IAEA
International Atomic Energy Agency

Machine learning concepts and methods for hydrological post-processing and forecasting (invited)

Georgia Papacharalampous[1,*], Hristos Tyralis[2,3]

[1] Department of Engineering, Roma Tre University, Via V. Volterra 62, 00146 Rome, Italy

[2] Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Polytechniou 5, 15780 Zographou, Greece

[3] Hellenic Air Force General Staff, Hellenic Air Force, Mesogion Avenue 227–231, 15561 Cholargos, Greece

[*] Correspondence: papacharalampous.georgia@gmail.com

Technical Meeting on Artificial Intelligence for Nuclear Technology and Applications

#AI4Atoms Virtual Event

25–29 October 2021

Introduction

- Forecasting of hydrometeorological and hydroclimatic variables (hereafter referred to as "**hydrological forecasting**") is performed at various temporal scales and horizons according to the requirements of technical frameworks.
- These frameworks support a variety of **engineering and environmental services**; therefore, achieving improvements (e.g., in terms of accuracy) in hydrological forecasting leads to various **societal and environmental benefits**.

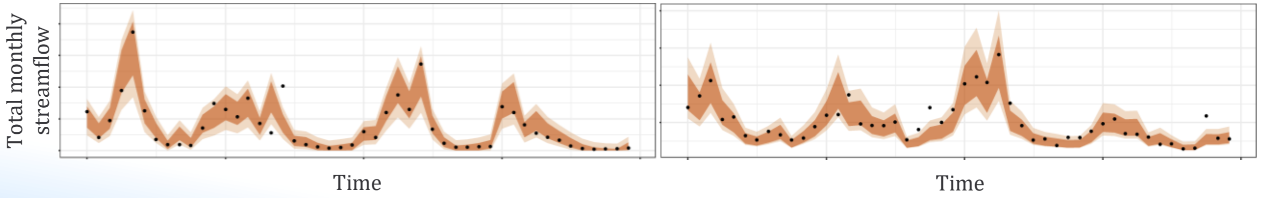
Introduction

- Hydrological forecasting experiments help us understand **how forecastable** the various hydrometeorological and hydroclimatic variables are and **why**.

Introduction

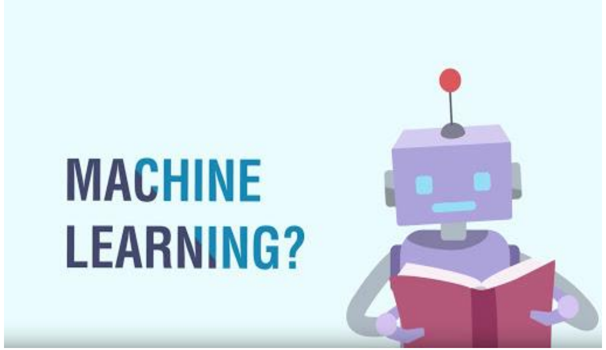
- The various **hydrological forecasting frameworks** can take very different forms depending on (a) some general requirements for the output, (b) the targeted variable, (c) the temporal scale and (d) the horizon, among others.
- Much high-quality work has been conducted so far towards:
  - ✓ **Proposing** such frameworks and their more general blueprints;
  - ✓ **Improving** such frameworks (e.g., in terms of reliability or applicability);
  - ✓ **Adapting** such frameworks (e.g., by adding new components to them) for meeting new requirements.
- Characteristic examples of **adapted hydrological forecasting frameworks** are those relying on (i) process-based catchment models, (ii) meteorological or climatological forecasts, and (iii) **hydrological post-processing** for issuing probabilistic instead of mean-value streamflow forecasts.

Examples of probabilistic forecasts obtained through hydrological post-processing



Introduction

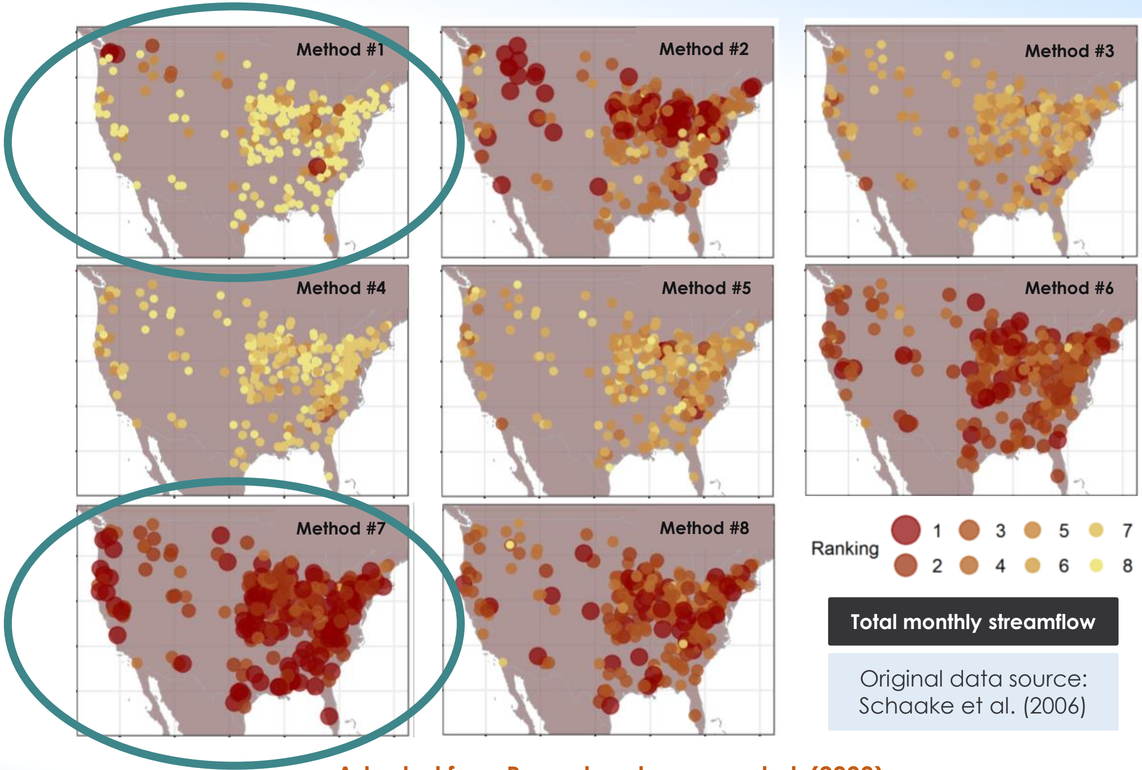
- ***What about machine learning?*** Machine learning methods (see, e.g., the comprehensive lists and descriptions provided by Hastie et al. 2009; James et al. 2013) are increasingly investigated for hydrological forecasting (see, e.g., the review by Tyralis et al. 2019b; see also the daily streamflow forecasting methods in Papacharalampous et al. 2019b; Tyralis et al. 2021).
- **Still, several useful and realistic machine learning concepts are currently underexploited in hydrological forecasting and forecastability investigations.**
- Here, we extensively discuss some of these concepts, together with related **key findings** and **implementation examples**.
- In these examples, the proposed concepts and **machine learning methods** have been merged with **large hydrological datasets** and **largely interpretable methods** (i.e., stochastic and process-based catchment models).
- The benefits from such mergings are also extensively discussed.



