

Input Data Preparation: Methodology and Challenges

Colin Mazengarb,
Mineral Resources Tasmania



APPENDIX ONE

INTRODUCTION

This report documents the input sources and the methods used for construction of the spatial inputs consumed by the ANUGA modelling process. While it was initially presumed that data compilation would be a relatively straightforward GIS exercise, it has proven to be far more complicated process requiring a detailed summary. Issues encountered include the considerable size of the datasets that challenged both the hardware and software, errors discovered in supplied data that required a considerable level of manual intervention, and constraints of the ANUGA software itself.

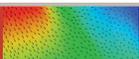
ANUGA INPUT CONSTRUCTION

ANUGA uses a finite volume modelling method for solving shallow water wave equations and requires four main inputs, described below:

1. boundary condition hydrographs representing tsunami scenarios
2. mesh resolution boundaries
3. elevation model
4. surface roughness model
5. gauge locations

Boundary condition hydrographs

Boundary condition hydrographs have been previously generated for a range of tsunami scenarios by Geoscience Australia (Burbidge *et al.* 2008) and are located along the continental shelf at the 100m mark (Figure 1). The project team have chosen to accept these inputs without further question but realise that they are based on assumptions and computer models that may change with further research. To this end, MRT is actively collaborating with a PhD candidate at the Karlsruhe Institute of Technology in Germany to better understand the tsunami generation potential of the Puysegur Trench and the uncertainties in magnitude and likelihood of the tsunamis that could be produced. Preliminary results involving modelling a range of variations of rupture geometries indicate that expected wave height has a significant standard deviation about the mean (Schäfer, *et al.*, 2016).



