlt6} , b_{mlt12} 는 각각 6월 21일과 12월 21일의 융설계수이다. 여기에서 융설계수는 자연유역에서는 약 1.4 ~ 6.9 mm H₂O/day/°C로 알려져 있으며, 도시유역은 보행자와 차량의 운행 등 인위적인 요인에 따라 큰 영향을 받지만 대체로 3.0 ~ 8.0 mm H₂O/day/°C의 범위를 가지는 것으로 알려져 있다(Jeong et al., 2005). J.Y. Park (2007)에 따르면 수증기의 압력이 포화 수증기압보다 낮을 경우, 즉 상대습도가 100% 미만인 경우 영하의 기온에서도 얼음이 승화가 지속적으로 일어난다고 하였다. 즉 얼음의 승화는 기압과, 수증기압과의 관계(상대습도)에 따라 정도가 달라질 수 있는 것이다. 따라서 본 연구에서는 눈의 융해에 직접적으로 영향을 미칠 것으로 예상되는 기온 및 일사량 변수와 눈의 승화에 영향을 미칠 것으로 예상되는 상대습도를 기상변수로 선택하여 관측된 적설량과의 관계들을 살펴보았다. 이를 위해 2000년부터 2020년까지의 기간 동안 모든 관측소에서 관측된 1,867개의 사상들을 수집하고, 적설량이 감소하며 변화하는 일사량, 기온, 습도의 상태를 살펴보았으며 Fig. 2에 몇 가지 표본을 나타내었다. Fig. 2에서 비어있는 원은 밤 시간(18시~09시)에 기록된 적설량이며, 채워진 원은 낮 시간(09시~18시)에 기록된 적설량이다. Fig. 2에서 알 수 있듯 대부분의 밤 시간의 적설량은 증가하거나, 일직선을 이루고 있다. 점들의 기울기 증가는 강설발생 중인 것을 나타내고, 일직선은 현재 상태 유지를

나타낸다. 즉, 눈이 오지 않는 밤 시간에는 거의 모든 사상에 서 눈의 양이 변화하지 않고 현재 상태가 유지되고 있음을 알 수 있었다. 또한 Fig. 2에 표시된 2번을 제외한 1번 ~ 6번까지의 기온과 습도를 보면, 습도가 감소하고 기온이 증가할 때 눈의 양이 감소하였다. 2번에서는 기온은 영상의 온도로 증가했으나, 습도가 감소하다 중간 지점에서 급상승하는 모습을 보이고 있으며, 이때는 눈의 양이 별로 감소하지 않았다. 7번도 마찬가지로 기온은 거의 영상에 근접하게 증가하고 있으나, 습도가 중간에 다시 상승하는 모습을 보이고, 적설량은 습도가 증가하면서 부터 줄어들지 않았다. 9번과 12번 또한 기온은 거의 영상에 가깝거나 그 이상이지만, 습도에 따라 적설량이 많이 감소하지 않는 모습을 보이고 있다. 그러나 13번을 보면 12번 보다 기온이 낮음에도 불구하고, 습도가 급하게 줄어들며 적설량도 크게 감소하는 모습을 보였다. 또한 2번, 5번, 8번, 10번, 11번 등에서 눈이 녹기 전에 기온과 습도가 먼저 변화하는 모습을 보였다.

3.1.2 융설 영향 인자 분석 결과

기존 연구와, 우리나라의 적설 관측 당시의 기상변수 거동을 살펴본 결과, 적설량은 일사량, 기온, 습도 등에 영향을 받는 것으로 예상할 수 있었으며, 밤 시간에 눈의 양이 감소하는 사례가 거의 없었다. 또한 눈의 양이 감소하기 전 시간대에서 기온은 먼저 상승하고, 습도는 하락하는 것을 파악할 수 있었다. 즉, 본 연구에서 3시간 단위로 구축한 데이터를 기반으로 하였을 경우 06시에서 15시 사이에 기온이 증가하고, 습도가 감소하는 동안 적설량은 09시부터 18시까지 감소하는 모습이 관찰됐다.

3.2 분석데이터 구축

다중회귀분석을 실시하기 위해, 09시부터 18시까지의 적설 변화량을 종속변수로 선정하였으며, 독립변수로 09시 적설량, 09시~18시까지의 일사량 합계, 06시~15시까지의 기온·습도 절편, 기온·습도 변화량을 이용하였다. Table 3에 각 변수별 시간범위 및 출처를 나타내었다.

