AI to support water management

Ms Yuliya Vystavna

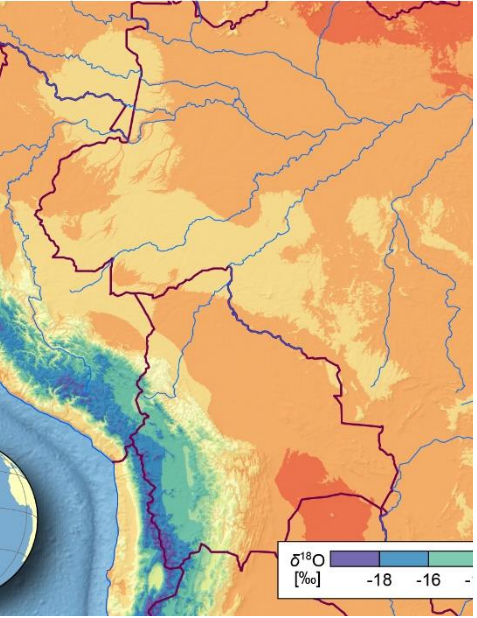
y.vystavna@iaea.org
Department of Nuclear Sciences and Applications
Division Physical and Chemical Sciences
Isotope Hydrology Section

The Isotope Hydrology Section (IHS)

- IAEA Director General
  - Secretariat of the Policy making Organs
  - Office of Public Information and Communication
  - Office of Internal Oversight Services
  - Office of Legal Affairs
  - Department of Technical Cooperation
  - Department of Nuclear Energy
  - Department of Nuclear Safety and Security
  - Department of Management
  - Department of Nuclear Sciences and Applications
    - Div. Physical and Chemical Sciences
      - Isotope Hydrology Section inc. Isotope Hydrology Laboratory
  - Department of Safeguards

- About 30 staff, incl. professional posts, laboratory technicians and support staff.
- IHS is responsible for the design and implementation of all activities of the IAEA's Water Resources Programme.

Objectives: to provide Member States with science-based information and technical skills on isotope hydrology that will help them to better assess and manage their water resources



Isotope Hydrology at the IAEA

Coordinated Research Projects
Development of new field and lab methods to assess water in rivers, lakes, and aquifers for drinking, irrigation, and industrial use.

Global Monitoring Networks
IAEA global isotope monitoring networks for hydrological, climate, food and environmental data and studies. GNIP (precipitation) and GNIR (rivers)

Capacity Building and Training.
Data synthesis and publications; teaching and training materials both in person and online (e-courses).

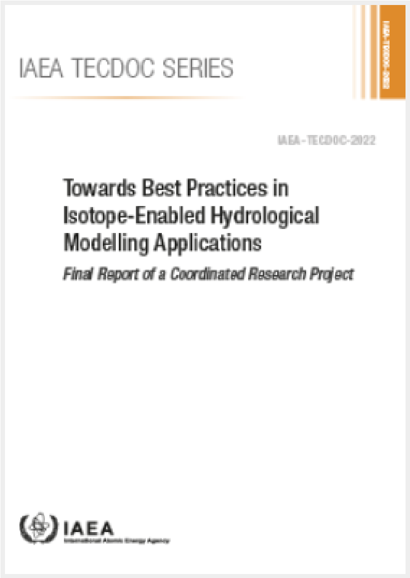
Technical Cooperation Projects
Provide technical support to TC Dept for equipment, experts, and training courses both in person and virtually.

Laboratory Support for Member States
Provides direct laboratory support to Member States, through a variety of mechanisms, including CRPs, TC projects and training courses.

IAEA

Guidelines on modelling

Final Report of a Coordinated Research Project



IAEA-TECDOC-2022

English IAEA-TECDOC-2022 ¦ 978-92-0-119023-9

52 pages ¦ 13 figures ¦ € 22.00 ¦ Date published: 2023

Download PDF (2.17 MB)

Get citation details

https://www.iaea.org/publications/15290/towards-best-practices-in-isotope-enabled-hydrological-modelling-applications

National Water Analysis Capacity
Achieving SDG6

Functioning and effective national water laboratories are an essential component of any resilient social system and underpins the achievement of SDG6 in a variety of ways including:

