



Statistical analysis of hazen-williams C and influencing factors in multi-regional water supply system

Kim, Bumjun^a · Kim, Gilho^{b*} · Kim, Hung soo^c

^aKorea Infrastructure Safety and Technology Corporation, Goyang, Korea

^bKorea Institute of Civil Engineering and Building Technology, Goyang, Korea

^cDept. of Civil Engineering, Inha University, Incheon, Korea

Paper number: 16-002

Received: 5 January 2016; Revised: 24 February 2016 / 14 March 2016; Accepted: 14 March 2016

Abstract

In case of the application of Hazen-Williams C for design, operation or maintenance of water supply system, field situations always should be reflected on the factors. In this study, the relationships between C factors and influencing factors are analyzed using statistical techniques with 174 measured C factor data collected in periodic inspection for safety diagnosis in multi-regional water supply systems. To analyze their relationships, cross analysis, one-way ANOVA, correlation analysis were conducted. Analysis results showed that C factors had high correlations with both of elapsed year and pipe diameter and were relatively highly affected by coating material among influencing factors with the categorical type. On the other hand, elapsed year, pipe diameter and water type were meaningful influencing factors according to the results of multiple regression analysis. The Cluster analysis revealed that C factors had a tendency of being fundamentally classified on the basis of the elapsed year of about 20 years and the pipe diameter of 1500mm. Although C factors were generally greatly affected by elapsed year, size of pipe diameter relatively had a large influence on values of them in case of large diameter pipes. Lastly, It can be suggested that C factor estimation formulas using multiple regression analysis and clustering analysis in this study, can be applied as decision standards of C factor in multi-regional water supply systems.

Keywords: clustering analysis, Hazen-Williams C, influencing factor, multi-regional water supply system, multiple regression analysis, statistics techniques

광역상수도 유속계수와 영향인자에 관한 통계적 분석

김범준^a · 김길호^{b*} · 김형수^c

^a한국시설안전공단 시설물정보센터, ^b한국건설기술연구원 수자원·하천연구소, ^c인하대학교 사회인프라공학과

요 지

유속계수는 상수도 설계, 운영, 유지관리 등의 과정에서 항상 현장 실정이 고려된 값이 사용되어야 한다. 본 연구는 광역상수도 정밀안전진단 과정에서 174개 실측자료를 바탕으로 유속계수와 주요 인자 간의 관계를 통계적 기법을 활용하여 분석하였다. 이들 관계를 분석하기 위해서 교차 분석, 일원배치 분산분석, 상관분석 등을 수행하였으며, 그 결과 유속계수와 사용연수 및 관경이 높은 상관관계를 나타내고 여러 범주형 자료 형태 영향인자 중에 내부도장재가 상대적으로 유속계수에 많은 영향을 주는 것으로 검토되었다. 반면 다중회귀분석의 결과에서는 사용연수, 관경 및 수종이 중요한 영향인자인 것으로 검토되었다. 군집분석 결과, 유속계수는 기본적으로 사용연수 약 20년, 관경 1500mm를 기준으로 분류되는 경향이 있었으며, 유속계수는 전반적으로 사용연수에 많은 영향을 받으나 대구경 관에서는 상대적으로 사용연수보다 관경에 많은 영향을 받는 것으로 검토되었다. 마지막으로 본 연구에서는 회귀분석과 군집분석을 사용하여 유속계수 산정식들을 제안하였으며, 이러한 추정식들은 추후 광역상수도의 유속계수 결정 및 사용 시에 판단기준이 될 수 있을 것으로 판단된다.

핵심용어: 유속계수, 영향인자, 광역상수도, 통계적 기법, 다중회귀분석, 군집분석

*Corresponding Author.

E-mail: kgh0518@kict.re.kr (G. Kim)

1. 서론

인류의 시작 때부터 물은 생존에 필수 요건이며, 이를 이용하기 위해 상수도는 인류의 발전과 함께 하였다. 국가기반시설인 상수도는 국토의 동맥 역할을 하는 시설물로 수십 년 전부터 국가 주도적으로 건설되고 있으며, 현재에도 계속 건설 중이다. 더욱이 삶의 질이 향상됨에 따라 물의 사용이 많아짐에 따라, 안정적인 공급이 중요시되고 그 필요성은 더 높아지고 있다. 더불어 수십 년이 지난 기존 상수도 시설물의 유지관리 또한 매우 중요한 상수도 사업의 한 과정으로 여겨지고 있다. 이때, 관로시설의 설계, 시공 및 유지관리에서 매우 필수적인 인자는 관의 마찰계수이며, 우리나라에서는 주로 이마찰계수로 Hazen-Williams 공식의 유속계수 C를 사용한다. Hazen-Williams C 사용빈도에 비해 우리나라에서 이에 관한 연구는 미진한 실정으로, 본 연구에서는 광역상수도 정밀안전진단 시 측정된 유속계수 174개의 자료를 가지고 유속계수의 영향인자 및 유속계수 특성을 분석하였다.

그동안 유속계수와 관련하여 Son (1996)은 국내에서 측정된 유속계수를 분석하여 국외 마찰계수 변화 형태와 유사한 경향을 보이나 관의 통수능이 외국에 비해 빨리 저하된다고 하였으며, 단계적(stepwise) 회귀분석을 통하여 유속계수와 영향인자와의 관계를 분석하였다. 유속계수 추정할 수 있는 산정식은 국외에서는 Mononobe (物部)(한국수자원공사, 1995), Walski, et al. (1988), 국내에서는 Kim, et al. (1996)이 비선형식으로 제안된 바가 있다. 영향인자 검토와 관련하여 통계학적 기법을 활용한 사례는 다양한 분야에 응용되고 있으며, 상수도 분야에서는 Chung, et al. (2003)은 상수도관의 노후도에 영향을 미치는 환경인자들을 선정하여 상관성 분석, 다중회귀분석 등 통계학적 분석을 통하여 영향인자 간의 상관관계를 평가하였다. 또한 Park, et al. (2009)은 상수도 누수위치 좌표를 이용하여 누수위치들 간의 공간적 선형관계에 대한 계층적 군집분석을 수행, 최적 군집분석 방법을 제시하고 분할된 구역들에 대한 상수관망 유지관리의 우선순위를 결정하였다. 그리고 Lee, et al. (2009)에 의해서 다중회귀분석 및 군집분석과 인공신경망을 결합하여 교각 세굴심을 예측하는 연구 사례도 있었다.

이에 본 연구에서는 유속계수 영향인자 검토와 관련한 기존 연구방법 및 결과 등을 참고하여 교차분석, 일원배치 분산분석, 상관분석, 다중회귀분석, 군집분석 등 통계적 기법을 적용하여 광역상수도 유속계수 특성을 파악하였으며, 이때의 자료는 한국시설안전공단에서 실시한 정밀안전진단 과정에서 생산된 것을 사용하였다. 또한, 유속계수를 산정하여 사용할 수 있도록 다중회귀분석을 통한 산정식을 제시하고 향후 실무

에서 광역상수도 관로의 유속계수 결정시 활용될 수 있도록 하였다.

2. 연구방법

2.1 자료수집

1995년 시설물 안전관리에 관한 특별법이 시행된 이후 한국시설안전공단에서는 광역상수도에 대해 정밀안전진단을 실시하고 있으며, 정밀안전진단 시에는 대상 관로에 대해 유속계수를 측정하고 있다. 유속계수 산정은 Hazen-Williams 공식을 이용하며, 여기서 손실수두 Δh 를 구하기 위해 Bernoulli 방정식을 이용한다. 대상관로의 제원과 실측결과를 이용하여 Hazen-Williams 공식의 관경, 유량, 측정거리를 구하며, Δh 는 Bernoulli 방정식에서 동수경사, 즉 압력수두와 위치수두를 활용하여 구한다. 여기서 압력수두는 고정밀도의 수압계를 통해서, 수압계 설치 E.L. (Elevation Level) 차를 나타내는 위치수두는 수준측량 통해서 결정된다.