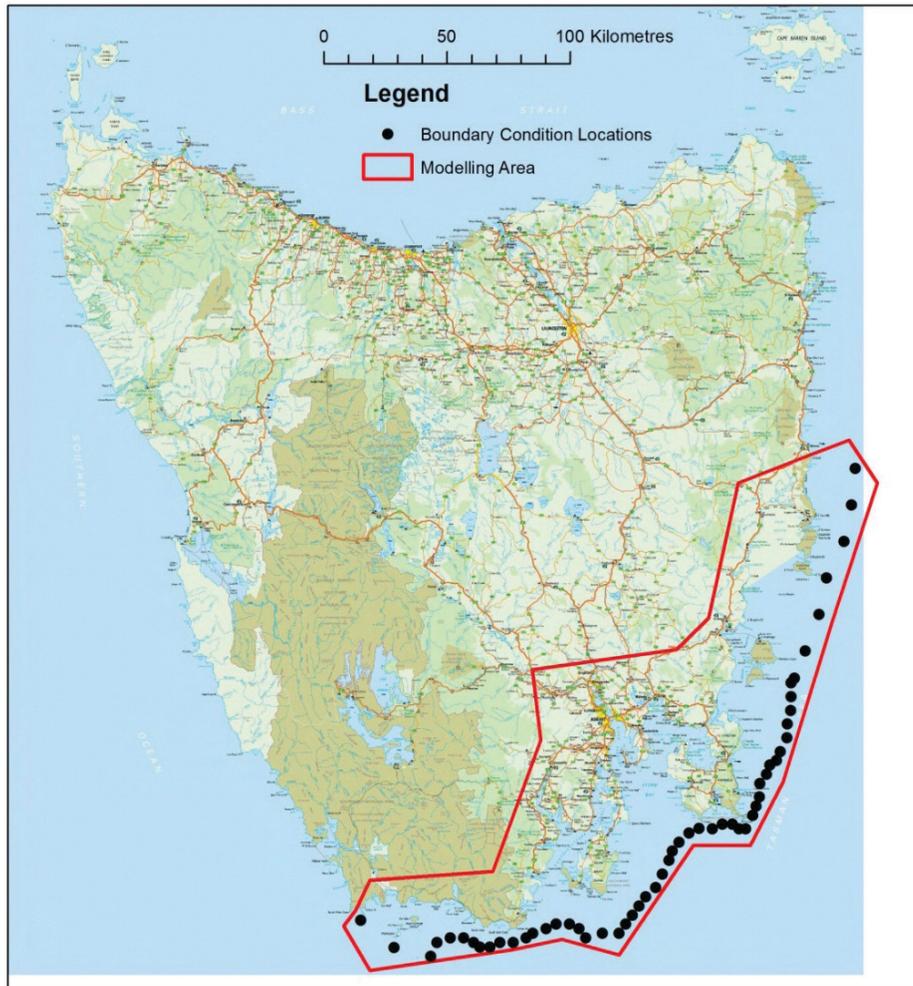


FIGURE 1: Model area with location of boundary condition hydrographs

Mesh Boundaries

ANUGA performs its calculations within a triangular mesh environment, either structured (equally spaced triangles similar to a raster) or unstructured mesh, where triangles are created using the Delaunay method. In this study an unstructured mesh is used as it allows us to vary the triangle dimensions across the study area. The construction of the mesh is performed within ANUGA based on a series of input GIS polygons (shapefiles) which define the triangle sizes. High mesh densities (small triangles) are assigned in priority areas where detail is required to delineate hazard boundaries, such as within low lying coastal settlements. Lower mesh density triangles (larger triangles) can be assigned to areas where detail is not required.

In constructing a mesh in a large area, such as that in this study, the limitations of the computing hardware must be considered in order to reduce the processing overhead and achieve acceptable run times (days rather than weeks). The two main constraints are the total number of triangles in the mesh (a question of size) and the size of the smallest triangle considered (the smallest triangle determines the size of the time step).

While the initial polygon shapefile contained a number of topologically correct features (no overlaps or underlaps) at least two issues arose. Some of the polygon shapes were too complex, having incorporated natural features such as shorelines into their boundaries. The triangle dimensions created by the ANUGA meshing process are controlled by the node spacing and

therefore “sliver” triangles (at times $<1 \text{ m}^2$) were formed in the vicinity of mesh zone boundaries where polygons were too complex. As previously explained, incorporating these into the model would significantly reduce the time step and consequently greatly increase the processing time. As such, the polygons were simplified in order to ensure a suitable node spacing. Some difficulties were also encountered regarding the relative proximity of different polygon resolutions. After some trial and error, finer resolutions were always nested within coarser, and intermediate resolution polygons were not omitted. In addition, the spacing between each intermediate polygon (i.e. the distance between the inner and outer boundary) was set at a minimum dimension of the desired triangle size in order to prevent undesirable slivers forming. The downside of this approach is that the interior of larger islands contained unnecessary detail, adding to processing overhead.

While most of the GIS preparation was undertaken using *ArcGIS desktop basic* software, some functions in other software (*Global Mapper*) were employed in order to achieve the desired outputs.

In this study a series of mesh boundaries have been created (Figure 2):

- Coarse resolution: 80 000 m² (maximum 400 m x 400 m triangles)
 - Ocean areas $> 30 \text{ m}$ depth.
 - Land areas above 15 m AHD (this is expected to be well above maximum run-up distance for expected tsunamis).
- Medium resolution: 20 000 m² (maximum 200 m x 200 m triangles)
 - Ocean areas between 10 m and 30 m depth.
- Fine resolution: 1 250 m² (maximum 50 m x 50 m triangles)
 - Non-urban areas of coastline from +10 m to -10 m AHD
- Very-fine resolution: 200 m² (maximum 20 m x 20 m triangles).
 - Urban areas of interest to be modelled and plotted in detail. The resolution is required to reflect flow paths on roads between buildings in built up areas and to yield clearer inundation plots than previously produced.

- Extra-fine resolution: 100 m² (maximum 10 m x 10 m triangles)
 - An urban area, at Blackmans Bay, was chosen specially to test inundation on an elevation model that includes buildings in 3D as extracted from LiDAR data (Figure 3).

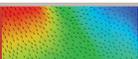
Construction of the Elevation Model

A variety of data sources were integrated in order to build a complete model of elevations within the modelled area. The initial intention was to use the elevation data created by Geoscience Australia (Van Putten et al. 2009) wherever possible, and replace those parts where subsequent more detailed information had become available. However, an inspection of the dataset revealed a number of significant problems and the 2009 dataset was discarded.

