IAEA
International Atomic Energy Agency

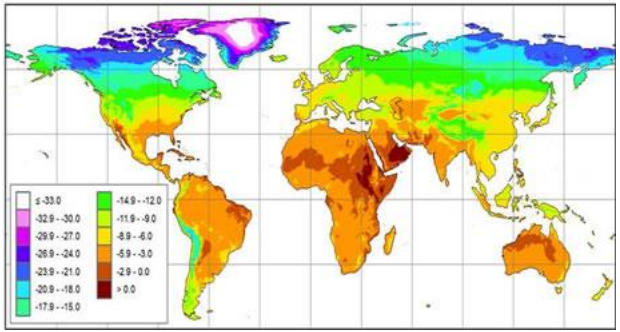
Summary of the Working Group on AI for Water and Environment

David Soto, Astrid Harjung, Yuliya Vystavna, Oliver Kracht

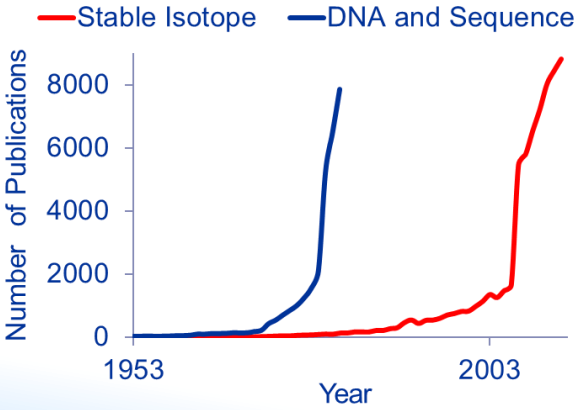
Technical Meeting on Artificial Intelligence for Nuclear Technology and Applications
#AI4Atoms Virtual Event
25–29 October 2021

State of the Art

- Stable (^{16}O, ^{17}O, ^{18}O, ^{1}H, ^{2}H) and radioactive (^{3}H) water isotopes are known as ***powerful tools to track the path of water molecules throughout the hydrological cycle***, and…
- The GNIP (IAEA/WMO) is world's largest and most comprehensive collection of isotope data in atmospheric waters since 1960s (140,000 isotope records with a large potential to be used in AI)



1. Discovery (1953 – 1990's)
2. Integration (1990's – 2020)
3. Innovation (2020 –)



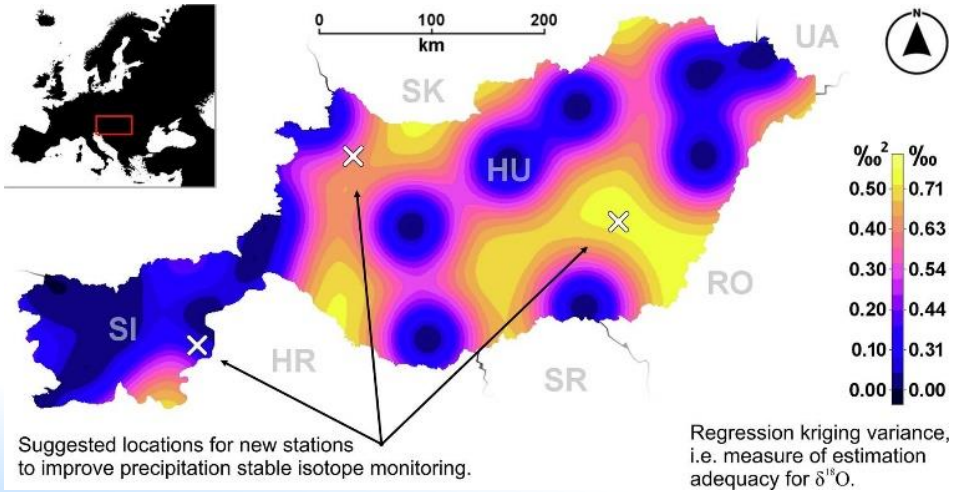
Data repository, which includes:

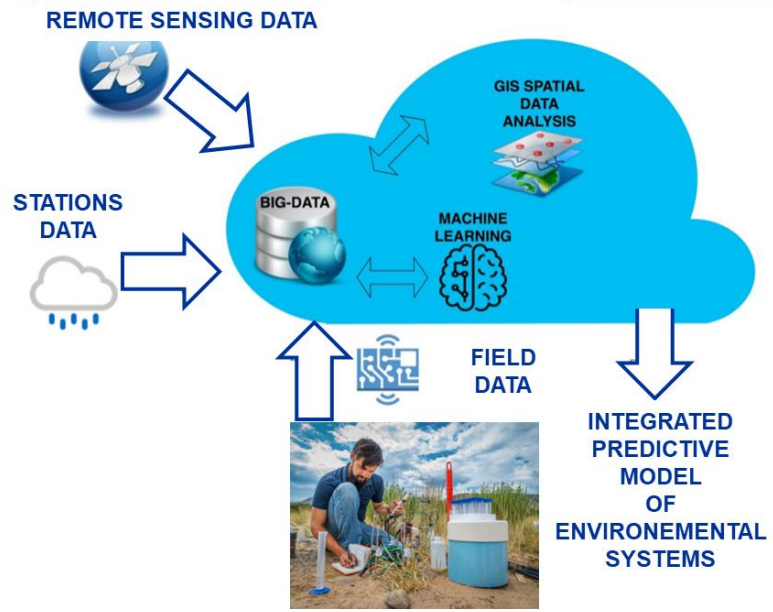
- **Source attribution**
- **Open access**
- **Longevity**
- **Sustainability**

Presentations by G. Bowen and C. Bataille

Current Challenges

- Long-term systematic observations
- Common isotope data base
- Interdisciplinary approach
- Use of best praxis including artificial intelligence





From presentations by G. Bowen, P. Vreca and C. Bataille

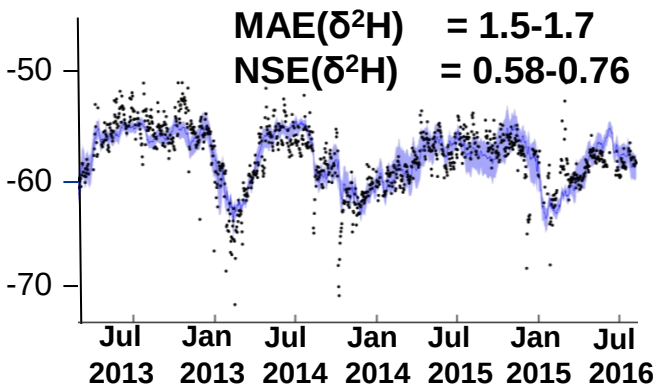
AI / ML approaches applied in isotope data

- 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 and high frequency data**

