Method Article

Python program for spatial reduction and reconstruction method in flood inundation modelling

Yuerong Zhou*, Wenyan Wu, Rory Nathan, Quan J. Wang

Department of Infrastructure Engineering, Faculty of Engineering and Information Technology, The University of Melbourne, Victoria, Australia

A B S T R A C T

Fast and accurate modelling of flood inundation has gained increasing attention in recent years. One approach gaining popularity recently is the development of emulation models using data driven methods, such as artificial neural networks. These emulation models are often developed to model flood depth for each grid cell in the modelling domain in order to maintain accurate spatial representation of the flood inundation surface. This leads to redundancy in modelling, as well as difficulties in achieving good model performance across floodplains where there are limited data available. In this paper, a spatial reduction and reconstruction (SRR) method is developed to (1) identify representative locations within the model domain where water levels can be used to represent flood inundation surface using deep learning models; and (2) reconstruct the flood inundation surface based on water levels simulated at these representative locations. The SRR method is part of the SRR-Deep-Learning framework for flood inundation modelling and therefore, it needs to be used together with data driven models. The SRR method is programmed using the Python programming language and is freely available from https://github.com/yuerongz/SRR-method.

- The SRR method identifies locations which are representative of flood inundation behavior in surrounding areas.
- The representative locations selected following the SRR method have sufficient flood data for developing emulation models.
- Flood inundation surfaces can be reconstructed using the SRR method with a detection rate of above 99%.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

DOI of original article: 10.1016/j.envsoft.2021.105112
* Corresponding author.
E-mail address: yuerongz@student.unimelb.edu.au (Y. Zhou).
Specifications table

| Subject Area: | Engineering |
|---------------|-------------|
| More specific subject area: | Environmental Engineering and Modelling – Flood inundation modelling |
| Method name: | Spatial Reduction and Reconstruction for flood inundation modelling (SRR) |
| Name and reference of original method: | N/A |
| Resource availability: | https://github.com/yuerongz/SRR-method |

Method details

**SRR Method**

Flood studies are of increasing focus in recent years due to the risk of potential economic and life losses, covering research topics such as global flood hazard simulation [1,5,15,16,23,26] and local ensemble flood forecasting [11–13,27]. Flood inundation models are one of the important tools that are widely applied for the assessment of flood risks. Many fast flood inundation models have been developed [3,4,6,8, 17,21,25,29,30] as the high computational burden of traditional two-dimensional (2D) hydrodynamic models limits their use in large scale or real-time flood modelling applications. Most of the fast models are developed based on a volume-filling strategy to distribute flood water across the modelling domain and assume constant water level in each region/cluster. Thus, they are not well suited to modelling the temporal dynamics of floods. Several studies [7,10,18,28] have found that artificial neural networks (ANNs), including modern deep learning (DL) models, are good at capturing the temporal evolution of floods in simulations, but are still inefficient in representing spatial extent of flood surfaces [31].

This article introduces the spatial reduction and reconstruction (SRR) method developed as part of the spatial reduction and reconstruction - deep learning (SRR-DL) framework for flood inundation modelling introduced in the *Environmental Modelling & Software* paper related to this article by Zhou et al. [31]. The SRR method consists of two major modules: SRR-RL and SRR-Reco. The SRR-RL module is used to identify representative locations (RLs) where water level information can be used to accurately represent flood inundation across the entire model domain. The SRR-Reco module is a flood inundation mapping tool which constructs the flood surface across the modelling domain based on water level information at RLs. The RL water level information needs to be simulated using data driven models such as deep learning (DL) models. Both modules use a function called SRR-Search, which searches for the local thalwegs (drainage paths) from a given starting location. An overview of the SRR method is presented in Fig. 1.

Auxiliary data required in the SRR method include digital elevation models (DEM) and the maximum inundation extent in the modelling domain. The maximum inundation extent can be generated using a two-dimensional (2D) hydrodynamic model or approximated by applying a very high water level over the DEM. The DEM is used to provide the basic topology for searching, and the maximum inundation extent is used to determine the starting locations for the thalwegs using the SRR-Search function. A simple resampling method is used to select starting locations from the maximum inundation extent. The SRR method is validated using a real-world river system - the Burnett River downstream of the Paradise Dam in Queensland, Australia.

1. **SRR-Search function**

The SRR-Search function is designed to search for the thalweg (drainage path) from each given starting point. Although there are several existing stream channel network extraction methods
Y. Zhou, W. Wu and R. Nathan et al. / MethodsX 8 (2021) 101527

available [9,14,20,22,24], these methods delineate the entire stream channel network in a given region, instead of identifying drainage path from a given starting point, which is required for the SRR method. In addition, these stream channel network extraction methods often do not have associated program tools that are freely available online. Therefore, the SRR-Search function is developed as part of the SRR method.

