Hydrologically Informed Machine Learning for Rainfall-Runoff Modelling: Towards Distributed Modelling

Herath Mudiyanselage Viraj Vidura Herath[1], Jayashree Chadalawada[1], Vladan Babovic[1]

[1]Department of Civil and Environmental Engineering, National University of Singapore, 117576, Singapore

Correspondence to: Vladan Babovic (vladan@nus.edu.sg)

Abstract. Despite showing a great success of applications in many commercial fields, machine learning and data science models in general, show a limited use in scientific fields including hydrology. The approach is often criticized for lack of interpretability and physical consistency. This has led to the emergence of new paradigms, such as Theory Guided Data Science (TGDS) and physics informed machine learning. The motivation behind such approaches is to improve the physical meaningfulness of machine learning models by blending existing scientific knowledge with learning algorithms. Following the same principles, in our prior work (Chadalawada et al., 2020), a new model induction framework was founded on Genetic Programming (GP) namely Machine Learning Rainfall-Runoff Model Induction Toolkit (ML-RR-MI). ML-RR-MI is cable of developing fully-fledged lumped conceptual rainfall-runoff models for a watershed of interest using the building blocks of two flexible rainfall-runoff modelling frameworks. In this study, we extend ML-RR-MI towards inducing semi-distributed rainfall-runoff models. This effort is motivated by the desire to address the decreasing meaningfulness of lumped models which tend to particularly deteriorate within large catchments where the spatial heterogeneity of forcing variables and watershed properties are significant. Henceforth, our machine learning approach for rainfall-runoff modelling titled Machine Induction Knowledge Augmented - System Hydrologique Asiatique (MIKA-SHA) captures spatial variabilities and automatically induces rainfall-runoff models for the catchment of interest without any subjectivity in model selection. Currently, MIKA-SHA learns models utilizing the model building components of two flexible modelling frameworks. However, the proposed framework can be coupled with any internally coherent collection of building blocks. MIKA-SHA's model induction capabilities have been tested on the Red Creek catchment near Vestry, Mississippi, United States. MIKA-SHA builds and tests many model configurations using the model building components of the two flexible modelling frameworks and quantitatively identifies the optimal model for the catchment of interest. In this study, MIKA-SHA is utilized to identify two optimal models (one from each flexible modelling framework) to represent the runoff dynamics of Red Creek catchment. Both optimal models achieve high-efficiency values and good visual match with the observed runoff response of the catchment. Further, the resulted model architectures are compatible with previously reported research findings and fieldwork insights of the watershed and are readily interpretable by hydrologists.

1 Introduction

Understanding the underlying environmental dynamics occurring within watersheds is an essential and fundamental task in hydrology. Hydrological models play a key role in capturing the discharge dynamics of watersheds. Irrespective of considerable advance over past decades, there is still some scope to advance state of art in hydrological knowledge to fully describe the functioning of a watershed upon a rainfall event owing to the highly complex, interdependent, and non-linear behaviours of governing physical phenomena. So far, no hydrological model structure can perform equally well over the entire range of problems (Fenicia et al., 2011; Beven, 2012a). This leads to different research directions seeking different hydrological models based on different modelling strategies (Beven, 2012c). Hydrological models are expected not only to have good predictive power but also to be interpretable in capturing relationships among the forcing terms and catchment response which may lead to the advancement of scientific knowledge (Babovic, 2005, 2009; Karpatne et al., 2017). Therefore, the final goal of any successful hydrological model must be based on a physically meaningful model architecture along with a good predictive performance.

Data science methods (machine learning techniques) have shown limited success in many scientific fields including hydrology compared to the level of success in many commercial fields (Karpatne et al., 2017). Although the data-driven models are often performing better in terms of predictive capabilities than traditional physics-based, conceptual, and empirical hydrological models (e.g. Nearing et al., 2018; Kratzert et al., 2019a), they may contribute little towards the advancement of scientific discovery due to the lack of interpretability of the model configurations (Karpatne et al., 2017). Recently, a novel modelling paradigm called Theory Guided Data Science (TGDS) (Karpatne et al., 2017) or physics informed machine learning (Physics Informed Machine Learning Conference, 2016) has emerged to enhance the explainability of machine learning models or data science models in general. Here, the existing body of knowledge is blended with machine learning algorithms to induce physically consistent models.

In this contribution, following the above-mentioned modelling paradigm, we introduce a novel model induction engine called Machine Induction Knowledge Augmented - System Hydrologique Asiatique (MIKA-SHA) for automatic induction of semi-distributed rainfall-runoff models for a catchment of interest. This work is motivated by the success of our previously introduced (Chadalawada et al., 2020) model induction toolkit titled Machine Learning Rainfall-Runoff Model Induction Toolkit (ML-RR-MI). ML-RR-MI is capable of inducing fully-fledged lumped conceptual rainfall-runoff models for a watershed of interest. We use the term "hydrologically informed machine learning" to refer that the existing body of hydrological knowledge is used to govern the machine learning algorithms to induce physically consistent model configurations. The proposed framework uses Genetic Programming (GP) as its learning algorithm, whereas the model building modules of two flexible rainfall-runoff modelling frameworks namely FUSE (Clark et al., 2008) and SUPERFLEX (Fenicia et al., 2011; Kavetski and Fenicia, 2011) represent the elements of existing hydrological knowledge.

By being a TGDS approach, the top priority of MIKA-SHA (ML-RR-MI also) remains as the induction of readily interpretable rainfall-runoff models with high prediction accuracies using GP. However, the specific objectives of the current study involve 1) Incorporation of spatial heterogeneities of catchment properties and climate variables into the rainfall-runoff modelling while maintaining the model parsimony of induced models, 2) Adoption of a quantitative model selection approach to select an optimal model with appropriate complexity instead of "simpler the better" paradigm used in ML-RR-MI. The approach addresses the common hydrological issues, such as equifinality, subjectivity, and uncertainty, in the context of semi-distributed modelling and machine learning. This study is a part of the larger ongoing research effort of using hydrologically informed machine learning for automated model induction.

The remaining of this text is arranged as follows. Section 2 provides a brief discussion on fundamental approaches in hydrological modelling. Section 3 discusses machine learning applications in water resources engineering and Sect. 4 discusses the physics informed machine learning and its applications in water resources. The specific modelling strategies used in the current study are presented in Sect. 5. The proposed model induction framework is introduced in Sect. 6. An application of the proposed framework is given in Sect. 7. The last section (Sect. 8) discusses the research findings and the conclusions of the present study. Additional details are presented in the Appendix.

2 Fundamental Approaches in Hydrological Modelling

A general discussion on different hydrological modelling paradigms some of which are used in the present study is presented in this section. An in-depth discussion of each paradigm is beyond the scope of this paper.

2.1 Physics-based Models vs. Conceptual Models vs. Data Science Models

Physics-based models and conceptual models are founded in scientific principles and theories to describe different hydrological processes. These models are following an approach where a hypothesis is assumed initially, and the observations are used to accept or reject it. While small-scale physics are used within the physics-based models, conceptual models consist of a collection of reservoir units that approximate the moisture storage within the basin.

The first reported physics-based model (digitally-simulated) was introduced by Freeze and Harlan (1969). At the time, its usage was greatly limited due to computation demand and intensive data requirements. The ideal solution to understanding and prediction of any environmental dynamic would be through physics-based models if and only if the body of knowledge is sufficient enough to fully describe the behaviour of those environmental processes. However, this is not the situation in hydrology and water resources science in general. For example, the use of the Darcy-Richards equation to represent subsurface flow may not be accurate if the soil properties are not uniform (Beven, 2012c).

In earlier applications, conceptual models were also referred to as Explicit Soil Moisture Accounting (ESMA) models (O'Connell, 1991). Due to the conceptual representation instead of small-scale physics utilized in physics-based models, the complexities of conceptual models are largely reduced when compared to physics-based models. As the conceptual components are derived from known physics but in a simplified manner, conceptual models can provide good process representation and reasonable physical meaningfulness in the model configurations. However, the parameters of conceptual models are not directly linked to the physically measurable quantities as in physics-based models. Hence, it is often required to use calibration schemes to identify the appropriate combination of model parameter values. In practice, there might be different combinations of such parameter values which may result in the same level of model performance. This phenomenon is commonly known as equifinality (structural and measurement uncertainties and lumping may also cause for equifinality) which raises the important question of "are we getting the right results for the right reasons?" (Beven, 2012a). Equifinality is one of the most crucial factors to be addressed in conceptual modelling.

On the other hand, with the advancement in the computer power and acquisition of data through remote sensing and geographical information systems, data science models gained more attraction in many fields. Especially, within the last two decades, there is an increase in data science model applications, such as machine learning models in hydrological modelling (Babovic and Abbott, 1997; Babovic, 2005; Yaseen et al., 2015). The data science models utilize the available data to build input-output relationships which provide actionable models with good predictive power. Physics-based, conceptual and data science models depend on data to a different extent. For example, physics-based models require the measured physical quantities to use as model parameters, whereas the model parameters of conceptual models are derived through a calibration process using measured input-output data. Similar to the conceptual models, data science models utilize the measured input-output data for the model training process.

While theory-based models (physics-based and conceptual models) are frequently admired by the community due to its interpretability which may lead to better understand catchment dynamics, they often experience poorer predictive power than data science models. At the same time, simplistic applications of data-driven models which often result in higher prediction accuracies than the physics-based and conceptual models may suffer serious difficulties with interpretation as they are unable to provide basic hydrological insights (Chadalawada et al., 2020). This dichotomy led to the evolution of two major communities in water resources engineering: those who work with theory-based modelling and those who deal with machine learning techniques, which appear to be working quite separately (Todini, 2007; Sellars, 2018). Recently, a novel modelling paradigm called Theory Guided Data Science (TGDS) (Karpatne et al., 2017) or physics informed machine learning has emerged by combining the strengths of both theory-based models and data science models (Keijzer and Babovic, 2002; Babovic, 2009). Further details of this paradigm and its application in water resources are described in Sect. 4.

2.2 Fixed Models vs. Flexible Models

Two types of modelling approaches can be identified within the conceptual modelling: the models based on a single hypothesis (fixed models) and the models based on multiple hypotheses (flexible models). Fixed models are built around a general model architecture that gives satisfactory model performances over a fairly broad range of watersheds and meteorological conditions. Rainfall-runoff models, such as NAM (Nielsen, 1973), TOPMODEL (Beven et al., 1995), SACRAMENTO (Burnash, 1995), and ARNO (Todini, 1996) belong to this category. Computational efficiency due to standardization, easy interpretability of connections among model parameters and basin characteristics benefit in model explanation and regionalization. These are the main reasons for the popularity of fixed models in hydrological modelling. At the same time, it is quite improbable for a model to perform equally well in completely different climates and geological regions. Further, the adaption of constitutive functions through the addition of specialized modules is often required in fixed models to facilitate the ensemble of processes over a range of watersheds (Fenicia et al., 2011). One alternative to handle this matter would be to test many fixed models on any single catchment to identify the most suitable, which, may be a considerable and cumbersome task. In addition to that, the unavailability of publicly available computer codes for most of the fixed models makes this approach challengeable. In a recent study (Knoben et al., 2019; 2020), an open-source toolbox including computer codes of 46 fixed conceptual models has been developed to facilitate the above-mentioned approach.

In contrast to fixed modelling, flexible modelling frameworks provide more granularity in the model building by allowing the hydrologist to customize the model structure to suit the intended task. These flexible modelling frameworks provide model building blocks that can be arranged in different ways to test many hypotheses about catchment dynamics instead of the one fixed hypothesis in fixed models. Such robust quality of any modular modelling framework allows the modeller to consider the uniqueness of the area of their application. RRMT (Wagener et al., 2001), FUSE (Clark et al., 2008), MMS (Leavesley et al., 2008), SUPERFLEX (Fenicia et al., 2011; Kavetski and Fenicia, 2011), SUMMA (Clark et al., 2015a, 2015b), and RAVEN (Craig et al., 2020) are some widely used flexible modelling frameworks.

The high degree of transferability of flexible modelling frameworks is an aiding factor in proceeding in the direction of a unified hydrological theory at a watershed level. Simultaneously due to the dynamic modularity and high level of granularity, constructing a suitable model for the watershed of concern may require significant effort and expert knowledge. Hence, a hydrologist with novice knowledge would require to test many model structures beforehand selecting an optimal model which is time demanding and computationally intensive, in consequence, hinders the opportunity to use the flexible modelling frameworks in their full potential. Further, the selection of a model configuration without testing a large number of possible combinations may introduce a high level of subjectivity into the model building phase. In a recent paper (Addor and Melsen, 2019) based on more than 1500 peer-reviewed research articles, concluded that the model selection in hydrological modelling is more often driven by legacy rather than the adequacy. In such situations, model selection may be governed by the factors, such as the popularity of model, easiness, prior experience instead of the appropriateness of the model for the intended task

which may result in biased research findings. Therefore, we find a requirement to automate the model building phase to remove the subjectivity and consider many configurations without direct human involvement.

2.3 Lumped Models vs. Distributed Models

Hydrological models are broadly classified into lumped and distributed models based on how they treat the spatial variabilities of catchment properties and climate variables. Lumped models ignore the spatial heterogeneity and recognize the whole watershed as a single unit. Such models use catchment average variable values as model inputs. Most of the present-day conceptual models belong to this category. Ease and simplicity of use have made them a popular hydrological modelling approach. However, especially when the catchment size increases, the meaningfulness of the lumped values decreases and hence the inferences made on the basis of a lumped model may be accurate but not be reasonable or realistic. Further, the observations reveal a lack of consistency among different watersheds which leads to having an insufficient understanding of macro-scale patterns in hydrological behaviours across basins. Namely, there is a possibility that macro-scale patterns of catchments are governed by the heterogeneity (Nearing et al., 2020a). In addition to that, if the modeller's requirement lies within the catchment (e.g. discharge at a particular location within the catchment), then the only option would be to adopt a distributed model where the spatial variabilities are considered in its modelling process. As stated in Fenicia et al. (2016), three distinct steps can be identified in any kind of distributed model building. The first step is to implement a spatial discretization scheme. Spatial discretization can be achieved by using either regular grids, irregular grids, and subcatchments, or Hydrological Response Units (HRUs). The next step is to define the model structure and the connections between the spatial elements. The final step is to achieve model parsimony through the specification of model parameters and state constraints.

The majority of distributed models are physics-based models. They discretize the watershed into regular or irregular grids and use small-scale physics to model the fluxes through the spatial elements (commonly attributed to as fully distributed models). In the early stages of development, researchers believed that more data about the catchment properties and climate variables would be available with the advancement of technology and hence thought of including such data into hydrological modelling with the intention of achieving improvements in model simulations (Beven, 2012c). This helped fully distributed modelling to attract a lot of attention among hydrologists. From its earliest applications like System Hydrologique Europeen model (SHE) (Abbot et al., 1986a, 1986b) hydrological community has invested heavily in these fully distributed hydrological (physics-based) models (e.g. Development of US National Water Model (Salas et al., 2018)).

One way of addressing the so-called uniqueness of the place as a major issue to deal with hydrological modelling (Beven, 2020) is to use distributed models. At the early stages of distributed modelling, the approach was constrained due to the lack of data and computational power (Wood et al., 2011; Beven, 2012c; Fatichi et al., 2016). Hence, it was thought that this approach would gain success with the advancement of technology. Until today, however, the distributed models have not achieved the expected outcome (Beven, 2020). This points out that the problem lies not only in the lack of local information

but also due to the issues in how processes are represented within the distributed model (Beven, 2020). The high complexity and the huge demand for the input data, such as topography, geology, soil, and land use are the main limitations of fully distributed models. The more granular approach requires a large number of model parameters which often leads to over-parameterization. Too many model coefficients may result in good fitting however may also result in transfer functions to be physically unrealistic (Beven, 2012b). A comprehensive review of applications, challenges, and future trends of fully distributed modelling in hydrology is presented in Fatichi et al. (2016).

