Ward, P. J., Jongman, B., Aerts, J. C. J. H., Bates, P., Botzen, W. J. W., Diaz-Loaiza, A., ... Winsemius, H. C. (2017). A global framework for future costs and benefits of river-flood protection in urban areas. *Nature Climate Change*, *7*(9), 642-646. https://doi.org/10.1038/nclimate3350

A global framework for future costs and benefits of river flood protection in urban areas

Supplementary Information

Ward, P.J., Jongman, B., Aerts, J.C.J.H., Bates, P.D., Botzen, W.J.W., Diaz Loaiza, A., Hallegatte, S., Kind, J.M., Kwadijk, J., Scussolini, P., Winsemius, H.C.

Table of Contents

1. Selection of RCP/SSP combinations

In this study, we used Representative Concentration Pathways (RCPs)[1] and Shared Socioeconomic pathways (SSPs)[2] to represent future climate and changes in future socioeconomic conditions, respectively. In total, there are 4 RCPs and 5 SSPs, leading to a matrix of 20 combinations of projections[2].

In the main text of this letter, we focus on four combinations that explore a wide range of plausible future outcomes; the selection of these scenarios is described in the following paragraphs. We have also chosen to display the globally aggregated results of benefits, costs, and B:C ratios, in a pathway matrix (Supplementary Table 1), in which all possible combinations are shown. Note that the NPV can also be calculated from this matrix, by subtracting the costs from the benefits (both are already discounted). Moreover, in the Supplementary Dataset, we provide the results for all scenario combinations, and for all individual GCMs, per sub-national unit. Hence, all of the possible model and scenario combinations are available for users interested in a specific RCP/SSP/GCM combination.

We use the selection of 3 scenarios also described in Ref. 3, namely: RCP2.6/SSP1; RCP6.0/SSP3; and RCP8.5/SSP5. To this selection, we added a fourth scenario, namely RCP4.5/SSP2. The scenarios were selected to provide a broad range of plausible futures. Using the results matrix in Table 1, we checked whether this selection of RCP/SSP combinations does indeed give a good coverage of the scenario space. Examining the globally aggregated B:C ratios (for the optimise objective), we see that RCP8.5/SSP5 represents the highest ratios, RCP6.5/SSP3 represents the third lowest ratios, and the B:C ratios for RCP2.6/SSP1 (6.7) and RCP4.5/SSP2 (5.7) represent an even spread around the median B:C ratio of 6.2.

The rationale of the combination of each RCP/SSP combination is described below. The descriptions for RCP2.6/SSP1, RPC6.0/SSP3, and RCP8.5/SSP2 are taken directly from Ref. 3.

RCP2.6/SSP1: SSP1 implies that we move to the use of more green energy sources and reduction of emissions. Therefore RCP 2.6 is a plausible match with SSP1, in which it is assumed that the world makes relatively good progress towards sustainability.

RCP4.5/SSP2: SSP2 represents a middle of the road future, often termed ‘current trends continue’. In this world, trends typical of recent decades continue, with some progress towards achieving development goals, reductions in resource and energy intensity at historic rates, and slowly decreasing fossil fuel dependency. In combination with RCP6.0 or RCP8.5, this has often been used to represent a ‘Business as Usual’ Scenario. In light of the Paris Agreement of the United Nations Framework Convention on Climate Change, we have chosen to combine SSP2 with RCP4.5 to examine potential climate impacts if the goals of accelerating and intensifying actions and investments for climate change mitigation are achieved to some extent, but less so than under RCP2.6.

RCP6.0/SSP3: SSP3 can match with a large variety of emission scenarios. We have selected a quite conservative RCP to match with SSP3. SSP3 is an important scenario to include, as it gives very contrasting socio-economic conditions compared to other scenarios in developing countries. It is therefore a useful scenario to consider if the interest is in developing regions. The extremely high

population growth numbers in developing countries make this scenario important for a study on the socioeconomic impacts of floods.

RCP8.5/SSP5: SSP5 assumes that we continue relying on fossil fuels a lot. This implies that we may expect high emissions, and therefore the combination with RCP8.5 is very plausible. In terms of climate change, this is the most extreme scenario. It may result in both increase in flood hazard, in particular in regions where more excessive rainfall is expected, while it results in less severe flood hazard in regions that are becoming more drought prone as a result of climate change.

2. Sensitivity and robustness assessment

In this letter, we demonstrate the results of our global model framework for assessing costs and benefits of flood protection in urban areas by means of dikes. For reasons of conciseness and clarity, in the main text we show the results based on a limited number of assumptions and model parameters. Firstly, we only show results based on our middle-cost estimate. Secondly, we only show the results for the 4 selected combinations of RCPs and SSPs. Thirdly, all the results shown in the main text are averaged over the 5 GCMs used in this study. Fourthly, we only display the results using a discount rate of 5% per year. Finally, in order to carry out the experiments, we need an estimate of the current flood protection standard for each sub-national unit; for this paper we used values from the FLOPROS database[4]. We acknowledge that each of these assumptions introduces uncertainty. Therefore, we have also carried out a large number of additional simulations to assess the sensitivity of the results to these assumptions and parameters, and to assess the overall robustness of the results. These additional simulations are mentioned briefly in the main text; in this section we provide additional information. It is important to note that since the aim of this letter is to demonstrate the framework, we do not present a full uncertainty analysis across all model parameters. To do this, we believe that large multi-modelling and collaborative projects are required, like those carried out for the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). A full uncertainty analysis would require using more than the 1 hydrological model used here, and also perturbing all of the hydrological model parameters, as well as using different Digital Elevation Models, river network datasets, and so forth. Moreover, it would involve using multiple datasets of projected future changes in socioeconomic conditions, and ideally different impact models. Such an exercise is beyond the scope of this letter. However, the framework has been set up such that it can be integrated with other datasets and models when they become available.

In this section, we elaborate on the sensitivity and robustness analyses referred to in the main text of the letter. The following analyses are presented: (a) sensitivity to different RCP/SSP combinations; (b) sensitivity to different GCMs; (c) sensitivity to different discount rates; (d) sensitivity to different cost estimates; (e) sensitivity to assumed baseline protection standards; and (f) overall robustness. Each of these aspects is discussed in the following sub-sections. Note that we hereby focus mainly on results for the 'optimise' objective. Also, we have made all results available in a Supplementary Dataset for: all RCP/SSP combinations; each individual GCM; different discount rates; high, middle, and low cost estimates; and different assumptions on assumed baseline protection. This allows users who are interested in the uncertainty in a particular part of the modelling chain to examine our results in more detail.

2.1. Sensitivity to different RCP/SSP combinations

As explained in Supplementary Information 1, in this letter we used four combinations of RCPs and SSPs to represent future climate and changes in future socioeconomic conditions, respectively. The scenarios were selected to provide a broad range of plausible futures. In Supplementary Table 1 we show the globally aggregated B:C ratios for all different combinations of RCP and SSP (20 combinations in total). The results in this table are under the same assumptions as the results shown in Table 1 in the main text.

We see that for the 'optimise' objective, there is a large difference between the RCP/SSP combinations in both the benefits and the costs. The costs at the global scale range from USD 22 billion per year (RCP2.6/SSP3) and 78 billion USD per year (RCP8.5/SSP5). Averaged across all SSPs, the results show that the costs at the global scale increase as the CO_2-equivalent concentration increases (RCP2.6 = USD 40 billion per year; RCP4.5 = USD 45 billion per year; RCP6.0 = USD 47 billion per year; RCP8.5 = USD 55 billion per year). This means that globally, higher greenhouse gas concentrations will lead to higher adaptation costs as a result of a generally larger increases in flood hazard (note that there are also areas where higher greenhouse gas emissions lead to reduction in flood risk). For all combinations of RCP/SSP, on average, the benefits far outweigh the costs, leading to B:C ratios ranging from 3.6 (RCP2.6/SSP3 and RCP4.5/SSP3) to 10.2 (RCP8.5/SSP5). The B:C ratios are also larger for the RCPs with higher CO_2 concentrations, since the larger increase in hazard under those RCPs means that flood damage is higher, and therefore the potential avoided damage is also higher. Averaged across all RCPs, the costs at the global scale follow the projected increases in GDP to 2080 between the different SSPs (i.e. highest costs in SSP5 and lowest costs in SSP3, which are the SSPs with the highest and lowest global GDP growth respectively).

Under the 'constant absolute risk' objective, B:C ratios exceed 1 for all RCPs combined with SSPs 1, 2, 4, and 5. However, for SSP3, which represents a more fragmented world, the B:C ratios are less than 1, because under this scenario, economic prospects are poor and thus the benefits of adaptation low. Under the 'constant absolute risk' objective', B:C ratios exceed 1 for all combinations of RCPs and SSPs.

2.2. Sensitivity to different GCMs

It has been shown in several studies that future hydrological simulations are sensitive to the choice of GCM used to force the hydrological model[5,6]. Therefore, we have chosen to display results averaged across the five bias-corrected GCM simulations[7] used in the ISIMIP project. All results for each individual GCM are provided in the Supplementary Dataset. In Supplementary Figure 5, we show the projected return periods of protection required to achieve the 'optimise' objective in 2080 for each GCM individually, for the 4 combinations of RCPs and SSPs studied throughout the letter. The results in Supplementary Figure 5 are under the same assumptions as the results shown in Table 1 in the main text. For areas shown in white, there is no protection standard that provides a positive NPV; in other words there is no protection standard with a B:C ratio that exceeds 1. These results are further summarised in Supplementary Figure 6, which shows the number of GCMs for each RCP/SSP combination for which a protection standard can be reached whereby the NPV is positive (i.e. whereby the B:C ratio exceeds 1).

These results show that whilst there are differences between the simulations forced by the different GCMs, the overall patterns are consistent. For example, Supplementary Figure 6 shows that the results are consistent between GCMs across all different RCP/SSP combinations in terms of whether an optimal protection standard can be achieved in most parts of North America, large parts of northern Europe and Australia, much of East and Southeast Asia, the Indian Subcontinent, parts of Central Africa, and along the Nile Valley. Reference to Supplementary Figure 5 shows that for these regions, the exact return period for this 'optimise' objective varies between GCMs, even though there is still rather large

consistency in many regions. However, as stated in the main text, the aim of this framework is not to exactly define the required protection standards for project implementation, but rather to identify priority regions, help to initiate dialogue with stakeholders, and highlight potential first order savings through increased structural flood protection. Given the consistency between GCMs (and indeed combinations of RCPs and SSPs) shown in Supplementary Figure 5, in terms of those regions where dikes could reduce flood risk whilst providing a positive NPV, we believe that the results meet this aim.

2.3. Sensitivity to different discount rates

Both the benefits and costs of adaptation are influenced by the selected discount rate. Many countries have official discount rates that are used for the screening and planning of different kinds of projects. Our model framework therefore allows for the calculation of benefits and costs for any discount rate. In the main text, we show results using a discount rate of 5% per year. In this subsection, we describe the sensitivity of the results to using alternative discount rates of 3% and 8% per year.

Results at the globally aggregated scale can be found in Supplementary Table 2. The B:C ratios are higher using a discount rate of 3% per year, and lower using a discount rate of 3% per year. Using a discount rate of 3% per year, the B:C ratios exceed 1 for all four RCP/SSP combinations studied. This is even the case for RCP6.0/SSP3 under the 'constant absolute risk' objective, for which the B:C ratio is lower than 1 when the discount rate of 5% per year is used. Using a discount rate of 8%, this RCP/SSP combination results in B:C ratios less than 1 under both the 'constant absolute risk' and the 'constant relative risk' objectives. Overall, the spatial pattern is robust when comparing results for a discount rate of 5% per year (Figure 2) with discount rates of 3% per year (Supplementary Figure 7) and 8% per year (Supplementary Figure 8).

2.4. Sensitivity to different cost estimates

Compared to the sensitivity of the optimal return periods and B:C ratios to the different GCMs, combinations of RCPs and SSPs, and different discount rates, the sensitivity of the results to the use of different cost estimates is relatively high. This uncertainty is important to consider as protection costs are difficult to estimate, since they depend on factors including future material and labour costs.

The optimal protection standards for the middle-cost estimates are shown in Figure 2, and those for the low- and high-cost estimates are shown in Supplementary Figures 9 and 10 respectively. For the low-cost estimate (Supplementary Figure 9), positive NPVs are achieved for the 'optimise' objective in most regions, including many parts of South America and Africa where this is not the case using middle-cost estimates. For the high-cost estimate (Supplementary Fig. 10), the general spatial pattern remains, albeit with lower protection standards and fewer sub-national units where positive NPVs can be achieved. This shows that it is important to ascertain good estimates of dike construction costs for particular regions in which a user is interested. The model framework has therefore been designed in a flexible way so that construction cost estimates can easily be changed if such information is available for a particular region. Nevertheless, the results also show that there are many regions that still show a consistent pattern in terms of whether the benefits of adaptation through dikes would exceed the costs, even when taking into account this uncertainty between the low- and high-cost estimates.

2.5. Sensitivity to assumed baseline protection standards

In order to estimate current risk and to estimate the dike height required to provide the different protection standards assessed in the future scenarios, it is necessary to assume a current protection standard that is in place for each sub-national unit. For this letter, we use current protection standards from the FLOPROS database[4]. FLOPROS provides modelled protection standards at the sub-national scale. It is the only global database of flood protection standards, and is the first attempt to systematically map these standards worldwide. The database has been validated against actual protection standards in place in several regions in Ref. 4.However, the number of regions for which such data are available is scarce, and therefore the FLOPROS protection standards are subject to uncertainty.

Whilst our model framework allows the current protection standard to be adjusted if better information is available at the local scale, we here present an assessment of the sensitivity of the results per sub-national unit to the current protection standards used in this study. We do this by re-running the analyses assuming: (a) current flood protection to be half that stated in the FLOPROS database (FLOPROS-halved); and (b) current flood protection to be double that stated in the FLOPROS database (FLOPROS-doubled).

The results at the globally aggregated scale are shown using protection standards stated in FLOPROS in Table 1, FLOPROS-halved in Supplementary Table 3, and FLOPROS-doubled in Supplementary Table 4. Both the costs and the benefits are higher when using FLOPROS-halved, and lower when using FLOPROS-doubled. However, for the 'optimise' objective, the resulting B:C ratios are very similar between the different assumptions on current protection. For the 'constant absolute risk' and 'constant relative risk' objectives, the changes in B:C ratios between the different assumptions on current protection standards is larger, but still small in all cases. For FLOPROS, FLOPROS-halved, and FLOPROS-doubled, the globally aggregated B:C ratios exceed 1 in all cases, except for RCP6.0/SSP3 for the 'constant absolute risk objective'. The latter has a B:C ratio lower than 1 for FLOPROS, FLOPROS-halved, and FLOPROS-doubled.

We also show comparisons of the protection standards for the 'optimise' objective for current protection standards assuming FLOPROS (Figure 2), FLOPROS-halved (Supplementary Figure 17) and FLOPROS-doubled (Supplementary Figure 17). The accompanying B:C ratios for these protection standards are shown assuming FLOPROS (Supplementary Figure 2), FLOPROS-halved (Supplementary Figure 18) and FLOPROS-doubled (Supplementary Figure 20). These figures show that the results are very robust to the use of the different assumptions on current protection, in terms of their influence on the B:C ratio and the order of magnitude of the optimal flood protection standard. Of course, the individual benefits, costs, and therefore NPV will change (as was shown for the globally aggregated results above), but in terms of the overall pattern of areas where adaptation through dikes may be feasible, the results are robust.

2.6. Overall robustness

In Supplementary Information 2.1 to 2.5 we have assessed the sensitivity of our model results to several assumptions and parameters. Next to the assessment of robustness shown in Figure 4 of the main text, we here assess the overall robustness of the results to all assumptions and parameters used in this letter

(Supplementary Figure 21). To do this, we took the B:C ratios per sub-national unit for each combination of the model parameters used in this study (2700 combinations in total), namely: (a) 5 GCMs; (b) 5 SSPs; (c) 4 RCPs; (d) 3 cost estimates (high, middle, and low); (e) 3 discount rates (3, 5, and 8% per year); and (f) 3 estimates of current protection standards (FLOPROS, FLOPROS-halved, and FLOPROS-doubled). We then counted the number of simulations per sub-national unit for which the B:C ratio exceeds 1, and calculated this as a percentage of the total number of simulations (i.e. 2700).

The results show that despite the large uncertainties discussed above, there are many regions where the modelling framework provides robust results in terms of whether a protection standard can be achieved by 2080 for which the benefits exceed the costs. The results show both areas where there is high confidence that this could be achieved, and areas where there is high confidence that this could not be achieved. For the former regions, this information can help to identify regions in which this adaptation strategy may be most effective and/or feasible, and can provide stakeholders with first order estimates of the savings that could be made through increased structural flood protection. Such information can help to initiate and stimulate dialogue, although when moving to implementation, detailed studies should be performed using local models and data. For the latter regions, the results suggest that more attention is required for other flood risk management strategies.

3. Inundation model benchmarking

In order to evaluate the extent to which the GLOFRIS inundation maps are able to represent flood hazard, several model benchmarking experiments have been carried out. In this section, we first briefly summarise the results of benchmarking experiments carried out using the GLOFRIS inundation maps in past studies (Supplementary Information 3.1). Given the limited geographical scope of these past benchmarking exercises, and the fact that they did not specifically assess the ability of GLOFRIS to simulate inundation extent in urban areas, we have carried out extensive further benchmarking for this letter. The approach, results, and discussion of the new benchmarking experiments are described in Supplementary Information 3.2.

3.1. Past benchmarking experiments

In past studies, the GLOFRIS inundation maps have been benchmarked for several regions. A first visual benchmarking was carried out by comparing them to maps of the Dartmouth Flood Observatory (DFO, http://floodobservatory.colorado.edu/) for South Asia and South East Asia[8]. Maps showing maximum inundation extents per year (over a 30 year period) for South Asia and South East Asia were simulated using the GLOFRIS inundation downscaling routine. Then, the maximum inundation extents over the 30 year period were extracted, and compared visually to the coarse flood extent maps of DFO, based on a documentation of past floods. This validation exercise was limited, since the GLOFRIS 30 year flood extent is not fully equivalent with the DFO observed flood extent. Therefore, we did not use quantitative performance metrics, but rather visually compared patterns of flooding. As a first-cut validation, it showed that the simulated 30 year flood extent and the DFO flood extents in urban areas around the large rivers (e.g. Ganges, Brahmaputra, Chao Phraya, Irrawaddy, and Indus) showed broadly similar areas of inundation, although floods due to backwater effects are not represented since our model is not hydrodynamic. In the same paper[8], the GLOFRIS inundation module was also used to simulate the major 2004 flood event in Bangladesh. The simulated daily flood extents (over a 1-month period) were then compared visually to satellite imagery over the same period, again supplied by DFO. The results are described in detail in Ref. 8. Underestimations of extents do occur in regions where other flood processes besides river flooding play a dominant role, and in some regions due to the limited horizontal and vertical resolution and accuracy of the digital elevation model. The use of the inundation maps for Bangladesh for benchmarking is made difficult by the fact that there are regions in which inundated areas are clearly missing due to the presence of clouds.

In Ref. 3, the most recent GLOFRIS inundation maps (the ones used in this study) were compared to benchmarked inundation maps from more localised models, using several standard performance metrics (hit rate, false alarm ratio, critical success index; see Supplementary Information 3.2.1). The benchmarking exercise was carried out for three case studies for a 100 year flood, namely: the Thames River (UK), the Severn River (UK), and the Neue Luppe and Mulde Rivers (Saxony, Germany). Results for these cases are described and further elaborated in Supplementary Information 3.2.3.

3.2. New benchmarking experiments

The qualitative benchmarking experiments carried out to date are limited in several ways. Firstly, they only cover a limited number of case studies, all located in northwest Europe. Secondly, they do not specifically assess the ability of GLOFRIS to simulate flood extents in urban areas, which are the focus of this study. Thirdly, the past experiments have not assessed the extent to which the differences in inundation extents between the GLOFRIS and benchmark datasets lead to differences in simulated flood impacts. Therefore, for this letter we extend the benchmarking exercise described in Ref. 3.Firstly, we add more case studies using modelled or observed inundation maps from other sources; the case studies are in Europe, the USA, and Thailand. Secondly, we specifically calculate the performance metrics in urban areas, and compare model performance in these areas to model performance across the entire geographic domain of each case study. To do this, we first calculate the performance metrics for all cells within the case study area, and then we calculate the same metrics only for the cells that are classed as urban in the GLOFRIS exposure dataset (i.e. cells for which urban density exceeds zero, which are also the cells for which we calculate urban damage). These benchmarking exercises are carried out at two spatial resolutions: the spatial resolution of the GLOFRIS inundation maps, and the spatial resolution of the original benchmark datasets. Thirdly, we use both the GLOFRIS and benchmark inundation maps to assess the potential impacts of each case study in terms of the maximum potential damage. This gives an indication of the size of the uncertainty in simulated flood impacts that results from the use of the GLOFRIS inundation maps. We compare this uncertainty to several other model uncertainties (the use of FLOPROS, different GCMs, RCPs, and SSPs) for which we have carried out robustness analyses (see Supplementary Information 2). By doing this, we are able to examine the relative magnitude of the uncertainties in flood impact due to the use of GLOFRIS compared to other uncertainties in the modelling framework. Finally, we compare our model performance metrics with those in benchmarking studies carried out using the SSBN-Bristol model[9] and the European scale inundation model of the European Commission Joint Research Centre (JRC model)[10]. The approach used is described in Supplementary Information 3.2.1; the benchmark datasets are described in Supplementary Information 3.2.2, and the results are presented and discussed in Supplementary Information 3.2.3.

3.2.1. Approach

To carry out the benchmarking, we extend the approach described in Ref. 3. This involves comparing the inundation extents simulated using GLOFRIS, with inundation maps from other sources, i.e. the benchmark datasets. In Ref. 3, this was carried out by comparing our maps for return periods of 100 years to simulations of 100 year floods derived from local models. The areas studied were the Thames and Severn Rivers in the UK, and the Neue Luppe and Mulde Rivers in Germany. In the current letter, we also use these same case studies, and add several case studies in the USA for which benchmark inundation maps for given return periods are available. Furthermore, we also compare the GLOFRIS inundation maps with maps of inundation extent from specific flood events, based on satellite imagery.

All of the benchmark inundation maps have a higher spatial resolution than the GLOFRIS inundation maps. Each cell shows whether a cell is flooded (1) or not (0). In order to be able to compare the two

datasets, they must both be in the same resolution. This can be achieved by either resampling the benchmark inundation maps to the resolution of the GLOFRIS inundation maps (benchmarking at GLOFRIS resolution), or by resampling the GLOFRIS inundation maps to the resolution of the benchmarking inundation maps (benchmark at original resolution). For this study, we carried out both of these approaches.

For the benchmarking at GLOFRIS resolution, we first resampled each of the benchmark inundation maps to a horizontal resolution of 30" x 30". During resampling, cells that are inundated in more than 50% of the corresponding cells in the higher resolution benchmarking dataset, are classed as inundated in the resampled (30" x 30") dataset. To ensure that comparability is valid, we first removed river sections with a Strahler order smaller than 6 from the benchmark inundation maps, as these are neither taken into account in GLOFRIS, nor in the cost and benefit assessment framework. The GLOFRIS inundation maps were also reclassed to show cells that are flooded (1) or not (0).

For the benchmarking at original resolution, we resampled the GLOFRIS inundation maps to the higher spatial resolution of the benchmark inundation maps. To do this, we used a nearest neighbour resampling method. This means that if a GLOFRIS cell is shown as (non-) flooded, it will appear as (non-) flooded in all of the higher resolution cells in the resampled dataset. Then, the benchmark and GLOFRIS inundation maps are compared with each other. First, we visually plotted the area that is inundated in both datasets, the GLOFRIS dataset only, and the benchmark dataset only. We then calculated the following performance metrics: hit rate, false alarm ratio, and critical success index. Each of these metrics is described below. We first calculated these performance metrics for all cells within the case study area, and then we calculated the same metrics only for the cells that are classed as urban in the GLOFRIS exposure dataset.

Hit rate: The hit rate indicates how well a model dataset replicates a benchmark dataset, without penalising for over-prediction. Essentially, it indicates the fraction of inundated cells in the benchmark dataset that are also inundated in the GLOFRIS dataset. It is calculated as:

$$HR = \frac{A_G \cap A_B}{A_B} \quad \text{(Eq. 4)}$$

where, *HR* is the hit rate, A_G is the number of inundated cells in the GLOFRIS dataset, and A_B is the number of inundated cells in the benchmark dataset. *HR* ranges from 0 to 1, whereby a hit rate of 1 indicates that all of the inundated cells in the benchmark dataset are inundated in the GLOFRIS dataset, and a hit rate of 0 indicates that none of the inundated cells in the benchmark dataset are inundated in the GLOFRIS dataset.