본 연구는 2000년 이후에 실시된 18개의 광역상수도 정밀안전진단 시 위와 같은 방법으로 고정밀도의 측정장비를 사용하고 신뢰성을 가지는 것으로 판단되는 유속계수 174개의 실측 유속계수를 선정하여 사용하였다. 수집된 유속계수 자료의 분포특성은 다음 Fig. 1과 같으며, 측정된 유속계수의 관경 및 사용년수별 자료 수는 Table 1과 같다. 유속계수 자료의 정적 분포모형을 찾기 위해 정규분포, Gumbel 분포, Weibull 분포에 대해서 Chi-Square (χ^2), Kolmogorov-Smirnov (K-S), Cramer Von Mises (CVM), Probability Plot Correlation Coefficient (PPCC) 4가지 방법의 적합도 검정을 실시하였으며, 실시결과 다음 Fig. 1과 같이 측정된 유속계수는 정규분포와 Weibull 분포를 따르는 것으로 검토되었다.

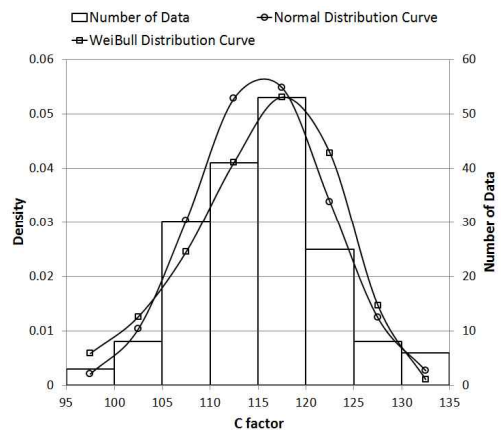


Fig. 1. Characteristics of Measured C factor

Table 1. Pipe diameter and elapsed year of measured C factors

Pipe Diameter (mm)	Number of Obs.	Elapsed Year	Number of Obs.
~ 500	20	~ 5	1
500~1,000	54	5~10	12
1,000~1,500	46	10~15	79
1,500~2,000	36	15~20	34
2,000~	18	20~25	28
		25~30	20

또한, 광역상수도 관로의 영향인자 및 특성을 분석하기 위해 유속계수에 영향을 주는 관로의 일반적 특성과 과거 연구 결과들을 고려하여 유속계수 영향인자로서 관의 내면재질을 나타내는 관종 및 내부도장재, 수질을 나타내는 수종, 그리고 관경과 사용년수를 택하였으며, 측정된 유속계수에 대한 각 인자 자료를 수집하여 분석하였다. 이때, 수질부식성을 나타내는 LI (Langerlier Index)의 자료를 수집·분석해야 하나, 유속계수 측정 당시 LI의 자료를 모두 수집하기 어렵고, 광역상수도의 용수 대부분은 수질부식성을 가지므로 수질을 대표하는 자료로 수종 자료를 수집하였다.

2.2 연구이론 및 방법

2.2.1 교차분석, 일원배치 분산분석, 상관분석

유속계수의 특성을 분석하기 위해서는 무엇보다도 유속계수와 각 영향인자 간의 관계를 파악함이 중요하다. 이에 측정된 유속계수와 영향인자로 선택한 관종, 내부도장재, 수종, 관경, 사용년수 간에 관계를 통계적 기법을 이용하여 분석하고, 변수 간에 상호 연관성 등이 있는지를 확인하였다. 이때, 관종, 내부도장재, 수종의 변수는 범주형 자료이며 관경, 사용년수, 유속계수는 연속형 자료로, 이들 상호 간에 관계를 분석하기 위해서 교차분석, 일원배치 분산분석(Oneway ANOVA), 상관분석 등의 통계적 방법을 활용하였다.

교차분석은 두 범주형 자료를 분석하는데 사용하는 방법으로 관종, 내부도장재, 수종 간에 관계를 파악하였으며, 카이제곱검정(Chi-Square Test)을 실시하여 변수가 변함에 따라 변수의 결과가 변하는지, 즉 독립여부를 파악하였다. 또한, 연속형 변수인 관경, 사용년수, 유속계수 간에 상관성을 파악하기 위해 상관분석을 실시하였다. 두 개 이상의 모집단들의 평균이 서로 동일한지 여부를 검정하고자 할 때 분산분석(ANOVA)이라 하며, 독립변수, 즉 요인이 하나인 것에 대한 분석법을 일원배치 분산분석이라 한다. 일원배치 분산분석을 통하여 유속계수와 관련된 범주형 변수와 연속형 변수와의 관계를 파악하였다.

2.2.2 다중회귀분석

사용년수 및 관경 등의 영향인자가 유속계수에 어느 정도의 영향을 미치며, 여러 조건에서 이러한 영향인자 값들이 어느 정도 정확도의 유속계수를 산정할 수 있는지에 대해 분석하기 위해서 다중회귀분석을 수행하였다. 다중회귀분석은 2개 이상의 독립변수를 사용하여 종속변수 y 를 설명하는 방법으로, 독립변수가 추가된 것을 제외하고 단순회귀분석과 동일하다. 종속변수 y 를 설명하는데 n 개의 독립변수인 x_1, \dots, x_n 을 도입할 때 다중회귀모형은 Eq. (1)과 같이 정의된다. 여기서 y 는 종속변수, β 는 회귀계수, x 는 독립변수, n 은 독립변수 개수, ϵ 은 오차를 의미한다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (1)$$

독립변수를 앞서 수집된 자료, 즉 사용년수, 관경, 관종, 내부도장재, 수종으로 하고, 종속변수를 유속계수로 두고 분석을 수행하였으며, 여기서 명목변수인 관종, 내부도장재, 수종 등은 측정 대상을 고유한 특성에 따라 분류하거나 구분하여 회귀분석에 포함해 분석할 수 있도록 더비변수(dummy variable)로 수치화, 즉 명목척도 데이터로 변환하여 분석하였다.

여러 독립변수를 대상으로 보다 타당성 있고 합리적인 회귀식을 유도하기 위해서는 전진선택법(forward selection method), 후진제거법(backward elimination method), 단계선택법(stepwise selection method) 등을 고려할 수 있는데, 본 연구에서는 모든 독립변수, 즉 측정된 모든 영향인자를 가지고 다중회귀분석을 실시하고, 추가적으로 단계선택법을 사용하여 변수를 조정하여 분석하였다. 단계선택법은 전진선택법과 후진제거법을 개선한 방법으로 독립변수의 추가와 제거를 적절히 조합하여 최선의 회귀식을 도출하는 방법이다. 또한, 도출된 다중회귀식에서 독립변수, 즉 영향인자들의 종속변수인 유속계수에 대한 상대적 영향력을 수치적으로 비교하기 위해서 표준화 회귀계수(standardized regression coefficient)를 참고하였다. 표준화 회귀계수는 회귀모형에서의 각 변수들이 평균이 0이고 표준편차가 1을 갖도록 조정되었을 때의 회귀계수를 말하는데, 해당 독립변수에 의하여 설명되는 종속변수의 비율을 의미한다.

2.2.3 군집분석

유속계수는 영향인자의 특성에 따라 분류될 수 있으므로, 본 연구의 분석결과에 따라 유속계수에 큰 영향을 미치는 인자를 선택하여 군집분석을 실시하고 그 유속계수의 특성을

분석하였다. 군집분석은 많은 객체들을 일정한 속성에 따라 몇 개의 군집으로 분류하여, 같은 군집에 속한 객체들의 유사성과 서로 다른 군집에 속한 객체간의 상이성을 규명하고자 하는 통계분석법이다. 그리고 군집이란 모임을 구성하고 있는 구성원간의 상호유사성 또는 관련성을 중요시하는 개념이다. 본 연구에서는 유속계수에 큰 영향을 미치는 인자인 사용년수와 관경을 기준으로 유속계수를 2~7개의 군집으로 분류하였으며, 분류된 군집별 유속계수의 특성을 분석하였다. 또한 군집별로 다중회귀분석을 실시하여 영향인자와 유속계수와의 관계를 분석하고 유속계수 산정식을 도출하였다.

