Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR032248

Special Section:
Advancing flood characterization, modeling, and communication

# Upskilling Low-Fidelity Hydrodynamic Models of Flood Inundation Through Spatial Analysis and Gaussian Process Learning

**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 new hybrid surrogate model for predicting the dynamic evolution of flood inundation extent is proposed
- The hybrid model significantly improves the accuracy of flood inundation extent predictions compared to a low-fidelity model
- The computational cost is substantially reduced compared to a high-fidelity model

**Correspondence to:**
N. Fraehr,
nfraehr@student.unimelb.edu.au

**Citation:**
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*, e2022WR032248. https://doi.org/10.1029/2022WR032248

Received 23 FEB 2022
Accepted 1 JUL 2022

**Abstract** Accurate flood inundation modeling using a complex high-resolution hydrodynamic (high-fidelity) model can be very computationally demanding. To address this issue, efficient approximation methods (surrogate models) have been developed. Despite recent developments, there remain significant challenges in using surrogate methods for modeling the dynamical behavior of flood inundation in an efficient manner. Most methods focus on estimating the maximum flood extent due to the high spatial-temporal dimensionality of the data. This study presents a hybrid surrogate model, consisting of a low-resolution hydrodynamic (low-fidelity) and a Sparse Gaussian Process (Sparse GP) model, to capture the dynamic evolution of the flood extent. The low-fidelity model is computationally efficient but has reduced accuracy compared to a high-fidelity model. To account for the reduced accuracy, a Sparse GP model is used to correct the low-fidelity modeling results. To address the challenges posed by the high dimensionality of the data from the low- and high-fidelity models, Empirical Orthogonal Functions analysis is applied to reduce the spatial-temporal data into a few key features. This enables training of the Sparse GP model to predict high-fidelity flood data from low-fidelity flood data, so that the hybrid surrogate model can accurately simulate the dynamic flood extent without using a high-fidelity model. The hybrid surrogate model is validated on the flat and complex Chowilla floodplain in Australia. The hybrid model was found to improve the results significantly compared to just using the low-fidelity model and incurred only 39% of the computational cost of a high-fidelity model.

**Plain Language Summary** Floods are the most common type of natural disaster and therefore it is important to predict when and where flooding occurs. This is normally done using a complex computer model that divides the area of interest into small subareas and then calculates how the water moves between each subarea. However, to predict flooding accurately over large areas, it is necessary to use millions of small subareas and it takes a long time to calculate the movement of flood water between subareas. To mitigate this issue, this study proposes an alternative approach based on a simpler computer model. This simpler model uses larger subareas to predict flooding, which makes the model less accurate but much faster. To compensate for the reduced accuracy, the results are corrected using an advanced computer method that is calibrated to predict the relationship between the predictions made using the complex and simpler models. The new approach is used to predict flooding on a large, flat floodplain in Australia. The predictions show a significant improvement compared to just using the simpler computer model. Furthermore, the calculations only take about 39% of the time taken by a complex model with the small subareas, but the accuracy is similar.

## 1. Introduction

Floods are some of the most destructive natural disasters in the world and they are projected to become more severe and frequent with climate change (IPCC, 2021). During a flood event normally dry areas are inundated until a maximum inundation extent is reached (flooding period), whereafter the water recedes back to the normal state (recession period). Capturing the dynamics of this behavior is of great importance for risk management and has led to the development of advanced hydrodynamic models. Hydrodynamic models can represent different levels of complexity and precision. For simulating the dynamics of flood inundation, two-dimensional hydrodynamic models that numerically solve the depth-averaged Navier-Stokes equations on a high-resolution grid is normally applied (Teng et al., 2017). These high-resolution hydrodynamic models are often referred to as high-fidelity models, where the fidelity refers to the model's degree of realism (Razavi et al., 2012). However,

the high precision of high-fidelity models comes at an expense of high computational cost, which makes them unfeasible in many practical applications such as ensemble and real-time modeling (Teng et al., 2017; W. Y. Wu et al., 2020). To address this issue, computationally efficient approximation methods named surrogate models have been developed (Razavi et al., 2012).

Many different types of surrogate models have been considered and can be divided into three groups: conceptual, emulator, and low-fidelity models (McGrath et al., 2018; Razavi et al., 2012; Teng et al., 2017). Simplified conceptual models utilize simple hydraulic concepts to make predictions and can provide useful estimates for the maximum or final flood inundation extent (McGrath et al., 2018; Teng et al., 2017). However, their capability to predict the dynamical behavior of the flood events is limited (McGrath et al., 2018; Teng et al., 2017).

Emulator models, also known as response surface surrogates or meta models (Razavi et al., 2012), are data-driven models that are trained to predict observations or results from high-fidelity models. Emulators are capable of mapping complex non-linear relationships, and, once trained, have a high computational efficiency (Razavi et al., 2012). However, emulators are not physics-based models, and it is not straightforward to employ an emulator to approximate high spatial-temporal dimensional data from a high-fidelity flood inundation model. To deal with the hysteresis of system behavior, it is usually necessary to incorporate timeseries data. For emulators, this is often done by time-shifting input variables to provide information on previous and future timesteps. This is a simple approach to provide the emulator model with a sense of memory, but each time-shifted input creates a new input to the model, and thereby increases the dimensionality the input data (Brahim-Belhouari & Bermak, 2004; Brahim-Belhouari & Vesian, 2001; Zahura et al., 2020). Consequently, emulator models are often limited to just predicting the maximum flood inundation extent (e.g., Devi et al., 2019; Kim & Han, 2020; Lin et al., 2020) rather than predicting a timeseries of flood behavior.

However, recently emulator-based surrogate models have been developed to incorporate timeseries data and to predict the dynamic flood inundation extent (Chu et al., 2020; Kabir et al., 2021; Xie et al., 2021; Zhou et al., 2021). These studies predict flood inundation using numerous individual emulator models. Each of the models are independent and predict flooding at a specific location in the floodplain. The number of individual models varies with model application. For example, Kabir et al., 2021 used 150, Zhou et al. (2021) used 125, Chu et al. (2020) used 14227 and Xie et al. (2021) used 14278. Using many single models is impractical and does not account for the spatial correlation of flood inundation behavior (Chu et al., 2020). To address this issue, new methods have been proposed, such as the parallel partial approach by Gu and Berger (2016) and Ma et al. (2019) where correlation parameters are shared between individual Gaussian Process (GP) emulator models. Even so, dealing with spatial correlation is an issue that persists and needs to be addressed when employing emulator models.

Low-fidelity models represent the last type of surrogate models. These are physics-based models similar to high-fidelity models, but with reduced complexity. Model complexity is reduced by changing the numerical accuracy, adopting simplified assumptions for the governing scheme, or applying a simpler model type (e.g., using a one-dimensional instead of two-dimensional model) (Asher et al., 2015; Razavi et al., 2012). Due to the reduced complexity, low-fidelity models have a lower computational demand than high-fidelity models, but at the cost of reduced accuracy (Fernández-Godino et al., 2017, 2019; Liu et al., 2018; Park et al., 2017). In comparison to emulator models, low-fidelity models can more easily incorporate hysteresis and spatial dimensionality but with a higher computational burden.

Emulator and low-fidelity models both have their strengths and weaknesses, thus a combination of these two or a hybrid model utilizing both surrogate model types, is an appealing approach. However, as mentioned previously emulator models have issues dealing with the spatial correlation inherent in hydrodynamic behavior, thus many single models are used for individual locations across a floodplain. This is often impractical and can lead to discontinuity between the estimates derived for neighboring grid cells. To reduce the number of emulator models, dimensionality reduction techniques such as feature selection methods have been introduced to identify key locations in a floodplain (e.g., Zhou et al. (2021)). An alternative way of reducing dimensionality of spatial-temporal data is to extract key features in the form of patterns or trends (feature extraction methods). A common feature extraction method is Empirical Orthogonal Function (EOF) analysis, which has been used in areas of remote sensing, climate science and oceanography (e.g., Aires et al., 2014, 2020; Alvarez & Pan, 2016; Chang et al., 2020; Ghosh et al., 2021; Golestani & Sørensen, 2013; Jolliffe & Cadima, 2016; Marques et al., 2009). EOF analysis

reduces the spatial-temporal data into pairs (modes) of spatial patterns (EOF) and temporal variability functions, termed expansion coefficients (EC) (Jolliffe & Cadima, 2016; Zhang & Moore, 2015). When ranked, each mode explains a descending proportion of the variance in the data, and the dimensional reduction is achieved by using only the first few significant modes to explain most of the variance in the data set (Jolliffe & Cadima, 2016; Zhang & Moore, 2015). In addition, EOF analysis is reversible, meaning that the data set can be both decomposed to and reconstructed from the ECs and EOFs (e.g., Aires et al., 2014).

EOF analysis can be used for downscaling data from low-resolution to high-resolution, thus making it appealing for use with low- and high-fidelity flood inundation modeling. For this reason, Carreau and Guinot (2021) recently predicted high-resolution water depths and discharge using a hybrid surrogate approach that combined a low-resolution hydrodynamic model with Artificial Neural Network (ANN) emulator models to predict ECs from a high-resolution hydrodynamic model. Carreau and Guinot (2021) demonstrated the value of using EOF analysis and emulator models to downscale the results from low-fidelity models, and they obtained higher resolution predictions of water depth and discharge for flooding events in urban environments. They derived the "low-fidelity model results" by averaging the high-fidelity results over selected subdomains. While this approach suited their evaluation purposes, in practice the low-fidelity model results need to be derived independently from the high-fidelity model to avoid the computational burden involved, and this will most likely introduce additional uncertainty to the low-fidelity results. It is also worth noting that they developed individual EOF analyses and hybrid models specific to different flow problems. To ensure consistency, the EOF analysis should be performed once for the entire data set of flood events, and the same hybrid model should be able to simulate the full duration of various flood events on a real-world topology with complex flow patterns and dynamically changing inundation extents.

An emulator, such as the ANN used by Carreau and Guinot (2021), is well suited to describe the complex functional relationships that exists between the ECs. Nevertheless, in recent years a probabilistic treatment of predictions has increased in popularity and with it, interest in Gaussian Process (GP) models. This is due to the ability of a GP model to characterize uncertainty by predicting both the mean and standard deviation of the associated errors (Schulz et al., 2018). GP models have already been used in numerous studies to predict wave height/water level (Ma et al., 2019; Malde et al., 2016; Parker et al., 2019), timeseries behavior (Brahim-Belhouari & Bermak, 2004; Contreras et al., 2020; Hachino & Kadirkamanathan, 2011), and timeseries with ECs as input (Avendaño-Valencia et al., 2017), and they have been used widely in multi-fidelity modeling (Fernández-Godino et al., 2017, 2019; Park et al., 2017; Toal, 2015). However, GP models become very computationally demanding when dealing with large datasets due to the difficulties encountered when inverting large covariance matrices (Bauer et al., 2017; Burt et al., 2019). Flood inundation events can have long timeseries consisting of several thousand timesteps, thereby making it computationally infeasible to use the GP model. Fortunately, Sparse Gaussian Processes (Sparse GP) offer means to this issue. The Sparse GP models use a number of inducing variables to approximate the full GP and thereby reduce the computational demand (Leibfried et al., 2021). Despite the promising aspects of the Sparse GP models, their applications to real-life problems are still limited, and this study therefore aims to investigate approaches that are suited for practical applications of this type of emulator models.

