Estimating flood probability Bands Using Flood Event Data

Problem presented by

Adam Baylis

Environment Agency

ESGI0 was jointly hosted by
Smith Institute for Industrial Mathematics and System Engineering
and the University of Oxford

with additional financial support from
Engineering and Physical Sciences Research Council
European Journal of Applied Mathematics
Oxford Centre for Collaborative Applied Mathematics
Warwick Complexity Centre

Report author

Ilaria Prosdocimi (University of Bath)

Executive Summary

Estimates of the frequency at which floods are expected to happen in any location in England based on gauged streamflow data are made available by the Environment Agency and can be used to assess the amount of funding provided by Defra to develop local flood protection schemes and their potential benefits. Nevertheless in some areas it might be the case that a better assessment of the frequency at which floods occur in a specific location can be obtained by using local information such as the timing of flood occurrences in the most recent years. This type of data is less detailed than what is obtained by systematically gauging a stream and comes essentially in the form of a binary information in which only the information that a flood occurred in a given year is used. This type of data can be directly modeled by means of a Binomial distribution and the Bayesian inference framework can be used to fully use the fact that more often than not there is some prior knowledge on the likelihood of flooding in an area. In particular the Beta distribution can be employed as a prior distribution for the probability of flooding, and this results in a known distribution for the posterior distribution, which can then be estimated directly using simple arithmetic. Further the solution can be implemented in common spreadsheet software, for example Microsoft Excel. The importance of using a suitable prior and of using all possible information available are showcased and discussed, in particular in relation to practical guidelines that could be given when using this type of sparse and local data.

Version 1.0
November 5, 2017
iv+15 pages

Contributors

Emily Kawabata (University of Edinburgh)
Georges Kesserwani (University of Sheffield)
Ilaria Prosdocimi (University of Bath)
Paul Sharkey (University of Lancaster)
Nick Watkins (London School of Economics and Political Science)

Contents

1 Introduction: the industrial challenge

For flood management purposes the Environment Agency (EA) defines four flood probability bands which form the basis of the calculations behind the potential benefits and prioritisation of flood defense schemes and the total funds which DEFRA might transfer to local authorities to implement flood defense schemes. These bands are defined by the probability of at least one flood happening in any given year in the area of interest. Flooding here is defined as the event of fluvial water covering substantial parts of land which are normally dry at any point in time during a year. These probabilities are derived as the return period of a flood, i.e. the amount of years that on average it would take for one flood to occur in an area. For example, if an area is expected to flood on average every 30 years this would correspond to a probability of flooding in any given year of $1/30 = 0.0\bar{3}$. The specification of the bands is given in Table 1

Table 1: Risk bands specification

Band name	Lower bound	Upper bound
Very significant	1 in 30	1 in 1
Significant	1 in 100	1 in 30
Moderate	1 in 1000	1 in 100
Low	0	1 in 1000

These calculations rely on the assumption that the risk of flooding is constant throughout time (i.e. that the risk of flooding is the same in any given year) and do not take into account the case in which multiple floods happen in a given year, but simply assess the probability of at least a flood happening in a year.

Currently, it is possible to assess the flood band associated to any location in England using the Flood Map for Planning, a national map produced by the EA and available at `https://flood-map-for-planning.service.gov.uk/`. Although the Flood Map for planning has been developed to be a reliable tool which can be used to assess the flooding risk in any location in the country, it is possible that for some specific area the resulting estimates are not reliable and the actual number of flooding occurrences is higher than what would be expected under the large scale averages provided by the map. The EA therefore challenged the MathForesees scientists to develop a method and a tool which would assess the flood band which a specific area belongs to using only historical records of floods occurrence in the area. The method should in particular be practically easy to implement so that it can be accessible to communities which might claim only a small sum to design smaller scale flood protection schemes and would not have enough resources to employ a flood hydrologist to make a traditional flood risk assessment study.

2 From the challenge to a statistical model

The type of data which the EA imagines could be available to local communities who wish to use historical records to re-assess the flood risk of a specific area is a set of dates in which properties have been flooded in the last years. This is a very coarse dataset which can be further simplified and represented as a series of 0 (for years in which no flood occurred) and 1 (for years in which at least a flood did occur). It is further assumed that the risk of flooding is unchanged from year to year, a very strong and possibly unrealistic assumption, which is needed to make the problem easier to characterise. Finally, we are interested in finding θ, a parameter which indicates the probability of an area being flooded in a given year, i.e.:

$$\Pr[\text{The area is flooded}] = \theta \qquad \Pr[\text{The area is not flooded}] = 1 - \theta.$$

In other words θ represents the rate at which the area gets flooded, or the proportion of years in which flooding occurrs. Typically we would have information on the flooding occurrence over a certain period of time covering a total of n years, for which we know that in x years at least one flood has occurred. A useful distribution which can be used to characterize the total number of occurrences of an event over a certain number of trials is the Binomial distribution, defined as:

$$\Pr[\text{Observing x floods over n years}] = \binom{n}{x}\theta^x(1-\theta)^{n-x}.$$