군집방법은 무수히 많으나 크게 계층적 방법(hierarchical method)과 비계층적 방법(non-hierarchical method)으로 구분된다. 계층적 방법은 사전에 군집수 K를 정하지 않고 단계적으로 서로 다른 군집결과를 제공하는 것이며, 비계층적 방법은 사전에 군집수 K를 정한 후 각 객체를 K개 중의 하나 군집에 배정하는 것이다. 본 연구에서는 K-means 알고리즘을 사용하여 군집분석을 실시하였으며, 비계층적 군집방법 중 가장 널리 사용되는 것으로 K개 군집의 중심좌표를 고려하여 각 객체를 가까운 군집에 배정하는 반복적 알고리즘이다. K는 군집 수 또는 중심점의 수를 의미하며, 사용자에 의해 K개의 중심이 정해지면 각 점들은 가장 가까운 중심점에 할당되어 K개의 군집을 형성한다. 각 군집 내의 유사성 거리는 최대한 가깝게 형성하며 각각의 군집 간의 유사성 거리는 최대한 멀게 형성한다. 본 연구에서는 객체 간의 유사성 거리는 보편적으로 많이 사용되는 유클리드 거리(euclidean distance)를 사용하였으며, 변수값의 단위에 따라 상이한 결과를 초래하고 이러한 편차가 객체간의 거리를 완전히 좌우 때문에 변수값들을 표준화(standardization)하여 유클리드 거리 측정하였다.

3. 영향인자 통계적 분석 및 결과분석

3.1 교차분석, 일원배치 분산분석, 상관분석 결과

교차분석을 통해 범주형 자료인 관중, 내부도장재, 수중간의 관계를 파악하였으며, 카이제곱검정(Chi-Square Test) 결과, 유의확률이 모두 0.000으로 변수 간에 서로 독립, 각 변수에 따라 결과에 차이가 있는 것으로 분석되었으며, 상세 교차표는 Table 2에 나타내었다. 일원배치 분산분석을 통하여 유속계수와 관련된 범주형 변수와 연속형 변수와의 관계를 파악하였으며, Table 3은 범주형 변수와 연속형 변수와의 검정통계량(F)에 대한 유의확률을 나타낸 것으로, 변수 간에 유의확

률이 0.05 보다 작은 경우는 범주형 자료에 따라 연속형 자료 결과가 차이가 발생함을 뜻한다. 마지막으로 연속형 변수인 관경, 사용년수, 유속계수 간에 상관성을 파악하기 위해 상관분석을 실시하였으며, 각 변수 간에 상관계수는 다음 Table 4에 나타내었다. 관경과 사용년수의 상관계수 r은 -0.021로 상관성이 거의 없고, 관경과 유속계수의 상관계수는 0.353, 사용년수와 유속계수의 상관계수는 -0.567로 $0.3 < |r| < 0.7$ 범위에 있어 선형관계에 있는 것으로 분석되었다(Won과 Jung, 2013).

Table 2. Relationship table between variables

Pipe Type	Internal Coating	Water Type			Average	
		Raw	Treated	Industrial		
SP	Coaltar	No.	52	36	1	89
		Diameter	1,618	988	600	1,352
		Elapsed Year	20.71	17.67	29.00	19.57
		C factor	113.44	115.07	96.00	113.90
	Epoxy	No.	19	31	2	52
		Diameter	1934	1244	1350	1500
		Elapsed Year	12	12.9	13.50	12.60
		C factor	119.65	118.05	122.50	118.81
DCIP	Mortar	No.	-	33	-	33
		Diameter	-	511	-	511
		Elapsed Year	-	15.12	-	15.12
		C factor	-	113.83	-	113.83
SP		No.	71	67	3	141
		Diameter	1,703	1,106	1,100	1,406
		Elapsed Year	18.38	15.46	18.67	17.00
		C factor	115.10	116.45	113.67	115.71
DCIP		No.	-	33	-	33
		Diameter	-	511	-	511
		Elapsed Year	-	15.12	-	15.12
		C factor	-	113.83	-	113.83
Total		No.	71	100	3	174
		Diameter	1,703	910	1,100	1,236
		Elapsed Year	18.38	15.35	18.67	16.64
		C factor	115.10	115.58	113.67	115.35

Table 3. F-statistic by oneway ANOVA

Continuous Variable \ Categorical Variable	Pipe Type	Internal Coating	Water Type
Pipe Diameter	0.000	0.000	0.000
Elapsed Year	0.091	0.000	0.002
C factor	0.163	0.000	0.828

Table 4. Correlation coefficient by correlation Analysis

Continuous Variable	Pipe Diameter	Elapsed Year	C factor
Pipe Diameter	1.000	-0.021	0.353
Elapsed Year	-0.021	1	-0.567
C factor	0.353	-0.567	1

Hazen-Williams 공식에서 마찰을 나타내는 유속계수 C와 관경 D는 서로 반비례에 관계에 있으나, 상관분석 결과로는

양의 상관관계에 있다. 공식에서의 반비례 관계는 손실수두 h_f 가 일정했을 때의 관경과 유속계수의 관계이다. 그러나 손실수두 값은 관경이 커지면 작아지고 관경이 작아지면 커지는 특성이 있다. 또한, 공식에서 관경과 유속계수, 손실수두와 유속계수는 반비례한다. 결국 유속계수는 관경의 영향보다는 관경이 변함에 따라 변화하는 손실수두의 영향을 더 많이 받기 때문에 관경과 유속계수 간에는 비례관계가 된다. 이는 Re (레이놀즈 수)와 f (Darcy-Weisbach 공식의 마찰계수)의 관계를 나타내는 Moody 선도나 Walski, et al. (1988) 등의 과거 제시된 제안식에 의해서도 확인할 수 있다.

특히, Table 2는 유속계수와 관련된 변수 간의 관계를 교차분석 및 일원배치 분산분석을 통해 나타난 결과로 상수관로의 일반적 특성이 보이는 부분도 있으며, 그 밖에 다음과 같은 결과를 도출판단할 수 있었다. 상수도관로의 일반적 특성으로 관경이 DCIP의 내부도장재는 모두 콘크리트모르타르로 되어 있고, SP가 DCIP보다 상대적으로 관경이 상당히 크며, 이에 따라 DCIP는 모든 수종이 모두 정수, 송수관로로 쓰이고 있었다. SP의 내부도장재는 콜타르에나멜과 액상에폭시로 되어 있으며, 콜타르에나멜이

액상에폭시보다 많다. 이는 제품화 시기에 따른 것으로 액상에폭시가 콜타르에나멜보다 최근에 사용되었기 때문에 판단된다. SP 내부도장재에 따라 콜타르에나멜은 상대적으로 수종이 원수, 즉 도수관로에 많이 쓰였으며, 반대로 액상에폭시는 송수관로에 많이 사용되었다. 앞서 언급한 바와 같이 SP의 내부도장재의 제품화 시기 때문에 액상에폭시가 상대적으로 콜타르에나멜보다 사용년수는 작고, 유속계수는 큰 것으로 분석되었다. 또한, 수종에 따라 도수관로가 송수관로에 비해 관경과 사용년수가 큰 것으로 나타났으나, 수종이 농업용수, 즉 농업관로는 자료수가 3개로 분석 시 큰 의미가 없는 것으로 판단된다.