False alarm ratio: The false alarm ratio is a measure of model over-prediction. Essentially, it indicates the proportion of cells that are inundated in the GLOFRIS dataset but are not inundated in the benchmark dataset. It is calculated as:

$$FAR = \frac{A_G/A_B}{(A_G \cap A_B + A_G/A_B)} \quad \text{(Eq. 5)}$$

where *FAR* is the false alarm ratio. *FAR* ranges from 0 to 1, whereby a value of 0 indicates that none of the inundated cells in the GLOFRIS dataset are false alarms, and a value of 1 indicates that all of the inundated cells in the GLOFRIS dataset are false alarms.

Critical Success Index: The critical success index extends on the hit rate and the false alarm ratio to create a combined score that penalises for both under-prediction and over-prediction. It is calculated as:

$$CSI = \frac{A_G \cap A_B}{A_G \cup A_B} \quad \text{(Eq. 6)}$$

where *CSI* is the critical success index. *CSI* ranges from 0 to 1, whereby 0 indicates no match between the GLOFRIS and benchmark datasets and 1 indicates no match between the GLOFRIS and benchmark datasets.

Since the ultimate aim of GLOFRIS is to assess flood impacts, and not only flood hazard, we carried out an additional experiment to assess the potential impacts of each case study in terms of the maximum potential damage. To do this, we carried out the impact assessments at the spatial resolution of the original benchmark datasets. First, we resampled and reprojected the lower resolution GLOFRIS inundation extent maps to the same resolution and projection as the benchmark inundation maps. Then, we used the GLOFRIS impacts module to assess the maximum potential damage for each case study (using both the GLOFRIS and benchmark dataset). Since we only have data on inundation extent for most of the benchmark datasets, rather than inundation depths, we calculated the maximum potential damage, instead of the actual damage. Finally, we assessed the percentage difference in maximum potential damage per case study, when using the GLOFRIS and benchmark inundation maps.

3.2.2. Benchmark inundation maps

As mentioned above, for this letter we carried out benchmarking against both inundation maps for a given return period, and inundation maps for specific flood events. Each benchmark inundation map is described in the following paragraphs.

Benchmarking against return period inundation maps

Thames and Severn Rivers (UK)

Benchmark inundation maps of the Thames and Severn Rivers in the UK were provided by the Environment Agency of England and Wales. The GLOFRIS inundation maps were already benchmarked against these datasets in Ref. 3. Here, we carried out the same analysis, but also calculating the performance metrics in urban areas only. Both hazard maps in the UK represent inundation conditions assuming no flood protection, which is representative of the GLOFRIS inundation schematisation. The datasets represent catchment-wide flood extents for floods with a return period of 100 years. The catchment sizes are ca. 16,000 km^2 and 11,000 km^2 for the Thames and Severn respectively, and both rivers flow through a large variety of different kinds of urban areas, ranging from villages and small to medium towns and cities, right up to the major conurbation of London. As such, these cases allow us to benchmark model performance for medium sized rivers through a range of urban fabrics. The

inundation maps are comprised of several data sources; in descending priority, these are: (a) an observed 1 in 100 year flood outline; (b) the output from detailed 1D hydraulic models built with airborne laser terrain data with a resolution of ~2 m and a vertical accuracy of <10 cm; and (c) a 2D hydraulic model built using airborne interferometric Synthetic Aperture Radar terrain data with 5 m spatial resolution and 0.5-1 m vertical accuracy. The data were delivered as a polygon shapefile and resampled to a 30" × 30" grid. These benchmark datasets were compared with the 100 year return period GLOFRIS inundation map.

Neue Luppe and Mulde River (Saxony, Germany)

The benchmark inundation maps of the Neue Luppe and Mulde Rivers in Saxony, Germany, were provided by the Saxony State Office for Environment, Agriculture and Geology. The GLOFRIS inundation maps were already benchmarked against these datasets in Ref. 3. Here, we carried out the same analysis, but also calculating the performance metrics in urban areas only. The inundation maps for Saxony are different from the Thames and Severn maps, as the former do include flood protection measures. They are based on local, high resolution datasets, terrain models, and inundation models. They represent flood extents for floods with a return period of 100 years. The reaches used in this study flow through many villages and small towns, plus the medium sized city of Leipzig. Again, this allows us to benchmark model performance for medium sized river reaches through several urban fabrics. These benchmark datasets were compared with the 100 year return period GLOFRIS inundation map.

Flint, Mississippi, and Susquehanna Rivers (USA)

Benchmark inundation maps for the Flint River (at Albany, Georgia), Mississippi River (at Saint Paul, Minnesota), and Susquehanna River (at Harrisburg, Pennsylvania) were provided by the United States Geological Survey (USGS). These were selected from a set of 10 sites in the eastern USA. Of those 10 sites, only the case studies were selected where the river reach has a Strahler order of 6 or above, and where the reach runs through an urban area. These case studies represent very small river reaches of ca. 7.7 km (Flint), 10.1 km (Mississippi), and 40 km (Susquehanna). As such, they pose an extremely strong challenge for global flood models. For these cases, we used benchmark and GLOFRIS inundation maps for floods with a return period of 100 years.

The simulations for the Flint River were carried out using recent digital elevation model data and the US Geological Survey finite-element surface-water modelling system for two-dimensional flow in the horizontal plane (FESWMS-2DH)[11]. The model employs a mesh with resolutions between 6-12m. Inundation depths for this mesh were first rasterised onto a grid of 5m x 5m, before being resampled to 30" x 30" as described previously. Along this stretch, the Flint River flows through the city of Albany (Georgia), which has a population of ca. 77,000. In the Albany area, the natural state of the Flint River has been changed by the construction of dams, a human-made peninsula, and numerous bridges, which are incorporated in the model.

The simulations for the Mississippi River were carried out using the HEC-RAS model[12], a 1D step-backwater hydraulic model. For these inundation maps, the hydraulic analyses were carried out using steady-state flow computations. The model employs a grid of ca. 1m. Along this stretch, the Mississippi

River flows through the city of St. Paul (Minnesota), which has a population of ca. 300,000. In the St. Paul area, four major levees have been constructed within the inundation mapping reach used for the benchmarking study, which are incorporated in the model.

The simulations for the Susquehanna River were carried out using the HEC-RAS model[13], again using steady-state flow computations. The model employs a mesh with resolutions between 1-2m. Inundation depths for this mesh were first rasterised onto a grid of 5m x 5m, before being resampled to 30" x 30" as described previously. Along this stretch, the Susquehanna River flows through the city of Harrisburg (Pennsylvania), which has a population of ca. 50,000. In the Harrisburg area, various human-made drainage structures have been constructed. To account for these, nine bridges, an in-channel dam, and a levee system are incorporated in the model.

Benchmarking against event-based inundation maps

St. Louis flood (Mississippi River, USA)

The benchmark inundation map for the St. Louis flood of 1993 is derived from satellite imagery. This flood on the upper reaches of the Mississippi River had huge impacts across the US Midwest, and led to extensive flooding in the city of St. Louis, Missouri, which has a population of ca 315,000, and is the centre of the Greater St. Louis area (population of almost 3 million). We used an inundation extent map from the NASA Earth Observatory[a]. This map is provided as a false-colour image, which combines infrared, near infrared, and green wavelengths from the Thematic Mapper (TM) instrument of the Landsat 5 satellite. We then converted this false-colour image to an inundation map by unsupervised classification in ArcGIS. We reclassified dark blue and pink cells to water. The dark blue cells in the image represent cells where water was present at time that the image was captured, and the pink cells represent cells where flood waters have retreated to reveal newly exposed soil. The flood is estimated to have had a return period slightly higher than 100 years[14]. We therefore compared it to our 100 year return period inundation map from GLOFRIS. Before carrying out the analysis, we first removed the smaller local tributaries, leaving the main branches of the Mississippi, Missouri, and Illinois Rivers, since the former were not in flood stage at the time of the event. The reach of the Mississippi used in this case study has a length of ca. 170 km, and all of the rivers run through the urban area of St. Louis.

Chao Phraya River flood (Thailand)

The benchmark inundation map for the Chao Phraya River flood of 2011-2012 is derived from satellite imagery. This massive flood caused huge damage over large parts of Thailand, including 813 fatalities and tens of billions of dollars of damage[15]. It affected many major cities and towns, including Bangkok, Ayutthaya, and Nakhon Sawan. We used an inundation map derived from satellite imagery, taken from Ref. 16.This inundation map shows cells detected as inundated over the period 14 August 2011 to 3 March 2012, and was developed in Ref. 16 using inundation maps downloaded from the Dartmouth Flood Observatory (DFO) website[17]. These inundation maps are derived from the MODIS (Moderate Resolution Imaging Spectroradiometer) near real-time global flood mapping project of NASA Goddard's

[a] https://earthobservatory.nasa.gov/IOTD/view.php?id=5422

Office of Applied Science, at a resolution of ca. 250m x 250m. The data from Ref. 16 are confined to the lower plains of the Chao Phraya River, below Nakhon Sawan, as defined by the 30 m elevation contour, and a set back of 25 km from the coast is used. The set back is used to exclude the low lying areas on the coast which were mainly affected by flooding coasts by a combination of high tide and high discharge. The flood has been estimated to have had a return period between 10-20 years by Ref. 18, although this same study states that this estimate may be low-biased. We therefore compared it to our 25 year return period inundation map from GLOFRIS.

3.2.3. Benchmarking results and discussion

In this section, we begin with an overview of the key findings, both in terms of the hazard benchmarking and the impact benchmarking (Supplementary Information 3.2.3.1). This is followed by more detail on the individual case study results (Supplementary Information 3.2.3.2).

3.2.3.1. Overview of benchmarking results

To the best of our knowledge, the collection of case studies presented here represents the most comprehensive attempt to date to benchmark the results of a global flood model against local datasets. The inundation benchmarking performance metrics for all case studies are summarised in Supplementary Table 5.

The differences between the benchmarking at GLOFRIS resolution and at original resolution are generally small. Benchmarking at the resolution of the original dataset is extremely stringent, and comes with its own problems, because the version of GLOFRIS used in this study was specifically designed to run at 30” x 30”, as this is the resolution at which the impacts module operates (due to the availability of exposure data at this resolution). Therefore, it is to be expected that GLOFRIS will over-or underestimate in some locations compared to the higher resolution benchmark dataset. If the exposure data used in the impact module had a higher resolution, we would also have run the GLOFRIS inundation module at higher resolution. It is therefore encouraging that model performance is similar at higher resolutions for which GLOFRIS was not originally designed. Since GLOFRIS is applied at a resolution of 30” x 30”, we believe that the benchmarking exercise at this resolution is the most appropriate, but results at both resolutions are included for transparency.

In general, the model captures between ca. two-thirds and almost nine-tenths of the inundated area in the benchmark dataset. The simple averaged HR across the eight case studies is 0.70 for the benchmarking at GLOFRIS resolution. For the benchmarking at original resolution, the simple averaged HR is slightly lower, at 0.66. The HR for the Severn case study is notably lower for the latter (0.51) than the former (0.63). At the same time, in general the model does not produce excessively large false alarms. The simple average across the case studies is 0.35 for the benchmarking at both resolutions. However, the range is large between the different case studies, namely 0.13 to 0.52 at GLOFRIS resolution and 0.18 to 0.53 at original resolution. For the Saxony case, over half of the inundated cells are false alarms. Therefore, it is also vital to assess how these differences in inundated areas between the GLOFRIS and benchmark inundation maps translate into differences in flood impacts, as discussed later in this section.

The performance metrics are similar when calculated for urban areas only, compared to the metrics for the entire geographical domain (Supplementary Table 5). This lends support to the application of GLOFRIS for simulating flood damage in urban areas, which is key for this letter. In urban areas only, across the eight case studies there is no difference in the simple average HR when calculated across the entire geographical domain compared to urban cells only; this is the case for both benchmarking at GLOFRIS and original resolution. The simple average FAR is slightly higher for urban cells only compared to the entire geographical domain, although the differences are minimal: 0.37 compared to 0.35 for benchmarking at GLOFRIS resolution, and 0.38 compared to 0.36 for benchmarking at original resolution. For the Thames and Severn, the difference in CSI is 0.01 or less. For the Neue Luppe and Mulde Rivers (Saxony), the CSI is ca. 11% better in urban areas only at GLOFRIS resolution, and 5% better at original resolution , whilst for the St. Louis it is ca. 7% better across the entire geographical domain (for both benchmarking resolutions). The inundation maps for the small reaches in the USA only include flooding in urban areas. The only case study with a substantially lower CSI in urban areas only compared to the entire domain is the Chao Phraya river, by ca. 17% at GLOFRIS resolution and 19% at original resolution. This difference can be partially explained by the presence of flood protection measures in the Thonburi area of Bangkok, as discussed in the specific case study description in Supplementary Information 3.2.3.2.

For the Thames and Severn case studies, benchmarking exercises have also been carried out in previous papers using the SSBN-Bristol model[9] and the JRC model[10]. The JRC model has also been benchmarked for the Saxony case study. The SSBN-Bristol and JRC models use more physically based approaches to simulate inundation and operate at a higher resolution; the former solves 2D hydrodynamic equations, whilst the latter uses a 2D hydraulic modelling scheme. Another advantage of the SSBN-Bristol model is that it uses global terrain data that have been corrected for vegetation bias and bias in urban areas. Therefore, it is useful to examine how the performance of our simpler model scheme compares against these more complex schemes. Whilst these benchmarking exercises used the same model performance metrics, the exact methodological approaches were different, and therefore a direct comparison is difficult. However, broadly speaking the performance metrics presented in Ref. 9 using the SSBN-Bristol model are somewhat better than those presented for GLOFRIS in this letter. As this model solves 2D hydrodynamic equations and operates at a higher resolution, we would expect it to perform better in hydrodynamically complex areas. Our performance metrics are generally of the same order of those described for the JRC Europe model.

Ultimately, the important question is whether the selected inundation model is fit for the purpose for which it is applied. When carrying out a scenario modelling exercise such as the one carried out for this letter, an important consideration is whether the model provides reasonable performance but also produces inundation maps within a reasonable time-frame and computational cost. An important consideration herein is whether the model can simulate inundation with high enough skill so that the flood impact results do not deviate excessively from impact results based on benchmark datasets. To test this, we calculated the maximum potential damage per case study using the GLOFRIS and benchmark inundation maps, whereby the GLOFRIS maps were first resampled to the resolution of the benchmark dataset. The percentage difference in this maximum potential damage using the GLOFRIS

and benchmark inundation maps is shown in Supplementary Table 6. For most cases, the differences range between an increase of about a third to a decrease of about a third. The exception is Saxony, for which the maximum potential damage is 71% higher when using the GLOFRIS inundation map compared to the benchmark dataset. To understand what these differences mean for the overall modelling framework, we compare the magnitude of these differences with the magnitude of the uncertainty introduced by several other model parameters discussed in Supplementary Information 2.

First, for the sub-national unit in which each of the case studies belongs, we assessed the difference in baseline EAD when using different assumptions on the current flood protection standard. To do this, we compared EAD using the FLOPROS-halved and FLOPROS-doubled assumptions, compared to EAD using FLOPROS (see Supplementary Information 2.5). Whilst the maximum potential damage is not the same metric as the EAD (which is also based on the use of depth-damage functions and integrates across different return period floods), it is an indicator of the percentage change in impact resulting from the differences between the GLOFRIS and benchmark datasets, and is therefore a relevant metric for comparison. Reference to Supplementary Table 6 shows that the percentage difference in maximum potential damage due to the use of the different inundation maps is much smaller than the difference in EAD introduced by using the FLOPROS-halved and FLOPROS-doubled assumptions. Even for the Saxony case study, the difference falls within these bounds. In Supplementary Information 2.5, we show that whilst the individual benefits, costs, and NPV change substantially when using the different assumptions on current flood protection, the overall patterns are robust in terms of areas where adaptation through dikes may be feasible. Since the differences in impacts between the simulations carried out using GLOFRIS and the benchmark inundation datasets are much smaller than those caused by the different assumptions on current protection standards, this shows the overall conclusions to be robust considering potential errors in the GLOFRIS inundation maps.

Second, we assessed the change in EAD between baseline conditions and 2080 when using different GCMs, RCPs, and SSPs. For each GCM, we extracted the EAD averaged across the four RCPs in 2080, using current socioeconomic conditions. We then calculated the change in EAD for each GCM compared to baseline EAD, and the standard deviation of this change across the 5 GCMs; this result is reported per case study in Supplementary Table 6. The results show that the percentage difference in maximum potential damage due to the use of the different inundation maps is much smaller than the standard deviation of the change from baseline to future across the different GCMs. We then performed a similar analysis for each RCP, in which we extracted the EAD for each RCP averaged across the five GCMs in 2080, again using current socioeconomic conditions. For most case studies, the percentage difference in maximum potential damage due to the use of the different inundation maps is much smaller than the standard deviation of the change from baseline to future across the different RCPs. Finally, we performed a similar analysis for each SSP, in which we extracted the EAD for each SSP, assuming current climate conditions. For all case studies, the percentage difference in maximum potential damage due to the use of the different inundation maps is much smaller than the standard deviation of the change from baseline to future across the different SSPs; in many cases this difference is of an order of magnitude or more. In Supplementary Information 2, we assessed the sensitivity of our results to the different GCMs, RCPs, and SSPs, and found that the overall patterns are robust in terms of areas where

adaptation through dikes may be feasible. Since the differences in baseline impacts between the simulations carried out using GLOFRIS and the benchmark inundation datasets are much smaller than those caused by the different assumptions on GCMs, RCPs, and SSPs, this again shows the overall conclusions to be robust considering potential errors in the GLOFRIS inundation maps.

There are some regions in which GLOFRIS clearly underestimates inundation area compared to the benchmark dataset. One of these areas is around the river-coast interface (deltas, estuaries), as seen in the Thames and Severn cases (see specific case study descriptions in Supplementary Information 3.2.3.2 for detail). This is because GLOFRIS currently does not include coastal flooding or coastal boundary conditions in these regions. Work is ongoing to couple the GLOFRIS model with: (a) the Catchment-based Macro-scale Floodplain model (CaMa-Flood)[19] to simulate water levels globally including a 1D flow equation (a local inertial equation); and subsequently (b) the Global Tide and Surge Reanalysis (GTSR) dataset[20], which is the world's first dataset of global coastal water levels based on hydrodynamic modelling. This will allow us to include boundary conditions along the coast and backwater effects in order to improve the representation of hazard in these regions. Another example is in heavily vegetated areas (like parts of the Mulde River in the Saxony case study), where the bare earth elevation is much lower than in our DEM. This issue could be addressed by improving the DEM underlying the inundation routine. For example, the SSBN-Bristol model uses a DEM that has been corrected for vegetation bias and systematic bias in urban areas9.

Examples of regions in which GLOFRIS overestimates inundation extent include several areas protected by levees (see for example the specific case study descriptions for the Chao Phraya river and Mississippi case studies Supplementary Information 3.2.3.2). Whilst these are not included directly in our inundation model, we do account for the presence or absence of flood protection measures in our calculation of risk, which is the main premise of this paper. Nevertheless, it would be very useful for the modelling community to develop databases of individual flood protection measures (dikes, levees, etc) around the world. These could then potentially be parameterised in the kinds of global models described in Ref. 9, such as the SSBN-Bristol model or CaMa-Flood. GLOFRIS also shows larger inundation extents than the benchmark datasets in narrow floodplains, such as the upper catchments of the Severn River and Saxony case studies. This could be improved in future model simulations by increasing the resolution of the inundation model, and also incorporating a hydraulic simulation engine. However, at present the best available exposure datasets with globally consistent coverage are also at a resolution of 30" x 30" or lower, and therefore this is the scale at which our risk assessments are carried out. A future improvement could be to develop globally applicable models, whereby global models serve as input to more detailed local hydrodynamics models. Inundation maps from these models could then be combined with higher resolution and quality data on exposure where available, such as object-based data available in OpenStreetMap.

It is, of course, always possible to further develop models to make them more physically realistic, and there will also always be other possible models that, at particular times and in particular places, do a better job. The key point is whether the model being used is fit for the purpose in hand, and whether the conclusions are robust to potential errors resulting from the use of that model. We have demonstrated that the differences in simulated impacts resulting from the use of the GLOFRIS

inundation maps (compared to benchmark inundation maps) are much smaller than the differences in impact results due to the use of different assumptions on flood protection standards, and the use of different GCMs, RCPs, and SSPs. Therefore, we believe that the benchmarking results demonstrate that the GLOFRIS model used here has sufficient skill for the purpose to which it is here put. Further improvements for other applications, requiring more accurate local flood inundation estimates may be found by: introducing better terrain data in which effects of vegetation and urban built-up areas are improved; the use of a 1D hydraulic scheme that solves the diffusion wave scheme so that backwater effects can be included; and improvements in global channel dimension estimates.

3.2.3.2. Benchmarking results for individual case studies

When interpreting the benchmark results, it is important to remember that this is not the same as a true model validation exercise. For the case studies in which we compare the GLOFRIS model output to the output of more local benchmark models, it should be noted that those local benchmark models cannot be considered as the truth. In reality, both the GLOFRIS and the benchmark datasets contain errors. Moreover, for the case studies where we benchmark GLOFRIS results for return period floods against event-based benchmark datasets, there is an additional challenge. The inundation maps from GLOFRIS show a flood extent for the return period across an entire geographical domain. For an event-based flood map, the actual return period will never be the same across the entire reach.

Thames River (UK)

Maps of the GLOFRIS and benchmark inundation extents can be found in Supplementary Figure 22 and 23 (at GLOFRIS and benchmark resolution respectively). The HR for the entire catchment is 0.65 (for both the benchmarking at GLOFRIS and original resolution). Visual inspection shows that a large proportion of the inundated cells that are missed by GLOFRIS are located downstream in the Thames Estuary. The benchmark inundation dataset includes flood hazard from both the river and the sea, whilst GLOFRIS only includes river flooding, and therefore it is to be expected that the benchmark dataset will show more inundation in this area. The FAR is not affected by this issue, and is 0.29 at GLOFRIS resolution and 0.44 at original resolution. The higher FAR for the benchmarking at original resolution is caused by cells in the upstream catchment whose floodplain is significantly lower than 30” in width. When the performance metrics are calculated for urban areas only, they only change marginally, and are the same when rounded to 2 significant digits (in Supplementary Table 5); for this case the model performs equally well in urban areas compared to the entire geographical domain. Simulated maximum potential damage using the GLOFRIS inundation map is 9% higher than using the benchmark inundation map. This is much smaller than the difference introduced through the use of the alternative assumptions on current flood protection standards, and the standard deviation of change in future EAD due to the use of different GCMs, RCPs, and SSPs.

Severn River (UK)

For the Severn River, calculating the performance metrics for the entire catchment results in a similar HR (0.63), but a higher FAR (0.46), when carried out at GLOFRIS resolution. This can be explained by the resolution of 30” x 30” used for the GLOFRIS benchmarking, as discussed in Ref. 3. Basically, the Severn

River floodplains in the upper catchments are generally much narrower than 30", and therefore when the benchmark inundation maps for these areas are resampled to 30" x 30", no inundation is given in the resulting maps. When carried out at original resolution, the HR is lower (0.51), but the FAR is similar (0.45). For the FAR, the reason is similar; the use of nearest neighbour resampling to the resolution of the benchmark dataset means that the inundated area is wider than the floodplains in the lower resolution model. Visual inspection of Supplementary Figure 24 emphasises this point, as it can be seen that the overestimation of GLOFRIS tends to be in these upper catchments. In Ref. 3, the GLOFRIS inundation maps were therefore further downscaled to 3' × 3', by performing a further volume spreading across a Height-Above-Nearest-Drain (HAND) map using Google Earth Engine. This led to a much lower FAR of 0.18. The main area of underestimation in GLOFRIS is in the area around the Bristol Channel. As with the Thames case, this can be explained by the fact that the benchmark datasets include coastal flooding. When calculating the performance metrics for the urban areas only, there is hardly any change; for this case the model performs equally well in urban areas compared to the entire geographical domain. Simulated maximum potential damage using the GLOFRIS inundation map is 12% lower using the benchmark inundation map. This is much smaller than the difference introduced through the use of the alternative assumptions on current flood protection standards, and the standard deviation of change in future EAD due to the use of different GCMs, RCPs, and SSPs.

Neue Luppe and Mulde River (Saxony, Germany)

The performance metrics for this case study for the benchmarking at GLOFRIS resolution are similar to those for the Severn River. The HR is 0.65, but the FAR is high, at 0.52. Again, this can be explained by the narrow floodplains, especially in the upper catchments of the Mulde River (Supplementary Figure 26 and 27). When the GLOFRIS inundation maps were further downscaled to 3' x 3' in Ref. 3, the FAR decreased to 0.28. The HR is higher (0.73) when carried out at the resolution of the original benchmark datasets, with a similar FAR (0.53). Visual inspection of Supplementary Figure 26 and 27 shows that there are some wider floodplain areas in the more downstream reaches of the Mulde that are underestimated by GLOFRIS. For this case study the performance metrics are better for the urban areas only, compared to those for the entire catchment. As a result of the large number of false alarms, this case study has by far the largest percentage difference in simulated maximum potential damage using the GLOFRIS inundation map compared to using the benchmark inundation map, namely +71%. The is of the same order of magnitude as the influence on EAD of using flood protection standards equal to half the standards listed in FLOPROS (compared to using the standards listed in FLOPROS). It is also of the same order of magnitude as the standard deviation of change in future EAD due to the use of different RCPs, although lower than the influence of using the different GCMs and SSPs.

