ADVANCING GEOSMIN PREDICTIVE MODELS THROUGH CONTINUOUS MONITORING OF THE TER RIVER IN CATALONIA, SPAIN

WATERTOP CONFERENCE

CARMEN ESPINOSA ANGONA

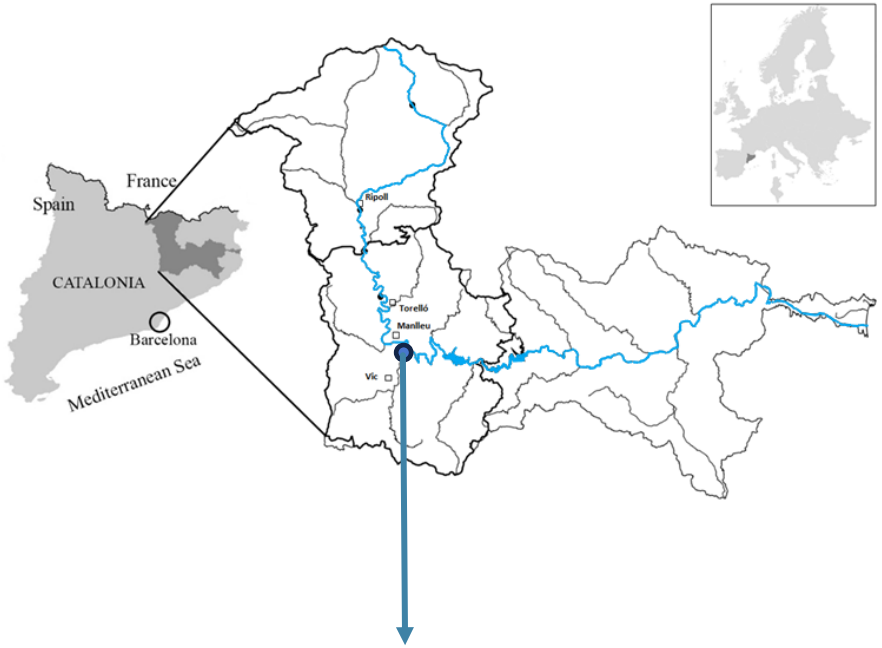
FEBRUARY 22ND, 2024

AVSA
Aigües Vic

CONTEXTUALIZATION – TER RIVER

- Located in the NE of Catalonia
- 208 km-long and 3,010 km^2 of catchment area
- Affected by environmental fluctuations typical of the Mediterranean climate
- Several anthropogenic activities affect water flow and quality:

  (i) small and frequent hydropower weirs
  (ii) livestock farming and intensive agriculture
  (iii) a large reservoirs system



CONTEXTUALIZATION – AIGÜES VIC S.A.

AVSA Aigües Vic | ATCA Abastaments, tractaments i control d'aigües | AVENTEC Enginyeria i tecnologia

73 years of history linked to the integral water cycle in Osona

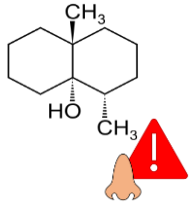
What we do:

- Supply and supply of drinking water in high and low
- Laboratory
- Engineering
- Digital Solutions
- Hydraulic works

CONTEXTUALIZATION – GEOSMIN EPISODES

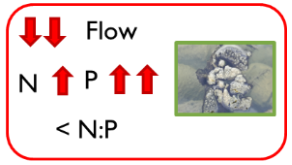
Geosmin: Metabolite produced by cyanobacteria and actinomycetes. It generates bad taste and odor of water.