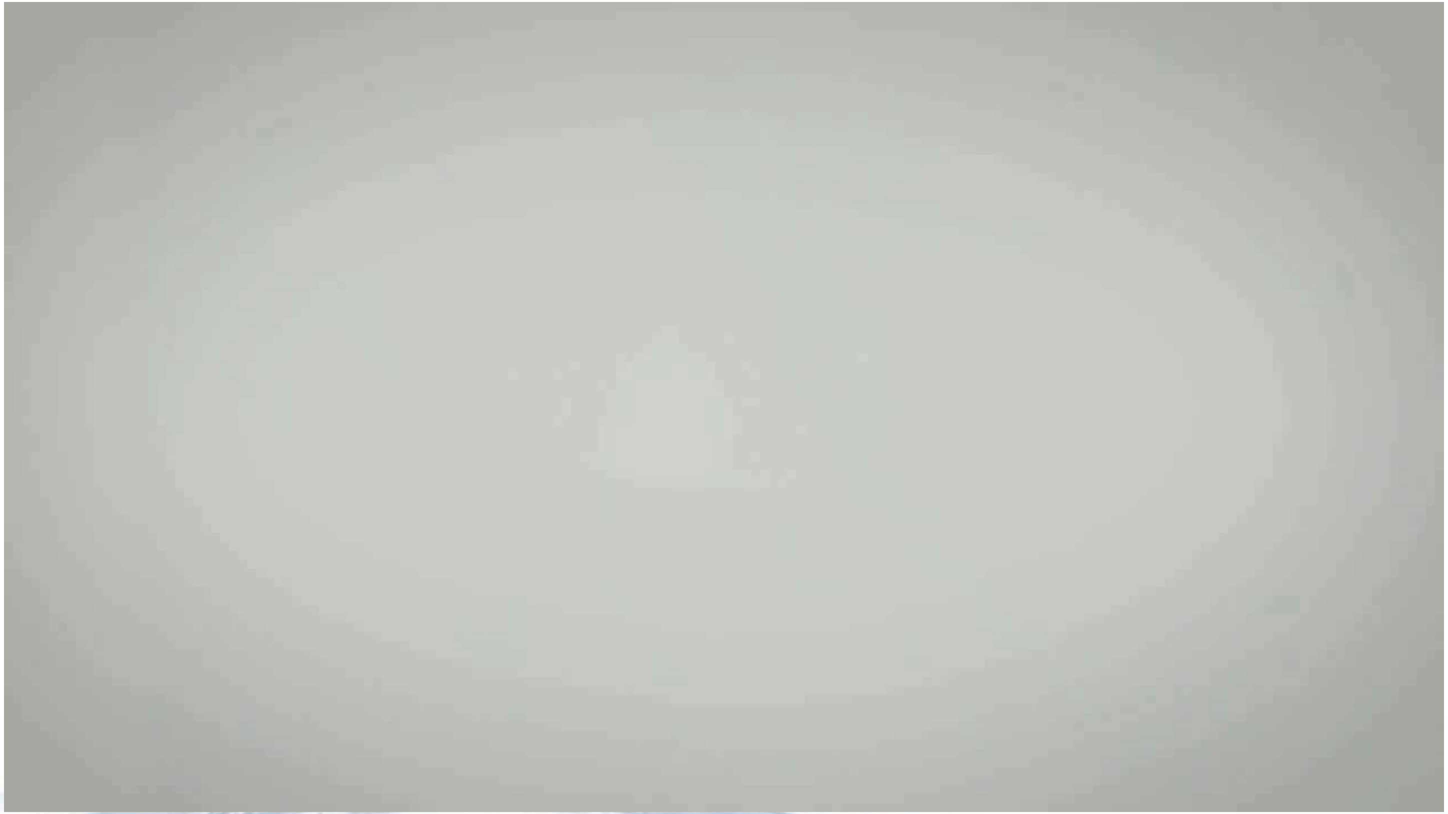
- Providing a specific focus for **capacity development** actions and activities both scientific and technical
- Allowing countries **to generate the necessary data** and information needed to properly manage both surface water and groundwater resources
- Creating **opportunities for innovation** in the way in which water is evaluated and managed
- **Supporting the legislative, regulatory and policy frameworks** necessary for effective water governance.

These actions are consistent with the five accelerators of SDG6 that have been identified - these are finance, governance, innovation, data and information and capacity development.

