Prof. Jim Freer

Editor

Hydrology and Earth System Sciences

May 18th, 2015

Dear Prof. Freer,

We would like to acknowledge the revision of our work entitled "**Propagation of hydro-meteorological uncertainty in a model cascade framework to inundation prediction**". Again, we thank you for your constructive comments.

We have digested the concerns from the reviewer and added some new information to clarify these points. We believe that this new version of our manuscript is indeed better and thank the reviewers and yourself for your effort and time in this revision. In the following lines we explain how (i.e. by writing our reply in red) and where (i.e. by giving line numbers) the raised points have been addressed in the revised manuscript. We hope that this new version proves to be worth for publication in HESS.

Best wishes,

Dr. Adrián Pedrozo-Acuña on behalf of all authors

Editor Decision: Publish subject to minor revisions (Editor review)

Comments to the Author: The paper is now getting close to be acceptable for publication. However the 2nd reviewer has identified some omissions and also some inconsistencies which need to be addressed before this happens. These should not add too much to the paper length but the justification for differences in model skill has to be scientifically justified and the overall aims of the uncertainty cascade and what it does and doesn't do. So if the authors can address these points, respond and complete the modifications necessary we should be able to move to publication, best wishes, Jim

Reviewer 3: Propagation of hydro-meteorological uncertainty in a model cascade framework to inundation prediction – REVIEW

This paper considers the propagation of uncertainty through a cascading model system, linking a Numerical Weather Prediction model with hydrological and 2D hydrodynamic models. The paper is well written and the topic will be of interest to a wide ranging audience, although I am not entirely sure what specifically this work contributes to scientific progress. **This should be more clearly specified by the authors**.

R: In order to provide a clearer specification of the contribution of our work, we have modified the abstract and conclusions to highlight its overall purpose.

In the abstract, we have acknowledged this fact in the following manner, at Page 1 in particular adding the sentences shown here in bold:

"This investigation aims to study the propagation of meteorological uncertainty within a cascade modelling approach to flood prediction. The methodology was comprised of a Numerical Weather Prediction Model (NWP), a distributed rainfall-runoff model and a 2D hydrodynamic model. **The quantification of uncertainty was carried out in a hindcast scenario, removing non-behavioural ensemble members at each stage, based on the fit with observed data.** The selected extreme event corresponds to a flood that took place in the Southeast of Mexico during November 2009, for which field data (e.g. rain gauges; discharge) and satellite imagery were available. Uncertainty in the meteorological model was estimated by means of a multi-physics ensemble technique, which considers variations in the specific setup options to determine a given precipitation. Precipitation fields from the meteorological model were employed as input in a distributed hydrological model, and resulting flood hydrographs were used as forcing conditions in the 2D hydrodynamic model. **This enabled the assessment of uncertainty and its propagation, from a modelled rainfall event to a predicted flooded area and depth.** Moreover, the evolution of skill within the model cascade shows a complex aggregation of errors between models, suggesting that in valley-filling events hydro-meteorological uncertainty has a larger effect on inundation depths than that observed in estimated flood inundation extents."

In this resubmission, the paper has been substantially improved and the authors have addressed many of the previous reviewers' comments adequately, although I do have some questions:

The research aims to quantify uncertainty in a hindcast scenario, removing non-behavioural ensemble members at each stage based on the fit with observed data. In the first instance (NWP predictions), a Nash Sutcliffe (NS) value of >0.3 is accepted as behavioural, while the hydrographs were rejected if the score fell below 0.6. How were these limits defined? Justification should be given, particularly as the choices that are made will have a significant influence on the perceived uncertainty in the model chain.

R: We thank the reviewer for this comment. It should be noted that the change in the spatial scale from the meteorological model to the distributed hydrological model, involves some sort of downscaling issue. Transferring information from a large scale (atmospheric domain) to a smaller scale (catchment), involves downscaling and in this sense, we modified the model performance criteria for the hydrologic model to consider only the members with NSC>0.6. The more relaxed criteria used at the NWP stage is thought in order to incorporate predictions with a wide range of skill, from which the error may be propagated to both the hydrologic and hydrodynamic scale of the model chain.

The uncertainty in the hydrological model parameters are defined by calibrating the model to a series of past events. Some more information would be useful. For instance, what rainfall input was used during this calibration?

R: The utilised rainfall input corresponds to that recorded for those events at the same 4 weather stations that are within the river catchment, which location is shown in the top panel of Figure 1. This has been acknowledged in Page 11 Line 30 to Page 12 Line 2.

Also, what was the advantage in defining 6 sets of parameter values from various events rather than simply using the 2009 event and accepting any parameter sets that provided hydrographs that lay within the specified threshold? This is particularly relevant as some of the calibrated NS scores were very poor (e.g. 0.155), while also the 2009 event was significantly larger than any of the others.

R: In the hydrological model, the definition of six sets of plausible parameters from past flood events (Table 3) is thought to reduce the dimensionality of the parameter calibration problem (see Gupta et al., 2009). This procedure was preferred over a GLUE analysis, as the investigation was aimed to the understanding of the propagation of uncertainty along the model chain. In other words, to evaluate how an error originated in the meteorological model propagates to the definition of an inundated area and depth. Although simplistic, it is important to report on how an error originated in the first model of the chain is propagated to a result of interest to decision-makers (flood map).

On the other hand, it is reflected that the use of six sets of parameters from past flood events enable a multi-response validation, which in due course allow the assessment of the overall modelling performance. Additionally, it should be borne in mind that there are too many sources of uncertainty in the modelling process within a cascade of models that cannot easily be disaggregated. Thus, it is necessary to make assumptions about how to represent uncertainty, and there are sufficient degrees of freedom in doing so, such that different methods based on different types of assumptions (including purely qualitative evaluations) cannot easily be accepted or rejected.

There is no representation of uncertainty in the hydrodynamic model. This feels like a fairly major omission given the attempt to establish a framework for quantifying uncertainty in extreme events. There are many sources of uncertainty in hydrodynamic models, and I feel the exclusion of all of them needs some further justification. Alternatively, could sensible parameter ranges be estimated using a Monte Carlo approach, rejecting parameter ranges based on NS scores as done for the other model components?

R: As the reviewer points out, there are many sources of uncertainty that arise in producing fluvial flood risk maps. Some of these have to do with the natural variability in the occurrence of floods; others have more to do with the limited knowledge available about the nature of flood runoff and flood wave propagation including the geometry and infrastructure of flood plains.

Indeed, several investigations confirm that there is significant uncertainty associated with flood extent predictions using hydraulic models (e.g. Aronica et al., 1998, 2002; Bates et al. 2004; Pappenberger et al., 2005, 2006, 2007; Romanowicz and Beven, 2003). These uncertainties may be ascribed to differences in spatio-temporal resolutions used in the numerical model, or the hydraulic roughness that is determined for the river and flood plain; and little guidance exists on the magnitude of such effects. This opens the door to complex questions of scaling and dimensionality. However, in our study, a more detailed consideration of the different sources of uncertainty in the hydraulic model was not pursued, this is due to the fact that the numerical setup of the hydraulic model is built following published guidelines for an accurate representation of our problem (see Asselman et al. 2008).

For instance, high quality topographic and bathymetric data were employed (LiDAR derived DEM and field survey) for the construction of the numerical representation of both, the river and the floodplain. It is reflected that this enables us to build the discussion on how an uncertainty generated at the meteorological stage of the model chain propagates and influences a resulting flooded area and depth.

This justification has been included in the manuscript at Page 15 Line 1 – Line 10.

Complementary references that have been included:

Aronica, G, Hankin, B.G., Beven, K.J., 1998, Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, Advances in Water Resources, 22(4), 349-365

Aronica, G., Bates, P.D. and Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrological Processes, 16, 2001- 2016.

Asselman, N., Bates P., Woodhead S., Fewtrell T., Soares-Frazão S., Zech Y., Velickovic M., de Wit A., ter Maat J., Verhoeven G., Lhomme J. 2008. Flood Inundation Modelling – Model Choice and Proper Application, Report T08-09-03, FLOODsite Project.

Bates, P. D., Horritt, M. S., Aronica, G. and Beven, K J, 2004, Bayesian updating of flood inundation likelihoods conditioned on flood extent data, Hydrological Processes, 18, 3347-3370

Pappenberger, F., Beven, K.J., Hunter N., Gouweleeuw, B., Bates, P., de Roo, A., Thielen, J., 2005, Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfallrunoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). Hydrology and Earth System Science, 9(4), 381-393.

Pappenberger, F, Matgen, P, Beven, K J, Henry J-B, Pfister, L and de Fraipont, P, 2006, Influence of uncertain boundary conditions and model structure on flood inundation predictions, Advances in Water Resources, 29(10), 1430-1449,doi:10.1016/j.advwatres.2005.11.012

Pappenberger, F., Beven, K.J., Frodsham, K., Romanovicz, R. and Matgen, P., 2007. Grasping the unavoidable subjectivity in calibration of flood inundation models: a vulnerability weighted approach. Journal of Hydrology, 333, 275-287.

Romanowicz, R. and Beven, K. J., 2003, Bayesian estimation of flood inundation probabilities as conditioned on event inundation maps, Water Resources Research, 39(3), W01073, 10.1029/2001WR001056.

A tidal boundary is mentioned briefly in the site description, however, no further information is provided. Is this boundary condition influential to the model? How was this boundary calculated?

R: The astronomical tide (microtidal in nature with tidal range <1 m) is determined using the monthly tidal forecast at a nearby point, published by CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada) for those dates (http://predmar.cicese.mx/calmen.php).

This information has been included in the manuscript at Page 14 Line 13 – Line 16

Propagation of hydro-meteorological uncertainty in a model cascade framework to inundation prediction

J. P. Rodríguez-Rincón[1], A. Pedrozo-Acuña[1*] and J. A. Breña-Naranjo[1]

[1]{National Autonomous University of México, Institute of Engineering, D.F., Mexico}

*Correspondence to: A. Pedrozo-Acuña (APedrozoA@ii.unam.mx)

Abstract

~~The purpose of this~~This investigation ~~is~~aims to study the propagation of meteorological uncertainty within a cascade modelling approach to flood ~~mapping~~prediction. The methodology was comprised of a Numerical Weather Prediction Model (NWP), a distributed rainfall-runoff model and a ~~standard~~ 2D hydrodynamic model. The ~~cascade~~quantification of ~~models is used to reproduce an~~ uncertainty was carried out in a hindcast scenario, removing non-behavioural ensemble members at each stage, based on the fit with observed data. The selected extreme event corresponds to a flood ~~event~~ that took place in the Southeast of Mexico~~,~~ during November 2009~~. The event was selected as high quality~~, for which field data (e.g. rain gauges; discharge) and satellite imagery ~~are~~were available. Uncertainty in the meteorological model ~~(Weather Research and Forecasting model)~~ was ~~evaluated through the use~~estimated by means of a multi-physics ensemble technique, which considers variations in the specific setup options to determine a given precipitation~~ event. The resulting precipitation~~. Precipitation fields ~~are used~~from the meteorological model were employed as input in a distributed hydrological model, ~~enabling the determination of different hydrographs associated to this event. Lastly, by means of a standard 2D hydrodynamic model,~~ and resulting flood hydrographs ~~are~~were used as forcing conditions ~~to study~~in the ~~propagation of the meteorological~~ 2D hydrodynamic model. This enabled the assessment of uncertainty ~~to an estimated inundation~~and its propagation, from a modelled rainfall event to a predicted flooded area~~. Results show the utility of the selected modelling approach to investigate error propagation within a cascade of models~~ and depth. Moreover, the evolution of skill within the model cascade shows a complex aggregation of errors between models, suggesting that in valley-filling events hydro-meteorological

uncertainty ~~affects~~has a larger effect on inundation depths ~~in a higher degree~~ than that observed in estimated flood inundation extents.

1 Introduction

Hydro-meteorological hazards can have cascading effects and far-reaching implications on water security, with political, social, economic and environmental consequences. Millions of people worldwide are forcibly displaced as a result of natural disasters, creating political tensions and social needs to support them. These events observed in developed and developing nations alike, highlight the necessity to generate a better understanding on what causes them and how we can better manage and reduce the risk.

The assessment of flood risk is an activity that has to be carried out under a framework full of uncertainty. The source of these uncertainties may be ascribed to the involvement of different, and often rather complex models and tools, in the context of environmental conditions that are at best, partially understood (Hall, 2014). In addition to this, flooding events are dynamic over a range of timescales, due to climate variability and socio-economic changes, among others, which further increases the uncertainty in the projections. Therefore, numerous types of uncertainties can arise when using formal models in the analysis of risks.

Uncertainty is often categorised between aleatory and epistemic (Hacking, 2006): aleatory is an essential, unavoidable unpredictability, and epistemic uncertainty reflects lack of knowledge or the inadequacy of the models to represent reality. In the context of any modelling framework, epistemic uncertainties may be ascribed to the definition of model parameters and to the model structure itself (limited knowledge).

In a technological era characterised by the advent of computers, there is an increased ability of more detailed hydrological and hydraulic models. Their use and development has been motivated as they are based on equations that have (more or less) physical justification; and allow a more detailed spatial representation of the processes, parameters and predicted variables (Beven, 2014). However, there are also disadvantages, these numerical tools take more computer time and require the definition of initial, boundary conditions and parameter values in space and time. Generally, at a level of detail for which such information is not available even in research studies. Moreover, these models may be subjected to numerical problems such as numerical difussion and instability. All of these disadvantages can be interpreted as sources of uncertainty in the modelling process.

Due to wide range of uncertainty sources in the flood risk assessment process, it is of great interest to investigate the propagation and behaviour of these different uncertainties from the start of the modelling framework to the result. The size of registered damages and losses in recent events around the world, reveal the urgency of doing so, even under a context of limited predictability.

In September 2013, severe floods were registered in Mexico as a result of the exceptional simultaneous incidence of two tropical storms, culminating in serious damage and widespread persistent flooding (Pedrozo-Acuña et al., 2014a). This unprecedented event is part of a recent set of extreme flood events over the last decade caused by record-breaking precipitation amounts across Central Europe (Becker and Grünewald, 2003), United Kingdom (Slingo et al., 2014), Pakistan (Webster et al., 2011), Australia (Ven den Honert and McAneney, 2011), Northeastern US (WMO, 2011), Japan (WMO, 2011) and Korea (WMO, 2011). In all cases, the immediate action of governments through the implementation of emergency and action plans was required. The main aim of these interventions was to reduce the duration and impact of floods. In addition, risk reduction measures were designed to ensure both a better flood management and an increase in infrastructure resilience.

One key piece of information in preventing and reducing losses is given by reliable flood inundation maps that enable the dissemination of flood risk to the society and decision makers (Pedrozo-Acuña et al., 2013). Traditionally, this task requires the estimation of different return periods for discharge (Ward et al., 2011) and their propagation to the floodplain by means of a hydrodynamic model. There is currently a large range of models that can be used to develop flood hazard maps (Horrit and Bates, 2002; Horrit et al., 2006).

The aforementioned accelerated progress of computers has given way to the development of model cascades to produce hydrological forecasts, which make use of rainfall predictions from regional climate models (RCMs) with sufficient resolution to capture meteorological events (Bartholomes and Todini, 2005; Demerrit et al., 2010). Within this approach, the coupling of different operational numerical models is carried out, using numerical weather prediction (NWP) with radar data for hydrologic forecast purposes (Liguori and Rico-Ramirez, 2012; Liguori et al., 2012), or NWP with hydrological and hydrodynamic models to determine inundation extension (Pappenberger et al., 2012; Cloke et al., 2013; Ushiyama et al., 2014).

