06 – Development of an automated tool to undertake a national flood risk assessment

Patrick Rath[1], Richard Bingham[1], Sandie Mann[2], Ruth Ellis[2]
[1] *RPS Consulting Engineers, Elmwood House, 74 Boucher Road, Belfast, BT12 6RZ*
[2] *Scottish Environmental Protection Agency (SEPA), Silvan House, SEPA 3rd Floor, 231 Corstorphine Rd, Edinburgh EH12 7AT*

Abstract
The Flood Risk Management (Scotland) Act 2009 (FRM Act) requires the Scottish Environmental Protection Agency (SEPA) to prepare a flood risk assessment for each flood risk management district. In December 2011 SEPA produced a National Flood Risk Assessment (NFRA) for Scotland which provided a high level screening tool which assessed the relative adverse consequences of flooding across the country. Section 10 of the FRM Act requires SEPA to review and, where appropriate, update the first NFRA published in December 2011 by December 2018.

The purpose of the NFRA is to identify locations where a significant flood risk exists, or is likely to occur in the future, by assessing the potential adverse consequences of any future flood, and the impacts of significant historic flood events, to human health, the environment, cultural heritage and economic activity. The overall purpose of the 2018 NFRA (NFRA2) and subsequent cycles is to continually improve the understanding of flood hazard and risk across Scotland and prime further development in flood maps and strategies. A specific focus of the 2018 update has been embedding climate change within the assessment, and enhancing the consideration of community impacts to support the delivery of future focused, socially just flood risk management in Scotland.

A GIS analysis tool was developed which enables a more automated and consistent approach to undertake the NFRA2. The algorithms (written in python) follow SEPAs national flood risk assessment methodologies and are capable of assessing all receptors and flood hazard data on a national scale. The receptor data input includes Residential Properties (2.5 million), Non-Residential Properties (0.5 million), Community Facilities (40,000), Community Utilities (8,000), Roads (0.8 million road segments), Agricultural Land (1.1 million land parcels) as well as national environmental sites, cultural heritage sites, airport runways, agricultural buildings and intensive farming installations.

The flood hazard data used in the analysis comprised:

1. Fluvial depth, velocity and extents from 47 hydrometric areas (varying resolutions).
2. Pluvial depth, velocity and extents from 78 regional models (varying resolutions).
3. National pluvial flood extents.
4. National coastal depths and extents.
5. Wave overtopping depths and extents from 28 coastal areas (varying resolutions).
6. Three Integrated Catchment Study (ICS) depth, velocity and extents

The volume and diversity of the data was a significant challenge and the production of the previous national flood risk assessment was extremely cumbersome and difficult to maintain national consistency. The goal was to nationalise the flood hazard data such that any scale of assessment is possible. A conceptual model workflow was designed which formed the overarching process to achieve a national scale output of flood risk across Scotland. All tools were written in Python and the modular

nature of the tool enables assessment in part or entirety. The three main modules to the NFRA2 model are:

1. Pre-processing hazard and receptor data such that the receptor and flood hazard data are trained and combined into a more nationally consistent data structure.
2. Receptor Analysis, where receptors are assessed for flood risk and damages are calculated.
3. Grid Analysis, where the receptor flood data is transposed to 500x500m grids to help visualise and analyse national flood risk.

Tool parameters (MCM codes (Penning-Rowsell, *et al*., 2015)) can be updated and various thresholds for defining risk may be tuned/refined. The tool can also be adapted and distributed to local authorities who can undertake flood risk assessments using the most recent SEPA methodologies, whilst reducing the risk of human error and maintaining national consistency. This recent development harnesses the flexibility and control that can be achieved through the use of GIS and scripting, and opens many more opportunities for implementing similar procedures in the industry.

