Exercise 4 - Probabilistic hazard mapping and uncertainty

This exercise is part of series designed to teach students how floodplain inundation can be simulated by numerical models and how flood risk maps can be produced from simulation results within the KULTURisk methodology framework. In this case the flood inundation model is LISFLOOD-FP (hereafter lisflood) however most two-dimensional hydraulic models can be used in much the same way. This exercise is the 4th in the series. It looks at alternative ways of representing flood risk and begins to explore how uncertainty can be estimated and communicated using hazard maps.

This exercise follows a simplified methodology of that presented in the following paper and it is recommended that you read this before the exercise and refer to it throughout the exercise.

Neal, J., Keef, C., Bates, P., Bevan, K., Leedal, D., 2013. Probabilistic flood risk mapping including spatial dependence. *Hydrological processes*, **27**, pp. 1349 – 1363

Task 1: Probabilistic hazard map for the flood river valley

Background information:

Data from numerical flood modelling simulations can be used to create flood hazard maps in many ways. Commonly, a deterministic approach is taken, as carried out in Exercises 2 and 3. Here a flood hazard map was produced based one set of results which represent a flood of a certain size. In the case of Exercises 2 and 3, the flood extent was representative of a flood expected to occur approximately once every 100 years. If suitable data exists, a probabilistic approach can be taken instead where a map is produced showing the annual probability of flooding is shown for each cell. If created using a flood modelling program then many simulations are run with a wide range of water input conditions representing those that might be expected over a long period of time. The results of all of these simulations are then combined to show the annual probability of each cell being inundated. In their paper, Neal et al. (2013) used stage height data collected over a period of 32 years, to produce a probability of inundation map for Carlisle, England, covering up to 1 in 1000 year events. They used the relationship between stage-height and recurrence period to create numerous lisflood input files with a wide range of flow rates. To predict the inundation dynamics over a longer period of time (1000 years) than field recorded data was available for, a statistical model was fitted to all available field data above a set “extreme event” threshold and used to predict the typical recurrence period of more extreme flow events. This statistical model was then used to create a simulated dataset comprising all the maximum flow rates which were above the extreme threshold which could be expected over a period of 1000 years on the main river channel and its tributaries (Figure 1), with realistic spatial dependence between the channel and tributaries (this will be explained further in task 2). These flow rates were them fed into lisflood to predict 1000 years of flooding on the river.

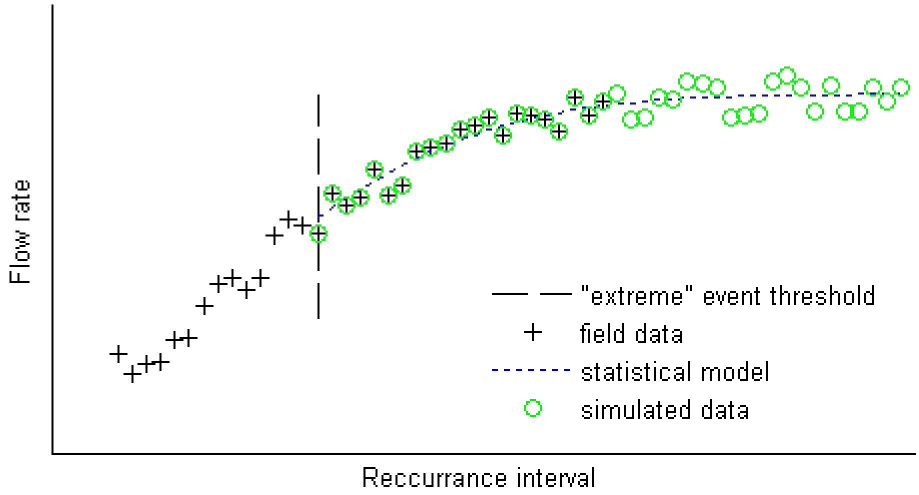

Figure 1 Illustration of how a statistical model can be fitted to field data recorded over a short period of time and then used to create a simulated dataset representing a longer period of time

Data provided:

Flow-rate data was recorded in the field for the main channel and its two tributaries for the fictional Flood River Valley over a period of 26 years (these tributaries were not considered in previous exercises). A statistical model was fitted to this flow-rate data to predict extreme events and 1000 years of events were simulated from the model. The 100 largest events (based on flow on the main channel) were simulated in lisflood and the results will be used to provide the best estimate using the available field data of the annual probability of flooding in the Flood River Valley.