5 – 10 ng/L



- Difficult to remove with conventional treatments
- High cost
- Difficult direct quantification
- Seasonal occurrence

Anthropogenic pressures + **Climate change**



Drought

Osona.com
Nacióddigital

Dilluns 10 de juliol de 2017

PORTADA OPINIÓ FOTOS AGENDA POLÍTICA SOCIETAT NACI

TORELLÓ

Preocupació a Torelló per l'olor i el gust de l'aigua

- Sorea garanteix que l'aigua subministrada a la ciutat és potable
- Alguns torellonencs avisen de la diferència a través de Twitter

Noticias

Episodi de Geosmina - Març 2016

«Atrás

Imprimir

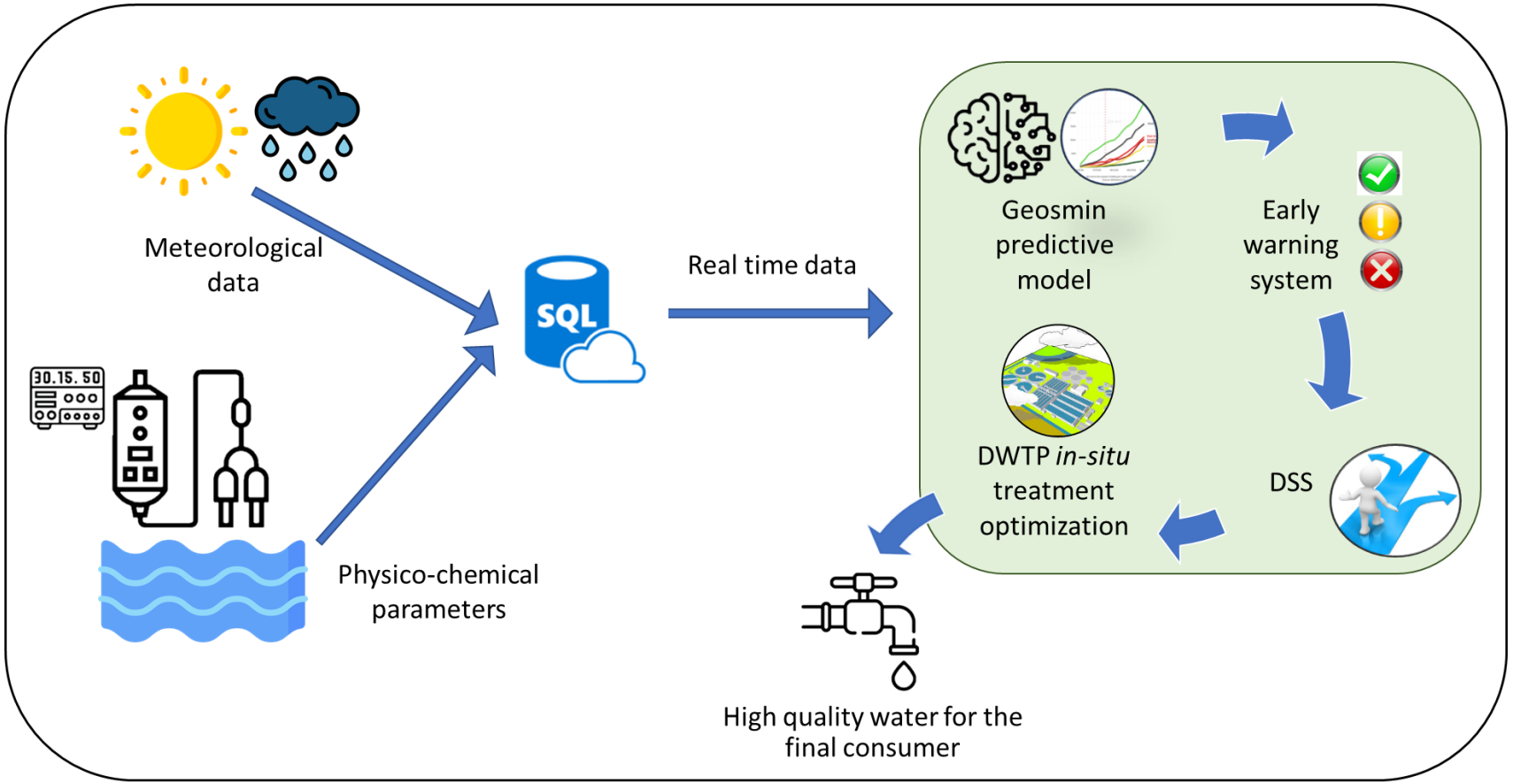
La Geosmina és un compost orgànic característic pel seu aroma terrós que **deixa mal gust a l'aigua però no és perjudicial per la salut**.

