인천지역 사업장 방류수 중 수질오염물질 배출 특성 연구 A Study on Characteristics of Pollutants from the Industrial Wastewater Effluent in Incheon Area

이진숙**** · 여인환* · 박홍경* · 이효정* · 김현주* ·
곽완순* · 이창근* · 이성모* · 이진호** · 동종인***,†

Jin Sook Lee**** · In Hwan Yeo* · Hong Gyeong Park* · Hyo Jeong Lee* · Hyun Ju Kim* ·
Wan Soon Kwack* · Chang Keun Lee* · Sung Mo Lee* · Jin Hyo Lee** · Jong In Dong***,†

*인천광역시 보건환경연구원 · **서울특별시 보건환경연구원 · ***서울시립대학교 환경공학과

**Incheon Metropolitan City Institute of Health and Environment*

***Seoul Metropolitan Government Research Institute of Public Health and Environment*

****Department of Environmental Engineering, University of Seoul*

(Received October 20, 2018; Revised October 30, 2018; Accepted November 8, 2018)

Abstract : To enhance supervision for discharging of water pollutants, the Ministry of Environment has been expanding the number of water pollutants (51 species) and specific hazardous substances (29 species) in legal regulations annually. The present discharging state of water pollutants and specific hazardous substances was investigated and data mining technique as well as statistical analysis were conducted to draw discharging characteristics of water pollutants from the industrial wastewater effluent in incheon area by the industrial category. Upon investigation, 1,1-dichloroethylene, benzene, carbon-tetrachloride, vinylchloride, acrylonitrile, PCP, organic phosphorus, 1,2-dichloromethane, PCE had not ever detected or were detected only once within a investigation period. The 95% tiles of both 1,4-dioxane and CN concentration were 3.409 mg/L and 0.465 mg/L respectively and these values were in less than effluent quality standard for others except clean area. As a result of PCA, principal components of water pollution by industrial wastewater effluent in incheon area consist of PC1, which is associated with both Ni and Zn such as plating facilities, PC2, which is associated with Fe such as metal processing facilities, PC3, which is associated with Cu such as copper rolling and pressing processing facilities, PC4, which is associated with Cr such as Cr discharging facilities. As a result of classification for the industrial category using by SVM, LR, ANN techniques, ANN is the most accurate of algorithms and classification possibility means probability for grasping discharging characteristics of water pollutants from the industrial wastewater effluent in incheon area by the industrial category.

Key Words : *Industrial Wastewater Effluent, Effluent Pollutants, Specific Hazardous Substances, Principal Component Analysis, Data Mining, Classification*

요약 : 환경부는 현재 배출허용기준 적용 수질오염물질 관리대상 항목은 51종, 폐수배출시설 적용기준이 있는 특정수질유해 물질 항목은 29종으로 해마다 관리항목을 점차 확대하고 있으며 새로운 수질오염물질에 대한 관리를 강화하고 있다. 본 연구에서는 인천지역 사업장 방류수에 대해 업종별로 수질오염물질 및 특정수질유해물질을 조사하여 배출현황을 파악하고 기존의 통계학적 방법 이외에 데이터 마이닝 기법을 이용하여 검출된 오염물질에 대해 업종별 배출 특성을 도출하고자 하였다. 특정수질유해물질 조사 결과, 1,1-dichloroethylene, benzene, carbon-tetrachloride, vinylchloride, acrylonitrile, PCP, 유기인, 1,2-dichloromethane, PCE 등은 조사기간 중 불검출 및 단 1회만 검출되었고 1,4-dioxane의 95%tile 농도값은 3.409 mg/L, CN의 95%tile 농도값은 0.465 mg/L로 청정지역을 제외한 나머지 지역 배출허용기준(각각 4.0 mg/L, 1 mg/L) 이내로 조사되었다. PCA 분석 결과, 인천지역 수질오염에 대한 사업장 방류수의 주요인으로 PC1은 니켈, 아연과 관련된 도금업 사업장 요인, PC2는 철과 관련된 금속가공 사업장 요인, PC3은 구리압연압출공정 등 구리 배출 사업장 요인, PC4는 크롬 배출 사업장 요인으로 판단된다. SVM, LR, ANN과 같은 분류 알고리즘을 이용하여 사업장 방류수의 업종별 분류 정확도를 평가한 결과, ANN 알고리즘이 가장 높은 정확도를 나타내었고 업종별 분류 가능성은 인천지역 사업장 방류수의 업종별 수질오염물질 배출 특성 추출 가능성 및 개연성을 의미한다.

주제어 : 사업장 방류수, 수질오염물질, 특정수질유해물질, 주성분 분석, 데이터 마이닝, 분류 기법

1. 서론

하천 등 수생태계 내 수질오염원의 하나인 사업장 방류수는 생활하수에 비해 배출량이 상대적으로 적고 처리효율이 높지만 오염 농도가 높고 오염 부하량이 절대적으로 많

아 철저한 관리가 요구된다. 환경부는 사업장 방류수에 의한 수질오염을 방지하기 위해 설치허가, 배출허용기준, 토지이용규제, 폐수종말처리장 운영 등의 제도를 이용하여 관리하고 있다. 하지만 산업이 발달함에 따라 공공수역으로 방류되는 일반수질오염물질 및 특정수질유해물질의 검출

† Corresponding author E-mail: jjdong@uos.ac.kr Tel: 02-6490-2862 Fax: 02-2244-2245

빈도 및 독성은 해마다 증가하여 인간과 생태계에 악영향을 미치고 있으므로 이를 예측하고 관리해야 하는 필요성이 점차 증대되고 있다. 또한 생활하수는 시간이 흐르면 생분해되는 경우가 대부분이지만 사업장 방류수는 생물학적 처리가 어렵고 미량 성분도 인체에 치명적이며 수체에 독성이 강한 유해물질이 다양하게 존재하기 때문에 더욱 엄격한 관리가 필요하다.