An effective alternative for both lumped and fully distributed models would be the semi-distributed models where separate conceptual models are assigned to functionally distinguishable land segments (Boyle et al., 2001). In the semi-distributed modelling approach, each model operates individually and there are no interconnections or dependencies with other models in the network. This and the use of conceptual models instead of small-scale physics make this approach several orders less complex than the fully distributed models. However, semi-distributed models are much more complex than lumped conceptual models, because they consider the spatial variabilities of catchment properties and climate variables, resulting in more meaningful inferences gained through the model. In early applications (Boyle et al., 2001), subcatchments were identified as functionally distinguishable land segments. But, with the popularity of the Hydrological Response Unit (HRU) concept, HRUs were used as functionally distinguishable land elements (e.g. Fenicia et al., 2016). In a semi-distributed model, the total catchment response consists of the routed sum of the individual model responses of each spatial element. Spatial Tools for River basins and Environment and Analysis of Management options (STREAM) (Aerts et al., 1999) and Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) can be categorized as semi-distributed models.

3 Machine Learning in Water Resources

Machine learning or data science in general, have become an irreplaceable tool, not only in commercials but also in many scientific fields. They have shown superior performances in many applications including language translation, object tracking, autonomous driving, and character recognition (Karpatne et al., 2017). Data-driven techniques started to gain a lot of attention among the hydrologists within the last two decades. Artificial Neural Networks (ANN), Evolutionary Computation (EC), Wavelet-Artificial Intelligence models (W-AI), Support Vector Machines (SVM), and Fuzzy set are the most popular data science techniques in hydrological modelling (Yaseen et al., 2015). Each of these techniques has its strengths and weaknesses. The scope of this paper does not discuss different data-driven techniques in detail. Instead, interested readers are directed to review papers by Govindaraju (2000), Yaseen et al. (2015), Mehr et al. (2018), and the textbook by Hsieh (2009).

Machine learning models have shown encouraging performances in a range of water resources applications, such as rainfall-runoff modelling (Minns and Hall, 1996; Khu et al., 2001; Babovic and Keijzer, 2002; Chiang et al., 2004; Kratzert et al., 2018, 2019a, 2019b), streamflow forecasting (Babovic et al., 2000b; Nourani et al., 2009; Meshgi et al., 2014, 2015; Humphrey et al., 2016; Karimi et al., 2016), estimation of missing data (Elshorbagy et al., 2002), error correction (Sun et al., 2012), water

quality modelling (Savic and Khu, 2005; Singh et al., 2011; García-Alba et al., 2019), sediment transport modelling (Babovic and Abbott, 1997; Afan et al., 2014; Safari and Mehr, 2018), reservoir management (Giuliani et al., 2015), prediction of climate variables (Dahamsheh and Aksoy, 2013; Ferreira et al., 2019), because of their ability to capture noise complexity, non-linearity, non-stationarity and dynamism of data (Yaseen et al., 2015). Certainly, if we are only interested in better forecasting results then, the machine learning models might be the preferred choice over the conceptual or physics-based models due to their better predictive capability. Another major advantage of a machine learning model is that it requires much less human effort to develop and calibrate than a physics-based model (Nearing et al., 2020a).

Data-driven techniques have made it possible to develop actionable models with high prediction accuracy without depending on domain knowledge. At the same time, this very nature of data-driven models has become the main point of criticism especially in scientific fields including hydrology. They are regularly quoted as black-box models where the user has little or no knowledge about how the model makes its predictions. Karpatne et al. (2017) offer two reasons for the limited success of data-driven models in scientific fields. The first reason is the limited availability of labelled instances for the model training which makes it harder to extrapolate model predictions beyond the available labelled data. The second reason is associated with the objectives of the scientific discovery where the final goal is not only to have actionable models but also to convey a mechanistic awareness of underlying operations which may lead to the advancement of scientific knowledge. Further, data science models, such as Deep Learning (DL) models have shown better performances in hydrograph predictions than the traditional approaches in ungauged catchments (Kratzert et al., 2019a). At the same time, a recent paper (Beven, 2020), questions the performance of a DL model in ungauged catchments when the geological characteristics are not well defined within the model. According to this paper, DL models have not solved the ungauged catchment problem and they have just achieved higher efficiency values than the traditional approaches.

Nearing et al. (2020a) argue that there is a danger for the hydrologic community in not recognizing the potential of machine learning offers for the future of hydrological modelling. The authors argue that machine learning models can capture catchment similarities by providing good results even for the catchments which were not used for the training of those models. This implies the capability of machine learning models in developing catchment scale theories that traditional models were unable to do so well. Further, the authors reject the most common criticism on machine learning models (the lack of explainability) by stating that even the accuracy of process representation in physics-based models is questionable due to their poorer prediction accuracies, criticizing only on machine learning models is unfair and meaningless. Despite having a huge potential within machine learning models, the state of art machine learning capabilities have not been tested in hydrological modelling and they expect even distributed hydrological models are to be developed primarily on machine learning in near future. Beven (2020) highlights the importance of the interpretability of DL models and suggests more direct incorporation of process information into such models. Further, he points out that machine learning models should also need to pay attention to similar

issues associated with traditional modelling approaches like data and parameter uncertainties and equifinality. A brief discussion of two widely used machine learning techniques in hydrology is presented below.

3.1 Artificial Neural Networks (ANN)

ANNs (McClelland and Rumelhart, 1986) are the most popular machine learning technique in many commercial and scientific fields including hydrology. ANN is a computing model inspired by the functionality of neurons in a human brain, that is widely used to compute and process complex functional units. A wide range of successful applications, such as clustering, pattern recognition, classification, and identifying non-linear relationships have made ANNs a popular data-driven modelling technique. Typically, ANN architecture consists of three components i) input layer with one or more input nodes ii) One or more hidden layers with the activation function iii) Output layer with one or more output nodes (Yaseen et al., 2015). Successful ANN applications in water resources engineering include rainfall-runoff modelling (Minns and Hall, 1996; Chiang et al., 2004), streamflow estimation (Nourani et al., 2009; Humphrey et al., 2016), water quality modelling (Singh et al., 2011; García-Alba et al., 2019), groundwater modelling (Nayak et al., 2006; Gholami et al., 2015), data assimilation (Babovic et al., 2000a; Vojinovic et al., 2003), estimation of climate variables (Dahamsheh and Aksoy, 2013; Ferreira et al., 2019), flood and drought forecasting (Chang et al., 2014; Dehghani et al., 2014) and sediment transport modelling (Afan et al., 2014).

Deep learning (DL) is a new direction in ANN research that is widely used for clustering and regression tasks in many disciplines including hydrology. There is no definite definition for DL models, but neural networks with large multilayer architectures (large depth) that work with big, raw data are generally referred to as DL models (Shen, 2018). DL models are capable of extracting abstract features from raw data automatically via the hidden layers. Two of the well-established classes in DL are Convolutional Neural Networks (CNNs) for clustering tasks, such as computer vision and image analysis, and Recurrent Neural Networks (RNNs) for regression tasks, like modelling sequential data and time series analysis (Hu et al., 2018). Long Short-Term Memory (LSTM) is the most successful RNN architecture which utilizes gates and memory cells to retain state information of sequential data. Hence, LSTMs are more suitable for hydrological modelling applications, such as rainfall-runoff modelling. The state of art DL capabilities have not yet been tested in hydrological modelling and there are only a few DL applications so far (Shen et al., 2018). Successful DL applications in hydrology include rainfall-runoff modelling (Hu et al., 2018; Kratzert et al., 2018, 2019a, 2019b; Fan et al., 2020; Nearing et al., 2020b; Xiang et al., 2020), soil moisture modelling (Xiaodong et al., 2016), precipitation forecasting (Kumar et al., 2019), groundwater estimation (Afzaal et al., 2019) and uncertainty estimation (Gude et al., 2020).

3.2 Genetic Programming (GP)

Genetic Programming is an evolutionary computation algorithm (Koza, 1992) inspired through the basic principle of Darwin's theory of evolution. GP is capable of automatic generation of computer programs and falls under the supervised machine learning category. The most distinct feature of GP over the other machine learning techniques is its ability to produce explicit

mathematical expressions of input-output relationships. As a result, GP is referred to as a grey box data-driven technique and differentiates it from the other black box data-driven approaches, like ANNs. Other than that, its conceptual simplicity, the ability of parallel computing, and the capability of obtaining the near-global or global solution make GP a powerful machine learning technique.

There are different variants of GP like Monolithic GP (MGP), Multigene genetic programming (MGGP), Gene expression programming (GEP), Linear GP (LGP), and Grammar-based GP (GGP) (Mehr et al., 2018). Despite variants, the fundamental operations are quite similar. GP generates the structure of its solutions (GP individuals) by arranging mathematical functions, input variables, and random constants. These are known as the building blocks of the GP algorithm. The algorithm starts with a randomly generated set of candidate solutions for the task at hand. The performance of each candidate is then assessed using a user-defined objective function. Individuals are selected by assigning higher chances of selection for better individuals (based on objective function value) to create offspring by apply genetic operators (crossover, mutation, and elitism). The new set of offspring becomes the candidate solutions in the next generation. This process is repeated until the algorithm meets its termination criteria (usually a maximum number of generations). The candidate solutions evolve towards the global optimum when the GP algorithm curtails the error margin between the simulated values of its individuals and measured observations (Babovic and Keijzer, 2000).

Successful GP applications in water resources engineering can be found in rainfall-runoff modelling (Khu et al., 2001; Babovic and Keijzer, 2002; Babovic et al., 2020), streamflow prediction (Meshgi et al., 2014, 2015; Karimi et al., 2016), water quality modelling (Savic and Khu, 2005), groundwater modelling (Datta et al., 2014), reservoir management (Giuliani et al., 2015), sediment transport (Babovic and Abbott, 1997; Safari and Mehr, 2018), climate variables and soil properties modelling (Bautu and Bautu, 2006; Elshorbagy and El-Baroudy, 2009).

4 Physics Informed Machine Learning

One promising way forward which may bridge the gap between physics-based and machine learning modelling communities would be to couple the existing hydrological knowledge to guide machine learning models (Babovic and Keijzer, 2002; Babovic, 2009). This recent paradigm is presently referred to as Theory Guided Data Science (TGDS) (Karpatne et al., 2017) or Physics Informed Machine Learning (Physics Informed Machine Learning Conference, 2016). This modelling paradigm aims to simultaneously address the limitations of data science and physics-based models and induce more generalizable and physically consistent models. There are five ways of incorporating basic scientific knowledge with data-driven models (Karpatne et al., 2017): (i) theory-guided design of data science models, (ii) theory-guided learning of data science models, (iii) theory-guided refinement of data science outputs, (iv) learning hybrid models of theory and data science and (v) augmenting theory-based models using data science. A typical physics informed machine learning model may follow one or more of the above mention approaches to bring together scientific knowledge and data science techniques. Although, there are

few reported explainable artificial intelligence utilizations in hydrological modelling in past (e.g. Cannon and Mckendry, 2002; Keijzer and Babovic, 2002; Fleming, 2007), there is an increasing trend of adopting theory-guided machine learning models for recent water resources applications (McGovern et al., 2019), such as hydroclimatic model building (Snauffer et al., 2018), automated model building (Chadalawada et al., 2020) and hydrologic process simulation (Solander et al., 2019). Even though there are attempts in almost every machine learning technique to incorporate existing hydrological knowledge into the basic frameworks, in the sequel, we only discuss such attempts in ANNs and GP.

Relative to the GP, ANNs suffer the most severe consequences of lack of interpretability of resulted models (Mehr et al., 2018). An effective solution for this would be the use of augmented versions of neural networks where the existing theoretical knowledge is used to govern the learning algorithm to enhance the interpretability of induced models. Brunton et al. (2016), Raissi et al. (2017) and Rudy et al. (2017) used Physics Informed Neural Networks (PINN) in time series analysis to derive governing partial differential equations. Prediction of extreme rainfall events was carried out by Cannon (2018) using a neural network architecture constrained by physical laws. Wang et al. (2020) introduced a deep learning framework called Theory Guided Neural Networks (TGNN) for subsurface flow modelling where the governing equations, physical constraints, engineering controls and expert knowledge are used to guide the ANN model. Please refer to Fleming et al. (2014) and Xu et al. (2019), for further theory-guided neural network utilization in water resources.

Although the physics informed machine learning was only recently identified as a new modelling paradigm in the context of GP, there were attempts over past two decades to blend the hydrological knowledge into basic GP framework to induce more physically reliable hydrological models. To achieve physical consistency and dimensional accuracy of GP induced models, researches developed few enhanced versions of the GP algorithm by incorporating the existing hydrological knowledge. Declarative bias and preferential bias were incorporated with the model-building phase of GP to reduce physical contraventions and to achieve dimensional accuracy of induced equations (Babovic and Keijzer, 1999, 2002; Keijzer and Babovic, 2002). At the initialization stage, declarative bias forces to sample only the dimensionally correct solutions (a hard constraint on dimensional correctness) while preferential bias guides the algorithm towards the dimensionally correct solution (a soft constraint on dimensional correctness) and allows all solutions to induce. Authors have reported that this augmented version of GP resulted in fast convergence through the reduction of solution space and achieved more parsimonious and regularize expressions than traditional GP. Dimensionally aware GP was utilized to extract hydraulic formulae from measurements by Babovic et al. (2001).

The inclusion of high-level theoretical concepts in sediment transport modelling with GP resulted in equal or superior performances than the traditional modelling with human expert knowledge (Baptist et al., 2007; Babovic, 2009). Another augmented version of GP was used for the identification of predominant processes in hydrological system dynamics by Selle and Muttil (2011). A reservoir model, a cumulative sum and delay function, and a moving average operator were incorporated as basic hydrological insights into the GP function set by Havlicek et al. (2013), to develop a rainfall-runoff prediction

programme called SORD. They were able to achieve superior performances in terms of prediction accuracy with SORD than to ANNs and GP without above-mentioned special functions. GP was used as a model induction algorithm in Chadalawada et al. (2017), to optimize both model architecture and parameters to automatically induce most appropriate Tank model structure for a watershed of interest. Here, the hydrological knowledge is incorporated as special functions inspired through the Sugawara Tank model template (Sugawara, 1979). In our prior work (Chadalawada et al., 2020), an automatic lumped conceptual rainfall-runoff model induction toolkit was developed using GP and the building blocks available in two modular modelling frameworks (FUSE and SUPERFLEX) were used as the components of hydrological insights.

5 Methods

Considering the uniqueness of the place is an important aspect of hydrological modelling (Beven, 2020). The use of distributed modelling concepts and flexible modelling frameworks are two available toolsets to incorporate the spatial heterogeneity into the model building phase. Due to the limited success and higher-order complexity of fully distributed models, the semi-distributed modelling concept is used for the current study where a network of functionally distinguishable conceptual models from flexible modelling frameworks is developed to represent the watershed dynamics. As a result of the higher granularity and flexibility provided by the flexible modelling frameworks, even with a lumped application, one can try thousands of possible model architectures for a catchment of interest. This may rise to millions of possible model combinations in the context of semi-distributed modelling which makes it almost impossible to test them manually. Further, the selection of a model configuration without testing alternative model configurations would become highly subjective and may require considerable expert's knowledge and time. Therefore, we see a necessity to automate the model building phase to overcome these limitations. Hence, in this work, a novel model induction toolkit called Machine Induction Knowledge Augmented-System Hydrologique Asiatique (MIKA-SHA) is proposed to induce an optimal semi-distributed model for a catchment of interest.

GP has been selected as the machine learning technique here due to its ability to optimize both model configuration and model parameters together. It is interesting to note that, most state of art GP utilizations in water resources (Oyebode and Adeyemo, 2014; Mehr et al., 2018), GP is still utilized as a short-term prediction mechanism which is analogous to ANN applications. In our contribution, we explore the full potential of GP by inducing fully-fledged rainfall-runoff models where the hydrological insights are introduced through the integration of process understanding by including model building components from existing flexible modelling frameworks into the function set of GP algorithm. Our earlier work (Chadalawada et al., 2020), presented the capacity of this modelling approach (ML-RR-MI) as a lumped conceptual model induction toolkit. In the current study, this framework is extended to induce semi-distributed rainfall-runoff models. As per the taxonomy defined in Karpatne et al. (2017), our framework falls under the hybrid TGDS category. Currently, MIKA-SHA learns models utilizing the model building components of two flexible modelling frameworks. A brief discussion of the two flexible modelling frameworks used

in the current study is given below. However, the proposed framework can be coupled with any internally coherent collection of building blocks.

