Understanding the effects of spatial rainfall patterns on flood events

Maths Foresees & Turing Gateway Study Group

Report of Environment Agency Challenge

A. Gregory[1], O. Bokhove and P. Sertyesilisik [2], M. Branicki and J. Vanneste[3], C. Cafaro[4], A. Chen[5], S. Patidar[6], A. Osojnik and O. Sheridan-Methven[7],

[1]Imperial College, [2]University of Leeds, [3]University of Edinburgh, [4]University of Reading, [5]University of Exeter, [6]Heriot-Watt University, [7]University of Oxford

March 26, 2018

Challenge presented by: Adam Baylis

Contents

1 Introduction and motivation

In flood modelling the rainfall distribution is often assumed to be uniform or is only varying on coarse scales, e.g. is constant on each $5 \times 5\text{km}^2$ grid, leading to a fairly even rainfall spread across catchments. Given that rainfall and especially summer rainfall tends to be more localised [13, 20], such coarse or uniform rain distributions are unrealistic and can lead to several problems in flood risk assessment. As an example, consider the observed rainfall over the Thames estuary for July 2014 in Fig. 1. This radar image shows extremely localized rainfall over Canvey Island that led to severe flooding in the surrounding area. If the rainfall had been spread evenly across the region, as is often assumed in flood modelling, it would have been less likely to lead to any flooding. The issue is that when the spatial distribution of rainfall is assumed or smoothed to be too uniform then it cannot capture more localized precipitation events. Consequently, the flood risk in some areas may be over- or underestimated, which leads to inefficient use of vital resources, with some communities missing out on necessary flood protection while other communities are overprotected. There are certain techniques that can be employed to alleviate this rainfall distribution problem, known as "continuous simulation" and "Monte Carlo analysis" [14, 18]. The question is whether the extra cost and time involved in applying these techniques is worth the added confidence. The challenge for mathematicians is to find an approach that gets some of the benefits of a more detailed approach, with only a minimal increase in the modelling costs.

Improved understanding of rainfall patterns can be crucial in improved forecasting of extreme flood events. Just as the temporal variations in these patterns are important in being able to forecast clustered flood events over only short time periods, potentially causing cumulative flood damage, combined spatial and temporal rainfall variations across the entire catchment can also lead to cumulative extreme flood damage. Consider for example heavy rainfall in the upper catchment of a river, causing a flood peak in a hydrograph propagating downstream. Consider either a new weather pattern or the same weather pattern causing another heavy rainfall event at a later time but further downstream near a city. When the latter event causes a second flood peak coalescing entirely or partially with the first flood peak, then an extreme, cumulative flood event can unfold. At present, there is a lack of understanding the extreme risk of such a spatiotemporal rainfall distribution and their resulting flood risk. Flood models often do not deal very well with such complicated rainfall distrbutions, and certainly not when model rainfall is uniformly distributed over the catchment [14].

We aim to develop a methodology to improve flood predictions for complex spatio-temporal distributions of rainfall in a catchment with only a minimal increase in the modelling resource and time. Providing a full picture of the rainfall distribution across a large river catchment is prohibitively expensive in terms of computational resources and time. We will outline a tractable approach achieved by using pattern recognition to reduce the dimensionality and by sampling spatially varying extreme rainfall distributions using novel Monte Carlo techniques.

To aid the approach in this study, we used a simplified modelling environment, inspired by and used in the mathematical design of the tabletop flood-demonstrator "Wetropolis" [4]. Wetropolis conceptualises the science of flooding and rainfall in an idealised river catchment. Physically, it is comprised of a river channel, possibly with a canal running in parallel, a reservoir for water storage, a porous flow cell with observable groundwater flow analogous to an upper catchment of porous moorland and random "rainfall", the latter which occasionally leads to flooding in the idealised urban area of Wetropolis. Rain falls either in the reservoir or in the moor. We use a idealised mathematical model version of Wetropolis to sample rainfall intensity. It enables us to define an impact metric, based on a weighted rainfall intensity, and a damage function relating to flood extent, i.e. a threshold water level within the city area of the model. It thus allows us to formulate a minimisation problem for spatiotemporal multi-variate rainfall distributions in a clear manner.

The challenge posed above, on the use of the spatial distribution of rainfall in flood models with only a minimal increase in the modelling resource and time, was presented to the study group by Dr. Adam Baylis of the Environment Agency (EA). The EA is an executive non-departmental public body, sponsored by DEFRA (the Department for Environmental, Food and Rural Affairs), responsible for protecting and improving the environment, including management of flood risk.