환경부는 수질오염물질의 관리를 위해 현재 배출허용기준 적용 관리대상 항목은 51종, 폐수배출시설 적용기준이 있는 특정수질유해물질 항목은 29종으로 해마다 점차 확대하고 있으며 다양하고 새로운 오염물질에 대한 관리를 강화하고 있다.¹⁾ 유해물질에 대한 통합관리의 필요성이 제기되면서 2007년 12월부터는 생태독성 기준이 추가되었다. 또한 2018년 1월부터 환경부가 사업장의 특정수질유해물질 배출량을 조사 및 공개함으로써 자율적인 저감을 유도하고 수질개선 및 수생태계 보호를 위해 「특정수질유해물질 배출량 조사제도」를 실시하였다.

사업장 내 생산공정이 복잡해지고 이로 인해 발생하는 사업장 방류수 내 수질오염물질이 다양화되고 있는 추세이지만 각 지자체들은 이를 검토할 전문적인 시스템이 없기 때문에 서류만 검토하고 허가를 내주면서 문제점이 발생하기도 한다. 사업자가 허위신고를 하더라도 법규와 규정 등 이를 제재할 아무런 규제가 없다는 점도 문제점으로 떠오르고 있다. 이러한 이유로 사업자들은 폐수처리시설 설치 시 인·허가신청서에 수질오염물질 배출항목을 고의로 누락시키는 사례가 발생되기도 한다.²⁾

수질오염물질에 관한 연구동향을 살펴보면, 국립환경과학원은 업종별 최종방류수에서 수질오염물질의 배출현황을 조사하여 폐수배출시설 인허가 및 지도점검 개선방안 마련에 필요한 기초 자료를 마련하고자 하였으며³⁾ Lee 등⁴⁾은 산업단지 폐수종말처리장의 특정수질유해물질의 유입 및 방류 현황을 조사하여 유입수는 특례지역기준대비, 방류수는 청정지역기준대비 농도수준을 나타내었다. 또한 Kim 등⁵⁾은 광주지역 사업장 방류수 중 특정수질유해물질의 배

출특성 연구를 수행하여 항목별 검출빈도와 농도수준을 조사하였고 일원배치 분산분석으로 업종별 특정수질유해물질의 상관성을 도출하였다.

또한 최근 환경 분야에서 다양하게 시도되고 있는 데이터마이닝 기법에 관한 연구동향을 살펴보면, Kang 등⁶⁾은 2006년부터 2008년까지의 MODIS 자료를 이용하여 동아시아 지면피복분류 개선에 대한 연구를 수행하였고 Yun 등⁷⁾은 대기 중 미세먼지 통합 예보모형 개발연구에서 기상자료에 대해 신경망 알고리즘을 적용하는 등 다양한 데이터마이닝 기법을 활용하였다.

인천의 폐수 배출사업장 개수, 폐수발생량, 폐수방류량은 각각 3,089개소, 343,810 m³/일, 100,646 kg/일로 전국 기준 각각 6.1%, 6.9%, 2.7%를 차지하고 있다. 폐수 배출사업장을 규모별로 구분하면 1종(2000 m³/일 이상), 2종(700~2000 m³/일), 3종(200~700 m³/일), 4종(50~200 m³/일) 사업장이 각각 12개소, 32개소, 57개소, 135개소이며, 나머지 2,853개소가 5종 사업장(50 m³/일 이하)에 해당한다. 또한 특정수질 유해물질 배출사업장 개수, 특정폐수방류량은 각각 1,051개소, 62,398.7 m³/일로 전국 기준 각각 15.1%, 2.2%를 차지하고 있다(2015년 12월 말 기준).⁸⁾

본 연구에서는 인천지역 사업장의 방류수에 대해 업종별로 수질오염물질 및 특정수질유해물질을 조사하여 배출현황을 파악하고 기존에 활용된 통계학적 방법뿐만 아니라 데이터마이닝 기법을 새롭게 적용하여 검출된 수질오염물질에 대해 업종별 분류 및 배출 특성을 도출하고자 하였다. 이러한 결과를 바탕으로 향후 인천지역 내 폐수 배출사업장에 대한 수질오염물질 및 특정수질유해물질의 효율적인 관리를 위해 활용 가능한 기초자료를 마련하기 위함이다.

2. 연구내용 및 방법

2.1. 조사대상

2017년 1월부터 2018년 7월까지 인천지역 내 폐수 배출

Table 1. The category and number of investigated wastewater-discharging facilities

Category	Number	Category	Number	Category	Number
Car washing	288	Semiconductor-related product manufacturing	12	Car maintenance	5
Plating	120	Laundry	11	Seawater treatment	5
Printed circuit board manufacturing	99	Oil storage and purification	11	PCB manufacturing	4
Wastewater treatment	81	Meat processed food manufacturing	10	Scrap iron unloading	4
Soil stone collecting and processing	54	Gas manufacturing	6	Drug manufacturing	4
Sea-sand washing	28	Electronic components manufacturing	6	Primary metal manufacturing	3
Power plant facility	22	Grill washing	5	Sugars manufacturing	3
Metallic product manufacturing	20	Waterwork business	5	Painting and film treatment	3
Cosmetics manufacturing	16	Dye and stain manufacturing	5	Feed manufacturing	3
Food manufacturing	14	Medical care	5	Others (etc. facilities)	109
Construction	12	Automobile parts manufacturing	5	Total	978