Maximum water depth data for these 100 lisflood simulations are provided for this task (output as res.max ascii grid files)[1]. These consist of 48 x 76 cell grids of data with the value in each cell representing the maximum water depth predicted during the simulation. These have been imported into a MatLab and Excel file for ease of analysis. In the case of the MatLab file these are contained in a structure variable called water_depthBE containing 100 matrices called max1 to max100. The Excel file water_depthBE.xls contains 100 worksheets, each of which represents one maximum water depth file.

Use the maximum water depth data provided to create a probabilistic flood map for the Flood River Valley

This should be a single 48 x 76 grid of values between zero and 100 representing the number of times over the 100 simulations that each cell has been inundated. This can be converted into a grid of the probability each year of a particular cell being inundated by dividing by 1000.

With a little knowledge of MatLab or Excel you should be able to identify how many times each cell has been inundated over the course of the 100 simulations. However, if you are unsure of how to do this then

[1] These water depths files were actually produced by scaling the simulated dataset of flow rates from Neal et al (2013) to give suitable amount of flooding in the flood river valley.

some hints are provided below and if these are not sufficient then MatLab code/Excel formula to perform this task are provided in the answers document. Note: an error about file formats and extensions will pop up when you open these excel files (because they were created by MatLab); just hit "Yes" to open them anyway.

- Hints for MatLab:

Use a for loop to scroll through the 100 water depth files (`water_depthBE.(['max',num2str(xx)])`), identify all the cells which have been inundated (`~=0`) and sum the results to represent the probability

The function "imagesc" will draw a colour coded image of you grid.

- Hints for Excel:

The function "COUNT" can be used to identify any cells with numeric data in and all sheets can be selected using the syntax "Sheet1:Sheet100!"

Conditional formatting can be used to colour-code your worksheet to create a hazard map

Task 2: Investigate the effect of spatial dependence

Background information:

In addition to creating a probabilistic hazard map, Neal et al. (2013) also investigated the effect of the spatial dependence of flow rates on the tributary rivers in Carlisle as well as the uncertainty in their predicted flood probabilities. In task 2 we will look at spatial dependence.

Commonly when creating hazard maps, channels and their main tributaries are either considered separately or the flows on all are expected to be equally extreme. For example, a hazard map will be created representing the case of a 1 in 100 year flow rate on the main channel and its tributaries at the same time. In reality, although there will be a relationship between the flows on individual channels, this relationship is unlikely to be perfect due to difference in rainfall patterns, run-off generation etc. This is the "spatial dependence" considered by Neal et al. (2013). In addition to reproducing the relationship between stage height and recurrence interval on individual channels, the statistical model which Neal et al. (2013) fitted to the stage height data also aims to replicate the relationship between stage heights on all three channels, i.e. how likely is it that the stage height of tributary 1 is high when then stage height of the main channel is high. This meant that the probability map they created incorporated a realistic spatial dependence between flows on the different channels, rather than representing a worst case scenario situation where each channel experiences its most extreme flow rates simultaneously. Neal et al. (2013) compared the extent of flooding caused by a 1/100 year event using realistic spatial dependence (i.e. using their best estimate probability map) with that simulated assuming 1/100 year events on all three channels. They found that as expected, using the 1/100 year flow rates on all of the channels created more flooding than that predicted for a 1 in 100 event using the probabilistic map with realistic spatial dependence (Figure 2).

