Spatio-temporal generation of large-scale hazardous events that may cause flooding

Thesis

Spatio-temporal Generation of Large-Scale Hazardous Events that may cause Flooding

PhD Thesis

November 4, 2022

Dirk Diederen

Discipline: Engineering

Establishment: HR Wallingford (ARC of The Open University)

This research is part of the 'System-Risk' project and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 676027

HR Wallingford
Working with water

Summary

This thesis treats the generation of weather scenarios that may cause flooding, generally referred to as events, within the context of flood risk. A key concept for the risk approach is that not only hazardous events are generated, but that realistic probabilities, or frequencies, are connected to the events. The purpose of this thesis is to provide some methodological improvements on flood hazard generators and to bring into discussion some useful concepts with regards to spatio-temporal challenges.

The flood risk approach is largely data driven. The methods used in this thesis start from obtained data sets, provided by others. General computer algorithms and statistical models are used to generate hypothetical, or 'synthetic' data sets. These synthetic data sets are investigated and analysed using computational code.

The starting point is the generation of a large synthetic set of pan-European river discharge events, spread out over several hundreds of locations in Europe. Some methodological advances are provided, which allow moving discharge waves to be tracked throughout all major river basins in Europe. A key point in the used methodology is to capture the spatial dependence between events occurring at different locations, which will be referred to as the static spatio-temporal approach. Compared to a local approach, where each location is considered individually, the gains of considering spatial dependence seem rather clear. However, it appears that the static spatio-temporal approach does not work well with an event-based approach, since this methodology implicitly puts boundaries on the procedure of spatio-temporal event identification.

Therefore, the next step was the development of a generator that could provide large synthetic sets of precipitation events, over the entire Atlantic sea and Europe. A key point in the methodology used here is that not only the events are dynamic, as their movement is tracked, but also the event descriptors are dynamic. Hence, here the dynamic spatio-temporal approach is introduced, which implies the generator methodology moved beyond generation at a fixed set of locations. The methodological framework is established, which allows a first version of such a dynamic generator, with the potential to be applied globally. And some exploration is provided of methodological extensions that allow to treat multiple variables concurrently in a compound framework.

With the newly introduced complexity of the dynamic generator, the final step was to start generating big synthetic data sets and to thoroughly test what the generator produced. A comprehensive sensitivity analysis provides some first insights into the behaviour of the generator, which allow to understand two key concepts. First, with the dynamic spatio-temporal approach, spatial coherence of extremes comes naturally. This contrasts with the static spatio-temporal approach where spatial coherence of extremes has to be assumed prior to the modelling and is typically done by the application of a spatial process. Second, for any location, the dynamic spatio-temporal approach allows

to more directly include events occurring in the area surrounding a location to compute extremes at that particular location. Relatively short data records are a standard limitation for the risk approach, whereby this 'dynamic expansion of information' may be of help.

The provided methodological advances and new concepts may help the way forward to the generation of global hazard events that may cause flooding. Large-scale, coherent event sets allow interactions and system behaviour to be studied, which is a main requirement to be able to compute the system risk. In addition, an interesting outlook is provided for future research. The dynamic spatio-temporal approach may in the future be able to provide not only spatially coherent extremes, but also temporally coherent extremes. This could be a first step towards credible global flood risk time series, which could be a very useful tool, in the current predicament of global climate change.

Contents

Publications

[1] Dirk Diederen. *Discharge waves in major European rivers for pan-European flood risk assessment*. Copernicus Publications. 2019. URL: https://doi.org/10.5446/40894.

[2] Dirk Diederen and Ye Liu. "Dynamic spatio-temporal generation of large-scale synthetic gridded precipitation - With improved spatial coherence of extremes". In: *Stochastic Environmental Research and Risk Assessment* 34.9 (2020), pp. 1436–3259. URL: https://doi.org/10.1007/s00477-019-01724-9.

[3] Dirk Diederen et al. "Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment". In: *Natural Hazards and Earth System Sciences* 19.5 (2019), pp. 1041–1053. URL: https://doi.org/10.5194/nhess-19-1041-2019.

1

Introduction

1

1.1. Background

1.1.1. What is flood risk?

Flood events cause large damages worldwide [27]. To decide on mitigating measures, a comprehensive understanding is required of the risk. Large investments, like in infratructure and in other urban capital, generally involve long time spans. Structures are designed to have a lifespan of multiple years or decades. Before investing in structures, we would like to know what conditions the structure may have to face. This contrasts with the current state of forecasting systems, which generally are capable of forecasting a few days or weeks at maximum with reasonable certainty. Forecasting can be expected to improve, but will not reach years or decades any time soon. So, with risk analysis, the question of what exactly will happen in the future (forecasting) is reduced to the question of what to expect, which may be considered a more feasible question to ask for long-term investments. Therefore, we deviate from the attempt to predict exactly what will happen to the investment, but we resort to scenarios. For a good quality, scenario-based analysis, not only should all (relevant) scenarios be considered, but it is important to include the probability of occurrence of scenarios. In fact, the probability of a particular scenario has a large influence on the relevance of that scenario.

Flood risk is defined to be the sum of consequences of all possible flood scenarios, where for each scenario the consequences are scaled by the probability of occurrence. With 'Flood Risk Assessments' (FRA), we attempt to approximate the flood risk. It may be decomposed into three parts; hazard, exposure and vulnerability. First, the hazard part, is the component related to the sources of flooding. Floods can arise from different sources (fluvial, pluvial and coastal). These sources can be studied separately, although it may be interesting to understand how these sources interact, which is referred to as multi-source or compound flooding. Second, the exposure part, is the component related to societal assets that are affected, such as population and infratructure. The hazard scenarios are used to drive flood models, which show how much flooding occurs when and where. Third, the vulnerability part, is the component related to the consequences of flooding. It is typically decomposed into casualties and damages.

The focus of this thesis is within the hazard part, in which scenarios are typically referred to as events. As extreme floods occur rarely it is necessary to consider a specialist branch of statistics, extreme value theory, to gain understanding of the hazard component of flood risk.

1.1.2. The flood hazard

The flood hazard consists of all possible scenarios of flooding, with for each scenario the associated probability of occurrence. The probability of occurrence can be directly converted to the frequency of occurrence, which is a standard in engineering. Typical terms used are return periods and associated return level estimates, which are the estimated values that correspond to the return periods.

Computers can deal with a limited amount of information. Therefore, it is impossible to consider all, i.e. an unlimited amount of, possible scenarios. It is only possible to consider a discrete set of (many) scenarios. The computational factor is limiting at present, and was more so in the past. With the increase in computational power, large-scale flood risk as-

sessments have become feasible in the 21st century. So, now that the right tools have become available, it seems to be the right time to start tackling the problem of large-scale flooding, and, thereby, to investigate the effect of the system behaviour on the flood risk.

A flood hazard assessment can only be valid for a particular system state. Since systems are never stationary (i.e. things change), the possibility and probability of scenarios are both time dependent and therefore the flood hazard itself is time dependent. Time is discretised, under the assumption that a system state remains stationary for the time of the discretised interval. Events inform specifically about the system state in which they occur. Typically, the aim is to obtain the flood hazard for a small time interval, but the information used may fall outside that time interval. We normally try to infer the hazard in the current system state from past events, which are events that contain information about a system state in the past. This is particularly problematic when using very extreme events, such as events that occurred 10.000 years ago.

In this thesis, trends in the flood hazard, i.e. changes in system states, are not considered. Climate change is out of scope, as the stochastic process is not extrapolated into the future. However, the methods developed here may be relevant for future flood hazard studies, which would require extensions of methodology. In Section (7.2), an outlook will be provided, demonstrating how to approach climate change in future research, by considering flood hazard time series.

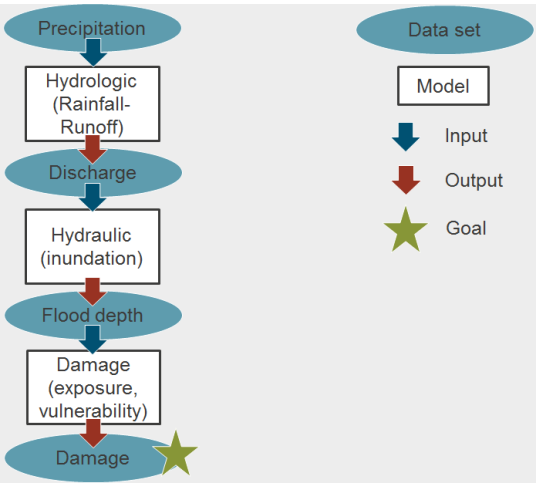


Figure 1.1: A simple flood risk model chain.

1.1.3. A chain of models - where to start?

Typically, for FRA, a chain of models is applied. It covers the entire risk cascade from hazardous extreme events down to flood damages (for example, expected annual damage) or to casualties resulting from inundation.

A simplified example of a flood risk model chain is displayed in Fig. (1.1). The objective is to obtain an estimation of the flood risk (green star). The arrows indicate different steps to get to the green star. Red arrows represent a transformation of variables, performed by input-output models such as hydrological models, hydraulic models or damage models.

A first argument for the choice of flood risk model chain set-up is the intended application of the flood risk analysis. For example, the application can be cost-benefit analysis, which is an important tool for decision makers. For cost-benefit analysis the flood risk in different system states has to be explored. For example, to explore the cost-benefit of the construction of a dike, a flood risk analysis can be performed with and without the dike. The costs could be the construction and (yearly) maintenance of the dike, and the benefit could be the (yearly) reduction

in flood risk after building the dike. The model chain would have to include an inundation model, such that the dike can be modelled and different results are obtained with or without the dike. Another example could be how changes in land use affect the flood risk, for which a hydrological rainfall-runoff model would have to be included.

After determining which model steps are required for the intended application, a route through the model chain has to be chosen. As can be observed in Fig. (1.2), there are multiple options to get to the green star with different degrees of complexity and/or computational effort. Yellow arrows are introduced, which represent the statistical step, in which the stochastic process is simulated. In the statistical step, methods are applied to an observed data set (for example, 25 years) to create a large synthetic data set (for example, a synthetic 10.000 years). In this thesis, the data concerning what has happened in the past will be referred to as 'observed data' and the generated scenarios will be referred to as 'synthetic data'. What happens exactly in these yellow arrows will be further discussed in more detail in Chapter (3). First we will consider the implications of the choice of route.

Route A might be the simplest route, where statistical methods are directly applied to observed flood damage data to obtain an approximation of the flood risk. A disadvantage of this very short route could be the relatively low availability of flood damage data, whereas statistical methods typically improve with more data. This route can be found in actuarial studies based on insurance claims data.

Route B starts from inundation data (flood depth). The (modelled) observed damages are calculated with a flood damage model. If more flood depth data is available than damage data, this additional model step could address the (potential) shortage of observed data in route A. However, is the modelled data as reliable as the observed?

Route C is similar to route B, but here statistical methods are directly applied to the flood depth data. A synthetic 10.000 years of data may be run through the damage model. These 10.000 years of synthetic data will contain a larger number of extreme scenarios than the 25 years of observed data, which is great. However, should the (damage) model be asked to perform calculations outside the range in which it has been calibrated?

Route D is often applied in practice, since the discharge in rivers can be constantly monitored and discharge data are therefore widely available. Statistics are directly applied to observed river discharge. Something unrealistic may happen with this particular variable, e.g. can we reliably extrapolate the river discharge beyond bank-full condition using a standard extreme value distribution?

Route E is again slightly more complex, since it starts from precipitation. Precipitation data is widely available, but yet another (hydrological) input-output model is required. In this case, statistics can be applied to the modelled river discharge, rather than to observed discharge, which can give a larger spatial coverage (for example, ungauged basins). However, again, should we extrapolate river discharge beyond bank-full?

Route F is the interesting alternative to route E, where the difference is in the first step. Instead of using the observed (25 years of) precipitation data to force the hydrological model, the synthetic (10.000

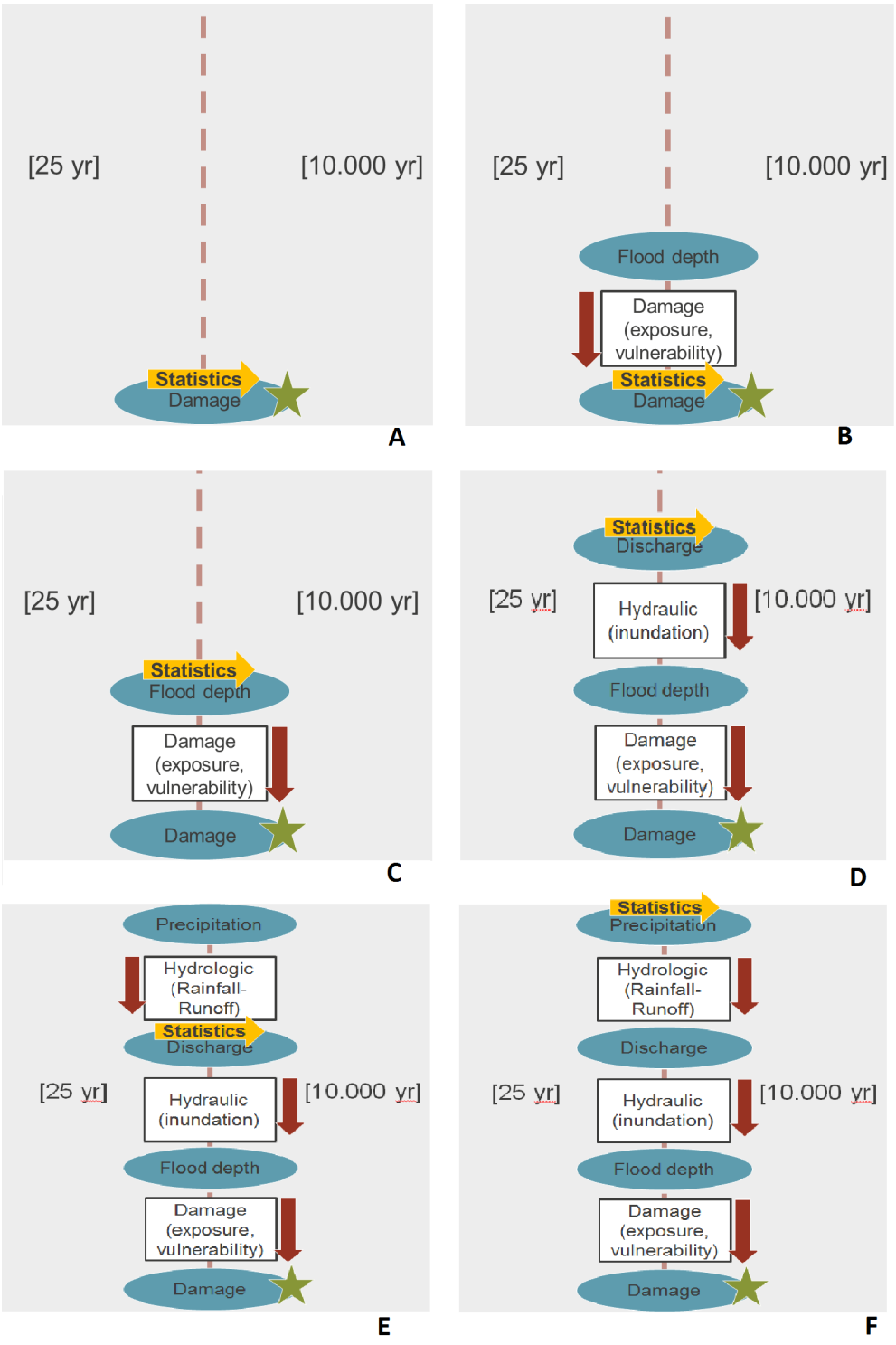


Figure 1.2: Different routes through the flood risk model chain. These different routes should lead to the same result, the green star, which would be a flood risk estimate.

years of) precipitation data are used. However, should now also the hydrological model be asked to perform calculations outside the range in which it has been calibrated?

Different routes through the model chain are supposed to lead to the same result (a flood risk estimate). However, in practice they will not. This gives a second argument for the choice of set-up of the flood risk model chain, which is based on the quality of the entire route.

At this point, it could be useful to shortly discuss the outdated use of 'design events'. In each layer of the model chain, events can be defined and probabilities, or frequencies, can be associated to these events. So, one could identify a design event in the first layer, e.g. a precipitation event with a 100-year return period, and then run this event through the model chain. The outcome could be a damage event, and in the old practice of design events, the original probability would then be assigned to this damage event (100-year return period). However, it is important to realise that probabilities of events are generally not portable across model transformations. For example, a precipitation event with a peak that has a 100-year return period may fall on a catchment which has a particular level of wetness that occurs once every 10 years, the combination of which may lead to a discharge event that has a peak that has a 1000 year return period. Or a dike, weakened by a sustained period of drought, may fail when a peak-discharge with a 25-year return period occurs, which may cause an inundation extent that has a 500-year return period. Decision making benefits from a large catalogue of events/scenarios of results of flooding, i.e. in terms of damage or loss of life. Therefore, the entire range of possible hazard events has to be inferred from available data to be subsequently evaluated. In summary, we should not consider design events, but should aim to develop methods to generate comprehensive sets of boundary conditions for full flood risk assessments (yellow arrows).

In this thesis, the focus lies not so much with the routes, but specifically with the statistical step (yellow arrows), where the routes provide context for this step. The statistical step will be applied to discharge in Chapter (4) and to atmospheric variables (mainly precipitation) in Chapters (5) and (6). How to generate synthetic data sets? What methods are currently available and how to improve them?

1.2. Research context

1.2.1. Project: System-Risk

System-Risk – a large-scale systems approach to flood risk assessment and management. System-Risk is a Marie-Skłodowska-Curie European Training Network which aims on developing and implementing a systems approach for large-scale flood risk assessment and management and provides a framework for training and career development of 15 Early Stage Researchers.

Particularly, national and regional policy development adhering to the solidarity principle anchored in the European Flood Directive as well as the insurance industry require tools to assess and manage flood risk at large scales, from the larger river basin to the European scale. Yet, the majority of flood research has centred on small- to meso-scale catchments and, to date, such requirements have usually been addressed by piecing together small-scale solutions.

Today, increased data availability, new numerical algorithms and dramatically higher computer performance enable large-scale analyses and models which were not feasible a few years ago. System-Risk performs leading-edge research with spotlights on three essential pillars of flood risk research. First, the risk chain, considering the complete risk chain from the Sources through the Pathways to the Receptors and Consequences. Second, interactions, augmenting the 'Source-Pathway-Receptors and Consequences' model by putting interactions centre stage and, in this way, replacing the traditional linear approach of the risk chain by a more realistic approach with interdependent linkages between physical and societal processes which finally shape the spatio-temporal flood risk. Third, temporal dynamics, investigating the time-varying nature of flood risk and its components on different time scales as for instance hours to days when flood defence failures change flood probabilities, months to years when people learn from damaging floods and improve private precaution and decades to centuries when human settlements in floodplains evolve.

System-risk is composed of 15 Early Stage Research Projects (ESRs) organised in three scientific work packages (WP) complemented by training, dissemination and management. The scientific WPs beam the 'risk chain – interactions – dynamics' spotlights on the 'Atmosphere-Catchment-System' (WP1), 'River-Dike-Floodplain System' (WP2) and the 'Socio-Economic System' (WP3). System-risk brings together research and training at ten leading centres of flood research in Europe and embeds partnerships with eight partners from the industry and administration in five countries. This Project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 676027.

Source www.system-risk.eu

1.2.2. Study: topic

Probabilistic approach As extreme floods occur rarely it is necessary to consider a specialist branch of statistics, extreme value theory, to gain understanding of the probability component of flood risk. In this Early Stage Research project (ESR4) in WP1 of the System-Risk project, the focus lies with the flood hazard. The flood hazard can be defined to consist of all potential scenarios of flooding that may occur. The idea is that many flood scenarios could potentially occur, from which only some actually do occur. So, given that only a few flood scenarios occurred in reality, and even fewer observed with complete data record, we have to address the problem of sparsity of extreme events. Namely, the observed data hold only a fragmented picture of the actual flood hazard, whereas we are trying to get the whole picture. The statistical model that belongs to this approach is called the full stochastic process.

Systems approach Methods that are currently in use to address sparsity have mostly been developed on a per-location basis, i.e. are local. They address sparsity in time by fitting distributions to data records per location. This approach has always been sensible, because data records have traditionally been available per location, typically for a limited number of locations only. An example can be found in the

current UK national FRA, which is a mosaic of the results of locally executed FRA. When putting together the individual, local FRA, the results are fragmented. This means that these methods do not allow system interactions to be studied. They cannot provide a whole, coherent picture of the flood risk and, since system behaviour influences the flood risk, they cannot be accurate.

The generation of hypothetical scenarios, derived from observed data, is the specific aspect of focus of this ESR project (4) 'Development of a spatio-temporal weather generator for large-scale floods'. To address sparsity not only in time but concurrently in space, methodology is to be developed for spatio-temporal weather generation. The weather generator is the tool required to generate hypothetical scenarios that are coherent in space and in time, which is a requirement to be able to incorporate a systems approach in FRA.

Large-scale The weather generator should not work with data records per location, but is to be developed for large-scale, gridded data sets. Flood models that are cut off at country borders or catchment boundaries are inappropriate, because rainfall and catchments do not respect such boundaries. The weather generator has to provide the opportunity to get the whole picture of the flood hazard using the new generation of data: big data.

Interaction and changes The weather generator can be used as a tool to study the system-risk in river basins. For example, the system interactions of dike failures. In combination with hydrological and hydraulic models it would allow the production of coherent, large-scale (continental) summary maps of the flood hazard. Such maps have broad application and are required by the EU floods directive. These maps will be spatially coherent, rather than fragmented maps compiled from separate, local analyses. They will include the effects of the system behaviour of floods, the importance of which is described in more detail in the other ESR sub-projects of the System-Risk project.

Finally, with the use of atmospheric data, the weather generator will allow the cost-benefit analysis of human interventions in river basins, such as the construction or adjustment of dikes or changes in land use. In short, it allows to study the influence of changes in one part of the system on other parts of the system.

Title Development of a spatio-temporal weather generator for large-scale floods.

1.3. Academic contributions

1.3.1. A framework to generate hazardous flood events

Based on the upcoming literature review in Chapter (2), a new, generally applicable framework for event-based flood hazard generation will be proposed in Chapter (3).

1.3.2. The limitations of analysis on a per-location basis

Measurements are traditionally collected per location. Gauging stations at rivers constantly monitor water levels, which are transformed to river discharge using rating curves. Precipitation rates are captured with a wide range of methods, from simple buckets to more advanced methods (such as acoustic or image based devices). With the availability of per-location data, a large range of methods has

been devised for local probabilistic analysis (LPA), defined in Appendix (A.1.4).

In this thesis, event-based methodology is considered, which will be explained in Section (3.1.2). When an event occurs at a particular location, nearby locations are likely to be affected, whereas locations far away are less likely to be affected. A dependence structure exists between the occurrence of events at different locations. This spatial dependence structure gave rise to the development of the multi-site type of analysis, which has been extensively studied in (approximately) the last 20 years. In this thesis, multi-site analysis is referred to as 'static spatio-temporal probabilistic analysis (SSTPA)', defined in Appendix (A.1.5). The general idea of SSTPA is to connect what occurs at a predefined set of locations.

When aiming to describe hazardous events at a predetermined set of locations, the problem arises that events do not occur at all locations simultaneously. The event identification metholodogy can be adjusted to produce events that occur at all locations simultaniously. But these events are far from what anyone imagines spatio-temporal events to be. Should it be part of the same event what happens simultaneously on the moon?

For SSTPA, the requirement of a data point for each event at each location, originates from that methods for multivariate analysis generally require complete matrices. If statisticians manage to develop multivariate methods that can handle incomplete matrices, event-based SSTPA can continue to be developed. However, whether this is feasible remains to be seen, because the literature points out that it is already quite challenging to address all (relevant) patterns in a complete matrix (see Section (2.3)), where incompleteness of a matrix adds severe, additional complexity.

In Chapter (4), the limitations of event-based SSTPA will be discussed. Details can be found in Section (4.5.2).

1.3.3. A framework to explore the performance of generators

A framework to explore the performance of generators will be proposed in Chapter (6). This framework proposes that generators are checked for generic generator objectives. It allows to go beyond focussed performance checks, in which only the particular generator objectives are checked for which the generator was optimised. With this framework, different generators can theoretically be compared and per generator, it can be more clearly mapped what the trade-offs are regarding several aspects, such as the particular generator methodology, the chosen settings, and more.

1.3.4. Spatially coherent extremes - why the spatial process may not be scientific

With the general concept in mind that events are dynamic, moving around through space in time, extremes at neighbouring locations are generally expected to show a high degree of similarity. However, when fitting distributions to extremes per location, the distributions that emerge are not guaranteed to be smooth at all, which can be explained by the sparsity of extreme events.

With the expectation of spatial smoothness of extremes in mind and within the context of fitting distributions per location, methods have been developed to force spatial smoothness. Such methods are referred to as a 'spatial

process' and are typically applied to distribution parameters. First an expectation of spatial smoothness is formulated and then methods are incorporated to force this expected spatial smoothness. Such an approach is appriopriate in a practically orientated world of engineering, but appears to lack scientific rigour.

The newly developed type of generator, that will be discussed in Chapter (5), will be demonstrated to be able to produce spatially coherent extremes in Chapter (6), without forcing spatial smoothness. The quality of the generator methodology and of the synthetic events will be up for discussion. This developed generator will require plenty of improvement in the future. However, the generation of synthetic events may provide a more scientific approach than the application of a spatial process, since spatial coherence is not forced but instead emerges as a consequence of the applied methodology. In short, the discussion may be moved from how much smoothening to apply towards whether the generated (dynamic) events are realistic. Details can be found in Section (6.5.2).

1.3.5. Dynamic expansion of information

Natural phenomena travel. Discharge waves move through a river network, clouds and associated precipitation fields move through the atmosphere, etcetera. Thus, physical phenomena are dynamic by nature. A dynamic approach, in which the spatio-temporal footprints of events are captured and described by spatio-temporal descriptors, is different from a multi-site approach. A definition of spatio-temporal probabilistic analysis is provided in Appendix (A.1.5).

The confidence decision makers can have in flood risk analyses is limited by the uncertainty inherit in risk analyses. A major source of uncertainty is the limited availability of information, i.e. 'evidence'. Merz and Blöschl [61] introduced a framework for potential methods to address the limitation of information. They proposed a framework of expansion of information, distinguishing between temporal, spatial and causal expansions. Although the combination of temporal and spatial expansion of information may suggest it has already been incorporated, this framework does not include spatio-temporal coherence of information. The reason for missing this coherence is that time and space are addressed separately in the original framework. Therefore, an addition to the original framework is proposed, namely dynamic expansion of information.

Using dynamic spatio-temporal probabilistic analysis (DSTPA), for each location, information is utilised from the entire region of influence of that particular location. This principle will become more clear by the example analysis in Section (6.3.4), Fig. (6.10). Details can be found in Section (6.5.1).

1.4. Practical contribution

1.4.1. From gauge data towards gridded data sets

To enable the capturing of spatio-temporal events, physical phenomena have to be tracked in space and time. The feasibility of a spatio-temporal approach is determined by temporal and spatial coverage, which is different for each variable. Moreover, the higher the resolution of the data set, the more is to be gained from a tracking method. Physical phenomena could be tracked in a dense network of gauges, but gridded data sets generally have a higher

resolution (density). With the increasing availability, quality and resolution of remote sensing of data, it is likely that gridded data sets will be considered for FRA more and more in the future.

Gridded data sets have particular resolutions in space and time, which, because of the different dimensions, are not directly comparable. To determine which dimension is the limiting factor, the celerity can be used, which is the speed with which physical phenomena propagate through a medium (such as water or air). The celerity is different for different variables, such that different combinations of spatial and temporal resolutions are required for different variables. Discharge waves propagate through (the larger) channels and rivers with celerities ranging between 1-10 m/s (roughly 100-1000 km/day), which can be used as a scaling factor. Discharge data with a spatial resolution of about 100 km would require a temporal resolution which is at least daily. In contrast, precipitation fields move around with celerities that are an order of magnitude higher, such that precipitation data with a spatial resolution of about 100 km would require a temporal resolution which should be hourly. With the increase in computational power since the 2nd millennium and the recent releases of large-scale, gridded data sets, new and novel methods of event identification and description are required.

However, it should be noted that these types of analysis are currently computationally expensive and may produce large data quantities (many terabytes). Fully-probabilistic risk analysis can be explored, but to apply sensitivity analysis on top of the fully-probabilistic approach is, though feasible, computationally still very demanding and, therefore, limited. Gridded data is used throughout this thesis and so this thesis was limited by the computational capacity that was available.

1.4.2. The tracking of discharge waves

In Chapter (4), an event-based, multi-site approach will be used to capture the spatial dependence of local discharge peaks in a gridded, pan-European data set. It is not trivial to capture a spatial dependence structure within a multi-site framework. Two main groups of methods are currently applied in the literature; block-based methods and event-based methods. An overview of studies sorted by these two methods is given in Table (2.3).

When using a block-based method, a comparison can be made between descriptors occurring at different locations but within the same block of time. The main benefit of this approach is its simplicity. It is mainly usable for large temporal blocks, in which travelling times are relatively small, such that it is likely that, within the temporal length of each block, complete spatio-temporal events are captured at the different locations considered. However, with an increasing level of detail (i.e. higher resolution, smaller temporal blocks), it becomes more and more unlikely that entire events are captured within each block, so it becomes more and more likely that events are split between different blocks.

Event-based studies are designed to address the level of detail were block-based methods do not suffice. However, when using an event-based method, connecting what happens at different locations may become difficult. What if the events occur at different times (time lags)? Therefore, across locations, which events are to be connected? To address these simple ques-

tions, events have to be matched across locations. This could be a manual job, with the application of expert judgement. However, when considering a large number of locations and/or descriptors, this approach will not be a feasible, such that matching approaches that make use of algorithms are required. A simple matching approach would be to use dynamic time windows (which differ from blocks in that there may be gaps between the time windows), which can be used specifically when time lags between events occurring at different locations are small. The time windows can be lagged across locations, when the time lags between events at different locations are well-known and fairly constant. When such assumptions no longer hold, more sophisticated approaches are required.

A highlight in this analysis is the proposed matching method, where local events are connected by the tracking of discharge waves in all major river basins, thereby not having to make any assumptions regarding time lags between different locations.

1.4.3. A large-scale dynamic spatio-temporal weather generator

In Chapter (5) a new type of dynamic generator is developed. In Section (5.3), the pioneering development work provides insight on the feasibility of the new approach. In Section (5.4), the possibility of multi-source analysis is demonstrated using the dynamic type of generator with a few extensions.

The dynamic type of generator has overcome the limitations of static spatio-temporal probabilistic analysis, shortly discussed in Section (1.3.2). It has academic benefits, already shortly discussed in Sections (1.3.4) and (1.3.5). Future research will have to point out if the introduction of dynamic spatio-temporal probabilistic analysis comes with new limitations.

1.5. Outline of the thesis

In Chapter (2), the existing literature is reviewed. First, a high level view is applied on how studies fit into the model chain. Second, the literature is organised based on how the studies dealt with space and time. Third, it is analysed which statistical methods were used in the studies.

In Chapter (3), as a result coming forth from the literature review, a framework for event-based flood hazard generation is proposed. This chapter supports Section (1.3.1).

In Chapter (4), a static spatio-temporal probabilistic analysis (SSTPA) is applied to discharge for the European continent. Events at different locations are matched using a new wave tracking methodology. A problem with event-based SSTPA is identified and discussed. This chapter supports Sections (1.3.2), (1.4.1) and (1.4.2).

In Chapter (5), a dynamic spatio-temporal probabilistic analysis (DSTPA) is applied to atmospheric variables for a large chunk of the world. The spatio-temporal footprints of precipitation fields are captured and summarised with spatio-temporal descriptors. The descriptors are then used partly for classification and partly for multivariate, statistical analysis. From the resulting synthetic descriptor sets, synthetic events are reconstructed and placed in space-time continua, where the observed atmospheric configuration is more or less respected. This chapter supports Sections (1.4.1) and (1.4.3). The dynamic spatio-temporal methodology is further expanded by considering multiple

sources (precipitation, pressure and wind).

In Chapter (6), a framework is introduced to check the performance of generators. Local probabilistic analysis and static spatio-temporal analysis are contrasted with dynamic spatio-temporal probabilistic analysis. A broad range of performance indicators is introduced and subjected to a sensitivity analysis with a few generator settings. This chapter supports Sections (1.3.3) to (1.3.5).

In Chapter (7), the work is concluded. A summary is provided of the academic contributions, where the connections between these contributions are described. An outlook for future research is provided, with a preliminary proof-of-concept of large-scale, flood hazard time series. Finally, the main conclusions of the different chapters are connected and combined.

2

Literature review

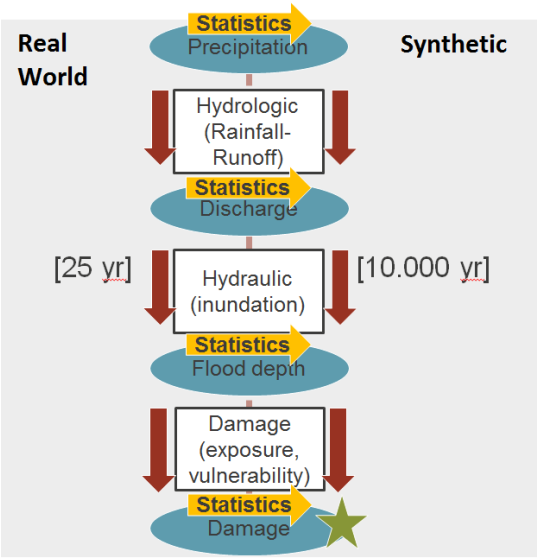


Figure 2.1: All possible routes through a flood risk modelling chain.

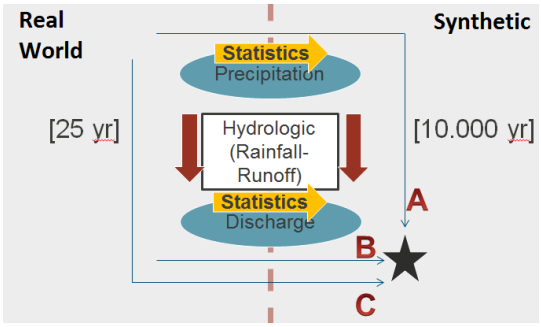


Figure 2.2: Different routes to obtain a fluvial flood hazard estimate comprising discharge scenarios (black star): A.) statistical methods are applied to observed precipitation data (and/or other atmospheric variables), after which the resulting synthetic data are used to drive a hydrological model, B.) statistical methods are applied directly to the observed discharge data, C.) observed precipitation (and/or other atmospheric variables) are fed into a hydrological model, after which statistical methods are applied to the (real-world) modelled discharge data.

2.1. Flood risk model chain

2.1.1. Set-up of the model chain

One of the first decisions to be made in a flood risk analysis is the choice of modelling chain and the route through the chain. This was introduced in Section (1.1.3). For this thesis, no input-output models are applied. However, the set-up of the modelling chain determines to which variable statistics are applied. So which model chains are used in the literature?

An example of a model chain for fluvial flooding is displayed in Fig. (2.1). For the analysis of fluvial flood risk, most studies apply statistical methods to discharge [2, 17, 31, 53, 61, 64, 65, 77, 102, 103, 105]. Others apply statistical methods to precipitation [6, 13, 17, 21, 25, 39, 47, 85, 86, 103], thereby introducing the additional requirement of setting up a hydrological model.

2.1.2. Routes, methods and the issue of validation

All Roads Lead to Rome

Flood risk requires the full spectrum of hazard scenarios that may occur, with associated probabilities. The discretised approximation of this full spectrum of hazard scenarios will be referred to as 'the synthetic data' throughout this thesis. Studies that aim to generate synthetic data sets run into the issue of validation. This is a difficult topic, since synthetic data sets have no observed data ('ground truth') to be directly validated against. However, what can be done is a comparison of statistical products. It would strengthen the credibility of an analysis when different approaches would lead to similar results. So which different approaches can be found in the literature?

Three ways are distinguished to vary the general approach. First, different methods are used. Basinger, Montalto, and Lall [6] and Hundecha, Pahlow, and Schumann [47] generated synthetic precipita-

Route	Studies
A and B	Cameron et al. [17], Cowpertwait et al. [25], Winter et al. [103]
C	Pappenberger et al. [64], Winsemius et al. [102], Alfieri et al. [2], Paprotny and Morales-Nápoles [65]

Table 2.1: Overview of studies applying routes A,B and/or C in Fig. (2.2).

tion and validated their product by checking if the synthetic precipitation showed the same distribution as lower resolution block maxima. Eastoe and Tawn [31] did a similar check for discharge, where their main focus was the use of frequency models. Cowpertwait et al. [25] compared both synthetic precipitation and synthetic, modelled discharge to the observed. Generally, the simpler methods (like block maxima) are used as a 'dummy ground truth' for validation. Note that, within the context of extreme value analysis, even when using the simplest methodology available, models are used (such as distributions). Second, different routes are taken through the model chain. Within the hazard component of fluvial flood risk assessment, different routes have been taken in the literature which meet in synthetic discharge data. Fig. (2.2) shows three main routes taken through this step. An overview of routes taken is displayed in Table (2.1). Cameron et al. [17], Cowpertwait et al. [25], and Svensson, Kjeldsen, and Jones [86] generated synthetic precipitation (in combination with other variables), which they ran through a hydrological model to obtain synthetic, modelled discharge (route A), the distributions of which they compared statistically to the distributions fitted directly to the observed discharge data (route B). Within the context of large-scale (continental) flood risk assessments, several studies have explored route C. Alfieri et al. [2], Pappenberger et al. [64], Paprotny and Morales-Nápoles [65], and Winsemius et al. [102] used distributed precipitation data to drive distributed hydrological models, after which they captured the distributions of the resulting discharge for all sites (grid-cells). Vorogushyn et al. [93] concluded that current methods simply assemble local results to form the larger picture, thereby 'contravening the fundamental principles of the flood risk system functioning because they largely ignore basic interactions and feedbacks between atmosphere, catchments, river-floodplain systems, and socio-economic processes'. They called for an 'Evolutionary leap in large-scale flood risk assessment'. Third, the settings with which the synthetic data is produced are varied. This approach is applied throughout the literature and leads to examination in the form of sensitivity analyses. In addition, different set-ups along the modelling chain can provide synthetic data sets that can be compared. An example can be found in Winter et al. [103].

Within the context of large-scale flood hazard studies, two developments would be helpful. First, the development of a spatially consistent, large-scale generator of discharge events, which would extend existing methodology by capturing and modelling the dependence structure (joint distributions) of large-scale events. This would be the second step (yellow arrow labelled statistics) of route C. Second, the development of a consistent, large-scale (spatio-temporal) precipitation gen-

erator, which would be the first step (yellow arrow labelled statistics) of route A.

2.1.3. Continuous versus event-based

Two modes are distinguished to run the models in the flood risk chain. First, continuous mode, for which initial conditions have to be formulated only once for a run, such that models run a long time span continuously. Running the models in continuous mode has the advantage that pre-event conditions such as, for example, soil moisture state or pre-event river flow are explicitly simulated by the model chain. Simulation in continuous mode requires continuous forcing and can be computationally expensive. This procedure results in continuous output data, in which there will be the occasional flooding. Boughton and Droop [12] provided a review of 'continuous simulation for design flood estimation' some time ago. Falter et al. [33] provided a proof-of-concept to run an entire flood risk model chain in continuous mode. Second, simulation in event-based mode. Event-based simulation requires an event catalogue and initial conditions for each event, introducing the challenge to include the influence of antecedent conditions [8, 9, 67]. This procedure results in the direct output of (flood) events. Some examples of extensive analysis of flood events can be found in Fischer, Schumann, and Bühler [35], Tarasova et al. [87], and Tarasova et al. [88].

Another matter, which has to be clearly distinguished from the mode of model simulation, is the mode of the methodology to obtain the synthetic data. This is where the switch in model chain forcing is made from observed (reality) to synthetic (hypothetical/scenarios). Here, two groups of methods are separated. First, continuous-based generation, for which statistical methods are directly applied to continuous data series. This procedure results directly in synthetic, continuous forcing. A recent example can be found in Brunner and Gilleland [14]. Second, event-based generation, which is a procedure that results in a synthetic event catalogue. A recent example can be found in Quinn et al. [68]. These two groups may appear to directly correspond to the two modes of model simulation, but they correspond to the preceding step and switches can be made. For example, gaps may be filled between events, such that an event-based generation method produces synthetic data to run the model chain in continuous mode. Or, the opposite, events may be extracted from a continuous-based generation method, such that the model chain can be run event-based.

Several generators are used that are continuous-based. For example, in a recent review of precipitation generators, Vu et al. [96] compared five models for nine sites around the world and compare the generated continuous precipitation series statistically to observations at gauging stations. The methods in these precipitation generators are continuous-based, with the exception of the 'RainSim' generator [15], which is event-based and will be further discussed in Section (2.2.3). Event-based generators are slightly more rare and sometimes hard to spot. The title of Cameron et al. [17] may seem to suggest that they used a continuous-based generation method. However, they used precipitation event descriptors (duration and mean intensity) and generated a synthetic catalogue, after which which they fitted normalised storm profiles to reconstruct synthetic storms. They did not mention ex-

plicitly that they placed events in time, but they used a 'storm inter-arrival time' descriptor, which implies that they were concerned with putting their generated events on a (synthetic) time line. As mentioned, the option exists to fill the gaps between events (generally with non-event data, i.e. low values in the context of flooding). Although event-based, this is an example of where an event-based generator produced continuous, synthetic data.

In this thesis, the focus lies with event-based methods to generate synthetic data. Whether methods are to be applied to fill the gaps for the data to become continuous, depends on the variables considered. Generally, discharge waves can be well characterised as events and can therefore be modelled reasonably well in event-based mode, with perhaps not too much influence of the initial conditions in the river system. In contrast, whether precipitation events may trigger relatively small or large pluvial or fluvial flood events, will largely depend on the state of the catchment when the precipitation event occurs (i.e. will depend on the antecedent conditions). In this case, running the flood risk model chain in continuous mode may be more appropriate.

2.2. Time and space

2.2.1. Local probabilistic analysis

Historically, observations have been made at specific locations. For example, discharge gauging stations are placed at specific locations along rivers. Methods for local probabilistic analysis, defined in Appendix (A.1.4), were extensively developed in the previous century. In the UK, simple methods came up such as the flood index method [49], for which the mean 'flood' (discharge) $\bar{Q}$ was captured, which was multiplied with a location dependent correction factor to be used as a design 'flood' (discharge). The strength of this method is that it is a very pragmatic approach, easy to apply. Another example is the method of duration frequency curves, where the aim was to capture the dependence structure of duration, frequency and the magnitude of a specific variable with specific lines. For discharge, these would be discharge(peak)-duration-frequency (QDF-)curves. For precipitation, these would be intensity-duration-frequency (IDF-)curves.

In more recent methodology, joint distributions are considered. The motivation for this development would be that scatter is generally observed in joint-distributions. With significant scatter, the dependence structure cannot be appropriately modelled with a (few) curve(s). The modelling of dependence will be discussed in Section (2.3). However, to be able to model dependence, first dependence has to be captured. For that, data clusters have to be identified and described. Which methods of cluster identification and description are used in the literature?

A large group of studies applies the (simple) method of blocks to identify data clusters [2, 6, 13, 47, 61, 64, 65, 77, 85, 102]. The main benefit of this approach is that it is easy to implement and allows to rapidly capture basic distributions. For block data clusters, the most widely used descriptor is the maximum [2, 13, 17, 25, 31, 39, 47, 61, 64, 65, 77, 86, 102]. Some use sums [6] or apply their statistical methods to the original data values [6, 47, 85]. Maybe worthwhile to mention is the GRADE [70] methodology, since it is currently operational in the Netherlands to derive the design discharge in the country in the Rhine and

Meuse rivers. In this method, spatial images of daily means are reshuffled to form sequences of several days, that are different (and potentially more 'extreme') than observed before. This implies that (on a daily basis) no values can be generated that are more extreme than (ever) observed, which is unusual for generators. The sequencing is based on particular (local) indices of similarity, which means it may not be straightforward to extend the methodology to large spatial scale.

Another large group of studies applied event-based methods to obtain the data clusters [17, 21, 25, 31, 39, 53, 86, 105]. For events, a large range of descriptors can be found in the literature, of which only a few will be discussed. Most used peaks [31, 39, 105]. Other descriptors include, but are not limited to, the mean [17, 86], duration [17, 86] and original data values [21, 53]. Peaks-over-threshold (POT) [20] is the most widely used method in the literature and comprises of both steps in one, according to the framework of this thesis, which will be discussed in Section (3.3). First, a threshold is set, where events are the data clusters exceeding the threshold. Second, the events are described by their peaks (maximum values).

Patterns, which will be be discussed in Section (2.3.1), are to be captured between data clusters (and their descriptors). In this process, data clusters are compared, which means that they have to be comparable. The process of grouping comparable data clusters is referred to in this thesis as 'classification'. Cameron et al. [17] divided the precipitation events into classes of different duration. Cooley, Nychka, and Naveau [21] separated their gauge data based on climate, for the fitting of spatially coherent distribution parameters. Hundecha, Pahlow, and Schumann [47] aimed to address seasonality by splitting their data set based on month of occurrence. Cowpertwait et al. [25] made a distinction between convective and stratiform storm types, which they used for their spatio-temporal point process and model fitting. Some studies addressed the seasons by splitting summer and winter events ([54, 77, 86]). Implicitly, all studies that compare data clusters on a per-location basis, use location for classification. Mentioning this may seem redundant at this stage, but, when moving to dynamic events in Section (2.2.3), the (exact) location will no longer be used for classification. This is why the classification procedure is extra important for dynamic events.

An overview of block-based and event-based methodology for LPA is provided in Table (2.2).

Table 2.2: Studies applying local probabilistic analysis. A general distinction is made between studies applying a block-based method and studies applying an event-based method. Acronyms can be found in Table (A.1).

Type	Method	Variable	Descriptors	Study
blocks	annual	discharge	M	Cameron et al. [17]
	annual	discharge	M	Merz and Blöschl [61]
	daily	rainfall	V	Hundecha, Pahlow, and Schumann [47]
	annual	rainfall	M	Hundecha, Pahlow, and Schumann [47]
	daily	rainfall	V	Srikanthan and Pegram [85]
	annual, monthly, daily	rainfall	Long list, see paper	Srikanthan and Pegram [85]
	daily	rainfall	V	Basinger, Montalto, and Lall [6]
	daily, annual	rainfall	S, NW	Basinger, Montalto, and Lall [6]
	annual	discharge	M	Eastoe and Tawn [31]
	annual	precipitation	M	Ghosh and Mallick [39]
	annual	discharge, storage, depth	M	Pappenberger et al. [64]
	annual	rainfall	M	Cowpertwait et al. [25]
	annual	discharge	M	Cowpertwait et al. [25]
	annual	discharge	M	Svensson, Kjeldsen, and Jones [86]
	annual	discharge	M	Winsemius et al. [102]
	annual	discharge	M	Alfieri et al. [2]
	annual	precipitation	$M(S$, 3-day)	Bracken et al. [13]
	annual	discharge	M	Paprotny and Morales-Nápoles [65]
	3-day	discharge	M	Schneeberger and Steinberger [77]
events	thresholds	precipitation	μ, D, $IEAT$	Cameron et al. [17]
	thresholds	precipitation	V	Cooley, Nychka, and Naveau [21]
	thresholds (unclear)	discharge	$P \pm 3$	Keef, Tawn, and Svensson [53]
	thresholds	discharge	P	Eastoe and Tawn [31]
	thresholds	precipitation	P	Ghosh and Mallick [39]
	disks (?)	rainfall	V, R	Cowpertwait et al. [25]

thresholds	rainfall	μ, D, $IEAT$, SMD_0, Dis_0	Svensson, Kjeldsen, and Jones [86]
thresholds (runs)	discharge	P	Wyncoll and Gouldby [105]
NR	discharge	P	Diederen et al. [28]

2.2.2. Static spatio-temporal probabilistic analysis

A logical step after local probabilistic analysis is to extend single-site analysis to multi-site, thereby addressing the challenge of spatial dependence [50]. Connecting what happens at different sites has received a lot of attention in the last two decades, and still does. To recall from the previous Section (2.2.1), two main groups of cluster identification methods were distinguished for local (temporal) probabilistic analysis: block-based and event-based. When extending the analysis to include space, things become slightly more complicated, so that for spatio-temporal probabilistic analysis three main groups can be distinguished: 'blocks'; 'block-events'; and 'dynamic events', which will be discussed in the next Section (2.2.3).

Merging single-site blocks to multi-site blocks may seem conveniently straightforward. For example, Ghosh and Mallick [39] apply this approach to monthly (blocks of) precipitation data. However, the limitations of this approach become evident with a decreasing temporal length of the blocks, such as weekly, daily or even hourly. For example, a discharge wave travels from site A to B with a 1-day time lag. Using multi-site blocks, the maximum in day 1 at site A may be compared with the maximum in day 1 at site B, and day 2 at A may be compared with day 2 at B. Since the wave is travelling, observations at site A and B may respectively show 'high-low' and 'low-high'. The blocks method would fail to capture the positive correlation (high-high) that could be found when properly taking the time lag into account (i.e. by comparing the maximum of day 1 at site A with the maximum of day 2 at site B). Instead, it finds two data points that show negative correlation (high-low,low-high). The probability that time lags are captured within a multi-site block depends on the (temporal) size of the time lag as compared to the (temporal) size of the block. The size of the time lag depends on the celerity of the physical phenomenon and on the distance between the sites. Examples for (moving) precipitation (fields) would be similar as for discharge, except that the celerity of precipitation fields is significantly higher. Given a study of a phenomenon with a particular celerity, blocks may be reasonable for near sites but may not work for distant sites. It can be concluded that the performance of multi-site, block-based methods becomes poorer with an increasing spatial scale (i.e. large distances) and/or an increasing temporal resolution (so not so much monthly or yearly blocks, but when moving to daily, hourly or sub-hourly). Wilks [101] provided results where their stochastic generator 'under-represents' the 1-day lag correlations of daily precipitation amounts. Srikanthan and Pegram [85] present better results for this particular aspect, since they showed that they managed to capture the 1-day lag correlations of daily precipitation amounts reasonably well. However, with reasonable results for a spatial domain of about $500km$ diameter at a daily resolution, it is unclear if this method can be extended to a larger domain or to hourly resolution. Several comparisons of precipitation generators of the type block-based SSTPA can be found in the literature [19, 60, 84, 89, 100]. The performance of these generators is discussed based a large range of statistical indicators (such as distributions, annual cycles, wet/dry spells, etc.). Although some of these reviews include checks of spatial dependence (using pair-

wise correlation), they do not carefully consider the capturing of dependence, which is problematic for all block-based approaches.

