# Report: A Data Scientist’s Guide to Stream-flow Prediction

## Introduction

* Title: "A Data Scientist’s Guide to Streamflow Prediction"
* Authors: Martin Gauch and Jimmy Lin
* Affiliation: David R. Cheriton School of Computer Science, University of Waterloo, ON, Canada
* Abstract Summary:
  + Data-driven science in geophysical disciplines, especially hydrology.
  + Focus on hydrologic rainfall-runoff models for flood forecasting and streamflow prediction.
  + Aims to help data scientists understand the problem, hydrologic concepts, and key details.

## The Problem in a Nutshell

* Overview of streamflow prediction:
  + Predicting water flow in rivers based on past streamflow, meteorological variables, and geophysical data.
  + Temporal granularity (hours to years) and its significance.
* Importance of accurate streamflow predictions during extreme events and climate change.

## A Lot More Nuance

* Understanding what is predicted:
  + Delineation of catchments, gauged vs. ungauged basins, nested sub-basins.
* Differentiating between runoff and stream-flow, their significance, and measurement uncertainties.

## Understanding the Types of Input Data

* Forcing and geophysical data:
  + Forcing: Time series of meteorological data (e.g., precipitation, temperature).
  + Geophysical data: Static information (e.g., land cover, soil, elevation).
* Handling spatial distribution of data for lumped, semi-distributed, and distributed models.
* Mention of publicly available stream-flow datasets, such as CAMELS.

## Understanding How they’re Actually Predicting

* Spectrum of Models:
  + Process-Based Models: Modeling hydrologic processes.
  + Data-Driven Models: Machine learning models for predictions.
* Classification based on spatial distribution: lumped, semi-distributed, distributed models.
* Model transferability in time (simulation, forecasting, hindcasting) and space (spatial validation).

**Report: Fourcastnet: A global data driven high resolution weather model using adaptive Fourier neural operators**

Title: FOURCASTNET: Revolutionizing Weather Forecasting with High-Resolution Data-Driven Models

**Summary:** FourCastNet is a groundbreaking global data-driven weather forecasting model that provides high-resolution predictions. It offers vital insights into forecasting surface wind speed, precipitation, and atmospheric water vapor. FourCastNet matches and surpasses existing models' accuracy for different variables, enabling faster forecasts. This model is invaluable for various applications, from wind energy planning to predicting extreme weather events.

**Introduction:**

* The history of numerical weather prediction and its critical role in various sectors.
* The evolution of weather forecasting from manual calculations to electronic computers.
* The recent emergence of data-driven Deep Learning (DL) models as a promising addition to traditional Numerical Weather Prediction (NWP) models.

**Training Methods:**

* Introduction to the ERA5 dataset provided by the ECMWF, a comprehensive atmospheric reanalysis dataset.
* Focus on training FourCastNet to forecast near-surface wind velocities and 6-hourly total precipitation.
* Highlight the practical significance of these variables and the model's potential in overcoming limitations present in NWP models.

**Results:**

* FourCastNet's forecasting skill for surface wind speeds at a 0.25◦ resolution.
* Emphasis on the model's ability to forecast wind speeds accurately up to 96 hours in advance.
* The remarkable performance of FourCastNet in tracking extreme weather events, such as super-typhoons and hurricanes.

**Conclusion:**

* A concise summary of the report's key findings.
* Emphasis on FourCastNet's potential to revolutionize weather forecasting with its high-resolution and high-speed capabilities.
* The model's implications for various sectors, including energy planning and disaster preparedness.

**FOURCASTNET: DATA-DRIVEN HIGH-RESOLUTION ATMOSPHERIC MODELING AT SCALE**

Title: FOURCASTNET: Data-Driven High-Resolution Atmospheric Modeling at Scale

**Introduction**: FOURCASTNET, or Fourier ForeCasting Neural Network, represents a significant leap forward in the field of atmospheric modeling. It is a state-of-the-art (SOTA) deep-learning-based weather emulator with the potential to revolutionize global weather forecasting and climate modeling. This report delves into the core aspects and implications of FOURCASTNET, which is designed to address the challenging demands of climate science and weather prediction.

**Challenges in Climate Science and Weather Prediction:**

* Model Complexity: Climate science involves intricate physical processes described by hundreds of partial differential equations (PDEs) and complex parameterizations.
* Computational Cost: Achieving high resolution to capture fine-scale atmospheric phenomena incurs exponential increases in computational power requirements.
* Scalability and Performance: Current models are not optimized to harness modern supercomputing substrates, such as GPUs, and consume substantial energy resources.