The starting locations for the SRR-Search function are determined from the maximum inundation extent. Each grid cell in the provided inundation extent are first classified into dry (0) or wet (1). Then, the values for each grid cell in the eight surrounding grid cells are summed up and saved in the assessing cell ($N_{wet}$). An example of the assessment of one grid cell is shown in Fig. 2. This process results in a new raster with $N_{wet}$ values. By finding grid cells with $N_{wet}$ smaller than four ($N_{wet} < 4$), the coordinates of the center of these cells are the starting locations which are used by SRR-Searching function.

The SRR-Search function takes the DEM and the identified starting locations where the searches start as inputs, then the thalwegs are delineated by searching for elevation drops. The search terminates in the termination zones; these are either the main river channel or the lower boundary of the model domain. Special termination values need to be used to mark the grid cells of termination

---

**Fig. 1.** Overview of the SRR method.
zones in the DEM. The outputs of the function are lines in *ESRI Shapefile* format, starting from the given starting locations to end points located along the edges of the termination zones.

The computational algorithm of the SRR-Search function consists of three components: thalweg search, error handling and thalwegs selection. The thalweg search algorithm searches for thalwegs from given starting locations. Since the thalwegs found by the thalweg search algorithm can have issues such as early termination and having ‘thalweg knots’ (see section on error handling below), the error handling algorithm is developed to correct these thalwegs. After error handling, the thalwegs are similar to the stream channel networks in the targeted area but are all connected to the given starting locations. Finally, a thalwegs selection algorithm is used to reduce the total number of thalwegs and make sure only the most representative thalwegs are selected. The pseudocode for the SRR-Search function is presented in Fig. 3. The three components of the SRR-search function are presented below.

1.1 Thalweg search

The pseudocode for the thalweg search algorithm is presented in Fig. 4. The manner in which the algorithm works to find the thalwegs is illustrated in Fig. 5. The search algorithm first takes the coordinates of a given starting location and extracts the elevation information from the 9-by-9 grid cell area surrounding this starting location from the DEM. This 9-by-9 DEM block is then turned into a 3-by-3 block using the minimum resampling function. After that, the zonal minimum direction is defined by taking the direction from the center cell to the cell with the minimum value in the 3-by-3 block, as illustrated in step 4 of Fig. 5. Following the zonal minimum direction, three neighboring grid cells next to the starting location are selected. The selection of the three cells based on the zonal minimum direction follows the rule described in Fig. 6. The elevations in these three cells are then extracted from the DEM. The grid cell with the minimum elevation is the next point in the
ththalweg. By marking the current starting cell as ‘searched’ and taking the new point as the next starting location, the above process is repeated until meeting one of the following three stopping criteria: (1) all surrounding grid cells have been searched, (2) there exist at least one surrounding grid cell belonging to one of the termination zones, and (3) there exist at least one surrounding grid cell reaching the boundary of the model domain.

After the above process is applied to all the starting locations, the resulting collection of thalwegs are saved and passed on to the error handling algorithm.

1.2 Error handling

A collection of thalwegs is obtained for all given starting locations using the thalweg search algorithm. However, these thalwegs have two issues, which are illustrated in Fig. 7. The first issue is that some of the thalwegs are terminated due to the first stopping criterion mentioned above, and therefore these thalwegs do not end in any termination zones or at the model domain boundary as shown in Fig. 7(a). These thalwegs need to be corrected so they lead to a termination zone or the model boundary. The second issue is that although some thalwegs terminate in termination zones, there are ‘thalweg knots’ where the thalweg meets itself when the search passes through elevation depressions or flat areas in the model domain, as shown in Fig. 7(d). These thalwegs also need to be corrected.

To solve the two issues with the thalwegs, the error handling algorithm is used. There are two solutions for issue 1 as shown in Fig. 7: (1) if there exists another thalweg which passes at least one point in the current thalweg, the current thalweg is connected to the segment of the other thalweg at the shared point; (2) if the condition mentioned in the first solution is not met, the thalweg search algorithm is used again and the re-starting location is set to be the point in the current thalweg which has the lowest elevation, so a new segment of the thalweg can be identified. For issue 2, the ‘thalweg knot’ is replaced by a segment of corrected thalweg. The corrected thalweg is a shortcut through the points in the current thalweg. Starting from the most downstream point of the ‘thalweg knot’, each point is connected to the most upstream point which appears in the eight surrounding grid cells of the current point. All the points between these two points are deleted from the current thalweg. All identified thalwegs are checked and corrected repeatedly until there are no issues in all thalwegs. The pseudocode for the error handling algorithm is presented in Fig. 8.