사업장의 지도·점검을 위해 시, 군·구가 채수한 978개소의 사업장 방류수에 대해 검사 의뢰된 수질오염물질 및 특정수질유해물질 분석 결과를 조사하였다. 조사대상 사업장을 32개 업종으로 구분하면 Table 1과 같다. 사업장 중 세차시설(car washing)이 288개소(29.6%)로 가장 많았으며 도금업(plating) 120개소(12.3%), 인쇄회로기판 제조업(printed circuit board manufacturing) 99개소(10.1%), 폐수처리업(wastewater treatment) 81개소(8.3%), 토사석 채취및가공업(soil stone collecting and processing) 54개소(5.5%), 해사세척업(sea-sand washing) 28개소(2.9%), 발전시설(power plant facility) 22개소(2.2%), 금속제품제조업(metallic product manufacturing) 20개소(2.0%), 화장품제조업(cosmetics manufacturing) 16개소(1.6%), 식품제조업(food manufacturing) 14개소(1.4%) 순으로 나타났다. 조사대상 사업장 방류수의 일반항목 농도 수준[$\text{min.} \sim \text{max. (mean} \pm \text{SD)}$]은 BOD 0.1~719.7 (24.4 $\pm$ 82.4) mg/L, COD_{Mn} 0.1~778.0 (40.7 $\pm$ 69.8) mg/L, SS 0~389.2 (16.3 $\pm$ 29.0) mg/L, T-N 0.1~352.7 (12.3 $\pm$ 28.8) mg/L, T-P 0~52.2 (0.6 $\pm$ 2.7) mg/L로 조사되었다.

2.2. 조사항목 및 방법

본 연구에서는 중금속 7종(Cr^{6+} , Cu, Pb, As, Hg, Se, Cd), 화학적으로 유사한 특성을 가지는 염화유기화합물, 방향족화합물 등 VOCs 12종(1,1-dichloroethylene, 1,2-dichloroethane, 1,4-dioxane, dichloromethane, benzene, bromoform, carbon tetrachloride, acrylonitrile, vinyl chloride, chloroform, tetrachloroethylene, trichloroethylene) 및 polychlorinated biphenyl (PCB), diethylhexylphthalate (DEHP), 유기인, CN, phenol,

pentachlorophenol (PCP) 등 폐수배출시설 적용기준이 있는 특정수질유해물질 25종 및 배출허용기준이 있는 수질오염물질에 대해 조사하였다. 시료 채취, 보관 및 분석은 수질오염공정시험기준(환경부고시 제2017-4호)⁹⁾에 따르면 수질오염물질 항목별 분석 방법을 Table 2에 나타내었다. 분석값에 대한 모든 통계 및 데이터마이닝 분석은 R (version 3.5.1), Matlab (version R2018a) 프로그램을 이용하여 수행하였다.

2.3. 통계 및 데이터마이닝을 이용한 사업장 방류수 배출 특성 연구

2.3.1. 주성분 분석

주성분 및 요인분석은 다변량 통계분석의 여러 가지 기법 중에서 가장 기본이 되는 분석 방법이다. 주성분 분석(Principal Component Analysis, PCA)은 변수 간에 상관성을 갖는 여러 종류의 특성치에 대한 정보를 큰 손실 없이 요약 표현이 가능하며, 이들 속에 내재하고 있는 상호 독립적인 고유의 패턴을 상호 연관한 특성치들의 변화 양상으로부터 도출할 수 있다. 소수의 특징적인 변량을 합성하여 다변량으로 구성되는 데이터로부터 주성분을 산출하고, 차원이 축소되어 데이터 분석이 용이해진 주성분을 사용하여 전체 다변량 데이터의 경향을 분석하게 된다. 주성분들은 원시 데이터(raw data)의 분산을 최대한 보존하는 방법으로 구한다.¹⁰⁾

2.3.2. 분류 기법

분류(Classification) 기법은 데이터마이닝에서 많이 사용되고 있는 기법으로 대용량의 데이터로부터 의미 있는 패턴이나 규칙을 추출하는 분석 방법이다. 즉, 모든 객체들을

Table 2. Analysis methods for water pollutants

Compound	Methods
COD _{Mn}	KMnO ₄ method at 100℃
BOD	20℃ 5days cultivation & check DO before and after incubation DO
SS	Vacuum Filtration (Glass Fiber Filters, GF/C)
T-N, T-P	Continuous Flow Analysis (CFA) (AASU-2000, BLTEC)
Hg	pyrolysis, amalgamation & AA (DMA-80, MILESTONE)
Cu, Cr, Mn, Fe, Ni, Zn, As, Se, Ba, Pb, Cd	ICP/MS (Inductively coupled plasma / Mass Spectrometer) (7800 ICP-MS, Agilent)
Cr^{6+}	Ion chromatography (IC) (944 Professional+UV/VIS Detector Vario, Metrohm)
TCE, PCE, benzene, dichloromethane, carbon tetrachloride, 1,1-dichloroethylene, 1,2-dichloroethane, chloroform, vinyl chloride, acrylonitrile, bromoform, naphthalene, toluene, xylene, 1,4-dioxane	HS (Headspace) & GC/MS (Gas Chromatograph / Mass Spectrometer) (SCION SQ GCMS (456-GC), Bruker)
formaldehyde	HS&GC/MS (Derivatization) (SCION SQ GCMS (456-GC), Bruker)
PCB	LLE (Liquid Liquid Extraction)-GC/MS (SCION SQ GCMS (456-GC), Bruker)
DEHP	LLE (Liquid Liquid Extraction)-GC/MS (SCION SQ GCMS (456-GC), Bruker)
organic phosphorus	LLE (Liquid Liquid Extraction)-GC/MS (SCION SQ GCMS (456-GC), Bruker)
phenol	LLE (Liquid Liquid Extraction)-GC/MS (SCION SQ GCMS (456-GC), Bruker)
PCP	LLE (Liquid Liquid Extraction)-GC/MS (SCION SQ GCMS (456-GC), Bruker)
CN, phenols, ABS	Continuous Flow Analysis (CFA) (AutoAnalyzer3, BRAN LUEBBE)