This study proposes a new hybrid Low-fidelity, Spatial analysis, and Gaussian Process (LSG) model to provide accurate flood inundation predictions in a computationally efficient manner. The model uses a low-fidelity model as a transfer function to capture the dynamics and spatial correlation of a flood event. The key spatial and temporal features of the low-fidelity model outputs are extracted through EOF dimension reduction techniques, thereby enabling the use of a Sparse GP model to refine predictions of the dynamic evolution of the flood inundation extent. The LSG model is applied to the simulation of complex flow patterns resulting from flood events in a flat extensive floodplain, which provides a challenging application for the model. The aim of the LSG model is to emulate a high-fidelity model and provide comparable results. For this reason, the performance of the LSG model is assessed by comparing to high-fidelity model results for the chosen study area.

This paper is organized as follows. In Section 2 the LSG model is presented, including the methodology for the EOF analysis and Sparse GP model. In Section 3 the case study for the Chowilla floodplain is outlined with the available data and tests performed. Then in Section 4 the results from the case study are presented, followed by discussion and conclusion in Sections 5 and 6, respectively.

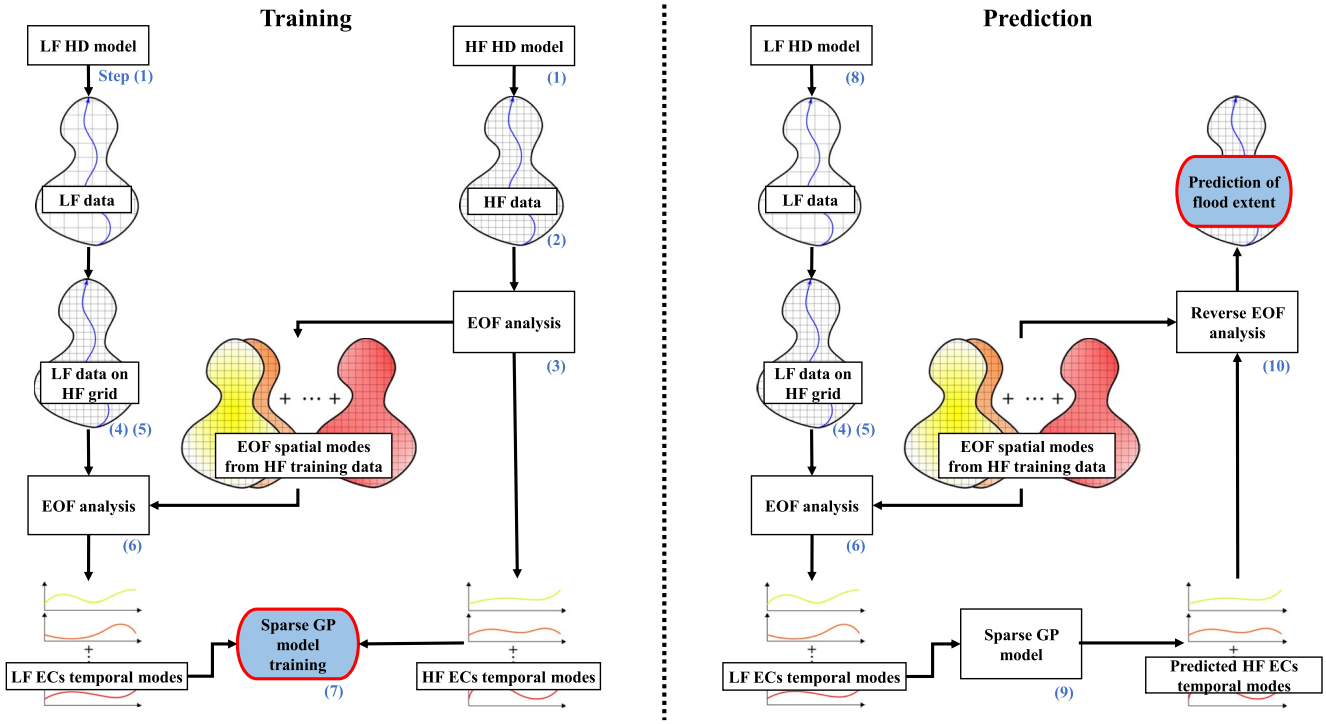


**Figure 1.** Process of training and prediction for the LSG model to simulate flood inundation extent. Blue ovals indicate the output of each process. Numbers in blue correspond to the steps in Table 1.

## 2. LSG Model

The LSG model is a surrogate approach that provides high-fidelity estimates of the dynamic behavior of flood inundation. It consists of a low-fidelity hydrodynamic model and a Sparse GP emulator model, where the Sparse GP model is used to convert the low-fidelity data to high-fidelity data via conversion of ECs from an EOF analysis. In this study the only difference between the low- and high-fidelity models is the degree of spatial resolution adopted, where the lower spatial resolution of the low-fidelity model reduces the accuracy of the predictions.

The workflows for training and prediction are illustrated in Figure 1 and Table 1. EOF analysis is performed on the high-fidelity data, thereby reducing the spatial-temporal data to EOF spatial maps and ECs temporal functions. The low-fidelity data is first converted to the same computational grid as the high-fidelity model, thus enabling the derivation of low-fidelity ECs through the use of the high-fidelity EOFs. Finally, the low-fidelity ECs is used as input and the high-fidelity ECs is used as output to train the Sparse GP model. Once the Sparse GP model is trained, the LSG model can be applied to new flood events to predict the dynamic flood inundation extent without the need to run a high-fidelity model. A detailed description of the workflows is given in the following sections with reference to the steps outlined in Figure 1 and Table 1.

### 2.1. EOF Analysis of Hydrodynamic Data

EOF analysis consists of reducing the dimensionality of spatial-temporal data by creating modes of spatial maps (i.e., EOFs) and temporal functions (i.e., ECs), where each mode is orthogonal to all others (Jolliffe & Cadima, 2016; Zhang & Moore, 2015).

Prior to the EOF analysis, the low- and high-fidelity models are used to simulate several different inundation events that span a wide range of inundation behavior from no flood to extreme flood scenarios **(Step 1)**. This will enhance the output space coverage of the Sparse GP model and improve prediction accuracy for new unseen events (Maier et al., 2010; W. Wu et al., 2013).

As the inundation extent is the focus of this study, the outputs from the low- and high-fidelity models are converted to binary values (1 for flooded and 0 for dry) **(Steps 2 and 5)**. The threshold for flooding is chosen to

**Table 1**
*Step-By-Step Workflow for Training and Prediction Using the LSG Model to Be Read in Conjunction With the Process Diagram in Figure 1*

| Training | | | |
|---|---|---|---|
| Step | Task | Result from task | Purpose of task |
| 1 | Run low- and high-fidelity model for training events. | Training data set for the Sparse GP model. | Running the low- and high-fidelity model for identical events enables the training of the Sparse GP model. |
| 2 | Convert high-fidelity data to binary values. | New binary representation of the high-fidelity model data. | Ensures only the flood extent is represented in the high-fidelity data. |
| 3 | Perform EOF analysis on binary high-fidelity data. | Spatial EOF modes and temporal ECs for high-fidelity data. | Reduces dimension of spatial-temporal high-fidelity data set, so it can be used to train Sparse GP model. |
| 4 | Spatially convert low-fidelity data to the high-fidelity model grid. | New spatial representation of the low-fidelity data. | Changing the spatial representation facilitates the use of the high-fidelity EOF spatial modes in step 6. |
| 5 | Convert low-fidelity data to binary values. | New binary representation of the low-fidelity model data. | Ensures only the flood extent is represented in the low-fidelity data. |
| 6 | Derive low-fidelity ECs using high-fidelity EOF spatial modes. | Temporal ECs modes for low-fidelity data. | Reduces dimension of spatial-temporal low-fidelity data set, so it can be used to train Sparse GP model. |
| 7 | Train Sparse GP model using low-fidelity ECs as inputs and high-fidelity ECs as outputs. | Optimized Sparse GP model. | Enables the Sparse GP model to convert low-fidelity ECs to high-fidelity ECs. |
| Prediction | | | |
| Step | Task | Result from task | Purpose of task |
| 8 | Run low-fidelity model for new event and follow step 4–6. | Temporal ECs for new event. | Creates a new input for the Sparse GP model. |
| 9 | Predict high-fidelity ECs using trained Sparse GP model. | Predicted high-fidelity ECs. | The predicted high-fidelity ECs is needed to reconstruct the inundation prediction in high-resolution. |
| 10 | Inverse EOF analysis using high-fidelity EOF spatial modes and predicted high-fidelity ECs | High-resolution prediction of flood inundation extent. | Upskills low-fidelity model prediction of flood inundation. |

*Note*. EC, expansion coefficients.

be 3 cm to ignore insignificant flooding and reduce numerical errors. The binarization facilitates the grouping of the grid cells into the three categories "Always dry" (AD), "Always flooded" (AF), and "Temporary flooded" (TF) based on their change of state over time. The state of the AD and AF cells remain constant over time and are therefore left out of the EOF analysis. The final step before the EOF analysis is to remove the temporal mean from the binary timeseries of each of the TF cells (detrending) and to apply a weighting according to the cell size. As hydrodynamic model grids can have cells of varying sizes (unstructured grids), this weighting ensures that larger grid cells are given higher weights, as they account for a larger proportion of the inundated area. If the cells have the same size (structured grids), the weighting can be disregarded as all cells would be given the same weight.

Let $HF$ be a $T \times P$ matrix, where each row is a timestep $t$ for $t = 1,\ldots,T$, and each column $p$ is a TF cell in the high-fidelity model for $p = 1,\ldots,P$. The EOF analysis is performed via singular value decomposition of the $HF$ matrix and follows Equation 1 **(Step 3)**. The EOF analysis is performed using the sklearn.decomposition.PCA module in the Scikit-learn machine learning package in Python programming language (Pedregosa et al., 2011).

$$\begin{aligned} HF &= EOF_{HF} \cdot \mathrm{U} \cdot \mathrm{D} \\ &= EOF_{HF} \cdot EC_{HF} \\ &\approx \sum_{k=1}^{K} EOF_{HF}(k,:) \cdot EC_{HF}(:,k) \end{aligned} \tag{1}$$

where $EOF_{HF}$ is a $T \times P$ orthogonal matrix where each row corresponds to a spatial map, and $EC_{HF}$ is a $T \times T$ matrix of column-wise temporal functions. $U$ and $D$ are $T \times T$ matrices, where $D$ is diagonal, containing respectively the eigenvectors and eigenvalues $\lambda$ of the covariance matrix from the EOF analysis. To enhance computational efficiency, only the first 100 EOF and ECs modes are derived. This is sufficient to ensure the significant modes are obtained.