5.1 SUPERFLEX

SUPERFLEX (Fenicia et al., 2011; Kavetski and Fenicia, 2011) framework facilitates hydrologists to test many different hypotheses about the functioning of the watershed of interest using the model building components (reservoirs, junctions, and lag functions) available in the framework. The water storages within the catchment, such as soil moisture, interception, groundwater, and snow along with their release of water are represented through reservoir units. Junction elements conceptualize the merging and splitting of different fluxes in catchment dynamics (e.g. Hortonian flow, evaporation). Channel routing (delays in flow transmission) is described using lag functions. A number of constitutive functions are available to describe lag function characteristics and storage-discharge relationships of storage units (reservoirs). SUPERFLEX applications in rainfall-runoff modelling are found in van Esse et al. (2013), Fenicia et al. (2014, 2016), and Molin et al. (2020).

5.2 FUSE

Clark et al. (2008) developed Framework for Understanding Structural Errors (FUSE) to examine the effect of model structural differences on rainfall-runoff modelling. FUSE conceptualizes the functioning of a catchment using a two-zone model architecture: an unsaturated zone (upper soil layer) and a saturated zone (lower soil layer). The model building modules of FUSE involve the choice of upper and lower soil configurations and parameterization for different hydrological processes, such as evaporation, percolation, interflow, surface runoff, and baseflow. The modeller has the freedom of selecting these model building modules from four rainfall-runoff models (TOPMODEL, ARNO/VIC, SACRAMENTO, and PRMS) which are known as parent models. For more details and applications of FUSE, please refer to Clark et al. (2010) and Vitolo (2015).

5.3 Performance Measures

MIKA-SHA consists of a performance measures library including the majority of the widely adopted performance matrices (Chadalawada and Babovic, 2017). In the present study, we have selected four absolute performance measures namely volumetric efficiency (Criss and Winston, 2008), Kling-Gupta efficiency (Gupta et al., 2009), Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) and log Nash-Sutcliffe efficiency (Krause et al., 2005) from the MIKA-SHA's performance measures library to evaluate the simulated discharge values against the measured discharge values. The four selected objective functions are sensitive to different regions of measured and simulated runoff signatures and their details are given in Table 1. The four selected objective functions are used in the multi-objective optimization scheme of MIKA-SHA.

Table 1: Absolute performance measures used in the current study

Name	Equation	Sensitivity	Optimum
Volumetric Efficiency (VE)	$VE = 1 - \frac{\lvert\sum_{t=1}^{N}(Q_{ot} - Q_{st})\rvert}{\sum_{t=1}^{N} Q_{ot}}$ N: Time steps, Q_{ot}: Observed streamflow, Q_{st}: Simulated streamflow	Water balance	1
Kling-Gupta Efficiency (KGE)	$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$ r: Linear correlation coefficient, $\alpha = \frac{\sigma_s}{\sigma_o}$, $\beta = \frac{\mu_s}{\mu_o}$, σ: Standard deviation, μ: Mean	Flow variability	1
Nash-Sutcliffe Efficiency (NSE)	$NSE = 1 - \frac{\sum_{t=1}^{N}(Q_{ot} - Q_{st})^2}{\sum_{t=1}^{N}(Q_{ot} - \overline{Q_{ot}})^2}$ $\overline{Q_{ot}}$: Mean of observed discharge values	High flows	1
Log Nash-Sutcliffe Efficiency (logNSE)	$logNSE = 1 - \frac{\sum_{t=1}^{N}(logQ_{ot} - logQ_{st})^2}{\sum_{t=1}^{N}(logQ_{ot} - log\overline{Q_{ot}})^2}$ log: Natural logarithm	Low flows	1

By nature, the genetic programming algorithm drives its total population towards the global or near-global solution which results in a set of possible solutions instead of one solution. In the context of rainfall-runoff model induction, such possible solutions may represent different model structures (different hypotheses about catchment dynamics). Hence, it is often required to use a model selection scheme to select the optimal model from the competing models. The more straight forward approach to select an optimal model from a set of equally performing models would be to select the model with least complexity in terms of the model structure (most parsimonious model). However, in the current study, we employ the following relative performance measures to evaluate the model performance of each model relative to other competing models at the optimal model selection stage. Details about how each performance measures are used within the proposed framework are given in Sect. 6.

Standardized Signature Index Sum (SIS)

The Standardized Signature Index Sum value (Ley et al., 2016) is a relative performance measure which quantifies how well a model captures the observed flow duration curve (FDC) relative to the other competitive models. Models with negative SIS

values indicate better than average performance in capturing observed FDC and vice versa. In SIS calculation, both observed and simulated FDCs are divided into four flow regimes based on flow exceeding probabilities and calculate the absolute difference in observed and simulated cumulative discharges in each region. Then, four separate Z-score values (representing 4 regions) are assigned to each model based on the mean and standard deviation of all models considered. The algebraic sum of those four Z-score values becomes the SIS value of the model.

$$Z_{sa} = \frac{|x_{sa}| - \overline{x_a}}{\sigma_a} \quad (1)$$

$$SIS_a = Z_{sFHV} + Z_{sFMV} + Z_{sFMS} + Z_{sFLV} \quad (2)$$

where $|x_{sa}|$: modulus of the signature index where, s: model, a: FDC signature based on Flow Exceeding Probability (FEP) (FHV: FEP less than 2%, FMV: FEP between 2% and 20%, FMS: FEP between 20% and 70%, FLV: FEP greater than 70%) and x: value, $\overline{x_a}$ and σ_a : average and standard deviation of $|x_{sa}|$, Z : standard score.

Cross sample entropy value (Cross-SampEn)

Cross-SampEn value is a derivation from the commonly used Sample Entropy value (Richman and Moorman, 2000). Sample Entropy is a complexity measure of data series which has its origin in information theory. Sample Entropy value gives an idea about the complexity of the data series based on the information content in a mathematical way. Cross-SampEn value also follows the same concept but is used to measure the correlation between two series by matching patterns from one series with another. A low Cross-SampEn value indicates that the two series are more similar to each other. More details about Cross-SampEn can be found in Delgado-Bonal and Marshak (2019).

Dynamic Time Warping (DTW) distance

Dynamic Time Warping (Sakoe and Chiba, 1978) is a similarity measure between two time series which includes warping of their time axes to find the optimal temporal alignment between the two. DTW distance is derived as an alternative to the commonly used Euclidean distance. Two identical time series with a small-time shift may ending up with a large Euclidean distance and may consider them as two dissimilar time series. The DTW method captures them as two similar time series as it ignores the shift in the time axes. A low DTW distance indicates more similarity between the two time series compared. Details and applications of the DTW method can be found in Salvador and Chan (2007), Giorgino (2009) and Vitolo (2015).

Model parsimony

Here, the model parsimony is evaluated in terms of the number of associated model parameters of each model. One model is considered more parsimonious than another model if the number of model parameters of the former is lower than the later.

6 MIKA-SHA

Chadalawada et al. (2020) introduced a new hydrologically informed rainfall-runoff model induction toolkit based on GP (ML-RR-MI) capable of developing lumped conceptual hydrological models utilizing model building components of FUSE and SUPERFLEX frameworks. The unique feature of ML-RR-MI is that it utilizes the existing body of hydrological knowledge to govern the GP algorithm to induce physically sound and consistent models with high prediction accuracies. The building blocks of the two flexible modelling frameworks are used as the elements of incorporated hydrological knowledge of ML-RR-MI. The model building components of SUPERFLEX consist of reservoir units, lag functions and constitutive functions to represent storage-discharge relationships and characteristics of lag functions. In the FUSE framework, building blocks include the selection of upper and lower layer architectures, flux equations to represent surface runoff, percolation, evaporation and the presence of interflow and flow routing. These building blocks are incorporated as special functions (named as "FUSE" and "SUPERFLEX") into the function set of ML-RR-MI along with basic mathematical functions. ML-RR-MI optimizes both model structure and model parameters simultaneously and selects an optimal model for the catchment of interest without any direct human involvement.

Successful application of ML-RR-MI toolkit motivated the present research to extend its modelling capabilities towards distributed hydrological modelling. Hence, we have developed an automatic model induction toolkit for semi-distributed rainfall-runoff models. In the present contribution, a new function called "DISTRIBUTED" has been incorporated to the GP function set along with "FUSE", "SUPERFLEX" and other mathematical functions. The "DISTRIBUTED" function represents the semi-distributed models (GP individuals) within MIKA-SHA. The parse tree representation of the "DISTRIBUTED" function is shown in Fig. 1. As it can be seen, "DISTRIBUTED" function uses either "FUSE" or "SUPERFLEX" functions as its function arguments depending on the selected model inventory by the user. The length of the function arguments of "DISTRIBUTED" function depends on the number of Hydrological Response Units (HRUs) within the watershed. The last two arguments are the lag parameters which are used to route HRU's outflow into subcatchment outlet (Lag_HRU) and subcatchment's outflow into catchment outlet (Lag_Sub). Here, the routing module is based on two-parameter Gamma distribution with shape parameter equals to 2.5. Nodes from depth = 2 to depth = maximum allowable tree depth, are the function arguments of either "FUSE" or "SUPERFLEX" functions. R (R Core Team, 2018) programming language has been used to implement MIKA-SHA. The workflow diagram of the MIKA-SHA is given in Fig. 2. Details about each module of MIKA-SHA (data preprocessing, model identification, model selection, and uncertainty analysis) are given in the sequel.

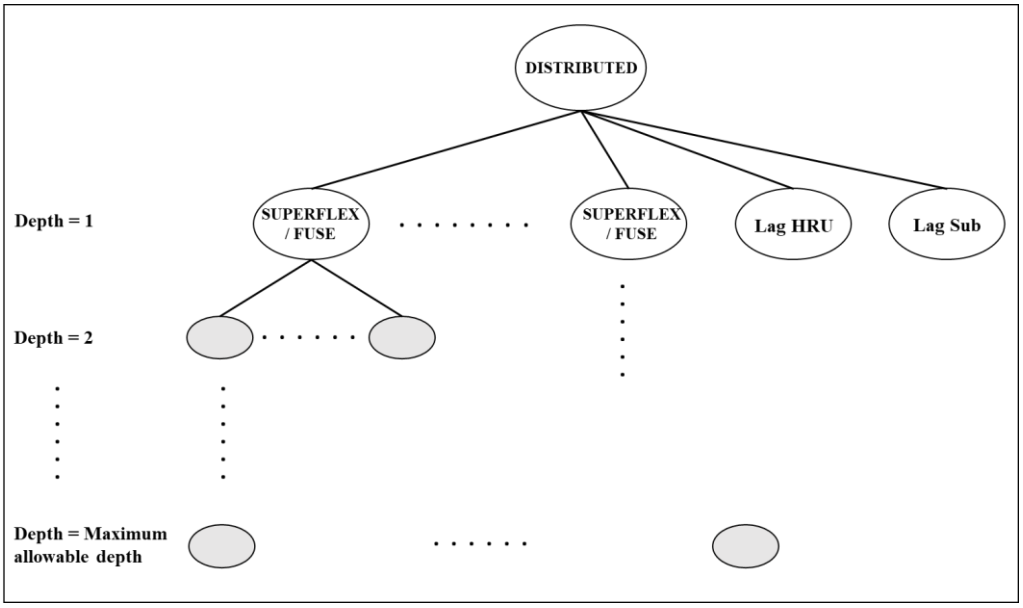

Figure 1: Parse tree representation of the DISTRIBUTED function in MIKA-SHA

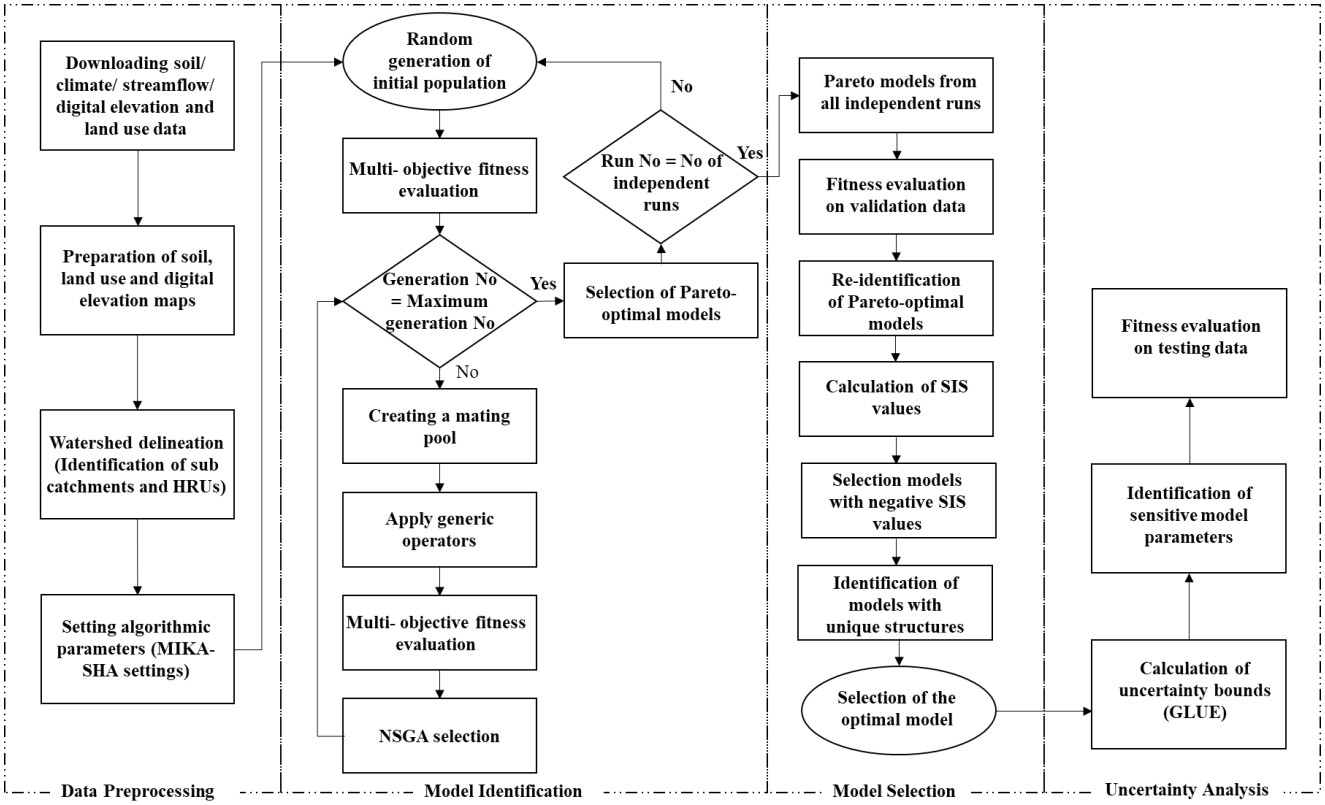

Figure 2: Workflow diagram of MIKA-SHA

6.1 Data Preprocessing

Data preprocessing stage involves quality control of forcing terms (precipitation, potential evaporation and temperature) and streamflow data, identification of subcatchments and HRUs through watershed delineation, and preparation of subcatchment averaged forcing terms vectors. MIKA-SHA uses QGIS software (QGIS, 2020) to prepare the required Digital Elevation Maps (DEM), land use maps, geological maps and soil maps for watershed delineation. Then, the SWAT+ plugin of QGIS software is used for the watershed delineation. HRUs can either be identified based on the topography, soil type or geology of the catchment of interest.