The "no free lunch" theorem and large-scale benchmarking

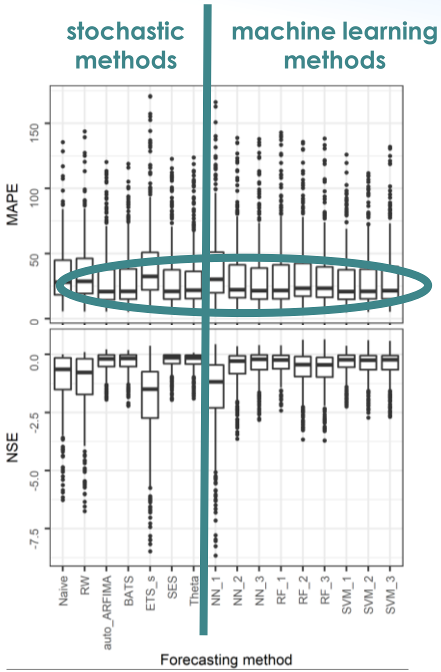
- Among a pool of reasonable algorithmic choices for solving a specific problem type (e.g., annual river flow forecasting), there is no way to know in advance which one will perform the best for one particular problem case (e.g., any annual river flow forecasting case study).
- There is a theorem behind the above statement, which is known as the "no free lunch" theorem (Wolpert 1996).
- This theorem implies that single-case studies cannot stand as empirical proofs that a prediction method performs better than others.
- An optimal selection of prediction methods can be supported by **large-scale benchmarking**, which requires:
  - ✓ **large datasets** comprising many and diverse problem cases to be studied;
  - ✓ **multiple automatic, computationally convenient and fast prediction models**;
  - ✓ **benchmarks** (e.g., simple or more interpretable models).

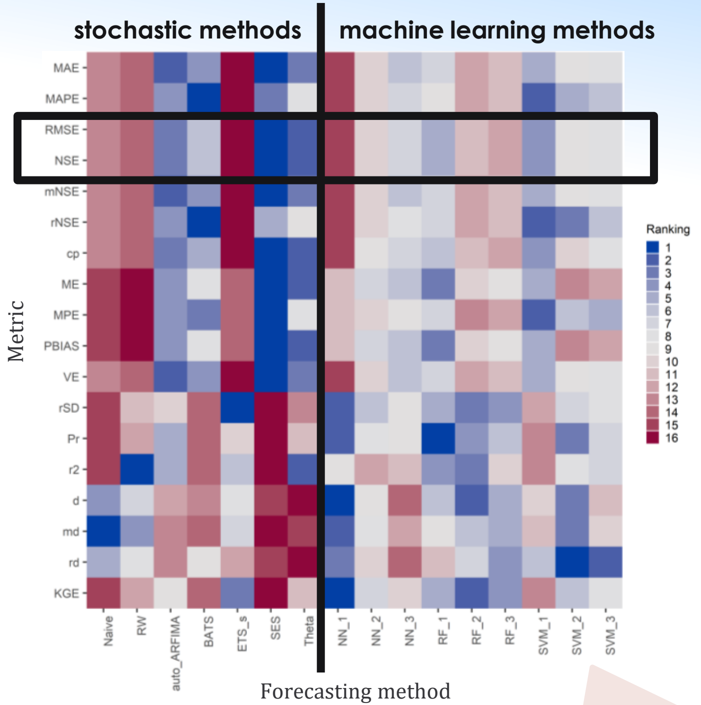
The "no free lunch" theorem and large-scale benchmarking



Adapted from Papacharalampous et al. (2020)

Large-scale comparisons for selecting forecasting methods





Average-case rankings summarizing forecasting performance over 405 geographical locations

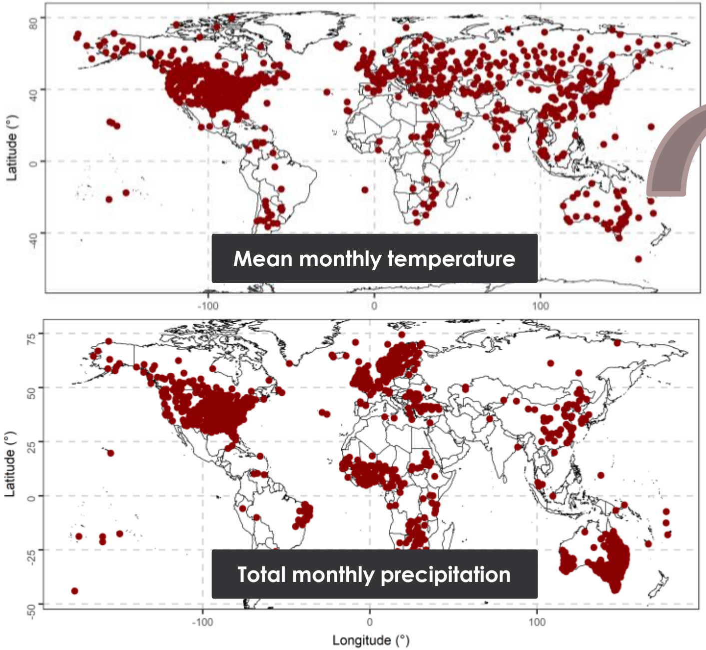
Further reading: Papacharalampous et al. (2019a)

Mean annual streamflow

Original data source: GRDC (2017)

Large-scale comparisons for selecting forecasting methods

Further reading: Papacharalampous et al. (2018b)



Original data sources: Peterson and Vose (1997), Lawrimore et al. (2011)

Which features are important for getting good forecasts?

Seasonality
Autocorrelation
Exogenous relationships
Long-range dependence
Shifts and trends

Examples

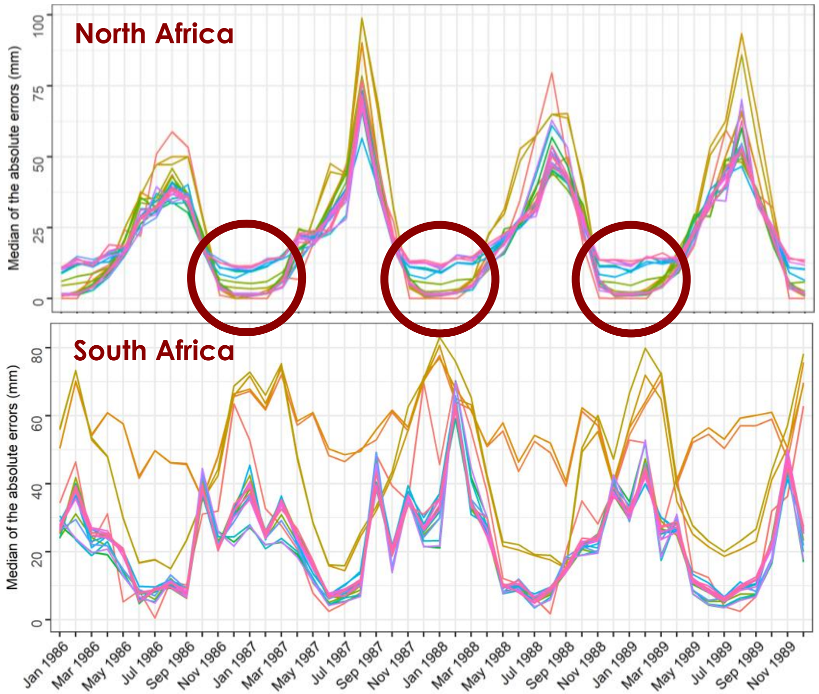
- Seasonal ARFIMA performs only slightly better than seasonal SES.
- Seasonal SES performs notably better than Prophet, and comparably to seasonal exponential smoothing with a trend term.