미리 정의된 범주를 가진 클래스들 중에 하나로 할당하는 기법으로 지지도 벡터 기계(Support Vector Machine, SVM), 로지스틱 회귀분석(Logistic Regression, LR), 인공신경망(Artificial Neural Network, ANN)과 같은 다양한 분류 기법이 있다.¹¹⁾

2.3.2.1. 지지도 벡터 기계

SVM은 통계적 학습이론 기반의 패턴 분류 알고리즘이다. SVM은 주어진 데이터 사이에서 찾아낸 최적 초평면(Optimal hyperplane)을 이용하여 두 개의 클래스를 구분한다. 초평면을 중심으로 두 범주를 나누어주는 개체를 지지도 벡터(Support vector)라고 하고, 이것들의 거리를 최적화시키는 것이 중요하다. 분류 초평면은 아래와 같은 식 (1)로 나타낼 수 있다.

$$< w \cdot x > + b = 0 \quad (1)$$

여기서 w 는 가중치 벡터이고 b 는 편향값, x 는 입력벡터이다. 주어진 입력벡터 x 의 클래스를 찾기 위해 식 (2)와 같이 최적의 결정함수를 찾는다.

$$f(x) = \text{sign}(< w \cdot x > + b) \quad (2)$$

SVM에서는 특정 공간에 대해 선형 분리가 불가능한 경우, 원래의 입력 공간을 새로운 고차원의 특징 공간으로 사상(mapping)하여 복잡한 비선형 형태를 가지는 데이터를 각 클래스로 분류할 수 있다. 일반적으로 많이 사용되는 커널 함수(Kernel function)로는 Polynomial kernel, Gaussian radial basis function kernel, Sigmoid kernel이 있다.¹¹⁾

2.3.2.2. 로지스틱 회귀분석

LR은 확률 모델로서 독립변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계 기법이다. 종속변수가 범주형 데이터를 대상으로 하며 입력 데이터가 주어졌을 때 해당 데이터의 결과가 특정 분류로 나뉘기 때문에 분류 기법으로 활용된다. 로지스틱 모형 식은 독립변수에 상관없이 종속변수 또는 결과값이 항상 범위 [0,1] 사이에 있도록 한다. 이는 오즈비(odds ratio)(식 (3))를 로짓(logit) 변환(식 (4))함으로써 얻어지고 그 형태는 식 (5)와 같다.¹¹⁾

$$\text{odds} = \frac{p(y=1|x)}{1-p(y=1|x)} \quad (3)$$

$$\text{logit}(p) = \log \frac{p}{1-p} \quad (4)$$

$$\text{logistic function} = \frac{e^{\beta_0 + \beta_1 X_1}}{1 + e^{\beta_0 + \beta_1 X_1}} \quad (5)$$

2.3.2.3. 인공신경망

ANN은 많이 사용되는 분류 기법 중 하나로 복잡한 구조를 가진 데이터 사이의 관계나 패턴을 찾아내는데 유리하고 범주형 및 연속형 변수 모두를 다룰 수 있다. 일반적으로 입력층(Input layer), 출력층(Output layer), 그리고 하나 이상의 은닉층(Hidden layer)으로 구성된다. 신경망은 반복적으로 입력된 정보에 대하여 입력 신호의 가중치를 목적에 맞도록 변화시킴으로써 예측 모형을 학습시킨다. 각 신경망의 노드마다 다른 가중치를 가지며 신경망에 의해 계산된 실제 출력값과 목표값과의 오차를 최소화하기 위해 신경회로망의 가중치를 반복적으로 갱신한다. 오차 제곱의 합으로 정의된 비용함수(cost function) E 의 일반식은 아래 식 (6)으로 나타낼 수 있다.^{11,12)}

$$E = \frac{1}{2} \sum_{i=1}^N (o_i - d_i)^2 \quad (6)$$

3. 결과 및 고찰

3.1. 특정수질유해물질 조사 결과

특정수질유해물질은 각 항목별 위해성 및 독성이 다르므로 절대농도를 서로 비교하기보다는 수질기준 대비 최대값의 비가 중요한 의미를 갖는다. 그러므로 항목별 농도 분포 및 90%tile 농도값을 청정지역 배출허용기준과 비교하여 Table 3에 나타내었다. 조사기간 중 검출되지 않은 특정수질유해물질은 1,1-dichloroethylene, benzene, carbon-tetrachloride, vinylchloride, acrylonitrile, PCP, 유기인 등 7종이었다. 또한 조사기간 중 단 1회만 검출되어 검출농도가 최대값에 해당되는 항목은 1,2-dichloromethane (N.D~0.006 mg/L), PCE (N.D~0.004 mg/L) 등 2종이었다.

조사기간 중 90%tile 농도값이 청정지역 배출허용기준을 초과하는 항목은 1,4-dioxane, CN 등 2종이었다. 청정지역 배출허용기준 대비 90%tile 농도값 비율이 가장 높은 항목은 1,4-dioxane (24.24)이었고, 95%tile 농도값은 3.409 mg/L로 청정지역을 제외한 나머지 지역 배출허용기준(4.0 mg/L) 이내로 조사되었다. CN은 청정지역 배출허용기준 대비 90%tile 농도값 비율이 1.12 수준으로 나타났고, 95%tile 농도값은 0.465 mg/L로 청정지역을 제외한 나머지 지역 배출허용기준 (1 mg/L) 이내로 조사되었다.

