Science Time Series:
Deep Learning in Hydrology

Junyang He
Prof. Geoffrey Fox

Background

AI for Science

- Traditional science is driven by theories and differential equations. Predictions are made based on these established formulas.
- New approach to science is driven by big data. Sophisticated DL models can learn the “hidden variables” and find the complex formula behind a natural phenomenon.

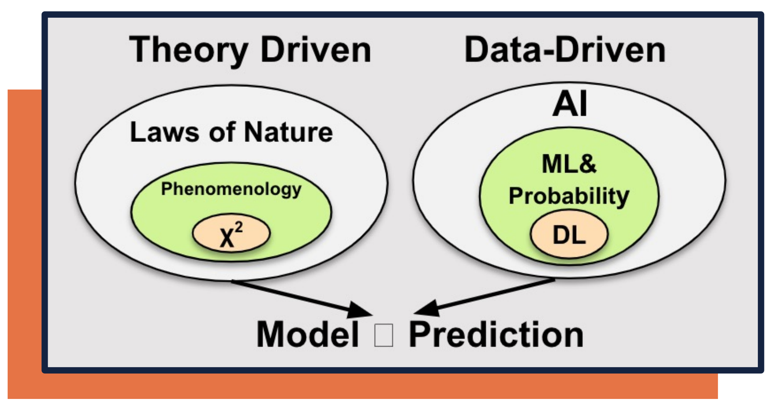

Fig. 1 Theory driven vs. data-driven science.

Science Time Series (Spatial Bag)

- **Science Time Series** consist of time dependent data collected at different geographic locations along with static features of those locations.
- The **Spatial Bag** [Fig. 2] describes the structure of Science Time Series data.
- Studies analyzing Science Time Series data can focus on forecasting the future or predict one target from the other (seq 2 seq).

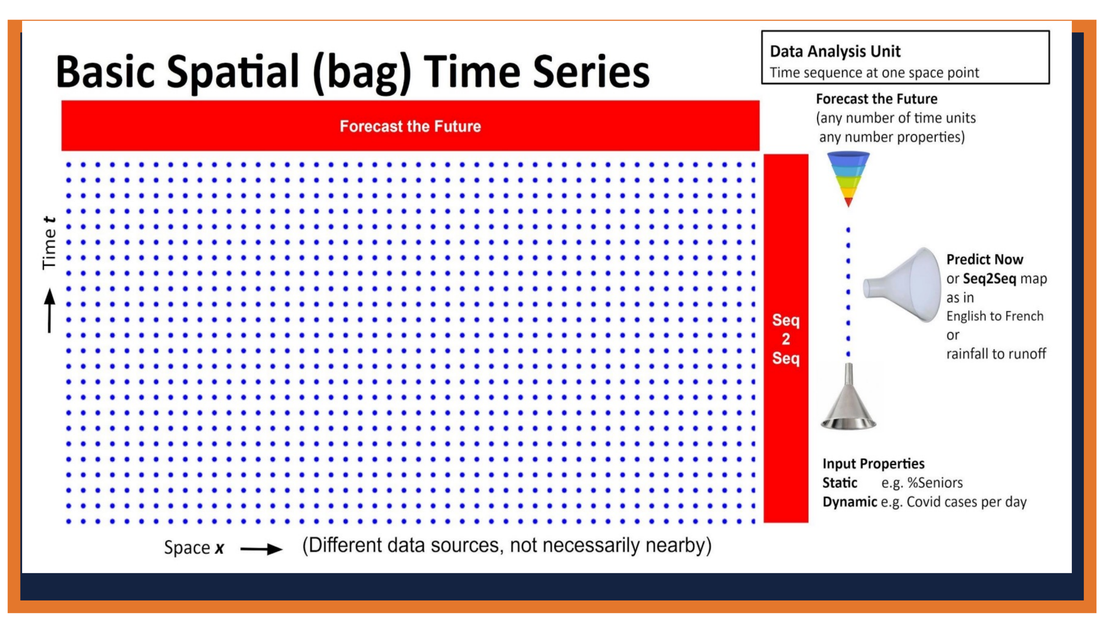

Fig. 2 Spatial bag structure.

