

Forecasting Model for Flood Risk at Bo Region

S. H. Kwon* · H. S. Oh**†

*Department of Statistics, Hannam University

**Department of Industrial and Management Engineering, Hannam University

보 지역 홍수 위험도 예측모형 연구

권세혁* · 오현승**†

*한남대학교 경상대학 비즈니스통계학과

**한남대학교 공과대학 산업경영공학과

During a flood season, Bo region could be easily exposed to flood due to increase of ground water level and the water drain difficulty even the water amount of Bo can be managed. GFI for the flood risk is measured by mean depth to water during a dry season and minimum depth to water and tangent degree during a flood season. In this paper, a forecasting model of the target variable, GFI and predictors as differences of height between ground water and Bo water, distances from water resource, and soil characteristics are obtained for the dry season of 2012 and the flood season of 2012 with empirical data of Gangjungbo and Hamambo. Obtained forecasting model would be used for keep the value of GFI below the maximum allowance for no flooding during flooding seasons with controlling the values of significant predictors.

Keywords : Bo Region, Empirical Analysis, General Linear Model, GFI

1. 서론

보를 설치한 주변 지역은 수위를 일정 수준 유지하는 보 특성으로 인하여 갈수와 달리 호우 시 홍수 위험에 노출되게 된다. 홍수 시 보에 유입되는 물의 양이 많아도 일정 수위를 유지하기 위하여 물을 배출하지만 보 주변 지역의 물의 양은 보의 존재로 인하여 배출되는 속도가 현저히 떨어지게 되고 이로 인하여 보 주변 지역의 홍수 위험 정도가 높아지게 된다. 그러므로 호우기의 보 지역의 홍수 위험도는 보 수위와 주변 지역의 물의 근원인 관정 수위의 차이는 물론 물의 유입과 배출에 영향을 주는 토양과 지질 인자들에 의해 결정된다[3].

지하수위 변동 분석 사례를 보면 독일 Dresden 지역의

홍수 동안 평균 최고 지하수위 이상의 지하수위, 지속기간, 지표면과 최고치 지하수위와의 차이, 그리고 지하수위 상승 경사에 관하여 수행된 바 있다. 그 연구에서는 지하수 홍수의 강도를 지하수위 상승 시 경사율과 지하수위 최고치와 지표면 사이의 최소 차이로 정의하였다[5]. 본 연구에서는 국내 기상 및 수문 현황 조건을 고려하여 호우기에는 평균 최고 지하수위 산정을 기간별 지하수위 분포의 백분율을 활용하여 홍수 위험도를 산정하였다[8].

보 주변 지역을 호우기에 홍수로부터 보호하려면 홍수 위험도를 예측하고 홍수 안정성이 보장되는 수준 이하로 홍수 위험도를 유지할 수 있도록 인자 수준을 결정하는 예측모형을 도출하는 것이 필요하다. 예측모형의 목표변수는 홍수 위험도이고 설명인자는 보 지역 물 배수에 영향을 미치는 변인인 보 지역 토양 및 지질 특성과 보 물 유입량과 배출량, 그리고 보 수위 관정 수위의 차이 등 수위에 의한 변인이다.

본 연구에서는 실증분석을 통하여 갈수기와 호우기에 적용할 수 있는 통합 홍수 위험도 예측모형 도출이 가능한지 추론하였다. 그리고 서로 다른 두 지역에 모두 적용될 수 있는 단일 예측모형 도출이 가능한지도 검정하였다. 제 2장에서는 실증분석 데이터로 사용된 함안보, 강정보 지역 관측 데이터를 기술하였으며, 갈수기와 호우기의 예측모형 도출과 모형의 활용방안을 제 3장에 정리하였다. 그리고 제 4장에는 결론과 예측모형 활용 방안을 기술하였다.