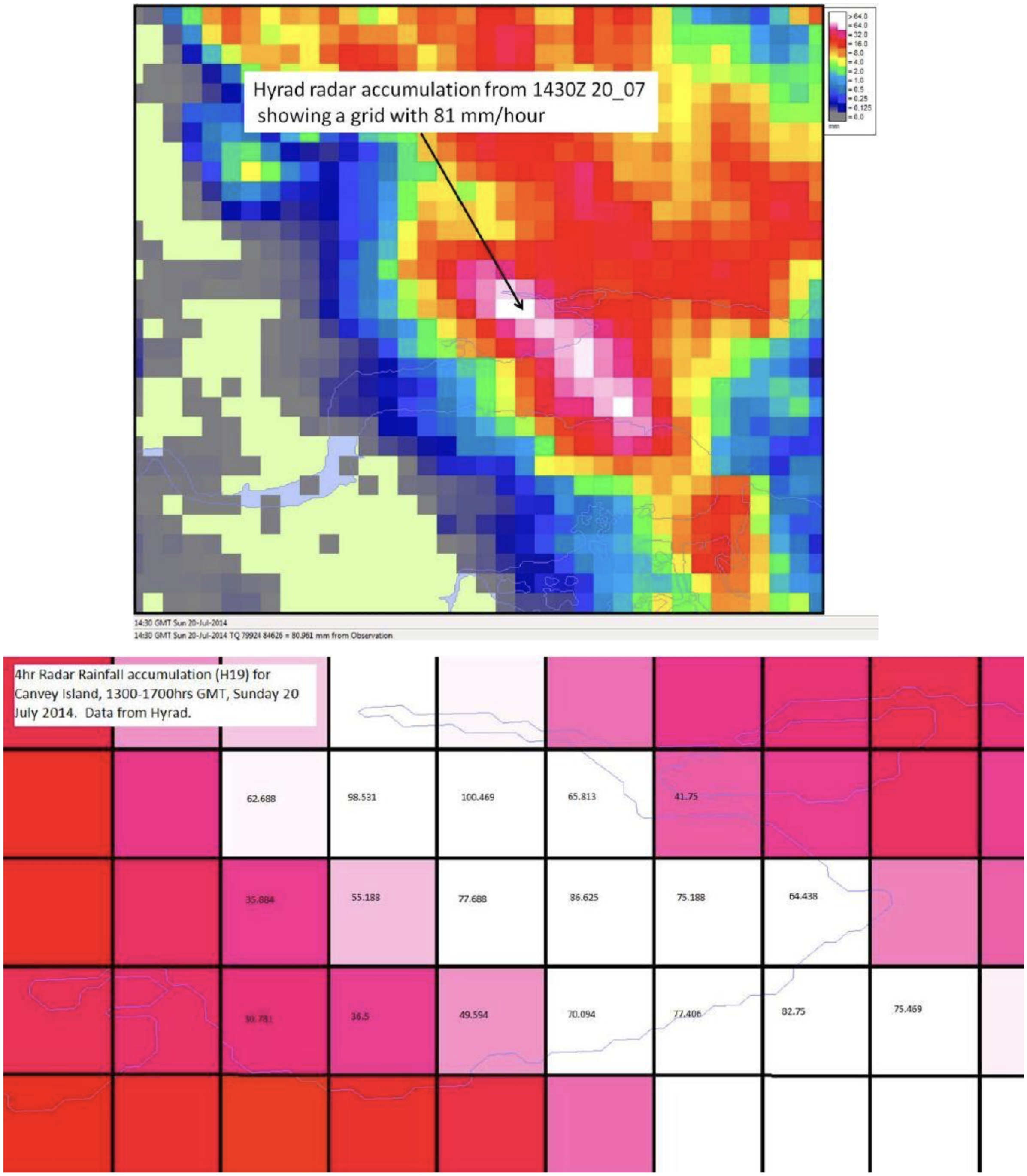

Figure 1: Hyrad radar hourly accumulation precipitation data over Canvey Island July 20th 2014 with each square having a size of $500 \times 500\text{m}^2$ in an overview (top), with the Thames estuary outlined underneath, as well as a zoom-in over Canvey Island (bottom) with accumulated rainfall in mm (Source: Environment Agency, 2014).

1.1 Spatially varying rainfall patterns

To investigate the effects of spatio-temporal rainfall distributions, we will consider a hierarchy of test cases to simulate rainfall, in one dynamic setting, including:

(1) spatially uniform rainfall as a reference case mimicing what is often done in the current flood forecasting practice;

(2) spatially varying rainfall (e.g. as for Canvey Island [3]); and,

(3) both temporally and spatially varying rainfall (e.g. as for the Medway [14]).

1.2 Probabilistic forecasting of flood events

Probabilistic forecasting is commonplace for flood prediction [7, 18, 9]. In this setting it is traditionally done with ensemble-based methods. Ensemble-based methods of probabilistic flood forecasting can incorporate spatio-temporal variation in rainfall patterns naturally. Once we can simulate spatially and temporally varying rainfall patterns, this incorporation is visualised via the following chain of events:

$$\textbf{Input: Sampling rainfall events} \longrightarrow \underbrace{\textbf{Black-box flood model}}_{expensive!} \longrightarrow \textbf{Output: Postprocess / forecast}.$$

The components of this chain are as follows:

Sampling rainfall events: The first component describes the sampling of rainfall events from a spatio-temporal distribution. The distribution that one samples from here is primarily dependent on the assumptions that one makes about the rainfall patterns in a certain catchment. Typical behaviour of such rainfall patterns should be known and a distribution could be inferred from this. Clearly, the more complicated the distribution, e.g. with more degrees of freedom, the more challenging the forecasting problem becomes. The challenge arises because with a fixed-sized and finite ensemble of simulations it is difficult to find a good approximation to the likelihood of rare flood events. Our optimisation algorithm concerns this component of the event chain.

Black-box flood model: The next component of the forecasting process contains most of the leg-work. One must usually run some black-box flood model with inputs (source term, topography) over a time interval and in space over the catchment / domain. This is by far the most computationally expensive part of the forecasting process. In some cases this black-box model will be the shallow water equations (SWEs) although this is a particularly expensive choice [10, 12, 5]. In this report we concentrate on a simpler one-dimensional flood model (used in the design of Wetropolis) for the forecasting component.

Postprocess and forecast: The final component usually takes the form of a statistical postprocessing / forecasting method. For example, given an ensemble of simulations derived from the last part (black-box model), one can compute statistics of the flood forecast via Monte Carlo methods. Here, a variety of verification / scoring techniques can be used to evaluate the quality / performance of the forecast [11].

There are various ways to reduce the computational complexity of the above chain to arrive at feasible forecasts. Variance reduction is a popular choice when using Monte Carlo in the estimating statistics to forecasts; multilevel methods have raised significant interest in the last decade. We refer also to a feasibility study of Maths Foresees' network, which looked into applying this particular variance reduction technique to flood ensemble forecasts [12].

2 Methodology

Given that the most expensive part of the forecast process is the black-box flood model, it is in our interest to only generate simulations of this model with optimised samples of spatio-temporal rainfall (the first component identified above). In this section, we define and apply an optimization problem to generate trained samples of certain catchment's spatio-temporal rainfall patterns that lead to predictions for rare flood events. The optimisation problem will be chosen to be a linear function of the spatio-temporal rainfall function r, involving a spatio-temporal function f which we train on the data. Since these are shown to be optimised samples for each catchment, one may only need to run the black-box model once for a "best guess" forecast or a handful times for the corresponding probabilistic forecast.