결과적으로 범주형 자료인 관경, 내부도장재, 수종 간에는 서로 독립적이지 않고 상관성을 가지고 있으며, 이러한 범주형 자료는 변수에 따라 연속형 자료, 특히 관경과 사용년수 결과에 차이를 발생시키는 것으로 분석되었다. 즉, Table 3의 일원배치 분산분석 결과와 같이 관경은 관경, 내부도장재 및 수종, 사용년수는 내부도장재 및 수종과 상호 연관성이 있다는 것이다.

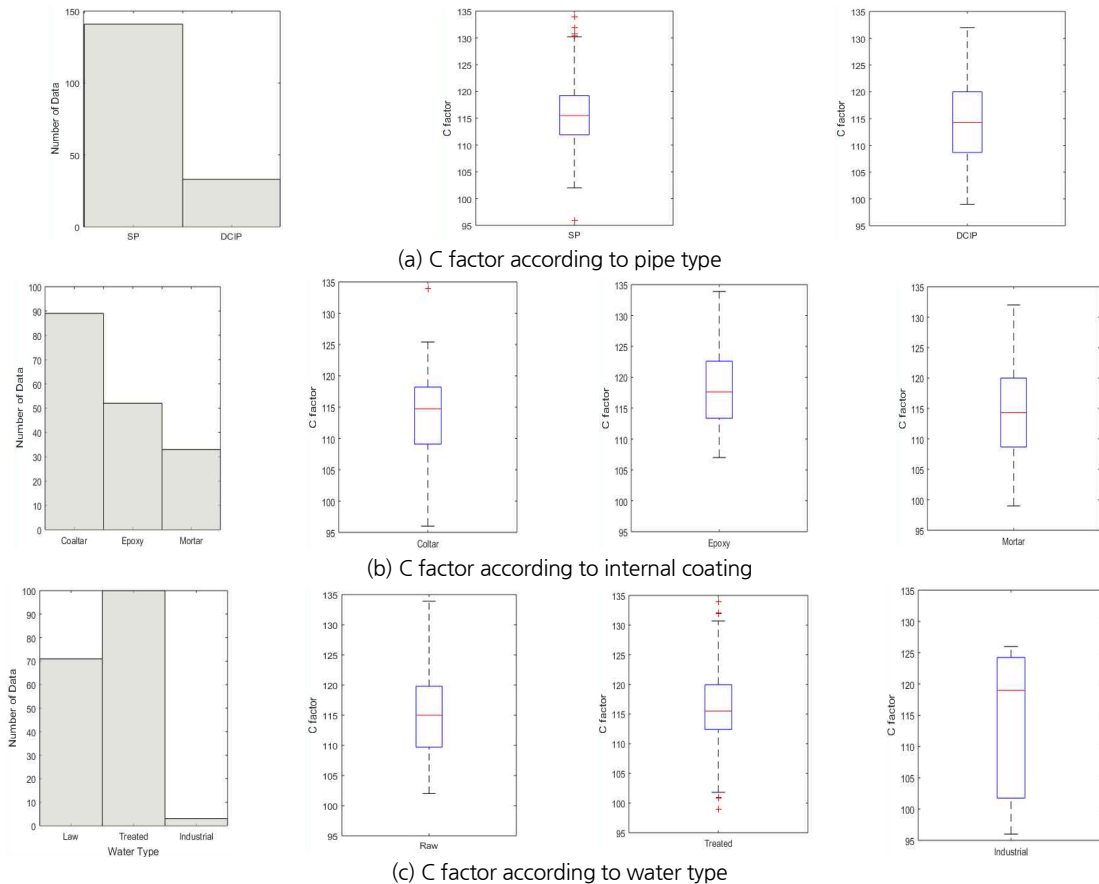


Fig. 2. C factor according to categorical variables

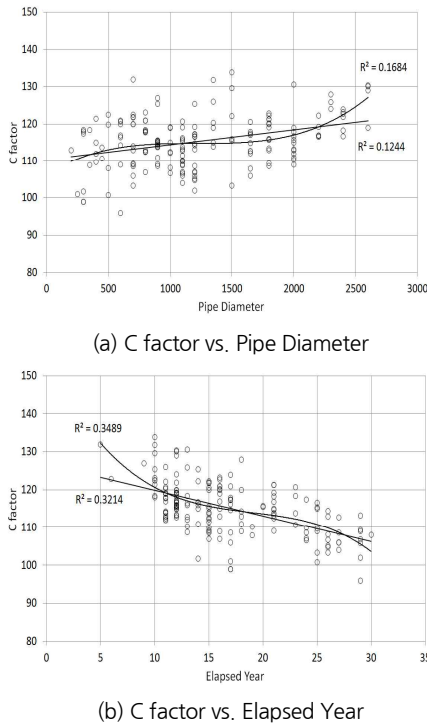


Fig. 3. Relationship elapsed year, Pipe Diameter and C factor

앞서 유속계수와 관련된 모든 변수 간에 관계에 대해 통계적으로 분석하고 결과를 도출하였으나, 무엇보다도 중요한 것은 영향인자와 유속계수와 관계로, 그 결과를 앞의 Table 3과 Table 4에 정리하였다. 앞의 결과와 Fig. 2와 같이 범주형 변수 중 내부도장재만 그 종류에 따라 유속계수에 결과에 영향을 주는 것으로 분석되었다. 사용년수 및 관경과 유속계수는 앞의 Table 4의 상관관계 결과와 다음 Fig 3과 같이 선형관계가 있다. 이는 관의 유체와의 접촉면적, 시간에 따른 조도, 마찰 증가의 원인과 앞의 일원배치 분산분석 결과와 같이 범주형 자료와 연속형 변수인 관경 및 사용년수와의 상호 연관성 등의 복잡한 영향인 것으로 판단된다.

3.2 다중회귀분석 결과

정밀안전진단 시 측정된 174개의 유속계수 자료와 관종 등의 영향인자를 이용하여 다중회귀분석을 실시하고, 그 결과를 활용하여 유속계수의 특성 분석 및 산정식을 도출하였다. 영향인자인 사용년수, 관경, 관종, 내부도장재, 수종을 독립변수로, 측정된 유속계수를 종속변수로 두고, 범주형 변수인 관종, 내부도장재, 수종은 더비변수로 수치화하여 분석하였으며, 그 결과인 다중회귀모형은 Eq. (2)와 같다. 여기서, Y는 사용년수(year), D는 관경(mm), PT

는 관종(SP: 1, DCIP: 2), IC는 내부도장재(폴타르에나멜: 1, 액상에폭시: 2, 콘크리트모르타르), WT는 수종(원수: 1, 정수: 2, 농업용수: 3)을 의미한다.

$$C = -0.658 Y + 0.006 D + 2.545 PT - 1.167 IC + 2.633 WT + 114.063 \quad (2)$$

다중회귀모형이 의미가 있는지의 판단, 즉 종속변수인 유속계수를 설명하는데 5개의 독립변수 중 적어도 하나의 변수는 유의한지 확인(Kim, et al., 2009)하는 유의성 검증 결과, 유의확률은 0.000으로 위 모형의 귀무가설을 기각하는 것으로 검토되었다. 또한, 앞서 통계적 분석 결과에서 독립변수 간에 서로 상관관계가 있는 것으로 분석되어, 산정된 위 회귀계수에서 큰 오차를 포함할 우려가 있으므로 다중공선성이 있는지를 검토하였다. 분산팽창계수 VIF(Variance Inflation Factor)가 5~10 이상이면 다중공선성이 있는 것으로 판단하는데, 본 다중회귀모형의 VIF는 여러 독립변수 중 관경이 최대 4.772로 다중공선성이 없는 것으로 검토되었다.