3.2. 업종별 주요 수질오염물질 배출 현황

업종별로 구분하여 사업장 방류수의 항목별 배출 오염도를 조사하였다. 시료개수가 많은 상위 10개소 중 세차업, 토사석 채취및가공업, 해사세척업, 발전시설, 금속제품제조업, 화장품제조업의 주요 조사항목은 COD_{Mn}, BOD, SS, T-N, T-P, ABS 등 일반항목 6종으로 아래 평가에서는 제외하였다.

Table 3. Analysis of water quality and evaluation of specific hazardous substance from the industrial wastewater effluent

Compound	Detection rate (%) ($N_{\text{detected}}/N_{\text{total}}$)	Concentration (mg/L)					Effluent standard** (mg/L)	Ratio***
		Median	Mean	90%tile	95%tile	Max.		
Cu	92.8 (372/404)	0.0362	0.3254	0.9608	1.5458	4.4610	1	0.96
Pb	48.3 (101/209)	N.D*	0.0036	0.0045	0.0254	0.1600	0.1	0.05
As	77.6 (118/152)	0.0005	0.0041	0.0012	0.0235	0.0629	0.05	0.02
Se	48.1 (37/77)	N.D	0.0025	0.0021	0.0051	0.1269	0.1	0.02
Hg	60.5 (75/124)	N.D	N.D	0.0002	0.0002	0.0009	0.001	0.20
Cd	42.7 (61/143)	N.D	0.0002	0.0005	0.0011	0.0020	0.02	0.03
Cr ⁶⁺	30.0 (69/230)	N.D	0.015	0.018	0.034	1.030	0.1	0.18
1,2-Dichloromethane	1.3 (1/77)	-	-	-	-	0.006	0.03	-
1,4-Dioxane	71.4 (70/98)	0.045	0.682	1.212	3.409	12.732	0.05	24.24
Dichloromethane	15.7 (16/102)	N.D	0.004	0.011	0.017	0.171	0.02	0.55
Bromoform	6.3 (4/64)	N.D	0.0059	N.D	0.0153	0.236	0.03	0.00
Chloroform	36.8 (42/114)	N.D	0.0207	0.0417	0.1031	0.4360	0.08	0.52
PCE	1.1 (1/87)	-	-	-	-	0.0200	0.02	-
Trichloroethylene	5.9 (6/101)	N.D	0.0067	N.D	0.0060	0.4810	0.06	0.00
PCB	33.3 (4/12)	N.D	2.1E-06	9.5E-06	0.00001	0.00001	0.0005	0.02
DEHP	21.9 (14/64)	N.D	0.0020	0.0030	0.0049	0.0807	0.02	0.15
CN	84.5 (235/278)	0.007	0.115	0.224	0.465	5.684	0.2	1.12
Phenol	61.5 (8/13)	0.0003	0.0011	0.0015	0.0045	0.0089	0.1	0.02

* N,D (Not Detected) : Detection limit less than

** Effluent standard : Effluent quality standard for clean area

*** Ratio : the ratio of 90%tile to the Effluent quality standard for clean area

3.2.1. 도금업, 인쇄회로기판 제조업, 식품제조업

금속가공 시설과 같은 도금업은 중금속 오염도가 주요 조사항목이다. 도금업은 120개소로 전체 조사대상의 12.3%를 차지하였으며 평가대상은 Cr⁶⁺, Cu, Pb, Ni, Zn, Fe, Cr, CN 등 총 8종이다. Ni의 농도가 중간값 0.2397 mg/L로 상대적으로 높았고 일반적으로 도금업에서 Ni와 함께 배출된다고 알려진 Cr, Cu, CN도 각각 중간값 0.0335 mg/L, 0.0688 mg/L, 0.0590 mg/L로 일정수준의 농도로 검출되었다. 인쇄회로기판제조업은 99개소로 전체 조사대상의 10.2%를 차지하였으며 평가대상은 Cu, Pb, Ba, As, Mn, Fe, CN 등 총 7종이다. Cu와 Fe이 각각 중간값 0.4648 mg/L, 0.8900 mg/L로 상대적으로 높게 검출되었다. 식품 제조업은 14개소로 전체 조사대상의 1.4%를 차지하였으며 평가대상은 Cu, Pb, As, Se, DEHP, CN 등 총 6종이다. Cu와 As가 각각 중간값 0.0064 mg/L, 0.011 mg/L로 상대적으로 높게 검출되었다. 도금업, 인쇄회로기판 제조업, 식품제조업의 조사항목별 농도분포는 Fig. 1에 나타내었다. 업종별로 각각 두 개의 그래프로 나타냈는데 왼쪽의 (a), (b), (c)는 전체 농도값의 분포를 box-plot과 scatter-plot으로 동시에 나타냈고 오른쪽의 (a)-1, (b)-1, (c)-1은 작은 스케일의 y축을 사용하여 일부 낮은 농도값 자료만 scatter-plot(up-scale)으로 나타냈다. box-plot의 + 표시(빨간색)는 outlier를 의미한다.