In line three of Equation 1 the data is represented by the first $K$ significant modes. The modes account for a decreasing proportion of the variance, meaning the majority of the variance in the data set is described in the first $K$ modes, where $K \ll T$. The remaining modes are considered noise and do not contain meaningful information about the data set. The error involved in using only the first $K$ modes to reconstruct the high-fidelity data set is considered minimal, thus, it is only $EC_{HF}(:,1{:}K)$ that needs to be predicted using the Sparse GP model. The significant modes are found using North's test (see Equation 2), which states that modes are significant if the difference between the eigenvalues of two modes are bigger than the error limits (North et al., 1982). Furthermore, all modes chosen should have eigenvalues above one (Kaisers Rule) to ensure the modes provide more information than just using the original individual input variables (Kaiser, 1960).

$$\Delta\lambda > \lambda\sqrt{2/T} \tag{2}$$

After the $EC_{HF}$ is derived, the next step is to prepare the low-fidelity data as input for the Sparse GP model. The low-fidelity model has a lower spatial resolution than the high-fidelity model, but by converting the low-fidelity data to the high-fidelity model grid (using the same spatial representation as the high-fidelity data) the $EOF_{HF}$ matrix can be used to derive the ECs for the low-fidelity data set **(Step 4)**. This approach obviates the need to derive EOF spatial modes for the low-fidelity data as it makes use of the high-fidelity EOFs derived the one time in **Step 3** from the high-fidelity data. Additionally, this spatial conversion ensures the ECs for all flood events for both the low- and high-fidelity data are derived using the same basis of EOF spatial modes. The spatial conversion is performed using a nearest neighbor method, where each high-fidelity cell is assigned the value of the closest low-fidelity cell for all timesteps by using the Euclidean distance between the $x$-$y$ coordinates. This method is chosen as it is independent of the grid structure and resolution of the low- and high-fidelity model.

As for the high-fidelity data set, only the TF cells are used in the EOF analysis for the low-fidelity data, thereby creating a new $T \times P$ matrix named $LF$ consisting of the low-fidelity data. The low-fidelity data is detrended and weighted in the same manner as for the high-fidelity data. This pre-processing enables the derivation of the ECs for the low-fidelity data utilizing the orthogonality of the $EOF_{HF}$ matrix in Equation 3 **(Step 6)**.

$$EC_{LF} = LF \cdot EOF'_{HF} \tag{3}$$

where $EC_{LF}$ is a $T \times T$ matrix of temporal functions derived for the low-fidelity data set and $EOF'_{HF}$ is the transpose of the $EOF_{HF}$ matrix.

Once both the $EC_{LF}$ and $EC_{HF}$ are derived, they can be used as input and output to train the Sparse GP model.

### 2.2. Sparse Gaussian Process (Sparse GP) Model

The $EC_{HF}(:,1{:}K)$ are predicted using individual Sparse GP models, thereby creating a total of $K$ models. The models are assumed to be fully independent due to the orthogonality of the $EC_{HF}$ in the EOF analysis. The number of models developed here is significantly reduced compared to the approach of building an emulator for each grid cell in the high-fidelity model. The Sparse GP models are implemented in Python using the GPflow package (Matthews et al., 2017), which has the advantage of utilizing GPU calculations for optimization of the model to reduce computational time. All descriptions under Section 2.2 are linked to **Step 7** in Table 1.

#### 2.2.1. General Concepts of the GP and Sparse GP Models

A GP model can predict non-linear complex relationships with statistical confidence by assuming that the relationship between input and output follows a Gaussian distribution of functions, explained by the mean and variance (see Equation 4 below) (Rasmussen & Williams, 2006).

$$GP(x) \sim \mathcal{N}\left(m(x), k\left(x, x'\right)\right) \tag{4}$$

where $m(x)$ is the mean function, which is normally assumed to be zero (Rasmussen & Williams, 2006), and $k(x,x')$ is the covariance function (popularly referred to as a "kernel") that is used to generate the covariance matrix. The kernel controls the variance of the prediction, and numerous kernel functions have been developed (Rasmussen & Williams, 2006). Different kernel functions may lead to different results, and therefore initial tests have been carried out using the most commonly used kernel functions including Radial Basis Function, Matern 3/2, Matern 5/2, and Exponential. The Exponential kernel has been found to provide the most robust performance

given the $EC_{LF}$ and $EC_{HF}$ as input and output, respectively. The Exponential kernel (see Equation 5) is a special case of the Matern kernel, with 1/2 roughness parameter and double lengthscale.

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{x - x'}{2l}\right) + \sigma_n^2 \tag{5}$$

where $\sigma_f^2$ is the signal variance, $l$ is the lengthscale, $x-x'$ is the Euclidean distance between inputs points, and $\sigma_n^2$ is the noise variance. The terms $\sigma_f^2$ and $l$ represent the hyperparameters of the GP that are optimized by maximum likelihood estimation. However, this requires inversion of the covariance matrix that has a computational requirement of $\mathcal{O}(T^3)$. This makes the GP model optimization infeasible when dealing with timeseries data that can have several thousand input samples (Bauer et al., 2017; Leibfried et al., 2021).

To deal with the high computational demand of full GP models, approximation methods called Sparse GP models have been developed (Bauer et al., 2017; Leibfried et al., 2021). Sparse GP models approximate the full GP via introduction of $M$ inducing points, which reduces the computational requirement to $\mathcal{O}(TM^2)$ (Snelson & Ghahramani, 2006; Titsias, 2009). The adaption of Equation 4 to accommodate the use of inducing points is shown in Equation 6.

$$SPGP(x) \sim \mathcal{N}\left(y|k_x' K_M^{-1} \overline{y}, K_{xx} - k_x' K_M^{-1} k_x + \sigma_n^2 I\right) \tag{6}$$

where $k_x$ is $k(x, \overline{x})$, $K_M$ is $k(\overline{x}, \overline{x})$, and $K_{xx}$ is $k(x,x')$. The variables $y$ and $x$ are the observation and input points, respectively, where $\overline{y}$ and $\overline{x}$ are the inducing points for the observations and input points. The observation inducing points ($\overline{y}$) can be removed via integration by assuming a prior distribution following the full GP, which is reasonable as $\overline{y}$ is expected to follow $y$ (Snelson & Ghahramani, 2006). Consequently, inducing points only need to be found for the input points.

Several types of Sparse GP models have been developed (Bauer et al., 2017; Leibfried et al., 2021; Titsias, 2009). Among them, the variational inference based Sparse GP model has the attractive feature that it improves with an increasing number of inducing points, and provides a good approximation to the full GP (Bauer et al., 2017). Therefore, the variational inference based Sparse GP model is chosen in this study to predict the relationship between $EC_{LF}$ and $EC_{HF}$. For more information on the Sparse GP model, the reader is referred to Burt et al. (2019) and Leibfried et al. (2021).

#### 2.2.2. Training of Sparse GP Models

The training of the Sparse GP models is performed using the maximum likelihood method, where the maximum likelihood estimates of the hyperparameters, $\sigma_f^2$ and $l$, and inducing points are obtained using the L-BFGS-B optimization algorithm. Each individual Sparse GP model is trained using all modes of the $EC_{LF}(:,1{:}K)$ as input and only one mode $EC_{HF}(:,k)$ as output **(Step 7)**. This ensures the Sparse GP models are optimized to the specific mode $k$ utilizing all the information available in the low-fidelity data. The input and output ECs timeseries are standardized to a mean of 0 and variance of 1 before being incorporated in the Sparse GP models to ensure numerical stability. A single lengthscale is optimized across all input dimensions in the Sparse GP models, as Automatic Relevance Detection with individual lengthscales for each input dimension can lead to overfitting of GP models (Cawley & Talbot, 2010).

The optimization process can have several local optima, and therefore the choice of initial conditions is important (Bauer et al., 2017; Rasmussen & Williams, 2006). The lengthscale describes how far away from an input sample that information can be used, and often a good initial choice of the lengthscale lies within the boundaries of the input sample values. The initial value of the lengthscale for each Sparse GP model is chosen as the absolute average value of the input values. This has shown to be a robust choice and ensures a good optimization. The signal variance $\sigma_f^2$ is optimized using an initial guess of 1, which is the default value for most applications.

Selecting the number and location of the inducing points is not straightforward. The number of inducing points depend on the number and distribution of the input data. When choosing the number of inducing points, the number should be significantly less than the number of input points to leverage the computational advantage of the sparse approximations. The ratio depends on the amount and distribution of the input data. The initial locations of the inducing points are chosen by initially distributing them linearly from the minimum to maximum value of the input, as this ensures a fast and robust optimization.

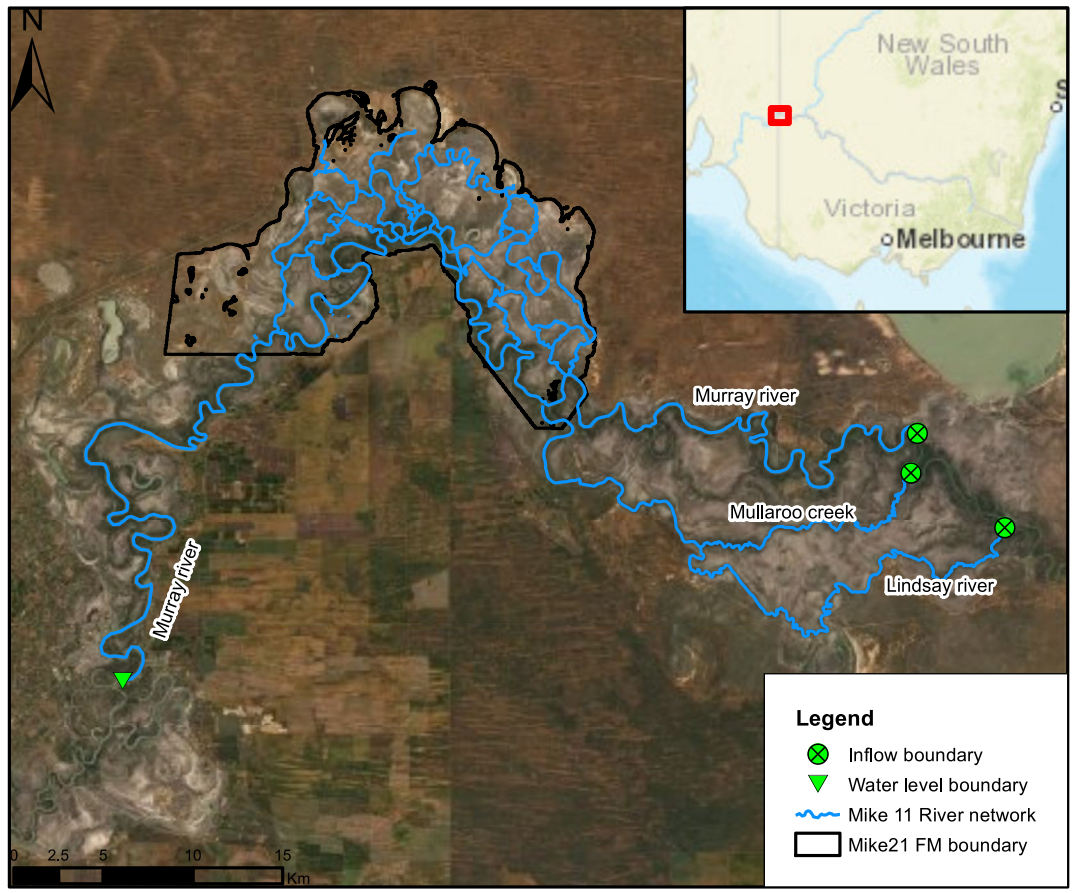


**Figure 2.** Study area and boundary locations for the MIKE 11 and MIKE 21 models (ESRI, 2021).

In addition, to further reduce the risk of being stuck in local optima in the optimization process, only the inducing points are optimized initially while the hyperparameters are fixed, as suggested in a previous study (Bauer et al., 2017). Thereafter, the hyperparameters are optimized with the inducing points fixed.