각 독립변수, 즉 영향인자가 다중회귀모형에 포함될 정도의 의미가 있는지를 확인하기 위해서 각 독립변수 회귀계수에 대한 검증을 수행하였으며, 그 결과인 유의확률(P-value)은 Table 5와 같다. 사용년수, 관경 및 수종은 유의확률이 0.05 이하로 의미 있는 독립변수이며, 관종 및 내부도장재는 0.05 이상으로 위 다중회귀모형에 포함될 정도의 의미가 있지 않은 것으로 검토되었다. 또한, 종속변수인 유속계수에 영향인자인 독립변수의 상대적 영향력을 비교하기 위해서 다음 Table 5와 같이 표준화 회귀계수를 분석하였다. 위 독립변수 회귀계수에 대한 유의확률의 결과와 같이 사용년수와 관경이 유속계수에 큰 영향을 미치는 것으로 나타났으며, 관종, 내부도장재, 수종은 유속계수에 많은 영향을 미치지 못하는 것으로 검토되었다.

Table 5. P-value of regression coefficient and standardized regression coefficient

Independent Variable	P-value of Regression Coefficient	Standardized Regression Coefficient
Elapsed Year	0.000	-0.546
Pipe Diameter	0.000	0.491
Pipe Type	0.243	0.144
Internal Coating	0.290	-0.130
Water Type	0.011	0.199

앞서 일원배치 분석결과에서 범주형 영향인자 중 내부도장재만 종류에 따라 유속계수 결과에 영향을 주는 것으

로 분석되었으나, 다중회귀분석에서는 독립변수 회귀계수에 대한 유의확률과 표준화 회귀계수를 비교한 결과, 수종이 타 범주형 자료보다 유속계수에 영향을 미치는 것으로 분석되었다. 이는 일원배치 분석은 내부도장재와 유속계수만 분석하여 서로간 상관관계를 나타내지만, 다중회귀분석은 여러 영향인자와 유속계수를 분석한 결과로 여러 영향인자 중 유속계수 결과에 영향을 주는 정도를 분석하기 때문에 서로 다를 수 있다. 또한 내부도장재는 앞서 설명한 바와 같이 제품화 시기가 달라 사용년수와 관련성이 있으며, 다중회귀분석에서는 내부도장재의 이러한 특성이 사용년수에 반영되었기 때문으로 판단된다.

앞서 연속형 변수인 사용년수 및 관경과 유속계수의 상관분석 결과와 위의 다중회귀분석 결과에 따라 사용년수와 관경은 유속계수에 가장 큰 영향을 미친다. 사용년수 및 관경과 유속계수의 관계는 앞의 Fig. 3에서 확인할 수 있다. 사용년수를 독립변수로 한 단순회귀분석 결과 결정계수 R^2 이 0.321, 관경을 독립변수로 한 단순회귀분석 결과 결정계수 R^2 이 0.124로, 상관분석 결과와 같이 사용년수가 관경보다 유속계수 결과에 영향을 많이 미치는 것으로 검토되었다. 여러 단순 비선형 회귀곡선 중 3차식이 가장 사용년수와 관경을 잘 표현하는 것으로 검토되었으며, 이는 사용년수가 상대적으로 작아지거나 관경이 커질수록 유속계수가 크게 증가하고 반대로 사용년수가 상대적으로 크거나 관경이 작아질 경우 유속계수가 크게 감소함을 나타낸다. 그러나 결정계수 R^2 이 선형 회귀분석 결과와 큰 차이가 없는 것으로 검토되어, 회귀모

형의 신뢰성 및 단순성을 고려하여 본 연구에서 선형회귀 분석을 수행하였다.

5개의 영향인자로부터 변수 선택법 중 단계선택법을 사용하여 독립변수를 선택한 결과, 위 독립변수의 회귀계수에 대한 유의확률 결과와 동일하게 사용년수, 관경, 수종이 독립변수로 채택되었다. Son (1996)이 같은 단계선택법을 통해서 유속계수와 관경사이에 상관관계가 없다는 결과와 다른 결과가 도출되었는데, 이는 수집된 유속계수와 영향인자 등의 자료 상의 차이 것으로 판단된다. 본 연구에서 단계선택법에 선택된 독립변수(사용년수: Y (year), 관경: D (mm), 수종: WT), 더하여 유속계수에 가장 영향을 많이 미치는 사용년수와 관경만을 독립변수로 하여 다중회귀분석을 실시한 결과는 다음 Eq. (3)과 Eq. (4)와 같다.

$$C = -0.624Y + 0.005D + 2.368WT + 115.557 \quad (3)$$

$$C = -0.675Y + 0.004D + 121.770 \quad (4)$$

앞서 여러 경우의 독립변수를 가지고 다중회귀분석을 실시하여 유속계수와 영향인자간의 관계 및 특성을 분석하고 다중회귀모형을 제시하였다. 이러한 다중회귀모형을 평가하고 모형을 통한 유속계수 산정 결과의 정확도를 파악하기 위해서는 측정된 유속계수와 다중회귀모형에 의해 계산된 유속계수의 비교가 필요하다. 일반적으로 결정계수 R^2 을 사용하여 모형을 평가하며, 그 결과는 다음 Table 6과 같고 측정된 유속계수와 계산된 유속계수간의 결과 그래프는 다음 Fig. 4와 같다.

Table 6. Determination coefficient of regression model by independent variables

	All Variables	Year, Diameter	Year, Diameter, Pipe Type
R^2	0.460	0.438	0.438
	Year, Diameter, Internal Coating	Year, Diameter, Water Type	
R^2	0.438	0.456	

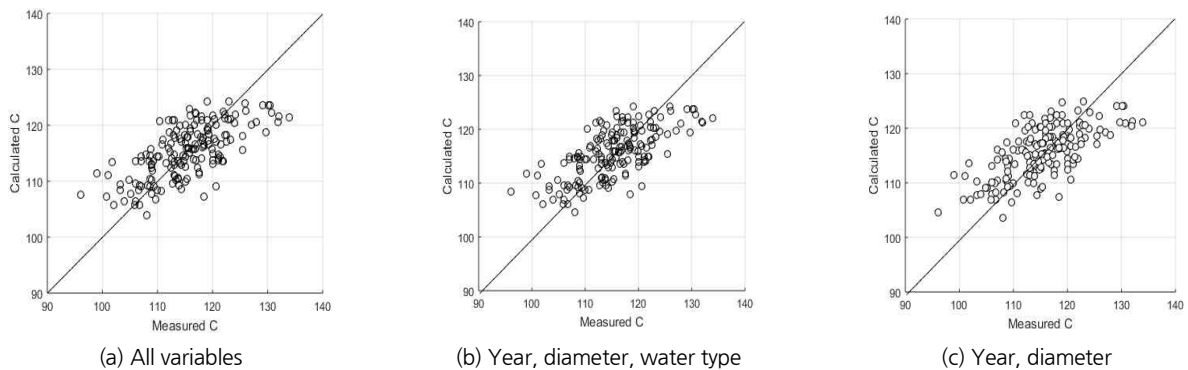


Fig. 4. Regression analysis results by independent variables

독립변수에 따른 다중회귀모델 평가 결과, 앞서 분석한 독립변수 회귀계수에 대한 유의확률과 표준화 회귀계수 비교 결과와 유사하게 경과년수, 관경, 수중을 독립변수로 하였을 때 측정된 유속계수와 가장 유사, 즉 산정 정확도가 뛰어난 것으로 나타났으며, 관중과 내부도장재는 평가결과인 R²에 거의 영향이 없는 것으로 분석되었다. 물론 모든 독립변수를 택하였을 때, R²이 가장 크게 나왔지만, 이는 독립변수가 많을수록 R²이 크게 나오는 특성이 있기 때문인 것으로 판단된다. 단순히 사용년수와 관경만의 독립변수로 한 다중회귀모델도 경과년수, 관경, 수중을 독립변수로 한 다중회귀모델의 결과와 큰 차이 없는 것으로 분석되었다.