3.2.2. 폐수처리업

고농도 유해화학물질을 처리하는 폐수처리업은 여러 종

류의 중금속 및 휘발성유기화합물의 발생농도가 다른 배출 시설에 비하여 상대적으로 높아 최적방지기술 도입 및 적정 운영이 중요하다. 폐수처리업은 81개소로 전체 조사대상의 8.3%를 차지하였고 조사항목은 1,4-dioxane, benzene, DEHP, naphthalene, phenols, toluene, xylene 등 방향족탄화수소 7종, 1,1-dichloroethylene, 1,2-dichloroethane, dichloromethane, bromoform, carbon-tetrachloride, chloroform, epichlorohydrin (ECH), vinylchloride, TCE, PCE 등 할로젠화탄화수소 10종, Cr⁶⁺, Cu, Pb, Ni, Ba, As, Se, Hg, Zn, Mn, Fe, Cd, Cr 등 중금속 13종 및 기타 항목으로 acrylonitrile, CN, F, formaldehyde, 유기인(organic phosphorus, O-P)이며 조사항목별 농도분포는 Fig. 2에 나타내었다. 할로젠화탄화수소 중 1,4-dioxane과 phenols의 90%tile 농도값은 각각 1.7838 mg/L, 0.0350 mg/L, 방향족탄화수소 중 chloroform과 dichloromethane의 90%tile 농도값은 각각 0.0309 mg/L, 0.0126 mg/L, 중금속 중 Cu, Fe, Mn, Ni의 90%tile 농도값은 각각 0.0096 mg/L, 1.1743 mg/L, 0.4723 mg/L, 0.1460 mg/L, 기타 항목으로 F의 90%tile 농도값은 2.0 mg/L로 나타나 일정수준의 농도로 검출되었다. 항목 그룹별로 각각 두 개의 그래프로 나타냈는데 왼쪽의 (a), (b), (c), (d)는 전체 농도값의 분포를 box-plot과 scatter-plot으로 동시에 나타냈고 오른쪽의 (a)-1, (b)-1, (c)-1, (d)-1은 작은 스케일의 y축을 사용하여 일부 낮은 농도값 자료만 scatter-plot (up-scale)으로 나타냈다.

Fig. 1. Concentration distribution of water pollutants from the industrial wastewater effluent : (a) Plating, (b) Printed circuit board manufacturing, (c) Food manufacturing, and (a)-1, (b)-1, (c)-1 are made of Y-axis, which have smaller maximum values.

Fig. 2. Concentration distribution of water pollutants from the effluent of wastewater treatment facilities : (a) Aromatic hydrocarbons, (b) Halogenated hydrocarbons, (c) Heavy metals (d) Others(etc. components), and (a)-1, (b)-1, (c)-1 are made of Y-axis, which have smaller maximum values.

3.3. 사업장 방류수 수질오염물질 배출 특성 추출 및 분류

3.3.1. 주성분 분석을 이용한 통계 분석

수질오염물질은 변수 상호 간의 상관성이 내재되어 있어 단상관 분석만으로는 수질오염의 주요 영향인자(배출원)를 파악하기 어렵다. 따라서 본 연구에서는 인천지역 공공수계의 수질오염에 영향을 미치면서 서로 독립적인 복합요인을 파악하기 위하여 주성분 분석(PCA)을 수행하였다. 조사 대상은 총 32개 업종이며 해당되는 사업장 방류수는 조사 대상 978개소 중 주요 중금속류 분석이 수행된 사업장 방류수 434개이다. PCA의 주요 변수로는 상대적으로 높은 검출빈도를 보이거나 고농도로 조사된 주요 중금속류 10종(Cr^{6+} , Cu, Pb, Ni, As, Zn, Mn, Fe, Cd, Cr)을 선정하였다. 중금속류로 제한한 이유는 공분산 행렬로부터 유도된 주성분은 측정단위의 크기에 좌우되기 때문에 척도(scale)의 영향을 줄이기 위함이다. 결측치(NA)에 대한 전처리하는 중간값의 surrogate 변수(대체값)를 사용하였고 공분산 행렬을 이용한 선형 결합의 PCA를 수행한 결과는 Table 4와 Fig. 3에 나타내었다. Table 4는 원데이터의 분산 중 주성분들의 설명력을 보여주고 있는데 각 주성분별 기여도가 PC1은 43.1%, PC2는 21.0%, PC3은 15.3%, PC4는 10.3%로 조사되었으며 최종적으로 중금속류 10종이 4개의 주성분으로 차원 축소되어 약 89.7%의 기여도를 보였다. Fig. 3에서 PCA를 수행하여 도출된 각 주성분의 특성을 파악하면 인천지역 수질오염에 대한 사업장 방류수의 주요인으로 PC1은 Ni,

Zn과 관련된 도금업 사업장 요인, PC2는 Fe과 관련된 소규모 철강, 금속가공 및 기계 관련 사업장 요인, PC3은 구리압연압출공정 등 Cu 배출 사업장 요인, PC4는 Cr 배출 사업장 요인으로 판단된다. PC1과 PC4는 요인 간 완전한 독립성이 이루어지지 않았는데 Cr이 도금업을 포함하여 다양한 업종의 사업장에 배출되기 때문으로 판단된다. 본 연구에서 도출된 결과만으로 특정 업종에서 특정수질유해물질을 포함한 수질오염물질의 배출여부를 단정하기에는 통계적으로 부족하다. 실제 현장에서는 지하수 및 배관망 등으로부터의 오염, 원료의 구성성분, 제품생산 공정 중 반응에 따른 부산물 등 다양한 원인과 경로를 통해 사업장 방류수 중 미량의 특정수질유해물질이 검출될 가능성이 있다. 따라서 사업장 방류수 중 수질오염물질의 정확한 검출원인을 파악하기 위해서는 방류수 수질검사와 함께 사용원료, 제품생산 공정, 폐수 발생과 흐름 등 전반적인 공정분석이 병행되어야 한다고 판단된다.

Table 4. Importance of components

Compound	PC1	PC2	PC3	PC4
Standard deviation	1.0884	0.7598	0.6485	0.5325
Proportion of variance	0.4309	0.2099	0.1529	0.1031
Cumulative proportion	0.4309	0.6408	0.7938	0.8970
Identification	Ni, Zn	Fe	Cu	Cr

Fig. 3. Results of principal component analysis of heavy metals data for the industrial wastewater effluent in Incheon area.

