AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR033836

Special Section:
Advancing flood characterization, modeling, and communication

Development of a Fast and Accurate Hybrid Model for Floodplain Inundation Simulations

Niels Fraehr[1], Quan J. Wang[1], Wenyan Wu[1], and Rory Nathan[1]

[1]Department of Infrastructure and Engineering, Faculty of Engineering and Information Technology, The University of Melbourne, Melbourne, VIC, Australia

Key Points:

- A flood inundation model is developed, based on Low-fidelity hydrodynamic modeling, Spatial analysis, and Gaussian Process learning (LSG)
- The LSG model predicts water depths with a mean Root Mean Square Error (RMSE) of 4 cm and a standard deviation of 5 cm
- The LSG model is 12 times faster than a traditional high-resolution hydrodynamic model

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

N. Fraehr,
nfraehr@student.unimelb.edu.au

Citation:

Fraehr, N., Wang, Q. J., Wu, W., & Nathan, R. (2023). Development of a fast and accurate hybrid model for floodplain inundation simulations. *Water Resources Research*, *59*, e2022WR033836. https://doi.org/10.1029/2022WR033836

Received 9 OCT 2022
Accepted 14 MAY 2023

Abstract High computational cost is often the most limiting factor when running high-resolution hydrodynamic models to simulate spatial-temporal flood inundation behavior. To address this issue, a recent study introduced the hybrid Low-fidelity, Spatial analysis, and Gaussian Process learning (LSG) model. The LSG model simulates the dynamic behavior of flood inundation extent by upskilling simulations from a low-resolution hydrodynamic model through Empirical Orthogonal Function (EOF) analysis and Sparse Gaussian Process learning. However, information on flood extent alone is often not sufficient to provide accurate flood risk assessments. In addition, the LSG model has only been tested on hydrodynamic models with structured grids, while modern hydrodynamic models tend to use unstructured grids. This study therefore further develops the LSG model to simulate water depth as well as flood extent and demonstrates its efficacy as a surrogate for a high-resolution hydrodynamic model with an unstructured grid. The further developed LSG model is evaluated on the flat and complex Chowilla floodplain of the Murray River in Australia and accurately predicts both depth and extent of the flood inundation, while being 12 times more computationally efficient than a high-resolution hydrodynamic model. In addition, it has been found that weighting before the EOF analysis can compensate for the varying grid cell sizes in an unstructured grid and the inundation extent should be predicted from an extent-based LSG model rather than deriving it from water depth predictions.

Plain Language Summary Every year, lives are lost, and infrastructure is destroyed due to floods. This highlights the need for fast and accurate flood predictions to inform flood forecasting and risk assessments. However, predicting flood inundation in high resolution is often not practically feasible due to the high computational cost involved in running complex computer models. Simplified computer models can be used to provide faster flood predictions, but they lack the accuracy provided by complex models. To address this issue, this study evaluates an alternative method based on the combination of a fast simple model together with an advanced spatial feature matching method. The advanced spatial feature matching method is used to convert the predictions obtained from the simple model to accurate predictions of flood inundation depth and extent. The new approach is applied to a large floodplain in Australia and different adaptations are explored to optimize the procedure and ensure robust performance. The new approach is compared to the use of a traditional complex model and a previous approach that only predicted inundation extent. The new approach shows similar accuracy to the traditional complex model while being 12 times faster, thereby making it more practically useful for flood risk assessments.

1. Introduction

Each year, flooding causes massive destruction of infrastructure and loss of lives all around the world. Taking Australia as an example, the cost of the 2011 Queensland floods was AU$2.38 billion (Australian Institute for Disaster Resilience, 2012), and the 2022 February–March floods in Queensland and New South Wales caused damages of AU$4.8 billion (The Insurance Council of Australia, 2022). Another example is the 2022 floods in Pakistan that have killed around 1,500 people and displaced more than 33 million people (Goldbaum & ur-Rehman, 2022). With future climate prediction, the recurrence of such flooding events is only expected to increase (IPCC, 2021; Kirezci et al., 2020), highlighting the need for effective modeling techniques to assist risk assessment, design of new infrastructure and real-time forecasting.

Flood inundation is traditionally modeled using high-resolution two-dimensional (2D) hydrodynamic models that simulate the physical processes of the flood event from a set of boundary conditions (Bates, 2022; Razavi et al., 2012; Teng et al., 2017). These models can simulate flood inundation accurately with a high degree of

realism and hence are often referred to as high-fidelity models (Razavi et al., 2012). However, due to the degree of detail needed (high resolution) and the complex nature of flood events, the computational costs of high-fidelity models are often too high for these models to be used for real-time modeling and flood risk assessment through ensemble modeling, where hundreds or thousands of model realizations are needed (Teng et al., 2017; Wu et al., 2020). To improve the computational efficiency of high-fidelity models, researchers have explored parallel and high-performance computing (Neal et al., 2009; Sanders & Schubert, 2019), graphics processing unit technologies (Ming et al., 2020; Morales-Hernández et al., 2021), and more efficient solution algorithms (Bates & De Roo, 2000; Sridharan et al., 2021). Although these methods have been shown to improve the computational efficiency of the simulations, the use of high-fidelity hydrodynamic models to simulate flood inundation of large regional-sized domains (>100 km^2, Bentivoglio et al. (2022)) with high resolution (1–100 m) continues to present computational challenges for practical applications involving real-time ensemble forecasts. To address this issue, researchers have developed surrogate models to provide approximate flood inundation simulations with a lower computational burden than high-fidelity models (Razavi et al., 2012).

Various types of surrogate models have been developed, and they can be divided into three categories based on the model structure: Conceptual, Low-fidelity, and Emulator models. Conceptual models are based on simple hydraulic concepts and are normally very fast, but they cannot capture dynamic behavior (e.g., Lhomme et al. (2008), Nobre et al. (2016), and Teng et al. (2019)). Low-fidelity models are physics-based hydrodynamic models which are faster but of lower accuracy compared to high-fidelity models (e.g., Altenau et al. (2017), Bates and De Roo (2000), Bomers et al. (2019), and Liu et al. (2019)). Emulator models are data-driven models, which are very fast and are able to predict complex relationships accurately; however, they cannot capture spatial correlation and are often restricted to low-dimensional data (e.g., Chu et al. (2020), Kabir et al. (2021), Xie et al. (2021), and Zhou et al. (2021)). Additional information on each surrogate type can be found in the literature reviews by Asher et al. (2015), Bates (2022), McGrath et al. (2018), Razavi et al. (2012), and Teng et al. (2017).

All three types of surrogate models have advantages and limitations. This has led to the concept of developing hybrid approaches that combine the benefits of multiple models whilst overcoming some of the limitations. One of the most recent hybrid models is the Low-fidelity, Spatial analysis, and Gaussian Process learning (LSG) model developed by Fraehr et al. (2022). The LSG model accurately simulates the dynamic behavior (e.g., the rising and recession components, and hysteresis) of flood inundation at a lower computational cost than high-fidelity models. The LSG model first uses a low-fidelity model to simulate flood inundation on a coarsely discretized grid. Due to the coarse resolution of the low-fidelity model, the simulation time is significantly faster than using a high-fidelity model, but the accuracy is also reduced. Thus, the primary purpose of the low-fidelity model in the LSG methodology is to capture the temporal and spatial dependencies of flood behavior in a computationally efficient way. While the low-fidelity simulation step is fast, the accuracy of the simulations needs to be improved to make the predictions useful for practical purposes. This can be done by developing a relationship between low- and high-fidelity model predictions based on a training data set, and then using this relationship to upskill the accuracy of the low-fidelity simulations (e.g., Yang et al. (2022) and Carreau and Guinot (2021)). In the LSG model, the upskilling of low-fidelity predictions is carried out by first applying Empirical Orthogonal Function (EOF) analysis to reduce the dimensionality of the low-fidelity data into a small number of independent features. The dimension reduction of the EOF analysis facilitates the training of a Sparse Gaussian Process (GP) model to convert the key low-fidelity features to high-fidelity features. GP models have been used in numerous studies and perform well in describing non-linear relationships (e.g., Contreras et al. (2020), Ma et al. (2019), and Parker et al. (2019)), but they are computationally demanding to optimize for large data sets (Bauer et al., 2016; Burt et al., 2019). For this reason, a Sparse GP model is used in the LSG methodology as it provides a high level of computational efficiency by approximating the full GP model by a set of assumptions (Leibfried et al., 2021). After the conversion through the Sparse GP model, the predicted high-fidelity features can be used to reconstruct flood inundation surfaces with high resolution and accuracy without needing to run a high-fidelity model.

Although Fraehr et al. (2022) demonstrated the good performance of the LSG model, the methodology was only applied to the simulation of inundation extent and timing of a flood event and not the water depth. Information on water depth is important to correctly represent the degree of hazard associated with a predicted inundation extent of a flood event (Hunter et al., 2007), and is a key indicator in warning systems and for estimating flood losses (Antony et al., 2021; Chang et al., 2019; Zischg et al., 2018). For those reasons, further development of the LSG model to predict the temporal-spatial distribution of water depths in inundated areas would make the LSG model substantially more useful for risk assessment. However, continuous variables, like water depth, are generally

harder to predict accurately than binary (i.e., wet/dry) indicators of inundation. This means that deriving inundation extent from water depth predictions instead of directly predicting inundation extent might affect the accuracy of the LSG model. Accordingly, there is value in extending the development of the LSG model to accommodate water depth predictions to determine whether there is any reduction in the accuracy of the simulations.

Furthermore, in the study by Fraehr et al. (2022) the LSG model was only applied to a coupled 1D-2D hydrodynamic model that had a structured quadratic grid (i.e., one which provided predictions at regular spacing throughout the model domain without increasing grid resolution in areas of interest). A coupled 1D-2D model uses a 1D component to simulate the flow and water depth in the mainstream of the river, and a 2D component to simulate inundation of the floodplain. Although this type of hydrodynamic model has been shown to provide good performance historically, modern hydrodynamic inundation models tend to be fully 2D and utilize unstructured (i.e., irregular or flexible) grids to describe complex geometries (Bates, 2022; Teng et al., 2017). Enabling the LSG model to accommodate unstructured grids will thus expand the possible applications of the model and strengthen its relevance to current practice.

In unstructured grids, cell sizes vary across the model domain, and it is a common practice to take these differences into account to ensure good model performance. Varying cell sizes are commonly seen in climate science and oceanography, where latitude-longitude grids with converging meridians are used. EOF analysis is also used in these research areas, where it is normal to compensate for irregular grids by applying weights according to the size of the grid cells before the EOF analysis (Baldwin et al., 2009; Hannachi et al., 2007). The general idea behind area-weighting is to ensure that larger cells are not valued equally to smaller cells. However, the general principle of area-weighting might not be applicable in flood inundation modeling as larger cells are normally located on the floodplain and smaller cells in the river regions, noting that flow behavior on floodplains is generally more gradually varying than flow within the main channel. Adopting a weighting scheme in the LSG model that is directly proportional to grid cell area could be problematic as this would give small cells located in regions of rapidly varying flow low weight, which may potentially reduce the accuracy of the model. Therefore, the effect that weighting has on the performance of the LSG model needs to be investigated to determine how the LSG model can be successfully applied to unstructured grids.

In this study, we further develop the methodology of the LSG model to simulate water depth and thereby strengthen the model's capabilities for the challenges encountered in practical applications. To investigate how the new developments affect the accuracy, the predictions of flooding extent obtained from the revised LSG model are compared to both a high-fidelity model as well as the original LSG model. The versatility of the LSG model is explored by applying the LSG model to a 2D hydrodynamic model with an unstructured grid to simulate flood inundation on the flat and complex Chowilla floodplain in Australia. In this application, we use a low-fidelity model that is considerably coarser than the model used previously in the study by Fraehr et al. (2022) to test the upskilling and speed-up capabilities of the LSG methodology even further. Accordingly, our objectives in this study are to (a) advance the capability of the LSG model to predict the temporal-spatial distribution of the water depth in inundated areas rather than just the flood extent and (b) explore the utility of adopting a weighting scheme to compensate for variable grid sizes in unstructured grids.