1. INTRODUCTION

RPS was commissioned by the Scottish Environmental Protection Agency (SEPA) to develop an automated GIS analysis tool to produce the 2018 National Flood Risk Assessment (NFRA2). The Flood Risk Management (Scotland) Act 2009 (FRM Act) requires SEPA to prepare a flood risk assessment for each flood risk management district. In December 2011 SEPA produced a National Flood Risk Assessment (NFRA) for Scotland which provided a high level screening tool which assessed the relative adverse consequences of flooding across the country. The purpose of the NFRA is to identify locations where a significant flood risk exists, or is likely to occur in the future, by assessing the potential adverse consequences of any future flood, and the impacts of significant historic flood events, to human health, the environment, cultural heritage and economic activity. The overall purpose of the NFRA2 (and subsequent cycles) is to continually improve the understanding of flood hazard and risk across Scotland and prime further development in flood maps and strategies. A storyboard of the Scottish NFRA can be access via the following link:
https://sepaweb.maps.arcgis.com/apps/Cascade/index.html?appid=323aefe6abcf4f859acabca202c30f9b

1.1 NFRA2 Overview

The NFRA2 assesses the impact of flooding on human health, economic activity, the environment and cultural heritage. The methodology - written by SEPA - focuses on assessing potential adverse consequences on the mandatory set of receptors. The receptor groups are as follows:

- Residential properties
- Non-residential properties
- Community facilities
- Community utilities
- Cultural heritage
- Environment
- Transport (roads, rail and airport runways)
- Agriculture

The NFRA2 methodology provides guidance on the calculation of damages associated with the various receptors and how raw receptor datasets should be prepared for the assessment. Where receptors do not inherit a damage figure (such as Environment, Utilities and Cultural Heritage), their impact is captured by assigning a direct risk score. The flood data captured in the receptor analysis is transposed to a 500m square grid to summarise flood damages and impact. The 2011 NFRA used a 1km square grid, but the 2018 grid size allows a more target assessment of flood risk whilst still being a high level, strategic, national assessment. The gridded outputs are then subject to a flood risk scoring system which also takes into account a number of influencing factors. Final NFRA outputs comprise national flood risk grids for all flood sources and receptors separately and combined, with an additional analysis that assesses the impact of climate change. The aim of the project was to streamline the assessment through automation whilst minimising user intervention and input.

1.2 SEPA Community Engagement

The 2018 National Flood Risk Assessment will be published publically by 22nd December 2018 and is the product which guides the identification of Potentially Vulnerable Areas (PVAs). PVAs are the primary unit for managing strategic flood risk management activities. Throughout the process of developing the NFRA, SEPA has engaged stakeholders and members of the public. SEPA engaged Local Authorities heavily in the identification of PVAs as the Local Authorities could bring extra local knowledge within those areas. Local Authorities also have to use PVAs in their Local Flood Risk Management Plans.

SEPA ran a national public consultation on the PVAs between the 1st May to the 31st July 2018. This allowed stakeholders, members of the public or anyone who was interested to comment on the NFRA methodology as well as the proposed PVA. There was also the option for consultation responders to provide additional supporting information such as photographs and flood repots. The public consultation was advertised in many ways to ensure people were aware of it and had the chance to respond. This included providing Local Authorities with briefing and media material which they could use to inform the public in their area, local press and radio advertising and advertising through social media, SEPA website and SEPA newsletters. SEPA also presented the NFRA method at a flood community engagement event where members of the public and local flooding community groups could find out more about the NFRA. In total SEPA received 263 responses to the consultation which is a 417% increase in responses since the last PVA consultation in 2011. There was general agreement with the PVAs proposed however due to the views received, SEPA made some minor amendments. A revised set of PVAs was then submitted to Scottish Ministers for approval. Once approved, the final set of PVAs will be published alongside the NFRA by the 22nd December 2018.

2. IMPLEMENTING AUTOMATED PROCEDURES

Undertaking a national flood risk assessment is a highly intensive and dynamic process. For example, there are many variables involved in the calculation of property damages related to flooding including property type, depth of flood water and floor level. The repetitive tasks associated with conducting a national flood risk assessment are laborious and extremely vulnerable to human induced errors and inconsistencies, particularly when the tasks are carried out by more than one operator. Designing a set of tools that implements the NFRA2 methodology offers a cost-efficient method of assessment with higher quality outputs. Consequently, SEPA's NFRA2 methodology has been implemented in an automated fashion using ESRI's Geographical Information System software (ArcMap).