GLOWAL Network of Water Analysis Laboratories

https://www.iaea.org/services/networks/glowal

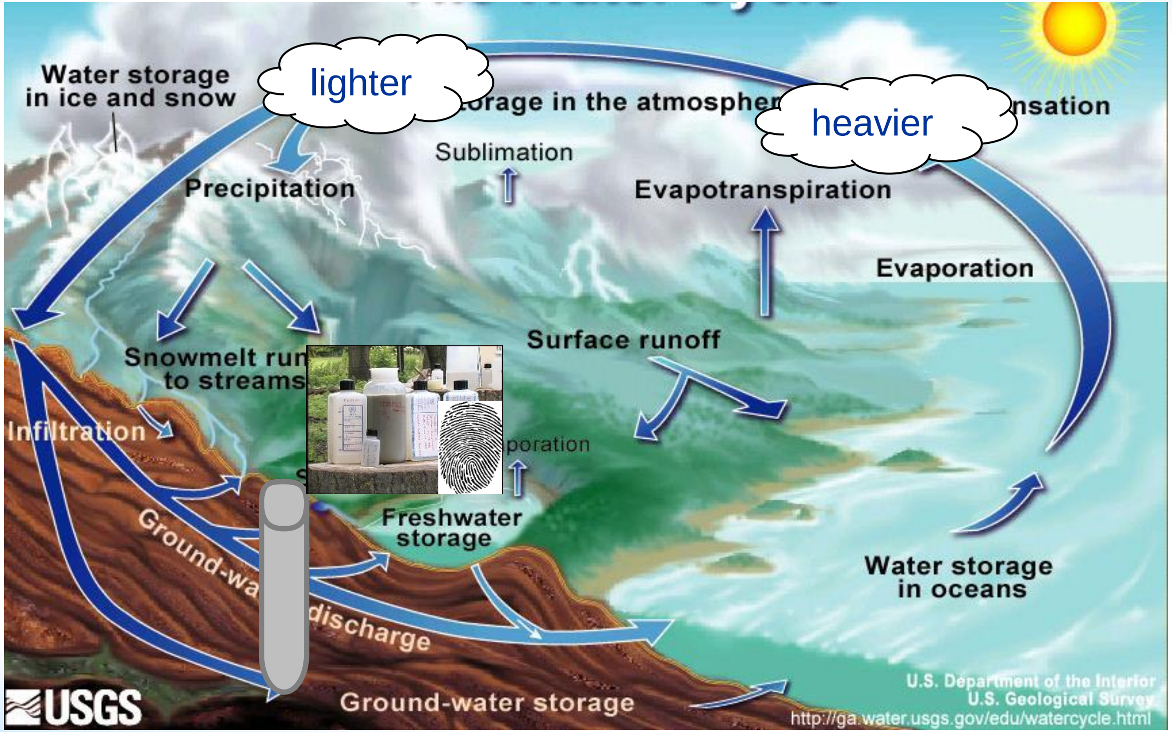
How do isotopes help understand the water cycle?



Short video

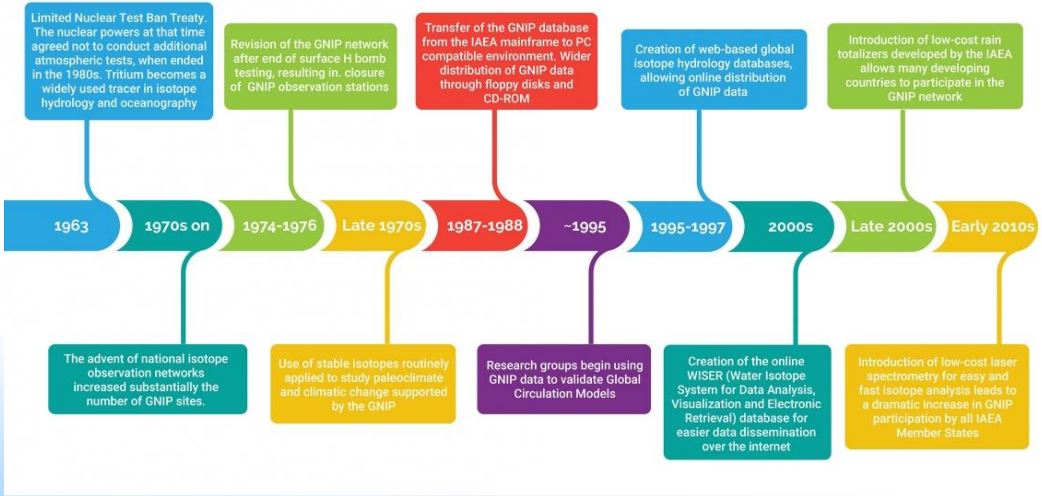
Isotope Hydrology

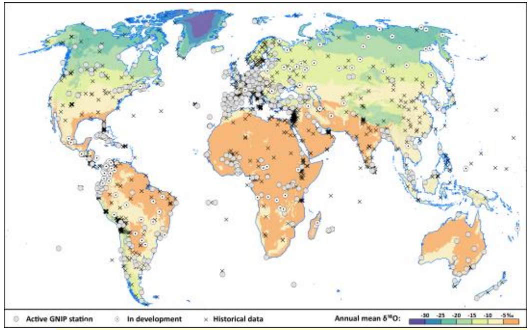
Stable (^{16}O, ^{17}O, ^{18}O, ^{1}H, ^{2}H) and radioactive (^{3}H) water isotopes are ***powerful tools to track water molecules throughout the hydrological cycle***



GNIP – 60 years

- Started 1960 by the IAEA in cooperation with the WMO, has become the world's largest and most comprehensive collection of isotope data in atmospheric waters.
- ~ 140 K isotope records from ~1100 stations (350 active) in collaboration with many contributors.