The paper is organized as follows. In Section 2, the methodology of the LSG model and the new developments are presented. The evaluation and case study for the application is presented in Section 3 and Section 4, respectively, followed by the results in Section 5. Finally, Sections 6 and 7 provide discussion and conclusions based on the results.

2. Methodology of the LSG Model

The overall concept of the LSG model is to rapidly derive accurate inundation estimates using information previously obtained from a small number of low- and high-fidelity model simulations, a process that avoids the computational burden of running a detailed hydrodynamic model for every set of new boundary conditions. This significantly enhances the computational efficiency without great loss of accuracy and represents a practical means of providing rapid estimates of complex flood behavior. The LSG model predicts inundation by upskilling inundation predictions from a low-fidelity model simulation. The upskilling is done by using EOF analysis to reduce the dimensionality of the spatial-temporal inundation behavior through EOF analysis. Essentially, EOF analysis is a means to identify a modest number of independent components that are representative of the spatial

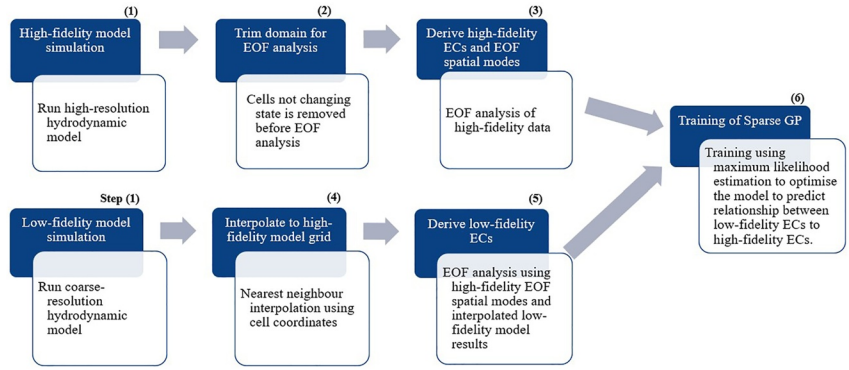

Figure 1. Workflow of training of the Low-fidelity, Spatial analysis, and Gaussian Process learning model.

and temporal patterns of inundation behavior. This is necessary as data-driven emulator models are not well suited to capturing spatial correlation and perform best when applied to low-dimensional data. After the EOF analysis, a Sparse GP model is used to convert the low-fidelity temporal components to high-fidelity temporal components. Finally, the high-resolution inundation prediction is obtained by reconstructing the hydrodynamic results through reverse EOF analysis using the predicted high-fidelity temporal components together with the high-fidelity spatial components.

In the following, we describe the details of the methodology of the LSG model for predicting water depth, and at the end of this section, we explain how the methodology differs from the one previously proposed by Fraehr et al. (2022).

2.1. Training of LSG Model

Before the LSG model can be used to predict flood inundation, it needs to be set up and trained. There are 6 steps involved in the LSG model training as shown in Figure 1. Steps 1, 2, and 3 and Steps 1, 4, and 5 involve deriving key spatial-temporal components through EOF analysis for the high- and low-fidelity data, respectively. The key temporal components are thus used in Step 6 to train the Sparse GP model. The details of the individual steps are explained in the following sections.

2.1.1. Step 1: Creation of Training Data Set

A training data set is needed to facilitate the EOF analysis and training of the Sparse GP model. To create a training data set for the LSG model, high- and low-fidelity models have to be set up for the specific study area. First, the high-fidelity model is set up and calibrated. Second, the low-fidelity model is created, normally by simplifying the high-fidelity model. Thus, the training data set is created by first running the high-fidelity model for a large number of flood events, and then running the low-fidelity model for the same flood events. The training events must span a wide range of inundation behavior to ensure the model performance is sustained for new events not included in the training.

2.1.2. Step 2: Trim Model Domain for EOF Analysis

When simulating flood inundation over a computational grid, some cells never get flooded. These cells do not contain any valuable information and therefore only slow down the EOF analysis in Step 3. Thus, by trimming the spatial domain to only contain cells that change state (changes in water depth) during the training events, the EOF analysis can be performed more efficiently (noting that the training events must cover the full range of conditions expected in the future to ensure all potential flood-prone areas are included after the trimming.). The trimming is carried out by categorizing the cells into two groups, namely: “dry” and “wet” cells. The wet cells are those whose water depth varies throughout the training events and are therefore the only ones included in the EOF analysis. A threshold of 3 cm water depth is applied to differentiate dry and wet cells and reduce noise.

2.1.3. Step 3: Reducing Dimensionality of High-Fidelity Data Through EOF Analysis

The high dimensionality of the spatial-temporal high-fidelity data set cannot easily be captured using a Sparse GP model. To address this issue, EOF analysis is used to reduce the dimensionality of the data set while maintaining the spatial and temporal correlation. EOF analysis deconstructs spatial-temporal data sets into pairs of spatial and temporal components. Each pair of spatial and temporal components is referred to as a mode, where the spatial components are called spatial maps (EOFs) and the temporal components are called expansion coefficients (ECs). All modes are orthogonal to the others (i.e., they are fully independent of one another), and the EOF analysis aims to find a linear combination of modes that maximizes the variance of the data set (Jolliffe & Cadima, 2016).

To find the modes, we first assemble a $T \times N$ matrix called D_{HF} containing the simulated water depths from the high-fidelity model. T is the number of timesteps in the training data set and N is the number of wet cells found through categorization in Step 2. Each column in D_{HF} is detrended by subtracting the temporal mean. This ensures centering of the data to a mean of zero.

2.1.3.1. Applying Weighting in the EOF Analysis

The next step in the EOF analysis is to decide whether to perform weighting or not. Weighting is included by multiplying D_{HF} with a vector containing the weights for each cell included in the EOF analysis.

As described in the introduction, weighting according to cell sizes is normally used to compensate for varying grid cells in unstructured grids in the areas of climate science and oceanography (Baldwin et al., 2009; Hannachi et al., 2007). The purpose of the weighting is to ensure larger cells, which account for a larger proportion of the model domain, are weighted higher. However, in hydrodynamic modeling of flood inundation, the smaller grid cells are usually located in the regions of rapidly varying flow (e.g., rivers), and it is important to simulate flow behavior in these areas precisely. Weighting according to cell sizes could therefore affect the accuracy of the LSG model as the river regions would be given a low weight. On the other hand, not including weighting might not represent flood behavior in the larger cells on the floodplain correctly, which could thus affect the accuracy of the inundation predictions. To examine the importance of weighting, this study builds two versions of the LSG model, one weighted and one unweighted, and applies them to simulate inundation on an unstructured grid (See Section 3).

2.1.3.2. Performing EOF Analysis and Deriving Significant Modes

Finally, after deciding to apply weighting or not, the modes are found via singular value decomposition of D_{HF}, following Equation 1.

$$D_{\mathrm{HF}} = U_{\mathrm{HF}} \cdot C_{\mathrm{HF}} \approx \sum_{k=1}^{K} U_{\mathrm{HF}}(k, :) \cdot C_{\mathrm{HF}}(:, k) \tag{1}$$

where U_{HF} is a $T \times N$ matrix where each row is an EOF spatial map, C_{HF} is a $T \times T$ matrix where each column corresponds to an EC temporal function.

After retrieving the modes, they are ranked according to the proportion of the data set's variance they explain. The dimension reduction of the EOF analysis thereby lies in selecting only a few (K) significant modes that describe the majority of the variance in the data set. These modes are chosen by satisfying both North's test, where a mode is considered significant if its eigenvalue lies outside the error limits of the eigenvalue for the previous mode (North et al., 1982), and Kaiser's Rule, where the eigenvalue should be above 1 (Kaiser, 1960). Once the K significant modes are found, they can be used to reconstruct D_{HF} with little loss of information. To assist the understanding of the EOF analysis, an example is given in Text S1 in Supporting Information S1.

2.1.4. Step 4: Interpolate Low-Fidelity Data to High-Fidelity Grid

The low-fidelity ECs can be derived using the high-fidelity spatial modes (See Step 5). This ensures the same basis of EOF spatial modes is used for both the high- and low-fidelity ECs and is a more computationally efficient process compared to performing EOF analysis from scratch due to the time-consuming derivation of the covariance matrix and eigenvalue decomposition. To facilitate the low-fidelity ECs derivation, the low-fidelity water surface elevations (water depth + terrain elevation) are interpolated to the high-fidelity grid using a nearest-neighbor method. After the interpolation, the areas where the terrain elevation of the high-fidelity cell is higher than the interpolated low-fidelity water surface elevation are assumed dry. This reduces the extent of the

interpolated low-fidelity results, and initial tests have shown that it helps minimize overestimation of the inundation extent for the LSG model.

2.1.5. Step 5: Derive Low-Fidelity ECs

The low-fidelity ECs are derived by inserting the interpolated low-fidelity data into a $T \times N$ matrix called D_{LF}. Since the low-fidelity data has been converted to the high-fidelity grid in Step 4, the D_{LF} and D_{HF} matrices have the same dimensions. Detrending is performed by subtracting the high-fidelity temporal mean derived for D_{HF} in Step 3. The high-fidelity temporal mean is used to detrend the low-fidelity data to ensure comparability between D_{LF} and D_{HF}, and as this mean is used in Step 9 for reconstruction of the high-resolution flood inundation. Finally, after detrending, the optional weighting is applied and the high-fidelity EOF spatial maps U_{HF} are used to derive the low-fidelity ECs in Equation 2.

$$C_{\mathrm{LF}} = D_{\mathrm{LF}} \cdot U'_{\mathrm{HF}} \tag{2}$$

where C_{LF} is a $T \times T$ matrix containing the low-fidelity ECs, and U'_{HF} is the transpose of U_{HF}. This approach is applied in a similar way by Zhao et al. (2022) to calibrate precipitation fields.

2.1.6. Step 6: Training of the Sparse GP Model

Once both the low- and high-fidelity ECs have been derived for the training data set, they can be used to train a Sparse GP model to predict the high-fidelity ECs from the low-fidelity ECs.

GP models assume the relationship between input and output follows a Gaussian distribution of functions, and by doing so, can predict non-linear relationships with statistical confidence (see Equation 3) (Rasmussen & Williams, 2006).

$$\mathrm{GP}(x) \sim \mathcal{N}\left(m(x), k\left(x, x'\right)\right) \tag{3}$$

$$k\left(x, x'\right) = \sigma_f^2 \exp\left(-\frac{x - x'}{2l}\right) + \sigma_n^2 \tag{4}$$

where $\mathrm{GP}(x)$ is the Gaussian function, $m(x)$ is the mean function, $k(x, x')$ is the covariance function (kernel), and x is the input variable, in this case, the low-fidelity ECs in C_{LF}. In the kernel function, σ_f^2 is the signal variance, l is the lengthscale, σ_n^2 is the noise variance and $x - x'$ is the Euclidean distance between inputs. As the data has been detrended in Steps 3 and 5, the mean function can be assumed to be zero (Rasmussen & Williams, 2006). For the kernel, an Exponential kernel function is used to describe the covariance (see Equation 4), following Fraehr et al. (2022).

The reason for using a Sparse GP in the LSG model instead of the standard full GP model is due to the high computational demand of the full GP model. A GP model is optimized using maximum likelihood estimation, which requires an inversion of the covariance matrix. This inversion has a computational demand of $\mathcal{O}\left(T^3\right)$ and makes the full GP model infeasible to be used for large data sets (Bauer et al., 2016; Leibfried et al., 2021). To address this issue, the Sparse GP model uses a number M of inducing variables, which should be significantly less than the T number of timesteps in the training data set. The inducing points are a reduced set of input variables that are optimized to approximate the full GP model, and thereby reduces the cost of the matrix inversion to $\mathcal{O}\left(T M^2\right)$ (Snelson & Ghahramani, 2006; Titsias, 2009). The Sparse GP model chosen for the LSG model is based on variational inference, as this has been shown to improve with an increasing number of inducing variables (Bauer et al., 2016).

In the LSG model, individual Sparse GP models are used to predict each significant mode of the high-fidelity ECs, resulting in a total of K models (See Figure 2). Each Sparse GP model receives all low-fidelity ECs as input and predicts one high-fidelity EC as output. All inputs and outputs are standardized to zero mean and unit variance before incorporating them into the Sparse GP models.

