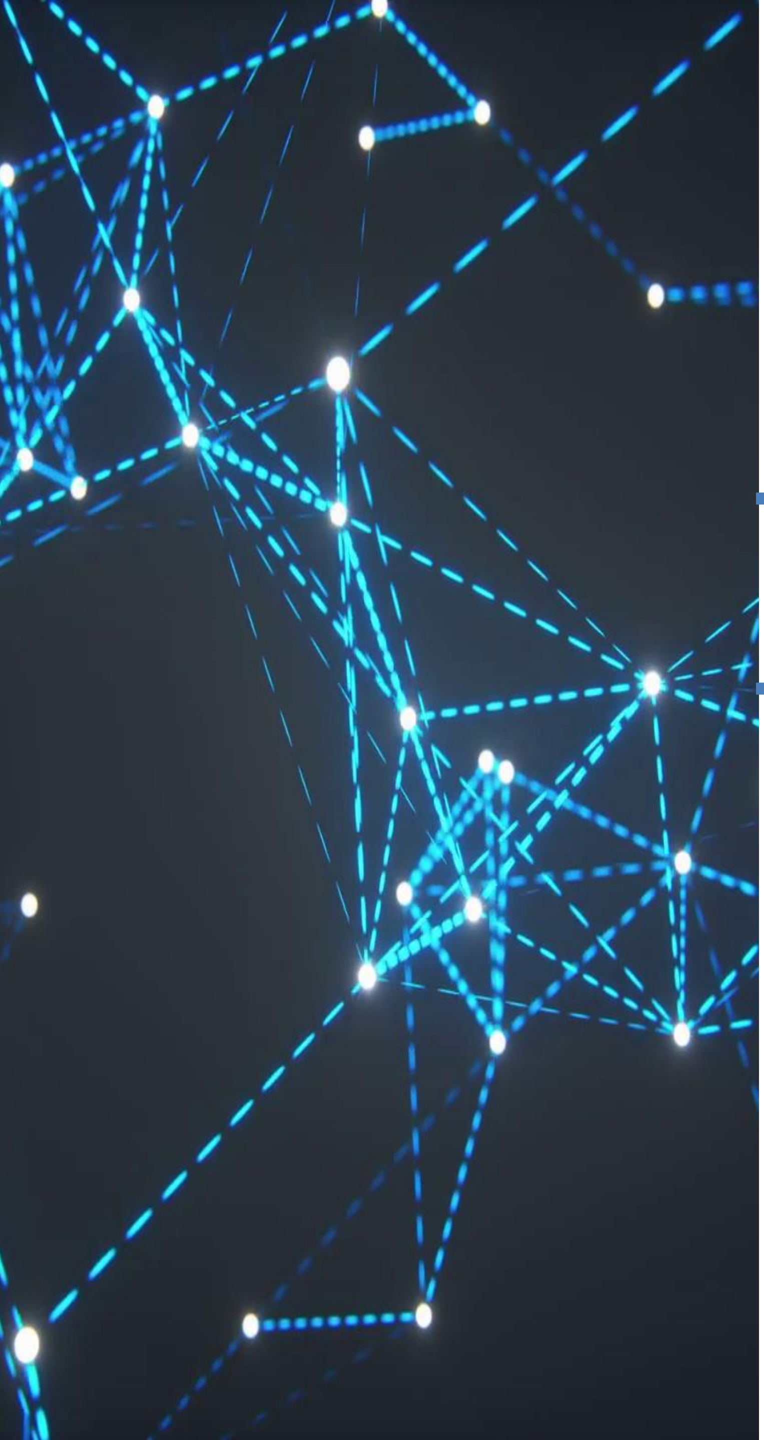


EGU General Assembly 2023

HS7.4: Steps towards future hydroclimatic scenarios for water resources management in a changing world

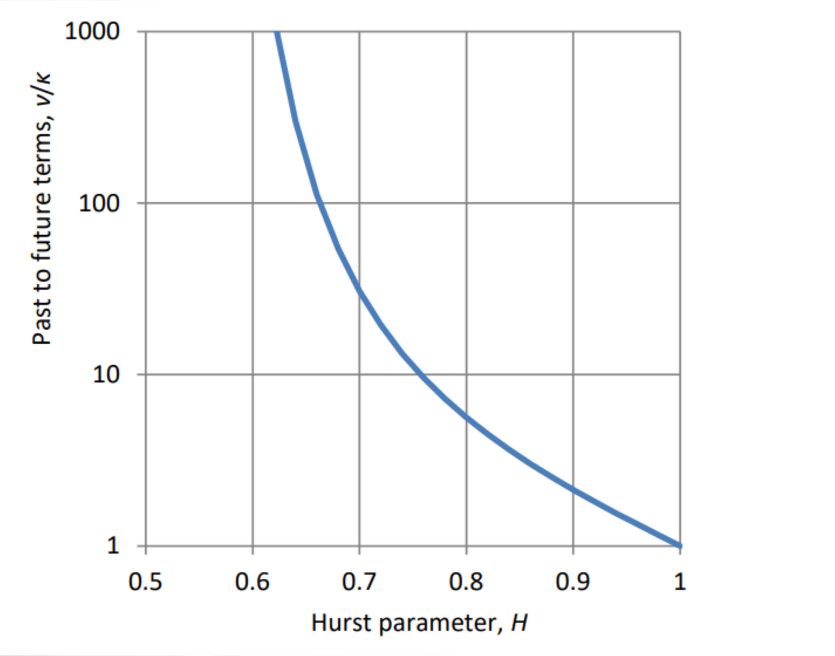
Comparison of Stochastic versus Deep Learning methods for prediction of hydroclimatic time series

Authors: Nikolaos Tepetidis, Theano Iliopoulou, Panayiotis Dimitriadis, and Demetris Koutsoyiannis

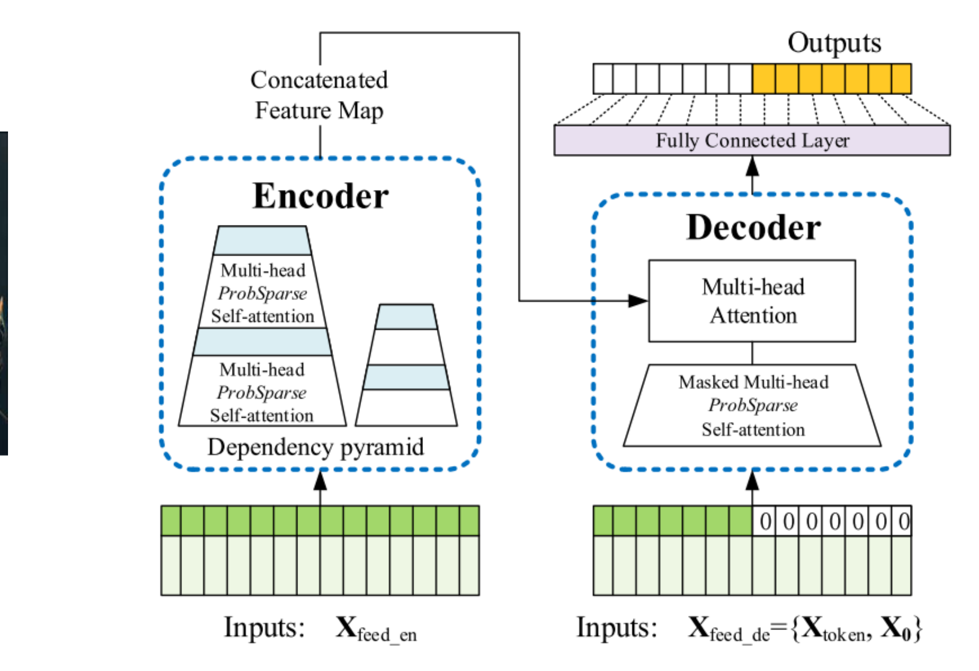
Affiliation: Department of Water Resources and Environmental Engineering
School of Civil Engineering
National Technical University of Athens, Greece

Methods

- Stochastic:
  - Computing the necessary past terms n to estimate a representative average of future mean for a predetermined period of length κ. **Simple** and **Theoretically substantiated** on stochastic processes.