The use of RCMs in climate impact studies on flooding has been reported by Teutschbein and Seibert (2010) and Beven (2011), noting that despite their usefulness, the spatial resolution of models (~25km) remains coarse to capture the spatial resolution of precipitation. This is particularly important, as higher resolution is needed to effectively model the hydrological processes essential for determining flood risk. To overcome this limitation, the utilisation of dynamic downscaling in these models has been significantly growing (Fowler et al., 2007; Leung and Qian, 2009; Lo et al., 2008).

Significant challenges remain in the foreseeable future, among these, the inherent uncertainties in the predictive models are likely to have an important role to play. For example, it is well known that the performance skill of NWPs deteriorates very rapidly with time (Lo et al., 2008). To overcome this, the long-term continuous integration of the prediction has been subdivided into short-simulations, involving the re-initialisation of the model to mitigate the problem of systematic error growth in long integrations (Giorgi, 1990; Giorgi, 2006; Qian et al., 2003). Moreover, the use of ensemble prediction systems to obtain rainfall predictions for hydrological forecasts at the catchment scale is becoming more common among the hydrological community as they enable the evaluation and quantification of some uncertainties in the results (Buizza 2008; Cloke and Pappenberger, 2009; Bartholmes et al. 2009). In these studies, an ensemble is a collection of forecasts made from almost, but not quite, identical initial conditions.

A key question that arises when using a cascade modelling approach to flood prediction or mapping is: how uncertainties associated to meteorological predictions of precipitation propagate to a given flood inundation map? Previous work has been devoted to the examination of uncertainties in the results derived from different ensemble methods, which address differences in the initial conditions in the NWP or even differences in using a single model ensemble vs. multi-model ensemble (Pappenberger et al. 2008; Cloke et al., 2013; Ye et al., 2014). However, less attention has been paid to the behaviour of errors within a model chain that aims to represent a flood event occurring at several spatial scales. In order to understand how errors propagate in a chain of models, this investigation evaluates the transmission of uncertainties from the meteorological model to a given flood map. For this, we utilize a cascade modelling approach comprised by a Numerical Weather Prediction Model (NWP), a rainfall-runoff model and a standard 2D hydrodynamic model. This numerical framework is applied to an observed extreme event registered in Mexico in 2009 for which satellite imagery is available. The investigated uncertainty is limited to the model parameter definition in the NWP model, by means of a multi-physics ensemble technique considering several multi-physics parameterization schemes for the precipitation (Bukosvky and Karoly, 2009). The resulting precipitation fields are used to generate spaghetti plots by means of a distributed hydrological model, enabling the propagation of meteorological uncertainties to the flood hydrograph. Hence, the resulting hydrographs represent the runoff associated to each precipitation field estimated with the NWP. In order to complete the propagation of the uncertainty through the cascade of models to the flood map, the hydrographs are used as forcing in a standard 2D hydrodynamic model.

On the other hand, it is acknowledged that each of the other models (hydrological and hydrodynamic) within the model cascade, will introduce other epistemic and random uncertainties to the result. In order to reduce their influence, the numerical setup of both these models is constructed with the best available data (e.g. LiDAR for the topography) and following recent guidelines for the assessment of uncertainty in flood risk mapping (Beven et al. 2011). In this way, the uncertainty associated to the meteorological model outputs is propagated through the model cascade from the atmosphere to the flood plain. Thus, the aim of this investigation is to study the uncertainty propagation from the meteorological model (due to model parameters), to the determination of an affected area impacted by a well-documented hydro-meteorological event.

This work is organised as follows: Section 2 provides a description of both, the study area and the extreme hydro-meteorological event, which are employed to test our cascade modelling approach; Section 3 introduces the methodology, incorporating a brief description of the selected models setup. Additionally, we incorporate a description of the multi-physics ensemble technique used to quantify and limit the epistemic uncertainty in the NWP model. The resulting precipitation fields, hydrographs and flood maps are compared with available field data and satellite imagery for the event. In Section 4, a discussion of errors along the model cascade, is also presented with some conclusions and future work.

2 Case Study

The selected study area is within the Mexican state of Tabasco, which in recent years has been subjected to severe flooding as reported by Pedrozo-Acuña et al. (2011; 2012). This region comprises the area of Mexico with the highest precipitation rate (2000-3000 mm/year), which mostly occurs during the wet season of the year between May and December. The rainfall climatology is also influenced by the incidence of hurricanes and tropical storms arriving from the North.

In this paper, the extreme hydro-meteorological event selected for the analysis corresponds to that registered in the early days of November 2009 in the Tonalá river. As it is shown in Fig.1, the river is located in the border of Tabasco and Veracruz and during the event, the substantial rainfall intensity provoked its overflowing leaving extensive inundated areas along its floodplain. Top panel of Fig. 1 shows the geographical location of the catchment, with an area of 5,021 km^2, as well as the location of 18 weather stations installed within the region by the National Weather Service. The event was the result of heavy rain induced by the cold front #9, which persisted for four days along Mexico's Gulf Coast, forcing more than 44,000 people to evacuate their homes and affecting more than 90 communities. High intensities in rainfall were recorded in rain gauges from the 31st October to 3rd November, with cumulative daily precipitation values reporting more than 270 mm. The river is approximately 300 km long and before discharging into the Gulf of Mexico, the stream receives additional streamflow from other smaller streams such as Agua Dulcita in Veracruz, and Chicozapote in Tabasco. The bottom panel of the same Figure illustrates the lower Tonalá River, where severe flooding was registered as it is shown in the photographs on the right. The yellow, blue and red dots on the panel represent the location at which the photographs were taken.

The hydrometric data in combination with the satellite imagery for the characterisation of the affected areas, enabled an accurate investigation of the causes and consequences that generated this flood event. The high quality of the available information, allowed the application of a cascade modelling approach comprised by state-of-the-art meteorological, hydrological and hydrodynamic models. This numerical approach is utilised with the intention to carry out an assessment of the modelling framework, with particular emphasis on the propagation of the epistemic uncertainty from the meteorological model to the spatial extent of an affected area. Such investigation paves the road towards a more honest knowledge transfer to decision-makers, whom consider the reliability of the model results.

3 Methodology and Results

The methodology is comprised of a Numerical Weather Prediction Model (NWP), a distributed rainfall-runoff model and a standard 2D hydrodynamic model. It is anticipated that the selected modelling approach will support the advance of the understanding of the connections among scales, intensities, causative factors, and impacts of extremes. This model cascade with state-of-the-art numerical tools representing a hydrological system, enables the development of a framework by which an identification of the reliability of simulations can be undertaken. This framework is utilised to explore the propagation of epistemic uncertainties from the estimation of precipitation in the atmosphere to the identification of a flooded area. Therefore, the aim is not to reproduce an observed extreme event, but to investigate the effects of errors in rainfall prediction by a NWP on inundation areas.

The proposed investigation is important as uncertainties are cascaded through the modelling framework, in order to provide better understanding on how errors propagate within models working at different temporal and spatial scales. It is acknowledged that this information would enhance better flood management strategies, which would be based on the honest and transparent communication of the results produced by a modelling system constrained by intrinsic errors and uncertainties.

3.1 Meteorological model

Simulated precipitation products from numerical weather prediction systems (NWPs) typically show differences in their spatial and temporal distribution. These differences can considerably influence the ability to predict hydrological responses. In this sense, in this study we utilise the advanced research core of the Weather Research and Forecasting (WRF) model Version 3.2. The WRF model is a fully compressible non-hydrostatic, primitive-equation model with multiple nesting capabilities (Skamarock et al., 2008).

As it is shown in **Fig. 2**, the model setup is defined using an interactive nested domain inside the parent domain. This domain is selected in order to simulate more realistic rainfall, with the inner frame enclosing the Tonalá river catchment within a 4 km resolution. The 4 km horizontal resolution is considered good enough to compute a mesoscale cloud system associated to a cold front. It is shown that this finer grid covers the central region of Mexico, while in the vertical dimension, 28 unevenly spaced sigma levels were selected. The initial and boundary conditions were created from the NCEP Global Final Analysis (FNL) with a time interval of 6 hours for the initial and boundary conditions. Each of the model simulations was reinitialised every two days at 1200 UTC, considering a total simulation time from the 27th October 2009 until the 13th November 2009.

Epistemic uncertainty is considered in the WRF model by means of the sensitivity of the results for precipitation, due to variations in the model setup. For this, we utilise a multi-physics ensemble technique proposed by Bukovsky and Karoly (2009), where the sensitivity of simulated precipitation in the model results is examined through variations in the specific setup options by means of twenty three different combinations. The comparison of computed precipitation fields against real measurements from weather stations within the catchment, enabled the quantification of uncertainty in the meteorological model for this event. **Table 1** shows a summary of the different multi-physics parameters used in the WRF model to generate the physics ensemble. As it is shown on this table, there is a large discrepancy in the model skill results in all 23 simulations Error metrics reported in this table are computed using information from all available stations within the numerical domain; which comprised stations that are outside the area of the catchment. It is demonstrated that only 13 of these model runs report a positive Nash-Sutcliff Coefficient (NSC), which indicates a better accuracy for those realisations. In contrast, model runs with negative NSC were dismissed for the numerical reproduction of the event, as this condition is a clear indicator that the observed mean is a better predictor than the model.

Therefore, meteorological model runs that comply with a criteria defined by a NSC>0.3 and a Correlation coefficient (Cor)>0.8 (for the whole numerical domain) are utilised to investigate the propagation of meteorological uncertainties through the modelling framework. This criteria narrows down the meteorological model runs to 12, which will be cascaded to the hydrological model stage to attain streamflow predictions. In this approach, the selected 12 multi-physics ensemble runs of the model represent a plausible and equally likely state of the system in the future.

Fig. 3 illustrates the cumulative precipitation curves computed for each of the 23 model runs of the multi-physics ensemble at four different stations located within the catchment. In this figure differences in the spatial distribution and intensity of precipitation are evident. Moreover, the selected 12 members by the criteria (NSC>0.3 and Cor>0.8) are illustrated by the blue solid lines, while the grey solid lines show those members that were rejected by it. Notably, dismissed members tend to underestimate the amount of precipitation in all four locations that are presented in this figure. For completeness, the rainfall measurements at each meteorological station are also shown by the black solid line, while the red dotted line depicts the mean value of the selected model runs to be propagated through the model cascade. If the 12 selected members are considered in the estimation of ensemble metrics at each station, it is shown that at Station No. 27075, the spread of the estimated cumulative precipitation curves is limited and quantified by a NSC=0.917 and a NRMSE = 10.7%, indicating a good skill of the selected WRF precipitation estimates at this point. In contrast, at Station No. 27007 the spread of the cumulative precipitation is large and characterised by a NSC=0.766 and a NRMSE=19.4%, showing less skill in the model performance than that observed in the previous case. The observed differences of estimated precipitation for this event, highlight the importance of incorporating ensemble techniques in the reproduction of precipitation with this type of models.

Fig. 4 illustrates the cumulative precipitation fields computed for each of the 12 selected members of the multi-physics ensemble, where differences in the spatial distribution and intensity of precipitation were evident. These results suggest that for this event, the precipitation field estimated with the WRF was highly sensitive to the selection of multi-physics parameters. To revise in more detail the performance of the WRF in reproducing this hydro-meteorological event, the estimated cumulative precipitation by each of the selected 12 members of the multi-physics ensemble was compared against measurements at the eighteen weather stations located within and close to the Tonalá catchment.

Table 2 presents a summary of the most well-known error metrics calculated at each weather station and for each member of the ensemble. Among these are the: Normalised Root-Mean Square Error (NRMSE), BIAS, Nash-Sutcliffe Coefficient (NSC), and the Correlation coefficient (Cor). The columns show the local value of each coefficient for a given member of the ensemble (M1, …, M12). As shown in all columns (i.e. member runs), the error metrics have a great spatial variability, hence, indicating the regions of the study area where the model performs better. To illustrate the performance of this ensemble technique at each weather station, the ensemble average of these error metrics is introduced in the last column and indicated by < >. Again, the spatial variability of the metrics is evident. The two bottom rows in each sub-table correspond to the average of the ensemble averages for the whole catchment and for the all the stations. It is shown, that when the average of all stations is taken into account, the skill decreases. However, in this investigation the error that is of interest is the one corresponding to the average of those weather stations located within the catchment, as these will be used as input in the hydrological model. This will enable the propagation of errors in the meteorological model within the model cascade. For clarity, in the same table the stations within the catchment are highlighted in blue.

A question that has been seldom explored in the literature, is how the uncertainty in the prediction of the precipitation (i.e. errors described in this section), cascade into an estimated flood hydrograph determined by a distributed hydrological model. In this sense, the next step in this work, considers the non-linear transfer of rainfall to runoff using a distributed rainfall-runoff model. For this, we employ each one of the selected 12 precipitation fields derived from the WRF as input to determine the associated river discharge with the hydrological model.

3.2 Hydrological model

The hydrological model used in this study was applied to the Tonalá River catchment in an early work presented by Rodríguez-Rincón et al. (2012). This numerical tool was developed by the Institute of Engineering – UNAM (Domínguez-Mora et al., 2008), and comprises a simplified grid-based distributed rainfall–runoff model. The model has been previously applied with success in other catchments in Mexico (e.g. Pedrozo-Acuña et al., 2014b).

The model is based on the method of the Soil Conservation Service (SCS) with a modification that allows the consideration of soil moisture accounting before and after rainfall events. The parameters that are needed for the definition of a runoff curve number within the catchment are the hydrological soil group, land use, pedology and the river drainage network. **Fig. 5** shows for the Tonalá River catchment, the spatial definition of the river network (center panels) and the runoff curve (right panels). For the numerical setup of the hydrological model, we employ topographic information from a LiDAR data set, from which a 10m resolution Digital Elevation Model (DEM) is constructed.

There are two main hypothesis that underpin the SCS curve number method. Firstly, it is assumed that for a single storm and after the start of the runoff, the ratio between actual soil retention and its maximum retention potential is equal to the ratio between direct runoff and available rainfall. Secondly, the initial infiltration is hypothesised to be a fraction of the retention potential.

Thus, the water balance equation and corresponding assumptions are expressed as follows:

$$P = P_e + I_a + F_a \qquad (1)$$

$$\frac{P_e}{P_a - I_a} = \frac{F_a}{S} \qquad (2)$$

$$I_a = \lambda S \quad (3)$$

Where P is rainfall, P_e effective rainfall, I_a is the initial abstraction, F_a is the cumulative abstraction, S is the potential maximum soil moisture retention after the start of the runoff and λ is the scale factor of initial loss. The value of λ is related to the maximum potential infiltration in the basin.

Through the combination of equations (1) - (3) and expressing the initial abstraction (I_a) by 0.2*S we have:

$$P_e = \frac{\left(P - 0.2S\right)^2}{P + 0.8S} \quad (4)$$

where, the value of S [cm] is determined by:

$$S = \frac{2450 - \left(25.4CN\right)}{CN} \quad (5)$$

CN is the runoff curve number, as defined by the Agriculture Department of the USA (USDA, 1985). Values for this parameter vary from 30 to 100, where small numbers indicate low runoff potential while larger numbers indicate an increase in runoff potential. Thus, the permeability of the soil is inversely proportional to the selected curve number. Another parameter that allows the modification of the curve number is the soil water potential given by *Fs*, following *S*=*S***Fs*.