3.3.2. 분류 기법을 이용한 데이터마이닝

본 연구에서는 SVM, LR, ANN의 분류 알고리즘을 이용하여 주요 중금속류 10종(Cr⁶⁺, Cu, Pb, Ni, As, Zn, Mn, Fe, Cd, Cr)의 농도값에 대한 업종별 분류를 수행하고 인천지역 사업장 방류수의 배출 특성에 따른 업종별 분류 가능성을 평가하였다. 조사대상은 총 32개 업종이며 해당되는 사업장 방류수는 조사대상 978개소 중 주요 중금속류 분석이 수행된 사업장 방류수 434개이다. 이 중 대표적으로 도금업, 폐수처리업, 인쇄회로기판 제조업에 대해 예측데이터와 실측데이터 간의 분류 알고리즘별 교차분할표(32 by 32)를 산출하였다. Table 5~7은 분류 알고리즘별 오분류표(Confusion matrix)를 나타내는데 오분류표 값을 활용하여 분류 정확도(Accuracy), 민감도(Sensitivity), 특이도(Specificity) 등의 분류 예측모형을 평가하는 지표를 계산할 수 있다. 정확도는 전체 자료 중 실제값과 예측값이 일치하는 정도를 의미한다. 민감도는 실제값이 TRUE인 자료 중 예측값이 적중한 정도를 의미하고 특이도는 실제값이 FALSE인 자료 중 예측값이 적중한 정도를 나타낸다. 이렇게 계산된 분류 정확도, 민감도, 특이도 값을 알고리즘별로 Fig. 4에서 비교 분석하였다. 결측치(NA)에 대한 전처리하는 항목별 전체 업종에 대한 중간값의 surrogate 변수(대체값)를 사용하였고 scale 함수를 이용하여 입력 데이터의 분포를 표준화(정규화) 하였다. SVM의 경우, 커널(kernel) 함수는 “radial”, 오분류 비용(cost)값은 1, gamma 값은 0.1, LR의 경우, 최대반복수(maxit)는 1000번, ANN의 경우, 활성화함수(activation function)은 “logistic”, 은닉층의 노드 수(hidden)는 13으로 수행하였다. 분석결과를 Fig. 4에 나타내었는데 도금업, 폐수처리업, 인쇄회로기판 제조업의 분류 정확도에서 ANN 알고리즘이 각각 92.4%, 89.2%, 93.6%로 나타나 3개의 분류 알고리즘 중 정확도가 가장 높았다. 3개 업종에서 ANN 알고리즘의 민감도는 각각

Fig. 4. Performance evaluation of algorithms : (a) Accuracy $\{(TP+TN)/(TP+TN+FP+FN)\}$ (b) Sensitivity $\{TP/(TP+FN)\}$, (c) Specificity $\{TN/(FP+TN)\}$.

Table 5. Results of SVM algorithm : Confusion matrix for (a) Plating, (b) Wastewater treatment, (c) Printed circuit board manufacturing

(a) n=120				(b) n=78				(c) n=99			
		Predicted				Predicted				Predicted	
		Positive	Negative			Positive	Negative			Positive	Negative
Actual	Positive	106 (TP*)	14 (FN*)	Actual	Positive	65	13	Actual	Positive	62	37
	Negative	141 (FP*)	173 (TN*)		Negative	40	316		Negative	20	315

* TP : True Positive, FN : False Negative, FP : False Positive, TN : True Negative

Table 6. Results of LR algorithm : Confusion matrix for (a) Plating, (b) Wastewater treatment, (c) Printed circuit board manufacturing

(a) n=120				(b) n=78				(c) n=99			
		Predicted				Predicted				Predicted	
		Positive	Negative			Positive	Negative			Positive	Negative
Actual	Positive	67	53	Actual	Positive	69	9	Actual	Positive	54	45
	Negative	97	217		Negative	50	306		Negative	37	298

Table 7. Results of ANN algorithm : Confusion matrix for (a) Plating, (b) Wastewater treatment, (c) Printed circuit board manufacturing

(a) n=120				(b) n=78				(c) n=99			
		Predicted				Predicted				Predicted	
		Positive	Negative			Positive	Negative			Positive	Negative
Actual	Positive	107	13	Actual	Positive	72	6	Actual	Positive	94	5
	Negative	20	294		Negative	10	346		Negative	29	306

89.2%, 89.2%, 93.6%, 특이도는 각각 93.6%, 97.2%, 92.2%로 나타나 3개의 분류 알고리즘 중 ANN 알고리즘이 가장 우수한 것으로 나타났다. 업종별 분류 가능성은 인천지역 사업장 방류수의 업종별 수질오염물질 배출 특성 추출 가능성 및 개연성을 의미한다. 산업이 발달함에 따라 공공수역으로 방류되는 수질오염물질의 검출빈도 및 독성은 증가하고 있으며 이에 따라 사업장 방류수의 배출특성을 예측하고 사전에 관리해야 하는 필요성이 점차 증대되고 있다. 본 연구에서 이용된 수질자료는 결측치(missing data)를 포함하고 있어 데이터마잉 기법을 적용하는데 다소 어려움이 있었다. 하지만 데이터마잉 기법을 통해 업종별 분류 가능성을 검토하였고 충분히 높은 분류 정확도 및 예측성을 기대할 수 있었다. 향후 수질자료를 빅데이터(Big data) 형태로 구축하고 분류 기법 이외에 다양한 데이터마잉 기법을 활용하여 수질오염물질에 대한 업종별 배출특성을 추출하고 이러한 정보를 활용한다면 각 지자체의 사업장에 대해 업종별 배출특성에 맞는 최적화된 수질 관리방안을 마련할 수 있을 것으로 판단된다.