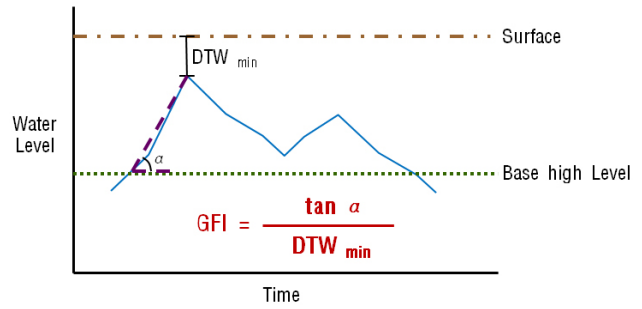
2. 관측자료의 분석

지하수 자동관측 데이터를 분석하여 지하수위 변동에 따라 호우기와 갈수기로 구분하여 홍수 위험도 예측이 필요하다고 판단하였다. 호우기는 강우 등으로 인하여 지하수위의 변동이 있는 시기로 설정하였으며 갈수기는 강우가 거의 없어 지하수위 변동이 적은 동절기를 포함하여 설정하였다. 본 연구의 실증분석에 사용된 데이터는 낙동강 지역의 2개의 보에서 측정하였다. 보의 위치 정보는 정부의 보 안내 페이지에서 찾을 수 있다[7]. 창녕의 함안보 지역의 경우에는 갈수기 데이터는 2012년 3월~6월, 호우기 데이터는 2012년 7월~10월에 수집하였고, 신천·백천 지역의 강정보 지역은 호우기 데이터는 2012년 5월부터 10월, 갈수기 데이터는 2012년 11월부터 2013년 2월까지 수집하였다[8].

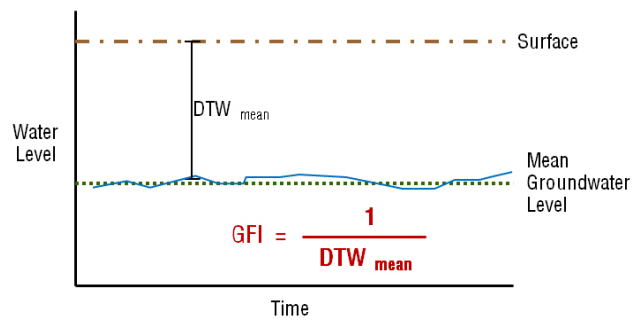
예측모형의 목표변수인 홍수 위험도의 측정값인 보 지역 홍수 강도(GFI; Ground water Flood Intensity)는 호우기와 갈수기에 따라 다르게 정의되어야 한다. 호우기는 GFI를 지하수위 상승 시 경사를 $\tan(\alpha)$ 과 지하수위 최고치와 지표면 사이의 최소 차이(DTW_{min} ; Depth To Water)의 관계로 측정하였고(<Figure 1> 참조) 갈수기에는 GFI를 담수 이후의 갈수기 평균 지하수위(DTW_{mean}) 역수로 측정하였다(<Figure 2> 참조).

갈수기와는 달리 호우기에는 지하수위 상승 시 GFI 값을 결정하는 경사율은 지하수위 상승 시점을 판단할 수 있는 기준 수위에 따라 달라지므로 상승 시점에 대한 정의가 필요하다. 본 연구에서는 지하수위 상승 시점을 기저고수위(Base High Level)로 정의하고 측정기간 동안 지하수위 분포의 80% 백분위 값을 기저고수위로 설정하였다.

보 주변 지역의 관정의 지하수위에 영향을 미치는 정량 인자를 정리하면 다음과 같다. 지하 관정과 하천과의 거리, 해수면 기준 지하수위의 위치, 물의 주 근원이 낙동강 수위와의 차이가 관정 지하수위에 영향을 미치고 관정의 지하수위가 홍수 위험도 관측값이 GFI에 영향을 미친다. 관정과 하천과의 거리가 멀수록 GFI는 낮아지고 해수



<Figure 1> GFI in a Flood Season



<Figure 2> GFI in a Dry Season

<Table 1> Quantitative Variables

Variables	Effect
distance between ground water and river(DBGR)	-
shortest distance between ground water and river (SDBGR)	-
ground water level based on sea level(GWL)	+
height difference between ground altitude of ground water and Nakdong river level(HDBGN)	-
height difference between ground water level and Nakdong river level(HDGN)	-

면기준 지하수위가 클수록 GFI는 높아진다. 그리고 물의 주 근원이 낙동강 수위와 지하수위 차이가 낮을수록 GFI는 커진다. <Table 1>에 예측모형의 목표변수 GFI에 영향을 미치는 정량 인자들을 리스트 하였고 각 인자들이 GFI에 영향을 미치는 양/음 관계를 부호로 표시하였다.

보의 물 배출에 영향을 주는 보 지역 토양 특성인 침식등급, 지형, 배수등급, 모암, 경사, 유효 토심, 표토토성, 심토 토성, 수문학적 토양군에 대한 측정값은 정밀토양도를 활용하였고[6], 토지이용 상태는 환경부가 제공하는 환경공간정보서비스[9]의 토지피복도를 활용하였다. 사전 일요인 분석분석을 실시하여 토양 특성과 토지이용 상태 중 GFI에 영향을 주는 변인을 선택하고 선택된 변인의 수준도 GFI에 유의한 영향을 미치도록 조정하였다. <Table 2>에 선택된 정성변인들과 GFI 예측모형 도출을

위하여 사용하게 될 각 변인의 수준(이진형, 앞의 수준 = 0, 나중 수준 = 1)을 정리하였다.