The model includes a parameter to reproduce the effects of evaporation on the ground saturation (F_o). This parameter is useful when the event to be reproduced lasts for several days; however, due to the duration of this event it is assumed equal to 0.9 in all cases. The computation of the runoff in the basin is carried out through the addition of the runoff estimated in each cell to then construct a general hydrograph (See Rodríguez-Rincón et al. 2012). With regards to the definition of values for the other two free parameters in the hydrological model (λ and *Fs*), a traditional calibration process is implemented. For this, we utilise flood hydrographs from past extreme events (2001, 2005, 2007, 2008, 2009 and 2011) observed in this river. For these events, we employ as rainfall input the registered precipitation at the same 4 weather stations that are within the river catchment, which location is shown in the top panel of **Figure 1**. Therefore, we determine six sets of free parameters that are good enough to represent the rainfall-runoff relationship in this catchment. The selected sets of values are illustrated in **Table 3,** where the correlation coefficient and NSC are also reported for each of the years. It is shown that in all the events, the selected set of parameters ensures a good correlation against the observed discharge which is given by Cor>0.7, as well as a positive NSC (accuracy).

It is well known that both the amount and distribution of rainfall can significantly affect the final estimated river discharge (Ferraris et al. 2002; De Roo et al., 2003; Cluckie et al., 2004). In consequence, the propagation of meteorological uncertainty to the rainfall-runoff model is carried out using the 12 WRF rainfall precipitation ensembles as an input in the hydrological model, considering the six sets of free parameters reported in **Table 3**. This procedure enabled the generation of 72 hydrographs that could represent the 2009 event with different skill. Error metrics of all the computed hydrographs are reported in **Table 4.**

For completeness, **Fig. 6a** illustrates the 72 computed hydrographs for the Tonalá River catchment in relation to the measured river discharge for the 2009 event (blue dashed line). It is shown that if all 72 hydrographs are taken into account, uncertainty bounds are significant. Indeed, this illustrates the interaction of the meteorological uncertainty with that coming from the setup of the hydrological model (definition of free parameters). However, the purpose of this study is to investigate in a model cascade framework, how errors in the meteorological prediction stage propagate down to a predicted inundation. In this sense, we narrow down the number of hydrographs shown in **Fig. 6a,** by selecting only those with a Cor>0.7 and NSC>0.6., as reported in **Table 4** only 31 out of 72 (shown in bold) follow this condition. **Fig. 6b** displays the 31 selected hydrographs along with the measured discharge for the 2009 event. Although there is a reduction in the uncertainty bounds, it tis shown that errors in the predicted rainfall are indeed propagated to the hydrological model, which employs a finer spatial resolution (1 km). It has been established that, in some cases, an error in the meteorological model can be compensated by an error in the hydrological model and vice-versa. To illustrate this in more detail, average values of the calculated error metrics for the 31 selected hydrographs are estimated and reported in **Table 4**, with NSC=0.79, Cor=0.96 and BIAS=1.11. Values of the NSC for selected hydrographs in **Table 4** illustrate the resulting differences in skill resulting from the combination of different setups in the hydrological model with the multi-physics ensemble. For instance, in the rows corresponding to the parametes determined for the 2011 event, member M12 indicates a NSC=0.738 showing a poorer skill at reproducing the river discharge with the precipitation derived from this member, in comparison to that registered for member M2 with NSC=0.938. The change in the values of the NSC indicates that results from the regional weather model can be enhanced or weakened by the performance of the hydrological model.

The utilisation of the 31 selected hydrographs in a 2D hydrodynamic model enables the study of the propagation of errors within the cascade of models. In particular, for estimating the flood extent during this extreme event.

3.3 Flood inundation model

Several 2D hydrodynamic models have been developed for simulating extreme flood events. However, any model is only as good as the data used to parameterise, calibrate and validate the model. 2D models have been regarded as suitable for simulating problems where inundation extent changes dynamically through time as they can easily represent moving boundary effects (e.g. Bates and Horritt, 2005). The use of these numerical tools has become common place when flows produce a large areal extent, compared to their depth and where there are large lateral variations in the velocity field (Hunter et al., 2008).

In this study, given the size of the study area the modelling system utilised is comprised by the flow model of MIKE 21 flexible mesh (FM). This numerical model solves the two dimensional Reynolds-averaged Navier–Stokes equations invoking the approximations of Boussinesq and hydrostatic pressure (for details see DHI, 2014). The equations are solved at the centre of each element in the model domain.

The numerical setup is based on a previous work on the study area (Pedrozo-Acuña et al. 2012), with selected resolutions for the elements of the mesh with a size that guarantees the proper assimilation of a 10 m DEM to characterise the elevation in the floodplain. The topographic data has been regarded as the most important factor in determining water surface elevations, base flood elevation, and the extent of flooding and, thus, the accuracy of flood maps in riverine areas (NRC, 2009). Therefore, the elevation data used in this study corresponds to LiDAR data set provided by INEGI (2008). The choice of a 10-m DEM is based on recommendations put forward by the Committee on Floodplain Mapping Technologies, NRC (2007) and Prinos et al. (2008), as such a DEM ensures both accuracy and detail of the ground surface. The model domain is illustrated in **Fig. 7**, along with the numerical mesh and elevation data, it comprises the lower basin of the Tonalá River and additional main water bodies. The colours represent the magnitude of the elevation and bathymetric data assimilated in the numerical mesh, where warm colours identify high ground areas and light blues represent bathymetric data. The integration of high quality topographic information in a 2D model with enough spatial resolution, enables the investigation of the propagation of the meteorological uncertainty to the determination of the flood extent. Moreover, as it is illustrated in **Fig. 7** the numerical mesh considers three boundary conditions. These are input flow boundary where the hydrograph from the rainfall-runoff model is set (red dot); the Tonalá's river mouth, where the astronomical tide occurs for the period of the event (27th October – 12th November 2009) (yellow dot) and the Agua Dulcita river set where a constant discharge of 100 m^3/s is introduced (blue dot). The astronomical tide (microtidal in nature with tidal range <1 m) is determined using the monthly tidal forecast at a nearby point, which is published by CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada) and it is available at (http://predmar.cicese.mx/calmen.php).

On the other hand, hydraulic roughness is a lumped term known as Manning's coefficient that represents the sum of a number of effects, among which are skin friction, form drag and the impact of acceleration and deceleration of the flow. The precise effects represented by the friction coefficient for a particular model depend on the model's dimensionality, as the parameterisation compensates for energy losses due to unrepresented processes, and the grid resolution (Bates et al., 2014). The lack of a comprehensive theory of "effective roughness" have determined the need for calibration of friction parameters in hydraulic models. Furthermore, the determination of realistic spatial distributions of friction across a floodplain in other studies, have showed that only 1 or 2 floodplain roughness classes are required to match current data sources (Werner et al., 2005). Indeed, this suggests that application of complex formulae to establish roughness values for changed floodplain land use are inappropriate until model validation data are improved significantly. Therefore, in this study hydraulic roughness in the floodplain is assumed to be uniform and different from the main river channel, in this sense two values for the Manning number are used, one for the main river channel ($M=32\ m^{1/2}s^{-1}$) and another for the floodplain ($M=28\ m^{1/2}s^{-1}$).

It should be noted that several investigations confirm that there is significant uncertainty associated with flood extent predictions using hydraulic models (e.g. Aronica et al., 1998, 2002; Bates et al. 2004; Pappenberger et al., 2005, 2006, 2007; Romanowicz and Beven, 2003). These uncertainties may be ascribed to differences in spatio-temporal resolutions or the hydraulic roughness that is used in the hydraulic model. In this investigation, however, a more detailed analysis of the different sources of uncertainty in the hydraulic model is not implemented. The numerical setup of the hydraulic model is built following published guidelines for an accurate representation of the case study (see Asselman et al. 2008), which enables us to build the discussion on how an uncertainty generated at the meteorological stage of the model chain propagates and influences a resulting flooded area and depth.

In order to assess whether the 2D model is able to reproduce the flood extent observed in 2009, numerical results of flood extent are compared against the affected area determined from a SPOT image (resolution of 124m). This practice is widely used in the literature to evaluate the results from inundation models and to compare its performance (Di Baldassare et al, 2010b; Wright et al., 2008).

Fig. 8a introduces the result of the hydrodynamic simulation for each of the 31 selected hydrographs, which resulted from the utilisation of the rainfall-runoff model using as input the WRF multi-physics ensemble output. The illustrated flood map summarises the 31 different possibilities of the inundation area that could result from the characterisation of precipitation with the WRF model. Each of these flood maps can also be associated to a probability enabling the representation of a probabilistic flood map, shown in the figure. This allows the identification of the areas highly vulnerable to flooding from this event. Additionally, **Fig. 8b** introduces the infrared SPOT satellite image of the 12th of November 2009, which is used for comparison against the produced flood maps derived from running the 31 hydrographs as inputs in the 2D model. Notably, in the numerical results, the blue area identifies the region of the domain that is most likely to be flooded (90%), the comparison of this area with the observed inundation in the satellite image, show a good skill of the model chain at reproducing the registered flood in the study area.

Despite the variability in the estimated peak discharge utilised as input in the different hydrodynamic runs, inundation results show similar affected areas in all realisations (only with small differences in its size). This is verified in the results shown in **Fig. 9a,** where the relationship between peak discharge of the 31 hydrographs, is plotted against the size of the maximum-flooded area. The distribution of points in this graph clearly indicates that although there are differences in the estimated peak flow (see histogram in **Fig. 9b**), in most cases the resulting size of the inundated area is similar. Histogram plot shown in **Fig. 9c** indicates a clear concentration numerically derived flooded areas with a size larger than 130 km^2. Indeed, the mean value of the maximum-flooded estimated area is 138.94 km^2, while the standard deviation is 16.09 km^2.

These results support that the hydraulic behaviour in all hydrodynamic simulations was indeed very similar, regardless of the peak discharge of the hydrograph. It is reflected that this may be the result of induced hydrodynamics by a valley-filling flood event, which is identified with the relatively high floodplain area-to-channel-depth ratios in all simulations. Hence, all possible hydrographs generated with the hydrological model show similar levels of lateral momentum exchange between main channel and floodplain. For this reason, the predictive performance of all hydrodynamic simulations used to reproduce the inundation extent appears to be good (see **Table 5**).

The estimation of several error metrics in these results was performed using binary flood extent maps, where the comparison is based on the generation of a contingency table, which reports the number of pixels correctly predicted as wet or dry. From this, measures of fit such as: BIAS, False Alarm Ratio (FAR), Probability of Detection (POD), Probability of False Detection (POFD), Critical Success Index (CSI) and the True Skill Statistics (TSS) are estimated. **Table 5** introduces the results for all 31 members and error metrics. Clearly, there is little variability in the performance of the model for each of the runs, showing that there has been a small propagation of the error to the flood map. The ensemble average of these quantities is also illustrated in the last column of the table, where values of BIAS=1.013, FAR=0.189, POD=0.819, POFD=0.180; CSI=0.686 and TSS=0.639 are reported. As noted before, these results indicate an apparent good skill of the model chain at reproducing the flood extension, due to the incidence of this extreme event. It should be borne in mind, however, that some misclassification errors may also be included in the observed flooded area due to specular reflections that may classify some wet vegetation as water or open water as dry land. In consequence, flood extent maps should be used with caution in assessing model performance (Di Baldassare, 2012). This is particularly true during high-magnitude events where the valley is entirely inundated, such as the case study of this investigation where small changes in lateral flood extent may produce large changes in water levels.

In this sense, it has been argued that flood extent maps are not useful for model assessment (Hunter et al., 2005) and high water marks are more useful to evaluate model performance. Unfortunately, for the case study information of inundation depths was not available. Despite this fact, a further revision of simulated inundation depths is also carried out. For this, 10 points distributed within the numerical domain are selected. These are illustrated by the coloured dots in **Fig. 10**, along with the values of mean water depth in all the 31 simulations (red solid line). In all cases, a high variability in the estimated inundation depth on the floodplain is depicted (with values varying between 1.5 and 3m). This result supports that in the case of valley-filling flood events, there is a higher sensitivity to errors in the vertical dimension of the flood.

In one hand, this demonstrates that the geomorphological characteristics of the site (e.g. low-lying area, smooth slopes in the river channel and floodplain) are dominant in the accurate determination of the magnitude of an inundated area, regardless of the peak discharge. This implies that for this type of rivers and when predicting inundation extent, it may be more important to have a good characterisation of the river and floodplain (e.g. high quality field data and a LiDAR derived DEM), than a good characterisation of the rainfall-runoff relationship.

Current approaches to flood mapping, have pointed out that in order to produce a scientifically justifiable flood map, the most physically-realistic model should be utilised (Di Baldassarre et al., 2010). Nevertheless, even with these models the amount of uncertainty involved in the determination of an affected area is important and should be quantified.

4 Discussion and Conclusions

Flood risk mapping and assessment are highly difficult tasks due to the inherent complexity of the relevant processes, which occur in several spatial and temporal scales. As pointed out by Aronica et al. (2013), the processes are subject to substantial uncertainties (epistemic and random), which emerge from different sources and assumptions, from the statistical analysis of extreme events and from the resolution and accuracy of the DEM used in a flood inundation model.

By acknowledging that all models are an imperfect representation of the reality, it is important to quantify the impact of epistemic uncertainties on a given result. The numerical approach utilised in this investigation enabled an assessment of a state-of-the art modelling framework, comprised by meteorological, hydrological and hydrodynamic models. Emphasis was given to the effects of epistemic uncertainty propagation from the meteorological model to the definition of an affected area in a 2D domain. Ensemble climate simulations have become a common practice in order to provide a metric of the uncertainty associated with climate predictions. In this study, a multi-physics ensemble technique is utilised to evaluate the propagation of epistemic uncertainties within a model chain. Therefore, the assessment of hydro-meteorological model performance at the three stages is carried out through the estimation of skill scores.

Fig. 11 presents a summary of the propagation of two well-known error metrics, BIAS (top panel) and NSC/TSS (bottom panel). These metrics were selected, as they enable a direct comparison of their values at each of the stages within the model cascade. In both metrics, the evolution of the confidence limits is illustrated by the size of the bars. Their evolution from the meteorological model to the hydrological model, show an aggregation of meteorological uncertainties with those originated from the rainfall-runoff model. However, the skill is considerably improved from a mean value of 0.65 in the meteorological model, to 0.793 in the hydrological model. In the last stage of the model chain (hydrodynamic model), the confidence limits of the results, show an apparent improvement in model skill. However, it should be noted that this may be ascribed to the complex aggregation of errors in valley-filling events, which is verified in the observed sensitivity of the simulated inundation depths. The mean value of the skill is reduced to TSS=0.639. The results provide an useful way to evaluate the hydro-meteorological uncertainty propagation within the modelling cascade system.