Note that to use the Binomial distribution we need to make the additional assumption that the occurrence of flooding in a given year is independent of the occurrence of flooding in any other year. This therefore implies that we are equally likely to see a flooding event in any point in time in the n years for which information is available. We are interested in making statistical inference on θ, the probability of flooding occurring in any given year, which can then be used to make an assessment on the probability band to which an area belongs. A natural estimate for θ would be the ratio between the number of years in which a flood was observed over the total number of years for which information is available: x/n. This natural estimate in fact corresponds to the maximum likelihood estimator, which is found as the value of θ which maximizes:

$$L(x, n; \theta) = \pi(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}. \tag{1}$$

Following standard practice, an analytical solution to equation 1 is found as the zero of the first order derivative of the log-likelihood function ($\log(L(x, n; \theta))$). This corresponds to $\hat{\theta}_{ML} = x/n$. To make inference on properties of the parameter of interest θ we can then rely on classic results for maximum likelihood estimators, which are unbiased, efficient and asymptotically normally distributed:

$$\hat{\theta}_{ML} = \frac{x}{n} \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right). \tag{2}$$

Given that θ is unknown it is not possible to directly calculate the variance of the normal distribution in equation (2) so the following approximation is generally used to make inference on the parameter of interest θ:

$$\hat{\theta}_{ML} = \frac{x}{n} \sim N\left(\theta, \frac{\hat{\theta}_{ML}(1-\hat{\theta}_{ML})}{n}\right) \tag{3}$$

Using the distribution in equation (3) it is possible to investigate whether an area belongs to a specific risk band by constructing an hypothesis test. For example one could investigate whether an area belongs to the very high risk area using an hypothesis test such as:

$$H_0 : \theta \geq 1/30 \qquad \text{versus} \qquad H_1 : \theta < 1/30.$$

Using the traditional frequentist approach is nevertheless somewhat unsatisfactory in this case as the analyst would be forced to do a series of nested hypothesis tests. Further the asymptotic results on which maximum likelihood theory relies would require a fairly large n, especially in the case in which θ is small as it is likely to be the case in this context. We therefore pursued a Bayesian inference for the problem at hand, see Gelman et al. (2013) among others for an introduction to Bayesian methods. In this particular case a Bayesian approach to statistical inference can be particularly beneficial in light of the following considerations:

- we are specifically interested in the distribution of θ as we would wish to make an assessment on the flood risk band in which an area is likely to belong;
- a Bayesian framework provides a more natural representation of the uncertainty in the parameters, which is the main object of interest for the specific problem; and
- we wish to make use of prior information on flooding behaviour, on which we are likely to have some loose information, possibly based on external sources such as the national flood maps or some geophysical characteristics of the area under study.

Further, by using the Beta distribution to characterize the prior information on θ, it is possible to obtain an analytical formula to represent the posterior distribution of θ, which is simply a Beta distribution with parameter values which can be computed from the prior assumptions and the observed data (this is shown in equation (8)). The Beta distribution is defined on the $[0, 1]$ domain and is therefore a natural candidate to characterise a parameter which must take values in $[0, 1]$: the highly flexible nature of the Beta distribution, is also advantageous for since it allows for very varied shapes of the prior distribution. Moreover, the Beta distribution is a conjugate prior of the Binomial distribution with unknown parameter θ, which means that the distribution of the posterior distribution of θ is found to be the same as the prior distribution, i.e. a Beta distribution. The use of a conjugate prior allows for simple analytical solutions to the problem of estimating the posterior of θ, and avoids the use of resampling techniques often used in Bayesian inference which might make the estimation slower and more complicated to implement for non-specialists.

The probability density function (pdf) of a Beta distribution is defined on [0,1] and parametrised by two parameters α and β. If $T \sim \text{Beta}(\alpha, \beta)$ the pdf of the distribution, denoted as $\pi(\theta)$, is defined as:

$$\Pr(T = \theta) = \pi(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \quad \text{for } 0 \leq \theta \leq 1; \alpha, \beta > 0. \tag{4}$$