The metadata statement for the 2009 elevation dataset is comprehensive in describing the data sources but not the methods employed to merge them. The data was loaded into an ESRI *Terrain* model (a triangulated irregular network) in order to visualise the data. The analysis shows that the 2009 elevation model is composed of regularly spaced arrays of xyz points of varying density that have been combined into a single file. In offshore areas, the regular spacing of points indicates that it was created by an undocumented interpolation process(es), it is therefore derived data and not the original input points. This method is of concern as potential mistakes in the interpolation process cannot be reversed if source data is not supplied.

The most significant issues are:

- a. Obvious elevation errors in proximity to the coastline where some points offshore have positive values and conversely some points onshore have negative values, and whose values are well outside the tidal range. In one instance the coastline is effectively about 1 km from its correct position.
- b. Some of the offshore areas have unrealistic morphology and do not conform to reliable bathymetric information contained on navigational charts.
- c. There is a serious mismatch of elevations about an arbitrary offshore boundary that probably relates the join between two (interpolated?) datasets.



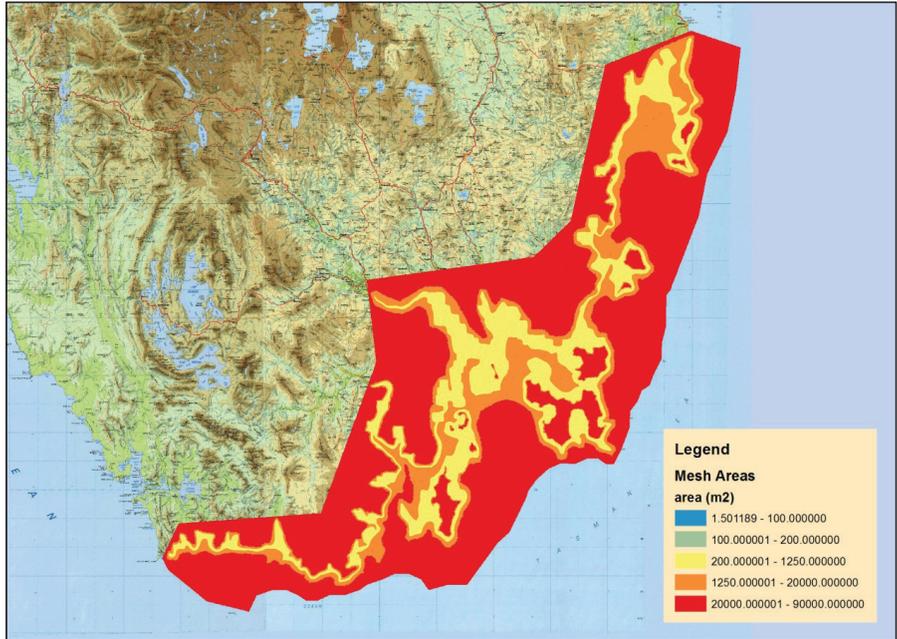


FIGURE 2 Variation in mesh resolution across the study area

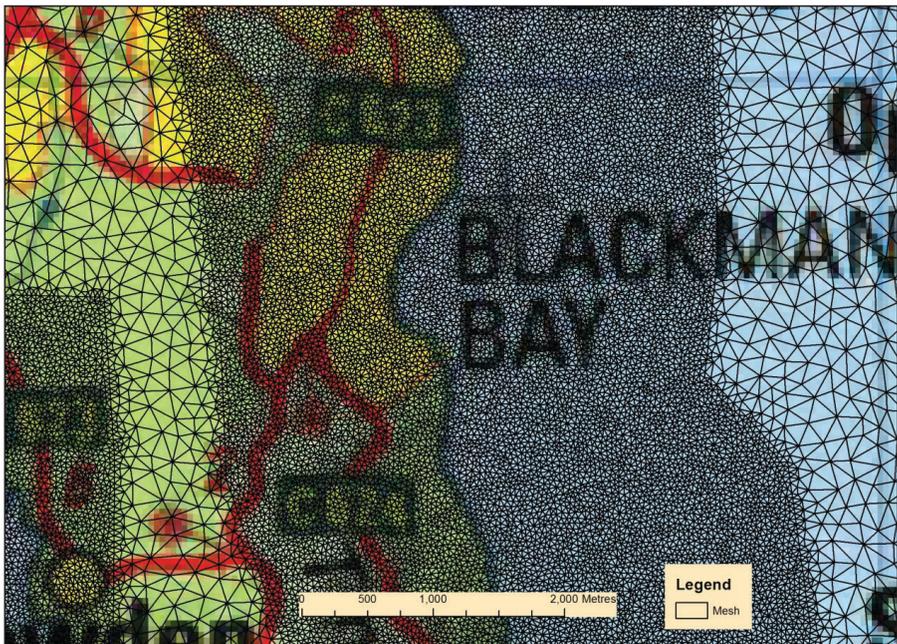


FIGURE 3 Detail of mesh resolution boundaries around Blackmans Bay where extra fine mesh sizes have been used