Another group of studies applies a local event-based method, after which concurrent values are chosen for all other locations. This type of analysis may be referred to as 'block-events'. An overview of such studies is provided in Table (2.3). The block-events are event-based in time, but block-based in space. As a result, these block-events are different from blocks in that not the entire time series of data is sliced into (regular) blocks, but that there are (irregular) time intervals in between the blocks. For these studies, about the same limitations apply as for the block-based studies concerning the capturing of time lags (i.e. they are not well-suited for large spatial scale and/or high temporal resolution). However, with block-events the option that time windows can be be moved back or forward in time, allows to better capture the relevant data clusters (and time lags). On a higher level of detail; if time lags are known and 'constant' between different sites, blocks may be shifted forward or backward per location. This may work decently well for discharge within a single river basin, but not for precipitation, since precipitation fields move in different directions (i.e. time lags can be positive or negative). In addition, this manoeuvrability decreases with an increasing number of sites addressed.

What 'blocks' and 'block-events' have in common, is that a set of sites of interest is predefined, where for the statistical analysis each site is considered in each block. Therefore, these types of (multi-site) analysis are referred to as 'static spatio-temporal probabilistic analysis' (SSTPA), as defined in Appendix (A.1.5). The spatial extent of each block-event is equal to the entire spatial domain. This is a limitation for all event-based SSTPA, that was shortly discussed in Section (1.3.2) and will be elaborated in Section (4.5). So studies that apply event-based SSTPA are limited to the use of blocks or block-events. However, this is not really what practitioners have in mind when they think of events. Events are generally thought of as something that occurs then here, then there, etcetera. Events are often conceptualised to be dynamic.

2.2.3. Dynamic spatio-temporal probabilistic analysis

In the context of flooding, it is hard to find studies that work with dynamic events, i.e. that apply 'dynamic spatio-temporal probabilistic analysis' (DSTPA), as defined for this thesis in Appendix (A.1.6). DSTPA is an event-based approach, where 'dynamic events' are conceptualised to be clusters of related data that freely occur at different locations in space and time with (potentially) time varying spatial extents (and therefore spatially varying durations). Physical phenomena, such as moving precipitation fields, often show a dynamic, progressive behaviour, rather than a static behaviour. Rainfall organization and movement within basins is an essential control of flood response and in particular of hydrograph timing [87]. DSTPA allows dynamic behaviour to be captured, thereby potentially providing a more physically meaningful event definition (OBJ3 that will be formulated in Section (3.2)). In addition, for a particular site of interest, moving events allow inference from data occurring outside that particular site, which was shortly discussed in Sec-

Study	Method
Keef, Tawn, and Svensson [53]	Identification of local event peaks (POT), expansion to a week by day of peak ± 3 days, concurrent values at all other sites.
Wyncoll and Gouldby [105]	Identification of local event peaks (POT), concurrent values at all other sites.
Schneeberger and Steinberger [77]	Identification of local blocks of three days, concurrent values at all other sites.

Table 2.3: A sample of a few studies that applied block-events.

tion (1.3.4) and which will be elaborated in Section (6.5.1).

To capture dynamic events, one has to let go of the (static) idea that a set of sites of interest can be predefined and that events will occur exactly at those predefined sites. One explanation for the lack of studies may be that a key ingredient for DSTPA is the tracking of progressive, physical phenomena. Tracking requires (preferably gridded) data sets of sufficiently high spatial and, in particular, temporal resolution. Discharge is typically available for a limited set of discharge gauges. For precipitation, data sets with hourly resolution have become available fairly recently [72].

A number of studies performed (subroutines of) DSTPA. In the study area of droughts, which is inversely related to floods, some first steps of DSTPA were applied several years ago: Andreadis et al. [3], Corzo Perez et al. [22], Haslinger and Blöschl [44], Sheffield et al. [82], and Vidal et al. [92]. In particular, Corzo Perez et al. [22] studied gridded discharge output from a global hydrological model and captured low discharge anomalies. They tracked drought events by capturing connections in space using an algorithm that looked at neighbouring pixels and subsequently using overlap between time steps. The result was a catalogue of observed dynamic events of low discharge. It should be mentioned that drought events move much more slowly than flood events and that, therefore, drought events can be tracked in data sets with a much lower temporal resolution than required to track flood events. Compared to low discharge anomalies, precipitation fields move around at higher speeds, such that higher resolution data products are required.

Within the context of flood related DSTPA, Cowpertwait [23] extended the single-site pulse process proposed by Rodriguez-Iturbe, Cox, and Isham [71] to the spatio-temporal Neyman-Scott pulse process (STNSPP). The original concepts in these studies were further extended and developed into a well-established rainfall generator called 'RainSim' Burton et al. [15]. An analytical model was used to generate synthetic precipitation events, in the form of disks with particular radii. The concepts they built on originated from the previous century, in which computational power was not what it is today. Therefore, it seems a sensible approach that they

drew synthetic events from an analytical model, rather than built synthetic events from observed events in a discretised manner. STNSPP was applied to place the generated events in space and time. With a specific calibration procedure (using numeric optimisation), they focussed on placing these synthetic events in such a way that the continuous, synthetic data generated respected particular statistical properties of the observed continuous data. This was achieved by formulating a set of statistical properties as an objective function. Vallam and Qin [89] concluded that RainSim compares favourably to three other generators (of the type SSTPA) concerning the capturing of spatial dependence. However, what mainly sets this approach aside from other approaches, is that there is no statistical process at the heart of this method in which the statistics of events are directly addressed. This statistical process will be discussed in the next Section (2.3).

2.3. Statistical methods

2.3.1. General patterns

Multivariate statistics refers to techniques used to capture patterns in matrices. The columns in the matrix represent stochastic variables, which in this thesis will be event descriptors. To clarify this with an example, in Section (4.3.2) we will construct a matrix in which each column represents peaks of discharge events at a particular site. With 298 sites and 428 identified events, this matrix has 298 columns and 428 rows. In Section (5.3.2) we construct multiple matrices, where each matrix represents a particular class of precipitation events. These matrices have 3 columns; peak, volume and extent - and have a variable numbers of rows, depending on the number of events captured per class.

For a particular matrix, three base patterns are often considered in practice. First, marginal distributions, which are captured per column; Section (2.3.2). Second, dependence between columns, which is captured in row-wise direction; Section (2.3.3). Third, dependence between rows, which is captured in column-wise direction; Section (2.3.4). Finally, an overview is provided in Section (2.3.5), where statistical patterns are listed according to study in Table (2.4).

2.3.2. Marginal distributions

Description One of the main patterns to capture for statistical analysis is the distribution of each variable [20]. This is done similarly in univariate analysis, where there is only one variable, and in multivariate analysis, where there are multiple variables. For multivariate analysis, matrices of descriptors are defined, but distributions are still fitted per column per matrix.

Generalised extreme value distribution The generalised extreme value distribution is a distribution for entire populations, with three model parameters; location, scale and shape. It is typically used in the literature to describe populations of block maxima.

Generalised Pareto distribution The generalised Pareto distribution is a distribution for tail-end populations, with a threshold (location) and two model parameters; scale and shape. It is often used in the literature to describe populations of peaks of events, which are typically obtained using the peaks-over-threshold event identification method.

2.3.3. Row-wise dependence

Description Row-wise dependence refers to joint distributions [32]. For an analysis to be considered multivariate, this pattern has to be addressed.

Copulas A popular theme in the literature is the applications of copulas [38, 62]. An approach is copular when the matrix of descriptors is marginally transformed to uniform, i.e. the values within every column are transformed to a uniform distribution with range $X =]0, 1[$. This allows the marginal distributions and the row-wise dependence structure to be separately studied.

Heffernan and Tawn (2004) In this thesis, Heffernan and Tawn [46] (HT04) provide the model of choice for extremal dependence. The essence of this model will be described in Section (3.4.4).

2.3.4. Column-wise dependence

Description Three different column-wise dependence patterns may be distinguished.

First, short-term dependence, referring to auto-dependence. Whether or not this pattern is present may be studied using auto-correlation (internally correlating the values of a particular stochastic variable to the previous values of that same stochastic variable). If strong auto-correlation patterns are found, methods for dependent sampling should be considered (for example, Markov chains). This pattern may be avoided by capturing (nearly) independent events, i.e. making sure that low auto-correlation exists in the descriptors, which would then allow random sampling.

Second, medium-term dependence, referring to a repeating pattern (like seasonality). Whether or not this pattern is present may be studied by splitting the data sample, for example by season. A pragmatic way of dealing with this pattern could be to take season into account at the classification step.

Third, long-term dependence, referring to a trend in the data series. Since systems change, observations may hold significant trends. If that is the case, the matrix may have to be de-trended, to find appropriate distributions for the particular system state under study. However, since trends are not easy to capture and are unsure in extrapolation, such an additional data transformation may introduce additional uncertainty.

2.3.5. Discussion

To explore the range of hypothetical scenarios, most studies fit marginal distributions to observed descriptors (with the notable exception of Cowpertwait [23] and follow up work like RainSim). For block maxima, the generally used distribution is the Generalised Extreme Value (GEV) distribution [13, 17, 31, 39, 47, 65, 86], of which the Gumbel [2, 64, 65] and the Fréchet [61] are family types. For event-based methods, the generally used distribution is the Generalised Pareto (GP) distribution [17, 21, 31, 39, 47, 77, 86]. Cameron et al. [17] randomly sample duration, mean intensity and inter-event-arrival-time separately, thereby assuming complete independence. The large-scale studies so far capture distributions per pixel [2, 64, 65, 102].

Event-based studies tend to avoid short-term column-wise dependence (autocorrelation), by aiming to capture independent events. Having mid-term dependence (seasonality) in the multivariate matrix is typically avoided by splitting the data

set by month or season. Long-term dependence may be neglected with the assumption of stationarity. Several studies use a multivariate dependence model. Bracken et al. [13] and Ghosh and Mallick [39] use copulas. Keef, Tawn, and Svensson [53], Schneeberger and Steinberger [77], and Wyncoll and Gouldby [105] use the HT04 model. In contrast, continuous-based studies that do address short-term column-wise dependence (autocorrelation) generally do not address the row-wise dependence. Two examples are Srikanthan and Pegram [85] and Basinger, Montalto, and Lall [6]. Only Ghosh and Mallick [39] tick the box for all three patterns: marginal distributions, row-wise dependence and column-wise dependence. However, they analyse monthly data, which shows only weak temporal dependence. It is unclear if their methodology can readily be applied to data with a higher temporal resolution, which would introduce a lot of additional difficulty.

To simultaneously capture statistical patterns is challenging, such that patterns are prioritised based on the application. In Table (2.4), an overview is provided of statistical methods applied in the literature. Distributions are always important for extreme value analysis. For some applications, dependence between marginals (columns) has the second priority, like case studies of spatial dependence, in which the marginals may represent data at different sites. For other applications, the row-wise dependence is important, like case studies of short term sequences or of climate change. Generally, studies address a combination of distributions with either a row-wise dependence model (like copulas or HT04) or a column-wise dependence model (mainly Markov chains). To develop advanced methods that address all three main patterns, for a multivariate matrix in which all three patterns are prevalent, is a current challenge for the statistical community. The current stance of this challenge is important for applied studies (such as flood risk assessments), because it provides feedback on (and limitations for) the methodology applied to set up the multivariate matrices.

Within the context of large-scale studies, Alfieri et al. [2], Pappenberger et al. [64], Paprotny and Morales-Nápoles [65], and Winsemius et al. [102] address distributions per pixel, but do not consider either dependence structures. This observation has pointed towards an event-based approach for large-scale discharge in this thesis, for which the multivariate approach will be extended to high dimensionality.

Table 2.4: An overview of statistical methods applied, Section (2.3.5). Acronyms can be found in Table (A.1).

Study	Marginal distributions	Dependence (Row-wise)	Dependence (Column-wise)	Space-time
Cameron et al. [17]	GP(ML)			
	GEV(ML)			
Cooley, Nychka, and Naveau [21]	GP (ML, SGP)			
Merz and Blöschl [61]	Weibull			
Keef, Tawn, and Svensson [53]	Gumbel	HT04		
Hundecha, Pahlow, and Schumann [47]	Gamma, GP	AR		
	GEV			
Srikanthan and Pegram [85]	Gamma, Exp, NST	DW	MC	
Basinger, Montalto, and Lall [6]			MC	
Eastoe and Tawn [31]	GEV, GP			PP
Ghosh and Mallick [39]	GEV	copula		
	GP	copula	MC	PP
Pappenberger et al. [64]	Gumbel(L)			
Cowpertwait et al. [25]				STNSPP
	Emp			
Svensson, Kjeldsen, and Jones [86]	Exp, GP			
	GEV (L)			
Winsemius et al. [102]	?			
Alfieri et al. [2]	Gumbel			
Wyncoll and Gouldby [105]	Emp, GP	HT04		
Bracken et al. [13]	GEV(ML, SGP)	copula		
Paprotny and Morales-Nápoles [65]	Gumbel(ML)			
Diederen et al. [28]	GP(PML)	NP, HT04		

2.4. Conclusions

To calculate flood risk estimates, a chain of models is used (Section (2.1.1)). Different statistical methods and different routes to move through the chain are possible, where the transition from observed (reality) to synthetic (scenarios) is made for different variables (Section (2.1.2)). Validation of synthetic data is a general issue. Alternative methods and routes cannot provide a 'baseline truth', but a comparison between different approaches may provide confidence in analyses. However, references to comparisons as 'validation' appear regularly in the literature, with little contextual discussion, such as on routes and methods. A general discussion could be useful that focusses on the quality of the flood risk assessments resulting from the different methods and routes. An example can be found in Winter et al. [103]. To advance this discussion, methodology should be further developed and extended for all possible routes. This would be useful for both continuous-based and event-based methodology (Section (2.1.3)). In this thesis, the focus lies with methodology for event-based generation methodology.

Most studies, including recent studies, have focussed on the analysis of local data (Section (2.2.1)). A number of studies has connected what happens at multiple sites by applying a SSTPA (Section (2.2.2)). With SSTPA, the small (regional) scale has been well explored, but not so much the large (continental) scale. Event-based methods for a regional scale SSTPA cannot simply be extended to a continental scale. Event-based SSTPA is limited inherently by the fact that the spatial domain of every event is fixed to be the entire spatial domain considered (by definition, hence 'static'). The answer to the limitation of event-based SSTPA could be to turn towards event-based DSTPA (Section (2.2.3)). New data products with large spatial coverage (which are typically gridded, but which may also be highly dense networks of gauges) have started to allow the capturing of the dynamic spatio-temporal (wave-like) behaviour of the physical phenomena that cause flooding. However, only a limited of number of studies has applied (parts of the methodology required for) event-based DSTPA. Some DSTPA event identification methods were found in the drought risk literature. Overall, the progressive (wave-like) behaviour of physical phenomena, that move through a large spatial domain and that lead to flooding, remains barely addressed. This calls for the development of DSTPA.

In most studies a statistical step (Section (2.3)) lies at the heart of the methodology. Generated synthetic data should show the same statistical patterns as the observed data, since the observed data should be a likely subset of the (large set of) synthetic data. Three base patterns were identified, which the literature pointed out are hard to address simultaneously. Therefore, the choice of patterns to address appears to be intimately related to the application. In this thesis, the HT04 model was chosen for extremal dependence (Section (2.3.3)).

These observations have driven the research topic, which is the development of a large-scale, spatio-temporal generator of hazardous events. Insights gained from the literature review have helped develop a general framework of event-based generation, which will be presented in Section (3.3). No studies were found that address the large scale using the full chain of event-based methodology, that will be defined in this framework. Within the

context of event-based flood risk assessment, this thesis will deliver synthetic forcing with increased physical realism and will allow the study of the system response to large-scale floods. Therefore, the research topic (Section (1.2.2)) is logically embedded in the Marie-Curie project System-Risk (Section (1.2.1)), in which the influence of system behaviour on flood risk is studied.

3

Methodology

3.1. Introduction

3.1.1. The full stochastic process and generators

For decision making, different tools are used in practice. For relatively short-term decision making, forecasts may be used. In the context of flooding, (weather) forecasts could be produced in the order of magnitude of days or, at maximum, weeks ahead in time with a usable amount of certainty. After a certain point ahead in time, the uncertainty bands of (ensemble) weather forecasts start to widen, as a result of time integration and thereby the time integration of errors that are inherit in any weather model. Sometimes the question is raised of why to do risk analysis, given the large amount of uncertainty inherit in risk analyses. The simple answer would be that it is required to do risk analysis because of uncertainty, specifically because of the wide spread in uncertainty bands of forecasts that try to go far ahead in time. So, long-term decision making, in the order of magnitude of months to decades, will be applied on the basis of scenarios and can be expected to remain so in the foreseeable future. Large works of infrastructure, that may be built for the purpose of flood protection or mitigation, are designed to last for multiple decades if not centuries.

For flood hazard analysis, a stochastic process is simulated. This is an artificial construct. The general idea is that many hypothetical scenarios could potentially occur, of which certain scenarios actually will occur (reality). Given a context of scenario-based analysis, it is sensible to assign probabilities to scenarios. If no probabilities were to be assigned, decision makers would have no other option than to weigh every scenario in equal measure. This can be considered to correspond to a probabilistic approach with uniform probabilities. A uniform distribution is generally not a suitable distribution, since more extreme scenarios will occur with lower frequencies. Even though a substantial amount of uncertainty may be associated to risk analyses in general, and, in particular, to the proposed probabilities (or frequencies), it should be kept in mind that any probabilities assigned with extreme value theory will be better than uniform probabilities.

By fitting a stochastic process, we are trying to generate a more complete picture from a fragmented picture. As the name suggests, the full stochastic process is supposed to be the full picture, which implies that the statistics derived from the stochastic process, e.g. the mean or the extreme quantiles, should be smooth. Smoothness may be looked for both in space and in time Therefore, the fitting of the stochastic process involves smoothening techniques. A simple example is the fitting of a (smooth) distribution to a (rough) set of observed data records. Or the reverse thought: a particular (small) draw from a smooth stochastic process will be quite rough, but, if one keeps drawing, the drawn samples together become more and more smooth until eventually the full stochastic process is approximated. Analytical models represent an infinite amount of draws, but, given limited computational power and storage, at some point the stochastic process, as represented by a generator, will have to be accepted as sufficiently smooth. In this process, sparsity in the tail-end of the distributions will be pushed farther and farther towards the extreme, where extremely low probabilities in the tail-end should eventually render this sparsity of insignificant consequence. So, (a sufficient amount

of) smoothness can be considered a requirement for a stochastic process.

The full stochastic process is formulated based on available evidence, typically mainly in the form of observed data. Given a particular set of evidence, an infinite amount of stochastic processes could be fitted. So, from all available options, how do we choose a stochastic process? The general concept applied in this thesis is based on likelihood. This concept works approximately as follows. Many stochastic processes are considered and, given each considered process, the (approximate) probabilities of the draw (observed data) are calculated. The stochastic process is chosen of which the observed data is the most likely draw, i.e. 'maximum likelihood'. This implies we are dealing with a particular form or inverse modeling. Incorporation of evidence hopefully helps to limit the set of possible fits of stochastic processes and, thereby, leads to a clear maximum signal in terms of likelihood. This implies that incorporation of additional (relevant) evidence should lead to a reduction in equifinality and, as a consequence, a reduction in uncertainty. So, incorporation of more evidence should generally make a stronger case, which implies that as much (relevant) evidence should be incorporated as possible.

3.1.2. Event-based probabilistic analysis

To unveil the full stochastic process, the development of a generator is required. In Section (2.1.3), it was explained that generator methodology can be continuous-based or event-based. A large part of the discussion on which type of methodology to use, requires a dive into complex methodology, in which the details matter. However, for now, it will be attempted to explain why an event-based is chosen in this thesis by discussing a simple example of time series analysis.

Fig. (3.1) demonstrates a few typical ways of obtaining a set of descriptors from a particular bit of observed data, in this case a simple time series of river discharge. Fig. (3.1)a shows how a yearly maximum is extracted from a year of (daily) data records. As can be observed, much information is lost, as the resolution of the descriptor is very coarse as compared to the variability present in the observed data. Fig. (3.1)b shows how monthly maxima are extracted. The level of detail captured by the descriptors is vastly improved compared to the yearly. Some temporal dependence between subsequent maxima can already be observed in the resulting sample of maxima. An increased level of complexity will have to be dealt with, since, for a generator developed on the basis of these maxima, methods to capture seasonality may have to be included. Fig. (3.1)c shows how daily maxima are extracted. It is a step farther with the idea in mind that a more detailed description is better. However, it would probably be difficult - if not impossible - for a statistical generator to reproduce this type of data. A model using time integration would probably be more appropriate. This is the type of model that is used in the field of (ensemble) forecasting, more appropriately used for relatively short-term decision-making, as explained in Section (3.1.1). Fig. (3.1)d shows that events are defined and that, in this case, their peaks are extracted. By coincidence, the sample seems to resemble the monthly peaks and seems to contain a similar level of detail, but an important difference can be observed. In the context of flooding, the most important parts of the observed

data would be the large discharge values, i.e. the discharge waves. With event-based statistical methodology, the focus lies with these values. And if multiple separate discharge waves occur in a single month, not just one is picked, but all discharge waves in that month can be considered. Discharge waves could be more easily reconstructed from these event peaks than from the maxima. Producing some synthetic peaks and then drawing simple triangular discharge waves, with the largest value being the event peak, would already be a (half-) decent effort. On the downside, additional complexity can be found compared to the maxima. For example, for a generator based on these peaks, a secondary statistical process would be required to place reconstructed discharge waves back on a (synthetic) timeline.

Improving methodology and tackling the complexity of event-based probabilistic analysis is a key point in this thesis. All methodology in the upcoming chapters will be event-based. Having established that choice, the next question comes up. What objectives are to be achieved using an event-based generator?

3.2. Objectives of a generator

3.2.1. Objective 1 - Introducing the unobserved

Sparsity In the context of flooding, events are the (clusters of) extreme highs of system forcing, i.e. when much water is thrown into the system. In a limited sample of observed data, there will hopefully be only a few hazardous events, so they will be sparse. For higher extremes, hazardous events will be more sparse, because higher extremeness implies that, in a data sample of a given temporal length, there will be fewer events.

Addressing sparsity with a generator If a system can be described by a stochastic process, and the interest of the application lies with extremes, sparsity will play an important role. According to the model of the stochastic process, reality is just a particular sample drawn from the stochastic process. Each individual sample will be sparse, especially in the extremes. The draw from a stochastic process of one particular sample, will be different from the next sample drawn. The sparse data sample of observations of reality is the basis on which the full stochastic process is inversely modeled. Theoretically, it should be possible to address sparsity with the generator approach, as a (sufficiently large) synthetic data sample is supposed to represent the full stochastic process at a particular (discretised) point in time. As explained in Section (3.1.2), the methodology applied in this thesis will be event-based, so the generators will be designed to generate synthetic events.

Sufficient synthetic data For a particular synthetic data set, the more synthetic events are generated, the farther sparsity is pushed to the extremes. Therefore, the more extreme the events of interest are, the more synthetic events have to be generated. Very high extremes, such as the entire world being flooded, may appear to be interesting, when focussing only on the consequences. However, the more extreme a particular event, the lower the associated probability. Since risk is defined as the sum of consequences of scenarios multiplied by the probabilities of the scenarios, very low probabilities will become dominant at some stage. This means that when sparsity has been pushed to the domain where (low) probabilities are dominant, it

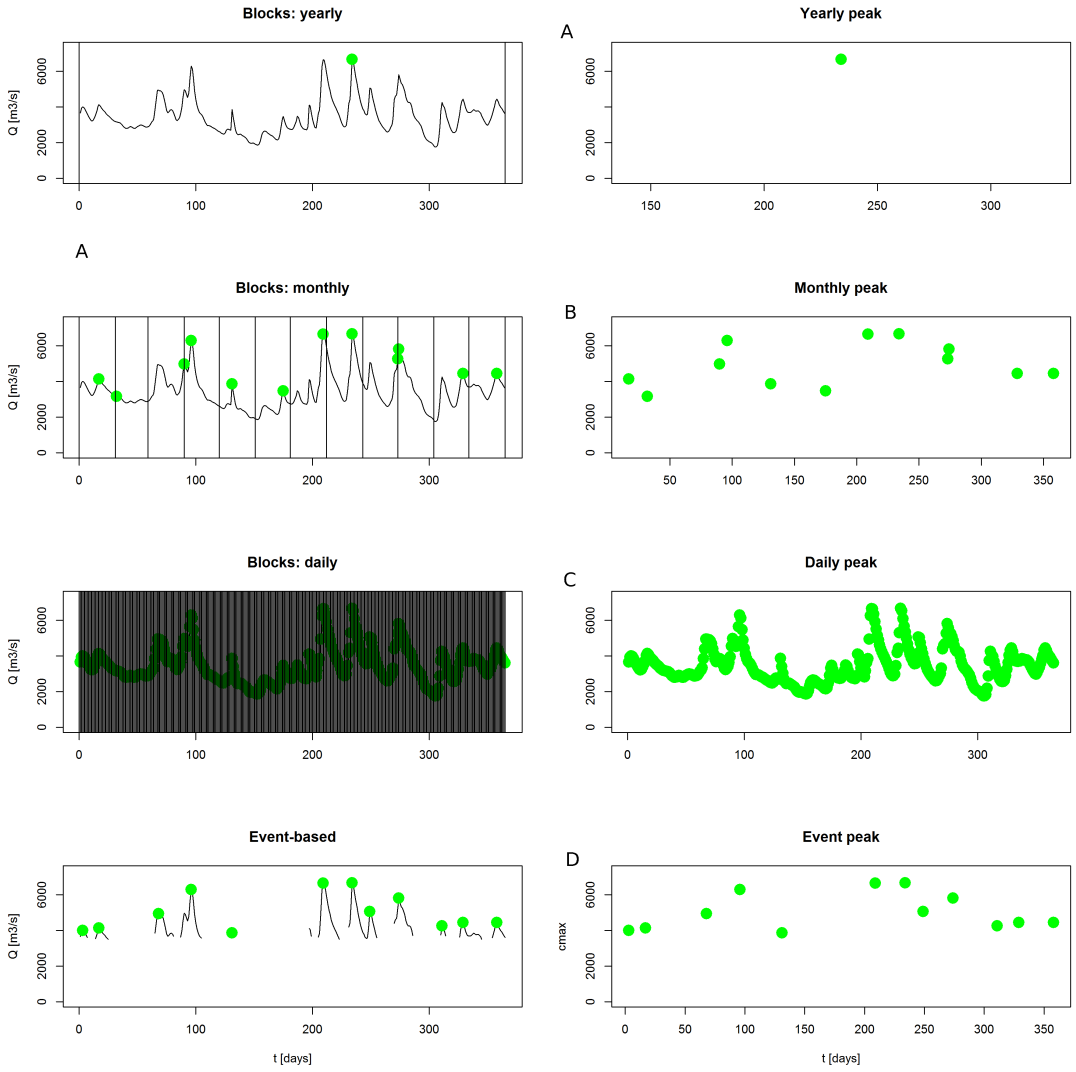


Figure 3.1: From block-based to event-based time series analysis. With yearly blocks, much information is ignored and lost. With monthly blocks, a time series is captured that is a mixture of extremes and low values. With daily blocks, a time series is captured that will be hard to reproduce by a stochastic model. With an event-based approach, a time series is captured of details of the extremes only, which is a good starting point for the stochastic model.

may be concluded that sufficient synthetic data has been generated. This is fortunate, because in practice, the generator approach will always hold some sparsity for high extremes, since computational limitations only allow the generation of a finite amount of synthetic events.

Unlimited degrees of freedom? Objective 1 - Introducing the unobserved, can be considered the main objective for any generator. However, is it sufficient to simply generate events that have not been observed before? No, generators should not be allowed to freely generate random synthetic events. Particular constraints should be formulated, guided by additional objectives. Synthetic data has to follow the statistical patterns in the observed and the synthetic data should be physically plausible, i.e. follow physical constraints.

3.2.2. Objective 2 - Reproducing patterns

Disclaimer The first type of generator constraint is that the synthetic data produced by a generator should follow the patterns found in the observed. The patterns will be discussed at the hand of event descriptors. The same principle that applies for the descriptors, applies for the event layer and the continuous data layer, which will be introduced in Section (3.3), and displayed in Fig. (3.2).

Different statistics may be used to capture patterns. In fact, there is no limit to the amount of patterns to follow, as any pattern found should be followed. The limitations are the capacity and capability of the architects of the generator. Therefore, the following list of patterns that is going to be discussed is non-exhaustive. However, hopefully, it will provide a general idea of the requirement of the reproduction of patterns.

Data format matrix At present, advanced methods exist for multi-variate statistics, in which matrices are the key data format. Although more complex data formats exist, such as arrays, matrices appear to currently prove challenging enough. Therefore, in this thesis, the data format by which data is described will be a matrix. When describing the following patterns, it will be indicated how the patterns relate to the matrix format.

The overall objective of reproducing patterns To a relatively small matrix of observed descriptors, a statistical model is fitted, from which a large matrix of synthetic descriptors is sampled. How this is done will be elaborated in Section (3.4). Overall, if all patterns are captured and reproduced by a generator, it should be possible to 'hide' the matrix of observed descriptors in the matrix of synthetic descriptors, without anyone being to able to tell where it is. More generally, patterns can be regarded to be appropriately reproduced when it is possible to 'hide' the observed data in the synthetic data.

Marginal distributions A first pattern to capture may be the pattern of marginal distributions. It is a pattern that occurs per column in the matrix. The row-wise order of entries does not matter for the distribution. To each separate column, a distribution is fitted. Dependent on the type of variable and the method event identification, a different distribution may be appropriate.

By fitting a distribution, the marginal population is smoothened. This can easily be visualised by sampling from a distribution and plotting a histogram. As the

amount of sampled data increases, the histogram may be plotted with more bins and will become increasingly smooth.

Dependence structure A second pattern to capture may be the dependence structure. This pattern works across different columns and is captured row-wise. The order of rows does not matter for the dependence structure. The dependence structure describes how variables (one variable per column) are dependent on each other. If there is dependence, then, when a particular first variable is large, it may be more likely that a particular second variable is also large or, reversely, more likely to be small. With no dependence between the two variables, the magnitude of the second variable does not depend on the first or the other way around. The dependence structure is as high-dimensional as the number of columns in the matrix.

The dependence structure may vary over regions in the multidimensional parameters space of the variables. It may be appropriate to fit different statistical models to different regions. For extreme values theory in particular, tail-end dependence may be distinguished from the dependence in the general populations. Therefore, it may be appropriate to fit one statistical model to capture the dependence in central regions and another statistical model for tail-end dependence.

Temporal dependence A third pattern to capture may be the temporal dependence. This pattern works across different rows and is typically captured per column, where it is assumed that each row represents a different time window or a different subsequent event. So, for temporal dependence, the row-wise order of entries does matter. Statistical models are required that allow the sampling of entire rows, where a new row sample is dependent on the (multiple) previous row sample(s).

Seasonality A fourth pattern to capture may be seasonality. This pattern does not work across subsequent rows or across columns. However, the matrix can typically be related to a (real or synthetic) time line, often per row. It is a pattern that can be found when variables are seasonally dependent, as they often are. For example, precipitation patterns can be quite different in summer than in winter, maybe less persistent but with higher intensity.

Seasonality may be found on different temporal scales, for which different terminology may be appropriate. For example, for a particular variable, magnitudes may vary with day or night. Or, specifically in relation to floods, magnitudes may depend on an 'el niño' or a 'la niña' year.

3.2.3. Objective 3 - Respecting physical constraints

Generators and models of physical phenomena A generator is to be used to create synthetic scenarios. They can be created in different forms, e.g. as events or as continuous data. These scenarios typically comprise boundary conditions and/or initial conditions.

Models of physical phenomena require boundary conditions and initial conditions. On the basis of these conditions, time integration of physically-based governing equations, such as the mass balance and the momentum balance, will be applied for a particular spatial domain. Although such models contain errors, which may accumulate in time due to the time integration, the use of physically-based gov-

erning equations provides physical realism.

In contrast, the statistical techniques used in a generator do not guarantee physical realism. In fact, a generator that is allowed to freely generate based on a statistical process, can be expected to create unrealistic scenarios. When unrealistic scenarios are used to drive models of physical phenomena, the model outcomes will also be unrealistic. Therefore, it is important to formulate physical constraints and implement them in the generator.

Constraining a generator No general pysical constraints can be formulated for generators. Which physical constraints are appropriate is specific to the type of data to generate. For example, physical constraints will be different for river discharge and for precipitation.

Physical constraints can have different levels of complexity. They could simply be based on common sense. For example, in a relatively short period of time, there should not be more rainfall in particular area than that there is water on the planet. Or, physical constraints may require expert knowledge, i.e. insight in geophysical behaviour of extreme events that comprise the flood hazard. Hurricanes should not be allowed by a generator to travel too far over land, since they typically gain strength over the ocean and reduce in strength over land.

From a practical point of view, it may make sense to apply an iterative strategy. Generator algorithms may already be complex in order to satisfy objectives 1 and 2 and may quickly increase in complexity when applying physical constraints. It is recommended to start by developing a simple generator. Then, to analyse the generated data and see what does not make sense physically. Then, to come up with additional constraints, and repeat.

3.3. A framework for event-based generation

3.3.1. Overview

The literature review, see Chapter (2), led to the deduction of a framework for event-based generation, used troughout this thesis. The framework can be seen as a tool to reach the objectives stated in Section (3.2).

The framework is displayed in Fig. (3.2) and decomposes the process of event-based generating into five general steps: (1) event identification; (2) event description; (3) multivariate statistics; (4) event reconstruction; and (5) space-time placement and gap filling.

3.3.2. Event identification

In this step, the subset of interest is extracted from the original data set, where the subset should exist of separated data clusters. In the next chapters, event-based methodology is used, so these clusters represent events. To be able to fit the literature to the framework in Section (3.5), these clusters can also be blocks, for which the data set would be sliced into temporal blocks, like years, months, weeks or days, etc.

3.3.3. Event description

In this step descriptors are captured for each event. The more descriptors used, the higher the level of detail that can be captured. Descriptors are captured for two purposes. First, to be able to divide the data into classes, where within each class the events are (supposed to be) comparable. Second, to be used for multivariate analysis. The subset of available descriptors to be used in the multivariate analysis

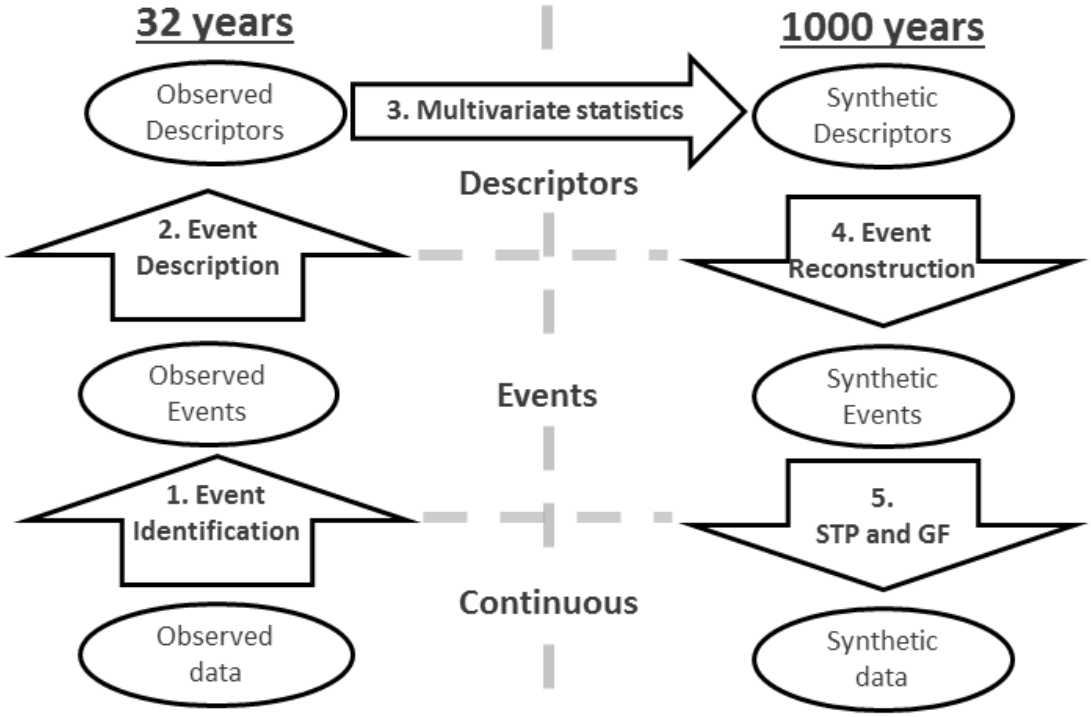


Figure 3.2: A framework for the generation of synthetic continuous data. STP refers to the spatio-temporal process and GF to gap-filling. 1000 years is indicative for a large (infinite) amount of synthetic data.

is determined by the event reconstruction method.

3.3.4. Multivariate statistics

In this step, statistics are applied to the captured descriptors. The aim of this step is to generate a large matrix of synthetic descriptors, of which the matrix of observed descriptors should be a likely subset. The synthetic matrix should have many more rows than the observed matrix, but the same number of columns. If this step is applied successfully, it should be possible to hide the small matrix of observed descriptors somewhere within the large matrix of synthetic descriptors. This would imply that patterns were successfully captured and simulated. Since this step is at the heart of the analysis, but kept constant throughout this thesis, it will be elaborated in Section (3.4), in anticipation of the subsequent chapters.

3.3.5. Event reconstruction

In this step, synthetic events are generated from the catalogue of simulated synthetic descriptors. The higher the level of detail, the better the result of this step will be, but also the more challenging this step will be. A higher level of detail can be obtained with a higher resolution and by using more descriptors. Potentially, different descriptors can be used for different classes of events.

3.3.6. Space-time placement and gap filling

For studies in which a subset of data was analysed in step 1, which are the studies where the clusters are events, the gaps between the reconstructed synthetic events may be filled to obtain synthetic continuous data.

When simulating high resolution blocks, continuous data is directly ob-

tained (with the resolution of the blocks).

3.4. Multivariate statistics

3.4.1. General

The statistical step can be performed in different ways. Multivariate statistics, where statistics are applied to matrices, is the current state of the art, and proves sufficiently challenging. This type of statistics is applied in this thesis. The general purpose of the statistical step is to generate a large matrix of synthetic event descriptors that is similar, but more comprehensive, than the matrix of observed event descriptors.

In this thesis, the statistical step is implemented with two sub-steps. First, statistical models are fitted, which may address distributions, row-wise dependence or column-wise dependence. Second, a discrete synthetic data set is simulated from the fitted statistical models.

3.4.2. Marginals

Multivariate statistical analysis deals with the data form of matrices. Marginal distributions are univariate distributions, fitted per column of a matrix.

3.4.3. General dependence structure - kernel density

In this thesis, a multivariate kernel density model is applied for the bulk of the populations, i.e. where samples are non-extreme. Multivariate kernel density estimation is a non-parametric way to estimate the probability density function of random variables. For the samples that are extreme in one or multiple dimensions, a separate model is used to capture the extremal dependence structure.

3.4.4. Extremal dependence structure - HT04

Heffernan and Tawn (2004) In this thesis, Heffernan and Tawn [46] provide the model of choice in this study for extremal dependence (HT04). The original HT04 model was proposed for the conditional distribution of all variables X_i where $i = 1, 2, ..., d$ given one of the variables is above a large threshold. More specifically, if the marginal distribution of each variable X_i is transformed to a standard Laplace distribution, denoting the transformed variables as Y_i, then $Y_{-i}|Y_i = y_i \; a_{-i}y_i + y_i^{b-i} \cdot Z_{-i}$ for all values $y_i > u$ and $i \in \{1, 2, ..., d\}$, where the $(d-1)$-dimension random variable Y_{-i} represents all but the i-th margin of the original full joint distribution; parameter a_{-i} and b_{-i} are the location and scale parameter vector of length $(d-1)$; the $(d-1)$-dimensional random variable Z_{-i} is a non-degenerate multivariate distribution that is independent of the value of Y_i, referred to as the residual distribution; and constant u is a high threshold above which the HT04 structure is assumed to apply with negligible error.

HT04 is a semi-parametric approach, in which the dependence at the extreme tail region is based on extrapolation (in the conditional mean and standard deviation) of the dependent variable conditional on the conditioning variable. Two HT04 model fits are required for each pair of marginals, with either marginal as the conditioning marginal Y_i and the other as the dependent marginal Y_{-i}. Each fit holds two parameters, a and b, after which a residual Z is calculated from each observed data point. The data used to fit the model are the pairs where the conditioning marginal Y_i is larger than a fitting threshold ζ_f. With an infinite number of samples drawn

from HT04, each model fit would result in as many pair-wise lines as there are data points. However, for simulation a set of these lines is used, since HT04 should be applied only if the largest marginal in the set is above a particular simulation threshold.

To explain why HT04 is used, two margins of a joint distribution (X and Y) are considered. They can be assumed to follow the same distribution without loss of generality. Quantity u is a very high quantile of the distribution of X and of Y. An exceedance probability $P(X > u)$ and a conditional exceedance probability $P(X > u|Y > u)$ are considered.

$P(X > u)$ converges to 0 as u goes to infinity. $P(X > u|Y > u)$ also converges to a (not necessarily 0) probability as u goes to infinity. In particular, the convergence rates of these probabilities in the tail-ends of the respective distributions are of importance, as the relative rate of convergence decides the type of extremal dependence between X and Y.

1. When $P(X > u)$ and $P(X > u|Y > u)$ converge at the same rate, this implies (near) independence.

2. When $P(X > u|Y > u)$ converges more slowly than $P(X > u)$, this implies a positive association.

3. When $P(X > u|Y > u)$ converges faster than $P(X > u)$, this implies a negative association.

So, first, HT04 is used, because it can capture all these types of extremal dependence simultaneously. Second, for the tail region, HT04 is very flexible in terms of capturing different types of extremal dependence [58]. Third, with HT04 the estimation of the model parameters does not suffer from the curse of dimensionality, whereby the number of observation data available for fitting the model decreases dramatically as the dimension of the problem grows.

HT04 was recently applied for fluvial flooding [53, 57, 77] and for coastal flooding [42, 105], in which the model fitting procedure is described in more detail.

3.4.5. Simulation

Simulation is performed by random sampling (Monte-Carlo). For the bulk of the populations, sampling is straightforward, making use of the kernels. For the tail-ends, conditional sampling is applied. One variable is sampled in the tail-end region (extreme), as defined by a marginal quantile threshold, and the rest of the variables are deduced making use of the HT04 model. Dependent on the HT04 model fit and on how extreme the sampling variable is, the dependent variables may also be extreme or may be non-extreme. Sampling is done until the number of envisioned synthetic sets is reached, where each synthetic set has the same number of samples as the observed set. These synthetic sets are combined to a single set, comprising the chosen synthetic length of time, e.g. 10.000 years.

3.5. Literature according to the framework

To demonstrate that the framework is generally applicable for event-based studies, a number of studies was fitted to the framework, displayed in Table (3.1). A slight modification was applied to be able to include studies that are block-based, where the more general term for event identification or block slicing would be clustering.

Table 3.1: A number of studies sorted according to the general framework of this thesis; Fig. (3.2). Acronyms can be found in Table (A.2).

Study	Variables	Clustering	Descriptors	Classification	Statistics
Cameron et al. [17]	1. precipitation (gen)	thresholds	μ, D, $IEAT$	loc, duration	GP (ML)
	2. discharge (val)	annual	M	Loc	GEV (ML)
Cooley, Nychka, and Naveau [21]	prec.	thresh.	V	Loc, climate	GP (ML, SGP)
Merz and Blöschl [61]	discharge	annual	M	Loc	Weibull
Keef, Tawn, and Svensson [53]	discharge, precipitation	thresholds (unclear)	$P \pm 3$	Loc[1]	Gumbel, HT04
Hundecha, Pahlow, and Schumann [47]	1. rainfall (gen)	daily	V	Loc[1], month	Gamma, GP, AR
	2. rainfall (val)	annual	M	Loc	GEV
Srikanthan and Pegram [85]	1. rainfall (Gen)	daily	V	Loc	Gamma, Exp, NST, DW, MC
	2. rainfall (Val)	annual, monthly, daily	Long list, see paper	Loc	
Basinger, Montalto, and Lall [6]	1. rainfall (Gen)	daily	V	Loc	MC
	2. rainfall (Val)	daily, annual	S, NW	Loc	
Eastoe and Tawn [31]	1. discharge	annual	M	Loc	GP, PP
	2. discharge	thresholds	P	Loc	GEV
Ghosh and Mallick [39]	1. precipitation	annual	M	Loc, month	GEV, copula
	2. precipitation	threshold	P	Loc	GP, copula, MC, PP
Pappenberger et al. [64]	discharge, storage, depth	annual	M	Loc	Gumbel (L)
Cowpertwait et al. [25]	1. rainfall, temperature (Gen)	?	V, R	Storm type (convective, stratiform)	STTP (NS)

	2. rainfall (Val)	annual	M	Loc	Emp
	3. discharge (Val)	annual	M	Loc	Emp
Svensson, Kjeldsen, and Jones [86]	1. rainfall, soils moisture, evaporation, discharge (Gen)	thresholds (rain)	μ, D, $IEAT$, SMD_0, Dis_0	Loc, summer-winter	Exp, GP
	2. discharge (Val)	annual	M	Loc	GEV (L)
Winsemius et al. [102]	discharge	annual	M	Loc	?
Alfieri et al. [2]	discharge	annual	M	Loc	Gumbel
Wyncoll and Gouldby [105]	discharge	threshold (runs)	P	Loc[1]	Emp, GP, HT04
Bracken et al. [13]	prec. (Gen)	annual	M (S, 3-day)	Loc[1]	GEV (ML, SGP), copula
Paprotny and Morales-Nápoles [65]	discharge	annual	M	Loc	Gumbel(ML)
Schneeberger and Steinberger [77]	discharge	3-day, annual	M	Loc[1], Summer-winter	GP (ML), GEV (L), HT04 (Laplace)
Diederen et al. [28]	discharge	NR	P	Loc[1]	GP (PML), NP, HT04

[1]Partial classification: marginals per site, lumped dependence model for all sites.

4

The limitations of static spatio-temporal event generation

4.1. Introduction

Flood events cause large damages worldwide [27]. Flood risk assessments (FRAs) are required for long-term planning, for example, for investments in infrastructure and other urban capital. Following the definition of risk [34], simply put as probability of damage, FRA requires an approximation of the risk curve under stationary climate conditions and a current distribution of asset values. Typically, for FRAs a chain of models is applied, covering the entire risk cascade from hazardous extreme events down to flood damages or casualties resulting from inundation (for example, expected annual damage, loss of life). The risk curve represents the probability of damages and is approximated by the evaluation of a comprehensive catalogue of hazard scenarios. The chain can be run in continuous mode [10, 12, 16, 33], or with separate events [42, 94]. To drive the chain of models, boundary forcing is required. This typically comprises a large catalogue of synthetic forcing data, with models conditioned on observations.

Widespread flooding can potentially cause large damage in a short time-window. Continental events and, for instance, maximum probable damages are of interest. In particular, the (re)insurance industry wants to know the chance of a widespread portfolio of assets getting affected in a short time-window. With the increase in computational power, continental FRAs have recently become feasible [2, 29, 66, 80, 95, 97]. Vorogushyn et al. [93] call for new methods for large FRA to enable the capturing of system interactions and feedbacks. The focus in this case study is on methodology required for the generation of a large catalogue of synthetic continental discharge event descriptors for fluvial FRA.

River discharge waves may cause the exceedance of bank-full conditions or may cause dikes to fail. They are dynamic, i.e. show a wave-like behaviour. Travel times of discharge waves in large river basins can be long, i.e. time lags between discharge peaks at different locations can be large. With large travel times, a new discharge wave may be generated upstream, while the previous discharge wave has not yet reached the river mouth. Furthermore, discharge waves in river basins are triggered by atmospheric events that may span across multiple river basins. Finally, discharge waves in different river basins may be related to a single atmospheric event, but do not occur at the same time, since catchments have different response times. With an increasing spatial domain, dynamic events start overlapping in time and merge into a space-time continuum. For a continental FRA, the challenge arises how to define observed continental river discharge events and how to simulate synthetic continental river discharge events while retaining the observed statistical properties in space (spatial dependence/coherence).

A distinction can be made between two groups of event identification methods: methods based on time blocks and methods based on dynamic events. Time blocks were previously discussed in Section (2.2.1). Dynamic events are defined as events with spatially varying time-windows, which are based on the discharge values. As described above, for large spatial domains small dynamic events at different locations may overlap in time and form one single long-lasting spatio-temporal event. Hence, a practical definition of dynamic space-time-windows is required.

Outline In Section (4.2), the methodology will be introduced, comprising the used data set and the (reduced) framework. In Section (4.3), Pan-European discharge waves, which are characterised by significant time lags between peaks at distant locations, will be analysed in the space-time continuum. A new method of dynamic event identification will be applied, where the aim is to capture discharge events in each major European river basin, after which a block-based time-window method will be used to merge them to spatially-coherent, pan-European events. The pan-European discharge events will be described by their peaks, with which a stochastic event-based generator of event descriptors will be parameterised. Using the generator, synthetic descriptor sets will be simulated, after which the statistical properties of the synthetic sets will be compared to those of the observed. In Section (4.4.2), the sensitivity of the generated synthetic descriptor sets will be investigated with regards to the coverage of the peaks and the capturing and retaining of the (spatial) dependence structure. In Section (4.5), applicability and the main limitations of the methodology will be discussed.