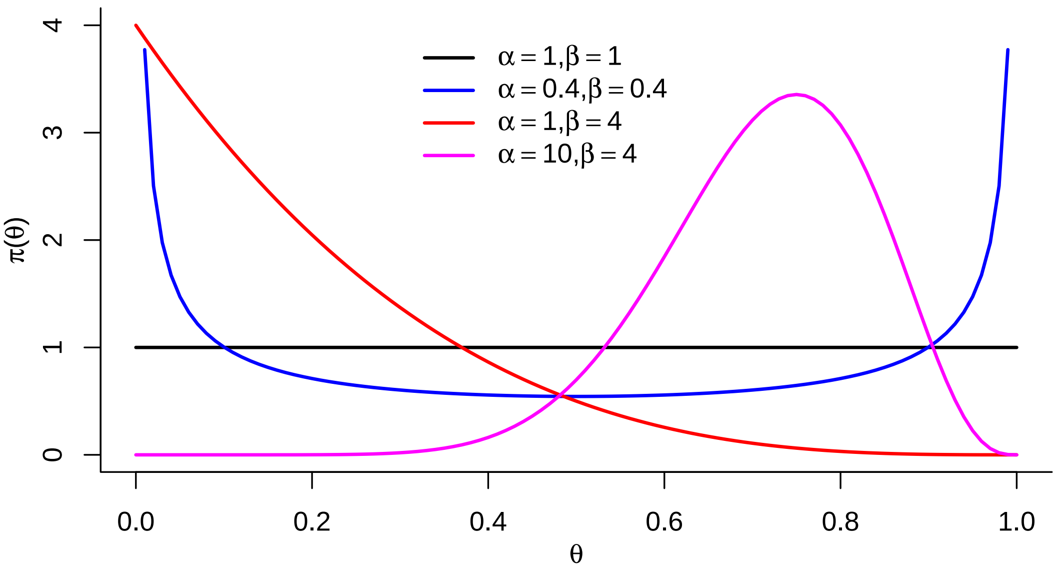

Figure 1: The density of the Beta distribution for the parameter of interest $\theta = \Pr{The area is flooded}$ for several different α and β values

The expected value and variance for a Beta distribution are found to be:

$$E[T] = \frac{\alpha}{\alpha+\beta} \qquad \text{and} \quad V(T) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{5}$$

The Beta distribution is extremely flexible and can take several shapes, as shown in Figure 1.

The functional form of the posterior distribution can be derived by the direct application of Bayes theorem, which states that:

$$\pi(\theta|x) = \frac{\pi(\theta)\pi(x|\theta)}{\pi(x)}, \tag{6}$$

where $\pi(\theta|x)$ indicates the posterior distribution of the parameter of interest θ given (i.e. after we have observed) the flood occurrence data; $\pi(x|\theta)$ indicates the likelihood of the observed data given the parameter θ and corresponds to the likelihood traditionally maximised in frequentist approaches; $\pi(x)$ indicates the marginal likelihood; and $\pi(\theta)$ indicates the prior distribution of the parameter of interest, i.e. the information we have about the parameters before observing the occurrence data. As mentioned above we take $\pi(x|\theta)$ to be a Binomial distribution and $\pi(\theta)$ to be a Beta(α, β) distribution, where α and β are the parameters which characterize the prior distribution of the parameter of interest θ and are referred to as hyper-parameters. Note that $\pi(x)$ is unknown, but we can write

$$\pi(\theta|x) \propto \pi(\theta)\pi(x|\theta) \tag{7}$$

and use the fact that, by definition, $\pi(\theta|x)$ must integrate to 1. So we find,

$$\begin{aligned}\pi(\theta|x) &\propto \pi(\theta)\pi(x|\theta) \\ &= \theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{x}(1-\theta)^{n-x} \\ &= \theta^{\alpha+x-1}(1-\theta)^{n+\beta-x-1}, \end{aligned} \tag{8}$$

which is proportional to a Beta$(\alpha+x, n+\beta-x)$. The characterisation of the posterior distribution of θ can then be easily found by simply combining the hyper-parameters α and β with the relevant information of the observed data. We also find that

$$E[\theta|x] = \frac{\alpha+x}{\alpha+n+\beta} \quad \text{and} \quad V(\theta|x) = \frac{(\alpha+x)(n+\beta-x)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}. \tag{9}$$

From the equations above we can see that as $n \to \infty$, the information in the prior becomes less and less relevant and is replaced by the mean of the observations x/n. Similarly, as $n \to \infty$ the value of the posterior variance decreases and tends to the variance used in frequentist inference. On the other hand, in the case of flood band risk estimation, it is likely that n and x will not be very large numbers and the choice of the hyper-parameters α and β will play an important role in the final estimation.

3 Some thoughts on the prior $\pi(\theta)$

The unadventurous analyst could think that the "safest" Bayesian inference could be obtained by taking a non-informative vague prior which assigns uniform probability to all the range of parameter values θ, taking $\pi(\theta) = \text{Beta}(1,1)$. This corresponds to a Uniform$(0,1)$ distribution, as seen in Figure 1. In terms of the problem at hand this corresponds to assuming, before having any information on the location of interest, that any of the flooding probabilities in $[0,1]$ are equally likely to be appropriate to describe the probability of flooding in the area of interest. This is quite extreme, as it assigns equal probabilities to the case in which flooding is a very common or very rare event. In reality, in most cases it is true that flooding is a fairly rare event, which is not likely to happen every year. It would therefore be advisable to use some more informative prior distribution which can represent the common understanding that flooding is likely to be an event which happens with moderate to low probability. It is enlightening to realise how the Beta(1,1) is a poorly specified prior when one thinks that if we know that for an area no flooding has been observed for the last n years, the expected value of the posterior distribution of θ, as derived from the formulas in eq. 9, would be $\text{E}[\theta|x] = 1/(n+2)$. In the (common) case in which n is not very large, this results in posterior probabilities with a large proportion of mass for high values of θ, which would result in the area under study to be placed in the very high risk band. This is somewhat counter-intuitive because if an area has not flooded for some years we would be inclined to think that there is some evidence for it to be located in the low risk band. This is caused by the fact that the small sample size can not compensate for the fact that the prior

