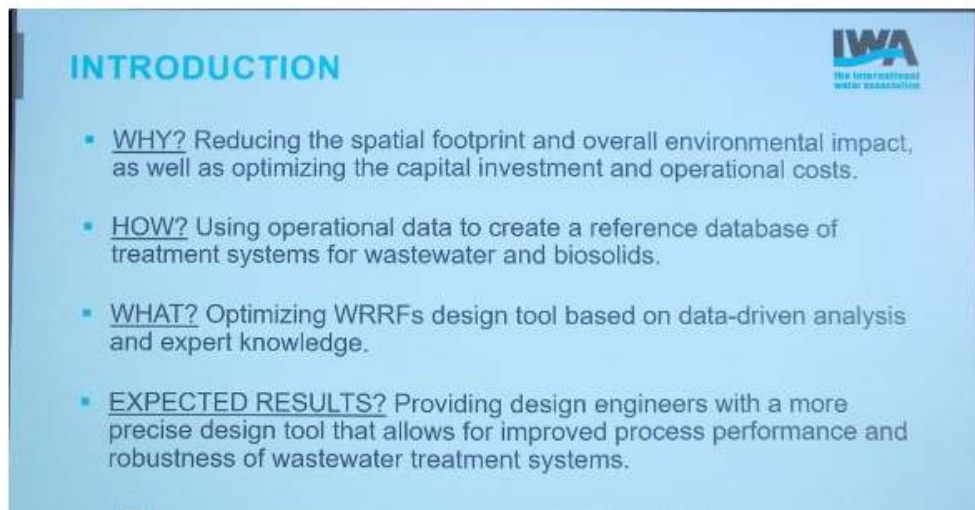


TS 1.7 Lessons from utility operations

: (4) DREAM - Data for Reengineering and Evaluation of Algorithms and Models : Mining of Operational Data to Optimize a Design Tool for Wastewater Treatment Systems

발표자 : Mathieu Delahaye



- 프로젝트의 목표

SUEZ를 통한 폐수 처리 플랜트나 식수 처리 플랜트 설계 최적화하는 것이다. 기존의 툴은 1980년대나 1990년대의 지식을 바탕으로 하고 있어, 개선 필요성이 있었다. 데이터를 수집해 방정식을 만들고 설계 툴을 업데이트하는 기존의 방식보다는, 머신러닝과 딥러닝 시스템을 사용해 툴을 최적화하는 프로젝트 진행하고자 한다.

→ 운영 중인 플랜트의 관찰 데이터를 더 정확하게 반영할 수 있는 설계 도구를 만들고자 한다.