BIAS and NSC/TSS error metrics (**Fig. 11**) revealed discrepancies between observations and simulations throughout the model cascade. For instance, an increase in the NSC from the rainfall to the flood hydrograph it implies that the hydrological model is more sensitive (wider uncertainty bars) to its main input (precipitation) than the WRF model is to the set of micro-physics parameterisations. On the other side, the uncertainty bounds in the hydrological model imply a high sensitivity of hydrographs to both, errors from the meteorological model and its numerical setup with free parameters (amplifying the uncertainty). This is observed in the spaghetti plot shown in **Fig. 6a**, where large uncertainty bounds were identified. In order to reduce errors from the interaction of uncertainties coming from both models, these bounds were reduced with the selection of 31 hydrographs that comply with Cor>0.7 and NSC>0.6 (see **Fig.6b**). It is reflected that the estimated error in the meteorological model may reflect a spatial scaling issue (comparing observations from rain gauges to simulations at the meso-scale).

Results concerning predictions of inundation extent indicate an apparent good skill of the model chain at reproducing the flood extension. The propagation of uncertainty and error from the hydrological model to the inundation area revealed that is necessary to assess model performance not only for flood extension purposes, but also to estimate inundation depths, where results indicate a higher variability (e.g. increase in the error). This last modelling step is quite important given the consequences for issuing warning alerts to the population at risk.

The similar magnitude in inundation extents of all numerical results indicated the predominance of a valley-filling flood event, which was characterised by a flooded area strongly insensitive to the input flood hydrograph. While this can be explained by the limited effect that the volume overflowing the riverbanks and reaching the floodplain will have on the maximum inundation area, the difference between the observed and the simulated flooded area remains important (TSS=0.639).

It should be pointed out, that this methodology contains more uncertainties that were not considered or quantified in the generation of flood extent maps for this event. To quantify the epistemic uncertainty in the larger scale (i.e. atmosphere), a mesoscale numerical weather prediction system was used along with a multi-physics ensemble. The ensemble was designed to represent our limited knowledge of the processes generating precipitation in the lower troposphere. It was shown that a large amount of uncertainty exists in the NWP model, and such uncertainty is indeed propagated over the catchment and floodplain. Members of the ensemble were shown to differ significantly in terms of their cumulative precipitation, spatial distribution, river discharge, inundation depths and areas. Therefore, epistemic uncertainties from each step in this model cascade can be aggregated up to the final output.

The evaluation of the skill in the model cascade shows further potential for improvements of the modelling system. Consequently, future work is planned to include the remaining uncertainties as adopted by, e.g. Pedrozo-Acuña et al. (2013). Special attention should be paid to the interaction between hydro-meteorological and hydrological uncertainty, as well as flood extent estimation in catchments with different morphological setting. The assessment of the error propagation within the model cascade is seen as a good step forward, in the communication of uncertain results to the society. However, as shown in this work, an improvement in model prediction during the first cascade step (rainfall to runoff) can be reverted during the second cascade step (runoff to inundation area) with important consequences for early warning systems and operational forecasting purposes. Finally, the proposed numerical framework could be utilised as a robust alternative for the characterisation of extreme events in ungauged basins.

Acknowledgements

The authors thank the financial support from the Institute of Engineering, UNAM, through internal and international grants. The authors gratefully acknowledge the comments and suggestions made by two anonymous referees and Prof. Jim Freer, handling editor of this manuscript.

References

Aronica, G. T., Apel, H., Baldassarre, G. D. and Schumann, G. J.-P. 2013. HP – Special Issue on Flood Risk and Uncertainty. Hydrol. Process., 27: 1291. doi: 10.1002/hyp.9812

Aronica, G, Hankin, B.G., Beven, K.J., 1998, Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, Advances in Water Resources, 22(4), 349-365

Aronica, G., Bates, P.D. and Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrological Processes, 16, 2001- 2016.

Asselman, N., Bates P., Woodhead S., Fewtrell T., Soares-Frazão S., Zech Y., Velickovic M., de Wit A., ter Maat J., Verhoeven G., Lhomme J. 2008. Flood Inundation Modelling – Model Choice and Proper Application, Report T08-09-03, FLOODsite Project.

Bartholmes, J., Todini, E. 2005. Coupling meteorological and hydrological models for flood forecasting, *Hydrol. Earth Syst. Sci*., 9, 333-346, doi:10.5194/hess-9-333-2005.

Bartholmes, J., Thielen, J., Ramos, M., Gentilini, S., 2009. The European flood alert system EFAS – Part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrology and Earth System Sciences, 2: 141–153.

Bates, P. D., Horritt, M. S., Aronica, G. and Beven, K J, 2004, Bayesian updating of flood inundation likelihoods conditioned on flood extent data, Hydrological Processes, 18, 3347-3370

Bates, P.D., Horritt, M.S. 2005. Modelling wetting and drying processes in hydraulic models. In Bates, P.D., Lane, S.N. and Ferguson, R.I. (eds), Computational Fluid Dynamics: applications in environmental Hydraulics, John Wiley and Sons, Chichester, UK

Bates, P.D., Pappenberger, F., Romanowicz, R.J. 2014. Uncertainty in Flood Inundation Modelling. In Beven, K.J., and Hall, J. (eds.), Applied Uncertainty Analysis for Flood Risk Management, Imperial College Press, World Scientific, London, UK

Becker, A. & Grünewald, U. 2003. Flood risk in central Europe. Science 300, 1099.

Beven, K.J. 2011. I believe in climate change but how precautionary do we need to be in planning for the future? Hydrological Processes 25: 1517–1520.

Beven, K.J. 2014. Use of Models in Flood Risk Management. In Beven, K.J., and Hall, J. (eds.), Applied Uncertainty Analysis for Flood Risk Management, Imperial College Press, World Scientific, London, UK

Beven, K., Leedal, D., McCarthy, S., Lamb, R., Hunter, N., Keef, C., Bates, P., Neal, J. and Wicks, J. 2011. Framework for Assessing Uncertainty in Fluvial Flood Risk Mapping. FRMRC Research Report SWP1.7

Buizza R. 2008.The value of probabilistic prediction. Atmospheric Science Letters, 9: 36–42.

Bukovsky, M. S. and D. J. Karoly. 2009. Precipitation simulations using WRF as a nested regional climate model. *Journal of Applied Meteorology and Climatology*, 48(10): 2152-2159.

Cloke, H. L. and Pappenberger, F. (2008). Evaluating forecasts for extreme events for hydrological applications: an approach for screening unfamiliar performance measures, *Meteorol. Appl*., 15(1), 181–197.

Cloke, H.L., Pappenberger, F. 2009. Ensemble flood forecasting: A review. Journal of Hydrology, doi:10.1016/j.jhydrol.2009.06.005

Cloke, H. L., Wetterhall, F., He, Y., Freer, J. E. and Pappenberger, F. 2013. Modelling climate impact on floods with ensemble climate projections. Q.J.R. Meteorol. Soc., 139: 282–297. doi: 10.1002/qj.1998

Cluckie I., Han D., Xuan Y. 2004. Preliminary Analysis on NWP-Based QPF over UK domain. Deliverable 4.2, FLOODRELIEF Project, URL: http://projects.dhi.dk/floodrelief/

Committee on Floodplain Mapping Technologies, NRC. 2007. Elevation data for floodplain mapping. Washington, DC: National Academic Press.

Committee on FEMA Flood Maps; Board on Earth Sciences and Resources/Mapping Science Committee; NRC. 2009. Mapping the Zone: Improving Flood Map Accuracy. Washington, DC: National Academic Press.

CONAGUA. *Atlas digital del Agua México 2010*, Sistema Nacional de Información del Agua (2010). ftp://ftp.conagua.gob.mx/.

Cuo, L., T. C. Pagano, and Q. J. Wang, 2011: A review of quantitative precipitation forecasts and their use in short to medium range streamflow forecasting. J. Hydrometeor., 12, 713–728, doi:10.1175/2011JHM1347.1.

Demeritt, D., Nobert, S., Cloke, H., and Pappenberger, F. 2010. Challenges in communicating and 5 using ensembles in operational flood forecasting, *Meteorol. Appl.*, 17, 209–222.

De Roo A., Gouweleeuw B., Thielen J., Bartholmes J. et al. 2003. Development of a European flood forecasting system. International Journal of River Basin Management 1(1): 49–59

DHI. MIKE 21 FM Flow model, Scientific documentation. 2014, DHI Group, Horslhome

Di Baldassarre G. 2012. Floods in a Changing Climate: Inundation Modelling. International Hydrology Series, Cambridge University Press, Online ISBN:9781139088411, doi: http://dx.doi.org/10.1017/CBO9781139088411

Di Baldassarre G., Schumann G., Bates P.D., Freer J.E., Beven K.J. 2010. Floodplain mapping: a critical discussion of deterministic and probabilistic approaches, Hydrological Sciences Journal, 55:3, 364-376, doi: 10.1080/02626661003683389.

Di Baldassarre, G., Montanari, A., Lins, H., et al. 2010b. Flood fatalities in Africa: from diagnosis to mitigation. Geophysical Research Letters, 37, L22402, doi:10.1029/2010GL045467

Domínguez M. R., Esquivel G. G., Méndez A. B., Mendoza R. A., Arganis J. M. L., Carrizosa E. E., 2008. Manual del Modelo para pronóstico de escurrimiento. Instituto de Ingeniería. Universidad Nacional Autónoma de México. ISBN 978-607-2-00316-3.

Ferraris L., Rudari R., Siccardi F. 2002. The uncertainty in the prediction of flash floods in the Northern Mediterranean environment. Journal of Hydrometeorology 3: 714–727

Fowler HJ, Blenkinsop S, Tebaldi C. 2007a. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology 27: 1547–1578.

Giorgi, F. 1990. Simulation of regional climate using a limited area model nested in a general circulation model, J. Clim., 3, 941– 963.

Giorgi, F. 2006. Regional climate modeling: Status and perspectives, J. Phys. IV, 139, 101–118.

Hacking, I. 2006 The emergence of probability, 2nd edn. New York, NY: Cambridge University Press.

Horritt M.S., Bates P.D. 2002. Evaluation of one-dimensional and two-dimensional models for predicting river flood inundation. Journal of Hydrology 268: 87–99.

Horritt M.S., Bates P.D., Mattinson M.J. 2006. Effects of mesh resolution and topographic representation in 2D finite volume models of shallow water fluvial flow. Journal of Hydrology 329: 306–314. DOI:10.1016/j.jhydrol.2006Ð02Ð016.

Hunter, N. M., Bates, P. D., Horritt, M. S., et al. 2005. Utility of different data types for calibrating flood inundation models within a GLUE framework. Hydrology and Earth System Sciences, 9(4), 412–430.

Hunter, M., Bates, P.D., Neelz, S., Pender, G., Villanueva, I., Wright, N.G., Liang, D., Falconer, A., Lin, B., Waller, S., Crossley, A.J., Mason, D.C., 2008, Benchmarking 2D hydraulic models for urban flooding, Water Management, 161, Issue WM1, 13-30.

INEGI. 2008. Nube de Puntos LIDAR ajustada al Terreno, Bloque conformado por las cartas 1:50,000: E15A75, E15A76, E15A85, E15A86 del Instituto Nacional de Estadística, Geografía e Informática, México.

Jankov, I., W. A. Gallus, et al. The Impact of Different WRF Model Physical Parameterizations and Their Interactions on Warm Season MCS Rainfall. *Weather and Forecasting* 20, (2005): 1048-

Leung, L. R., and Y. Qian, 2009. Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. Geophysical Research Letters, 36, L03820, doi:10.1029/2008GL036445.

Liguori, S., Rico-Ramirez, M.A. 2012. Quantitative assessment of short-term rainfall forecasts from radar nowcasts and MM5 forecasts. Hydrological Processes, vol 26., pp. 3842-3857

Liguori, S., Rico-Ramirez, M.A., Schellart, A., Saul, A. 2012, Using probabilistic radar rainfall nowcasts and NWP forecasts for flow prediction in urban catchments. Atmospheric Research, vol 103., pp. 80 – 95

Lo, J. C. F., Z. L. Yang, and R. A. Pielke Sr., 2008. Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model. Journal of Geophysical Research, 113, D09112, doi:10.1029/2007JD009216.

Milly P.C.D., Wetherland, R.T., Dunne, K.A., Delworth, T.L. 2002. Increasing risk of great floods in a changing climate. Nature, Vol.415, 514-517, doi :10.1038/415514a

Pappenberger, F., Beven, K. J., Hunter, N. M., Bates P. D., Gouweleeuw, B. T., Thielen, J., de Roo. A. P. J. 2005. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), Hydrology and Earth System Sciences, 9(4), pp. 381-393. doi:10.5194/hess-9-381-2005

Pappenberger, F., Beven, K.J., Hunter N., Gouweleeuw, B., Bates, P., de Roo, A., Thielen, J., 2005, Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfallrunoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). Hydrology and Earth System Science, 9(4), 381-393.

Pappenberger, F, Matgen, P, Beven, K J, Henry J-B, Pfister, L and de Fraipont, P, 2006, Influence of uncertain boundary conditions and model structure on flood inundation predictions, Advances in Water Resources, 29(10), 1430-1449,doi:10.1016/j.advwatres.2005.11.012

Pappenberger, F., Beven, K.J., Frodsham, K., Romanovicz, R. and Matgen, P., 2007. Grasping the unavoidable subjectivity in calibration of flood inundation models: a vulnerability weighted approach. Journal of Hydrology, 333, 275-287.

Pappenberger, F., J. Bartholmes, J. Thielen, H. L. Cloke, R. Buizza, and A. de Roo ~~(~~, 2008~~)~~, . New dimensions in early flood warning across the globe using grand-ensemble weather predictions, Geophys. Res. Lett., 35, L10404, doi:10.1029/2008GL033837.

Pappenberger, F., Dutra, E., Wetterhall, F., Cloke, H. 2012. Deriving global flood hazard maps of fluvial floods through a physical model cascade. Hydrology and Earth System Sciences, 16 4143–56.

Pedrozo-Acuña, A., Breña-Naranjo, J.A., Domínguez-Mora, R., 2014. The hydrological setting of the 2013 floods in Mexico. Weather. Vol.69, No.11, 295-302 Wiley and Sons. doi: 10.1002/wea.2355

Pedrozo-Acuña A., Mariño-Tapia I., Enriquez Ortiz C., Medellín Mayoral G., González-Villareal F.J. 2011. Evaluation of inundation areas resulting from the diversion of an extreme discharge towards the sea: case study in Tabasco, Mexico. Hydrological Processes, 26, (5), 687–704.

Pedrozo-Acuña, A., Rodríguez-Rincón, J.P., Arganis-Juárez, M., Domínguez-Mora, R. and González Villareal, F.J. 2013. Estimation of probabilistic flood inundation maps for an extreme event: Pánuco River, México. Journal of Flood Risk Management, doi: 10.1111/jfr3.12067

Pedrozo-Acuña, A., Ruiz de Alegria-Arzaburu, A., Mariño-Tapia, I., Enriquez, C., González-Villareal, F.J. 2012. Factors controlling flooding at the Tonalá river mouth (Mexico). Journal of Flood Risk Management, Vol.5 (3) pp 226-244. doi: 10.1111/j.1753-318X.2012.01142.x

Pedrozo-Acuña, A. Mejía-Estrada P.I., Rodríguez-Rincón, J.P., Domínguez-Mora, R., González-Villareal, F.J., Flood Risk From Extreme Events in Mexico, 11th International Conference on Hydroinformatics, 2014b.

Prinos P., Kortenhaus A., Swerpel B. & Jiménez J.A. 2008. Review of flood hazard mapping. Floodsite Report No. T03-07-01, 54.