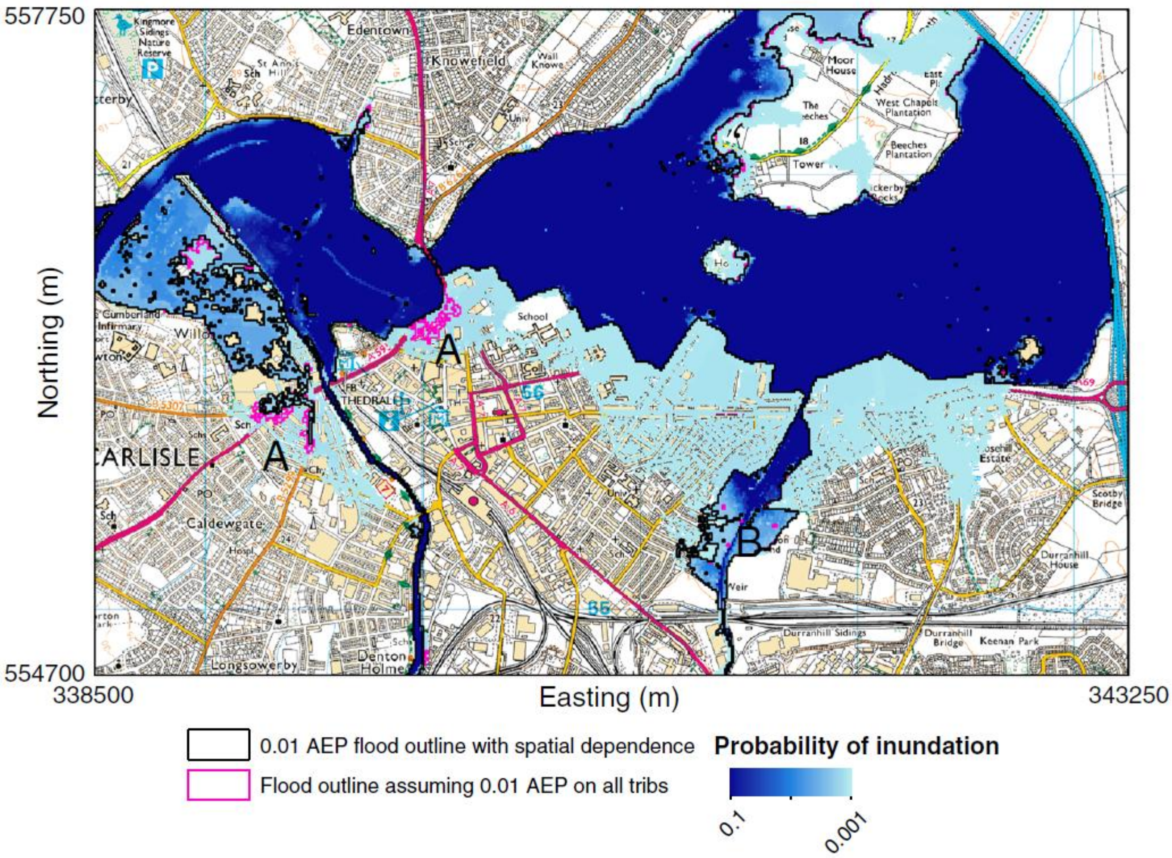

Figure 2 from Neal et al. (2013)

Data provided

The flow rates which were input to lisflood to simulate the 100 flood events for task 1 have been provided in the Excel file best_estimateQs.xls and the MatLab variable best_estimateQs. In each case there are 3 columns of data consisting of the maximum flow rate used for the main channel and its two tributaries in each of the 100 best estimate simulations. The first column contains flux values for the main channel and the second and third columns for the 1st (south of main channel) and 2nd (north of main channel) tributaries respectively. All of the files necessary to run a simulation in lisflood for the Flood River Valley, including channel information for the main channel and its two tributaries are also provided in the folder named lisflood_data.

Use lisflood to simulate an event with a 1 in 100 year flow rate on all three channels and compare the extent of flooding to that expected for the 1 in 100 year flood predicted using all 100 simulations from task 1 which include realistic spatial dependence

a) **Simulate the "worst case scenario" 1/100 year flood event in lisflood and identify the extent of flooding**

- Find the 10th largest flow rate predicted for each of the channels. This is the flow rate which would be equalled or exceeded 10 times in 1000 years, so flooding should be expected to be at least this bad once every 100 years; hence it is commonly termed a 1 in 100 year flood. Put another way it has an annual probability of exceedance (AEP) of 10/1000 or 0.01. Hint: in both MatLab and Excel there are functions named "sort".

- Open the file flood_valley_worst.bdy in your favourite text editor (this file is provided in the folder lisflood_data). This is a .bdy input file for lisflood, representing input hydrographs for the main channel (“channel”) in the flood valley and its two tributaries (trib1 and trib2). For each channel there are 3 rows of data which tell lisflood the flow rate at 3 times (in seconds) during the simulation. Flow rates start at 0 m^3s^{-1} at zero seconds, rise linearly to a maximum at 50,000 seconds and then drop back to 0 m^3s^{-1} at 100,000 seconds, the end of the simulation. Modify this file so that the 1 in 100 year flow rates identified in part a) are put in place of the flow rate values at 50,000 seconds (currently set to 10 m^3s^{-1}) and then simulate this event in lisflood. (See exercise 1 or 2 if you do not know how to do this).