La geosmina és produïda per diferents tipus de microbis, incloent-hi cianobacteris (algues blaves) i actionabacteris (especialment streptomyces) que s'alliberen al medi quan aquests moren.

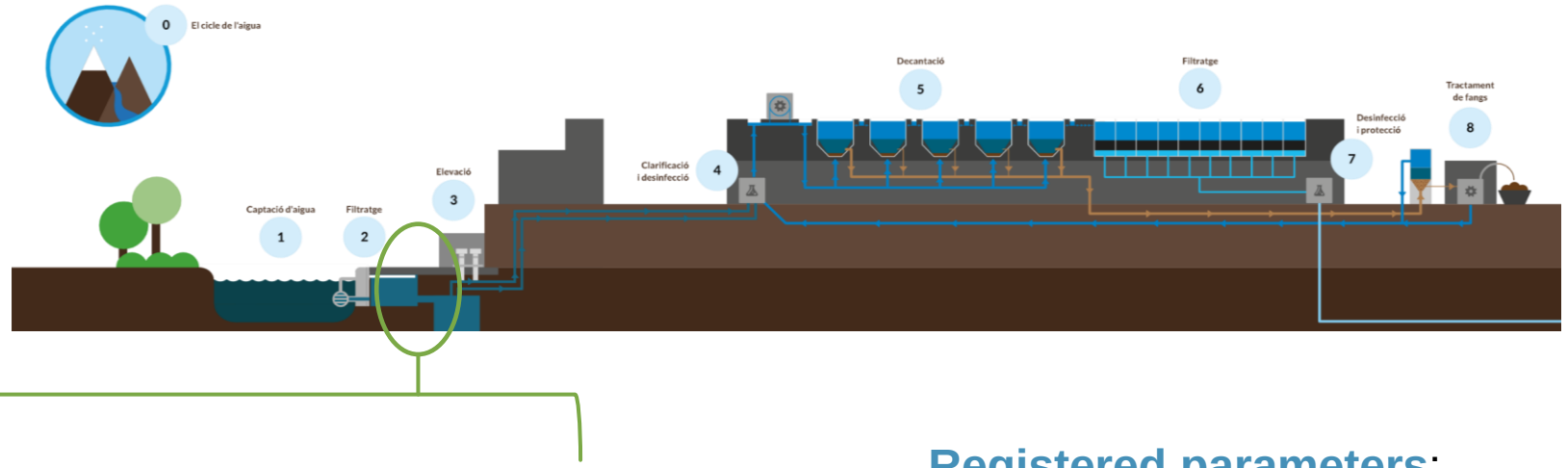
Need for tools capable of predicting geosmin episodes.