Rainfall-Runoff Modeling

- Model the relationship between precipitation and streamflow at various catchments [Fig. 3].
- Perform global-scale modeling with Deep Learning and big data.
- **Catchment** – a place where water aggregates (Eg. lakes, rivers).
- **Streamflow** – displacement of water in and out of a catchment.

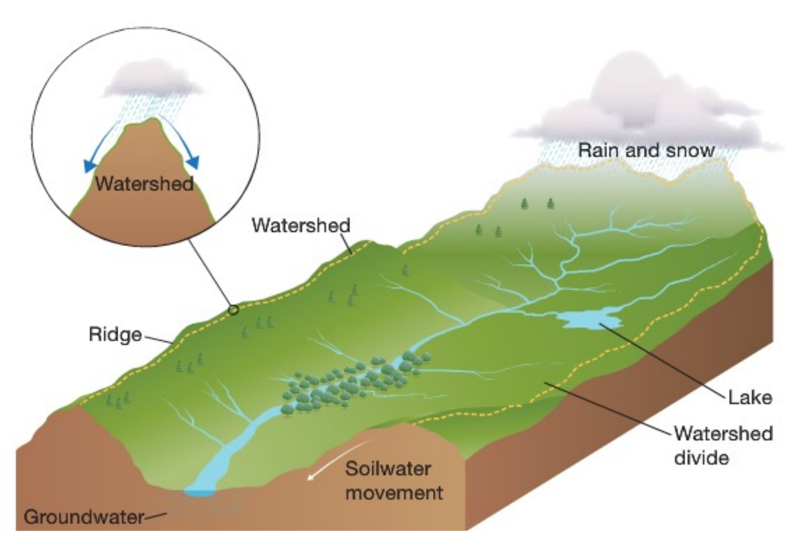

Fig. 3 Water cycle.

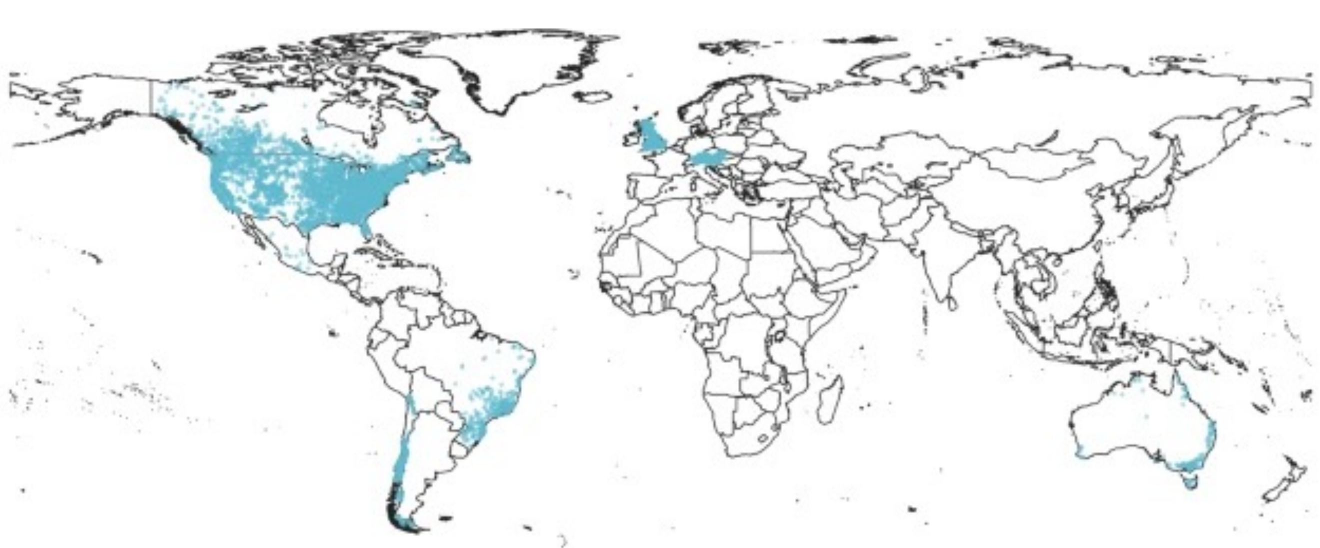

Fig. 4 CARAVAN data coverage.