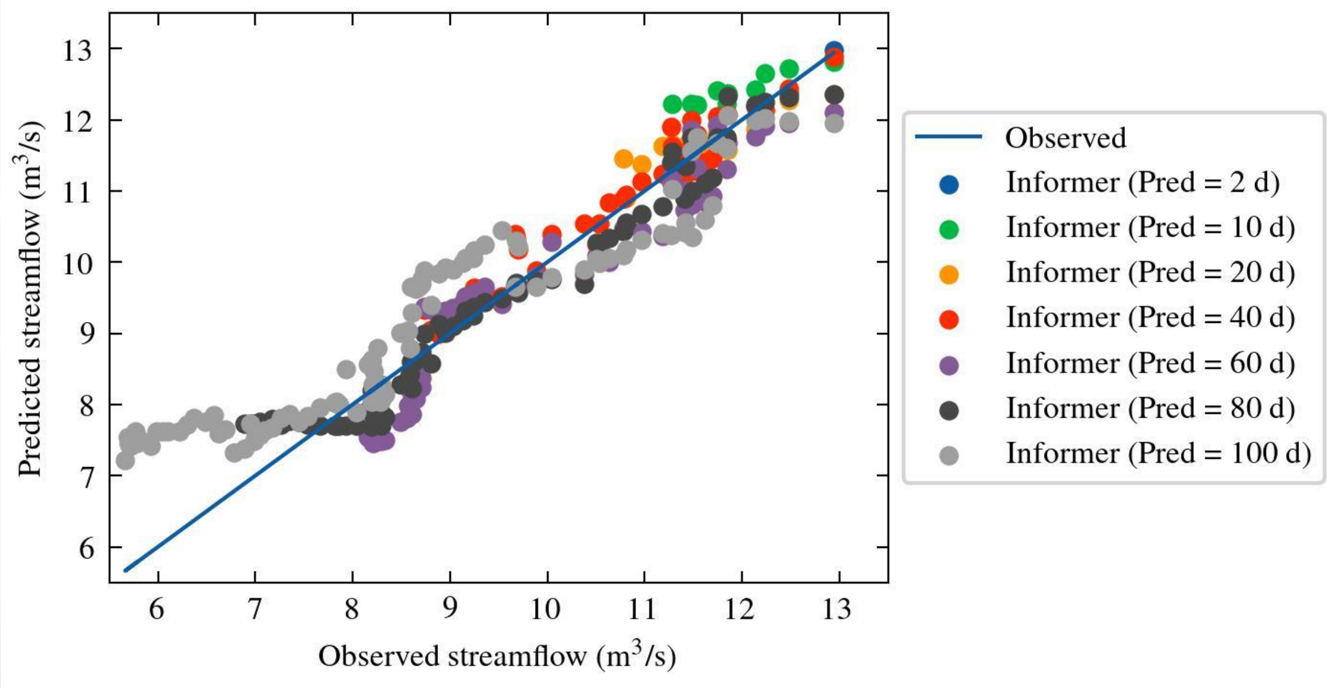


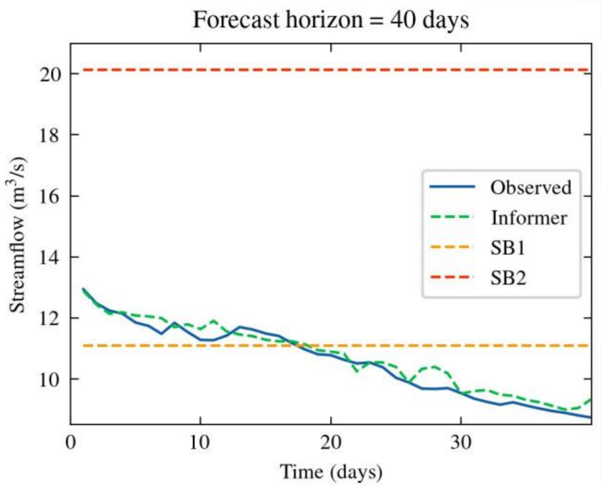
- Informer:
  - Transformer based model for long range predictions:
    Reduce computational cost with **ProbSparse attention** and generates the output sequence in **one forward pass**.

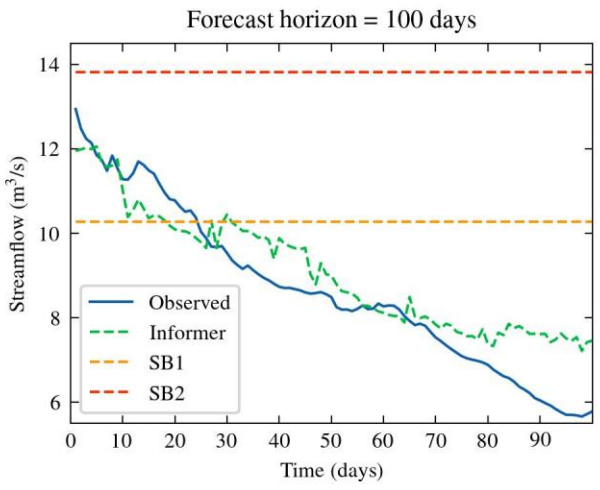


Results on case study: River Test - flow and climatic data

- Informer model has shown pretty encouraging results for time series forecasting tasks both in short and long forecast horizon.
- Informer outperforms both stochastic approaches used.

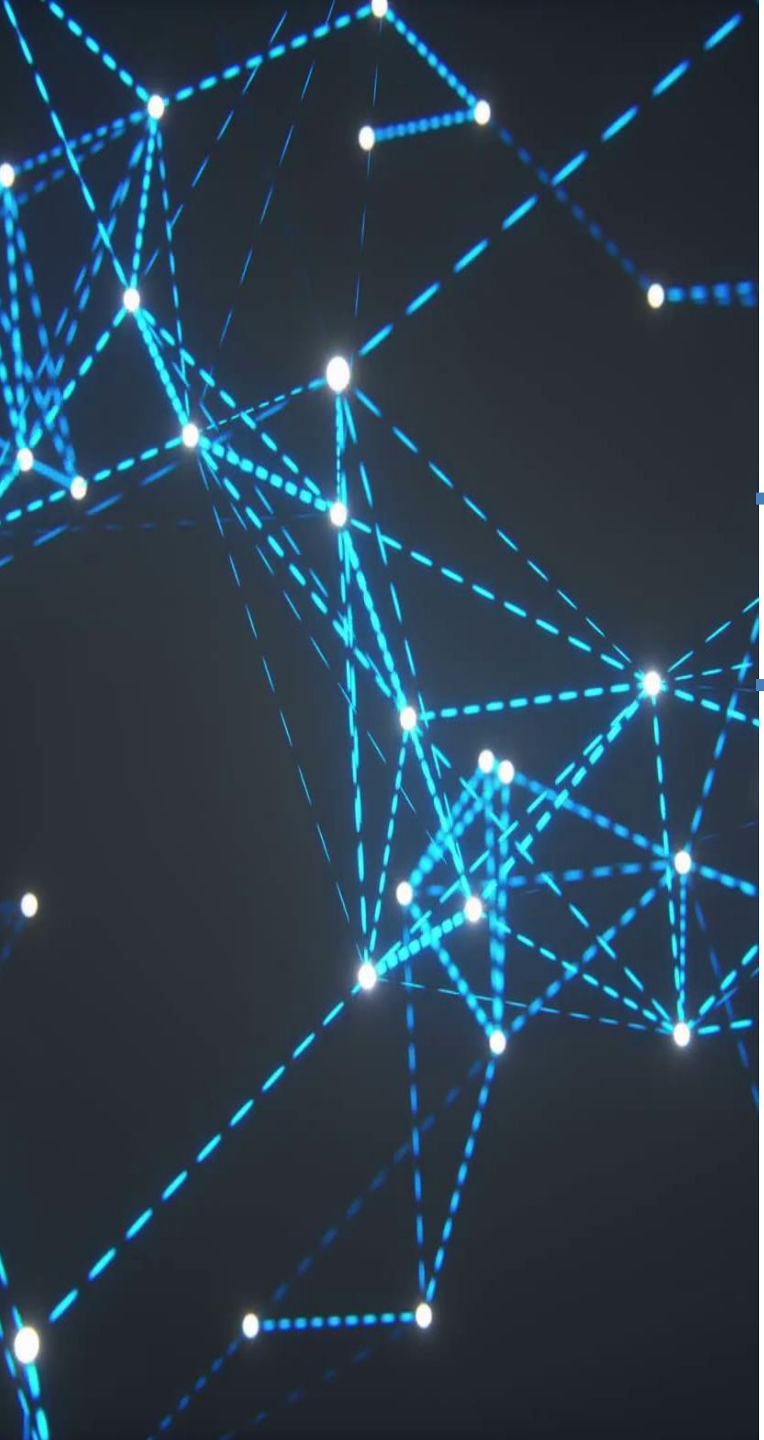






End of Presentation

Thank you for your ‘Probabilistic’ attention!