Many of the processes involved in the national flood risk assessment naturally lend themselves to automation (e.g. extraction of mean flood depths over numerous properties, flood types and scenarios). A key part of optimisation and allowing for greater levels of automation is the development of robust, standardised input data architecture. Standardising datasets, using unique logical naming conventions, and complying with a robust spatial data infrastructure offers greater opportunities to design efficient algorithms that are much less susceptible to unforeseen errors and exceptions. Making use of in-memory computation increases speed but may exhaust hardware resources, so the consideration of the available hardware specification is strongly recommended.

The design of an automated procedure usually begins with a conceptual design. The conceptual workflow design stage allows a holistic approach to understanding the connections and validates the opportunity for automation. ArcMap model builder was used at the outset to build conceptualised models and test the outputs, before moving to more advanced methods. Python scripting (Arcs default scripting language) allows much greater control over the assessment algorithms for each receptor. Using scripting techniques also offers greater flexibility over input parameters that affect monetary damage calculations. The approach has been designed by the project team to account for MCM refinements and updates, changes and upgrades to the receptor data and changes to the best available flood hazard information for coastal, fluvial and pluvial sources.

This paper provides a summary of the automated workflow for the NFRA2, introduces how each module works, details the purpose of each model and the interconnectivity of each tool within the overarching conceptual workflow. Finally, there are a number of recommendations for future improvement of the models and for future working by SEPA.

3. NFRA2 SCRIPTING METHODOLOGY

The NFRA2 toolbox provided as the final deliverable implements the NFRA2 methodology in an automated process. The process comprises three main stages/modules conceptualised in Figure 1.

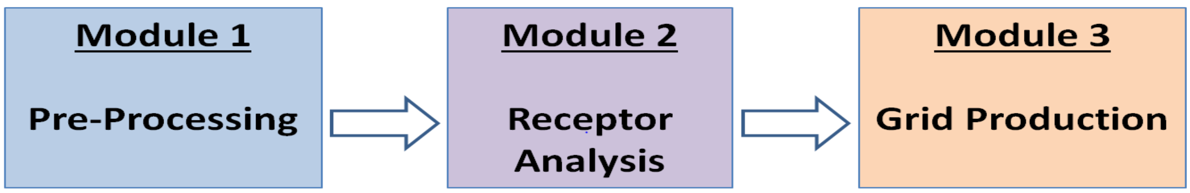


***Figure 1**: Simplified NFRA2 Workflow*

Note: A more detailed workflow for the NFRA2 developed tool is provided at the end of this paper.

3.1 Module 1 – Pre-procesing

In Module 1, all input receptor and flood hazard data required a level of 'training' if a national scale approach were to be viable. It was therefore necessary to design and execute pre-processing algorithms for each receptor type. For example, in many of the receptor datasets there were fields/attributes that were not required for the analysis. Consequently, the team agreed the receptor attributes that should be maintained in the outputs for future interrogation. All pre-processed receptor feature classes are suffixed with "_PP_YYYYMMDD", indicating it is eligible for insertion to the next module. A list of the receptors included in the NFRA2 along with naming conventions and national feature counts are provided in **Table 1**.

***Table 1:** NFRA2 Pre-Processed Receptors*

Receptor Feature Class	Shape Type	Feature count
Agri_Intensive_Farming_PPC_PartA_6_9_PP_YYYYMMDD	Polygon	99
Agricultural_Land_PP_YYYYMMDD	Polygon	1,057,965
Communities_Geographic_PP_YYYYMMDD	Polygon	5,248
Community_Facilities_PP_YYYYMMDD	Polygon	38,486
Community_Facilities_Point_PP_YYYYMMDD	Point	38,486
Cultural_Heritage_PP_YYYYMMDD	Polygon	76,324
Environment_PP_YYYYMMDD	Polygon	2,061
NFRA2Grid_PP_YYYYMMDD	Polygon	500,880
PDS_2017_NRP_Point_PP_YYYYMMDD	Point	488,740
PDS_2017_NRP_Polygon_PP_YYYYMMDD	Polygon	488,740
PDS_2017_RP_Point_PP_YYYYMMDD	Point	2,582,346
PDS_2017_RP_Polygon_PP_YYYYMMDD	Polygon	2,582,346
Rail_PP_YYYYMMDD	Polyline	6,985
Roads_PP_YYYYMMDD	Polyline	840,563
Transport_Airports_PP_YYYYMMDD	Polygon	40
UtilitiesDataset_PP_YYYYMMDD	Polygon	7,837
UtilitiesDataset_Point_PP_YYYYMMDD	Point	7,837