이러한 다중회귀모델은 상수도 관로를 설계할 때 유속계수 결정에 사용될 수 있으며, 사용 중인 관망의 수리검토 시 유속계수의 초기값 또는 여러 사유로 산정된 유속계수의 비교·검토 기준으로 사용될 수 있다. 다음 Fig. 5는 유속계수 사용 시 활용키 위해 사용년수와 관경에 따른 유속계수를 나타낸 것으로 Eq. (4)을 그래프로 표현한 것이다. 회귀모형의 신뢰성 및 단순성을 고려하여 Eq. (4)을 사용하였으며, 설계 시 목표년수에 따른 관경별 유속계수를 판단할 수 있다. 본 다중회귀모델은 현장 측정 자료를 바탕으로 만들어진 모델로 현재 사용되고 있는 유속계수를 대변하므로, 설계 시 사용할 경우에는 현장상황이나 통수능력 등의 안전율을 고려하여 모델에서 계산된 유속계수 값보다 작은 값을 사용하는 것이 바람직하다.

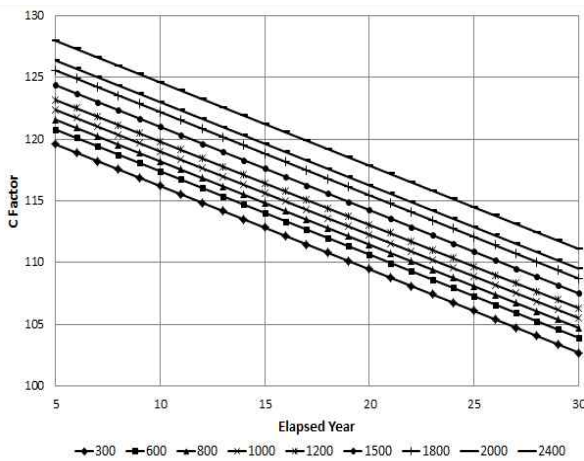


Fig. 5. C factor prediction by equation (4)

3.3 군집분석 결과

통계적 분석을 통해 유속계수에 가장 큰 영향을 미치

는 인자는 사용년수와 관경임을 확인하였고, 이 변수들을 사용하여 유속계수를 여러 개의 군집으로 분류하고, 그 군집별 특성 및 유속계수의 특성을 분석하였다. 군집분석은 가장 널리 사용되는 비계층적 방법 중 K-means 알고리즘을 사용하였고, 우선 비계층적 방법에서 중요한 군집수를 결정하기 위해 군집중심 간의 거리의 합을 아래 Fig. 6과 같이 결정하였다. 군집이 많을수록 이 거리의 합은 감소하기 때문에 본 연구에서는 거리의 합이 급속히 감소하기 시작하는 군집을 고려하여 군집수를 선택하였다. 또한, 자료수가 174개로 너무 많은 군집수는 각 군집의 데이터 수를 과도하게 작게 하는 경향이 있으므로 최종적으로 유속계수를 3개에서 5개의 군집으로 구분하여 분석하였다.

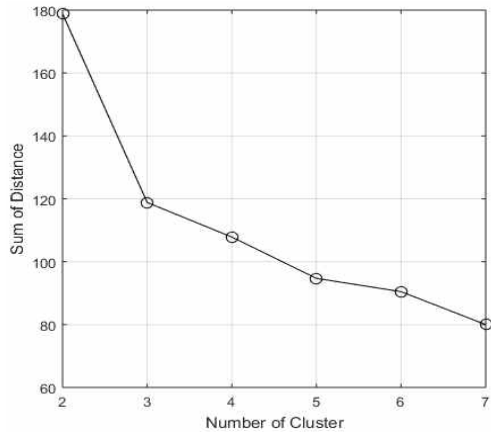


Fig. 6. Sum of distance by number of cluster

앞서 언급한 군집수의 고려 없이 우선 2개에서 7개의 군집으로 분류하여 분석하였으며, 그 결과는 Fig. 7과 같다. 군집수마다 결과 차이가 발생하고 있으나, 크게 사용년수 약 20년과 관경 약 1,500mm를 기본으로 분류되는 경향이 있다. 특히 4군집 이후에는 20년 이하의 유속계수만 더 세분화되어 분류되는 것으로 분석되었다. Fig. 7과 같이 3개의 군집은 사용년수 약 20년을 기준으로 20년 이상 1개 군집, 20년 미만은 관경에 따라 2개 군집으로 분류되었으며, 3개의 군집에서 사용년수 20년 이상의 1개의 군집(Cluster 2)이 2개의 군집(Cluster 2, Cluster 3)으로 나뉘어 4개의 군집이 형성되었다. 마지막으로 4개의 군집에서 사용년수 약 20년 이하, 관경 약 1,500mm 이하의 군집(Cluster 1)이 사용년수에 따라 세분화되어 2개의 군집(Cluster 3, Cluster 4)으로 나뉘어 5개의 군집이 형성되었다.