Qian, J.-H., A. Seth, and S. Zebiak 2003. Reinitialized versus continuous simulations for regional climate downscaling, Mon. Weather Rev., 131, 2857–2874.

Rodríguez-Rincón, J.P., Pedrozo-Acuña, A., Domínguez Mora, R., Reeve, D.E., Cluckie, I. 2012. Probabilistic estimation of flood maps: An ensemble approach. FloodRisk2012, The 2nd European Conference on FLOODrisk Management.

Romanowicz, R. and Beven, K. J., 2003, Bayesian estimation of flood inundation probabilities as conditioned on event inundation maps, Water Resources Research, 39(3), W01073, 10.1029/2001WR001056.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.-Y., Wang, W., Powers, J.G. 2008. *A description of the Advanced Research WRF version3*. NCAR Technical Note NCAR/TN475+STR.

Slingo, J., Belcher, S., Scafie, A., McCarthy, M., Saulter, A., McBeath, K., Jenkins, A., Huntingford, C., Marsh, T., Hannaford, J., Parry, S. 2014. The recent storms and floods in the UK. Report Met Office and CEH.

Teutschbein C, Seibert J. 2010. Regional climate models for hydrological impact studies at the catchment scale: a review of recent modelling strategies. Geography Compass 4: 834–860.

USDA-SCS. 1985. National Engineering Handbook, Section 4 - Hydrology. Washington, D.C.: USDA-SCS.

Ushiyama, T., Sayama, T., Tatebe, Y., Fujioka, S., Fukami, K., 2014. Numerical simulation of 2010 Pakistan Flood in the Kabul river basin by using lagged ensemble rainfall forecasting, Journal of Hydrometeorology, Vol. 15, 193-211 pp., doi: 10.1175/JHM-D-13-011.1.

Ven den Honert, R. C. & McAneney, J. 2011. The 2011 Brisbane floods: Causes, impacts and implications. Water 3, 1149–1173.

Wang, W., Bruyere, C., Duda, M., Dudhia, J., Gill, D., Lin, H. C., & Mandel, J.. *ARW version 3 modeling system user's guide*. Mesoscale & Miscroscale Meteorology Division. National Center for Atmospheric Research (July 2010), http://www. mmm. ucar. edu/wrf/users/docs/user_guide_V3/ARWUsersGuideV3. pdf.

Ward, P. J., De Moel, H., and Aerts, J. C. J. H. 2011 How are flood risk estimates affected by the choice of return-periods? Nat. Hazards Earth Syst. Sci. 11 3181–95.

Webster, P. J., Toma, V. E. & Kim, H. M. 2011. Were the 2010 Pakistan floods predictable? Geophys. Res. Lett. 38, L04806.

Werner, M.G.F., Hunter, N. and Bates, P.D. 2005. Identifiability of distributed floodplain roughness values in flood extent estimation, J. Hydrol., 314, 139–157.

World Meteorological Organization, 2011. Provisional Statement on the Status of the Global Climate; available http://www.wmo.int/pages/mediacentre/press_releases/gcs_2011_en.html

Wright, N. G., Asce, M., Villanueva, I., et al. (2008). Case study of the use of remotely sensed data for modeling flood inundation on the River Severn, UK. Journal of Hydraulic Engineering, 134(5), 533–540.

Ye, J., He, Y., Pappenberger, F., Cloke, H. L., Manful, D. Y. and Li, Z. (2014), Evaluation of ECMWF medium-range ensemble forecasts of precipitation for river basins. Q.J.R. Meteorol. Soc., 140: 1615–1628. doi: 10.1002/qj.2243

Table 1. Ensemble members defined for the multi-physics WRF ensemble

Ensemble member	Micro-Physics	surface layer physics	Cumulus physics	Feedback /sst_update	RMSE	NSC	Cor	Bias	Criteria NSC >0.3, Cor >0.8
1	WSM5	5-Layer TDM	Kain-Fritsch Eta	off/off	445.23	-0.25	0.94	0.44	reject
2	WSM5	5-Layer TDM	Kain-Fritsch Eta	off/on	262.73	0.44	0.97	0.98	select
3	WSM5	5-Layer TDM	Kain-Fritsch Eta	on/off	250.51	0.49	0.97	1.01	select
4	WSM5	5-Layer TDM	Kain-Fritsch Eta	on/on	257.35	0.43	0.97	1.05	select
5	WSM5	5-Layer TDM	Betts-Miller-Janjic	off/on	502.47	-0.65	0.97	0.28	reject
6	WSM5	5-Layer TDM	Betts-Miller-Janjic	on/on	520.58	-0.77	0.97	0.25	reject
7	WSM5	Noah	Kain-Fritsch Eta	off/off	233.04	0.42	0.96	1.18	select
8	WSM5	Noah	Kain-Fritsch Eta	off/on	236.14	0.33	0.96	1.24	select
9	WSM5	Noah	Kain-Fritsch Eta	on/off	359.11	0.17	0.90	0.56	reject
10	WSM5	Noah	Kain-Fritsch Eta	on/on	245.31	0.41	0.96	1.12	select
11	WSM5	Noah	Betts-Miller-Janjic	off/off	486.26	-0.49	0.98	0.33	reject
12	WSM5	Noah	Betts-Miller-Janjic	off/on	486.02	-0.49	0.97	0.34	reject
13	WSM5	Noah	Betts-Miller-Janjic	on/off	535.00	-0.82	0.97	0.23	reject
14	WSM5	Noah	Betts-Miller-Janjic	on/on	543.78	-0.87	0.96	0.23	reject
15	Thompson	5-Layer TDM	Kain-Fritsch Eta	off/off	216.70	0.60	0.97	1.09	select
16	Thompson	5-Layer TDM	Kain-Fritsch Eta	off/on	236.64	0.50	0.97	1.15	select
17	Thompson	5-Layer TDM	Kain-Fritsch Eta	on/off	238.89	0.57	0.96	0.97	select
18	Thompson	5-Layer TDM	Kain-Fritsch Eta	on/on	275.24	0.50	0.96	0.89	select
19	Thompson	5-Layer TDM	Betts-Miller-Janjic	off/on	571.49	-1.15	0.96	0.16	reject
20	Thompson	5-Layer TDM	Betts-Miller-Janjic	on/off	572.27	-1.14	0.95	0.16	reject
21	Thompson	5-Layer TDM	Betts-Miller-Janjic	on/on	502.47	-0.65	0.97	0.28	reject
22	Thompson	Noah	Kain-Fritsch Eta	off/off	238.06	0.38	0.96	1.25	select
23	Thompson	Noah	Kain-Fritsch Eta	off/on	234.03	0.48	0.97	1.13	select

Table 2. Error Metrics in the estimation of precipitation by members of the multi-physics ensemble (blue rows indicate the stations located within the Tonalá catchment)

Root-Mean Square Error (RMSE) and Normalised RMSE per Station considering Ensemble average													
Station No.	Multi-physics ensemble member												<Nor_RMSE> %
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	
30167	**210.26**	**96.56**	**144.62**	**104.42**	**106.84**	**76.31**	**160.48**	**129.88**	**101.03**	**210.95**	**164.85**	**86.80**	**13.96**
27003	**544.34**	**578.19**	**564.46**	**474.81**	**427.30**	**516.95**	**458.25**	**484.05**	**568.20**	**572.30**	**385.17**	**479.47**	**35.13**
27007	**234.90**	**246.00**	**198.01**	**135.27**	**129.43**	**207.93**	**126.51**	**197.32**	**246.90**	**328.28**	**132.09**	**191.81**	**19.44**
27015	**96.68**	**129.89**	**151.02**	**194.33**	**235.76**	**179.69**	**152.06**	**152.60**	**118.97**	**116.87**	**260.49**	**188.20**	**24.01**
27074	173.37	211.87	191.22	197.46	78.94	148.88	174.92	247.65	187.98	207.39	123.09	157.21	17.19
27073	227.47	201.91	228.62	256.39	281.38	245.68	186.21	219.36	159.34	147.79	247.69	223.88	46.46
27075	**87.04**	**119.26**	**104.10**	**100.82**	**151.17**	**64.92**	**76.45**	**147.30**	**85.75**	**105.68**	**52.14**	**68.67**	**10.72**
27076	140.53	160.28	141.95	124.03	108.33	130.53	191.75	162.59	226.04	236.09	129.78	150.84	17.14
27077	89.10	113.42	83.60	225.48	252.24	207.73	254.20	282.40	110.77	83.93	203.01	192.86	30.57
27039	333.50	204.36	197.48	295.84	302.19	261.39	264.08	321.66	172.86	152.14	257.59	430.63	73.28
27054	123.18	30.77	45.28	113.16	119.18	77.41	106.84	112.68	118.83	127.43	110.06	106.67	34.75
27060	70.69	56.23	59.51	33.42	40.13	30.04	78.07	93.80	88.46	80.36	56.73	66.31	19.88
27024	160.33	137.81	140.76	120.58	127.54	73.57	148.27	136.47	145.12	167.79	153.26	151.87	85.04
27084	68.72	71.32	54.58	53.56	106.93	65.65	61.06	72.31	61.46	62.96	50.14	50.92	19.02
7365	172.91	117.44	103.02	252.03	139.79	163.49	301.52	216.38	179.67	129.71	271.88	210.11	24.52
27011	143.70	162.77	143.61	107.82	77.55	86.15	128.03	143.69	106.59	116.49	86.81	81.27	106.83
27036	81.46	60.69	27.36	61.69	19.14	35.64	23.58	45.89	22.13	40.23	39.22	55.55	12.04
27008	158.85	72.82	74.96	131.34	134.94	100.16	102.82	149.97	66.67	79.36	97.87	254.33	19.68
										Average {Rel_RMSE} catch.			**23.14**
										Average {Rel_RMSE} all			33.87

BIAS per Station and Ensemble Average													
Station No.	Multi-physics ensemble member												<BIAS>
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	
30167	**0.71**	**0.90**	**0.81**	**1.07**	**1.12**	**0.99**	**0.80**	**0.85**	**0.91**	**0.71**	**1.23**	**1.06**	**0.93**
27003	**0.51**	**0.48**	**0.50**	**0.58**	**0.62**	**0.54**	**0.59**	**0.57**	**0.49**	**0.49**	**0.66**	**0.58**	**0.55**
27007	**0.72**	**0.71**	**0.79**	**0.91**	**0.91**	**0.78**	**1.13**	**1.26**	**0.73**	**0.61**	**0.90**	**0.80**	**0.85**
27015	**1.21**	**1.32**	**1.40**	**1.50**	**1.61**	**1.46**	**1.37**	**1.37**	**1.24**	**1.21**	**1.68**	**1.48**	**1.40**
27074	0.82	0.76	0.79	0.78	1.08	0.86	0.81	0.71	0.80	0.77	0.88	0.83	0.82
27073	1.74	1.65	1.74	1.83	1.91	1.80	1.58	1.70	1.47	1.44	1.80	1.72	1.70
27075	**0.92**	**0.85**	**0.88**	**0.88**	**1.20**	**0.96**	**0.90**	**0.80**	**0.89**	**0.86**	**0.98**	**0.93**	0.92
27076	0.86	0.82	0.86	0.91	0.95	0.89	0.79	0.84	0.73	0.71	0.89	0.85	0.84
27077	1.12	1.17	1.10	1.48	1.54	1.44	1.54	1.60	1.20	1.14	1.42	1.40	1.35
27039	2.41	1.87	1.84	2.26	2.29	2.11	2.13	2.36	1.73	1.64	2.09	2.84	2.13
27054	1.89	1.08	1.24	1.82	1.87	1.54	1.76	1.81	1.84	1.91	1.79	1.77	1.69
27060	1.42	1.33	0.72	1.08	1.20	1.05	1.47	1.57	1.54	1.49	1.32	1.39	1.30
27024	3.34	2.96	3.03	2.76	2.88	2.07	3.16	2.98	3.11	3.45	3.17	3.17	3.01
27084	1.32	1.35	1.17	1.23	1.61	0.78	1.27	1.36	1.27	1.29	1.07	1.01	1.23
7365	1.43	1.20	1.09	1.63	1.32	0.72	1.78	1.55	1.43	1.26	1.68	1.51	1.38
27011	3.57	3.91	3.55	2.93	2.33	2.49	3.33	3.58	2.91	3.09	2.56	2.45	3.06
27036	1.36	1.25	1.09	1.28	0.97	1.15	0.95	1.20	1.06	1.16	1.15	1.24	1.15
27008	1.37	1.07	1.05	1.29	1.31	1.20	1.21	1.35	0.99	0.93	1.19	1.62	1.22
										Average {Rel_RMSE} catch.			**0.94**
										Average {Rel_RMSE} all			1.42

Continuation of Table 2. Error Metrics in the estimation of precipitation by members of the multi-physics ensemble (blue rows indicate the stations located within the Tonalá catchment)

Nash-Sutcliff Coefficient per Station and Ensemble average													
Station No.	Multi-physics ensemble member												<NSC>
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	
30167	0.72	0.94	0.87	0.93	0.93	0.96	0.84	0.89	0.94	0.72	0.83	0.95	**0.88**
27003	0.16	0.05	0.09	0.36	0.48	0.24	0.40	0.33	0.08	0.07	0.58	0.34	**0.26**
27007	0.70	0.67	0.78	0.90	0.91	0.76	0.91	0.79	0.66	0.41	0.90	0.80	**0.77**
27015	0.88	0.78	0.70	0.50	0.27	0.57	0.70	0.69	0.81	0.82	0.11	0.53	**0.61**
27074	0.84	0.76	0.80	0.79	0.97	0.88	0.84	0.67	0.81	0.77	0.92	0.87	0.83
27073	-0.27	0.00	-0.28	-0.61	-0.94	-0.48	0.15	-0.18	0.38	0.46	-0.50	-0.23	-0.21
27075	**0.94**	**0.89**	**0.91**	**0.92**	**0.82**	**0.97**	**0.95**	**0.83**	**0.94**	**0.91**	**0.98**	**0.96**	**0.92**
27076	0.87	0.83	0.86	0.90	0.92	0.88	0.75	0.82	0.65	0.62	0.89	0.85	0.82
27077	0.82	0.70	0.84	-0.17	-0.46	0.01	-0.48	-0.83	0.72	0.84	0.05	0.15	0.18
27039	-4.41	-1.03	-0.90	-3.26	-3.44	-2.32	-2.39	-4.03	-0.45	-0.13	-2.23	-8.02	-2.72
27054	-0.46	0.91	0.80	-0.23	-0.36	0.42	-0.10	-0.22	-0.36	-0.56	-0.16	-0.09	-0.03
27060	0.60	0.75	0.72	0.91	0.87	0.93	0.51	0.29	0.37	0.48	0.74	0.65	0.65
27024	-7.99	-5.64	-5.93	-4.08	-4.69	-0.89	-6.68	-5.51	-6.36	-8.84	-7.21	-7.06	-5.91
27084	0.67	0.64	0.79	0.80	0.20	0.70	0.74	0.63	0.73	0.72	0.82	0.82	0.69
7365	0.50	0.77	0.82	-0.07	0.67	0.55	-0.54	0.21	0.45	0.72	-0.25	0.25	0.34
27011	-16.74	-21.76	-16.72	-8.99	-4.17	-5.38	-13.08	-16.74	-8.76	-10.66	-5.47	-4.67	-11.09
27036	0.61	0.78	0.96	0.78	0.98	0.93	0.97	0.88	0.97	0.91	0.91	0.82	0.87
27008	0.60	0.92	0.91	0.72	0.71	0.84	0.83	0.64	0.93	0.90	0.85	-0.03	0.73
										Average {Rel_RMSE} catch.			**0.63**
										Average {Rel_RMSE} all			-0.63