Flint, Mississippi, and Susquehanna Rivers (USA)

These three case studies represent short river reaches, and therefore provide a very strong challenge for global flood models. The maps can be found in Supplementary Figures 28-33. The Flint and Mississippi case studies in particular represent very short reaches, of just 7.7 km and 10.1 km respectively. Nevertheless, the benchmarking results show reasonable performance, as the percentage differences in simulated maximum potential damage using the GLOFRIS and benchmark inundation map are much

smaller than the influence of the alternative assumptions on current flood protection standards, and the standard deviations of change in future EAD due to the use of different GCMs, RCPs, and SSPs (with the exception of the Mississippi for the change in EAD due to the use of different RCPs).

For all three of these case studies, the results are only for urban areas, since all of the inundated cells are in areas classed as urban. For the benchmarking at GLOFRIS resolution, the HRs range between 0.67 (Flint) and 0.83 (Mississippi), and for the benchmarking at original resolution they range between 0.60 (Susquehanna) and 0.72 (Mississippi). At GLOFRIS resolution, the FARs for the Flint and Mississippi Rivers are relatively high, at 0.44 and 0.47 respectively. For the Mississippi, the FAR is also relatively high for the benchmarking at the resolution of the original dataset. This can be explained by the overestimation of GLOFRIS in the area at the centre of the case study (Supplementary Figure 30 and 31). This part of the city of St. Paul is protected by a levee that is parameterised in the benchmark inundation model. This area would have flooded if it was not for this levee, and since levees are not included in the GLOFRIS inundation module, this explains the difference here. However, when calculating flood risk, we do consider the presence of flood protection, by truncating the risk curve. If this levee were included in the inundation module directly, the FAR would be ca. 0.23. For the Flint river, the FAR is lower when the benchmarking is carried out at the resolution of the original dataset. Moving to the somewhat longer reach of the Susquehanna River, we see higher performance metrics, with an overall CSI of 0.59 at GLOFRIS resolution and 0.53 at original resolution.

St. Louis flood (Mississippi River, USA)

Along this reach of the Mississippi, Missouri, and Illinois Rivers, the return period of flooding during the 1993 event was fairly constant, at ca. 100 years[14]. Therefore, it is reasonable to compare this event with our GLOFRIS 100 year return period inundation map on the main reaches. The benchmarking results show strong agreement between the GLOFRIS and benchmark datasets at both benchmarking analysis resolutions (Supplementary Figure 34 and 35). For the entire domain, the HR is high at 0.87 at GLOFRIS resolution, and 0.80 at original resolution, with a low FAR of 0.17 at GLOFRIS resolution and 0.19 at original resolution. When calculated for urban areas only, there is little change in the performance metrics. As a result, simulated maximum potential damage using the GLOFRIS inundation map is just 11% higher than using the benchmark inundation map. This is much smaller than the difference introduced through the use of the alternative assumptions on current flood protection standards, and the standard deviation of change in future EAD due to the use of different GCMs, RCPs, and SSPs.

One area of overestimation in GLOFRIS can be seen in the southern section of the reach. In this area a levee can be seen in the original image, which explains this. Most of the underestimation in GLOFRIS occurs around the confluence of the Mississippi and Illinois rivers, which can be explained by the fact that our volume-spreading algorithm does not simulate backwater effects. Still, this case study shows GLOFRIS to be performing well for a large flood in a highly urbanised area.

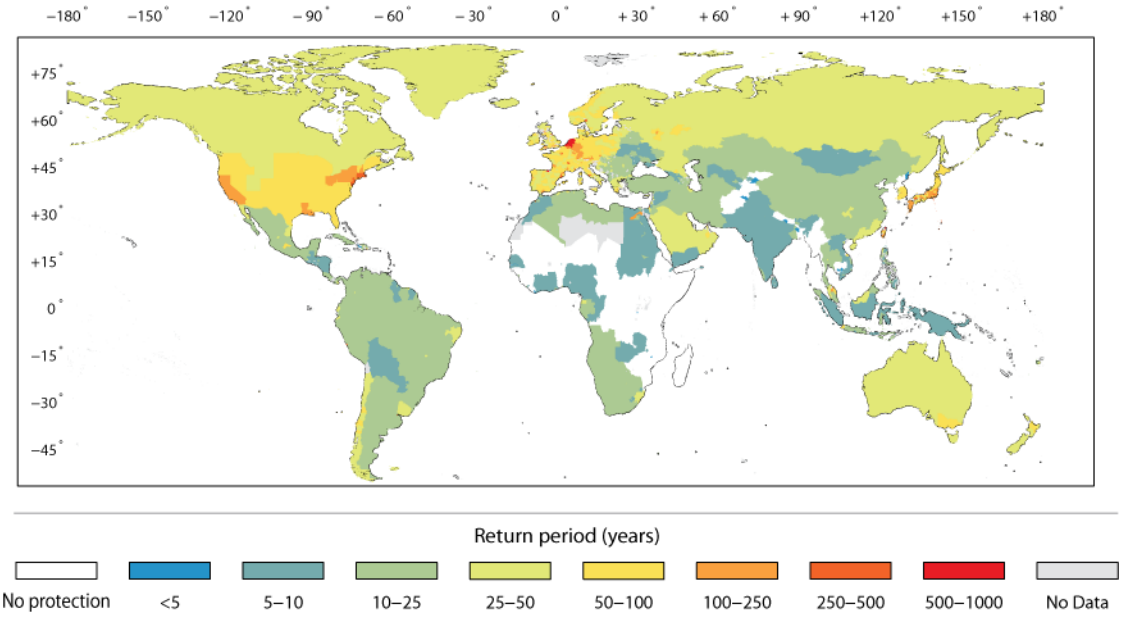
Chao Phraya River flood (Thailand)

The long duration flood event in Thailand was the result of complex hydroclimatic processes across the entire catchment. As such, it is difficult to determine a representative return period for this flood, and

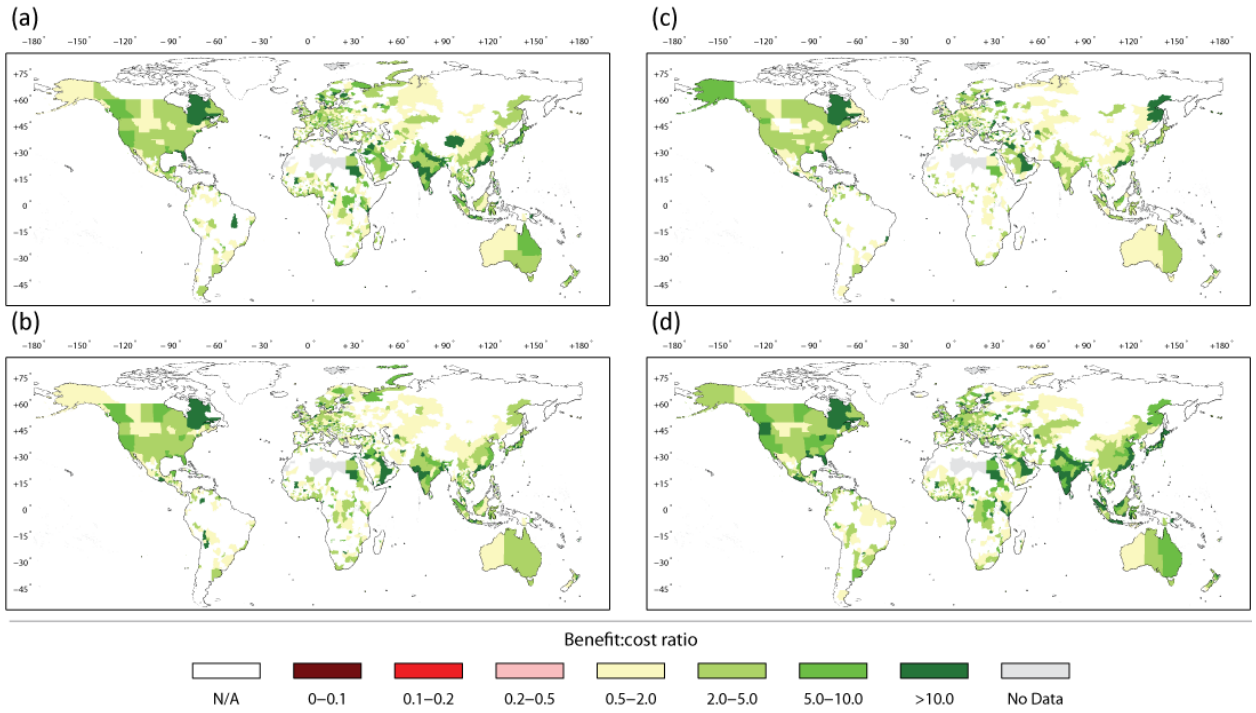
the actual flood period is expected to be much less constant than for the main reach of the Mississippi River in the St. Louis case study. We have chosen to compare the benchmark inundation data to the GLOFRIS 25 year return period inundation map, although we stress that this will lead to large differences in some areas. Given these caveats, the performance metrics are encouraging. Across the entire domain, we see a HR of 0.65 (at both GLOFRIS and benchmark resolution) and a FAR of 0.29 (GLOFRIS resolution) or 0.31 (original resolution). Simulated maximum potential damage using the GLOFRIS inundation map is 36% higher than using the benchmark inundation map. This is much smaller than the difference introduced through the use of the alternative assumptions on current flood protection standards, and the standard deviation of change in future EAD due to the use of different GCMs, RCPs, and SSPs.

The HR is very similar when calculated for urban areas only, but the FAR shows a greater increase, to 0.42 (GLOFRIS resolution) or 0.45 (original resolution). Closer inspection of the results (Supplementary Figure 36 and 37) reveals that this increase can largely be attributed to the fact that GLOFRIS shows extensive inundation in the Thonburi area of Bangkok (along the western bank of the Chao Phraya River in central Bangkok), whilst the benchmark dataset shows very limited areas of flooding in this area. In reality, parts of Thonburi were flooded up to chest height, but not to the spatial extent suggested by GLOFRIS. However, this was one of the parts of the cities where concern for a massive flood was the greatest, as the defences were almost overwhelmed. In GLOFRIS, we account for these protection standards by using flood protection standards from FLOPROS to truncate the risk curve, thereby assuming no damage below the assumed protection standard. For the Thonburi area, the protection standard in FLOPROS is ca. 30 years. Therefore, even though we overestimate this inundation area, this is corrected for in the calculation of risk. Moreover, the fact that the Thonburi area almost faced severe flooding in this ca. 25 year event, suggests that the estimated flood protection standard in this region is rather good. If we manually assume that no flood impacts occurred in this area, the percentage difference in simulated maximum potential damage using the GLOFRIS inundation map compared to the benchmark inundation map reduces to +22%. It should also be noted that areas upstream from central Bangkok were allowed to flood so as to protect downstream areas[21]. This could explain the fact that the benchmarking map in Supplementary Figure 36 and 37 shows underestimation in the more upstream reaches, and overestimation downstream (especially to the west of Bangkok).

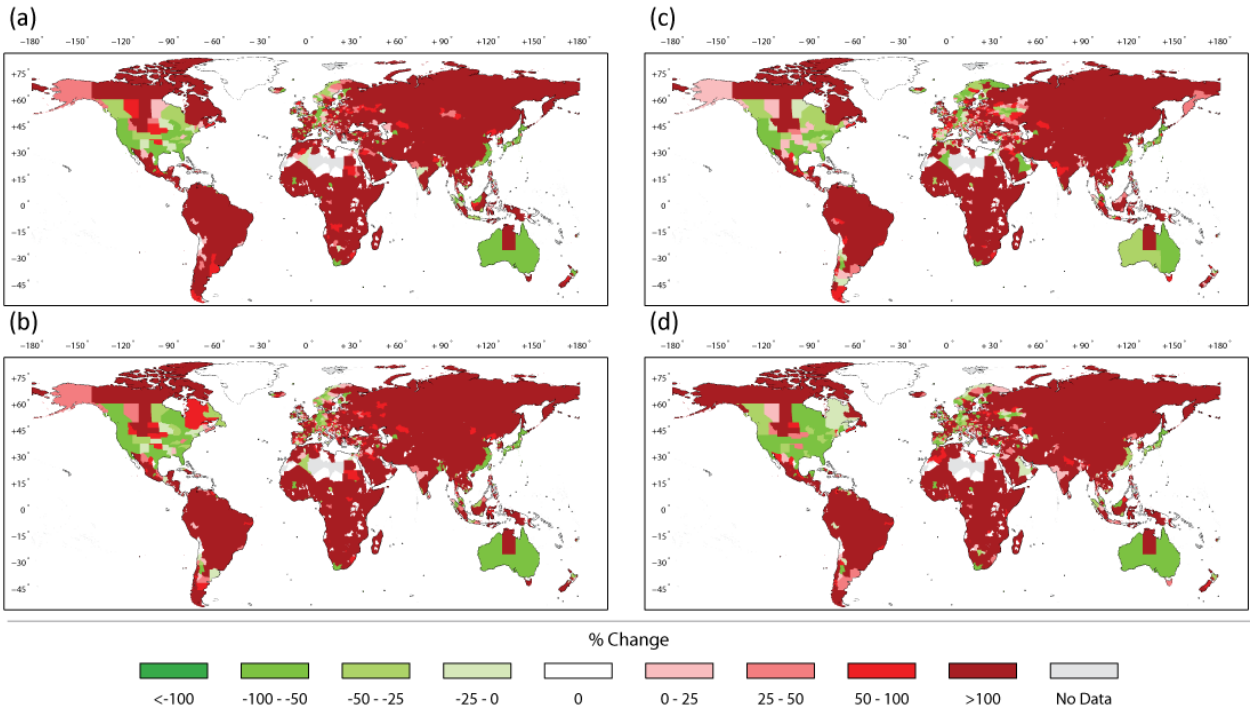
4. Supplementary Figures



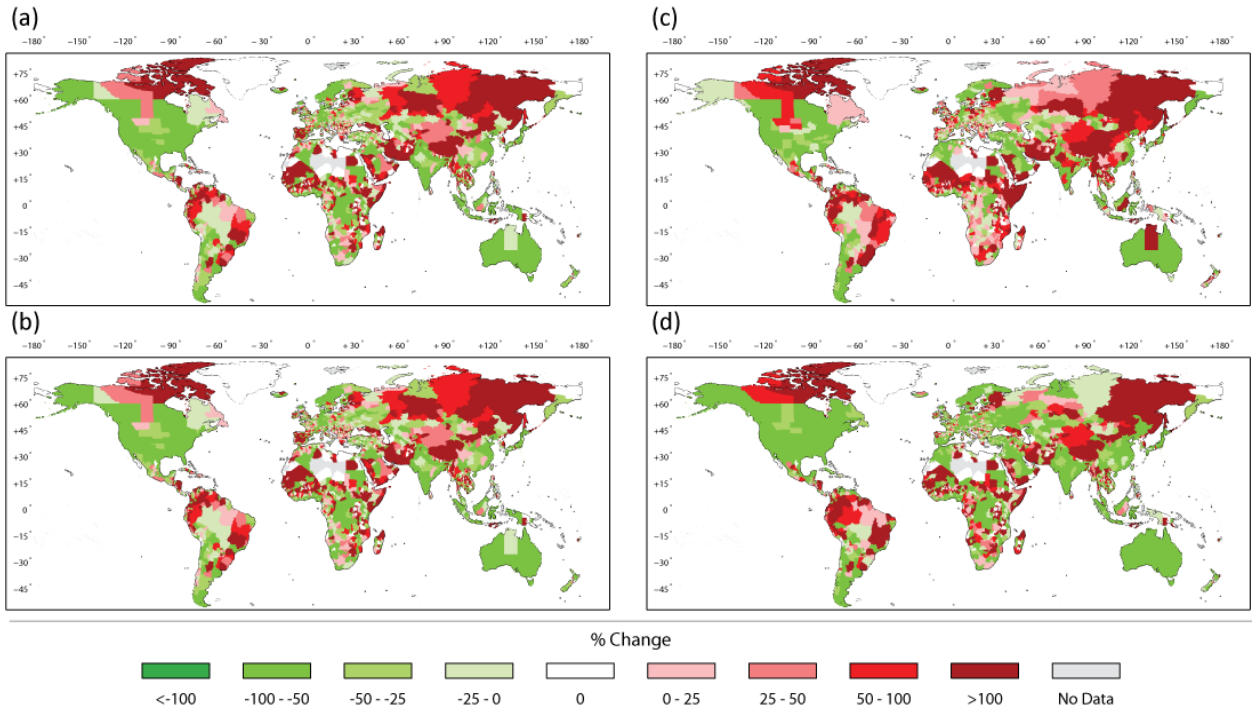
Supplementary Figure 1: Estimated current flood protection standards in urban areas at sub-national level, from the FLOPROS database[4].



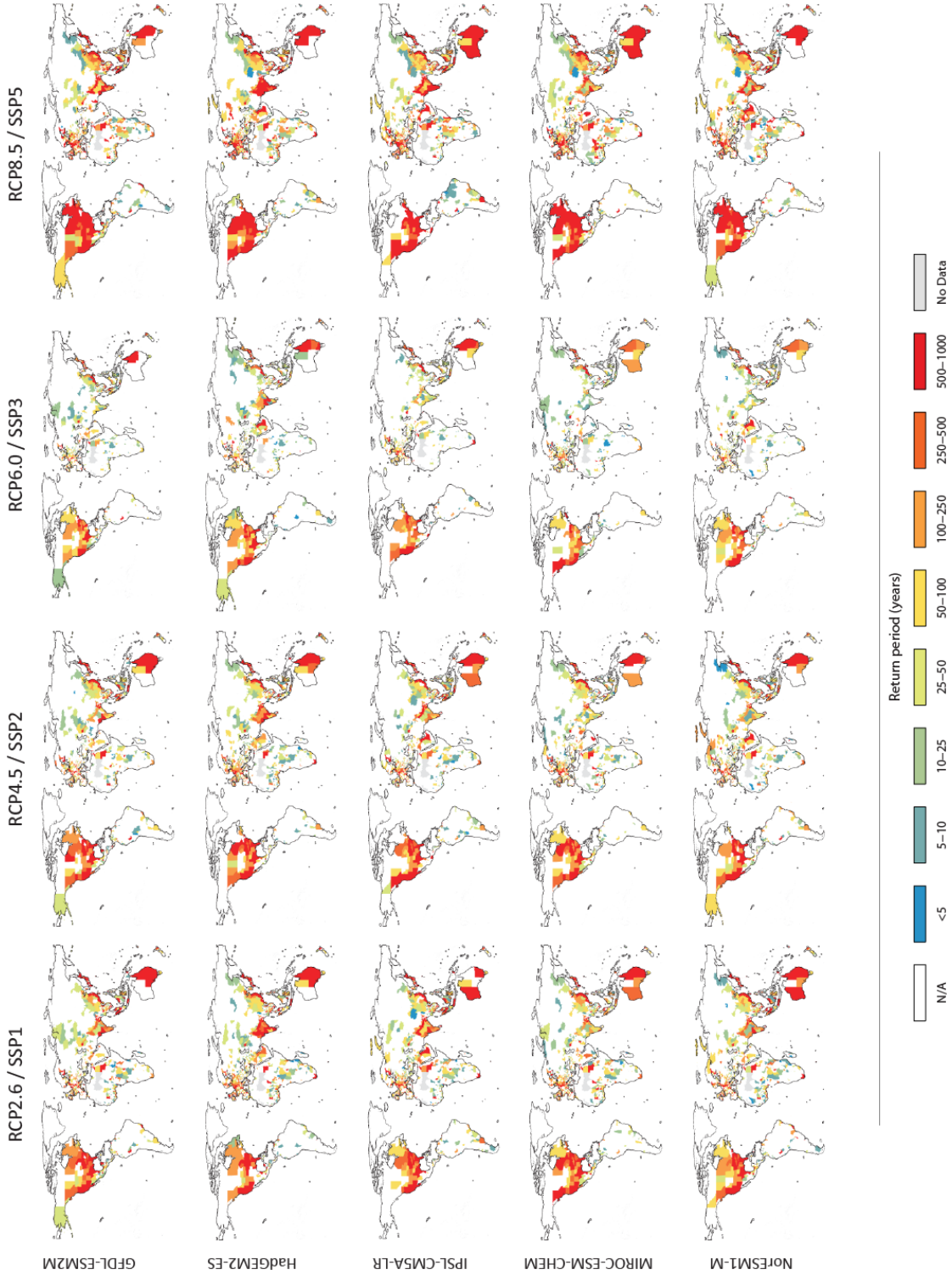
Supplementary Figure 2: Average B:C ratio (averaged across GCMs) per sub-national unit for the optimise objective, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



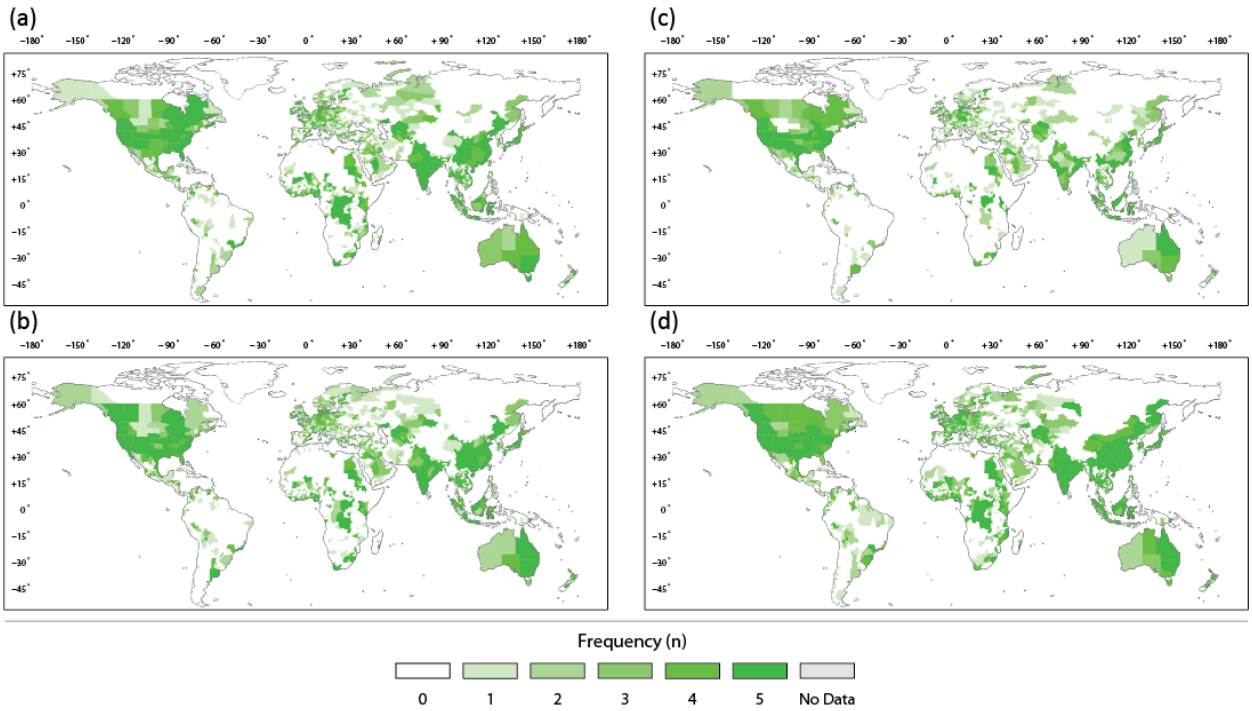
Supplementary Figure 3: Change in future EAD (in 2080) compared to current EAD, assuming the optimal flood protection standards shown in Figure 2. Results are averaged across the five GCMs, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