4.2. Methodology

4.2.1. Data

A gridded discharge reanalysis data set is used, which was obtained with the well-established LISFLOOD model [90] and which covers the major river networks in Europe. This data set resulted from a hydrological model driven by a climate reanalysis data set for the period 1990 to 2015. It has a spatial resolution of 5x5 km and a daily temporal resolution. A high temporal resolution is critical for river discharge waves to be tracked in the extended river network.

In order to keep the computational costs reasonable, the network is reduced to the major streams and tributaries. This means that, although the input data was two-dimensional in space (x,y), only the network of 1-dimensional rivers (s) is considered. For high-order small streams to be included, a higher spatial and temporal resolution would be required for wave tracking. Although the data set is derived from a modelled reanalysis data set, the used subset of data will be referred to as 'observed data' as it comprises observations of reality, contrasting with 'synthetic data', which comprises data values of what may hypothetically occur.

4.2.2. Generator objectives

There are three main generator objectives:

1. introduce the unobserved (GO_1),
2. reproduce patterns (GO_2) and
3. respect physical restraints (GO_3).

These generator objectives were described in more detail in Section (3.2).

The aim in this chapter is to provide a methodology (of the type SSTPA) to generate a large catalogue of synthetic, discharge descriptor sets at many different locations throughout Europe.

4.2.3. Framework

The framework for the generation of synthetic peak sets consisted of three consecutive steps, see Fig. (4.1). This is a reduced version of the generic framework introduced in Section (3.3). First, the identification of continental events in the continuous data on the entire pan-European river

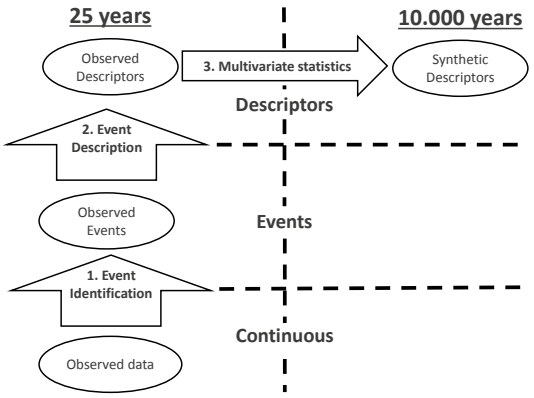


Figure 4.1: The applied framework, which comprises three steps. It is a reduced version of the framework displayed in Fig. (3.2).

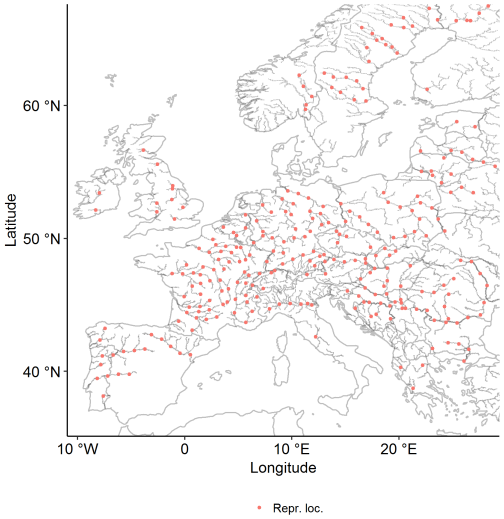


Figure 4.2: The network of major European rivers and a subset of 298 representative locations.

network (OBJ1). To achieve this, first local (single location) events were identified, for which a new method of time series analysis, 'Noise Removal' (NR), was applied at every location (grid-cell) in the river network. These local events were connected to neighbouring locations to obtain river basin events, to be subsequently merged to pan-European events, which span across multiple river basins. Second, the description of the pan-European events (OBJ1). To reduce the dimension (number of locations) for statistical analysis while trying to maintain an acceptable spatial coverage, 298 representative locations were selected within the network of major European rivers, see Fig. (4.2). At these representative locations, the continental discharge events were described by their local peaks. Third, the generation of a large catalogue of synthetic descriptor using multivariate statistical analysis (OBJ2). A multivariate dependence model was fitted to the catalogue of observed descriptors covering 25 years, retaining the observed spatial correlation structure. Finally, the fitted statistical model was used to simulate a large catalogue of synthetic event descriptors, characterised by spatial coherence and comprising a synthetic period of 10,000 years.

4.3. A generator of river discharge peaks

4.3.1. Event identification

Single-location events When using the popular 'Peaks-Over-Threshold' method (POT) per location, all events below a particular threshold are dropped. This is appropriate for event identification only when events show a homogeneous 'extremeness-per-location'. However, when studying discharge waves moving through the river network by 'extremeness-per-location', a heterogeneous behaviour can be expected. Relatively extreme events upstream may become less extreme moving downstream when the lower part of the river basin is not activated. Or, in contrast, relatively non-extreme events at different upstream branches can generate a relatively extreme event at confluences downstream due to wave superposition. To address the heterogeneity, a new noise removal algorithm was developed to capture

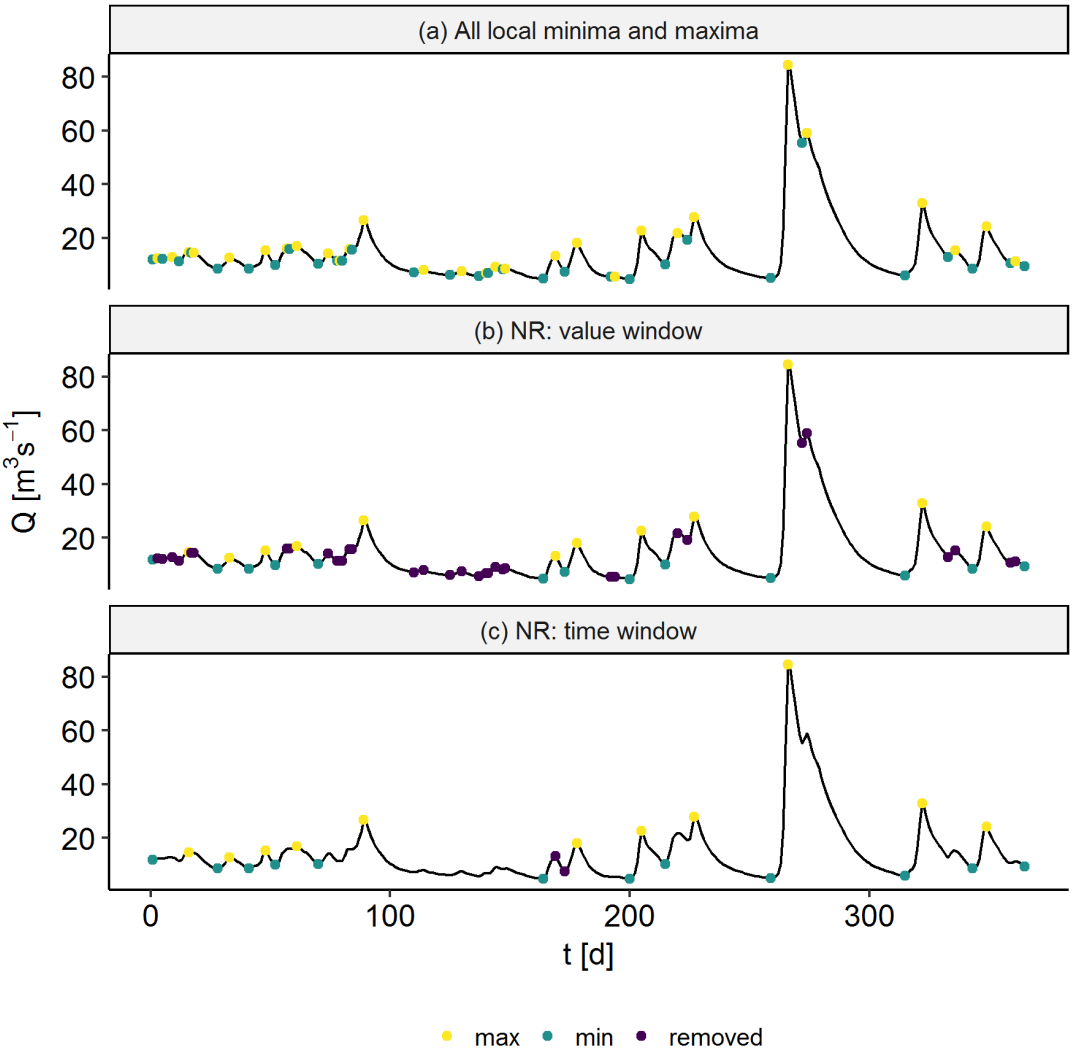


Figure 4.3: (a) All local minima and maxima. (b) Removal of noise using the value-window. (c) Removal of noise using the time-window.

local events which manages to eliminate small local peaks that are part of a bigger event (noise), while retaining small events that may be spatially connected to larger events upstream or downstream. This is a key feature to the wave tracking, which will be introduced in the next section.

In the procedure of NR the following variables are used:

- X is an arbitrary variable,
- μ_j are the local minima of X,
- M_k are the local maxima of X,
- $Y = \mu_1, M_1, \mu_2, .., \mu_{n-1}, M_{n-1}, \mu_n$ is a series of alternating minima and maxima,
- δ_y is the NR value-window (setting),
- t_μ are the times at which there is a local minimum,
- $T = (t_{\mu_1}, t_{\mu_2}, .., t_{\mu_n})$,
- δ_t is the NR time-window (setting),

First, all local minima μ and maxima M are identified, defined as the points where the sign of the increment changes from negative to positive and vice versa, see Fig. (4.3)a. Second, small perturbations are

identified as noise and are removed, where the following algorithm is applied:

1. Define a series $Y = (\mu_1, M_1, \mu_2, .., \mu_{n-1}, M_{n-1}, \mu_n)$ and calculate $dY = |\mu_1 - M_1, M_1 - \mu_2, .., M_{n-1} - \mu_n|$.
2. Either calculate the 'NR value-window' $\delta_y = f_y \times max(dY)$, where f_y is a fraction to set, or set δ_y directly.
3. Find i by selecting the smallest difference in value $dY_i = min(dY)$. If $dY_i < \delta_y$, remove Y_i and Y_{i+1} from Y, then recalculate dY. This step is repeated until there is nothing left to remove.

An example of the NR value-window filtering is displayed in Fig. (4.3)b. Third, to make sure that two local minima are not too close in time, the following algorithm is applied:

1. Define a series $T = (t_{\mu_1}, t_{\mu_2}, .., t_{\mu_n})$ and calculate $dT = (t_{\mu_2} - t_{\mu_1}, t_{\mu_3} - t_{\mu_2}, .., t_{\mu_n} - t_{\mu_{n-1}})$.
2. Either calculate the 'NR time-window' $\delta_t = f_t \times max(dT)$, where f_t is a fraction to set, or set δ_t directly.
3. Find i by selecting the smallest difference in time $dT_i = min(dT)$. If $\mu_i < \mu_{i+1}$, $j = i + 1$, else $j = i$. If $M_{j-1} < M_j$, $k = j - 1$, else $k = j$. Remove Y_{2j-1}, Y_{2k} and T_j, then recalculate dT. Repeat this step until there is nothing left to remove.

An example of the NR time-window filtering is displayed in Fig. (4.3)c. Fourth, a local event can be chosen to last from minimum to minimum, or to be only the time step in which the peak occurs, or something in between.

The NR value-window fraction was set relatively low $f_y = 0.01[-]$, such that many small local events were retained. However, by setting the fraction low, small perturbations (noise) made it difficult to spatially separate events. This was ameliorated by using the NR time-window $\delta_t = 10d$, ensuring a minimal amount of time between local minima. The choice of NR parameters will be elaborated in Section (4.4).

River basin events River discharge waves propagate through the network in the downstream direction, introducing time lags between the moments the waves pass at different locations. Time lags are difficult to estimate, because the celerity of river discharge waves can be highly nonlinear. The wave celerity is a function of the hydraulic depth and changes in a nonlinear way when overbank flow occurs and floodplains become inundated. When using gauge data (point-observations), combining local events to events that span multiple locations, time lags are typically addressed using time-windows. The gridded data set used in this case study allowed us to try a new method to combine local events to river basin events, which will be referred to as 'wave tracking'. Each location in the river network is physically connected to its neighbouring locations, which allows waves to be tracked throughout the entire river network. Wave tracking is robust to non-linearities in the wave celerity, and therefore it allows to better address time lags, so that, when peaks at different locations are compared in Section (4.3.3), it is made sure that they are of the same discharge wave.

To track river discharge waves, the following procedure was applied. First, lo-

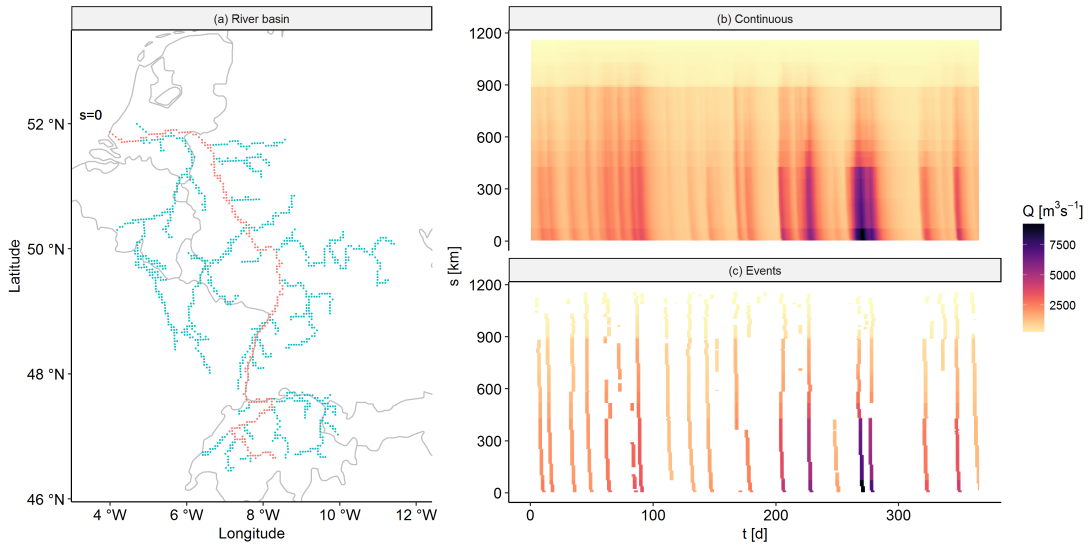


Figure 4.4: a) A particular branch of the river Rhine. b) The continuous discharge data on the river branch, where the river mouth is located at $s = 0km$ and the head water is located around $s = 1100km$. c) Events on the river branch. The polygons (i.e. separated islands of data) are discharge waves moving through the river branch.

cal events were separated by applying NR to time series at every location in the river network, where of each local event the day of the peaks ± 1 day was retained. Second, separate events per river branch were identified by capturing the polygons in the branch's space-time image; Fig. (4.4). The settings of the NR were adjusted by trial and error to try to obtain consistent polygons in space (low noise removal), but separated in time (high noise removal). Third, the events of different river branches were merged, based on overlap of event time coordinates at the confluences. This procedure resulted in a variable number of tracked discharge waves per river basin.

Pan-European events Precipitation events, which are the main driving source of river discharge events, span across different river basins. Therefore, large discharge events in adjacent river basins are likely to be correlated. To account for this correlation, events had to be defined that included discharge waves across different river basins (in this case study: pan-European events). Since discharge waves do not span across different river basins (by definition), such events should be connected to each other in a different way. Discharge waves in different basins are not synchronised, which adds additional complexity. In order to obtain a method to construct pan-European events, which on the one hand considers discharge waves in river basins and on the other hand accounts for trans-basin dependence, a combined approach of wave tracking and 'global time-windows' is proposed.

The following procedure was adopted. First, subsequent global time-windows were set up with a length of $\delta_l = 21d$, which resulted in 428 global time-windows in the period 1990-2015 (i.e. 428 pan-European events). The length of the global time-windows will be discussed in Section (4.5.2). Second, to each global time-window complete, tracked discharge waves were assigned. To do this, each discharge wave was represented by its first

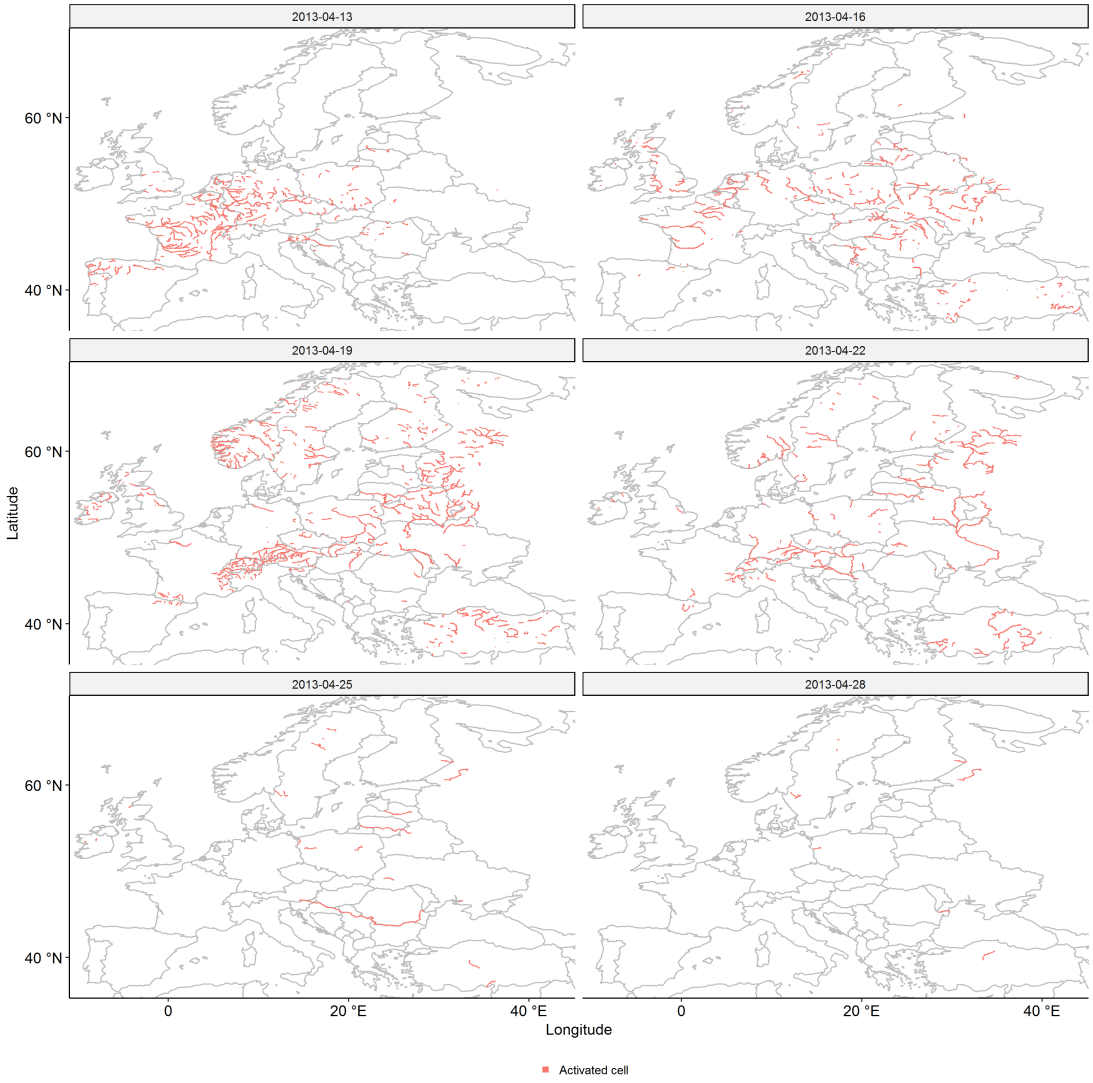


Figure 4.5: Daily snapshots of a Pan-European event with a large spatial extent.

time coordinate, i.e. the day when the discharge wave started somewhere (upstream) in the river basin. The discharge wave was then assigned to the global time-window in which this day occurred. If, per river basin, multiple discharge waves were assigned to a particular global time-window, only the discharge wave with the largest discharge value was retained. This procedure resulted in 428 pan-European events. An example of a pan-European event is displayed in Fig. (4.5).

4.3.2. Event description

The aim in this case study was to describe the pan-European events by their peak discharge, at 298 representative locations on the river network. However, the pan-European events did not yield discharge peaks at all representative locations for each event, i.e the observed descriptor matrix had gaps. To be able to capture the spatial dependence structure in Section (4.3.3), gaps had to be filled by assigning 'auxiliary values'. This will be further discussed in Section (4.5.2).

The following procedure was applied. At locations where an event occurred, the discharge peak was extracted. Where no event occurred (36% of the entries in the observed descriptor matrix), the gaps were filled using auxiliary values. Per repre-

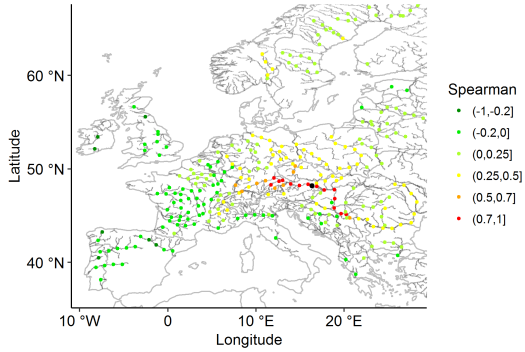


Figure 4.6: Correlation of descriptors at all representative locations versus descriptors at Vienna (black dot).

sentative location (i.e. column-wise), a number of local time-windows were set up in between the peaks of identified events, corresponding to the number of gaps between those respective peaks. Within each of these local time-windows, the maximum value was selected as auxiliary value. This procedure resulted in a (complete) observed descriptor matrix.

Fig. (4.6) shows the correlation of the descriptors with those at Vienna. It can be observed that the highest correlation was found at nearby locations within the same river basin. However, significant correlation was found in nearby locations that were not in the same river basin, which confirmed the importance of identifying events spanning multiple river basins.

In order to align with the corresponding literature in statistical models for multivariate extreme values, in the next Section (4.3.3) columns of the observed descriptor matrix will be referred to as margins and the large values in each column will be referred to as the upper tails of the marginal distributions.

4.3.3. Multivariate statistics

Marginal distributions Generalised Pareto Distributions (GPDs) [20] were fitted to the upper tail of the marginal distributions, i.e. for each column in the observed descriptor matrix. The issue of threshold choice for GPDs is well-discussed in the literature [63]. After comparing the model fits, the empirical quantile $q_m = 0.94$ was used as marginal threshold for the GPD at each location. This threshold was found by trial and error. The quality of the marginal GPD fits was tested with a standard method, comparing the empirical quantiles and probabilities against the modelled, including checks of the tolerance intervals.

Dependence model To be able to capture the dependence between sets of descriptors (i.e. rows in the observed descriptor matrix), the marginals were transformed to the uniform space. This transformation is applied in many other analyses, for example, copulas [38, 62]. Values below the marginal threshold, used to fit the GPDs, were transformed using the empirical marginal distribution and values above the marginal threshold were transformed using the GPDs of the previous (marginal) step. A model was applied with two different components to capture the dependence structure, one for the non-extreme part and one for the extreme part.

The dependence structure of the non-extreme part was captured using a non-parametric, multivariate kernel density model with Gaussian kernels. The (entire) uniform marginals were transformed to the normal space, with the mean $\mu = 0$ and the standard deviation $sd = 1$. Bandwidths H for the kernels were selected using the method of Silverman [83]. Silverman's rule (see also [18]) is defined as

$$H = \left(\frac{4}{n(m+2)}\right)^{2/(m+4)} \hat{\Sigma} \qquad (4.1)$$

where m is the number of variables, n is the

sample size, $\hat{\Sigma}$ is the empirical covariance matrix.

To capture the dependence of the extreme part the HT04 model was chosen, which is described in detail in Section (3.4.4). To fit HT04, the (entire) uniform marginals were transformed to the Laplace space [51]. HT04 model fits were obtained in the Laplace space using maximum likelihood, with each marginal as conditioning variable and all other marginals as dependent variables, resulting in a total of 298*297 model fits, where the fitting threshold $q_d = 0.9$ was chosen, which was a trade-off between variance and bias.

Simulation The observed uniform descriptor matrix was split into a 'non-extreme' part, and an 'extreme' part. Each row in which not a single descriptor exceeded an extremal simulation threshold $q_s = 0.98$ was determined to be non-extreme (23%), the rest (77%) was determined to be extreme (somewhere). For the non-extreme sets, re-sampling was done from the non-parametric model. For the extreme sets, re-sampling was done from HT04, where the model fit was used of the marginal that was the largest by quantile in the set. All sets were re-sampled $N = T_{sim}/T_{obs}$ times, where T_{obs} is the duration of the observed data (25 years) and T_{sim} is the duration of the synthetic data (10.000 years). After the simulation, the marginals of the synthetic descriptor matrix were transformed to respect the fitted GPDs. This implies that the synthetic marginals were forced to have the same distribution as the corresponding observed marginals. However, by forcing this transformation, the dependence structure was slightly distorted.

4.4. Focussed performance check of the discharge generator

4.4.1. Performance indicators

General dependence structure Using multivariate extreme value analysis, the observed descriptor matrix was extended with synthetic data, obtaining a (large) synthetic descriptor matrix. The patterns in the larger synthetic descriptor matrix had to match the patterns found in the smaller observed descriptor matrix. The focus was on two main patterns; marginal distributions (a column-wise pattern) and dependence structure (a row-wise pattern). To respect the fitted marginal distributions and, simultaneously, retain the dependence structure, is challenging. There is no perfect method for these two objectives. The choice was made to respect the distributions fitted to the observed marginals, for which the synthetic marginals were transformed to follow the corresponding observed distributions. Therefore, the dependence structure in the synthetic data had to be compared to that in the observed.

To further investigate the dependence structure, Fig. (4.7) shows a sample of the observed descriptors versus the synthetic descriptors. It can be observed that the distributions of the individual descriptors were filled up reasonably well (GO1), while retaining the observed dependence structure reasonably well (GO2), as the simulated descriptors follow the trends in the observed data.

Fig. (4.8.1) shows the pair-wise, spatial correlation structure between descriptors at different locations. Rather than choosing the distance between locations along the river branch, geo-spatial distance was chosen, such that locations could be compared not only within river basins,

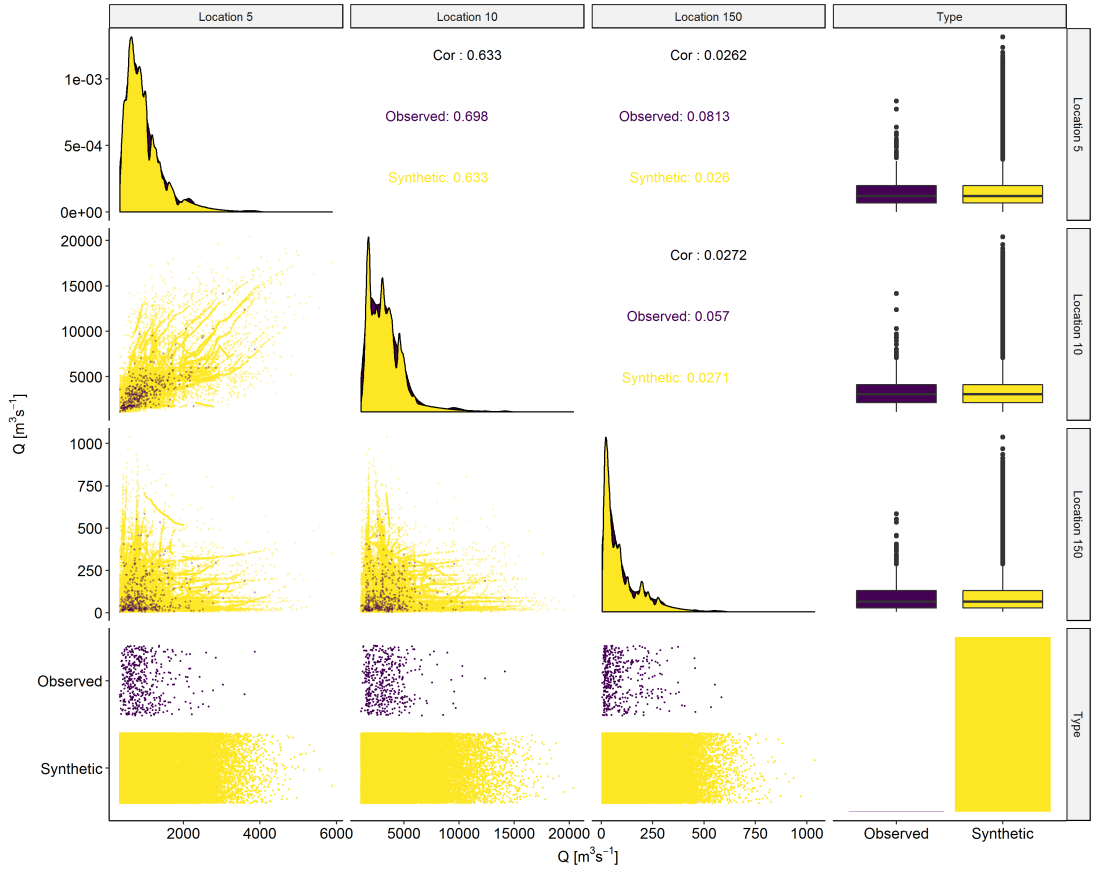
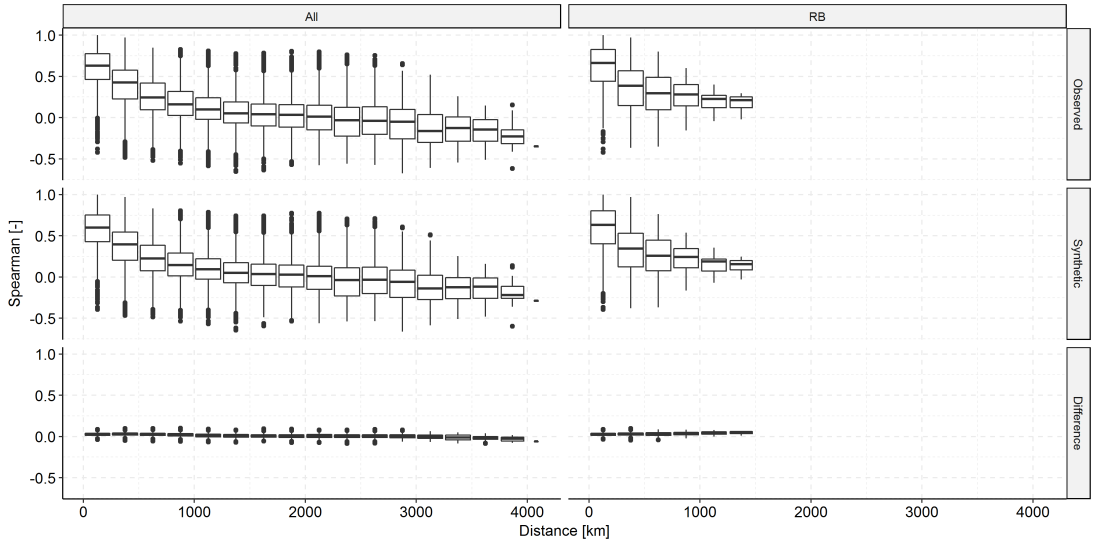


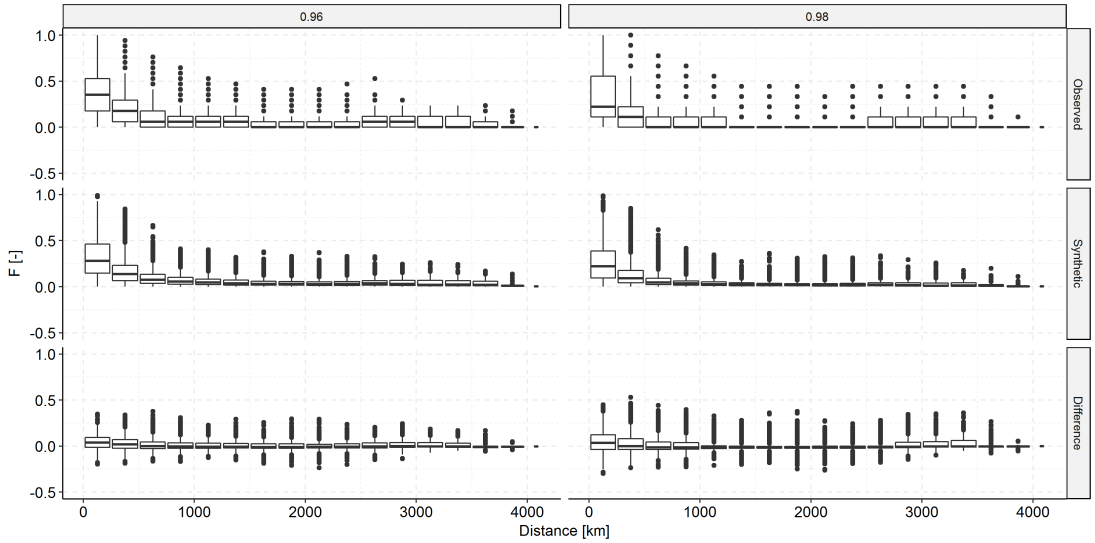
Figure 4.7: Observed (purple) versus synthetic (yellow) descriptors (discharge peaks) at three locations. In the diagonal, distributions of observed and synthetic descriptors per location are compared using box-plots. Below the diagonal, pair-wise scatter plots are displayed. Above the diagonal, pair-wise correlations are displayed.

but also across different river basins. The Spearman correlation coefficients of the observed descriptors and the synthetic descriptors agree very well (GO2), which indicates that the general spatial dependence structure is similar in the observed descriptor matrix and in the synthetic descriptor matrix. The difference indicates an overall slightly higher (positive or negative) correlation in the observed descriptor matrix. In both the observed and the synthetic descriptor set, a shift from positive to negative correlation can be observed around $2000 - 2500km$, which may be related to large atmospheric patterns.

Extremal dependence structure Following up on the general check of correlation between the entire descriptor sets, it was specifically checked if the tail-end correlations were captured. Fig. (4.8.2) shows that the general behaviour of co-occurrence of extremes was relatively well captured in the dependence model (GO2). The general pattern in the synthetic descriptors is reasonably similar to the pattern in the observed descriptors. A small positive bias can be observed, which shows that the dependence model slightly underestimated the frequency of joint occurrence of extremes. The zero difference generally falls within the lower quartile. Moreover, the higher the quantile for which the exceedance was checked, the fewer quantile exceedances to count, which lead to a larger spread in the difference between the



(4.8.1) Spatial correlation of the observed descriptors versus the synthetic descriptors, summarised by pair-wise Spearman correlation. The upper panel shows the correlation in the observed data, the middle panel shows the correlation in the synthetic data and the lower panel shows the difference between the observed and the synthetic correlation for each pair. The left column shows the correlation between all pairs of locations, right shows only the pairs that are in the same river basin.



(4.8.2) Spatial extremal correlation of the observed descriptors versus the synthetic descriptors. For a selection of high quantiles, the fraction F was counted of events where extremes at both locations exceeded the respective quantile divided by the total number of quantile exceedances. The upper panel shows the fractions in the observed data, the middle panel shows the fractions in the synthetic data and the lower panel shows the pair-wise difference between the observed and the synthetic fractions.

Figure 4.8: Spatial correlations.

observed and the synthetic.

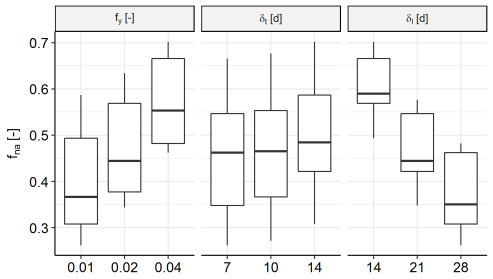
4.4.2. Sensitivity

Coverage of peaks In Section (4.2.3), the objective was stated to capture the spatial dependence structure between peaks of discharge events at different locations spread out through Europe (OBJ1), for which events had to be identified. In the event identification procedure, three main parameters were used for the identification of pan-European events: the value-window fraction f_y and time-window δ_t of the noise removal method, and the length of the global time-windows δ_l for the pan-European events (Section (4.3.1)). The following trade-off was considered. For each pan-European event, one discharge peak could be assigned to each location. However, depending on the length of the global time-window, there may be no river basin event to assign to the global time-window, i.e. a missing discharge peak, or multiple river basin events may be assigned from which only one discharge peak could be retained per global time-window per location. Therefore, a relatively large global time-window lead to the underestimation of the frequency of discharge waves in river basins, whereas a relatively small global time-window lead to a large percentage of missing local events at the representative locations. Since this was a continental analysis, the fraction of missing values in the observed descriptor matrix f_{na} was relatively large and therefore decisive for our choice of parameter settings.

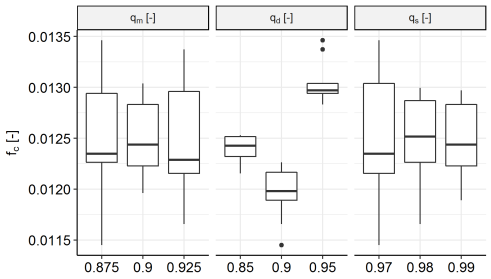
Fig. (4.9.1) shows the sensitivity of f_{na}, which is the percentage of missing peaks in the observed descriptor matrix (Section (4.3.1)). When more noise was removed, i.e. larger f_y and δ_t, events had a larger fraction of missing peaks, i.e. larger f_{na}. In contrast, a larger δ_l lead to a smaller fraction of missing peaks f_{na}, since the chance was larger for an event to occur at a particular location given more time. When comparing the sensitivity of the three parameters, it can be observed that the outcome is relatively stable with regard to the choice of f_y and δ_t, whereas the percentage of missing peaks $f_n a$ could vary quite a lot with δ_l. Our final choices are $f_y = 0.01[-]$, $\delta_t = 10d$ and $\delta_l = 21d$. A lower δ_l would have caused too many peaks being missing, which would have lead to unreliable estimation of the dependence model.

Dependence structure In Section (4.2.3), the objective was stated to generate a large catalogue containing synthetic discharge peaks, filling up the observed distributions while retaining the observed dependence structure (OBJ2). Two key features were defined for the quality of the generated synthetic catalogue. First, it had to contain descriptions of a much larger variety of hypothetical (synthetic) events than those identified in the observed data (GO1), which was achieved by sampling a large number of synthetic descriptors and transforming them to follow the same distributions as fitted to the observed. Second, the dependence structure of the synthetic catalogue needed to agree with that of the observed, since the observed descriptor catalogue had to be a likely subset of the synthetic (GO2), which was achieved with the dependence model, of which the results were demonstrated in Fig. (4.8). The sensitivity of the results was further investigated for GO2, using a summary descriptor of Fig. (4.8.1).

Fig. (4.9.2) shows the sensitivity of f_c, which is the mean absolute difference



(4.9.1) Sensitivity of f_{na}, which is the fraction of missing values in the observed descriptor matrix, to the three main settings used for the identification of events in Section (4.3.1). f_y is the NR value-window fraction, δ_t is the NR time-window, δ_l is the length of the global time-windows.



(4.9.2) Sensitivity of f_c, which is the mean of the absolute differences in correlation between the synthetic and the observed descriptor sets for all locations, to the three main settings used for the statistical analysis in Section (4.3.3). q_m is the quantile threshold for the GPDs, q_d is the quantile threshold for HT04, q_s is the quantile threshold for the simulation from HT04.

Figure 4.9: Sensitivity.

in Spearman correlation between the synthetic and the observed descriptor sets for all locations. No clear trend was found for both q_m, which is the quantile threshold for the GPDs, and q_s, which is the quantile threshold for the simulation from HT04. A local minimum was found for q_d, which is the quantile threshold to select the observed descriptors to which the HT04 model was fitted.

A recent, more comprehensive study of the sources of uncertainty in a probabilistic flood risk model was provided by Winter et al. [104], who used the framework provided by Hall and Solomatine [43].

4.5. Discussion

4.5.1. Applicability to pan-European FRA

The generated synthetic descriptor catalogue can be used to drive an event-based chain of models, which may cascade from a hydraulic model of the river network coupled with an inundation model to damage and/or life safety models. To drive an inundation model, synthetic discharge events have to be reconstructed from the synthetic descriptor sets in the catalogue, which corresponds to what would be step 4 in Fig. (3.2), moving down from synthetic descriptors towards synthetic events. This step comprises fitting discharge hydrographs to the synthetic descriptors and assigning time lags. The catalogue of synthetic discharge event descriptors was provided at 298 locations for a synthetic period of 10.000 years (with stationary climate conditions). Both the number of locations and the number of synthetic years can be expanded to provide a more detailed coverage. This catalogue can be used to generate discharge hydrographs to drive a pan-European inundation model for continental, event-based flood risk assessment.

4.5.2. Limitations in multivariate statistics

Historically, observations have been made at specific locations, for example, discharge gauge stations at certain locations along rivers. Therefore, most event

	L_1	L_2	L_3	L_4	L_5	L_6
Set 1	P	P	P	P	A	A
Set 2	A	P	P	A	P	A
Set 3	P	A	P	P	P	A
Set 4	P	P	A	P	P	P

Table 4.1: L_1-L_6 are different locations. Sets 1-4 describe a discharge event. Generally, dynamic discharge events do not occur at all locations, such that peaks (P) cannot be identified for all locations. Therefore, auxiliary values (A) have to be used to fill in the gaps.

identification methods are designed for local probabilistic analysis of discharge waves, starting with the identification of 'local events', i.e. events at certain locations, based on temporal dynamics [88]. When addressing spatial dependence using an event-based approach, the difficulty arises that discharge waves will not occur at all gauged locations within a reasonable time-window. The larger the spatial domain in which discharge waves are considered, the more likely it is they are spread out in time. Therefore, an extraction of a dynamic event from a space-time continuum, trying to obtain local peaks for all locations, will lead to a matrix of incomplete peak sets. This is problematic, because current statistical methods for multivariate event generation cannot handle a matrix with missing components [53]. Therefore, 'auxiliary values', i.e. values that do not represent discharge peaks, are required in order to fill up the gaps (see Table (4.1)). Different methods exist to assign auxiliary values, for different purposes. Gouldby et al. [42] analysed different coastal flood variables with an event-based approach, where they adopted concurrent values at all locations where particular thresholds had not been exceeded (i.e. no local event). Keef, Tawn, and Svensson [53] relaxed the time constraint, where they considered the values at all locations within a -3 to +3 days time-window. Since a large number of locations was considered, with large time lags, neither of these methods were appropriate. Therefore, auxiliary values were found using local time-windows, where these time-windows depended on the gaps per location.

4.5.3. Limitations in reconstruction of events

Difficulty can be expected in that the synthetic descriptor sets partially consist of synthetic discharge peaks and partially consist of synthetic auxiliary values. For the synthetic peaks, hydrographs can be reconstructed by fitting a typical (triangular) hydrograph shape to the synthetic peaks, whereas for the synthetic auxiliary values, it will not be entirely clear how to fit a hydrograph. Time lags for the synthetic descriptors sets could be borrowed from the corresponding observed descriptors sets. However, this would implicitly use the assumption that travel times or wave celerities are independent of magnitude, which is not a great assumption to make. These difficulties are not specific to this analysis, but apply for all analyses in which an event-based approach is combined with descriptors per location.

4.6. Conclusions

A new 'noise removal' method and a 'wave tracking' method were used with which discharge waves were successfully identified and tracked in all major European river basins. Using global time-windows, these river basin events were clustered to pan-European events. The dependence structure, between discharge peaks of daily discharge at 298 different locations on the river network of major European rivers, was captured with a mixture multivariate dependence model. A catalogue of spatially coherent synthetic event descriptors was created, containing 10.000 years of synthetic discharge peaks with a dependence structure that is similar as in the observed data, thereby showing spatially coherence. This catalogue can potentially be used as a starting point for the exploration of the range of possible scenarios of pan-European flooding and associated probabilities, which is the foundation of flood risk assessment.

However, several limitations were encountered, which require awareness for the usage of the catalogue. A major limitation of event-based SSTPA was described in Section (4.5.2). The a priori choice of a set of locations at which events are to be described leads to an incomplete matrix of event descriptors. Current statistical methodology is not compatible with incomplete matrices, such that the matrix has to be compromised by, first, adjusting the event identification procedure and, second, filling the gaps in the descriptor matrix with auxiliary values (i.e. non-specific event descriptors), such as time-window maxima. Since this particular case study was executed on a large spatial scale, there were many gaps in the matrix, which meant the matrix was severely compromised. Given the problem of gaps, it can be concluded that event-based SSTPA is not spatially scalable, which is a problem when the methodology is to be applied on a large spatial scale. Limitations in the reconstruction of hydrographs were described in Section (4.5.3). In particular, it would be difficult to reconstruct hydrographs from a matrix which holds inaccurate event descriptors as a result of the gap filling.

As a result of these limitations, the analysis stopped at the generation of the descriptor set. This implies that the original framework, displayed in Fig. (3.2), was reduced to the framework used in this case study, displayed in Fig. (4.1), by dropping the last two steps: event reconstruction; and space-time placement and gap filling. The limitations drove the decision to develop a new approach: event-based, dynamic, spatio-temporal probabilistic analysis (DSTPA), which will be applied to atmospheric variables in Chapter (5).

5

Development of a dynamic spatio-temporal weather generator

5.1. Introduction

Large-scale flood events cause devastating impact on the human society worldwide [27]. Probabilistic flood risk analysis (FRA), based on large sets of synthetic flood event data, provides a scientific approach to study the consequence of such flood events [42]. Stochastic precipitation generators [1, 4–7, 13, 15, 16, 19, 23, 24, 30, 37, 41, 45, 47, 60, 89, 96, 98, 101] can be used to generate synthetic precipitation data, by which hypothetical flood scenarios (i.e. scenarios that have not occurred) can be explored.

Flood risk analysis (FRA) based on large sets of synthetic flood event data provides a scientific approach to study the consequence of such flood events [42]. Three different approaches can be distinguished, with an increasing level of complexity.

First, for small and individual regions of interest, weather generators were used on a single-site basis. Cameron et al. [16], Eagleson [30], and Hebson and Wood [45] applied a local, event-based, probabilistic analysis, where the single-site was represented by catchment average precipitation. A recent overview in which current weather generators were tested for local weather generation was provided by Vu et al. [96], exploring different climatic regions individually. They tested local performance only, as they did not include any tests for spatial dependence.

Second, the extension from single-site analysis to multi-site received a lot of attention in the last two decades, addressing the challenge of spatial dependence [50]. For example, Wilks [101] described a local daily precipitation generator which he extended to multi-site. Wilby, Tomlinson, and Dawson [98] applied a similar approach, where they separated their gauging stations into sub-regions. Fowler et al. [36] provided an extension by linking sub-regional multi-site models to capture the historical spatial monthly cross-correlation between the sub-regions. Serinaldi [78] used copulas for the capturing of spatial dependence. Bárdossy and Pegram [5], Davison, Padoan, and Ribatet [26], Hundecha, Pahlow, and Schumann [47], Srikanthan and Pegram [85], Verdin et al. [91], and Zhang et al. [106] provided further extensions or variations of methodology. Serinaldi and Kilsby [79] describe a method to obtain meta-Gaussian daily precipitation fields for all of Austria, which may be interpreted as a step in which a significant increase in number of sites is considered, outlined on a grid. What these studies have in common is that a set of sites of interest was predefined, where for the probabilistic analysis each site was considered in each time step. This type of analysis will be referred to as 'static spatio-temporal probabilistic analysis' (SSTPA). SSTPA assumes that the essence of the physical phenomenon studied can be captured in snapshots, i.e. that images can be defined with relevant values for all locations studied. Keef, Tawn, and Svensson [53] and Quinn et al. [68] define events in time series, but then take concurrent values at all other sites. This means events are delineated in time but not in space, i.e. events occurs everywhere in a particular time window, which makes it suitable for SSTPA. None of these precipitation generators of the type SSTPA use spatio-temporal events, which may be explained by the difficulty of incorporating an approach that makes use of spatio-temporal events into a SSTPA [28]. An overview of stochastic precipitation generators of the type SSTPA was recently provided by Ailliot et al. [1].

Third, the contrasting approach of 'dynamic spatio-temporal probabilistic analysis' (DSTPA), which was applied in this chapter. DSTPA is an event-based approach, where 'dynamic events' are conceptualised to be clusters of related data that occur at different locations in space and time with (potentially) time varying spatial extents (and therefore spatially varying durations). Physical phenomena, such as moving precipitation fields, often show a dynamic, progressive behaviour, rather than a static behaviour. Rainfall organization and movement within basins is an essential control of flood response and in particular of hydrograph timing [87]. DSTPA allows dynamic behaviour to be captured, thereby potentially providing a more physically meaningful event definition (GO3 that will be formulated in Section (5.2.3)). In addition, for a particular site of interest, moving events allow inference from data occurring outside that particular site, which will be discussed later in Section (6.5.1). A number of studies performed (subroutines of) DSTPA. In the study area of droughts, which is inversely related to floods, some studies applied DSTPA several years ago [3, 22]. Corzo Perez et al. [22] studied gridded discharge output from a global hydrological model and captured low discharge anomalies. They connected events in space using an algorithm that looks at neighbouring pixels. The result was a catalogue of observed dynamic events of low discharge. Compared to low discharge anomalies, precipitation fields move around at higher speeds, such that higher resolution data products are required. Within the context of flood related DSTPA, Cowpertwait [23] extended the single-site pulse process proposed by Rodriguez-Iturbe, Cox, and Isham [71] to the spatio-temporal Neyman-Scott pulse process (STNSPP). The original concepts in this paper were further extended and developed into a well-established rainfall generator called 'RainSim' [15]. Using RainSim, synthetic, dynamic events are generated from an analytical model (disks of rainfall) and placed in space and time using the STNSPP. Vallam and Qin [89] concluded that RainSim compares favourably to three other generators (of the type SSTPA) concerning the capturing of spatial dependence.