1.3 Thalwegs selection

The thalweg selection algorithm is used to finally select the most representative thalwegs out of the corrected thalwegs. The pseudocode for this algorithm is presented in Fig. 9. First, the thalwegs that consist of less than 20 points are eliminated. These thalwegs do not play a significant role in representing the flood inundation because they are too short. Then, thalwegs ending outside of the main river channel are also deleted as they are not considered in the SRR method. The remaining thalwegs are grouped based on their end points. To identify the representative thalwegs from each group, a parameter named SRR thalwegs group selection ratio ($R_{SRR-thalweg}$) is used. The default
Fig. 5. Thalweg search algorithm graphic example.
Fig. 6. Selections of the three cells based on different zonal minimum directions.

selection ratio is 1/200. This ratio determines the total number of representative thalwegs to be selected for each group, as shown in Eq. (1).

\[ N_{\text{thal}} = 1 + \text{ceiling}(R_{\text{SRR-thalweg}} \times (N_{\text{thal-all}} - 1)) \]  

where, \( N_{\text{thal}} \) is the number of representative thalwegs in the current group, \( N_{\text{thal-all}} \) is the total number of thalwegs in the group, \( \text{ceiling()} \) is a function that takes the smallest integral value that is greater than or equal to the number provided.

In each group of thalwegs, the longest thalweg is first selected. Then, starting from the first and last points of the longest thalweg, the DUPLEX method developed by Kennard and Stone [19] is used to select among the starting points of the remaining thalwegs. Then, the thalwegs, the starting points of which are selected using the DUPLEX method, are identified as representative thalwegs in the group, which are saved as lines in ESRI Shapefile format called SRR thalwegs.

1.4 Applying SRR-Search function to define mainstream centroid line

Apart from identifying SRR thalwegs, the SRR-Search function is also used to define the mainstream centroid line. The starting point of the river centroid line needs to be manually selected and used as an input to the function. Usually, the location can be selected from the upstream inflow boundary of the model. The DEM of the model domain can be aggregated to a higher resolution to make the output line smoother. The aggregation rate is decided according a trial-and-error process to delineate a desirable river channel centroid line.

2. SRR-RL module

The SRR-RL module is designed to find the key locations in the model domain where the water levels will be modelled, referred to as the representative locations (RLs) in this paper. The RLs are
Fig. 7. Examples of the two issues in thalwegs constructed using the thalweg search algorithm and the solutions to these issues.
selected within the river channel where there are sufficient data with which to develop data-driven models. The total number of RLs is significantly smaller than the total number of grid cells in the model domain. Therefore, the number of models needed to model the flood inundation for the entire model domain is significantly reduced. This will not only increase model development efficiency, but also simplify the applications of data-driven emulation model to flood inundation modelling.

The pseudocode of the SRR-RL module is presented in Fig. 10. Two types of RLs are selected. The first type is RLs-CL, which are RLs selected along the centroid line of the main river channel at regular intervals. The interval between the points is one of the parameters in the SRR-RL module. This parameter is determined based on the length of sub-reaches of the river channel. Since these sub-reaches are assumed to be straight, the more meandering a river is, the more RLs-CL will be identified.

The second type of RLs is RLs-Side, which are selected on both sides of the main river channel where flood water exits the river and eventually flows to floodplains. These locations are identified using the SRR-Search function. First, the last points of SRR thalwegs are extracted to form the group of thalweg-RLs. Then, the RLs-Side selection is performed among thalweg-RLs using the DUPLEX method developed by Kennard and Stone [19] with a selection ratio. The total number of RLs-Side should be around twice the number of RLs-CL to have a similar density of RLs selected on both sides of the river channel. At last, RLs-CL and RLs-Side are combined and saved as an output of the SRR-RL module.
**Inputs:**  DEM file, water levels at RLs, RLs coordinates, SRR thalwegs, lower boundary water level (if available)