EGU General Assembly 2023

HS7.4: Steps towards future hydroclimatic scenarios for water resources management in a changing world

Comparison of Stochastic versus Deep Learning methods for prediction of hydroclimatic time series

Authors: Nikolaos Tepetidis, Theano Iliopoulou, Panayiotis Dimitriadis, and Demetris Koutsoyiannis

Affiliation: Department of Water Resources and Environmental Engineering
School of Civil Engineering
National Technical University of Athens, Greece

Introduction and Motivation

- Time series prediction is a fundamental task that involves predicting future values based on its past values.
- In civil engineering field hydroclimatic time series prediction can be crucial because it can inform the design and management of critical infrastructure systems.
- Accurate predictions of hydroclimatic variables such as precipitation, temperature and streamflow can bring up to date decisions on the design, construction and operation of these systems, as well as help to mitigate the impacts of natural disasters such as floods and droughts.
- Deep Learning techniques, such as transformer models, are receiving great scientific attention and increasingly gaining popularity with promising results in time series prediction.
- Motivation: Developing accurate time series prediction models can lead to better decision making and ultimately, improved outcomes in many domains.
- This work aims to assess the prediction capacity of transformer models compared to benchmark predictions provided by stochastics models.

Methods

- Methods used in this work for prediction:
  - Transformer based model → Informer

    1) Transformers (Vaswani et al., 2017) came to overcome some limitations of existing models, such as difficulty in modeling **long-term dependencies** and inefficiency in handling high-dimensional data.

    2) The **attention** mechanism in Transformers allows them to selectively focus on relevant inputs, making them efficient in handling long-term dependencies.

    3) Informer aims to reduce the computational cost through **ProbSparse** attention and generates the entire predicted output sequence in one forward pass.

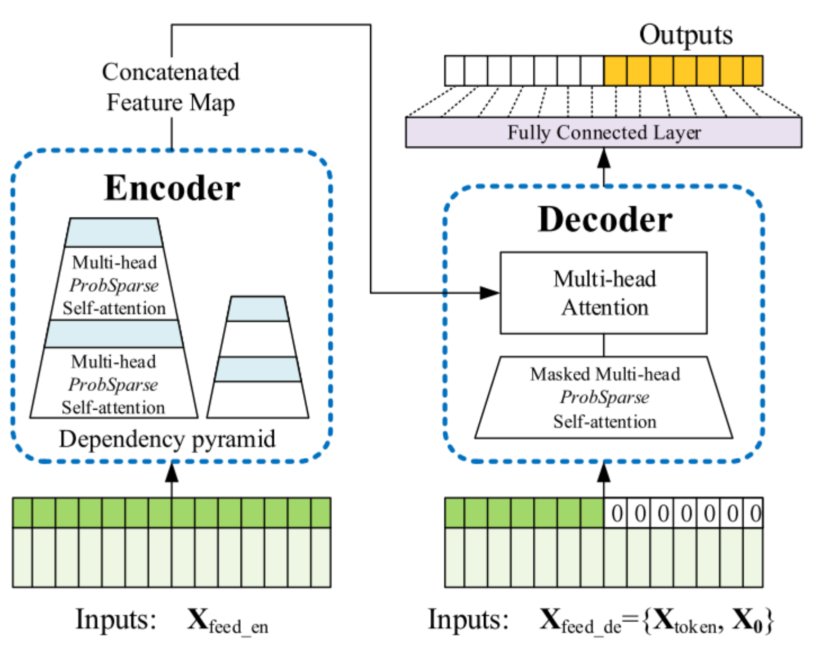
  - Stochastic based model → Stochastic Benchmark 1 (SB1), Stochastic Benchmark 2 (SB2)

    Computing the necessary past terms n to estimate a representative average of future mean for a predetermined predicted period of length κ.

Informer model (1/2)

- The **Informer** model introduced by Haoyi Zhou et al. (2021) is a deep learning model specifically designed for time series prediction tasks.
- The model is based on the **transformer** architecture, but it includes several unique components that enable it to efficiently incorporate multi-scale information and external factors.
- Informer features a **Multi-Encoder Multi-Decoder** (MEMD) structure that allows for the efficient extraction of features at multiple time scales.
- The use of **Autoregressive Decoding**, make the model able to generate the output sequence in one forward pass.

Informer Architecture



Informer model (2/2)

- Informer uses probabilistic sparsity attention mechanism (**ProbSparse attention**) instead of the traditional dot-product attention mechanism used in the original Transformer model.
- This mechanism enables the model to attend to a subset of the input features rather than the entire input, **reducing the computational cost** of the attention mechanism.
- In ProbSparse attention, the attention weights are computed based on the similarity between the query and a subset of the keys (two components that are used to compute the attention weights), rather than all the keys.

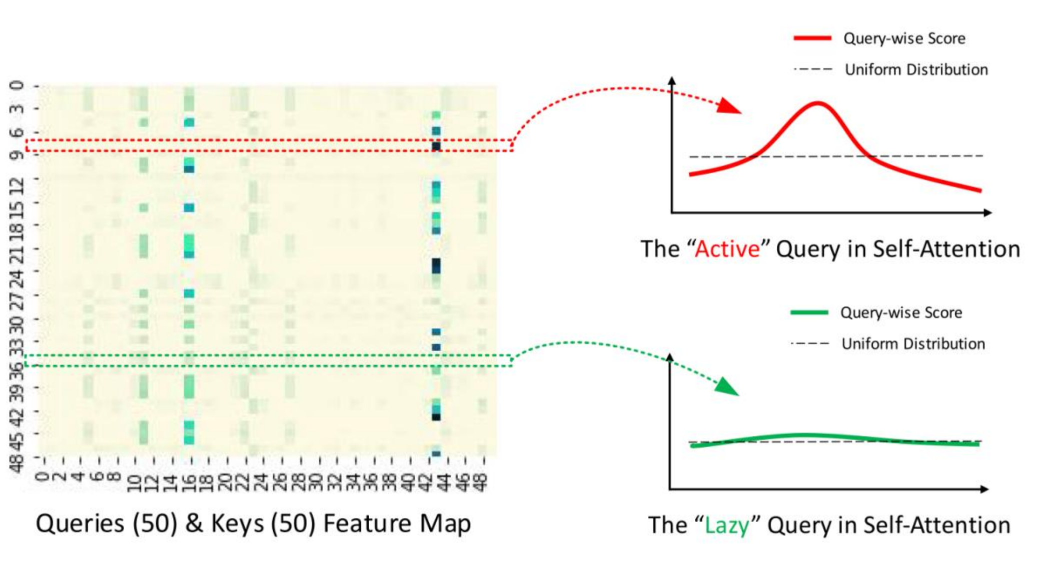


Illustration of ProbSparse Attention

Stochastic model (1/2)

- For the estimation of the local future mean at future prediction period of length κ (conditional on the present and past values of the discrete process $\underline{x}^i$), i.e.,

$$\underline{\mu}_\kappa := \mathrm{E}\left[\frac{1}{\kappa}\left(\underline{x}_1 + \cdots + \underline{x}_\kappa\right) | \underline{x}_0, \underline{x}_{-1}, \dots\right]$$

- As described in Koutsoyiannis (2021) approach, we select only the past $0 \leq \nu \leq n$ values, i.e.,

$$\underline{\hat{\mu}}_\nu := \frac{1}{\nu}\left(\underline{x}_0 + \underline{x}_{-1} + \cdots + \underline{x}_{-\nu+1}\right)$$

that minimizes the square error, i.e.: $A(\kappa, \nu) := \mathrm{E}\left[(\underline{\hat{\mu}}_\nu - \underline{\hat{\mu}}_\kappa)^2\right]$.