**Data-Driven Solutions for Atmosphere Dynamics:**

* Model Complexity: FOURCASTNET leverages data-driven surrogates to overcome model biases and learn parameterizations, thus simplifying the model complexity.
* Computational Cost: By exploiting GPU compute through deep learning-based surrogates, fourcastnet significantly improves performance and reduces computational cost.
* Scalability and Performance: Optimized implementations enable scalable AI through data and model parallelism, addressing issues of scalability and performance.

**Key Features of FOURCASTNET:**

* State-of-the-Art Data-Driven Model: FOURCASTNET is a cutting-edge data-driven model capable of simulating and predicting crucial atmospheric variables with exceptional skill.
* Significant Computational Speedup: FOURCASTNET enables inference at a speed 80,000 times faster than traditional models, facilitating the use of larger ensembles.
* Scalable to Exascale Computing: This model scales to around 4,000 GPUs, paving the way for exascale weather and climate computing.

**Variables Utilized by FOURCASTNET:**

* Forecasting Strategy: FOURCASTNET adopts a strategy where it forecasts atmospheric variables by recursively advancing in time. It initially focuses on a subset of atmospheric variables to characterize the state and plans to expand to include all variables.
* Input Vector X(t): The model's input sample is an array with dimensions of 21 x 720 x 1440, which is used both for training and inference. It predicts the target X(t + dt) from X(t) in an autoregressive forward step in time.

**Conclusion:** FOURCASTNET offers a promising solution to the challenges faced in climate science and weather prediction. By combining the power of data-driven models, high-resolution simulations, and improved performance, it unlocks the potential for profound advancements in atmospheric science. FOURCASTNET's scalability and exceptional computational speed make it a vital tool for addressing complex climate and weather dynamics, ultimately enabling more well-informed actions and policies.

**Earthformer: Exploring Space-Time Transformers for**

**Earth System Forecasting**

**Summury:** Earthformer, a groundbreaking space-time Transformer for Earth system forecasting, is introduced as a powerful alternative to conventional physical simulation-based models. While Earth system forecasting has traditionally relied on complex numerical models, Earthformer leverages Deep Learning and the Transformer architecture to make predictions based on large-scale Earth observation data. This report explores the key innovations behind Earthformer, its impressive performance, and its potential impact on the field of Earth system forecasting.

**Introduction:** The Earth's complex system with its various weather and climate patterns profoundly affects our daily lives. Accurate forecasting of Earth system variables like weather and climate has significant socioeconomic implications. However, the operational models for Earth system forecasting have not significantly evolved in the past five decades, constraining their ability to incorporate emerging geophysical observations and large-scale Earth observation data. Deep Learning (DL) offers an alternative approach, training models on Earth observation data to learn intrinsic physical rules and improve predictions. While DL has shown promise in weather forecasting applications, the complex and chaotic nature of Earth systems necessitates new DL architectures.

**Challenges in Earth System Forecasting:** Earth systems are highly chaotic, high-dimensional, and spatiotemporal, making accurate forecasting challenging. Traditional DL models rely on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to capture spatiotemporal patterns. However, these architectures may not fully capture the complexities of Earth systems, as they assume the persistence of spatial and temporal inductive biases.

**The Potential of the Transformer Architecture:** The Transformer architecture, originally designed for natural language processing, has made significant advances in various domains. However, its direct application to Earth system forecasting poses challenges due to the O(N^2) attention complexity and high-dimensional Earth observation data.

**Introducing Earthformer:** Earthformer is proposed as a space-time Transformer tailored to Earth system forecasting. Key innovations include:

* **Cuboid Attention:** To address the attention complexity challenge, Earthformer employs Cuboid Attention. It decomposes the input data into non-overlapping cuboids and applies cuboid-level self-attention in parallel. This significantly reduces the overall complexity while capturing different correlations.
* **Global Vectors:** Earthformer introduces global vectors that attend to all local cuboids, enabling them to understand the general dynamics of the system and share information.

**Experimental Validation:** Earthformer's design is validated through experiments on synthetic datasets, including MovingMNIST and a new N-body MNIST dataset. The results highlight:

* **Cuboid Attention:** Stacking cuboid attention layers with the Axial attention pattern proves to be efficient and effective.
* **Global Vectors:** The addition of global vectors consistently improves performance without increasing computational cost.
* **Encoder-Decoder Hierarchy:** Adding hierarchy to the architecture enhances performance.

**State-of-the-Art Performance:** Earthformer is compared with other baselines on real-world benchmarks for precipitation nowcasting and ENSO forecasting, SEVIR and ICAR-ENSO. The experiments show that Earthformer achieves state-of-the-art performance on both tasks.

**Broader Impact and Future Directions:** Earthformer's deterministic nature can be enhanced to model uncertainty, and the incorporation of physical knowledge into the model is an exciting research direction for the future. Earthformer represents a significant advancement in Earth system forecasting, leveraging cutting-edge DL techniques to improve the accuracy and reliability of weather and climate predictions.