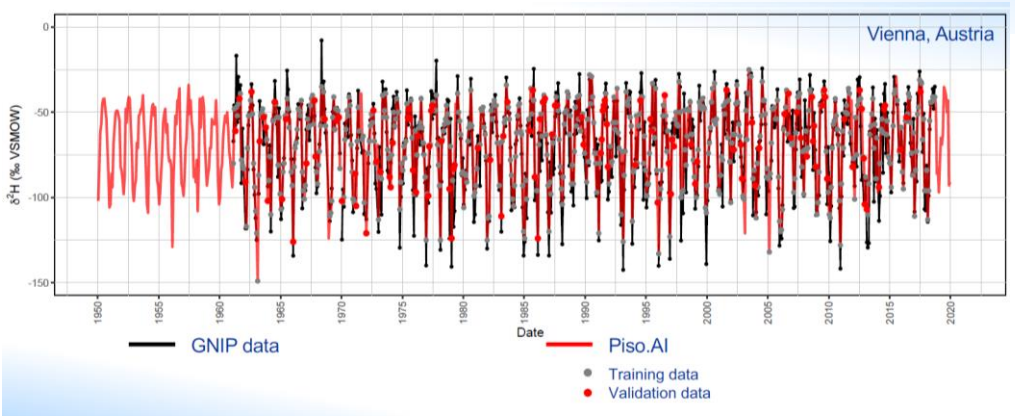
All presenters

Time series

LSTM on stream isotopes



ML on GNIP data to provide time series for Europe

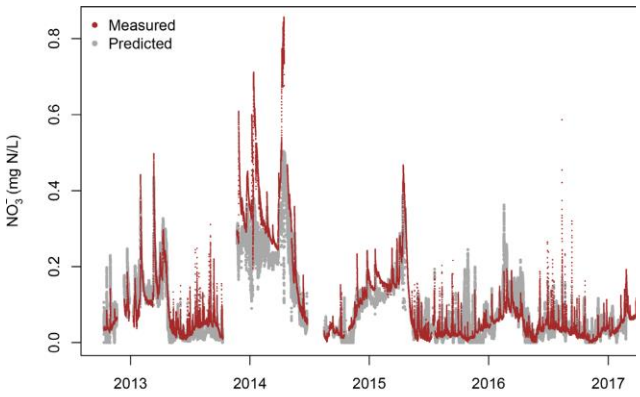


Results outperforming physical-based models and classical statistical methods.