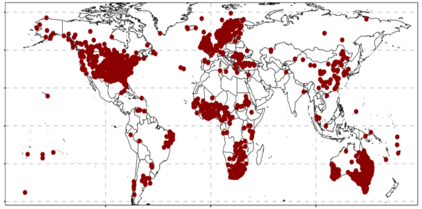
Analogous conclusions can be drawn for:

- **Mean annual temperature**
- **Mean annual precipitation**
- **Mean annual streamflow**
- **Daily streamflow**

Further reading: Papacharalampous et al. (2018a), Papacharalampous et al. (2019a), Tyralis et al. (2021)

Large-scale comparisons for characterizing forecastability





Total monthly precipitation

Forecasting method

naive, rw_1, rw_2, rw_3, rw_4, arfima_1, arfima_2, arfima_3, arfima_4, bats_1, bats_2, bats_3, bats_4, bats_5, bats_6, ses_1, ses_2, ses_3, ses_4, theta_1, theta_2, prophet_1, prophet_2, prophet_3

North Africa
Medians of the Nash Sutcliffe efficiency values
Minimum: 0.49
Maximum: 0.71

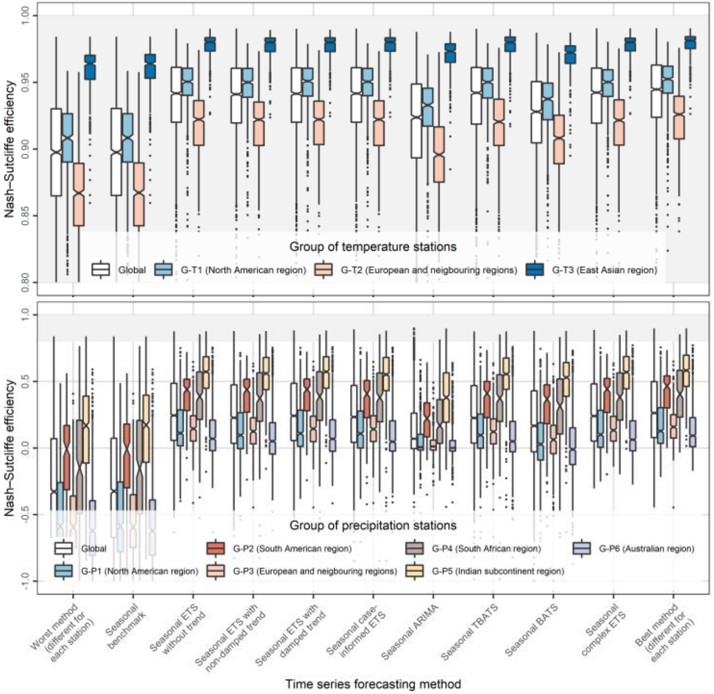
South Africa
Medians of the Nash Sutcliffe efficiency values
Minimum: –0.20
Maximum: 0.30

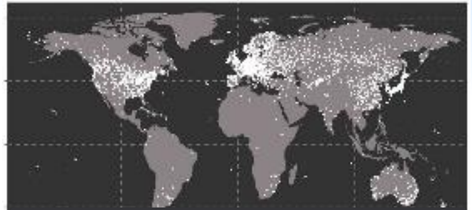
Further reading: Papacharalampous et al. (2018b)

Original data source: Peterson and Vose (1997)

Large-scale comparisons for characterizing forecastability

Further reading: Papacharalampous et al. (2022)





Mean monthly temperature

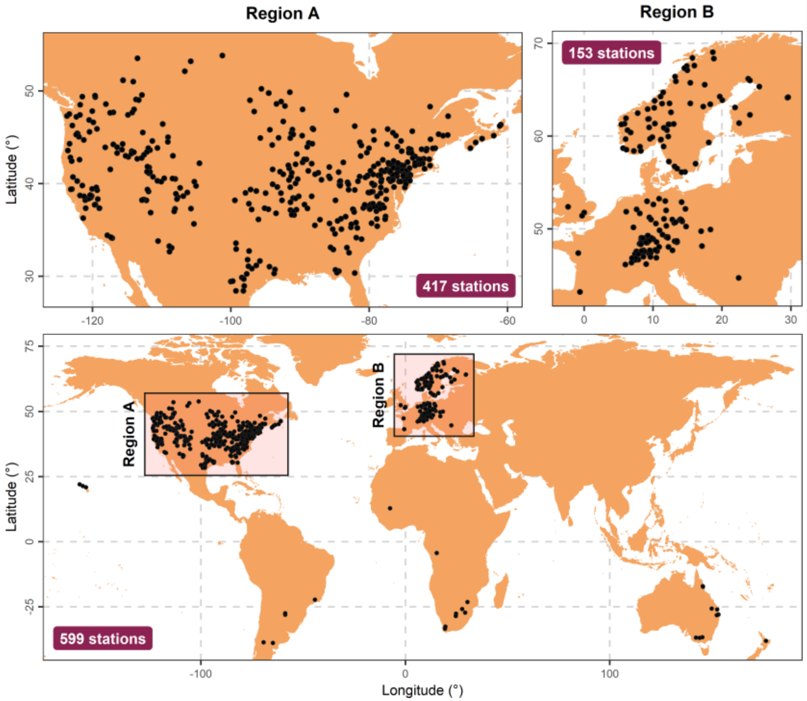
Total monthly precipitation

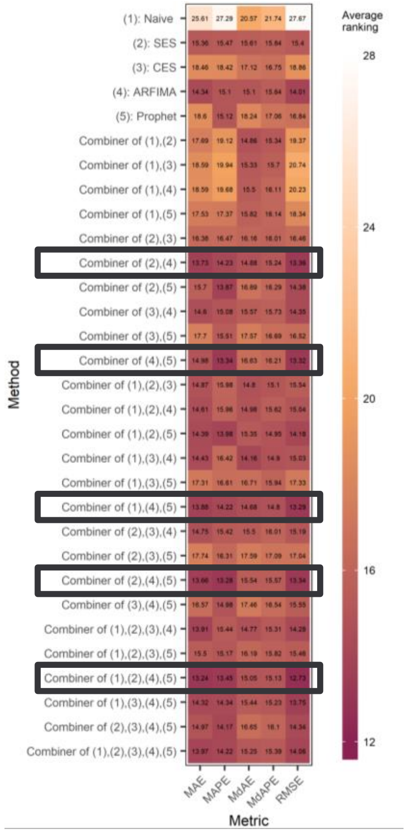
+ Global-scale characterizations of mean monthly river flow forecastability

Original data sources: Peterson and Vose (1997), Menne et al. (2018), Do et al. (2018)

Ensemble learning for benefitting from multiple methods

IAEA





Further reading: Papacharalampous and Tyralis (2020)

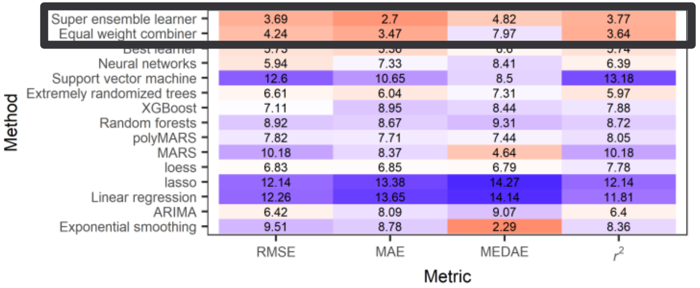
Mean annual streamflow

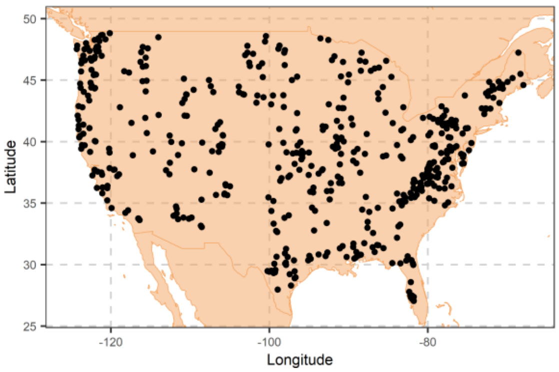
Original data source: Do et al. (2018)