METHODS

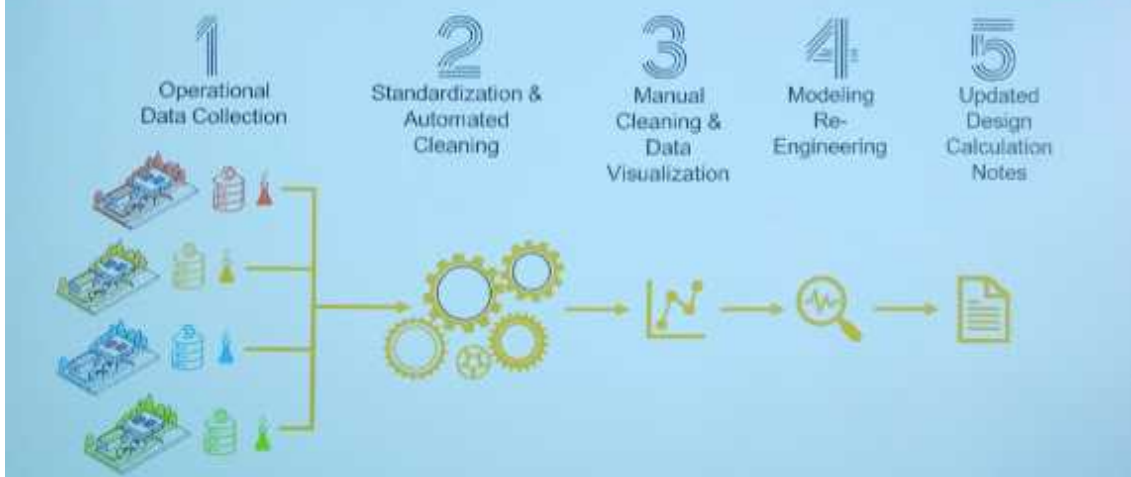


- Multi-year operational data available from 50+ WRRFs designed, built and managed by SUEZ were collected world-wide
- The DREAM project took advantage of this large amount of data available by applying data-driven models and expert knowledge as support tools for the optimal design of future plants.



- 전 세계적으로 SUEZ에 의해 설계, 건설, 관리된 50개 이상의 물 재이용 시설(WRRFs)에서 다년간의 운영 데이터가 수집된다.
- DREAM 프로젝트는 이러한 대량의 데이터를 활용하여 데이터 기반 모델과 전문가 지식을 지원 도구로 적용하여 향후 공장의 최적 설계를 지원한다.
- 각기 다른 지리적 위치와 기후 조건에서 다양한 폐수를 수집하였으며, 이에 따라 방대한 데이터베이스를 구축한다.

METHODS



- 프로젝트는 총 5단계로 진행된다.

1. 데이터 수집

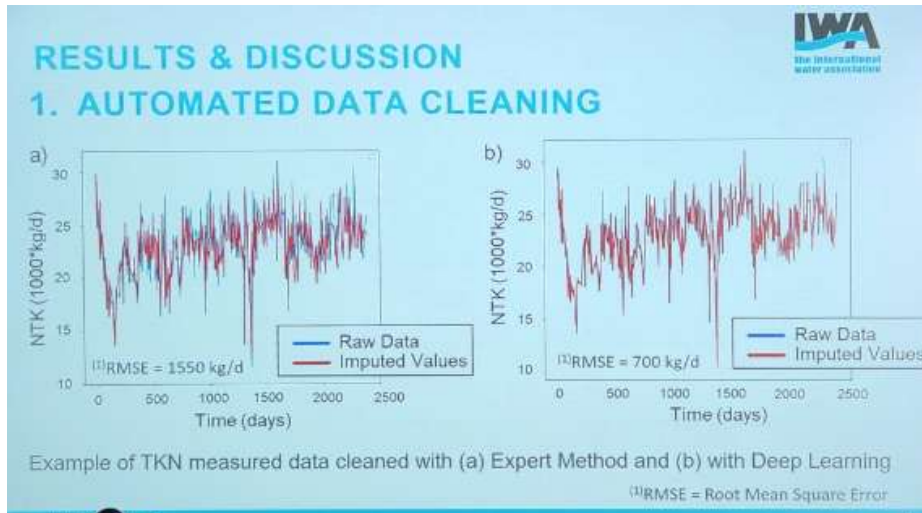
1. 운영자와 협력해 데이터를 확보하고 다양한 국가와 지역에서 데이터를 수집
2. SCADA 시스템에서 데이터를 얻는 것 뿐만 아니라 실험실 데이터도 수집