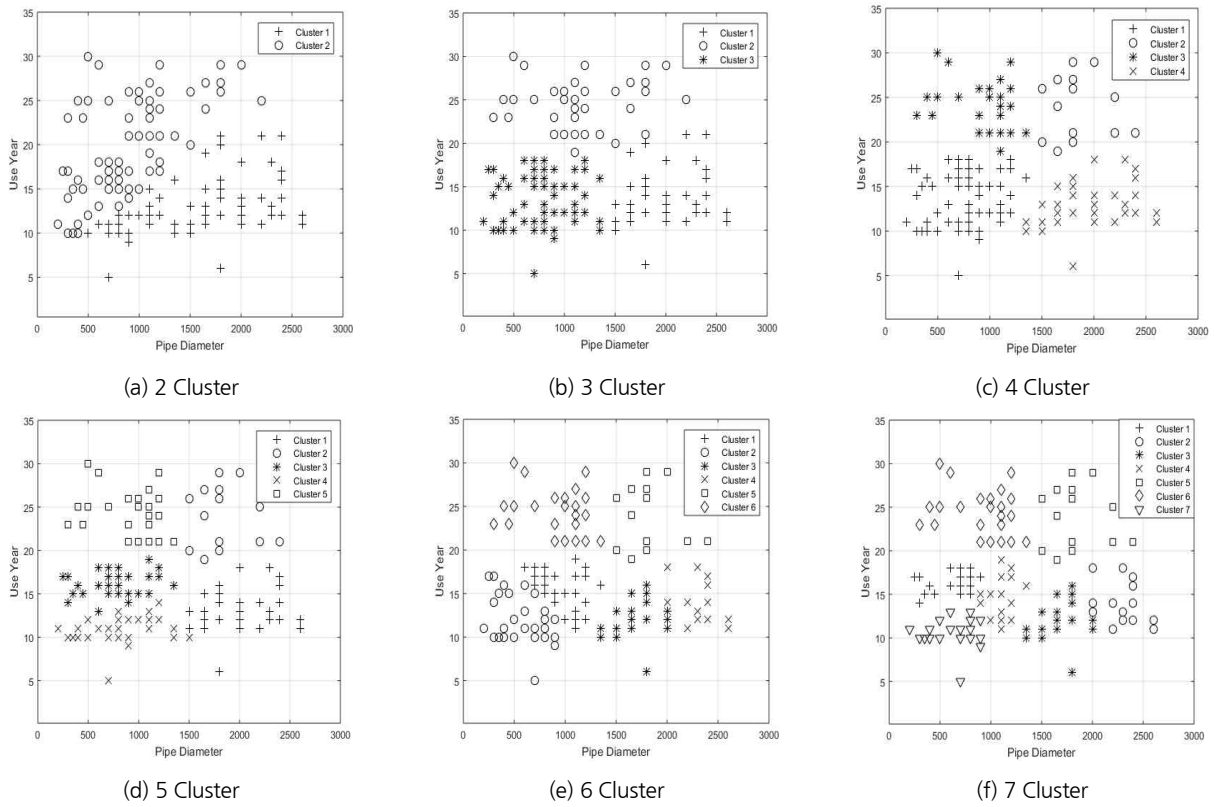


Fig. 7. C Factor classification results by number of cluster

Table 7. C factor characteristics by classified clusters

Number of Cluster	No	Elapsed Year		Pipe Diameter		C Factor (Average)	
		Range	Average	Range	Average		
3 Cluster	Cluster 1	47	6~21	13.57	1,500~2,600	2,005	119.62
	Cluster 2	47	19~30	24.79	300~2,200	1,197	110.37
	Cluster 3	80	5~18	13.66	200~1,350	808	115.77
4 Cluster	Cluster 1	76	5~18	13.82	200~1,350	780	115.39
	Cluster 2	18	19~29	24.06	1,500~2,400	1,856	113.67
	Cluster 3	34	19~30	24.53	300~1,350	974	109.40
5 Cluster	Cluster 4	46	6~18	12.59	1,350~2,600	1,943	120.36
	Cluster 1	40	6~18	12.90	1,500~2,600	2,025	119.51
	Cluster 2	18	19~29	24.06	1,500~2,400	1,856	113.67
	Cluster 3	43	13~19	16.00	250~1,350	753	112.99
	Cluster 4	40	5~14	11.10	200~1,500	909	119.43
	Cluster 5	33	21~30	24.70	300~1,350	970	109.38

분류된 군집별로 사용년수, 관경 및 유속계수 군집 특성을 분석한 결과, 앞의 상관분석의 결과와 동일하게 사용년수가 유사할 경우 관경이 클수록 유속계수가 커지며, 관경이 유사할 경우 사용년수가 작을수록 유속계수가 커지는 것으로 분석되었으며, 그 결과는 다음 Table 7과 같다. 예를 들어 4군집의 경우, 사용년수가 약 24년으로 유사한 Cluster 2와 Cluster 3, 그리고 약 13년으로 유사한 Cluster 1과 Cluster 4에서 보는 바와 같이 관경이 클수록

유속계수가 커짐을 확인할 수 있다. 또한, 관경이 약 1,900mm로 유사한 Cluster 2와 Cluster 4에서 보는 바와 같이 사용년수가 클수록 유속계수가 작아짐을 확인할 수 있다. Table 7에서의 값들은 군집분석에 의해 형성된 각각의 군집에 속한 경과년수, 관경 및 유속계수의 범위 및 평균값을 나타내며, Fig. 8은 4군집에서 각 군집의 평균을 분석한 결과를 나타낸다.

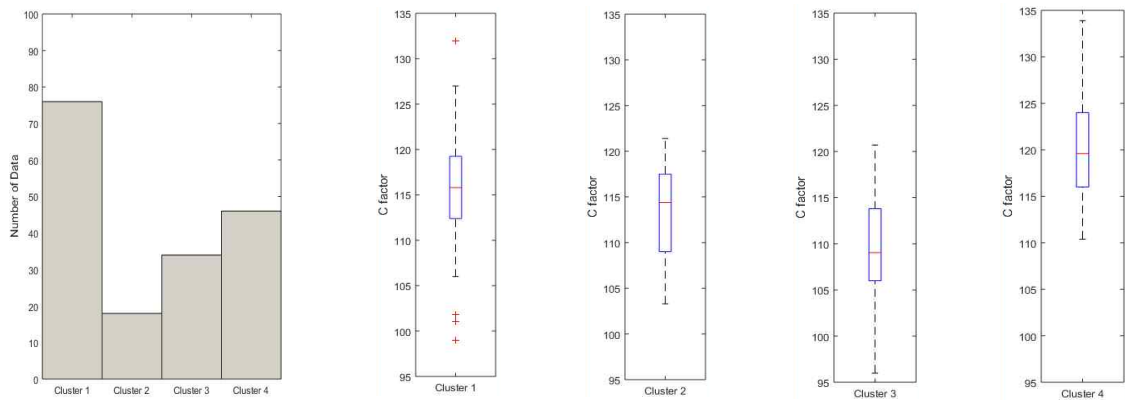


Fig. 8. C factor analysis results of 4 cluster

각 분류된 군집별로 사용년수와 관경의 인자가 유속계수에 어느 정도의 영향을 미치는지 알아보기 위해서 군집별로 다중회귀분석을 실시하고 표준화 회귀계수를 구하여 상대적 영향력을 비교하였다. 결과를 쉽게 알아볼 수 있도록 사용년수와 관경이 각 2개씩 구분되는 4군집에 대한 표준화 회귀계수를 구하였으며, 그 결과는 다음 Table 8과 같다.

Table 8. Regression formular and standardized regression coefficient of 4 cluster

Cluster	Elapsed Year	Pipe Diameter	Regression Formula	Standardized Regression Coefficient	
				D	Y
Cluster 1	13.82	780	$C=0.005D-1.109Y+126.915$	0.222	-0.492
Cluster 2	24.06	1,856	$C=0.008D-0.792Y+117.155$	0.407	-0.527
Cluster 3	24.53	974	$C=0.000D-1.133Y+137.047$	0.008	-0.606
Cluster 4	12.59	1,943	$C=0.004D-0.479Y+118.612$	0.242	-0.177

Cluster 2, 3과 같이 사용년수가 긴 군집의 유속계수는 관경보다 사용년수의 영향을 많이 받고, 특히 관경이 작은 Cluster 3은 사용년수의 영향이 큰 것으로 분석되었다. Cluster 1과 같이 사용년수와 관경이 작은 군집의 유속계수는 사용년수와 관경에 비슷한 영향을 받으나 Cluster 4와 같이 사용년수가 짧고 관경이 큰 경우는 관경의 영향을 많이 받는 것으로 분석되었다.

앞의 다중회귀분석의 결과와 동일하게 전반적으로 사용년수가 유속계수에 영향을 많이 받으나 관경이 커짐에 따라 관경의 영향정도가 커지는 것으로, 관경이 작아짐에 따라 사용년수의 영향정도가 커지는 것으로 검토되었다. 이는 과거 Son (1996)이 관경이 커질수록 유속계수와 관경

이 무관해진다고 한 결과와 상이한 결과이다.

앞서 다중회귀분석의 결과에서 사용년수와 관경을 독립변수로 하는 다중회귀모델은 신뢰성을 가지는 것으로 판단되어 본 절에서는 비슷한 특성을 가지는 군집별로 다중회귀분석을 수행하여 유속계수의 산정식을 도출하였으며, 각각의 군집별 다중회귀모델은 다음 Table 9와 같다.

Table 9. Regression formula by number of cluster

Number of Cluster		Regression Formula
2 Cluster	Cluster 1	$C=0.003D-0.668Y+122.269$
	Cluster 2	$C=0.006D-0.770Y+122.226$
3 Cluster	Cluster 1	$C=0.007D-0.501Y+112.585$
	Cluster 2	$C=0.005D-1.144Y+133.316$
	Cluster 3	$C=0.005D-1.147Y+127.175$
4 Cluster	Cluster 1	$C=0.005D-1.109Y+126.915$
	Cluster 2	$C=0.008D-0.792Y+117.155$
	Cluster 3	$C=0.000D-1.133Y+137.047$
	Cluster 4	$C=0.004D-0.479Y+118.612$
5 Cluster	Cluster 1	$C=0.009D-0.043Y+101.299$
	Cluster 2	$C=0.008D-0.729Y+117.155$
	Cluster 3	$C=0.007D-0.353Y+113.205$
	Cluster 4	$C=0.008D-2.744Y+142.831$
	Cluster 5	$C=0.000D-1.267Y+140.341$

구해진 다중회귀모델을 평가하고 모델에 의한 산정 결과의 정확도를 파악하기 위해서, 결정계수를 통하여 군집별 모델을 평가하였다. 그 결과는 Table 10과 같고 유속계수와 계산된 유속계수 간의 결과 그래프는 다음 Fig. 9와 같다. 앞서 군집분류 없이 사용년수와 관경을 독립변수로 하는 다중회귀모델의 R²값은 0.438로 Table 10에서 보는 바와 같이 군집을 2개로 분류한 결과와 같으며, 3개 이상으로 분류하여야 좀 더 개선된 결과로 산정될 수 있는 것으로 검토되었다. 그리고 3개와 4개의 군집분류를 통한 산정 결과는 유사하며, 5개의 군집분류를 통한 산정이 좀 더 정확한 결과를 얻을 수 있었다.