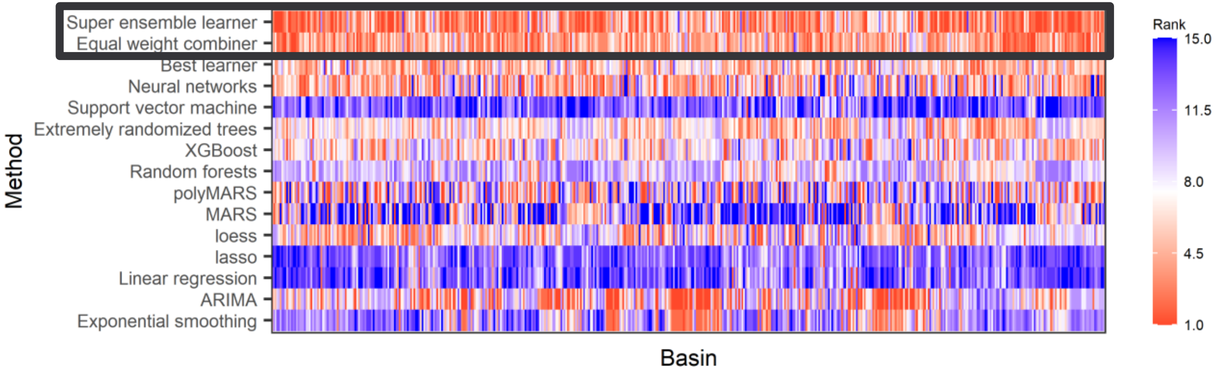
Ensemble learning for benefitting from multiple methods

IAEA

Further reading: Tyralis et al. (2021)



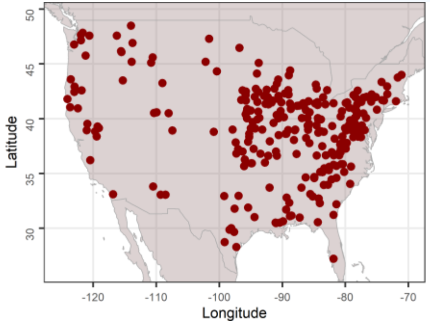




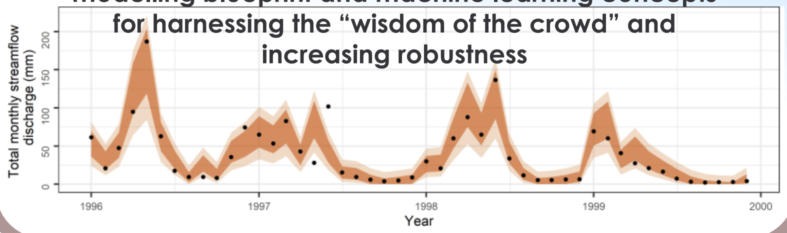
Daily streamflow

Original data sources: Newman et al. (2015), Addor et al. (2017)

Ensemble learning for benefitting from multiple methods

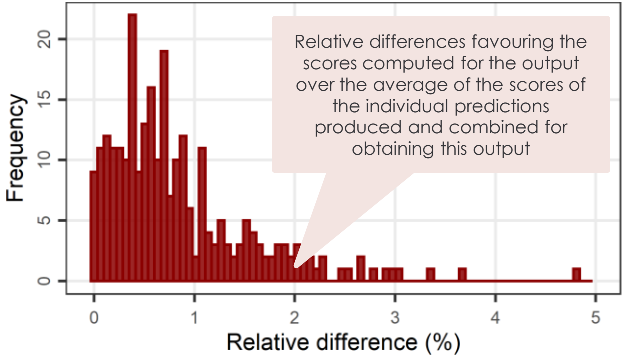


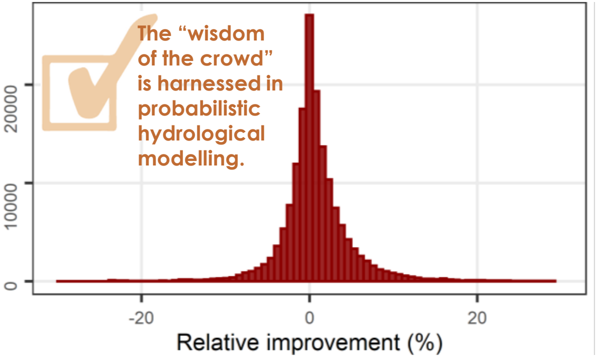
Combining key concepts from a hydrological modelling blueprint and machine learning concepts for harnessing the "wisdom of the crowd" and increasing robustness



Total monthly streamflow

Hydrological modelling blueprint by Montanari and Koutsoyiannis (2012)





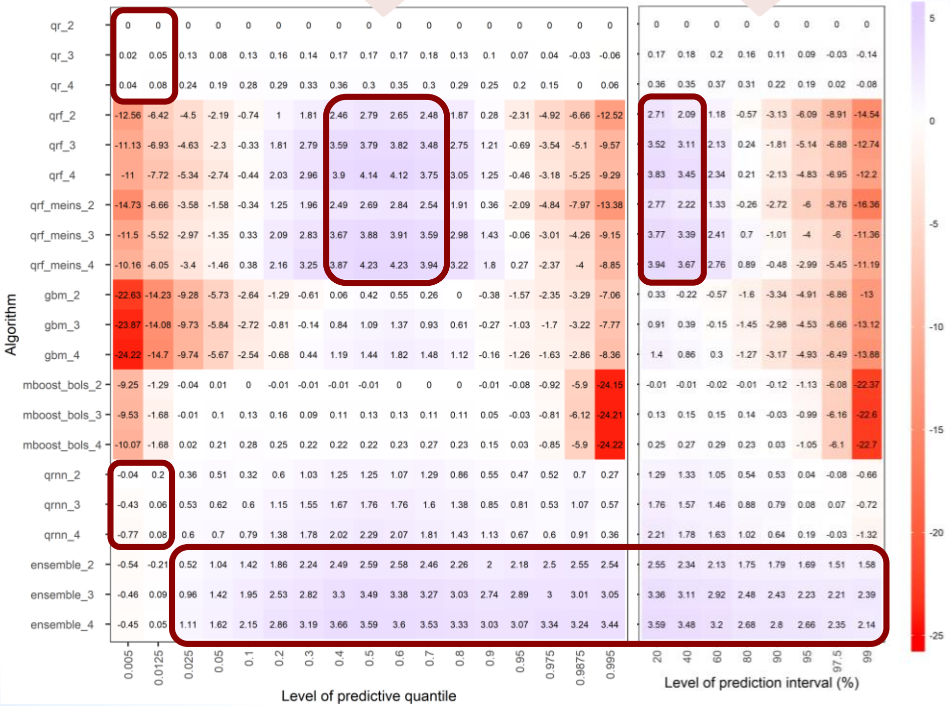
Original data source: Schaake et al. (2006)

Further reading: Papacharalampous et al. (2020)