1. Categorize selected SRR thalwegs into:
   a) Thalwegs connected to existing RLs
   b) Thalwegs not connected to any RLs
2. Assign water levels to all points along SRR thalwegs in category (a)
3. Apply 2D linear interpolation to get water levels at the end points of SRR thalwegs in category (b) with
   i. Water levels at RLs
   ii. Water levels along SRR thalwegs of category (a)
   iii. Water level at the lower boundary (at the current time, if provided)
4. Apply 2D linear interpolation to get water levels in the model domain with
   i. Water levels obtained from step 3
   ii. Water levels at RLs
   iii. Water levels along SRR thalwegs in category (a)
   iv. Water level at the downstream boundary (at the current time, if provided)
5. Compare reconstructed flood surface with DEM, and remove areas where water levels being lower than elevations

**Outputs:**  Flood inundation surface

---

*Fig. 11.* Pseudocode of SRR-Reco module.
3. SRR-Reco module

The SRR-Reco module, representing SRR Reconstruction module, is developed to reconstruct the flood inundation water surface from water levels at RLs (e.g. simulated using data-driven models). The pseudocode of the SRR-Reco module is presented in Fig. 11. First, the SRR thalwegs constructed by SRR-Search function are categorized into (a) thalwegs connected to RLs, and (b) thalwegs not connected to any RLs. Then, for thalwegs in category (a), the water levels at locations of all the points along the thalweg are set to be the same as the water level modelled at their RLs. For thalwegs in category (b), the end point of each thalweg is retrieved and the water levels at these points are interpolated using a 2D linear interpolation method [2] based on the water levels at (1) RLs, (2) locations along thalwegs of category (a), and (3) the lower boundary of the modelling domain. In the next step, the water levels obtained from the last step are included in the 2D linear interpolation together with the water levels at the three types of locations mentioned above, to obtain the flood water surface for the entire model domain. The water surface is then compared with the DEM so areas where the water level is lower than the elevation are eliminated. At last, the adjusted water surface is saved as an output from the SRR-Reco module in geo-tiff format. The resolution of the reconstructed flood inundation surfaces is determined by the resolution of the DEM used.

Method validation

The SRR method is validated within the SRR-DL framework for flood inundation modelling in the *Environmental Modelling & Software* paper related to this article. In that work, this Python program of the SRR method is used to parameterize the flood inundation in a real-world river system – the Burnett River reach downstream of the Paradise Dam in Queensland, Australia. Overall, the SRR method is proven to be an efficient tool in the SRR-DL framework for rapid flood inundation modelling. Some limitations of this method include reduced accuracy in locations (1) that are away from the selected RLs and (2) where overbank flood water can create ‘dead storages’ during the flood receding period. For detailed discussion on the performance of the method as part of the SRR-DL framework, please refer to the related study by Zhou et al. [31].

Direct submission or co-submission

Co-Submission. Refers to doi.org/10.1016/j.envsoft.2021.105112.

Acknowledgements

We thank Hydrology and Risk Consulting (HARC) for providing the TUFLOW 2D hydrodynamic model configuration and SunWater for their permission to use the Burnett River as a case study. Wenyan Wu acknowledges the support of the Australian Research Council via the Discovery Early Career Researcher Award (DE210100117). Also, for a range of open source software and libraries used in this study, including Python programming language (version 3.7), Python packages (Numpy, Scipy, Fiona), QGIS, and Geospatial Data Abstraction Library (GDAL), we would like to thank the developers and the open source community for all contributions.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Full responsibility for the editorial process for this article was delegated to Editor-in-Chief of MethodsX, Régis Braucher.

References

[1] G.D. Baldassarre, F. Nardi, A. Annis, V. Odongo, M. Rusca, S. Grimaldi, Brief communication: comparing hydrological and hydrogeomorphic paradigms for global flood hazard mapping. Nat. Hazards Earth Syst. Sci. 20 (5) (2020) 1415–1419.
P.D. Bates, A. De Roo, A simple raster-based model for flood inundation simulation, J. Hydrol. 236 (1–2) (2000) 54–77.

P.D. Bates, M.S. Horritt, T.J. Fewtrrel, A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation simulation, J. Hydrol. 387 (1–2) (2010) 33–45.

P.D. Bates, J. Neal, C. Sampson, A. Smith, M. Trigg, Progress toward hyperresolution models of global flood hazard, in: Risk Modeling for Hazards and Disasters, Elsevier, 2018, pp. 211–232.

A. Bernini, M. Franchini, A rapid model for delimiting flooded areas, Water Resour. Manag. 27 (10) (2013) 3825–3846.

L.C. Chang, M. Amin, S.-N. Yang, F.J. Chang, Building ANN-based regional multi-step-ahead flood inundation forecast models, Water 10 (9) (2018) 1283.