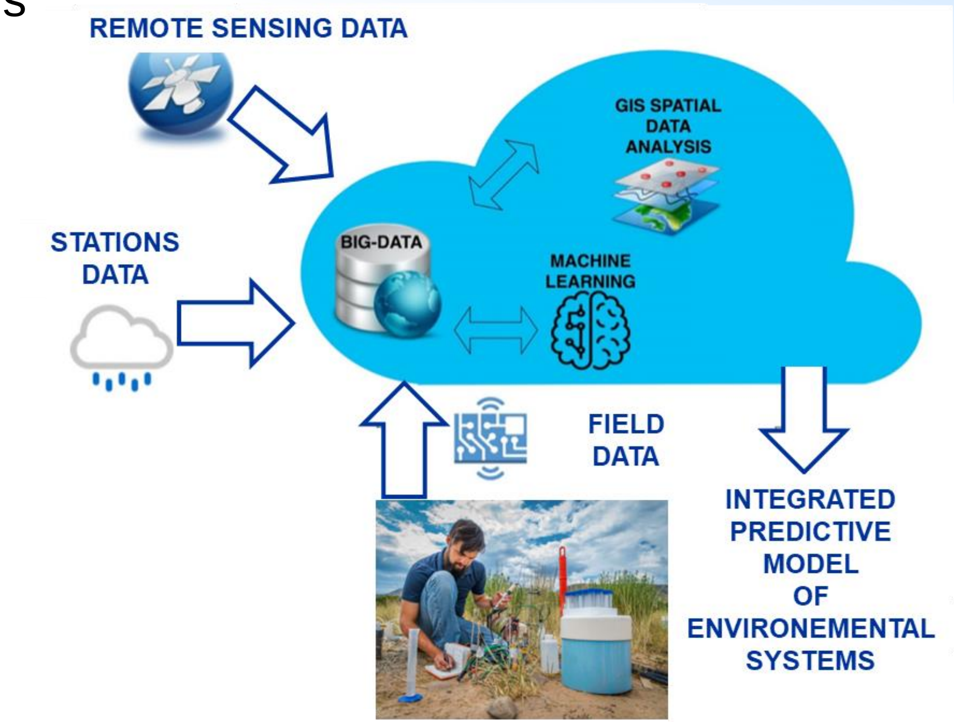




Isotopes in Precipitation Network (GNIP) for water and climate

Challenges

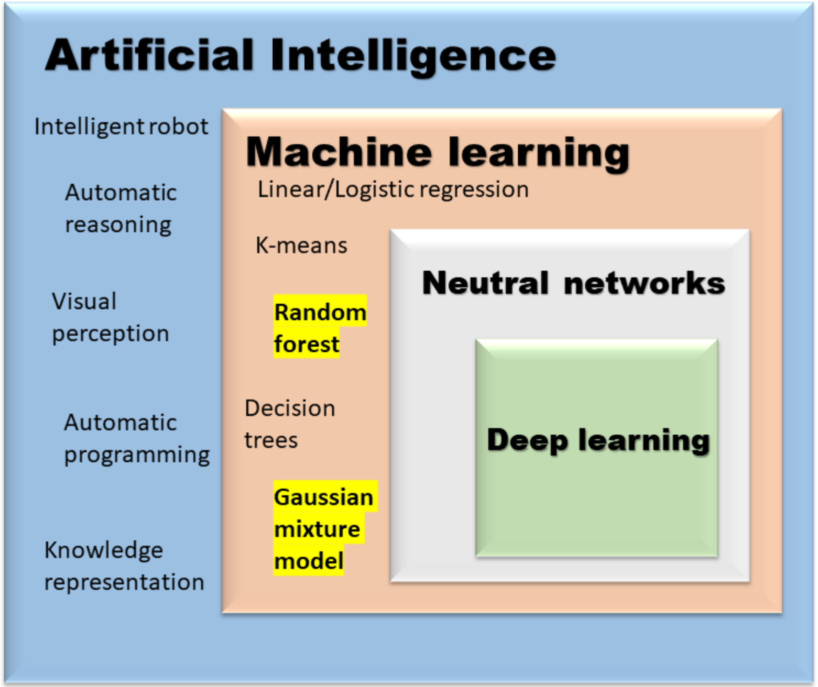
- Long-term systematic observations
- Common isotope data base
- Interdisciplinary approach
- Data quality
- High frequency – spatial and temporal