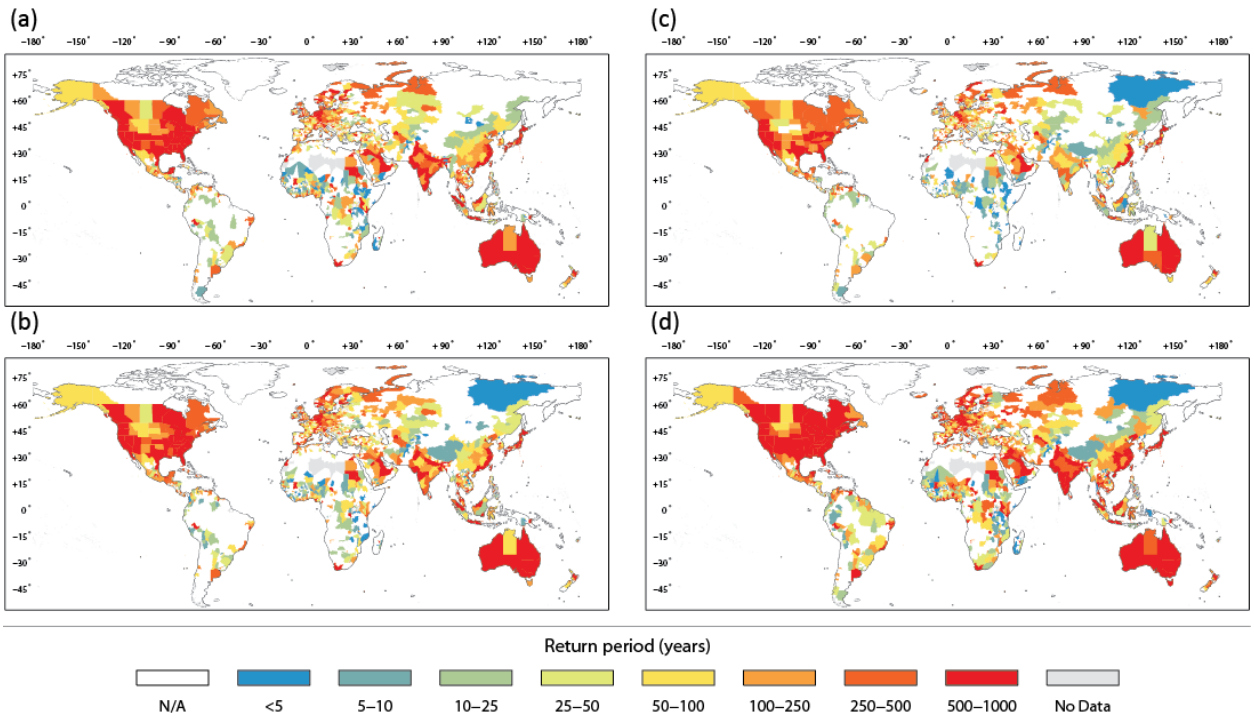
Supplementary Figure 4: Change in future EAD as a percentage of future GDP (in 2080) compared to current EAD as a percentage of current GDP, assuming the flood protection standards shown in Figure 2. Results are averaged across the five GCMs, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



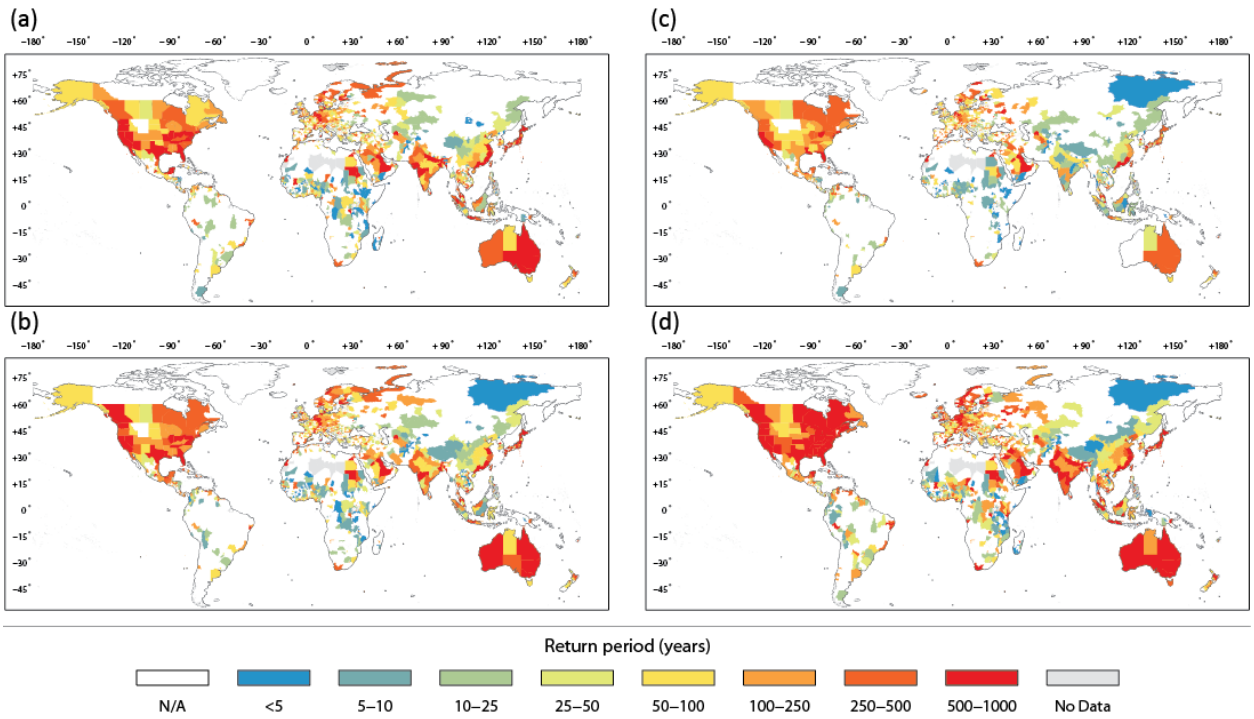
Supplementary Figure 5: Protection standards at sub-national level in 2080 that meet the 'optimise' objective. Results are shown per column for each of the following RCP/SSP combinations: RCP2.6/SSP1; RCP4.5/SSP2; RCP6.0/SSP3; and RCP8.5/SSP5. Results are shown per row for each of the GCMs used in this study. Sub-national units in which no increase in protection standards provides a positive NPV are indicated by N/A. Results are shown assuming middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year



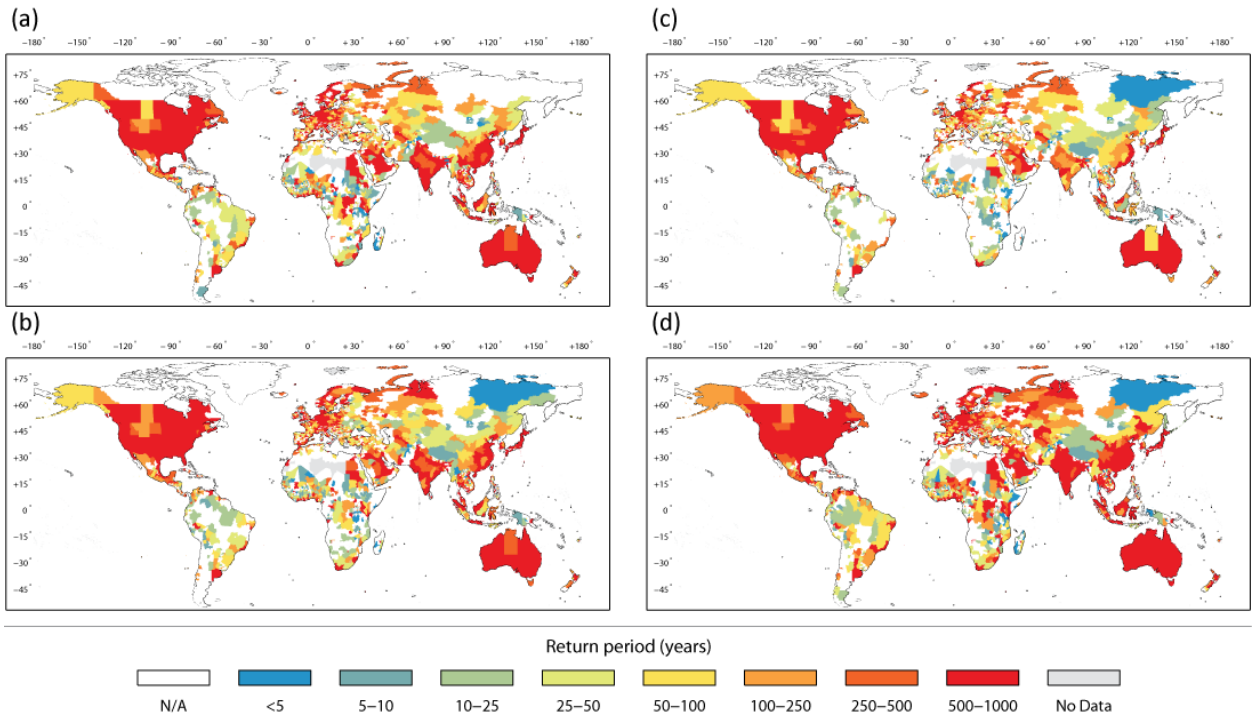
Supplementary Figure 6: The number of GCMs (frequency, n) for which a protection standard can be reached in 2080 for the 'optimise' scenario whereby the NPV is greater than zero (i.e. whereby the B:C ratio exceeds 1). The results in this figure summarise those shown per GCM is Supplementary Figure 5. Results are shown at sub-national level for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results are shown assuming middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year



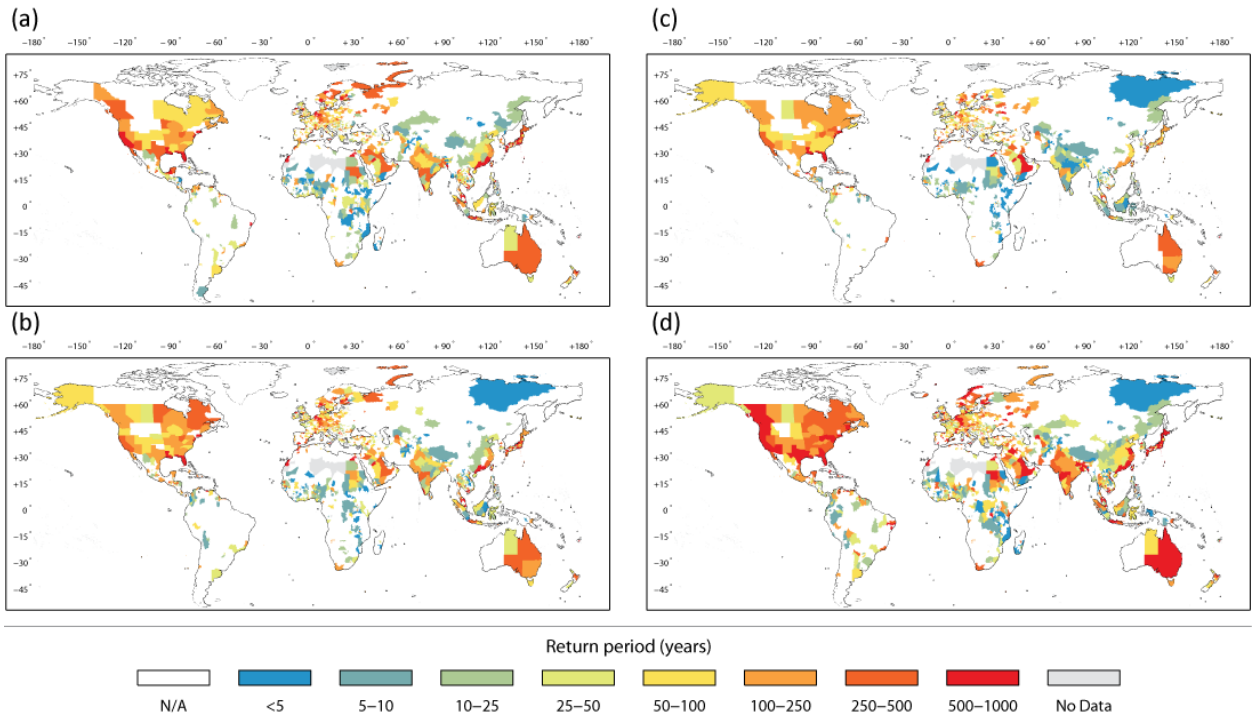
Supplementary Figure 7: Protection standard per sub-national unit in 2080 providing highest net present value (NPV), using middle-estimate investment costs. Results are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. The average return period is shown across the five GCMs. Sub-national units in which no increases in protection standards provide a positive NPV are indicated by N/A. Results are shown for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 3% per year.



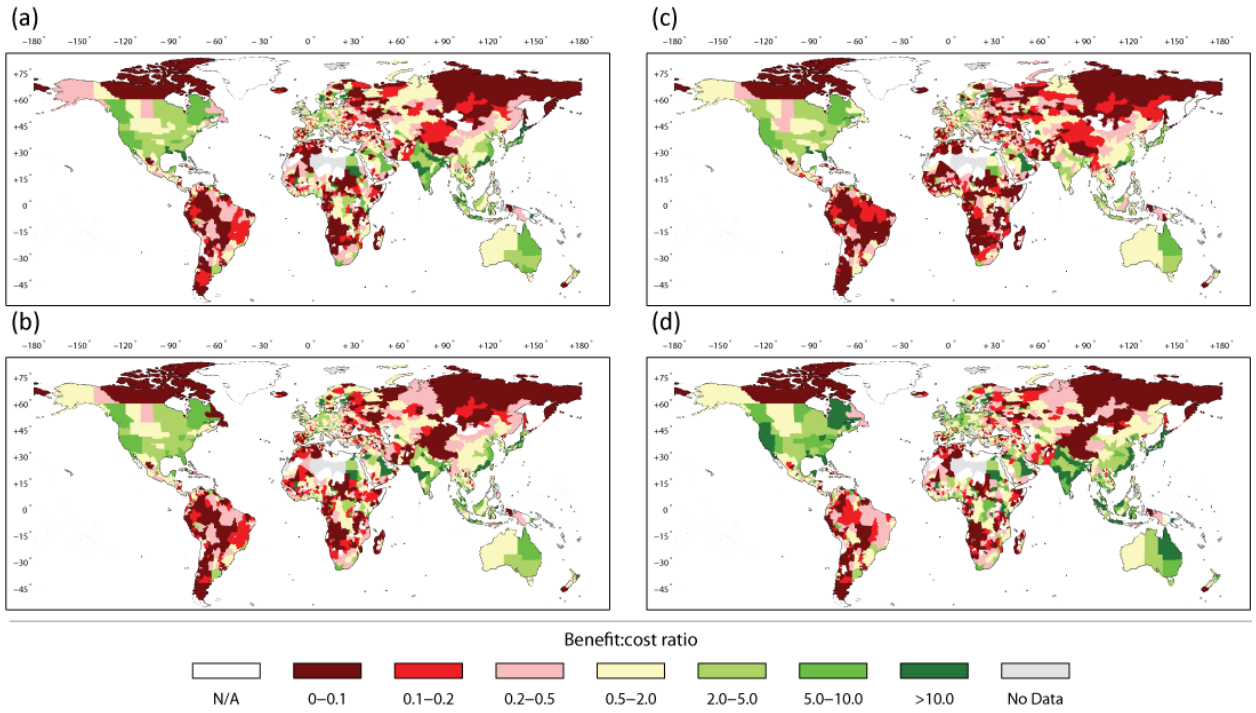
Supplementary Figure 8: Protection standard per sub-national unit in 2080 providing highest net present value (NPV), using middle-estimate investment costs. Results are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. The average return period is shown across the five GCMs. Sub-national units in which no increases in protection standards provide a positive NPV are indicated by N/A. Results are shown for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 8% per year.



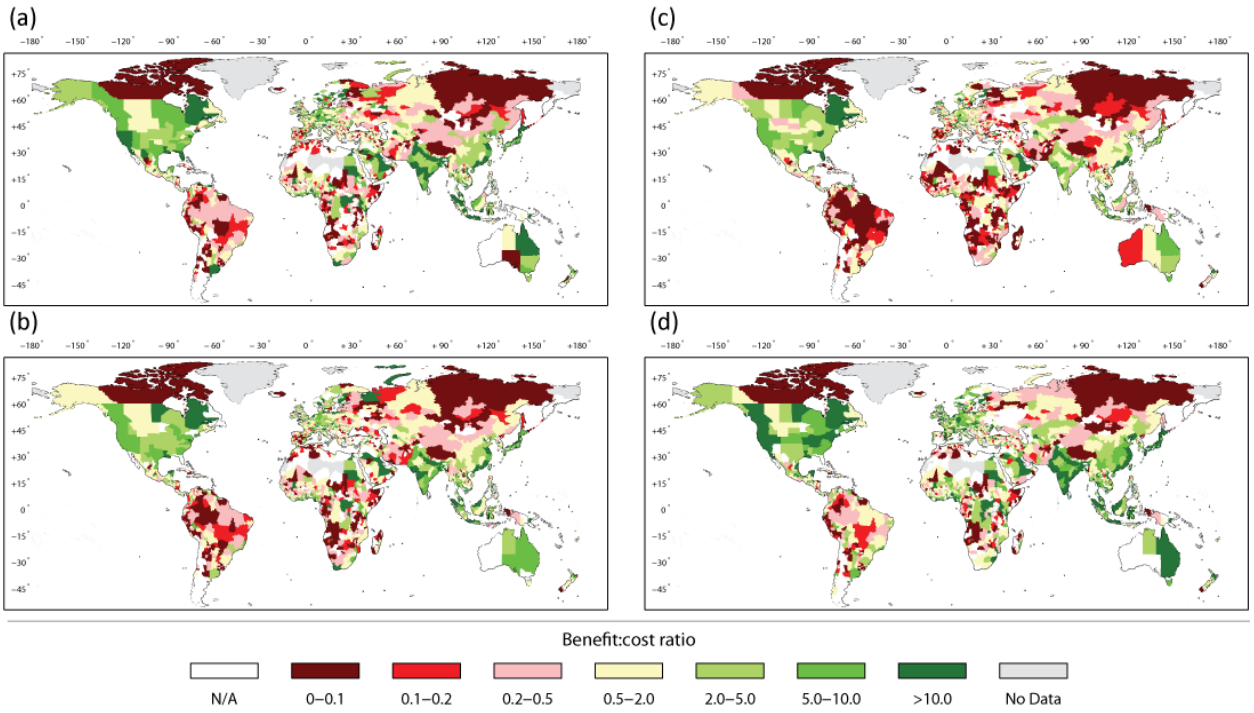
Supplementary Figure 9: Protection standard per sub-national unit in 2080 providing highest net present value (NPV), using low-estimate investment costs. Results are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. The average return period is shown across the five GCMs. Sub-national units in which no increases in protection standards provide a positive NPV are indicated by N/A. Results are shown for the following assumptions: low-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



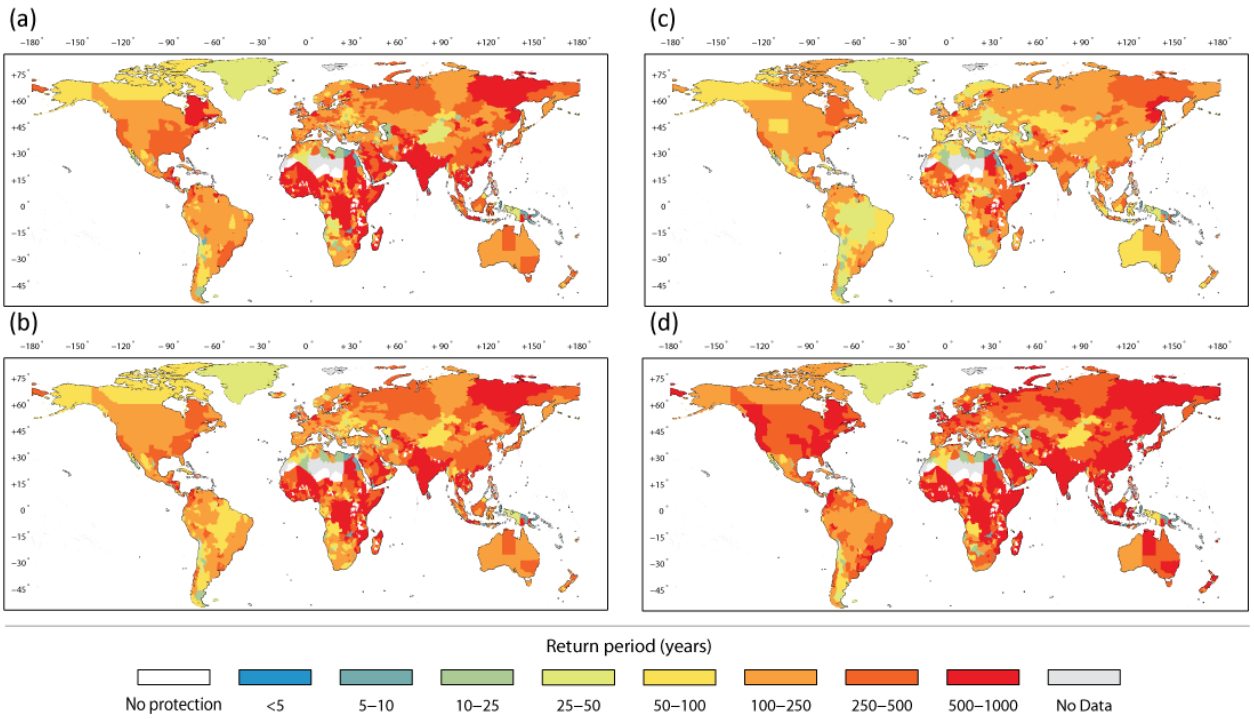
Supplementary Figure 10: Protection standard per sub-national unit in 2080 providing highest net present value (NPV), using high-estimate investment costs. Results are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. The average return period is shown across the five GCMs. Sub-national units in which no increases in protection standards provide a positive NPV are indicated by N/A. Results are shown for the following assumptions: high-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



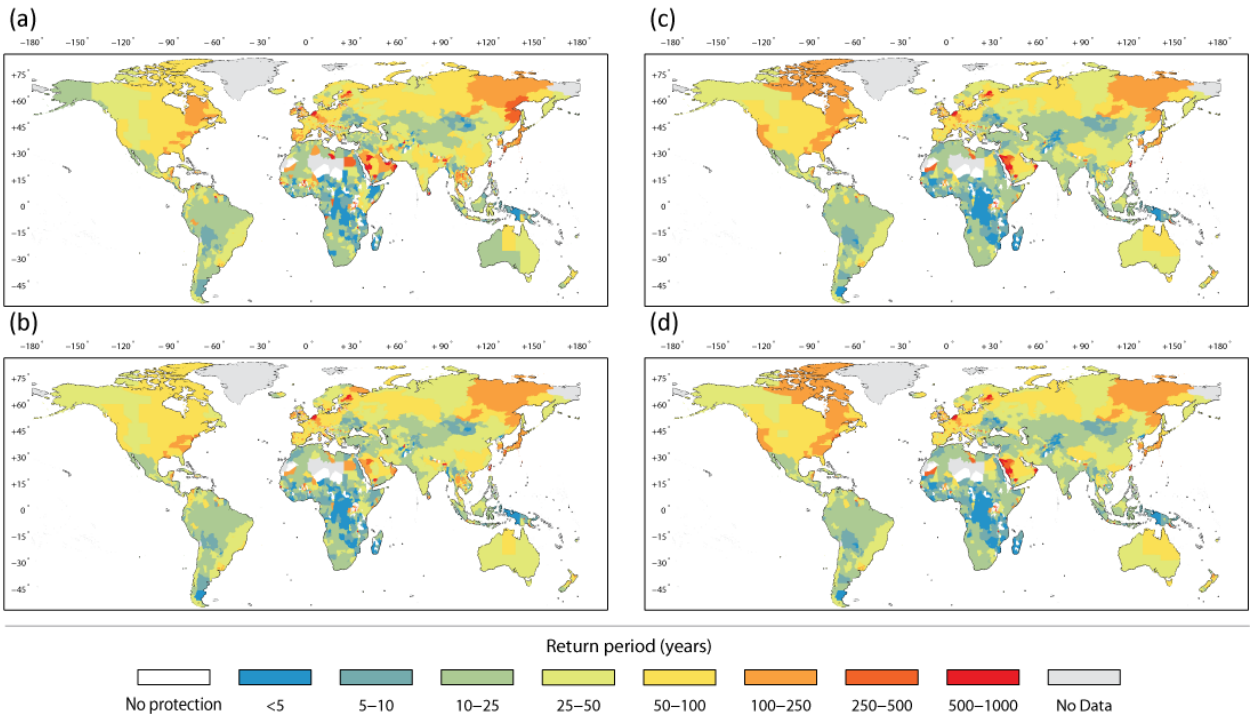
Supplementary Figure 11: Average B:C ratio (averaged across GCMs) per sub-national unit for the constant absolute risk objective, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