Hydrology Data

- Dataset 1: **CAMELS** (1858 catchments, 3 nations).
- Dataset 2: **Caravan** (6830 catchments, 4 continents) [Fig. 4].
- 20 years of daily time series data (Eg. precipitation, temperature, streamflow).
- Static features for each catchment (Eg. coordinates, soil texture, vegetation).

LSTM Model

- Special type of RNN network commonly used in Time Series Analysis.
- Model [Fig. 5] consists of a Dense encoder layer, two LSTM layers, and a Dense decoder layer, all with layer size of 320.
- MSE loss function, evaluated with Normalized Nash Sutcliffe Efficiency (NNSE).

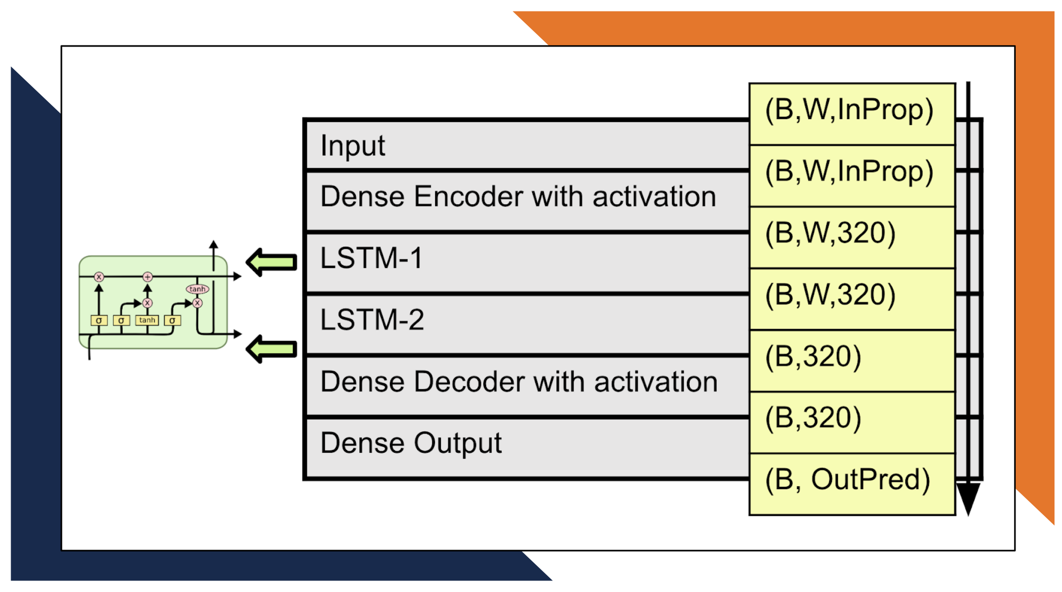

Fig. 5 LSTM Model setup.

Benchmark Model Training

- **Input:** time series targets (excluding streamflow) + static features.
- **Prediction:** time series targets (including streamflow).
- **Encodings:** linear spatial temporal encodings + annual Fourier temporal encodings.