2.1 Wetropolis as black-box flood model

Wetropolis is a flood demonstrator built for a public outreach and engagement project of the Maths Foresees' network [4], see its overview in Fig. 2. It offers a suitable model environment for our study. Wetropolis fits on a table-top of size $1.4 \times 1.4\text{m}^2$. It consists of a winding river channel with a constant upstream inflow at $x = 0$m using a coordinate $x \in [0, L]$ following the centre of this winding channel, random inflow from an outflow pipe of a reservoir centred at $x_1 = 0.925$m as well as random inflow from a porous moor with an overflow strip centred at $x_2 = 2.038$m. Currently, $L = 5.21$m while in the original mathematical design $L = 4.21$m. The walled river channel has a rectangular section of width $w_r = 0.05$m with a higher wall on one side and a lower wall on the other side of height 0.015m running into a slanted 0.1m-wide flood plain on the right when facing downstream. The latter plain slopes 0.005m at right angles from the channel centre line upward into a wall. The river slope is 1 : 100.

In the mathematical model facilitating Wetropolis' design, the dynamics was modelled as follows:

- river flow was modelled using the one-dimensional St. Venant equation with only a rectangular cross-section of width w_r for the river depth $h_r(x,t)$ as function of the distance x and time t;
- the level $h_{res}(t)$ of the reservoir was modelled using mass conservation with rain inflow and outflow using a rectangular weir equation (e.g., [16]) across the width of the reservoir;
- the moor was modelled using a nonlinear diffusion equation for porous media flow with depth- and width-averaged ground water level $h_m(y,t)$ as function of a centre-line coordinate y and time, cf. [1]; coordinate $y \in [0, L_y]$ lies normal to the river channel at $x = x_2$; rain falls uniformly on the moor, on one side there is a wall with no through-flow and on the other side at $y = 0$ the water flows out at a base level via a weir equation across the width of the moor; and,
- Wetropolis' model contains a city located between the coordinates $x \in [x_c^-, x_c^+]$ with a threshold bank level of 0.02m, above which the city floods.

Rainfall is random. It varies both in location and rain amount. Rain location is fourfold: either in the reservoir, both in reservoir and moor, in the moor or not. Rain amount is $(10, 20, 40)\%$ or 90% during a "Wetropolis" day of wd = 10s. Both rain location and amount are drawn randomly from a discrete distribution with outcomes $(3/16, 7/16, 5/16, 1/16)^2$. In the table-top set-up these are drawn visually every 10s when (two) steel balls fall through two skew-symmetric Galton boards [4].

Here, we will consider a simplified version of the above one-dimensional mathematical model. We keep the river channel with along-channel coordinate $x \in [0, L]$, but have replaced the reservoir and groundwater cell with two point sources of rainwater supply at two distinct coordinates x_1 and x_2 along the catchment. So neither the reservoir nor the ground water are modelled with the detail given above. Instead rainwater supply is drawn randomly using Richardson's 1981 model of precipition, with a two-state Markov chain-exponential distribution of rain [6]. We have thus simplified the spatial pattern of the rainfall at only $x = x_1$ and $x = x_2$, two degrees of freedom instead of the reservoir level $h_{res}(t)$ and instead of the continuum of multiple groundwater levels $h_m(y,t)$. The city located between the coordinates $x \in [x_c^-, x_c^+]$ remains the target area of interest for our flood forecasts. Such a simplified model can be easily numerically discretized and can run the wd = 10s of real-time in less than a second on a laptop. Hence, we can quickly and efficiently model flood events to test the methodology proposed in the present study.

The *kinematic model* for river flow in a rectangular channel is a simplification of the St. Venant or cross-sectionally averaged shallow-water equations with parametrized bottom drag. The latter St. Venant equations

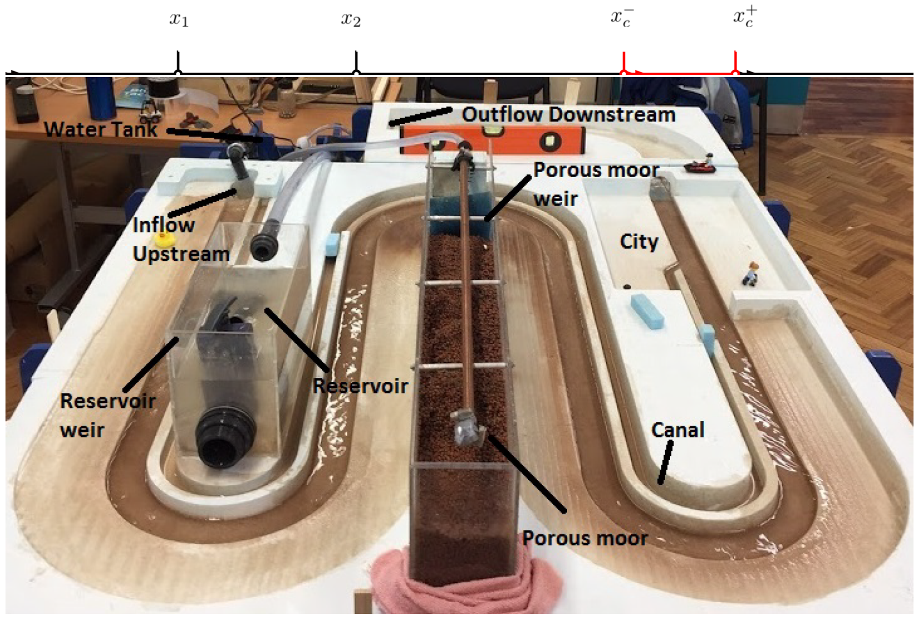

Figure 2: Schematic of the Wetropolis flood demonstrator in one dimension (top scale) and a photograph (bottom). Flow of water goes from left to right, with an outflow boundary at the right end of the catchment (top scale). The curvilinear x–coordinate following the river channel is given as a line above the schematic on which the rainfall point sources are located at the coordinates x_1 and x_2. The city is indicated in red by the zone within the interval $x \in [x_c^-, x_c^+]$.