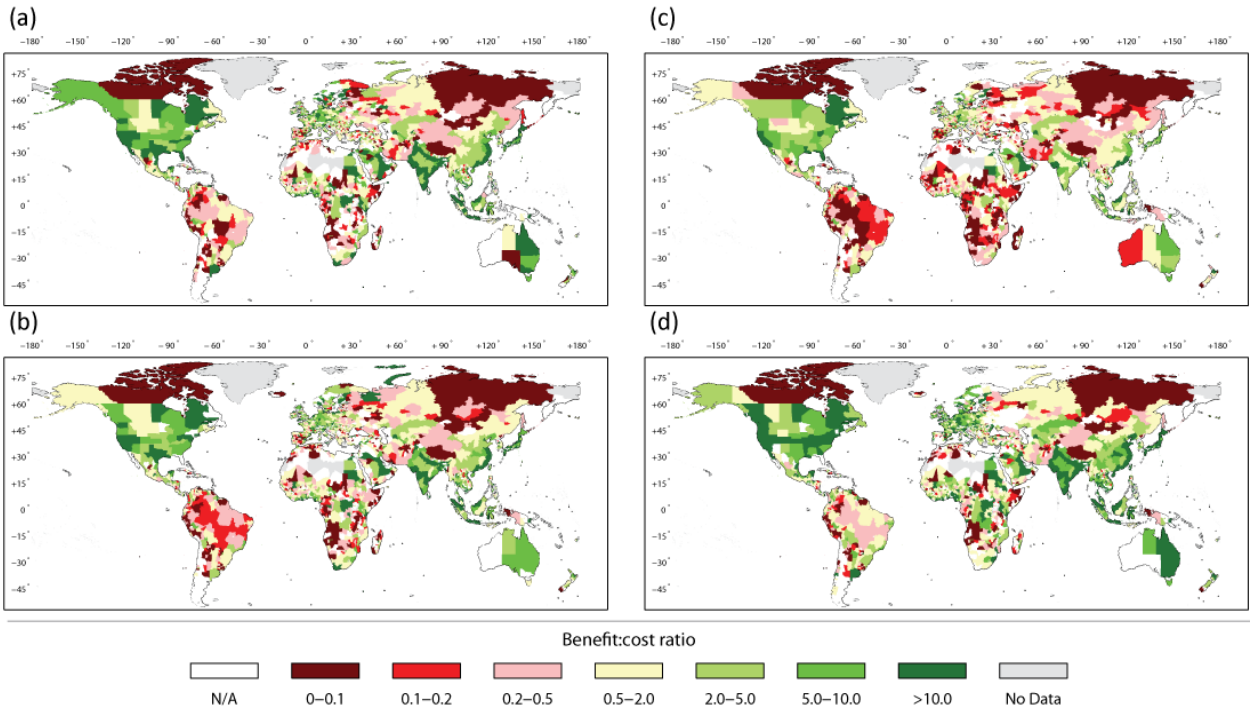
Supplementary Figure 12: Average B:C ratio (averaged across GCMs) per sub-national unit for the constant relative risk objective, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



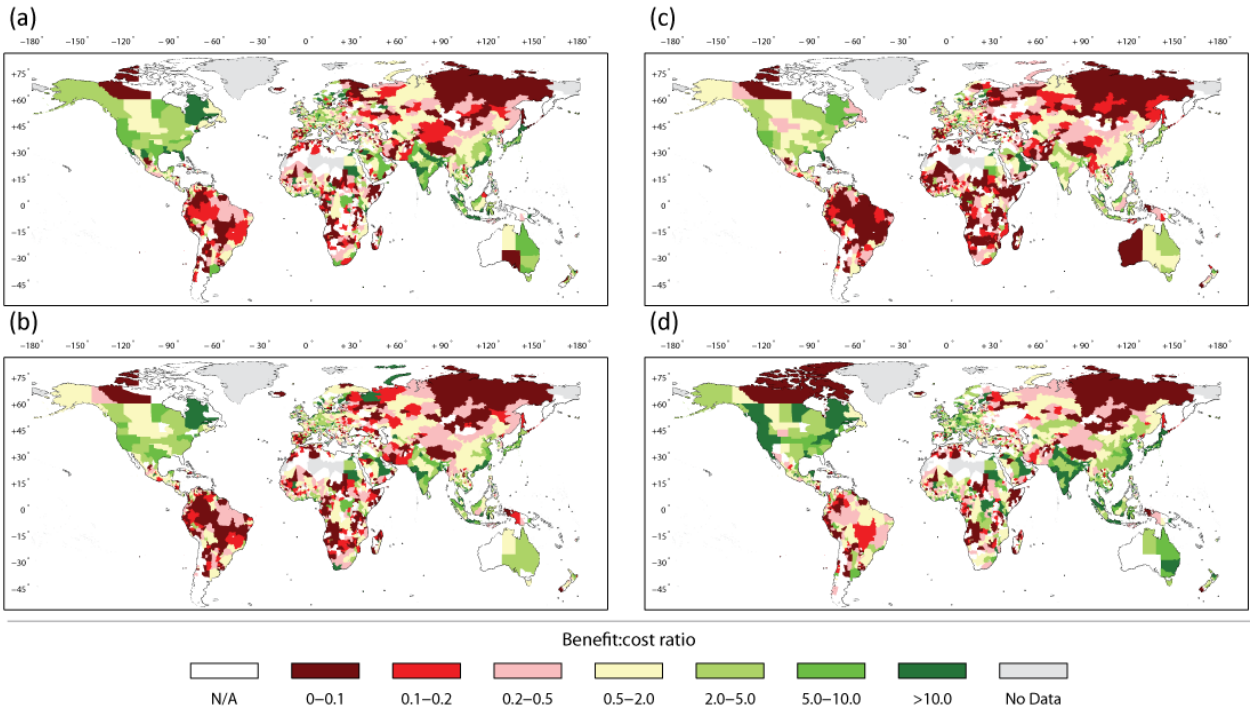
Supplementary Figure 13: Simulated flood protection standards per sub-national unit required by 2080 for the constant absolute risk objective. The flood protection standards shown here are averaged across the 5 GCMs, and are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



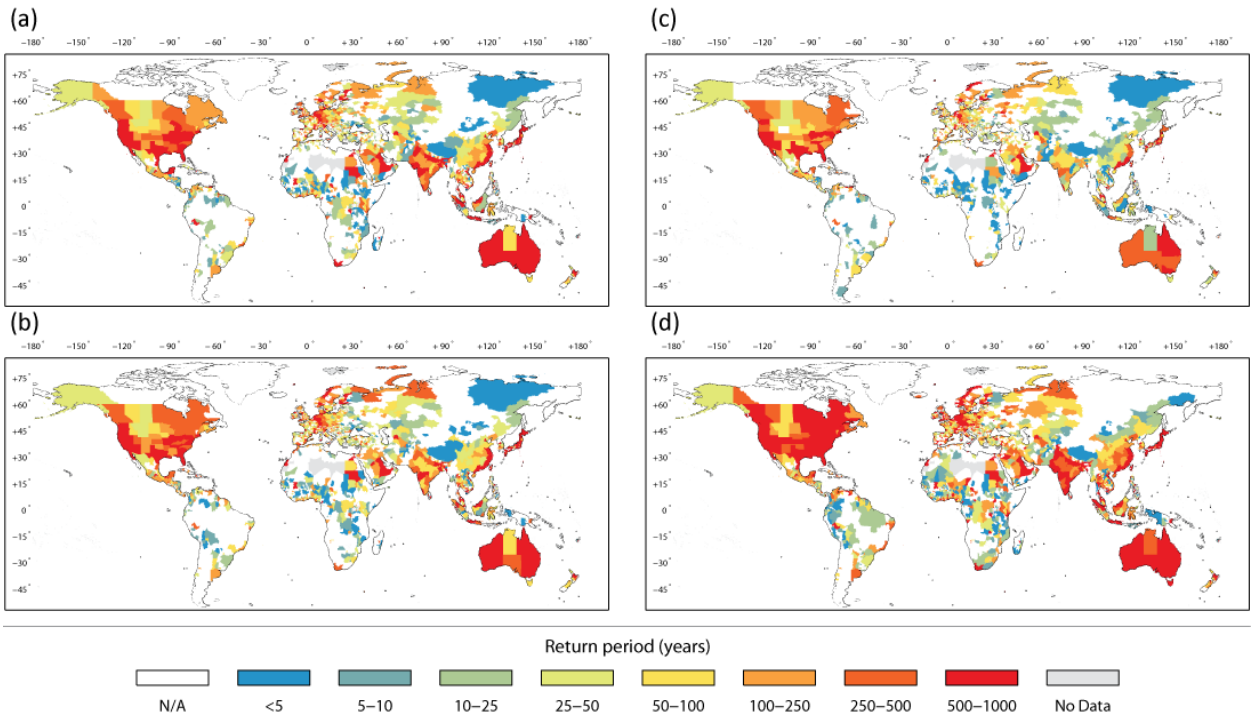
Supplementary Figure 14: Simulated flood protection standards per sub-national unit required by 2080 for the constant relative risk objective. The flood protection standards shown here are averaged across the 5 GCMs, and are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year.



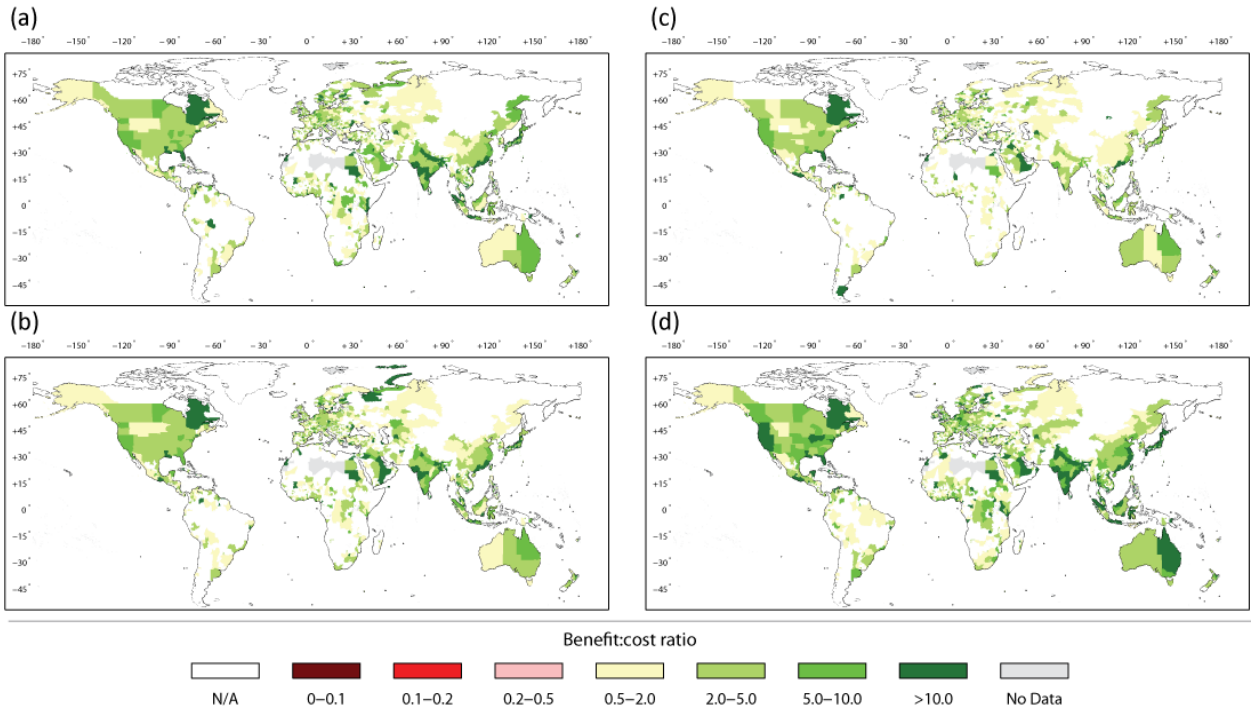
Supplementary Figure 15: Average B:C ratio (averaged across GCMs) per sub-national unit for the constant relative risk objective, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 3% per year.



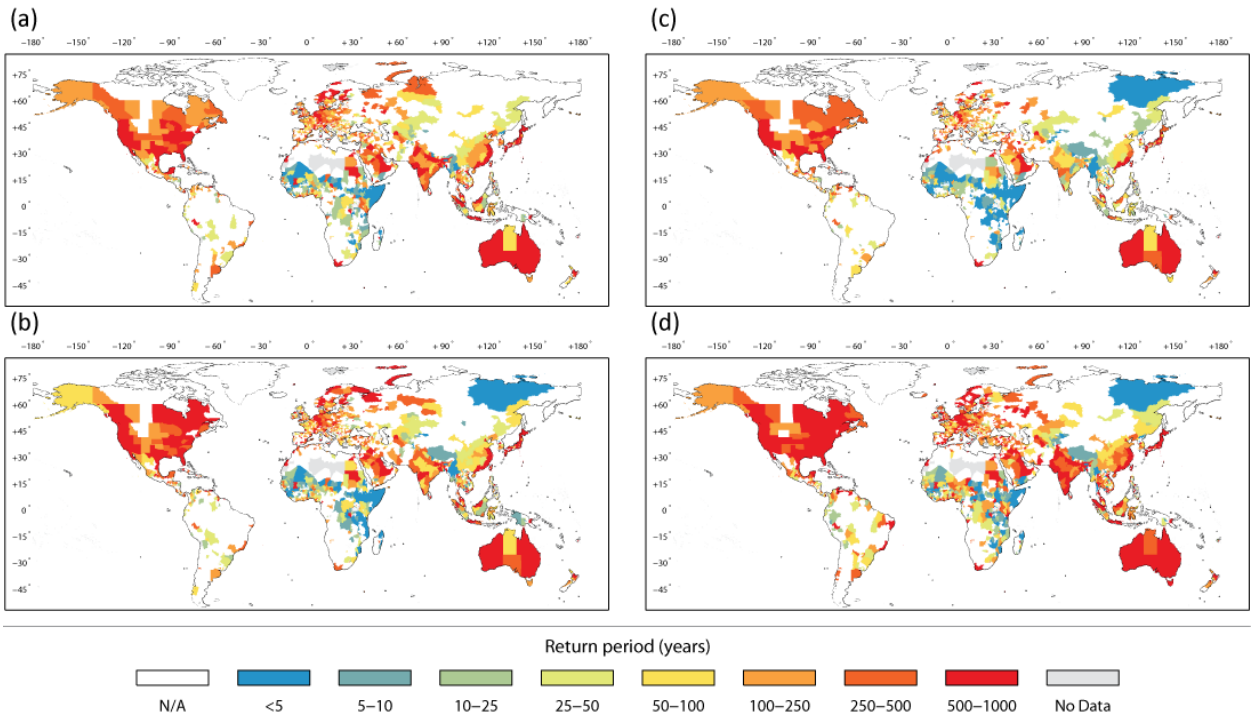
Supplementary Figure 16: Average B:C ratio (averaged across GCMs) per sub-national unit for the constant relative risk objective, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 8% per year.



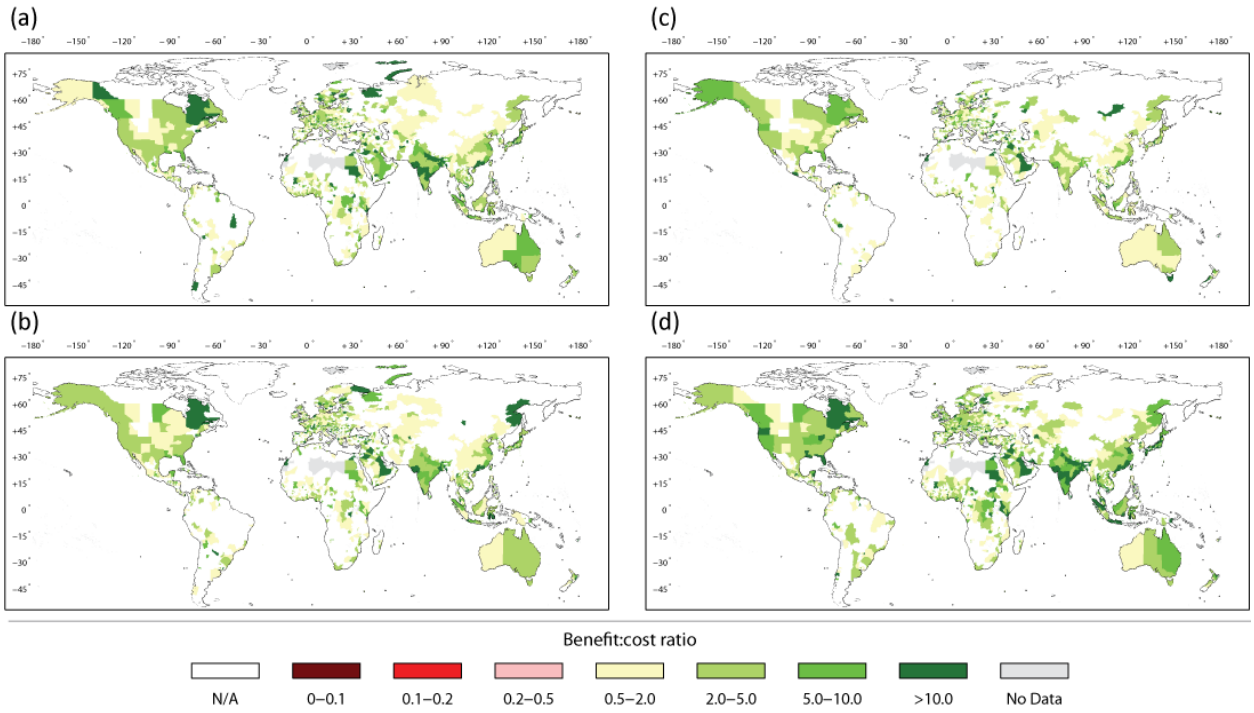
Supplementary Figure 17: Protection standard per sub-national unit in 2080 providing highest net present value (NPV), assuming a current protection standard equal to half of the protection standard stated in FLOPROS. Results are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. The average return period is shown across the five GCMs. Sub-national units in which no increase in protection standard provides a positive NPV are indicated by N/A. Results are shown for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year



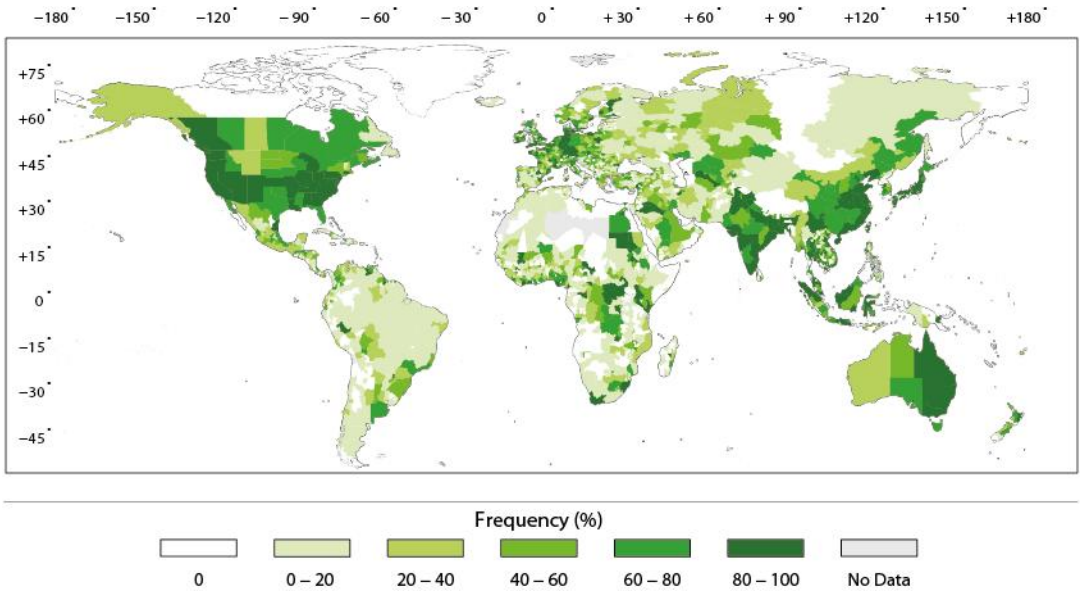
Supplementary Figure 18: Average B:C ratio (averaged across GCMs) per sub-national unit for the optimise objective, assuming a current protection standard equal to half of the protection standard stated in FLOPROS. Results are shown for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year



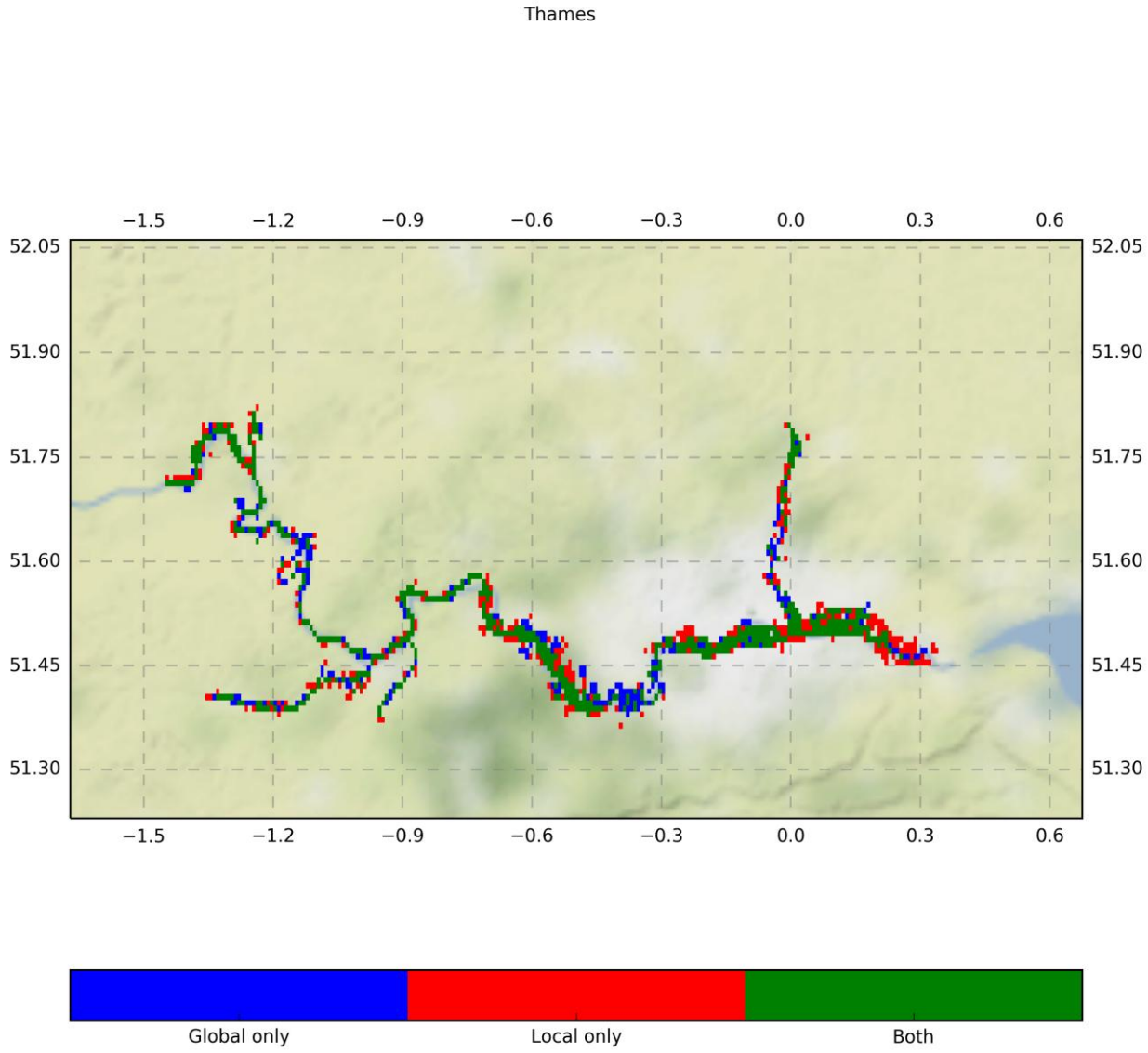
Supplementary Figure 19: Protection standard per sub-national unit in 2080 providing highest net present value (NPV), assuming a current protection standard equal to double the protection standard stated in FLOPROS. Results are shown for: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. The average return period is shown across the five GCMs. Sub-national units in which no increase in protection standard provides a positive NPV are indicated by N/A. Results are shown for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year