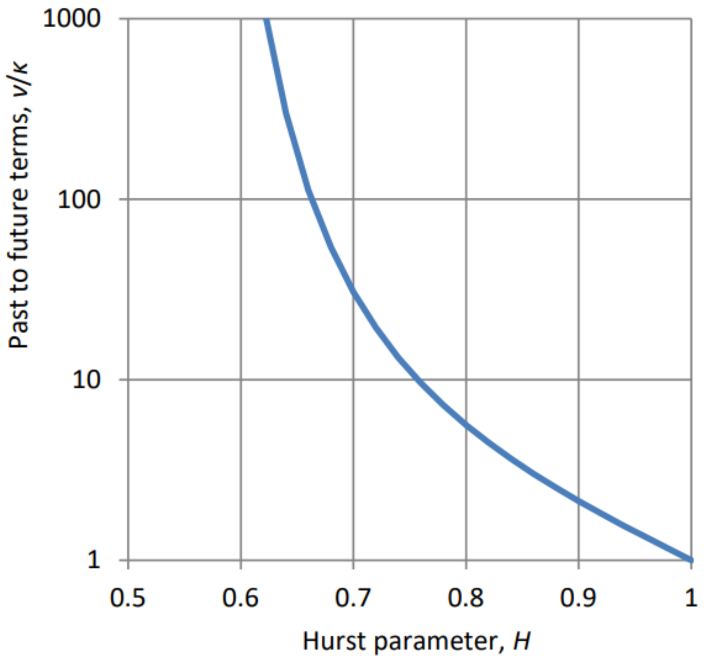
- It can be shown that the standardized mean squared error is:

$$A(\kappa, \nu) = \left(\frac{1}{\kappa} + \frac{1}{\nu}\right)\left(\kappa\, \gamma(\kappa) + \nu\, \gamma(\nu) - (\nu + \kappa)\, \gamma(\nu + \kappa)\right)$$

Stochastic model (2/2)

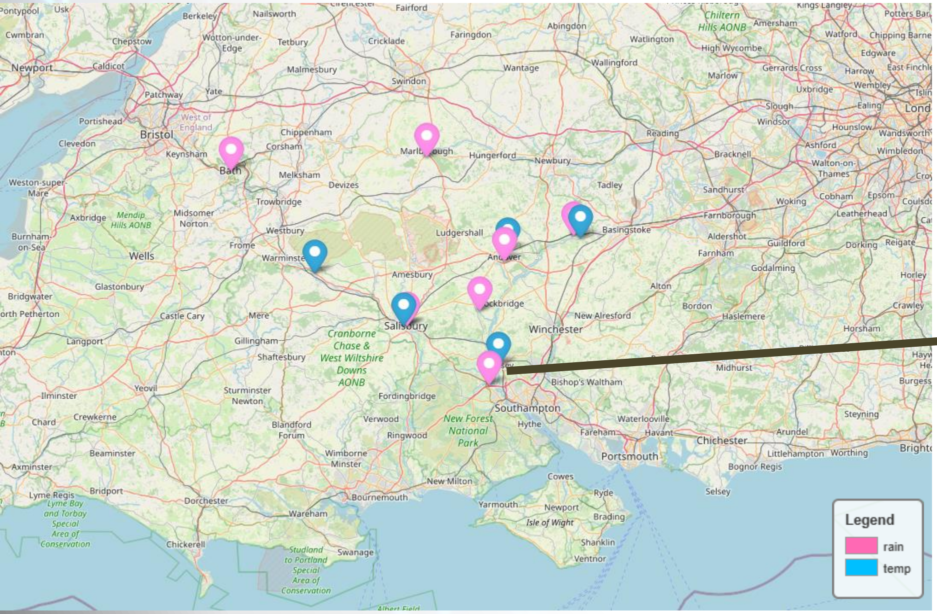
- For the simplest Hurst-Kolmogorov process, for which $\gamma(\kappa) = \lambda^2(\kappa/\alpha)^{2H-2}$ we get the value of ν that gives min A, compared to κ, Koutsoyiannis (2021):

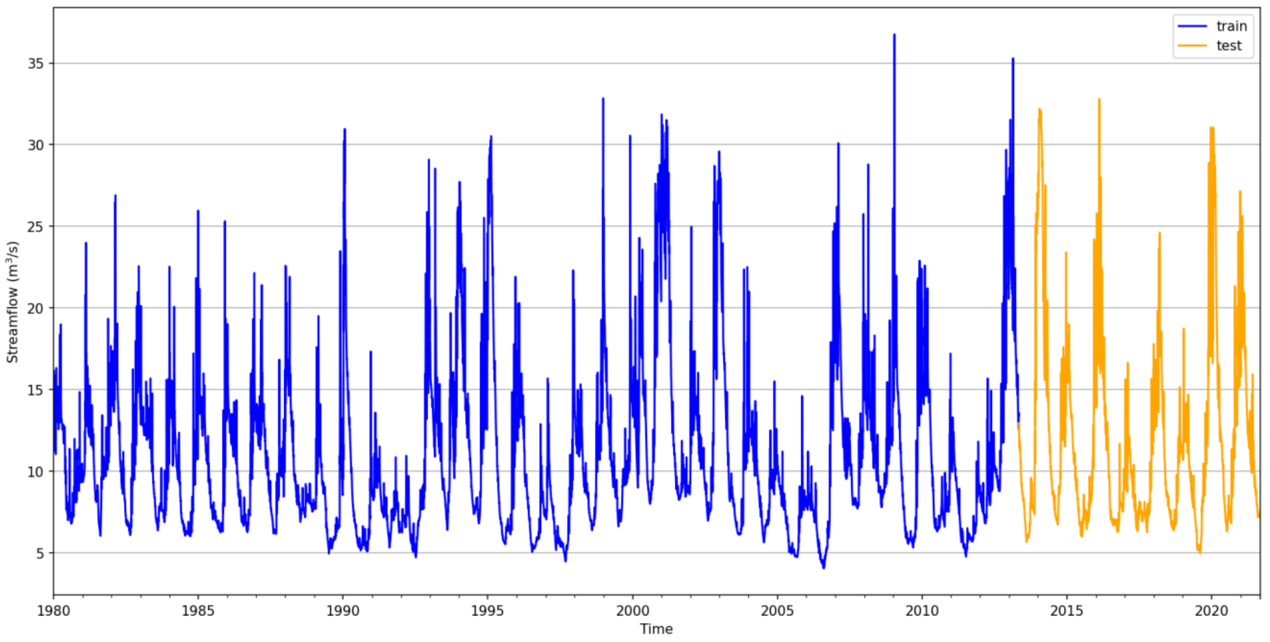
$$\boxed{\nu = \frac{\kappa}{(\max(0,\ \ 2.5H - 1.5))^{2.5}}}$$