Table 10. Determination coefficient of regression model by number of cluster

Number of Cluster	2 Cluster	3 Cluster	4 Cluster	5 Cluster
R ²	0.438	0.479	0.468	0.556

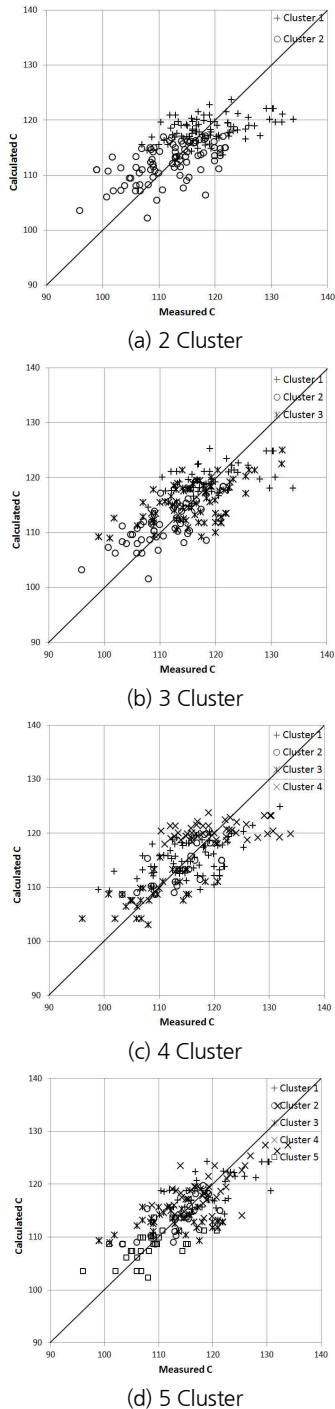


Fig. 9. Regression analysis results by number of cluster

군집분류 후 분류된 군집에서 다중회귀분석을 수행하는 과정은 자료의 수에 많은 영향을 받는다. 군집수를 많이 나눌수록 좀 더 좋은 결과가 도출되지만, 금회 분석에서는 174개의 제안된 자료수로 인해 너무 많은 군집수는 각 군집의 데이터 수를 과도하게 작게 하는 경향이 있고 모델의 유의성 등에 문제가 발생하기 때문에 최대 5개의 군집으로 제한하였다. 위와 같이 통계적 분석은 자료가 많을수록 신뢰성 있는 결과는 도출할 수 있으므로, 계속해서 유속계수 실측자료를 수집하여 분석함이 필요하다.

4. 결론

본 연구에서는 한국시설안전공단에서 광역상수도 정밀안전진단 시 측정된 신뢰성 있는 자료 174개의 자료를 활용, 통계적 분석을 실시하고 유속계수의 영향인자 및 특성을 분석하여, 향후 광역상수관로의 유속계수를 결정함에 있어서 활용될 수 있도록 하였다. 이러한 본 연구의 결론을 정리하면 다음과 같다.

1) 유속계수와 영향인자 간의 분석 결과, 범주형 인자 중 내부도장재가 유속계수의 결과에 영향을 미치며 연속형 인자인 사용연수($r=-0.567$)와 관경($r=0.353$)은 유속계수와 상관관계가 있는 것으로 분석되었다.

2) 다중회귀분석 중 각 독립변수 회귀계수에 대한 유의확률을 분석한 결과, 사용연수, 관경 및 수중은 의미 있는 독립변수로 분석되었지만 관중 및 내부도장재는 다중회귀모형에 포함될 정도의 의미가 있지 않은 것으로 검토되었다. 또한 표준화 회귀계수 검토 결과, 사용연수와 관경은 유속계수에 큰 영향을 미쳤으며 관중, 내부도장재 및 수중은 유속계수에 큰 영향을 미치지 못하는 것으로 검토되었다.

3) 영향인자에 따른 다중회귀모델 평가 결과, 경과연수, 관경, 수중을 독립변수로 하였을 때 측정된 유속계수와 가장 유사한 것으로 검토되었다. 또한, 제시된 다중회귀모델은 현재 사용되고 있는 유속계수를 대변하므로, 설계시 사용할 경우에는 안전율을 고려하여 모델에서 계산된 유속계수 값보다 작은 값을 사용토록 한다.

4) 유속계수의 군집분석 결과 군집수마다 결과 차이가 발생하고 있으나 크기는 사용연수 약 20년과 관경 약 1,500mm를 기본으로 분류되는 경향이 있었으며, 분류된 군집의 특성을 분석한 결과 사용연수가 유사할 경우 관경이 클수록 유속계수가 커지며, 관경이 유사할 경우 사용연수가 작을수록 유속계수가 커지는 것으로 분석되었다.

5) 분류된 군집에 대해서 다중회귀분석을 실시한 결과, 전반적으로 유속계수는 사용년수에 영향을 많이 받으나, 관경이 커짐에 따라 사용년수보다 관경이 유속계수에 영향을 많이 미치고, 관경이 작아짐에 따라 관경보다는 사용년수가 유속계수에 영향을 많이 미치는 것으로 분석되었다.

유속계수 측정자료를 통계적으로 분석하여 도출된 결과는 추후 유속계수 사용에 있어 유용한 판단기준이 될 수 있을 것으로 판단된다. 마지막으로 좀 더 신뢰성 있는 결과 도출을 위해서는 무엇보다도 지속적인 유속계수 측정 및 자료수집을 통해 신뢰성 있는 자료를 확보하는 것이 필요하다.

References

- Chung, W. S., Lee, H. D., Kim, I. T., Yu, M. J. (2003). Correlation Analysis of Environmental Factor for Drinking Water Pipe Deterioration, Journal of the Korean Society of Water and Wastewater, Vol. 17, No. 2, pp. 242-249.
- Jeon, C. H. (2012). Data Mining Technique and Application, Hannarae Academy, pp 64.
- Kim, J. K., Park, J. H., Park, H. J., Lee, J. J., Jeon, H. S., Hwang, J. S. (2009). Statistics for Engineer, Free Academy, pp. 247.
- Kim, J. H., Kim, Z. W., Lee, H. D., Kim, S. H. (1996). Development of Rehabilitation and Management Techniques for Old Water Distribution System, Journal of Korea Water Resources Association, Vol. 29, No. 3, pp. 197-205.
- Lee, C. H., Ahn, J. H., Lee, J. H., Kim, T. W. (2009). Prediction of Scour Depth Using Incorporation of Cluster Analysis into Artificial Neural Networks, Journal of Korean Society of Civil Engineers, Vol. 29, No. 2B, pp. 111-120.
- Lee, H. D., K-water (1995). Development of Decision-Support System for Water Pipeline Rehabilitations.
- Park, S. W., Im, K. C., Choi, C. L., Kim, K. L. (2009). Hierarchical Clustering Analysis of Water Main Leak Location Data, Journal of Korea Water Resources Association Vol. 42, No. 3, pp. 177-190.
- Son, K. I. (1996). Predicting Flow Resistance Coefficients in Water Supply Mains, Journal of Korea Water Resources Association Vol. 29, No. 4, pp. 223-231.
- Walski, T. M., Sharp, W. W., Shields, F. D. (1988). Prediction Internal Roughness in Water Mains, Miscellaneous Paper EL-88-2, US Army Engineering Waterways Experiment Station.
- Won, T. Y., Jung, S. W. (2013). SPSS PASW Statistics 18.0 Statistical Research Analysis, Hannarae Academy, pp. 339.