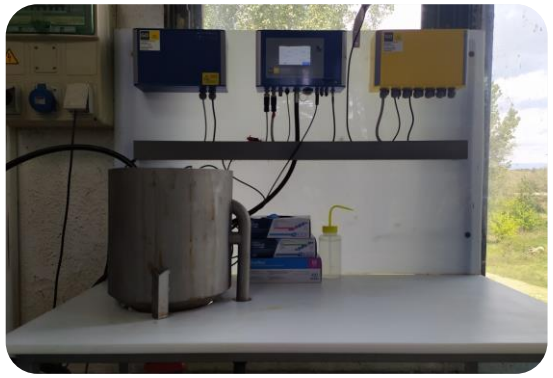
OBJECTIVE

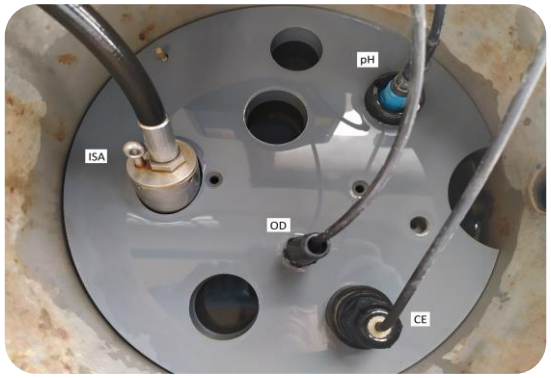
Improving the competitiveness of water treatment plants through the design and implementation of a continuous monitoring device for geosmin prediction and treatment optimization.



MATERIAL AND METHODS





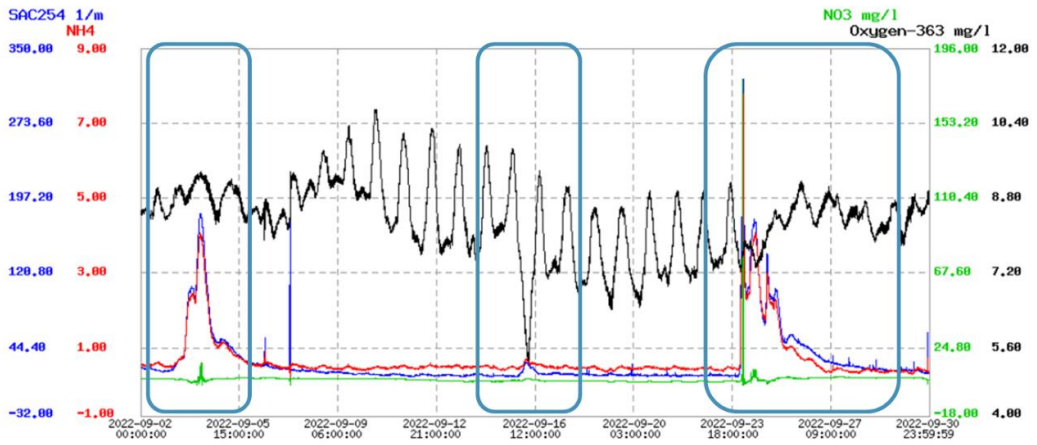


Registered parameters:

- pH
- Electrical conductivity
- Dissolved oxygen
- Temperature
- Nutrient concentration (NO_3^-, NH_4^+, NO_2^-, PO_4^{3-})
- Organic material

