[Slide 2]
Hello and welcome to the second presentation on flooding being given as part of the METEOR project. My name is Chris Sampson. I am a flood scientist and I led much of the flood modelling that has been done as part of this project. The first presentation that we gave introduced flood modelling and some of the core concepts; this presentation focuses on the national scale models that were produced for this project, and how the data produced by those models might be appropriately used.

[Slide 3]
During this presentation we are going talk about some of the key national data sets that went into the model, the methods used to employ the data and we are going to look at how the data might be applied.

[Slide 4]
I'd like to make a couple of key observations about national scale models. The first is that these models require a consistent methodology because many end users of national scale models need to be able to compare between locations and that is very difficult to do if you produce different models using different types of data in different locations. The second is that in the territories that the METEOR project is operating in, local observational data is usually not available. This means that the models have to operate using open access nationally available data. In many cases that means globally available datasets, and this does have implications for the accuracy and precision of the model because they are limited by the quality of those data.

[Slide 5]
Now perhaps the most fundamental data model is the terrain data – it is absolutely critical. Because of the limited availability of local scale terrain datasets, we had to use remotely sensed global data. This type of data is collected by satellite, but it is not collected by satellite specifically for the purposes of flood modelling and this means that it requires a fairly comprehensive amount of processing in order to make it usable within a flood model.

[Slide 6]
The data that we used for the METEOR project was the MERIT digital elevation model. This is a global terrain dataset that we helped to develop with colleagues in Japan that has a resolution of 3 arcseconds (which is approximately 90 m) and is based on US NASA SRTM Shuttle Radar Topography Mission data set. It has had a lot of processing done to it in order to remove many of the biases and noise that exist within SRTM. The result of all of this work is that the MERIT terrain dataset has significantly improved elevation accuracy in flat regions, and of course when we're worried about flooding it is the flatter part of the landscape that are at risk and therefore these improvements are felt very keenly in this application.

[Slide 7]
This slide here shows you quite how significant these improvements can be. This is the Mekong Delta that we looking at here, and on the left-hand side you can see the raw uncorrected SRTM data, and on the right you can see the processed MERIT

data. You can see very clearly that a load of striping error that is present in the raw data has been removed. If you run a flood model on a terrain dataset that has stripes like you can see in the left-hand panel then you get some very strange results. This indicates the importance of performing this sort of analysis on the data that goes into the model.

[Slide 8]
Now the next key dataset that goes into a flood model is the hydrography dataset. Hydrography just refers to the river network and associated characteristics such as river width and river upstream area. The basic hydrography network is derived from the elevation model and conditioned with other open source data where available.

[Slide 9]
The hydrography dataset that we use in this project is MERIT Hydro, so it is related to the elevation model that we just talked about but it was then fused with a number of open source datasets including OpenStreetMap data and Landsat imagery to better constrain the location of rivers throughout the landscape. MERIT Hydro today represents the best open-source global hydrography dataset that is available, and it is used widely by the scientific community.

[Slide 10]
Now when you are building flood models at national scales it is very important to automate the procedure. Traditional small-scale flood models were constructed manually. That means that the input data was collected manually and surveyed manually; it was then entered into the computer software manually, the model was calibrated manually, executed manually and the resulting output data was processed manually. Now whilst that is a perfectly reasonable approach for a local reach-scale engineering project, it is of course not feasible to do at the national scale. Therefore, national models require an automated framework that brings in all of the various input data and model components and constructs and executes the models on its own. The full details of our framework are laid out in the accompanying scientific literature, but a schematic for the model is on the slide here. You can see that conceptually it's quite simple. It takes the inputs in terms of terrain data and river topography. It also takes inputs in terms of river flows and it uses these to build the input conditions for the model. The model is then executed and the flood layers are produced.

[Slide 11]
So with that said, I want to give a very quick run through of some of the key methods that the model builder uses. I can't go into the full detail on this presentation but hopefully this will give a sense of what the model is doing. In terms of the terrain data, the global dataset is processed into 1-degree tiles that then have a buffer added to each side to allow the resulting simulation output files to be blended back together to create a seamless national data layer.

[Slide 12]
A traditional reach scale model would use a nearby river gauge to estimate extreme flows for that area. Unfortunately for most of the terrain we were simulating as part

of this project there were no local river gauges. That means we had to use a regionalised method in order to estimate appropriate extreme flows to place into the model. The method is known as the Regional Flood Frequency Analysis method and it uses river discharge data directly from thousands of river gauges around the world. The method links the flow behaviour of each of these gauges to predictor variables such as climate zone, upstream area, rainfall and slope. It then performs a clustering analysis on these gauges and their predictor variables in order to generate a model that can be deployed anywhere on the planet.

[Slide 12; advance image]
Validation work performed on this method has shown that it has similar levels of skill at predicting extreme flows to ungauged hydrological models. This suggests that the approach is entirely appropriate for estimating extreme flows for areas where we don't have observations.