<Table 2> Qualitative Variables

Type	Class
drainage class	good / poor
effective soil depth	< 100 / > = 100
land lay	flatland / mountain
surface soil type	silty loam / etc.
subsoil type	silty clay loam / etc.
hydrological soil type	A, B / C, D
ground coverage	field / rice paddy and cultivation under structure

예측변인 데이터 수집 시 포인트와 그리드 개념으로 측정하였다. 포인트 데이터는 지하수위 관측점이 위치한 지하수 관측점에 대한 속성 자료는 관측점 자료를 공간적 분포를 시킨 후 정밀토양도와 토지피복도 도면에 중첩시켜 각 관정별 데이터를 추출하였다. 관측점 인근 100m 제곱 셀의 최대 면적비율을 차지하는 속성 자료는 수치지형도를 참조하여 추출하였다.

그리드 데이터 추출은 분석 및 적용될 지하수위 데이터의 현실성과 현장성을 고려하기 위한 것으로 포인트 데이터 추출 방법과 동일하나 100m 제곱 셀 격자를 구성한 후 각 관정이 속해 있는 셀의 최대면적 속성을 추출하는 차이가 있다.

3. 예측모형 도출

본 연구에서 예측모형 도출을 위하여 일반선형모형을 활용하였으므로 모형 적합성을 높이기 위하여 추정 전에 홍수 위험도 포함한 측정형 변수에 대한 Anderson Darling AD-통계량 정규성 검정을 실시하였다[1]. 정규성 검정결과 정규분포를 따르지 않는 변수들은 Box-Cox 정규분포 변환을 실시하여 예측모형에 활용하였다[2]. 각 변수에 대한 정규분포 변환 결과는 <Table 3>에 정리하였다.

관정과 하천 수위 차이, 관정과 하천과의 거리는 정량 예측변인이고 토양특성 및 토지이용실태 변인은 정성적 변인이므로 이를 동시에 고려하여 예측모형을 설정하기 위하여 일반선형모형을 적용하여야 한다. 정량변인과 정성변인의 결합효과, 정성변인들의 교호작용을 고려하면 모형의 복잡성으로 추정과 변수선택의 어려움이 있으므로 정량변인들은 회귀분석, 정성요인은 분산분석을 개별적으로 실행하여 유의한 변인들만으로 일반선형모형을 추정하였다[4].

<Table 3> Box-Cox Normality Transformation

Variables	Transformation
Ground water Flood Intensity	log
distance between ground water and river	log
shortest distance between ground water and river	square root
ground water level based on sea level	log
height difference between ground altitude of ground water and Nakdong river level	log
height difference between ground water level and Nakdong river level	square root

3.1 갈수기 예측모형

목표변수 $\ln(GFI)$ 와 수위차이와 거리차이 등의 정량 변인들의 회귀계수 추정, 다중공선성 진단, 제거(Backward) 방법에 의한 유의한 변수선택(유의수준 5%), 이상치 진단(스튜던트 잔차 기준 ± 2) 결과를 <Table 4>에 정리하였다. <Table 4>에는 함안보와 강정보 지역에서 포인트와 그리드 데이터를 활용하여 목표변수 $\ln(GFI)$ 에 유의한 영향을 정량변인들의 추정 회귀계수(셀의 값), 표준화 회귀계수(셀의 괄호 안 값), 결정계수를 제시되어 있다.

<Table 4> Regression Results(Dry Season)

Independent variables	(Standardized) Beta			
	Hamanbo		Gangjungbo	
	Point	Grid	Point	Grid
Intercept	-1.05	-0.97	-27.7	-27.2
log(GWL)	1.87 (1.28)	1.86 (1.22)	9.31 (1.26)	9.07 (1.21)
log(HDBGN)	-1.89 (-1.39)	-1.91 (-1.36)	-0.94 (-0.95)	-0.94 (-0.99)
R^2	78.3%	78.4%	98.1%	98.3