MRT has therefore completely rebuilt the elevation dataset using mostly publicly available data that largely supersedes the 2009 GA dataset. One of the key offshore datasets used by GA was provided under license from the Australian Hydrographic Office (AHO), which prohibited GA from passing on to MRT. Fortunately, MRT was able to purchase this data under a license agreement from AHO.

The underlying philosophy of the data compilation methodology was to populate areas with the most reliable and accurate information available. This task was performed in a GIS environment, using ESRI ArcGIS Desktop 10.1, Basic license level, with the 3D Analyst extension enabled. Despite having a reasonably powerful computer for its time (HP Z600 Workstation, Intel 2.13GHz processor, 8 core, 16 Gb RAM with a SSD drive), the huge data volumes involved restricted the performance of the hardware and limited the geoprocessing tools that could be used. ESRI *terrain* geoprocessing tools were found to be the most practical, time efficient and robust methods to perform edits such as masking (deleting unwanted areas or erroneous points) on very large datasets.

Data were acquired by MRT in various digital formats, and using a range of datums and coordinate systems, all of which required conversion to a common datum and projection (GDA94 MGA zone 55, AHD (Tasmania)). All of the data sources were imported into feature classes within separate *file geodatabases* in order to do this task and temporary *terrain* models were built for each data source as a quality assurance exercise and to rapidly visualise problem areas. Masking was necessary to ensure spatially overlapping datasets were not mixed.

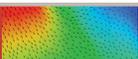
The creation of merged point cloud type elevation datasets is not without its problems. For instance, in areas of complex terrain, shallow water and sparse data will all affect the reliability of the modelling. To partly address this issue, additional artificial elevation points were added at regular intervals along the mapped shoreline to ensure that the land – water interface was sufficiently defined. However, unrealistic interpolation may have occurred, for example, in small embayments where bathymetric data is sparse and the artificial shoreline points interpolated horizontally across the bay. In situations where data is more plentiful, the use of breaklines along the shoreline (something