Fig. 2. Changes in Snow Depth and Weather Factors

Fig. 2. (Continued)

Table 3. Information of Variables

No.	Type	Name	Time Range	Source
1	Dependent variable	Melted Snow depth	09 ~ 18	ASOS Data From Korea Meteorological Administration
2	Independent variable	Snow depth	09 O'clock	
3		Sum of all solar radiation	09 ~ 18	
4		Rate of temperature change	06 ~ 15	
5		Temperature-intercept	06 ~ 15	
6		Rate of humidity change	06 ~ 15	
7		Humidity-intercept	06 ~ 15	

이때 기온과 습도의 변화량은 다음 Eq. (8)을 이용하여 산출하였으며, 기온과 습도의 절편은 Eq. (9) 방법을 이용하여 산출하였다.

$$a = \frac{\sum_{i=1}^n (x_i - x_{mean})(y_i - y_{mean})}{\sum_{i=1}^n (x_i - x_{mean})^2} \quad (8)$$

여기서 a는 기온 또는 습도의 기울기이며, x는 시간, y는 기온 또는 습도이다.

$$b = y_{mean} - (x_{mean} \cdot a) \quad (9)$$

여기서 b는 기온 또는 습도의 절편을 나타낸다.

3.3 다중공선성 진단

다중 회귀분석 방법을 이용할 때, 독립변수의 개수를 다수 이용하게 되는 경우가 발생할 수 있다. 이러한 경우를 고차원 자료(High-dimensional-data)라고 하는데, 고차원 자료를 이용한 회귀분석의 경우 일부 독립변수의 회귀계수가 추정되지 않는 상황이 발생할 수 있다. 다중회귀분석은 “독립변수들 간에 상관관계가 존재하지 않는다.”라는 가정하게 회귀식이 구축되기 때문이다. 그러므로 수집된 독립변수들 사이의 강한 상관관계로 인한 다중공선성(Multicollinearity) 문제가 심각한 회귀식이 개발된 경우 예측값이 실측값과 큰

차이를 나타낼 가능성이 크다(Kim et al., 2018). 다중공선성을 해소하기 위해서는 연구자의 직관에 의존하여 일부 변수만을 선택하거나 혹은 통계적 변수선택 절차를 적용할 수도 있으나 실제로 어떤 방법으로 알맞은 변수를 선택할지에 대한 논의는 아직까지 지속되고 있다(Kim et al., 2018). 일반적인 다중공선성의 진단 지표로는 분산팽창계수(Variance Inflation Factor, VIF)를 주로 사용하고 있으며, 분산팽창계수가 1이면 다중공선성이 존재하지 않는 것을 의미하고 5 이상의 경우 다중공선성 문제 의심, 10 이상인 경우 다중공선성 문제 발생이라는 기준이 대표적이다. 각 지점별 분산팽창계수를 검토한 결과, 모든 지점에서 10을 넘는 사례가 존재하였으며 몇몇 지점의 변수에 대해서는 VIF 값이 15를 넘기도 하였다. 일반적으로 다중공선성 문제가 발생하면 비록 수정된 결정계수(R_a^2) 값이 1에 가깝게 나타나더라도 개별 추정량의 통계적 유의성이 없게 나타는 것으로 판단한다. 따라서 본 연구에서는 객관적인 성능 평가를 위해 수정된 결정계수 외에, 회귀식 구축에 이용되지 않은 표본들의 RMSE를 분석하였다. 다중공선성 진단 결과를 Table 4에 나타내었다.

4. 회귀분석

4.1 분석 절차

우리나라의 적설량은 목측으로 기록하고 있으므로, 적설 변화량을 파악할 수 있게 총 적설량이 5 cm 이상인 강설 사상들 중 1 cm 이상 눈이 녹은 경우를 표본으로 추출하였다.