**Conclusion:** Earthformer's introduction demonstrates a remarkable step forward in the field of Earth system forecasting. Its innovative use of the Transformer architecture, coupled with efficient attention mechanisms, showcases its potential to revolutionize the way we forecast Earth system variables. By addressing the challenges in Earth system forecasting, Earthformer holds promise in improving the accuracy of critical predictions, benefiting both research and society at large.

**Flood forecasting with machine learning models in an**

**operational framework**

**summury:** Google's operational flood forecasting system, launched in 2018 with a primary focus on riverine floods in large, gauged rivers, has since expanded geographically and showcased the potential of machine learning in flood prediction and warning systems. The system encompasses four key subsystems: data validation, stage forecasting, inundation modeling, and alert distribution, with machine learning applied to two of these. Stage forecasting utilizes Long Short-Term Memory (LSTM) networks and Linear models, while flood inundation modeling leverages Thresholding and Manifold models, with all models demonstrating high-performance metrics during historical data evaluation. The system's operational deployment in India and Bangladesh during the 2021 monsoon season was instrumental in issuing over 100 million flood alerts to safeguard affected populations and territories. Ongoing and future endeavors include expanding system coverage to additional flood-prone areas and enhancing modeling capabilities for improved flood forecasting.

**Introduction:** Floods are a global natural threat causing substantial fatalities and economic damages annually. Operational flood warning systems play a pivotal role in reducing risks and damages in regions prone to flooding. Google's Flood Forecasting Initiative, launched in 2017, aims to leverage machine learning to provide highly accurate flood forecasts on a global scale. This paper presents Google's operational flood warning system and its deployment in India and Bangladesh during the 2021 monsoon season, highlighting the use of machine learning in two crucial modeling components: river stage forecasting and flood inundation modeling.

**Conclusions:** Google's operational flood forecasting system has proven to be a lifeline for populations residing in flood-prone areas. During the 2021 monsoon season, the system issued over 100 million flood alerts, directly benefiting individuals, relevant agencies, and emergency organizations. Machine learning models for river stage forecasting, particularly the LSTM model, outperformed conventional linear models. Flood inundation modeling using Thresholding and Manifold models showcased their efficacy, with the Thresholding model outperforming hydraulic models and exhibiting practical advantages in terms of computational resources. Continuous improvement in the system's effectiveness and a dedication to operational performance assessment are essential for its ongoing success, contributing to the enhancement of flood warning systems and scientific knowledge in this domain.

Google's operational flood forecasting system, initiated in 2018 with a primary focus on riverine floods in large, gauged rivers and later expanded geographically, is a remarkable example of leveraging machine learning in flood prediction and warning. Comprising four integral subsystems - data validation, stage forecasting, inundation modeling, and alert distribution - the system incorporates machine learning techniques in two specific domains. The stage forecasting component utilizes Long Short-Term Memory (LSTM) networks and Linear models, while flood inundation modeling employs Thresholding and Manifold models. These models have exhibited high-performance metrics when assessed on historical data, making them well-suited for operational use. During the 2021 monsoon season, the flood warning system was deployed in India and Bangladesh, covering flood-prone regions encompassing a vast area and population. Over 100 million flood alerts were issued, underscoring the system's significant real-world impact. Ongoing and future work includes expanding the system's coverage to additional flood-prone regions and enhancing modeling capabilities to further improve accuracy and performance in flood forecasting.

Keywords: Riverine floods, operational flood warning system, machine learning, stage forecasting, flood inundation, India, Bangladesh.

**A Rainfall‐Runoff Model With LSTM‐Based Sequence‐to‐Sequence Learning**

**Title**: A Rainfall-Runoff Model with LSTM-Based Sequence-to-Sequence Learning

**Summary**: This report presents a study on predicting hourly rainfall-runoff using advanced deep learning techniques, specifically the Long Short-Term Memory (LSTM) model and the sequence-to-sequence (seq2seq) structure. The research focuses on two Midwestern watersheds, Clear Creek and Upper Wapsipinicon River in Iowa, and evaluates the model's performance against traditional machine learning models. The LSTM-seq2seq model outperforms other methods and shows great potential for short-term flood forecasting.

**Introduction**: Rainfall-runoff modeling is crucial for understanding and predicting floods and their environmental impact. Various modeling approaches, from physical models to machine learning methods like support vector machines and artificial neural networks (ANNs), have been employed. This report explores the application of LSTM and seq2seq learning to improve hourly rainfall-runoff predictions, aiming to enhance flood preparedness and response.

**Conclusion**: The LSTM-seq2seq model, designed for continuous hourly rainfall-runoff predictions, has demonstrated its effectiveness. It surpasses traditional models and shows potential for application in different watersheds, particularly those with limited physical data. The report highlights the promise of deep learning techniques in hydrological applications.