read

$$\partial_t A + \partial_x (Au) = Q \equiv Q_1 \delta(x - x_1) + Q_2 \delta(x - x_2) \tag{1a}$$

$$\partial_t u + u \partial_x u + g \partial_x h_r = -\underline{g(\partial_x b + C_m^2 u|u|/R(h_r)^{4/3})} \tag{1b}$$

with cross-section $A(x,t) = w_r h_r(x,t)$, averaged velocity $u(x,t)$, average water depth $h_r(x,t)$ relative to the lowest point in a river cross-section, bottom topography $b(x)$, acceleration of gravity $g = 9.81\text{m/s}^2$, hydraulic radius $R(h) = \text{wet area/wetted perimeter} = w_r h/(2h + w_r)$ with Manning friction coefficient $C_m \in [0.01, 0.15]$ (for tabulated values of C_m, see [16]) and rainfall $r(x,t)$ with volumetric discharge terms Q_1 and Q_2 expressed as cubic metres per second, localised at x_1 and x_2 via two delta functions $\delta(\cdot)$. In Wetropolis $b(x) = -Sx$ with slope $S = 0.01$. The kinematic approximation arises when the underlined terms in (1) form the dominant balance, such that

$$u = R(h_r)^{2/3} \sqrt{-\partial_x b}/C_m. \tag{2}$$

Substitution of (2) into the continuity equation (1a) while using the definition of A yields the kinematic river flow equation

$$\partial_t A + \partial_x F(A) = r(x,t) \equiv Q_1 \delta(x - x_1) + Q_2 \delta(x - x_2) \tag{3}$$

with flux $F = F(A) \equiv AR(A/w_r)^{2/3} \sqrt{-\partial_x b}/C_m$, or

$$\partial_t (w_r h_r) + \partial_x \left(w_r h_r R(h_r)^{2/3} \sqrt{-\partial_x b}/C_m \right) = Q_1 \delta(x - x_1) + Q_2 \delta(x - x_2). \tag{4a}$$

This is a hyperbolic equation with positive characteristic since $\partial_A F/\partial A > 0$ or $(1/w_r)\partial_{h_r} F \partial h_r > 0$ with flux $F(h_r, C_m, \partial_x b, w_r)/w_r = w_r h_r R(h_r)^{2/3} \sqrt{-\partial_x b}/C_m$. Hence, only an upstream boundary condition is required, expressed in terms of an influx

$$Q_0(t) = A(0,t)u(0,t) = F\left(h_r(0,t), C_m, \partial_x b|_{x=0}, w_r\right) \tag{4b}$$

as well as an initial condition

$$h_{r0}(x) = h_r(x,0). \tag{4c}$$

A better approximation is to balance the free surface gradient $g\partial_x(h+b)$ with the friction term, which yields an advection-diffusion equation (cf. [2][1]).

The *numerical discretization* of (4) consists of a straighforward first-order finite-volume Godunov discretisation with upwind numerical flux (cf. [15])

$$\begin{aligned} h_k^{n+1} =& h_k^n - \frac{\Delta t_n}{\Delta x_k}\Big(\big(h_k^n R(h_k^n)^{2/3}\sqrt{-\partial_x b}/C_m|_{x_{k+1/2}}\big) - \big(h_{k-1}^n R(h_{k-1}^n)^{2/3}\sqrt{-\partial_x b}/C_m|_{x_{k-1/2}}\big)\Big) \\ &+ \Delta t_n Q_1^n \delta_{k,x_1} + \Delta t_n Q_2^n \delta_{k,x_2}, \end{aligned} \tag{5}$$

in which we have partitioned the line into $k = 1,\dots,N_k$ elements of finite volumes with (varying) time interval $\Delta t_n = t^{n+1} - t^n$ between time levels t^n and t^{n+1} for $n = 0,1,\dots$ and approximation $\Delta x_k h_k^n \approx \int_{x_{k-1/2}}^{x_{k+1/2}} h_r(x,t^n)\,\mathrm{d}x$ with the k^{th}–cell width $\Delta x_k = x_{k+1/2} - x_{k-1/2}$ based on cell edges $x_{k+1/2}$ and $x_{k-1/2}$. The modified "Kronecker delta" symbol is defined as follows: $\delta_{k,x_1} = 1$ when x_1 lies in cell k and zero otherwise. At the initial time $n = 0$, we take $\Delta x_k h_k^0 \approx \int_{x_{k-1/2}}^{x_{k+1/2}} h_r(x,0)\,\mathrm{d}x$.

The *random rainfall distributions* are based on Richardson's approach [19]. In the first approach we assume the distributions to be independent at the two locations. Per location, the distribution consists of a two-state Markov chain determining whether the day is dry, a state with value 1, or wet, a state with value 2 with associated transition probabilities. Given the transition probability $P(1|2)$ of a wet day with state 2 following a dry day with state 1 and the transition probability $P(2|2)$ of a wet day following a wet day, the remaining transition probabilities are given by:

$$P(1|1) = 1 - P(1|2) \quad \text{and} \quad P(2|1) = 1 - P(2|2). \tag{6}$$

Hence, if the day was dry then the chance of it remaining dry is $1 - P(1|2)$ while the chance of it becoming wet is $1 - P(2|2)$. In the case that the previous day was in the wet state, the precipitation Y_d at the next day n_d is drawn from an exponential distribution

$$f_e(y_d) = \lambda e^{-\lambda y_d} \quad \text{such that} \quad Y_d \sim \ln\left(U(n_d)\right)/\lambda \tag{7}$$

with uniform distribution $U(n_d)$. Here, the Markov-chain-exponential values $P(1|2) = 0.226$ and $P(2|2) = 0.475$ as well as $\lambda = 0.282$ used are based on a fit to weather data from Spokane, Washington, cf. [19].