The flood hazard data also had its own complexities mostly in relation to folder structure, naming conventions, different model types and varying cell resolutions. Discussions with SEPA resulted in an agreed 5m grid cell resolution for all flood data and allowed national scale flood hazard mapping to be produced in a single raster dataset for each source and return period. The purpose of merging all hazard data into a national dataset is to facilitate the full national scale flood risk assessment and take account of multiple sources of flood data simultaneously. As the flood hazard mapping from all sources was only available in discrete parcels, an optimized raster merge procedure was undertaken to minimise any geographical shift between hazard rasters. The output national scale flood hazard mapping ensures a consistent and logical naming convention which is then interpreted by the receptor models (Table 2).

***Table 2:** NFRA2 Pre-Processed Flood Hazard Mapping*

Depth Rasters	Velocity Rasters	Flood Extents
SOURCE_DEPTH_10_D	SOURCE_VELOCITY_10_D	SOURCE_EXTENT_10_D
SOURCE_DEPTH_30_D	SOURCE_VELOCITY_30_D	SOURCE_EXTENT_30_D
SOURCE_DEPTH_50_D	SOURCE_VELOCITY_50_D	SOURCE_EXTENT_50_D
SOURCE_DEPTH_100_D	SOURCE_VELOCITY_100_D	SOURCE_EXTENT_100_D
SOURCE_DEPTH_200_D	SOURCE_VELOCITY_200_D	SOURCE_EXTENT_200_D
SOURCE_DEPTH_1000_D	SOURCE_VELOCITY_1000_D	SOURCE_EXTENT_1000_D
SOURCE_DEPTH_30_D_2080	SOURCE_VELOCITY_30_D_2080	SOURCE_EXTENT_30_D_2080
SOURCE_DEPTH_200_D_2080	SOURCE_VELOCITY_200_D_2080	SOURCE_EXTENT_200_D_2080

3.2 Module 2 – Receptor Analysis

In Module 2, receptors are assessed against the flood hazard data. Each receptor analysis shares some similarities, but mostly each is bespoke and unique to the receptor. The input spatial datasets are populated with the flood hazard data (mean flood depth) which leads to the associated damage calculations or risk score assignment (depending on receptor group). The receptor models are modular, such that the basic model of any receptor will conduct a depth extraction assessment on one return period. A full receptor assessment repeats the analysis for each return period. Once all scenarios have been assessed, damage scripts are executed to calculate damages associated with the flooded depth/area/length. This modular approach, enables the assessment to be cut-off and re-run from any point. For example, a change to a damage parameter does not necessarily mean the entire assessment needs repeated. The depth extraction is a computationally intensive procedure and it was considered more efficient to allow the damages to be calculated separately.

There was a requirement to post process the community facilities and utilities due to the uncertainty of flood impact to these receptors. The process involves assessing if a facility/utility has been assigned a flood depth greater or equal to 0.2m and a flooded area greater or equal to 5%. Those that satisfy this criterion retain their flood depth information and those that do not are assigned a new identifier. An automated approach which involved clipping and calculating the percentage area flooded and applying an expression with the above criteria enabled this process to be carried out over all return periods in the receptor datasets.

As part of the NFRA2, there was a requirement to produce an output which has removed duplicate receptors. The following receptor summaries are captured in the output non-duplicate grids in later stages:

- Counts: Homes, Businesses, Airports, Community Facilities, Utilities Infrastructure
- Damage totals: Homes, Businesses, Agriculture, Roads
- Lengths: Roads, Rail (both flooded and impacted)
- Areas: Environment, Cultural Heritage

A standalone script was developed to undertake the removal of duplicate receptors in a fully automated process.

3.3 Module 3 – Grid Production

Module 3 transposes the data within the receptor datasets to the NFRA2 grid, calculating various statistics (depending on the receptor) and introduces scoring categories indicating level of risk. Risk scores are calculated using conditional statements based on the statistics extracted. The automated workflow includes additional functionality where these statements and conditions can be easily modified. Module three also includes tailored scripts for producing influenced score grids, PVA grids and cluster analyses.