Each Sparse GP model is optimized using maximum likelihood estimation of the hyperparameters: σ_f^2, l and the inducing variables. The number of inducing variables is found by a trial-and-error approach. For the application in this paper, 2% of the number of input samples has shown to be a sufficient number of inducing variables, but for smaller data sets, a larger proportion is most likely necessary. The optimization process for the hyperparameters

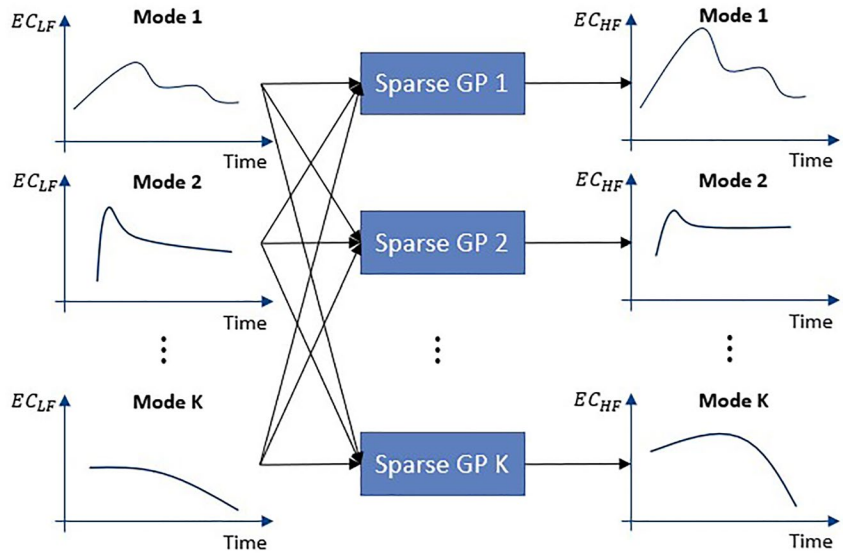

Figure 2. Expansion coefficients conversion using Sparse Gaussian Process models.

is performed using the L-BFGS-B optimization algorithm. More details of the initialization and optimization of the hyperparameters can be found in Fraehr et al. (2022).

2.2. Prediction Using the LSG Model

After finalizing the training phase, the LSG model can be used to predict flood inundation by following the workflow in Figure 3. In the prediction workflow, Steps 1, 4, and 5 from the training workflow are grouped together in Step 7 and involve deriving the low-fidelity temporal components through EOF analysis. In Step 8 the Sparse GP model is used to convert the low-fidelity components to high-fidelity components and finally, in Step 9 the flood inundation is reconstructed in high-resolution.

2.2.1. Step 7: Run the Low-Fidelity Model and Derive the Low-Fidelity ECs

In the prediction phase, only the fast low-fidelity model needs to be run to simulate the flood inundation. This is what makes the LSG model more computationally efficient than using a high-fidelity model. The low-fidelity is run for a new unseen event and the simulation results are converted to low-fidelity ECs, following the process described in Steps 4 and 5 (Note: It is still the high-fidelity EOF spatial maps from the training data that are used to derive the low-fidelity ECs. This ensures the ECs are comparable between events.).

2.2.2. Step 8: Predict High-Fidelity ECs Using the Sparse GP Model

The low-fidelity ECs are used as input to the Sparse GP models to predict the high-fidelity ECs, which can then be used to reconstruct the flood inundation. The Sparse GP model provides both a mean and variance of the predictions (Rasmussen & Williams, 2006).

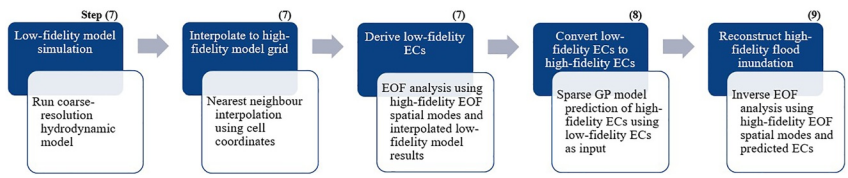

Figure 3. Workflow of prediction using the Low-fidelity, Spatial analysis, and Gaussian Process learning model.

2.2.3. Step 9: Inverse EOF Analysis to Reconstruct Flood Inundation

By assembling the predictions of high-fidelity ECs into a matrix $\widehat{C_{\mathrm{LSG}}}$, a high-fidelity prediction of flood inundation is obtained by reversing the EOF analysis to reconstruct the temporal-spatial inundation data from temporal functions of ECs and EOF spatial maps using Equation 5.

$$\widehat{D_{\mathrm{LSG}}} = \sum_{k=1}^{K} U_{\mathrm{HF}}(k, :) \cdot \widehat{C_{\mathrm{LSG}}}(:, k) \tag{5}$$

where $\widehat{D_{\mathrm{HF}}}$, after re-adding the high-fidelity temporal mean subtracted before the EOF analysis in Step 3, is the LSG model's temporal-spatial prediction of flood inundation.

When reconstructing the inundation, the first K significant modes do not explain all the variance in the data set. This means that there are minor deviations (noise) in the water depth prediction causing otherwise dry areas to appear to be inundated by insignificant shallow water depths. To address this issue, a threshold of 3 cm water depth is applied. This alleviates the problem of the LSG model predicting insignificant flooding in some cells. Finally, the dry cells identified in Step 2 are added to the $\widehat{D_{\mathrm{LSG}}}$ matrix to reconstruct the full prediction of flood inundation.

2.3. The LSG Model for Directly Predicting Flood Extent

Although the extent of flood inundation is usually derived from water depth predictions, it can also be predicted directly, as is the case in the previously proposed LSG model in Fraehr et al. (2022). Predicting the flood extent directly might result in higher accuracy for the flood extent, as this bypasses the process of deriving the flood extent from water depth predictions and the potential error from this derivation. To examine how the accuracy of predicting the flood extent directly compares to deriving it from water depth predictions, we construct both types of LSG models (See Section 3 for further details on the comparison).

In the following, we show the differences in the LSG model for direct extent prediction purposed by Fraehr et al. (2022) compared to the approach presented here to predict water depth. Only the steps that differ from the workflow in Sections 2.1 and 2.2 are presented.

2.3.1. Step 2: Convert to Binary Values Before Trimming the Model Domain

The high-fidelity model results in the training data set are converted to binary values (1 for flooded, 0 for dry) by applying a threshold of 3 cm water depth. This binarization facilitates categorizing the cells into three groups: dry, always wet, and temporary wet (TW). Only the TW cells change state when using binary values and are therefore the only ones included in the EOF analysis.

2.3.2. Step 3: Perform EOF Analysis Only on Temporarily Flooded Cells

With the new categories of cells, the EOF analysis is performed only on TW cells. This results in a D_{HF} matrix that is $T \times N_{\mathrm{TW}}$. N_{TW} is the number of TW cells identified. The EOF analysis is still performed using Equation 1.

2.3.3. Step 4: Interpolation of Low-Fidelity Binary Data

In Fraehr et al. (2022) the interpolation of the low-fidelity data to the high-fidelity grid is performed using binary values. Binary values cannot easily be related to terrain elevation, and for that reason, there was no filtering of areas where the water surface elevation of the low-fidelity model was below the terrain elevation of the high-fidelity cell. However, in the LSG model for direct extent prediction used in this study, we include filtering based on the water surface elevation as in Step 4 for the LSG model for water depth predictions (See Section 2.1.4). This is an improvement over the previous methodology and ensures the same high-fidelity model cells are used in both the water-depth and extent-based versions of the LSG model.

2.3.4. Step 9: Binary Threshold to Reconstruct Data

As mentioned in Section 2.2.3, not all of the variance in the data set is explained via the K significant modes. When reconstructing the flood extent, this results in noise, so the values do not completely reconstruct to 1 (flooded) or 0 (dry). To address this issue and convert the predictions to fully binary values, a threshold of 0.5 is

used. The full prediction of the flood extent is thus reconstructed by adding the always wet and dry cells identified in Step 2 to the $\widehat{D_{\mathrm{HF}}}$ matrix.

3. Evaluation of LSG Model for Water Depth Predictions

In this study, a new LSG model for water depth predictions is proposed. However, as discussed in Step 3 of the methodology (See Section 2.1.3.1), it needs to be examined how weighting according to grid cell sizes affects the accuracy of the LSG model when applied to an unstructured grid. Applying weights will give the larger cells normally located on the floodplain a higher weight than the smaller cells in the river regions. The rivers are normally the source of flooding and giving a smaller weight to these areas could therefore potentially reduce the accuracy of the LSG model. We examine this by creating two versions of the LSG model, one with weighting that we call LSG-WD (Weighted) and one without weighting that we call LSG-WD (Unweighted). Both of the models are evaluated in their ability to provide comparable inundation simulations to a high-fidelity model.

Besides the two LSG models for water depth predictions, we also create a LSG model for direct flood extent predictions following the methodology in Fraehr et al. (2022). To ensure that the same high-fidelity cells are used in all the LSG models tested in this study, we made one slight change to the methodology and adopted the same interpolation strategy as used to predict water depths where water surface elevation below the terrain elevation is assumed dry (See Section 2.3.3 for further details). We name this model LSG-EXT (Weighted) and use this to examine if the accuracy of the flood extent predictions is influenced by the use of water depth predictions compared to predicting the extent directly, as in the previous version of the LSG model (noting that only a weighted version is used for direct extent prediction as this was advocated by Fraehr et al. (2022) for unstructured grids, although it was not tested). If the accuracy is significantly higher by directly predicting the extent, it might be worth considering using two LSG models, one for predicting flood extent directly and one for predicting water depth for those areas predicted as being flooded.

3.1. Evaluation of Water Depth Predictions

The LSG models' ability to predict water depth is evaluated using Root Mean Square Error (RMSE) in Equation 6:

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y^{\mathrm{LSG}}(t) - y^{\mathrm{HF}}(t)\right)^2} \tag{6}$$

where y^{LSG} is the LSG prediction and y^{HF} is the high-fidelity simulation.

Furthermore, the results are plotted as a scatter plot to examine if the LSG model generally over- or under-predicts the water depth compared to the high-fidelity model. The low-fidelity model simulation will be used as a benchmark for comparison.

3.2. Evaluation of Inundation Extent Predictions

The overall prediction of inundation extent is evaluated using the same metrics as used by Fraehr et al. (2022), that is Relative RMSE (relRMSE), Relative Peak Value Error (relPeakValErr) and Relative Peak Time Error compared to the peak period (relPeakTimeErr) in Equations 7–9:

$$\mathrm{relRMSE} = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(A_{\mathrm{LSG}}(t) - A_{\mathrm{HF}}(t)\right)^2}}{\frac{1}{T}\sum_{t=1}^{T}A_{\mathrm{HF}}(t)} \tag{7}$$

$$\mathrm{relPeakValErr} = \frac{\overline{A_{\mathrm{LSG}}^{\mathrm{peak},5\%}} - \overline{A_{\mathrm{HF}}^{\mathrm{peak},5\%}}}{\overline{A_{\mathrm{HF}}^{\mathrm{peak},5\%}}} \tag{8}$$

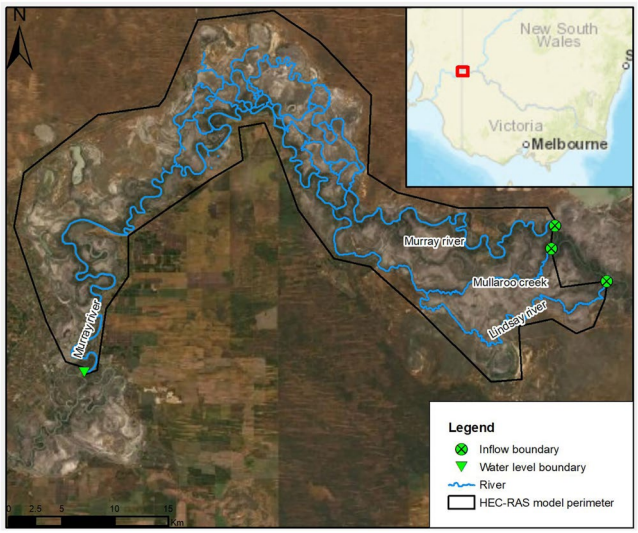

Figure 4. Overview of the Chowilla floodplain study site and perimeter of the Hydrologic Engineering Center's River Analysis System model (ESRI, 2022).

$$\text{relPeakTimeErr} = \frac{\overline{t_{\text{LSG}}^{\text{peak},5\%}} - \overline{t_{\text{HF}}^{\text{peak},5\%}}}{\max\left(t_{\text{HF}}^{\text{peak},5\%}\right) - \min\left(t_{\text{HF}}^{\text{peak},5\%}\right)} \tag{9}$$

where A_{LSG} is the inundation extent predicted using the LSG model, A_{HF} is the inundation extent from the high-fidelity simulation, $t_{\text{LSG}}^{\text{peak},5\%}$ and $t_{\text{HF}}^{\text{peak},5\%}$ are the timesteps of the peak period, and "peak, 5%" indicates only the 5% peak values are used. A value close to 0 indicates a good prediction for relRMSE, relPeakValErr, and relPeakTimeErr.