Presentation by C. Bataille

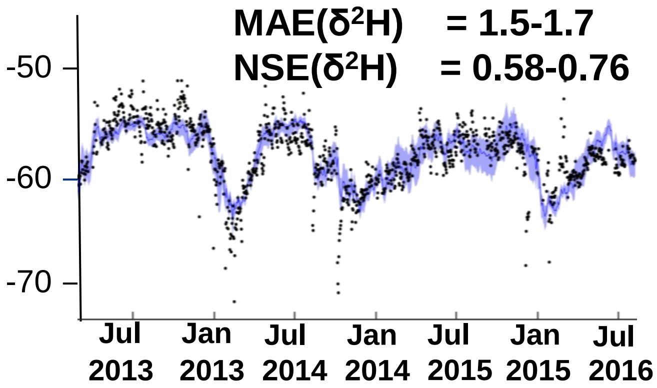
Machine learning approaches applied in isotope data

- Support vector algorithm (SVM)
- Genetic Programming (GP)
- Neural networks
- Long Short-Term Memory (LSTM) - Improved stream isotope simulations (**Time Series**)
- Multiple models combined to Super ensemble learner
- **Random Forest (RF) – spatial interpolation, high frequency data, explainable AI**

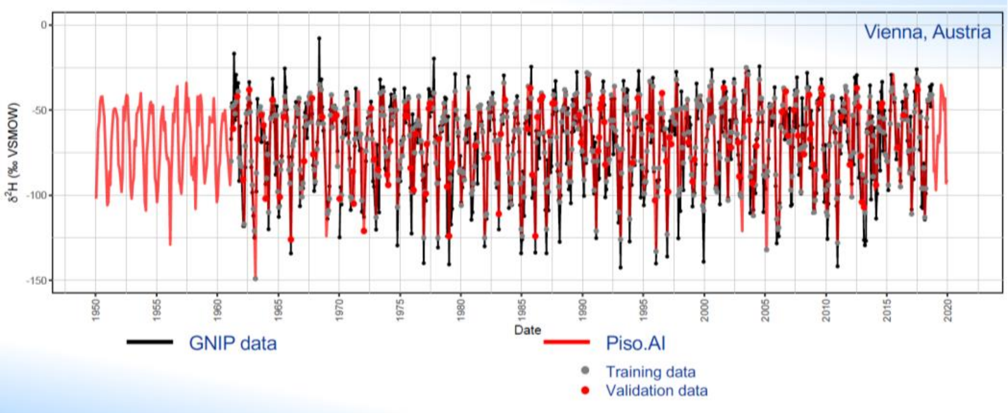


Time series

LSTM on stream isotopes



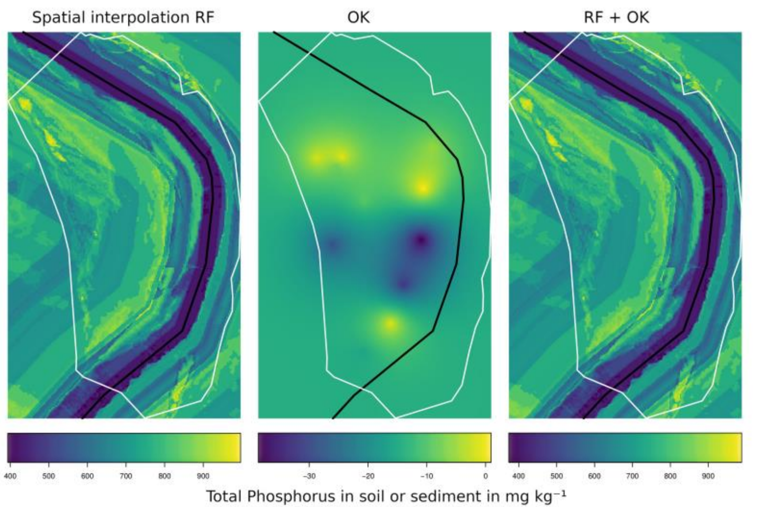
ML on GNIP data to provide time series for Europe