L.C. Chang, H.Y. Shen, F.J. Chang, Regional flood inundation nowcast using hybrid SOM and dynamic neural networks, J. Hydrol. 519 (2014) 476–489.

J. Chorowicz, I. Ichoku, S. Riaza, Y.J. Kim, B. Cervelle, A combined algorithm for automated drainage network extraction, Water Resour. Res. 28 (5) (1992) 1293–1302.

H. Chu, W. Wu, Q. Wang, R. Nathan, J. Wei, An ANN-based emulation modelling framework for flood inundation modelling: application, challenges and future directions, Environ. Model. Softw. (2019) 104587.

H. Cloke, F. Pappenberger, Ensemble flood forecasting: a review, J. Hydrol. 375 (3–4) (2009) 613–626.

D. Demeritt, H. Cloke, F. Pappenberger, J. Thelen, J. Bartholmes, M.H. Ramos, Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting, Environ. Hazards 7 (2) (2007) 115–127.

M. Gomez, S. Sharma, S. Reed, A. Mejia, Skill of ensemble flood inundation forecasts at short-to-medium-range timescales, J. Hydrol. 568 (2019) 207–220.

R.A. Heine, C.L. Lant, R.R. Sengupta, Development and comparison of approaches for automated mapping of stream channel networks, Ann. Assoc. Am. Geogr. 94 (3) (2004) 477–490.

Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, ..., S. Kanae, Global flood risk under climate change, Nat. Climate Change 3 (9) (2013) 816–821.

J. Hoch, M.A. Trigg, Advancing global flood hazard simulations by improving comparability, benchmarking, and integration of flood inundation models, Environ. Res. Lett. (2018).

J.M. Johnson, D. Munasinghe, D. Eyelade, S. Cohen, An integrated evaluation of the national water model (NWM)–Height above nearest drainage (HAND) flood mapping methodology, Nat. Hazards Earth Syst. Sci. 19 (11) (2019) 2405–2420.

S. Kabir, S. Patidar, X. Xia, Q. Liang, J. Neal, G. Pender, A deep convolutional neural network model for rapid prediction of fluvial flood inundation, J. Hydrol. (2020) 125481.

R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1) (1969) 137–148.

B. Lashermes, E. Foufoula-Georgiou, W.E. Dietrich, Channel network extraction from high resolution topography using wavelets, Geophys. Res. Lett. 34 (23) (2007).

J. Neal, G. Schumann, P. Bates, A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas, Water Resour. Res. 48 (11) (2012).

P. Passalacqua, P. Tarolli, E. Foufoula-Georgiou, Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape, Water Resour. Res. 46 (11) (2010).

C.C. Sampson, A.M. Smith, P.D. Bates, J.C. Neal, L. Affieri, J.E. Freer, A high-resolution global flood hazard model, Water Resour. Res. 51 (9) (2015) 7358–7381.

P. Soille, C. Gratin, An efficient algorithm for drainage network extraction on DEMs, J. Vis. Commun. Image Repres. 5 (2) (1994) 181–189.

J. Teng, J. Vaze, D. Dutta, S. Marvanek, Rapid inundation modelling in large floodplains using LiDAR DEM, Water Resour. Manag. 29 (8) (2015) 2619–2636.

P.J. Ward, B. Jongman, P. Salamon, A. Simpson, P. Bates, T. De Groeve, M.A. Trigg, Usefulness and limitations of global flood risk models, Nat. Clim. Change 5 (8) (2015) 712–715.

W. Wu, R. Emerton, Q. Duan, A.W. Wood, F. Wetterhall, D.E. Robertson, Ensemble flood forecasting: current status and future opportunities, Wiley Interdiscip. Rev. Water (2020).

S. Xie, W. Wu, S. Mooser, Q. Wang, R. Nathan, Y. Huang, Artificial neural network based hybrid modeling approach for flood inundation modeling, J. Hydrol. (2020) 125605.

X. Zheng, D.R. Maidment, D.G. Tarboton, Y.Y. Liu, P. Passalacqua, GeoFlood: large-scale flood inundation mapping based on high-resolution terrain analysis, Water Resources Research 54 (12) (2018) 10,013-010,033.

X. Zheng, D.G. Tarboton, D.R. Maidment, Y.Y. Liu, P. Passalacqua, River channel geometry and rating curve estimation using height above the nearest drainage, JAWRA J. Am. Water Resour. Assoc. 54 (4) (2018) 785–806.

Y. Zhou, W. Wu, R. Nathan, Q.J. Wang, A rapid flood inundation modelling framework using deep learning with spatial reduction and reconstruction, Environ. Model. Softw. (2021) 105112.