In the second approach, building on the above, we introduce correlations between the two sites, as follows. A 4×4 matrix P_{ij} with $i,j = 1,2,3,4$ of correlated transition probabilities is built. On the previous day, we now have four states: dd) both locations are dry (state 1), dr) location x_1 is dry but location x_2 is wet (state 2), rd) location x_1 is wet and location x_2 is dry or rr) both locations are wet, with corresponding states 2. The transition probabilities to wet states are then determined as follows:

$$dd \to rr: \quad 1 + P_{33} \quad \text{and} \quad 1 + P_{33} \tag{8a}$$
$$dr \to rr: \quad 1 + P_{31} \quad \text{and} \quad 1 + P_{11} \tag{8b}$$
$$rr \to rr: \quad 1 + P_{11} \quad \text{and} \quad 1 + P_{11} \tag{8c}$$
$$rd \to rr: \quad 1 + P_{11} \quad \text{and} \quad 1 + P_{13}. \tag{8d}$$

We now use one shared uniform distribution U_1 on a day as well as occasionally a second one U_2. When it has only been wet at location x_2 (dr-case) then $Y_{d2} \sim -\ln(3U_1/4 + U_2/4)\lambda$. When it has only been wet at location x_1 (rd-case) then $Y_{d1} \sim -\ln(U_1)\lambda$. When it has only been wet at both locations then $Y_{d1} \sim -\ln(U_1)\lambda$ and $Y_{d2} \sim -\ln(3U_1/4 + U_2/4)\lambda$.

In the third approach we took artificially simulated data relating to daily rainfall at different sites over part of Ashdown Forest in Sussex. This dataset is part of an example dataset provided with the R-package RGLIM-CLIM [6], designed for modelling and simulation of univariate or multivariate daily weather sequences at single

[1]An asymptotic justification thereof was given by Gavin Esler at the second General Assembley of the Maths Foresees' network in Edinburgh, September 2016, cf. [17].

or multiple sites. According to the manual, the dataset was generated by simulating a different generalised linear model fitted, with appropriate modifications. The data exhibit many typical features of rainfall sequences in temperate climates. We used data from sites G3 and G6 as realisations of variables $Y_{d1}^{n_d}$ and $Y_{d2}^{n_d}$, for which approximately 14% of the values are missing (at random). Due to the time restriction of the study group, we simply omitted the dates with missing rainfall data, potentially making this dataset unreliable. If we were to repeat this, we would fit the missing data using GLM methods implemented in the RGLIMCLIM software, as described in the example in [6].

For either of these three apporaches, let us now denote rainfall random variable on a particular day n_d as $Y_{d1}^{n_d} \sim Y_{d1}$ and $Y_{d2}^{n_d} \sim Y_{d2}$, and their realisations as $y_{d1}^{n_d}$ and $y_{d2}^{n_d}$ respectively. Then we can define the rainfall function as a piecewise constant function over each Wetropolis day of 10s, given by

$$r(x,t) = \begin{cases} Q_1 = s_r y_{d1}^{n_d}, & \text{if } x = x_1 \text{ and } t - 10n_d \in [0, 10\text{s}) \\ Q_2 = s_r y_{d2}^{n_d}, & \text{if } x = x_2 \text{ and } t - 10n_d \in [0, 10\text{s}) \\ 0, & \text{otherwise} \end{cases}$$

for all $n_d \in \mathbb{N}$. The manually determined scale factor $s_r = (w_v L_y R_0)/3$ adjusts the Spokane to the Wetropolis precipitation settings based on Wetropolis design values with the width and length of the moor $w_v = 0.095$m and $L_y = 0.925$m and the rainfall $R_0 = 2.048$mm/s.

2.2 The damage and impact metric

Our primary goal is to define a mathematical approach to assess the impact of spatio-temperal rainfall distributions on flooding of a target area. Here, the target area chosen is a city region where the impact would be most severe. First, there is the damage metric. Second, there is the idea that this metric is a complicated function of the precipitations, complicated because we need to solve the hydrological model to evaluate it. We explore whether we can replace this complicated function by a simpler one, a linear functional of the rainfall, involving a spatio-temporal function $f(x,t)$ which we train on data. The damage functional $\mathscr{D}[r]$ to evaluate the impacts of rare flood events depends on the spatio-temporal rain function $r(x,t)$, indirectly via $h_r(x,t)$ through the hydrological model, and is defined as

$$\mathscr{D}[r] = \int_0^T \int_{x_c^-}^{x_{\text{city}}^+} \max\left(h_r(x,t) - h_c, 0\right)\, dx\, dt \tag{9}$$

over a sufficiently long period T. The impact functional depends on the rain r and the function f as follows

$$\mathscr{D}_{est}[r|f] = \int_0^T \int_0^L f(x,t)\, r(x,t)\, dx\, dt. \tag{10}$$

The mathematical approach is now formulated as the following minimisation problem

$$G = \min_f \left\{ \|\mathscr{D}[r] - \mathscr{D}_{est}[r|f]\|^2 \right\}. \tag{11}$$

This damage functional (9) is a function of the flood water depth $h_r(x,t)$, over the area of interest: the city. In practice, rather than being based on the water depth, one would desire the damage functional to be primarily a function of the financial cost resulting from damage to the areas of interest in a catchment. However there is good reason to believe the two are positively correlated; a case study into certain catchments found a close coupling between the water depth from flooding and the financial cost avoided from installing barriers preventing floods [8].