6.2 Model Identification

As shown in Fig. 2, model identification stage starts with a randomly generated set of candidate model structures (semi-distributed model structures made out from the special functions, basic mathematical functions and random constants) to capture the runoff dynamics of the catchment of interest. These model structures (GP individuals) may differ from each other in terms of model structural components and parameter values. Then, the performance of each individual is assessed on a user-defined multi-objective criterion. Next, the individuals are selected to create a mating pool which is used to create offsprings (child population) through the application of genetic operators, such as crossover and mutation. The selection scheme (explained later) ensures that individuals with higher performance values in terms of the objective functions used have a higher chance of selection. This way GP algorithm optimizes both model configuration and associated parameters of the GP individuals simultaneously. Finally, the same selection scheme is utilized to select individuals to form the next generation (next parent population) from both individuals of the child population and parent population (combined population). This evolution process continues through the generations until the algorithm reaches the maximum number of generations. Here, the optimization algorithm is repeated for a user-specified number of iterations (independent runs) to cover the solution space to a greater extent. The output of the model identification stage consists of a set of non-dominated models (Pareto-optimal models) based on the selected objective criteria. More details about the basic steps involved at this stage are given in Chadalawada et al. (2020).

MIKA-SHA relies on a multi-objective optimization framework founded on Non-dominated sorting genetic algorithm-II (NSGA-II) (Deb et al., 2002) at model identification stage, using desired objective functions from the objective function library. Each individual (candidate solution) in the population is evaluated on each objective function separately. Based on the objective function values, each individual is assigned a non-domination rank and a crowding distance value. The ranks are identified based on the Pareto-optimality concept. For example, all the individuals with non-domination rank 2 are dominated by individuals with rank 1. However, individuals with rank 2 are not dominated by any other individuals with a higher rank (lower the rank better the individual). On the other hand, crowding distance measures how an individual located relative to the other individuals of the same rank (more the distance better the individual – more diversity). Therefore, at the selection phase of the algorithm, when two individuals are randomly selected and they have different ranks, the individual with the lower rank

is selected. If both of them have the same rank, then the individual with higher crowding distance is selected. Having said that, identification of the best performing model from Pareto front of non-dominated solutions for a watershed of interest is not a trivial matter. The explanatory power of the performance measure used to assess the prediction accuracy of model simulations has a direct impact on the optimal model selection (Chadalawada and Babovic, 2017).

6.3 Model Selection

Model selection stage starts with the best models of each independent run (front 1 models of final generation) derived through the GP framework at the model identification stage. The quantitative optimal model selection process is streamlined as follows.

1. Performance evaluation using the same multi-objective criterion on validation data for all identified models from the model identification stage.
2. Re-identification of Pareto-optimal models based on both calibration and validation fitness values.
3. Calculation of Standardized Signature Index Sum (SIS) of each Pareto-optimal model.
4. Selection of Pareto-optimal models with SIS scores below zero over the calibration and validation period.
5. Identify unique model structures (hereinafter referred to as competitive models) from the models in step 4. If there is more than one model with the same model structure, the model with the most negative SIS value is selected.
6. Quantitative selection of the optimal model to represent catchment dynamics based on three relative measures: Cross sample entropy value (Cross-SampEn), Dynamic Time Warping (DTW) distance and model parsimony based on the number of associated model parameters. Competitive models are ranked according to each measure (lower the value better the performance) and the model with the lowest sum up rank is selected as the optimal model for the watershed of concern.

6.4 Uncertainty Analysis

Once the optimal model is identified for the catchment of interest its uncertainty and sensitivity analysis are performed using Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) as described below.

1. A random subset of model parameters of the selected optimal model structure is changed uniformly within their parameter range (in this case between 0 and 1 as all parameter ranges are normalized within MIKA-SHA framework) while keeping the remaining model parameters at their calibrated values. NSE is used as the likelihood estimation. If the model parameter set provides an NSE value greater than the likelihood threshold, the parameter set, its NSE value and the simulated discharge are recorded (known as behavioural models).
2. Repeat the above step until the number of behavioural models reaches a user-defined value.
3. For each time step, simulated discharge values of all behavioural models are sorted in ascending order. Then, a weight is assigned to each model (NSE value itself is used as the weight). Finally, the Cumulative Probability Distribution Function (CDF) of the weights is calculated at each time step.

4. For each time step, a relationship diagram is obtained by taking CDF as the x-axis and simulated discharge at the y-axis. From the diagram, corresponding simulated discharge values of 95% and 5% quantile of CDF are selected as the upper and lower bounds of the 90% confidence band.
5. Percentage of observed discharge values (both in calibration and validation period) which fall within the 90% confidence band is used to measure the uncertainty estimation capability of the selected optimal model.
6. If the uncertainty estimation capabilities are satisfactory, the model performance of the optimal model is tested for an independent time frame (testing period) which is not used in model selection or identification stages. If the uncertainty estimation is not satisfactory, then, all the above steps are to be repeated with the next best competitive model.
7. Sensitivity scatter plots are drawn for each model parameter using the parameter values of behavioural models. The shape of the scatter plot (the x-axis – normalized parameter range, the y-axis – NSE values) is used to identify the degree of sensitivity of each model parameter.

The main target of MIKA-SHA is to induce physically consistent semi-distributed rainfall-runoff models through the incorporation of spatial heterogeneity data and existing hydrological knowledge with its machine learning algorithm. However, the following measures are taken to handle general hydrological modelling issues, such as overfitting, subjectivity and equifinality.

- Avoid overfitting – Limit tree growth, consider validation fitness and model parsimony in the optimal model selection, use of internally coherent special functions, evaluate the fitness of the optimal model on an independent data set.
- Remove subjectivity – Automated process ensures no direct human involvement, test many hypotheses before selecting an optimal model.
- Handling equifinality – Optimal model selection is based on many absolute and relative performance matrices.

7 Application of MIKA-SHA

This section aims to demonstrate how MIKA-SHA works through a one case study using a watershed in the United States. MIKA-SHA was used to identify two optimal model configurations for the catchment of interest using SUPERFLEX and FUSE model building libraries separately.

7.1 Study Area

The Red Creek watershed near Vestry, Mississippi (Fig. 3) was selected to test the semi-distributed model induction capabilities of MIKA-SHA. Red Creek watershed is located in the Eastern United States and basin details are summarized in Table 2. Soil and land use data of Red Creek catchment (resolution of 30 m x 30 m) were downloaded from the United States

Department of Agriculture's (USDA's) Geospatial Data Gateway (USDA's Geospatial Data Gateway, 2020), whereas the Digital Elevation Data (DEM) at 30 m resolution were obtained from the Shuttle Radar Topography Mission (SRTM) data from United States Geological Survey (USGS) EarthExplorer (USGS EarthExplorer, 2020). The whole watershed was divided into three subcatchments (Sub 1, Sub 2 and Sub 3) for the current application. HRUs were identified based on the topography of the area and three HRUs namely, Hill (slope band %> 10), Floodplain (slope position threshold = 0.1) and Plateau (slope band % < 10) were selected. The HRU details are given in Table 3.

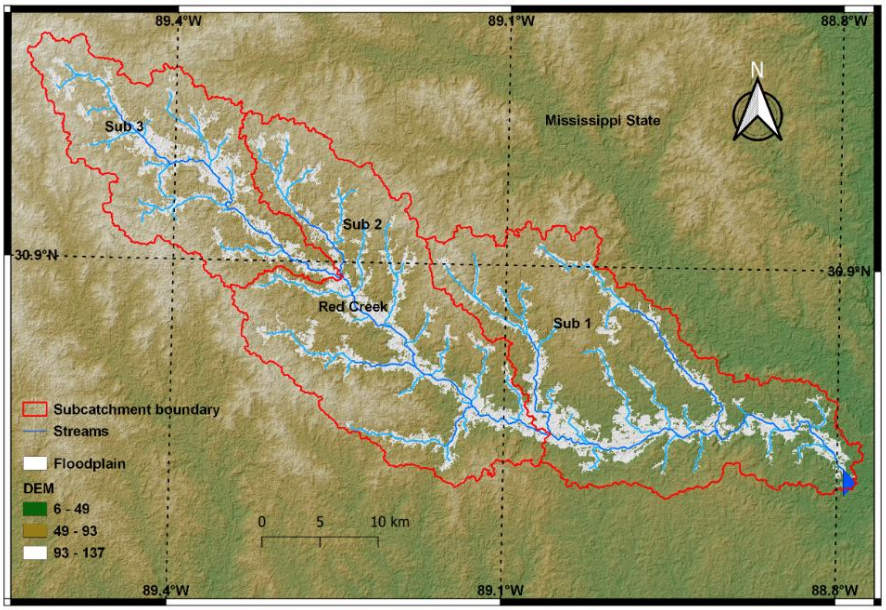

Figure 3: Red Creek catchment, Vestry, Mississippi, United States (map was generated through SWAT+ plugin in QGIS software using Shuttle Radar Topography Mission (SRTM) DEM data and USDA's Geospatial Data Gateway soil and land use data)

Eleven years (from 1 January 2004 to 31 December 2014) of forcing terms and discharge data of Red Creek catchment were used for model spin-up (1 January 2004 – 31 December 2004), model calibration (1 January 2005 – 31 December 2009), model validation (1 January 2010 – 31 December 2012) and model testing (1 January 2013 – 31 December 2014). Catchment average daily data of potential evaporation, temperature and streamflow were downloaded from CAMELS dataset (Station 02479300) (Newman et al., 2015). The spatial distribution of daily precipitation data were considered and lumped at the subcatchment scale (three precipitation time series for the three subcatchments). Precipitation data were downloaded from the Dayment dataset (Dayment, 2020) which provides daily weather parameters (resolution: 1 km x 1 km), over North America. The time series diagrams of precipitation, potential evaporation, temperature and streamflow of Red Creek watershed are displayed in Fig. 4. Once the relevant data were processed, the user can set the algorithmic parameters of MIKA-SHA, which eventually

decide the computation power and time required for the model induction. Table 4 summarizes the algorithmic setting of MIKA-SHA used in the current study.

585 **Table 2: Catchment details**

Parameter	Details
Drainage area	1144.2 km^2
Outlet coordinates	30.73611^0, -88.78111^0
Sub catchment area %	Sub 1 – 39.0%, Sub 2 – 37.9%, Sub 3 – 23.1%
Floodplain/ Upslope	23.6% / 76.4%
Annual average discharge	1.755 mm/day
Annual average potential evaporation	3.689 mm/day
Annual average temperature	19.57 ^{0}C
Annual average precipitation	4.201 mm/day
Average slope	5.85 m/km
Forest fraction	0.89

Table 3: Area percentages of topography based HRUs

Sub Catchment	Hill	Floodplain	Plateau
1	10.4%	22.4%	67.2%
2	15.3%	23.3%	61.4%
3	14.3%	26.1%	59.6%

Table 4: Algorithmic settings of MIKA-SHA

Option	Setting
Number of independent Runs	20
Size of population	2000
Termination criteria	Generation number = 50
The randomized method used for initialization	Ramped Half and half
Special functions/ Mathematical functions	SUPERFLEX, FUSE, DISTRIBUTED/ +, -, /, *
Input variables – SUPERFLEX	Precipitation, temperature, potential evaporation
Input variables – FUSE	Precipitation, potential evaporation
Dependent variable	Streamflow
Number of objective functions used	4

Normalized range of constants	0 to 1
Depth of parse trees- initial/ maximum	SUPERFLEX – 3/5, FUSE – 2/4
The mating pool selection strategy	Tournament selection with 4 competitors at once
Genetic operator probability: mutation	
Constant/ Tree/ Separation/ Node	0.5/0.5/0.3/0.3
Genetic operator probability: crossover	0.7
Count of CPUs used for parallel computation	40 units
Level of parallel computation	Performance evaluation level
Likelihood threshold - GLUE	NSE = 0.6
Behavioural models - GLUE	5000

590

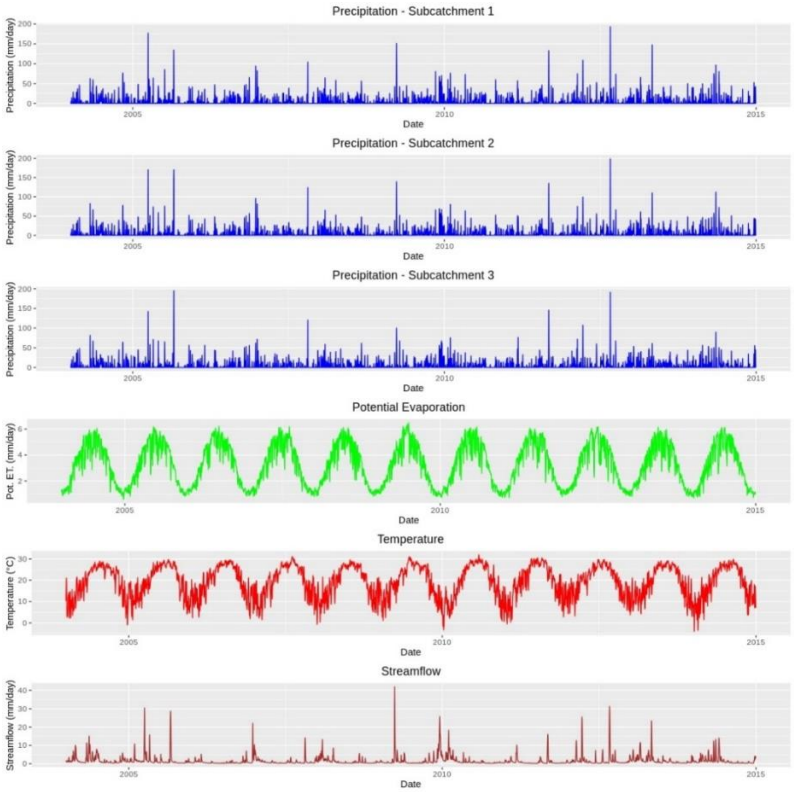

Figure 4: Forcing terms and streamflow data of Red Creek catchment

7.2 Results

7.2.1 MIKA-SHA Models induced using SUPERFLEX Building Blocks

Adhering to the methodology given in Sect. 6, three competitive models were identified. Their relative rank scores are presented in Table 5. Hence, model M2 (hereinafter referred to as SUPERFLEX_TOPO_M2) was identified as the optimal model architecture capturing the basin dynamics of Red Creek watershed. The model architecture of SUPERFLEX_TOPO_M2 is given in Fig. 5. Hillside structure of the SUPERFLEX_TOPO_M2 consists of two reservoirs connected in parallel: a fast-reacting soil reservoir (FR) and a riparian reservoir (RR). The model structure also consists of two half-triangular delay functions. The discharge of the FR incorporates a power function relationship with its storage. The model structure representing the floodplain differs from the hillside structure by the inclusion of a snow reservoir (WR). Plateau area is based on one reservoir configuration with an unsaturated soil reservoir (UR). The discharge storage relationship of UR is governed by the modified logistic function. Further, a lag function is connected with the base flow of the UR.

The performance matrix for calibration, validation and testing periods of SUPERFLEX_TOPO_M2 is given in Table 6. The high values of all four absolute performance measures suggest that SUPERFLEX_TOPO_M2 is competent in capturing the catchment dynamics of Red Creek basin. The model shows consistent behaviour throughout the calibration, validation and testing periods. Hence, we may expect no overfitting issues with training data (calibration data). Figure 6 illustrates the simulated hydrograph of SUPERFLEX_TOPO_M2 along with the observed hydrograph of the watershed. As can be seen, the simulated discharge signature matches the observed discharge signature reasonably well. It is noteworthy that SUPERFLEX_TOPO_M2 underestimates the peak discharges in some instances. Figure 7 illustrates the observed FDC of the watershed and the simulated FDCs of SUPERFLEX_TOPO_M2 for calibration, validation and testing periods. As it can be observed modelled FDCs nearly follow the measured FDC both in medium and high flow regimes but diverge slightly at low flow regime. Uncertainty analysis reveals that 75% of the observed discharge data lie within the 90% uncertainty bounds of SUPERFLEX_TOPO_M2. The sensitivity scatterplots of the model parameters of SUPERFLEX_TOPO_M2 along with the model parameters details are provided in the Appendix.