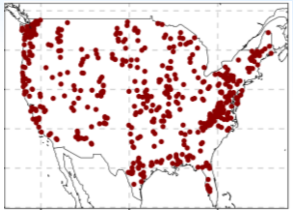
Ensemble learning for benefitting from multiple methods

Relative improvements in terms of quantile score

Relative improvements in terms of interval score



Daily streamflow



Original data sources: Newman et al. (2015), Addor et al. (2017)

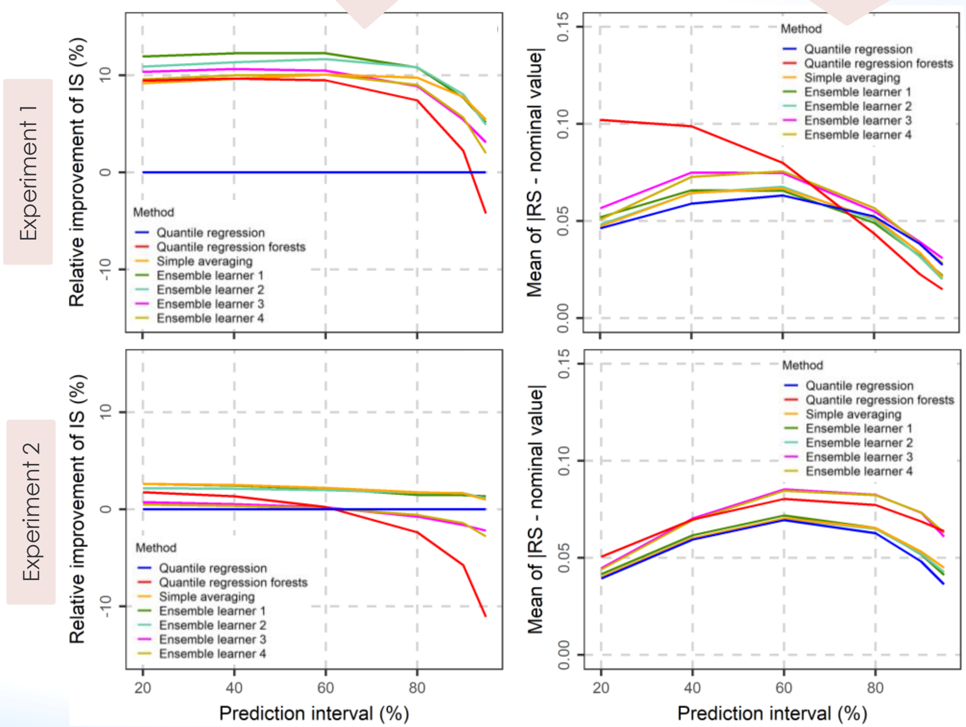
Further reading: Papacharalampous et al. (2019b)

+ Extensive discussions and opinions on the integration of process-based catchment models and machine learning quantile regression algorithms

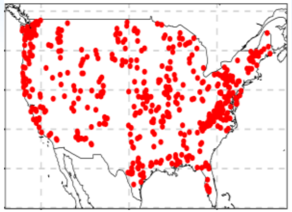
Ensemble learning for benefitting from multiple methods

Relative improvements in terms of interval score

Ensemble means of the absolute differences characterizing reliability



Daily streamflow



Original data sources: Newman et al. (2015), Addor et al. (2017)

Further reading: Tyralis et al. (2019a)

The proposed ensemble learning algorithm builds on fundamental concepts from the fields of hydrology, machine learning, and statistics.

Machine learning inspired adaptations of interpretable models

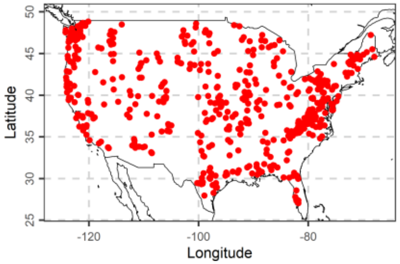
IAEA

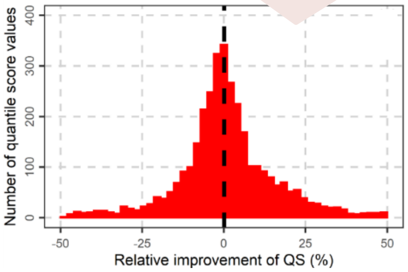
Further reading: Tyralis and Papacharalampous (2021)

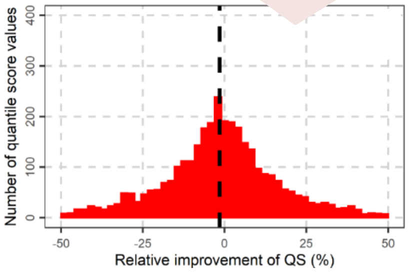
Process-based catchment models have learned to predict streamflow quantiles.

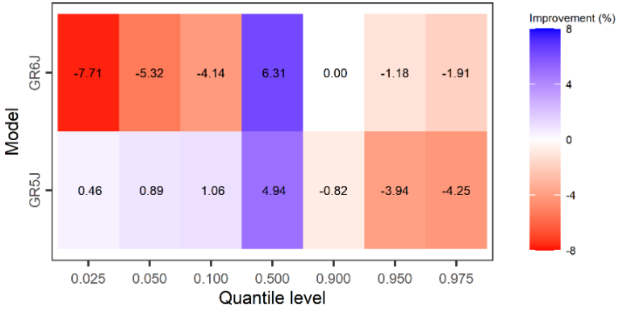
Relative improvements provided by GR5J with respect to GR4J

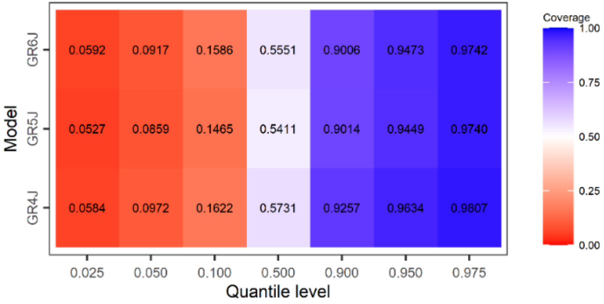
Relative improvements provided by GR6J with respect to GR4J