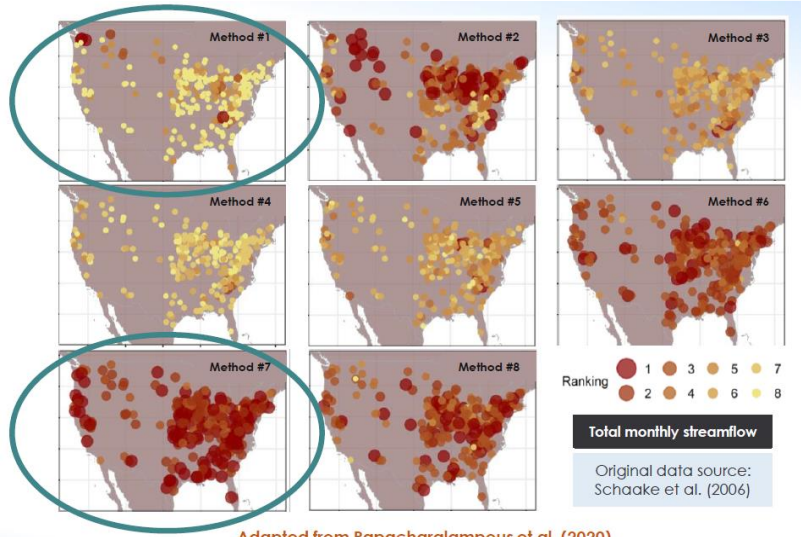
Presentations by A. Smith and D. Nelson

Time series

High frequency data



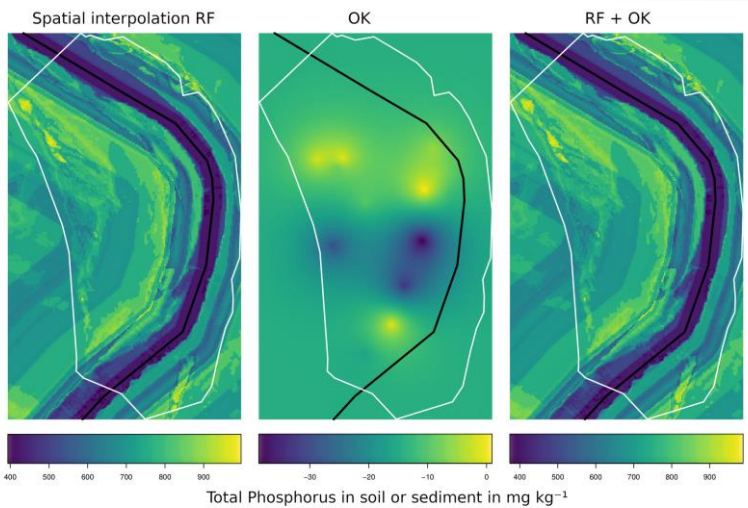
Forecasting

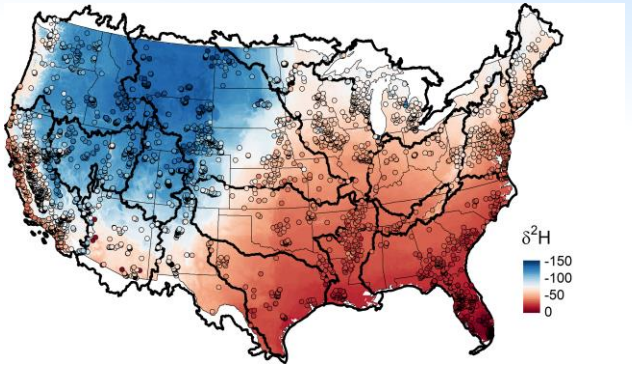


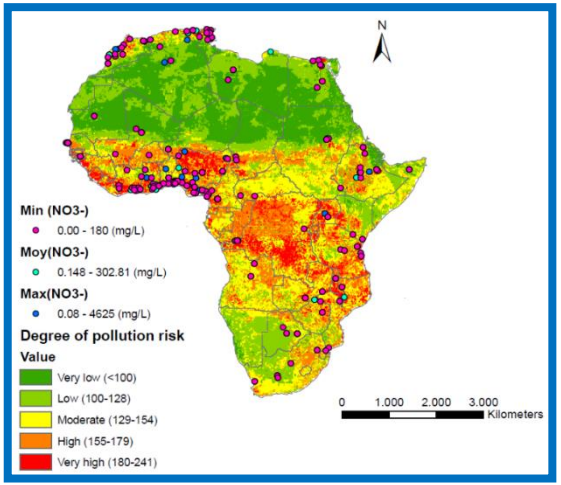
Adapted from Papacharalampous et al. (2020)

Presentations by G. Papacharalampous & H. Tyralis and M. Green

Random forest spatial interpolation







Coverage of the scarce data areas and seasons

Presentations by M. Pucher, M. Stahl, I. Ouedraogo and Y. Jameel

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 – availability, quality, and quantity of data limitations
- Potential for over-fit of highly uncertain data
- 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)

Accelerating Progress—IAEA's Role

Recommendations from experts:

- The use of a **central database** can support high quality data easily.
- Strengthen the role of organizing the **database structure and data contribution** – how to attribute data providers in times of BigData
- Facilitate **data sharing** among experts and stakeholders
- **Coordination** of databases and networks – permanent competence team is needed
- Guiding and organising efforts to build databases and networks (including quality control); i.e., manuals, guidance, best practices, etc.

Accelerating Progress—IAEA's Role

Recommendations from experts:

- Laboratories and data users need to be aware of **Quality Control** – assist "new" labs with a set of guidelines and standards as starter package.
- Advertise **data quality activities** to labs producing isotope data and groups providing isotope databases
- **Consortium** of people allocated from IAEA and member states should participate on a longer time frame working towards including datasets
- **Training needs**: integrated training -> process-based understanding with strong numerical/theoretical skills

Expected Outcomes

- Consolidation and expansion of global network activities IAEA (GNIP, GNIR) in active response to novel Big-Data challenges
- Potential CRP:
  - **ML to apply across the groundwater-soil-plant-atmosphere continuum**
  - **ML to expand knowledge to ungauged basins**
  - **ML to understand data and processes under extreme climate situations (droughts, floods)**
- Continuation of Quality Control of isotope analysis in MS labs (water, nitrate)
- Potential contributions from experts to Special Issue on using AI to advance understanding of hydrological processes (opinion paper and guidelines)

IAEA
International Atomic Energy Agency

Thank you!

Plenary: Ms Dörthe Tetzlaff
Participants (in order of presentation):

- Ms Jodie Miller
- Mr Gabriel Bowen
- Ms Polona Vreča
- Mr Clement Bataille
- Mr Aaron Smith
- Ms Georgia Papacharalampous
- Mr Hristos Tyralis
- Mr Daniel Nelson
- Mr Matthias Pucher
- Mr Mason Stahl
- Mr Yusuf Jameel
- Mr Issoufou Ouedraogo
- Mr Joel Podgorski
- Mr Mark Green
- Mr Juan Antonio Torres-Martínez
- Mr Alexander Mhizha

16 experts from 12 countries