4. 결론

본 연구는 인천지역 내 폐수 배출사업장의 방류수에 대해 업종별로 수질오염물질 및 특정수질유해물질의 배출현황을 파악하고 업종별 오염물질의 배출 특성을 추출하고자 추진되었으며 다음과 같은 결론을 도출하였다.

1) 특정수질유해물질 조사 결과, 1,1-dichloroethylene, benzene, carbon-tetrachloride, vinylchloride, acrylonitrile, PCP, 유기인, 1,2-dichloromethane, PCE 등은 조사기간 중 불검출 및 단 1회만 검출되었고, 90%tile 농도값이 청정지역 배출 허용기준을 초과하는 항목은 1,4-dioxane, CN 등 2종이었다. 1,4-dioxane의 95%tile 농도값은 3.409 mg/L로 청정지역을 제외한 나머지 지역 배출허용기준(4.0 mg/L) 이내로 조사되었다. CN의 95%tile 농도값은 0.465 mg/L로 청정지역을 제외한 나머지 지역 배출허용기준(1 mg/L) 이내로 조사되었다.

2) 업종별 주요 수질오염물질 배출 현황을 보면, 도금업의 방류수는 Ni의 농도가 상대적으로 높았고 Cr, Cu, CN도 일정수준의 농도로 검출되었다. 인쇄회로기판제조업의 방류수는 Cu와 Fe이 상대적으로 높게 검출되었다. 식료품 제조업의 방류수는 Cu와 As가 상대적으로 높게 검출되었다.

3) 고농도 유해화학물질을 처리하는 폐수처리업의 방류수는 할로젠화탄화수소 중 1,4-dioxane과 phenols, 방향족탄화수소 중 chloroform과 dichloromethane, 중금속 중 Cu, Fe, Mn, Ni, 기타 항목으로 F가 일정수준의 농도로 검출되었다.

4) PCA 분석 결과, 인천지역 수질오염에 대한 사업장 방류수의 주요인으로 PC1은 Ni, Zn과 관련된 도금업 사업장

요인, PC2는 Fe과 관련된 금속가공 사업장 요인, PC3은 구리압연압출공정 등 Cu 배출 사업장 요인, PC4는 Cr 배출 사업장 요인으로 판단된다.

5) SVM, LR, ANN과 같은 분류 알고리즘을 이용하여 주요 중금속류 10종의 농도값에 대한 업종별 분류를 수행하고 분류 정확도를 평가한 결과, 인천지역 사업장 방류수의 배출 특성에 따른 업종별 분류 모형은 ANN 알고리즘이 가장 높은 정확도를 나타냈고 업종별 분류 가능성은 인천지역 사업장 방류수의 업종별 수질오염물질 배출 특성 추출 가능성 및 개연성을 의미한다.

수생태계의 위해성을 우선적으로 고려한다면 현행 배출 허용기준에서 업종별 배출특성을 반영하는 차등화 규제 방법을 적용해 볼 필요가 있다. 본 연구결과는 인천지역 내 폐수 배출사업장에 대한 수질오염물질 및 특정수질유해물질의 효율적인 관리를 위해 기초자료로 활용될 수 있을 것으로 기대된다.

Acknowledgement

본 연구는 2017, 2018년도 환경부 “환경 분야 시험·검사의 국제적 적합성 기반 구축” 사업에 따른 국고보조금이 일부 지원되어 이루어졌으며 이에 감사드립니다.

KSEE

References

1. Ministry of Environment (MOE), “Enforcement Regulation of the Water Environment Conservation Act” (Amended by Act No. 745, Jan. 18, 2018).
2. Nakdong river supervision board, “Research on the Status of Occurrence and Management Plan of Specific Hazardous Substances in Nakdong River Water System”(2016).
3. National Institute of Environmental Research (NIER), “A Study on Discharge of Hazardous Water Pollutants in Industrial Wastewater”(2012).
4. Lee, W. T., “Sampling Survey of Hazardous Water Pollutants in Industrial Wastewater Treatment Plants,” *J. Korean Soc. Water Environ.*, **37**(10), 590~595(2015).
5. Kim, S. H., “Study on Characteristics of Specific Hazardous Substances in the Industrial Wastewater Effluent,” *Anal. Sci. & Technol.*, **29**(3), 114~125(2016).
6. Kang, J. H., “Land Cover Classification over East Asian Region Using Recent MODIS NDVI Data (2006-2008),” *Atmos. Korean Meteorol. Soc.*, **20**(4), 415~426(2010).
7. Yun, H. Y., “A Study on PM10 Data Assimilation and Integrated Forecasting Model Development,” Anyang University (2017).
8. Ministry of Environment (MOE), “2017 Occurrence and Disposal of Industrial Wastewater”(2017).

9. Ministry of Environment (MOE), "Korea Standard Methods for the Examination of Water and Wastewater"(2017).
10. Kim, I. K., "Characterization of Water Quality in Changnyeong-Haman Weir Section Using Statistical Analysis," *J. Korean Soc. Environ. Eng. Water Environ.*, **38**(2), 71~78(2016).
11. Suh, D. H., "Prediction of Bladder Cancer Recurrence Using Classification Methods," *J. Korean Inst. Informat. Sci. and Eng.*, **39**(3), (2012).
12. Jeon, S. H., "Prediction of Fine Dust PM10 using Deep Neural Network," Chongnam National University(2018).
13. Ahn, T. U., "Study on Discharge Characteristics of Water Pollutants among industrial wastewater per industrial classification and the probability evaluation," *J. Korean Soc. Environ. Eng.*, **38**(1), 14~24(2016).
14. Jang, J. I., "Spatial characteristics of pollutant concentrations in the streams of shihwa lake," *J. Korean Soc. Environ. Eng.*, 289~299(2011).