[Slide 13]
For the rainfall component of the model we use intensity duration frequency (or IDF) curves the define design rainfall events. We simulated the 1-hour, the 6-hour, and the 24-hour rainfall events for each return period. Local data was used to characterise these IDF curves where available, and where local data was not available a regionalised method similar to the flow discharge method was used to estimate appropriate rainfall intensity values for each event duration.

[Slide 14]
As we already mentioned MERIT Hydro was used to define the location of the rivers and a global river width database that uses Landsat imagery was used to estimate the width of rivers. However, the one thing that cannot be estimated from remote sensing is the depth of rivers, and that is the one other component that is essential for the flood model because in order to know how big it is you need to know approximately how wide it is and approximately how deep it is in order to get an idea of how much volume it can convey before it starts to go out of bank and cause flooding. Because you can’t observe this information it has to be modelled, and we use a one-dimensional channel solver in order to do this.

Now conceptually this model is quite simple as it takes the one-in-two year flow from the flood frequency analysis and assumes this to be bankfull. The solver then calculates the appropriate channel depth, given the width and given the slopes from the elevation model, to yield a water surface that is level with the elevation model at that point. By assuming that the elevation of the terrain model at a point is equal to the bank height of the river, we have a way of calibrating the river conveyance within the model to carry that 2-year flow before it starts to flood. The one-in-two year flow is approximately the bankfull return period of a river in its natural state.

[Slide 15]
The automated model builder then has to decompose the river network into segments or reaches and generate a hydrograph or an inflow boundary condition for each of these reaches for the model to then simulate. The regionalised flood frequency analysis gives us the discharge for each boundary condition, and we use

an estimate of time to peak based on the distance to the furthest point in the catchment in order to estimate how long a hydrograph should last. Rainfall events for each model domain are defined using the IDF curves that we just mentioned.

[Slide 16]
Now that the framework has generated all of the boundary conditions, it has to execute. It does this using a full 2D hydrodynamic model based on the simplified inertial variant of the shallow water equations. I talked about these in a little bit more detail in the previous presentation, but the important thing to note about them is that the physics is simplified as far as possible in order to maximise the speed of the simulation and minimise the amount of computational power that it requires while still preserving enough precision and accuracy to give a good result. The other thing that is worth noting about the implementation of the hydraulic model in these national scale models is that for the river system it uses a sub-grid method. This method allows the width of the rivers to be decoupled from the grid. When you use the data, you will know that the grid scale is about 90 m. This method allows us to simulate rivers that are much smaller than that because actually a 90 m wide river is really quite a large river and it is necessary to simulate rivers that are much smaller than that because they can actually contribute a lot of the flood risk, and this method allows us to do that.

[Slide 17]
The final stage of the process after the simulations have run is to pull together all of the different reach simulations within each tile and merge them, and then merge each tile to create a national scale layer for each return period. Now this may sound trivial, but there will in fact be many thousands of simulations performed and so again it is very important that this process is automated and efficient. The final step is to take these output data and transform them into a standard GIS format, for example GeoTIFFs.

[Slide 18]
On the screen here is just an example of what some of the flood output might look like.

[Slide 19]
Given the context in which these models have been constructed, and the data limitations that they are faced with, it is very important to consider where they can and cannot be suitably applied.

[Slide 20]
On the table here you can see potential applications for this data for a number of different stakeholder types, be it the public, local government, insurance or commercial. Now you will see from this table that the information produced by these models is very useful for things like flood awareness programs, flood hazard communication and identifying areas and regions that are likely to be at risk. However, it has limitations when you start to reduce the scale at which you are trying to apply the information and it is clearly not appropriate to apply this information for individual property level decisions.

[Slide 21]
Another way to frame this is to consider application scale. If we start at the national scale, the data is clearly appropriate and suitable to generate a national scale hazard assessment. If you're looking to determine which catchments are risky, or the relative level of risk between catchments, the data should also be able to provide a good assessment of this in all but the very smallest of catchments. When we start to look at individual settlements the picture becomes a bit more mixed. For larger settlements such as towns and cities the data should be fine to assess whether or not that town or city is exposed to significant flood hazard. However, as we move to smaller settlements and small villages we may be reaching the point at which the scale is becoming too small for the model to have adequate skill. Once we get down to the individual building level, unfortunately the data is not appropriate to assess flood hazard for individual buildings, and that is fundamentally because the underlying global terrain and hydrography datasets are not precise enough for this kind of application.

[Slide 22]
If we move onto infrastructure a similar logic applies. The data can be used to assist in planning decisions and it can be used to determine individual locations and individual pieces of infrastructure that might require a more detailed engineering investigation. However, the data cannot be used to make design decisions and it cannot be used to define the specifications of individual pieces of infrastructure. So therefore there is a general picture that the data can be used to infer areas and regions, towns and settlements, that may be at greater risk and require more detailed local information, but it can't be used to make detailed design or engineering decisions at the individual building level.

[Slide 23]
The final side here is a list of key academic references for more detail on the methods and the data that are used in this model. These papers are available at the DOI links provided for each. Thank you.