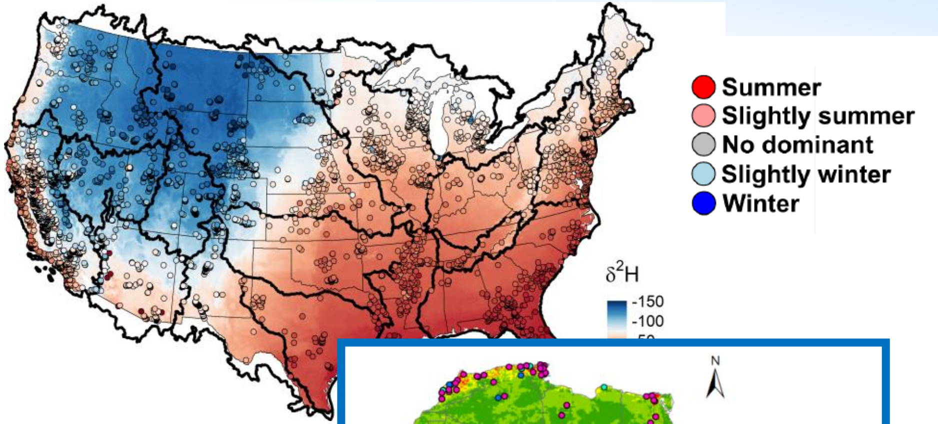


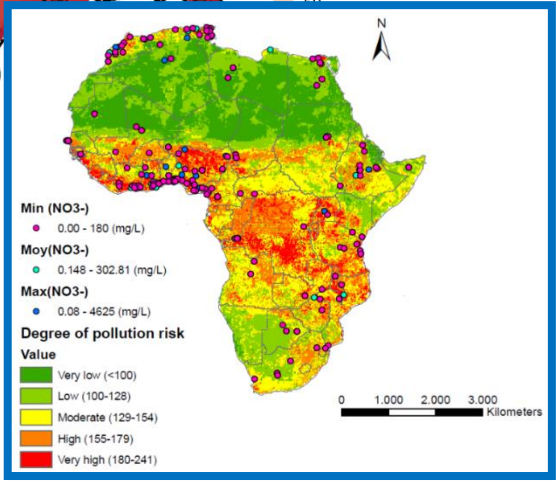
Results outperforming physical-based models and classical statistical methods.

Presentations by A. Smith and D. Nelson

Random forest spatial interpolation







Coverage of the scarce data areas and seasons

Presentations by M. Pucher, M. Stahl and Y. Jameel, Issoufou

Further Opportunities and Limitations

Opportunities

- Aid in spatially (and temporally) refining datasets
- Gap-fill and extend existing time-series
- Use to identify key regions → providing further hydrologic insights
- Gain insights into ungauged areas with more consistency
- Potential for large-scale models

Limitations and further outlooks

- Require robust datasets – <u>availability, quality, and quantity of data limitations</u>
- Potential for over-fit of highly uncertain data
- Careful use ***necessary***
- Predictive capability versus interpretability
- Difficult to assess predictive capability beyond the range of training and validation data set (as opposed to physical models)
- Need to hybridize traditional hydrology modelling with ML
- Continuation of information gained from machine learning (e.g. hidden states)