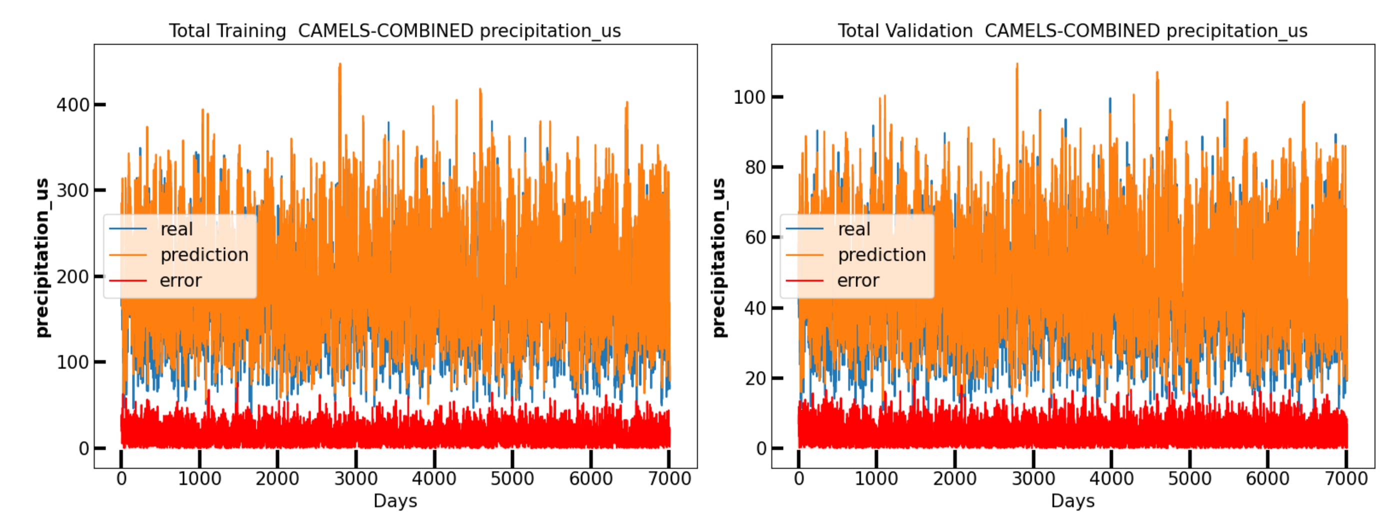

Fig. 6 CAMELS combined train / validation fit.

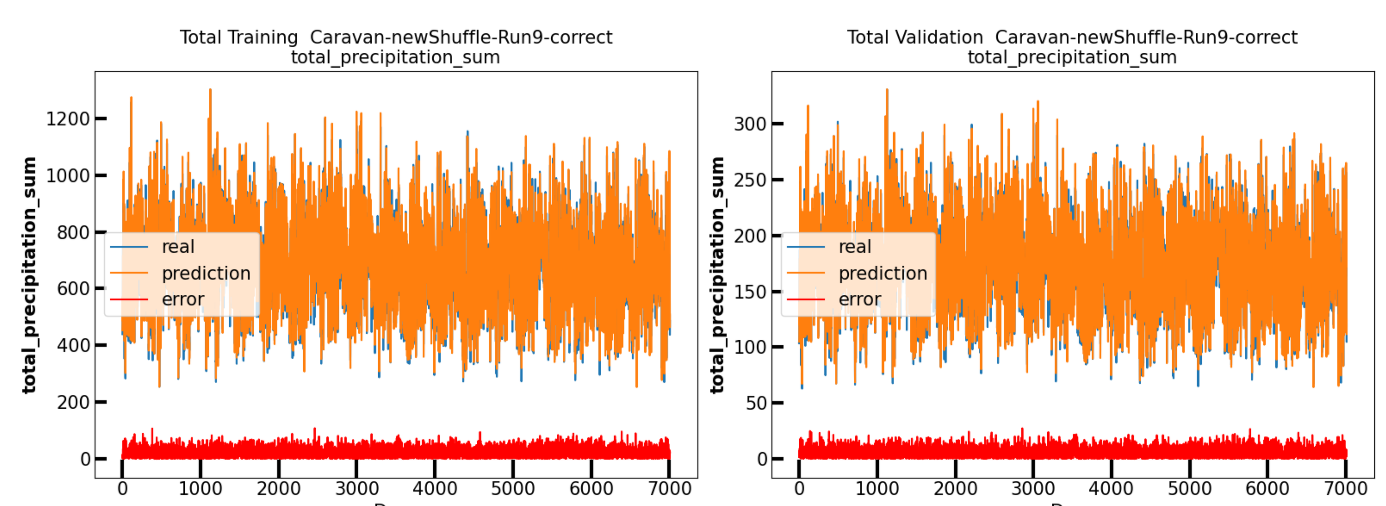

Fig. 7 Caravan combined train / validation fit.