홍수위험도에 유의한 예측변인은 해수면 지하수위(회귀계수 부호 양, 높아질수록 홍수 위험도 상승)와 관정지표표고와 낙동강 수위 차이(부호 음, 차이가 낮을수록 홍수 위험도 상승)였다. 두 보 지역의 동일한 예측변인들의 회귀계수 부호는 일치하나 회귀계수의 크기의 차이는 유의하였으므로 보 지역 특성에 따라 단일 예측모형 수립이 불가능함을 알 수 있다. 표준화 회귀계수는 크기는 보에 따라 추정 회귀계수의 크기의 순서가 다르므로 보 지역에 따라 동일 정량변인이라도 홍수 위험도에 대한 상대적 영향도가 다를 수 있다. 함안보 지역은 관정지표표고와 낙동강 수위 차이 변인이 강정보 지역은 지하수 해수면 변인이 다른 변인에 비해 보 지역 홍수 위험도에 더 큰 영향을 미친다. 그리고 그리드 데이터와 포인트

<Table 5> ANOVA Results(Dry Season)

Predictors	Hamanbo		Gangjungbo	
	Point	Grid	Point	Grid
drainage class	◎	◎	◎	◎
effective soil depth	◎	◎		◎
land lay			◎	
surface soil type				
subsoil type			◎	
hydrological soil type	◎			◎
ground coverage				

<Table 6> GLM Results of Point Data(Dry Season)

Independent variables	(Standardized) Beta	
	Hamanbo	Gangjungbo
Intercept	-0.27	-27.6
log(GWL)	1.45(0.99)	9.25(1.26)
log(HDBGN)	-1.60(-1.16)	-0.92(-0.93)
drainage class	0.96(0.46)	0.04(0.04)
effective soil depth	-0.63(-0.25)	
hydrological soil type	-0.41(-0.21)	
R^2	86.1%	98.2%

트 데이터의 결과는 유의한 변수도 동일하고 회귀계수 추정치가 통계적으로 동일하며 결정계수 크기도 같다.

<Table 5>에는 정성변인들에 대한 분산분석 결과를 제시하였다. 유의수준 5% 하에서 유의하게 선택된 변수들에 동그라미 표시를 하였다. <Table 5>에서 볼 수 있듯이 보 지역에 따른 차이와 데이터 유형에 따른 차이 모두 유의하였으므로 지역별, 데이터의 유형별 상이한 모형을 도출하는 것이 적절하다. 그러므로 단일 모형보다는 보 지역 특성에 맞는 홍수위험도 예측모형을 도출하는 것이 필요하다.

회귀분석 결과 유의한 정량변인들과 분산분석 결과 유의한 정성변인들을 고려하여 일반선형모형 추정 및 유의한 변수 선택 결과 그리드 데이터의 경우 함안보 지역에서는 정성변인 모두 유의하지 않게 되어 최종 예측모형은 포인트 데이터를 활용하여 도출하였다. 정성변인들은 수준이 2개인 이진형 변수이므로 더미변수로 일반선형모형을 실시하여 <Table 6>에 정리하였다. 셀의 값은 회귀계수 추정값이고 괄호 안의 값은 표준화 회귀계수 추정값이다.

3.2 호우기 예측모형

호우기에는 포인트 데이터만을 활용하여 최종 모형을

<Table 7> Regression Result(Point Data/Flood Season)

Independent variables	(Standardized) Beta	
	Hamanbo	Gangjungbo
Intercept	-7.93	-
log(GWL)	0.66(0.71)	-
log(HDGN)	0.58(0.31)	-
R^2	89.7%	-

<Table 8> ANOVA Results(Point Data/Flood Season)

Predictors	Hamanbo	Gangjungbo
drainage class		
effective soil depth		
land lay	◎	
surface soil type		◎
subsoil type		
hydrological soil type	◎	
ground coverage	◎	

추정하였다. 회귀분석 결과는 <Table 7>, 분산분석 결과는 <Table 8>에 정리하였다. <Table 7>에서 볼 수 있듯이 강정보의 경우 유의한 정량변인은 검정되지 않았으며 함안보의 경우 해수면지하수위는 갈수기와 동일하게 유의하나(추정 회귀계수 크기는 유의한 차이가 존재) 관정 지표표고와 낙동강 수위 차이 대신 지하수위와 낙동강수위 차이가 유의한 정량변인으로 검정되었다.

<Table 8>에서 볼 수 있듯이 갈수기와 다른 정성변인들이 유의한 변수로 선택되었다. 갈수기와 달리 유의한 변인인 표토토성(지질 지표면을 이루는 흙의 성질), 심토 토성(지질 지표면 밑에 있는 토층의 성질), 토지피복(토지 이용)은 물의 배수와 관련된 변인들이 홍수 위험도에 영향을 미침을 알 수 있다.

강정보의 경우 유의한 정량변인이 없으므로 최종예측 모형은 함안보 지역의 경우만을 도출하여 <Table 9>에 정리하였다. 갈수기의 예측모형과 유의한 변인이 다르고 회귀계수의 크기도 유의적으로 다름을 알 수 있다.