2. 데이터 표준화 및 형태 정리

1. 잘못된 데이터(예를 들어, 센서 오작동)가 들어온 경우 수동으로 정리하려면 매우 오래

걸리고 실수가 발생할 가능성이 큼

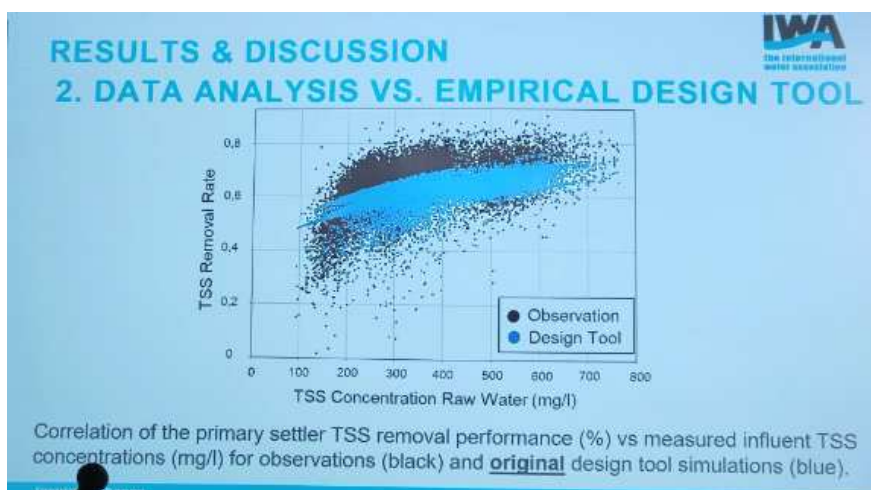
2. 딥러닝을 사용해 데이터를 자동으로 정리
3. 수동으로 데이터 정리
 1. 플랜트가 정상적으로 운영되지 않았던 시기, 유지보수 기간, 기능 장애가 발생했던 기간의 데이터들은 수동으로 틀에서 제외해야 했음
 2. 실제로 운영자들로부터 특정 기간 동안 플랜트에 문제가 있었음을 전달받았음
4. 데이터베이스 정리된 후, 최적모델 찾는 과정
5. 설계 틀에 사용되는 계산 노드 업데이트



Expert 방법으로 정제된 TKN 측정 데이터의 예시(a)와 딥러닝으로 정제된 예시(b)이다.

→ RMSE 확인해보면, Expert 방법으로 수동 정리한 데이터와 딥러닝을 통해 자동으로 정리한 데이터 사이의 오류 감소 확인할 수 있다.

- 해당 슬라이드는 TKN 농도와 부하에 관련된 것으로 모든 유형의 데이터를 대상으로 동일 작업 진행했다.



관찰된 데이터(검은색)와 초기 설계 도구 시뮬레이션(파란색)에 대한 1차 침전지의 TSS 제거 성능(%)과 측정된 유입 TSS 농도(mg/L)의 상관관계이다.

→ 원본 데이터를 확인해보면, 낮은 농도의 관측치와 설계 도구 결과 간의 차이가 존재한다.

RESULTS & DISCUSSION

3. MODEL RE-ENGINEERING

Search for trend correlations of errors with other process parameters using different types of data analytics models:

■ Linear Model:

$$TSS_{removal} = a + b * f_1([TSS]) + c * f_2([TSS/BOD]) + d * f_3([TSS/COD]) + e * f_4(v_{upflow}) + \dots$$

■ Logistic Model:

$$TSS_{removal} = a + \frac{b}{\left(1 + c * e^{d * f_1([TSS]) + e * f_2([TSS/BOD]) + g * f_3([TSS/COD]) + h * f_4(v_{upflow}) + \dots}\right)}$$

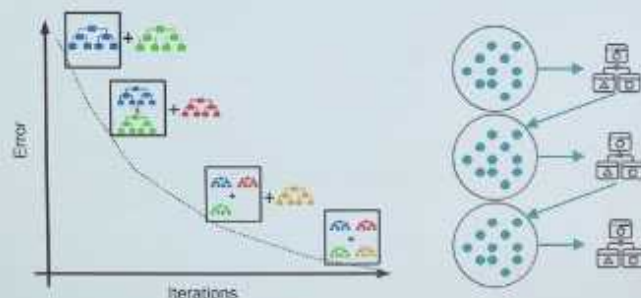
관측치와 초기 설계 도구 간의 차지를 줄이고자 프로젝트 시작했다.