MATERIAL AND METHODS

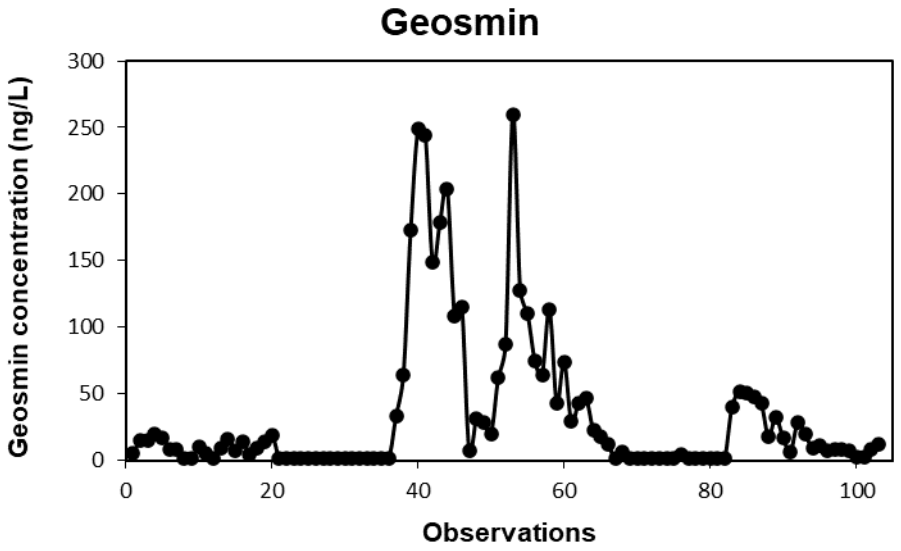
sensor name	unit	date	time	value	status
Conductivity	µS	03.02.2024	11:16:00	**446**	
Temperature	°C	03.02.2024	11:16:00	**7.3**	
NO2	mg/l	03.02.2024	11:11:37	**0**	
Intensity		03.02.2024	11:11:37	**23799**	
SAC254	1/m	03.02.2024	11:11:37	**4.07**	
NO3	mg/l	03.02.2024	11:11:37	**2.9**	
PO4	mg/l	03.02.2024	11:11:37	**0.6**	
NH4		03.02.2024	11:11:37	**0.1**	
Temperature 1-363	°C	03.02.2024	11:16:11	**6.4**	
Oxygen-363	mg/l	03.02.2024	11:16:11	**9.84**	
pH-363		03.02.2024	11:16:49	**8.24**	



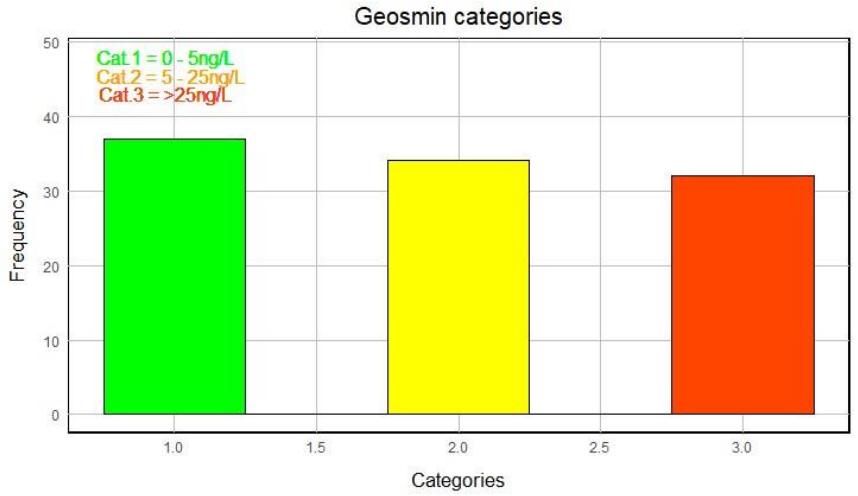
Database:

- Total complete observations: 103
- Variables included in the model
  - Physic-chemical: sensors data, (daily resolution), turbidity, suspended solids (≃ weekly resolution)
    - Calculated physic-chemical: DIN:SRP ratio, DIN:SRP ratio rate
  - Meteorological: Air temperature and Total rain (daily resolution)
    - Calculated meteorological: Temperature rate, Rain rate
  - River water flow (daily resolution)
    - Calculated : river water flow rate
  - Variable response: Geosmin

MATERIAL AND METHODS



Highest peak of geosmina in March.
Max. geosmin concentration = 260 ng/L.
Mean = 33,7 ± 56,3 ng/L.