Table 5: Optimal model selection details (Library – SUPERFLEX)

Model	Rank			
	Cross Sample Entropy	Dynamic Time Warping	Number of Model Parameters	Sum
M2	2	1	2	5.0
M3	3	2	1	6.0
M1	1	3	3	7.0

Table 6: Performance matrix of SUPERFLEX_TOPO_M2

Efficiency / Period	VE	KGE	NSE	logNSE
Calibration	0.748	0.911	0.922	0.845
Validation	0.724	0.932	0.919	0.838
Testing	0.759	0.933	0.879	0.881

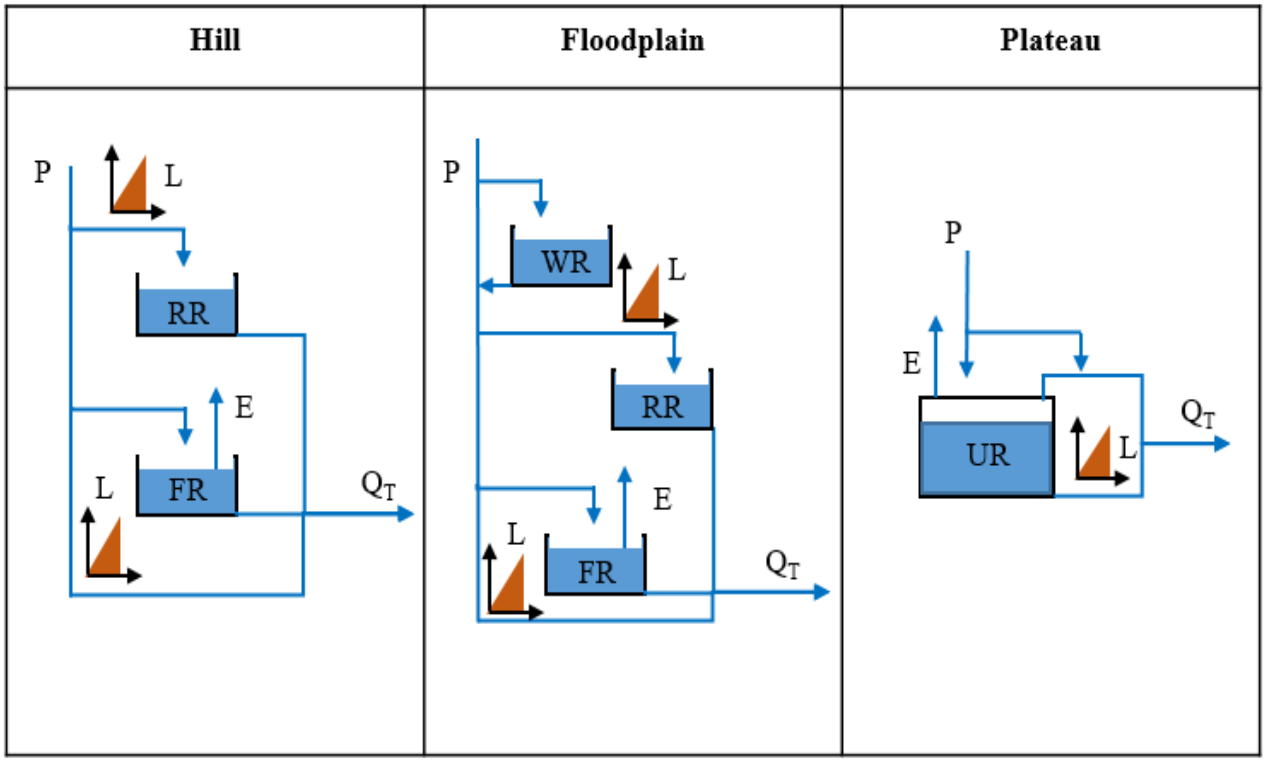

Figure 5: SUPERFLEX_TOPO_M2 model configuration (P: precipitation, E: evaporation, Q_T: total discharge, WR: snow reservoir, RR: riparian reservoir, FR: fast-reacting soil reservoir, UR: unsaturated soil reservoir, L: half-triangular lag function)

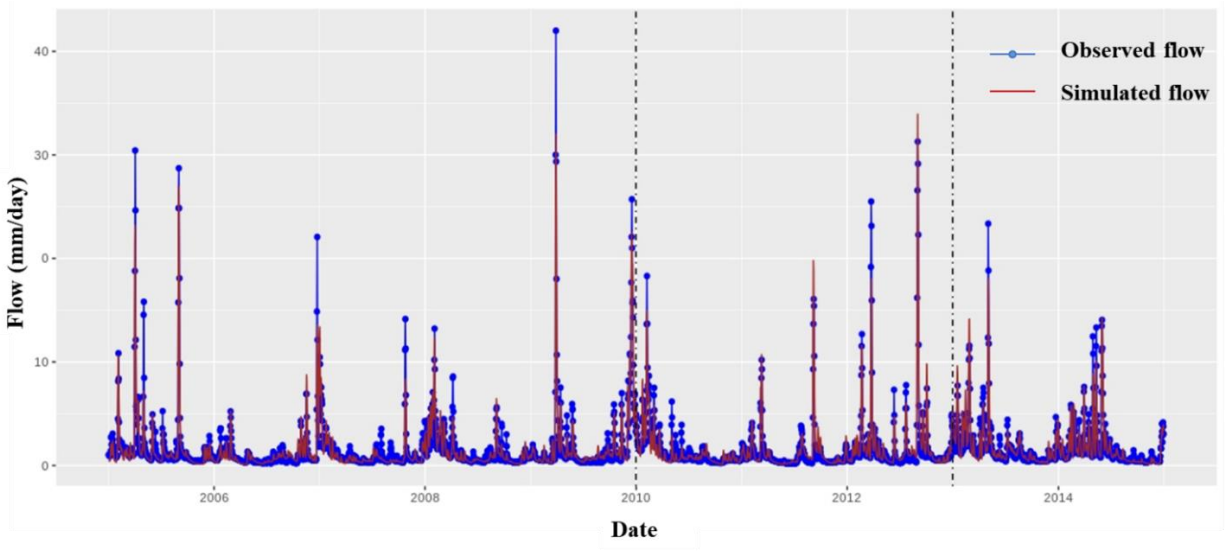

Figure 6: Hydrograph of SUPERFLEX_TOPO_M2

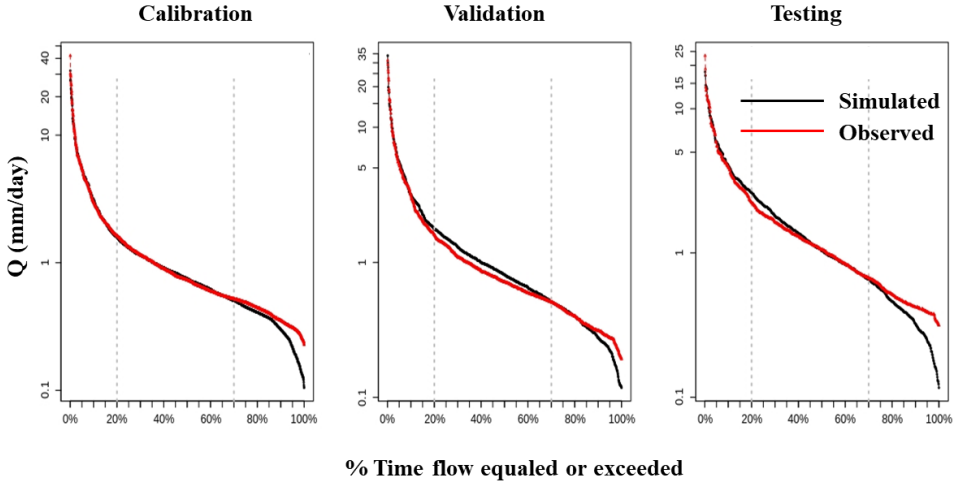

Figure 7: Flow Duration Curves of SUPERFLEX_TOPO_M2

Other than the inclusion of a WR in the floodplain model structure, both floodplain and hillside model structures share the same model architecture. Although WR is present, it is activated only if the temperature falls below a certain threshold value (calibrated value equals to 2^0C). During all 11 years of data used in the current study only less than 1% of time temperature falls below this threshold value. Hence, the effect of WR in floodplain model structure may be considered as negligible. Model structural components and the calibrated values suggest that the runoff generation in both hillside and floodplain areas respond quickly to precipitation. This quick response is reasonable in both areas due to the higher slopes at hillside and widening of the river across the floodplain in high flow events. Further, the main soil type in the floodplain area of Red Creek catchment

belongs to Smithton soil series which is characterized as soil with slow permeability and seasonally high-water table (Official Soil Series Descriptions, 2020). This may result in fast discharge generation dynamics, such as saturation excess overland flow. The constitutive function of the FR in both hillside and floodplain structure is the power function. This may help to capture the non-linear response of runoff generation. On the other hand, in plateau areas (around half of the total catchment area), one can expect higher residence times as the slopes are milder. This may lead to a delayed response in discharge to its forcing and may allow water to infiltrate more into subsurface layers. On top of that, the majority of the plateau area consists with McLaurin and Heidal soil types, which are characterized as sandy, well-drained soil types with moderate permeabilities (Official Soil Series Descriptions, 2020). Therefore, having a UR with a delayed base flow component as the plateau area model structure in SUPERFLEX_TOPO_M2 is meaningful. Finally, the choice of modified logistic function as the constitutive function may help the plateau area model structure to capture the threshold like behaviours (e.g. saturation excess overland flow) in catchment dynamics.

Out of the 34 model parameters included in SUPERFLEX_TOPO_M2, 10 model sensitive parameters can be recognized by analysing the shapes of sensitivity scatterplots. They are D_R and D_S in floodplain model structure, Beta_Qq_UR, Ce, Smax_UR, D_S, mu_Qq_UR and K_Qb_UR in plateau area model structure and two lag parameters (Lag_HRU and Lag_Sub). Please refer to the Appendix for more details about model parameters. Majority of the model sensitive parameters are connected with plateau area model structure. This is acceptable as the plateau area has the largest spatial coverage in terms of the catchment area of Red Creek catchment under topography based HRU classification. As reported earlier, obtaining a high percentage of measured data within the uncertainty bands (75%) suggests that the SUPERFLEX_TOPO_M2 is capable of estimating the total output uncertainty satisfactorily.

7.2.2 MIKA-SHA Models induced using FUSE Building Blocks

Application of MIKA-SHA with FUSE library resulted in five competitive model structures using topography based HRU classification. The relative rank scores are given in Table 7. Model M1 (hereinafter referred to as FUSE_TOPO_M1) was selected as the optimal model as it gave the lowest sum up rank. FUSE_TOPO_M1's model configuration is shown in Fig. 8. Both the hillside model structure and the floodplain model structure have the same upper- and lower-layer architectures identical to ARNO-VIC model with a single state upper soil reservoir and a fixed size base flow reservoir. Plateau area model structure incorporates a lower zone configuration like ARNO-VIC model and an upper zone configuration similar to SACRAMENTO model with the upper layer broken up into tension and free storages. Surface flow from all three model structures is developed as saturation-excess overland flow and described using the flux equations in FUSE parent model TOPMODEL. Both hillside and floodplain model structures have the same percolation mechanism which allows water to percolate from the field capacity to saturation and described using the flux equations of PRMS model, whereas in plateau area percolation is controlled by the moisture amount in the saturated zone as in SACRAMENTO model. A root weighting evaporation model is used both in floodplain and plateau area model structures, while a sequential evaporation model is used

in hillside model structure. Interflow is allowed only in the plateau area model structure and the routing is allowed only in hillside model structure.

Table 7: Optimal model selection details (Library – FUSE)

Model	Rank			
	Cross Sample Entropy	Dynamic Time Warping	Number of Model Parameters	Sum
M1	2	1	1.5	4.5
M4	4	2	1.5	7.5
M6	1	3	6	10.0
M2	3	4	5	12.0
M3	5	5	3	13.0
M5	6	6	4	16.0

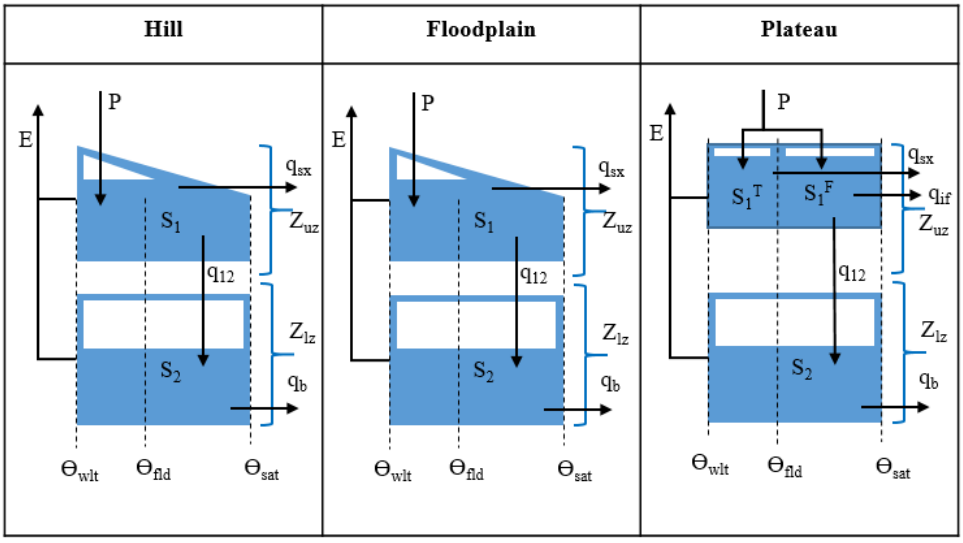

Figure 8: FUSE_TOPO_M1 model configuration (P: precipitation, E: evaporation, q_b: base flow, q_{sx}: surface flow, q_{12}: percolation, q_{if}: interflow, Z_{uz} and Z_{lz}: depth of unsaturated zone and saturated zone, S_1 and S_2: total water content in the unsaturated zone and saturated zone, S_1^F: free water content in the unsaturated zone, S_1^T: tension water content in the unsaturated zone, Θ_{wlt}: soil moisture at the wilting point, Θ_{fld}: soil moisture at field capacity, Θ_{sat}: soil moisture at saturation)

The performance matrix of FUSE_TOPO_M1 is given in Table 8. According to the high-efficiency values, simulated discharge of FUSE_TOPO_M1 shows a good match with the observed discharge data and a consistent performance throughout the calibration, validation and testing periods. Further, simulated hydrograph (Fig. 9) can capture the observed flow signature of

the watershed reasonably well. Simulated FDCs of FUSE_TOPO_M1 are presented in Fig. 10 along with observed FDC of the catchment. The simulated FDC at the calibration stage almost exactly follows the observed FDC and deviates slightly in validation and testing periods. Sensitivity scatterplots of model parameters of FUSE_TOPO_M1 and model parameter details are given in the Appendix. Ninety-four percent (94%) of the measured data fall between the 90% uncertainty bands of FUSE_TOPO_M1.

