Review of Manuscript NHESS – 2013 – 217

Analysis of the French insurance market exposure to floods: a stochastic model combining river overflow and surface runoff

D. Moncoulon et al.

First, the authors would like to thank the anonymous Reviewer #1 for his helpful and useful comments. The authors will take into account the recommendations of this reviewer and try to improve the overall redaction of the paper. This letter is a direct answer to this reviewer. The paper will be improved and re-submitted when we will receive the second review.

About the "deterministic" model

We will add more references about the rainfall / runoff model: especially to explain the choices in the modeling methods and to justify them.

The equation eq. 14 is not clear. It is easier to explain it with the following equations.

The river flow celerity has a constant value during the flood event. This value will correspond to the speed limit of the runoff / rainfall model: the maximum water speed is the time needed for water to over-cross a DTM grid. The river segment length is 50m and the time-step of the model has been fixed to 120 seconds. Thus, maximum celerity C is 0.41 $m.s^{-1}$.

$$C = \frac{dx}{dt}$$

The kinematic wave model (a simplification of the Barré-Saint Venant equations) can be expressed by a single equation:

$$\frac{\partial Q}{\partial t} + \frac{dA}{dQ}.\frac{\partial Q}{\partial x} = q$$

With

$$C' = \frac{dA}{dQ}$$

We have:

$$\frac{\partial Q}{\partial t} + C'.\frac{\partial Q}{\partial x} = q$$

If *C'* is a constant value, then:

$$Q_{out}(t + t_0) = \sum_{i=1}^{n} Q_{in}$$

With t_0 the time spent by river flow to over-cross each river segment. *C'* can be determined from the water flow celerity. Q_{in} is the sum of the water flow entering the river segment, i.e. the sum of water volume entering river segment during the time-step t. $Q_{out}(t+to)$ is the sum of water volume leaving the river segment at time *t+to*.

The description of the damage model

The damage model has been built with 2 important parameters to calibrate:

- The claim frequency, depending on the hazard intensity at the risk address;
- The destruction rate, applied to the insured value, in case of claim occurrence.

The first parameter is calibrated by fitting a logit model to the empiric hazard intensity / occurrence of claim (0 or 1) on the calibration events. The logit model equation:

$$logit(p) = \ln(\frac{p}{1-p})$$

The logistic regression model is thus:

$$\ln\left(\frac{P(X|1)}{P(X|0)}\right) = a_0 + a_1x_1 + \cdots + a_ix_i$$

With *p* a value between 0 and 1.

The second parameter is based on the square root function calibrated with the historic claim database. It is important to precise that this calibration is made separately for every risk type (residential, commercial, industrial and agricultural). The event set for calibration is selected among the 1995-2010 events. An event becomes a calibration event when both the representativeness of the claim database and the quality of the hazard simulation (based on the river flow model scores) are adequate for modeling. The calibration event set is thus an evolving selection of events. It will depend on the improvements of the modeling methods and the data loaded inside the claim database. The event set used for the calibration of the model used in the paper is the following: Aude November 1999, Marseille September 2000; Nice December 2000; Bretagne December 2000 and January 2001; Gard September 2002 and 2005; Rhône December 2003 and November 2008; Meurthe et Moselle October 2006 and Saint Tropez October 2009.

The damage model is a deterministic model except at the end of the whole process: a bootstrap method is applied on the commune level losses (compared to the historical losses) to determine a confidence interval based on the differences between simulations and claim extrapolations. Indeed, this is the only part of the model that is stochastic. The authors think that, if we explain this in the paper, we can call the model deterministic, in order to separate it from the probabilistic approach.

After calibration, the model is used to simulate the whole historical event set (more than 100 events). The error between simulation and claim extrapolation (which is our reference) is then calculated to validate the calibration process. Each time a new version of the model is created, the whole event set is simulated to validate the version.

The error simulation, the so-called, second level uncertainty, is of course simulated for each event in the probabilistic (or Monte Carlo) simulations.

About the F1 model

The distributions fitted to the historical dataset for the simulation of flood series are the following:

- Weibull;
- Log-normal;
- Gamma;
- Generalized extreme values (GEV).

To fit the distributions to the data, the “maximum likelihood” method has been used. The quality of the model has been evaluated by visual comparison with the empirical distribution (a graph “histogram versus probability density” for example has been used). Other automatized statistical tests have been used: χ^2, to validate the global shape of the distribution, and Anderson-Darling to validate the tail of the distribution. If two or more distribution models fit with the data, the choice has been made with the Akaike Information Criteria (AIC):

$AIC=-2\ln(L)+2k$

Where *L* is the likelihood of the model, *k* is the number of parameters in the model and *N* is the size of the sample.

The problem of the historical dataset depth is clearly one of the most important limits of this approach. We select the stations with at least 30 years of historical recording to fit the distributions. We are aware that the distribution presents many uncertainties above 100 or 200 years (return period). But the intensity of our flood events is not only based on the return period of the river flow but also on the occurrence of many overflows over a large territory. When we generate 1000 years of river flow, the probability to generate a high return period flow on a single station is low. But we will simulate extreme events, in terms of insurance losses, due to the number of rivers concerned by the event. When we generate a thousand fictive years, we consider all the events generated to have the same probability of occurrence. But our simulation method will generate a large majority of non-intense river flow. Thus, the number of extreme events will be lower and we expect it to be a good estimation of the reality.