The process of generating receptor risk grids is simplified in the step-by-step procedure below and is generally the same for all receptors:

1. Flood assessed receptor is queried for each scenario returning only those features that are impacted in the given return period.
2. The feature attributes are spatially joined to the NFRA2 grid with various cell statistics (count of properties, sum of damages etc.).
3. Calculate maximum risk score based on NFRA2 methodology criteria.
4. Climate change grid analysis.

Step 4 is an additional procedure required in the NFRA2 methodology where risk scores are assigned to each grid cell based on observed changes/differences between the non-climate change and climate change scenarios. The grid analysis and grid score criteria are stipulated in the NFRA2 methodology which is implemented in a similar way to the receptor risk scores. The script tool has the functionality of updating the thresholds and conditional statements provided they conform to python syntax.

3.4 NFRA2 OUTPUTS

The NFRA2 outputs comprise a number of grids with flood data, receptor statistics and risk scores that have been calculated, summarise and inextricably linked to the receptor assessment data. The main grid outputs are:

- Receptor Grids
- Source Score Grids
- Influenced Source Score Grids
- PVA Grids (Current and Future)
- Clustering

Receptor grids contain key flood statistics that filter into source score grids and source score grids which are the sum of all receptors and calculated risk scores for each return period. A source score grid is produced for each source of flooding. The scores are then influenced by a number of other factors that could exacerbate the impact of flooding within a given grid cell. Additionally, part of the NFRA2 process involves producing two final PVA risk grids (one based on current risk and another based on future climate change scenarios) to aid in identifying areas of significant flood risk. The challenge here was to provide user flexibility in choice of receptor and scenario whilst accounting for all sources simultaneously. Therefore, the tool interface allows the user to make these selections and produce PVA grids specific to a given input combinations of flood source, return period and receptor. **Figure 2** provides an illustration of the influenced source score grids along with the combined PVA score which is the PVA grid output.