Recently Vorogushyn et al. [93] called for new methods in large-scale FRA to enable the capturing of system interactions and feedbacks. Large-scale flood events are usually triggered by extreme precipitation fields with long duration, high intensity or both. Large-scale synthetic precipitation fields can be used to study system interactions and feedbacks, since in the atmosphere there are no direct human interventions such as in catchments, e.g. dikes; dams; reservoirs; changes in land use. Gridded re-analysis data [72] may provide the means to develop a next generation of stochastic weather generators of the type DSTPA, since for probabilistic analysis a long record of coherent data is required. This type of modelled input data can eventually be replaced by more direct precipitation observations, such as satellite-derived high-resolution precipitation products [55].

In addition to the generation of synthetic data of a single variable, recent developments in the literature show an increasing level of interest in combinations of multiple variables, i.e. in 'compound events' [48, 59, 69, 75]. Compound, event-based analysis makes sense for sources that

are dependent, but between which there is no direct causality. The methodology for of a single variable can be be extended to handle compound events, which introduces additional difficulty.

Outline In Section (5.2), the methodology will be introduced, comprising the used data set, the objectives and the framework. In Section (5.3), methodology will be introduced for DSTPA and applied to a single variable (precipitation). In Section (5.4), extensions of methodology for DSTPA will be introduced for compound variables (precipitation, wind and (low) pressure). In Section (5.5), the different steps will be centrally discussed and recommendations will be provided for improvement and extension of the methodology.

5.2. Methodology

5.2.1. Data

For this case study, the Climate Forecast System Reanalysis (CFSR) data set [73] was used, which is a gridded reanalysis product of several variables with global coverage. Precipitation rate, wind speed and pressure were selected, all of which have a .3125 degree spatial resolution, both in longitude and in latitude. The temporal resolution of these data sets is hourly, which was critical to be able to track moving fields. In order to reduce the computational burden, a spatial subset was taken of the data, as displayed in Fig. (5.1).

5.2.2. Generator objectives

There are three main generator objectives:

1. introduce the unobserved (GO_1),
2. reproduce patterns (GO_2) and
3. respect physical restraints (GO_3).

These generator objectives were described in more detail in Section (3.2).

The aim in this chapter is to provide a methodology (of the type DSTPA) to generate a large data set of synthetic, continuous data.

5.2.3. Framework

The step-by-step framework applied in this case study, which was introduced in Section (3.3) and displayed in Fig. (3.2). First, dynamic spatio-temporal events were identified. They were conceptualised to be coherent clusters of data, separated by the application of a mask to the data set. Second, of each event a wide range of descriptors was captured. Some descriptors were used to classify the events, others were used for multivariate analysis. Third, a non-parametric dependence model was fitted and an extremal dependence model, from which a large set of synthetic descriptors was sampled. The objectives in this step were to empirically fill up the marginal distributions while retaining the dependence structure between the variables, both in the non-extreme parameter space and the extreme. Fourth, duplicates of the observed event catalogue were created and the events were manipulated with the simulated, synthetic descriptors. This led to a large catalogue of synthetic events. Fifth, duplicates of the entire, continuous precipitation data set were created and, the original observed events were substituted with the created synthetic events.

5.3. A generator of precipitation

5.3.1. Event identification

Dynamic spatio-temporal precipitation events were identified using the following procedure.

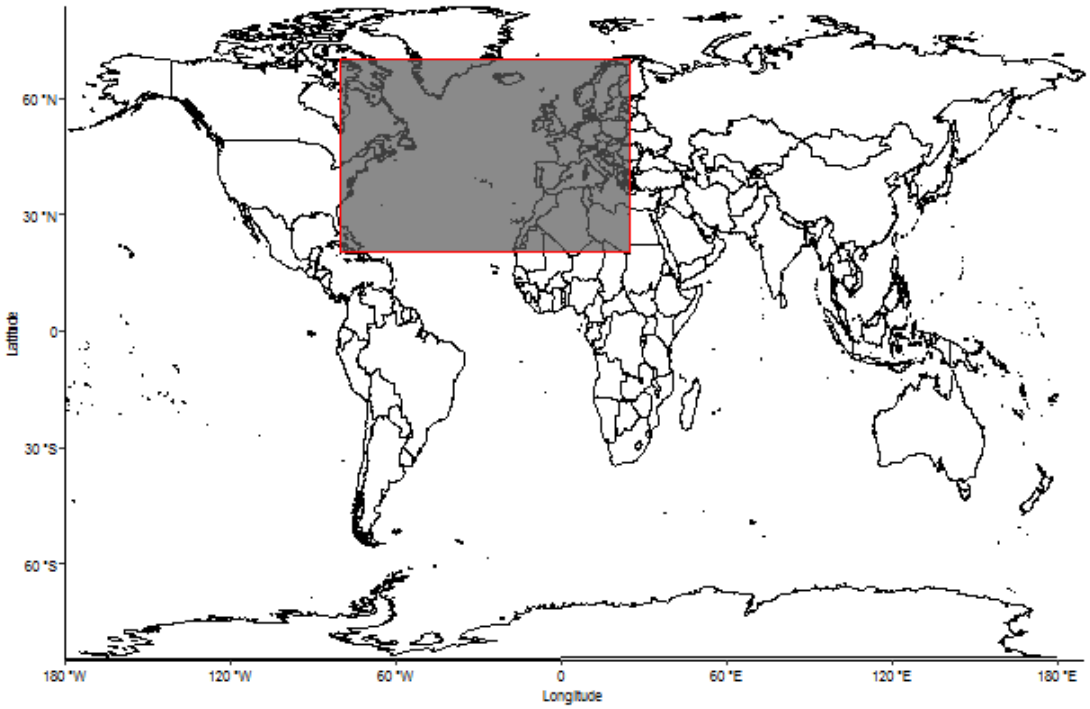


Figure 5.1: Spatial domain (BSL2) of the dynamic spatio-temporal generator used in this chapter. The British Isles are the target area under study. The spatial domain is chosen across the atlantic, since most precipitation events travel with the south-westerly winds.

Separation Clusters of data points first had to be separated, for which a global threshold was set. Data values below the global threshold were masked (i.e. set to NA). The threshold was set to $v = 1.8mmh^{-1}$, based on a trial and error procedure. The threshold needed to be sufficiently high such that clusters of precipitation could be separated into events, where the resulting events could be assumed to be independent. On the other hand, the threshold should not be too high to ensure a reasonable amount of data (and hence events) are available for fitting a statistical model. A lower threshold would allow the introduction of more variability (GO1), since a larger part of the observed data set would be included in the catalogue of observed events. This will be further discussed in Section (6.3.2).

Cluster labelling With the use of image analysis, polygons were captured in (the image of) each time step. Each polygon was assigned an initial unique event ID $\{ID_1, ID_2, ...\}$. Connections between (polygon) events were established. Across subsequent time steps, connections were established based on the spatial overlap of events, i.e. a connection was established between ID_i from time step t and ID_j from time step $t + 1$ if the two polygons intersected (i.e. overlapped). Implicitly, temporal connections were established over multiple time steps. For example, event ID_i at time step t was connected to event ID_j at time step $t + 1$, which was in turn connected to event ID_k at $t + 2$, and so on. Fourth, after all connections were established, all (directly and indirectly) connected polygons were merged and defined as one identified event.

Result Fig. (5.2) shows a sequence of moving precipitation events. The event identification procedure resulted in a cat-

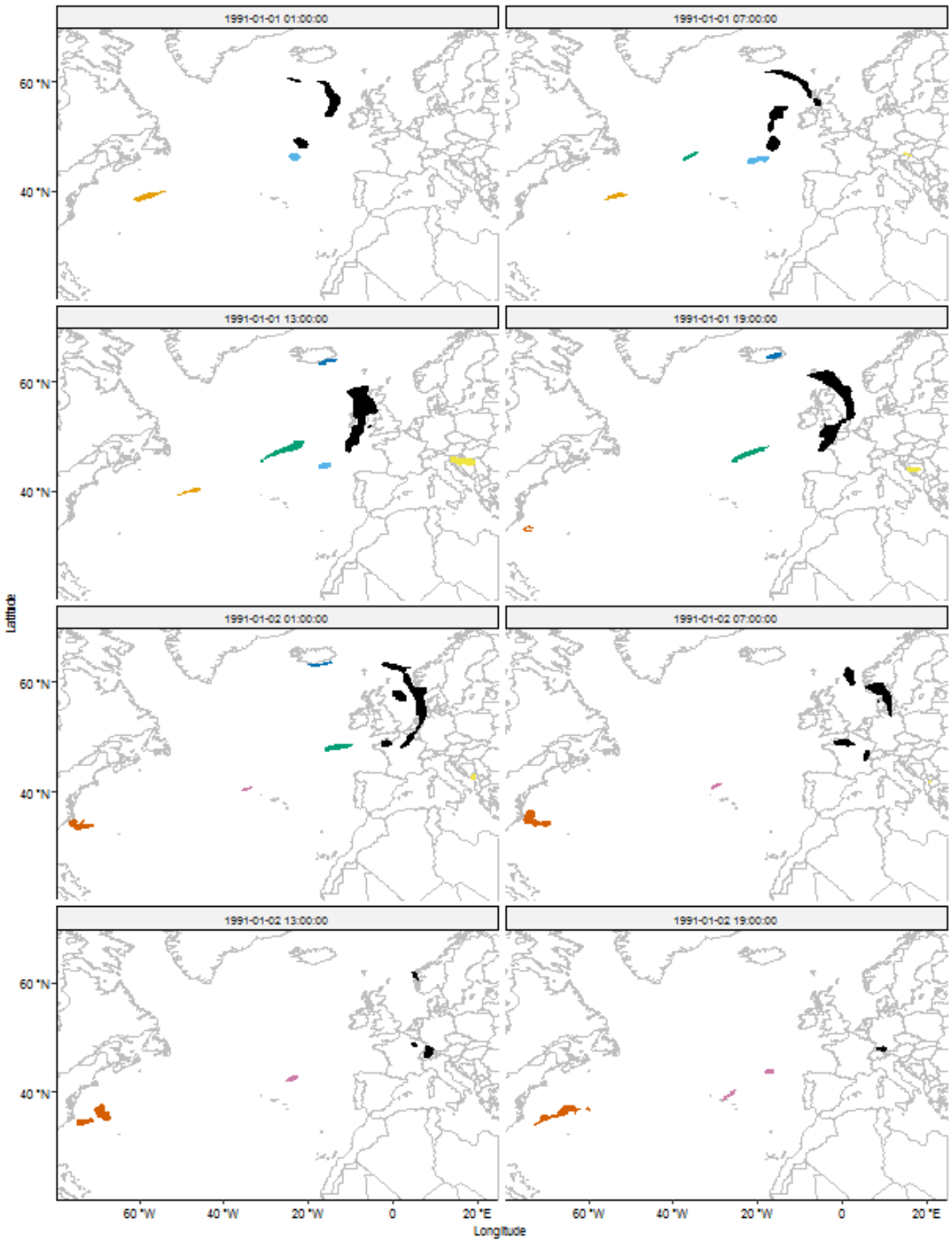


Figure 5.2: A sequence of a few moving observed precipitation events. Eight events are displayed, each represented by a unique colour.

alogue of 196810 observed events. Similar algorithms have been used for identifying dynamic spatio-temporal events in gridded data sets. For example, Corzo Perez et al. [22] identified dynamic events of low discharge anomaly for drought analysis.

5.3.2. Event description

Using statistical analysis, the statistical properties of the event descriptors will be modelled in Section (5.3.3). Instead of performing the statistical analysis over all events in a lumped fashion, the statistical properties are assumed to differ from one class of events to another. Hence, first a classification step was applied.

Classification For the classification procedure, two main decisions had to be made, namely the number of classes and which event descriptors to use for classification. For the former, from a statistical point of view, each class should hold sufficient data to be able to confidently fit joint-distributions, which implies that there should not be too many classes. Although no strict guidelines were available, the multivariate extremal dependence model would benefit from about 500-10000 events per class. With 196810 identified events, this implied a range of about 20-400 classes. From a physical point of view (GO3), each class should be physically different from the others (i.e. a class could represent hurricanes in region A, thunderstorms in region B, etc.), which implies there should be a sufficient number of classes. It was estimated that 196 separate classes should be sufficient to capture physically distinguishable classes. For the latter, a combination of four descriptors was considered. To be able to classify events according to region, the spatial centres of gravity of the events (cog_x and cog_y) were used . Two descriptors were included ($rat_{P/V}$ and $cogV_t$) that provided means to distinguish different physical behaviour of the precipitation events (GO3). The ratio of the peak precipitation rate divided by the precipitated volume $rat_{P/V}$ provides an indicator for the heterogeneity of events, where a large value points towards a relatively heterogeneous event and a small value points towards a relatively homogeneous event. The temporal centre of gravity of the precipitated volume $cogV_t$ (normalised, i.e. duration set to 1) provides an indicator of timing of precipitation release of events, where a small value indicates that events showed a large initial release and then slowly decayed, and vice versa. Several standard approaches were tested for cluster analysis, including hierarchical cluster analysis, partitional algorithms and self organising maps (SOM); see Wilks [99] for a comprehensive list of options. Finally, the technique of self organising maps [56] was used, because of its efficiency in handling large data sets.

Fig. (5.3) shows a sample of the resulting classes. Although different classes may exhibit similar distributions in a subset of the four classification descriptors, the classes were defined to ensure the joint distribution of the full set of classification descriptors were statistically distinguishable between one class and another.

Multivariate descriptor sets Having completed the classification, statistical analysis was performed separately for each class of events. The choice of descriptors for classification depended on their ability to produce distinct classes of comparable events. The choice of descriptors to which to apply the multivariate statistical

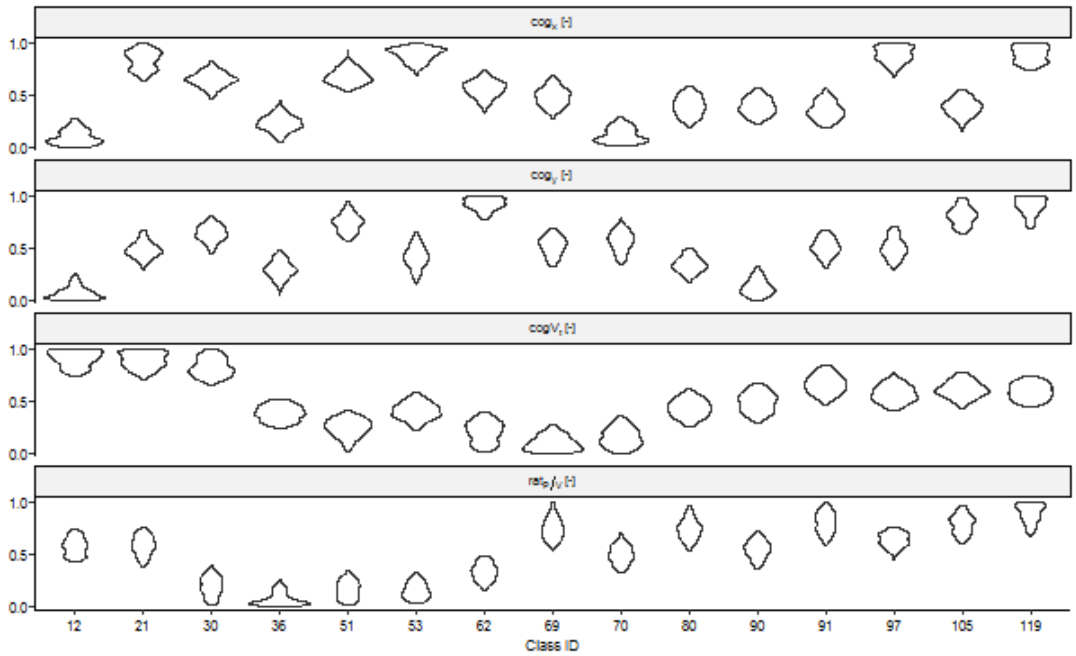


Figure 5.3: Density plots of the (normalised) descriptors used for classification. A sample of 16 different classes is displayed. cog_x and cog_y are the spatial centres of gravity, $cogV_t$ is the temporal centre of gravity of the volume, $rat_{P/V}$ is the ratio of peak over volume.

analysis depended on the reconstruction procedure, which will be discussed in Section (5.3.4). Therefore, the descriptors used for the multivariate statistical analysis were not the same as the descriptors used for classification. For all classes the same event descriptors were used: volume $V[m^3]$, which is the sum of all precipitation in the event; peak $P[mh^{-1}]$, which is the maximum precipitation rate in the event; and spatio-temporal extent $E[m^2h]$, which is the sum of spatial extent of all time steps in the event (i.e. if an event occurs at a particular grid-cell for three hours, the spatial extent of that grid-cell is counted three times).

Result The event description procedure resulted in 196 matrices of observed descriptors with three columns (P,V,E), one for each class, with a variable number of rows per matrix equal to the number of observed events belonging to the respective class.

In order to align with the corresponding literature in statistical models for multivariate extreme values, in the next Section (5.3.3) columns of these matrices will be referred to as margins and the large values in each column will be referred to as the upper tails of the marginal distributions.

5.3.3. Multivariate statistics

Joint distribution model To model the joint distributions of event descriptors within each separate class, a copula-like method [38, 62] was adopted, whereby the multivariate distribution was decomposed to the marginal distribution and the dependence structure. This involved fitting a univariate probability distribution for each margin as well as fitting models to capture the dependence structure between different margins. The concern was in particular with extreme events, where either one or some of the descriptors were very large. This was reflected in the proposed approach for both the marginal and the de-

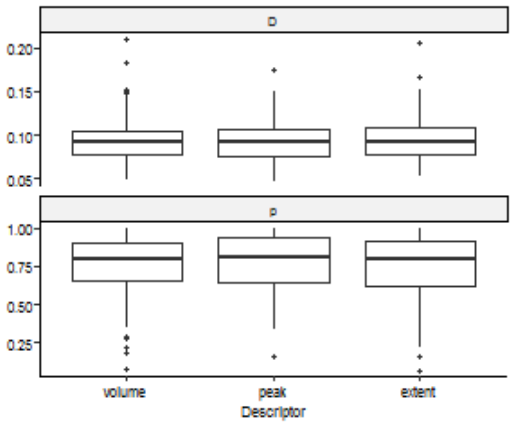


Figure 5.4: Standard Kolmogorov-Smirnov statistics are used to test the quality of the GPDs fitted to the marginals of the spatio-temporal descriptors. Above, the distance D is displayed which should be low. Below, the $p-value$ is displayed which should be close to 1.

pendence model.

Marginal distributions A mixture model was used for the marginal distributions, assuming values above a large threshold to follow a Generalised Pareto Distribution (GPD), where values below were captured by the empirical distribution. This type of mixture model is considered suitable for situations where attention is mostly focussed on the extrapolation of the distribution into high quantiles. A mixture model has been used for various applications of environmental data modelling [57, 105]. Scarrott and MacDonald [76] provided a detailed description regarding the formulation of threshold-based mixture models.

The GPD for the upper tail of each descriptor (i.e. each column of the observed descriptor matrix) was fitted using likelihood-based inference [20]. Based on standard threshold diagnosis suggested, the 90% quantile was deemed a reasonable threshold for fitting the GPD. This threshold was sufficiently high so that the limiting properties of the tail distribution could be approximated by the GPD and not too high that the variance of the GPD parameter estimation was prohibitive. With three event descriptors and 196 classes, the upper tail distributions were captured with 588 marginal GPDs.

Fig. (5.4) shows quality tests of the marginal GPD-fits, where the empirical quantiles and probabilities were compared against the modelled using standard Kolmogorov-Smirnov statistics. Overall, reasonable GPD fits were obtained.

Dependence structure Similar to the marginal model, the dependence structure was captured using a multi-dimensional mixture of two models. Such a multi-dimensional mixture was used previously by Bortot, Coles, and Tawn [11], who modelled coastal variables. First, the lower part of the distribution was modelled using a kernel density distribution [83], as there was no obvious choice of a parametric model to describe the dependence in this region. The use of the kernel bandwidth (rather than direct re-sampling) enabled us to draw values that were broadly similar but not identical to the observed data. Second, the upper tail region (i.e. where one or multiple variables were large) was modelled by the HT04 model, described in detail in Section (3.4.4).

The lower part of the joint distribution was captured using a multivariate kernel density based on Scott's rule-of-thumb choice of optimal bandwidth [83]. This was an improvement over the original rule-of-thumb proposed in Silverman (1982). Pairwise estimation of the HT04 parameters was performed based on pseudo-likelihood, separately for each class of pre-

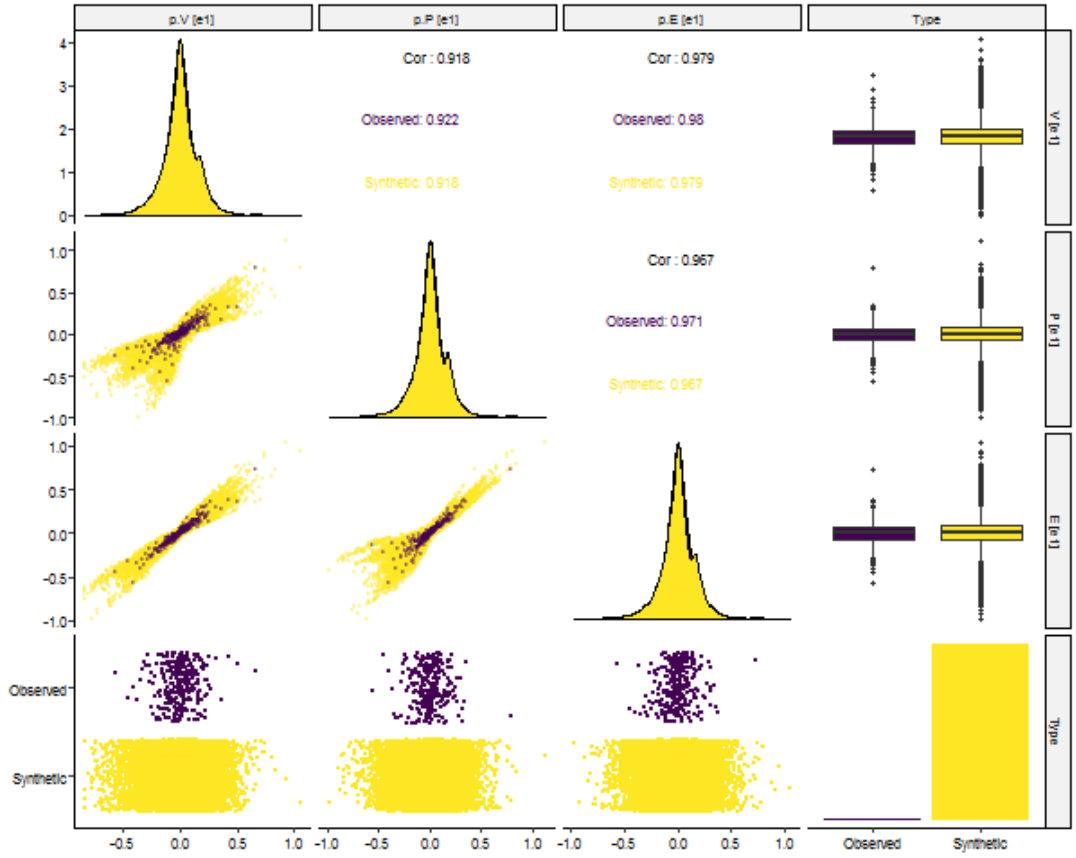


Figure 5.5: Pair-wise plots of the observed descriptor matrix and the synthetic descriptor matrix of class 6. Purple (dark) is the observed and yellow (light) is the simulated. V is the sum of all precipitation, P is the maximum precipitation rate and E is the spatio-temporal extent. In the diagonal, distributions of observed and synthetic descriptors are compared using box-plots. Below the diagonal, pair-wise scatter plots are displayed. Above the diagonal, pair-wise correlations are displayed. The data was transformed to the Laplace space.

cipitation events. Due to the complexity of the method, readers are advised to refer to the original paper for detailed steps in fitting the model. Similar to Heffernan and Tawn [46], the joint upper tail area, where multiple margins exceeded the threshold, was partitioned into subspaces such that the conditioning margin Y_i in the above formulation was always the largest margin in terms of quantile probability. Equal thresholds were used for all directions of conditioning as no substantial improvement in model fitting was found by allowing the threshold to be variable for different conditioning margins. The final threshold of choice was the 95% quantile of the Laplace distribution, based on the consideration of number of data points available above the threshold as well as the outcome of tests of independence between the residual distribution Z_{-i} and the values of the conditioning margin y_i.

With three descriptors and 196 classes, the dependence structure was captured with 196 multivariate kernel densities (one for each class) and 588 HT04 models (one for each conditioning descriptor in each class).

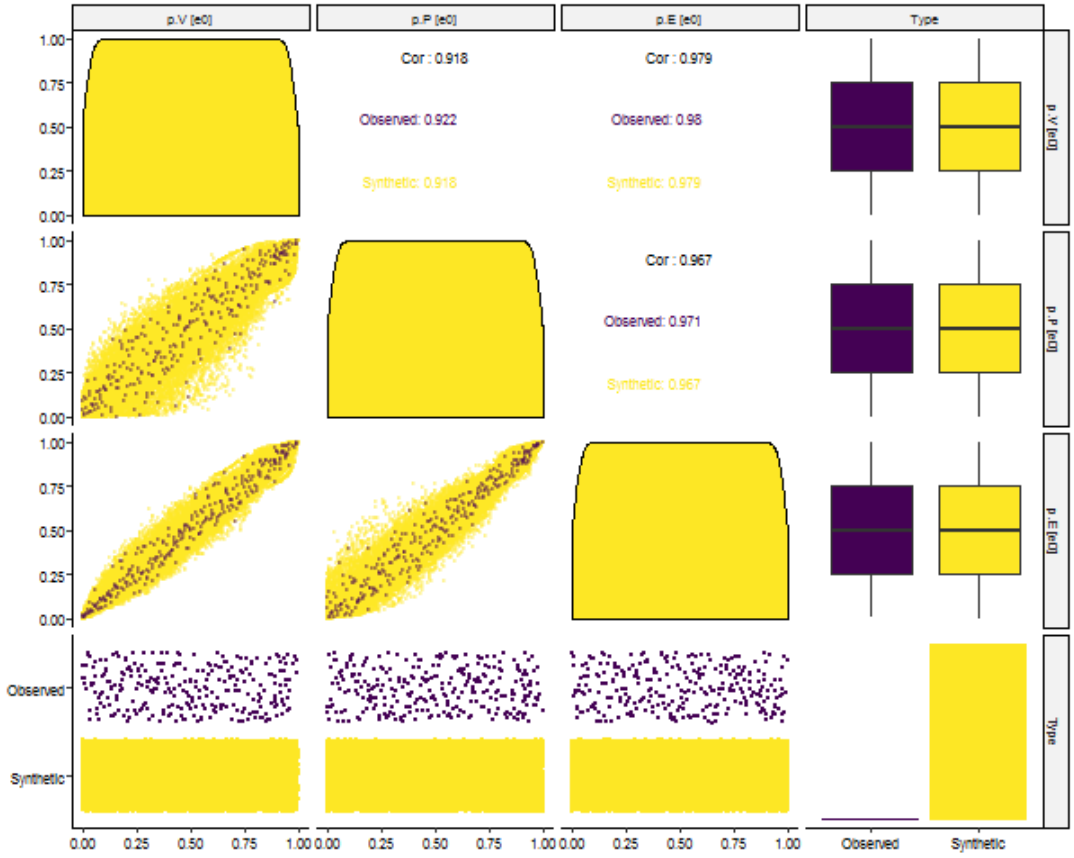


Figure 5.6: Pair-wise plots of the observed descriptor matrix and the synthetic descriptor matrix of class 6. Purple (dark) is the observed and yellow (light) is the simulated. V is the sum of all precipitation, P is the maximum precipitation rate and E is the spatio-temporal extent. In the diagonal, distributions of observed and synthetic descriptors are compared using box-plots. Below the diagonal, pair-wise scatter plots are displayed. Above the diagonal, pair-wise correlations are displayed. The data was transformed to the uniform space.

Simulation Similar to the model fitting, the simulation of the synthetic data was performed for each class of precipitation events separately. First, $n_{syn} = n_{obs}T_{syn}/T_{obs}$ descriptor sets were sampled from the multivariate kernel density distribution, where n_{obs} is the observed number of events in the class and T is the total duration of the respective data set. With the space-time placement procedure in mind, which will be discussed in Section (5.5.4), exactly $n_c = 31$ synthetic descriptor sets were sampled from each observed, such that 1000 years of synthetic event descriptors were simulated from 32 years of observed. Second, the synthetic descriptor sets in which the largest marginal cumulative probability exceeded the threshold $p_{qu} = 0.99$ were replaced them with simulated descriptor sets drawn from the corresponding HT04 model (conditional on the largest margin). As suggested by Heffernan and Tawn [46] a re-sampling method was used to draw samples from the residual distribution Z_{-i} in the HT04 model. With the reconstruction procedure in mind, which will be discussed in Section (5.5.4), this re-sampling scheme ensured that each synthetic descriptor set had a corresponding observed set. Third, the marginals of the

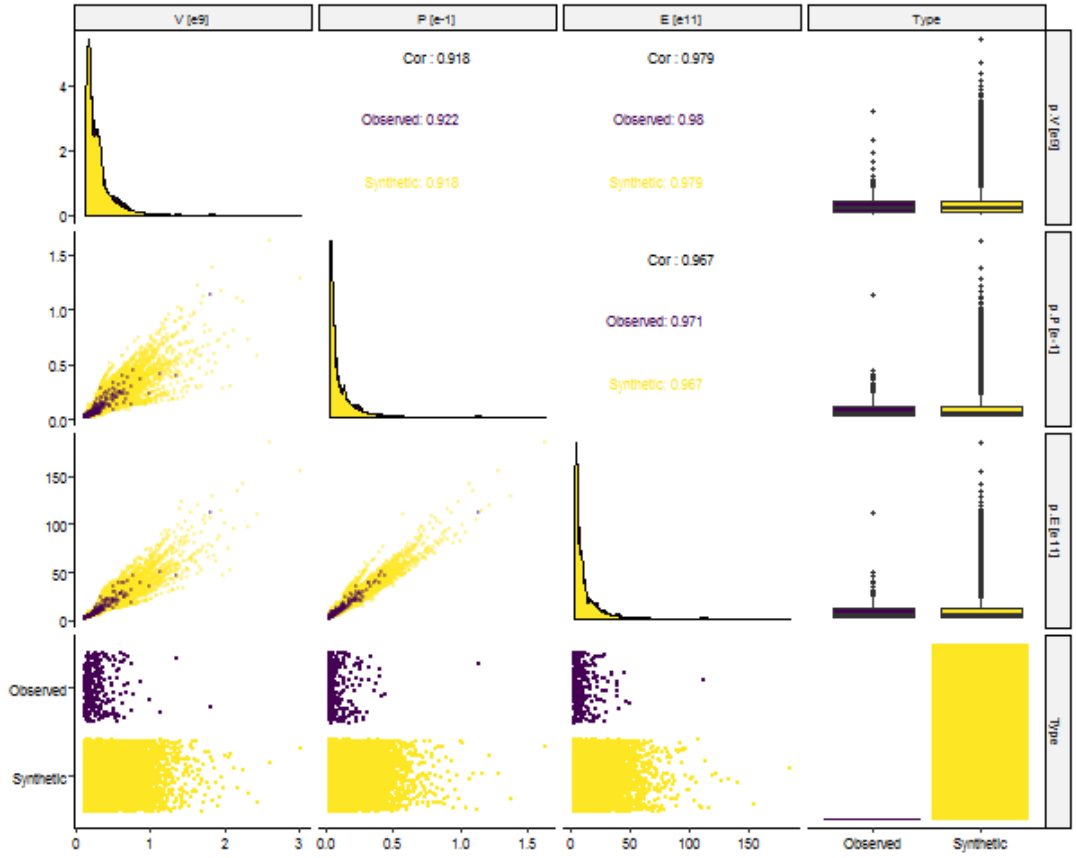


Figure 5.7: Pair-wise plots of the observed descriptor matrix and the synthetic descriptor matrix of class 6. Purple (dark) is the observed and yellow (light) is the simulated. $V[m^3]$ is the sum of all precipitation, $P[mh^{-1}]$ is the maximum precipitation rate and $E[m^2h]$ is the spatio-temporal extent. In the diagonal, distributions of observed and synthetic descriptors are compared using box-plots. Below the diagonal, pair-wise scatter plots are displayed. Above the diagonal, pair-wise correlations are displayed. No transormations were applied to the data.

synthetic descriptors were transformed to respect the mixed distributions fitted to the observed marginals. This implies that the synthetic marginals were forced to have the exact same distribution as the corresponding observed marginals. However, by forcing this transformation the dependence structure was slightly distorted.

Result The capabilities of the statistical dependence model were already tested and demonstrated in Section (4.3.3), where the model was stressed with application to a matrix that had more (298) columns and fewer (428) rows. Therefore, for demonstrative purposes, class 6 was picked by random selection in order to demonstrate the results.

Fig. (5.5) shows the direct result of the simulation, which was applied in the Laplace space. As can be observed, each variable centres around zero, which is typical for the Laplace space.

Fig. (5.6) shows the transformation of the variables to the uniform space. As can be observed, each variable falls in the range of zero to one, which is typical for the uniform space. Since the pairwise correlations are based on the rank (Spearman), they are not influenced by monotone transforma-

event	$V[m^3]$ $\times 10^{10}$	$P[mh^{-1}]$ $\times 10^{-2}$	$E[m^2h]$ $\times 10^{13}$
$D_{obs,a,1}$	4.01	1.77	1.02
$D_{syn,b,1}$	0.57	1.16	0.14
$D_{syn,c,1}$	7.18	2.57	1.93
$D_{obs,a,2}$	6.14	1.93	1.15
$D_{syn,b,2}$	1.46	1.41	0.18
$D_{syn,c,2}$	11.01	3.22	2.43

Table 5.1: Descriptor sets used for the reconstruction example displayed in Fig. (5.8). V is the total precipitated volume, P is the peak precipitation rate, E is the spatio-temporal extent.

tions.

Fig. (5.7) shows that with the synthetic descriptor sets a much wider variety of events was explored than with the observed (GO1) and that the observed descriptors were a likely subset of the synthetic (GO2), with similar correlations and identical marginal distributions (boxplots).

With one synthetic matrix for each observed matrix, the simulation procedure resulted in 196 synthetic matrices. Each row was a complete synthetic descriptor set and where each synthetic matrix had 31 times as many rows as the observed from the corresponding class, since 31 synthetic descriptors were to be sampled to reconstruct 31 synthetic data sets.

5.3.4. Event reconstruction

Required information For the reconstruction of each individual synthetic event, several pieces of information were used: the set of synthetic descriptors; the connected set of observed descriptors, from which the synthetic descriptors were sampled using the statistical models; the observed event, corresponding to the observed descriptor set, which was used as a starting point for reconstruction; and the continuous observed data set, to be able to assign values to synthetically expanded events. For each synthetic descriptor set, the starting point was the observed event corresponding to the observed descriptor set from which the synthetic set was re-sampled. The following procedure was applied.

Adjustment of extent First, the spatio-temporal extent of the observed event was adjusted, changing it from the observed extent E_{obs} to the synthetic extent E_{syn}. If $E_{syn} < E_{obs}$, the coordinates of the observed event were ranked by value and the coordinate with the smallest value was removed, which was repeated until the extent was reduced to the target value of E_{syn}. If $E_{syn} > E_{obs}$, neighbouring coordinates were added around the boundary of the observed event until the extent reached the target value of E_{syn}. Each round of addition was performed randomly such that all neighbouring coordinates on sharing the boundary with the event extend had an equal probability to be added.

Adjustment of volume and peak Second, within the extended or reduced extent, the original observed volumes were obtained for a given grid-cell X by multiplying the precipitation rate with the spatial extent of the grid-cell. The synthetic volumes for this grid-cell Y were constructed using

$$Y = m_y + (X - m_x)\frac{M_y}{M_x} \qquad (5.1)$$

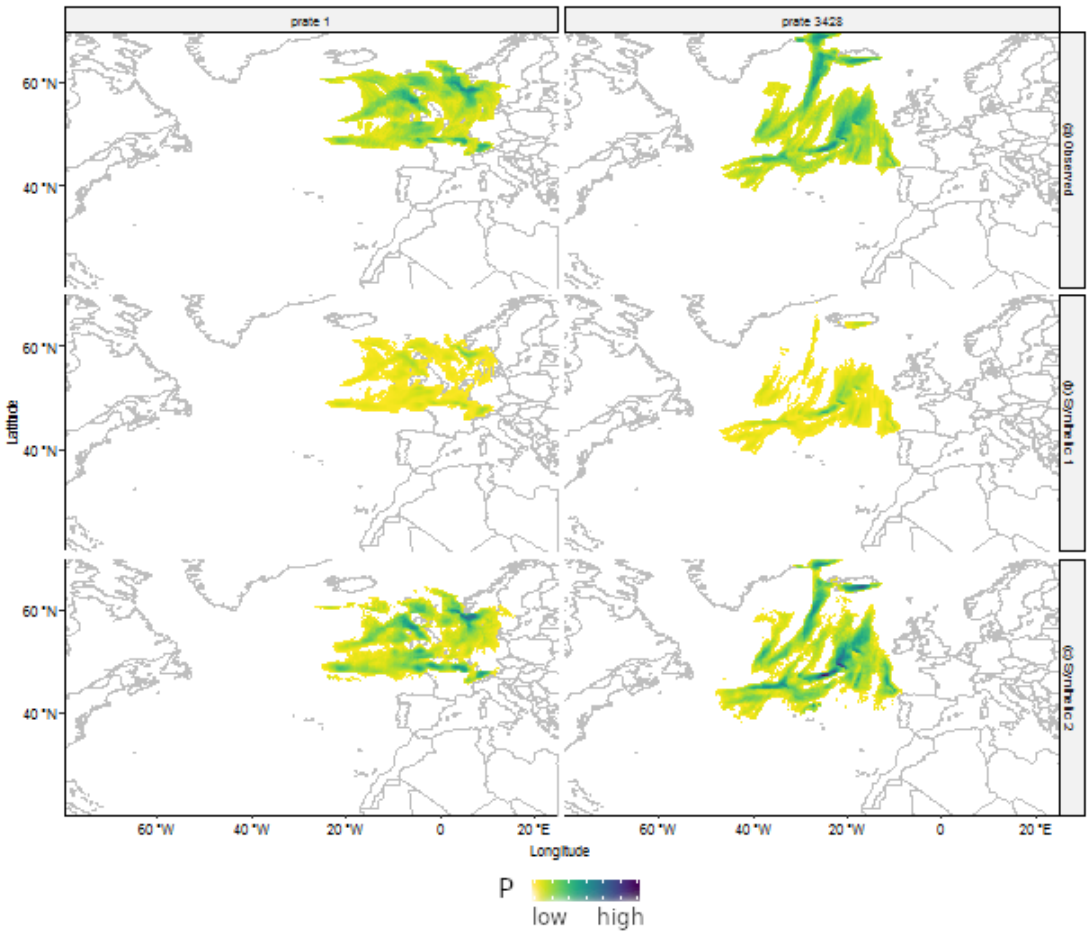


Figure 5.8: Reconstruction of two synthetic events of high precipitation rate, corresponding to the synthetic descriptor sets displayed in Table (5.1). a) The original observed events. b) Synthetic events 1, for which the observed events are reduced in magnitude. c) Synthetic events 2, for which the observed events are increased in magnitude. P is the sum of precipitation per grid-cell over the duration of the event. Indications of low and high are used, because the actual values of P depend on the grid-size.

where

$$m_y = M_y - (M_x - m_x)\frac{M_y - \mu_y}{M_x - \mu_x} \tag{5.2}$$

with

$$\mu_y = \mu_x \frac{V_{syn}}{V_{obs}} \tag{5.3}$$

and

$$M_y = M_x \frac{P_{syn}}{P_{obs}} \tag{5.4}$$

where $m_x = min(x)$, $M_x = max(x)$, $\mu_x = mean(x)$. This calculation preserves the shape of the distribution of the observed volumes with the mean and peak updated from μ_x to μ_y and from M_x to M_y respectively. Finally, the synthetic volumes per grid-cell were transformed back to precipitation rates using the spatial extent of the grid-cells.

Result Fig. (5.8) demonstrates two examples of the reconstruction of a synthetic event. Four sets of synthetic descriptors and the two connected observed descriptor sets are displayed in Table (5.1). The observed events, with corresponding set of observed descriptors D_{obs}, are displayed in Fig. (5.8)a. Two synthetic events are displayed in Fig. (5.8)b, with corresponding sets of synthetic descriptors D_{syn1}, which demonstrate a reduction in magni-

tude. Two synthetic events are displayed in Fig. (5.8)c, with corresponding sets of synthetic descriptors D_{syn2}, which demonstrate an increase in magnitude.

With 31 synthetic events reconstructed from every observed, the reconstruction procedure resulted in a catalogue of 6.101.110 synthetic events.

5.3.5. Space-time placement and gap filling

Space-time placement To form a 1000-year continuous gridded time series using the catalogue of synthetic events, the following procedure was applied. For each repetition of an observed event, a corresponding synthetic event was randomly sampled (without replacement). By the design made in Section (5.3.3), the number of corresponding synthetic events for each observed event was precisely 31, which meant that all synthetic events were used exactly once.

Gap filling The 31 repetitions of the 32-year observed time series data were stacked in order. The observed event, both extent and precipitation intensity, were replaced with that of the synthetic event. In particular, when the replacing synthetic footprint was larger than that of the observed event, the values in those coordinates that were not covered in the observed event were still overwritten. If the replacing synthetic footprint was smaller, the smallest value was retained and all other values were transformed linearly such that the largest value equalled the global threshold of event identification used in Section (5.3.1).

Result The space-time placement and gap gilling procedure resulted in a gridded, 1000-year continuous time series for the entire modelled region, with a discontinuity in between the subsequent repetitions. Figs. (5.9) and (5.10) display small samples of the observed continuous data versus the synthetic. The difference between the two data sets can be found where the observed events occurred and were replaced by synthetic events.

5.4. Extension for compound precipitation, wind and pressure

5.4.1. Event identification

For the compound generator, events were separated using multiple global thresholds; one for each individual variable. Fig. (5.11) displays the outcome of the event separation method. As can be observed, different variables show overlap in space and time.

In Section (5.3.1), for the clustering of events, connections were established based on the spatial overlap of polygons of a single variable across subsequent time steps. For compound events, additional connections were established based on the spatio-temporal overlap of the different variables.

The additional connections, as a result of the overlap between different variables, resulted in that the compound event identification produced different events than the single variable event identification procedure. Multiple clusters, that in the single variable approach were considered separate events, were clustered together. In particular, the low pressure fields often pasted together separate wind fields, that, as should be expected, can be found on the edges of the low pressure fields.

With three different variables, seven different event types were identified. Three event types consisted of events of the individual variables only. Three event types

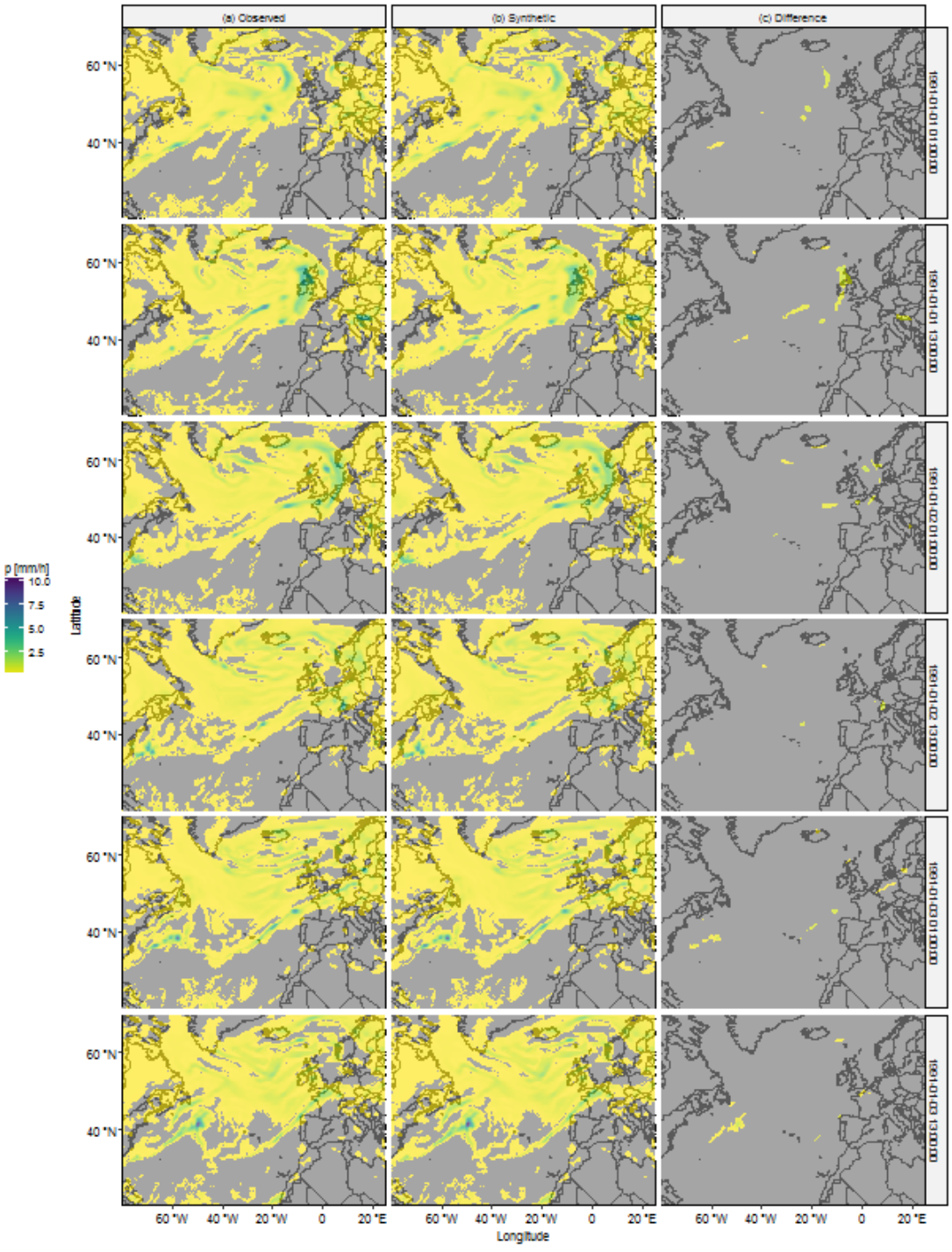


Figure 5.9: A sample is displayed of the precipitation rate p in different time steps of (a) the observed data, (b) the synthetic data and (c) the absolute difference between the observed data and the synthetic.

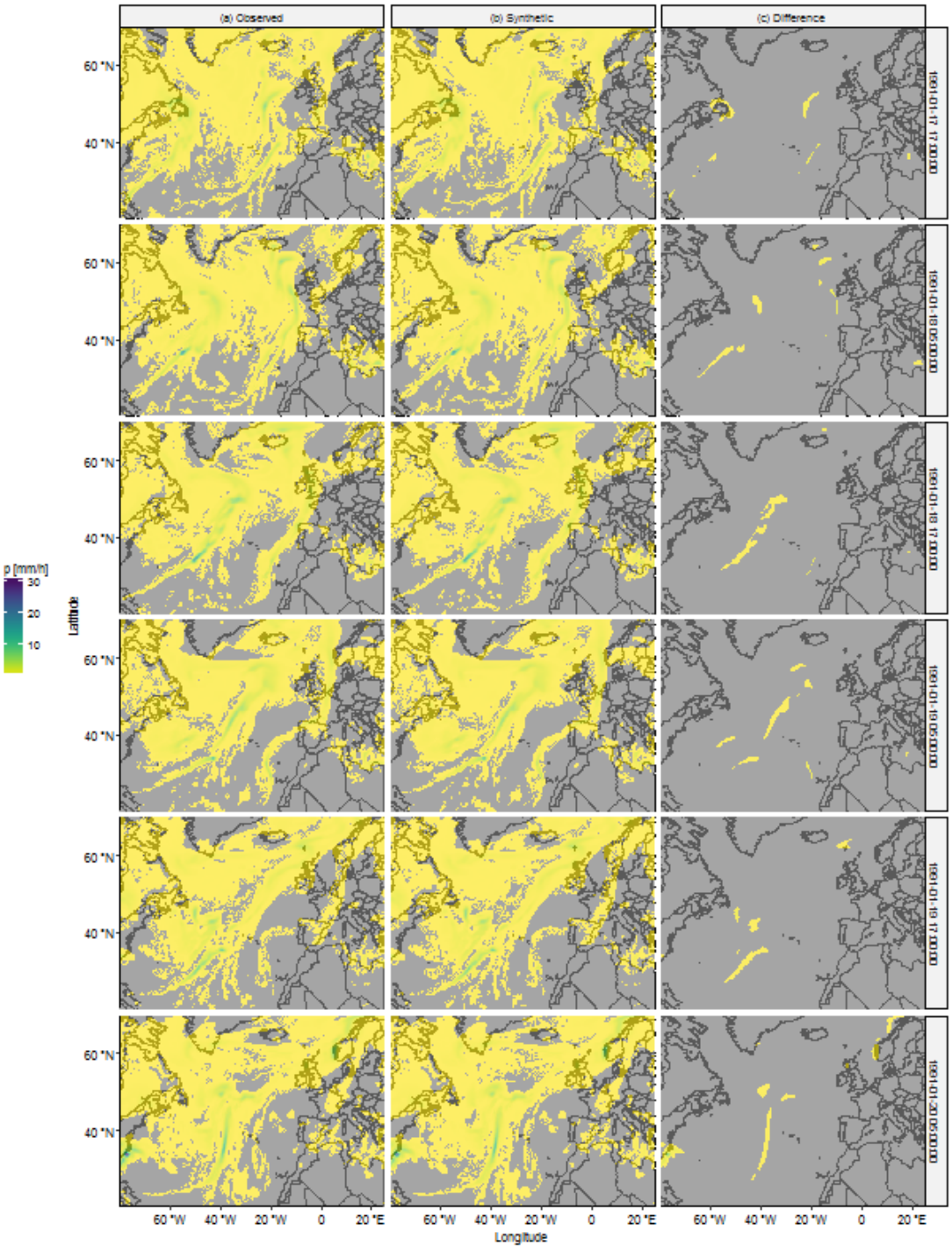


Figure 5.10: A sample is displayed of the precipitation rate p in different time steps of (a) the observed data, (b) the synthetic data and (c) the absolute difference between the observed data and the synthetic.