Case study: River Test

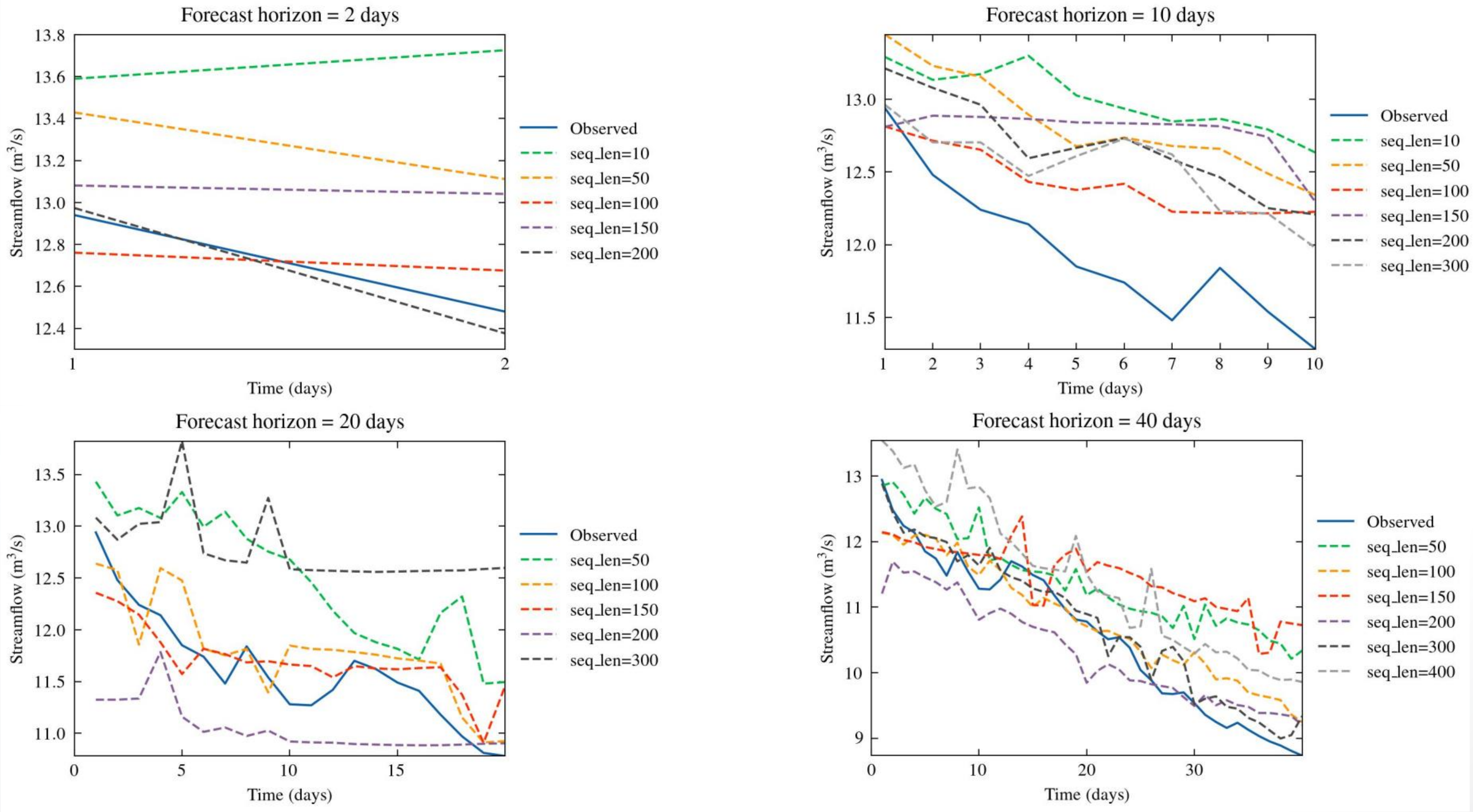
- The dataset used in this study contains the daily gauged flow rate (m^3/s) of the River Test and climatic data (precipitation and temperature) in Hampshire, England.
- The data covers the period during 1980-2021 (41 years).
- Prediction task was performed in Broadland station of River Test, where its flow time series depicted in the image below (train/test split = 0.8/0.2).





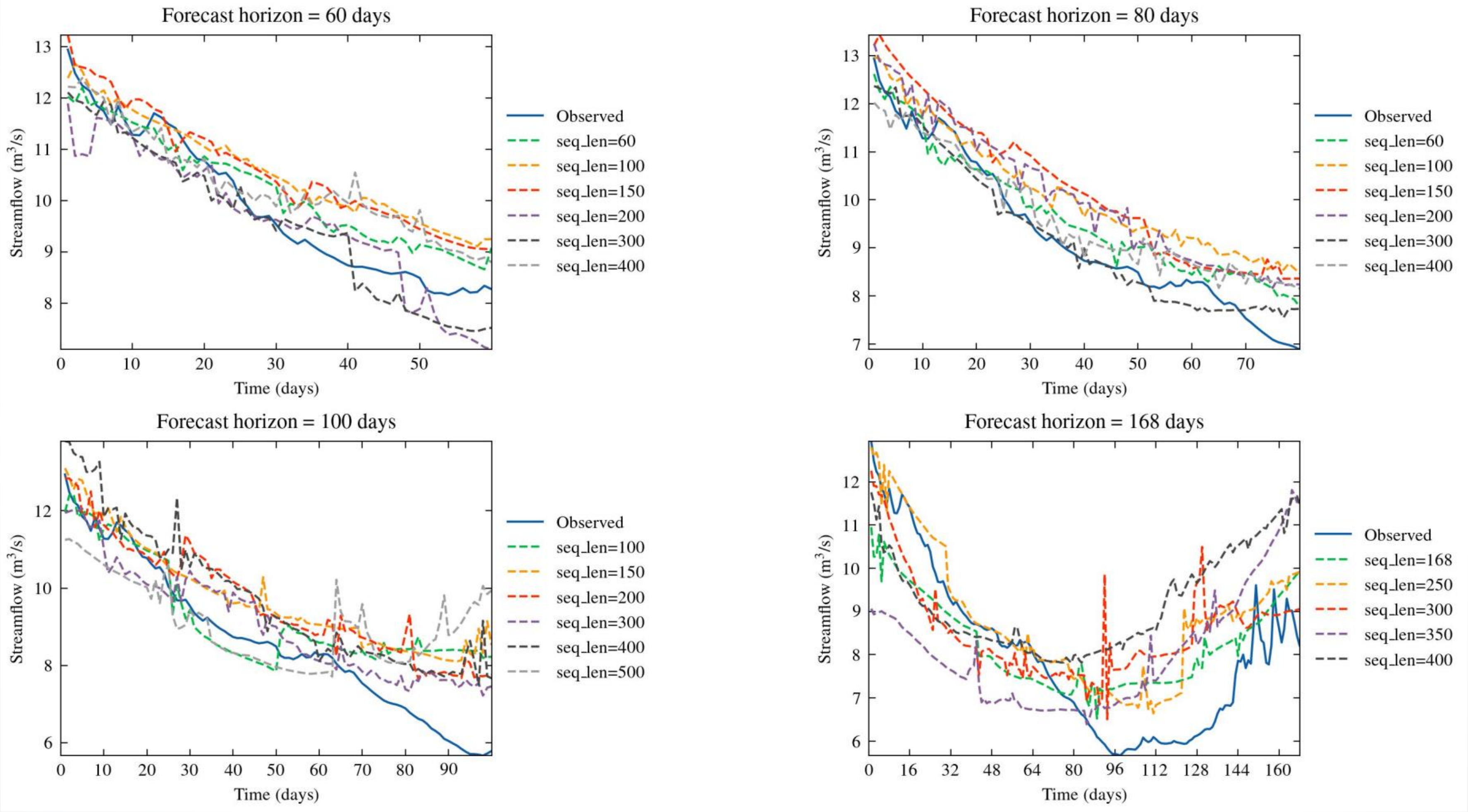
Informer optimization (1/2)

- Results on different scenarios for each forecast horizon.



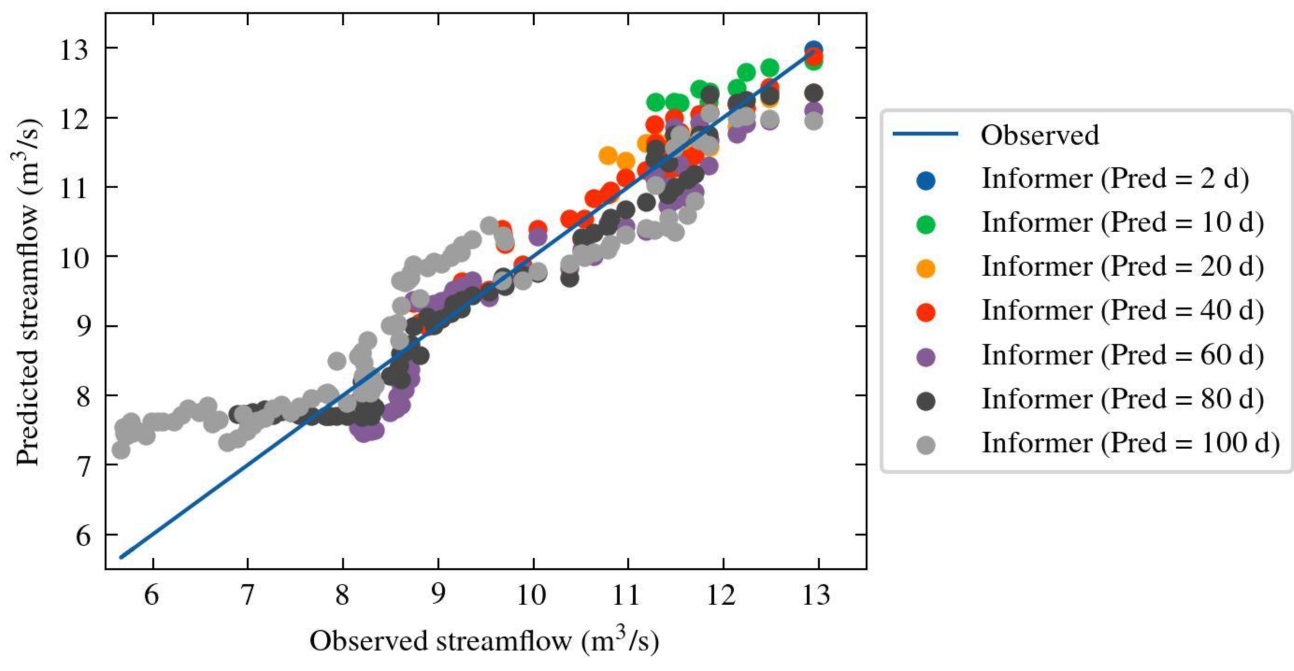
Informer optimization (1/2)