### 2.3. Reconstruction of Flood Extent Data Using Predicted ECs

Once the Sparse GP models are trained, the low-fidelity model can be run for new flood events **(Step 8)**, and the Sparse GP model can be used to predict $EC_{HF}$ **(Step 9)**. By reversing the EOF procedure, the data for the TF cells can be reconstructed using the $K$ significant modes, following Equation 1 **(Step 10)**. The flood data does not reconstruct fully from the EOF analysis, even if the $EC_{HF}$ is perfectly predicted, as not all modes are used. For this reason, the reconstructed flood data is converted to binary values by adopting a standardized threshold of 0.5 to differentiate between flooded and dry cells. To reconstruct the data set for all cells (AF, AD and TF), the AD and AF cells are added to the reconstructed TF cell data. This provides a high-resolution prediction of the dynamic flood inundation extent without the need to run a high-resolution high-fidelity model.

## 3. Application of the LSG Model

### 3.1. Study Area and Hydrodynamic Models

The LSG model is evaluated on the flat and complex Chowilla floodplain, which is located near the state border of New South Wales, Victoria, and South Australia in south-eastern Australia (see Figure 2). The Chowilla floodplain is adjacent to the Murray river, and includes several small creeks, wetlands, lakes, and billabongs that all contribute to the dynamic change of inundation in the area (Murray-Darling Basin Authority, 2021a). Flood events in the Chowilla floodplain can last several months due to the combination of a flat topography and low

gradient of the Murray River that together slows down the movement of water. Furthermore, the Murray River is roughly 2,500 km long (Murray-Darling Basin Authority, 2021b) and has a large catchment area (>1 million $km^2$ (Murray-Darling Basin Authority, 2022)). The Chowilla floodplain is located in the downstream part of the catchment thus resulting in long runoff times and extended periods with high flows. The study area is approximately 224 $km^2$.

To simulate flood inundation of the study area, a hydrodynamic model provided by the Murray–Darling Basin Authority (MDBA) is used. The model is calibrated to simulate the inundation in the Chowilla floodplain, and it is currently used by the MDBA to simulate the natural inundation extent (e.g., Nicol et al. (2020)). The model is a two-way coupled model, also known as a one-dimensional + two-dimensional (1D-2D) model, consisting of a MIKE 11 and a MIKE 21 FM model that are combined using the MIKE FLOOD framework (DHI, 2019). The MIKE 11 model simulates the water level and discharge in the river network based on the upstream inflow and downstream water level boundaries. The boundary conditions for the MIKE 11 model are obtained from the Bureau of Meteorology's (BoM) online water data platform (Bureau of Meteorology, 2021). The river bathymetry is incorporated through 796 cross-sections and Manning coefficients varying between 17–33 $m^{1/3}/s$. Additionally, the MIKE11 model includes 8 weirs, 15 culverts and 13 control structures (gates and overflow regulators that are kept steady throughout the simulations) that affect the flow in the river channels. The MIKE 21 model simulates the 2D surface flow on a quadratic grid with a spatially varying Manning coefficient of 17–33 $m^{1/3}/s$. There is no precipitation included, and a "no-flow" boundary is used along the edge of the MIKE 21 model, meaning that any changes to water on the floodplain are due to interactions with the MIKE 11 model.

In this study, both high- and low-resolution MIKE 21 models are used. These constitute the high- and low-fidelity models used in the EOF analysis, as discussed in Section 2.1. The dimensions of the grid cells in the high-fidelity model is 30 × 30 m, and in total 249,263 cells are required to represent the full model domain. The low-fidelity model has coarser grid cells of 100 × 100 m (28,935 cells in total) and is developed by averaging the elevation and roughness of the high-fidelity grid cells over the larger area.

### 3.2. Generating Training and Validation Data

The hydrodynamic models are used to simulate flood events for the Chowilla floodplain between 15 August 2010 and 15 January 2021. This period is selected based on the availability of historic data for specifying the boundary conditions and includes nine historic events with durations varying from 75 to 290 days. In this period, the average inflow discharge to the model from the Murray River is 171 $m^3/s$ but spans from a minimum of 21 $m^3/s$ to a maximum of 1,092 $m^3/s$, showing a great variability in the flow conditions. However, four of the nine historical events are too small to cause any significant inundation of the floodplain. This causes a problem for training the Sparse GP models, as a large number of events spanning a wide range of inundation behavior is needed to properly train the models. The training data should include extreme events with respect to the magnitude and the duration of their flood behavior. To ensure this, the observed inflow hydrographs and/or duration of the four small events were scaled to create 21 synthetic events. As a result, a total of 26 flood events (21 synthetic + 5 historic events) are available for model development and evaluation. A summary of the events characteristics is found in Appendix A.

The simulated inundation events are divided into training and validation datasets. Of the 26 events, 21 synthetic and 2 historic events were used for training, and the remaining 3 historic events were used for validation. The three validation events are unique historic events covering the periods 15 August 2010–01 June 2011, 01 March 2012–15 June 2012 and 28 May 2016–30 March 2017. These events are different in magnitude and dynamic flood evolution, and are numbered 1, 3 and 6, respectively (numbering is based on the chronological order of the historic events). The remaining historic events, including all scaled events, are used for training and consist of a total of 10,586 timesteps across all training events.

To ensure the same starting point and the stability of the simulations, all flood events are simulated using the same set of initial conditions, where a fixed timestep of 2 s is adopted for both the MIKE 11 and MIKE 21 models. This timestep was selected by the MDBA in model development to ensure model stability for the exchange between the 1D and 2D models during flooding and drying in the model. In addition, a warm-up period of 10 days is used to establish a relationship between the flood levels obtained by the 1D and 2D models. This warm-up period is

selected based on examination of initial model simulation results, and data from this warm-up period are removed before the EOF analysis.

It is important to have a fine temporal resolution of the hydrodynamic results to accurately describe the flood inundation but increasing the number of timesteps also increases the computational cost of training and prediction for the Sparse GP models. For the Chowilla floodplain the change in the floodplain inundation is relatively slow and therefore a timestep of 6 hr between saved datapoints is chosen. If the LSG model is applied on a more rapidly changing flood problem (e.g., local flash flooding), a higher frequency timestep would be needed.

### 3.3. Setup of Sparse GP Models for the Case Study

The setup and training of the Sparse GP model follow the procedure describe in Section 2. However, the number of modes found by the EOF analysis and the number of inducing variables is dependent on the training data.

For the case study, the number of significant modes ($K$) is found to be 52 modes via EOF analysis on the high-fidelity training data set. These modes explain 97.8% of the variance in the data set and are found by means of North's test (see Section 2.1). This means a total of 52 Sparse GP models are developed and trained for this case study.

The number of inducing points for each Sparse GP model is chosen to be 2% of the number of input samples. This percentage has shown to be sufficient to approximate the ECs in this study and is found via a trial-and-error approach with the training data, which is a commonly used approach (Burt et al., 2019).

### 3.4. Evaluation of the LSG Model

A number of evaluation metrics are used to evaluate the performance of the LSG model. The relative Root Mean Square Error (relRMSE) is used to capture the general performance of the LSG model and is calculated using Equation 7:

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

where $A^{LSG}$ is the prediction using the LSG model, and $A^{HF}$ is the inundation extent simulated using the high-fidelity model.

The prediction of the peak of a flood inundation event is important, as most areas will be inundated at that stage. To reduce the effect of smaller variations the average flood inundation extent of the top 5% highest values is compared by using the relative Peak Value Error (relPeakValErr) in Equation 8:

$$\text{relPeakValErr} = \frac{\overline{A^{LSG}_{peak,5\%}} - \overline{A^{HF}_{peak,5\%}}}{\overline{A^{HF}_{peak,5\%}}} \tag{8}$$

where $\overline{A^{LSG}_{peak,5\%}}$ and $\overline{A^{HF}_{peak,5\%}}$ are the average inundation extent for the 5% highest values obtained from the LSG and high-fidelity models, respectively. The reason for choosing the highest 5% of peak values and not a single timestep is that the peak can last several days, due to the long duration of the floods in the Chowilla floodplain. Tests using the 1%–10% highest values have been carried out, but the adoption of different percentages did not change the conclusions.

Another important parameter for flood prediction is the timing of the flood peak, as this is when the greatest impact on people and infrastructure is to be expected. The ability of the LSG model to predict the timing of the peak is assessed using the relative average peak time error compared to the peak period (relPeakTimeErr−1) for the top 5% highest values (See Equation 9), and the overall timing of the flood inundation prediction is determined using the relative average peak time error (relPeakTimeErr−2) compared to the rising limb of the flood event (See Equation 10).

$$\text{relPeakTimeErr} - 1 = \frac{\overline{t^{LSG}_{peak,5\%}} - \overline{t^{HF}_{peak,5\%}}}{\max\left(t^{HF}_{peak,5\%}\right) - \min\left(t^{HF}_{peak,5\%}\right)} \tag{9}$$

$$\text{relPeakTimeErr} - 2 = \frac{\overline{t^{LSG}_{peak,5\%}} - \overline{t^{HF}_{peak,5\%}}}{\overline{t^{HF}_{peak,5\%}} - t^{HF}_{rise,10\%}} \tag{10}$$

where $t^{HF}_{peak,5\%}$ and $t^{LSG}_{peak,5\%}$ are vectors containing the timesteps at which the top 5% highest flood inundation extent are registered (peak period), $\overline{t^{LSG}_{peak,5\%}}$ and $\overline{t^{HF}_{peak,5\%}}$ are the average timestep for the peak period for the LSG and high-fidelity models, respectively. $t^{HF}_{rise,10\%}$ indicates the start of the rising limb of the flood event, which is chosen to be at a 10% increase compared to the minimum flood extent.

The ability of the LSG model to predict the spatial location of the inundation is assessed using the Probability of Detection (POD) and Rate of False alarm (RFA) as shown in Equations 11 and 12.

$$\text{POD} = \frac{A_{detected}}{A_{detected} + A_{missed}} \tag{11}$$

$$\text{RFA} = \frac{A_{false\ alarm}}{A_{detected} + A_{false\ alarm}} \tag{12}$$

where $A_{detected}$ is the area detected as flooded or dry at a given timestep using both the high-fidelity and LSG models, $A_{missed}$ is flooded areas predicted using the high-fidelity model but which is dry using the LSG model, and $A_{false\ alarm}$ is the flooded areas predicted using the LSG model but not the high-fidelity model. Furthermore, $A_{detected}$, $A_{missed}$, and $A_{false\ alarm}$ are plotted on maps for the maximum inundation extent to inspect the locations of error. Bounds and values corresponding to a good prediction for all the evaluations metrics are shown in Appendix B, Table B1.