Definition of 3 categories depending on the risk – action to be carried out:

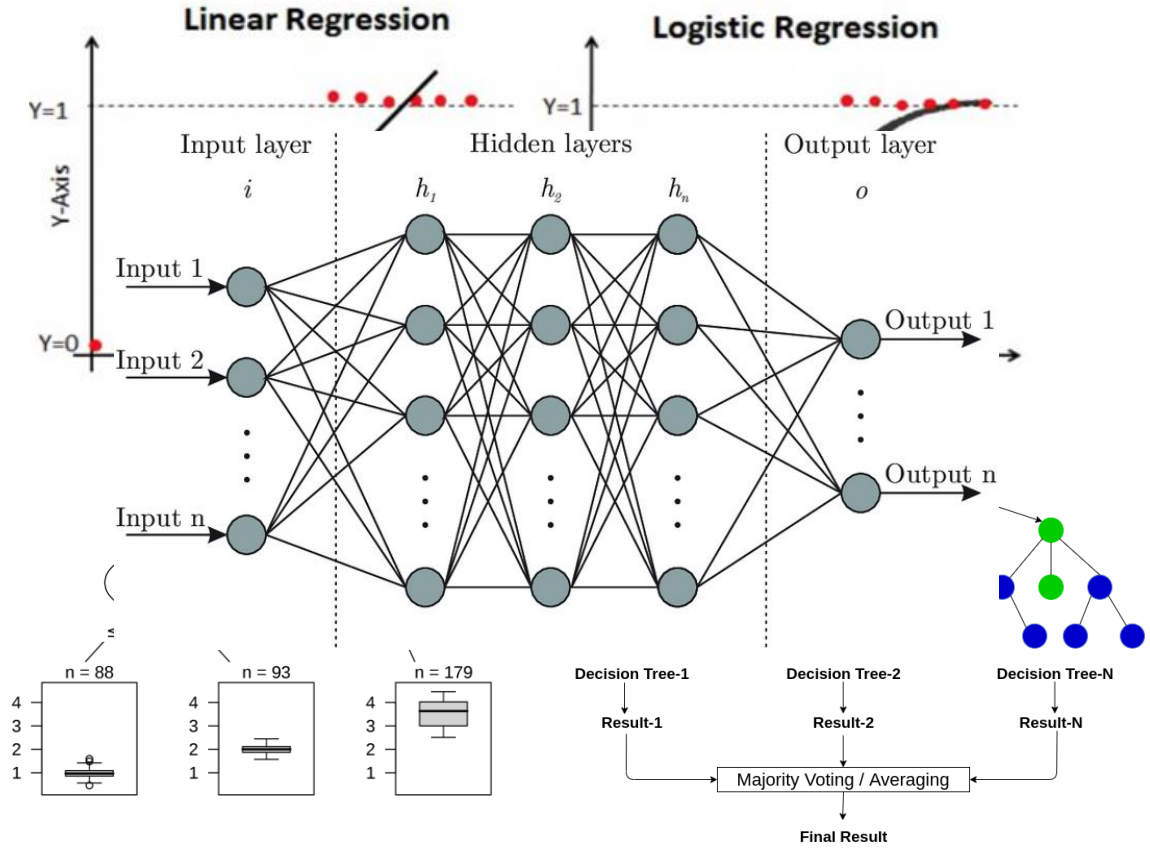
- **Cat. 1** = 0 – 5 ng/L
- **Cat. 2** = 5 – 25 ng/L
- **Cat. 3** = >25 ng/L

MATERIAL AND METHODS

- **Regression models evaluated**

- ✓ Simple and Multiple Linear regression (M/LR)
- ✓ Best subsets
- ✓ Ridge regression
- ✓ Logistic regression (LoR)
- ✓ Least Absolute Shrinkage and Selection Operator (LASSO) regression
- ✓ Elastic net
- ✓ Cross-validation
- ✓ Random Forest (RF)
- ✓ Gradient boosting (GB)
- ✓ Cubist model (CM)

- ✓ Artificial Neural Network (ANN)