The LSG model's ability to predict the spatial coverage of inundation is assessed using Probability of Detection (POD), Rate of False alarm (RFA), and Critical Success Index (CSI), following Equations 10–12 (Schaefer, 1990):

$$\text{POD} = \frac{A_{\text{detected}}}{A_{\text{detected}} + A_{\text{missed}}} \tag{10}$$

$$\text{RFA} = \frac{A_{\text{false alarm}}}{A_{\text{detected}} + A_{\text{false alarm}}} \tag{11}$$

$$\text{CSI} = \frac{1}{\frac{1}{\text{POD}} + \frac{1}{1-\text{RFA}} - 1} \tag{12}$$

where A_{detected} are those areas correctly detected as being inundated or flooded, A_{missed} are areas simulated to be inundated using the high-fidelity model but are predicted to be dry using the LSG model, and $A_{\text{false alarm}}$ are dry in the high-fidelity model simulation but predicted as being inundated using the LSG model. The POD and RFA evaluate under- and overestimation, respectively. The CSI is a comprehensive metric that combines the POD and RFA metrics to provide an overall evaluation of the model's ability to predict the inundation extent. A POD and CSI of 1 and RFA of 0 indicate a good model performance.

4. Data and Model Application

4.1. Study Site

The study site chosen for the evaluation of the LSG models is the complex and flat Chowilla floodplain (See Figure 4). The Chowilla floodplain is located in the lower part of the Murray-Darling basin that has a total catchment area of approximately 1 million km^2 (Murray-Darling Basin Authority, 2022). The area represented in the model domain is 740 km^2.

The Chowilla floodplain provides a challenging application for the LSG model, as it contains the Murray River and includes several local minor streams, billabongs, and lakes; in addition, flows in the Murray River are impacted by the operation of several weirs and culverts (Murray-Darling Basin Authority, 2021), which help regulate water for irrigation supply and environmental watering (South Australia - Department for Environment and Water, 2022).

All of these features contribute to the complex inundation dynamics of the floodplain, where flood inundation events can last several months due to the large upstream catchment area and shallow gradient of the Murray River.

4.2. Hydrodynamic Flood Inundation Models

4.2.1. High-Fidelity Model

The flood inundation in the Chowilla floodplain is simulated using a high-fidelity 2D hydrodynamic HEC-RAS model (Hydrologic Engineering Center's River Analysis System). HEC-RAS is a freely available flood modeling software developed by the US Army Corps of Engineers and simulates flood inundation using a diffusive wave model on an unstructured grid (US Army Corps of Engineers, 2021a). HEC-RAS uses a subgrid treatment to account for the hydraulic properties of the underlying terrain (Casulli, 2009; US Army Corps of Engineers, 2021b). Subgrid models are also known as porosity models and have been shown to perform well on

coarse grids (Forest, 2020; Sanders & Schubert, 2019), which is advantageous in the development of the coarser low-fidelity model (See Section 4.2.2). To the authors' knowledge, HEC-RAS is currently the only hydrodynamic modeling software that can apply subgrid treatment to an unstructured grid, thus making HEC-RAS particularly useful for exploring how the LSG model performs when simulating flood inundation using an unstructured grid, as described in Section 3.

The high-fidelity HEC-RAS model has three inflow boundaries (Murray River, Station no. 426200; Mullaroo Creek, Station no. 414211; and Lindsay River, Station no. 414212), and one water level outlet downstream at the Murray River Lock 5 upstream (Station no. A4260512). All boundaries rely on historical data retrieved from a publicly available water data platform (Bureau of Meteorology, 2022). The locations of the boundaries are shown in Figure 4.

The high-fidelity model simulates flooding using an unstructured grid with cell sizes varying from 25 m along rivers and structures (weirs and control structures) up to 100 m on the floodplain. The total number of grid cells in the model domain is 109,914 cells, and a total of 796 river cross sections have been incorporated into the model bathymetry, which also includes 22 weirs and control structures. The Manning n's roughness coefficient has been calibrated to 0.026 $s/m^{1/3}$ in the river region and 0.083 $s/m^{1/3}$ on the floodplain. The model was calibrated according to 6 water level stations located across the Chowilla floodplain and Landsat 7 satellite images. The high-fidelity model is run at a fixed 20 s timestep to ensure model stability. Further information on the setup and calibration of the high-fidelity model is given in Text S4 in Supporting Information S1.

4.2.2. Low-Fidelity Model

The low-fidelity model used in the LSG model is obtained by simply reducing the resolution of the high-fidelity model. Fraehr et al. (2022) showed that using a low-fidelity model with over 3 times larger cell sizes in the LSG model setup can provide comparable results to the high-fidelity model. In this study, we test the capabilities of the LSG model further by adopting an even coarser level of discretization. In the low-fidelity model, a grid cell size of 400 m is used along the rivers and on the floodplain, while the 25 m resolution around weirs and structures is preserved. This reduces the number of grid cells to 4,916, which is on average 1/20th of the high-fidelity model resolution.

Note that the only difference between the low- and high-fidelity model is the computational grid. The boundaries and roughness coefficients are not changed. This is the simplest way of developing the low-fidelity model, as no calibration is undertaken to account for the change in spatial resolution. This approach is adopted as we want to examine if the LSG model can upskill results despite having a poorly developed low-fidelity model. Due to the larger grid cells, the low-fidelity model can be run at a steady timestep of 1 min without showing signs of instability.

4.3. Flood Events for Training and Validation

The high- and low-fidelity models are run for a number of flood events to create a training data set, as described in Step 1 of the LSG model (See Section 2.1.1). For the Chowilla floodplain, historic boundary data is available for the period 15 August 2010 to 18 June 2022. In this period, 10 historic flood events have been identified. The duration of the events ranges from 75 to 306 days, with inundated areas ranging between 100 and 450 km^2.

In the inspection and initial simulations of the historic events, it was identified that only 6 of the 10 historic events resulted in significant inundation of the floodplain. For training the LSG model, a large training data set spanning a wide range of inundation behavior is needed. It was therefore decided to create synthetic events by scaling and extending the duration of the minor historic events. This procedure resulted in there being a total of 29 events for training and validation (6 historic and 23 synthetic). Each event is simulated using the high- and low-fidelity models, where flood information is saved every 6 hr. An overview of the duration and inundation extent for all the events is included in Text S3 in Supporting Information S1.

We use cross-validation to evaluate the performance of the LSG models. As mentioned, the data set contains both historic and synthetic flood events. The temporal pattern of each of the synthetic events is similar to the original historic event that was used to create it. This means a random cross-validation procedure cannot be applied for this application as that could result in similar events being included for both training and validation. We have therefore chosen to divide the 29 flood events into 10 groups based on the historic event from which they originate (See Table S2 and Figure S11 in Text S3 in Supporting Information S1). In the cross-validation, we train

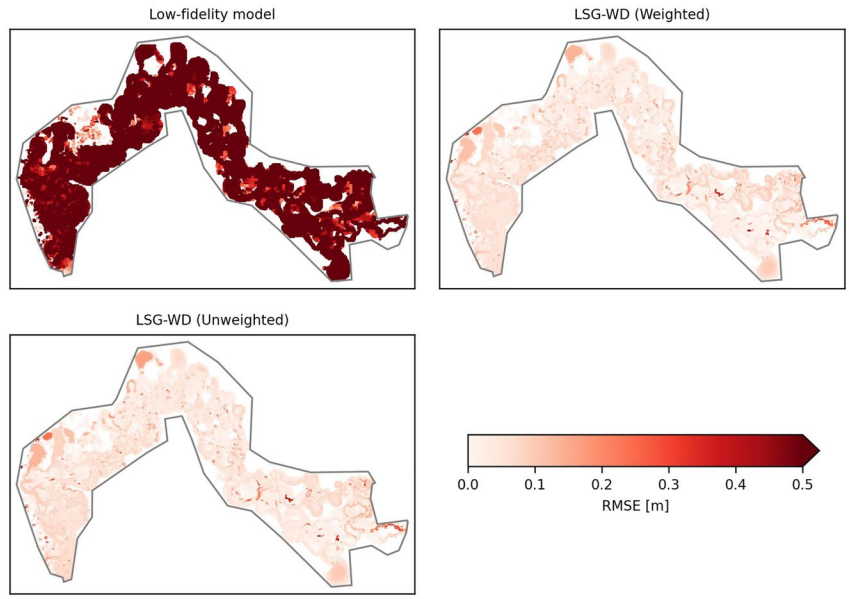

Figure 5. Average Root Mean Square Error for water depth predictions for all 29 simulated events using the low-fidelity, LSG-WD (Weighted) and LSG-WD (Unweighted) models compared to the high-fidelity model simulation.

the LSG models on 9 groups and use the remaining group for validation, resulting in a 10-fold cross-validation. This ensures events originating from the same historic event are not used for validation when they are included in the training data set.

5. Results

In this section, the water depth predictions using the developed LSG-WD (Weighted) and LSG-WD (Unweighted) models are compared to examine the importance of using weighting in the EOF analysis. Subsequently, the accuracy of the inundation extent of the new water depth-based LSG model is compared to the LSG-EXT (Weighted) for direct extent prediction, and finally, additional results of the difference in the EOF analysis and computational efficiency for the LSG models are presented.

5.1. Water Depth

The water depth predictions using the low-fidelity, LSG-WD (Weighted) and LSG-WD (Unweighted) models are compared to the high-fidelity model using RMSE in Figure 5. It is seen that the low-fidelity model on average has significantly higher RMSEs than the two LSG models over all the 29 simulated events. This is expected as the low-fidelity model is based on a considerably coarser grid resolution and is not calibrated. However, this also shows the power of the LSG methodology to significantly reduce errors compared to the low-fidelity model.

Comparing the LSG-WD (Weighted) and LSG-WD (Unweighted) model results does not show any significant differences. For both models, the overall mean RMSE is 4 cm, and the standard deviation is 5 cm. The highest errors are located close to the inflow boundaries in the eastern areas and locally near the model boundaries in the western and north-western parts of the Chowilla floodplain.