## 4. Results

### 4.1. Inundation Extent

The inundation extent for the low-fidelity, LSG and high-fidelity models is shown in Figure 3 for event 1 at three different timesteps. The timesteps are chosen according to the flooding, peak, and recession periods of the flood event (See Figure 4). The resolution of the low-fidelity model is coarse, and the floodplain topology is not well described. In general, the low-fidelity model significantly underestimates the flood inundation extent. This is unexpected, as models with a low-resolution are known to overestimate the flood inundation extent compared to models with a finer resolution (Chatterjee et al., 2008; Yu & Lane, 2006). One reason for this is related to the coupling of the 1D and 2D models. The low- and high-fidelity MIKE 21 models are coupled to the MIKE 11 model at the same location, but not necessarily at the same elevation. As the low-fidelity model is averaged over a larger area, the lower elevations in the river are smoothed out by the floodplain, thus resulting in a higher elevation of the grid cell and of the 1D-2D coupling. This means the river level in the MIKE 11 model has to reach a higher elevation before flooding on the floodplain occurs, and as a result, less water inundates the floodplain.

The LSG model can compensate for this underestimation and demonstrates clear improvement over the predictions from the low-fidelity model. The LSG model overestimates the inundation extent slightly, but in general shows a similar inundation extent to the high-fidelity model at all three timesteps in Figure 3. The performance of the LSG model compared to the high-fidelity model is assessed in detail in the following paragraphs.

The prediction of the LSG model is summarized as a timeseries of the inundation extent for the three validation events in Figure 4. For all three events the low-fidelity model underestimates the flood inundation extent but provides a similar evolution of the flood extent compared to the high-fidelity model. This demonstrates not only the low-fidelity model's ability to capture the dynamic features (timing) of the flood inundation events, but also the need for the Sparse GP models to correct the low-fidelity results.

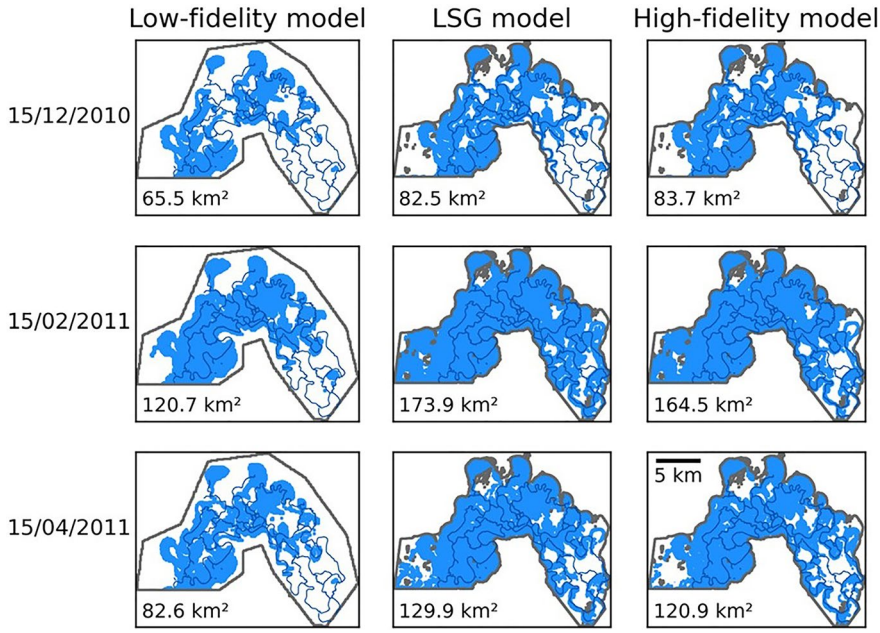


**Figure 3.** Flood inundation extent for validation event 1 simulated using the low-fidelity, Low-fidelity LSG, and high-fidelity models. Rivers are showed as dark blue lines, inundated areas are colored in light blue and the extent is showed in $km^2$ in the lower left corner of each subfigure.

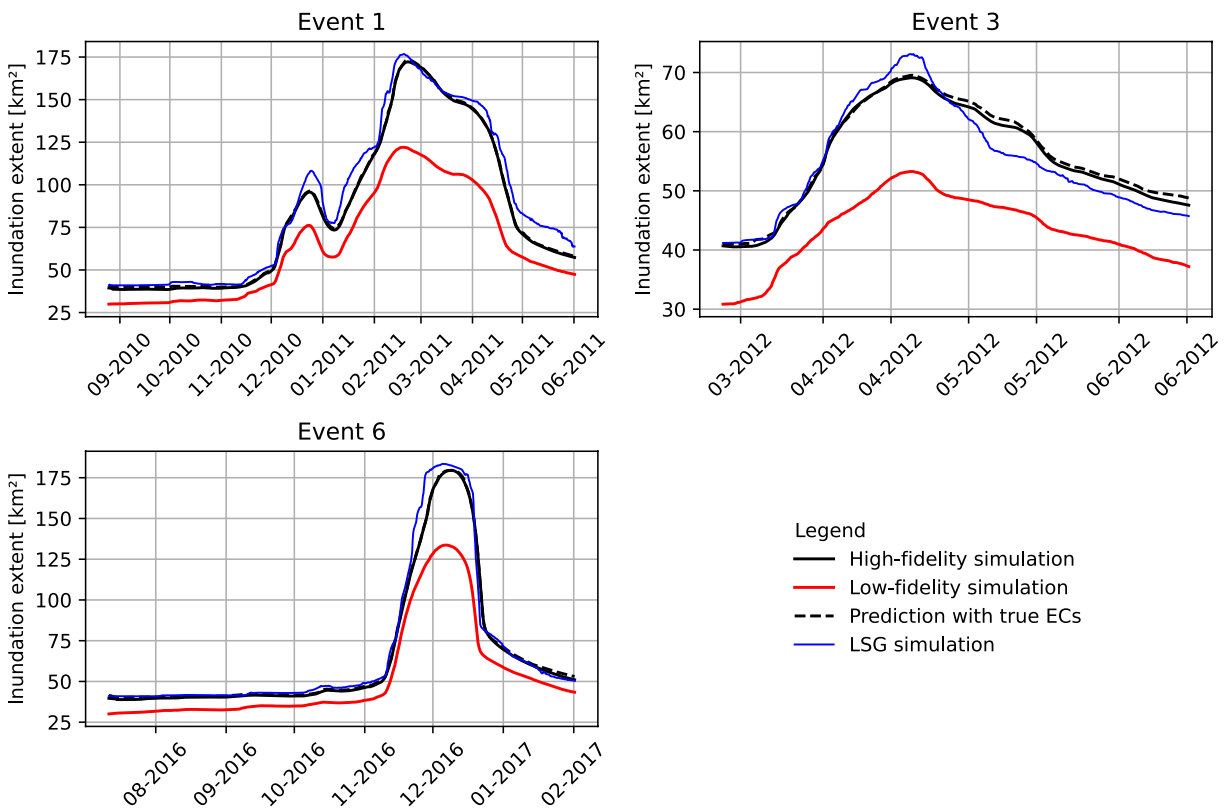


**Figure 4.** Inundation extent obtained using the high-fidelity and LSG models to simulate the three validation events.

**Table 2**
*Evaluation of the Relative Performance of the LSG Model Compared to the High-Fidelity Model to Simulate the Validation Events*

| Metric | Event 1 | Event 3 | Event 6 |
|---|---|---|---|
| relRMSE | 0.09 | 0.04 | 0.09 |
| relPeakValErr | 0.02 | 0.06 | 0.03 |
| relPeakTimeErr-1 | −0.25 | 0.06 | −0.25 |
| relPeakTimeErr-2 | −0.04 | 0.01 | −0.04 |

For event 1 the LSG model significantly improves the low-fidelity model predictions, especially during the first flat period and the rising limb before the first peak. The first smaller peak is overestimated, but for the second and larger peak, the LSG model performs well, and the peak and recession period are only slightly overestimated. For event 3 the LSG model performs significantly better than the low-fidelity model in predicting the rising limb. However, the peak is overestimated significantly, showing the same tendency as for the first smaller peak in event 1. The recession period for event 3 obtained from the LSG model is underpredicted, but it still shows an improvement compared the low-fidelity model. For the last validation event (Event 6), the LSG model predicts the flood inundation extent well from start to finish of the event, despite overpredicting the peak. This shows the LSG model does have the ability to correct the low-fidelity results and to predict a flood inundation extent that is similar to the high-fidelity model. The difference in the prediction accuracy between the validation events is a result of the differences between validation and training events, and more training events could potentially improve the performance of the LSG model.

Considering the evaluation metrics in Table 2, the relative RMSE (relRMSE) for event 3 is lower than that of the other two validation events. This is because event 3 shows signs of both over- and under-prediction, which on average evens out the errors. The peak value is overestimated for all three events (relPeakValErr $> 0$), but the relative error compared to the size of the flood event is low, especially for event 1 and 6. In general, both the relRMSE and relPeakValErr metrics show errors less than 0.10 compared to the high-fidelity model for all three validation events, which is considered a good performance.

The timing of the peak shows a similar tendency for both event 1 and 6, where the LSG predicts the peak earlier than the high-fidelity model, as indicated by the negative peak timing errors (relPeakTimeErr-1 and relPeakTimeErr-2). In the LSG model structure, the low-fidelity model is assumed to capture the dynamics of the event, where the key difference between the high- and low-fidelity models is the spatial resolution of the grid cells. Any systematic differences in timing errors could be compensated for by calibrating the roughness of the low-fidelity model to match the evolution of the flood inundation (Yu & Lane, 2006), or the results of the low-fidelity model could be shifted according to the average timing error in the training data. However, for event 3, the LSG model predicts the peak later than the high-fidelity model, and an adjustment of the low-fidelity model results would therefore not improve predictions for event 3.

### 4.2. Detection of Flooding

The POD and RFA obtained from the LSG model for the three validation events are shown in Figure 5. The results demonstrate that the ability of the LSG model to detect the spatial extent of inundation varies throughout the events. The POD is above 0.76 and the RFA is below 0.20 for the entire duration of all three validation events, and the POD shows better performance of the LSG model at the beginning of the events. Event 6 has a low point in the POD around 20 December 2016, which is due to a timing error of the falling limb of the flood event. The LSG model demonstrates high prediction accuracy for the POD of Event 6 until this point. The RFA varies throughout the events due to the general overprediction of the LSG model. Examining the timeseries behavior of POD and RFA is not typically done, as these metrics are generally used to characterize errors in the maximum flood inundation extent. The LSG model's ability to predict the dynamical flood inundation extent is therefore hard to compare to that of other surrogate models.

Considering the POD and RFA for the maximum inundation extent in Table 3, the LSG model performs well. The POD and RFA of the maximum inundation extent are comparable and are better than found in similar studies, which used surrogate models to predict flood inundation (e.g., Zhou et al. (2021) showed a POD of 0.99–0.999 and RFA of 0.046–0.067, and Xie et al. (2021) showed a POD of 0.955–1 and a RFA of 0.001–0.07).