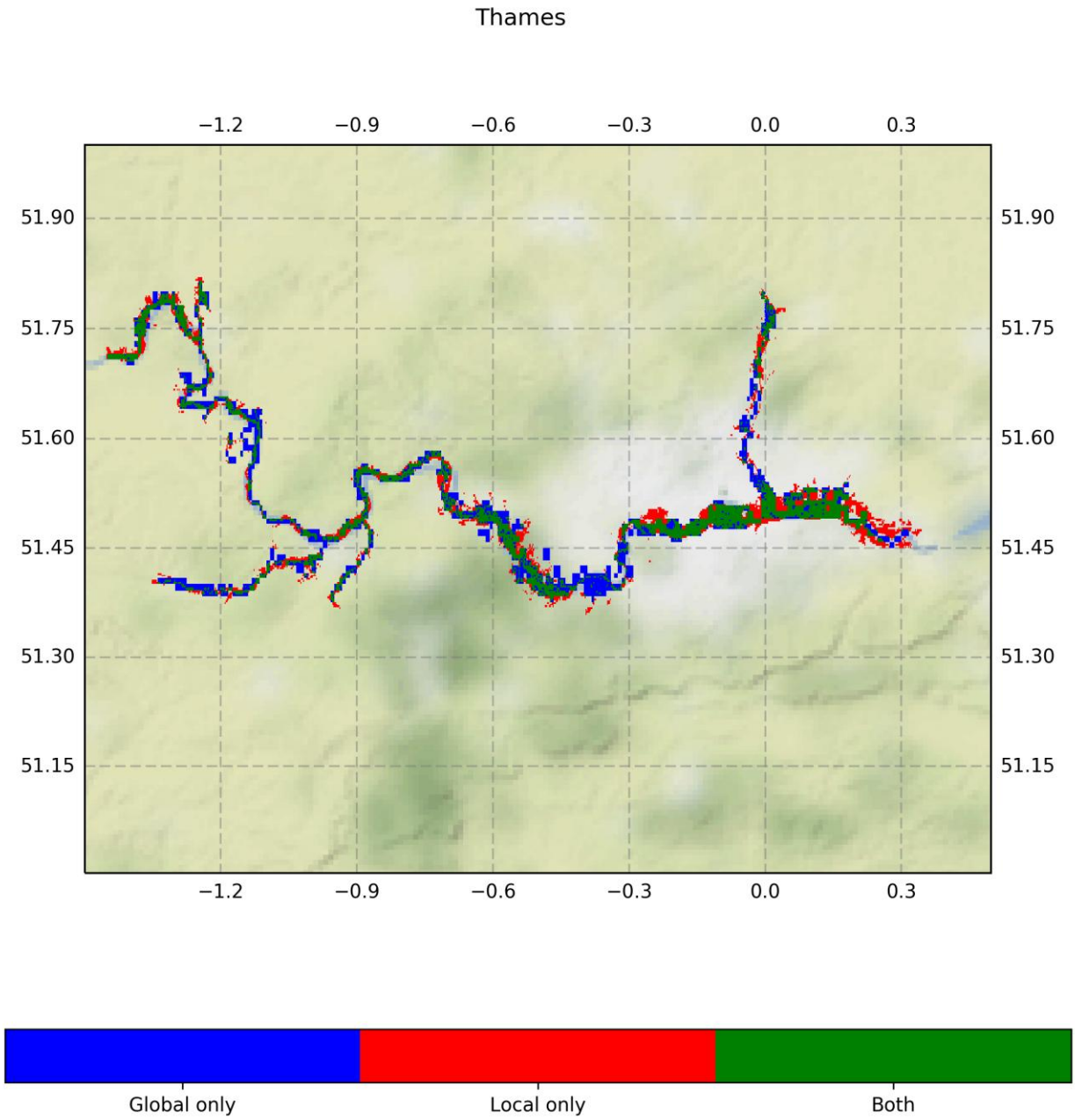
Supplementary Figure 20: Average B:C ratio (averaged across GCMs) per sub-national unit for the optimise objective, assuming a current protection standard equal to double the protection standard stated in FLOPROS. Results are shown for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. Results shown are for the following assumptions: middle-estimate investment costs, maintenance costs of 1% per year, and discount rate of 5% per year



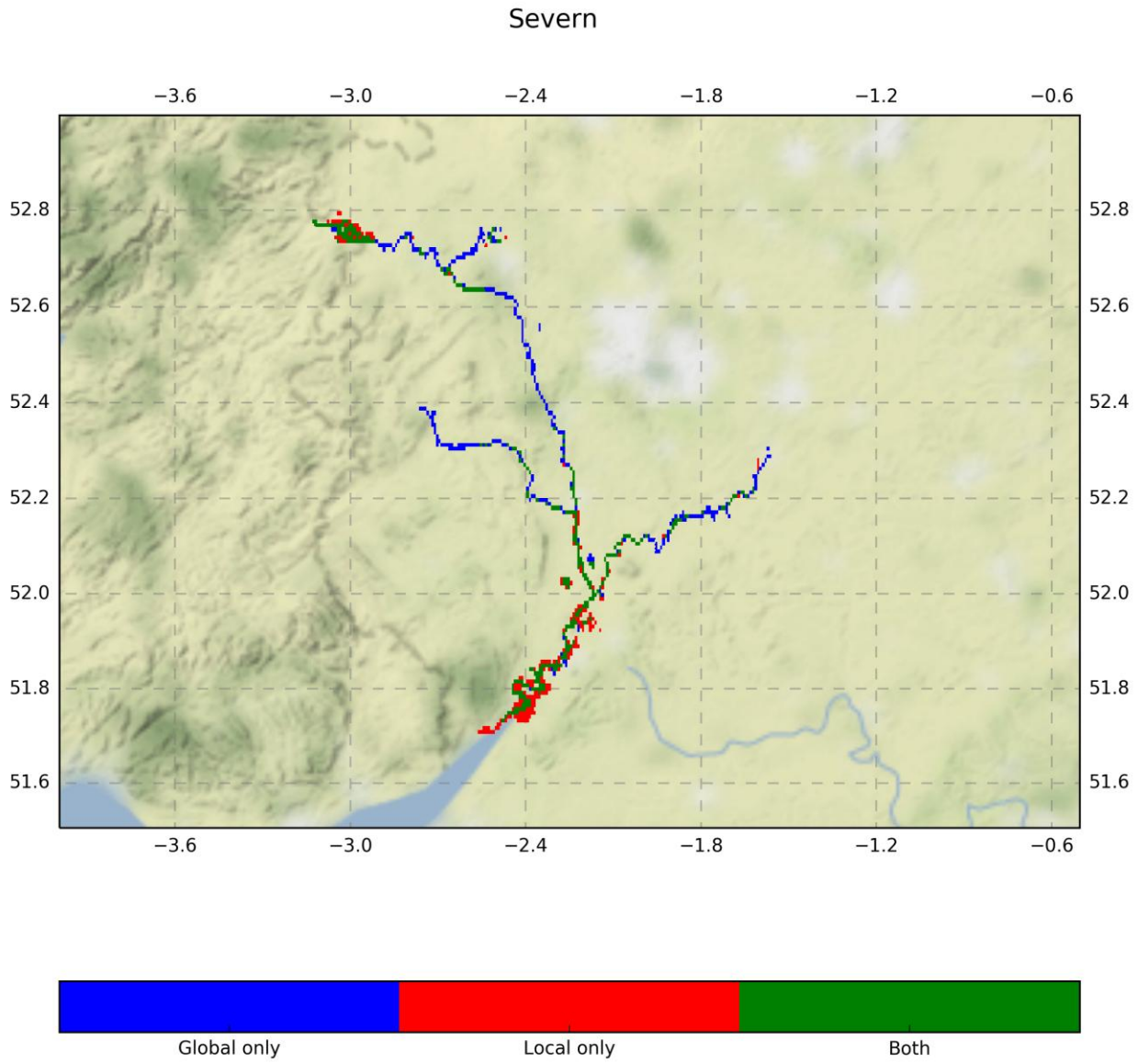
Supplementary Figure 21: Robustness analysis of B:C ratios for the optimise objective. For each sub-national unit, the percentage of the simulations (Frequency, %) is shown for which a B:C ratio exceeding 1 can be achieved under the optimise objective, across all possible combinations of the experiment parameters used in this study (2700 parameter combinations in total), namely: (a) 5 GCMs; (b) 5 SSPs; (c) 4 RCPs; (d) 3 cost estimates (high, middle, low); (e) 3 discount rates (3, 5, and 8% per year); and (f) 3 estimates of current protection standards (FLOPROS, FLOPROS standards halved, and FLOPROS standards doubled)



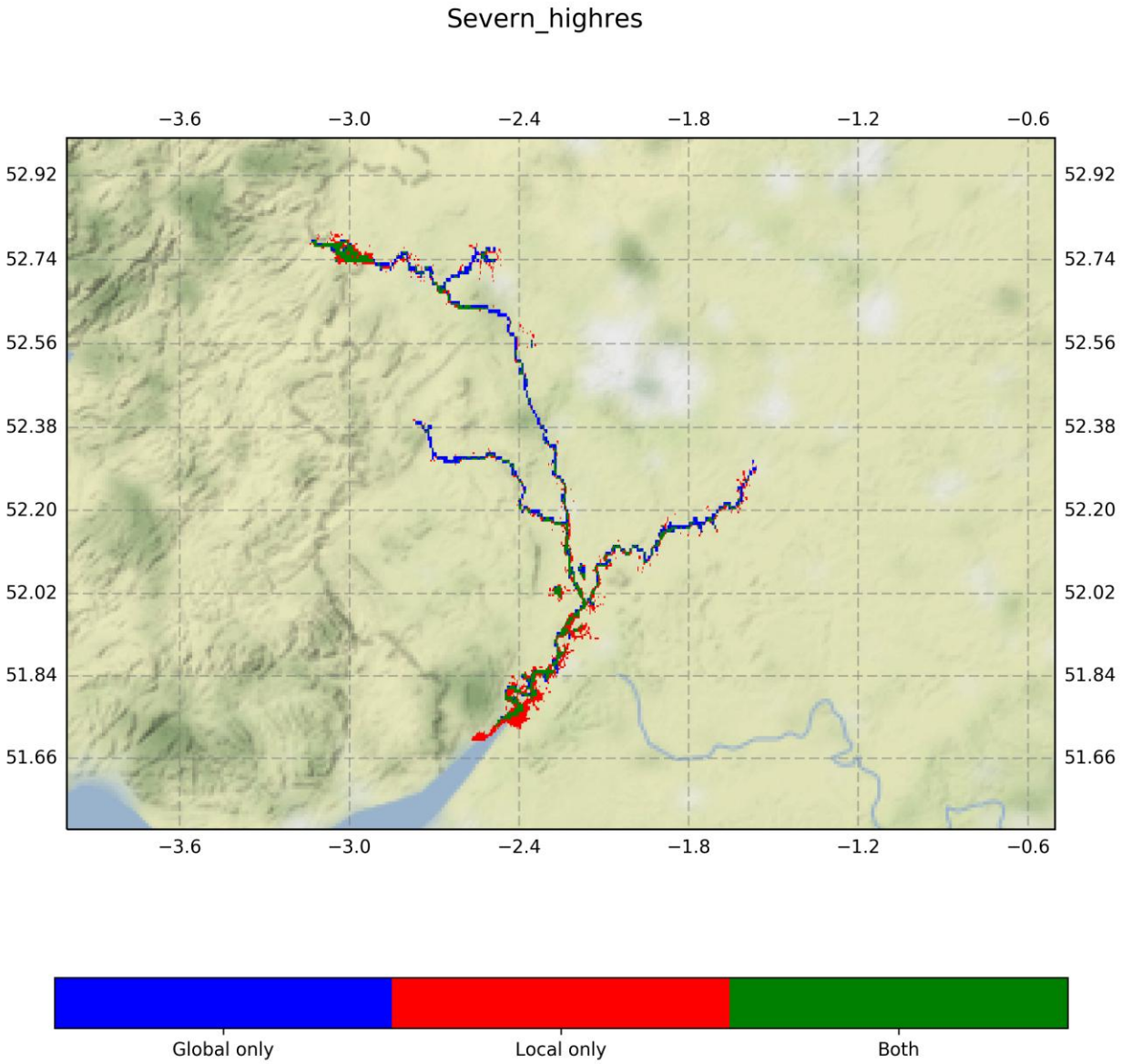
Supplementary Figure 22: Map showing the agreement between the GLOFRIS inundation extent and the Thames River (UK) benchmark data at 30" x 30". Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



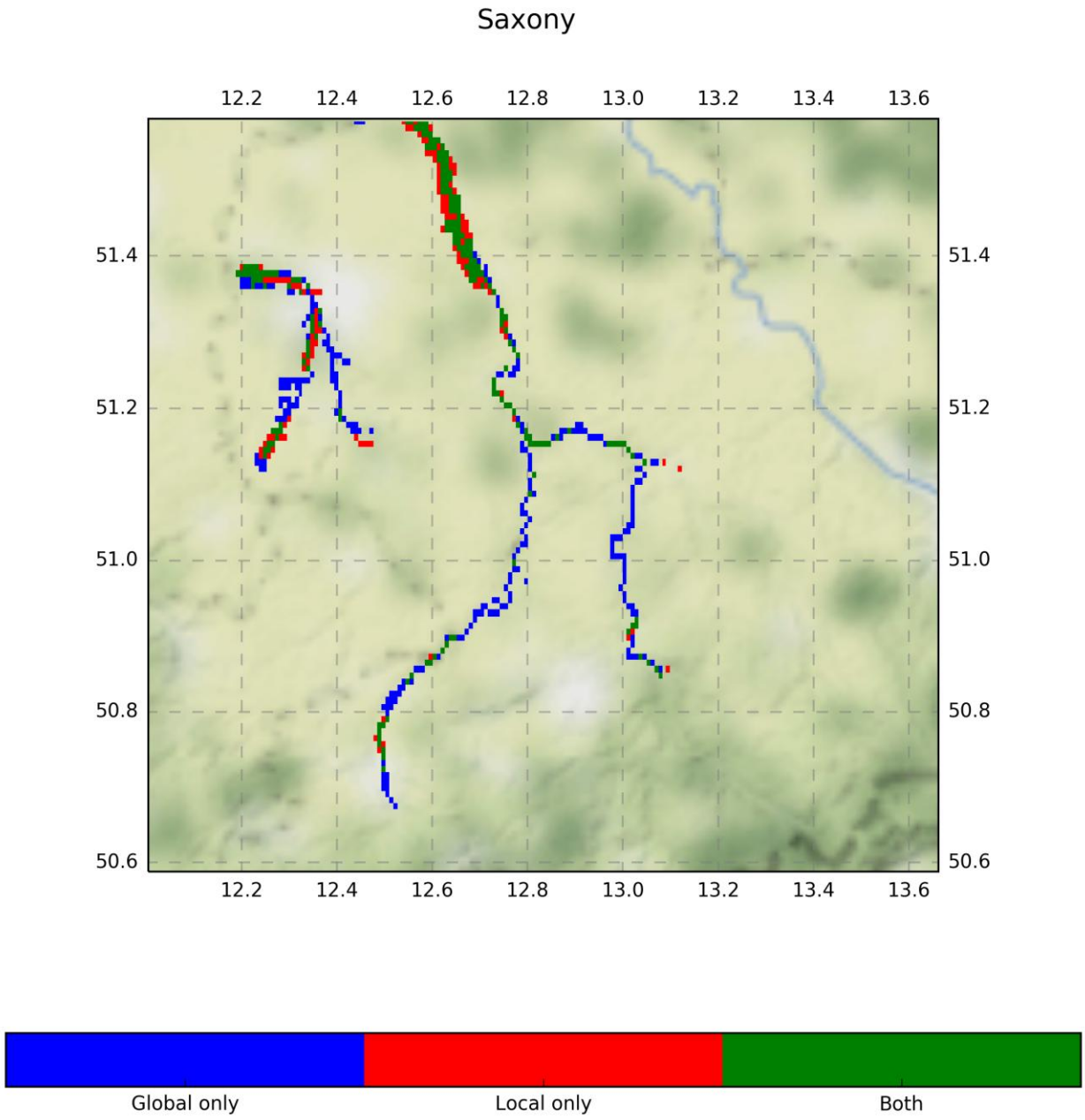
Supplementary Figure 23: Map showing the agreement between the GLOFRIS inundation extent and the Thames River (UK) benchmark data at the resolution of the original benchmark dataset. Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



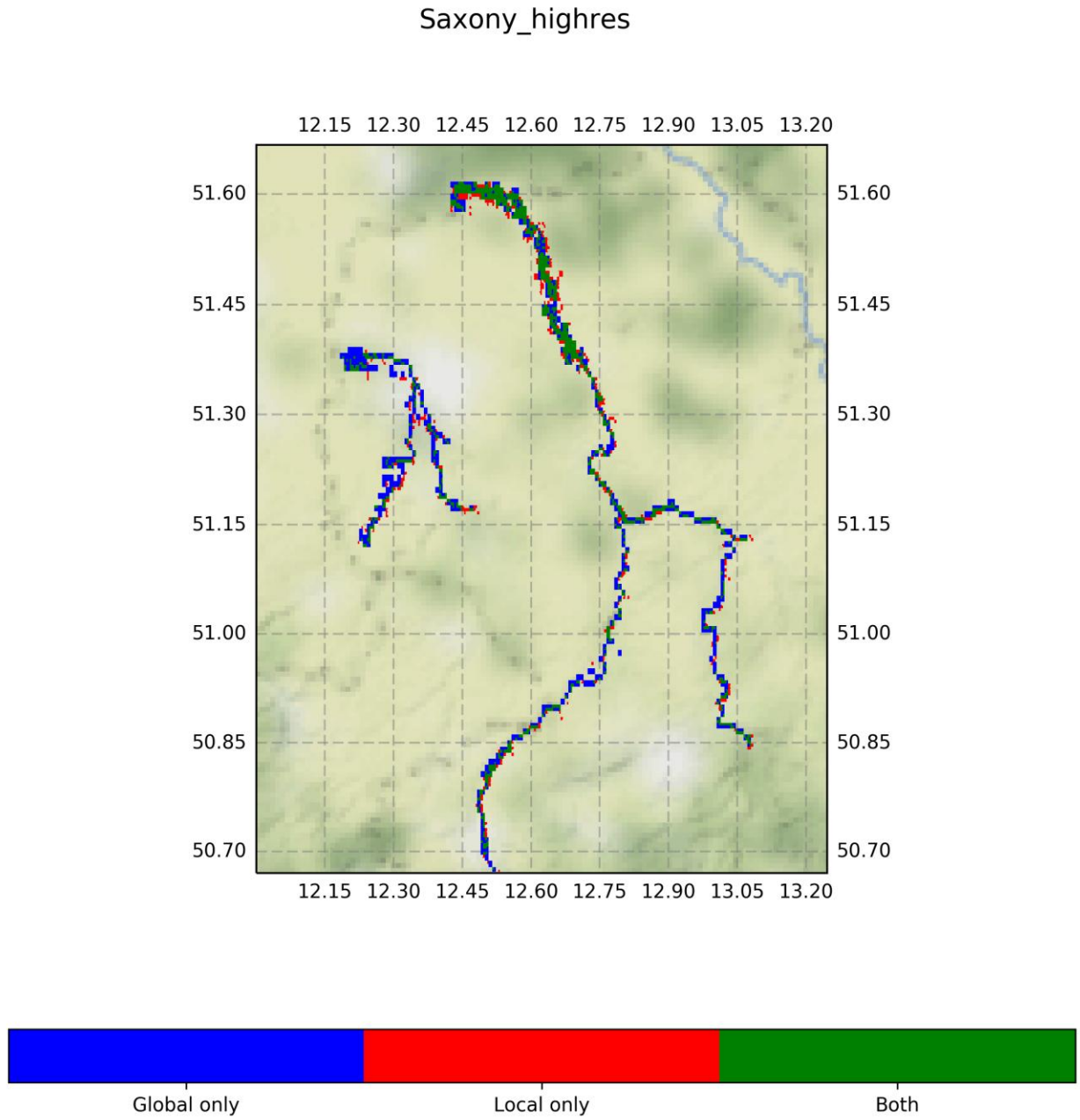
Supplementary Figure 24: Map showing the agreement between the GLOFRIS inundation extent and the Severn River (UK) benchmark data at 30" x 30". Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



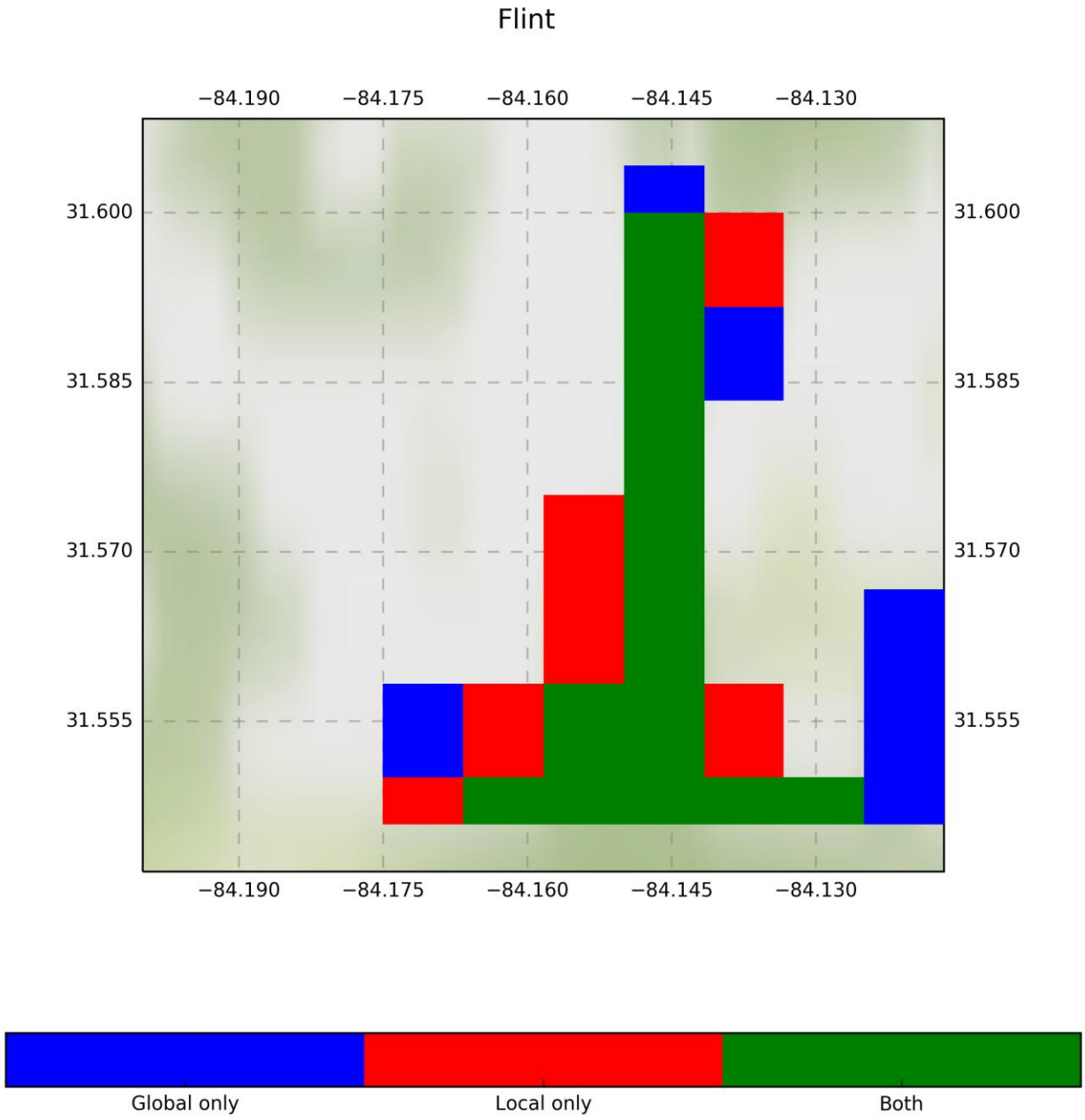
Supplementary Figure 25: Map showing the agreement between the GLOFRIS inundation extent and the Severn River (UK) benchmark data at the resolution of the original benchmark dataset. Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



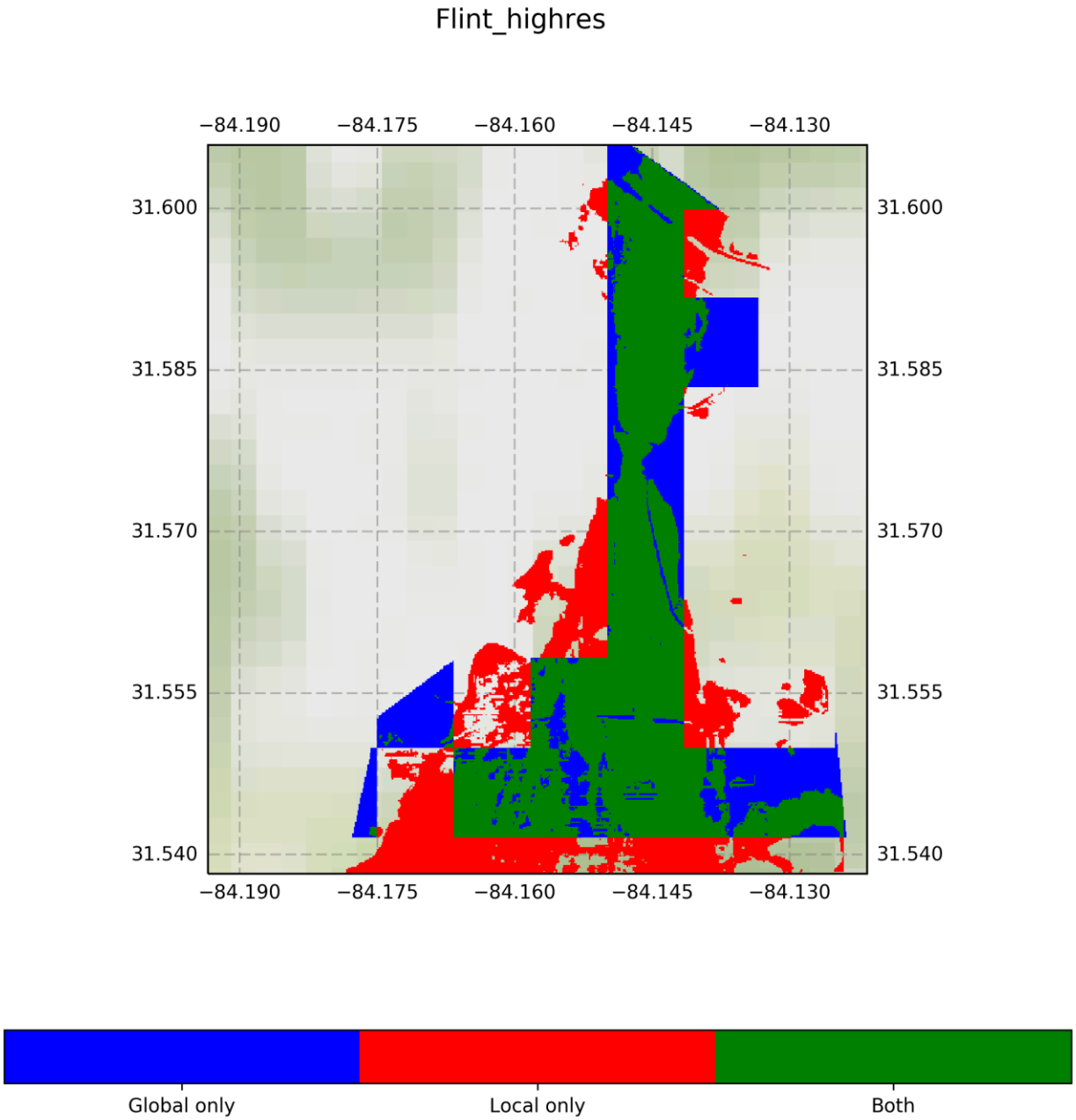
Supplementary Figure 26: Map showing the agreement between the GLOFRIS inundation extent and the Neue Luppe and Mulde Rivers (Saxony, Germany) benchmark data at 30" x 30". Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