Table 8: Performance matrix of FUSE_TOPO_M1

Efficiency / Period	VE	KGE	NSE	logNSE
Calibration	0.785	0.967	0.935	0.896
Validation	0.749	0.870	0.912	0.891
Testing	0.744	0.826	0.891	0.882

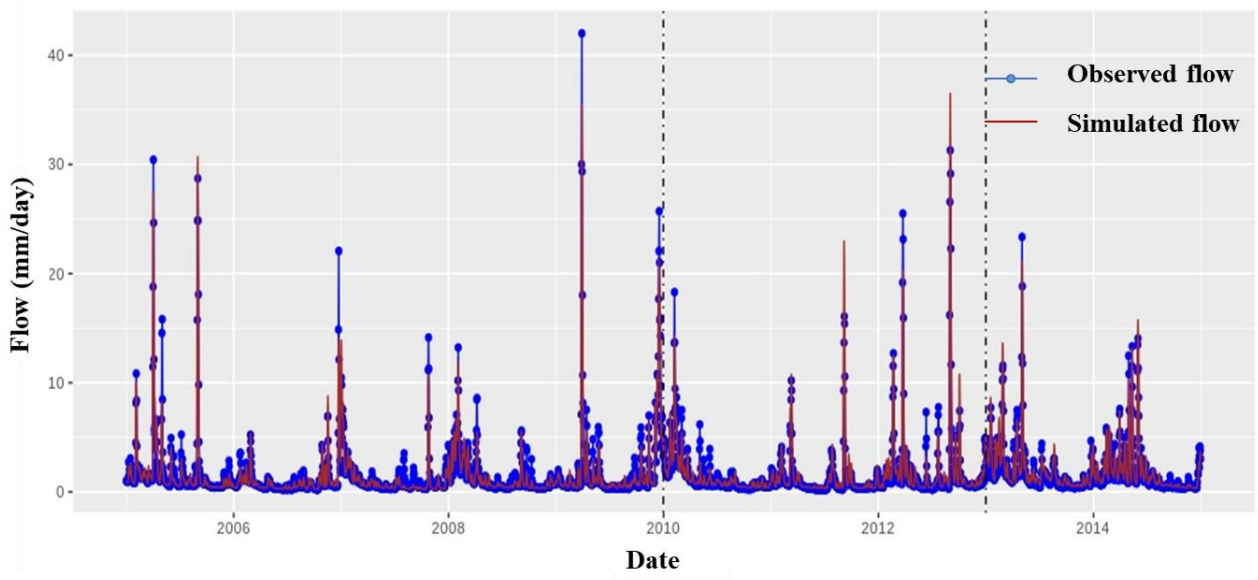

Figure 9: Hydrograph of FUSE_TOPO_M1

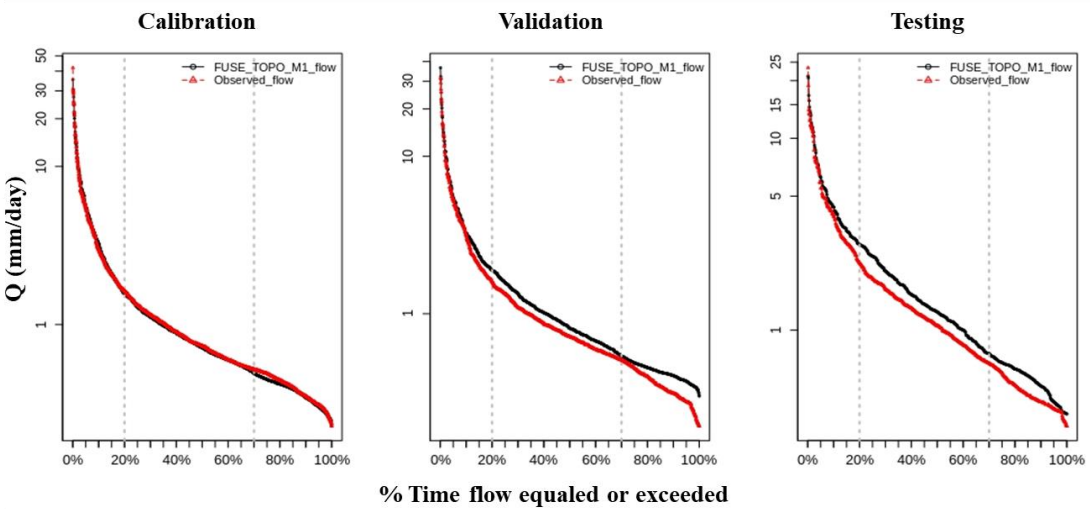

Figure 10: Flow Duration Curves of FUSE_TOPO_M1

It is interesting to note that, as in SUPERFLEX_TOPO_M2, both floodplain and hillside model structures are based on the same model configuration, whereas plateau area model structure has a different model configuration in FUSE_TOPO_M1's model architecture. This demonstrates the consistency of MIKA-SHA in capturing the similarities in runoff generation even with different model inventory libraries. Hillside model structure shows a delayed response compared to the floodplain as it allowed routing using two-parameter Gamma function. In the plateau area model structure, more subsurface type response in runoff generation can be expected as it incorporates an interflow component and its percolation is controlled by the moisture amount in the saturated zone. This goes in line with the soil properties of the Red Creek catchment. The lower layer architectures of all three model components of FUSE_TOPO_M1 include non-linear storages functions which may help capture the non-linear catchment dynamics of Red Creek catchment. Out of the 33 model parameters only 5 parameters can be identified as sensitive parameters. They are maxwater_2, baserte and qb_power in plateau model structure and the two lag parameters (Lag_HRU and Lag_Sub). This demonstrates a lesser dependency on model parameters compared to the total model performance in semi-distributed modelling owing to a large number of model parameters. FUSE_TOPO_M1 results in high value (94%) for the percentage of measured streamflow data within the confidence interval bands and hence shows a significant capability of estimating associated uncertainty.

8 Discussion and Conclusions

In this study, spatial heterogeneity of the catchment was incorporated into the model building process based on the topography of the area (i.e. three HRUs namely hills, floodplain and plateau were identified based on the topography of the area). The results obtained based on topography based HRUs, such as achieving high-efficiency values for the absolute performance measures and obtaining a good visual equivalent between measured and modelled hydrographs suggest that topography of the

catchment may have a strong impact on runoff generation. This demonstrates another potential utilization of the MIKA-SHA which is to use the toolkit to identify the runoff drivers of a catchment of interest. For example, one can also define the HRUs based on either geology or soil type of the catchment of interest and use MIKA-SHA to identify optimal model configurations. This way one can identify the relative dominance of runoff drivers towards the total catchment runoff response.

One of the major issues with machine learning models is the overfitting of the model to its training dataset. The consistent performances over the calibration, validation and testing periods of all selected optimal models through MIKA-SHA show no such issues in this case. Deterministic semi-distributed modelling would require/ rely a large number of model parameters, by comparison, a smaller number of model parameters which are sensitive towards the total model performance. Further, the values of two lag parameters associated with "DISTRIBUTED" function (Lag_HRU and Lag_Sub) were found to be crucial in achieving high model performances. As the research findings of MIKA-SHA demonstrate a logical match with previously reported research findings and fieldwork insights, it may be safe to assume that MIKA-SHA is capable of handling equifinality phenomenon satisfactorily (i.e. selected optimal models perform for the right reasons). Additionally, the quantitative model selection scheme of MIKA-SHA ensures the selected optimal model has the appropriate complexity to describe the dominant runoff generation processes of the catchment instead of selecting an optimal model only based on model parsimony.

More importantly, both SUPERFLEX_TOPO_M2 and FUSE_TOPO_M1 share similarities among their model configurations, such as the inclusion of the same model structure for hillside and floodplain HRUs, demonstration of more subsurface type delayed responses by plateau area model structures, use of non-linear constitutive functions to represent storage-discharge relationships. This consistency demonstrated by the MIKA-SHA in capturing the similar runoff dynamics across different model inventories shows that the toolkit is smart enough to mine knowledge from data which makes it feasible to depend on the induced model configurations beyond just statistical confidence.

Further, it was possible to find a reasonable match between the model structural components of optimal models and the catchment characteristics. For example, both hillside and floodplain model configurations demonstrate a quick runoff response while plateau area model configurations demonstrate a delayed subsurface type runoff response to their forcing terms. This matches the topographic characteristics (e.g. higher slopes at hillside and widening of the river across the floodplain in high flow events, high water table at floodplain, milder slopes in plateau area) and soil characteristics (plateau area: sandy, well-drained soil types with moderate permeabilities, floodplain: soil with slow permeability) of Red Creek catchment.

In this contribution, we introduce Model Induction Knowledge Augmented-System Hydrologique Asiatique (MIKA-SHA) for learning semi-distributed models where the spatial distributions of catchment properties and climate variables are taken into account. MIKA-SHA utilizes the existing hydrological knowledge to guide the machine learning algorithm which eventually results in physically meaningful hydrological models that can be readily interpretable by domain specialists. In the current

study, background hydrological knowledge is blended with the machine learning algorithm through the model building components of flexible rainfall-runoff modelling frameworks.

Results of this study indicate that the consideration of spatial distributions of forcing data and catchment properties gives more meaningful insights regarding the environmental dynamics occurring within the watershed. The unique and distinct feature of MIKA-SHA is that it could be coupled with any internally coherent collection of building blocks representing the elements of hydrological knowledge and use genetic programming to optimize both model architecture and model parameters simultaneously. This approach enables hydrologists to utilize flexible modelling frameworks to their full potential by trying many hypotheses before selecting an optimal model. MIKA-SHA is expected to be most valuable in circumstances where there may be a lack of experimental insights regarding the catchment of interest or human expert's knowledge.

We recognize the potential offered by machine learning algorithms towards hydrological modelling. However, simplistic black box type data-driven models may contribute to the development of accurate models with severe difficulties with interpretation may not serve towards the advancement of hydrological knowledge. Thus, the most promising way forward would be through the integration of existing hydrological knowledge with learning algorithms to induce more generalizable and physically consistent models. This was the motivation behind the development of the proposed MIKA-SHA framework which has been founded on both machine learning and hydrological theories. Therefore, we expect this work will strengthen the link between two leading, but historically, largely independent communities in water resource science and engineering: those working with physics-based process simulation modelling, and those working with machine learning. Finally, we expect more research studies on theory-guided machine learning to be directed towards the knowledge mining and automated model building in hydrological modelling.

Appendix A

Table A1: SUPERFLEX_TOPO_M2's model parameter details

Model parameter	Unit	Range	Symbol	Model structure		
				Hill	Flood	Plat.
K in Q = K*(S) from RR	t^{-1}	5e-2 - 4	K_Qq_RR	0.050	0.050	-
Fraction of inflow to RR	No units	0 - 1	D_R	0.593	0.132	-
K in $Q = K * S^{\alpha}$	$mm^{\alpha}\ t^{-1}$	1e-4 - 10	K_Qq_FR	0.019	0.019	-
Smoothing parameter for Poten. Evapo. of FR	No units	1e-2 - 2	m_E_FR	1.943	1.943	-
α in $Q = K * S^{\alpha}$	No units	1e-1 - 10	α_Qq_FR	2.369	2.226	-
Portion of inflow from Qq to Qb	No units	0 - 1	D_F	0.657	0.633	-
Evaporation multiplying parameter	No units	1e-1 - 3	Ce	2.060	2.536	1.418
Base of rising limb	t	1 - 10	Tlag	8.174	9.696	10
Portion of rainfall to FR	No units	0 - 1	D_S	0.850	0.802	0.012
Correction factor for snow	No units	5e-1 - 5	Cp_WR	-	0.711	-

Melt rate smoothing parameter	mm	1e-2 - 2	m_Q_WR	-	1.129	-
Potential melt rate coefficient	mm $^0C^{-1}$ t^{-1}	1e-2 - 10	Kq_WR	-	4.514	-
Snow forming temperature	^0C	0 - 10	Tp_WR	-	2.051	-
Snow melting temperature	^0C	0 - 4	Tm_WR	-	0.827	-
Runoff coefficient parameter	No units	1e-3 - 10	β_Qq_UR	-	-	9.500
Maximum reservoir capacity	mm	1e-1 - 1e4	Smax_UR	-	-	516.6
Smoothing parameter for Poten. Evapo. of UR	No units	1e-2 - 10	Beta_E_UR	-	-	4.767
State initial factor	No units	0 - 1	SiniFR_UR	-	-	0.894
Percolation coefficient	t^{-1}	1e-6 - 2	K_Qb_UR	-	-	$1x10^{-6}$
Parameter of Modified logistic curve	No units	1e-1 - 1	mu_Qq_UR	-	-	1
Max reservoir storage of IR	mm	1e-1 - 20	Smax_IR	-	-	-
Smoothing parameter for Poten. Evapo. of IR	No units	1e-3 - 1	m_QE_IR	-	-	-
Infiltration excess threshold	mm t^{-1}	1e-1 - 1e7	P_ED_max	-	-	-
Infiltration excess flow smoothing factor	mm t^{-1}	1e-3 - 10	m_P_ED	-	-	-
Time delay-HRU to subcatchment outlet	t	1e-2 - 5	lag_HRU	3.074		
Time delay-Subcatchment to catchment outlet	t	1e-2 - 5	lag_Sub	1.387		

Table A2: FUSE_TOPO_M1's model parameter details

Model parameter	Unit	Range	Symbol	Model structure		
				Hill	Flood	Plat.
Maximum total storage in upper soil layer	mm	25-500	maxwatr_1	192.9	375.1	500.0
Maximum total storage in lower soil layer	mm	50-5e3	maxwatr_2	5000	5000	361.8
Fraction total storage as tension storage	No units	0.05-0.95	fracten	0.399	0.189	0.082
1st baseflow reservoir's storage fraction	No units	0.05-0.95	fprimqb	-	-	-
Percolation rate	mm day^{-1}	0.01-1e3	percrte	169.19	153.98	-
Percolation exponent	No units	1-20	percexp	20	19.24	-
Fraction of percolation to tension storage	No units	0.05-0.95	percfrac	-	-	-
Range of the baseflow rate	No units	1e-3-1e3	baserte	838.5	14.9	149.0
Baseflow exponent	No units	1-10	qb_powr	4.369	1.188	8.993
Mean value:log-transformed topographic index	m	5-10	loglamb	8.306	9.184	9.683
Shape para: topo index gamma distribution	No units	2-5	tishape	2	5	4.1
Time delay in runoff	day	0.01-5	timedelay	0.01	-	-
Range of the fraction of roots in the upper layer	No units	0.05-0.95	rtfrac1	-	0.563	0.148
SAC percolation multiplier for dry soil layer	No units	1-250	sacpmlt	-	-	109.5
SAC percolation exponent for dry soil layer	No units	1-5	sacpexp	-	-	4.266
Interflow rate	mm day^{-1}	0.01-1e3	iflwrte	-	-	711.5
Baseflow depletion rate 1st reservoir	day^{-1}	1e-3-0.25	qbrate_2a	-	-	-
Baseflow depletion rate 2nd reservoir	day^{-1}	1e-3-0.25	qbrate_2b	-	-	-
Range of the maximum saturated area	No units	0.05-0.95	sareamax	-	-	-
Time delay-HRU to subcatchment outlet	day	0.01-5	lag_HRU	2.918		
Time delay-Subcatchment to catchment outlet	day	0.01-5	lag_Sub	3.180		

Table A3: Model configurations of competitive models – SUPERFLEX

Model	WR	IR	RR	UR	FR	SR	CR	Lag_RR	Lag_FR	Lag_SR
M1	H, F		H	H, F, P	H	H	P	F, P	H	-
M2	F	-	H, F	P	H, F	-	-	H, F	-	H, F,P
M3			H, F	P	H, F			H, F, P	H	H, F, P

H: Hill, F: Floodplain, P: Plateau

Table A4: Model configurations of competitive models – FUSE

Model	Upper Architecture			Lower Architecture			Surface runoff			Percolation			Evaporation			Interflow			Routing		
	H	F	P	H	F	P	H	F	P	H	F	P	H	F	P	H	F	P	H	F	P
M1	T/V	T/V	S	V	V	V	T	T	T	T/P	T/P	S	T/P/S	V	V	NA	NA	A	A	NA	NA
M2	T/V	T/V	S	S	V	V	V	T	T	T/P	T/P	S	T/P/S	V	V	NA	NA	A	A	NA	NA
M3	T/V	T/V	T/V	S	V	P	T	T	V	T/P	T/P	V	T/P/S	T/P/S	T/P/S	NA	NA	NA	NA	A	A
M4	T/V	T/V	T/V	V	V	T	V	T	V	S	T/P	S	V	T/P/S	T/P/S	A	NA	NA	NA	NA	NA
M5	T/V	T/V	T/V	S	V	T	V	T	V	S	T/P	S	V	T/P/S	T/P/S	A	NA	NA	NA	NA	NA
M6	T/V	P	T/V	V	P	S	V	V	T	S	S	T/P	T/P/S	T/P/S	T/P/S	A	NA	NA	A	NA	A

As calculated in T: TOPMODEL, V: VIC, P: PRMS, S: SACRAMENTO. A: Allowed, NA: Not allowed