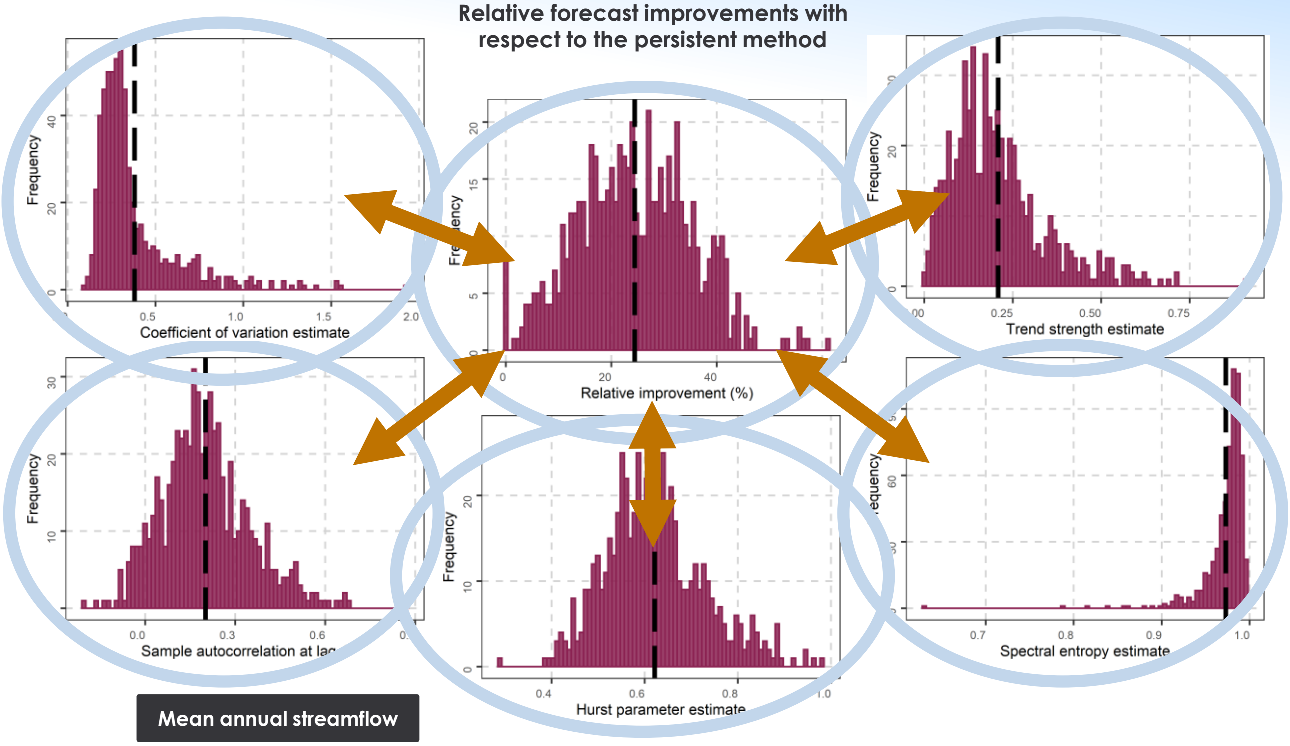




Daily streamflow

Original data sources: Newman et al. (2015), Addor et al. (2017)

Towards analysis-informed integrations of forecasting methods



Original data source: Do et al. (2018)

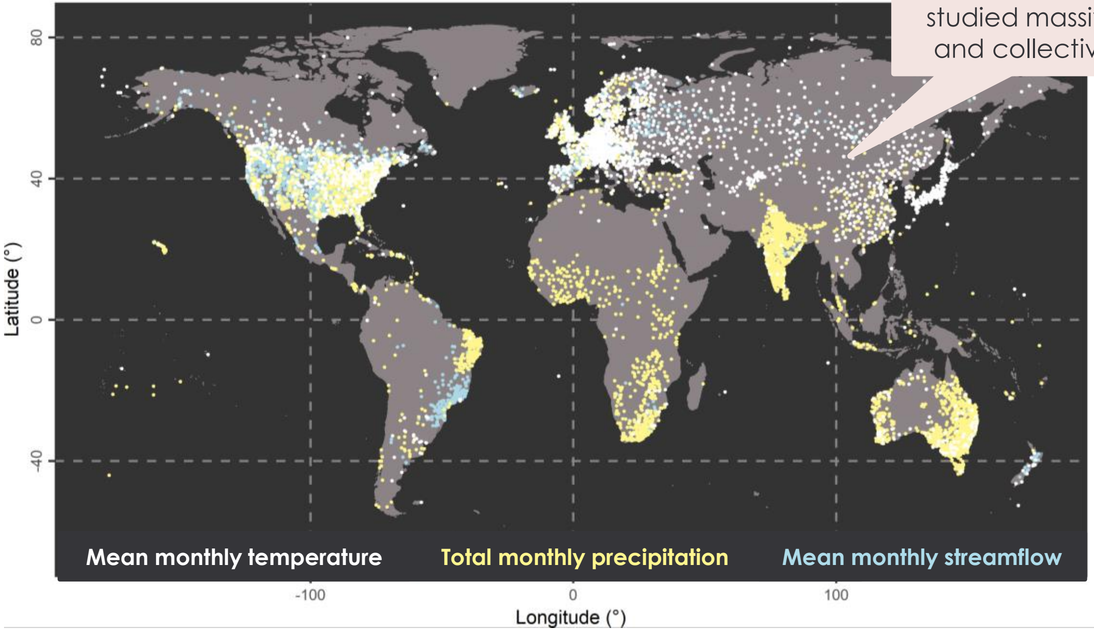
Further reading: Papacharalampous and Tyralis (2020)

Towards analysis-informed integrations of forecasting methods

Benefitting from approximately 60 diverse features

Further reading: Papacharalampous et al. (2021)

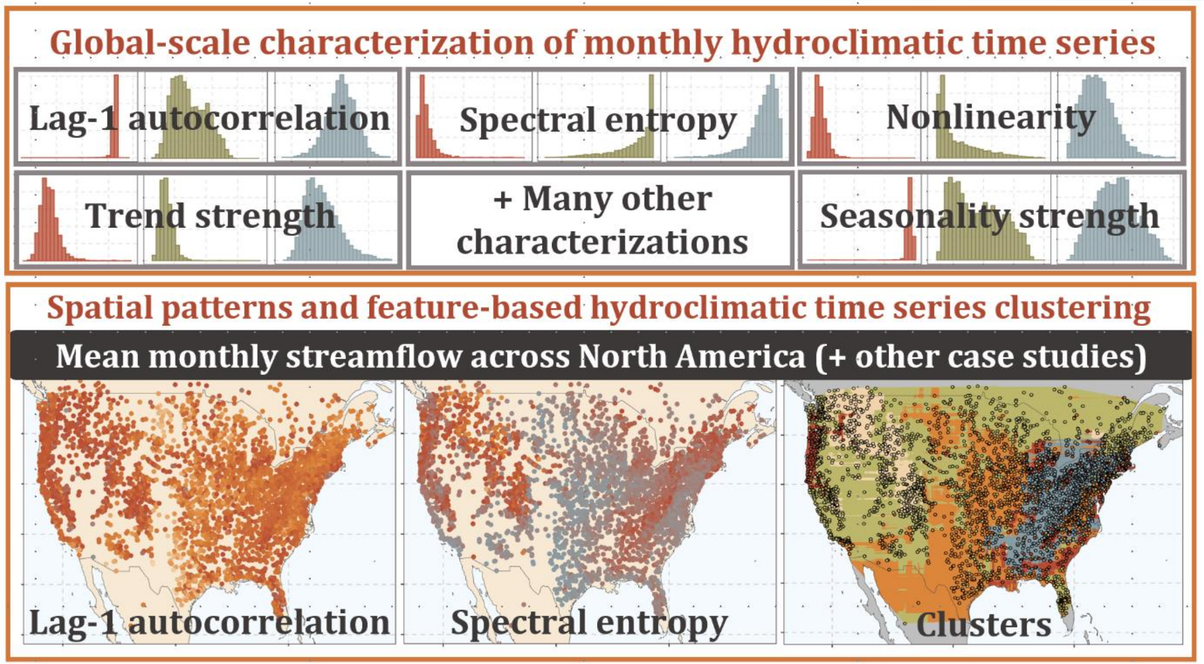
Features should be studied massively and collectively



Original data sources: Peterson and Vose (1997), Menne et al. (2018), Do et al. (2018)