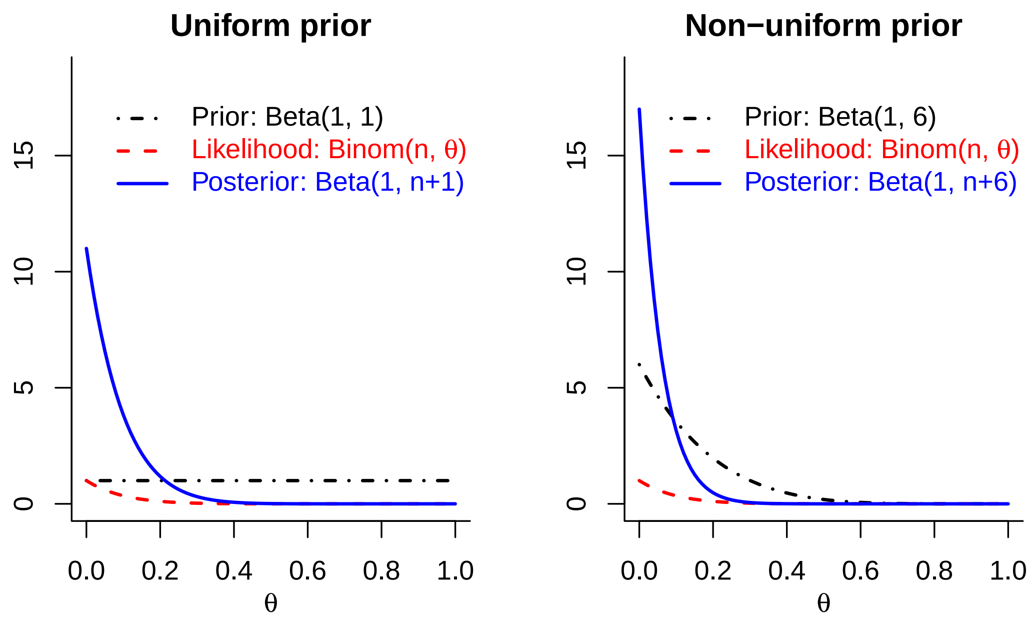

Figure 2: The likelihood function together with the priori distribution and the resulting posterior distribution $\pi(\theta|x)$. The sample used for the calculations had a total of $x = 0$ flood records in $n = 10$ years

distribution assigned a large mass to high values of θ. This is shown graphically in Figure 2 where the prior distribution $\pi(\theta)$, the likelihood function $\pi(x|\theta)$ and the posterior distribution $\pi(\theta|x)$ are shown together in each plot. Using a non-uniform prior in which more probability is assigned to smaller values of θ, as the one in the right hand side of Figure 2, results in a sharper posterior which assigns smaller probabilities to high θ values. Although the difference in the posterior estimate might seem minimal in this example, it is worth pointing out the importance of using suitable priors in the estimation procedure, especially in the case in which the available information only covers a few years.

It is conceivable that the Environment Agency should be able to construct sensible prior distributions for flooding in a given region based either on some large scale estimate of flood risk or on some geomorphological properties of the area (for example the proximity to a river and the presence of ponds or other forms of natural flood attenuation instances). On the other hand the use of very informative priors is likely to have a large influence in the final results, so the definition of the prior probability is likely to be a matter of discussion and possible disagreement between the Environment Agency and local communities. It is recommended that some care is taken in the choice of the parameters of the prior distribution and it is advisable to carry out a sensitivity analysis to explore the impact of the choice of prior on the final result.