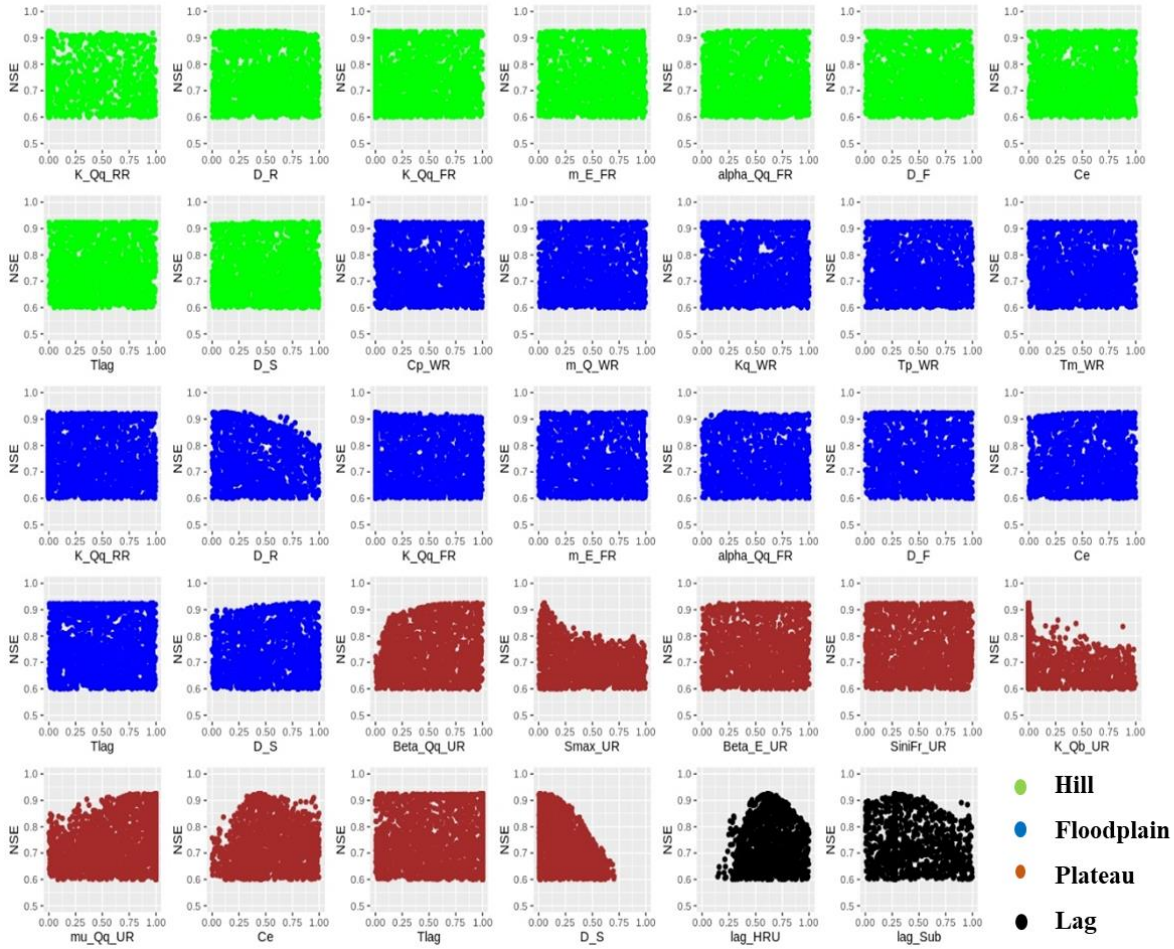

Figure A1: Sensitivity scatter plots of SUPERFLEX_TOPO_M2's model parameters

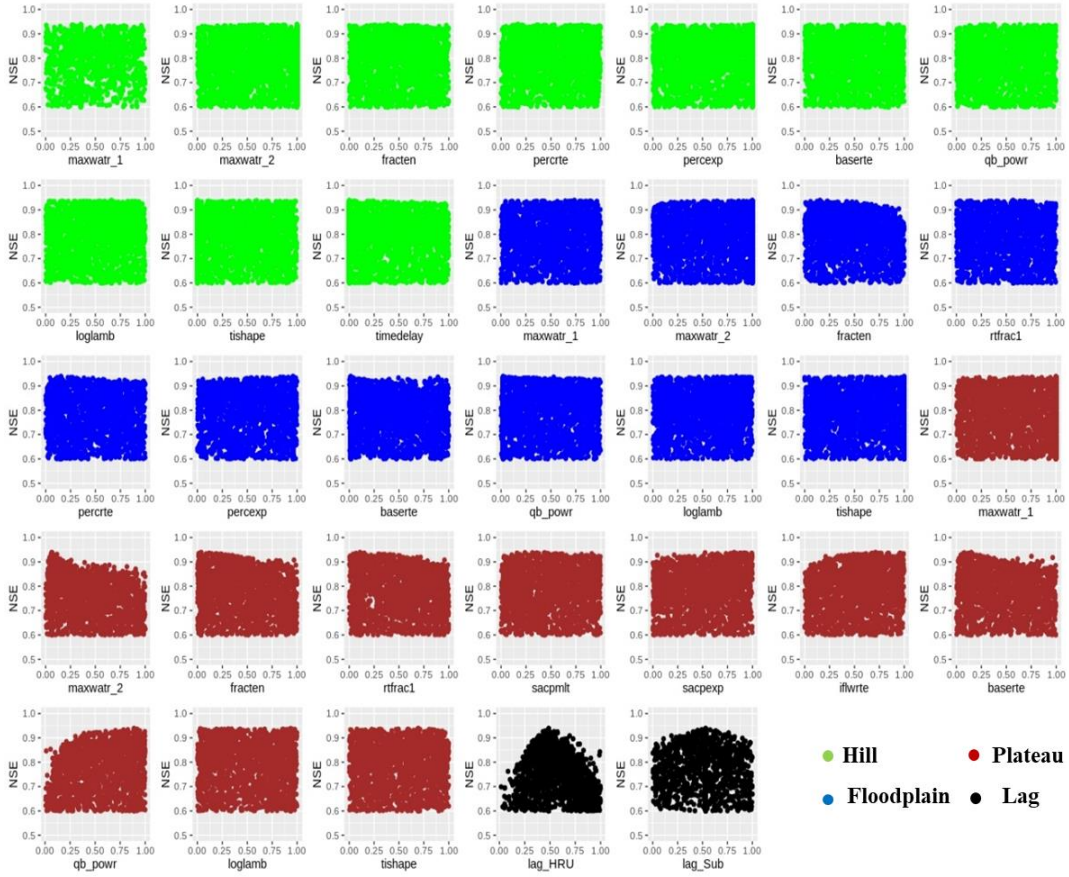

Figure A2: Sensitivity scatter plots of FUSE_TOPO_M1's model parameters

Data Availability

The data sets used in this study are publicly available through the references given in the manuscript.

Author Contribution

H. M. V. Vidura Herath: Conceptualization, Formal Analysis, Methodology, Software, Visualization, Writing – original draft

Jayashree Chadalawada: Conceptualization, Methodology, Software

Vladan Babovic: Conceptualization, Project administration, Supervision, Writing – review & editing

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

We greatly appreciate the support given by Dr. Vojtech Havlicek in R programming. Further, we acknowledge Dr. Fabrizio Fenicia for his assistance in SUPERFLEX framework.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell P. E. and Rasmussen, J.: An introduction to the European Hydrological System – Syst`eme Hydrologique Europ´een (SHE): 1. History and philosophy of a physically based distributed modelling system, Journal of Hydrology, 87, 45–59, 1986a.

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J.: An introduction to the European Hydrological System – Syst`eme Hydrologique Europ´een, (SHE): 2. Structure of a physically-based distributed modelling system, Journal of Hydrology, 87, 61–77, 1986b.

Addor, N., and Melsen, L. A.: Legacy, rather than adequacy, drives the selection of hydrological models, Water Resources Research, 55, https://doi.org/10.1029/2018WR022958, 2019.

Aerts, J., Kriek, M., and Schepel, M.: STREAM (spatial tools for river basins and environment and analysis of management options): 'set up and requirements', Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, 24(6), 591-595, doi:10.1016/s1464-1909(99)00049-0, 1999.

Afan, H. A., El-Shafie, A., Yaseen, Z. M., Hameed, M. M., Wan Mohtar, W. H. M., and Hussain, A.: ANN based sediment prediction model utilizing different input scenarios, Water Resour. Manage., 29, 1231–1245, doi: 10.1007/s11269-014-0870-1, 2014.

Afzaal, H., Farooque, A. A., Abbas, F., Acharya, B., and Esau, T.: Groundwater Estimation from Major Physical Hydrology Components Using Artificial Neural Networks and Deep Learning, Water, 12, 5, doi:10.3390/w12010005, 2019.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large Area Hydrologic Modeling and Assessment Part I: Model Development, Journal of the American Water Resources Association, 34, 73–89, 1998.

Babovic, V.: Data mining in hydrology, Hydrological Processes, 19, 1511-1515, 2005.

Babovic, V.: Introducing knowledge into learning based on genetic programming, Journal of Hydroinformatics, 11, 181–193, 2009.

Babovic, V., and Abbott, M. B.: The evolution of equations from hydraulic data Part II: Applications, J. Hydraul. Res., 35, 411–430, 1997.

Babovic, V., and Keijzer, M.: Forecasting of river discharges in the presence of chaos and noise, in: Coping with floods: Lessons learned from recent experiences, edited by Marsalek, J., nato arw series, Dordrecht, Kluwer, 1999.

Babovic, V., and Keijzer, M.: Genetic programming as a model induction engine, Journal of Hydroinformatics, *2*, 35–60, 2000.

Babovic, V., and Keijzer, M.: Rainfall runoff modelling based on genetic programming, Hydrology Research, 33, 331–346, 2002.

Babovic, V., Keijzer, M., and Bundzel, M.: From global to local modelling: a case study in error correction of deterministic models, in: Proceedings of the Fourth International Conference on Hydroinformatics, 4, Lowa City, USA, 2000a.

Babovic, V., Keijzer, M., and Stefansson, M.: Optimal embedding using evolutionary algorithms, in: Proceedings of 4th Int. Conference on Hydroinformatics, Cedar Rapids, USA, 2000b.

Babovic, V., Keijzer, M., Aguilera, D. R., and Harrington, J.: An evolutionary approach to knowledge induction: Genetic programming in hydraulic engineering, in: Proceedings of the world water and environmental resources congress, 64, Orlando, Florida, 2001.

Babovic, V., Li, X., and Chadalawada, J.: Rainfall–Runoff Modeling Based on Genetic Programming, in: Encyclopedia of Water: Science, Technology, and Society, 5 Volume Set, edited by Maurice, P., Wiley, New York, USA, 1081, 2020.

Baptist, M. J., Babovic, V., Uthurburu, J. R., Keijzer, M., Uittenbogaard, R. E., Mynett, A., and Verwey, A.: On inducing equations for vegetation resistance, Journal of Hydraulic Research, 45, 435-450, 2007.

Bautu, A., and Bautu, E.: Meteorological data analysis and prediction by means of genetic programming, in: Proceedings of the 5th workshop on mathematical modeling of environmental and life sciences problems, 35–42, Constanta, Romania, 2006.

Beven, K.: Down to basics: Runoff processes and the modelling process, in: Rainfall-runoff modelling: the primer, Wiley-Blackwell, West Sussex, United Kingdom, 1–22, 2012a.

Beven, K.: Predicting Hydrographs Using Models Based on Data in: Rainfall-runoff modelling: the primer, Wiley-Blackwell, West Sussex, United Kingdom, 83–118, 2012b.

Beven, K.: Beyond the Primer: Next Generation Hydrological Models, in: Rainfall-runoff modelling: the primer, Wiley-Blackwell, West Sussex, United Kingdom, 313–327, 2012c.

Beven, K.: Deep Learning, Hydrological Processes and the Uniqueness of Place, Hydrological Processes, in press, https://doi.org/10.1002/hyp.13805, 2020.

Beven, J. K., and Binley, M. A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6, 278–298, 1992.

Beven, K., Lamb, R., Quinn, P., Romanowicz, R., Freer, J.: Topmodel, in: Computer models of watershed hydrology, edited by Singh, V. P., Water Resource Publications, Colorado, 627–668, 1995.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z., and Smith, M.: Toward improved streamflow forecasts: Value of semidistributed modelling, Water Resources Research, 37, 2749-2759, doi:10.1029/2000wr000207, 2001.

Brunton, S. L., Proctor, J. L., and Kutz, J. N.: Discovering governing equations from data by sparse identification of nonlinear dynamical Systems, in: Proceedings of the National Academy of Sciences, 113, 3932–3937, 2016.

Burnash, R. J. C.: The NWS River Forecast System-Catchment modeling, in: Computer models of watershed hydrology, edited by Singh, V.P., Water Resource Publications, Colorado, 311–366, 1995.

Cannon, A. J.: Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes, Stochastic Environmental Research and Risk Assessment, 32, 3207–3225, 2018.

Cannon, A. J., and Mckendry, I. G.: A graphical sensitivity analysis for statistical climate models: Application to indian monsoon rainfall prediction by artificial neural networks and multiple linear regression models, International Journal of Climatology, 22, 1687–1708, https://doi.org/10.1002/joc.811, 2002.

Chadalawada, J., and Babovic, V.: Review and comparison of performance indices for automatic model induction, Journal of Hydroinformatics, 21, 13–31, 2017.

Chadalawada, J., Havlicek, V. and Babovic, V.: A Genetic Programming Approach to System Identification of Rainfall-Runoff Models, Water Resour Manage., 31, 3975–3992, doi:10.1007/s11269-017-1719-1, 2017.

Chadalawada, J., Herath, H. M. V. V., and Babovic, V.: Hydrologically informed machine learning for rainfall-runoff modeling: A genetic programming-based toolkit for automatic model induction, Water Resources Research, 56, https://doi.org/10.1029/2019WR026933, 2020.

Chang, L. C., Shen, H. Y., and Chang, F. J.: Regional flood inundation nowcast using hybrid SOM and dynamic neural networks, J. Hydrol., 519, 476–489, http://dx.doi.org/10.1016/j.jhydrol.2014.07.036, 2014.

Chiang, Y.M., Chang, L.C., Chang, F.J.: Comparison of static-feedforward and dynamic-feedback neural networks for rainfall–runoff modeling, J. Hydrol., 290, 297–311, http://dx.doi.org/10.1016/j.jhydrol.2003.12.033, 2004.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resources Research, 44, W00B02, https:// doi.org/10.1029/2007WR006735, 2008.

Clark, M. P., Hilary, K. M., Daniel, B. G. C., Kavetski, D., and Woods, R. A.: Hydrological field data from amodeller's perspective: Part 2: Process-based evaluation of model hypotheses, Hydrological Processes, 25(4), 523–543, 2010.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M.: A unified approach for process-based hydrologic modelling: 1. Modelling concept, Water Resources Research, 51, 2498–2514, doi:10.1002/2015wr017198, 2015a.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., and Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., and Marks, D. G.: The structure for unifying multiple modelling alternatives (SUMMA), Version 1.0: Technical description, NCAR Tech. Note NCAR/TN-5141STR, 2015b.

Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., Mai, J., Serrer, M., Sgro, N., Shafii, M., Snowdon, A. P., and Tolson, B. A.: Flexible watershed simulation with the Raven hydrological modelling framework, Environmental Modelling & Software, 129 (December 2019), 104,728, https://doi.org/10.1016/jenvsoft.2020.104728, 2020.

Criss, R. E., and Winston, W. E.: Do Nash values have value? Discussion and alternate proposals, Hydrological Processes, 22, 2723, 2008.

Dahamsheh, A., and Aksoy, H.: Markov chain-incorporated artificial neural network models for forecasting monthly precipitation in arid regions, Arab. J. Sci. Eng., http://dx.doi.org/10.1007/s13369-013-0810-z, 2013.

Datta, B., Prakash, O., and Sreekanth, J.: Application of Genetic Programming Models Incorporated in Optimization Models for Contaminated Groundwater Systems Management, Advances in Intelligent Systems and Computing EVOLVE – A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V, 183-199, doi:10.1007/978-3-319-07494-8_13, 2014.

Daymet: https://daymet.ornl.gov/, last access: 20 March 2020.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6, 182–197, 2002.

Dehghani, M., Saghafian, B., Nasiri Saleh, F., Farokhnia, A., and Noori, R.: Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation, Int. J. Climatol., 34, 1169–1180, http://dx.doi.org/10.1002/joc.3754, 2014.

Delgado-Bonal, A., and Marshak, A.: Approximate Entropy and Sample Entropy: A Comprehensive Tutorial, Entropy, 21, 541, doi:10.3390/e21060541, 2019.