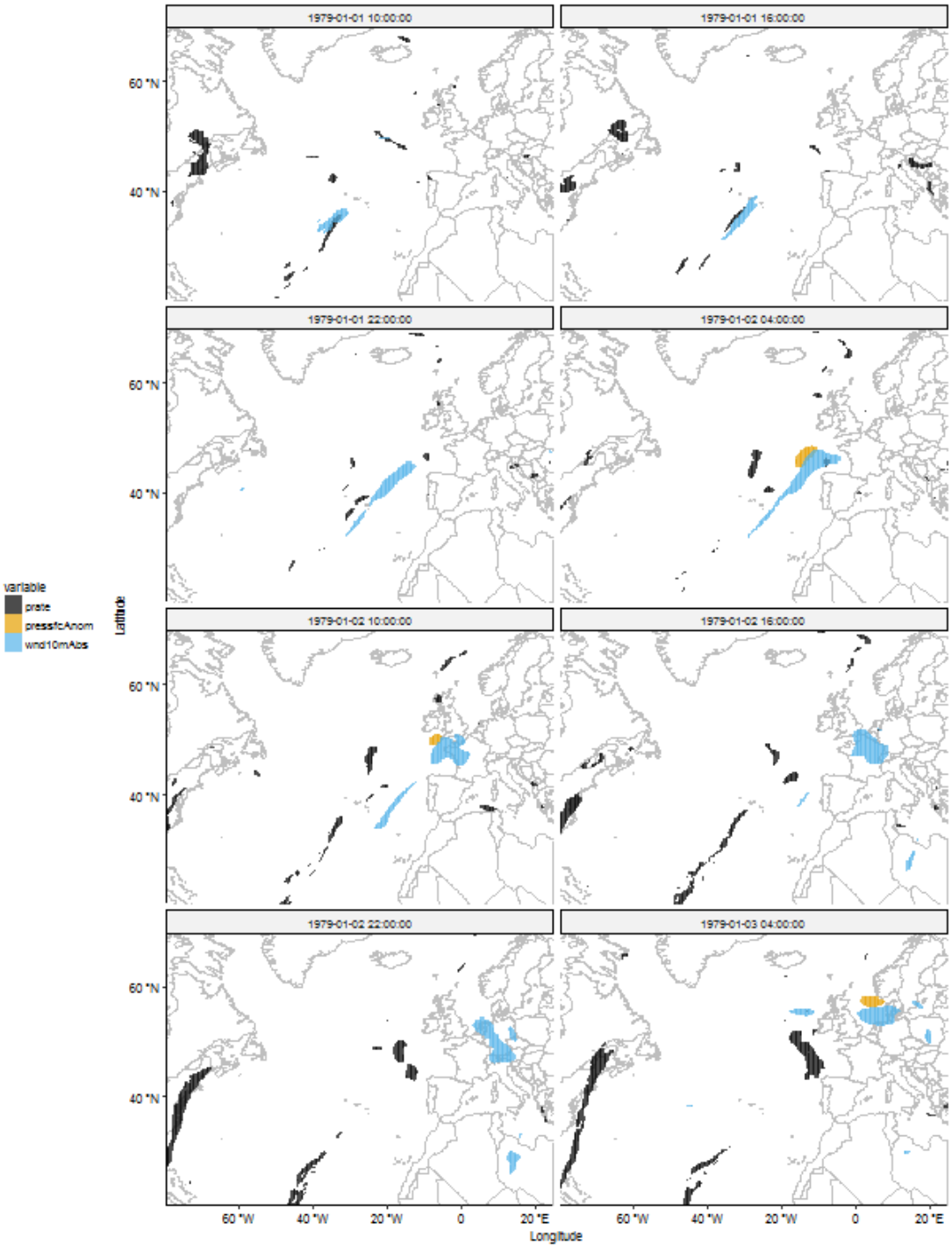


Figure 5.11: The tracking of movement of extremes is displayed for three variables. $prate$ (black) is the precipitation rate, $pressfcAnom$ (orange) is the low pressure anomaly, $wnd10mAbs$ (blue) is the wind speed at 10m high in any direction.

consisted of events of combinations of three variables. One event type consisted of events of all three variables combined. Each event type was handled separately in the subsequent steps in the generation process.

5.4.2. Event description

Per event type, events were classified using the same classification procedure as in Section (5.3.2). Per class, the same event descriptors were captured for the compound procedure as for the single variable procedure: peak (P), volume (V) and spatio-temporal extent (E). However, for the compound events, these event descriptors were captured for each variable in the event type considered.

5.4.3. Multivariate statistics

Per event type, per class, the multivariate structure was captured with the same multivariate model as in Section (5.3.3).

Fig. (5.12) displays the multivariate structure of a particular class of compound events that comprised of combinations of all three variables. The dependence structure was successfully captured and reproduced.

5.4.4. Event reconstruction

Events were reconstructed per variable, similar to the procedure used in Section (5.3.4), irrespective of the event type and class that the events belonged to. The reason for this was to keep the reconstruction procedure relatively simple, manageable and computationally not too expensive.

5.4.5. Space-time placement and gap filling

Events were placed back in space and time at their original locations, similar as in Section (5.3.5).

5.5. Discussion

5.5.1. Event identification

The key difference between the generation procedure for a single variable (in Section (5.3.1)) and for compound variables (in Section (5.4.1)) was found in the event identification procedure. Particular types of events, such as tropical cyclones, could potentially be relatively clearly delineated, providing similar event identification results in both a single and a compound variable framework. However, in a compound variable framework, (relatively large) low pressure systems may lead to the merging of precipitation and wind clusters, which, when considered in a single variable framework, would be considered separate events.

There is no commonly agreed definition of events - what some may consider to be a single event, could be considered to be multiple events by others. The general procedure is to (iteratively) separate and connect (cluster) according to a preconceived, subjective idea of what an event is. This is already the case for a single variable and the compound approach appears to contribute additional subjectivity. It can be observed in Fig. (5.11) that a compound variable approach may lead to different events than the single variable approach, even when the conceptual ideas (thresholds, overlap) are very much the same for both approaches. The subjective nature of the event identification procedure can be considered a hurdle for an objective, scientific approach, since science requires a certain consistency in results. However, event

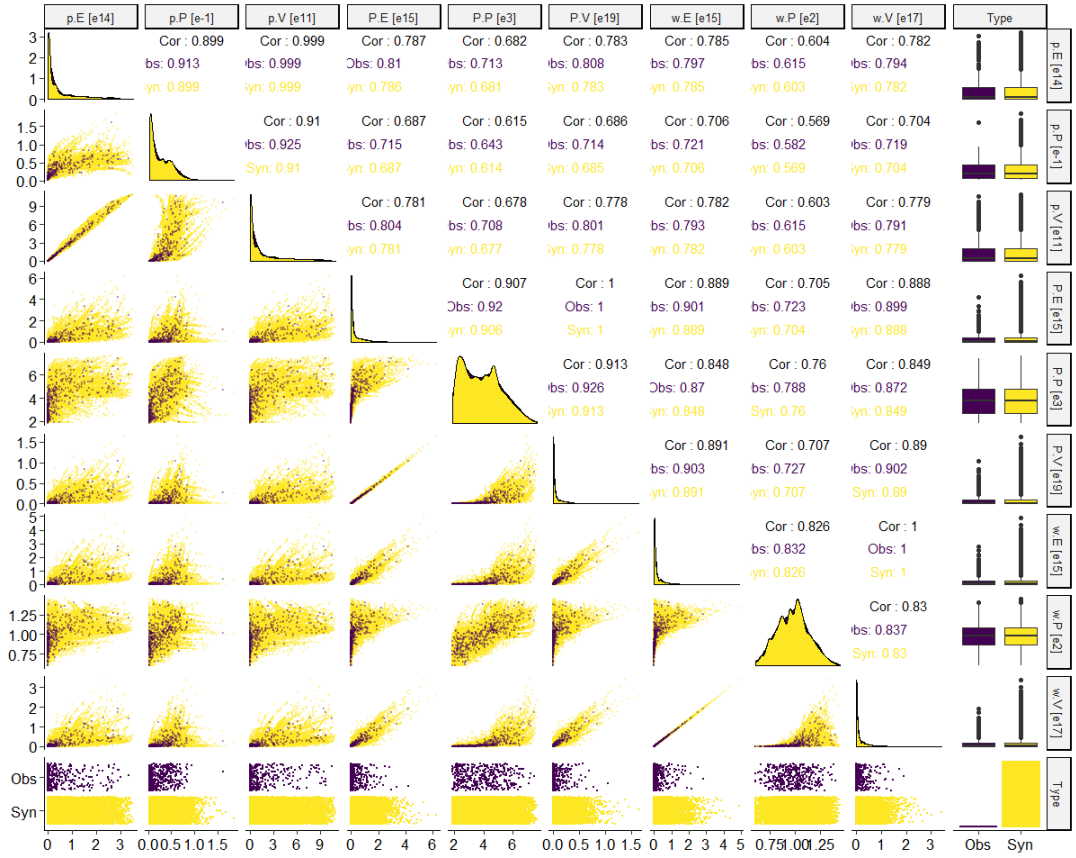


Figure 5.12: Pair-wise plots of the observed descriptor matrix and the synthetic descriptor matrix of class 6. Purple (dark) is the observed and yellow (light) is the simulated. In the diagonal, distributions of observed and synthetic descriptors are compared using box-plots. Below the diagonal, pair-wise scatter plots are displayed. Above the diagonal, pair-wise correlations are displayed. The labels are formatted as $A.B[C]$, where A is the variable, B is the descriptor type and C is the magnitude. The three variables are p [mm/h] the precipitation rate, P [hPa] the low pressure anomaly and w [m/s] the windspeed. The three descriptor types are E the spatio-temporal extent of the event, P the peak value of the event and V the sum of the event values over the entire extent and duration of the event.

identification is a required step to be able to apply event-based, probabilistic analysis, and therefore cannot be avoided.

For future research, there are several recommendations. First, from the dynamic spatio-temporal idea that events are physical phenomena that travel, pathways could be incorporated in the event identification procedure. For example, a single event may have to be split into two events if subclusters of the event can be distinguished that (at some point in time) move into or arrive from distinctly different directions. Second, the event identification method should be compatible with the reconstruction method such that any reconstructed events should be identified in its exact extent. The current reconstruction algorithm may cause parts of the events to break off when the synthetic event extent is reduced.

5.5.2. Event description

Classification A robust classification method should account for the difference

between precipitation events in their statistical properties as well as their physical characteristics. Statistical cluster analysis should incorporate physical descriptors about the class of events, whereas physical descriptors about event classes could probably be refined based on the results of the statistical classification method. This step is recommended to be executed as an iterative process. Seasonality can be expected to be effectively captured by classification and therefore it can be recommended for future research to consider temporal descriptors such as month of occurrence [47] or season [86].

5.5.3. Multivariate statistics

HT04 In Section (5.3), 196 classes were chosen, each of which had 3 marginals, providing a total of 588 marginal distributions. Follow-up studies may be expected to formulate a larger number of classes and, for each class, to describe events more accurately with a larger number of descriptors (which makes sense when more advanced reconstruction methods are developed that would make use of more descriptors). Therefore, with an expected increase in number of classes considered and an expected expansion of the dimensionality of the matrix in each class, HT04 will be a very practical model and can be recommended to be used in follow-up studies.

Fig. (5.12) displays the multivariate structure of a particular class of compound events that comprised of combinations of three variables. With a total of nine descriptors, additional complexity was found and captured for the compound procedure as compared to the multivariate structure of the single variable analysis, which comprised of only three descriptors; Fig. (5.7). In Section (4.3.3), the same multivariate dependence model was applied to a multivariate matrix with 298 columns, so the increase in dimensionality from three to six (or nine) could easily be handled and allows for further expansions of dimensionality in the future.

5.5.4. Event reconstruction, space-time placement and gap filling

Events were reconstructed using peak P, volume V and extent E. This was done similarly for each class of events. When the synthetic event required a reduction in extent as compared to the observed ($E_{syn} < E_{obs}$), the used procedure for the reduction in spatio-temporal extent (using the method of largest values) led to an issue. A single synthetic event could turn out to comprise multiple non-connected clusters of data; Fig. (5.8)b. So what should become a single synthetic event, could potentially be split into multiple synthetic events, when applying the same event identification procedure as applied to the observed data. Vice versa, when generating synthetic events by growing the extent of the observed and when placing the synthetic events in a synthetic space-time continuum at their original locations, occasionally separate synthetic events were merged. For future research, it is recommended that these issues are taken into account, since ideally the event definition should be consistent for both observed and synthetic events. In conjunction with improved event identification and classification methodology, it is recommended that more advanced (i.e. realistic) reconstruction methods be developed. Potentially, a different selection of descriptors can be used for each class. Examples of descriptors to include for reconstruction could be: (object) orientation, for which princi-

pal component analysis could be used; moments; duration; etcetera.

A parsimonious method was implemented for the space-time placement of synthetic events, by which copies of the observed data set were created in which the observed events were replaced by synthetic events based on their original location of occurrence. In particular, the spatio-temporal extent of events was made smaller or larger Fig. (5.8). This provided an increased spatial smoothness of extremes, as will be discussed in Chapter (6). To address spatio-temporal sparsity, additional randomness could be introduced concerning the 'when and where' that events may occur, There are several recommendations. First, pathways of events could be altered. When considering the moving precipitation fields as 3-dimensional objects (in x,y,t), this corresponds to an altering of the orientation of the objects and an altering of the shape of the objects, where the centre of gravity (in x,y,t) remains the same. Second, the centre of gravity of events could stochastically be moved around. The farther events are moved from their original location of occurrence, the more likely that synthetic events would start to overlap and the more likely that the general atmospheric pattern would be disturbed. Third, as a more rigorous version of the second option, a spatio-temporal point process (STPP) could applied, in which the catalogue of synthetic events could be used to build continuous data starting from an empty synthetic space-time continuum. For example, Cowpertwait et al. [25] apply a flexible STPP in which 'independent superposed point processes are used to allow for different storm types (for example, convective and stratiform).' However, if an STPP is applied to large-scale precipitation events, the challenge arises to place the synthetic events in space and time in such a way that the large-scale atmospheric patterns are well simulated. This can be expected to pose a significant challenge for a single variable approach and an even larger challenge for the compound variables approach.

5.6. Conclusions

For the development of the generator, the full event-based framework, as designed in Section (3.3) was applied. Simple methodology of the type DSTPA (Appendix (A.1.6)) was provided for every step in the framework. In general, simplistic choices were taken for two reasons, first, to keep the general flow of the analysis manageable; and, second, to keep the computational burden manageable.

The limitations that applied in Chapter (4), discussed in Section (4.5), which caused a reduction to the framework, were overcome with the dynamic spatio-temporal approach. In particular, first, with the dynamic approach, no gaps were created in the descriptor matrix (as in Section (4.5.2)), which implies that dynamic event identification and description are fully compatible with multivariate, statistical methodology. Second, as a result of working with a matrix of meaningful synthetic event descriptors, synthetic events were successfully reconstructed.

Extensions of methodology were provided to be able to apply the DSTPA, which was developed for a single variable, to multiple variables, by addressing compound events. In addition, a number of potential improvements was provided in Section (5.5). In particular, an important improvement would be the implementation of a process to induce variation in the spatio-temporal occurrence of the events. However, realistic atmospheric patterns should be obtained, which will be a significant challenge. Overall, it can be concluded that, with these potential improvements, the method appears to be promising.

This chapter addressed the challenge of the development of a large-scale weather generator that is capable of producing long time series of synthetic, continuous data containing spatio-temporally coherent events. In Chapter (6), the dynamic weather generator will be tested extensively, which requires inspection of the massive amounts of data generated, and a final, large-scale, flood hazard time series will be generated.

6

A generic framework to explore the performance of generators

6.1. Introduction

6.1.1. General introduction of generators

The starting point of a common probabilistic flood risk analysis is the determination of the flood hazard, introduced in Section (1.1.2). The flood hazard may be considered a collection on all hypothetical scenarios that may cause flooding, with associated probabilities or frequencies of occurrence. The collection is often developed based on observed scenarios that occurred in the past. The idea of considering all hypothetical scenarios and their probabilities can be implemented by the modelling of a full stochastic process. The collection of scenarios is obtained by the set up of a specific type of statistical model, referred to as a generator. A generator can be used to approximate the full stochastic process of a system in a particular (discretised) time interval, by generating much synthetic data that can be used to drive a chain of input-output models, as discussed in Section (1.1.3). Generators may be event-based, representing the scenarios to consider by (hazardous) events. Event-based generators typically make use of event descriptors, which are first used to summarise event characteristics and later, after being subjected to statistical models, to reconstruct synthetic events, as described in Section (3.3).

In Chapter (5), a generator was developed based on large-scale dynamic events, both for precipitation as a single source and for the multi-source combination of precipitation, (low) pressure and wind speed. In Section (5.5), a large number of recommendations was provided to improve the quality of the generator. However, before further improvements to the generator are to be implemented, it is important to explore the performance of the current generator, by examining thoroughly the data that it generates.

6.1.2. A subset of generator objectives

Generators are developed with a particular purpose in mind. A number of generator objectives is formulated, which is typically a specific subset of the generic set of generator objectives described in Section (3.2). This subset is catered to the particular purpose for which the generator will be used. Normally, generator methodology is used that focusses on the optimisation of the performance for this particular subset of generator objectives. Examples can be found in Table (2.4), which focusses in particular on the retention of statistical patterns (GO2). It is evident that, in (almost) all studies, methods are applied that optimise the retention of a subset of the statistical patterns formulated in this table. And note that the patterns addressed in this table are non-exhaustive. We appear to be far from the creation of the perfect generator algorithm that, while generating maximum variability (GO1), manages to capture and retain all relevant statistical patterns (GO2) and in doing so creates realistic events, i.e. respects all physical constraints (GO3). Thus, our priority is to focus the methodology on a subset of generator objectives.

In compliance with a focussed methodology, it is interesting - and also convenient - to perform a focussed performance check. A focussed performance check refers to a check on how a generator performs in those aspects that its methodology was designed for. A focussed performance check is a minimum requirement, that should be easy to pass in most cases. The literature typically stays on this level of check. An example

can be found in Chapter (4), published in [28], where a static spatio-temporal generator was developed with a focus on marginal (local) distributions combined with (spatial) dependence, for which the performance check focussed indeed on the two aspects that were optimised in the particular methodology of this generator - the extreme values and the (spatial) dependence structure. Similar examples can be found in [52, 68, 77].

Performance indicators can be defined for most generator objectives. It should be interesting to explore the performance of generators for those generator objectives that the methodology was not optimised for. In addition to how well a generator is performing in certain aspects, this would provide insight in what the shortcomings of a generator are.

6.1.3. Generator settings and sensitivity

Each performance indicator provides information on the performance of a generator in a particular aspect. In order to objectively compare performance across different generators, separate performance indicators could function as common standards. However, it is not the intention of this chapter to summarise the different performance indicators into a single grade for a particular generator. To summarise the different performance indicators of a particular generator into a single grade would require the appreciation of different aspects compared to one another, or 'weighing', which would be a subjective procedure.

Generators require configuration using a (potentially large) number of generator settings. For example, in Section (5.3), a number of generator settings was described. First, settings may be quite general: the spatial domain and temporal length of the observed data; and the amount of synthetic data to generate. Second, settings may be related to the specific generator methodology: method of event identification, method of classification and which event descriptors to use, method of reconstruction and which event descriptors to use. Third, settings may be numerical: for event identification, a threshold; for event description, a number of events per class; and for multivariate statistics, GPD thresholds, extremal dependence thresholds and simulation thresholds.

Changes to the generator settings may significantly alter the synthetic events. To provide some level of guidance on the compromises between different performance indicators for a particular generator, a sensitivity analysis may be carried out. It has been observed through this study and others, that tuning the generator to perform well in one aspect may lower the performance in other aspects. With a complex generator, users typically face multiple trade-offs in the settings. For example, [52] showed how changing the threshold of a GPD fit is a trade off between model bias and parameter uncertainty.

It is important to keep in mind that, to approximate the full stochastic process, a synthetic data set has to be generated that may comprise a large quantity of data. This applies in particular for generators of gridded data sets with high spatial and temporal resolutions and/or when the objective of the study is to look at rare events with small probabilities. For a sensitivity analysis, many synthetic data sets have to generated, which implies massive amounts of data have to be generated and analysed. Therefore, it is challenging to apply sensitivity analysis to generators of large data

quantities and the feasibility of sensitivity testing has to be investigated.

For sensitivity analysis, various methods have been established. In particular, a disctinction can be made between local and global analysis [74]. In this chapter, global sensitivity analysis will be applied. A grid of generator settings will be used for the sensitivity analysis, which is listed in Appendix (A.4.3). The main advantage of this method is that all parts of the generator setting space are investigated in equal measure. The main disadvantage is that, because of the limitations of computational power and storage, a limited coverage of the generator setting space will be obtained.

6.1.4. Chapter structure

The objective in this chapter is to establish a generic framework to help explore the performance of a generator. First, the methodology to apply the framework will be defined. Second, as a case study, the framework will be applied to the dynamic generator developed in Chapter (5). In particular, the performance of this generator will be explored for generator objectives that its methodology was not optimised for. In addition, the feasibility will be examined of applying a sensitivity test to a generator of large-scale, gridded, synthetic of high spatial and temporal resolution.

6.2. Methodology

6.2.1. Generator objectives

There are three main generator objectives:

1. introduce the unobserved (GO_1),
2. reproduce patterns (GO_2) and
3. respect physical restraints (GO_3).

These generator objectives were described in more detail in Section (3.2).

6.2.2. The role of performance indicators

To be able to evaluate how well the generator objectives are fulfilled, a number of performance indicators has to be set up. For the different generator objectives, different performance indicators are required.

First, for the introduction of the unobserved (GO_1), performance indicators are required that provide insight in the level of variability introduced. The goal of these performance indicators is to check how much the newly created synthetic data differs from the observed, in all required aspects, e.g. spatial event coverage of extremes or spatial and temporal smoothness of the stochastic process. Variability may typically be checked by use of distributions of event descriptors on a per-location basis, where, for flooding, the emphasis lies with the upper tail of distributions. In addition, since variability should be introduced not only in time but also in space, spatial coverage of events may be checked.

Second, to investigate if statistical patterns are retained (GO_2), performance indicators are required that provide insight in how well the observed data agrees with the synthetic data. Distributions allow to check if the event descriptors in the observed data are a likely subset of the (much larger) population of synthetic event descriptors. However, to evaluate the pattern of event occurances, additional performance indicators are required. For temporal patterns, it matters in which sequence events occur. For spatial patterns, it matters what happens at different locations in a given time window. Although time and space should not necessarily be considered separately, temporal and spatial patterns

may be relatively easily captured separately and may be described using correlations.

Third, to investigate if physical restraints are respected (GO3), performance indicators are required that demonstrate how realistic the events are that are used and produced in the generator. Reality of events is a generic concept, but can not easily be described in a generic fashion. These performance indicators will have to work with the specific type of data that is produced by each particular generator.

6.2.3. Sensitivity analysis

A sensitivity analysis may be carried out to explore the performance of a generator, for two reasons in particular. First, a sensitivity analysis may provide insight in the stability of the generator, by looking at how much the results vary when settings are adjusted. Second, a sensitivity analysis may provide guidance on how to choose parameters, by the testing of the quality of the generated data, as measured for the different aspects by the performance indicators.

For a risk analysis that considers the full stochastic process, all possible events are supposed to be covered. Therefore, a generator is supposed to be used to generate large quantities of data. With a sensitivity analysis, the generator is run many times with different generator settings and inputs, which means that very large quantities of data will be produced. This implies that compromises have to be made for feasibility and that the completeness of a generator sensitivity analysis will largely be limited by the available computational power and storage capacity.

6.3. Performance indicators applied to a dynamic precipitation generator

6.3.1. Overview

In this particular case study, the performance is explored of the (single-source) precipitation generator that was developed in Section (5.3). In the following sections, performance indicators will be defined and discussed in detail. The BSL2 domain is used for generation and the UK2 domain is used to check the results. For the correlation tests, a subset of locations in the UK2 domain will be used, displayed in Fig. (6.8), for the reasons of dimensionality and interpretability.

Performance indicators were organised into three aspects.

1. *Local distributions.* Local time series were cast into local distributions. A 'local distribution' is the marginal distribution of a joint probability distribution, where each margin relates to data at a specific location. For every location in the target domain, the means, event coverage, General Pareto distribution parameters and return level estimates (RLE) were investigated. Since these statistical indicators were captured for each grid-cell in a grid, spatial patterns emerged. The full stochastic process is supposed be spatially smooth, i.e. there should not be sudden changes in the parameter values or in the RLE when moving across the spatial domain, whereas the limited sample of observed data may be spatially rough. Therefore, spatial smoothness or roughness indicators were used. So, these indicators

should show that the statistics of the synthetic data were smoother, than the observed data.

A good generator needs to have similar local distributions of the entire populations between the observed and synthetic data. Due to sparsity, the generator should be allowed to produce slightly different patterns of local extremes. It should produce smooth local distributions, since that would mean that sparsity is addressed conform GO1. To summarise the performance indicators of the local distributions and of the spatial smoothness, the spatial means of the produced maps were used. These summary performance indicators were also used in the sensitivity analysis.

2. *Temporal and spatial correlations.* For a subset of locations, spatial correlation and temporal correlations were investigated. This was done across different parts of the data in different directions. The dimensions of time and space were treated as separate directions.

   Good performance was considered to be obtained when these performance indicators were similar for the synthetic data and the observed data.

3. *Spatio-temporal integrity.* The spatio-temporal aspect, where the spatio-temporal nature of events was investigated. One aspect of the spatio-temporal nature, the use of complete events, was investigated using the region of influence. The idea was that, if the region of influence shows much observed event data on the edges of the domain, this indicates that events have been cut off, i.e. were incomplete. Incomplete observed events were considered to be disturbances in classification and in distributions and also unlikely to allow realistic synthetic events to be constructed.

Good performance was considered to be obtained when these performance indicators were as low as possible for the observed data. For these indicators, this implies: a roughness as low as possible; the proportion of incomplete observed events as low as possible; and the similarity in spatial and temporal correlations between the observed and the synthetc as high as possible.

The performance indicators are elaborated in Sections (6.3.2) to (6.3.4). In each of these sections, first, maps are provided of the aspect under investigation and, second, the summarising performance indicators are provided as a function of the varied generator settings.

Note that these indicators perform checks on a per-location basis, whereas the methodology of the investigated generator is dynamic by nature, i.e. is based on distributions of descriptors of moving precipitation events. This means that the performance checks do not have the same focus as the optimisation in the generator methodology, as introduced in Section (6.1.2). So these performance checks go beyond the minimum requirement of the focussed checks.

6.3.2. Aspect 1: patterns of local distributions

Means. *To check GO2 for the bulk of the populations.*

Temporal means were used to check the bulk of the populations of precipita-

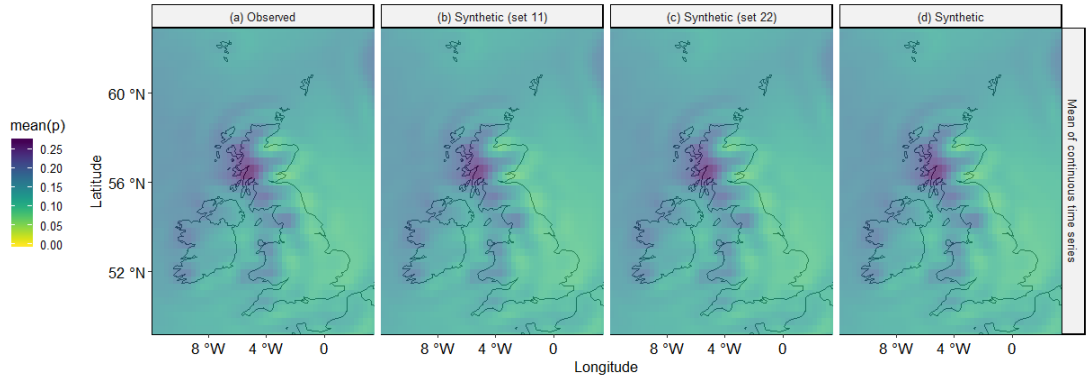


Figure 6.1: Means of precipitation rate p [mm/h] of the entire local populations (per grid-cell) for (a) the observed data set (20 years), (b,c) two generated synthetic data sets (20 years each) and (d) the entire series of all generated synthetic sets (1.000 years).

tions rates per location, for both the observed and the synthetic data. It is an important check to do, as the bulk populations of the synthetic data set should have temporal means that are similar to those of the observed data set. The reason is that sparsity, see Section (3.1.1), is mainly an issue for (unlikely) events, whereas sparsity should not be expected to have a significant effect on the bulk populations. Other statistics can be drawn and compared between the observed and the synthetic in a similar manner, e.g. variance, higher-order moments, quantiles, etc.

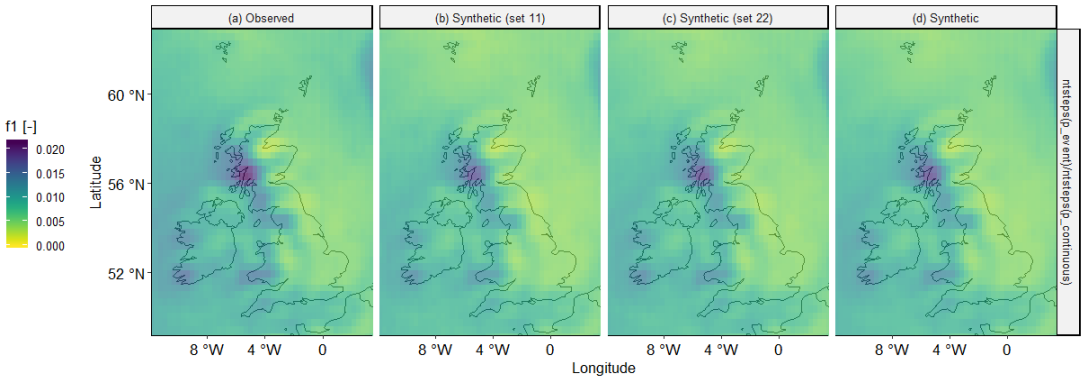
Fig. (6.1) displays the temporal means of precipitation rates for the BI, on a per-location basis. It can be observed that the temporal means were very similar for the observed data set, two synthetic data sets and the combined synthetic data sets. This test is always relevant but, in particular, will be more relevant with increased complexity of the generator, which will probably give less similar results.

Event coverage *To check GO1 and GO2 for the observed and synthetic event catalogues.*

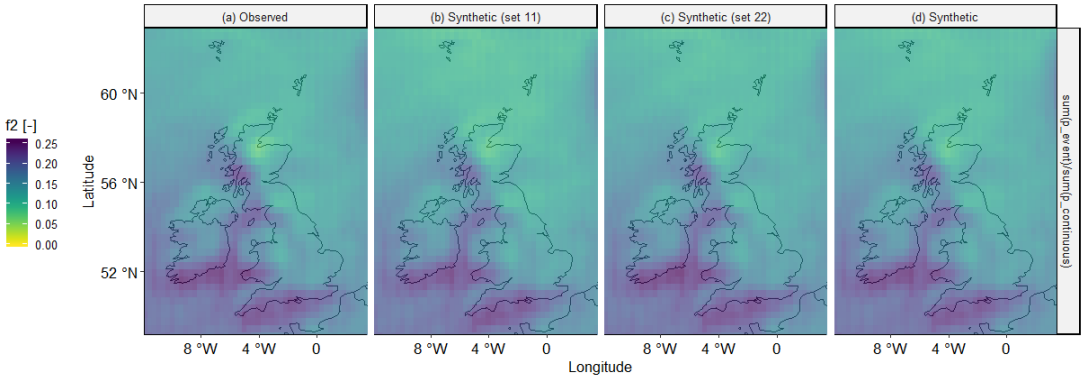
Although similar temporal means may be an indicator of good performance of the generator, the event coverage should also be taken into account. Fig. (6.2) displays the event coverage in the BI, considering the observed and the synthetic event catalogues.

For each location, Fig. (6.2.1) shows f_1, which is the fraction of the number of time steps in the event catalogue divided by the number of time steps in the continuous series. It generally shows the same patterns for the observed and the synthetic data, but it can be observed that $f1$ was slightly higher in the observed event catalogue than in the synthetic event catalogues. The event reconstruction algorithm proposed in Section (5.3.4) did not always succeed, which led to a lower number of events. Moreover, the event reconstruction algorithm was iterating towards the correct values of descriptors. For computational reasons, there was a tolerance on this iterative process, such that the reconstruction stops when the tolerance level was reached. When this happened, per event a slightly smaller spatio-temporal extent was obtained than was prescribed by the synthetic spatio-temporal extent descriptor.

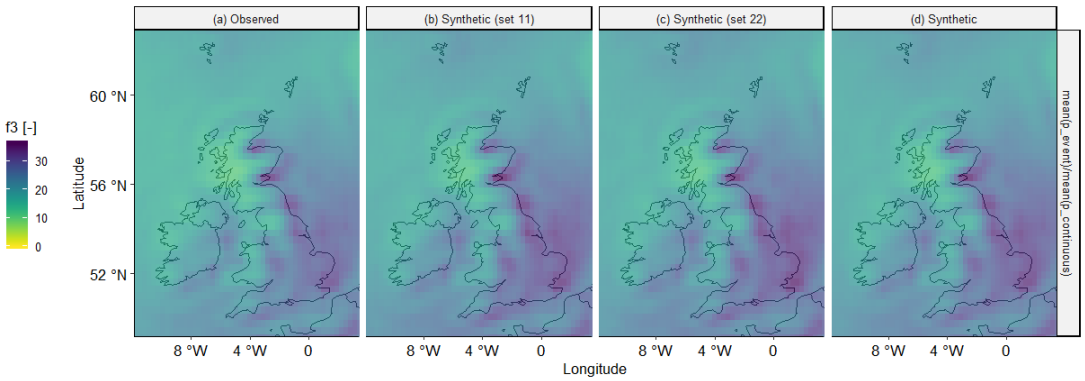
For each location, Fig. (6.2.2) shows f_2, which is the fraction of the total sum



(6.2.1) $f1$ [-] is the fraction of the number of time steps in the event catalogue divided by the number of time steps in the continuous series.



(6.2.2) $f2$ [-] is the fraction of the total sum of precipitation in the event catalogue divided by the total sum of precipitation in the continuous series.



(6.2.3) $f3$ [-] is the fraction of the mean of precipitation in the event catalogue divided by the mean of precipitation in the continuous series.

Figure 6.2: Per location (cell), fractions of descriptors are displayed for (a) the observed data set (20 years), (b,c) two generated synthetic data sets (20 years each) and (d) the entire series of all generated synthetic sets (1.000 years).

of precipitation in the event catalogue divided by the total sum of precipitation in the continuous series. Again, it generally shows the same patterns for the observed and the synthetic data, but it can be observed that $f2$ was slightly higher in the observed event catalogue than in the synthetic event catalogues. The difference was small and harder to spot than the difference in $f1$, which means that the observed event catalogue held a larger fraction of small event values than the synthetic event catalogue. About the same total sum of rainfall was put into the synthetic event catalogue as was in the observed event catalogue, which is an important indicator of performance.

For each location, Fig. (6.2.3) shows f_3, which is the fraction of the mean of precipitation in the event catalogue divided by the mean of precipitation in the continuous series. It generally shows the same patterns for the observed and the synthetic data, but it can be observed that $f3$ was slightly lower in the observed event catalogue than in the synthetic event catalogues. This confirmed that the synthetic event catalogue held a smaller fraction of small event values compared to the observed event catalogue.

Local (marginal) extreme value analysis. *To check GO1 for spatial sparsity of extremes, to check GO2 for the tail ends of the populations and to check GO2 for spatial coherence of extremes.*

First, in Figs. (6.3) to (6.6), the extremes in individual synthetic sets of 32 years each are compared to the extremes in the 32 years of observed data. The purpose is to check if the patterns in the extremes are retained per synthetic set, i.e. the patterns of the extremes in the observed data should agree reasonable well with the synthetic.

Figs. (6.3) and (6.4) show that the GPD fits of the synthetic sets and the observed data set are similar but not the same. Here it should be kept in mind that the scale and shape parameter can compensate for each other for lower extremes. Fig. (6.5) shows very similar patterns for the lower RLE, with a return period of 10 years, in the synthetic sets and the observed. Fig. (6.6) shows some similarity in the patterns for the higher RLE, with a return period of 1.000 years, but also some more distinct differences. This can be attributed to the higher degree of sparsity of higher RLE, which should indeed give the expectation of more variability over the different sets.

Second, in Fig. (6.7), the extremes in the entire collection of synthetic data, which comprises 1.000 years of data as an approximation of the full distribution, are compared to the extremes in the distributions fitted directly to the observed data per grid-cell. The purpose is to compare spatial coherence of extremes for the dynamic generator, with the direct GPD fits to the observed data per grid-cell as a baseline model.

To assess the smoothening of the extreme values in the synthetic data produced by statistical model $B1$ (the dynamic generator described in Chapter (5)), a statistical model $A1$ was used as a baseline reference for comparison. Statistical model $A1$ comprises local GPDs fitted to time series per pixel using Peaks-over-Threshold, i.e. the peaks of local events above a high threshold. The following procedure was applied.

1. Statistical models $A1$ and $B1$ were both fitted to the 32 years of observed

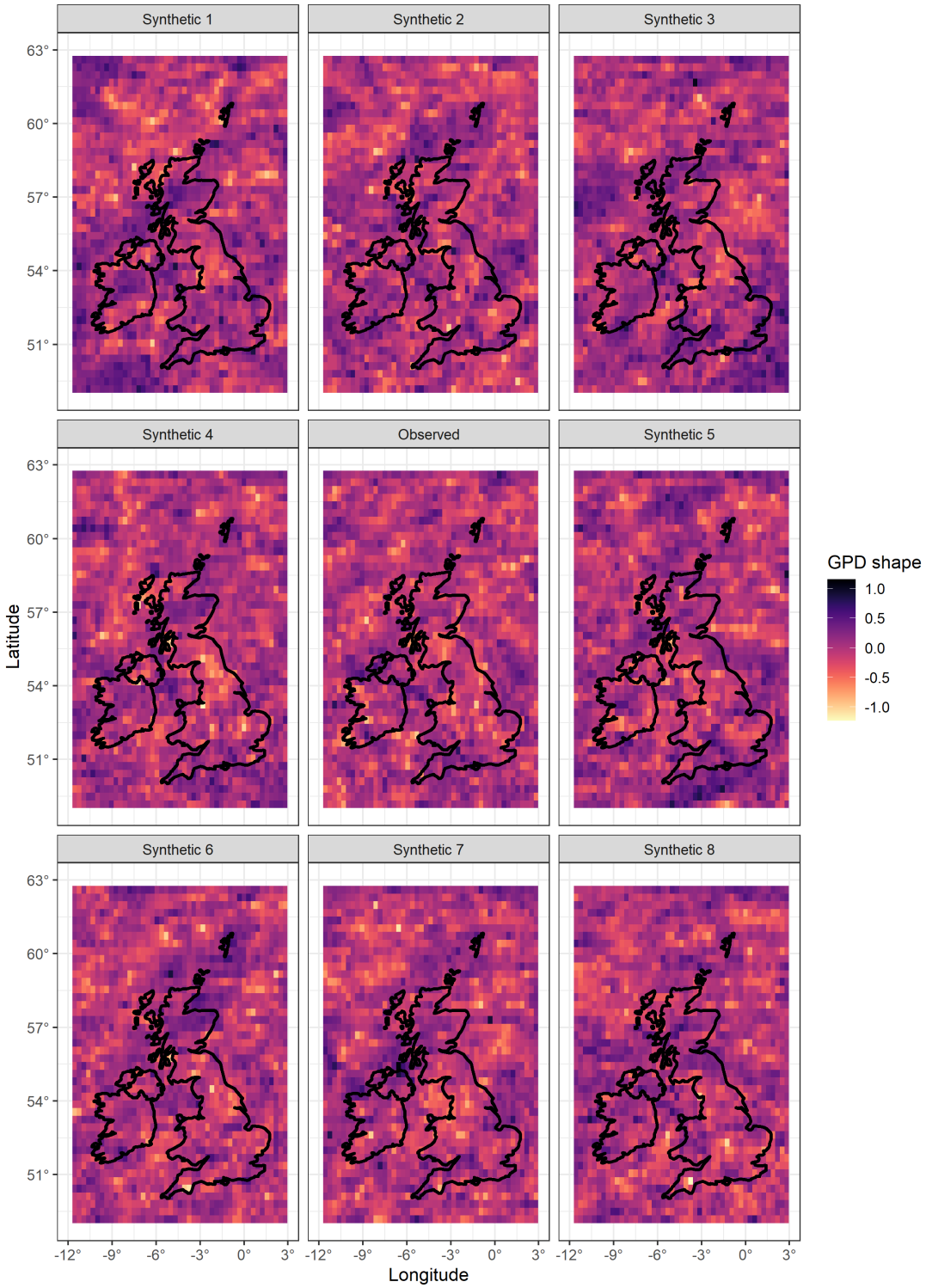


Figure 6.3: Shape [-] parameters of the GPDs fitted per grid-cell. The centre figure displays the fits directly to the 32 years of observed data. The surrounding figures display the fits to synthetic sets of 32 years each.

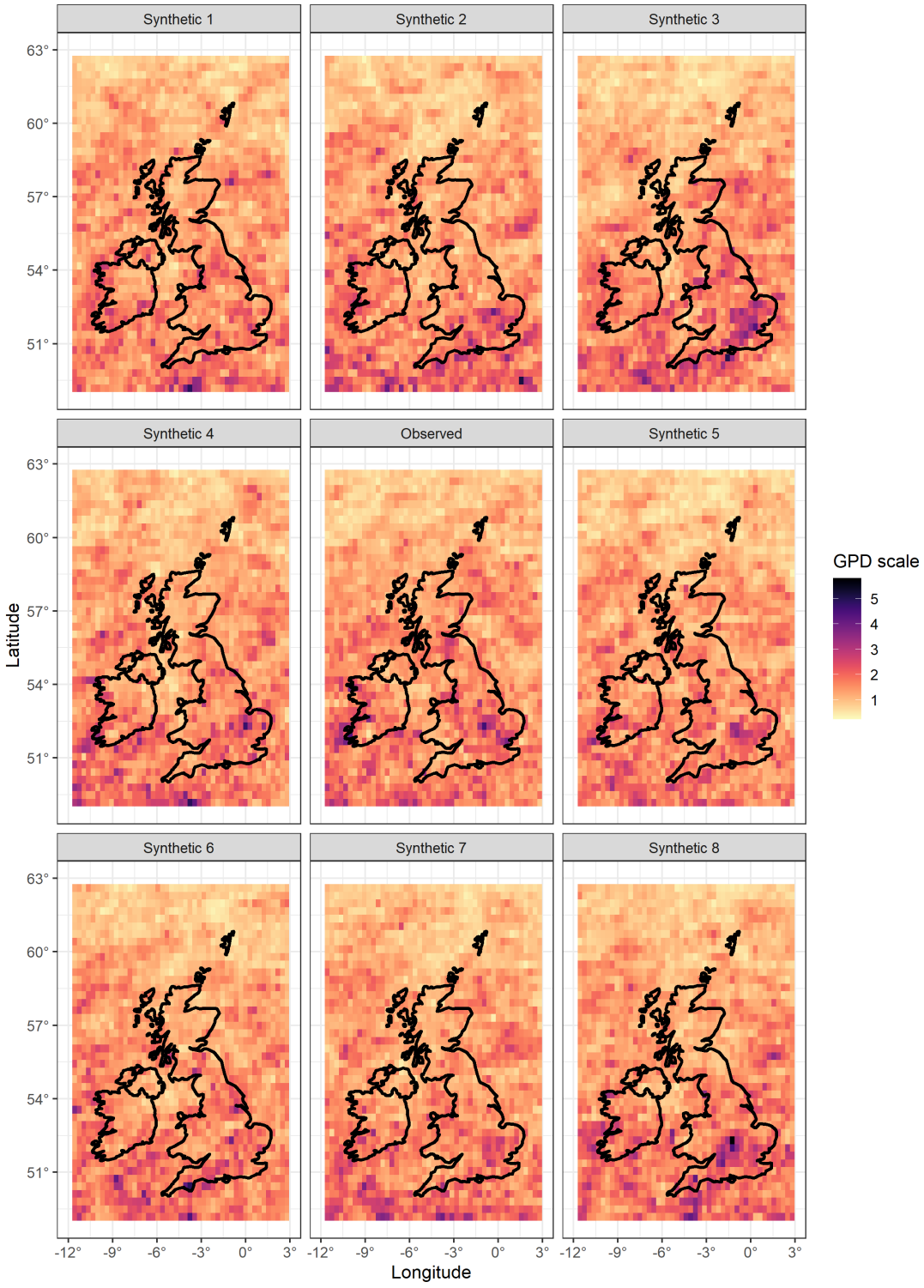


Figure 6.4: Scale [-] parameters of the GPDs fitted per grid-cell. The centre figure displays the fits directly to the 32 years of observed data. The surrounding figures display the fits to synthetic sets of 32 years each.

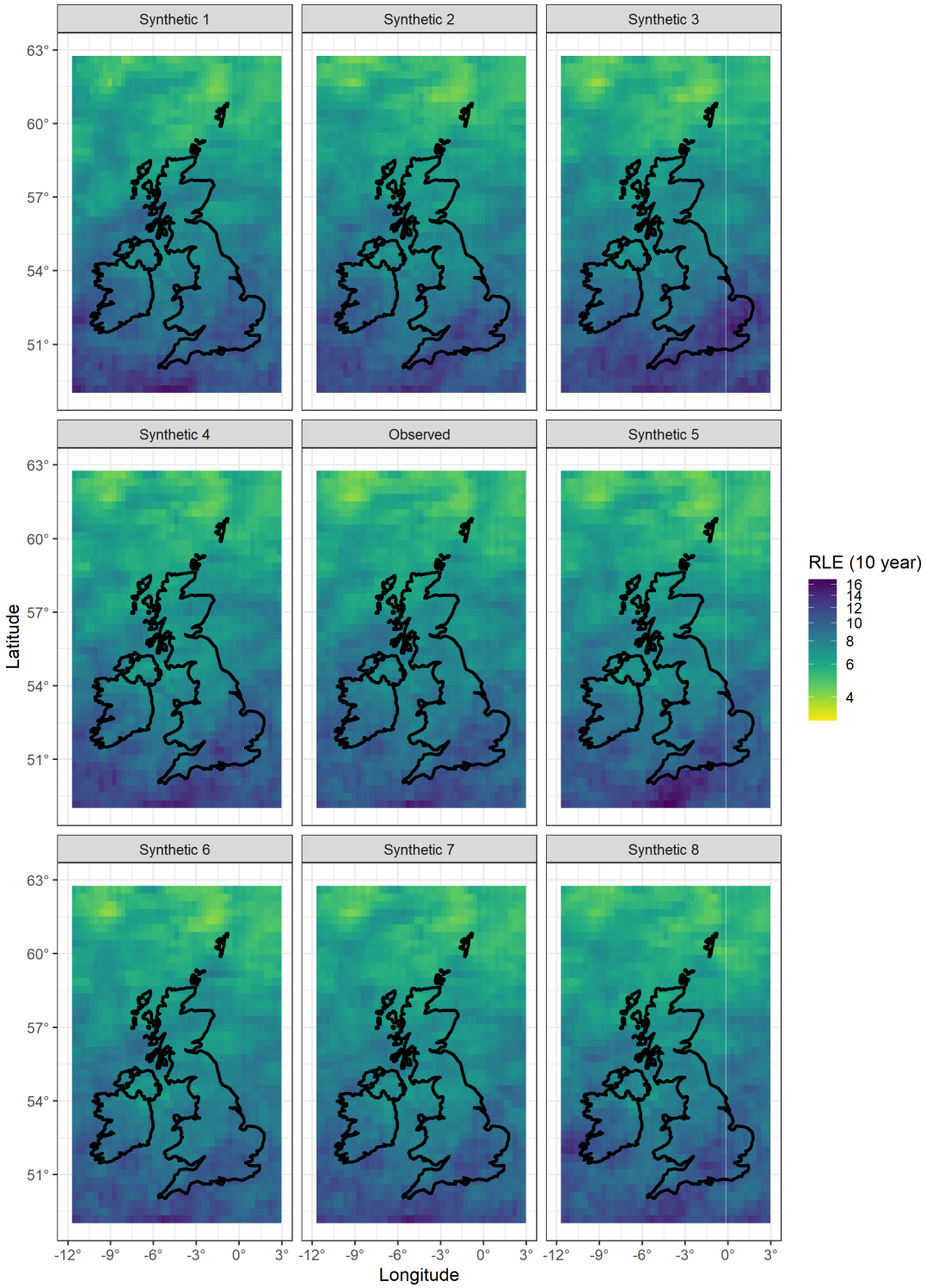


Figure 6.5: RLE of precipitation rate p [mm/h] corresponding to a 10 years return period. The centre figure displays the fits directly to the 32 years of observed data. The surrounding figures display the fits to synthetic sets of 32 years each.