Table 4. Result of Multi-Collinearity

Independent Variable	Result of Multi-Collinearity by Criteria (VIF Value)				
	Daegwallyeong	Gwangju	Seosan	Mokpo	Jeonju
Snow depth	2.5	3.5	3.8	3.8	5.3
Sum of solar radiation	12.3	12.6	10.2	8.6	11.3
Rate of temperature change	8.1	8.1	8.2	5.9	10.6
Temperature-intercept	6.2	8.9	5.7	5.1	7.3
Rate of humidity change	9.2	12.5	11.3	16.7	16.0
Humidity-intercept	8.6	8.5	7.3	10.9	13.1

첫째, Fig. 3의 1번과 같이 구축한 기상데이터를 이용해 Fig. 3의 2번과 같이 회귀분석에 이용될 표본을 구축한다.

- 09시 적설량 선택
- 09시부터 18시까지의 일사량 합계 계산
- Eqs. (6)과 (7)을 이용하여 기온과 습도의 변화량 및 절편 계산

둘째, 구축된 전체 표본들 중 무작위로 70%의 표본을 추출하여 다중회귀식을 도출한다.

셋째, 모형 구축에 사용되지 않는 표본의 융설 깊이를 산정한다.

- 모형 구축에 이용되지 않은 데이터들을(30%) 오차 검증 표본으로 추출한다.
- 다중회귀식을 이용하여 융설 깊이를 산정한다.

넷째, 산정된 적설량과, 관측된 적설량을 비교하여 오차를 분석한다.

- 09시 적설량과 산정된 융설 깊이의 차이를 이용해 18시의 잔여 적설량을 산정한다(Eq. (10)).

$$Snowdepth_{09} - Melting of Snowdepth_{09 \sim 18} = Snowdepth_{18} \quad (10)$$

- 산정된 값과 관측된 값 사이의 RMSE를 통해 오차를

분석한다.

Table 5에 각 지점별 사상 개수와 다중회귀분석에 이용된 표본 수 및 검증에 이용된 표본 수를 나타냈다. Table 5를 살펴보면 대관령이 총 686개의 사상으로 가장 많았으며, 광주가 176, 전주 130, 목포 103, 전주 87개로 타났으며, 각 지점별로 대관령은 213개, 광주 58개, 서산 46개, 목포 33개, 전주 29개의 표본으로 검증하였다.

5. 분석 결과

각 지점별 회귀계수를 Table 6에 나타내고, 수정된 결정계수 및 RMSE 값을 Table 7에 나타냈으며, Fig. 4에 예측값과 실측값 간의 차이를 도식화 하였다. Table 6의 수정된 결정계수를 보면 대관령이 0.589로 가장 낮고, 전주가 0.869로 가장 높게 나타났다. RMSE를 보면 대관령 지점이 약 2.68 cm로 가장 높고 전주가 2.33 cm로 두 번째를 나타냈으며, 서산이 1.73 cm로 가장 좋은 결과를 나타냈다. Fig. 4의 Box plot을 보면 대관령의 검증표본 총 213개 중 200개의 사상에서 약 5 cm 미만의 오차를 나타냈으며, 181개의 사상이 3 cm 미만의 오차를 보였다. 광주는 총 58개의 표본 중 51개의 사상에서 3 cm 미만의 오차를 보였다. 서산은

Fig. 3. Multiple Regression Analysis Process

Table 5. Number of Samples by Region

Independent Variable	Daegwallyeong	Gwangju	Seosan	Mokpo	Jeonju
Total	686	176	130	103	87
Learning Sample	473	118	84	70	58
Verification Sample	213	58	46	33	29

Table 6. Multiple Regression Coefficient

Variable	Area				
	Daegwallyeong	Gwangju	Gwangju	Mokpo	Jeonju
Snow Depth	-0.051	-0.249	-0.262	-0.242	-0.311
Sum of Solar Radiation	-0.519	0.400	-0.138	-1.131	0.145
Rate of Temperature Change	-0.207	-0.017	-0.109	-0.262	-0.127
Temperature-Intercept	-0.040	0.005	-0.047	-0.016	0.038
Rate of Humidity Change	-0.027	-0.019	-0.016	-0.005	-0.002

Table 7. Result of Multiple Regression Analysis

Area	Result of Multiple Regression Analysis	
Daegwallyeong	R_a^2 : 0.589	RMSE (Unit : cm) : 2.688
Gwangju	R_a^2 : 0.839	RMSE (Unit : cm) : 2.004
Seosan	R_a^2 : 0.845	RMSE (Unit : cm) : 1.732
Mokpo	R_a^2 : 0.769	RMSE (Unit : cm) : 2.064
Jeonju	R_a^2 : 0.869	RMSE (Unit : cm) : 2.335