<Table 9> GLM Result(Point Data/Flood Season)

Independent variables	(Standardized) Beta
	Hamanbo
Intercept	-5.88
log(GWL)	0.39(0.43)
log(HDGN)	0.80(0.43)
ground coverage	-0.66(-0.28)
R^2	92.8%

4. 결론

강정보, 함안보 지역 홍수 위험도 GFI 값을 예측하는 모형을 갈수기와 호우기로 나누어 도출하였다. GFI에 영향을 미치는 정량변인은 지하수위와 하천수위 차이와 지하수와 하천 간 거리 등을 고려하였으며 정성변인은 토질 특성과 토지이용실태를 활용하였다. 정성변인들은 사전 일원분산분석을 통하여 GFI에 유의한 변수들만 모형에 고려하였으며 수준은 2개로 하여 이진형 변수로 모형에 삽입하였다.

정성변인과 정량변인을 동시에 고려하면 모형이 복잡해지므로 정량변인들은 회귀모형, 정성변인들은 분산분석을 통하여 유의한 변수들을 선택하고 유의한 변수들만 삽입하여 예측모형을 추정하였다. 예측모형은 보 지역에 따라 다르고 갈수기와 호우기 기간에 따라 유의적으로 상이하였다. 유의한 변인도 상이하고 동일 변인이라도 회귀계수의 크기가 상이하고 GFI에 영향을 미치는 상대적 크기도 상이하였다.

호우기의 강정보 지역 최종 GFI 예측모형은 정량변인이 없어 갈수기와 호우기 함안보 지역 최종 GFI 예측모형 도출 결과를 <Table 10>에 정리하였다. 갈수기나 호우기 기간에 따라 GFI에 영향을 미치는 변인이 상이하며 동일하게 선택된 해수면 지하수위의 회귀계수 크기도 유의적으로 상이하다. 호우기에는 갈수기와 달리 물 배출과 물 유입에 관련된 변인의 영향이 크고 유의하다.

<Table 10> Final Forecasting Model(Hamanbo)

Independent variables	(Standardized) Beta	
	dry season	flood
Intercept	-5.88	-5.88
log(GWL)	1.45(0.99)	0.39(0.43)
log(HDBGN)	-1.60(-1.16)	-
log(HDGN)	-	0.80(0.43)
drainage class	0.96(0.46)	-
effective soil depth	-0.63(-0.25)	-
hydrological soil type	-0.41(-0.21)	-
ground coverage	-	-0.66(-0.28)
R^2	86.1%	92.8%

호우기에 보 지역을 홍수 위험으로부터 안전하게 통제하려면 해당 보 지역에 적합한 호우기 예측모형을 도출해야 한다. 토지피복과 같은 정성변인은 통제될 수 없는 인자이고 해수면 기준 지하수위는 강우량에 의해 결정되므로 실제 통제 가능한 변인은 낙동강 수위이다. 그러므로 보 지역 홍수 위험도를 안전 수준으로 유지하기 위한 최대 낙동강 수위를 추정된 예측모형으로부터 다른 변인들의 수준에 따른 적절한 값을 예측하고 통제해야 할 것이다. 향후 연구에서는 보다 많은 보 지역에 대한 호우기 GFI 예측모형을 도출하고 이를 일반화 하여 국내 적합 단일 예측모형을 도출할 필요가 있다.

Acknowledgement

This study has been partially supported by the 2013 University Research Fund of Hannam University, Daejeon, Korea.

References

- [1] Anderson, T. and Darling, D., Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, 1952, Vol. 23, p 193-212.
- [2] Box, G. and Cox, D.R., An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 1964, Vol. 26, No. 2, p 211-252.
- [3] Kwon, S.H. and Oh, H.S., Construction of Observational Locations for Measuring Water Quality in River Area. *Journal of Society of Korea Industrial and Systems Engineering*, 2012, Vol. 35, No. 3, p 187-191.
- [4] McCullagh, P. and Nelder, J.A., *Generalized Linear Model*, 2nd edition, Chapman and Hall/LRC, 1989.
- [5] Sommer, T. and Ullrich, L., *Influence of Flood Event 2002 on Groundwater. Research Report(in German)*. Environmental Office, City of Dresden, 68p., 2005.
- [6] National Academy of Agricultural Science, Detailed Soil Map, 2013.
- [7] <http://www.riverguide.go.kr>.
- [8] Water, K. *Research Report for Flood Risk at Bo Region*, 2013.
- [9] <http://egis.me.go.kr>.