MATERIAL AND METHODS

- **Model's options evaluated**
  - Different time lag: t0 (actual) → t2 (+2 sampling days)
  - Different % of train and test dataset: 70/30%, 80/20% and 90/10%
    - Split all the database at the same time
    - First classifying the concentrations in 3 categories (0-5, 5-25 and > 25ngL), making the split by categories and finally joining the 3 training and test databases → to ensure that both the train and the test have the same percentage of observations from each category.
  - Different set.seed values (random database split) → Average, SD and Cumulative SD
  - Geosmin real value and logarithmically transformed
  - Geosmin at time 0 as predictors or not in models at time +1 and +2
  - Geosmin at time -1 and -2 as predictors
  - Calculated variables included or not as predictors
  - Different values for the hyperparameters of the ANN

RESULTS

- **Regression models performance**

R^2 adj.: multiple det. coefficient

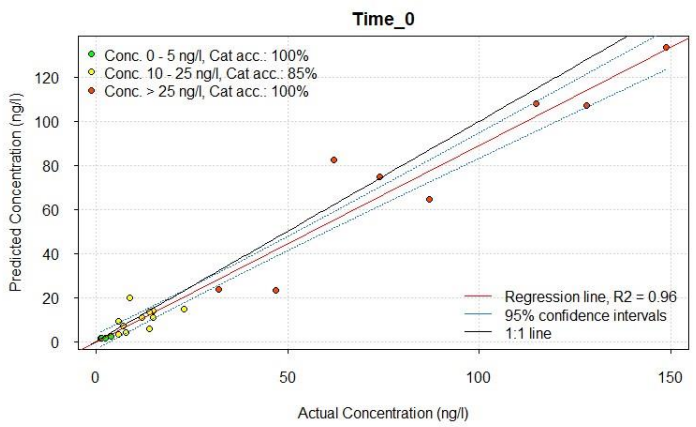
RMSE: Root Mean Square Error

- Best option:
  Logarithmically transformed, including calculated variables as predictors, split 70/30%

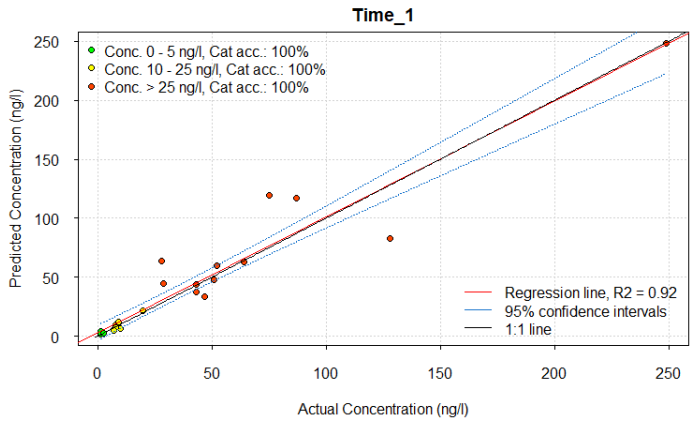
	GEOSMIN	
	R^2 adj	**RMSE**
MLR	0.358	0.246
Best subsets	0.300	0.326
Ridge regression	0.281	0.321
LASSO	0.255	0.334
Elastic net	0.288	0.314
Cross-validation	0.257	0.366
Random Forest	0.637	0.163
Gradient boosting	0.537	0.237
Cubist model	0.611	0.239
ANN	0.131	0.325

RESULTS

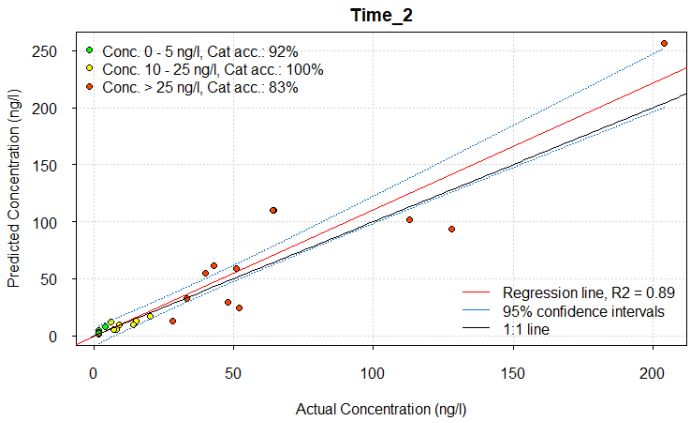
Two-tier modelling results



R^2 = 0.96
Accuracy by category: **85 – 100%**
Model: RF + Deeplearning



R^2 = 0.92
Accuracy by category: **100%** (overestimates cat. 3)
Model: RF + Stacked Ensemble