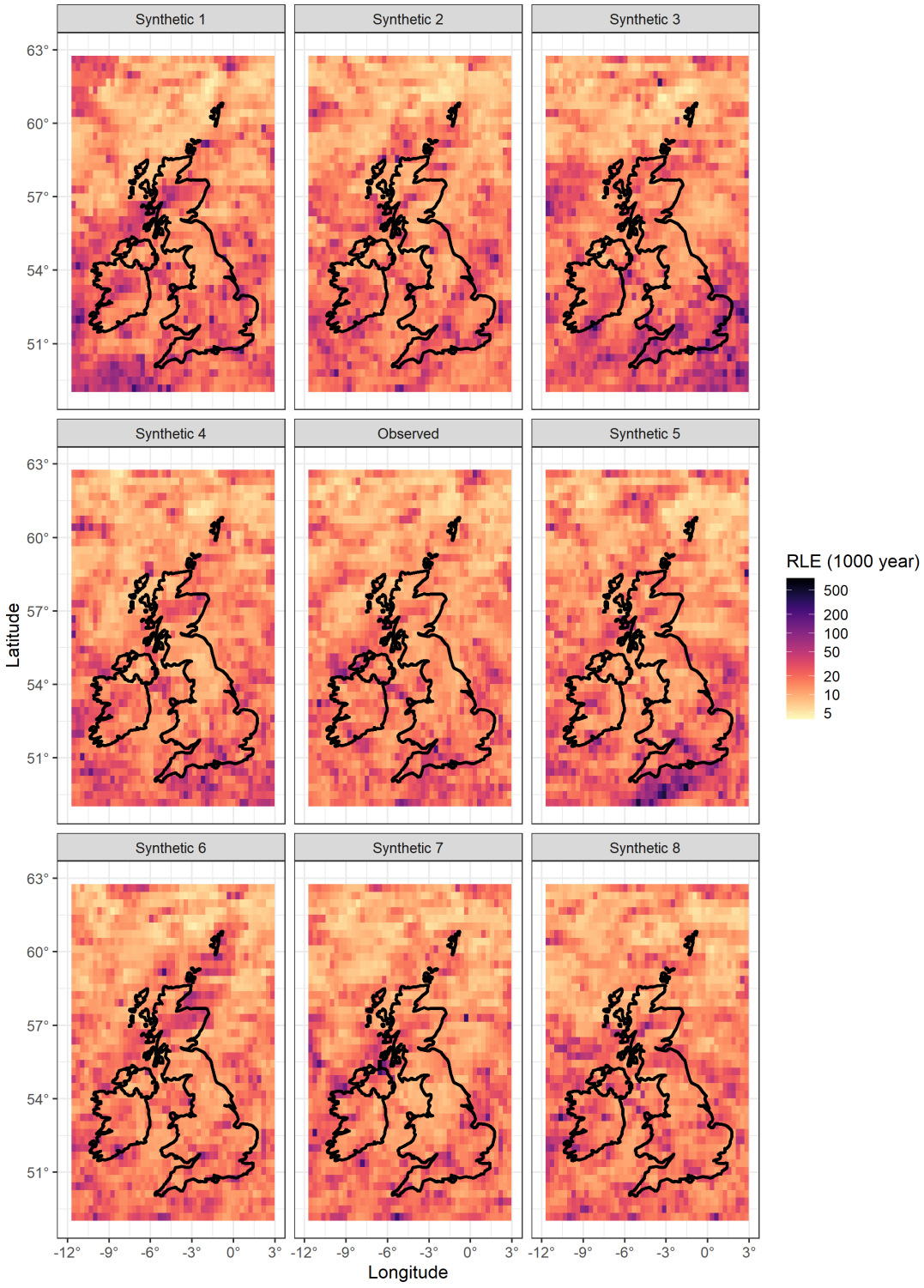
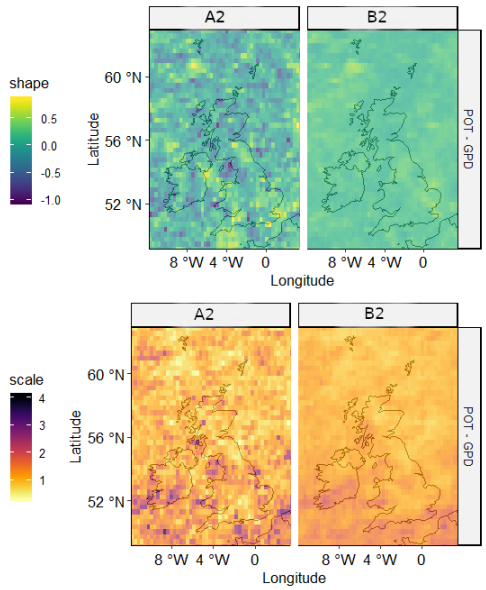
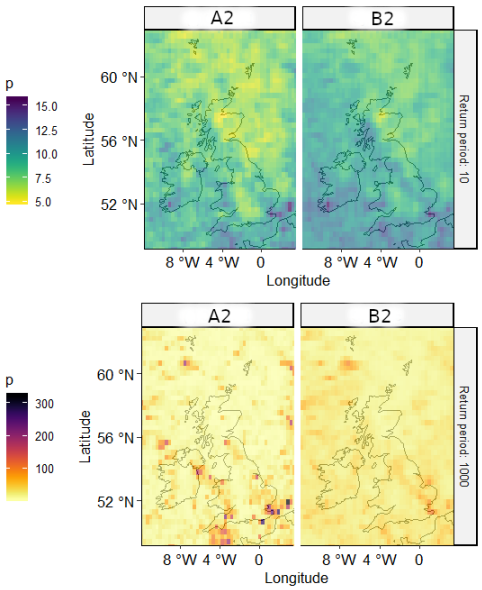


Figure 6.6: RLE of precipitation rate p [mm/h] corresponding to a 1.000 years return period. The centre figure displays the fits directly to the 32 years of observed data. The surrounding figures display the fits to synthetic sets of 32 years each.



(6.7.1) Shape [-] and scale [-] parameters of the GPDs fitted per grid-cell.



(6.7.2) RLE of precipitation rate p [mm/h] for two return periods.

(6.7.3) Spatial roughness ξ [-] of the GPDs.

(6.7.4) Spatial roughness ξ [-] of the RLE.

Figure 6.7: Above, local GPD parameters and RLE are displayed for statistical models $A2$ and $B2$. Below, spatial roughness is investigated for the GPD parameters and RLE.

data.

2. 1.000 years of synthetic time series were sampled and reconstructed from the statistical models $A1$ and $B1$.

3. Statistical model $A2$ (local GPDs per pixel using Peaks-over-Threshold) was fitted per pixel to the 1.000 years of synthetic time series sampled from statistical model $A1$, using the same GPD thresholds as used to fit model $A1$.

4. Statistical model $B2$ (local GPDs per pixel using Peaks-over-Threshold) was fitted per pixel to the 1.000 years of synthetic time series sampled from statistical model $B1$, using the same GPD thresholds as used to fit model $A1$.

Note that, when using a sample of a synthetic 1.000 years, the GPD parameters of statistical model $A2$ are close to being identical to the GPD parameters of statistical model $A1$. As the synthetic sample goes towards infinitely long, the GPD parameters of statistical model $A2$ become identical to the GPD parameters of statistical model $A1$.

Fig. (6.7.1) shows the resulting GPD parameters for each grid-cell within the spatial domain of the British Isles (BI). It is obvious that the GPD parameters of statistical model $A2$ were much rougher than those of statistical model $B2$. For the scale parameter, the GPD parameters for statistical models $A2$ and $B2$ on average appear to be similar. For the shape parameter, the GPD parameters for statistical model $A2$ were slightly lower on average than those for statistical model $B2$. This can be observed more clearly in the sensitivity analysis. The estimators of the shape and scale parameters are negatively correlated, so the RLE analysis provides an additional angle for observing any potential constrast between statistical models $A2$ and $B2$.

Fig. (6.7.2) shows the RLE in the BI. For the return period of 10 years the RLE of $A2$ show a similar pattern as the RLE of $B2$, but are lower on average. For the higher return period of 1.000 years, the RLE of $A2$ are generally lower on average than the RLE of $B2$, but contain very high local values. This becomes more obvious in the sensitivity analysis. This is potentially caused by the sparsity of extreme observations, i.e. a few observed extreme events may have largely influenced (upwards) the extrapolation of the tail distribution for particular grid-cells, whereas nearby grid-cells with similar physical conditions were not affected. Statistical model $A1$ works per grid-cell, and does not have the capability to take into account what happened around that grid-cell.

Fig. (6.7.3) shows the roughness of the GPD parameters in the BI. Here, the roughness (ξ) is defined as the difference between the maximum and the minimum value of a grid-cell and its 8 surrounding grid-cells. The GPD parameters fitted to statistical model $B2$ were spatially much smoother, i.e. more spatially coherent, than the GPD parameters of statistical model $A2$. This can be explained by a lower degree of event sparsity in the synthetic time series generated by statistical model $B1$, as compared to the synthetic time series generated by statistical model $A1$. Statistical model $B1$ does not work per grid-cell and has the capability to take into account what happened around that grid-cell. This capability is referred to as 'dynamic expansion of information' and is discussed in length in Section (6.5.1).

Fig. (6.7.4) shows the roughness of the RLE in the BI. The combinations of rough GPD shape and scale parameters in the RLE of $A2$ delivered a relatively smooth image of RLE for low return periods. The roughness in the scale and shape parameter respectively have canceled each other out to some extent. These RLE are closely linked with the high quantiles observed in the input data, which should be relatively smooth. However, the roughness of the GPD parameters becomes apparent in the RLE for high return periods, where the RLE are based on not only the observed high quantiles, but also on the extrapolation along the tail of the GPD distribution, where the shape parameter takes a quite dominant role. The smooth combinations of shape and scale parameters in the RLE of $B2$ delivered a relatively smooth image of RLE for both low and high return periods. The RLE were generally slightly higher in the synthetic data, but with fewer outstanding RLE spikes across the domain. This is most likely due to the capability of 'dynamic expansion of information' of statistical model $B1$.

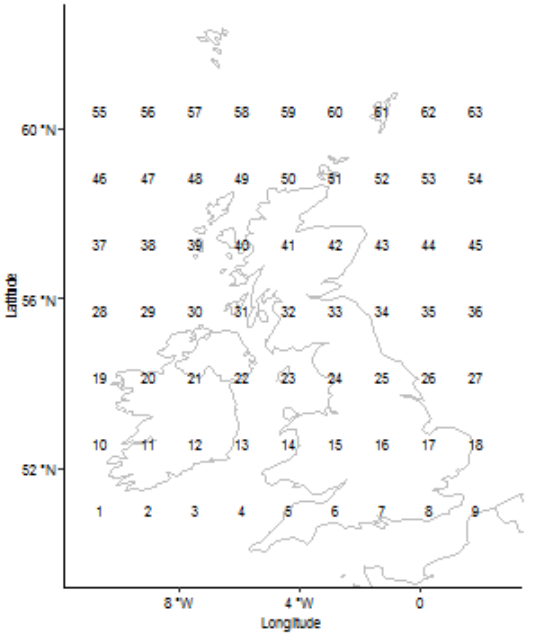


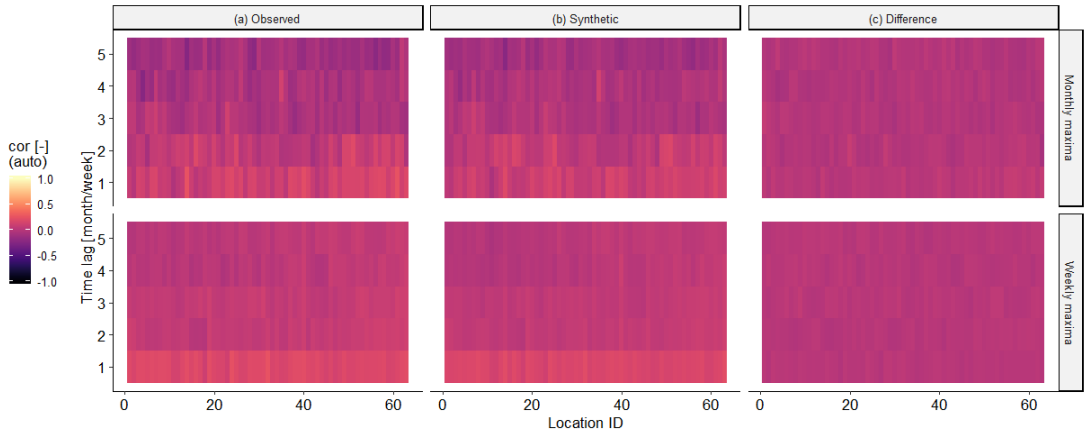
Figure 6.8: A subset of locations was chosen to be able to investigate particular statistical patterns. These locations represent individual grid-cells.

6.3.3. Aspect 2: patterns of occurrence

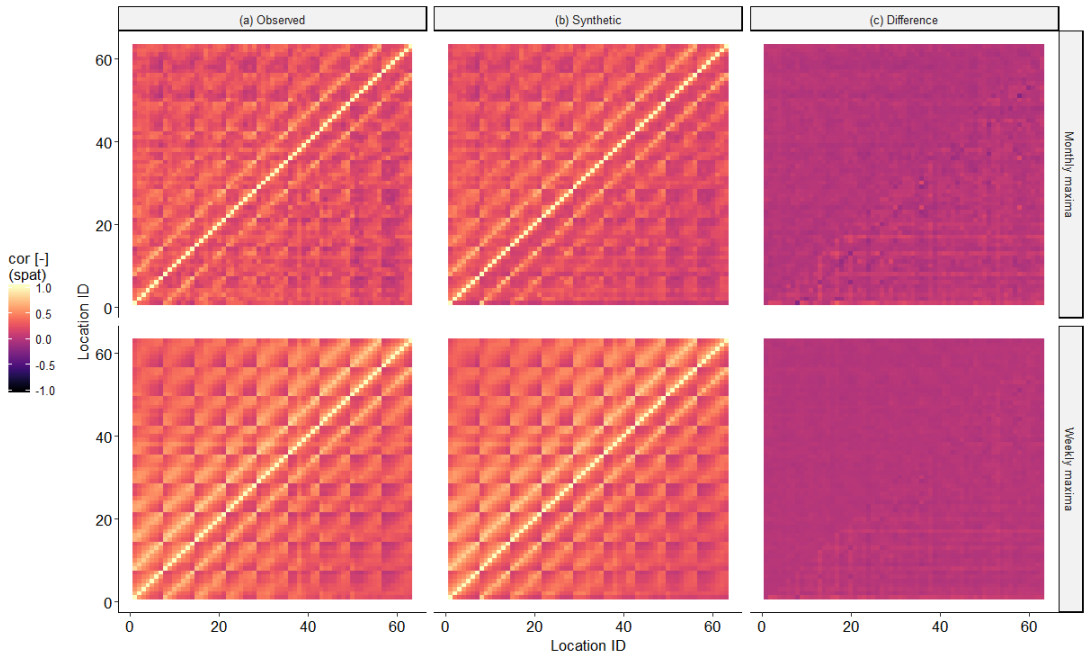
Temporal and spatial correlations *To check GO2 for simple patterns of occurrence.*

To check if statistical patterns were retained, correlations were captured for both the observed and the synthetic data, on a subset of locations; Fig. (6.8). Autocorrelations were calculated to serve as an indicator for temporal dependence per location and spatial correlations were calculated to serve as an indicator for spatial dependence, i.e. dependence between different locations.

Fig. (6.9) displays correlations found in the observed and the synthetic data, including the difference between the two per location or per pair of locations. Fig. (6.9.1) shows that autocorrelations show comparable patterns in the observed and the synthetic data set. The autocorrelations are slightly higher in the observed series than in the synthetic. This will become more clear in the sensitivity analysis. This means that this pattern is not perfectly captured with the current generator. Fig. (6.9.2) shows that spatial correlations show very comparable patterns in the observed and the synthetic data set. The difference is minimal, but again, in the observed data set the spatial correlations are slightly higher than in the synthetic data set. This will become more clear in the sensitivity analysis.



(6.9.1) Autocorrelations [-] at the locations defined in Fig. (6.8) for 1-5 steps time lag, captured from monthly and weekly maxima.



(6.9.2) Pairwise (spatial) correlations [-] between the locations defined in Fig. (6.8), captured from monthly and weekly maxima. Above the diagonal Spearman correlation is used and below the diagonal Pearson.

Figure 6.9: Correlations: (a) captured in the observed data, (b) captured in the generated data and (c) the difference between the two.

6.3.4. Aspect 3: spatio-temporal integrity

Region of influence *To check GO3 for complete dynamic events.*

Fig. (6.10) displays the number of time steps of observed events which have at least one coordinate in the BI. Fig. (6.10)a shows an example of a precipitation event that started close to the east coast of the United States (US) and travelled all the way to the BI and beyond. Since all data points in this observed event were used to generate synthetic events, the data points close to the US (marginally) influenced the RLE in the BI. Fig. (6.10)b shows the spatial extent of all observed events used by the generator for the generation of the synthetic time series within the BI, providing a first order impression of the region of influence of RLE in the BI. The subset was taken of observed events which have at least one coordinate (i.e. one point in x,y,t of precipitation) within the extent of the BI (red square). If synthetic events would be reconstructed from observed events by changing the precipitation rates only (i.e. no changes in spatio-temporal extent, no movement of the centres of gravity of events and no STPP), this would be the data that influences RLE in the BI. From this subset of events, the number of time steps was counted per grid-cell. As can be observed, with the generator approach, many data points observed outside the BI were used to generate events in the BI. Note that Fig. (6.10) is for illustrative purpose only, as it is highly dependent on the event identification procedure described in Section (5.3.1). A higher global threshold for event identification would lead to a smaller extent, and vice versa. A global threshold indicates a single threshold that is used to divide the entire observed data set into data that is part of the observed events (data above the global threshold) and data that is not part of the observed events (data below the global threshold).

6.4. Sensitivity analysis of a dynamic precipitation generator

6.4.1. Overview

A gridded sensitivity analysis was performed. This means that an array of settings was defined, with which the effect of three critical settings was investigated.

1. *Global threshold for event identification.* First, given the importance of event identification, five different global threshold settings were investigated. For the entire observed data set, data above this global threshold was considered to be part of an event, whereas data below this threshold was not used.

2. *Spatial domain of generator.* Second, three spatial domains were used to generate synthetic data, displayed in Fig. (6.11): a large, baseline domain (BSL2); a small domain (SML2); and the smallest domain (UK2), which is a generator domain which is equal to the domain in which a flood risk analysis is to be executed. So, to generate boundary conditions for a particular target domain (UK2), the generator made use of the same domain as the target domain, or the generator made use of a bigger spatial domain (BSL2 or SML2), the reason of which will be discussed in Section (6.5.1).

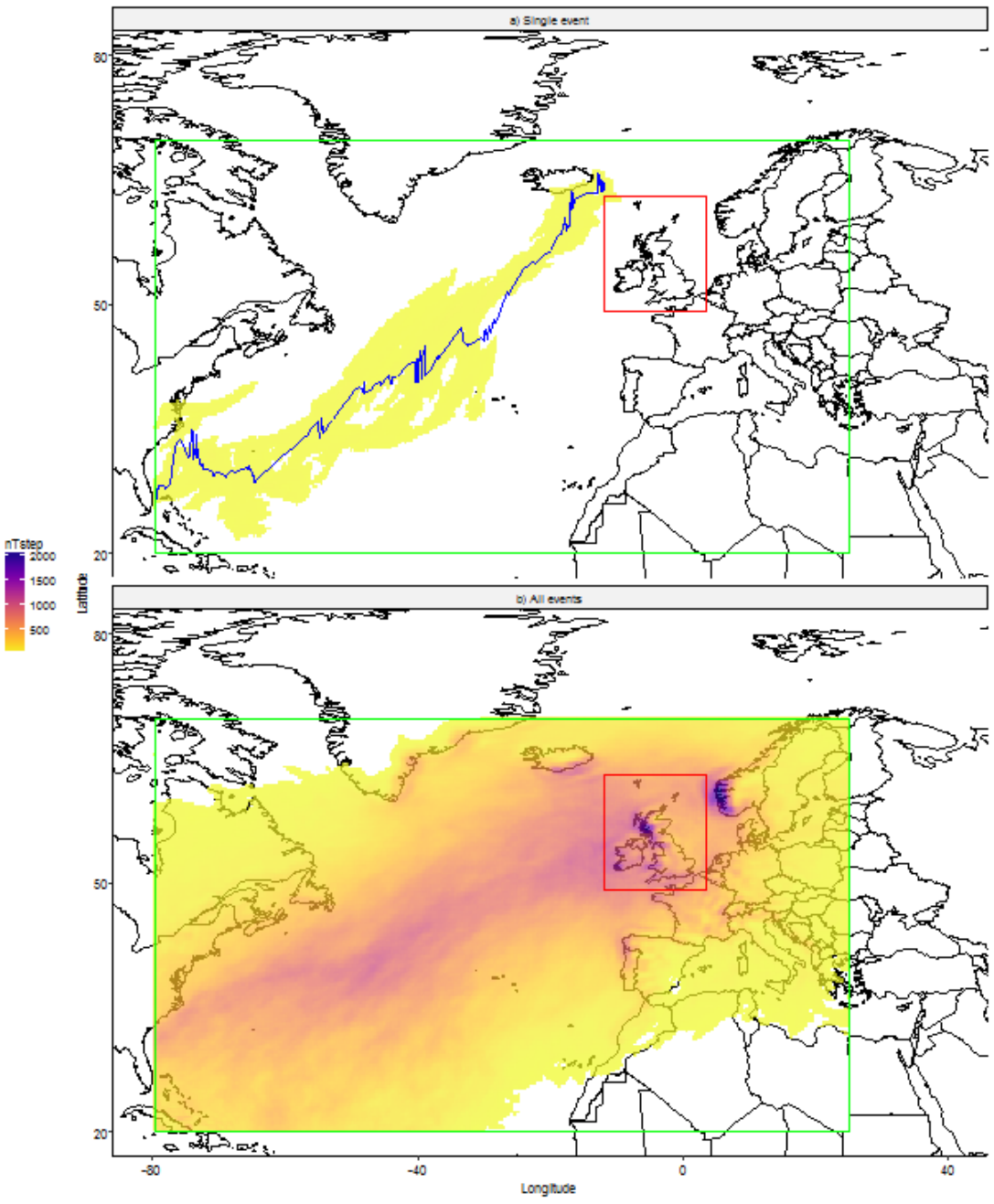


Figure 6.10: Per grid-cell, the number of time steps $nTstep$ [-] was counted, for (a) a single event that travels from the east coast of the US to the BI, and (b) all observed events which have at least one coordinate in the BI, which outlines an impression of the region of influence for the RLE of the BI using DSTPA.

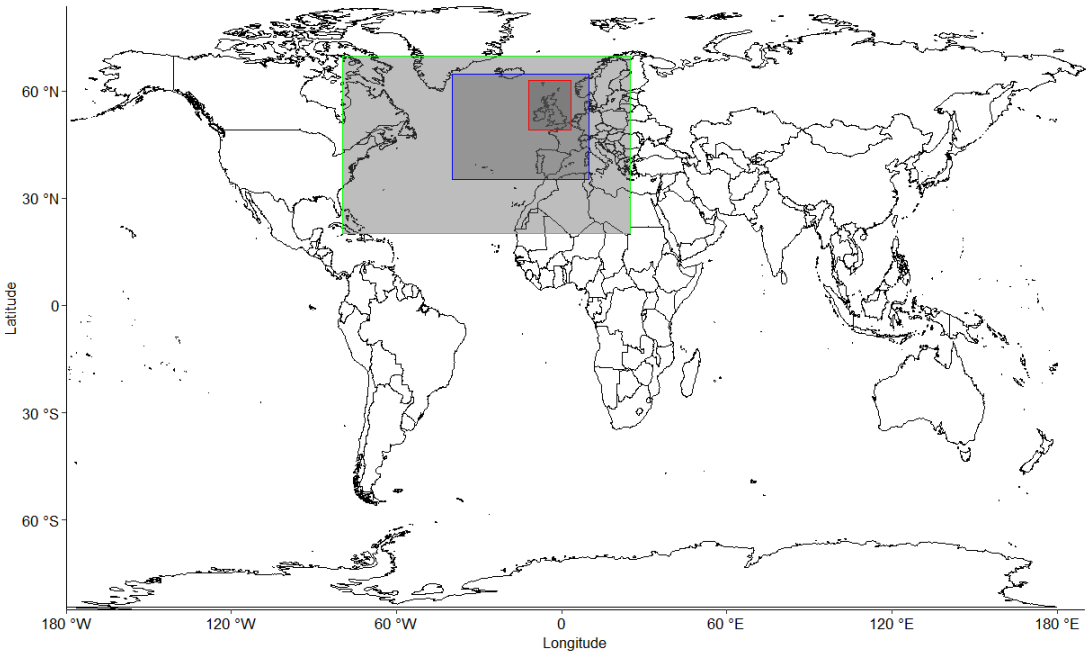


Figure 6.11: Three spatial domains were used to generate data: a big domain (green, BSL2), a medium domain (blue, SML2) and a small domain (red, UK2).

3. *Temporal length of input.* Third, varying temporal lengths of input data are used. The input data was divided into three periods: $P1 = 1981 - 2010$, $P2 = 1986 - 2005$, $P3 = 1991 - 2000$. With an assumption of stationarity and without any de-trending of the input data, the idea was that these periods would result in an approximation of the hazard around $1995 - 1996$, which is the midpoint of the three temporal ranges.

4. Several other generator settings were structurally adjusted such that the analyses could be carried out. A detailed overview of the settings can be found in Tables (A.5) to (A.7).

With the variations in these three settings, in total, $5 * 3 * 3 = 45$ generator runs were performed. Some other settings, that were not under investigation, had to be defined depending on the three critical settings, listed in Appendix (A.4.3). Throughout Section (6.4), each figure will be based on a total of 45 summary performance indicators. As such, the number of data points in each boxplot will be 45 divided by the number of subfigures.

So for this sensitivity analysis the performance indicators, described in Section (6.3), had to be summarized. This was done by taking the spatial means of the performance indicators, implying that each figure led to one summary performance indicator. These spatial means of the performance indicators are labelled in the figures by 'spatial mean (variable)'. In addition, roughness maps of each performance indicator were provided in Section (6.3), which were summarized similarly by spatial means. These spatial means of the roughness of the performance indicators are labelled in the figures by 'spatial mean (roughness)'.

Table (6.1) shows which performance indicators were investigated as a function of the varying settings. For conciseness, only results were described where, as a

	Global threshold	Spatial domain	Temporal input
Temporal means	X		
Event coverage	X		
Local extreme values	X	X	X
Temporal dependence			X
Spatial dependence	X	X	X
Completeness of events		X	

Table 6.1: An overview of the performance indicators investigated as a function of the varying settings.

function of the settings, there was some form of dependence or trend.

It should be mentioned that the sensitivity of many other components of the generator remains to be checked, for example settings with regards to classification, statistical models and reconstruction algorithms. These were not covered in this study because (a) they follow roughly the same procedure and (b) the amount of computation has become excessive for no obvious additional gain. The PI's already included in this chapter have served the purpose, which is to give a good demonstration of how this performance evaluation framework works for a statistical extreme weather generator.

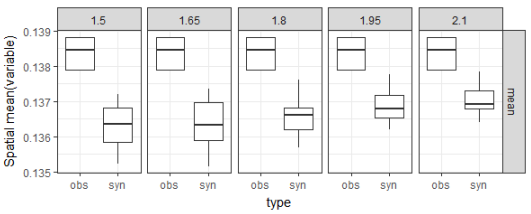


Figure 6.12: Sensitivity of mean of precipitation rate p [mm/h] to global threshold of Fig. (6.1).

6.4.2. Aspect 1: patterns of local distributions

Temporal means Temporal means were used to check the bulk of the populations of precipitations rates per location, for both the observed and the synthetic data. Fig. (6.12) shows that overall the temporal means are slightly higher for the observed data than for the synthetic, which indicates that there is not enough precipitation in the generated synthetic data. With a higher threshold, the temporal means of the synthetic become higher and the spread becomes smaller. Variation in the other settings showed no pattern and therefore they are omitted.

Considering the limited event coverage fractions that will be shown in Fig. (6.13), relatively similar maps of means should be expected. The reason for this is that the temporal means include not only event values but also the values below the gobal thresholds. The more variability is introduced, for which (high) event coverage fractions are important indicators, the more difficult it will be for the generator to retain similar means per location, and, therefore, the more indicative temporal means will be for the performance of the

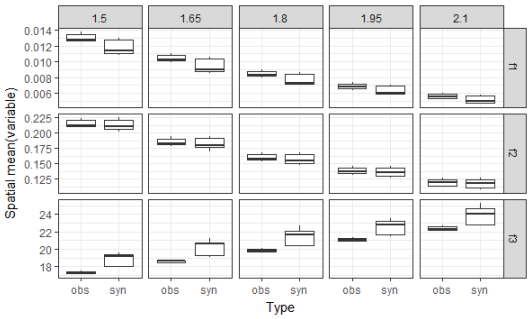


Figure 6.13: Sensitivity of event coverage to global threshold of Figs. (6.2.1) to (6.2.3). f_1 [-] is the fraction of the number of time steps in the event catalogue divided by the number of time steps in the continuous series. f_2 [-] is the fraction of the total sum of precipitation in the event catalogue divided by the total sum of precipitation in the continuous series. f_3 [-] is the fraction of the mean of precipitation in the event catalogue divided by the mean of precipitation in the continuous series.

generator.

Several extensions of methodology were proposed in Section (5.5). In particular, when a more daring space-time placement procedure is incorporated, not only will events be moved around, but, as a consequence, also small non-event values will be moved around. This means that an eye should be kept on the spatial patterns of local distributions, for which the temporal mean is a basic indicator.

Event coverage Fig. (6.13) shows that overall $f1$ and $f2$ decrease with a higher threshold, whereas $f3$ increases. This makes sense, because the fraction of time steps in the event catalogues $f1$ should decrease with a higher threshold, since fewer data points are assigned to events. As such, the fraction of sums of precipitation in the event catalogues $f2$ is to be expected to be lower with a higher threshold. And since a threshold is used, the fraction of means in the event catalogues $f3$ is to be expected to be higher with a higher threshold. It can be observed that $f1$ is slightly lower in the synthetic data than in the observed and that $f3$ is slightly higher. However, $f2$ shows reasonable similar results in the observed as in the synthetic data, with a slightly lower mean. This indicates that, although there is some room for improvement, about the right amount of precipitation is contained in the generated synthetic event catalogue. In addition, the spread of the distribution is wider in the synthetic than the observed, possibly due to the larger sample.

The generator described in Chapter (5) was developed to derive a stochastic process from an observed data set. Introduction of variability (GO1) was mainly achieved by replacing observed events with synthetic events. In order to be able to introduce variability everywhere, events are required for all locations. If variability is to be introduced in a spatially homogeneous way, the event coverage should be spatially homogeneous, implying the same fraction of event coverage for each location. Therefore, it could be argued that an event identification procedure should produce a map of spatially homogeneous event coverage.

To check whether the chosen event identification procedure in Section (5.3.1) led to spatially homogeneous event coverage, event coverage was considered with different fractions. The global threshold did a relatively good job in producing spatio-temporally coherent events. However, it did not produce a homogeneous event coverage, as certain areas, like hills or mountain ranges with effects such as orographic lifting, may receive relatively high precipitation rates and therefore more

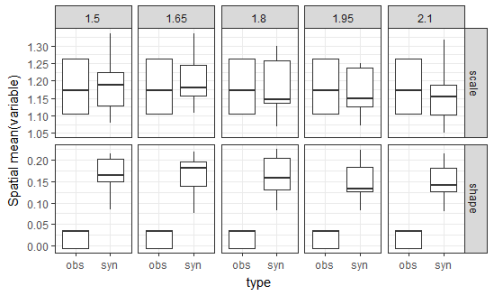
events, whereas other areas, like plains, may receive relatively low precipitation rates and therefore fewer events. A simple alternative to the global threshold would be to set thresholds varying per location. That could lead to exactly the same event coverage per location in the observed data set, either: by using the fraction of number of precipitation values, as in Fig. (6.2.1); by using the fraction of precipitation sums, as in Fig. (6.2.2); or by using the fraction of precipitation means, as in Fig. (6.2.3). However, forcing the event coverage to be homogeneous on a per-location basis would lead to fragmentation in the spatio-temporal events. This would be particularly the case in spatio-temporal events with relatively small precipitation values.

A discussion should be raised in the hydrological/statistical society on the topic of dynamic event identification, which should be the starting point for any event-based hazard analysis. First, what should be considered events and, second, how to develop algorithms to capture these envisioned events? Note that the development of sophisticated event identification algorithms would only make sense after a fruitful outcome of the discussion on event identification. Some requirements for events can potentially be used in the discussion. First, variability should be introduced in the entire data set (GO1). This may be interpreted to give a requirement for a relatively homogeneous event coverage. Second, events should be spatio-temporally coherent (GO3), which will require event identification procedures that go beyond identification on a per-location basis. Finally, sparsity, which was discussed in Section (3.2.1), should be expected to lead to a certain level of heterogeneity of the event coverage in the observed data set. Therefore, a homogeneous map of event coverage in the observed data set should not be a goal. However, sparsity would theoretically not be present in the full stochastic process and therefore a fully homogeneous map of event coverage of the synthetic data could be a goal.

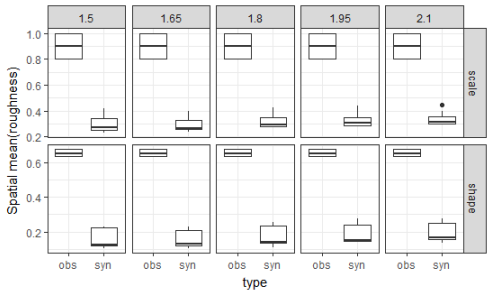
Local extreme value analysis In Fig. (6.14), the sensitivity of local GPD parameters (per grid-cell) is investigated. It shows that the scale parameter is roughly similar for the GPDs fitted to the observed and the synthetic data. For the GPDs fitted to the synthetic data, the shape parameters are generally higher and both parameters are spatially much smoother than the GPDs fitted to the observed.

Fig. (6.14.1) shows that both the scale and the shape parameter vary only slightly with a varying global threshold. Fig. (6.14.2) shows that the GPDs fitted to the synthetic data become slightly less smooth with a higher global threshold. Fig. (6.14.3) shows that the scale parameters are roughly equal on average for the GPDs fitted to the synthetic data than fitted the observed, but the shape parameter is much higher for the GPDs fitted to the synthetic. It shows that the scale parameter may be slightly higher or lower, depending on the chosen generator domain. The shape parameters appears to become closer to zero with a decreasing size of generator domain.

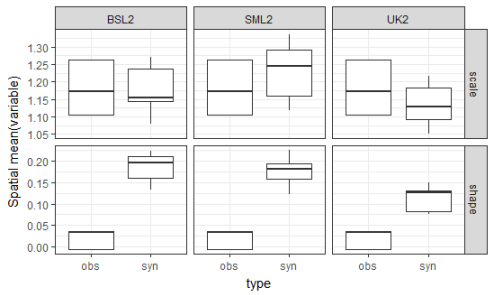
Fig. (6.14.4) shows that both the scale and the shape parameter are much smoother for the GPDs fitted to the synthetic data than fitted the observed. The GPDs fitted to the synthetic data are slightly less smooth with a decreasing size of generator domain. Fig. (6.14.5) shows that both the scale and the shape param-



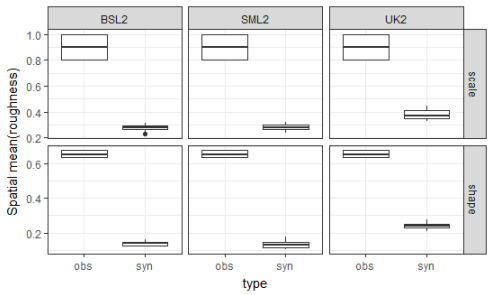
(6.14.1) Sensitivity of local GPD parameters to global threshold.



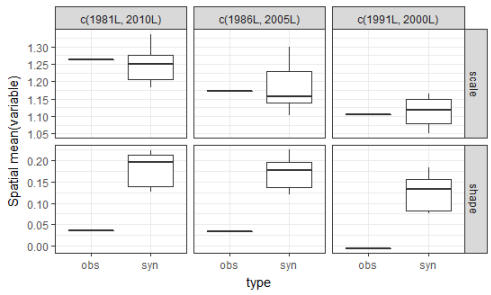
(6.14.2) Sensitivity of spatial roughness to global threshold.



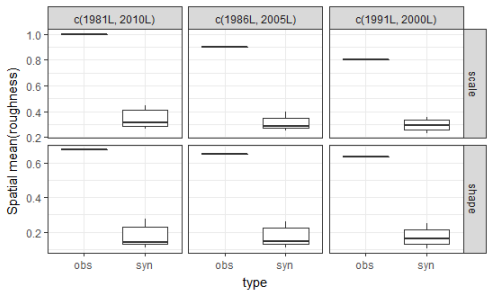
(6.14.3) Sensitivity of local GPD parameters to generator domain.



(6.14.4) Sensitivity of spatial roughness to generator domain.



(6.14.5) Sensitivity of local GPD parameters to period of observed data.

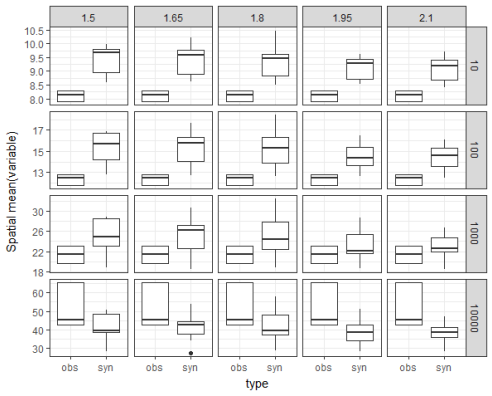


(6.14.6) Sensitivity of spatial roughness to period of observed data.

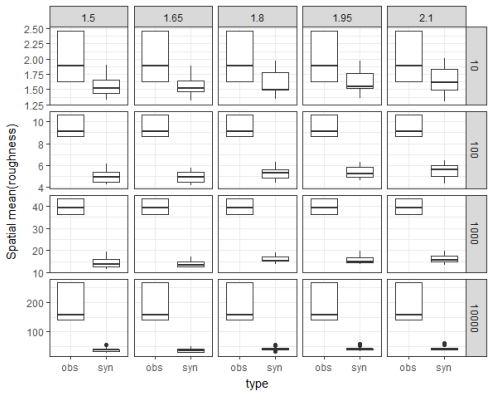
Figure 6.14: Sensitivity of: local GPD scale [-] and shape [-] parameters (per grid-cell), Fig. (6.7.1); and of the spatial roughness [-] of the local GPD parameters, Fig. (6.7.3).

eter decrease with a shorter period of input data. Fig. (6.14.6) shows that the GPDs fitted to the synthetic data remain about equally smooth with a varying period of input data.

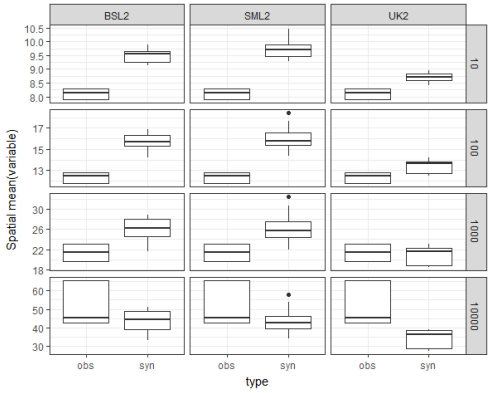
In Fig. (6.15), the sensitivity of local RLE (per grid-cell) is investigated. It shows that RLE calculated from the GPDs fitted to the synthetic data on average are higher for low return periods and lower for high return periods, compared to those calculated from the GPDs fitted to the observed. This appears largely to be a result of the higher degree of spatial coherence of the extremes in the synthetic data. Fig. (6.15.1) shows that the RLE calculated from the GPDs fitted to the synthetic data are slightly lower when picking a lower global threshold. Fig. (6.15.2) shows that the roughness of the RLE calculated from the GPDs fitted to the synthetic data is much lower, i.e. is smoother. Fig. (6.15.3) and Fig. (6.15.4) show similar behaviour as found in Fig. (6.14.3). No clear pattern is present so the sensitivity of RLE to input domain could be further investigated.



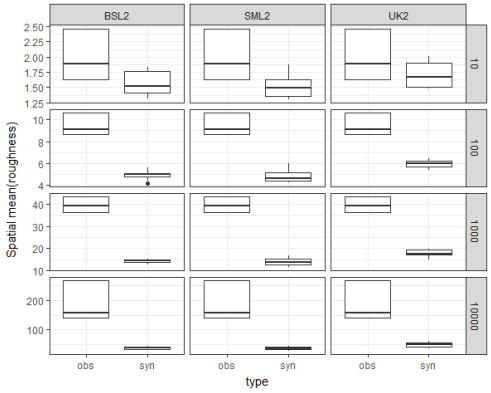
(6.15.1) Sensitivity of local RLE to global threshold.



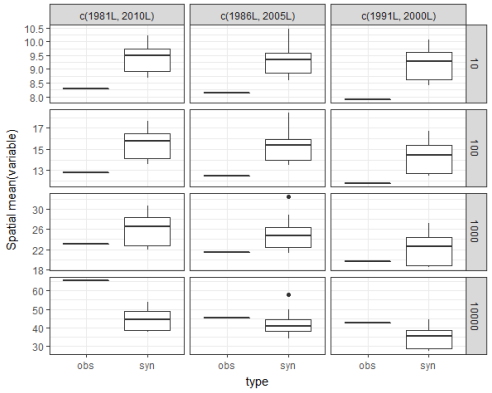
(6.15.2) Sensitivity of roughness of local RLE to global threshold.



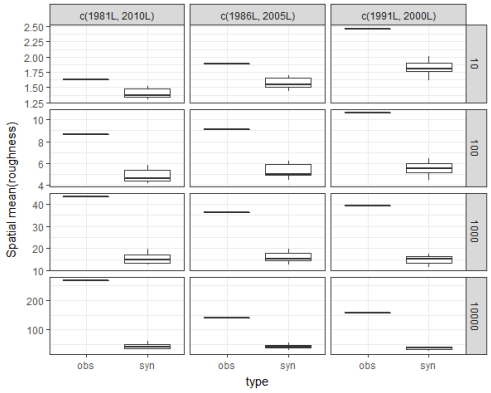
(6.15.3) Sensitivity of local RLE to generator domain.



(6.15.4) Sensitivity of roughness of local RLE to generator domain.



(6.15.5) Sensitivity of local RLE to period of observed data.



(6.15.6) Sensitivity of roughness of local RLE to period of observed data.

Figure 6.15: Sensitivity of: local RLE of precipitation rate p [mm/h] per grid-cell, Fig. (6.7.2); and of the spatial roughness [-] of the RLE, Fig. (6.7.4).

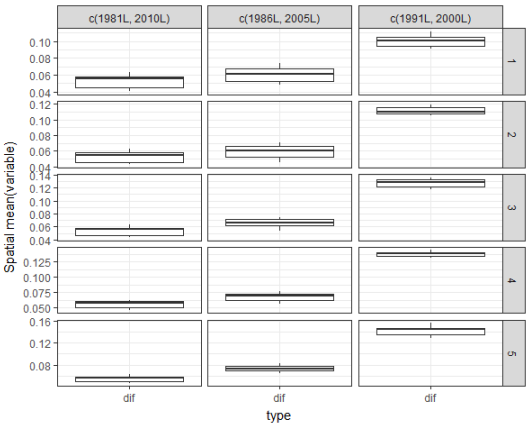


Figure 6.16: Sensitivity of autocorrelations [-] to period of observed data [y], Fig. (6.9.1).

Fig. (6.15.5) shows that the RLE calculated from the GPDs fitted to the synthetic data are slightly lower when using a shorter period of input data. Fig. (6.15.6) shows that the RLE corresponding to low return periods are rougher when using a shorter period of input data. However, this relationship appears to reverse for RLE corresponding to higher return periods.

The local extreme value analysis will be further discussed in a dedicated Section (6.5.2).

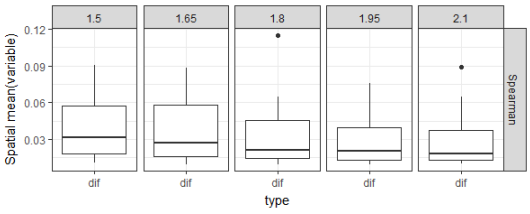
6.4.3. Aspect 2: patterns of occurrence

Temporal dependence Fig. (6.16) displays the difference in autocorrelations between the observed data and the synthetic. It shows that the differences in autocorrelations are significant and that the generator should be improved to take autocorrelation into account. It shows that autocorrelations are lower in the synthetic data than in the observed data when using a shorter period of input data, which indicates that they are not well captured by the generator.

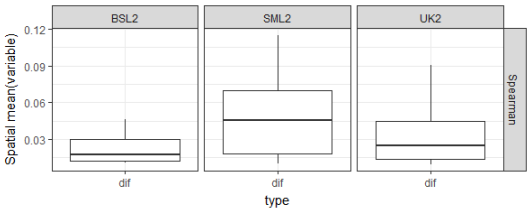
Temporal dependence between events was not taken into account in the generator. Typically, event identification procedures are designed to (aim to) capture temporally independent events, which means that no temporal sequencing (or spatio-temporal process) has to be implemented in the reconstruction procedure, thereby significantly simplifying the process. However, such independence can be accomplished using event identification procedures that result in temporally distant events, like annual blocks or like POT with large temporal spacing using temporal separation techniques or simply very high thresholds. Regardless of whether a generator implements techniques to capture temporal correlation, it is an important check to include since temporal sequencing can be important, in particular for precipitation [8].

Fig. (6.16) showed that temporal correlation was not well addressed with the generator developed in Chapter (5). In Section (5.5.4), a recommendation was made to incorporate a space-time placement process. It is expected that such a process should be able to appropriately address short term sequencing. Longer term sequencing, such as seasonality, could be incorporated by classification based on seasonality, as recommended in Section (5.5.2).

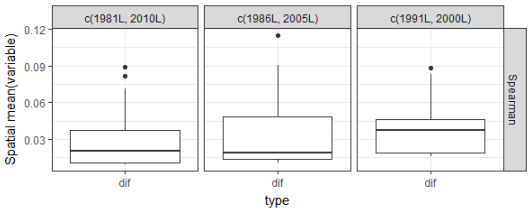
Spatial dependence Fig. (6.17) displays the difference in spatial correlations between the observed data and the synthetic. It shows that pair-wise spatial correlations are generally quite similar for the observed and the synthetic data, which means that spatial correlation is reasonably well captured by the generator. Fig. (6.17.1) shows that the differences are slightly smaller with an increasing global threshold. Fig. (6.17.2) shows no clear relation between spatial correlations and the size of



(6.17.1) Sensitivity of spatial correlations [-] to global threshold of precipitation rate p [mm/h].



(6.17.2) Sensitivity of spatial correlations [-] to generator domain [-].



(6.17.3) Sensitivity of spatial correlations [-] to period of observed data [y].

Figure 6.17: Sensitivity of spatial correlations; Fig. (6.9.2).

the input domain. The spatial correlations are the least well captured in the medium domain SML2. Fig. (6.17.3) shows no clear relation between spatial correlations and the period of input data. The means of boxplots are quite similar for $P1$ and $P2$. It could be that the differences become larger for $P3$, because the observed data record becomes too small and spurious correlations are captured in the observed data.

Spatial dependence across locations was not explicitly taken into account in the generator. Fig. (6.17) showed that spatial correlations were fairly well captured by the dynamic spatio-temporal generator.

It should be expected that spatio-temporal coherence of dynamic events largely takes care of the spatial dependence between different locations. However, a

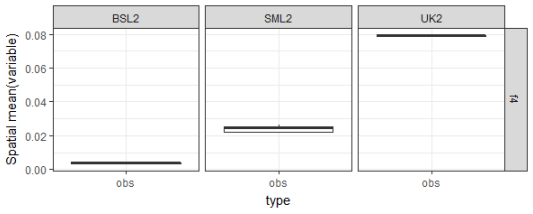


Figure 6.18: Sensitivity of an indicator of spatio-temporal integrity $f4$ [-] to generator domain of Fig. (6.10). $f4$ is the number of time steps in the observed event catalogue at the edge of the generator domain, divided by the total number of time steps in the observed event catalogue within the generator domain.

challenge lies with some of the expanded proposed procedures in Section (5.5.4). In particular, it will be challenging to retain spatial dependence across locations, if not only dynamic spatio-temporal events are reconstructed, but if also a space-time placement process is applied.

6.4.4. Aspect 3: spatio-temporal integrity

Completeness of events In Fig. (6.10)b, it can be observed that if a domain is chosen too small, incomplete dynamic events are used to obtain RLE in the BI. To quantify this effect $f4$ is used, which is the number of time steps in the observed event catalogue at the edge of the generator domain divided by the total number of time steps in the observed event catalogue within the generator domain.

Fig. (6.18) displays the number of time steps in the observed event catalogue at the edge of the generator domain, divided by the total number of time steps in the observed event catalogue within the generator domain.

As expected, Fig. (6.18) shows that $f4$ decreases with an increasing size of the generator domain. Therefore, the genera-

tor benefits from a generator domain that approximates the area of influence on the area of interest in which RLE are to be calculated. To use a bigger generator domain would be fine. To use a smaller generator domain implies that statistics are applied to incomplete dynamic events.

When considering a dynamic spatio-temporal generator, in which events freely move around, it would be best to formulate no spatial boundaries. A number of problems can be avoided by applying the generator on a global scale. No dynamic precipitation events would be incomplete. Classes of dynamic precipitation events, which should probably be obtained by classification according to region as proposed in Section (5.3.2), could be complete.

Computational limitations are the reason why this method was not applied at the global scale. Computational aspects are reported in Appendix (A.4). For efficient computation with the current code, a massive amount of RAM and disc space would be required for global scale computation. However, there are potential gains. For example, global precipitation hazard time series could be obtained, for which a proof-of-concept will be proposed in Chapter (7). Therefore, it should be worthwhile to put some large computational power to work to generate globally and, thereby, avoiding the issue of incomplete events.