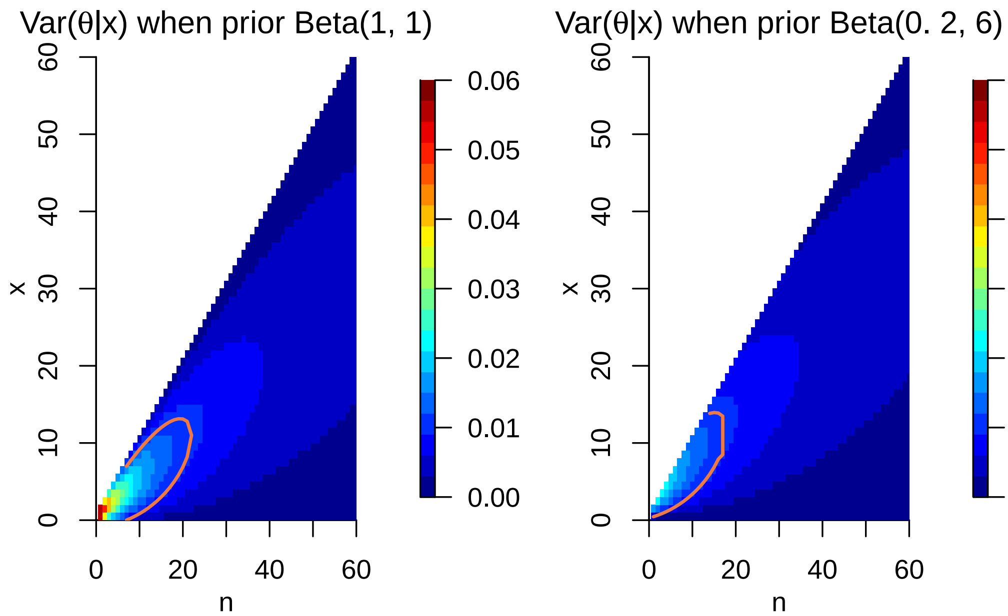

Figure 3: The variance of the posterior distribution for different sample sizes (n) and different number of occurances (x) using two different prior distributions. The contour lines such that $\text{Var}(\theta|x) < 0.01$ are also shown

4 Properties of the posterior estimation and their impact in decision making

One of the collateral questions that the EA raised in the challenge was whether it was possible to derive some general rule on what could be considered enough evidence from a local community to trigger a decision from the EA on considering to move an area to a different risk band. In other words the EA was asking what is the sample size that can be deemed to be enough evidence? The first point to make to answer this question is that the evidence brought by the communities should not be only about the number of floods which happened in the recent years but should also be about the period of time for which these flood occurrences are the only flooding which happened. Essentially since both x, the number of floods, and n the number of years for which information is available, are key part of the inferential process, an effort should be done to ensure that both items of information are known as precisely as possible. Prosdocimi (2017) discusses methods to estimate n as the unknown maximum of a population size, a problem known in statistics as the German Tanks problem, but it is recommended that communities should provide evidence of the time coverage (n) as well.

Assuming that the information on x and n has been retrieved reliably and that a suitable prior distribution has been chosen it is then possible to proceed to the estimation of the posterior distribution and use this to make inference. One can then fix some desirable properties which the posterior distribution should have for the data to be used as strong enough information to trigger an action from the

EA. Pham-Gia and Turkkan (1992) and Pezeshk (2003) discuss some methods to assess sample size needed to achieve some desirable properties for the posterior distribution. In particular, Pezeshk (2003) specifies that several of the methods discussed in the literature are well defined in principle but possibly difficult to implement in practice. It is therefore suggested that the sample size required in the analysis might be simply obtained by requiring that the posterior variance is small, i.e.

$$E[\text{var}(\theta|x)] < \epsilon,$$

where ϵ is a prespecified small constant. It is shown in Pham-Gia and Turkkan (1992) that when using an non-informative uniform prior (i.e. a Beta(1,1)) a sample size of $n = 22$ would be enough to achieve $E[\text{var}(\theta|x)] < 0.01$ for any value of x. If an informative prior $\pi(\theta)$ is used smaller sample sizes are needed to achieve prescribed variance of less than 0.01, as shown in Figure 3. Note that in Figure 3 it is also evident how the length of the record needed to achieve a pre-specified variance also depends on the number of flooding events recorded in the record. Another possible practical approach which could be used would be to require that the posterior probability that the area belongs to a certain flood band (i.e. that $\theta \in \Theta_0$, with Θ_0 an appropriately defined subset of [0,1]) is larger than a certain value. This would correspond to requiring posterior distributions with little variability and with a large proportion of probability mass spread over specific intervals. For example, in Table 2 the probability of θ to belong in each one of the flood risk bands is shown for four different flood occurrence samples when using a weakly informative prior $\text{Beta}(1, 3)$. The prior distribution was chosen arbitrarerly to be somewhat asymmetric with more probabilty mass in the left part of the distribution (the median of a Beta(1,3) is approximately 0.21). It is clear that in the case in which the information available covers $n = 30$ years it is much more informative to also know that $x = 3$ floods, rather than $x = 1$ flood, have occurred: the disparity between the probabilities for the $(1/30, 1]$ interval is very large. On the other hand, if the available record covers a longer time period of $n = 125$, even when it is known that $x = 3$ floods (and not $x = 1$ flood) have occurred this does not allow for a clear differentiation between the case in which $\theta \in (1/30, 1]$ and $\theta \in (1/100, 1/30]$. The plots in Figures 4 and 5 show the posterior distributions for the same estimation cases presented in Table 2: again it can be seen that when the available information covers a total of 30 years knowing that 3 floods occurred drastically changes the probability of $\theta > 1/30$ and the probability of the low and moderate risk classes. The large reduction in the probabilities connected to the low and moderate risk classes can be noticed also when the available information covers a period of time of $n = 125$ years, but in this case the very significant risk class still has some non-negligible probability attached to it. These type of tables and plots could be useful for the Environment Agency to make an assessment to whether the available information is enough to warrant a change in the flood risk band to which an area belong.