The improvement in accuracy of using the LSG models compared to the low-fidelity model is also evident from the boxplots showing the spread of RMSEs for each event in Figure 6. Considering Figure 6b the LSG-WD (Weighted) show a lower median RMSE for events 5b, 7b, 7c, 7f, 8d, 9a, 9c–9f, and 10b. The LSG-WD (Unweighted) show a lower median RMSE for the remaining events, although the difference between the two models is minimal. This shows that neither of the water depth-based LSG models outperforms the other in predicting water depth and suggests that weighting according to the grid sizes is of little importance when using

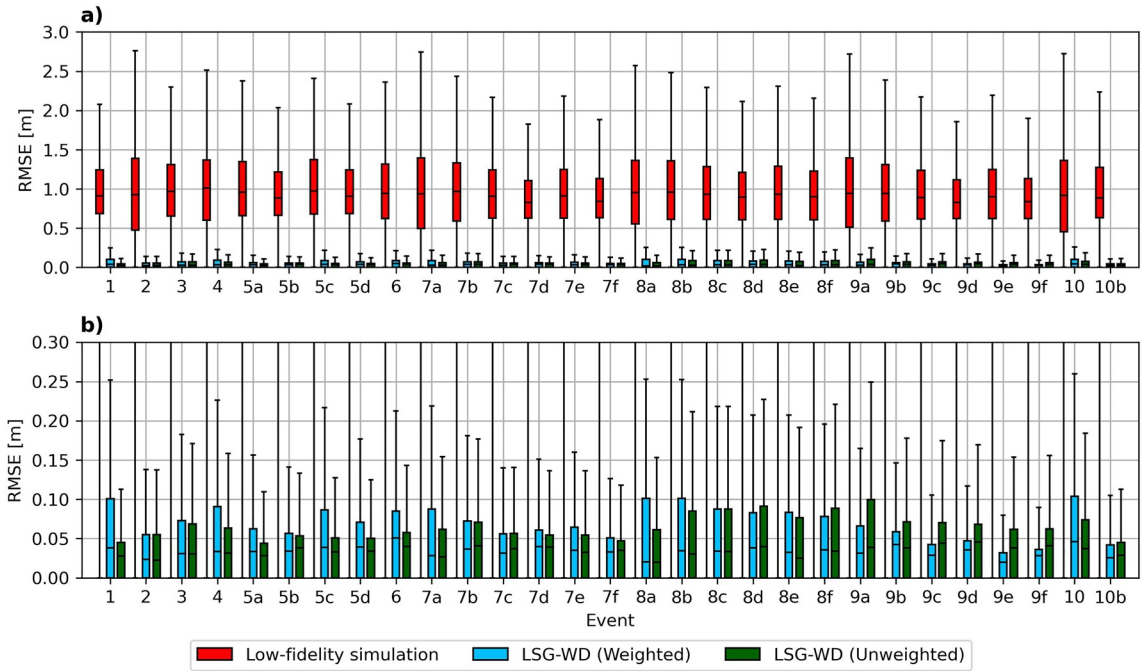

Figure 6. Boxplots of Root Mean Square Error (RMSE) between the low-fidelity, LSG-WD (Weighted) and LSG-WD (Unweighted) models and the simulated water depth using the high-fidelity model. (a) Shows the full range of RMSE, (b) highlights the differences using weighting and no weighting in the Empirical Orthogonal Function analysis. Outliers are not shown. An overview of the events is provided in Text S3 in Supporting Information S1.

the proposed LSG methodology to predict water depth on an unstructured grid. The reason for differences between the models for individual events is most likely due to numerical errors originating from the EOF analysis and Sparse GP model prediction.

The peak water depth is often of high concern in emergency response and flood risk assessments, and a flood inundation model should therefore be able to predict this accurately. The ability of the low-fidelity and LSG models to predict the peak water depth is evaluated by comparing the simulated peak water depth from these models to those using the high-fidelity model as shown in Figure 7. The figure shows the results of the peak water depth in all 109,914 cells for all 29 simulated events (as a density map). A total of 3,187,506 data points are compared for each model.

The low-fidelity model consistently overpredicts the water depth, both for shallow and deeper depths, with a large spread in the predicted values. On the other hand, the LSG models show good agreement with the high-fidelity model, illustrated by a coefficient of determination approximately equal to 1. The prediction errors are heteroscedastic, generally showing a narrower spread for large water depths and wider for shallower water depths.

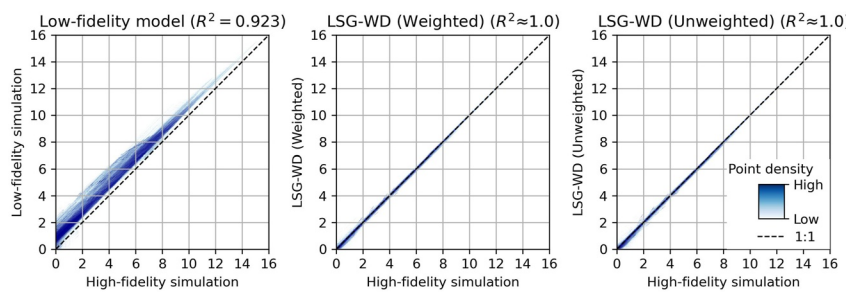

Figure 7. Peak water depth in each grid cell in the model domain predicted using the low-fidelity, LSG-WD (Weighted) and LSG-WD (Unweighted) models compared to the high-fidelity model simulation for all 29 simulated events. The density map shows light and dark blue colors to indicate low and high data point density, respectively.

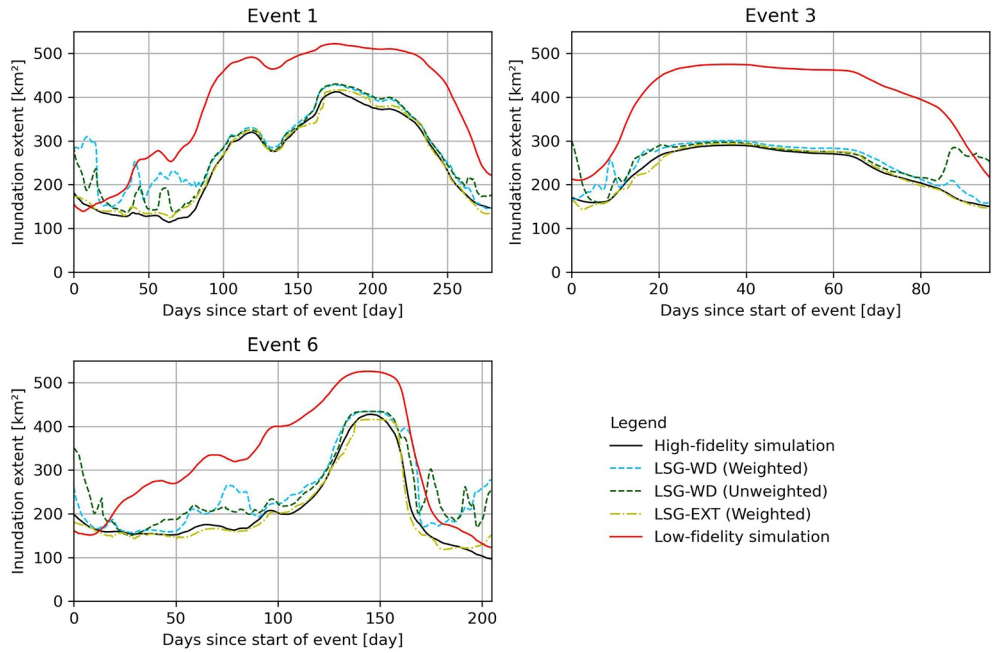

Figure 8. Inundation extent using the high-fidelity, low-fidelity, LSG-WD (Weighted), LSG-WD (Unweighted), and LSG-EXT (Weighted) models for three representative events 1, 3, and 6. The predicted inundation extents for the remaining events are included in Text S5 in Supporting Information S1.

This outcome is due to the flat topology of the Chowilla floodplain, which results in shallow inundation depths that vary over a narrow range over most of the areas. There are no distinctive differences between the LSG-WD (Weighted) and LSG-WD (Unweighted) models, as both models show a good ability to predict the peak water depth.

5.2. Inundation Extent

The inundation extents simulated using the high-fidelity model and the LSG models are shown in Figure 8 for three representative events. Figures showing the inundation extent for the remaining events are provided in Text S5 in Supporting Information S1. The inundation extent is found for the LSG-WD (Weighted) and LSG-WD (Unweighted) models by adopting a threshold of 3 cm water depth to differentiate the cells into flooded and dry. This follows the binary procedure adopted for mapping the flood extent in the LSG-EXT (Weighted) model (See Step 2 in Section 2.3.1). HEC-RAS has the ability to simulate partially flooded cells. However, in this study, it was decided to simply use binary values to identify wet and dry cells, to represent the results that would be obtained by other common hydrodynamic modeling software, such as MIKE21 (DHI, 2022) and TUFLOW (BMT, 2020) that do not have the capability for representing partially flooded cells.

The LSG models improve predictions significantly compared to using a low-fidelity model. The low-fidelity model overpredicts the inundation extent compared to the high-fidelity model, which is consistent with the degree of overprediction of water depths shown in Figure 7. This was expected of the low-fidelity model, as coarser grids tend to exhibit larger dispersion of the flood inundation extents (Chatterjee et al., 2008; Yu & Lane, 2006).

For inundation extents below approximately 300 km^2, the predicted inundation extent spuriously fluctuates for both the LSG-WD (Weighted) and LSG-WD (Unweighted) models, resulting in uncertain predictions. This is due to the threshold applied for converting the data to binary values. At low water depths, the entire cell can quickly change between flooded and dry, and this means that large areas can suddenly transition from a dry to a flooded state and vice versa. The predictions of the LSG-EXT (Weighted) model are less

Table 1
Flood Extent Evaluation Metrics for All 29 Simulated Events

Metric	Low-fidelity model	LSG-WD (Weighted)	LSG-WD (Unweighted)	LSG-EXT (Weighted)
relRMSE	0.69 (0.14)	0.19 (0.19)	0.18 (0.14)	**0.05 (0.02)**
relPeakValErr	0.48 (0.25)	0.10 (0.19)	0.10 (0.16)	**−0.02 (0.03)**
relPeakTimeErr	**0.88 (2.79)**	1.02 (3.25)	0.90 (3.56)	0.91 (1.75)
POD	**0.99 (0.00)**	0.99 (0.01)	0.99 (0.01)	0.98 (0.01)
RFA	0.17 (0.06)	0.06 (0.06)	0.06 (0.06)	**0.01 (0.00)**
CSI	0.82 (0.06)	0.94 (0.06)	0.93 (0.06)	**0.97 (0.01)**

Note. Results are shown as mean values over all events with standard deviations shown in parentheses. relRMSE, relPeakValErr, and relPeakTimeErr are based on the temporal evolution of the inundation extent as seen in Figure 8. POD, RFA, and CSI are spatial metrics based on the maximum inundation extent shown in Text S5 and Figures S31–S35 in Supporting Information S1. The model with the best performance for each metric is shown in bold.

variable as the distinction between flooded and dry areas is already incorporated in the setup and training of the model.

The relRMSE, relPeakValErr, and relPeakTimeErr evaluation metrics are displayed in Table 1. With reference to the relRMSE metric, the LSG-EXT (Weighted) model performs the best and has errors consistent with the values reported in Fraehr et al. (2022). Of the LSG-WD (Weighted) and (Unweighted) models, the performance is similar.

With reference to the relPeakValErr metric, the LSG-EXT (Weighted) model still outperforms the water depth-based models. Another interesting observation is that the LSG-WD (Weighted) and (Unweighted) models generally overpredict the peaks, whereas the LSG-EXT model underpredicts the peaks.

For the timing of the peaks, all the models provide predictions that are generally late compared to the high-fidelity model. The low-fidelity model shows the lowest mean timing error, although the LSG-EXT (Weighted) perform almost equally well to the low-fidelity model and has a lower standard deviation. All the models show a relatively high standard deviation for the relPeakTimeErr compared to the other metrics, thus indicating a large uncertainty. This is due to the flat topography of the Chowilla floodplain and the low gradient of the Murray River, resulting in flood events with long-lasting flat attenuated peaks, where minor uncertainties in the predictions can have a large influence on the exact timing of the peak. This is also evident in Figure 8, where the temporal evolution of the inundation extents of all the LSG models follows the high-fidelity model well in the vicinity of the peak. In addition, it might be expected that using an increasingly coarser low-fidelity model will result in the LSG models predicting consistently early peak timings due to the larger dispersion of the flood extent. However, this is not evident from the results, even though this study uses a low-fidelity model that is much coarser than that used in Fraehr et al. (2022).

The POD, RFA, and CSI metrics measure the LSG models' ability to predict the spatial coverage of the maximum inundation extent. In Table 1, the LSG-EXT (Weighted) model has a lower POD and RFA than the LSG-WD (Weighted) and (Unweighted) models. This is due to the LSG-EXT model's general underprediction of the flood extent. The LSG-WD (Weighted) and (Unweighted) models show similar performance and generally overpredict the inundation extent. This means they have a high probability of detecting a flooded area, but also a higher rate of false alarms.