Elshorbagy, A., and El-Baroudy, I.: Investigating the capabilities of evolutionary data-driven techniques using the challenging estimation of soil moisture content, Journal of Hydroinformatics, 11, 237–251, 2009.

Elshorbagy, A., Simonovic, S. P., and Panu, U. S.: Estimation of missing streamflow data using principles of chaos theory. Journal of Hydrology, 255, 123–133, 2002.

Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., and Jiang, J.: Comparison of Long Short Term Memory Networks and the Hydrological Model in Runoff Simulation, Water, 12, 175, doi:10.3390/w12010175, 2020.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Brian A. Ebel, B. A., Jones, N., Kim, J., Mascaro, G., Richard G. Niswonger, R. G., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and David Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, Journal of Hydrology, 537, 45-60, doi: 10.1016/j.jhydrol.2016.03.026, 2016.

Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resources Research, 47, W11510, https://doi.org/10.1029/2010WR010174, 2011.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: Is there a correspondence? Hydrological Processes, 28, 2451–2467, 2014.

Fenicia, F., Kavetski, D., Savenije, H. H., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modelling decisions, Water Resources Research, 52, 954–989, doi:10.1002/2015WR017398, 2016.

Ferreira, L. B., Cunha, F. F., Oliveira, R. A., and Filho, E. I.: Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach, Journal of Hydrology, 572, 556-570, doi: 10.1016/j.jhydrol.2019.03.028, 2019.

Fleming, S. W.: Artificial neural network forecasting of nonlinear Markov processes, Canadian Journal of Physics, 85, 279–294, https://doi.org/10.1139/p07-037, 2007.

Fleming, S. W., Bourdin, D. R., Campbell, D., Stull, R. B., and Gardner, T.: Development and operational testing of a super-ensemble artificial intelligence flood-forecast model for a Pacific Northwest River, JAWRA Journal of the American Water Resources Association, 51, 502–5125, https://doi.org/10.1111/jawr.12259, 2014.

Freeze, R. A., and Harlan, R. L.: Blueprint for a physically-based, digitally-simulated hydrologic response model, Journal of Hydrology, 9, 237–258, 1969.

García-Alba, J., Bárcena, J. F., Ugarteburu, C., and García, A.: Artificial neural networks as emulators of process-based models to analyse bathing water quality in estuaries, Water Research, 150, 283-295, doi: 10.1016/j.watres.2018.11.063, 2019.

Gholami, V., Chau, K.W., Fadaee, F., Torkaman, J., and Ghaffari, A.: Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers, Journal of Hydrology, 529, 1060–1069, 2015.

Giorgino, T.: Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package, Journal of Statistical Software, 31, 1–24, 2009.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., and Reed, P. M.: Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations, Journal of Water Resources Planning and Management, 142, 4,015,050, https://doi.org/10.1061/(asce)wr.1943-5452.0000570, 2015.

Govindaraju, R. S.: Artificial neural networks in hydrology. II: Hydrologic applications, Journal of Hydrologic Engineering, 5, 124–137, 2000.

Gude, V., Corns, S., and Long, S.: Flood Prediction and Uncertainty Estimation Using Deep Learning, Water, 12, 884, doi:10.3390/w12030884, 2020.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, 2009.

Havlič̆cek, V., Hanel, M., Maca, P., Kuraž, M., and Pech, P.: Incorporating basic hydrological concepts into genetic programming for rainfall-runoff forecasting, Computing, 95, 363–380, 2013.

Hsieh, W.W.: Machine learning in the environmental sciences, Cambridge University Press, Cambridge, United Kingdom: 2009.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., and Lou, Z.: Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation, Water, 10, 1543, doi:10.3390/w10111543, 2018.

Humphrey, G. B., Gibbs, M. S., Dandy, G. C., and Maier, H. R.: A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network, Journal of Hydrology, 540, 623–640, 2016.

Karimi, S., Shiri, J., Kisi, O., and Shiri, A. A.: Short-term and long-term streamflow prediction by using ‘wavelet-gene expression’ programming approach, ISH Journal of Hydraulic Engineering, 22, 148–162, 2016.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., and Kumar, V.: Theory-guided data science: A new paradigm for scientific discovery from data, IEEE Transactions on Knowledge and Data Engineering, 29, 2318–2331, doi: 10.1109/TKDE.2017.2720168, 2017.

Kavetski, D., and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resources Research, 47, W11511, https://doi.org/10.1029/2011WR010748, 2011.

Keijzer, M., and Babovic, V.: Declarative and preferential bias in GP-based scientific discovery, Genetic Programming and Evolvable Machines, 3, 41–79, 2002.

Khu, S. T., Liong, S.Y., Babovic, V., Madsen, H., and Muttil, N.: Genetic programming and its application in real-time runoff forecasting, Journal of the American Water Resources Association, 37, 439–451, 2001.

Knoben, W. J. M., Freer, J. E., Fowler, K. J., Peel, M. C., and Woods, R. A.: Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1.2: An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations, Geoscientific Model Development, 12, 2463-2480, doi:10.5194/gmd-12-2463-2019, 2019.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., and Woods, R. A.: A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments, Water Resources Research, 56, e2019WR025975. https://doi.org/10.1029/2019WR025975, 2020.

Koza, J. R.: Genetic programming: on the programming of computers by means of natural selection, 1, MIT press, Cambridge, Massachusetts, England, 1992.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resources Research, *55*, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019a.

Kratzert, F., Klotz, D., Schulz, K., Klambauer, G., Hochreiter, S., and Nearing, G.: Benchmarking a Catchment-Aware Long Short-Term Memory Network (LSTM) for Large-Scale Hydrological Modeling, Hydrol. Earth Syst. Sci. Dis., https://doi.org/10.5194/hess-2019-368, 2019b.

Krause, P., Boyle, D., and Base, F.: Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89–97, 2005.

Kumar, D., Singh, A., Samui, P., and Jha, R. K.: Forecasting monthly precipitation using sequential modelling, Hydrological Sciences Journal, 64, 690-700, doi: 10.1080/02626667.2019.1595624, 2019.

Leavesley, G. H., Markstrom, S. L., Viger, R. J., and Hay, L. E.: The Modular Modelling System (MMS): a toolbox for water- and environmental resources management, in: Hydrological Modeling in Arid and Semi-Arid Areas, edited by Wheater, H., Sorooshian, S., and Sharma, K. D., Cambridge University Press, 87-98, doi: 10.1017/CBO9780511535734.008, 2008.

Ley, R., Hellebrand, H., Casper, M. C., and Fenicia, F.: Comparing classical performance measures with signature indices derived from flow duration curves to assess model structures as tools for catchment classification, Hydrology Research, 47, 1–14, 2016.

McClelland, J. L., and Rumelhart, D. E.: Parallel distributed processing: Explorations in the microstructure of cognition, The MIT Press, Massachusetts, Cambridge, 1986.

Mcgovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., and Smith, T.: Making the black box more transparent: Understanding the physical implications of machine learning, Bulletin of the American Meteorological Society, 2175–2199, 2019.

Mehr, A. D., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A. M. A., and Yaseen, Z. M.: Genetic programming in water resources engineering: A state-of-the-art review, Journal of Hydrology, 566, 643–667, 2018.

Meshgi, A., Schmitter, P., Babovic, V., and Chui, T. F. M.: An empirical method for approximating stream baseflow time series using groundwater table fluctuations, Journal of Hydrology, 519, 1031–1041, 2014.

Meshgi, A., Schmitter, P., Chui, T. F. M., and Babovic, V.: Development of a modular streamflow model to quantify runoff contributions from different land uses in tropical urban environments using genetic programming, Journal of Hydrology, 525, 711–723, 2015.

Minns, A.W., and Hall, M. J.: Artificial neural networks as rainfall-runoff models, Hydrological Sciences Journal, 41, 399–417, 1996.

Molin, M. D., Schirmer, M., Zappa, M., and Fenicia, F.: Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: The case study of the Thur catchment, Hydrology and Earth System Sciences, 24(3), 1319-1345, doi:10.5194/hess-24-1319-2020, 2020.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models Part I—A discussion of principles, Journal of hydrology, 10, 282–290, 1970.

Nayak, P. C., Satyaji Rao, Y. R., and Sudheer, K. P.: Groundwater level forecasting in a shallow aquifer using artificial neural network Approach, Water Resources Management, 20, 77–90, https://doi.org/10.1007/s11269-006-4007-z, 2006.

Nearing, G., Yatheendradas, S., Crow, W., Zhan, X., Liu, J., and Chen, F.: The efficiency of data assimilation, Water resources research, 54 (9), 6374-6392, 2018.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resources Research, doi: 10.1029/2020WR028091, in press, 2020a.

Nearing, G. S., Sampson, A. K., Kratzert, F., and Frame, J. M.: Post-Processing a Conceptual Rainfall-Runoff Model with an LSTM, EarthArXiv preprint, https://doi.org/10.31223/osf.io/53te4, 2020b.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al.: Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous USA: Dataset characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, 2015.

Nielsen, S. A., and Hansen, E.: Numerical simulation of the rainfall-runoff process on a daily basis, Hydrology Research, 4, 171–190, 1973.

Nourani, V., Komasi, M., and Mano, A.: A multivariate ANN-wavelet approach for rainfall-runoff modelling, Water Resources Management, 23, 2877–2894, https://doi.org/10.1007/s11269-009-9414-5, 2009.

O'Connell, P. E.: A historical perspective, in: Recent Advances in the Modeling of Hydrologic Systems, edited by: Bowles, D. S., and O'Connell, P. E., Springer Science+Business Media, B.V., Kluwer, Dordrecht, 3–30, doi:10.1007/978-94-011-3480-4, 1991.

Official Soil Series Descriptions: https://soilseries.sc.egov.usda.gov/, last access:29 July 2020.

Oyebode, O. K., and Adeyemo, J. A.: Genetic programming: Principles, applications and opportunities for hydrological modelling, World Academy of Science, Engineering and Technology International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering, 8, 348–354, 2014.

Physics Informed Machine Learning Conference, Santa Fe, New Mexico, USA, 2016.

QGIS.org: QGIS Geographic Information System, Open Source Geospatial Foundation Project, http://qgis.org, 2020.

R Core Team.: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2018.

Raissi, M., Perdikaris, P., and Karniadakis, G. E.: Physics informed deep learning (Part II): Data-driven discovery of nonlinear partial differential equations, arXiv preprint arXiv:1711.10566, 2017.

Richman, J. S., and Moorman, J. R.: Physiological time-series analysis using approximate entropy and sample entropy**,** Am. J. Physiol. Heart Circul. Physiol., 278, 20139 20149, doi:10.1152/ajpheart.2000.278.6.H2039, 2000.

Rudy, S. H., Brunton, S. L., Proctor, J. L., and Kutz, J. N.: Data-driven discovery of partial differential equations, Science Advances, 3, 1602614, 2017.

Safari, M. J. S., and Mehr, A. D: Multigene genetic programming for sediment transport modeling in sewers at non-deposition with deposited bed condition, Int. J. Sedim. Res., https://doi.org/10.1016/j.ijsrc.2018.04.007, 2018.

Sakoe, H. and Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. on Acoust., Speech, and Signal Process, 26, 43–49, 1978.

Salas, F. R., Somos-Valenzuela, M. A., Dugger, A., Maidment, D. R., Gochis, D. J., David, C. H., Yu, W., Ding, D., Clark, E. P., and Nomanet, N.: Towards real-time continental scale streamflow simulation in continuous and discrete space, JAWRA Journal of the American Water Resources Association, 54, 7-27, 2018.

Salvador, S., and Chan, P.: Toward accurate dynamic time warping in linear time and space, Intelligent Data Analysis, 11, 561-580, doi:10.3233/ida-2007-11508, 2007.

Savic, D., and Khu, S. T.: Evolutionary computing in hydrological sciences, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., 331–348, Wiley, New York, USA, 2005.

Sellars, S.: "grand challenges" in big data and the earth sciences, Bulletin of the American Meteorological Society, 99 (6), ES95-ES98, 2018.

Selle, B., and Muttil, N.: Testing the structure of a hydrological model using genetic programming, Journal of Hydrology, 397, 1–9, 2011.

Shen, C.: A trans-disciplinary review of deep learning research and its relevance for water resources scientists, Water Resour. Res., https://doi.org/10.1029/2018WR022643, 2018.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F., Ganguly, S., Hsu, K., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., and Tsai, W.: HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community, Hydrology and Earth System Sciences, 22, 5639-5656, doi:10.5194/hess-22-5639-2018, 2018.

Singh, G., Kandasamy, J., Shon, H.K., and Cho, J.: Measuring treatment effectiveness of urban wetland using hybrid water quality – artificial neural network (ANN) model, Desalin. Water Treat., 32, 284–290, doi: 10.5004/dwt.2011.2712, 2011.

Snauffer, A. M., Hsieh, W. W., Cannon, A. J., and Schnorbus, M. A.: Improving gridded snow water equivalent products in British Columbia, Canada: Multi-source data fusion by neural network models, The Cryosphere, 12, 891–905, 2018.

Solander, K. C., Bennett, K. E., Fleming, S. W., and Middleton, R. S.: Estimating hydrologic vulnerabilities to climate change using simulated historical data: A proof-of-concept for a rapid assessment algorithm in the colorado river basin, Journal of Hydrology: Regional Studies, 26, 100,642, 2019.

Sugawara, M.: Automatic calibration of the tank model/l'etalonnage automatique d'un modele a cisterne, Hydrological Sciences Journal, 24, 375–388, 1979.

Sun, Y., Babovic, V., and Chan, E. S.: Artificial neural networks as routine for error correction with an application in Singapore regional Model, Ocean Dynamics, 62, 661–669, 2012.

Todini, E.: The ARNO rainfall-runoff model, J. Hydrol., 175, 339–382, doi:10.1016/S0022-1694(96)80016-3, 1996.

Todini, E.: Hydrological catchment modelling: past, present and future, Hydrology and Earth System Sciences, 11 (1), 468-482, 2007.

USDA's Geospatial Data Gateway: http://datagateway.nrcs.usda.gov/, last access: 21 March 2020.

USGS EarthExplorer: https://earthexplorer.usgs.gov/, last access: 20 March 2020.

van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D., and Lobligeois, F.: The influence of conceptual model structure on model performance: a comparative study for 237 French catchments, Hydrol. Earth Syst. Sci., 17, 4227–4239, https://doi.org/10.5194/hess-17-4227-2013, 2013.

Vitolo, C.: Exploring data mining for hydrological modelling, Ph.D. thesis, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom, 2015.

Vojinovic, Z., Kecman, V., and Babovic, V.: Hybrid approach for modeling wet weather response in wastewater systems, Journal of water resources planning and management, 129, 511-521, 2003.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S.: A framework for development and application of hydrological models, Hydrology and Earth System Sciences Discussions, 5, 13–26, 2001.

Wang, N., Zhang, D., Chang, H., and Li, H.: Deep learning of subsurface flow via theory-guided neural network, Journal of Hydrology, 584, 124700, doi: 10.1016/j.jhydrol.2020.124700, 2020.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L., Bierkens, M. F., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C.,

Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring earth's terrestrial water. Water Resources Research, 47 (5), https://doi.org/10.1029/2010WR010090, 2011.

Xiang, Z., Yan, J., and Demir, I.: A rainfall-runoff model with LSTM-based sequence-to-sequence learning, Water Resources Research, 56, e2019WR025326, https://doi.org/10.1029/2019WR025326, 2020.

Xiaodong, S., Ganlin, Z., Feng, L., Decheng, L., Yuguo, Z., and Jinling, Y.: Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model, Journal of Arid Land, 8, 734–748, doi: 10.1007/s40333-016-0049-0, 2016.

Xu, T., Longyang, Q., Tyson, C., Zeng, R., Neilson, B. T., and Tarboton, D. G.: Hybrid physically-based and deep learning modeling of a snow dominated mountainous karst watershed, in: Presentation at the American Geophysical Union fall meeting, San Francisco, CA, 2019.

Yaseen, Z.M., El-shafie, A., Jaafar, O., and Afan, H.A.: Artificial intelligence based models for stream-flow forecasting: 2000-2015, Journal of Hydrology, 530, 829-844, http://dx.doi.org/10.1016/j.jhydrol.2015.10.038, 2015.