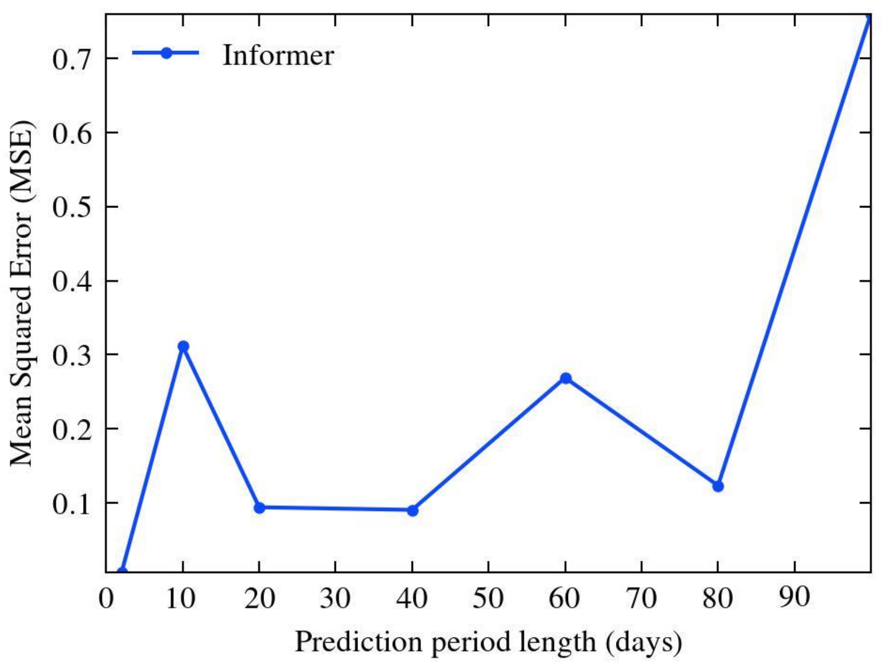
- Results on different scenarios for each forecast horizon.



Informer best results

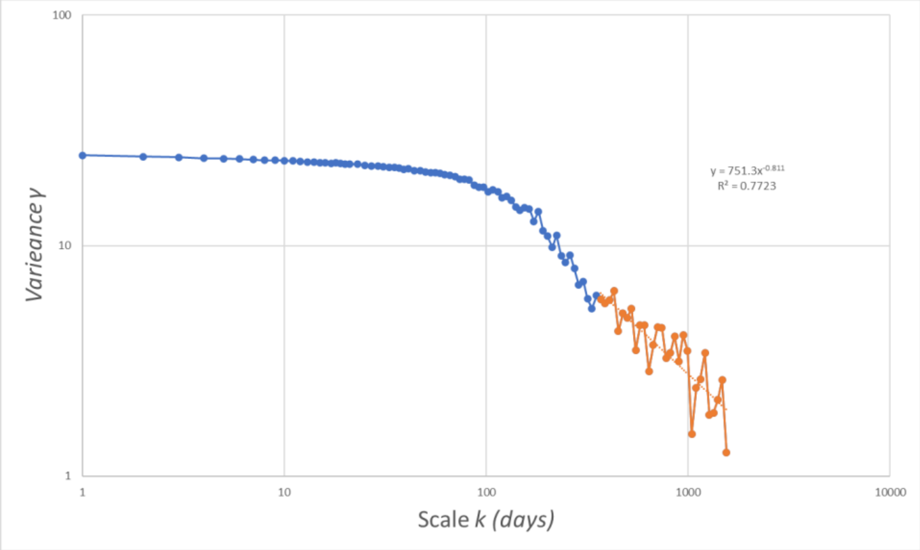
- Results of Informer model for which we achieved the best MSE score in every forecast horizon. Generally results are fairly satisfactorily, with the MSE being quite higher on 100 steps ahead.





Stochastic approaches

- SB1 assumptions: only the long-dependence was taken into account and not the fully climacogram, with a Hurst parameter $H = 0.7$.
- SB2 assumptions: parameters were taken same as in a similarly work of streamflow climacogram (Dimitriadis et al., 2022), with a Hurst parameter $H = 0.8$.



- SB1: the relation of future time steps κ, with the number of past time steps ν we use to calculate the average for our prediction.

κ (days)	2	10	20	40	60	80	100	168
ν (days)	64	320	640	1280	1920	2560	3200	5376

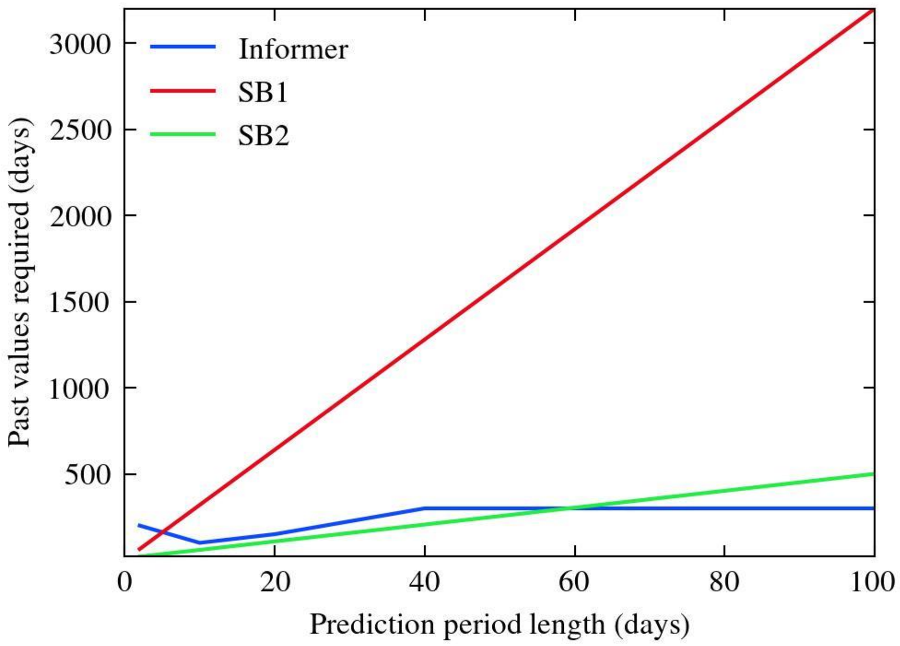
- SB2: the relation of future time steps κ, with the number of past time steps we use calculate average for our prediction.

κ (days)	2	10	20	40	60	80	100	168
ν (days)	20	59	108	206	304	402	500	803

✓ So, the larger the H parameter gets, the less past value is required for the prediction.