- Import the maximum water depth file (res.max) from the new simulation into MatLab[2] or open the file in Excel[3]. Create a new MatLab variable or Excel worksheet which has values of 1 for cells within the “worst case scenario” 1 in 100 year flood extent and values of 0 outside of this. Hint: in MatLab use the criteria x=y>0 and in Excel the function “COUNTIF” can be used to identify any cells which satisfy a specified criteria (e.g. “>0”)

b) **Identify the 1/100 year flood extent outline predicted by the 100 simulations**

As described above, the 1 in 100 year flood is one which is expected to be equalled or exceeded once in every 100 years, or 10 times in every 1000 years. Therefore the 1 in 100 year flood extent predicted using the results from all 100 simulations will include all of the cells in the probabilistic flood map created in task 1 with values of greater or equal to 10.

Create a new MatLab variable or Excel worksheet which has values of 1 for cells within the best estimated 1 in 100 year flood extent and values of 0 outside of this. Hint: in MatLab use the criteria x=y>9 and in Excel the function “COUNTIF” can be used to identify any cells which satisfy a specified criteria (e.g. “>9”)

c) **Compare the two flood extents**

Add the grid representing the worst case scenario 1/100 year flood outline to that representing the best estimate multiplied by 10. This should give you a grid with values of zero where flooding is never expected, 11 for cells predicted to be inundated during a 1 in 100 year flood using both methods, 1 where only the worst case method predicts flooding and 10 where only the best estimate method predicts flooding

How much extra flooding is caused by an inaccurate representation of the spatial dependence between flow rates the main channel and its tributaries?

---

[2] E.g. use the function provided “ascii_reader.m”: [dem, ncols, nrows, xllcorner, yllcorner, cellsize] = ascii_reader ('C:\MyFolder\MyFile.asc') or the MatLab function dlmread (see MatLab help files for more information).
[3] To open in Excel use File > Open and ensure “All files (*.*) is selected as the file type. Then choose “Delimited” and Tab or Space for the delimiter.

Task 3: estimate uncertainty:

Background information

As described above, statistical models can be fitted to field data to extrapolate the event-size recurrence-interval relationship to longer time periods allowing the realistic modelling of extreme flood events. However, the output from the statistical model will depend strongly on the field data to which it is fitted and these field-collected data points can only ever represent a sample of the real event-size recurrence-interval relationship due to the relatively short time period over which data is collected.

Bootstrapping allows estimation of the errors introduced from having to fit the statistical model to a sampled dataset. To estimate the uncertainty from their short record of stage-height data, Neal et al. (2013) created not one, but 100 probability hazard maps in total. The first map, created using the methodology described above utilised the whole field collected stage-height dataset. The other 99 maps were produced using an identical methodology, but starting from 99 different sub-sets of stage-height data created by sampling from the original complete data set. This allowed estimation of the variation in final calculated probabilities which results from having to use an incomplete or short record. Neal et al. (2013) produced a hazard map which indicated for each cell, the number of times over all 99 sets of simulations that it was predicted to be inundated by a 1 in 100 year flood. On top of this they highlighted the contours identifying the 10th and 90th percentile of the flood extent such that the area between these two contours represented the range of flooding predicted by 80% of the sets of simulations and as such could be considered as the 80% confidence interval (Figure 3). Put another way this indicates that in 80% of the sets of simulations, the extent of flooding predicted for a 1 in 100 year flood fell between these two contours.