Correlation Coefficient per Station and Ensemble average													
Station No.	Multi-physics ensemble member												<Cor>
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	
30167	**0.99**	**0.99**	**0.99**	**0.97**	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.97**	**0.98**	**0.99**
27003	**0.95**	**0.96**	**0.97**	**0.97**	**0.98**	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**
27007	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.95**	**0.98**	**0.97**	**0.97**
27015	**0.97**	**0.96**	**0.97**	**0.94**	**0.93**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.93**	**0.94**	**0.95**
27074	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0.98	0.98	0.98	0.99	0.99	0.98
27073	0.95	0.96	0.95	0.94	0.94	0.94	0.92	0.92	0.91	0.92	0.94	0.94	0.94
27075	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	**0.99**
27076	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.97	0.97	0.97
27077	0.96	0.95	0.96	0.96	0.95	0.96	0.95	0.95	0.97	0.97	0.95	0.96	0.96
27039	0.95	0.95	0.94	0.93	0.94	0.94	0.94	0.94	0.95	0.95	0.94	0.93	0.94
27054	0.91	0.96	0.94	0.93	0.93	0.94	0.91	0.92	0.91	0.90	0.93	0.93	0.93
27060	0.96	0.97	0.97	0.96	0.97	0.97	0.95	0.95	0.96	0.96	0.97	0.96	0.96
27024	0.91	0.93	0.92	0.90	0.91	0.95	0.89	0.90	0.89	0.89	0.94	0.94	0.91
27084	0.91	0.91	0.92	0.94	0.92	0.95	0.92	0.91	0.92	0.92	0.93	0.93	0.92
7365	0.93	0.93	0.94	0.92	0.94	0.97	0.91	0.92	0.91	0.92	0.91	0.92	0.93
27011	0.94	0.94	0.95	0.93	0.95	0.96	0.89	0.93	0.91	0.92	0.91	0.91	0.93
27036	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
27008	0.97	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.96	0.96	0.96	0.96
										Average {Rel_RMSE} catch.			**0.97**
										Average {Rel_RMSE} all			0.95

Table 3. Flood events in the Tonala River used in the calibration process of free parameters in the hydrological model, along with computed error metrics.

Event	Max Q (m3/s) Obs.	ʎ	Fs	Fo	Max Q (m3/s) Calc.	NSC	Cor	Bias
2001	577.98	0.2	0.1	0.9	584.79	0.529	0.764	1.112
2005	589.25	0.4	0.6	0.9	609.87	0.812	0.907	1.043
2007	538.50	0.2	1.8	0.9	543.87	0.483	0.780	0.902
2008	597.35	0.4	1.8	0.9	823.04	0.155	0.861	0.983
2009	1262.57	0.8	1.8	0.9	1424.56	0.910	0.962	0.942
2011	545.40	0.9	1.6	0.9	597.08	0.413	0.721	1.051

Table 4. Error metrics in the estimation of river discharge by the rainfall-runoff model using 6 parameter sets and 12 members of the multi-physics ensemble (those selected are shown in bold with NSC>0.6 and Cor>0.7).

Member No.	WRF Member	Hydrological Parameters	NSC	Cor	Bias
1	**M1**	**2001**	**0.733**	**0.884**	**0.852**
2	M2	2001	0.074	0.973	1.529
3	M3	2001	-0.035	0.974	1.564
4	M4	2001	-0.511	0.975	1.686
5	M5	2001	-0.638	0.441	1.485
6	M6	2001	-0.223	0.961	1.593
7	M7	2001	-0.192	0.961	1.579
8	M8	2001	-0.043	0.959	1.537
9	M9	2001	0.064	0.958	1.504
10	M10	2001	0.245	0.971	0.525
11	M11	2001	-1.503	0.944	1.832
12	M12	2001	-0.752	0.954	1.710
13	**M1**	**2005**	**0.639**	**0.901**	**0.742**
14	M2	2005	0.404	0.977	1.414
15	M3	2005	0.318	0.978	1.449
16	M4	2005	-0.077	0.977	1.569
17	M5	2005	-0.545	0.366	1.368
18	M6	2005	0.181	0.968	1.478
19	M7	2005	0.200	0.968	1.465
20	M8	2005	0.321	0.966	1.422
21	M9	2005	0.408	0.966	1.389
22	M10	2005	-0.081	0.960	0.426
23	M11	2005	-0.909	0.951	1.717
24	M12	2005	-0.264	0.961	1.595
25	M1	2007	0.376	0.914	0.601
26	**M2**	**2007**	**0.761**	**0.978**	**1.244**
27	**M3**	**2007**	**0.711**	**0.979**	**1.278**
28	M4	2007	0.444	0.976	1.395
29	M5	2007	-0.440	0.261	1.191
30	**M6**	**2007**	**0.633**	**0.974**	**1.306**
31	**M7**	**2007**	**0.647**	**0.974**	**1.293**
32	**M8**	**2007**	**0.722**	**0.973**	**1.251**
33	**M9**	**2007**	**0.771**	**0.972**	**1.219**
34	M10	2007	-0.508	0.952	0.322
35	M11	2007	-0.129	0.959	1.539
36	M12	2007	0.340	0.969	1.420
37	M1	2008	0.240	0.922	0.547
38	**M2**	**2008**	**0.837**	**0.978**	**1.186**
39	**M3**	**2008**	**0.797**	**0.978**	**1.220**
40	M4	2008	0.570	0.974	1.337
41	M5	2008	-0.479	0.209	1.132
42	**M6**	**2008**	**0.741**	**0.976**	**1.248**
43	**M7**	**2008**	**0.753**	**0.976**	**1.235**
44	**M8**	**2008**	**0.813**	**0.975**	**1.194**
45	**M9**	**2008**	**0.851**	**0.975**	**1.161**
46	M10	2008	-0.720	0.945	0.276
47	M11	2008	0.079	0.962	1.481
48	**M12**	**2008**	**0.495**	**0.972**	**1.361**
49	M1	2009	-0.036	0.838	0.494
50	**M2**	**2009**	**0.819**	**0.978**	**0.882**
51	**M3**	**2009**	**0.899**	**0.977**	**0.907**
52	**M4**	**2009**	**0.649**	**0.963**	**1.286**
53	M5	2009	0.060	0.811	0.580
54	**M6**	**2009**	**0.839**	**0.959**	**0.849**
55	**M7**	**2009**	**0.883**	**0.959**	**0.890**
56	**M8**	**2009**	**0.896**	**0.954**	**0.929**
57	**M9**	**2009**	**0.890**	**0.950**	**0.928**
58	M10	2009	-1.233	0.972	0.209
59	**M11**	**2009**	**0.638**	**0.938**	**1.236**
60	**M12**	**2009**	**0.885**	**0.946**	**1.042**
61	M1	2011	-0.247	0.949	0.396
62	**M2**	**2011**	**0.938**	**0.970**	**1.019**
63	**M3**	**2011**	**0.930**	**0.971**	**1.052**
64	**M4**	**2011**	**0.819**	**0.964**	**1.168**
65	M5	2011	-0.662	0.055	0.955
66	**M6**	**2011**	**0.890**	**0.978**	**1.133**
67	**M7**	**2011**	**0.899**	**0.979**	**1.120**
68	**M8**	**2011**	**0.931**	**0.979**	**1.079**
69	**M9**	**2011**	**0.945**	**0.978**	**1.047**
70	M10	2011	-1.136	0.931	0.195
71	M11	2011	0.433	0.967	1.364
72	**M12**	**2011**	**0.738**	**0.976**	**1.246**
<Ensemble average of selected members>			0.793	0.965	1.113

Table 5. Error metrics in the estimation of river discharge by the hydrodynamic model using the 31 members of the multi-physics ensemble.

Comparison of flooded areas between numerical results from running ensemble members vs. Observed																															
Error metrics	**Ensemble Member**																														**<Ensemble average>**
	M1	**M13**	**M26**	**M27**	**M30**	**M31**	**M32**	**M33**	**M38**	**M39**	**M42**	**M43**	**M44**	**M45**	**M50**	**M51**	**M52**	**M54**	**M55**	**M56**	**M57**	**M59**	**M60**	**M62**	**M63**	**M64**	**M66**	**M67**	**M68**	**M69**	**M72**	
BIAS	0.903	0.838	1.084	1.099	1.119	1.120	1.094	1.078	1.056	1.021	1.092	1.089	1.096	1.051	0.902	0.915	0.891	0.820	1.020	0.982	0.872	1.056	1.004	0.982	0.995	1.047	1.040	1.028	1.016	1.005	1.092	1.013
FAR: False Alarm Ratio	0.148	0.120	0.215	0.217	0.283	0.210	0.216	0.212	0.209	0.217	0.216	0.215	0.152	0.207	0.148	0.154	0.139	0.137	0.193	0.155	0.133	0.206	0.187	0.178	0.182	0.204	0.201	0.225	0.192	0.187	0.216	0.189
POD: Probability of Detection	0.770	0.737	0.851	0.861	0.849	0.849	0.858	0.849	0.836	0.751	0.857	0.854	0.848	0.833	0.769	0.775	0.751	0.810	0.823	0.845	0.756	0.847	0.816	0.807	0.814	0.833	0.831	0.821	0.821	0.818	0.857	0.819
POFD: Probability of False Detection	0.124	0.094	0.217	0.222	0.187	0.187	0.220	0.214	0.205	0.186	0.220	0.219	0.186	0.203	0.124	0.131	0.185	0.185	0.184	0.066	0.108	0.266	0.175	0.163	0.168	0.199	0.195	0.186	0.182	0.175	0.220	0.180
CSI : Critical Succes Index	0.679	0.670	0.690	0.695	0.711	0.711	0.694	0.691	0.685	0.709	0.693	0.692	0.710	0.685	0.679	0.679	0.706	0.654	0.687	0.708	0.677	0.620	0.687	0.687	0.690	0.686	0.687	0.619	0.688	0.688	0.693	0.686
True Skill Statistics	0.645	0.643	0.634	0.639	0.621	0.662	0.638	0.636	0.631	0.660	0.637	0.636	0.661	0.631	0.645	0.643	0.615	0.601	0.639	0.659	0.648	0.660	0.641	0.644	0.640	0.634	0.636	0.610	0.640	0.642	0.637	0.639

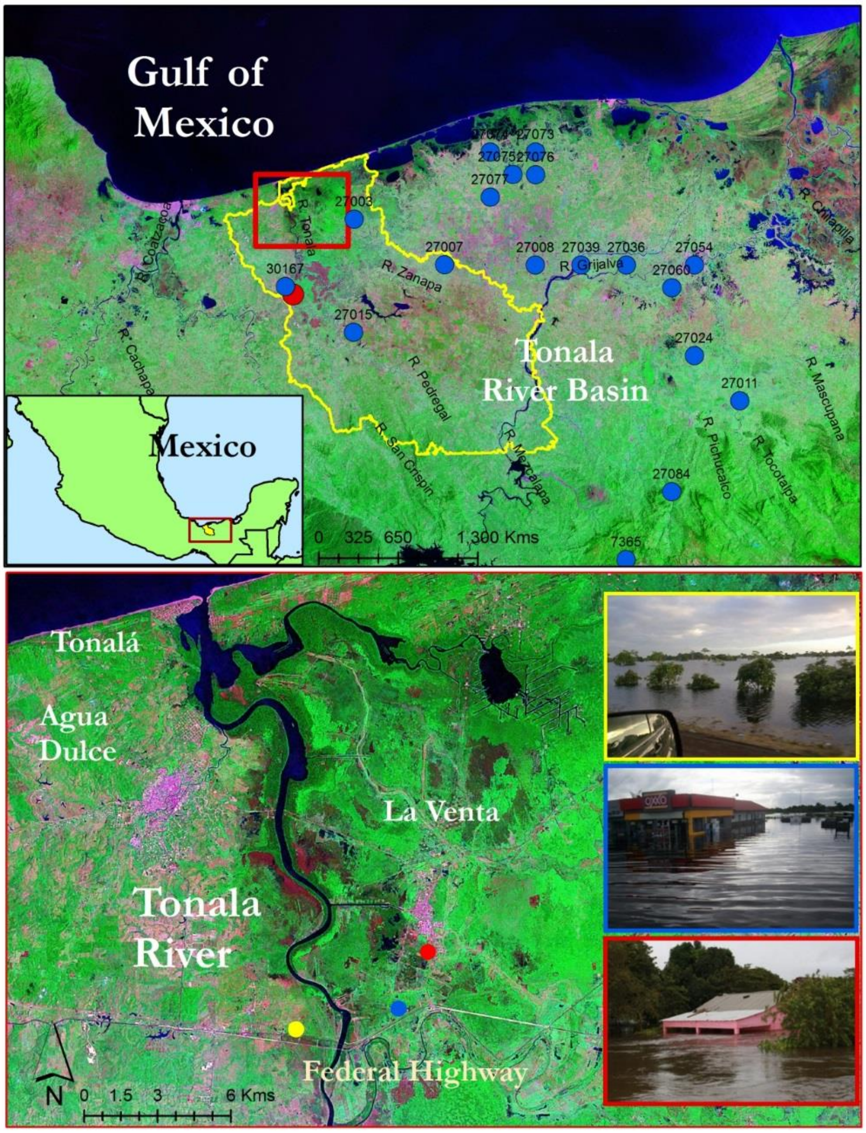

Figure 1. Top panel: Location of the Tonala River basin in Mexico, blue line represents the boundary limits of the catchment; blue dots illustrate the location of weather stations; red dot: streamflow gauge. Bottom panel: zoom of the study area and photographs of observed impacts; yellow, blue and red dots represent the location at which photos were taken.

Figure 2. Numerical setup of the WRF with a nested domain covering Mexico. Domain 1: 25km resolution; Domain 2: 4km resolution; the orange region illustrates the Tonalá catchment.