We try to estimate the important loss uncertainties for extreme quantiles. We have initiated a project to collect historical data on the major flood events, which could be used to fit statistical distributions (cf. Gaume and Payrastre recent works).

For the second question, I think the problem is due to the paper, which is not clear. We do not consider river flow at different stations for the same river to be independent. By using the Gaussian copula to generate spatial and temporal dependencies, we, on the contrary, consider them spatially and temporally dependent. But, when the monthly maximum river flow have been generated on the whole stations of the watershed, we do not verify the consistency of the hydrological values when we simulate the flood: the river flow, for example, at a station C, downstream of a tributary station B and the upstream station A will not necessarily verify: C= A + B.

About the F2 model

Yes indeed, the ETP is not represented. We have studied the impact of the ETP on the efficient rain for the majority of our 72h historical flood events. The table below shows the distribution of rain (r) and evapotranspiration (etp) and the ratio for the calibration events. For nearly 75% of the cases, the evapotranspiration on rainfall ratio is lower than 22% and for 50% of the cases, lower than 5% of the rainfall.

	r	etp	(etp/r)
Min	0,20	0,10	-
10ème	1,5	0,2	0,006
25ème	6,4	0,5	0,017
Médiane	18,0	1,1	0,050
75ème	43,1	2,0	0,222
90ème	80,5	3,1	1,063
99ème	218,4	4,9	12,000
Max	551,2	6,4	28,500
Moyenne	32,7	1,4	0,651

Yes the error distribution, or second level uncertainty, is calculated for each event of the F1 and F2 event set. Then, when we build the annual distribution, we select randomly the event and also, randomly, the loss percentile of the event. Thus, if we simulate 10 times the same event, we will have 10 different losses based on the distribution.

This is a misunderstanding of our method. We simulate 150 years of simulated rainfall with the TBM simulator (IRSTEA). Among these years, we detect the 72h events (which exceed the threshold of two year return period). We create a rainfall event database, used to simulate the hazard and damages with the deterministic model.

Then, as we did for F1 model, we generate 5000 years of 72h fictive rainfall by using the Gaussian copulas and taking into account both spatial and temporal dependencies. When a 72h fictive rainfall exceed the 2 years return period, we then look into the event database, select the event which has the same (or closest) 72h rainfall and randomly select its cost among the loss distribution.

About the combination of F1 and F2

Yes other methods were tested and we will improve this part of the paper in the reviewed version. In the F2 model, 5000 years are included.

Indeed, S is the maximum value of F2 losses and the F1 model is used for the tail of the distribution (extreme events). Eq. 15 refers only to F1 model, since the threshold S is the maximum value of simulated losses for the F2 model. Thus, no loss will exceed this value for F2.

We agree with your presentation of Eq. 16. We will change it to use the specific variable name, as you propose, T).

We agree, the paragraph 4.1 could be more useful in the method. We will propose a reviewed paper which will take it into account.

About the description of the MACIF portfolio

The ‘all perils’ is flood included. It means all Nat Cat perils: flood, drought, earthquakes, cyclonic winds, seasurge, etc.

We understand your remark about the useless of the individual risk distribution in the MACIF portfolio. We will simplify the portfolio description in the reviewed paper.

Example of the Argens floodplain

The model results for the June 2010 and November 2011 events in the Argens flood plain are the following:

Event losses (M€)	Simulation 10th percentile	Simulation 90th percentile	Historical losses (at the 6th of June 2013)
June 2010	138,5	610,1	462,7
November 2011	387,8	448,3	266,6

The historical losses for these recent events are not definitive. The market share of the claim database is evaluated to 41% for 2010 and 24% for 2011. The historical losses will be re-evaluated frequently in the incoming years, particularly for the November 2011 event. These events are not yet used for calibration and their estimation error are not used for the bootstrap. The model has overestimated the Argens flood losses in June 2010. This is especially due to the extreme elevation of water level in the rivers in a very short time. In Frejus, houses that have been built on the riverbanks have been totally destroyed by the flood and the mudslide. The calibration of the flood model is based on an historical claim database which did not contain any total destruction. Since 2010, the database has been significantly improved. These events will be taken into account in a future calibration. In 2011, the long duration of the event and its large spatial extension, has conducted the model to overestimate the flood losses.

The 1-10% of claim frequency inside the flood zone is the same range of frequency that we detect on the historical claim database for a given level of flood water. For example, for 1m of water level, we only find 10% of the insurance risks that will be damaged. Indeed, the particularities of the risk constructions are not taken into account in the database (an elevation of the first floor, a protection against the flood, a small elevation of the ground which is not taken into account in the DTM, etc.). Furthermore, the hazard model uncertainties are included in the claim frequency.

Conclusion

The conclusion will be written again to take into account your recommendations.

The authors- October, 3rd 2013