Towards analysis-informed integrations of forecasting methods

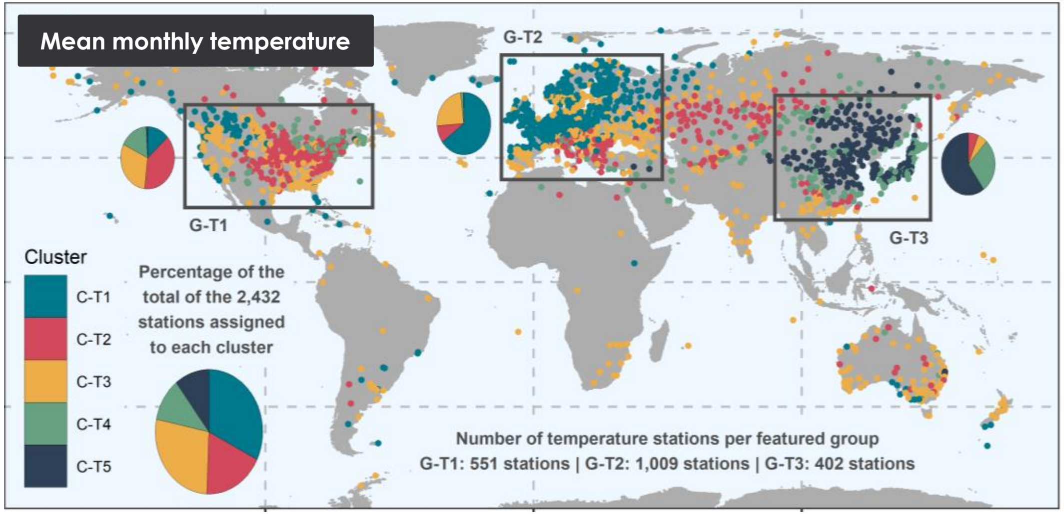
Further reading: Papacharalampous et al. (2021)



Original data sources: Peterson and Vose (1997), Menne et al. (2018), Do et al. (2018)

Towards analysis-informed integrations of forecasting methods

Temperature time series forecastability in terms of Nash-Sutcliffe efficiency in the different clusters

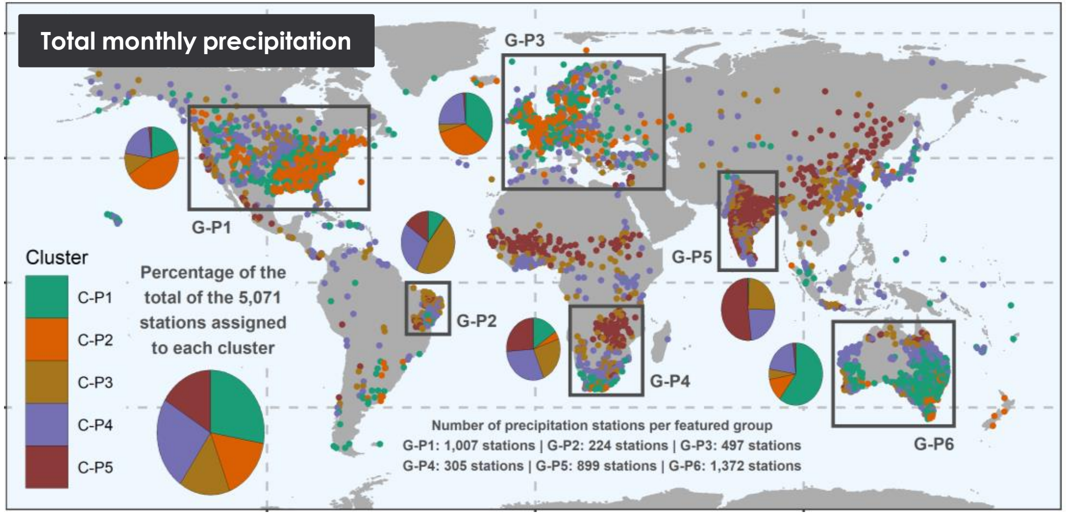


Original data source: Menne et al. (2018)

Further reading: Papacharalampous et al. (2022)

Towards analysis-informed integrations of forecasting methods

Precipitation time series forecastability in terms of Nash-Sutcliffe efficiency in the different clusters



Original data source: Peterson and Vose (1997)

Further reading: Papacharalampous et al. (2022)

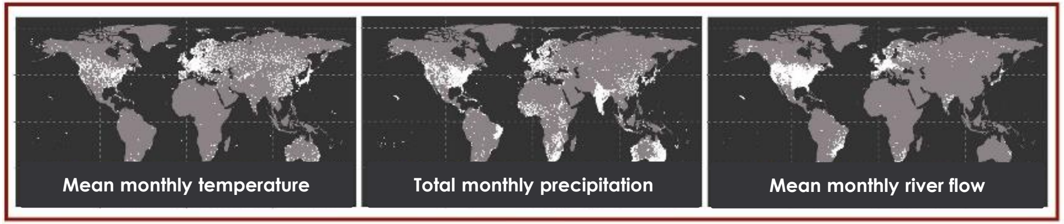
Towards analysis-informed integrations of forecasting methods

+ Benefitting from explainable machine learning

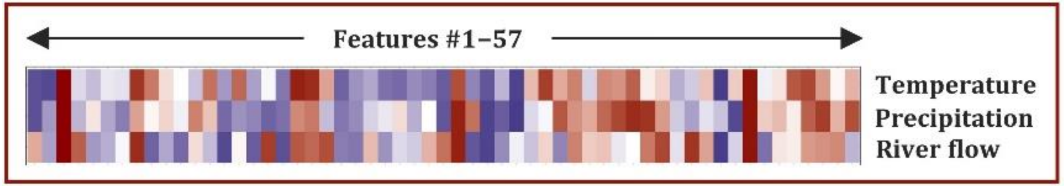
Further reading: Papacharalampous et al. (2022)

Forecasting at scale and forecastability assessment + Massive feature extraction → Comparisons of the features with respect to their usefulness in explaining forecastability

Global hydroclimatic datasets



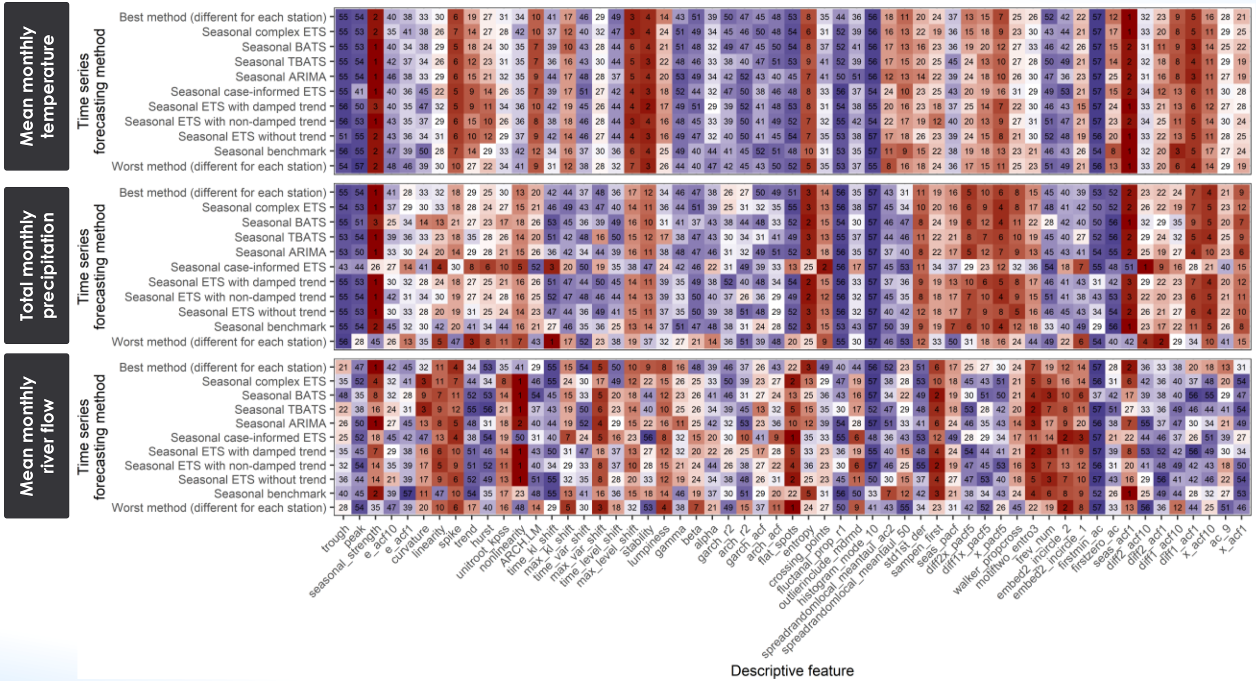
Rankings of the features from the most to the least informative ones