The optimisation problem posed to determine $f(x,t)$ is rather complicated and involves many degrees of freedom but there are a few simplifications possible. First, both the damage and impact functionals are only nonzero when there is a flood event, so the minimisation integral will become a sum of integrals over the duration of each storm event. Second, here a storm event is defined to occur when the river level at the start of the city at $x = x_c^-$ surpasses a certain flood threshold $h(x_{c^-}, t) > h_c = 0.02$m. Every such storm event has a duration $n_{d,0} - n_{d,1}$ Wetropolis days with an associated timeline, as indicated in Fig. 3, where $t_0 = n_{d,0} \cdot 10s$ and $t_1 = n_{d,1} \cdot 10s$ are respectively the start and end time of the flooding in the city. Third, the minimisation function and rain function are here chosen to depend only on two spatial locations such that: $f(x,t)r(x,t) \approx f_1(t)r(x,t)\delta(x - x_1) + f_2(t)r(x,t)\delta(x - x_2)$. Over all storm events we parametrise $f_{1,2}$ with three parameters each.

Fourth, given the simplification for f and that rainfall is resolved only per Wetropolis day, we also simplify the rainfall function over the period attributed to each flood event, i.e. $r(x,t)$ for $t \in [t_{-1}, t_1]$ with $t_0 - t_{-1} = 20\text{s} = 2\text{wd}$ (see Fig. 3), so that it can be characterised approximately with only three values. The additional time considered before the start of the flood, $t \in [t_{-1}, t_0)$, highlights that the damage is caused by rain events that started in the past. In practice the delay of the flood will depend on the level and the water levels but in the calculation only a gross estimate seemed to be required. This is based on bespoke training data from the Wetropolis model. It must be noted, however, that this is a crude assumption and does not account for multiple rain storms in quick succession causing long-lasting floods.

Note that the second, third and fourth simplifications are particular to our simplified and illustrative Wetropolis setup. Details are explained in the next section.

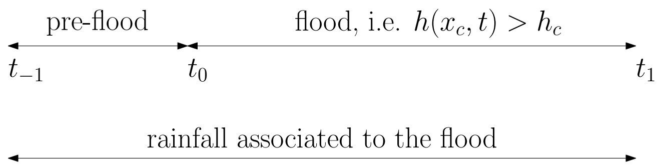

Figure 3: Timeline of a flood event.

2.3 Least squares minimisation problem

A reference case is to take a uniform function $f(x,t)$ amounting to equal weighting of extreme rainfall events, i.e. only counting those that caused (any) damage. The minimisation problem aims to find an optimised function $f(x,t)$ relating impact and damage functionals.

We have defined the localised functions in space f_1 and f_2 representing the temporal variation in rainfall at each of the two source coordinates, respectively, and thus the spatio-temporal variation in rainfall over the whole catchment and the entire duration of the record. For simplicity, we have chosen three coefficients per location, yielding six $\alpha_{i,j}$, $i = \{1,2\}$, $j = \{1,2,3\}$, in total making up both functions as follows $f_i(t) = \sum_{j=1}^{3} v_{i,j}(t)\alpha_{i,j}$. The three degrees of freedom in each function represents the 'start', 'middle' and 'end' of a given flood, determined over all storm events. Therefore

$$v_{i,j}(t) = \begin{cases} 1, & \frac{(j-1)(t_1 - t_{-1})}{3} \leq t - t_{-1} \leq \frac{j(t_1 - t_{-1})}{3}, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, the rainfall function $r(x,t)$ is piecewise constant at each location. During a particular storm that lasts for a few Wetropolis days, we only have daily resolution, whereas by the way we have defined f we might be interested in better/worse resolution than the daily rainfall over the time period attributed to the flood. We would like to make the rainfall resolution consistent over storm durations of any number of Wetropolis days. We achieve this via piecewise linear interpolation over rainfall at a particular location x_i during Wetropolis days related to the storm, this is over points $(10n_d, r(x, 10n_d)))$ such that $10n_d \in [t_{-1}, t_1]$ with n_d a positive integer, and t_{-1} and t_1 related to a particular storm. We denote the obtained interpolation function as $\tilde{r}_i(t)$ and define $\tilde{r}_{i,j} := \tilde{r}_i\left(t_{-1} + \frac{(j-1)(t_1 - t_{-1})}{n_r}\right)$ for $j = 1, 2, \ldots, n_r + 1$. The value of the positive integer n_r can be chosen depending on the desired resolution, but we choose $n_r = 3$. We then approximate rainfall at a location x_i via $r(x_i, t) \approx \sum_{j=1}^{3} \tilde{r}_{i,j} v_{i,j}(t)$, which further simplifies our impact function.

We now explain how to train $f_{1,2}$ around data of flood events and use it to predict future flood events whilst capturing spatio-temporal patterns in rainfall over a catchment. Given that we have a "training set" of N flood events via continuous simulation of Wetropolis, we can frame the above minimisation problem as a least squares minimisation problem.

Each recorded flood event, $m = 1, ..., N$, has an associated time period $T^m = [t_{-1}^m, t_1^m]$ and rainfall magnitude $r(x,t)\mathbb{1}_{T^m}(t)$ which is nonzero at the time of the storm. In detail, we derive the impact of storm m to be

$$I_m(f) = \int_0^L \int_0^T f(x,t) r(x,t) \mathbb{1}_{T^m}(t)\,\mathrm{d}t\,\mathrm{d}x \tag{12a}$$

$$= \int_0^L \int_{t_{-1}^m}^{t_1^m} f(x,t) r(x,t) \tag{12b}$$

$$= \sum_{i=1}^{2} \int_0^L \int_{t_{-1}^m}^{t_1^m} f_i(t) r(x,t) \delta(x - x_i)\,\mathrm{d}t\,\mathrm{d}x \tag{12c}$$

$$= \sum_{i=1}^{2} \int_{t_{-1}^m}^{t_1^m} f_i(t) r(x_i,t)\,\mathrm{d}t \tag{12d}$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{3} \alpha_{i,j} \int_{(j-1)(t_1^m - t_{-1}^m)/3}^{j(t_1^m - t_{-1}^m)/3} v_{i,j}(t) r(x_i,t)\,\mathrm{d}t \tag{12e}$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{3} \alpha_{i,j} \int_{(j-1)(t_1^m - t_{-1}^m)/3}^{j(t_1^m - t_{-1}^m)/3} r(x_i,t)\,\mathrm{d}t \tag{12f}$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{3} \alpha_{i,j} \int_{(j-1)(t_1^m - t_{-1}^m)/3}^{j(t_1^m - t_{-1}^m)/3} \sum_{k=1}^{3} \tilde{r}_{i,k} v_{i,k}(t)\,\mathrm{d}t \tag{12g}$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{3} \alpha_{i,j} \int_{(j-1)(t_1^m - t_{-1}^m)/3}^{j(t_1^m - t_{-1}^m)/3} \tilde{r}_{i,j}\,\mathrm{d}t \tag{12h}$$

$$= \frac{t_1^m - t_{-1}^m}{3} \sum_{i=1}^{2} \sum_{j=1}^{3} \alpha_{i,j} \tilde{r}_{i,j}. \tag{12i}$$