Fig. 4. Comparison of Observed and Estimated Snow Depth using Multiple Regression Analysis

Fig. 4. (Continued)

46개의 사상 중 42개, 목포는 33개 중 31개, 전주는 29개 중 24개의 데이터가 3 cm 미만의 오차를 나타냈다. 전체 검증 표본으로 계산하였을 경우, 총 379개의 표본 중 329개의 표본이 3 cm 미만의 오차를 나타냈다. 즉 검증 결과 86.8% 표본에서 3 cm 미만의 오차를 나타냈다.

비교적 큰 오차인 8 cm 이상의 오차는 목포가 1건 대관령이 5건이며 모두 융설 깊이를 과소 측정된 사례이다. 각 사례들의 기온을 살펴보면, 대관령의 경우 기온이 -18.3°C 에서 2.3°C 까지 급격히 상승하거나, -16.6°C 에서 3.3°C , 5.3°C 에서 3.3°C , -0.2°C 에서 9.5°C 등 기온이 급격히 상승한 사례들이었으며, 목포 지점의 경우 2005년 12월 23일 겨울철 날씨임에도 불구하고 4.9°C , 5.5°C , 7.9°C 를 기록한 사례였다.

6. 결 론

본 연구에서는 통계적 방법을 이용하여 일일 융설 깊이 예측 가능성에 대해 검토하고자 하였다. 이를 위해 우리나라 기상청에서 운영하는 중관기상관측소(Automatic Weather System, ASOS)에 기록된 데이터를 기반으로 강설이 5 cm 이상 기록된 사상들이 가장 많은 대관령, 광주, 서산, 목포, 전주를 연구 지점으로 선정하였다. 또한 우리나라의 적설은 목측관측이므로 눈이 1 cm 이상 녹은 사례들을 대상으로 다중회귀분석을 실시하였다. 독립변수로 적설량, 기온변화율, 기온 절편, 습도변화율, 습도절편, 일사량 등 6가지를 활용하였다. 예측력 평가를 위해 수정된 결정계수와 RMSE 값을 이용하였고, 그 결과 모든 지점에서 RMSE 값이 약 2.7 cm 미만으로 나타났다. 또한 대관령 지점에서 수정된 결정계수 값이 0.589로 다소 낮은 값을 나타냈지만, 다른 모든 지점에서 0.75 이상의 값을 나타내고, 최대 0.869의

값을 나타냈다. 오차 검증 결과 대부분의 눈이 녹는 사례들을 2.7 cm의 오차 내로 예측이 가능한 것으로 나타났으며, 본 연구에서 활용한 독립변수들은 기상청 예보 시스템을 통해 수집할 수 있으므로, 일일 융설량 예측에 활용 가능한 접근법으로 볼 수 있을 것으로 판단되지만 몇 가지 개선사항 또한 존재하였다. 먼저 다중공선성 진단에서 분산팽창요인(VIF) 값이 10 이상인 값이 다수 존재하여, 다중공선성 문제가 의심되나 명확히 해소하지 못하였다. 또한 독립변수는 고온의 기온이 지속되거나, 이례적으로 기온이 급상승 하는 등의 상황에서 다소 큰 오차가 발생하였다. 마지막으로 09시부터 18시까지의 융설 깊이를 일일 융설 깊이로 가정하여 분석하였으므로, 09시부터 18시 사이의 융설 깊이를 시간 단위로 예측할 수 없는 단점이 있다. 따라서 이러한 단점을 해소하기 위해 결정 트리 학습방법 등을 통해 기상 상화에 따라 Case를 구분하고, 독립변수의 차원이 축소된 분석 방법 등이 필요할 것이다. 또한 모형의 활용성을 증가시키기 위해, 시간 단위를 축소시킬 필요가 있다. 그럼에도 불구하고 통계적 방법을 이용하여 일일 융설 깊이 예측에 대한 가능성을 보였으므로, 향후 고도화된 분석기법을 적용한다면 현재 관측, 기록하고 있는 기상데이터를 활용해 보다 정확한 융설 깊이를 예측 할 수 있을 것으로 판단되며, 추가적으로 눈이 잘 녹지 않는 기후조건을 가진 지역을 파악해 제설함 등의 제설장비 투입 등에 대한 연구도 진행할 수 있을 것으로 판단된다.