6.5. Discussion

6.5.1. Dynamic expansion of information - a new alternative

One of the main sources of uncertainty in risk analysis is the limited sample of observed data, i.e. a limited amount of evidence. To expand a (relatively small) sample of local evidence, Merz and Blöschl [61] provided a framework for 'expansion of information' (EOI), in which they outlined three types: causal, temporal and spatial.

First, causal EOI. In Section (1.1.3), a simple flood risk modelling chain was introduced, which showed that the statistical step can be applied in different layers in the chain; Fig. (1.2). In the statistical step, multiple (compound) variables may be addressed in a single layer, as was done in Section (5.4). However, if causality exists between variables, then preferably the variables should be put into different layers in the flood risk modelling chain. The statistical step should be applied in a single layer, and dependent variables should be calculated through the modelling chain. In the context of precipitation, causal EOI implies that information on the causes of precipitation events is used - or information on the consequences. This means that information would be combined from what should be different layers in the flood risk chain. Therefore, causal EOI is not considered.

Second, temporal EOI. Given the relatively large amount of uncertainty in hazard analyses, typically all available temporal information at a particular location is used. Temporal expansion of information implies that not only all available information should be used, but that additional information should be gathered. E.g. old archaeological evidence, such as ancient flood marks, which can then be used as a proxy for peak water levels of extreme flood events. For temporal EOI, a big assumption is required: stationarity. This may be considered a problematic assumption, in particular in times of accelerated climate change. When an event occurs, it provides information about the system in which it occurs. For example, an event that occurred 100.000 years ago provides

information about the system of 100.000 years ago. This means that, if such an event would be incorporated to fit distributions of the current system, implicitly an assumption is made of 100.000 years of stationarity, i.e. no changes in the system. In general, given a hazard analysis of a system state (the system in a particular point in time), the closer on the time line an event is to the point in time of the studied system state, the more informative the event is. This means that, for a study of the current system state, recent data should be used. Since this statement directly contradicts the idea behind temporal EOI, the following question is raised: how long should the time window of evidence be - are there any alternatives to temporal EOI - what is a good compromise?

Third, spatial EOI. The general idea of spatial EOI is that time is substituted for space. So, as an example of spatial EOI, a spatial process may be applied to GPD parameters, such that spatially coherent GPD fits are produced. In Section (6.5.2), it was explained that the application of such a spatial process to model parameters may be considered an approach that requires quite strong and explicit assumptions, and that instead the sparsity of dynamic events should be addressed.

In Chapter (5), a generator was developed to simulate the stochastic process of a particular system state. In Section (6.4.2), two approaches to obtain RLE in the British Isles (BI) were described; a naive generator approach (denoted as $A1$, fitting extremal distributions to observed data) and a dynamic spatio-temporal generator (denoted as $B1$, fitting extremal distributions to synthetic data).

The input data to both generators was 32 years (of observed data), whereas the temporal length of the output was 1.000 years (of synthetic data). The spatial extent for both approaches was the same (BI). However, the synthetic data came from the dynamic spatio-temporal generator, in the procedure of which observed data (evidence) outside the BI was used. For each location, the dynamic spatio-temporal generator made use of more observed data than the local time series (per pixel). Therefore, as compared to traditional local analysis of extremes in time series, EOI is applied, but none of the three listed above (causal, temporal or spatial). To understand what type of EOI was applied, it should be considered which evidence (observed data) was used to calculate the RLE. In Section (6.4.4), it was roughly traced back which observed data was used to generate synthetic data in the domain UK2; Fig. (6.10). It is not temporal EOI, since the generator is based on the 32 years of observed data, i.e. no additional temporal evidence was used. It is different from spatial EOI, in that time was not substituted for space, but was included based on the movement of events through space in time. This introduces a new concept: 'dynamic expansion of information'.

6.5.2. Spatially coherent extremes - what is wrong with the spatial process

In Fig. (6.7.3), it was observed that the parameters of General Pareto distributions fitted to the synthetic data were spatially much more smooth than those fitted to the observed data. This may be explained by the use of dynamic spatio-temporal probabilistic analysis, Appendix (A.1.6), which benefits from dynamic expansion of information, as was discussed in Section (6.5.1). However, when using generators that work on a per-location basis, a spatial process is

normally used to obtain a similar effect - increased spatial smoothness.

The spatial process refers to particular methodology that works with static spatio-temporal probabilistic analysis, Appendix (A.1.5). In essence, extreme local distributions are fitted (on a per-location basis), but then the spatial process provides smoothness across nearby locations, based on the expectation that proximity should give comparable extreme value distributions. This aspect relates to GO1, since smooth local distributions show that variability is introduced with regards to the sparse, local sample. Spatial sparsity is another aspect that should be addressed. These aspects directly relate to what is aimed to be achieved in the overarching model, the stochastic process, described in Section (3.1.1).

In the literature, several authors applied spatial process by forcing spatial smoothness of GPD parameters using Bayesian statistics. Gilleland, Nychka, and Schneider [40] considered ozone: *'However, the number of observations is small for any one station, and one would expect significant uncertainty in the estimates because of too few observations (exceeding the threshold). Also, the individual models provide no obvious way to extrapolate to locations where ozone is not measured. One strategy to improve the accuracy and provide for spatial prediction is to include a spatial component that links the distribution for different stations. In this section, a hierarchical component is added that treats the parameters of the GPD as a smooth surface.'* Cooley, Nychka, and Naveau [21] considered precipitation: *'We anticipate that the log-transformed scale parameter $\varphi(x)$ will be sensitive to regional climate effects and build a model which describes its relationship with the latent spatial process. ... In contrast to the transformed scale parameter, we are less certain of the shape parameter's sensitivity to regional variables. Because the shape parameter is more difficult to estimate than the scale parameter, we start to model $\xi(x)$ as a single value and increasingly add complexity until we have a reasonable fit'* Sharkey and Winter [81] considered precipitation: *'This layer <process level> of the hierarchical model borrows strength across locations. We assume an underlying spatial process in the mean of the distribution of both GPD parameters, such that the parameters are more likely to be similar in neighbouring grid-cells. As well as these spatial effects, we can also incorporate fixed climate or physical effects in a grid-cell through the inclusion of covariates in the model.'*

In general, these authors think that a stochastic process of coherent events should be expected to produce spatially coherent extremes and therefore spatially smooth GPD parameters. Although slightly dependent on the spatial resolution of the data, this should be considered a reasonable expectation. With that thought in mind, all of these authors spatially smoothened the parameters of models fitted to descriptors of events.

With these approaches, it is quite common that the directly estimated parameters were intentionally altered to fit with an prior belief of smoothness in the marginal distribution parameters. The extent of alteration would be dependent on the assumption of a Bayesian prior distribution or a penalty term for roughness during the parameter estimation. In both cases, there is usually a quite explicit assumption about the expectation of smoothness, which one could argue is dif-

ficult to justify.

The roughness in the map of GPDs fitted to the observed data may be due to the sparsity of events in the observed data. So, rather than spatially smoothening the model parameters, which might require quite explicit and often ”hard-to-justify” assumptions, the spatial sparsity of events should be addressed, directly in conjunction with the temporal sparsity. The generation of synthetic dynamic spatio-temporal events provides a viable alternative in that regard.

6.6. Conclusions

Generators may generate massive amounts of data, in particular those of the dynamic type. The performance of generators can be explored by investigating how the generated, synthetic data relates to the observed data that it was generated from.

In Section (6.2), a generic framework was provided which can be used to explore the performance of generators. To check the performance of a generator, all aspects of performance should be considered and not just the particular aspects that the generator is trying to optimise. The framework was indicated to be non-exhaustive, but aims to help in the organisation of the performance check. It proposes to formulate generic generator objectives, where for each aspect performance indicators should be set up that then are to be tested with different generator settings, aiming to get an idea of the sensitivity of the generator.

In Section (6.3), the framework was tested using a generator of the dynamic type, developed in Chapter (5). To explore the performance of the generator, a plethora of diverse performance indicators was set up after which a sensitivity analysis of these performance indicators was executed in order to understand how the generator works with different settings. The indicators informed on several aspects of the generator's performance. First, the generator managed to replicate the event coverage in the observed data fairly well. Second, local tail-end distributions (Generalised Pareto distributions) were more spatially coherent when fitted to the generated synthetic data than to the observed. Third, local RLE of high return periods derived from the synthetic data were spatially more coherent than the ones derived directly from the observed data. Fourth, the general patterns in temporal and spatial correlations were fairly well captured, but some significant differences were found, which means that these patterns should be aimed to be improved when further extensions are applied to the generator. Fifth, spatio-temporal integrity is an issue when not using the entire global domain, in which case dynamic events and classes of dynamic events may be incomplete and therefore potentially slightly misleading.

In Section (6.4), a sensitivity analysis of the performance indicators was executed in order to understand how the generator works with different settings. As a tool for the exploration of generator performance, the sensitivity analysis was found to be feasible, under the condition that a large amount of computational power and storage capacity are available. The generator developed in Chapter (5) was found to be sensitive in several aspects. First, the event coverage was found to be highly dependent on the global threshold, which indicates that alternative methods of event identification should be discussed and developed. Second, local distributions were found to be sensitive to all settings, which indicates that the generator method should be refined and that a more extensive sensitivity analysis should be conducted in which parameters spaces should be found that produce stable results. Third, local RLE were found to be sensitive to all settings, which is a direct result of the sensitivity of the General Pareto distributions. Fourth, temporal correlations were found to be sensitive in particular to the period of input data and spatial correlations were found to be sensitive to all settings, which means that the methodology should be expanded to perform better on these aspects, in particular on seasonality. Fifth, the simple test of

spatio-temporal integrity was found to be sensitive to the size of the input domain, which indicates that the generator domain should be chosen sufficiently large, to be able to deliver spatio-temporal integrity for the smaller target domain.

In Section (6.5), two issues were discussed that relate to the performance of generators, but that cannot be captured within the framework itself. First, it was discussed that generators operate based on evidence (observed data), where the limited historical sample is one of the main sources of uncertainty in risk analysis. The performance of a generator is related to how well it deals with the generator objectives, for which the observed data is the basis. However, it also matters how well the generator is able to incorporate evidence, for which a framework of 'expansion of information' exists in the literature. An addition to this framework was provided, 'dynamic expansion of information', which comes with the newly developed type of 'dynamic generator'. The investigation of the region of influence showed that the extreme value analysis using the synthetic data was derived from much more than the data observed in the region for which the extreme value analysis was carried out. This is a direct result of the consideration of moving, dynamic events and hence the concept was introduced of dynamic expansion of information. When discussing the quality of a generator, it should be included how well the generator makes use of the different options of expansion of information. Second, the spatial process was discussed, which is a method used to provide spatial smoothness of extreme distributions, based on the expectation that the extreme distributions should be spatially smooth. The dynamic type of generator does not make use of a spatial process, but the local extreme value analysis revealed that the General Pareto distribution parameters fitted to the generated synthetic data were spatially much smoother than those fitted to the observed data. Where typically in the literature such smoothness is forced by use of a spatial process, which could be related to strong and explicit assumptions about the spatial smoothness. When such assumptions are hard to justify, the generation of synthetic dynamic events provides a viable alternative to the approach of the spatial process.

In summary, a generic framework to explore the performance of generators was provided, tested and discussed. It provides a tool to get an overview of the many different performance aspects that have to be considered, which inform on the quality of the generator and provide a certain amount of grip on the often massive amounts of synthetic data produced using generators.

7

Summary, outlook and conclusions

7.1. Summary

In Chapter (3), a framework was proposed for the event-based generation of synthetic data. This framework helps with the systemic development of event-based generators, to be used to provide an estimate of the stochastic process. It was applied in this thesis within the context of large-scale flood hazard sources.

In Chapter (4), the framework was applied to the river discharge in all major European river basins. This data was extracted from a (modelled) gridded discharge data set that covers the entire European continent. Methodology of the type event-based static spatio-temporal probabilistic analysis (SSTPA) was applied. Advances were made by tracking discharge waves and by successfully applying the statistical model to a a matrix of high dimensionality, thereby capturing the spatial dependence structure of discharge peaks. Three main limitations of event-based SSTPA were identified. First, the application of event-based SSTPA lead to gaps in the descriptor matrix. This is problematic for the statistical model and the reconstruction of events. Second, as a result of the first, it was concluded that this type of methodology is not spatially scalable. With a larger spatial domain, there will be an increasing amount of gaps. Third, with SSTPA, an increased density of observed data, i.e. using more locations, results in an increased dimensionality of the statistical model. This implies that, beyond the limiting dimensionality of the statistical model, event-based SSTPA does not benefit from an increasing density of observed data.

In Chapter (5), methodology was developed of the type event-based dynamic spatio-temporal probabilistic analysis (DSTPA), designed in particular to overcome the limitations of event-based SSTPA. The generator was developed to be applied to - and to benefit from - large, dense, gridded data sets. It was applied to a global precipitation data set. A subset of the global data set was used, but, with more computational resources, the methodology can be readily applied to the entire global data set and to data sets of increased resolution. By describing dynamic spatio-temporal events, the dimensionality of the matrices for multivariate analysis was significantly reduced and no gaps emerged, thereby affirming compatibility between the method of application and the statistical model used at the heart of the generator. In addition, the possibility of application to multiple compound sources was investigated and successfully applied, with the requirement of a few slight extensions of the methodology.

In Chapter (6), the performance of the dynamic precipitation generator was analysed. A framework of relevant checks was developed, for which simple performance indicators were provided. Using these performance indicators, a range of reasonable generator settings was explored. This analysis required significant computational effort and resources, which highlighted the need for simple but effective algorithms in the context of big data analysis. A non-standard category of spatio-temporal integrity emerged, which is related to the (new) spatio-temporal character of the dynamic generator. It was found that the generation of synthetic, dynamic events provides spatial smoothness of extremes. This eliminates the need for explicit assumptions (or prior belief) of the spatial smoothness in the model parameters from the application of SSTPA in combination with a spatial process, which is an-

other popular approach in the literature. Finally, the concept of dynamic expansion of information was introduced, which may play a key role the production of large-scale hazard time series.

7.2. Outlook - large-scale precipitation hazard time series

7.2.1. Ingredients

Hazards are dependent on the system state and therefore change in time, as the system evolves in time. In Section (3.1.1), it was explained that the stochastic process should be expected to be smooth both in space and in time. In Section (6.5.1), a fundamental problem with temporal expansion of information was introduced, related to the (in)validity of the stationarity assumption, where it was argued that not necessarily all temporal evidence should be included. Dynamic expansion of information was introduced, which can be used to incorporate additional evidence to a particular location by considering dynamic events that (could potentially) move over that location. To investigate if a stochastic process with spatially and temporally smooth return level estimates (RLE) can be obtained using a limited time window of evidence but with the help of dynamic expansion of information, the generator developed in Chapter (5) is applied, using the results of the sensitivity analysis applied in Chapter (6), to generate precipitation hazard time series.

7.2.2. Settings

A maximum amount of event coverage should be obtained using a lower threshold (OBJ1), but the threshold should not be chosen too low to be able to separate dynamic events, which is required for sufficiently large populations in the statistical models. The global threshold was set to be 1.8, which showed reasonable performance in all sensitivity tests.

The BSL2 domain was chosen based on the spatio-temporal integrity of events; Fig. (6.18). Computational power is the argument to chose a smaller spatial domain, but sufficient computational power was available for performing the analysis in the larger domain of the three investigated; Fig. (6.11).

A limited temporal coverage was used comprising 20 years of observed continuous precipitation data, which in the sensitivity analysis gave reasonable performance in retaining auto and spatial correlations. The temporal coverage was subsequently shifted by one year in order to obtain the inputs for precipitation hazard time series.

7.2.3. Local temporal smoothness

In Figs. (7.1) to (7.3), time series of RLE are displayed, corresponding to return periods of respectively 10, 100 and 1.000 years calculated from the observed and the synthetic data. It can be observed that, in general, the RLE calculated from the synthetic data are slightly higher than from the observed, with the exception of the locally very high RLE peaks, for which the RLE calculated from the observed data are much higher than from the synthetic. So a smoothening effect is apparent. It may be interpreted that the high peak RLE, found at particular locations in the observed data, are more spread out in the synthetic data over adjacent locations. This spatial smoothening effect was already evident in Fig. (6.7.4). However, now a smoothening effect can also be observed in time. This combination of spatial and temporal smoothening demonstrates

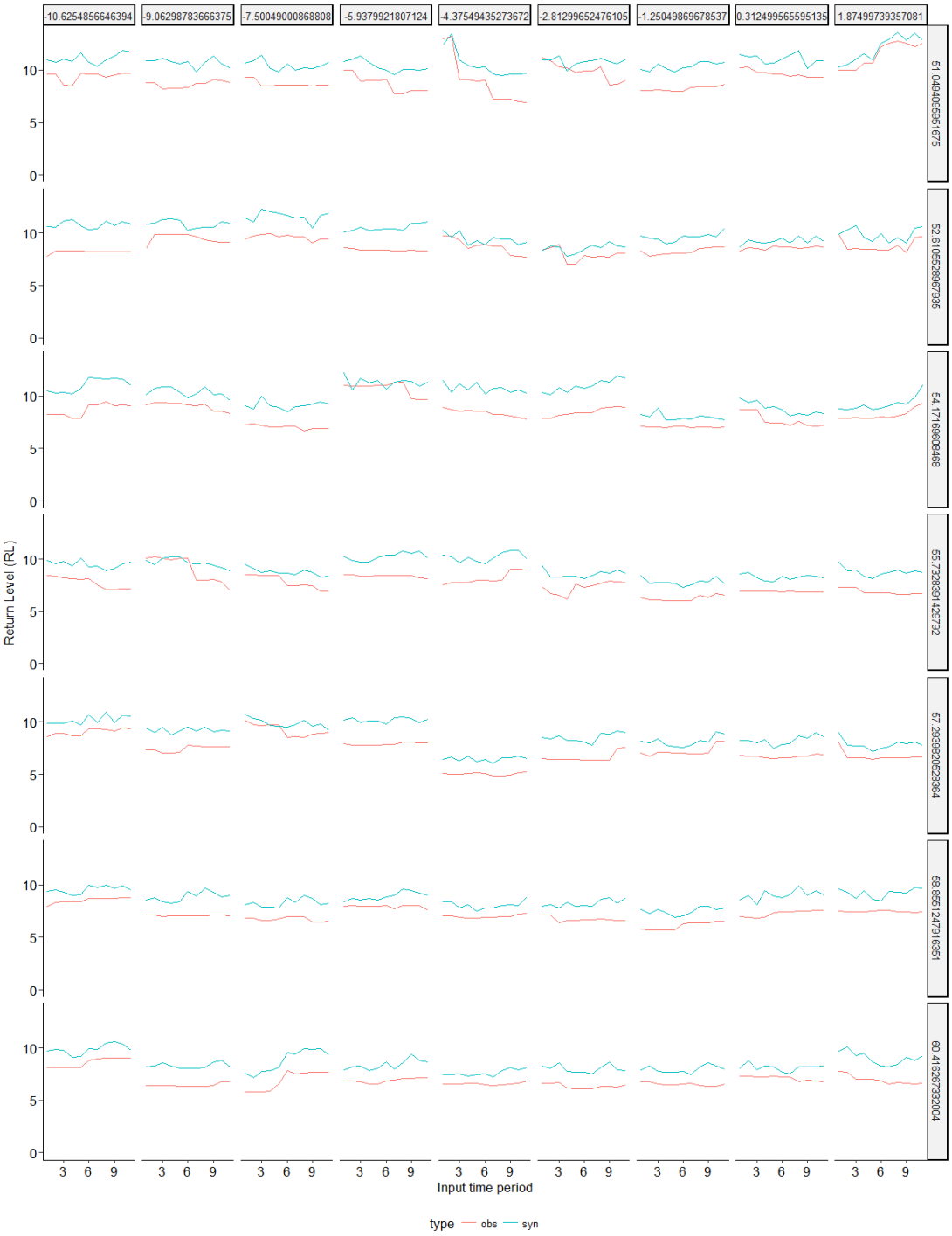


Figure 7.1: Time series of RLE of precipitation rate p [mm/h] corresponding to a return period of 10 years, for the locations defined in Fig. (6.8). The labels are the longitudes and lattitudes of the locations.

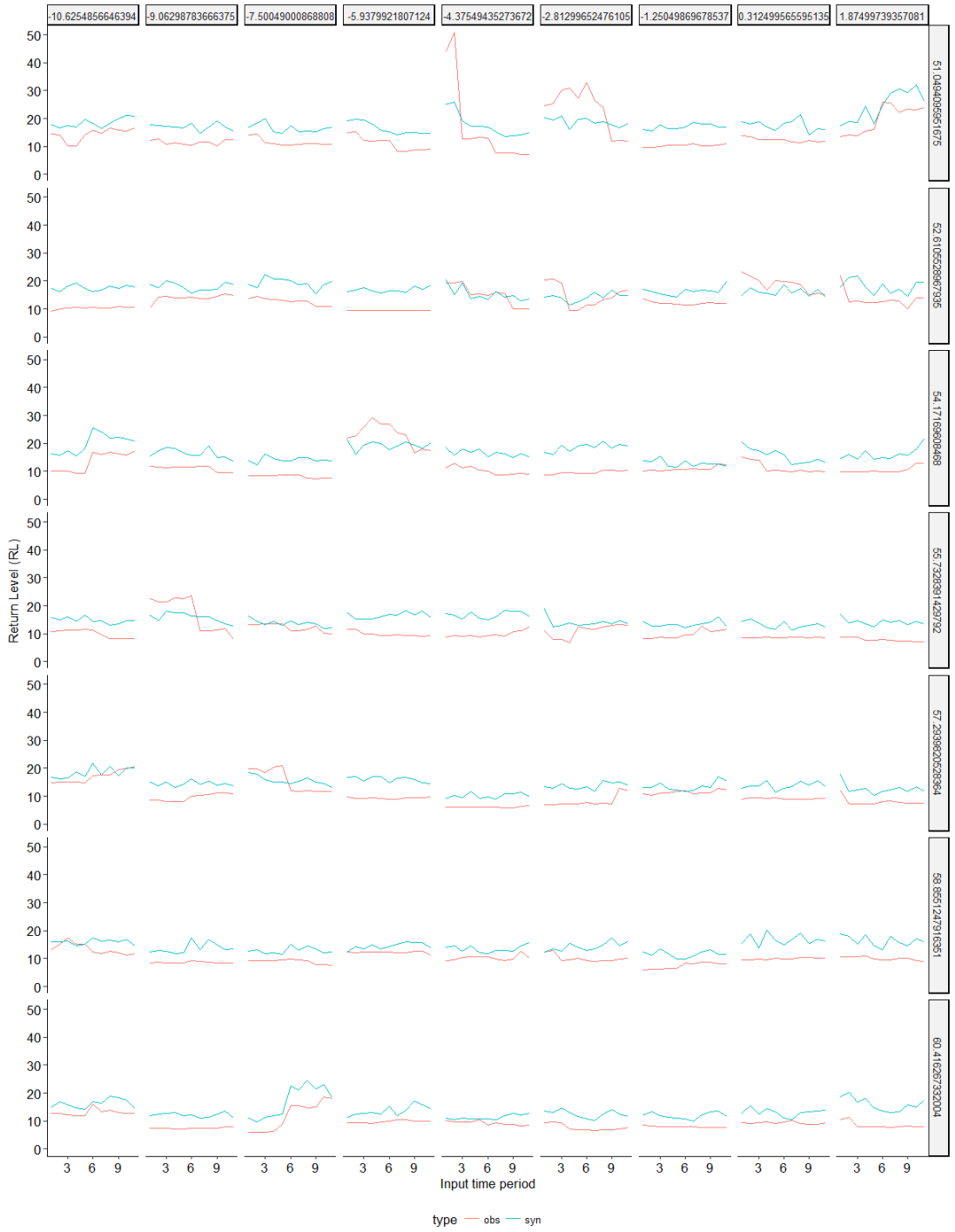


Figure 7.2: Time series of RLE of precipitation rate p [mm/h] corresponding to a return period of 100 years, for the locations defined in Fig. (6.8). The labels are the longitudes and lattitudes of the locations.

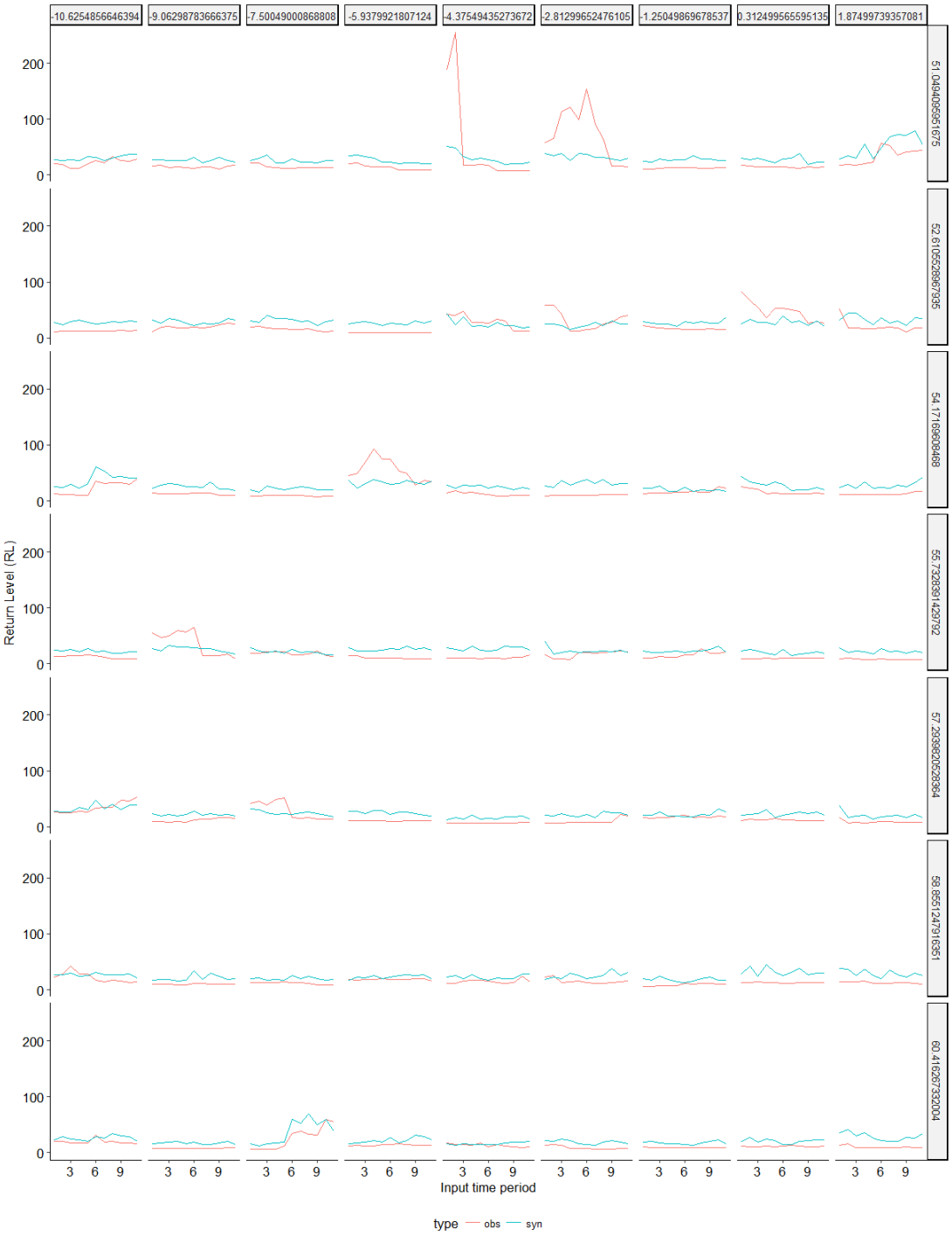


Figure 7.3: Time series of RLE of precipitation rate p [mm/h] corresponding to a return period of 1.000 years, for the locations defined in Fig. (6.8). The labels are the longitudes and lattitudes of the locations.

that a dynamic spatio-temporal generator provides spatio-temporal smoothening of RLE.

Note that the synthetic data suffers from vibrations, which implies a slight temporal instability in the synthetic data produced by the generator that should be addressed in future analyses. Moreover, the analysis presented here might reveal the effects on flood risk of changes in the real world system (e.g. non-stationarity related to climate change), but the results are also influenced by sample variability. Therefore, further tests would be needed to assess the causes of any changes shown in Figs. (7.1) to (7.3).

7.2.4. Maps of temporal smoothness

In Fig. (7.4), time series of RLE are displayed for the entire British Isles, corresponding to return periods of respectively 10, 100, 1.000 and 10.000 years for both the observed and the synthetic data. With appropriate transformations of the colourscale, it can be observed that the time series of the RLE from the synthetic data are more smooth, in line with the local time series in Figs. (7.1) to (7.3).

In Fig. (7.5), the differences in RLE between the (yearly) time steps is displayed. It can be observed that the differences between the RLE from the observed data are highly fragmented, whereas, in the differences between the RLE from the synthetic data, spatio-temporal patterns are emerging. These spatio-temporal patterns are volatile. Therefore, the results require improvement and should not yet be attempted to be interpreted. At this stage, it is difficult to detect clear trends or signatures of (multi-year) variations in these patterns, which means that it is not yet possible to link these patterns to physical processes. However, the simple fact that spatio-temporal patterns emerge, is a motivating finding. Although the patterns are not clear for the British Isles, it is likely that global spatio-temporal patterns would be more clear, which implies that in the first place this type of analysis should be carried out on the global scale, after which a finer spatial scale should be considered.

Fig. (7.6) displays two indicators of temporal smoothness/stability. In Fig. (7.6.1), it can be observed that the standard deviation of the return level estimate time series are more smooth in the synthetic data than in the observed. In the observed data, at locations where there was a few relatively small observed events, small standard deviations are found, which is not surprising. Large standard deviations can be found in areas where there are relatively big events, which correspond to the large jumps observed in the time series of the subset of locations displayed in Figs. (7.1) to (7.3). In Fig. (7.6.2), it can be observed that the mean of absolute differences between time steps shows a similar general picture as the standard deviation. However, this indicator is less sensitive to trends in RLE. It is smaller at many locations in the observed data, in particular for the RLE corresponding to the lower return periods. In addition, some areas can be identified in the synthetic data where the mean differences pick up a large signal, which probably means that the generator has some instability in these regions. This can also be observed by the noise in the signal in the time series of the subset of locations displayed in Figs. (7.1) to (7.3).

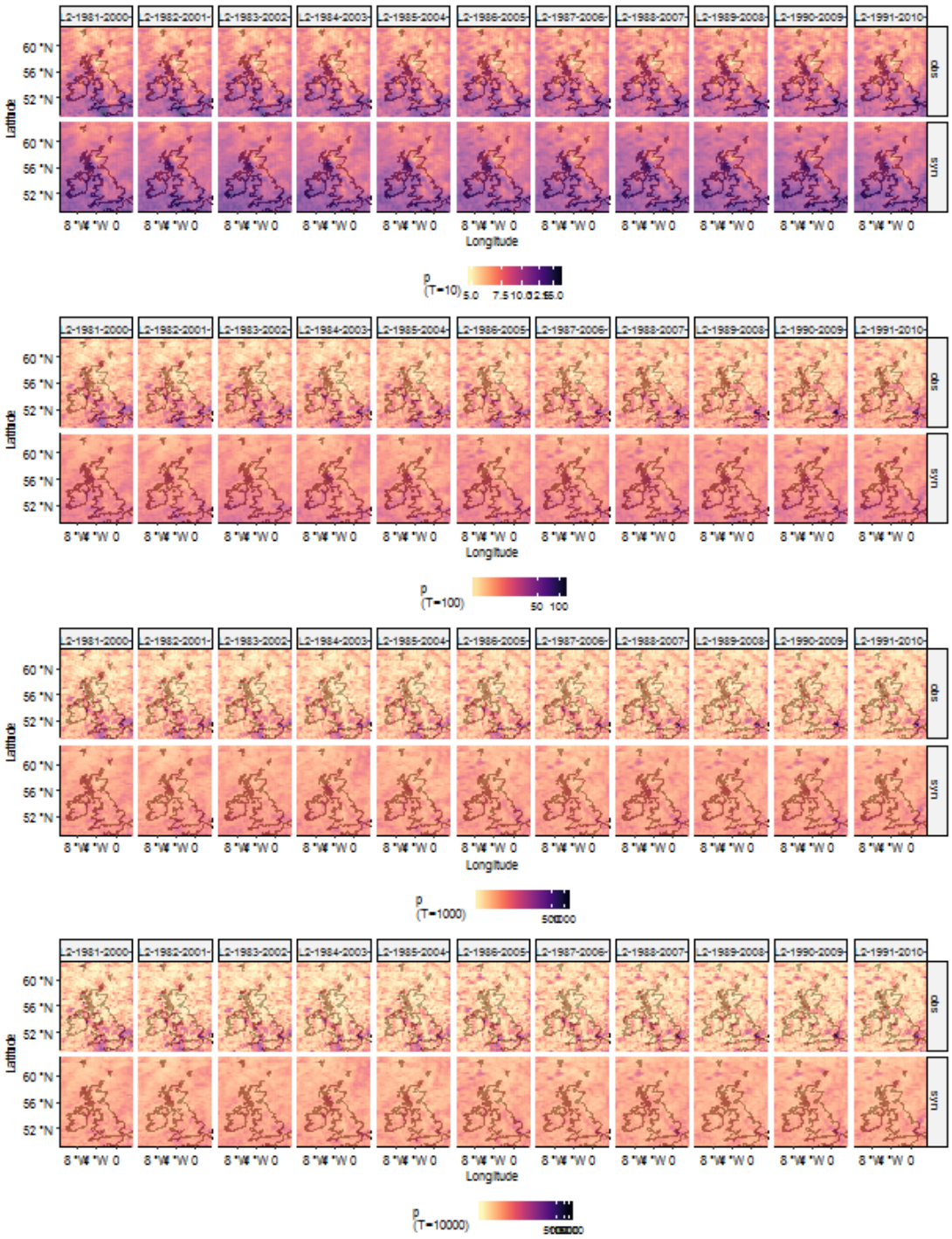


Figure 7.4: Maps of the RLE of precipitation rate p [mm/h] depicted in Figs. (7.1) to (7.3), per time interval.

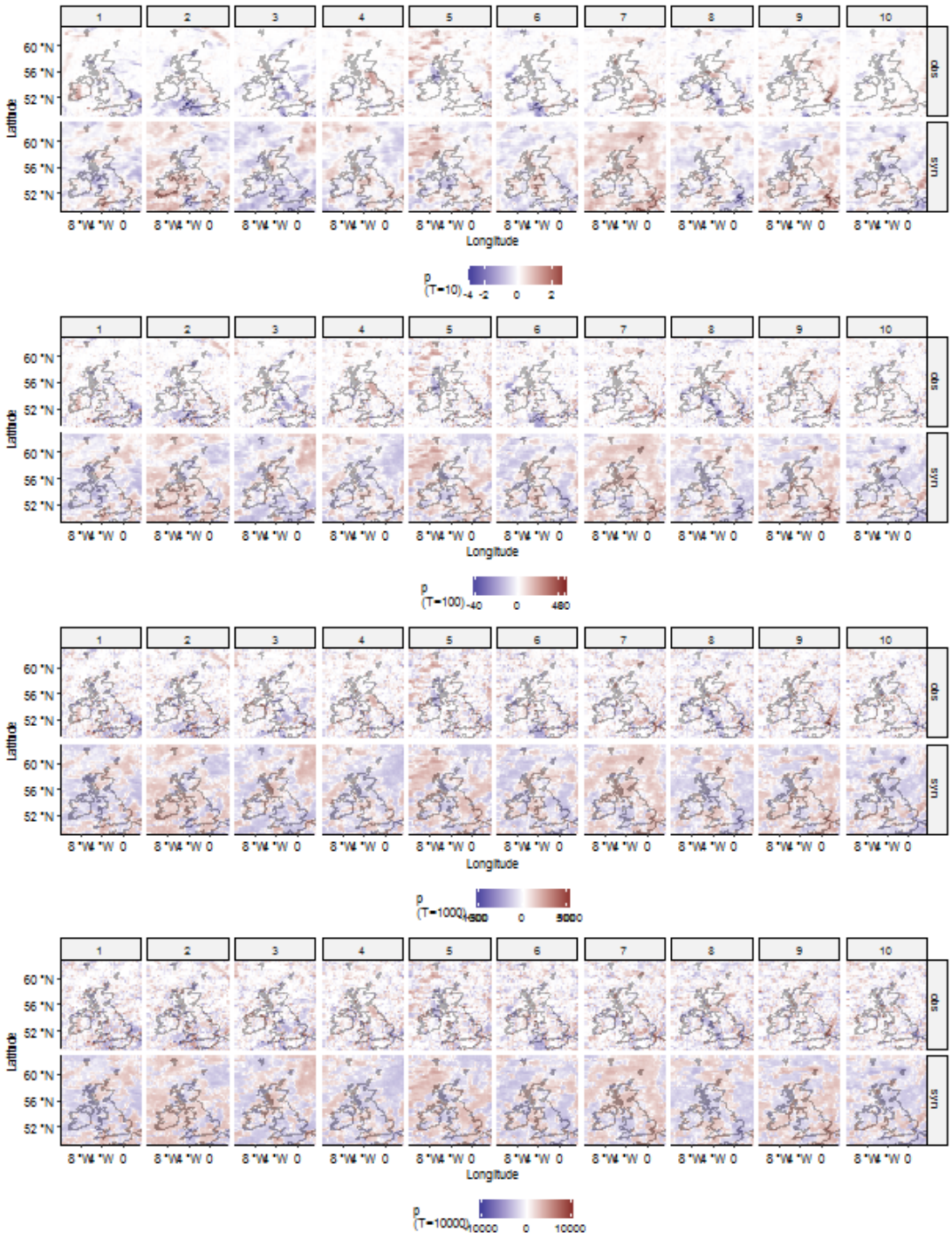
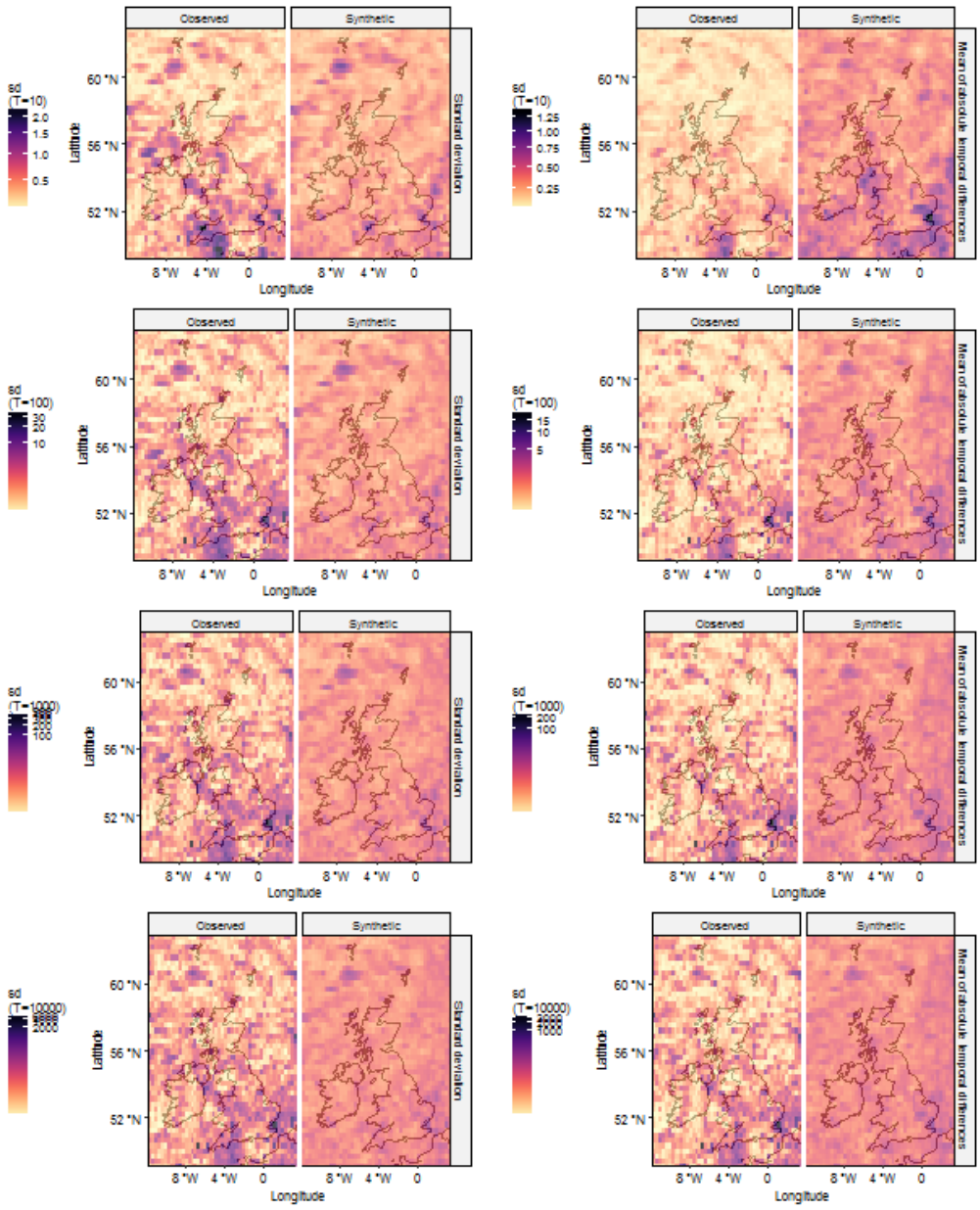


Figure 7.5: Maps of difference in RLE of precipitation rate p [mm/h] per subsequent time interval, for the maps depicted in Fig. (7.4).



(7.6.1) The standard deviation is displayed of RLE in the 11 time steps displayed in Fig. (7.4).

(7.6.2) The standard deviation is displayed of the absolute differences in RLE between timesteps displayed in Fig. (7.5).

Figure 7.6: Two indicators of temporal smoothness/stability of the time series of RLE of precipitation rate p [mm/h] displayed in Figs. (7.4) and (7.5).

7.3. Conclusions

A new generation of analyses - why we should move from static spatio-temporal analysis to dynamic spatio-temporal analysis In Chapter (3), a framework for event-based flood hazard generation was introduced. For event-based probabilistic analysis, three reasons were identified to move from static spatio-temporal probabilistic analysis to dynamic spatio-temporal probabilistic analysis. First, in Chapter (4), it was identified that for event-based static spatio-temporal probabilistic analysis a limited compatibility exists between methods of event description and statistical multivariate methods. This limitation led to the development of a new, dynamic type of generator, which was carried out in Chapter (5). Second, in Chapter (6), it was found that dynamic spatio-temporal probabilistic analysis leads to spatial smoothness of extremes, which can be considered a scientifically more rigorous method than the standard application of static spatio-temporal probabilistic analysis in combination with a spatial process. Third, it turns out that the application of dynamic spatio-temporal probabilistic analysis provides the added value of dynamic expansion of information, which may increase the credibility of probabilistic analyses by incorporating additional evidence, or, alternatively, may allow the use of a shorter temporal record for probabilistic analysis.

Outlook In Section (7.2), an outlook was provided of where this research may be going: large-scale precipitation hazard time series. In particular, the concept of dynamic expansion of information was used, which implies that, for a particular location, additional evidence could be incorporated using the dynamic type of generator. This suggested the possibility of using shorter periods of input data, which in turn enabled the estimation of hazard time series. Since the dynamic spatio-temporal generator is spatially scalable, it can be applied globally. Making use of these contributions, an outlook of global hazard time series analysis was proposed, supported by a proof-of-concept which made use of the ingredients provided in Chapters (5) and (6). Although it was not yet possible to link physical explanations to the results of this trial experiment in terms of temporal changes in the precipitation hazard, it was clear that spatio-temporal patterns started to emerge. The first next challenge would be to improve the stability of the stochastic process in time. The results suggest that global precipitation hazard time series are within reach. These global precipitation hazard time series could be used to calculate global pluvial (and, with rainfall runoff models, global fluvial) flood risk time series, which should be considered extremely relevant in the current predicament of global climate change. What is required now is a project team to improve the quality of the generator and a very big computer.

A

Appendix

A.1. Definitions

A.1.1. Flood

What is a flood? The word 'flood' is old and generally means 'flowing water' in different languages (Old-Dutch/Dutch: 'fluod/vloed'; Anglo-Saxon: 'flód'; Danish/Norwegian/Swedish: 'flod'; French: 'flot'; German: 'flut').

The old meaning of 'flowing water' is not particularly useful to understand the current usage of the word 'flood'. Unfortunately, 'flood' evolved into several directions, so that now several different definitions can be identified in the literature, used in a mixed fashion:

1. A rising *water level*. This can be a rising water level in the river (as the result of a discharge wave) or in the sea (tidal waves, flood/ebb; wind storm, surge).

2. A (high) *river discharge*. The river discharge is the water flowing within the river banks. Coupling 'flood' to high discharge is popular in the contemporary literature, where flood frequency analysis generally refers to (peaks of) discharge waves.

3. Water flowing where it normally does not. This refers to water flowing outside the river banks or coastal defence structures, i.e. *inundation depth/extent*. This definition appears to agree with the way most people understand the word flood.

Within the context of flood risk, where we are interested in damage/consequences, definition 3 would be the most useful. However, where possible, the variables of interest (*discharge, water level, inundation depth/extent*, etcetera) will be directly referred to.

A.1.2. Blocks versus events

For statistical analysis, data series are generally pre-processed to a particular set of descriptors. These descriptors describe particular clusters of data. We distinguish two generic types of clusters of data:

a. 'blocks': the entire data set is sliced into blocks of a particular temporal length,

b. 'events': the subset of interest is extracted from the data set.

Confusion may arise for several reasons:

i. If the definition of events is that they are the data clusters of interest within a particular data set, one might argue that all blocks of data are of interest, and that therefore each block constitutes an event. Here, events are defined to be a subset of the continuous data set, where the subset is not the entire data set.

ii. The fact that blocks are included in a thesis that focusses on event-based analyses. Studies based on blocks are widely found throughout the literature and therefore it was interesting to compare and contrast the methods.

iii. When block-like events are applied, like blocks of one week that are non-regularly spaced in time, which would here fall under the event-based approach. Hence, 'block-events'.

iv. Descriptions which combine *annual maximum* with *peak* are confusing, since 'annual' suggests a block-based approach, whereas 'peak' suggests an event-based approach. Generally, an annual maximum will correspond to a peak of an event, unless the event spans across two years, for example,

from 30 December to 5 January, in which case one of the two annual maxima may not be a peak. It is assumed that 'annual maximum peaks' are the largest peaks of events per year, where events are assigned to the year in which the peak occurred.

A.1.3. Return period and return level estimates

Reference is made to 'floods' (events) with a particular return period, for example, the '100-year flood'. Depending on:

- the variable of interest (*water level, discharge, inundation depth*),
- the block/event definition,
- the chosen descriptor of the block/event (*peak, volume, extent, duration*),

a particular event will have a different return period. In other words, the different descriptors of a particular flood event will have return level estimates that can be associated with different return periods. For example a particular event can have a peak that occurs once every 100 years and an extent that occurs once every 10 years. It is therefore concluded that a flood event does not have a unique return period and therefore, when using the notion of return period, it will be attempted to clearly explain which event descriptor is referred to.

Return level estimates are defined to be the values corresponding to the return periods.

A.1.4. Local probabilistic analysis

In this study, 'local probabilistic analysis' (LPA) is defined as an analysis where time series are analysed per location.

A.1.5. Static spatio-temporal probabilistic analysis

In this study, 'static spatio-temporal probabilistic analysis' (SSTPA) is defined as an analysis where the spatial extent of events is determined by the (pre-defined) locations at which data clusters, which are generally temporal blocks or local events, are described. In the literature, this type of analysis may be referred to as 'multi-site', as 'spatial-temporal' or as 'spatio-temporal'.

A.1.6. Dynamic spatio-temporal probabilistic analysis

In this study, 'dynamic spatio-temporal probabilistic analysis' (DSTPA) is defined as an analysis where the spatial extent of events: A.) is defined based on the location of occurrence of the physical phenomenon and B.) can vary in time for each event.

A.2. Acronyms

Table A.1: Acronyms used in Tables (2.2) and (2.4).

Type	Acronym	Meaning
General	Blocks	The entire data set is sliced into blocks of a particular temporal length.
	Events	The subset of interest is extracted from the data set.
Descriptors	V	Value
	M	Maximum
	P	Peak
	S	Sum
	μ	Mean
	D	Duration
	R	Radius
	SMD	Soil Moisture Deficit
	$IEAT$	Inter Event Arrival Time
	ρ	Cross-Correlation
	NW	Number of wet..
Marginal Distributions	Emp	Empirical distribution
	GP	Generalised Pareto distribution
	GEV	Generalised Extreme Value distribution
	Exp	Exponential distribution
	NST	Normal Scores Transform
Row-Wise Dependence	NP	Non-Parametric model
	AR	Auto-Regressive model
	HT04	Heffernan and Tawn 2004
Column-Wise Dependence	MC	Markov Chain
Space-Time	PP	Poisson process
	STPP	Spatio-Temporal Point Process
	NS	Neyman–Scott process
Model Fitting	L	L-moments
	ML	Maximum Likelihood

	PML	Penalised Maximum Likelihood
	AIC	Akaike Information Criterion
	SGP	Spatial Gaussian Process

Table A.2: Acronyms used in Table (3.1).