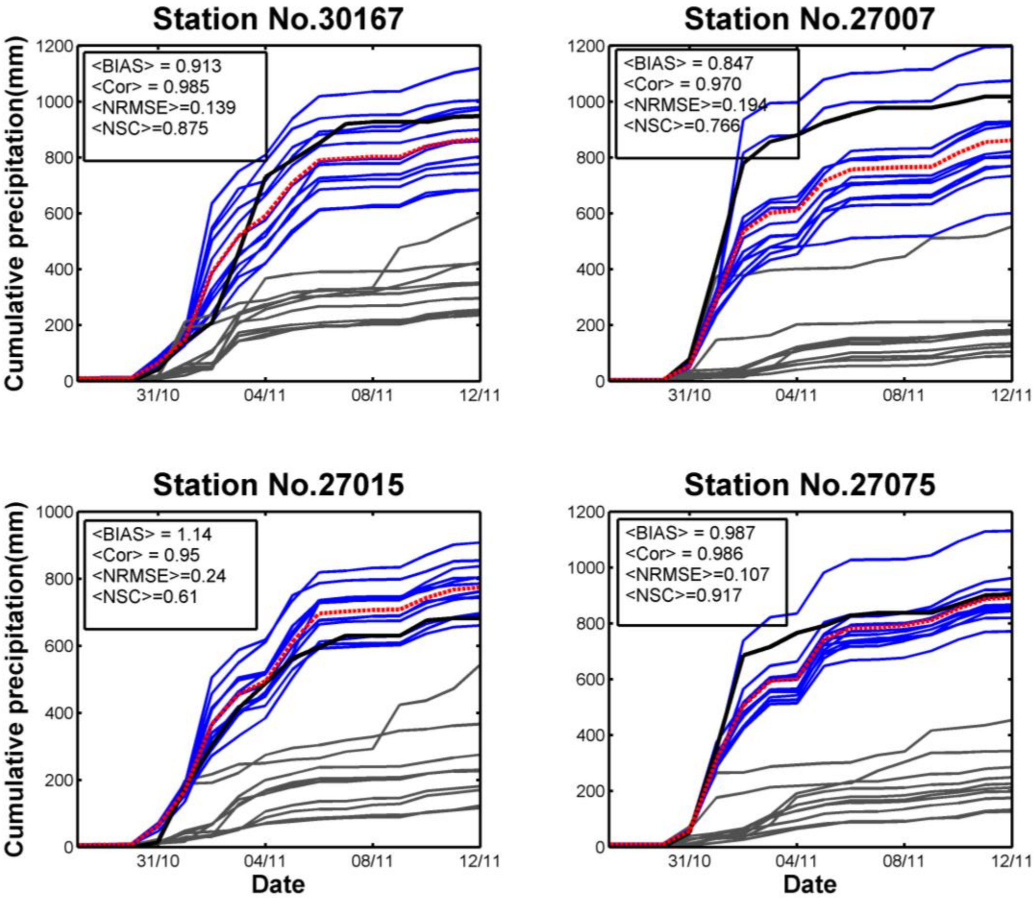

Figure 3. Comparison of cumulative precipitation estimated by the 23 model runs of the WRF multi-physics ensemble. Blue solid line: selected members with NSC> 0.3; grey solid line: disregarded members with NSC <0.3; red dotted line: mean of the selected 12 members; black solid line: measurements at each of the four weather stations from 27th October 2009 to 12th November 2009.

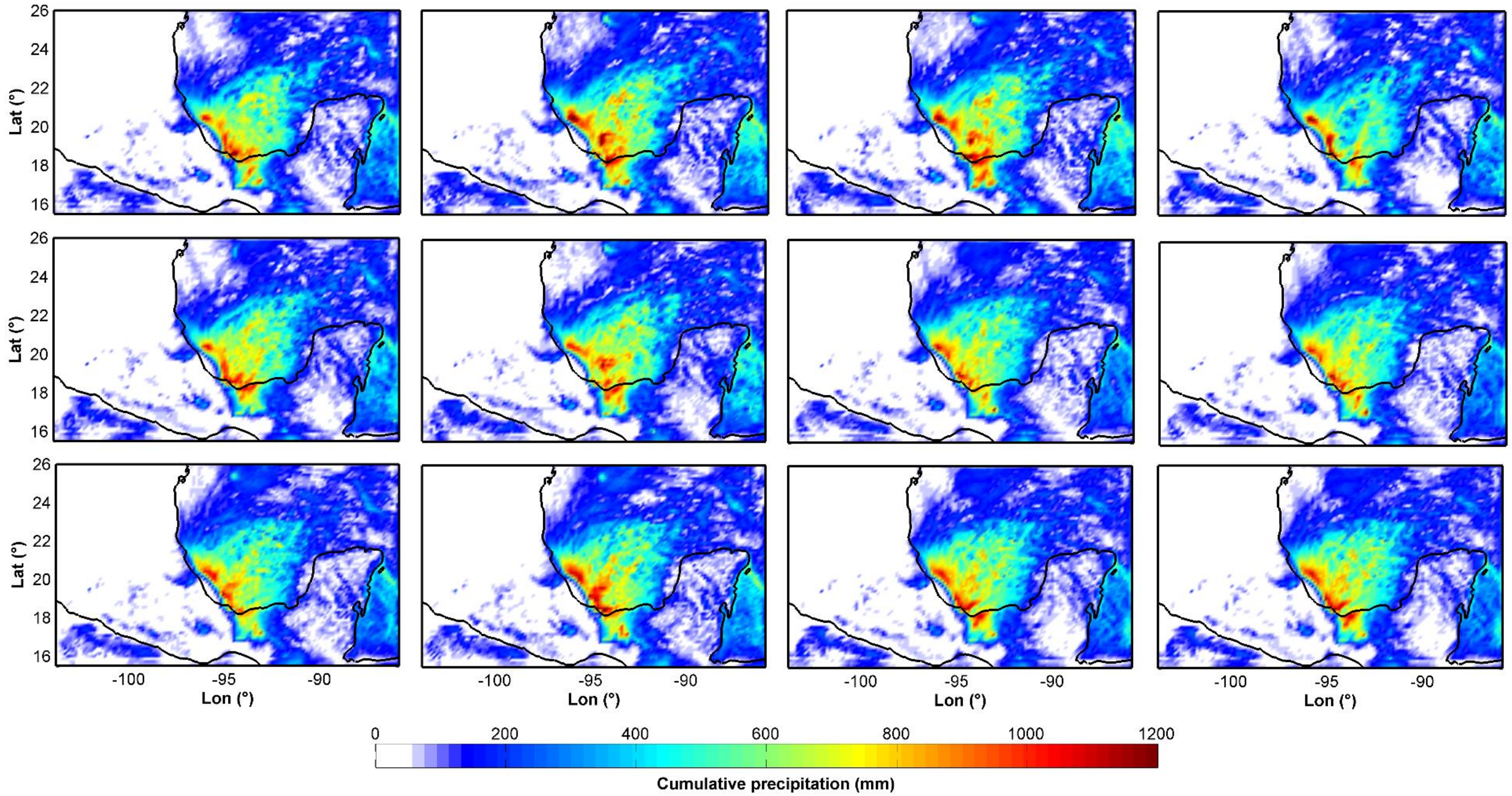

Figure 4. Cumulative precipitation fields estimated by the WRF model using the selected 12 members of the multi-physics ensemble (27th October 2009 – 12th November 2009).

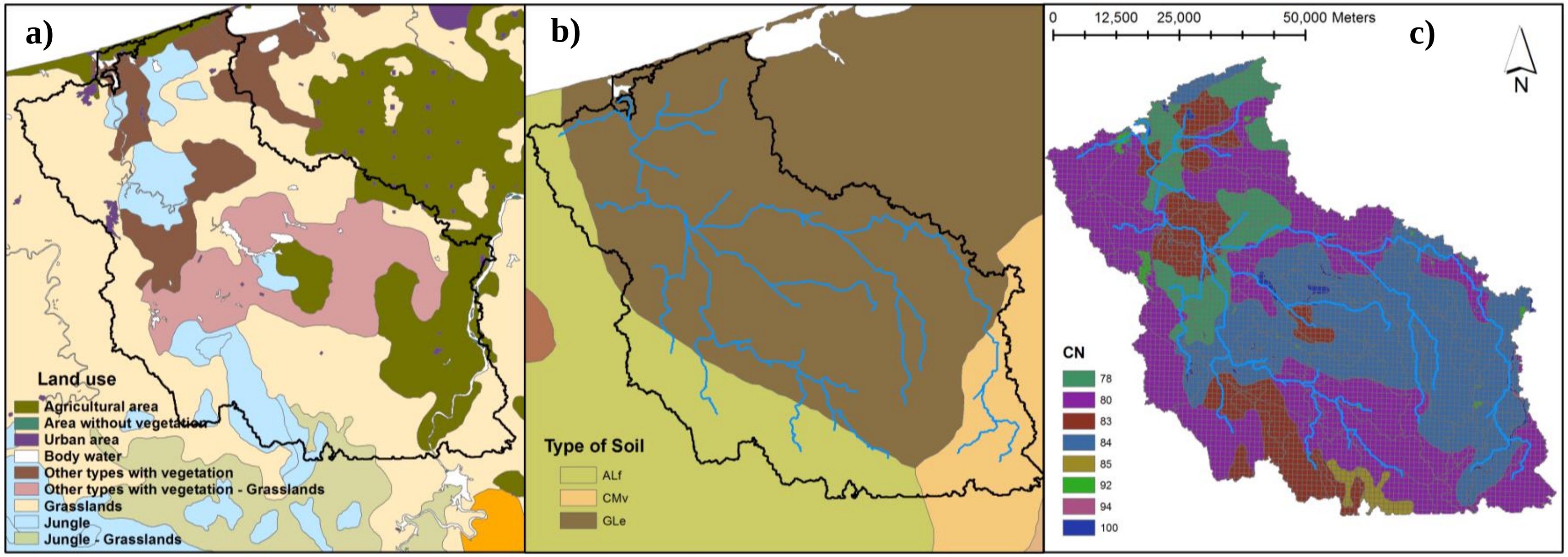

Figure 5. Input data parameters in the hydrological model; a) Land use; b) Pedology; c) River network, curve number and grid.

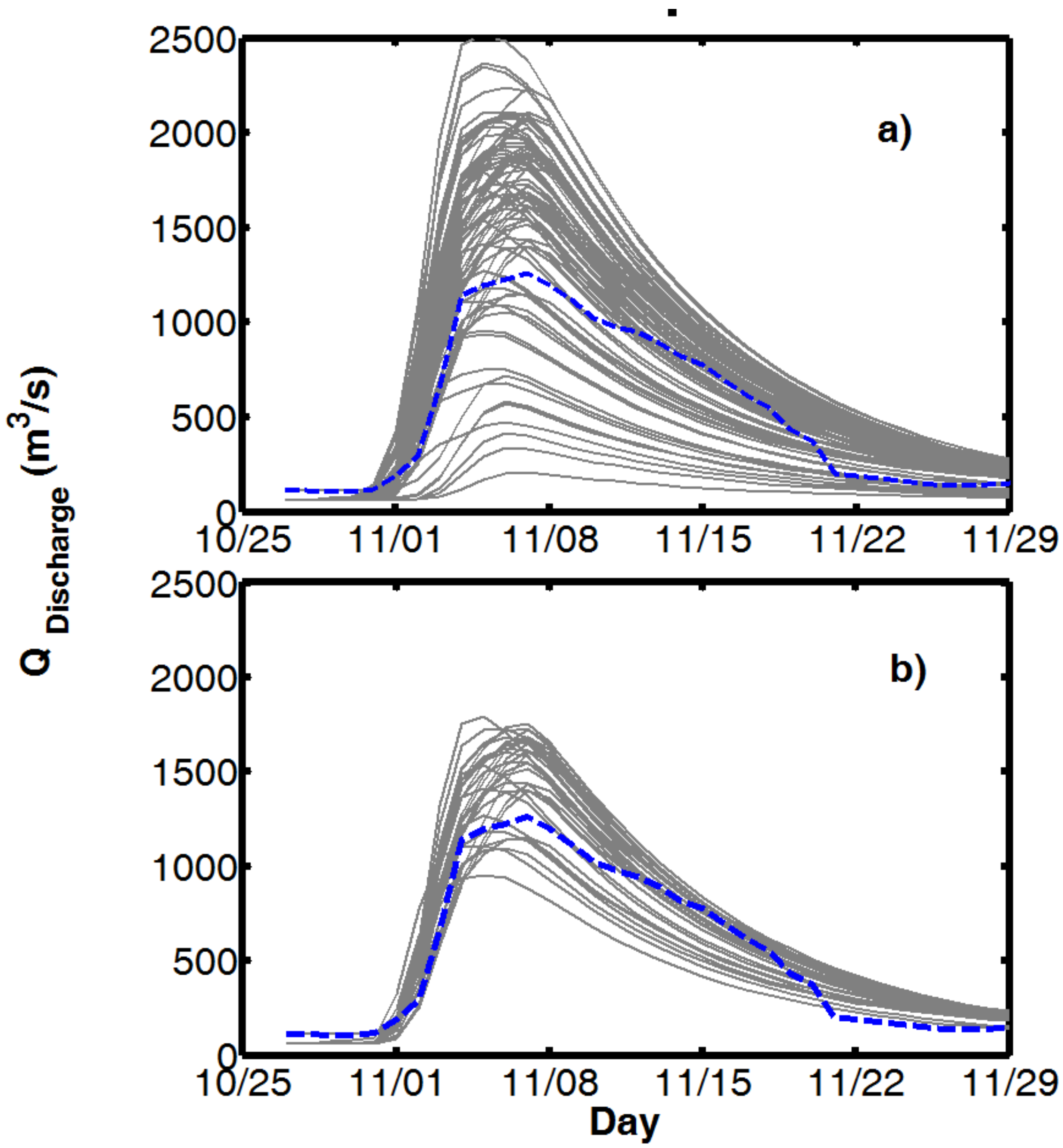

Figure 6. a) 72 hydrographs computed using the rainfall-runoff model with 6 sets of parameters and 12 WRF ensemble precipitation fields as input data; b) 31 selected hydrographs to serve as input in the hydrodynamic model; grey lines illustrate the ensemble members and the blue dashed line shows the measured river discharge for the event.

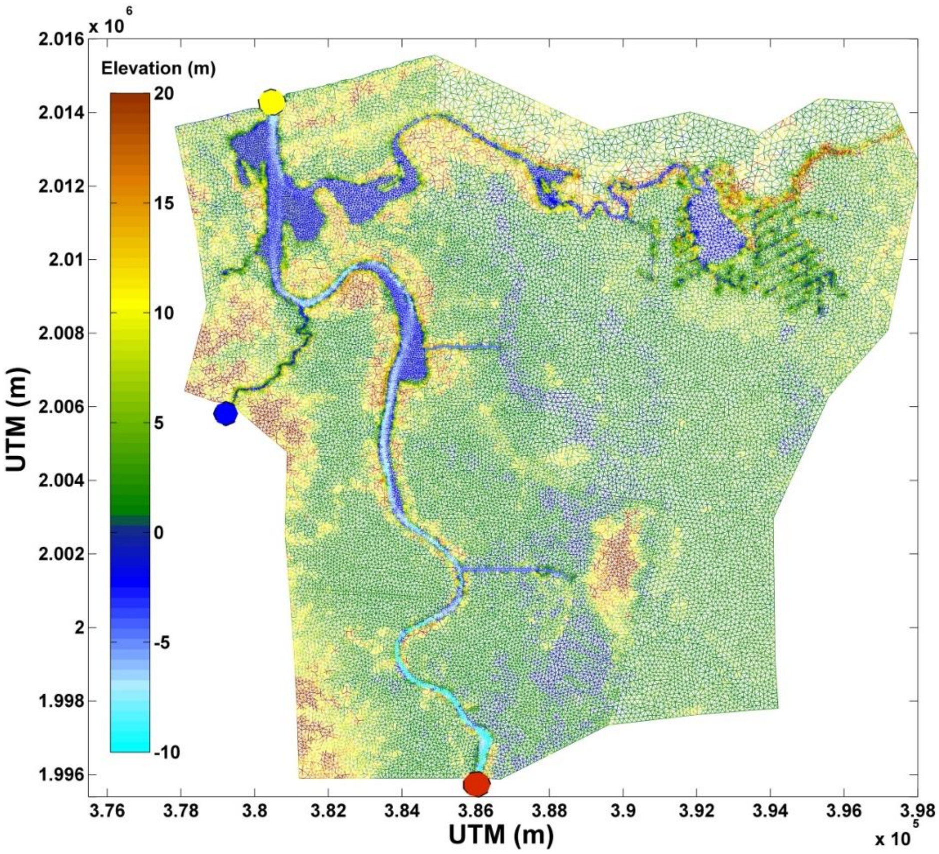

Figure 7. Model domain along with the numerical mesh and elevation data in the study area; Boundary conditions are represented by blue dot: Agua Dulcita river; red dot: input hydrograph; yellow dot: river-mouth.