감사의 글

본 결과물은 환경부의 재원으로 한국환경산업기술원의 지능형 도시수자원 관리사업의 지원을 받아 연구되었습니다(2019002950002).

References

- Elder, K., Rosenthal, W., and Davis, R.E. (1998). Estimating the spatial distribution of snow water equivalence in a montane watershed. *Hydrological Processes*, Vol. 12, No. 10-11, pp. 1793-1808.
- Guan, B., Molotch, N.P., Waliser, D.E., Jepsen, S.M., Painter, T.H., and Dozier, J. (2013). Snow water equivalent in the Sierra Nevada: Blending snow sensor observations with snowmelt model simulations. *Water Resources Research*, Vol. 49, No. 8, pp. 5029-5046.
- Jeong, J.U., Kim, D.G., Park, J.H., and Park, C.G. (2005). Runoff simulation take into consideration of snow pack and snow melt. *Proceedings of 2005 Annual Conference*, Korea Society of Civil Engineers, pp. 1290-1293.
- Kim, J.H., Park, J.H., Choi, C.H., and Kim, H.S. (2018). Development of regression models resolving high-dimensional data and multicollinearity problem for heavy rain damage data. *Journal of the Korean Society of Civil Engineers*, Vol. 38, No. 6, pp. 801-808.
- Kim, S.B., Ahn, S.R., Shin, H.J., and Kim, S.J. (2014). Assessment of snowmelt impact on Chungju Dam watershed inflow using Terra MODIS data and SWAT model. *Journal of the Korean Society of Civil Engineers*, Vol. 34, No. 2, pp. 457-467.
- Kim, S.B., Shin, H.J., Lee, J.W., Yu, Y.S., and Kim, S.J. (2011). Mapping technique for heavy snowfall distribution using Terra MODIS images and ground measured snowfall data. *Journal of the Korean Association of Geographic Information Studies*, Vol. 14, No. 4, pp. 33-43.
- Molotch N.P., Colee, M.T., Bales, R.C., Dozier, J. (2005). Estimating the spatial distribution of snow water equivalent in an alpine basin using binary regression tree models: The impact of digital elevation data and independent variable selection. *Hydrological Processes*, Vol. 19, pp. 1459-1479.
- Oh, Y.R., and Chung, G.H. (2018). Multiple regression models of snow damage prediction according to the snow damage vulnerability groups. *J. Korean Soc. Hazard Mitig.*, Vol. 18, No. 2, pp. 355-359.
- Park J.Y. (2007). Misunderstanding and understanding of the phase diagram for water: Water evaporation and ice sublimation in the atmosphere. *Journal of the Korean Chemical Society*, Vol. 51, No. 6, pp. 577-584.
- Park, S.H. (2007). *Regression analysis* (3rd ed.). Seoul: Minyoungsa.
- Roebber, P.J., Bruening, S.L., Schultz, D.M., and Cortinas Jr, J.V. (2003). Improving snowfall forecasting by diagnosing snow density. *Weather and Forecasting*, Vol. 18, No. 2, pp. 264-287.
- Shin, H.J., Kang, S.M., Kwon, H.J., and Kim, S.J. (2006). Snowmelt impact on watershed hydrology using climate change scenarios-Soyanggang-dam and Chungju-dam watersheds. *Proceedings of the KSRS Conference*, The Korean Society of Remote Sensing, pp. 198-201.
- Tarboton, D.G. (1994). *Measurements and modeling of snow energy balance and sublimation from snow*. Report Paper 61. Retrieved from https://digitalcommons.usu.edu/water_rep/61
- Winstral, A., Elder, K., and Davis, R.E. (2002). Spatial snow modeling of wind-redistributed snow using terrain-based parameters. *Journal of Hydrometeorology*, Vol. 3, No.5, pp. 524-538.
- Yu, I.S., Kim H.Y., Necesito, I.V., and Jeong, S.M. (2014). Assessment and improvement of snow load codes and standards in Korea. *Journal of the Korean Society of Civil Engineers*, Vol. 34, No. 5 pp. 1421-1433.

Received	October 21, 2020
Revised	October 21, 2020
Accepted	October 29, 2020