Column	**Acronym**	**Meaning**
Variables	Gen	Generation
	Val	Validation
Descriptors	V	Value
	M	Maximum
	P	Peak
	S	Sum
	μ	Mean
	D	Duration
	R	Radius
	SMD	Soil Moisture Deficit
	$IEAT$	Inter Event Arrival Time
	ρ	Cross-Correlation
	NW	Number of wet..
Classification	Loc	Location/site
Statistics		
Marginals	Emp	Empirical distribution
	GP	Generalised Pareto distribution
	GEV	Generalised Extreme Value distribution
	Exp	Exponential distribution
	NST	Normal Scores Transform
Dependence (row)	AR	Auto-Regressive model
	NP	Non-Parametric model
	HT04	Heffernan and Tawn 2004
Dependence (column)	MC	Markov Chain
Space-Time	PP	Poisson process
	STPP	Spatio-Temporal Point Process
	NS	Neyman–Scott process

Model Fitting	L	L-moments
	ML	Maximum Likelihood
	PML	Penalised Maximum Likelihood
	AIC	Akaike Information Criterion
	SGP	Spatial Gaussian Process

A.3. Assumptions Heffernan and Tawn (2004)

Denote the multivariate distribution with standard Laplace marginal distribution

$$\mathbf{Y} = \{Y_1, Y_2, ..., Y_d\} \tag{A.1}$$

Denote the sub multivariate distribution

$$\mathbf{Y_{-i}} = \{Y_1, Y_2, ..., Y_{i-1}, Y_{i+1}, ..., Y_d\} \tag{A.2}$$

The Heffernan-Tawn model assumes that vector functions a and b, both $R \rightarrow R^{(d-1)}$, exist such that the standardized variable

$$\mathbf{Z}|(Y_i = y_i) = \frac{\mathbf{Y_{-i}} - \mathbf{a}(y_i)}{\mathbf{b}(y_i)} \tag{A.3}$$

has the property that

$$\mathbf{lim}_{y_i \rightarrow \infty} \mathbf{Pr}\{\mathbf{Z} \leq \mathbf{z}|(Y_i = y_i)\} = \mathbf{G}(\mathbf{z}) \tag{A.4}$$

where the limit distribution G has non-degenerate marginal distributions. It follows from this assumption that, conditional on Y_i being larger than a fixed threshold u, as the threshold $u \rightarrow \infty$ variables Y_i and Z are independent for all Y_i. Here the same standardizing vector functions as Heffernan and Tawn [46] have been used, where

$$\mathbf{a}(y) = (a_1 y, a_2 y, ..., a_{i-1} y, a_{i+1} y, ..., a_d y)^T \tag{A.5}$$

and

$$\mathbf{b}(y) = (y^{b_1}, y^{b_2}, ..., y^{b_{i-1}}, y^{b_{i+1}}, ..., y^{b_d})^T \tag{A.6}$$

A.4. Implementation: big data system

A.4.1. Devices

	deviceNr	device	nCores	memtotal
1	1	compute05.hrw-uk.local	12	47.16
2	2	compute06.hrw-uk.local	12	47.16
3	3	compute08.hrw-uk.local	12	47.16
4	5	compute19.hrw-uk.local	16	62.87
5	6	compute20.hrw-uk.local	16	62.87
6	7	compute21.hrw-uk.local	16	62.87
7	8	compute22.hrw-uk.local	16	62.87
8	9	compute27.hrw-uk.local	28	62.81
9	10	compute28.hrw-uk.local	28	62.81

Table A.3: Devices used

A.4.2. Computational times

	scriptNr	script
1	1	CFSR/R_scripts/runtime/CFSR_all.R
2	2	CFSR/R_scripts/runtime/CFSR_inspection.R
3	3	R_scripts/generator/1_event_identification/1_eventsSeparate.R
4	4	R_scripts/generator/1_event_identification/2_eventsImageAnalysis.R
5	5	R_scripts/generator/1_event_identification/3a_eventsCluster_hours_stitch.R
6	6	R_scripts/generator/1_event_identification/3b_eventsCluster_months.R
7	7	R_scripts/generator/1_event_identification/3c_eventsCluster_compound.R
8	8	R_scripts/generator/2_event_description/1_featuresAll.R
9	9	R_scripts/generator/2_event_description/2_classification_som_hclust.R
10	10	R_scripts/generator/3_statistics/1_GPD.R
11	11	R_scripts/generator/3_statistics/2_transforms.R
12	12	R_scripts/generator/3_statistics/3_HT04.R
13	13	R_scripts/generator/4_simulation_features/1_simFaster.R
14	14	R_scripts/generator/4_simulation_features/2_varAnalysed.R
15	15	R_scripts/generator/5_reconstruction_events/1_reconstruct_events.R
16	16	R_scripts/generator/5_reconstruction_events/2_eventsPerVar.R
17	17	R_scripts/generator/6_reconstruction_continuous/1_reconstruct_contin.R
18	18	R_scripts/inspection/2_observed_versus_synthetic/0_data/0_arrayObs_to_timeSeries_hydra.R
19	19	R_scripts/inspection/2_observed_versus_synthetic/0_data/1_arraySyn_to_timeSeries_hydra.R
20	20	R_scripts/inspection/2_observed_versus_synthetic/0_data/2_merge_timeSeries_synSets_hydra.R
21	21	R_scripts/inspection/2_observed_versus_synthetic/0_data/3_subsetLoc.R
22	22	R_scripts/inspection/2_observed_versus_synthetic/1_local/1_2_3_b_returnlevels_POT_GPD.R
23	23	R_scripts/inspection/2_observed_versus_synthetic/1_local/4_correlation_temporal.R
24	24	R_scripts/inspection/2_observed_versus_synthetic/1_local/5_event_coverage.R
25	25	R_scripts/inspection/2_observed_versus_synthetic/2_spatial/1_spatial_correlations.R
26	26	R_scripts/inspection/2_observed_versus_synthetic/2_spatial/2_smoothness_images.R

Table A.4: Script numbers

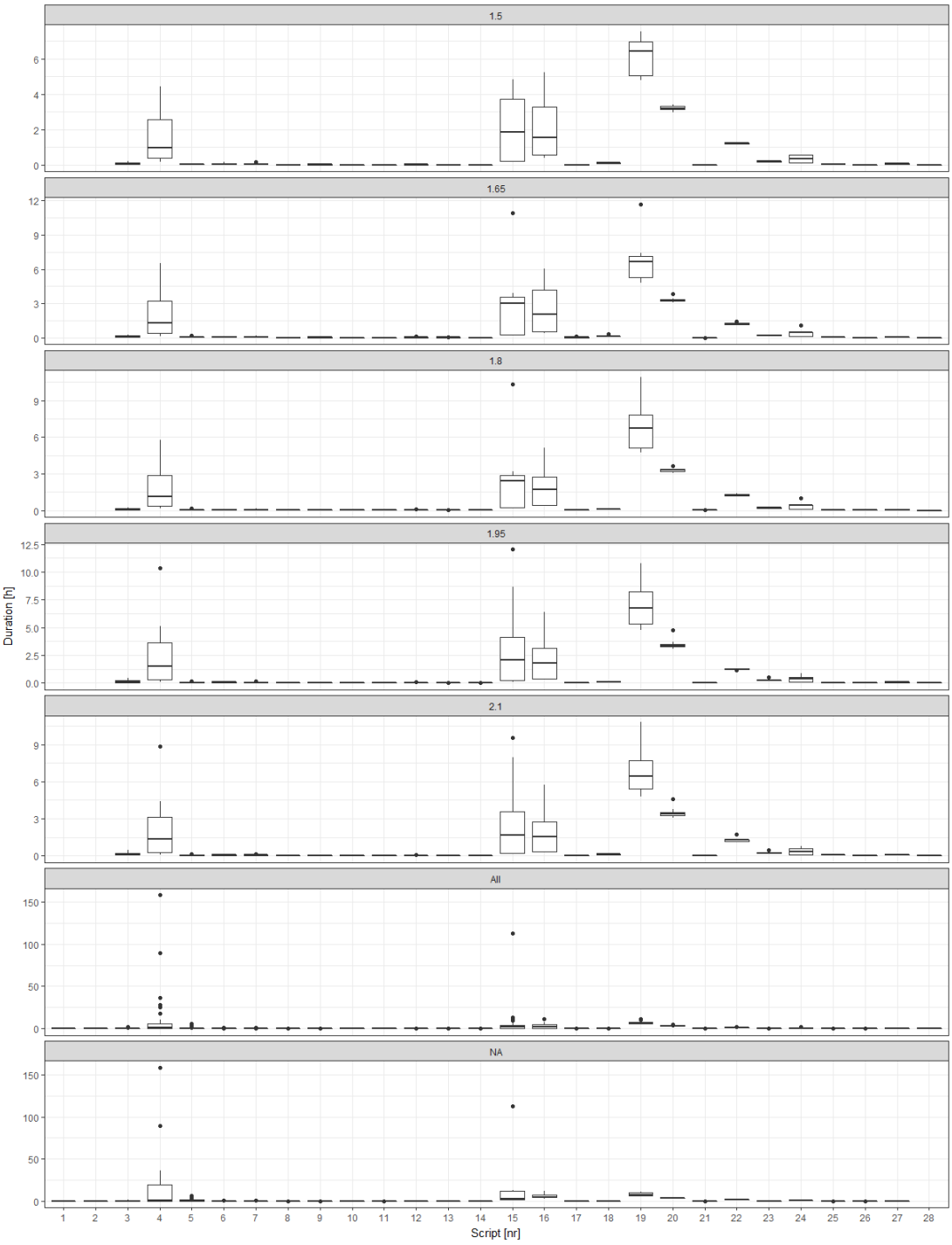


Figure A.1: Computational times for the different analyses performed (in Tables (A.5) to (A.7)), per script (in Table (A.4)).

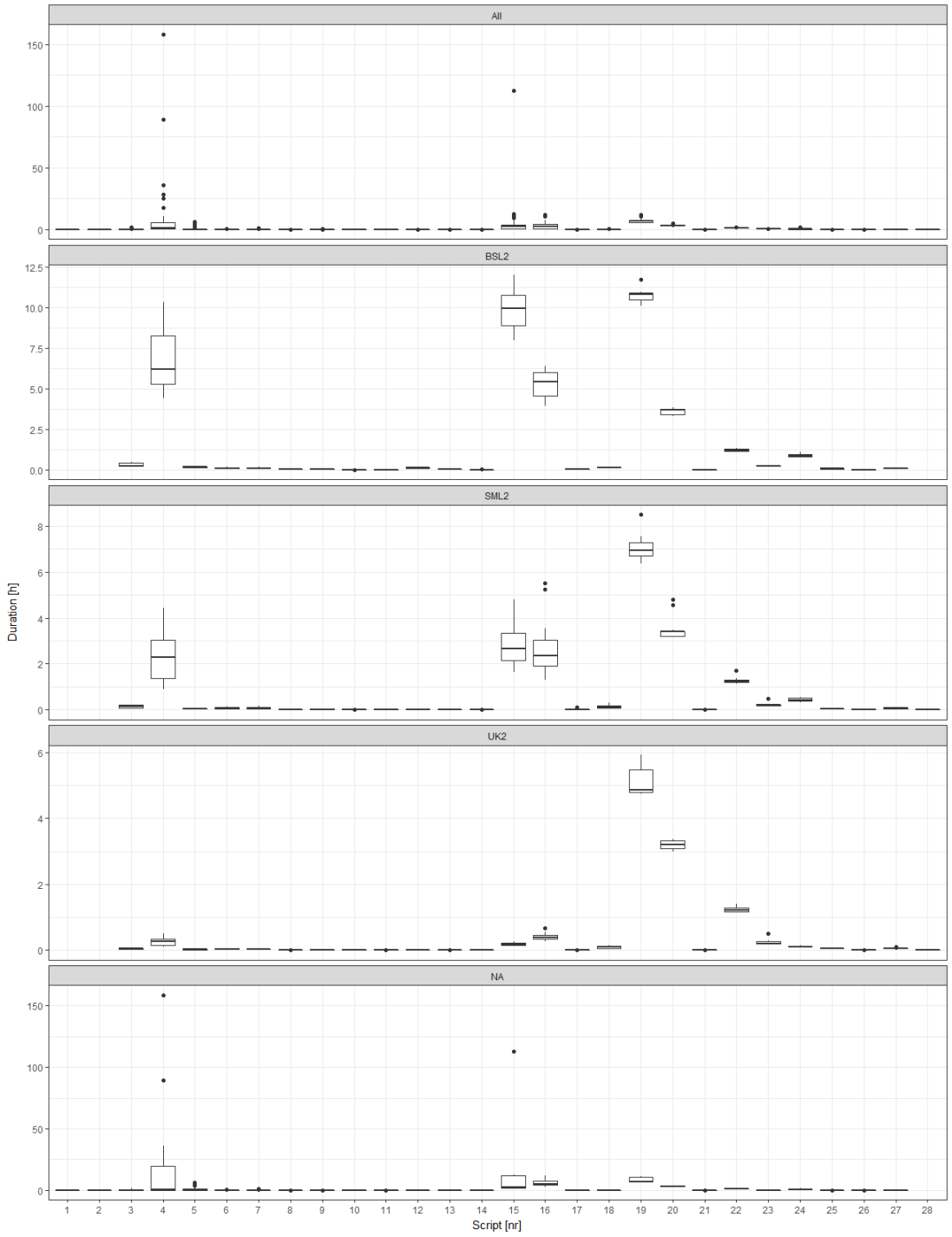


Figure A.2: Computational times for the different analyses performed (in Tables (A.5) to (A.7)), per script (in Table (A.4)).

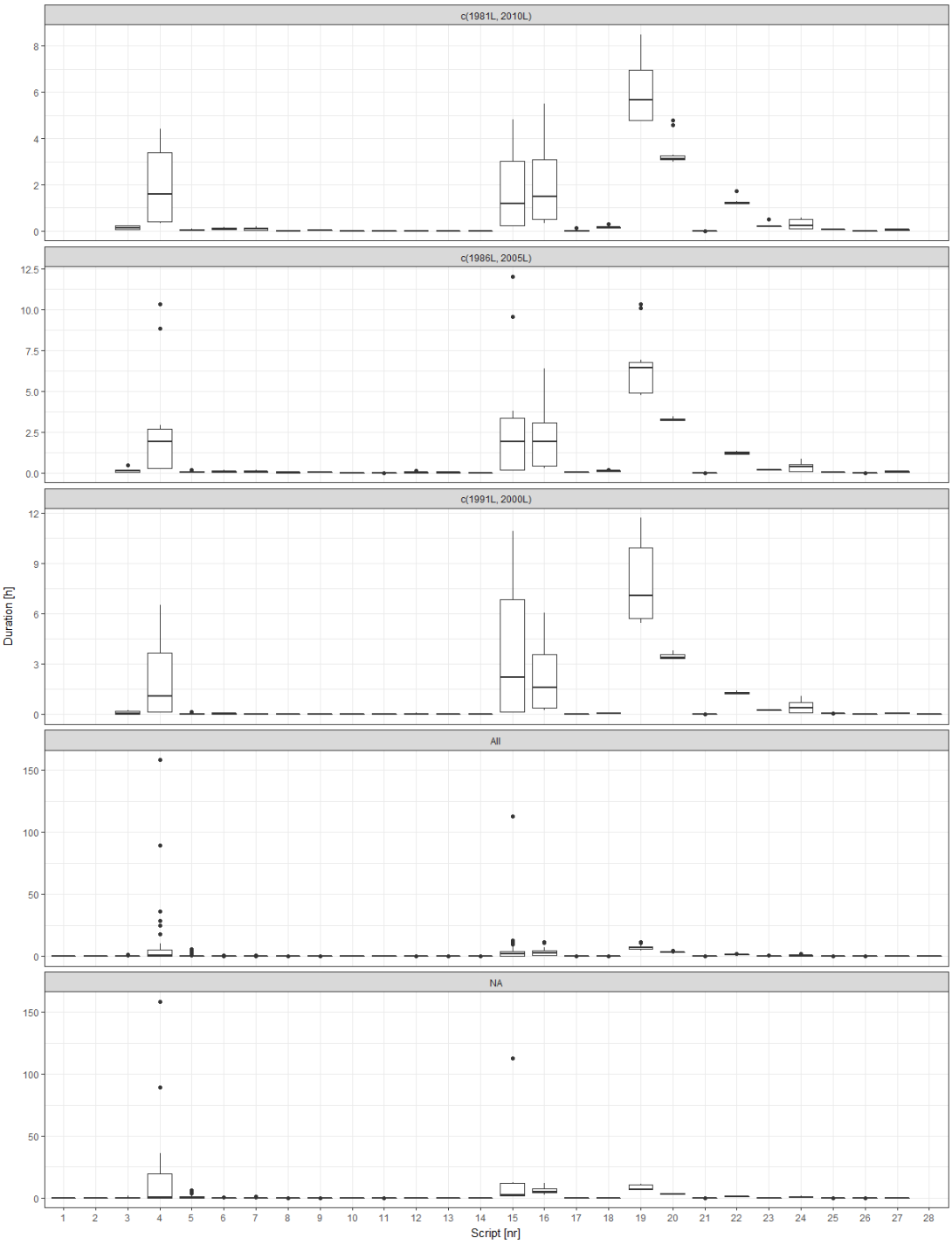


Figure A.3: Computational times for the different analyses performed (in Tables (A.5) to (A.7)), per script (in Table (A.4)).

A.4.3. Settings

	analV	variables	domains.local	domains.ST	settings.years	settings.months	settings.nSynSets
1	"V1-BSL2-1991-2000-1.5-R4"	"prate"	"UK2"	"BSL2"	c(1991L, 2000L)	c(1L, 12L)	100L
2	"V1-BSL2-1991-2000-1.65-R4"	"prate"	"UK2"	"BSL2"	c(1991L, 2000L)	c(1L, 12L)	100L
3	"V1-BSL2-1991-2000-1.8-R4"	"prate"	"UK2"	"BSL2"	c(1991L, 2000L)	c(1L, 12L)	100L
4	"V1-BSL2-1991-2000-1.95-R4"	"prate"	"UK2"	"BSL2"	c(1991L, 2000L)	c(1L, 12L)	100L
5	"V1-BSL2-1991-2000-2.1-R4"	"prate"	"UK2"	"BSL2"	c(1991L, 2000L)	c(1L, 12L)	100L
6	"V1-SML2-1991-2000-1.5-R4"	"prate"	"UK2"	"SML2"	c(1991L, 2000L)	c(1L, 12L)	100L
7	"V1-SML2-1991-2000-1.65-R4"	"prate"	"UK2"	"SML2"	c(1991L, 2000L)	c(1L, 12L)	100L
8	"V1-SML2-1991-2000-1.8-R4"	"prate"	"UK2"	"SML2"	c(1991L, 2000L)	c(1L, 12L)	100L
9	"V1-SML2-1991-2000-1.95-R4"	"prate"	"UK2"	"SML2"	c(1991L, 2000L)	c(1L, 12L)	100L
10	"V1-SML2-1991-2000-2.1-R4"	"prate"	"UK2"	"SML2"	c(1991L, 2000L)	c(1L, 12L)	100L
11	"V1-UK2-1991-2000-1.5-R4"	"prate"	"UK2"	"UK2"	c(1991L, 2000L)	c(1L, 12L)	100L
12	"V1-UK2-1991-2000-1.65-R4"	"prate"	"UK2"	"UK2"	c(1991L, 2000L)	c(1L, 12L)	100L
13	"V1-UK2-1991-2000-1.8-R4"	"prate"	"UK2"	"UK2"	c(1991L, 2000L)	c(1L, 12L)	100L
14	"V1-UK2-1991-2000-1.95-R4"	"prate"	"UK2"	"UK2"	c(1991L, 2000L)	c(1L, 12L)	100L
15	"V1-UK2-1991-2000-2.1-R4"	"prate"	"UK2"	"UK2"	c(1991L, 2000L)	c(1L, 12L)	100L
16	"V1-BSL2-1986-2005-1.5-R4"	"prate"	"UK2"	"BSL2"	c(1986L, 2005L)	c(1L, 12L)	50L
17	"V1-BSL2-1986-2005-1.65-R4"	"prate"	"UK2"	"BSL2"	c(1986L, 2005L)	c(1L, 12L)	50L
18	"V1-BSL2-1986-2005-1.8-R4"	"prate"	"UK2"	"BSL2"	c(1986L, 2005L)	c(1L, 12L)	50L
19	"V1-BSL2-1986-2005-1.95-R4"	"prate"	"UK2"	"BSL2"	c(1986L, 2005L)	c(1L, 12L)	50L
20	"V1-BSL2-1986-2005-2.1-R4"	"prate"	"UK2"	"BSL2"	c(1986L, 2005L)	c(1L, 12L)	50L
21	"V1-SML2-1986-2005-1.5-R4"	"prate"	"UK2"	"SML2"	c(1986L, 2005L)	c(1L, 12L)	50L
22	"V1-SML2-1986-2005-1.65-R4"	"prate"	"UK2"	"SML2"	c(1986L, 2005L)	c(1L, 12L)	50L
23	"V1-SML2-1986-2005-1.8-R4"	"prate"	"UK2"	"SML2"	c(1986L, 2005L)	c(1L, 12L)	50L
24	"V1-SML2-1986-2005-1.95-R4"	"prate"	"UK2"	"SML2"	c(1986L, 2005L)	c(1L, 12L)	50L
25	"V1-SML2-1986-2005-2.1-R4"	"prate"	"UK2"	"SML2"	c(1986L, 2005L)	c(1L, 12L)	50L
26	"V1-UK2-1986-2005-1.5-R4"	"prate"	"UK2"	"UK2"	c(1986L, 2005L)	c(1L, 12L)	50L
27	"V1-UK2-1986-2005-1.65-R4"	"prate"	"UK2"	"UK2"	c(1986L, 2005L)	c(1L, 12L)	50L
28	"V1-UK2-1986-2005-1.8-R4"	"prate"	"UK2"	"UK2"	c(1986L, 2005L)	c(1L, 12L)	50L
29	"V1-UK2-1986-2005-1.95-R4"	"prate"	"UK2"	"UK2"	c(1986L, 2005L)	c(1L, 12L)	50L
30	"V1-UK2-1986-2005-2.1-R4"	"prate"	"UK2"	"UK2"	c(1986L, 2005L)	c(1L, 12L)	50L
31	"V1-BSL2-1981-2010-1.5-R4"	"prate"	"UK2"	"BSL2"	c(1981L, 2010L)	c(1L, 12L)	34L
32	"V1-BSL2-1981-2010-1.65-R4"	"prate"	"UK2"	"BSL2"	c(1981L, 2010L)	c(1L, 12L)	34L
33	"V1-BSL2-1981-2010-1.8-R4"	"prate"	"UK2"	"BSL2"	c(1981L, 2010L)	c(1L, 12L)	34L
34	"V1-BSL2-1981-2010-1.95-R4"	"prate"	"UK2"	"BSL2"	c(1981L, 2010L)	c(1L, 12L)	34L
35	"V1-BSL2-1981-2010-2.1-R4"	"prate"	"UK2"	"BSL2"	c(1981L, 2010L)	c(1L, 12L)	34L
36	"V1-SML2-1981-2010-1.5-R4"	"prate"	"UK2"	"SML2"	c(1981L, 2010L)	c(1L, 12L)	34L
37	"V1-SML2-1981-2010-1.65-R4"	"prate"	"UK2"	"SML2"	c(1981L, 2010L)	c(1L, 12L)	34L
38	"V1-SML2-1981-2010-1.8-R4"	"prate"	"UK2"	"SML2"	c(1981L, 2010L)	c(1L, 12L)	34L
39	"V1-SML2-1981-2010-1.95-R4"	"prate"	"UK2"	"SML2"	c(1981L, 2010L)	c(1L, 12L)	34L
40	"V1-SML2-1981-2010-2.1-R4"	"prate"	"UK2"	"SML2"	c(1981L, 2010L)	c(1L, 12L)	34L
41	"V1-UK2-1981-2010-1.5-R4"	"prate"	"UK2"	"UK2"	c(1981L, 2010L)	c(1L, 12L)	34L
42	"V1-UK2-1981-2010-1.65-R4"	"prate"	"UK2"	"UK2"	c(1981L, 2010L)	c(1L, 12L)	34L
43	"V1-UK2-1981-2010-1.8-R4"	"prate"	"UK2"	"UK2"	c(1981L, 2010L)	c(1L, 12L)	34L
44	"V1-UK2-1981-2010-1.95-R4"	"prate"	"UK2"	"UK2"	c(1981L, 2010L)	c(1L, 12L)	34L
45	"V1-UK2-1981-2010-2.1-R4"	"prate"	"UK2"	"UK2"	c(1981L, 2010L)	c(1L, 12L)	34L

Table A.5: Settings sensitivity - part 1

	methods.sep	settings.prate.threshSep	methods.clust	methods.class	settings.featClassify	settings.nEventsPerClass
1	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
2	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
3	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
4	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
5	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
6	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
7	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
8	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
9	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
10	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
11	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
12	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
13	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
14	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
15	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	250L
16	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
17	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
18	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
19	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
20	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
21	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
22	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
23	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
24	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
25	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
26	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
27	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
28	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
29	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
30	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	500L
31	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
32	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
33	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
34	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
35	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
36	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
37	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
38	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
39	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
40	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
41	"threshGlob"	1.5	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
42	"threshGlob"	1.65	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
43	"threshGlob"	1.8	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
44	"threshGlob"	1.95	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L
45	"threshGlob"	2.1	"polygon"	"som"	c("ratPeakVol", "cogV_t", "cog_x", "cog_y")	1000L

Table A.6: Settings sensitivity - part 2

	settings.mqu	settings.dqu	settings.pqu	methods.rec	settings.featReconstruct
1	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
2	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
3	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
4	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
5	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
6	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
7	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
8	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
9	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
10	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
11	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
12	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
13	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
14	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
15	0.8	0.8	0.9	"R4"	c("volume", "peak", "extent")
16	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
17	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
18	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
19	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
20	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
21	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
22	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
23	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
24	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
25	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
26	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
27	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
28	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
29	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
30	0.85	0.85	0.9	"R4"	c("volume", "peak", "extent")
31	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
32	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
33	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
34	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
35	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
36	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
37	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
38	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
39	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
40	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
41	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
42	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
43	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
44	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")
45	0.9	0.9	0.9	"R4"	c("volume", "peak", "extent")

Table A.7: Settings sensitivity - part 3

References

[1] Pierre Ailliot et al. "Stochastic weather generators: an overview of weather type models". In: *Journal de la Société Française de Statistique* 156.1 (2015), pp. 101–113.

[2] Lorenzo Alfieri et al. "Advances in pan-European flood hazard mapping". In: *Hydrological Processes* 28.13 (2014), pp. 4067–4077. URL: https://doi.org/10.1002/hyp.9947.

[3] Konstantinos M Andreadis et al. "Twentieth-Century Drought in the Conterminous United States". In: *Journal of Hydrometeorology* 6.6 (2005), pp. 985–1001. URL: https://doi.org/10.1175/JHM450.1.

[4] Pragalathan Apputhurai and Alec G Stephenson. "Spatiotemporal hierarchical modelling of extreme precipitation in Western Australia using anisotropic Gaussian random fields". In: *Environmental and Ecological Statistics* 20.4 (2013), pp. 667–677. URL: https://doi.org/10.1007/s10651-013-0240-9.

[5] A Bárdossy and G.G.S. Pegram. "Copula based multisite model for daily precipitation simulation". In: *Hydrology and Earth System Sciences* 13.12 (2009), pp. 2299–2314. URL: https://doi.org/10.5194/hess-13-2299-2009.

[6] Matt Basinger, Franco Montalto, and Upmanu Lall. "A rainwater harvesting system reliability model based on nonparametric stochastic rainfall generator". In: *Journal of Hydrology* 392.3 (2010), pp. 105–118. URL: https://doi.org/10.1016/j.jhydrol.2010.07.039.

[7] Jules J Beersma and T Adri Buishand. "Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation". In: *Climate Research* 25.2 (2003), pp. 121–133. URL: https://doi.org/10.3354/cr025121.

[8] L Berthet et al. "How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments". In: *Hydrology and Earth System Sciences* 13.6 (2009), pp. 819–831. URL: https://doi.org/10.5194/hess-13-819-2009.

[9] Konstantinos Bischiniotis et al. "The influence of antecedent conditions on flood risk in sub-Saharan Africa". In: *Natural Hazards and Earth System Sciences* 18.1 (2018), pp. 271–285. URL: https://doi.org/10.5194/nhess-18-271-2018.

[10] Edoardo Borgomeo, Christopher L Farmer, and Jim W Hall. "Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments". In: *Water Resources Research* 51.7 (2015), pp. 5382–5405. URL: https://doi.org/10.1002/2014WR016827.

[11] Paola Bortot, Stuart Coles, and Jonathan Tawn. "The multivariate Gaussian tail model: an application to oceanographic data". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 49.1 (2000), pp. 31–049. URL: https://doi.org/10.1111/1467-9876.00177.

[12] W Boughton and O Droop. "Continuous simulation for design flood estimation—a review". In: *Environmental Modelling & Software* 18.4 (2003), pp. 309–318. URL: https://doi.org/10.1016/S1364-8152(03)00004-5.

[13] Cameron Bracken et al. "Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain". In: *Water Resources Research* 52.8 (2016), pp. 6643–6655. URL: https://doi.org/10.1002/2016WR018768.

[14] M. I. Brunner and E. Gilleland. "Stochastic simulation of streamflow and spatial extremes: a continuous, wavelet-based approach". In: *Hydrology and Earth System Sciences* 24.8 (2020), pp. 3967–3982. URL: https://doi.org/10.5194/hess-24-3967-2020.

[15] A Burton et al. "RainSim: A spatial–temporal stochastic rainfall modelling system". In: *Environmental Modelling & Software* 23.12 (2008), pp. 1356–1369. URL: https://doi.org/10.1016/j.envsoft.2008.04.003.

[16] D.S Cameron et al. "Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty)". In: *Journal of Hydrology* 219.3 (1999), pp. 169–187. URL: https://doi.org/10.1016/S0022-1694(99)00057-8.

[17] D. Cameron et al. "Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation)". In: *Hydrology and Earth System Sciences* 4.1 (2000), pp. 23–34. URL: https://doi.org/10.5194/hess-4-23-2000.

[18] J.E. Chacón, T. Duong, and M.P. Wand. "Asymptotics for general multivariate kernel density derivative estimators". In: *Statistica Sinica* 21.2 (2011), pp. 807–840. URL: https://doi.org/10.5705/ss.2011.036a.

[19] Jie Chen and François P Brissette. "Comparison of five stochastic weather generators in simulating daily precipitation and temperature for the Loess Plateau of China". In: *International Journal of Climatology* 34.10 (2014), pp. 3089–3105. URL: https://doi.org/10.1002/joc.3896.

[20] Stuart Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001. URL: https://doi.org/10.1007/978-1-4471-3675-0.

[21] Daniel Cooley, Douglas Nychka, and Philippe Naveau. "Bayesian spatial modeling of extreme precipitation return levels". In: *Journal of the American Statistical Association* 102.479 (2007), pp. 824–840. URL: https://doi.org/10.1198/016214506000000780.

[22] GA Corzo Perez et al. “On the spatio-temporal analysis of hydrological droughts from global hydrological models”. In: *Hydrology and Earth System Sciences* 15.9 (2011), pp. 2963–2978. URL: https://doi.org/10.5194/hess-15-2963-2011.

[23] Paul SP Cowpertwait. “A generalized spatial-temporal model of rainfall based on a clustered point process”. In: *Proceedings of the R. Soc. Lond. A* 450.1938 (1995), pp. 163–175. URL: https://doi.org/10.1098/rspa.1995.0077.

[24] Paul SP Cowpertwait. “A spatial-temporal point process model with a continuous distribution of storm types”. In: *Water Resources Research* 46.12 (2010). URL: https://doi.org/10.1029/2010WR009728.

[25] Paul Cowpertwait et al. “Regionalised spatiotemporal rainfall and temperature models for flood studies in the Basque Country, Spain”. In: *Hydrology and Earth System Sciences* 17 (2013). URL: https://doi.org/10.5194/hess-17-479-2013.

[26] Anthony C Davison, Simone A Padoan, and Mathieu Ribatet. “Statistical Modeling of Spatial Extremes”. In: *Statistical Science* 27.2 (2012), pp. 161–186. URL: https://doi.org/10.1214/11-STS376.

[27] B. Desai et al. *Making Development Sustainable: The Future of Disaster Risk Management, Global Assessment Report on Disaster Risk Reduction*. eng. 910; 333.7-333.9. ID: unige:78299. United Nations Office for Disaster Risk Reduction (UNISDR), 2015. URL: https://archive-ouverte.unige.ch/unige:78299.

[28] Dirk Diederen et al. “Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment”. In: *Natural Hazards and Earth System Sciences* 19.5 (2019), pp. 1041–1053. URL: https://doi.org/10.5194/nhess-19-1041-2019.

[29] Francesco Dottori et al. “Development and evaluation of a framework for global flood hazard mapping”. In: *Advances in water resources* 94 (2016), pp. 87–102. URL: https://doi.org/10.1016/j.advwatres.2016.05.002.

[30] P. S. Eagleson. “Dynamics of flood frequency”. In: *Water Resources Research* 8.4 (1972), pp. 878–898. URL: https://doi.org/10.1029/WR008i004p00878.

[31] Emma F Eastoe and Jonathan A Tawn. “Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series”. In: *Water resources research* 46.2 (2010). URL: https://doi.org/10.1029/2009WR007757.

[32] Brian S Everitt, Graham Dunn, et al. *Applied multivariate data analysis*. Vol. 2. Wiley Online Library, 2001. URL: https://doi.org/10.1002/9781118887486.

[33] Daniela Falter et al. “Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain”. In: *Journal of Hydrology* 524 (2015), pp. 182–193. URL: https://doi.org/10.1016/j.jhydrol.2015.02.021.

[34] Christopher B Field. *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press, 2012. URL: https://doi.org/.

[35] Svenja Fischer, Andreas Schumann, and Philipp Bühler. "A statistics-based automated flood event separation". In: *Journal of Hydrology X* 10 (2021), p. 100070. ISSN: 2589–9155. DOI: https://doi.org/10.1016/j.hydroa.2020.100070. URL: https://www.sciencedirect.com/science/article/pii/S2589915520300213.

[36] HJ Fowler et al. "A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change". In: *Journal of Hydrology* 308.1-4 (2005), pp. 50–66. URL: https://doi.org/10.1016/j.jhydrol.2004.10.021.

[37] Eva M. Furrer and Richard W. Katz. "Improving the simulation of extreme precipitation events by stochastic weather generators". In: *Water Resources Research* 44.12 (2008). URL: https://doi.org/10.1029/2008WR007316.

[38] Christian Genest and Anne-Catherine Favre. "Everything you always wanted to know about copula modeling but were afraid to ask". In: *Journal of hydrologic engineering* 12.4 (2007), pp. 347–368. URL: https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347).

[39] Souparno Ghosh and Bani K Mallick. "A hierarchical Bayesian spatio-temporal model for extreme precipitation events". In: *Environmetrics* 22.2 (2011), pp. 192–204. URL: https://doi.org/10.1002/env.1043.

[40] Eric Gilleland, Douglas Nychka, and Uli Schneider. "Spatial models for the distribution of extremes". In: *Hierarchical modelling for the environmental sciences: statistical methods and applications*. Oxford University Press on Demand, 2006, pp. 170–183. URL: https://doi.org/10.1086/523192.

[41] Vassilis Glenis et al. "A transient stochastic weather generator incorporating climate model uncertainty". In: *Advances in water resources* 85 (2015), pp. 14–26. URL: https://doi.org/j.advwatres.2015.08.002.

[42] Ben Gouldby et al. "Multivariate extreme value modelling of sea conditions around the coast of England". In: *Proceedings of the Institution of Civil Engineers - Maritime Engineering* 170.1 (2017), pp. 3–20. URL: https://doi.org/10.1680/jmaen.2016.16.

[43] Jim Hall and Dimitri Solomatine. "A framework for uncertainty analysis in flood risk management decisions". In: *International Journal of River Basin Management* 6.2 (2008), pp. 85–98. URL: https://doi.org/10.1080/15715124.2008.9635339.

[44] K Haslinger and G Blöschl. "Space-time patterns of meteorological drought events in the European Greater Alpine Region over the past 210 years". In: *Water Resources Research* 53.11 (2017), pp. 9807–9823. URL: https://doi.org/10.1002/2017WR020797.

[45] Charles Hebson and Eric F. Wood. "A derived flood frequency distribution using Horton Order Ratios". In: *Water Resources Research* 18.5 (1982), pp. 1509–1518. URL: https://doi.org/10.1029/WR018i005p01509.

[46] Janet E Heffernan and Jonathan A Tawn. "A conditional approach for multivariate extreme values (with discussion)". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66.3 (2004), pp. 497–546. URL: https://doi.org/10.1111/j.1467-9868.2004.02050.x.

[47] Yeshewatesfa Hundecha, Markus Pahlow, and Andreas Schumann. "Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes". In: *Water resources research* 45.12 (2009). URL: https://doi.org/10.1029/2008WR007453.

[48] Bart van den Hurk et al. "Analysis of a compounding surge and precipitation event in the Netherlands". In: *Environmental Research Letters* 10.3 (2015), p. 035001. URL: https://doi.org/10.1088/1748-9326/10/3/035001.

[49] Institute Of Hydrology. *Flood estimation handbook*. 1999. URL: https://doi.org/.

[50] Richard W Katz, Marc B Parlange, and Philippe Naveau. "Statistics of extremes in hydrology". In: *Advances in water resources* 25.8-12 (2002), pp. 1287–1304. URL: https://doi.org/10.1016/S0309-1708(02)00056-8.

[51] Caroline Keef, Ioannis Papastathopoulos, and Jonathan A Tawn. "Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model". In: *Journal of Multivariate Analysis* 115 (2013), pp. 396–404. URL: https://doi.org/10.1016/j.jmva.2012.10.012.

[52] Caroline Keef, Jonathan A Tawn, and Rob Lamb. "Estimating the probability of widespread flood events". In: *Environmetrics* 24.1 (2013), pp. 13–21. URL: https://doi.org/10.1002/env.2190.

[53] Caroline Keef, Jonathan Tawn, and Cecilia Svensson. "Spatial risk assessment for extreme river flows". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 58.5 (2009), pp. 601–618. URL: https://doi.org/10.1111/j.1467-9876.2009.00672.x.

[54] C Keef et al. *The risk of widespread flooding - Capturing spatial patterns in flood risk from rivers and coasts.* 2011.

[55] C Kidd et al. "Intercomparison of high-resolution precipitation products over northwest Europe". In: *Journal of Hydrometeorology* 13.1 (2012), pp. 67–83. URL: https://doi.org/10.1175/JHM-D-11-042.1.

[56] Teuvo Kohonen. *Self-organization and associative memory*. Vol. 8. Springer Science & Business Media, 2012. URL: https://doi.org/10.1007/978-3-642-88163-3.

[57] Rob Lamb et al. "A new method to assess the risk of local and widespread flooding on rivers and coasts". In: *Journal of Flood Risk Management* 3.4 (2010), pp. 323–336. URL: https://doi.org/10.1111/j.1753-318X.2010.01081.x.

[58] Anthony W Ledford and Jonathan A Tawn. "Statistics for near independence in multivariate extreme values". In: *Biometrika* 83.1 (1996), pp. 169–187. URL: https://doi.org/10.1093/biomet/83.1.169.

[59] Michael Leonard et al. "A compound event framework for understanding extreme impacts". In: *Wiley Interdisciplinary Reviews: Climate Change* 5.1 (2014), pp. 113–128. URL: https://doi.org/10.1002/wcc.252.

[60] R Mehrotra, R Srikanthan, and Ashish Sharma. "A comparison of three stochastic multi-site precipitation occurrence generators". In: *Journal of Hydrology* 331.1-2 (2006), pp. 280–292. URL: https://doi.org/10.1016/j.jhydrol.2006.05.016.

[61] Ralf Merz and Günter Blöschl. "Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information". In: *Water Resources Research* 44.8 (2008). URL: https://doi.org/10.1029/2007WR006745.

[62] Roger B Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007. URL: https://doi.org/10.1007/0-387-28678-0.

[63] Paul J Northrop, Nicolas Attalides, and Philip Jonathan. "Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 66.1 (2017), pp. 93–120. URL: https://doi.org/10.1111/rssc.12159.

[64] Florian Pappenberger et al. "Deriving global flood hazard maps of fluvial floods through a physical model cascade". In: *Hydrology and Earth System Sciences* 16.11 (2012), pp. 4143–4156. URL: https://doi.org/10.5194/hess-16-4143-2012.

[65] Dominik Paprotny and Oswaldo Morales-Nápoles. "Estimating extreme river discharges in Europe through a Bayesian network". In: *Hydrology and Earth System Sciences* 21.6 (2017), pp. 2615–2636. URL: https://doi.org/10.5194/hess-21-2615-2017.

[66] Dominik Paprotny, Oswaldo Morales-Nápoles, and Sebastiaan N Jonkman. "Efficient pan-European river flood hazard modelling through a combination of statistical and physical models". In: *Natural Hazards and Earth System Sciences* 17.7 (2017), p. 1267. URL: https://doi.org/10.5194/nhess-17-1267-2017.

[67] S Pathiraja, S Westra, and A Sharma. "Why continuous simulation? The role of antecedent moisture in design flood estimation". In: *Water Resources Research* 48.6 (2012). URL: https://doi.org/10.1029/2011WR010997.

[68] Niall Quinn et al. "The spatial dependence of flood hazard and risk in the United States". In: *Water Resources Research* (2019). URL: https://doi.org/10.1029/2018WR024205.

[69] N. Ridder, H. de Vries, and S. Drijfhout. "The role of atmospheric rivers in compound events consisting of heavy precipitation and high storm surges along the Dutch coast". In: *Natural Hazards and Earth System Sciences* 18.12 (2018), pp. 3311–3326. URL: https://doi.org/10.5194/nhess-18-3311-2018.

[70] RIZA Rijkswaterstaat, KNMI Marcel de Wit, and Adri Buishand. "Generator of Rainfall And Discharge Extremes (GRADE) for the Rhine and Meuse basins". In: (2007).

[71] Ignacio Rodriguez-Iturbe, David Roxbee Cox, and Valerie Isham. "Some models for rainfall based on stochastic point processes". In: *Proc. R. Soc. Lond. A* 410.1839 (1987), pp. 269–288. URL: https://doi.org/10.1098/rspa.1987.0039.

[72] Suranjana Saha et al. "The NCEP climate forecast system reanalysis". In: *Bulletin of the American Meteorological Society* 91.8 (2010), pp. 1015–1058. URL: https://doi.org/10.1175/2010BAMS3001.1.

[73] Suranjana Saha et al. *NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010*. Boulder CO, 2010. URL: https://doi.org/10.5065/D6513W89.

[74] Andrea Saltelli et al. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[75] P Samuels. "Flooding and multi-hazards". In: *Journal of Flood Risk Management* 11 (2018), S557–S558. URL: https://doi.org/10.1111/jfr3.12341.

[76] Carl Scarrott and Anna MacDonald. "A review of extreme value threshold estimation and uncertainty quantification". In: *REVSTAT–Statistical Journal* 10.1 (2012), pp. 33–60.

[77] Klaus Schneeberger and Thomas Steinberger. "Generation of Spatially Heterogeneous Flood Events in an Alpine Region—Adaptation and Application of a Multivariate Modelling Procedure". In: *Hydrology* 5.1 (2018), p. 5. URL: https://doi.org/10.3390/hydrology5010005.

[78] Francesco Serinaldi. "Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective". In: *Stochastic environmental research and risk assessment* 23.5 (2009), pp. 677–693. URL: https://doi.org/10.1007/s00477-008-0249-z.

[79] Francesco Serinaldi and Chris G Kilsby. "Simulating daily rainfall fields over large areas for collective risk estimation". In: *Journal of Hydrology* 512 (2014), pp. 285–302. URL: https://doi.org/10.1016/j.jhydrol.2014.02.043.

[80] Francesco Serinaldi and Chris G Kilsby. "A blueprint for full collective flood risk estimation: demonstration for European river flooding". In: *Risk Analysis* 37.10 (2017), pp. 1958–1976. URL: https://doi.org/10.1111/risa.12747.

[81] Paul Sharkey and Hugo C Winter. "A Bayesian spatial hierarchical model for extreme precipitation in Great Britain". In: *Environmetrics* 30.1 (2017), e2529. URL: https://doi.org/10.1002/env.2529.

[82] Justin Sheffield et al. “Global and continental drought in the second half of the twentieth century: Severity–area–duration analysis and temporal variability of large-scale events”. In: *Journal of Climate* 22.8 (2009), pp. 1962–1981. URL: https://doi.org/10.1175/2008JCLI2722.1.

[83] Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018. URL: https://doi.org/10.1201/9781315140919.

[84] R Srikanthan and TA McMahon. “Stochastic generation of annual, monthly and daily climate data: A review”. In: *Hydrology and Earth System Sciences* 5.4 (2001), pp. 653–670. URL: https://doi.org/10.5194/hess-5-653-2001.

[85] Ratnasingham Srikanthan and Geoffrey GS Pegram. “A nested multisite daily rainfall stochastic generation model”. In: *Journal of Hydrology* 371.1-4 (2009), pp. 142–153. URL: https://doi.org/10.1016/j.jhydrol.2009.03.025.

[86] Cecilia Svensson, Thomas R Kjeldsen, and David A Jones. “Flood frequency estimation using a joint probability approach within a Monte Carlo framework”. In: *Hydrological sciences journal* 58.1 (2013), pp. 8–27. URL: https://doi.org/10.1080/02626667.2012.746780.

[87] Larisa Tarasova et al. “Causative classification of river flood events”. In: *Wiley Interdisciplinary Reviews: Water* 0.0 (2019), e1353. URL: https://doi.org/10.1002/wat2.1353.

[88] L Tarasova et al. “Exploring controls on rainfall-runoff events: 1. Time-series-based event separation and temporal dynamics of event runoff response in Germany”. In: *Water Resources Research* (2018). URL: https://doi.org/10.1029/2018WR022587.

[89] Pramodh Vallam and Xiao Sheng Qin. “Multi-site rainfall simulation at tropical regions: a comparison of three types of generators”. In: *Meteorological Applications* 23.3 (2016), pp. 425–437. URL: https://doi.org/10.1002/met.1567.

[90] JM Van Der Knijff, Jalal Younis, and APJ De Roo. “LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation”. In: *International Journal of Geographical Information Science* 24.2 (2010), pp. 189–212. URL: https://doi.org/10.1080/13658810802549154.

[91] Andrew Verdin et al. “Coupled stochastic weather generation using spatial and generalized linear models”. In: *Stochastic environmental research and risk assessment* 29.2 (2015), pp. 347–356. URL: https://doi.org/10.1007/s00477-014-0911-6.

[92] J-P Vidal et al. “Evolution of spatio-temporal drought characteristics: validation, projections and effect of adaptation scenarios”. In: *Hydrology and Earth System Sciences* 16.8 (2012), pp. 2935–2955. URL: https://doi.org/10.5194/hess-16-2935-2012.

[93] Sergiy Vorogushyn et al. “Evolutionary leap in large-scale flood risk assessment needed”. In: *Wiley Interdisciplinary Reviews: Water* 5.2 (2018). URL: https://doi.org/10.1002/wat2.1266.

[94] S Vorogushyn et al. “A new methodology for flood hazard assessment considering dike breaches”. In: *Water Resources Research* 46.8 (2010). URL: https://doi.org/10.1029/2009WR008475.

[95] Michalis I Vousdoukas. “Developments in large-scale coastal flood hazard mapping”. In: *Natural Hazards and Earth System Sciences* 16.8 (2016), p. 1841. URL: https://doi.org/10.5194/nhess-16-1841-2016.

[96] Tue M Vu et al. “Evaluation of multiple stochastic rainfall generators in diverse climatic regions”. In: *Stochastic Environmental Research and Risk Assessment* 32.5 (2018), pp. 1337–1353. URL: https://doi.org/10.1007/s00477-017-1458-0.

[97] Philip J Ward et al. “Assessing flood risk at the global scale: model setup, results, and sensitivity”. In: *Environmental Research Letters* 8.4 (2013), p. 044019. URL: https://doi.org/10.1088/1748-9326/8/4/044019.

[98] RL Wilby, OJ Tomlinson, and CW Dawson. “Multi-site simulation of precipitation by conditional resampling”. In: *Climate Research* 23.3 (2003), pp. 183–194. URL: https://doi.org/10.3354/cr023183.

[99] Daniel S. Wilks. “Chapter 9 Methods for multivariate data”. In: *Statistical Methods in the Atmospheric Sciences*. Vol. 59. International Geophysics. Academic Press, 1995, pp. 359–428. URL: https://doi.org/10.1016/S0074-6142(06)80045-8.

[100] Daniel S Wilks and Robert L Wilby. “The weather generation game: a review of stochastic weather models”. In: *Progress in physical geography* 23.3 (1999), pp. 329–357. URL: https://doi.org/10.1177/030913339902300302.

[101] DS Wilks. “Multisite generalization of a daily stochastic precipitation generation model”. In: *Journal of Hydrology* 210.1-4 (1998), pp. 178–191. URL: https://doi.org/10.1016/S0022-1694(98)00186-3.

[102] H. C. Winsemius et al. “A framework for global river flood risk assessments”. In: *Hydrology and Earth System Sciences* 17.5 (2013), pp. 1871–1892. URL: https://doi.org/10.5194/hess-17-1871-2013.

[103] B. Winter et al. “Event generation for probabilistic flood risk modelling: multi-site peak flow dependence model vs. weather-generator-based approach”. In: *Natural Hazards and Earth System Sciences* 20.6 (2020), pp. 1689–1703. URL: https://doi.org/10.5194/nhess-20-1689-2020.

[104] B Winter et al. “Sources of uncertainty in a probabilistic flood risk model”. In: *Natural Hazards* 91.2 (2018), pp. 431–446. URL: https://doi.org/10.1007/s11069-017-3135-5.

[105] D Wyncoll and B Gouldby. “Integrating a multivariate extreme value method within a system flood risk analysis model”. In: *Journal of Flood Risk Management* 8.2 (2015), pp. 145–160. URL: https://doi.org/10.1111/jfr3.12069.

[106] Qiang Zhang et al. “Copula-based spatio-temporal patterns of precipitation extremes in China”. In: *International Journal of Climatology* 33.5 (2013), pp. 1140–1152. URL: https://doi.org/10.1002/joc.3499.