3 Results

In this section we will optimise f over a training set of flood events (continuous simulation of Wetropolis). This will then allow us to test the effectiveness of the method at constructing "best guess" forecasts of the damage of future flood events based on a testing set of rainfall storms. The effectiveness of the optimisation will be based on the improvement it offers over using an uniform f over both rainfall locations and temporally constant throughout each flood.

Both spatially correlated and uncorrelated scenarios of training and testing rainfall and flood data will be used to confirm if the first two test cases highlighted in subsection 1.1 (uniform and spatially correlated) hold. As the rainfall data will be simulated continuously over time, the third case (spatio-temporal rainfall) is also tested. The correlated rainfall data will be produced by using an R–package, *RGLIMCLIM*, with two rainfall sites [6].

The training sets of flood events are simulated over about 5000 Wetropolis days and the optimisation is also done over this period. The use of these optimised coefficients in forecasting mode is left for future research.

3.1 Spatially uncorrelated rainfall patterns

In the uncorrelated rainfall case, the fitted function f appears from Figure 4 to predict the damage from a particular storm event much more accurately than using an uniform f, especially in smaller flood events.

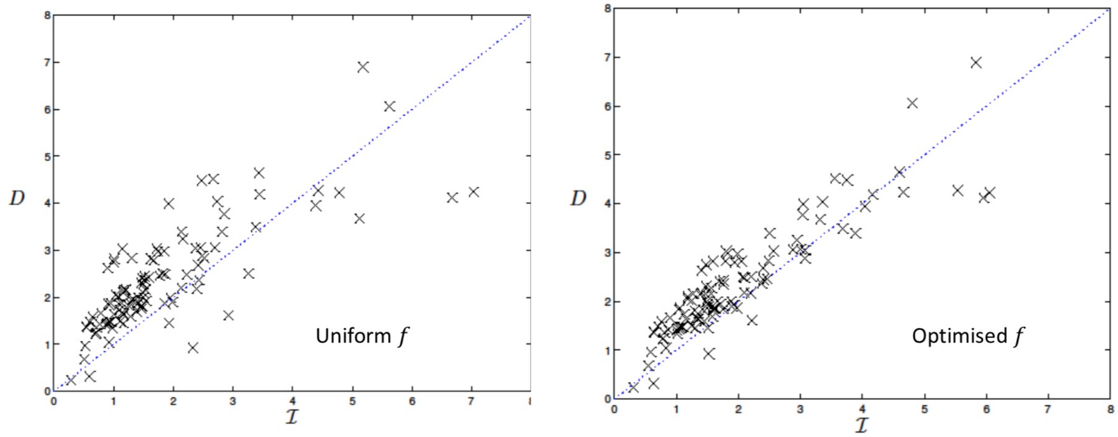

Figure 4: Predicted values of $\mathscr{D}_{est}(f)$ with uniform f (left panel) and optimised f (right panel) against the damage functional, $\mathscr{I} = \mathscr{D}$, for the testing set of flood events with spatially uncorrelated rainfall patterns.

3.2 Spatially correlated rainfall patterns

The test function f becomes more important in the spatially correlated rainfall case, and it has to be able to capture the spatial dependence of the two rainfall locations to be able to accurately predict the damage of the test flood events. One can see from Figure 5 that using the proposed method, fitting f around the training data, does lead to improved correlations of flood events better than that of using an uniform f. Whilst there is little change for smaller flood events, the root-mean-square-error (RMSE) from that of a perfect prediction is significantly improved in the case of using the optimised f because of the larger flood events. Two main outliers effect the cases when using an uniform f; these are much more constrained when using the optimised f.

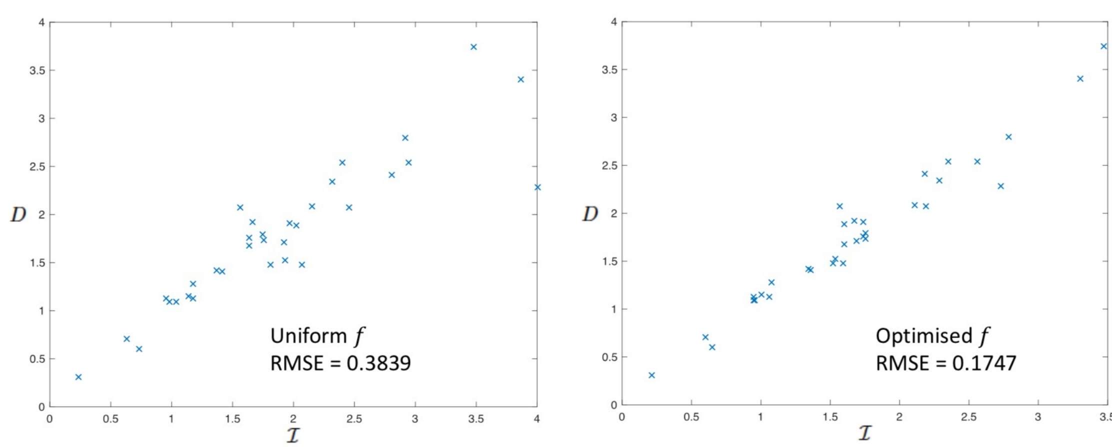

Figure 5: Same as in Figure 4, but for the testing set of flood events with spatially correlated rainfall patterns.