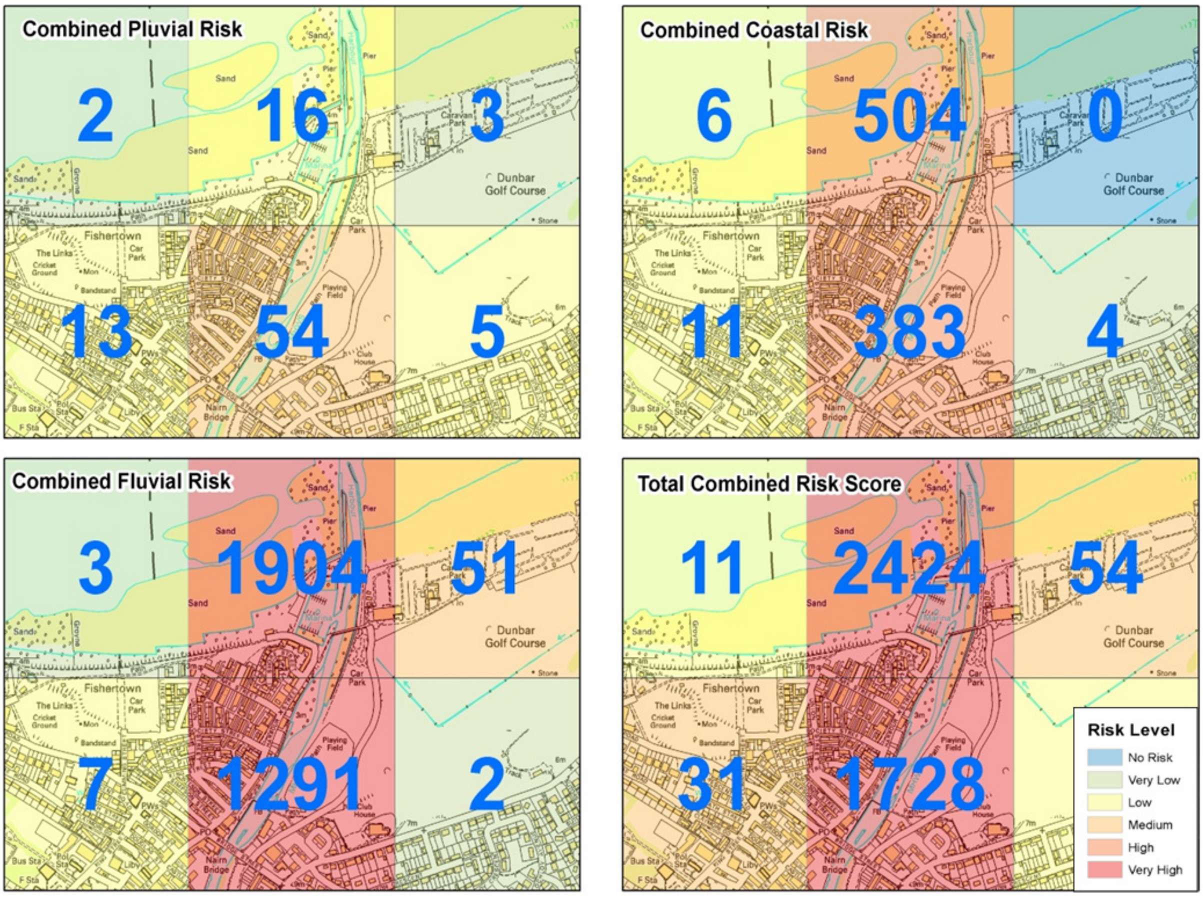


Figure 2: *PVA Grid Output*

Further to accounting for combined sources of flooding and receptor risk across numerous scenarios, an effective spatial clustering tool was designed to aid in identifying locations and boundaries of significant flood risk concentrations throughout Scotland. The tool has been optimised such that it processes clusters simultaneously at national scale and ensures that two (or more) growing clusters approaching one another are merged together to form a single cluster. The cluster analysis yields two outputs. The first output is a feature class representing the cluster extents and the second is a summary statistics table for each cluster. All fields included in the summary table and feature class are dynamic and depend on the user input to the cluster analysis. Figure 3 illustrates the cluster analysis outputs.