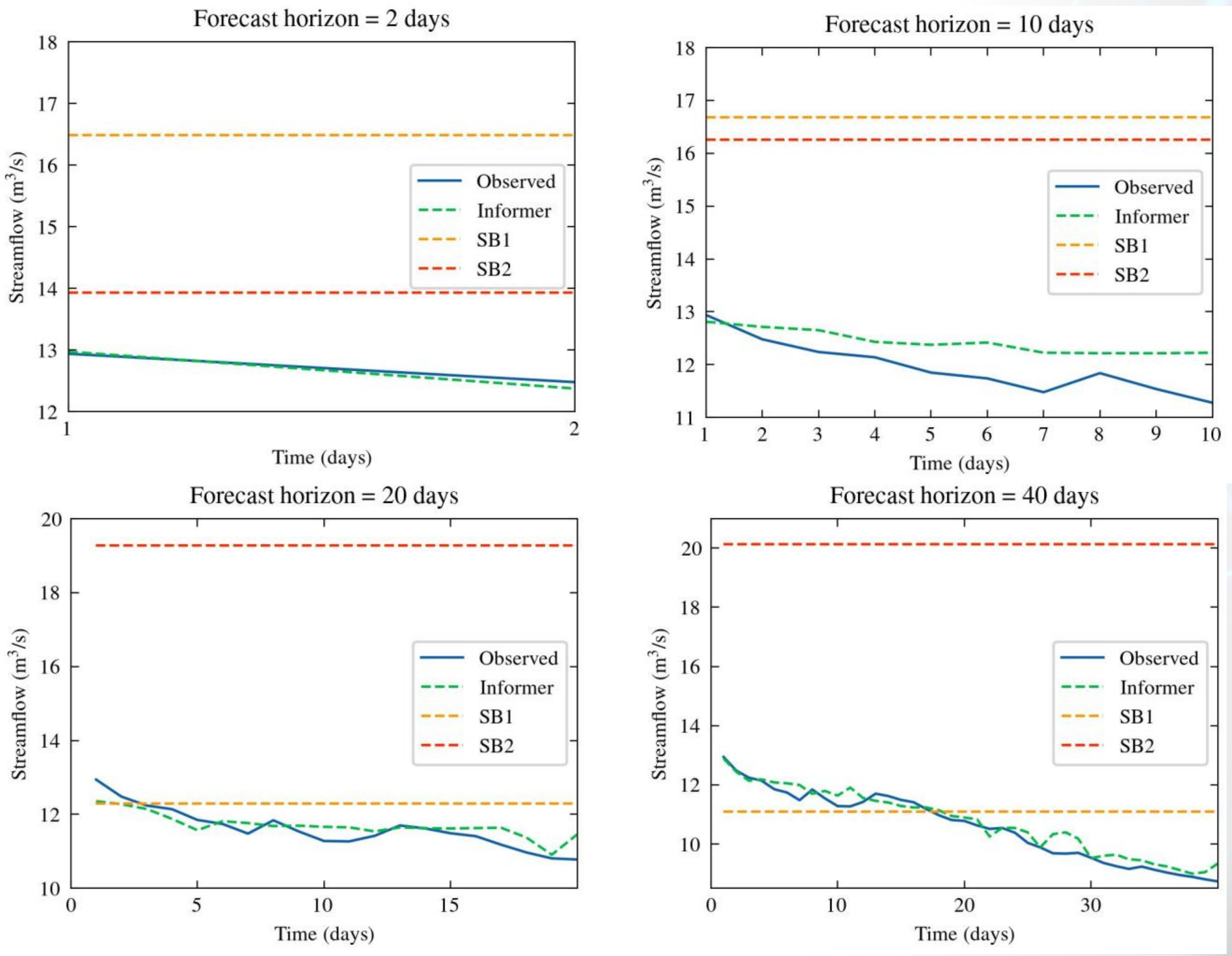
Predictability window

- Comparing the predictability time windows of the 3 models.
- Informer has similar behavior with SB2, while SB1 take into account more past values for its prediction.



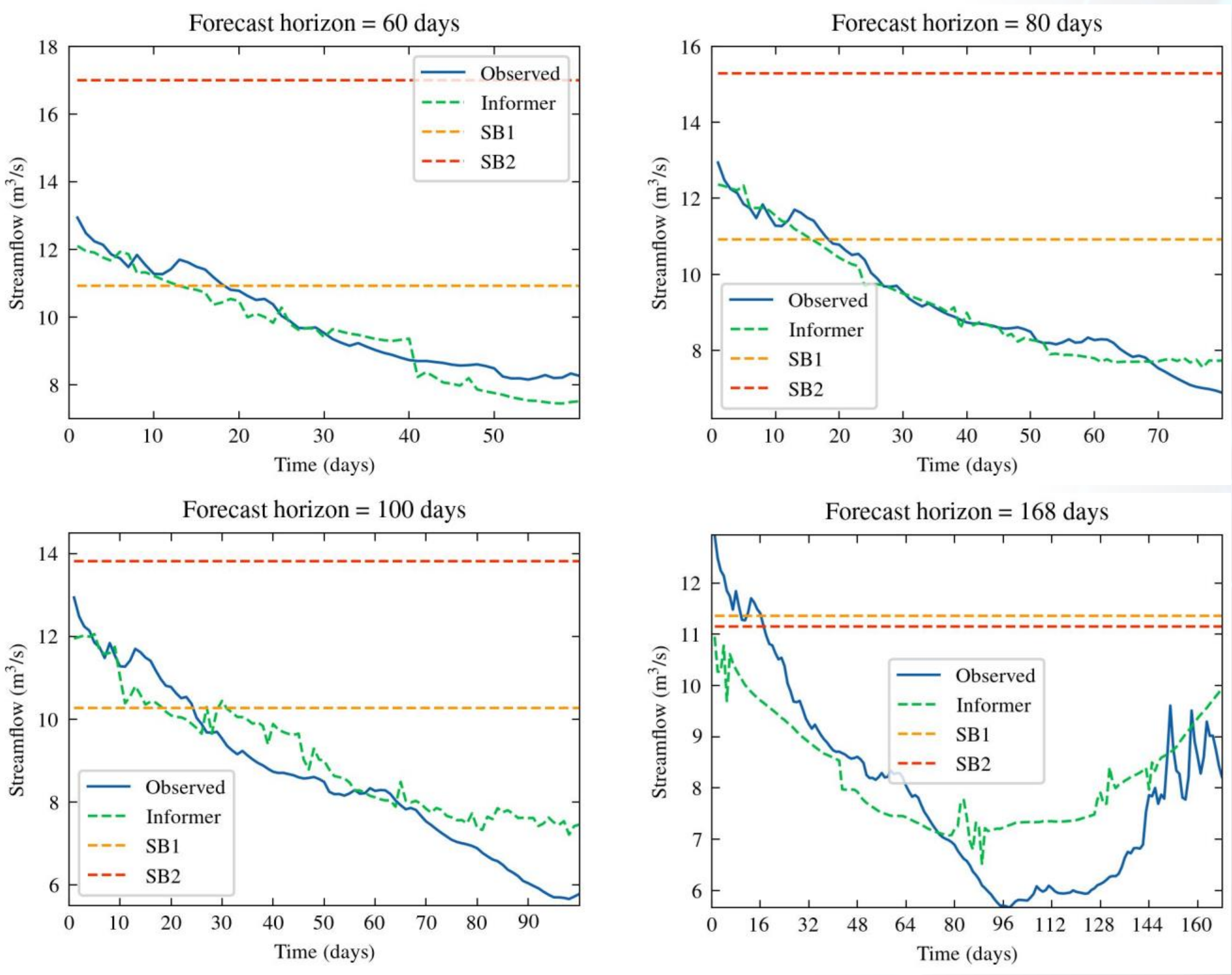
Comparison of results (1/2)

- Informer seems to outperform the two Stochastic models, while SB1 (taking into account more past values) having better behavior than SB2.



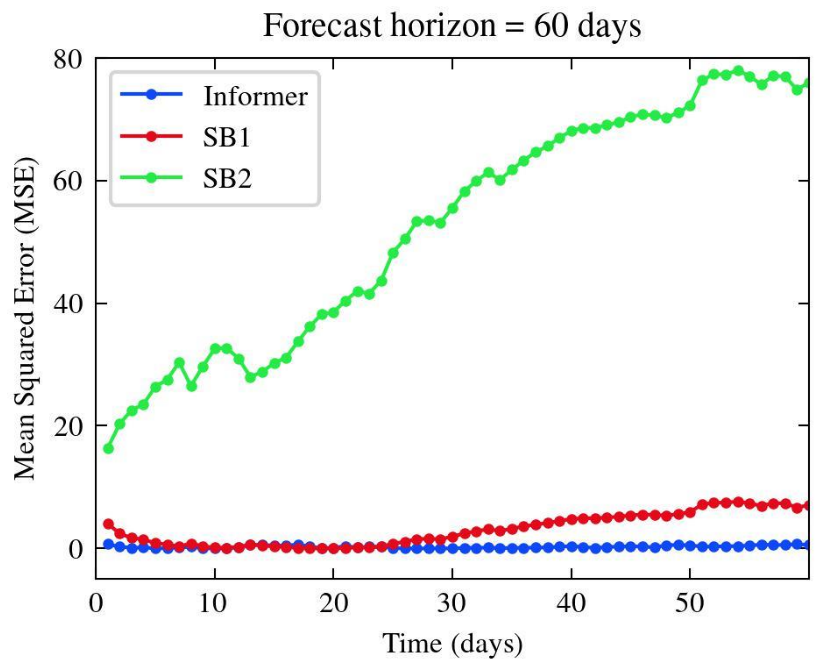
Comparison of results (2/2)