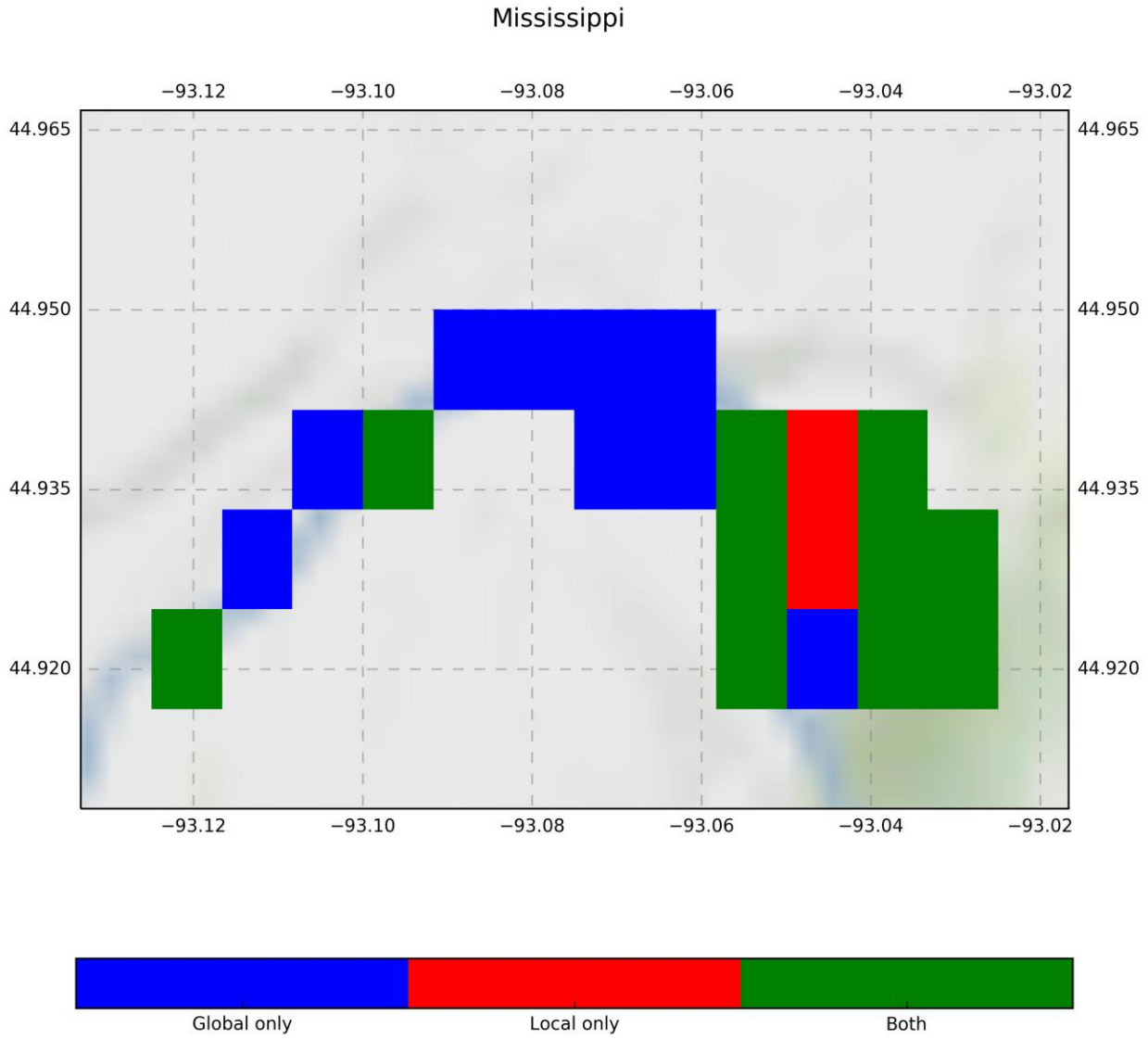
Supplementary Figure 27: Map showing the agreement between the GLOFRIS inundation extent and the Neue Luppe and Mulde Rivers (Saxony, Germany) benchmark data at the resolution of the original benchmark dataset". Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



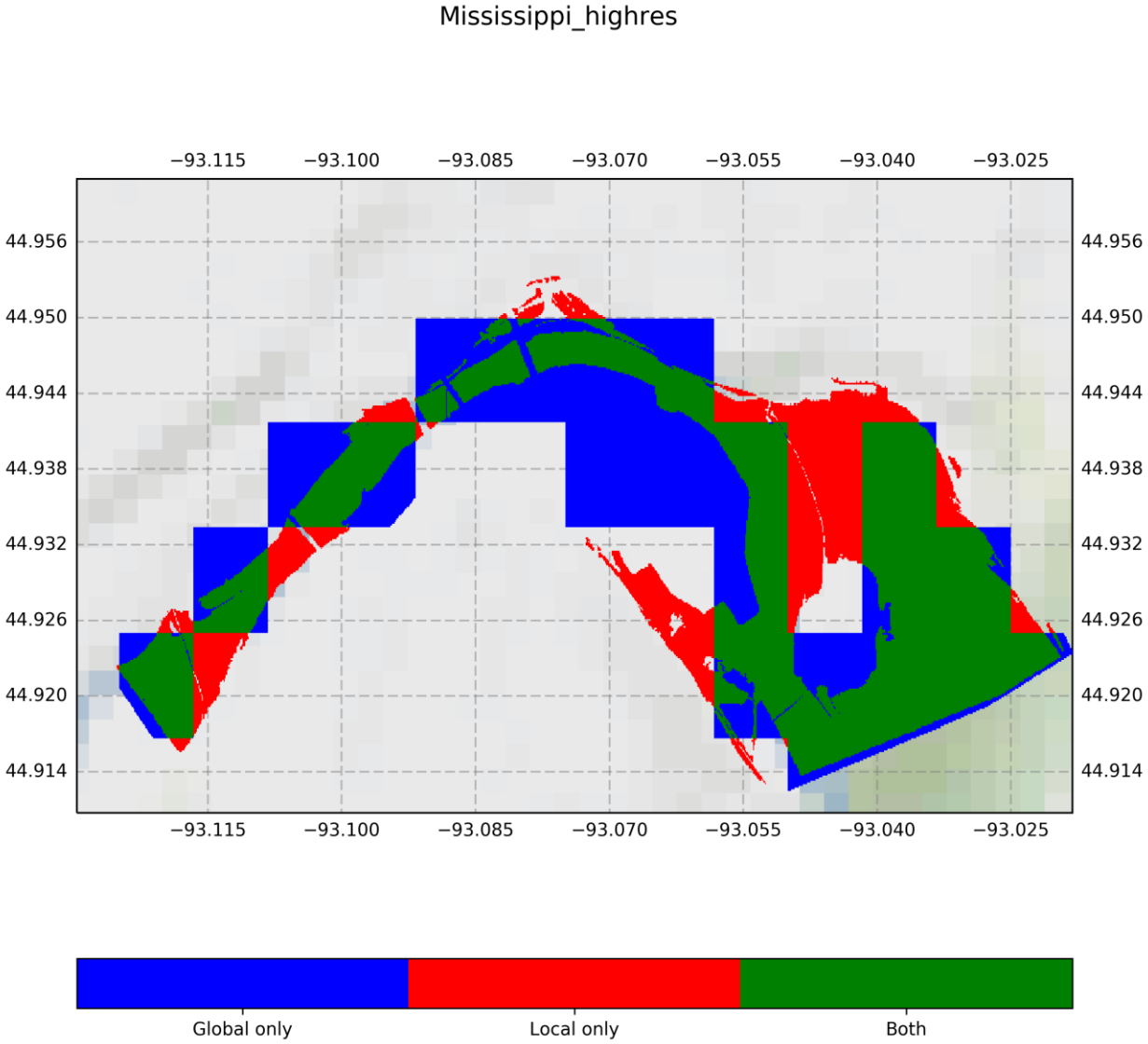
Supplementary Figure 28: Map showing the agreement between the GLOFRIS inundation extent and the Flint River (Georgia, USA) benchmark data at 30" x 30". Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



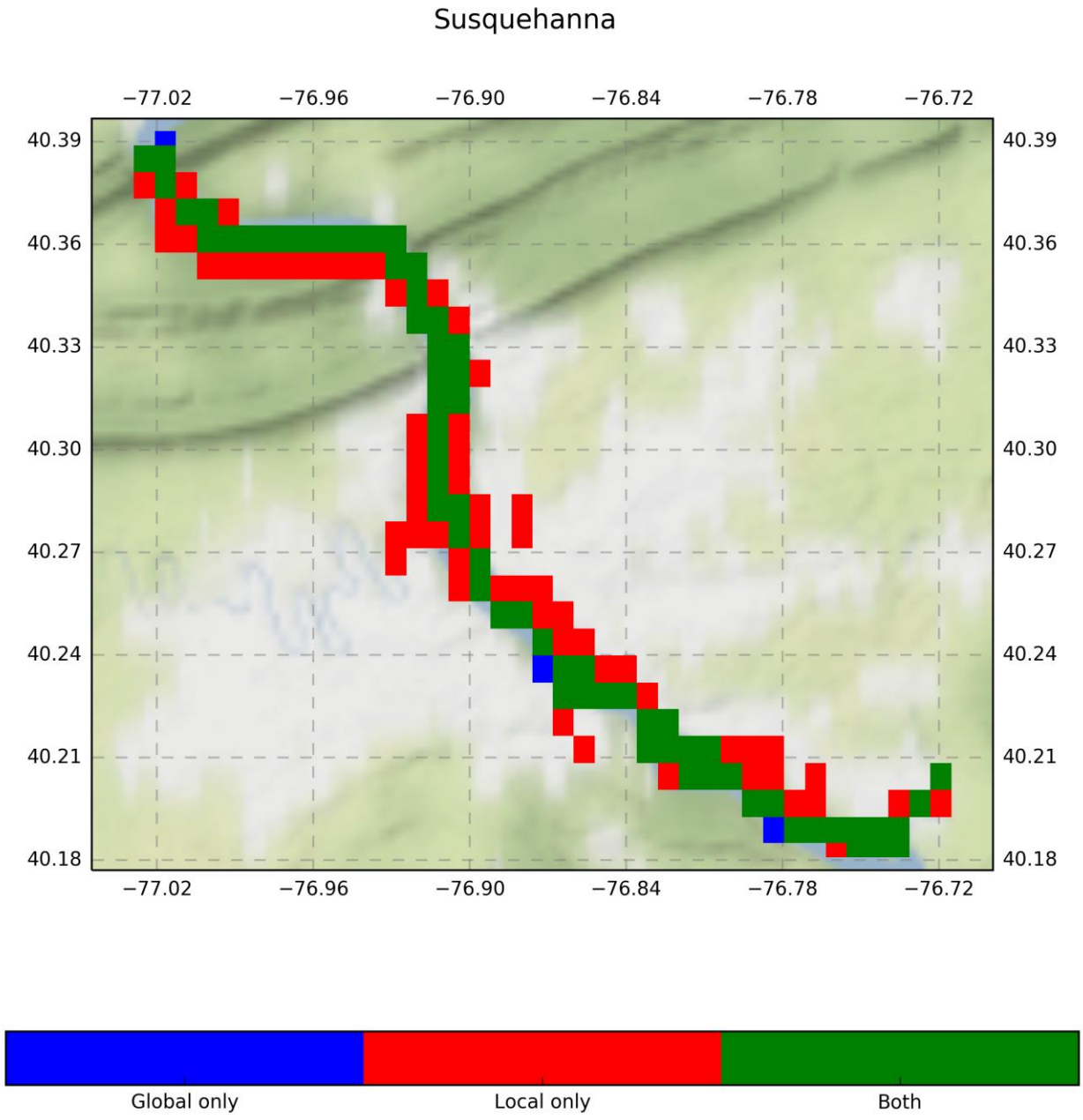
Supplementary Figure 29: Map showing the agreement between the GLOFRIS inundation extent and the Flint River (Georgia, USA) benchmark data at the resolution of the original benchmark dataset. Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



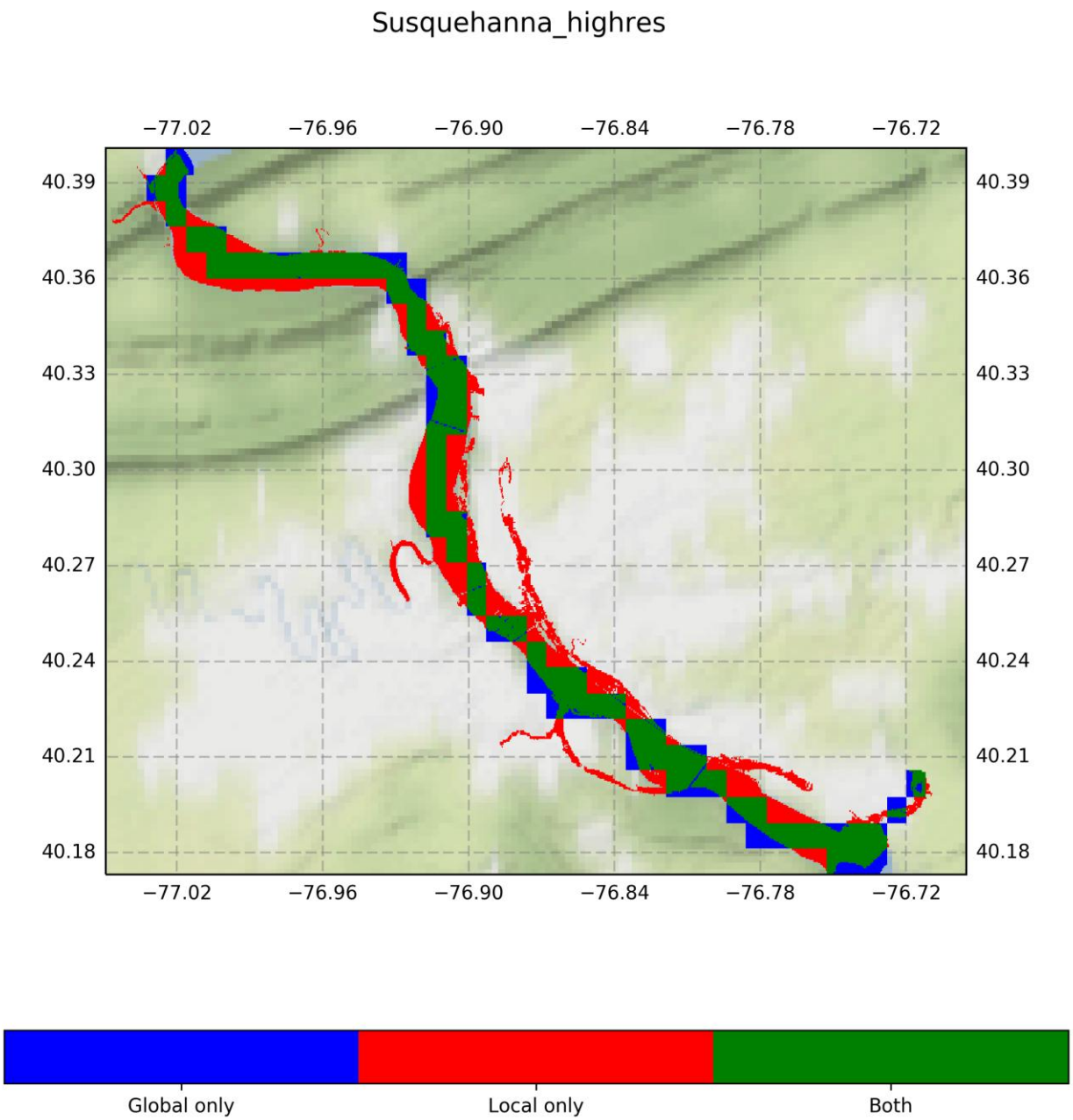
Supplementary Figure 30: Map showing the agreement between the GLOFRIS inundation extent and the Mississippi River (Minnesota, USA) benchmark data at 30" x 30". Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



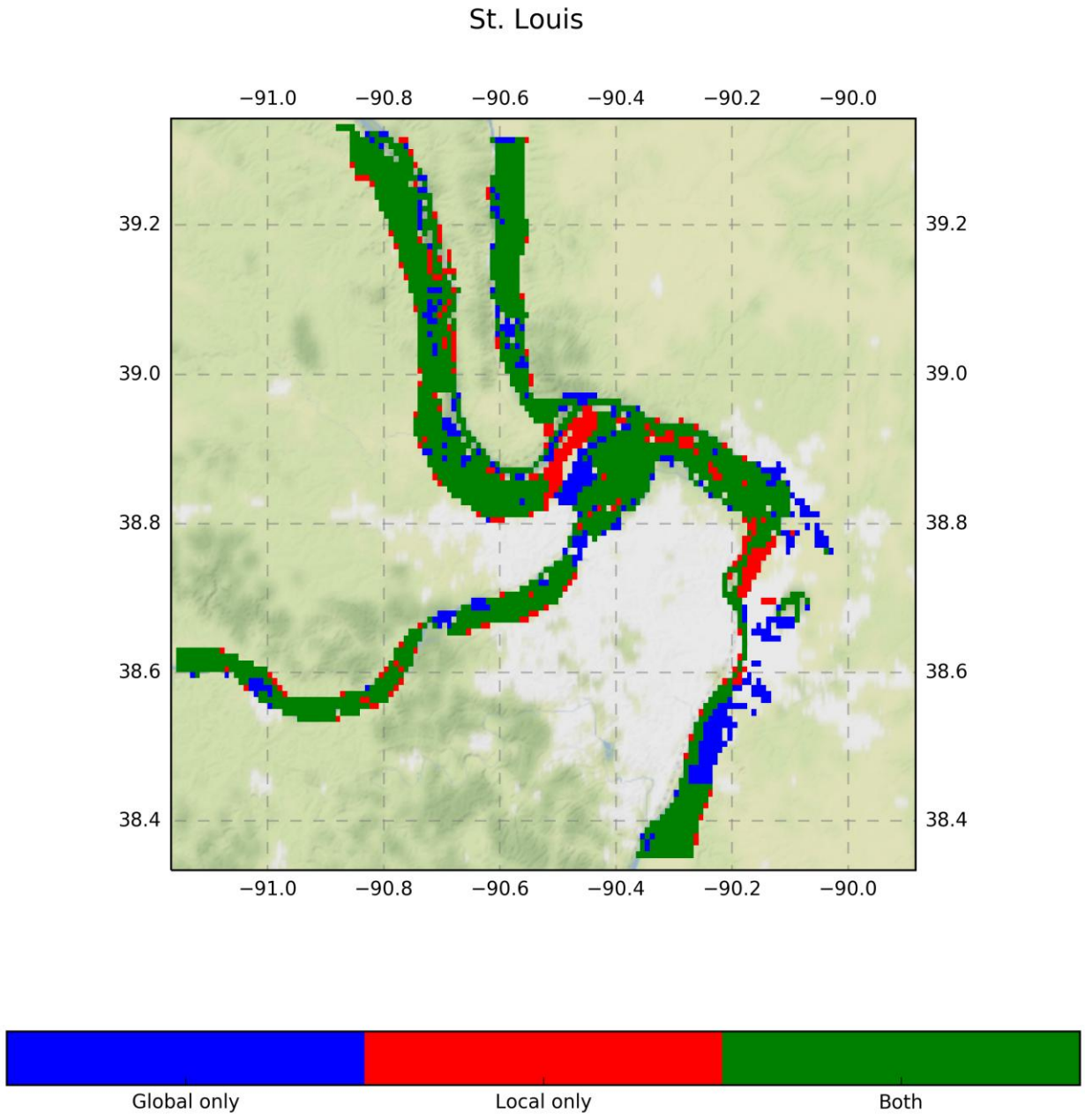
Supplementary Figure 31: Map showing the agreement between the GLOFRIS inundation extent and the Mississippi River (Minnesota, USA) benchmark data at the resolution of the original benchmark dataset. Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



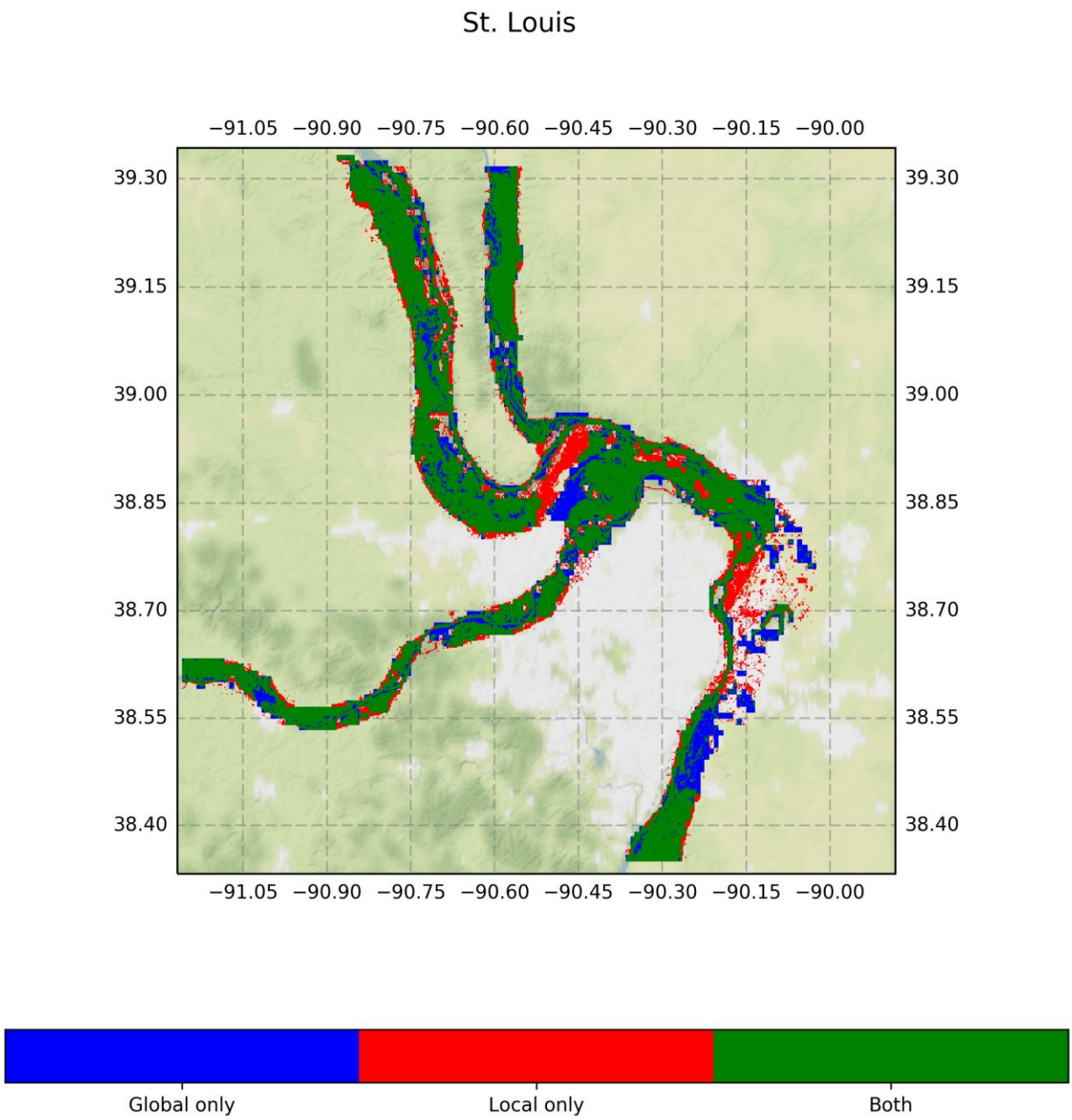
Supplementary Figure 32: Map showing the agreement between the GLOFRIS inundation extent and the Susquehanna River (Pennsylvania, USA) benchmark data at 30" x 30". Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