- 1차 침전 외의 모든 공정에도 동일 작업 수행
- 데이터를 사용해 3가지 모델 테스트
 - 1) 선형 모델 : 간단한 계산과 다양한 매개변수로 TSS 제거 예측함
 - 2) 로지스틱 모델 : 선형 모델보다 더 복잡한 계산 과정 가짐
 - 3) XGB 모델(머신러닝) : 결정트리의 수를 반복적으로 늘려서 품질 최적화 및 오류 최소화하는 방식

RESULTS & DISCUSSION

3. MODEL RE-ENGINEERING

■ Decision-Trees (XGB):



Architecture of a Decision-Tree Model (XGB). (Source: neptune.ia)

- 3) XGB 모델(머신러닝) : 결정트리의 수를 반복적으로 늘려서 품질 최적화 및 오류 최소화하는 방식

RESULTS & DISCUSSION

3. MODEL RE-ENGINEERING

Performance analysis of the different type of models:

Plant	Original	Training			Testing		
		Linear	Logistic	XGB	Linear	Logistic	XGB
A	6,20	3,54	3,03	2,88	3,67	3,68	3,21
B	5,00	4,39	4,41	4,03	4,42	4,45	4,36
C	4,11	4,35	4,33	3,60	4,35	4,34	4,07
D	6,13	3,93	4,06	3,21	4,33	4,50	5,71
E	8,47	5,63	5,44	2,99	6,49	6,34	6,58
F	14,18	5,06	5,04	3,60	5,08	5,06	5,04
G	7,32	4,31	4,11	3,53	4,35	4,12	6,37
ALL	6,73	4,29	4,30	3,33	4,52	4,53	5,98

Model performance expressed as Mean Absolute Error (%) on TSS removal rates.

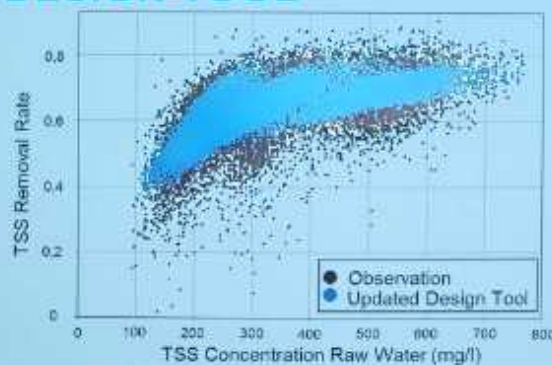
전체 데이터의 80%로 3가지 모델 훈련 후, 20%로 테스트 진행했고, 모델 성능은 각 플랜트의 TSS 제거율에 대한 평균 절대 오차(%)로 표현했다. 전 세계 7개 플랜트를 선택해서 해당 train 및 test 수행했다.

가장 복잡한 모델인 XGB 모델보다는, 로지스틱 모델을 선택하였다.

로지스틱 모델이 가장 균일한 결과 제공하였으며, XGB 모델은 일부 플랜트에서 오류가 매우 높게 나타났기 때문에 비균질하다고 판단하였다. 이는 기존의 데이터베이스에 기반해 학습한 것이 아닌 처음 접하는 데이터를 줄 경우, 그 결과가 완전히 잘못된 것이 되기 때문이다.

RESULTS & DISCUSSION

4. INTEGRATION OF UPDATED SEMI-EMPIRICAL MODEL IN DESIGN TOOL



Correlation of the primary settler TSS removal performance (%) vs measured influent TSS concentrations (mg/l) for observations (black) and for updated design tool simulations (blue).

관찰된 데이터(검은색)와 업데이트된 설계 도구 시뮬레이션(파란색)에 대한 1차 침전지의 TSS 제거 성능(%)과 측정된 유입 TSS 농도(mg/L)의 상관관계이다.

CONCLUSIONS



- The DREAM project has positively impacted the design of future WRRFs fully compliant with regulations with a smaller spatial footprint, reducing their environmental impact.
- Data analysis gives a better understanding of wastewater treatment systems in their geographical context that relate to both the wastewater composition and to local treatment regulations.
- Overall project outcome has allowed for optimization of robust semi-empirical models.