We will focus on two main approaches

Explanatory powers of AI

× explain processes and find determinants

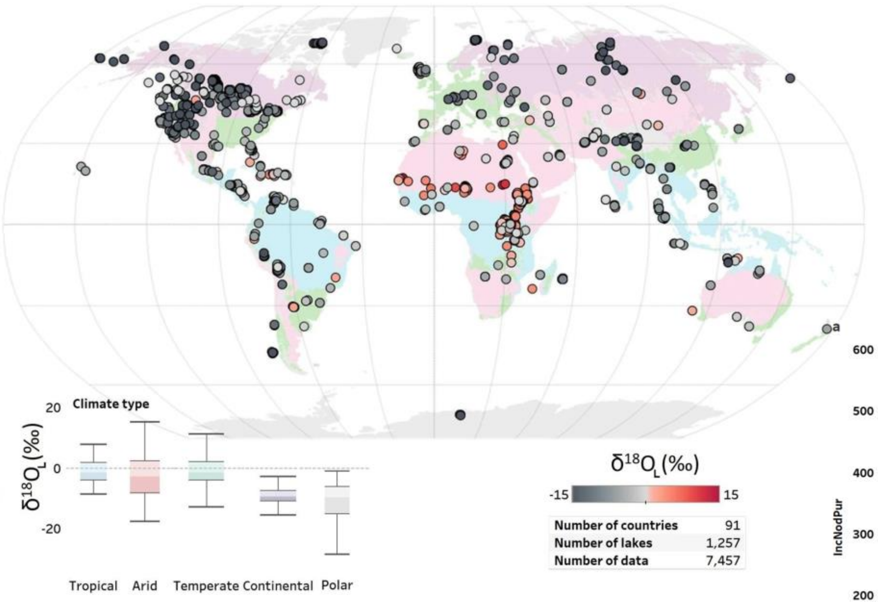
Predictive powers of AI

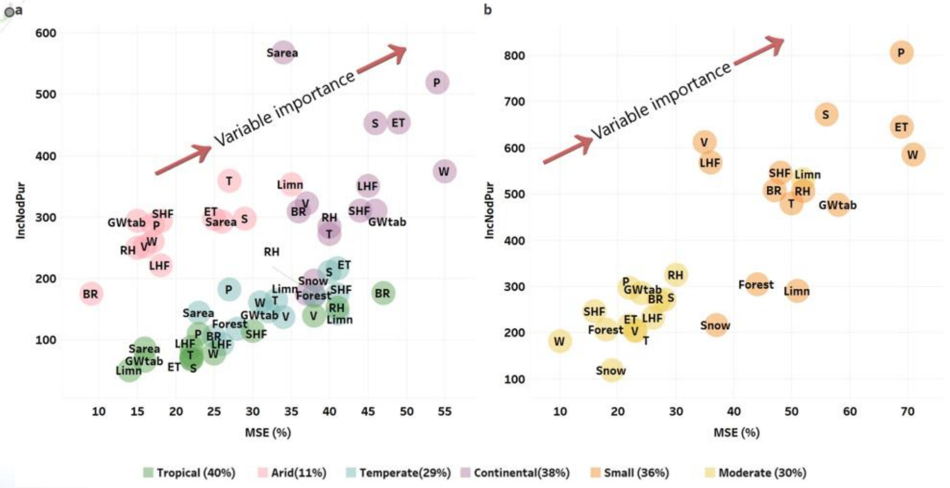
× find determinants and predict processes

Explanatory power of AI model: why lakes are disappearing?

Lakes are sensitive to climate change

- What factors impact lakes water balance?
- Which lake is the most sensitive?



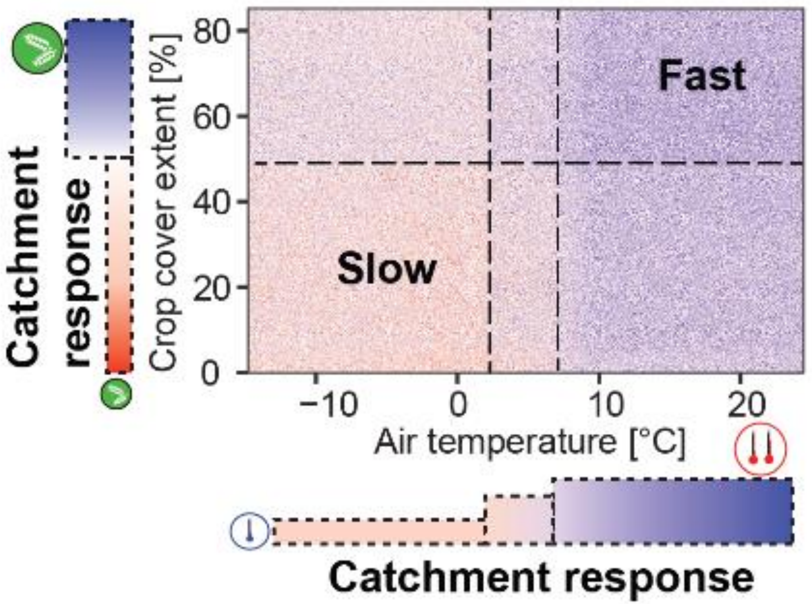


Vystavna et al. 2021 Nature Communication

Explanatory power of AI model: why lakes are disappearing? How it supports water management?

- Give the priority to water users
- Identify the evaporation hotspots
- Determine factors that can be "managed"
- Develop lakes restoration strategy

Prediction power of AI model: how catchments response on flood and drought



River catchments response on climate and land use changes differently:

- Fast response – flash floods and emerging droughts
- How interaction between land use impacts catchment response?
- Predict how future changes in climate and land use will impact the catchment response

Vystavna et al. In review Nature Water

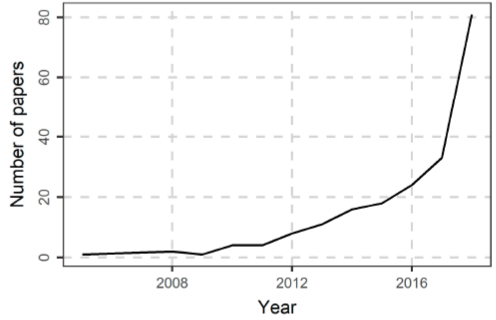
Prediction power of AI model: how catchments response on flood and drought? How to support water management?

- Identify the most responsive catchments (early warning systems/gauging)
- Manage land use to control your catchment response
- Determine factors impacting catchment response
- Develop river/catchment restoration strategy

What will we do today?