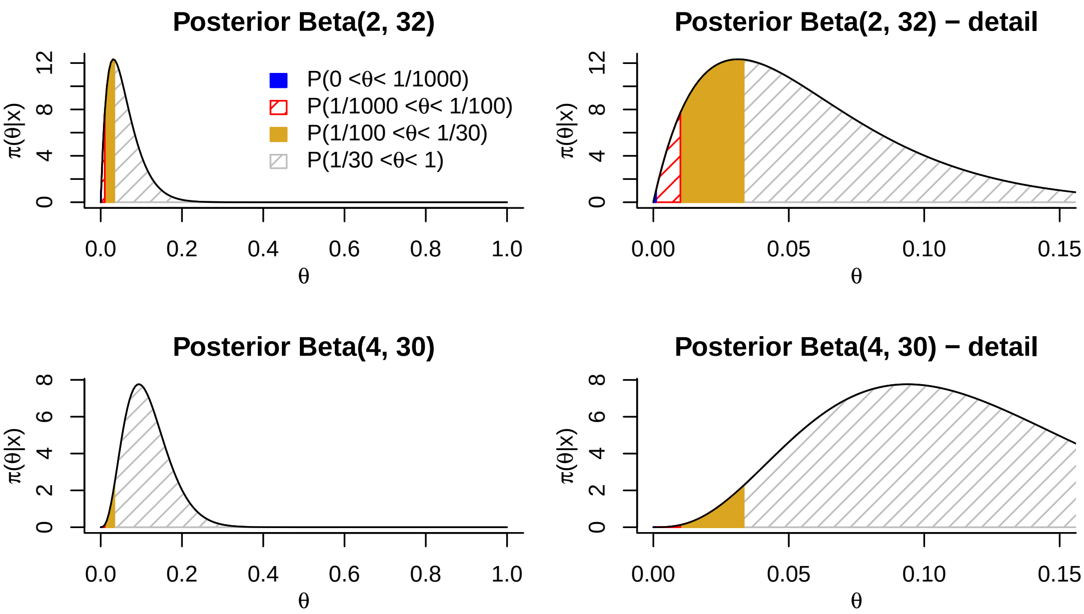

Figure 4: The posterior $\pi(\theta|x)$ divided in subsets corresponding to the flood risk bands used by the EA for the two cases covering a total of $n = 30$ years. Plots in the top panels show the posterior obtained when $x = 1$, plots in the bottom panels show the posterior obtained when $x = 3$. Beta(1,3) (vaguely informative) is used as prior distribution.

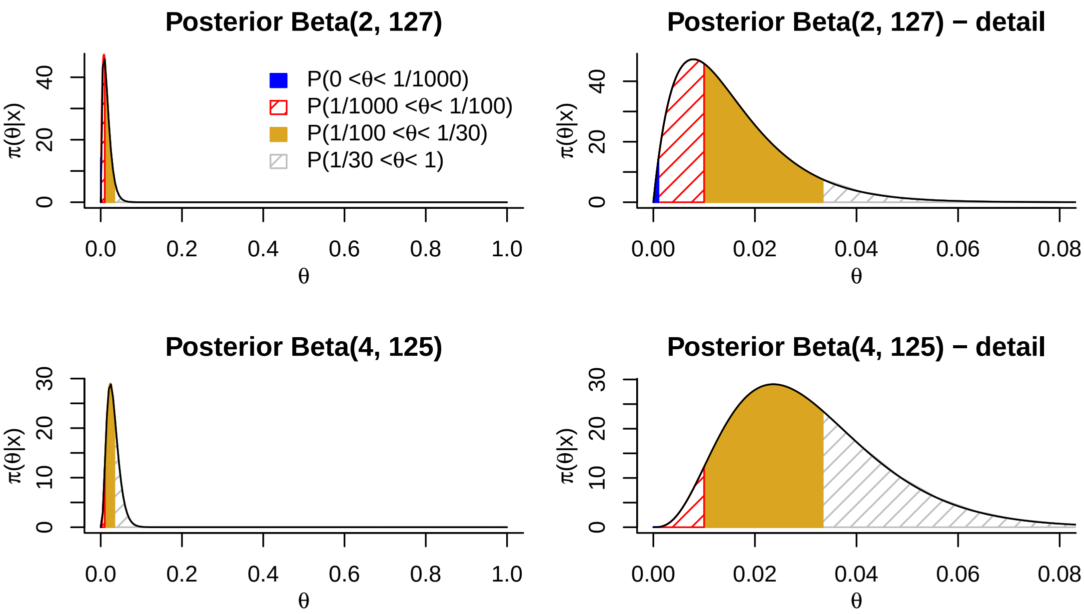

Figure 5: The posterior $\pi(\theta|x)$ divided in subsets corresponding to the flood risk bands used by the EA for the two cases covering a total of $n = 125$ years. Plots in the top panels show the posterior obtained when $x = 1$, plots in the bottom panels show the posterior obtained when $x = 3$. Beta(1,3) (vaguely informative) is used as prior distribution.