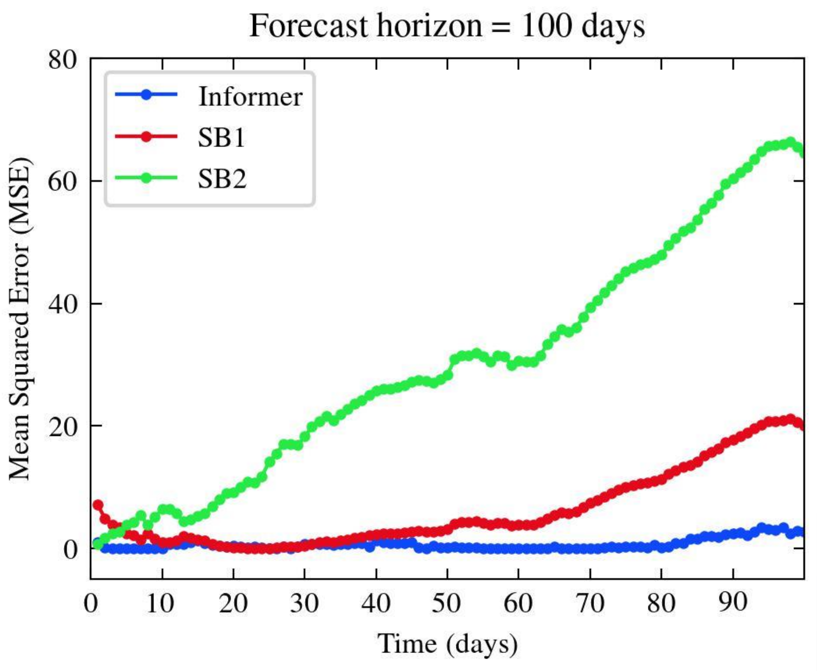
- Informer seems to outperform the two Stochastic models, while SB1 (taking into account more past values) having better behavior than SB2.



MSE on each step ahead

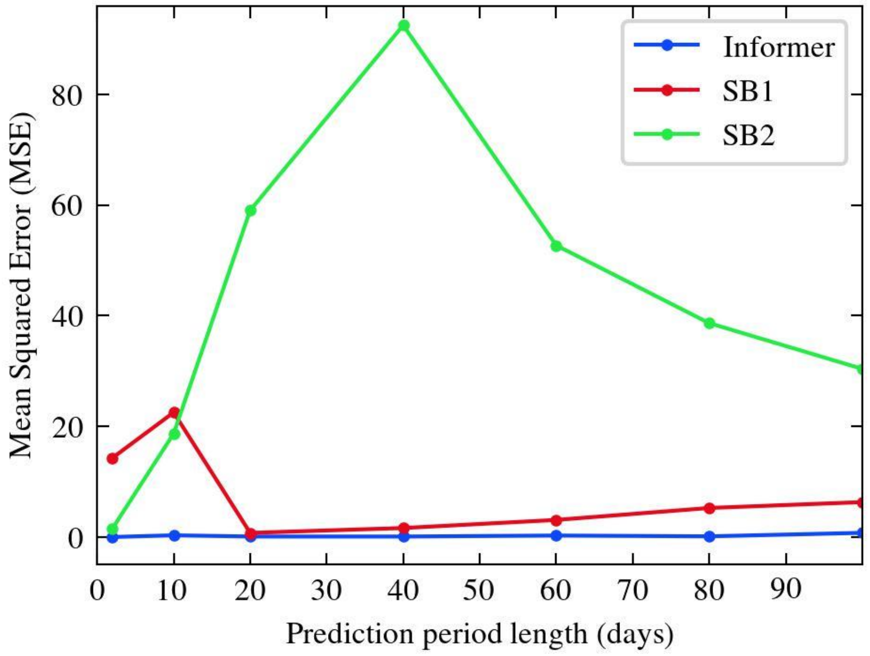
- As depicted in figures the prediction MSE is increasing as we move forward in time, for all the models.





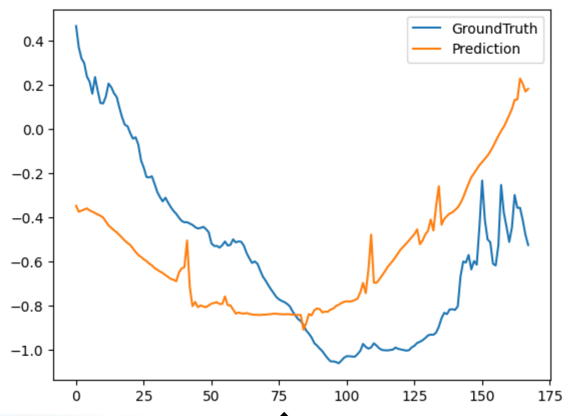
MSE compared to prediction length

- ❑ Informer model having approximately the same error in forecast horizons 2-80 days.
- ❑ SB1 and SB2 MSE show more fluctuation with respect to prediction horizon.

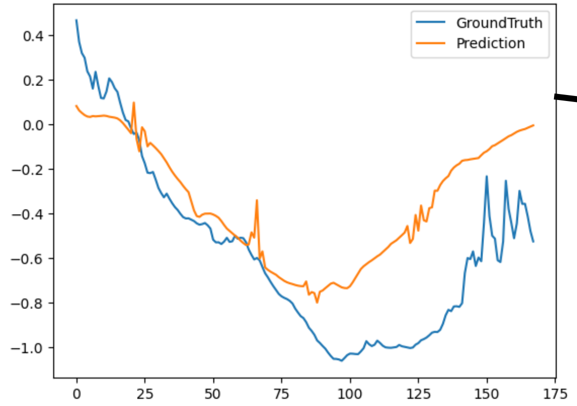


Multivariate forecasting

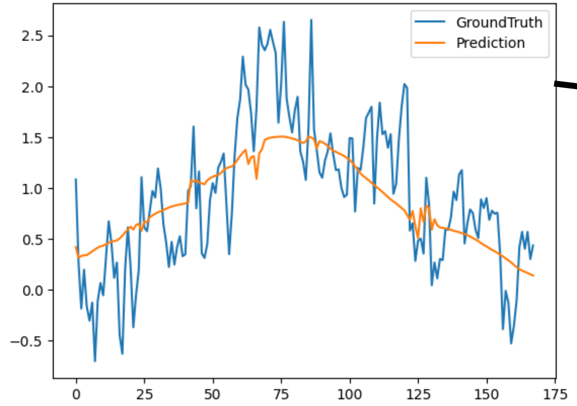
- Informer also supports multivariate time series forecasting, allowing making better predictions, when the forecasting task provides one or more auxiliary (explanatory) variables.



Results for univariate streamflow predictions



Results for multivariate streamflow predictions



Predictions of temperature (auxiliary variable)

Summary and Conclusions

1) Informer model is able for efficient multi-scale forecasting, both on short and long sequence forecasting.
2) Informer can generate the output sequence in a single forward pass and is suitable for both univariate and multivariate time series predictions.
3) From this work Informer seem to have great performance and significantly better than stochastic benchmark ones.
4) Stochastic approach is simple and theoretically substantiated on stochastic processes, appropriate for using them as benchmark.
5) Transformer technology shows great potential in improving the accuracy of time series prediction models.
6) Some cons of transformer procedure is the computational complexity, limited interpretability and scalability. In this regard, each time series should be approached and evaluated independently.

End of Presentation

Thank you for your 'Probabilistic' attention!

References

- Koutsoyiannis, D., Stochastics of Hydroclimatic Extremes—A Cool Look at Risk, Edition 0, National Technical University of Athens: Athens, Greece, 330p., 2021.
- P. Dimitriadis, Iliopoulou, T., and Koutsoyiannis, D., Theoretical framework for the stochastic synthesis of the variability of global-scale key hydrological-cycle processes and estimation of their predictability limits under long-range dependence, *IAHS 100th Anniversary – 11th IAHS-AISH Scientific Assembly 2022*, Montpellier, France, IAHS2022-610, International Association of Hydrological Sciences, 2022.
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106-11115. https://doi.org/10.1609/aaai.v35i12.17325.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In NIPS, 5998–6008.
- Data of River Test: https://nrfa.ceh.ac.uk/ , https://environment.data.gov.uk/hydrology/doc/reference.