The maximum inundation extent for the LSG models is shown in Figure 9 for three representative events. The results for the low-fidelity model are not shown due to the large overprediction of the low-fidelity model. The LSG-WD (Weighted) and (Unweighted) models both over- and underpredict the inundation extent, as indicated by the misses and false alarms. The largest difference between the models is evident for the (smaller) event 3, where the LSG-WD (Unweighted) model significantly overpredicts the inundation and the LSG-EXT (Weighted) model performs well with only minor areas of misses and false alarms. The LSG-WD (Weighted) model achieves a performance that lies between the two other models for event 3.

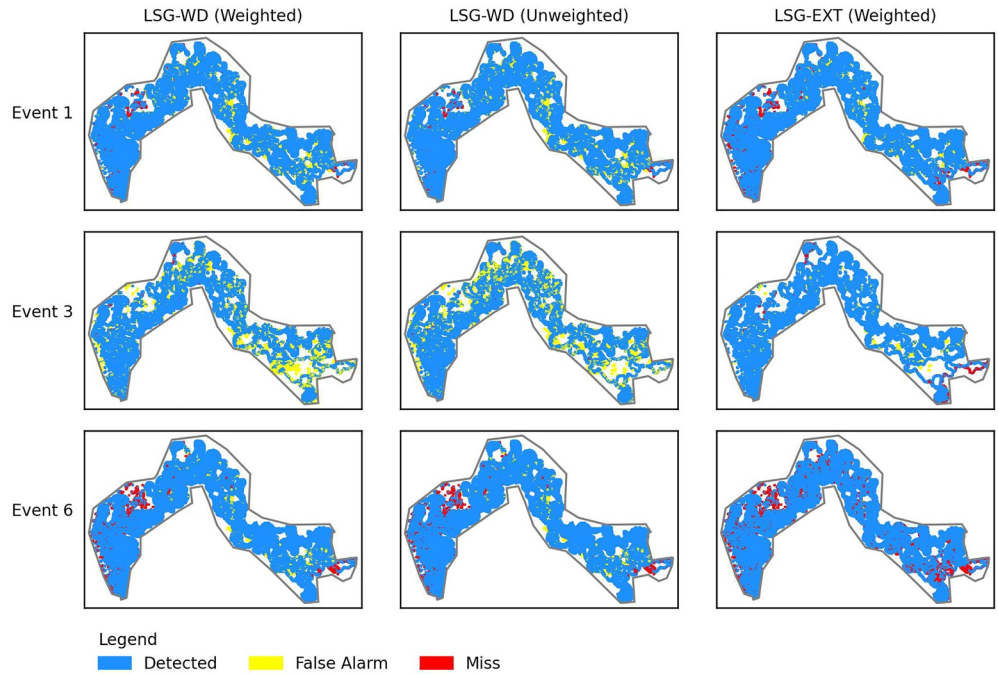

Figure 9. Detected, Misses, and False alarms for the LSG-WD (Weighted), LSG-WD (Unweighted), and LSG-EXT (Weighted) models for three representative events 1, 3, and 6. The detection results for the remaining events are included in Text S5 in Supporting Information S1.

5.3. Additional Results

5.3.1. Comparison of EOF Analysis for Each LSG Model

A different number of significant modes have been found in the EOF analyses undertaken for each of the three LSG models (LSG-WD (Weighted), LSG-WD (Unweighted) and LSG-EXT (Weighted)), as shown in Table 2. Individual EOF analyses have been performed for the training data used in each fold of the cross-validation (See Section 4.3), resulting in a total of 10 EOF analyses for each LSG model. The number of significant modes in each EOF analysis is found through North's test and Kaiser's rule, as described in Section 2.1.3.

The LSG-WD (Unweighted) tends to have the least number of significant modes. This is noteworthy as a model with fewer modes implies a lower degree of dimensionality to explain the majority of variance in the data set, and this has the benefit of requiring fewer features to be predicted using the Sparse GP model. Furthermore, the proportion of variance explained when performing EOF analysis on binary values in the LSG-EXT (Weighted) model is lower than the water-based LSG models, even though a similar number of significant modes are found. The reason for this is found in the methodology of the EOF analysis. The EOF analysis seeks to find a linear combination of ECs and EOFs to maximize the variance. Linear combinations of continuous values reconstruct poorly when predicting binary data, and therefore more modes are needed to explain the variance. Another way to think about this is in terms of variance between the cells. In the binary data set, one cell might be dry and another flooded, and therefore the values switch between 0 and 1. However, in the water depth-based data sets, the same two cells might have a water depth of 0.00 and 0.05 m, respectively, and thus the numerical difference is smaller when using water depths. This also explains why the LSG-WD (Weighted) model needs more modes than the LSG-WD (Unweighted) model, as some of the cells might have a higher weight, and the water depth thereby is multiplied by a large value, creating a big difference in values between the cells.

Table 2
Significant Modes and Explained Variance From the Empirical Orthogonal Function Analyses for Each Low-Fidelity, Spatial Analysis, and Gaussian Process Learning Model

LSG model	Number of significant modes	Variance explained
LSG-WD (Weighted)	42 (18)	99.7 (0.2)%
LSG-WD (Unweighted)	32 (9)	99.7 (0.3)%
LSG-EXT (Weighted)	37 (12)	90.6 (1.8)%

Note. Results are shown as means with the standard variation shown in parentheses of the EOF analyses performed for the 10-fold cross-validation.

Table 3
Computational Time for Flood Inundation Prediction of Event 3

	EOF analysis and sparse GP training	Prediction
High-fidelity	–	10 hr 43 min 34 s
Low-fidelity	–	54 min 49 s
LSG-WD (weighted)	17 min 29 s	54 min 54 s
LSG-WD (unweighted)	9 min 0 s	54 min 52 s
LSG-EXT (weighted)	5 min 45 s	54 min 51 s

5.3.2. Computational Efficiency

The simulations using the low- and high-fidelity models, as well as the EOF analyses and the model training and prediction steps with the Sparse GP models, have all been undertaken on a high-performance computer with a 3.70 GHz processor with Intel® Xeon® E-2288G CPU, 64 GB ram, 64 cores and a NVIDIA Quadro RTX 5000 graphic card. The simulations in HEC-RAS were undertaken using the "All available cores" option, which was found to be the most computationally efficient setting in the initial testing.

The computational times using the different models have been summarized in Table 3 for event 3. The tendency is similar for the other events. The advantage of using the LSG methodology over the high-fidelity model is clear, as the computational time is approximately 12 times faster. The time for training and prediction using the three versions of the LSG model varies. This is due to a different number of Sparse GP models being trained, because of a different number of significant modes found in the EOF analysis (See Section 5.3). When using event 3 for validation, a total of 75, 37, and 23 significant modes were found in the EOF analysis for the LSG-WD (Weighted), LSG-WD (Unweighted), and LSG-EXT (Weighted) models, respectively. If the LSG models all had an equal number of significant modes and thereby Sparse GP models, the training and prediction times would be similar. The difference in prediction time between the low-fidelity model and the LSG models is the prediction time of the Sparse GP models and the subsequent reconstruction of the inundation data set. This time is minimal compared to the low-fidelity simulation time. As the low-fidelity model needs to be run every time a prediction is required using the LSG methodology, it is worth further exploring the possibilities of further enhancing the efficiency of the low-fidelity model.

Another aspect when considering using the LSG model is the time used for the creation of the training data set in Step 1 of the model setup (see Section 2.1.1). In the training data set development, numerous simulations of both the low- and high-fidelity models are needed. In this study, a total of 29 events were simulated. This required approximately 24 computational days for the high-fidelity model and represents a large computational burden that needs to be overcome before the LSG model can be implemented. However, the training data set generation only needs to be undertaken once before being used to provide predictions, whereafter the full speed of the LSG model can be utilized.

6. Discussion

This study demonstrates that the LSG model is a powerful tool to upskill low-fidelity model simulations to emulate the results of a fully 2D hydrodynamic high-fidelity model on an unstructured grid. In this section, we discuss several points, including the importance of weighting according to grid cell size, the best method for predicting flood extent, and the future directions for the LSG model.

6.1. Importance of Weighting in EOF Analysis

Two water depth-based versions of the LSG model are explored in this study, one with weighting according to cell size before the EOF analysis (i.e., the LSG-WD (Weighted) model) and one without weighting (i.e., the LSG-WD (Unweighted) model). The purpose of the development of these models was to examine the value of using weights to compensate for the varying grid cell sizes in an unstructured grid. It was found that the water depth predictions for the two models are similar, and this suggests that weighting is of minor importance when applying the LSG model to simulate flood inundation on an unstructured grid. The predicted inundation extent using the two models is also similar. For the evolution of the total inundation extent the LSG-WD (Unweighted) model shows slightly better performance than LSG-WD (Weighted) as seen by the relRMSE, relPeakValErr, and relPeakTimeErr being closer to 0. However, the relRMSE, relPeakValErr, and relPeakTimeErr consider the summarized area and not the spatial location of the inundation. Correctly capturing flooded areas is of high importance for flood risk assessments. The LSG-WD (Weighted) model has a marginally higher CSI, and this suggests that the weighting helps the LSG model to capture the spatial extent of the inundation more accurately. However, the overall difference between the LSG-WD (Weighted) and (Unweighted) models is negligible.

The above outcome is not consistent with expectations as the EOF analysis is a measure of variability and treats all cells equally, regardless of their spacing and size. The weighting scheme provides a means to counteract the uniform influence of the individual grid cells and creates what is known as "intrinsic EOFs" (Baldwin et al., 2009; North et al., 1982), which are continuous spatial fields that are independent of the grid. The lack of improvement obtained by creating these more generalized EOF spatial fields is most likely due to the ability of the Sparse GP models to compensate for the differences in the EOF analyses and ensure a good conversion of the low-fidelity ECs to high-fidelity ECs, regardless of the weighting scheme. It is possible that using a simpler model to convert the low-fidelity ECs to high-fidelity ECs might have yielded a greater difference in the results.

Both water depth-based LSG models show spurious fluctuations in the predicted inundation extent for areas below 300 km^2. The inundation extent for the LSG-WD (Weighted) and LSG-WD (Unweighted) adopts a 3 cm water depth threshold to convert the results to binary values. This threshold was chosen to make the method and results comparable to the previous study by Fraehr et al. (2022) where a MIKE21 model was used. However, as mentioned in Section 5.2, HEC-RAS can accommodate partially flooded cells due to a subgrid treatment that accounts for the terrain variations within a cell. This capability has been tested to convert the water depths to a partially flooded cell area (see Text S2 in Supporting Information S1). The use of this option improves predictions significantly and remediates the spurious fluctuations evident in Figure 8 for the predicted inundation extents below 300 km^2 for both the LSG-WD (Weighted) and LSG-WD (Unweighted) models. However, using partially flooded cells is only possible when using hydrodynamic models that have subgrid solvers, such as HEC-RAS, and for that reason, it is not used for comparisons in this study.

Weighting according to grid cell size is simple, easy to implement and commonly used (Baldwin et al., 2009; Hannachi et al., 2007), though the results of this study show that applying this weighting scheme has minimal effect on the accuracy of the LSG model. Due to the simplistic nature of the weighting scheme applied, it is worth considering if a more sophisticated weighting scheme could increase the accuracy of the LSG model. However, as the accuracy of the LSG model is already high, it is likely that only minor improvements could be achieved; given the expected low return on effort (at least with this model configuration), it was decided to forego investing effort in developing a more sophisticated weighting scheme.

6.2. Flood Extent Derived From Water Depth Predictions Compared to a Direct Extent Prediction

The LSG-EXT (Weighted) model for direct extent prediction proposed by Fraehr et al. (2022) is significantly better for predicting the inundation extent than the water depth-based LSG model, as shown by the evaluation metrics in Table 1. As discussed in the previous section, the water depth-based LSG models adopt a 3 cm threshold to differentiate between flooded and dry areas and minor numerical differences may determine whether a cell is flooded or not. If the water depth in a cell is 4 cm, the water depth-based LSG models are only allowed a numerical error of 0.01 before the cell is predicted as dry. On the other hand, the LSG-EXT (Weighted) model predicts values between 0 and 1, with a threshold of 0.5 for flooding and drying. Thus, the LSG-EXT (Weighted) model accommodates larger numerical errors without it affecting the predicted inundation extent.