4 Conclusion and discussion

In this report, we have proposed a technique to take into account spatio-temporal variability of rainfall without needing the computational resources of full continuous Monte Carlo simulations. The idea is relate flood damage directly to the rainfall through a function that is trained on simulation data. This "optimisation" over the spatio-temporal function $f(x,t)$ was done before simulation of a black-box flood forecasting model which may be more efficient than traditional Monte Carlo sampling strategies. The method has been shown to work on a simple proof of concept using an idealised two-rainfall-site model, Wetropolis, by leading to increased accuracy of predictions to flood events in cases where rainfall is both spatially correlated and uncorrelated.

We have left out the uncertainty quantification aspect of the least squares minimisation problem used in this report for simplicity. This uncertainty quantification, by means of confidence interval around the least squares minimisation, for example, could be utilised in this method to capture probability bands on our forecast. This still only requires very few runs of our black-box flood model; once again showing the method is more efficient than traditional Monte Carlo sampling ideas.

Our approach in principle allows to *a priori* gauge the damage caused, here in city, given a certain spatio-temporal rainfall metric, using either observed or predicted rainfall $r(x,t)$.

Future extensons of this research and further recommendations are as follows:

- The use of the optimised coefficients f in forecasting mode is left for future research.
- In reality rain can fall over a large area, so the spatial variability in the minimiser $f(x,t)$ and the rainfall $r(x,t)$ has much more spatial variability. This can be tested in the actual Wetropolis model in which rain falls (uniformly) on the moor but is delayed by the groundwater dynamics, including previous hydrological history.
- The correlation between the damage and impact functionals includes the delay between rainfall and the reulting river level dynamics. Here it is modelled as a set delay Δt_r but this needs to be explored further.
- Our approach should be compared and explored for different rainfall scenerios and rainfall generators, including a comparisoon between the uncorrelated and correlated rainfall generators based on Richardson [19] and the Chandler [6] used in this study.

acknowledgements

We thank Ian Hewitt for proofreading a preliminary version of this report.

References

[1] G.I. Barenblatt. Scaling, self-similarity, and intermediate asymptotics. Cambridge University Press. 1996.

[2] P.D. Bates, M.S. Horritt, T.J. Fewtrell. A simple inertial formulayion of the shallow water equations for efficient two-dimensuonal flood inundation modelling. *J. Hydrology* **387**, 33–45. 2010.

[3] A. Baylis. Spatial rainfall distribution in flood modelling. Presentation Environmental modelling in Industry study group of EPSRC LWEC Maths Foresees' network. `http://www.turing-gateway.cam.ac.uk/event/tgmw41/programme`. April 3rd 2017.

[4] O. Bokhove, W. Zweers. Wetropolis flood demonstrator. `https://blogs.reading.ac.uk/dare/2017/07/25/wetropolis-flood-demonstrator/` & `www1.maths.leeds.ac.uk/mathsforesees/projects.html`. 2017.

[5] L. Cea, J. Puertas, L. Pena, and M. Garrido. Hydrological forecasting of fast flood events in small catchments with a 2D-SWE model. numerical and model experimental validation. In *World Water Congress*, pages 1–4. 2008.

[6] R. E. Chandler. *RGLIMCLIM: A Multisite Multivariate Daily Weather Generator Based on Generalized Linear Models, User Guide*. University College London, London, England. 2015.

[7] H. Cloke and F. Pappenberger. Ensemble flood forecasting: A review. *J. Hydrology*, 375:613–626. 2009.

[8] M. Dale, Y. Ji, J. Wicks, K. Myline, F. Pappenberger, and H. Cloke. Applying probabilistic flood forecasting in flood incident management. Technical Report SC090032, Environment Agency. 2013.

[9] D. Demeritt, H. Cloke, F. Pappenberger, J. Thielen, J. Bartholmes, and M-H. Ramos. Ensemble predicitions and perceptions of risk, uncertainty, and error in flood forecasting. *Environmental Hazards* 7, 115–127. 2007.

[10] A. Ern, S. Piperno, and K. Djadel. A well-balanced Runge–Kutta discontinuous galerkin method for the shallow-water equations with flooding and drying. *International Journal for Numerical Methods in Fluids* 58, 1–25. 2008.

[11] T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *J. Royal Statistical Society: Series B (Statistical Methodology)* 69(2), 243–268, 2007.

[12] A. Gregory and C. J. Cotter. Multilevel Monte Carlo Methods for Flood Risk Assessment: A feasibility study report. Technical report, Maths Foresees Network. 2017. `www1.maths.leeds.ac.uk/mathsforesees/projects.html`

[13] G. Jenkins M. Perry, and J. Prior. The climate of the UK and recent trends. Met Office Hadley Centre, Exeter, UK. 2008.

[14] R. Lamb, D. Faulkner, P. Wass, and D. Cameron. Have applications of continuous rainfall-runoff simulation realized the vision for process-based flood frequency analysis? *Hydrological Processes.*, 30(1), 2463–2481. 2016.

[15] R.L. LeVeque. Numerical Methods for Conservation Laws. Lectures in Mathematics, Birkhäuser. 1990.

[16] B.R. Munson, D.F. Young and T.H. Okiishi. Fundamentals of fluid mechanics. Wiley. 2005.

[17] D. Pritchard, and A. Hogg. The effects of hydraulic resistance on dam-break and other shallow inertial flows. *J. Fluid Mech.* 501, 179–212. 2004.

[18] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian data Assimilation*. Cambridge University Press, Cambridge. 2015.

[19] C.W. Richardson. Stochastic simulation of daily precipitation, temperature, and solar radition. *Water Res. Research* **17**, 181–190. 1981.

[20] M. Sanderson. Changes in the frequency of extreme rainfall events for selected towns and cities. Met Office report, Exeter, UK. 2010.