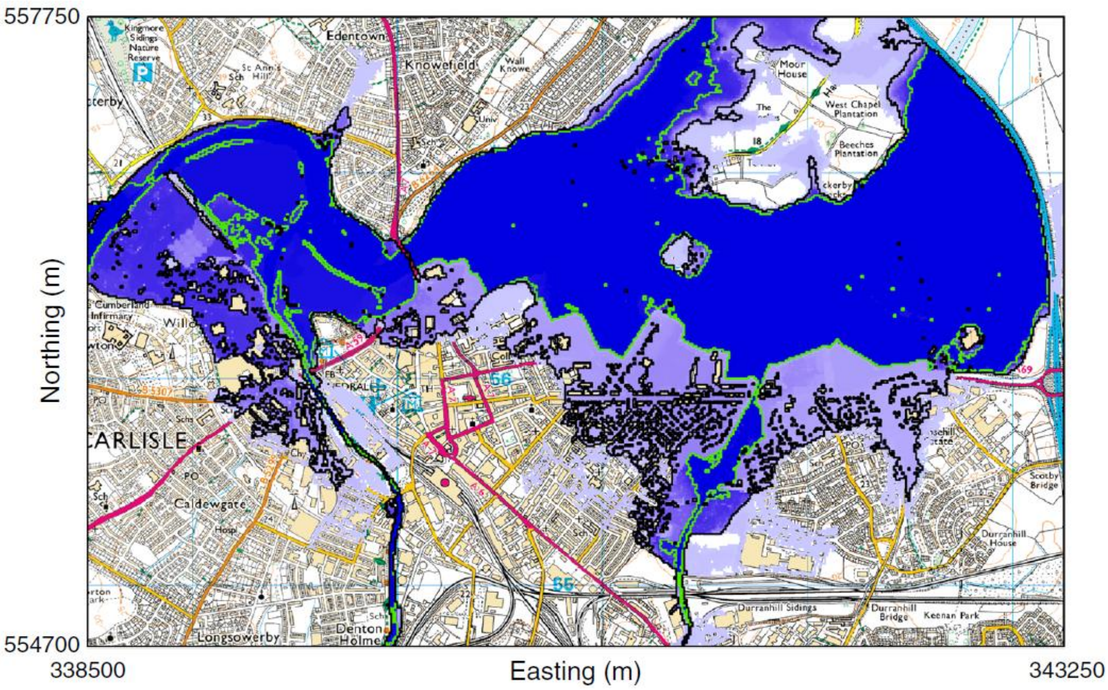

Figure 3. Copy of figure 14 from Neal et al., (2013). Colour indicates the number of times during the 100 simulations that an area is expected to be flooded during a 1 in 100 year flood event, darkest blue representing flooding in all simulation sets and lightest blue representing flooding in just one simulation set. The black line indicates the 90th percentile of the extent data, and the green line the 90th percentile.

Data provided:

The stage-height dataset used to calculate the probabilistic hazard map for the Flood River Valley in Task 1 was re-sampled 99 times and an identical methodology followed to produce a further 99 probabilistic flood hazard maps. From these 99 hazard maps, 99 grids representing the extent of the predicted 1 in 100 year flood were created using the same methodology as Task 2b. These flood extents have been provided in the format of a MatLab structure variable called AE10BS containing 99 logical matrices named BS2 to BS100 and an Excel file called AE01extentsBS with 99 worksheets. In both cases values of 1 indicate cells which would be inundated during the predicted 1 in 100 year flood event.

Create a map showing the uncertainty in the flood extent predicted for a 1 in 100 year flood event in the Flood River Valley - This should be in the form of a grid in which the cell value can be used to identify 1) those cells covered by the smallest 10% of flood extents, 2) the 80% confidence interval, 3) the maximum flood extent predicted and 4) those cells never expected to be inundated during a 1/100 year flood event

a) First, create a grid in which each cell value represents the number of times a cell has been included in the 100 year flood extent over all of the 99 simulation sets. Hint: in MatLab this can be carried out in a very similar to Task 1, in Excel it can be done in exactly the same way, using “COUNT” or “SUM”

b) Using the cell values in the grid created during part a) produce a new coloured grid which has different values for cells covered by the middle 80% of flood extents (values from part a) of 11 to 90), those covered by the smallest 10% of floods (values from part a) of 91 to 99), and those covered by only the largest 10% of flood extents (values from part a) or 1 to 10). I.e. a map which shows you the full range of flood extents predicted and highlights the 80% confidence interval

- Hint for MatLab: One way is to use a for loop to look at each cell in turn (e.g. `for xx=1:10; for yy=1:10`) and then use an if loop to specify a new value for the cell depending on its old value (`e.g. if temp(xx,yy)>3 && temp(xx,yy)<27; temp(xx,yy)=4`)

- Hint for Excel: Use a nested if statement to test for multiple values. For example the statement `=IF(B2=0,2,IF(B2<10,1,3))` produces a value of 2 if the value in cell B2 **is** zero, a value of 1 if the value is not zero but **is** less than 10, and a value of 3 if it is not zero and is also not less than 10 (i.e. if the value is in fact greater than 10)

Optional task for MatLab: Create a probabilistic flood hazard map indicating the 1 in 100 year flood extent and the uncertainty associated with this

For a more professional looking map the contour plotting function in MatLab can be used to create a probabilistic hazard map (from Task 1) which includes contour outlines for both the estimate of the 1 in 100 year flood extent (from Task 2b) and the uncertainty associated with this due to the short record (from Task 3).

First use the imagesc function to plot the probabilistic map. Then, holding the figure on, use the contour function to plot contours for the 1 in 100 year flood extent from the probability grid (10 in 1000 years) and the 10^{th} and 90^{th} percentile contours from the uncertainty grid. If you are unsure how to do this use the MatLab help files for more information, or check the answers document.

Acknowledgments

The development of this online course was supported by funding through the European Community's Seventh Framework Programme through the KULTURisk Project (ENV.2010.1.3.2-1) and work was carried out by colleagues at the University of Bristol, UK.