However, information on water depth is highly beneficial for risk assessments and can greatly assist in the identification of flood hazards. Flood inundation predictions should therefore be carried out using both an extent- and water depth-based LSG model. The extent-based LSG model should be used to predict the inundation extent, and the water depth-based LSG model should then be used to predict the water depth for those areas predicted as being flooded. The accuracy of the inundation extent estimates would be similar to the performance of the LSG-EXT (Weighted) model, and the water depth predictions would be similar to the LSG-WD (Weighted) model. Accordingly, these results are not shown here, but this approach is recommended for future implementations of the LSG model.

6.3. Future Directions for the LSG Model

The low-fidelity model used in this study is approximately 12 times faster than running a high-fidelity model. This is a substantial improvement in computational efficiency. However, in ensemble modeling used for risk assessment hundreds or thousands of model runs are needed (Nayak et al., 2018; Nester et al., 2012; Wu et al., 2020). Ensemble modeling would help uncover if the errors in the LSG model predictions are lower than the uncertainty

associated with the input boundaries. This would improve the confidence in using the LSG model as the LSG model would not be the biggest source of uncertainty in the inundation predictions.

Ensemble modeling imposes a high computational demand on the low-fidelity model, and therefore, further research into optimizing the efficiency of the low-fidelity model is needed. Options for improving the computational efficiency include simplifying the geometry by using still coarser grid cells, increasing the timestep, reducing model complexity by adopting simplifying assumptions, or using a more computationally efficient model or software (Razavi et al., 2012). HEC-RAS is currently the only hydrodynamic modeling software that utilizes subgrid treatment on an unstructured grid and for that reason was chosen in this study. As an example of other models to use, the LISFLOOD-FP model proposed by Bates and De Roo (2000) has been shown to be 20 times faster than HEC-RAS when running both models on the same grid and timestep (Shustikova et al., 2019). This suggests that LISFLOOD-FP may provide fast low-fidelity flood inundation estimates, although LISFLOOD-FP predicts flood inundation on a quadratic grid and does not have the same subgrid treatment as included in HEC-RAS, which might reduce its accuracy. Future studies should explore using other modeling software to examine the accuracy and computational efficiency of the LSG model across a variety of hydrodynamic modeling platforms.

Besides the computational efficiency of the LSG model, the low-fidelity model also affects the accuracy. Minimal attention has been given to the accuracy of the low-fidelity model in the development of the LSG model in both this study and the previous study by Fraehr et al. (2022). Considering the RMSE of the LSG models in Figure 5, the highest errors are located in a few local areas. This suggests that performance in these areas might be improved by locally improving the low-fidelity model. Another consideration is to calibrate the low-fidelity model using observations and/or the results of the high-fidelity model as this to some degree can counteract the dispersion of the flood inundation due to the coarser grid (Yu & Lane, 2006). In locations with available and regularly updated observations, it may be expected that the use of data assimilation could improve the accuracy of the LSG model (Abbaszadeh et al., 2022; Jafarzadegan et al., 2021). Only a few studies have explored the use of data assimilation in combination with hydrodynamic models. For example, Xu et al. (2017) used data assimilation together with a 1D river model, and both Jafarzadegan et al. (2021) and Muñoz et al. (2022) used data assimilation together with a 2D flood inundation model. Accordingly, incorporating data assimilation in the LSG model could provide accurate predictions in high resolution for future real-time forecasting applications. Future studies should therefore focus on improving the precision of the LSG model, as well as on increasing computational efficiency.

Data-driven models like the Sparse GP model are particularly good at describing complex non-linear relationships, such as those between the low- and high-fidelity ECs. In the initial tests, other data-driven models like the Multilayer Perceptron have been tested and shown to have similar performance to the Sparse GP model, although the Multilayer Perceptron did not provide an uncertainty estimate. Similarly, Carreau and Guinot (2021) used a simple Artificial Neural Network structure in their study to describe the relationship between ECs. The LSG model is therefore not limited to using the Sparse GP model, and other data-driven could be implemented.

The LSG model needs to be tested on other types of flooding behavior to ensure that the prediction accuracy of the LSG model is robust. This could include consideration of storm surge flooding in an estuary, urban flooding, and compound floods resulting from exogenous influences. Future applications will further examine the capabilities of the LSG model and help ensure it is a robust surrogate model for flood inundation.

7. Conclusion

Traditionally flood inundation is predicted using high-fidelity models that are accurate, but computationally expensive to apply. In a previous study, the hybrid LSG model has been proposed to predict the dynamic behavior of the flood inundation extent in a more computationally efficient way than the traditional high-fidelity models. This study shows how the LSG model can be further developed to predict the depths, as well as the spatial extent, of flood inundation.

The LSG model is evaluated by simulating flood inundation of the Chowilla floodplain using a HEC-RAS model with an unstructured grid. To compare the LSG model for water depth prediction to the previous LSG model developed for direct flood extent predictions, both models were used to simulate the flood inundation of the

Chowilla floodplain. The extent-based model exhibits significantly better separation of dry and flooded areas. However, including water depth in the inundation predictions has considerable potential to improve flood risk assessments. For that reason, it is recommended to use both extent- and water depth-based LSG models. This would ensure a high accuracy of both inundation extent and the water depth, thus making the LSG model a valuable tool to predict key flood hazard indicators and inform flood risk assessments.

When applying the LSG model as a surrogate for a high-fidelity model with an unstructured grid, the influence of different grid cell sizes can be accommodated by the adoption of a weighting scheme based on the cell area in the EOF analysis. To explore the importance of this weighting, two LSG models with and without weighting were developed. The results indicate that the weighting has minimal influence on both the water depth and flood extent predictions. This result highlights the robustness of the LSG model as it is able to compensate for variations in the model setup and provide accurate flood inundation predictions on an unstructured grid.

The LSG model is approximately 12 times faster than using a high-fidelity model and provides accurate predictions of the flood inundation depth and extent. In comparison to the high-fidelity model, the LSG model has a RMSE with a mean of 4 cm and a standard deviation of 5 cm for the water depth predictions. The larger errors are concentrated in local areas and could potentially be resolved by locally improving and/or calibrating the low-fidelity model.

Future studies of the LSG model should focus on the low-fidelity model development. The low-fidelity model is the most computationally demanding part of the LSG structure, and it has a great influence on prediction accuracy. Optimizing the low-fidelity model can therefore significantly influence the performance of the LSG model. In addition, it would be of interest to test the LSG model on a wide range of flood problems to evaluate the benefits of the approach in more detail.

Data Availability Statement

The LSG model is coded using Python (Version 3.9) and is available in Fraehr (2023) together with the data used to produce the results in this paper.

Acknowledgments
Niels Fraehr acknowledges support from The University of Melbourne via the Melbourne Research Scholarship and Wenyan Wu acknowledges support from the Australian Research Council via the Discovery Early Career Researcher Award (DE210100117). Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

References

Abbaszadeh, P., Muñoz, D. F., Moftakhari, H., Jafarzadegan, K., & Moradkhani, H. (2022). Perspective on uncertainty quantification and reduction in compound flood modeling and forecasting. *iScience*, *25*(10), 105201. https://doi.org/10.1016/j.isci.2022.105201

Altenau, E. H., Pavelsky, T. M., Bates, P. D., & Neal, J. C. (2017). The effects of spatial resolution and dimensionality on modeling regional-scale hydraulics in a multichannel river. *Water Resources Research*, *53*(2), 1683–1701. https://doi.org/10.1002/2016wr019396

Antony, R., Rahiman, K. U. A., & Vishnudas, S. (2021). Flood hazard assessment and flood inundation mapping—A review. In *Current trends in Civil Engineering* (pp. 209–218). https://doi.org/10.1007/978-981-15-8151-9_20

Asher, M. J., Croke, B. F. W., Jakeman, A. J., & Peeters, L. J. M. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, *51*(8), 5957–5973. https://doi.org/10.1002/2015wr016967

Australian Institute for Disaster Resilience. (2012). Queensland and Brisbane 2010/11 floods. Retrieved from https://knowledge.aidr.org.au/resources/flood-queensland-2010-2011/

Baldwin, M., Stephenson, D., & Jolliffe, I. (2009). Spatial weighting and iterative projection methods for EOFs. *Journal of Climate*, *22*(2), 234–243. https://doi.org/10.1175/2008JCLI2147.1

Bates, P. D. (2022). Flood inundation prediction. *Annual Review of Fluid Mechanics*, *54*(1), 287–315. https://doi.org/10.1146/annurev-fluid-030121-113138

Bates, P. D., & De Roo, A. P. J. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, *236*(1–2), 54–77. https://doi.org/10.1016/s0022-1694(00)00278-x

Bauer, M., Wilk, M. V. D., & Rasmussen, C. E. (2016). Understanding probabilistic sparse Gaussian process approximations. In *Proceedings of the 30th international conference on neural information processing systems*. https://doi.org/10.48550/arXiv.1606.04820

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrology and Earth System Sciences*, *26*(16), 4345–4378. https://doi.org/10.5194/hess-26-4345-2022

BMT. (2020). TUFLOW FV user manual - Build 2020.02. Retrieved from https://downloads.tuflow.com/_archive/TUFLOW_FV/Manual/TUFLOW_FV_User_Manual_2020.pdf

Bomers, A., Schielen, R. M. J., & Hulscher, S. (2019). Application of a lower-fidelity surrogate hydraulic model for historic flood reconstruction. *Environmental Modelling & Software*, *117*, 223–236. https://doi.org/10.1016/j.envsoft.2019.03.019

Bureau of Meteorology. (2022). Water data online. Retrieved from http://www.bom.gov.au/waterdata/

Burt, D., Rasmussen, C. E., & Wilk, M. V. D. (2019). Rates of convergence for sparse variational Gaussian process regression. In *Proceedings of the 36th international conference on machine learning*. https://doi.org/10.48550/arXiv.1903.03571

Carreau, J., & Guinot, V. (2021). A PCA spatial pattern based artificial neural network downscaling model for urban flood hazard assessment. *Advances in Water Resources*, *147*, 103821. https://doi.org/10.1016/j.advwatres.2020.103821

Casulli, V. (2009). A high-resolution wetting and drying algorithm for free-surface hydrodynamics. *International Journal for Numerical Methods in Fluids*, *60*(4), 391–408. https://doi.org/10.1002/fld.1896

Chang, L. C., Chang, F. J., Yang, S. N., Kao, I. F., Ku, Y. Y., Kuo, C. L., & Amin, I. (2019). Building an intelligent hydroinformatics integration platform for regional flood inundation warning systems. *Water*, *11*(1), 9. https://doi.org/10.3390/w11010009

Chatterjee, C., Förster, S., & Bronstert, A. (2008). Comparison of hydrodynamic models of different complexities to model floods with emergency storage areas. *Hydrological Processes*, *22*(24), 4695–4709. https://doi.org/10.1002/hyp.7079

Chu, H. B., Wu, W. Y., Wang, Q. J., Nathan, R., & Wei, J. H. (2020). An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environmental Modelling & Software*, *124*, 104587. https://doi.org/10.1016/j.envsoft.2019.104587

Contreras, M. T., Gironas, J., & Escauriaza, C. (2020). Forecasting flood hazards in real time: A surrogate model for hydrometeorological events in an Andean watershed. *Natural Hazards and Earth System Sciences*, *20*(12), 3261–3277. https://doi.org/10.5194/nhess-20-3261-2020

DHI. (2022). MIKE 21 flow model - Hydrodynamic module - User guide. Retrieved from https://manuals.mikepoweredbydhi.help/latest/Coast_and_Sea/M21HD.pdf

ESRI. (2022). World imagery. Retrieved from https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9

Forest, M. (2020). *HEC-RAS subgrid bathymetry theory and application*. Kleinschmidt. Retrieved from https://www.kleinschmidtgroup.com/ras-post/hec-ras-subgrid-bathymetry-theory-and-application/

Fraehr, N. (2023). Data from HEC-RAS models for training and validation in "Development of a fast and accurate hybrid model for floodplain inundation simulations" (Version 2) [Dataset]. The University of Melbourne. https://doi.org/10.26188/21235782

Fraehr, N., Wang, Q. J., Wu, W., & Nathan, R. (2022). Upskilling low-fidelity hydrodynamic models of flood inundation through spatial analysis and Gaussian Process learning. *Water Resources Research*, *58*(8), e2022WR032248. https://doi.org/10.1029/2022WR032248

Goldbaum, C., & ur-Rehman, Z. (2022). In *Pakistan's record floods, villages are now desperate Islands*. The New York Times. Retrieved from https://www.nytimes.com/2022/09/14/world/asia/pakistan-floods.html

Hannachi, A., Jolliffe, I. T., & Stephenson, D. B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, *27*(9), 1119–1152. https://doi.org/10.1002/joc.1499

Hunter, N. M., Bates, P. D., Horritt, M. S., & Wilson, M. D. (2007). Simple spatially-distributed models for predicting flood inundation: A review. *Geomorphology*, *90*(3–4), 208–225. https://doi.org/10.1016/j.geomorph.2006.10.021

IPCC. (2021). *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment Report of the intergovernmental panel on climate change*. C. U. Press.