The extent of the maximum inundation, as well as the detections, misses and false alarms from the LSG model, are shown in Figure 6. In general, there is a good agreement between the LSG and high-fidelity models considering the spatial inundation detection ability of the LSG model, although there are false alarms for all three validation events and misses for events 1 and 3. Events 1 and 6 are larger than event 3 and most of the floodplain

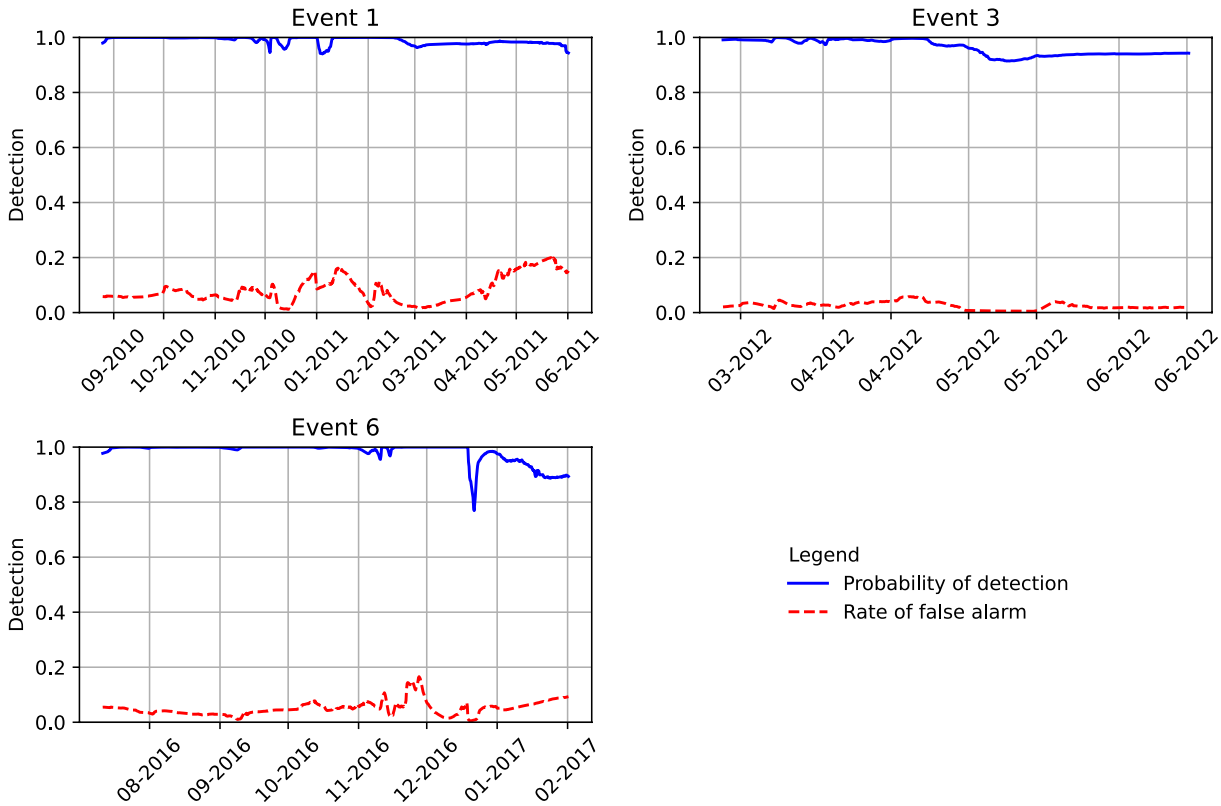


**Figure 5.** Probability of detection and Rate of false alarm for the three validation events.

is inundated at some point during these events. Given the "no-flow" boundary in the MIKE 21 model (described in Section 3.1), flood flows cannot escape by crossing the boundary, which results in a build-up of water on the floodplain. This means most cells will be inundated at some point during the events, and thereby detected in the maximum inundation extent.

The eastern and western parts of the floodplain show the biggest errors between the LSG model prediction and the high-fidelity simulation. These are also the areas that are normally the last to be inundated during a flood event in this floodplain, and inundation in these areas is thus harder to predict than in areas that always get inundated.

### 4.3. Computational Demand

The simulations are carried out on a High-Performance Computer with a 3.70 GHz Intel® Xeon® E-2288G CPU with 64 GB ram and an NVIDIA Quadro RTX 5000 graphic card for GPU calculations. The computational time of the low-fidelity model is approximately 39% of that of the high-fidelity model, see Table 4. The training and prediction time of the EOF analysis and the Sparse GP models is considerably shorter than that of running the low-fidelity model. Further reducing the complexity of the low-fidelity model would increase computational efficiency of the LSG model, but this is likely to also reduce the accuracy of model predictions. The nature of this trade-off is an aspect that needs further exploration.

**Table 3**
*Probability of Detection and Rate of False Alarm of the Maximum Flood Inundation Extent for the Three Validation Events*

| Parameter | Event 1 | Event 3 | Event 6 |
|---|---|---|---|
| POD | 0.99 | 1.00 | 1.00 |
| RFA | 0.03 | 0.06 | 0.02 |

## 5. Discussion

The results in Section 4 demonstrate the potential for the LSG model to provide fast and accurate predictions of flood inundation extent over time. The LSG model has been tested in its ability to successfully emulate a high-fidelity model. The high-fidelity model used in this study was calibrated by the MDBA and little attention has therefore been given to precision of the high-fidelity model compared to observations. However, as the LSG model

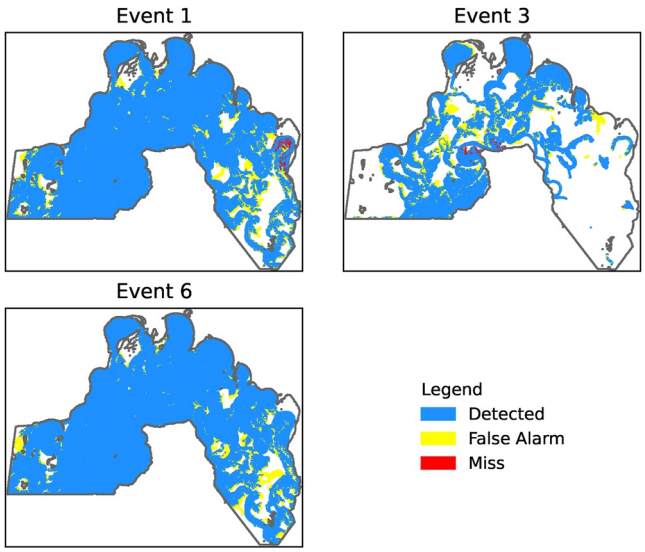


**Figure 6.** Detected, Misses and False alarms for the LSG model compared to the high-fidelity model for the maximum flood extent.

is compared to the high-fidelity model and not observations, the accuracy of the high-fidelity model does not affect the study results. For applying the LSG model to new real-world applications to replace a high-fidelity model, it is important to ensure the high-fidelity model is well calibrated and validated according to observational data.

1D-2D hydrodynamic models, such as the high- and low-fidelity model used in this study, are especially suitable for simulations that are focused on floodplain inundation and less on the river flow (Bates, 2022), but computational advances have made fully 2D models a more practical option, making them more feasible for flood inundation modeling. As mentioned in Section 2, the methodology presented in this paper is not limited to 1D-2D models with constant quadratic grid cells. For confirmation, tests have been carried out using a fully 2D hydrodynamic model with unstructured grid for the Edward-Wakool floodplain (a major anabranch and floodplain of the River Murray, located in southern New South Wales, Australia) and the results (not shown in this paper) are similar to the ones reported for the 1D-2D model for the Chowilla floodplain.

In the development of the low-fidelity model, little attention has been given to the model structure and parameters used. The grids cells in the low-fidelity model are simply averaged over a larger area than in the high-fidelity model. This is a fast, but also crude method to develop the low-fidelity model, as the model parameters are most likely sensitive to the spatial resolution. However, the results show that even using an uncalibrated and coarse low-fidelity model can result in reasonably accurate final predictions. This is due to the powerful transformation of the low-fidelity data through the EOF analysis and Sparse GP that successfully upskills the low-fidelity model results. Furthermore, in this study the low-fidelity model accounts for 99.7% of the computational burden of the LSG model. Although the low-fidelity model is approximately 2.5 times faster than the high-fidelity model, the hybrid model setup used in this study is not feasible for practical applications, such as ensemble and real-time modeling. In ensemble modeling, 10–100 of model realizations are normally used for uncertainty estimates and flood risk assessments (W. Y. Wu et al., 2020). This means the low-fidelity model needs to be several orders of magnitude faster than the high-fidelity model. It is therefore worth exploring possibilities of using an even simpler low-fidelity model structure. Simplifications of the low-fidelity model will compromise the accuracy, thereby creating trade-offs between accuracy and computational burden. In the case study considered, the low-fidelity model is simply a coarser version of the high-fidelity model. To reduce the number of grid cells an unstructured grid that adopts a fine resolution in the river and a coarser resolution on the floodplains could be applied. Additionally, a simplified governing physics scheme can be applied, such as the diffusive wave model used in programs like HEC-RAS and LISFLOOD-FP. This is interesting future directions for the LSG model and will be explored in future research.

One objective of this study was also to examine the Sparse GP model and its performance as an emulator. In training the Sparse GP models, it is essential that the training data includes events of different magnitudes and variability in the evolutional patterns of the flood inundation, so the training data covers the entire output space required (Maier et al., 2010; W. Wu et al., 2013). Once trained, Sparse GP models are able to handle large input datasets and describe the complex relationship between the low- and high-fidelity model for a flat complex floodplain. Inclusion of the Sparse GP model is an important component in achieving accurate predictions in this study and are considered to be an effective emulator for flood inundation simulation.

Besides the choice of low-fidelity and/or emulator model, an important aspect of surrogate modeling is the effort needed to setup the modeling framework. The setup of the LSG model can be tedious due to the need to generate suitable training data set. This is because numerous simulations with the high-fidelity model are needed to train the Sparse GP models and create a robust hybrid surrogate model that can be applied to future flood

**Table 4**
*Training and Prediction Time of the High-Fidelity Model Compared to the Low-Fidelity for Simulation of Validation Event 3*

| | High-fidelity model | Low-fidelity model | EOF analysis + Sparse GP models |
|---|---|---|---|
| Import and data conversion | – | – | 10 min |
| Training of Sparse GP | – | – | 11 min |
| Prediction | 1012 min | 396 min | 1 min |

problems. For this reason, the LSG model is mostly appropriate for a study area where a high-fidelity model and several relevant simulation results are already available, or for projects with a long time-horizon so the training data can be generated, such that the desirable gains in the computational efficiency after training can be achieved. Furthermore, the EOF analysis and Sparse GP model is undertaken using Python without a graphical user interface. To make the model more accessible for industry users, a simple modeling package with instructions for how to best derive low- and high-fidelity results and how to use the model could be developed, hence advancing the method from theory to more practical applications.

After the prediction of the inundation extent, the next natural step for the LSG model is to extend the methodology to predict other parameters such as water depth and discharge. This is important, as only predicting the inundation extent can misrepresent the severity of a flood (Hunter et al., 2007). The MIKE 21 hydrodynamic model already simulates these parameters but reconstructing continuous hydraulic variables using the EOF analysis is more complicated than reconstructing binary depth data. To reconstruct continuous hydraulic variables, boundary constraints on the EOF analysis may be required to avoid negative values, as suggested by Giordani and Kiers (2007). Alternatively, other dimension reduction techniques like Self-organizing Maps (Kohonen, 1982) or Auto-encoders (Hinton & Salakhutdinov, 2006) could be explored.