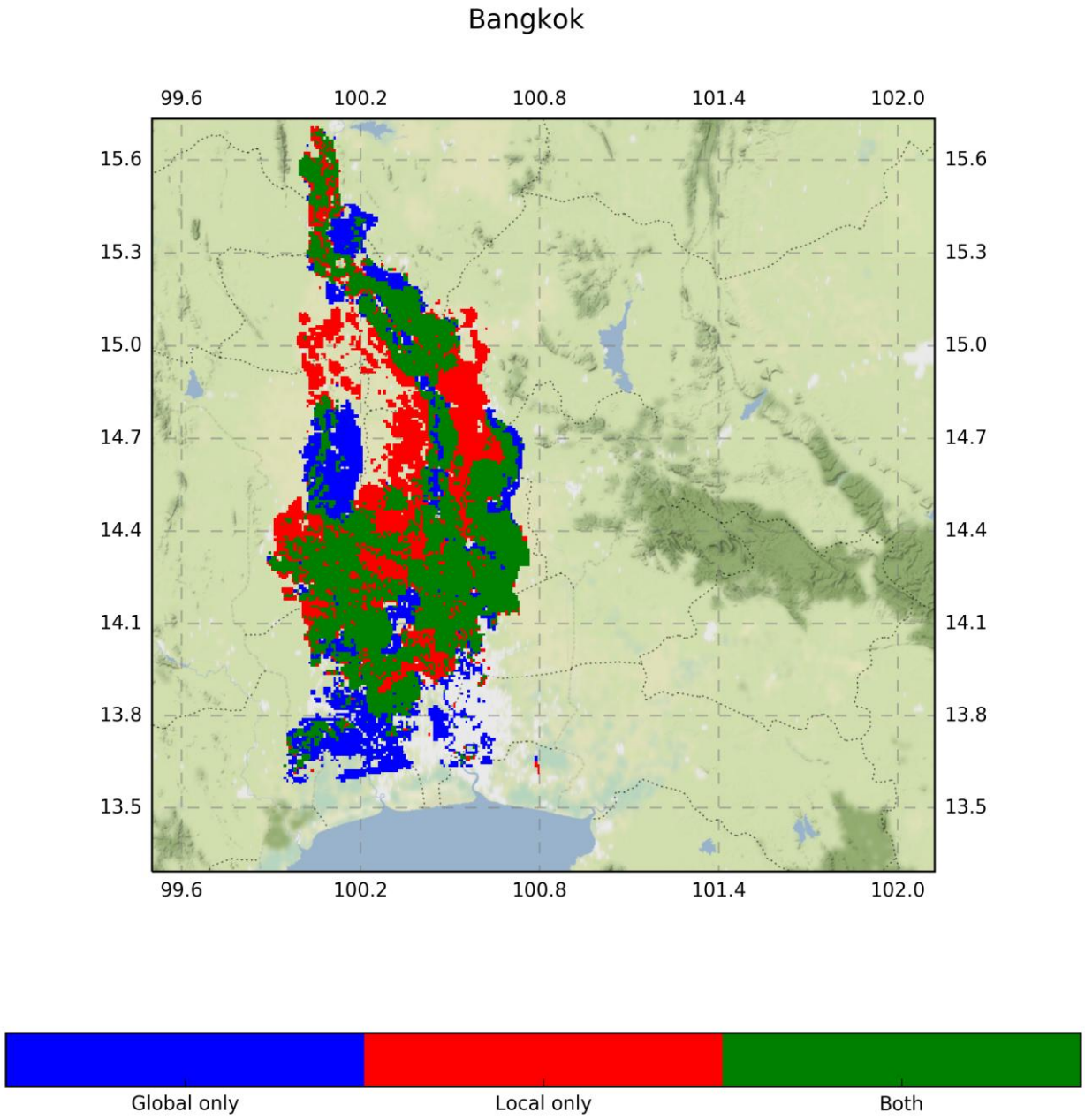
Supplementary Figure 33: Map showing the agreement between the GLOFRIS inundation extent and the Susquehanna River (Pennsylvania, USA) benchmark data at the resolution of the original benchmark dataset. Both maps are shown for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



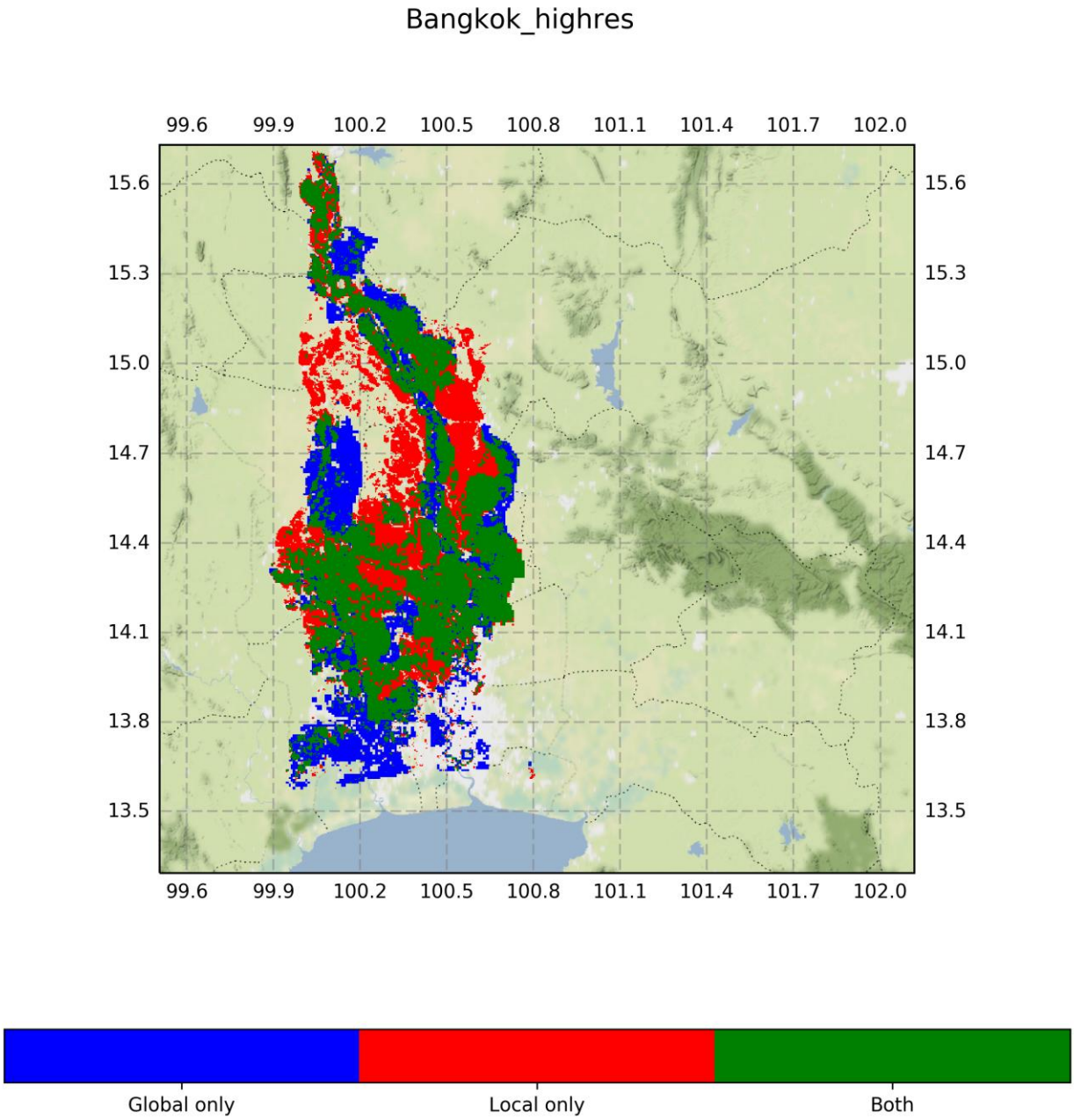
Supplementary Figure 34: Map showing the agreement between the GLOFRIS inundation extent and the St. Louis (USA) validation data at 30" x 30". The St. Louis data are for the flood event of 1993, and the GLOFRIS data are for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



Supplementary Figure 35: Map showing the agreement between the GLOFRIS inundation extent and the St. Louis (USA) validation data at the resolution of the original benchmark dataset. The St. Louis data are for the flood event of 1993, and the GLOFRIS data are for a return period of 100 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



Supplementary Figure 36: Map showing the agreement between the GLOFRIS inundation extent and the Chao Phraya River (Thailand) validation data at 30" x 30". The Chao Phraya data are for the flood event of 2011-2012, and the GLOFRIS data are for a return period of 25 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets



Supplementary Figure 37: Map showing the agreement between the GLOFRIS inundation extent and the Chao Phraya River (Thailand) validation data at the resolution of the original benchmark dataset. The Chao Phraya data are for the flood event of 2011-2012, and the GLOFRIS data are for a return period of 25 years. 'Global only' refers to cells that are only flooded in the GLOFRIS inundation map; 'Local only' refers to cells that are only flooded in the benchmark dataset; and 'Both' refers to cells that are flooded in both datasets

5. Supplementary Tables

Supplementary Table 1: Globally aggregated results for the 'optimise', 'constant absolute risk', and 'constant relative risk' adaptation objectives, for all combinations of RCP and SSP. The table shows the average results across the five Global Climate Models, under the following assumptions: middle-estimate investment costs; maintenance costs of 1% per year; and discount rate of 5% per year. We assumed that the construction of dikes begins in 2020 and is completed by 2050, and that by 2050 dikes are designed to the standard required for the climate at the end of the 21st century (2060-2099). Annual costs are based on the period 2020-2100.

	RCP2.6			RCP4.5			RCP6.0			RCP8.5		
Values in billion USD	B	C	B:C	B	C	B:C	B	C	B:C	B	C	B:C
Objective: optimise												
SSP1	316	47	6.7	381	53	7.3	420	56	7.5	529	65	8.2
SSP2	209	39	5.4	254	44	5.7	279	46	6.1	354	54	6.5
SSP3	78	22	3.6	95	26	3.6	105	27	3.9	136	34	4.1
SSP4	168	32	5.2	204	37	5.6	224	38	5.9	287	46	6.3
SSP5	484	58	8.3	579	64	9.0	635	67	9.4	799	78	10.2
Objective: Constant absolute risk												
SSP1	339	170	2.0	405	184	2.2	443	186	2.4	556	204	2.7
SSP2	231	164	1.4	276	177	1.6	303	180	1.7	381	199	1.9
SSP3	95	139	0.7	114	152	0.7	125	155	0.8	160	175	0.9
SSP4	185	148	1.3	223	161	1.4	243	164	1.5	309	183	1.7
SSP5	509	186	2.7	605	199	3.0	662	201	3.3	827	219	3.8
Objective: Constant relative risk												
SSP1	275	73	3.8	341	85	4.0	378	89	4.3	490	109	4.5
SSP2	181	68	2.7	225	80	2.8	251	84	3.0	329	104	3.2
SSP3	70	61	1.2	88	72	1.2	100	76	1.3	134	96	1.4
SSP4	149	73	2.1	187	84	2.2	207	88	2.3	272	108	2.5
SSP5	406	73	5.6	501	85	5.9	556	88	6.3	721	108	6.7

Supplementary Table 2: Globally aggregated results for the 'optimise', 'constant absolute risk', and 'constant relative risk' adaptation objectives, for the following scenarios: (a) RCP2.6/SSP1; (b) RCP4.5/SSP2; (c) RCP6.0/SSP3; and (d) RCP8.5/SSP5. The table shows the average results across the five Global Climate Models, under the following assumptions: middle-estimate investment costs; maintenance costs of 1% per year; and discount rates of 3%, 5%, and 8% per year. We assumed that the construction begins in 2020 and is completed by 2050, and that by 2050 dikes are designed to the standard required for the climate at the end of the 21st century (2060-2099). Annual costs are based on the period 2020-2100.

	Scenario			
	RCP2.6/SSP1	RCP4.5/SSP2	RCP6.0/SSP3	RCP8.5/SSP5
Objective: Optimise				
3% per year discount rate				
Benefits (USD billion per year)	596	479	205	1487
Costs (USD billion per year)	75	70	47	119
Benefit:Cost ratio	7.9	6.8	4.4	12.5
Net present value	521	409	158	1368
5% per year discount rate				
Benefits (USD billion per year)	316	254	105	799
Costs (USD billion per year)	47	44	27	78
Benefit:Cost ratio	6.7	5.7	3.9	10.2
Net present value	269	210	78	721
8% per year discount rate				
Benefits (USD billion per year)	152	120	48	390
Costs (USD billion per year)	27	25	14	47
Benefit:Cost ratio	5.6	4.9	3.4	8.4
Net present value	125	96	34	343
Objective: constant absolute risk				
3% per year discount rate				
Benefits (USD billion per year)	624	509	230	1524
Costs (USD billion per year)	231	241	211	297
Benefit:Cost ratio	2.7	2.1	1.1	5.1
Net present value	393	268	20	1227
5% per year discount rate				
Benefits (USD billion per year)	339	276	125	827
Costs (USD billion per year)	170	177	155	219
Benefit:Cost ratio	2.0	1.6	0.8	3.8
Net present value	169	99	-30	608
8% per year discount rate				
Benefits (USD billion per year)	168	137	62	411
Costs (USD billion per year)	120	125	110	155
Benefit:Cost ratio	1.4	1.1	0.6	2.7
Net present value	48	12	-48	257
Objective: constant relative risk				
3% per year discount rate				
Benefits (USD billion per year)	624	509	230	1524
Costs (USD billion per year)	231	241	211	297
Benefit:Cost ratio	2.7	2.1	1.1	5.0
Net present value	393	268	20	1227
5% per year discount rate				
Benefits (USD billion per year)	275	225	100	721
Costs (USD billion per year)	73	80	76	108
Benefit:Cost ratio	3.8	2.8	1.3	6.7
Net present value	202	145	24	613
8% per year discount rate				

Benefits (USD billion per year)	168	137	62	411
Costs (USD billion per year)	120	125	110	155
Benefit:Cost ratio	1.4	1.1	0.6	2.7
Net present value	48	12	-48	257

15

Supplementary Table 3: Globally aggregated results for the 'optimise', 'constant absolute risk', and 'constant relative risk' adaptation objectives, using baseline flood protection standards equal to half the protection standard stated in the FLOPROS database. The table shows the average results across the five Global Climate Models, under the following assumptions: middle-estimate investment costs; maintenance costs of 1% per year; and discount rate of 5% per year. We assumed that the construction of dikes begins in 2020 and is completed by 2050, and that by 2050 dikes are designed to the standard required for the climate at the end of the 21st century (2060-2099). Annual costs are based on the period 2020-2100

	Scenario			
Adaptation objectives	RCP2.6/SSP1	RCP4.5/SSP2	RCP6.0/SSP3	RCP8.5/SSP5
Objective: optimise				
Benefits (USD billion per year)	408	323	137	947
Costs (USD billion per year)	59	55	36	91
Benefit:Cost ratio	6.9	5.8	3.9	10.4
NPV (USD billion per year)	349	267	102	857
Objective: constant absolute risk				
Benefits (USD billion per year)	428	343	154	974
Costs (USD billion per year)	183	188	164	232
Benefit:Cost ratio	2.3	1.8	0.9	4.2
NPV (USD billion per year)	245	154	-10	742
Objective: constant relative risk				
Benefits (USD billion per year)	329	262	115	808
Costs (USD billion per year)	76	82	78	110
Benefit:Cost ratio	4.3	3.2	1.5	7.3
NPV (USD billion per year)	252	180	37	698

Supplementary Table 4: Globally aggregated results for the 'optimise', 'constant absolute risk', and 'constant relative risk' adaptation objectives, using baseline flood protection standards equal to double the protection standard stated in the FLOPROS database. The table shows the average results across the five Global Climate Models, under the following assumptions: middle-estimate investment costs; maintenance costs of 1% per year; and discount rate of 5% per year. We assumed that the construction of dikes begins in 2020 and is completed by 2050, and that by 2050 dikes are designed to the standard required for the climate at the end of the 21st century (2060-2099). Annual costs are based on the period 2020-2100

	Scenario			
Adaptation objectives	RCP2.6/SSP1	RCP4.5/SSP2	RCP6.0/SSP3	RCP8.5/SSP5
Objective: optimise				
Benefits (USD billion per year)	238	194	79	658
Costs (USD billion per year)	36	34	20	65
Benefit:Cost ratio	6.6	5.7	3.9	10.2
NPV (USD billion per year)	202	160	59	593
Objective: constant absolute risk				
Benefits (USD billion per year)	259	216	99	686
Costs (USD billion per year)	157	166	149	204
Benefit:Cost ratio	1.6	1.3	0.7	3.4
NPV (USD billion per year)	102	50	-49	482
Objective: constant relative risk				
Benefits (USD billion per year)	222	184	84	625
Costs (USD billion per year)	72	80	77	108
Benefit:Cost ratio	3.1	2.3	1.1	5.8
NPV (USD billion per year)	150	104	7	516

Supplementary Table 5: Model performance metrics for comparison of GLOFRIS and benchmark inundation maps. The model performance metrics shown are the hit rate (HR), false alarm ratio (FAR), and the critical success index (CSI). The metrics are calculated for the complete geographical domain for (complete area) and also only for urban areas within this geographical domain (urban area only).

	Complete area			Urban area only		
	HR	**FAR**	**CSI**	**HR**	**FAR**	**CSI**
Benchmarking at GLOFRIS resolution						
Thames (UK)	0.65	0.29	0.51	0.65	0.29	0.51
Severn (UK)	0.63	0.46	0.41	0.62	0.47	0.40
Saxony (Germany)	0.65	0.52	0.38	0.66	0.47	0.42
Flint (USA)		Urban only		0.67	0.44	0.43
Mississippi (USA)		Urban only		0.83	0.47	0.48
Susquehanna (USA)		Urban only		0.65	0.13	0.59
St Louis (USA)	0.87	0.17	0.73	0.86	0.23	0.68
Chao Phraya (Thailand)	0.65	0.29	0.52	0.62	0.42	0.43
Benchmarking at original resolution						
Thames (UK)	0.65	0.44	0.43	0.65	0.44	0.43
Severn (UK)	0.51	0.45	0.36	0.51	0.45	0.36
Saxony (Germany)	0.73	0.53	0.40	0.77	0.52	0.42
Flint (USA)		Urban only		0.64	0.31	0.50
Mississippi (USA)		Urban only		0.72	0.45	0.45
Susquehanna (USA)		Urban only		0.60	0.18	0.53
St Louis (USA)	0.80	0.19	0.67	0.77	0.24	0.62
Chao Phraya (Thailand)	0.65	0.31	0.51	0.62	0.45	0.41

Supplementary Table 6: Demonstration of the relative influence of differences in the GLOFRIS and benchmark inundation datasets on potential damage, compared to uncertainties associated with the use of FLOPROS, and different GCMs, RCPs, and SSPs. In the second column, we show the percentage difference in maximum potential damage per case study when calculated using the GLOFRIS inundation map compared to the benchmark inundation map. In the third and fourth columns, we show the percentage change in baseline EAD when using protection standards equal to half those stated in FLOPROS (FLOPROS-halved) and double those stated in FLOPROS (FLOPROS-doubled), compared to the baseline EAD using FLOPROS protection standards. In the final three columns, we show the standard deviation of the percentage change in EAD between baseline (current climate and socioeconomic conditions, FLOPROS protection standards) and 2080 for: 'All GCMs' (averaged across all RCPs, using current socioeconomic conditions); 'All RCPs' (averaged across all GCMs, using current socioeconomic conditions; and 'All SSPs' (using current climate conditions).

	% difference in maximum potential damage	% difference in baseline EAD compared to FLOPROS		St. dev. of % difference in EAD between baseline and 2080, across:		
	GLOFRIS v Benchmark	FLOPROS-halved	FLOPROS-doubled	All GCMs[a] (current exposure)	All RCPs[b] (current exposure)	All SSPs (current climate)
Thames (UK)	9	+71	-48	40	90	192
Severn (UK)	-12	+75	-46	144	208	202
Saxony (Germany)	71	+92	-40	101	60	106
Flint (USA)	-7	+75	-44	65	33	130
Mississippi (USA)	34	+78	-45	49	26	116
Susquehanna (USA)	-33	+74	-48	59	38	102
St Louis (USA)	11	+77	-43	44	61	119
Chao Phraya (Thailand)	36	+56	-47	173	75	806

[a]averaged across all RCPs
[b]averaged across all GCMs

6. Supplementary References

1. Van Vuuren, D.P. et al. The representative concentration pathways: an overview. Clim. Change 109, 5-31 (2011)
2. Van Vuuren, D.P. *et al*. A new scenario framework for Climate Change Research: scenario matrix architecture. *Clim. Change* **122**, 373-386 (2014)
3. Winsemius *et al*. Global drivers of future river flood risk. *Nature Clim. Change* **6**, 381-385 (2016)
4. Scussolini, P. *et al*., FLOPROS: an evolving global database of flood protection standards. *Nat. Hazard Earth Sys.* **16**, 1049-1061 (2016)
5. Sperna Weiland, F.C., Van Beek, L.P.H., Kwadijk, J.C.J. & Bierkens, M.F.P. Global patterns of change in discharge regimes for 2100. *Hydrol. Earth Syst. Sci* **16**, 1047-1062 (2012)
6. Alfieri, L., Feyen, L., Dottori, F. & Bianchi, A. Ensemble flood risk assessment in Europe under high end climate scenarios. *Global Environ. Chang* **35**, 199-212 (2015)
7. Hempel, S., Frieler, K., Warszawski, L., Schewe, J. & Piontek, F. A trend-preserving bias correction - the ISI-MIP approach. *Earth Syst. Dyn.* **4**, 219-236 (2013)
8. Winsemius H.C. *et al.* A framework for global river flood risk assessments. *Hydrol. Earth Syst. Sci.* **17**, 1871-1892 (2013)
9. Sampson, C.C., *et al.* A high-resolution global flood hazard model. *Water Resour. Res.* **51**, 7358-7381 (2015)
10. Alfieri, L*. et al.* Advances in pan-European flood hazard mapping. *Hydrol. Process. **28**, 4067-4077 (2014)*
11. Musser, J.W. & Dyar, T.R. *Two-dimensional flood-inundation model of the Flint River at Albany, Georgia: Atlanta* (U.S. Geological Survey Scientific Investigations Report 2007-5107, 2007)
12. Czuba, C.R., Fallon, J.D., Lewis, C.R., Cooper, D.F. *Development of flood-inundation maps for the Mississippi River in Saint Paul, Minnesota* (U.S. Geological Survey Scientific Investigations Report 2014-5079, 2014)
13. Roland, M.A. *et al*. *Flood-inundation maps for the Susquehanna River near Harrisburg, Pennsylvania* (U.S. Geological Survey Scientific Investigations Report 2014-5046, 2013)
14. Southard, R.E., Smith, B.J. *Flood of 1993 – Mississippi River near the Jefferson National Expansion Memorial (Arch), St. Louis, Missouri* (U.S. Department of the Interior & U.S. Geological Survey Fact Sheet FS-188-98)
15. Komori, D. *et al*., Characteristics of the 2011 Chao Phraya River flood in Central Thailand. *Hydrological Research Letters* **6**, 41-46 (2012)
16. Trigg, M.A., Michaelides, K., Neal, J.C., Bates, P.D. Surface water connectivity dynamics of a large scale extreme event. *J. Hydrol.* **505**, 138-149 (2013)
17. Brakenridge, G.R., Kettner, A.J., Slayback, D., Policelli, F. *Maps 100E020N and 090E020N. The Surface Water Record*. (Dartmouth Flood Observatory, http://floodobservatory.colorado.edu/index.html (2012)
18. Gale, E.L., Saunders, M.A. The 2011 Thailand flood: climate causes and return periods. *Weather* **68**, 233-237 (2013)
19. Yamazaki, D., Kanae, S., Kim, H., Oki, T. A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.* **47**, W04501 (2011)

20. Muis, S., Verlaan, M., Winsemius, H.C., Aerts, J.C.J.H., Ward, P.J. A global reanalysis of storm surges and extreme sea levels, *Nat. Commun.* **7**, 11969 (2016)
21. Marks, D. The urban political ecology of the 2011 floods in Bangkok: the creation of uneven vulnerabilities. *Pacific Affairs* **88**, 623-651 (2015)