Table 2: Probability that θ belong to given sets corresponding to the flood risk bands used by the EA for different flood occurrences sample. Beta(1,3) (vaguely informative) is used as prior distribution.

θ value	Band name	Sample properties			
		$n = 30$ $x = 1$	$n = 30$ $x = 3$	$n = 125$ $x = 1$	$n = 125$ $x = 3$
$\theta \in (1/30, 1]$	Very significant	0.6984	0.9766	0.0706	0.3793
$\theta \in (1/100, 1/30]$	Significant	0.2585	0.023	0.5628	0.5804
$\theta \in (1/1000, 1/100]$	Moderate	0.0425	3×10^{-4}	0.3591	0.0403
$\theta \in [0, 1/1000]$	Low	5×10^{-4}	4×10^{-8}	0.0075	1×10^{-10}

5 Implementing the estimation in Excel

One of the great advantages of using a conjugate prior and having therefore a posterior distribution which follows a known distribution is that the estimation procedure can be performed using non-specialised software such as Microsoft Excel. An Excel spreadsheet which makes it easy to input the prior information and the available data and produces information on the posterior has been developed and shared with the Environment Agency during the Study Group working week. Examples of the spreadsheet in action can be seen in Figure 6 and 7.

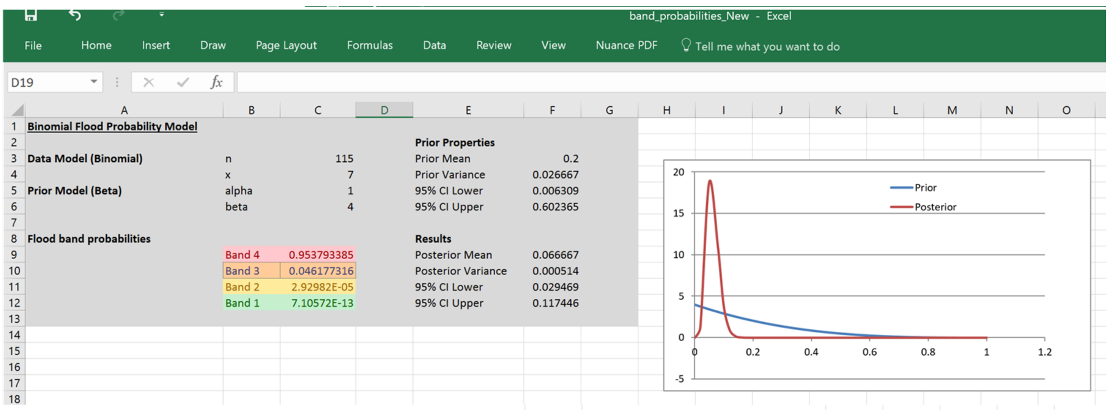

Figure 6: A possible Excel sheet implementation of the inferential procedure - in this case the prior is Beta(1, 4), the sample has $x = 7$ flooding records in $n = 125$ years.

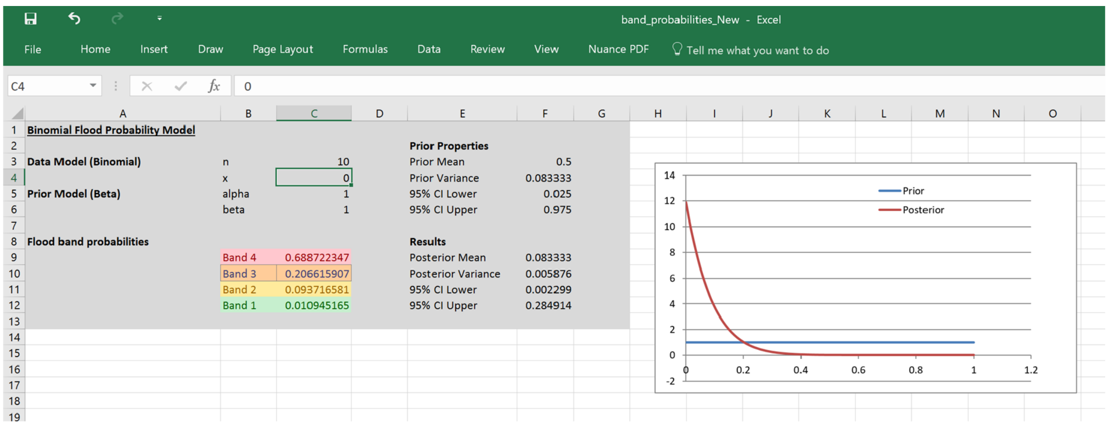

Figure 7: A possible Excel sheet implementation of the inferential procedure - in this case the prior is Beta(1, 1), the sample has $x = 0$ flooding records in $n = 10$ years.

A modified version of the spreadsheet could be made available to local communities which would then be able to make an assessment as to whether there is enough evidence to claim a change in the flood risk bands.