결론

- DREAM 프로젝트는 규정을 완전히 준수하면서 더 작은 공간적 차지로 환경 영향을 줄이는 미래의 물 재이용 시설(WRRFs) 설계에 긍정적인 영향을 미쳤다.
- 데이터 분석은 폐수 처리 시스템을 그 지리적 맥락에서 더 잘 이해할 수 있도록 하며, 이는 폐수 성분과 지역 처리 규정 모두와 관련이 있다.
- 전반적인 프로젝트 결과는 견고한 준경험적 모델의 최적화를 가능하게 했다.

→ 이전에 충분히 활용되지 않았던 방대한 데이터를 활용해 딥러닝을 통한 데이터 정리와 머신러닝을 통해 모델 최적화를 수행함으로써, 설계 도구를 실질적인 성능에 더 맞도록 최적화할 수 있었다.

1차 침전 외의 다른 생물학적 처리 공정에도 해당 작업을 수행함. 슬러지 처리 및 기타 처리에도 동일 작업 진행 중이다.

CAPEX(자본 지출), OPEX(운영 비용)를 예측하는 데 있어 최적의 설계를 목표로 하는 것이며, 화학물질 사용 줄이고, OPEX 개선할 수 있을 것으로 본다. 또한 폐수 처리 외에도 식수 처리에도 동일 작업 적용 중이다.

REFERENCES & ACKNOWLEDGEMENT



References:

- [1] SUEZ Water Treatment Handbook. (2024). SUEZ Degrémont, Paris, France. Updated regularly and accessible at suezwaterhandbook.com
- [2] Therrien J.-D., Nicolai N. & Vanrolleghem P.A. (2020). A critical review of the data pipeline: How wastewater system operation flows from data to intelligence. *Water Science and Technology*, 82, 2613-2634.
- [3] Rieger L., Takács I., Villez K., Siegrist H., Lessard P., Vanrolleghem P. A. & Comeau Y. (2010). Data reconciliation for wastewater treatment plant simulation studies—Planning for high-quality data and typical sources of errors. *Water Environment Research*, 82(5), 426-433.
- [4] Corominas L., Garrido-Baserba M., Villez K., Olsson G., Cortés U. & Poch, M. (2018). Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. *Environmental Modelling & Software*, 106, 89.

Acknowledgement: The authors would like to thank the operators of the participating wastewater treatment plants managed by SUEZ world-wide for access to operational process data.

질문)

1. 1차 침전의 효율성을 비교할 때 주로 구조적 요소나 원수의 차리 때문에 플랜트 간의 큰 차이가 발생함. 침전조의 구조적 요소와 원수의 차이로 인해 침전조가 다르게 작동하는 경우가 많음. 설계 도구가 머신러닝을 통해 최적화된 이후, 침전조의 설계에서의 작거나 큰 최적화의 원인을 데이터로부터 파악할 수 있나요?

→ 기존 플랜트 데이터를 통해 결과를 도출하면 같은 결과를 얻지 못할 수 있음. 현재 시점에서는 유효하지만 3~5년 후에는 다시 작업을 진행해야 할 수도 있음. 기후 변화로 인해 폐수의 성질이 달라질 수 있고, 탄소 질소 인의 비율이 변화할 수도 있음

굳이, 시간에 따라 고정하지 않음. 유럽과 중국의 폐수가 다르기 때문에 해당 프로젝트에서는 다양한 지리적 위치와 데이터를 사용했고 플랜트의 위치와 원수의 차이에 상관없이 설계 도구를 최적화하는 것을 목표로 함(온도와 폐수의 구성 요소를 고려하여 결과를 제공하고 있음)

2. XGB 모델 테스트할 때, 해당 모델을 어떻게 개선할 수 있나요? 예측 정확도를 높이기 위해서는 어떤 것이 필요한가요?

→ 데이터 범위의 부족이 문제였다고 봄. 데이터가 여전히 동일한 종류의 측정 범위 내에 있었기 때문에, 차트 밖의 데이터가 부족했음. 또는 테스트 단계에서 비정상적인 데이터를 사용했을 가능성도 존재

→ 평균 절대 오차의 차이가 크지 않았기 때문에 보다 더 간단한 로지스틱 모델 사용하려 함