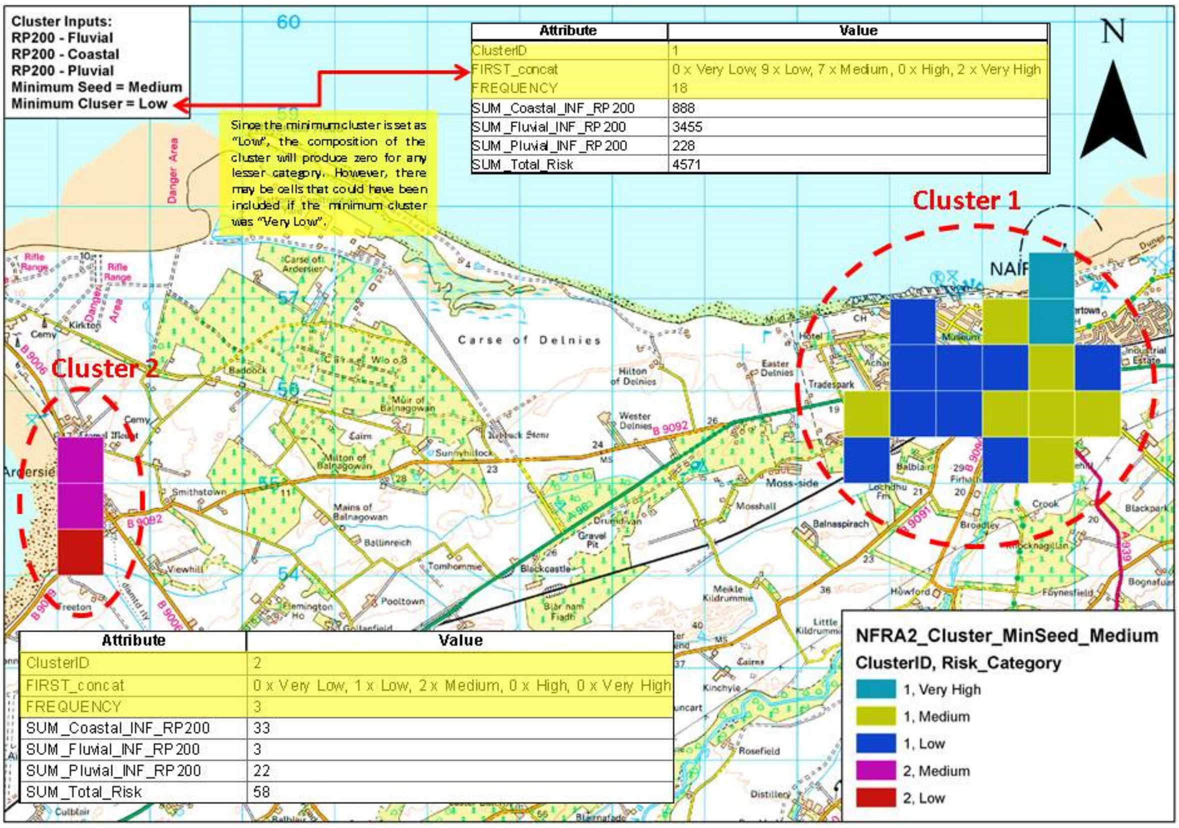


***Figure 3**: NFRA2 Cluster Analysis*

4. SUMMARY AND RECOMMENDATIONS

An automated approach was developed in close consultation with SEPA that can absorb all flood hazard data for Scotland and produce individual receptor risk information for all 2.5 million Residential Properties, 0.5 million Non-Residential Properties, 40,000 Community Facilities, 8,000 Utilities, 840,000 Road segments and over 1 million Agricultural land parcels in Scotland, as well as all national environmental sites, cultural heritage sites, airports and railway infrastructure in the country.

Fluvial, coastal and pluvial flood depths, velocities and extents are used for all scenarios (10yr, 25/30yr, 50yr, 100yr, 200yr, 1000yr, 30yrCC, 200yrCC) to assess individual and combined flood risk and calculate damages for all receptors. A conceptual model workflow was designed and models were initially built in ArcGIS model builder. These models were then converted and refined into a set of complex modular Python scripts capable of carrying out the NFRA2 methodology on a national scale and producing the required grid-based outputs which provide a visual pattern of flood risk across Scotland.

Initially, national runs of the data were executed while maintaining close collaboration for model development/improvement with SEPA by continually making minor amendments to various modules within the package. The final product that is delivered to SEPA is made up of 3 main modules:

1. Pre-processing hazard and receptor data
2. Receptor Analysis
3. Grid Outputs

The continual improvement and refinement of the automated workflow has enabled very fast runtimes, along with a great amount of functionality to be built into the models, to allow SEPA to update information and parameters based on best practice and best available scientific information. A major advantage of the NFRA2 toolbox is that they are adaptable to any scale of assessment, meaning they could be distributed to Local Authorities to enable nationally consistent flood risk assessments.

The main recommendations for future improvement of the automated workflow or for similar studies can be summarised as follows:

- More stringent flood hazard mapping quality protocols to be implemented. However, the standards the SEPA currently use are very adequate and this recommendation is only in relation to national flood risk assessment.

- Maintain a national flood hazard raster catalogue for all sources of flooding. This will facilitate efficient updates to master flood hazard datasets.

- All receptor data groups should be kept in merged datasets with an agreed schema. This reduces the complexity of models and prepares data for future analyses.