- Construct and simulate AI explanatory model
- Overview AI prediction model

➡ WE WILL GO IN THE FOREST



Random Forest in Water Sciences (published papers)

RANDOM FOREST

Machine learning algorithm used for classification and regression

Applied as supervised and data-driven machine learning algorithm

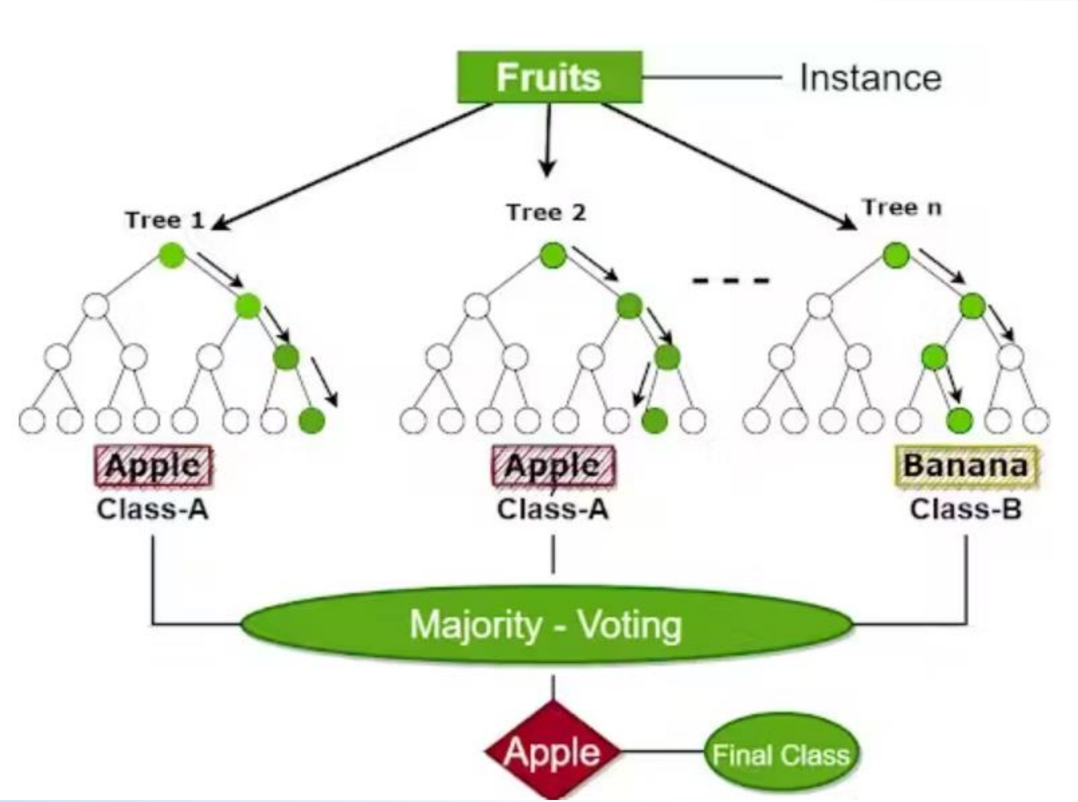
(you are training AI on data sets and control tuning of parameters)

AI model performance depends on tuning of parameters and

selected variables (data sets):

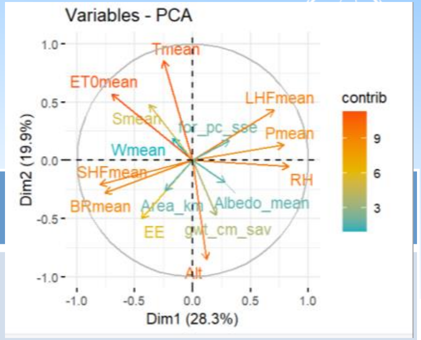
(i) number of trees,
(ii) number of possible directions for splitting at each node of each tree
(iii) number of examples in each cell, below which the cell is not split

RANDOM FOREST VISUALIZATION



- dependent variable
- respective observations of the predictor variables

RANDOM FOREST vs. PCA



RANDOM FOREST	PCA
Linear and non-linear regression, non-parametric	Only linear regression
Needs skills	Easier
Needs tuning	Doesn’t need tuning
Has explanatory and prediction power	Has mainly explanatory power
Needs larger database	Can be perform in smaller database
Visualization capacity	Visualization capacity
Performance determined by many parameters	Performance determined by several parameters
No overfitting and autocorrelation	Autocorrelation

Our tasks

- Download R
- Get R script and the sampled database
- Run R random forest model
- Explain results