In this study, the LSG model is applied to a floodplain that is particularly flat and extensive, which is a challenging example to consider when relating differences between high- and low-fidelity model predictions. The methodology as described is not restricted to this floodplain, or only fluvial flood problems. In theory, the LSG model could be applied to any flood inundation problem, or to other similar problems, such as downscaling remotely sensed data.

## 6. Conclusion

Accurate predictions of the dynamic behavior of flood inundation extent are of great importance to operational flood risk management. Traditional methods based on high-fidelity hydrodynamic models are known to provide accurate results, but at high computational cost. This has led to the development of surrogate models that can reduce computational cost whilst still maintaining an acceptable level of accuracy. However, current surrogate models have difficulties in handling the high spatial-temporal dimensionality of flood inundation data. The hybrid LSG surrogate model proposed in this study addresses this challenge. By focusing on the dynamic behavior of the flood inundation extent, the LSG model goes beyond the normal application of emulator surrogate models which generally only predict the maximum inundation extents.

The hybrid model consists of a low-fidelity hydrodynamic model to capture the dynamic and spatial correlation of the flood inundation event and a Sparse GP model to improve the accuracy of the low-fidelity model. The hydrodynamic model results are decomposed through EOF analysis into EOF spatial maps and ECs temporal function. This enables the Sparse GP model to transform the low-fidelity ECs into high-fidelity ECs, whereafter the predicted high-fidelity ECs are used to reconstruct the dynamic inundation extent with improved accuracy without actually running a computationally heavy high-fidelity model.

The LSG model is evaluated on the flat and complex Chowilla floodplain using three different historic events. Compared to just using a low-fidelity model, the LSG model significantly improves predictions of the flood inundation extent, thereby showing the benefit of using Sparse GP models to correct the low-fidelity results. The LSG model achieved a POD above 0.76 and a RFA below 0.20 for the entire duration of the validation events compared to the results obtained using the high-fidelity model. Furthermore, if only the maximum inundation extent is considered, then a POD $> 0.99$ and an RFA $< 0.05$ are achieved, which demonstrates high prediction accuracy of the LSG model.

The LSG model shows a good overall ability to capture the dynamic behavior of flood inundation, but it tends to overpredict the peak inundation extent (e.g., 1%–6% for the case study considered). Regarding the timing, the predictions follow the patterns of the high-fidelity model predictions, and there is no general tendency for the timing of the peaks to be over- or under-predicted. Once trained, the LSG model reduces the computational demand to 39% of that of the original high-fidelity model for the selected case study.

In future studies, the trade-offs between model simplicity and computational efficiency need to be investigated. The low-fidelity model is the most computationally demanding part of the hybrid model, meaning a reduction

in the low-fidelity model complexity could lead to significant reduction in the computational time, but this is expected to degrade the accuracy of the hybrid model. Another aspect to consider is to extend the methodology to estimate flood parameters such as water depth or velocity. These parameters are simulated using hydrodynamic models and are highly relevant in flood and hazard estimation. A surrogate model should therefore be able to estimate these parameters to be a fully comparable alternative to a high-fidelity model. Finally, as the methodology is not dependent on the case study, the hybrid model is applicable to other flood inundation problems (e.g., urban flooding, storm surge) and applications (e.g., downscaling of remote sensing data). New applications would therefore shed further light on the potential of the LSG model.

## Appendix A: Historic Events for Training and Validation

The flood events used for training and validation of the Low-fidelity, Spatial analysis, and Gaussian Process (LSG) model is shown in Table A1 and Figure A1. Data to simulate the events is obtained from Bureau of Meteorology's online water data platform (Bureau of Meteorology, 2021) for the three inflow boundaries, Murray river (Station no. 426200), Mullaroo creek (Station no. 414211) and Lindsay river (Station no. 414212), and the downstream water level boundary for the Murray river (Station no. A4260512). All boundary data is recorded as daily mean values of both discharge and water level. However, some days only contain a recorded water level for an inflow boundary location. To address this issue, polynomial functions have been fitted to describe the relation between water level and discharge for days with both variables recorded. These functions are used to calculate an

**Table A1**
*Flood Events Simulated Using the High- and Low-Fidelity Models for Training and Validation of the LSG Model*

| Event no. | Start | End | Inflow scaling factor | Extended duration | Validation event |
|---|---|---|---|---|---|
| 1[a] | 15 August 2010 | 01 June 2011 | 1 | – | **Yes** |
| 2[a] | 01 July 2011 | 15 October 2011 | 1 | – | No |
| 3 | 01 March 2012 | 15 June 2012 | 1 | – | **Yes** |
| 4 | 20 June 2012 | 01 November 2012 | 1 | – | No |
| 5a | 01 July 2013 | 01 December 2013 | 3 | – | No |
| 5b | 01 July 2013 | 01 December 2013 | 4 | – | No |
| 5c[b] | 01 July 2013 | 01 December 2013 | 3 | x2 | No |
| 5d[b] | 01 July 2013 | 01 December 2013 | 4 | x2 | No |
| 6[a] | 01 July 2016 | 01 February 2017 | 1 | – | **Yes** |
| 7a | 01 November 2017 | 15 January 2018 | 3 | – | No |
| 7b | 01 November 2017 | 15 January 2018 | 4 | – | No |
| 7c | 01 November 2017 | 15 January 2018 | 5 | – | No |
| 7d | 01 November 2017 | 15 January 2018 | 6 | – | No |
| 7e[b] | 01 November 2017 | 15 January 2018 | 5 | x2 | No |
| 7f[b] | 01 November 2017 | 15 January 2018 | 6 | x2 | No |
| 8a | 01 September 2019 | 01 December 2019 | 3 | – | No |
| 8b | 01 September 2019 | 01 December 2019 | 4 | – | No |
| 8c | 01 September 2019 | 01 December 2019 | 5 | – | No |
| 8d | 01 September 2019 | 01 December 2019 | 6 | – | No |
| 8e[b] | 01 September 2019 | 01 December 2019 | 5 | x2 | No |
| 8f[b] | 01 September 2019 | 01 December 2019 | 6 | x2 | No |
| 9a | 01 November 2020 | 15 January 2021 | 3 | – | No |
| 9b | 01 November 2020 | 15 January 2021 | 4 | – | No |
| 9c | 01 November 2020 | 15 January 2021 | 5 | – | No |

**Table A1**
*Continued*

| Event no. | Start | End | Inflow scaling factor | Extended duration | Validation event |
|---|---|---|---|---|---|
| 9d | 01 November 2020 | 15 January 2021 | 6 | – | No |
| 9e[b] | 01 November 2020 | 15 January 2021 | 5 | x2 | No |

*Note.* Bold indicates to highlight the few events that are different from the others.
[a]Only data for the Murray River is available for the inflow boundaries. Linear interpolation is used for the other inflow boundaries. [b]Start and end dates reflect original dates of the event. Events are extended by the factor in the extended duration column.

estimated discharge, for days with missing discharge recordings. For days with neither water level nor discharge recorded, the daily values are found using linear interpolation.

For 3 of the flood events, inflow data is only available for the Murray river, see Table A1. The discharge in the Murray river is main source for the flooding and on average a factor ~790 and ~10 higher than the discharge in the Lindsay river and Mullaroo creek, respectively. The difference between these 3 events compared to the remaining events is therefore considered negligible.

As both the low- and high-fidelity models is run with the same boundary conditions, these adaptations of the boundary values do not affect the results of the LSG model in this paper.

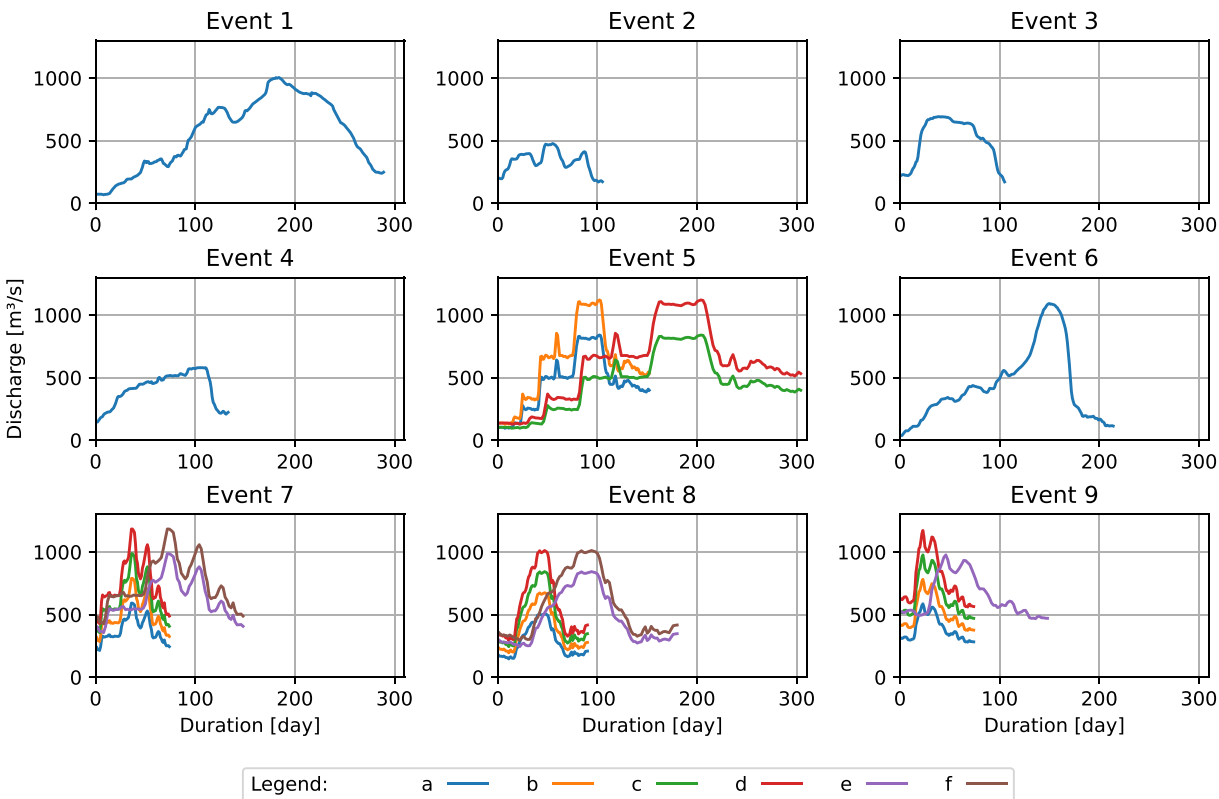


**Figure A1.** Inflow hydrographs for discharge in the Murray river during the historic and synthetic flood events. In the legend "a, b, … , f" refers to the event number in Table A1. Events without a letter corresponds to the "a" hydrograph.

## Appendix B: Evaluation Metrics

The evaluation metrics used in this paper can take a variety of values. In Table B1 is an overview of the possible values and what corresponds a good prediction.