6 Additional topics discussed during the study group

The methods outlined in the previous sections give a fairly simple and practical solution to the main challenge posed by the EA. Nevertheless the data which is effectively available to local communities might be of a slightly different form from what is assumed for the Binomial distribution and might be affected by other issues. A brief list of some of the topics discussed is given below.

Identification of trends and flood rich/flood poor periods
It is possible that the flood risk in a place is indeed not constant through time but might be affected by some external factors, for example climate change or changes to the upstream catchment. If a long enough record of years in which flooding has occurred or not occurred is available it could be possible to use logistic regression (either in a traditional frequentist framework or in some Bayesian approach) to test whether a change in the flooding risk can be detected. Another approach which might also be used to detect whether a significantly higher number of events has been recorded in recent years is that of point processes. Finally it has been mentioned that it is a well known fact that flood poor and flood rich periods naturally exist and this would most likely result in overdispersed binary data (i.e. data with a variance higher than what one would expect from a binary model). It could be useful to model this potential overdispersion and assess whether some form of clustering can indeed be detected in the data.

Non-constant risk
Beside the presence of flood-rich and flood poor periods, it could also be the case that some areas might be at higher or lower risk of flooding in some periods of time due to changes in the flood protection management in the catchment. For example, it could happen that a large flood in a given year would break some flood defences, thus making flooding more likely in the near future. Regardless of the cause of the change in the flooding risk, models which can accommodate for a change in the parameter θ throughout time would be beneficial. A possible avenue would be the use of state-dependent models which have been applied in communication research.

Number of floods/Multi-peaked events
The type of data used in the work presented till now is of a very coarse nature. The fact that only the binary information flood/no flood in one calendar year is used means that the case in which an area has been flooded several times in a year is treated in the same way as the case in which the area has flooded only once in a year. It is possible to model the number of flooding events rather than the binary information (if this information is available for all years) by using a Poisson distribution as the likelihood for the data. By using a Gamma as a conjugate prior a Negative Binomial is obtained as a posterior. This model could also be used to assess whether the number of floods is increasing by performing a (Bayesian) Poisson regression. If this modelling approach is to be pursued, it is of key importance

that the number of flooding events for a given year correspond indeed to hydrologically independent events, i.e. events which can be clearly deemed not to be the continuation of one another.

Climate change adaptation

It is generally accepted by the climate scientific community that anthropogenic changes to climate have an impact on the rainfall frequencies and intensities in the UK and this results in potential changes to the frequency of floods. To ensure that infrastructure and flood risk management strategies can be resilient to these changes, it is recommended the the estimated design floods are corrected by some factor which accounts for the projected effects of climate change (see for example Environment Agency, 2011). In practical terms this means that, if the 1-100 year event Q_{100} is estimated to become larger, structures should be designed using a corrected $\gamma * Q_{100}$ (with $\gamma > 1$). A similar approach could be used to correct a point estimate of the probability of flooding in an area: if the present day estimate is θ, the future estimate could be a corrected value $\gamma^* * \theta$. The value of γ^* would need to be estimated, possibly using the projected climate change scenarios. Another possible approach to develop more resilient flood management strategies could be to rethink the definition of the flood risk bands. As mentioned above it is projected that the 1-100 year event will become larger in the future: this would also correspond to the fact that what is now the 1-100 year event will be in the future an event which will occur at a higher rate than 1/100, say for example 1/80. This would mean that an area that was located in the "moderate" risk band (i.e. flooded with probability $1/1000 < \theta \leq 1/100$) would in the future be located in the "significant" risk band (i.e. flooded with probability $1/100 < \theta \leq 1/30$). By determining the realistic changes in the frequency to which the present 1/30, 1/100,1/1000 events will occur in the future it would then be possible to define some different risk bands based on the risk of the area in the future decades.

Both the methods sketched here are based on the understanding of future risk obtained by climate change projections at a large scale. On the other hand changes to flood risk in an area might be due to very local reasons and possible corrections to flood risk management strategies could be devised using some local knowledge on the nature of the changes which are happening in a specific area. Further, the topic of quantifying the change that is affecting an area is interwoven with the issue of the identification of trends discussed above. Ideally correction factors for potential changes in flood risk should be derived or at least harmonised with changes detected in the actual flood records.

References

Environment Agency (2011) Adapting to Climate Change: Advice for Flood and Coastal Erosion Risk Management Authorities. *Environment Agency, Bristol, UK.*

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin (2013) Bayesian Data Analysis, 3rd edition. *London: Chapman and Hall.*

Pezeshk (2003). Bayesian techniques for sample size determination in clinical trials: a short review. *Statistical Methods in Medical Research 12*(6),1135–1147. doi:10.1191/0962280203sm345oa.

Pham-Gia, T. and N. Turkkan (1992). Sample Size Determination in Bayesian Analysis. *Journal of the Royal Statistical Society. Series D (The Statistician) 41*(4), 389–397.

Prosdocimi, I. (2017). German tanks and historical records: the estimation of the time coverage of ungauged extreme events. *Stochastic Environmental Research and Risk Assessment. Forthcoming*, doi:10.1007/s00477-017-1418-8.