Static Encoding Experiment

- Tested the impact of input static features and encodings on model training.
- Observed that spatial-temporal encodings play a larger role on Science Time Series training accuracy than input static features.
- Found that model trained with input static features obtained from Principal Component Analysis, a common dimensionality reduction technique, performs just as well as model trained from raw static features.

Run Number	1	2	3	4	5
6 Time series	x	x	x	x	x
Linear space-time		x	x	x	x
Annual Fourier time			x	x	x
27 Static				x	x
11 extra Fourier, Legendre in time					x

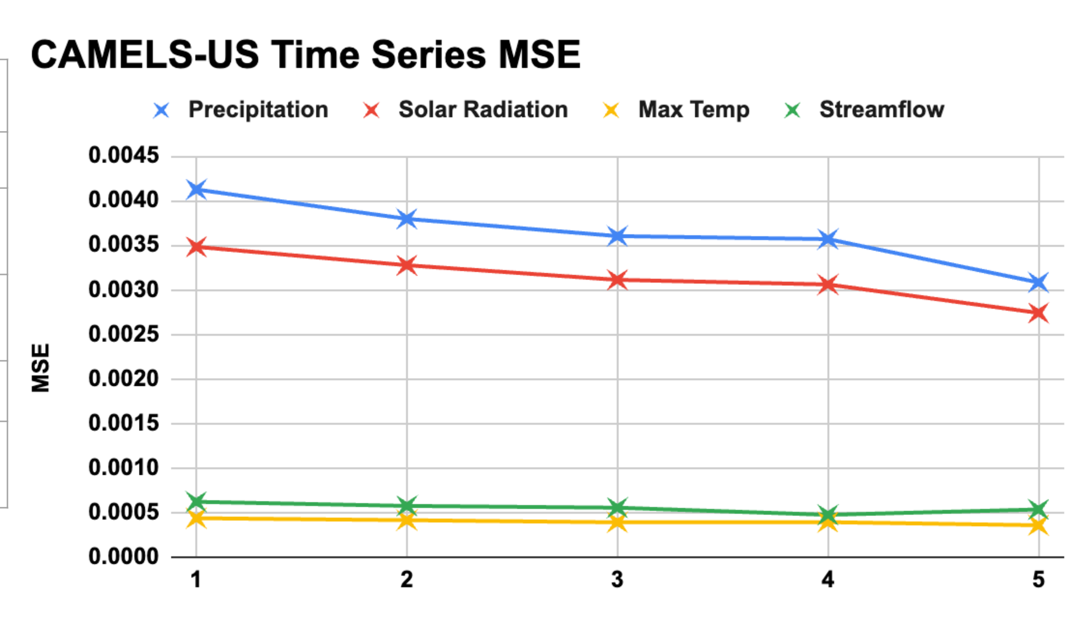

Fig. 8 Static Encoding Experiment.

Future Work

- Examine new time series large foundation models and test model fit on CAMELS / Caravan Hydrology data.
- Compare foundation model approach to traditional LSTM approach.

References

- C. K. R. P. K. W. M. M. Sitterson, J. and B. Avant, *An Overview of Rainfall-Runoff Model Types*. U.S. Environmental Protection Agency, Washington, DC, 2017.
- Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- G. Fox, J. Rundle, A. Donnellan, and B. Feng, “Earthquake nowcasting with Deep Learning,” *arXiv.org*, 18-Dec-2021, https://arxiv.org/abs/2201.01869.
- F. Kratzert, M. Herrnegger, D. Klotz, S. Hochreiter, and G. Klambauer, NeuralHydrology – Interpreting LSTMs in Hydrology. Cham: Springer International Publishing, 2019, pp. 347–362. [Online]. Available: https://doi.org/10.1007/978-3-030-28954-619
- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017.
- F. Kratzert, G. Nearing, N. Addor, T. Erickson, M. Gauch, O. Gilon, L. Gudmundsson, A. Hassidim, D. Klotz, S. Nevo, G. Shalev, and Y. Matias, “Caravan - a global community dataset for large-sample hydrology,” *Scientific Data*, vol. 10, no. 1, p. 61, Jan. 2023. [Online]. Available: https://doi.org/10.1038/s41597-023-01975-w