**Table B1**
*Evaluation Metrics and Bounds for Values They Can Take*

| Metric | Bounds | Good prediction | Notes |
|---|---|---|---|
| relRMSE | [0, 1] | 0 | |
| relPeakValErr | [−1, 1] | 0 | Negative and positive value indicates an under- and over-prediction, respectively. |
| relPeakTimeErr-1 | [−∞, ∞] | 0 | Negative and positive value indicate the peak being early or late, respectively |
| relPeakTimeErr-2 | [−∞, ∞] | 0 | Negative and positive value indicate the peak being early or late, respectively |
| POD | [0, 1] | 1 | |
| RFA | [0, 1] | 0 | |

## Data Availability Statement

The Python code, MIKE 21 model results and boundary data, together with the data generated to create the results presented in this paper, are available at https://doi.org/10.26188/19100996.v3. The programming is performed using Python (version 3.9). All necessary dependencies are open-source libraries and stated in the import section of the code. Additionally, a GitHub repository has been created for sharing the code and future updates to the LSG model (https://github.com/nfraehr/Hybrid_LSG_model).

**Acknowledgments**
We thank the Murray–Darling Basin Authority for providing the hydrodynamic model for the Chowilla floodplain. N. Fraehr acknowledges support from The University of Melbourne via the Melbourne Research Scholarship, and W. 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

Aires, F., Papa, F., Prigent, C., Cretaux, J. F., & Muriel, B. N. (2014). Characterization and space-time downscaling of the inundation extent over the inner Niger delta using GIEMS and MODIS data. *Journal of Hydrometeorology*, *15*(1), 171–192. https://doi.org/10.1175/jhm-d-13-032.1

Aires, F., Venot, J. P., Massuel, S., Gratiot, N., Pham-Duc, B., & Prigent, C. (2020). Surface water evolution (2001-2017) at the Cambodia/Vietnam border in the upper Mekong delta using satellite MODIS observations. *Remote Sensing*, *12*(5), 800. https://doi.org/10.3390/rs12050800

Alvarez, F., & Pan, S.-Q. (2016). Predicting coastal morphological changes with Empirical Orthogonal Function method. *Water Science and Engineering*, *9*(1), 14–20. https://doi.org/10.1016/j.wse.2015.10.003

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 modelling. *Water Resources Research*, *51*(8), 5957–5973. https://doi.org/10.1002/2015wr016967

Avendaño-Valencia, L. D., Chatzi, E. N., Koo, K. Y., & Brownjohn, J. M. W. (2017). Gaussian Process time-series models for structures under operational variability [methods]. *Frontiers in Built Environment*, *3*, 69. https://doi.org/10.3389/fbuil.2017.00069

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

Bauer, M., van der Wilk, M., & Rasmussen, C. E. (2017). Understanding probabilistic sparse Gaussian Process approximations. *Advances in Neural Information Processing Systems*, *29*, 1533–1541. https://doi.org/10.48550/arXiv.1606.04820

Brahim-Belhouari, S., & Bermak, A. (2004). Gaussian Process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, *47*(4), 705–712. https://doi.org/10.1016/j.csda.2004.02.006

Brahim-Belhouari, S., & Vesin, J. M. (2001). Bayesian Learning using Gaussian Process for time series prediction. (Cat. No. 01TH8563). In *Proceedings of the 11th IEEE signal processing workshop on statistical signal processing* (pp. 433–436). IEEE. https://doi.org/10.1109/ssp.2001.955315

Bureau of Meteorology. (2021). Water data online. Retrieved from http://www.bom.gov.au/waterdata/

Burt, D. R., Rasmussen, C. E., & van der Wilk, M. (2019). Rates of convergence for sparse variational Gaussian Process regression. In *International conference on machine learning* (pp. 862–871). PMLR. 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

Cawley, G., & Talbot, N. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*, 2079–2107. https://dl.acm.org/doi/10.5555/1756006.1859921

Chang, C. H., Lee, H., Kim, D., Hwang, E., Hossain, F., Chishtie, F., et al. (2020). Hindcast and forecast of daily inundation extents using satellite SAR and altimetry data with rotated Empirical Orthogonal Function analysis: Case study in Tonle Sap Lake Floodplain [Article]. *Remote Sensing of Environment*, *241*, 111732. https://doi.org/10.1016/j.rse.2020.111732

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

Devi, N. N., Sridharan, B., & Kuiry, S. N. (2019). Impact of urban sprawl on future flooding in Chennai city, India. *Journal of Hydrology*, *574*, 486–496. https://doi.org/10.1016/j.jhydrol.2019.04.041

DHI. (2019). MIKE flood. Retrieved from https://manuals.mikepoweredbydhi.help/2019/Water_Resources/MIKE_FLOOD_UserManual.pdf

ESRI. (2021). World imagery. Retrieved from https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9

Fernández-Godino, M. G., Park, C., Kim, N. H., & Haftka, R. T. (2017). Review of multi-fidelity models. arXiv preprint arXiv:1609.07196. https://doi.org/10.48550/arXiv.1609.07196

Fernández-Godino, M. G., Park, C., Kim, N. H., & Haftka, R. T. (2019). Issues in deciding whether to use multifidelity surrogates. *AIAA Journal*, *57*(5), 2039–2054. https://doi.org/10.2514/1.J057750

Ghosh, M., Singh, J., Sekharan, S., Ghosh, S., Zope, P. E., & Karmakar, S. (2021). Rationalization of automatic weather stations network over a coastal urban catchment: A multivariate approach [Article]. *Atmospheric Research*, *254*, 105511. https://doi.org/10.1016/j.atmosres.2021.105511

Giordani, P., & Kiers, H. A. L. (2007). Principal component analysis with boundary constraints. *Journal of Chemometrics*, *21*(12), 547–556. https://doi.org/10.1002/cem.1074

Golestani, M., & Sørensen, J. (2013). Empirical Orthogonal Function analysis of 2D current transects in the Fehmarn Belt (Vol. *5*). https://doi.org/10.1115/OMAE2013-10745

Gu, M., & Berger, J. O. (2016). Parallel partial Gaussian Process emulation for computer models with massive output. *Annals of Applied Statistics*, *10*(3), 1317–1347. https://doi.org/10.1214/16-aoas934

Hachino, T., & Kadirkamanathan, V. (2011). Multiple Gaussian Process models for direct time series forecasting [Article]. *IEEJ Transactions on Electrical and Electronic Engineering*, *6*(3), 245–252. https://doi.org/10.1002/tee.20651

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507. https://doi.org/10.1126/science.1127647

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.

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. In *Proceedings of the Institution of Civil Engineers-Water Management*, (Vol. 174(1), pp. 27–41). Thomas Telford Ltd. 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

Kim, H. I., & Han, K. Y. (2020). Linking hydraulic modeling with a machine learning approach for extreme flood prediction and response. *Atmosphere*, *11*(9), 987. https://doi.org/10.3390/atmos11090987

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. https://doi.org/10.1007/bf00337288

Leibfried, F., Dutordoir, V., John, S. T., & Durrande, N. (2021). A tutorial on sparse Gaussian Processes and variational inference. arXiv pre-print serverarxiv:2012.13962. https://doi.org/10.48550/arXiv.2012.13962

Lin, Q., Leandro, J., Wu, W. R., Bhola, P., & Disse, M. (2020). Prediction of maximum flood inundation extents with resilient backpropagation neural network: Case study of Kulmbach [Article]. *Frontiers of Earth Science*, *8*(8). 332. https://doi.org/10.3389/feart.2020.00332

Liu, H. T., Ong, Y. S., Cai, J. F., & Wang, Y. (2018). Cope with diverse data structures in multi-fidelity modeling: A Gaussian Process method [Article]. *Engineering Applications of Artificial Intelligence*, *67*, 211–225. https://doi.org/10.1016/j.engappai.2017.10.008

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. https://doi.org/10.48550/ARXIV.1909.01836

Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, *25*(8), 891–909. https://doi.org/10.1016/j.envsoft.2010.02.003

Malde, S., Wyncoll, D., Oakley, J., Tozer, N., & Gouldby, B. (2016). Applying emulators for improved flood risk analysis. *E3S Web of Conferences*, *7*, 04002. https://doi.org/10.1051/e3sconf/20160704002

Marques, W. C., Fernandes, E. H., Monteiro, I. O., & Möller, O. O. (2009). Numerical modeling of the Patos Lagoon coastal plume, Brazil. *Continental Shelf Research*, *29*(3), 556–571. https://doi.org/10.1016/j.csr.2008.09.022

Matthews, A. G. D. G., Van Der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., León-Villagrá, P., et al. (2017). GPflow: A Gaussian Process library using TensorFlow. *Journal of Machine Learning Research*, *18*(40), 1–6. Retrieved from http://jmlr.org/papers/v18/16-537.html

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

Murray-Darling Basin Authority. (2021a). 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. (2021b). Lower Murray. Retrieved from https://www.mdba.gov.au/water-management/catchments/lower-murray

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

Nicol, J., Frahn, K., Fredberg, J., Gehrig, S., Marsland, K., & Weedon, J. (2020). Chowilla icon site - floodplain vegetation monitoring 2019 interim report. Retrieved from https://pir.sa.gov.au/__data/assets/pdf_file/0005/360590/Chowilla_Icon_Site_-_Floodplain_Vegetation_Monitoring_2019_Interim_Report.pdf

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

Park, C., Haftka, R. T., & Kim, N. H. (2017). Remarks on multi-fidelity surrogates. *Structural and Multidisciplinary Optimization*, *55*(3), 1029–1050. https://doi.org/10.1007/s00158-016-1550-y

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. Retrieved from http://jmlr.org/papers/v12/pedregosa11a.html

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

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian Process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16. https://doi.org/10.1016/j.jmp.2018.03.001

Snelson, E., & Ghahramani, Z. (2006). Sparse Gaussian Processes using pseudo-inputs. *Advances in Neural Information Processing Systems*, *18*, 1257.

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

Titsias, M. (2009). Variational learning of inducing variables in sparse Gaussian Processes. In *Proceedings of the twelth International conference on Artificial Intelligence and statistics* (pp. 567–574). Proceedings of Machine Learning Research. Retrieved from http://proceedings.mlr.press/v5/titsias09a.html

Toal, D. J. J. (2015). Some considerations regarding the use of multi-fidelity Kriging in the construction of surrogate models. *Structural and Multidisciplinary Optimization*, *51*(6), 1223–1245. https://doi.org/10.1007/s00158-014-1209-5

Wu, W., May, R. J., Maier, H. R., & Dandy, G. C. (2013). A benchmarking approach for comparing data splitting methods for modeling water resources parameters using artificial neural networks. *Water Resources Research*, *49*(11), 7598–7614. https://doi.org/10.1002/2012wr012713

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 [Article]. *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

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

Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y. W., Morsy, M. M., & Behl, M. (2020). Training machine learning surrogate models from a high-fidelity physics-based model: Application for real-time street-scale flood prediction in an urban coastal community [Article]. *Water Resources Research*, *56*(10), e2019WR027038. https://doi.org/10.1029/2019wr027038

Zhang, Z., & Moore, J. C. (2015). Chapter 6—Empirical Orthogonal Functions. In Z. Zhang & J. C. Moore (Eds.), *Mathematical and physical fundamentals of climate change* (pp. 161–197). Elsevier. https://doi.org/10.1016/B978-0-12-800066-3.00006-1

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