Jafarzadegan, K., Abbaszadeh, P., & Moradkhani, H. (2021). Sequential data assimilation for real-time probabilistic flood inundation mapping. *Hydrology and Earth System Sciences*, *25*(9), 4995–5011. https://doi.org/10.5194/hess-25-4995-2021

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kabir, S., Patidar, S., & Pender, G. (2021). A machine learning approach for forecasting and visualising flood inundation information. *Proceedings of the Institution of Civil Engineers-Water Management*, *174*(1), 27–41. https://doi.org/10.1680/jwama.20.00002

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kirezci, E., Young, I. R., Ranasinghe, R., Muis, S., Nicholls, R. J., Lincke, D., & Hinkel, J. (2020). Projections of global-scale extreme sea levels and resulting episodic coastal flooding over the 21st century. *Scientific Reports*, *10*(1), 11629. https://doi.org/10.1038/s41598-020-67736-6

Leibfried, F., Dutordoir, V., John, S. T., & Durrande, N. (2021). A tutorial on Sparse Gaussian Processes and variational inference. arXiv pre-print server.

Lhomme, J., Sayers, P., Gouldby, B., Wills, M., & Mulet-Marti, J. (2008). *Recent development and application of a rapid flood spreading method* (pp. 15–24). CRC Press. https://doi.org/10.1201/9780203883020.ch2

Liu, Z., Merwade, V., & Jafarzadegan, K. (2019). Investigating the role of model structure and surface roughness in generating flood inundation extents using one- and two-dimensional hydraulic models. *Journal of Flood Risk Management*, *12*(1), e12347. https://doi.org/10.1111/jfr3.12347

Ma, P., Konomi, G. K. B. A., Asher, T. G., Toro, G. R., & Cox, A. T. (2019). Multifidelity computer model emulation with high-dimensional output: An application to storm surge. arXiv. https://doi.org/10.48550/ARXIV.1909.01836

McGrath, H., Bourgon, J.-F., Proulx-Bourque, J.-S., Nastev, M., & Abo El Ezz, A. (2018). A comparison of simplified conceptual models for rapid web-based flood inundation mapping. *Natural Hazards*, *93*(2), 905–920. https://doi.org/10.1007/s11069-018-3331-y

Ming, X., Liang, Q., Xia, X., Li, D., & Fowler, H. J. (2020). Real-time flood forecasting based on a high-performance 2-D hydrodynamic model and numerical weather predictions. *Water Resources Research*, *56*(7), e2019WR025583. https://doi.org/10.1029/2019wr025583

Morales-Hernández, M., Sharif, M. B., Kalyanapu, A., Ghafoor, S. K., Dullo, T. T., Gangrade, S., et al. (2021). TRITON: A multi-GPU open source 2D hydrodynamic flood model. *Environmental Modelling & Software*, *141*, 105034. https://doi.org/10.1016/j.envsoft.2021.105034

Muñoz, D. F., Abbaszadeh, P., Moftakhari, H., & Moradkhani, H. (2022). Accounting for uncertainties in compound flood hazard assessment: The value of data assimilation. *Coastal Engineering*, *171*, 104057. https://doi.org/10.1016/j.coastaleng.2021.104057

Murray-Darling Basin Authority. (2021). Chowilla floodplain report card 2019–20. Retrieved from https://www.mdba.gov.au/issues-murray-darling-basin/water-for-environment/chowilla-floodplain-report-card

Murray-Darling Basin Authority. (2022). Where is the Murray–Darling basin. Retrieved from https://www.mdba.gov.au/importance-murray-darling-basin/where-basin

Nayak, M. A., Herman, J. D., & Steinschneider, S. (2018). Balancing flood risk and water supply in California: Policy search integrating short-term forecast ensembles with conjunctive use. *Water Resources Research*, *54*(10), 7557–7576. https://doi.org/10.1029/2018WR023177

Neal, J., Fewtrell, T., & Trigg, M. (2009). Parallelisation of storage cell flood models using OpenMP. *Environmental Modelling & Software*, *24*(7), 872–877. https://doi.org/10.1016/j.envsoft.2008.12.004

Nester, T., Komma, J., Viglione, A., & Blöschl, G. (2012). Flood forecast errors and ensemble spread—A case study. *Water Resources Research*, *48*(10), W10502. https://doi.org/10.1029/2011wr011649

Nobre, A. D., Cuartas, L. A., Momo, M. R., Severo, D. L., Pinheiro, A., & Nobre, C. A. (2016). HAND contour: A new proxy predictor of inundation extent. *Hydrological Processes*, *30*(2), 320–333. https://doi.org/10.1002/hyp.10581

North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J. (1982). Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, *110*(7), 699–706. https://doi.org/10.1175/1520-0493(1982)110<0699:seiteo>2.0.co;2

Parker, K., Ruggiero, P., Serafin, K. A., & Hill, D. F. (2019). Emulation as an approach for rapid estuarine modeling. *Coastal Engineering*, *150*, 79–93. https://doi.org/10.1016/j.coastaleng.2019.03.004

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*(7), W07401. https://doi.org/10.1029/2011WR011527

Sanders, B. F., & Schubert, J. E. (2019). PRIMo: Parallel raster inundation model. *Advances in Water Resources*, *126*, 79–95. https://doi.org/10.1016/j.advwatres.2019.02.007

Schaefer, J. T. (1990). The critical success index as an indicator of warning skill. *Weather and Forecasting*, *5*(4), 570–575. https://doi.org/10.1175/1520-0434(1990)005<0570:Tcsiaa>2.0.Co;2

Shustikova, I., Domeneghetti, A., Neal, J. C., Bates, P., & Castellarin, A. (2019). Comparing 2D capabilities of HEC-RAS and LISFLOOD-FP on complex topography. *Hydrological Sciences Journal*, *64*(14), 1769–1782. https://doi.org/10.1080/02626667.2019.1671982

Snelson, E., & Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs. Advances in neural information processing systems. https://doi.org/10.5555/2976248.2976406

South Australia - Department for Environment and Water. (2022). *Environmental watering and monitoring*. Government of South Australia. Retrieved from https://www.environment.sa.gov.au/topics/river-murray/improving-river-health/wetlands-and-floodplains/chowilla-floodplain/environmental-watering-and-monitoring

Sridharan, B., Bates, P. D., Sen, D., & Kuiry, S. N. (2021). Local-inertial shallow water model on unstructured triangular grids. *Advances in Water Resources*, *152*, 103930. https://doi.org/10.1016/j.advwatres.2021.103930

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, *90*, 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Teng, J., Vaze, J., Kim, S., Dutta, D., Jakeman, A. J., & Croke, B. F. W. (2019). Enhancing the capability of a simple, computationally efficient, conceptual flood inundation model in hydrologically complex terrain. *Water Resources Management*, *33*(2), 831–845. https://doi.org/10.1007/s11269-018-2146-7

The Insurance Council of Australia. (2022). 2022 flood now third costliest natural disaster ever. Retrieved from https://insurancecouncil.com.au/resource/2022-flood-now-third-costliest-natural-disaster-ever/

Titsias, M. (2009). Variational learning of inducing variables in sparse Gaussian processes. In *Proceedings of the twelfth international conference on artificial intelligence and statistics, proceedings of machine learning research*. Retrieved from http://proceedings.mlr.press/v5/titsias09a.html

US Army Corps of Engineers. (2021a). Hydraulic reference manual. [Computer Program Documentation] (HEC-RAS - River Analysis System,Version 6.0).

US Army Corps of Engineers. (2021b). 2D modeling User's manual. [Computer Program Documentation] (HEC-RAS - River Analysis System,Version 6.0).

Wu, W. Y., Emerton, R., Duan, Q. Y., Wood, A. W., Wetterhall, F., & Robertson, D. E. (2020). Ensemble flood forecasting: Current status and future opportunities. *Wiley Interdisciplinary Reviews-Water*, *7*(3), e1432. https://doi.org/10.1002/wat2.1432

Xie, S., Wu, W., Mooser, S., Wang, Q. J., Nathan, R., & Huang, Y. (2021). Artificial neural network based hybrid modeling approach for flood inundation modeling. *Journal of Hydrology*, *592*, 125605. https://doi.org/10.1016/j.jhydrol.2020.125605

Xu, X., Zhang, X., Fang, H., Lai, R., Zhang, Y., Huang, L., & Liu, X. (2017). A real-time probabilistic channel flood-forecasting model based on the Bayesian particle filter approach. *Environmental Modelling & Software*, *88*, 151–167. https://doi.org/10.1016/j.envsoft.2016.11.010

Yang, Q., Wu, W., Wang, Q. J., & Vaze, J. (2022). A 2D hydrodynamic model-based method for efficient flood inundation modelling. *Journal of Hydroinformatics*, *24*(5), 1004–1019. https://doi.org/10.2166/hydro.2022.133

Yu, D., & Lane, S. N. (2006). Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 1: Mesh resolution effects. *Hydrological Processes*, *20*(7), 1541–1565. https://doi.org/10.1002/hyp.5935

Zhao, P., Wang, Q., Wu, W., & Yang, Q. (2022). Spatial mode-based calibration (SMoC) of forecast precipitation fields from numerical weather prediction models. *Journal of Hydrology*, *613*, 128432. https://doi.org/10.1016/j.jhydrol.2022.128432

Zhou, Y., Wu, W., Nathan, R., & Wang, Q. J. (2021). A rapid flood inundation modelling framework using deep learning with spatial reduction and reconstruction. *Environmental Modelling & Software*, *143*, 105112. https://doi.org/10.1016/j.envsoft.2021.105112

Zischg, A. P., Felder, G., Mosimann, M., Rothlisberger, V., & Weingartner, R. (2018). Extending coupled hydrological-hydraulic model chains with a surrogate model for the estimation of flood losses. *Environmental Modelling & Software*, *108*, 174–185. https://doi.org/10.1016/j.envsoft.2018.08.009

References From the Supporting Information

Chinchor, N. (1992). MUC-4 evaluation metrics. In *Proceedings of the 4th conference on Message understanding - MUC4 '92*. https://doi.org/10.3115/1072064.1072067

Fuqin, L., Jupp, D. L. B., Sixsmith, J., Wang, L., Dorj, P., Vincent, A., et al. (2022). GA Landsat 5 TM analysis ready data collection 3 geoscience Australia. Retrieved from http://pid.geoscience.gov.au/dataset/ga/130853

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Queensland Government. (2022). SILO - Australian climate data from 1889 to yesterday. Retrieved from http://www.longpaddock.qld.gov.au/silo

Sims, N. W., Garth, Overton, I., Austin, J., Gallant, J., King, D., et al. (2014). *RiM-FIM floodplain inundation modelling for the Edward-Wakool, lower Murrumbidgee and lower Darling river systems* (water for a healthy country flagship Report series, issue). CSIRO. Retrieved from https://publications.csiro.au/rpr/download?pid=csiro:EP143823&dsid=DS3

Szabo, S., Gácsi, Z., & Bertalan-Balazs, B. (2016). Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. *Landscape & Environment*, *10*(3–4), 194–202. https://doi.org/10.21120/LE/10/3-4/13

Xu, H. (2005). A study on information extraction of water body with the modified normalized difference water index (MNDWI). *Journal of Remote Sensing*, *9*, 589–595.