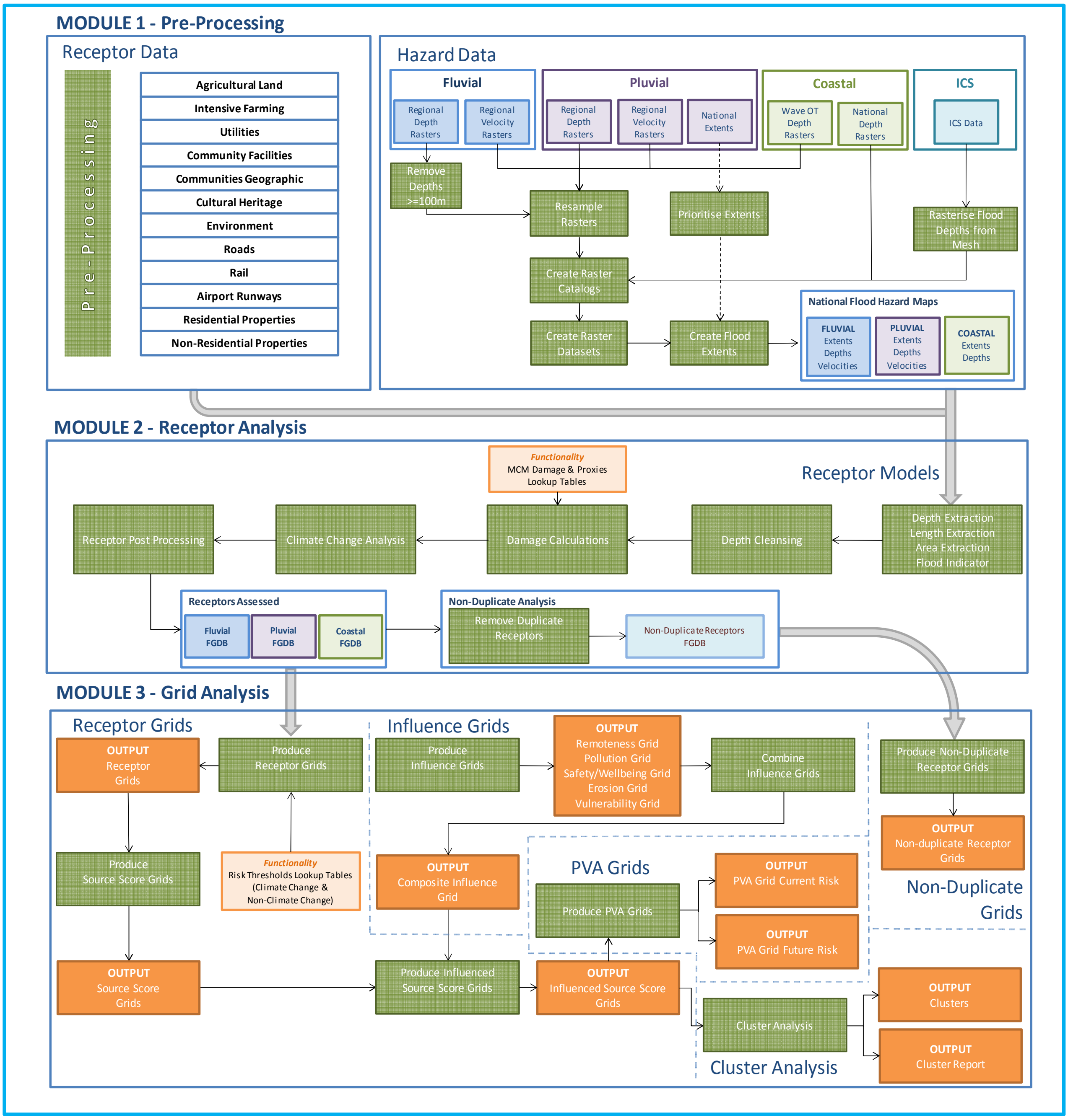
- More stringent quality control of spatial data to ensure point and polygon datasets are coincident.

- The national flood risk assessment could be further automated by integrating software with a higher performance and additional tools available than ArcGIS (such as FME).

- Alternatively, additional Python modules not currently available in the default Python package for ArcGIS (such as GDAL) supplied with ArcGIS may prove very useful. Doing this would require significant set-ups and installations carried out by a highly experienced programmer/computer scientist/ software engineer.

A final note should be aimed at the advanced technologies that are fast emerging and may offer substantial benefits in undertaking big data analyses, even in the context of national scale flood risk assessments. Open source packages offer utilisation of multiple Graphical Processing Units and harnessing such computational horse power would drastically increase efficiencies. These packages would potentially mean migrating data out of a spatial environment to enable the computational power before migrating back to a GIS interface. The use of these technologies is strongly recommended as an alternative or in combination with current software for the future.

5. DETAILED NFRA2 TOOL WORKFLOW



6. REFERENCES

Penning-Rowsell, E., Priest, S., Parder, D., Morris, J., Tunstall, S., Viavattene, C., *et al*. (2015). *Multi-Coloured Handbook*. Flood Hazard Research Centre, Middlesex University.