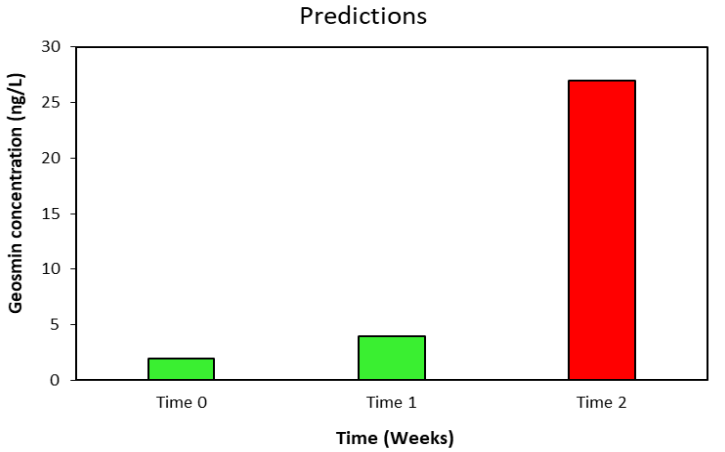
R^2 = 0.89
Accuracy by category: **83 - 100%** (general tendency of overestimation)
Model: RF + Deeplearning

SOLUTION

Application of the model to a new observation

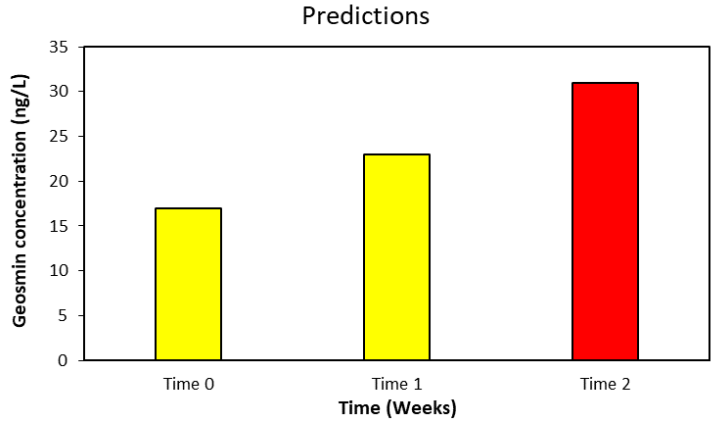
A

Time (Weeks)	Conc. (ng/L)	Category	Recommendations
0	2	1	Review necessary treatment
1	4	1	Notice to the population
2	27	3	PAC application (Y dose)



B

Time (Weeks)	Conc. (ng/L)	Category	Recommendations
0	17	2	Control of geosmin conc. at the DWTP outlet
1	23	2	PAC application (X dose)
2	31	3	PAC application (Y dose)



CONCLUSIONS AND NEXT STEPS

- ✓ Continuous and real-time information is available on the water quality of the Ter River.
- ✓ The predictive models developed have high precision for the three times evaluated.
- ✓ AVSA is capable of anticipating geosmin episodes, which are intensified by drought.
- ✓ During the last year of the project, the validation and improvement of the predictive models will be carried out.
- ✓ The decision-making protocol of the water treatment plant will also be optimized against the results predicted by the models.

THANKS A LOT FOR YOUR ATTENTION

CARMEN ESPINOSA ANGONA

cespinosa@aiguesvic.com / c.e.angona@gmail.com

AVSA
Aigües Vic