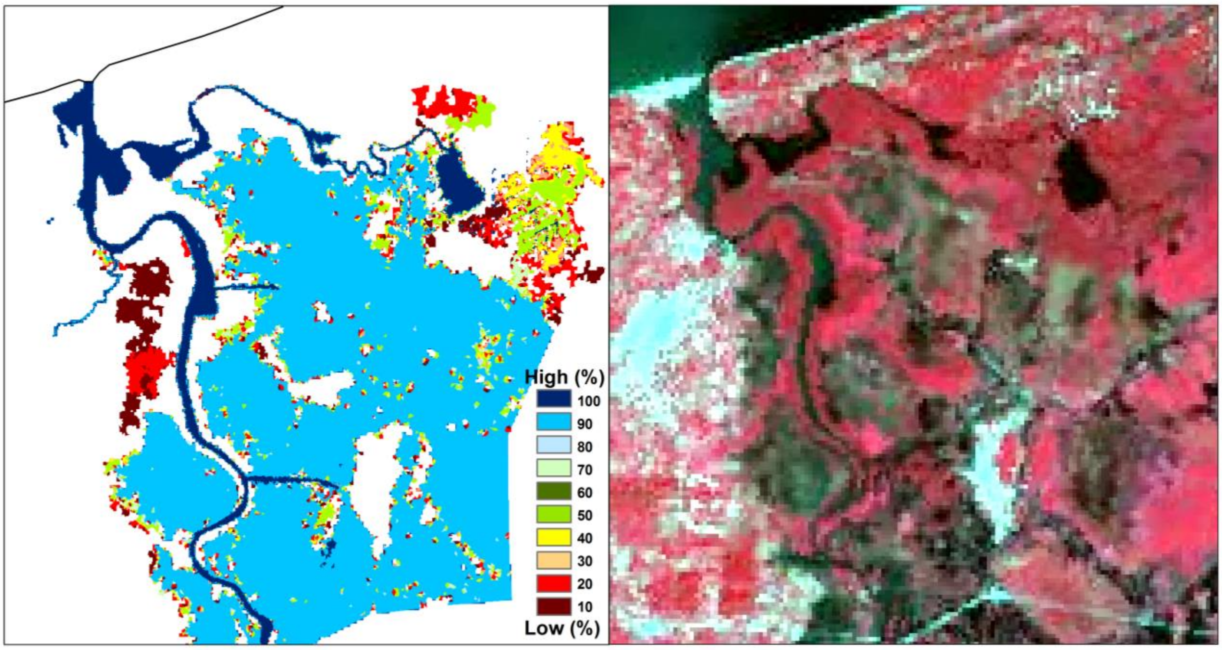

Figure 8. Data vs. model comparison of flood extent; a) Probabilistic flood map derived from the ensemble runs with the hydrodynamic model; b) Infrared SPOT image corresponding to the 15th November 2009.

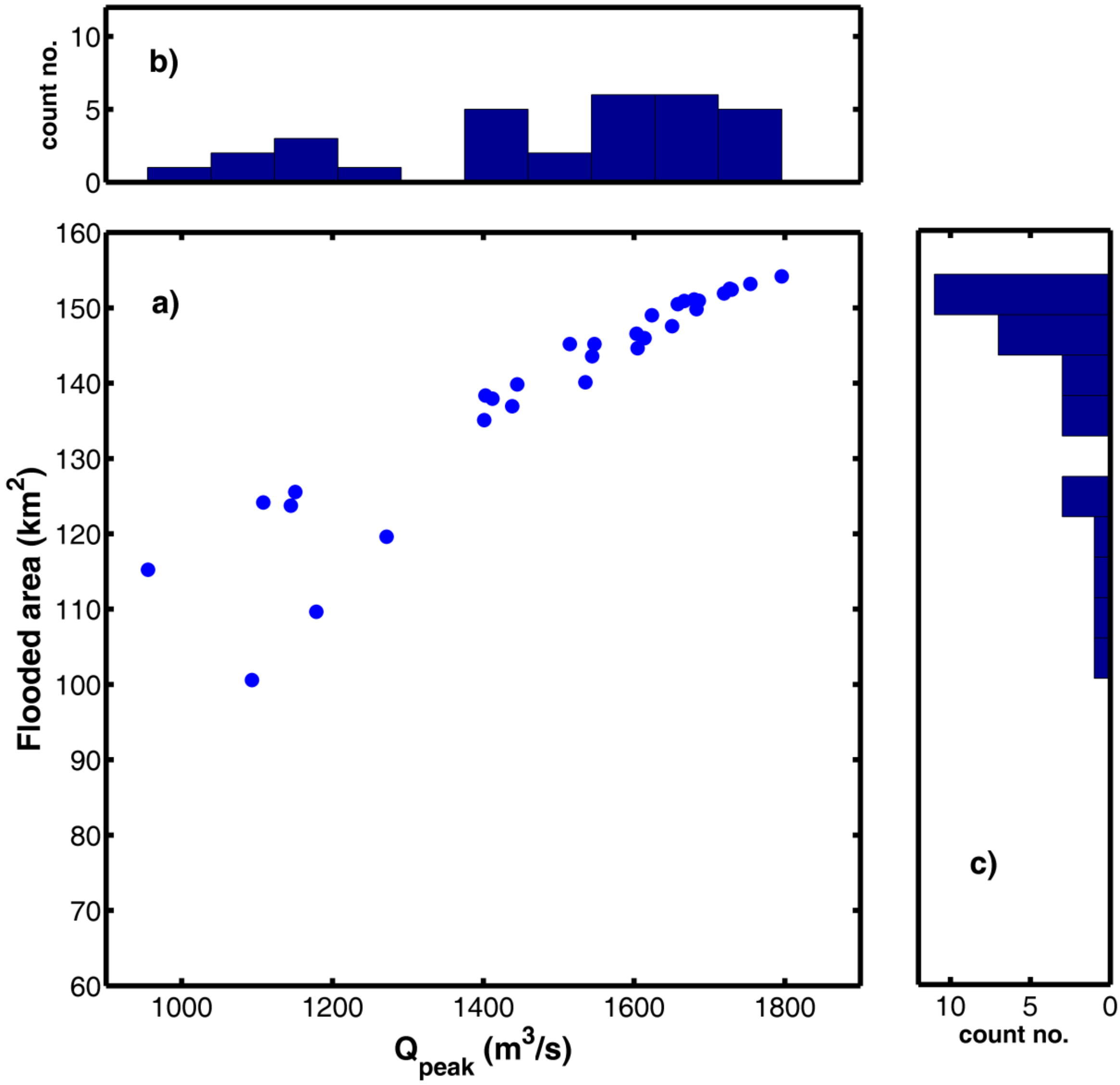

Figure 9. a) Maximum-flooded area vs. peak discharge estimated for all 31 hydrodynamic simulations of the 2009 flood event; b)Histogram of peak discharges; c) Histogram of estimated size of maximum-flooded area.

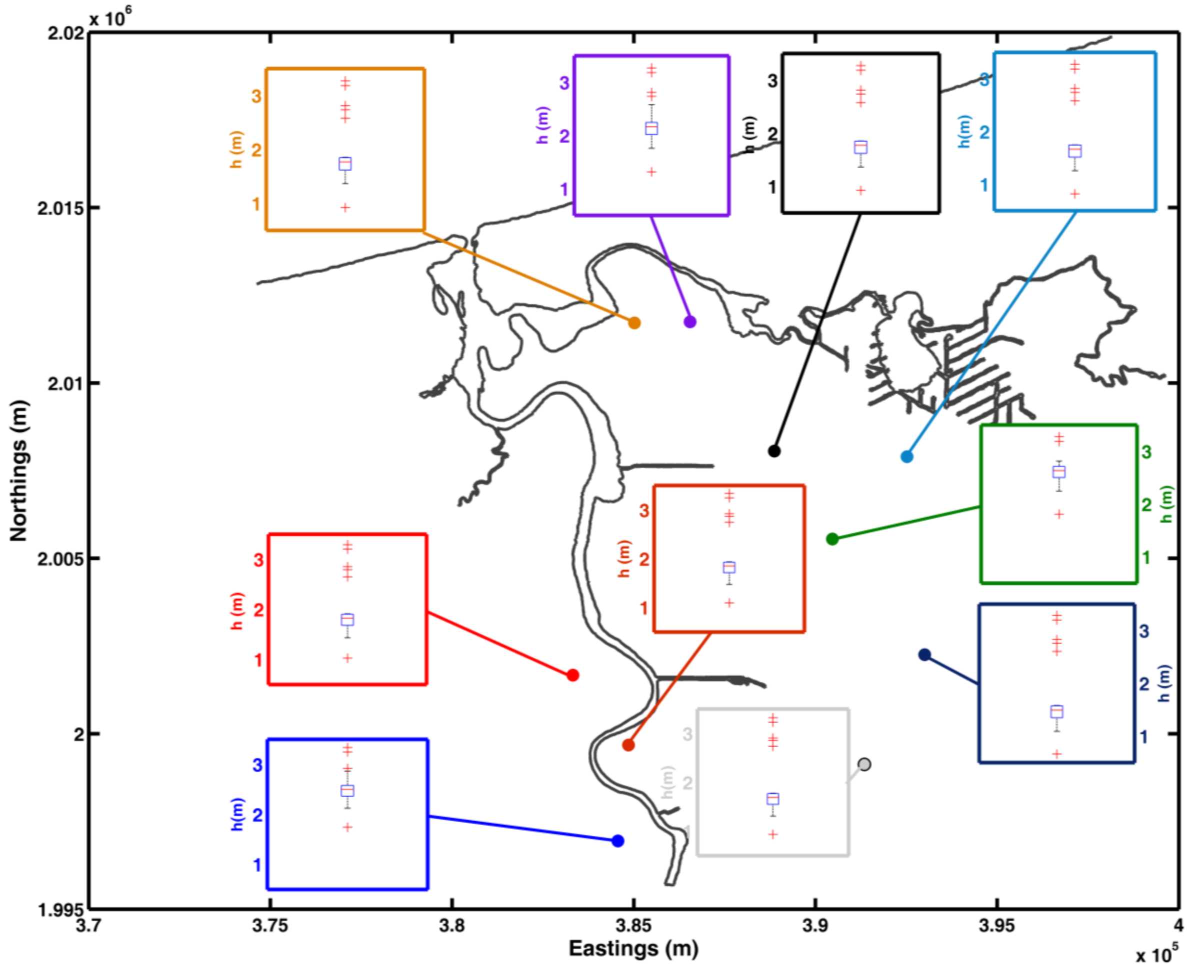

Figure 10. Estimated maxima inundation depths at different locations within the floodplain. Red line represents the median. Bars correspond to the standard deviation. Upper and lower limits of the box are the values of the 25th and 75th , respectively. Crosses depict outliers.

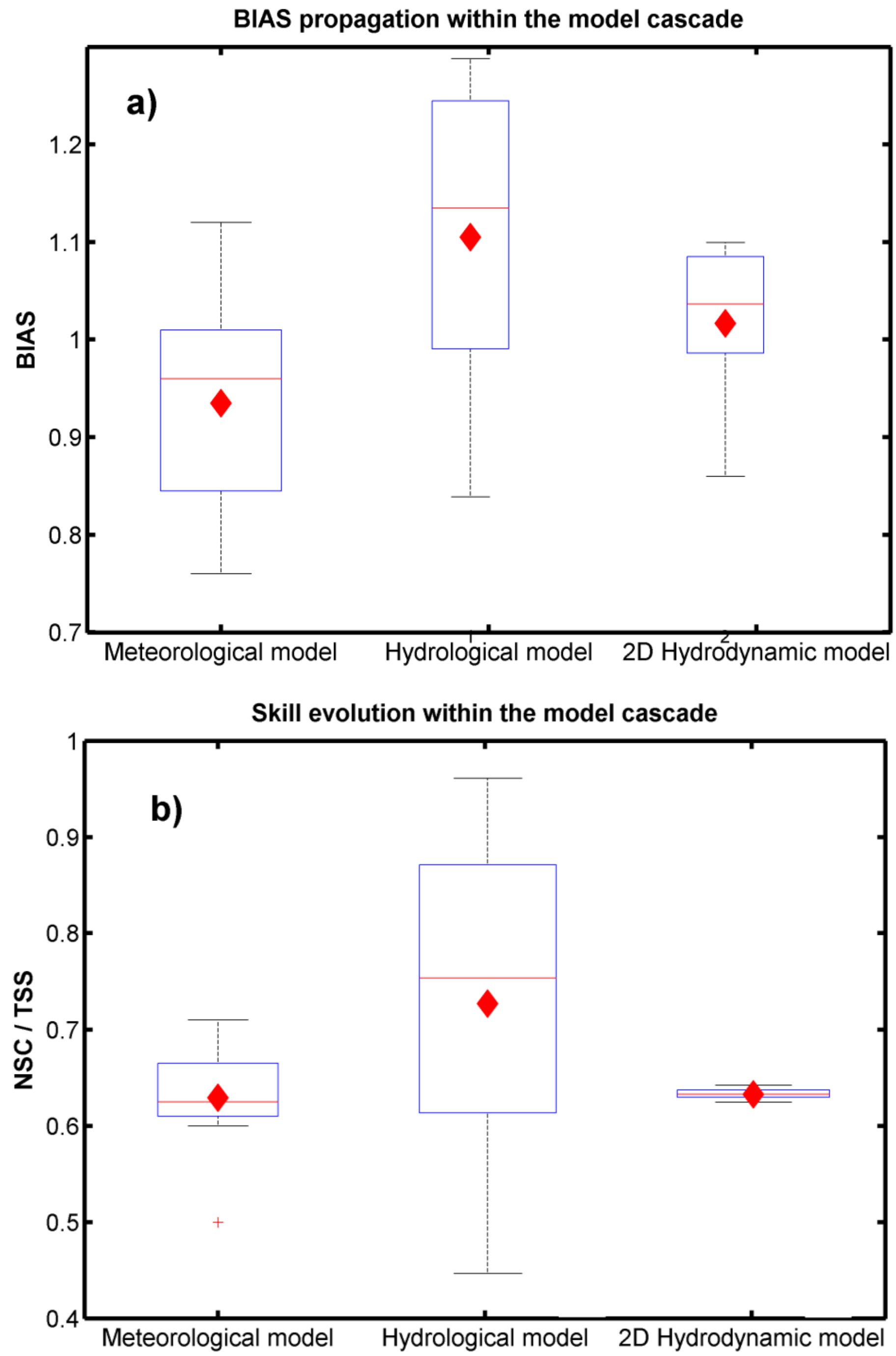

Figure 11. a) BIAS and b) Skill propagation within the model cascade (meteorological-hydrological-hydrodynamic); diamonds: corresponding ensemble mean value.