Preparing your manuscript

Florian Börgel[1], Sven Karsten[1]

[1]Leibniz-Institute for Baltic Sea Research Warnemünde,

Corresponding author: Florian Börgel, `florian.boergel@io-warnemuende.de`

Abstract

The abstract (1) states the nature of the investigation and (2) summarizes the important conclusions. The abstract should be suitable for indexing. Your abstract should:

- Be set as a single paragraph.
- Be less than 250 words for all journals except GRL, for which the limit is 150 words.
- Not include table or figure mentions.
- Avoid reference citations unless dependent on or directly related to another paper (e.g., companion, comment, reply, or commentary on another paper(s)). AGU's Style Guide discusses formatting citations in abstracts.
- Define all abbreviations.

Plain Language Summary

A Plain Language Summary (PLS) can be an incredibly effective science communication tool. By summarizing your paper in non-technical terms, you can explain your research and its relevance to a much broader audience. A PLS is required for submissions to AGU Advances, G-Cubed, GeoHealth, GRL, JAMES, JGR: Biogeosciences, JGR: Oceans, JGR: Planets, JGR: Solid Earth, JGR: Atmospheres, Space Weather, and Reviews of Geophysics, but optional for other journals. A PLS should be no longer than 200 words and should be free of jargon, acronyms, equations, and any technical information that would be unknown to people from outside your scientific discipline. Read our tips for creating an effective PLS.

1 Introduction

River runoff is an important component of the global water cycle as it comprises about one third of the precipitation over land areas (Hagemann, 2020). Moreover, accurate runoff forecasting, especially over extended periods, is pivotal for effective water resources management, as highlighted by studies such as Yang et al. (2018), Tan et al. (2018), and Fang et al. (2019).

Over the past decades, models for long-term runoff forecasting have been primarily bifurcated into physically based models and data-based models. While the former attempts to emulate intricate and nonlinear physical hydrological processes, the latter hinges on establishing statistical models that delineate the relationship between large-scale climate patterns and catchment runoff.

Machine Learning (ML) models, such as those employing artificial neural networks, support vector machines, adaptive neuro-fuzzy inference systems, and notably, Long Short-Term Memory (LSTM) neural networks, have gained traction for long-term hydrological forecasting due to their commendable performance (Humphrey et al 2016, Huang et al 2014, Ashrafi et al 2017, Yuan et al 2018, Xu et al 2021).

LSTM networks, an evolution of the classical Recurrent Neural Networks (RNNs), have shown stability and efficacy in sequence-to-sequence predictions, such as using climatic indices for rainfall estimation or long-term hydrological forecasting. However, a limitation of LSTMs is their inability to effectively capture two-dimensional structures, an area where Convolutional Neural Networks (CNNs) excel. Recognizing this, we introduce the ConvLSTM, which integrates the strengths of both LSTM and CNN, to extract spatiotemporal features from precipitation fields for predicting river runoff in the Baltic Sea catchment, summarized by 97 inidivual rivers.

Modeling the Baltic Sea is to a large part the result of the quality of the freshwater input, that is used for the simulation. Meier and Kauker(2003) showed that decadal salinity variations of about 1 $g\ kg^{-1}$ are caused, inter alia, by annual runoff variations. Further, Meier and Kauker (2003) showed that about 50 % of the decadal

salinity variability can be explained by variations in freshwater input into the Baltic Sea.

This paper delves into the application of deep learning, particularly ConvLSTM, to the challenging task of precipitation nowcasting, a domain yet to fully harness the potential of advanced machine learning techniques. We present ConvLSTM as a novel solution to this spatiotemporal sequence forecasting challenge, highlighting its advantages and potential future applications.

2 Methods

2.1 LSTM network

The Long Short-Term Memory (LSTM), a specialized form of Recurrent Neural Networks (RNNs), is specifically tailored for modeling temporal sequences. Its unique design allows it to adeptly handle long-range dependencies, setting it apart from traditional RNNs in terms of accuracy. This prowess in modeling long-range dependencies has been validated in various studies [12, 11, 17, 23]. The cornerstone of LSTM's innovation is its memory cell, c_tt, which functions as a repository of state information, also refered to as long-term memory. This cell is accessed, modified, and reset through several self-parameterized gates. As new inputs arrive, their information is accumulated into the cell if the input gate is activated. The forget gate f_t defines the percentage of the previous cell status that is stored c_{t-1}. The decision to propagate the latest cell output, c_t, to the final state, h_t, is governed by the output gate, o_t. A significant advantage of this architecture is the memory cell's ability to retain gradients. This mechanism addresses the vanishing gradient problem, where, as input sequences elongate, the influence of initial stages becomes harder to capture, causing gradients of early input points to approach zero. The LSTM's activation function, inherently recurrent, mirrors the identity function with a consistent derivative of 1.0, ensuring the gradient remains stable throughout backpropagation.

One LSTM cell hence maybe expressed as:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \circ c_{t-1} + b_i)$$
$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f)$$
$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \circ c_t + b_o)$$
$$h_t = o_t \circ \tanh(c_t)$$

2.2 ConvLSTM network

The FC-LSTM, while adept in many scenarios, falters in encoding spatial information when handling spatiotemporal data due to its reliance on full connections in both input-to-state and state-to-state transitions. Addressing this limitation, our design ensures that all inputs $X_1, ..., X_t$, cell outputs $C_1, ..., C_t$, hidden states $H_1, ..., H_t$, and gates i_t, f_t, o_t of the ConvLSTM are 3D tensors. The last two dimensions of these tensors represent spatial dimensions, specifically rows and columns. Conceptually, these inputs and states can be visualized as vectors positioned on a spatial grid.

In the ConvLSTM, the future state of a specific cell on this grid is determined by the inputs and past states of its neighboring cells. This spatial consideration is integrated by employing a convolution operator in both state-to-state and input-to-state transitions, as illustrated in Fig. 2. The foundational equations for ConvLSTM are:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i)$$
$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f)$$
$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o)$$
$$H_t = o_t \circ \tanh(C_t)$$

Considering the states as representations of moving entities, a ConvLSTM with a larger transitional kernel can capture rapid movements, while a smaller kernel is apt for slower motions. Drawing parallels with [16], the traditional FC-LSTM's inputs, cell outputs, and hidden states can also be envisioned as 3D tensors, with the last two dimensions being unitary. Thus, the FC-LSTM can be seen as a specific instance of ConvLSTM where all features are concentrated on a singular cell.

To maintain consistency in the dimensions of states and inputs, padding is essential before convolution operations. This padding, especially at boundary points, can be interpreted as leveraging the state of external surroundings for computations. Typically, all LSTM states are initialized to zero before the first input, symbolizing a lack of knowledge about future events. Similarly, zero-padding (as adopted in this study) on hidden states sets the external world's state to zero, implying no prior information about external conditions. This padding approach allows for differential treatment of boundary points, proving advantageous in many scenarios. For instance, in a system with a ball moving and bouncing off unseen walls, the existence of these walls can be inferred by observing the ball's repeated bounces, a deduction challenging if boundary points share the same state transition dynamics as inner points.

3 Acknowledgments

4 Open research

Phasellus interdum tincidunt ex, a euismod massa pulvinar at. Ut fringilla ut nisi nec volutpat. Morbi imperdiet congue tincidunt. Vivamus eget rutrum purus. Etiam et pretium justo. Donec et egestas sem. Donec molestie ex sit amet viverra egestas. Nullam justo nulla, fringilla at iaculis in, posuere non mauris. Ut eget imperdiet elit.

References