Towards analysis-informed integrations of forecasting methods

+ Benefitting from explainable machine learning

Further reading: Papacharalampous et al. (2022)



Take-home messages

- Hydrological post-processing and forecasting can be improved by exploiting **machine learning concepts** and methods.
- The same holds for hydro-forecastability assessments and interpretations.
- As long as their outputs are useful, hydrological forecasting methods do not have to be (but they can be) **interpretable**.
- There is no certainty and **"no free lunch"** in predictive modelling.
- **Large-scale benchmarking** and **ensemble learning** are ways to cope with this fact in a meaningful sense.
- By conducting large-scale benchmark tests, we can find:
  - ✓ **which forecasting methods perform well** (practically, better than others) in the long run; and
  - ✓ **which features are important** (practically, more important than others) for getting good forecasts in the long run.
- An interesting example is **methods with trends**.
- Such methods are getting much attention in the hydrological literature; however, they do not offer improvements (as individual methods) in terms of forecasting performance.

Take-home messages

IAEA

- Selecting individual forecasting methods is meaningful; however, preferably **multiple methods** should be integrated and combined for **maximizing the benefits** and **reducing the risks** from their use.
- Methods that would probably be discarded as individual ones based on their performance in the long run (e.g., methods with trends or naïve methods) might be proven important as parts of **ensemble methods**.
- The forecasts of **diverse methods** seem to complement themselves well in **ensemble learning** contexts.
- Further improvements could be achieved through **analysis-informed combinations** and **analysis-informed integrations** of many and diverse forecasting models.
- For achieving meaningful combinations and integrations in this regard, **many and diverse descriptive features** should be studied.
- A massive and collective examination of **hydroclimatic features** is also necessary for understanding **hydroclimatic forecastability**.
- Overall, by merging **machine learning concepts** and methods with large hydrological datasets and largely interpretable (e.g., stochastic or process-based catchment) models, new fruitful avenues open up for our field.

References

Addor N, Newman AJ, Mizukami N, Clark MP (2017) The CAMELS data set: Catchment attributes and meteorology for large-sample studies. Hydrology and Earth System Sciences 21:5293–5313. doi:10.5194/hess-21-5293-2017

Do HX, Gudmundsson L, Leonard M, Westra S (2018) The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. Earth System Science Data 10:765–785. doi:10.5194/essd-10-765-2018

Hastie T, Tibshirani R, Friedman JH (2009) The Elements of Statistical Learning: Data Mining, Inference and Prediction, second edition. Springer, New York. doi:10.1007/978-0-387-84858-7

James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning. Springer, New York. doi:10.1007/978-1-4614-7138-7

Lawrimore JH, Menne MJ, Gleason BE, Williams CN, Wuertz DB, Vose RS, Rennie J (2011) An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. Journal of Geophysical Research – Atmospheres 116(D1912). doi:10.1029/2011JD016187

Menne MJ, Williams CN, Gleason BE, Rennie JJ, Lawrimore JH (2018) The global historical climatology network monthly temperature dataset, version 4. Journal of Climate 31(24):9835–9854. doi:10.1175/JCLI-D-18-0094.1

Montanari A, Koutsoyiannis D (2012) A blueprint for process-based modeling of uncertain hydrological systems. Water Resources Research 48(9):W09555. doi:10.1029/2011WR011412

Newman AJ, Clark MP, Sampson K, Wood A, Hay LE, Bock A, Viger RJ, Blodgett D, Brekke L, Arnold JR, Hopson T, Duan Q (2015) Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. Hydrology and Earth System Sciences 19:209–223. doi:10.5194/hess-19-209-2015

Papacharalampous GA, Tyralis H (2020) Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. Journal of Hydrology 590:125205. doi:10.1016/j.jhydrol.2020.125205

Papacharalampous GA, Tyralis H, Koutsoyiannis D (2018a) One-step ahead forecasting of geophysical processes within a purely statistical framework. Geoscience Letters 5:12. doi:10.1186/s40562-018-0111-1

Papacharalampous GA, Tyralis H, Koutsoyiannis D (2018b) Predictability of monthly temperature and precipitation using automatic time series forecasting methods. Acta Geophysica 66(4):807–831. doi:10.1007/s11600-018-0120-7

Papacharalampous GA, Tyralis H, Koutsoyiannis D (2019a) Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. Stochastic Environmental Research and Risk Assessment 33(2):481–514. doi:10.1007/s00477-018-1638-6

Papacharalampous GA, Tyralis H, Langousis A, Jayawardena AW, Sivakumar B, Mamassis N, Montanari A, Koutsoyiannis D (2019b) Probabilistic hydrological post-processing at scale: Why and how to apply machine-learning quantile regression algorithms. Water 11(10):2126. doi:10.3390/w11102126

Papacharalampous GA, Tyralis H, Koutsoyiannis D, Montanari A (2020) Quantification of predictive uncertainty in hydrological modelling by harnessing the wisdom of the crowd: A large-sample experiment at monthly timescale. Advances in Water Resources 136:103470. doi:10.1016/j.advwatres.2019.103470

Papacharalampous GA, Tyralis H, Papalexiou SM, Langousis A, Khatami S, Volpi E, Grimaldi S (2021) Global-scale massive feature extraction from monthly hydroclimatic time series: Statistical characterizations, spatial patterns and hydrological similarity. Science of the Total Environment 767:144612. doi:10.1016/j.scitotenv.2020.144612

Papacharalampous GA, Tyralis H, Pechlivanidis IG, Grimaldi S, Volpi E (2022) Massive feature extraction for explaining and foretelling hydroclimatic time series forecastability at the global scale. arXiv:2108.00846

Peterson TC, Vose RS (1997) An overview of the Global Historical Climatology Network Temperature database. Bulletin of the American Meteorological Society 78(12):2837–2850. doi:10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2

Schaake J, Cong S, Duan Q (2006) US MOPEX data set. IAHS Publication 307:9–28

Tyralis H, Papacharalampous GA (2021) Quantile-based hydrological modelling. arXiv:2110.05586

Tyralis H, Papacharalampous GA, Burnetas A, Langousis A (2019a) Hydrological post-processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS. Journal of Hydrology 577:123957. doi:10.1016/j.jhydrol.2019.123957

Tyralis H, Papacharalampous GA, Langousis A (2019b) A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water 11(5):910. doi:10.3390/w11050910

Tyralis H, Papacharalampous GA, Langousis A (2021) Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms. Neural Computing and Applications 33:3053–3068. doi:10.1007/s00521-020-05172-3

Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. Neural Computation 8(7):1341–1390. doi:10.1162/neco.1996.8.7.1341

IAEA
International Atomic Energy Agency

Thank you!