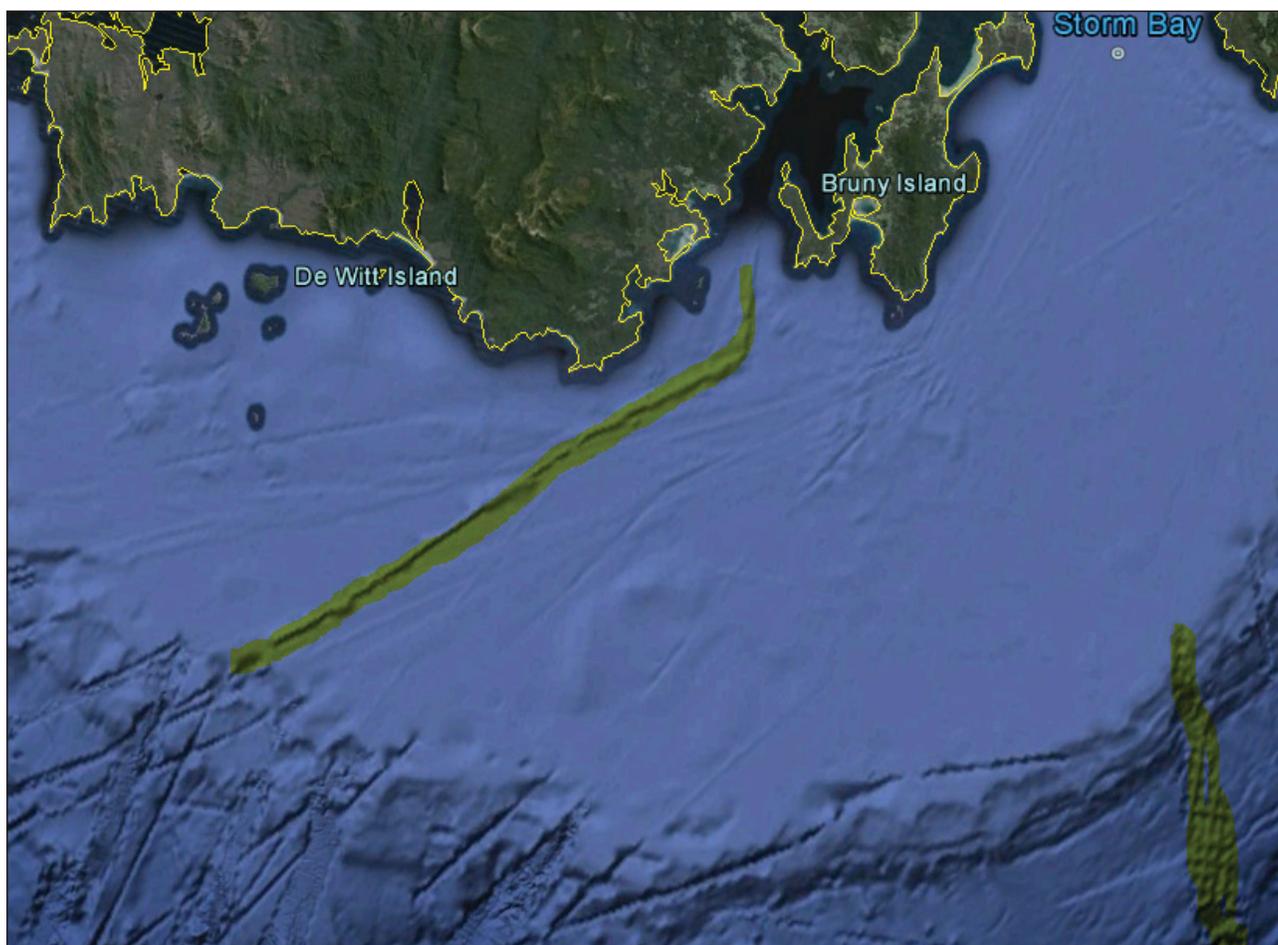
that ANUGA supports) could be used to control the interpolation. However, in data poor areas such as these they will not make much difference and it would take somewhat intensive desktop work to better control the interpolation. The use of breaklines was not adopted in this study for reasons of expediency as it is considered that, in any case, they may make little difference to the run up of the tsunami wave.

In reviewing other examples of tsunami modelling, it is noted that the *Service hydrographique et océanographique de la marine* agency (SHOM) in Brest, France is systematically undertaking investigations for coastal nuclear power stations (Maspataud, *et al.*, 2015). Their method involves creating DEMs (rasters) from a rich variety of overlapping data sources of varying accuracy and utilising the *Multi Level B-Spline* (MSB) method available in the SAGA GIS software (www.saga-gis.org). The MSB is an inexact interpolation algorithm that fits a smooth surface through scattered data while minimising local approximation error for each control point. Their study areas are relatively small compared with that in this project, and while the MSB method is probably very good, it would require a significant amount of tiling (and effort) in order to work within the constraints of the computer resources available. For this reason alone, the method was not adopted.

Each of the data sources employed will be described below, but first one important data source that was not used will be discussed.

The CSIRO swath data available from its data portal, is derived from multiple cruises over many years and consists of densely spaced (<1 m and irregular) xyz points of considerable data size. At the time of compilation, for some of the cruises, the data have not been tidally corrected and cannot be used in their present form. In addition these datasets appear to be internally noisy and inconsistent with other overlapping swaths. To contemplate using these data would require significant processing and smoothing to achieve an acceptable form using specialised software and expertise not available to MRT. We note that these data have been used in national bathymetric models and even on Google Earth imagery despite containing obvious elevation mismatches across the continental shelf (Figure 4). For these reasons the CSIRO data was not used to build the elevation model.





Bathymetric Data

Several data sources were available to construct the bathymetric model.

1. TasPorts swath data (acquisition date: 2010 – 2013)

A local and detailed swath dataset was obtained from TasPorts but did not come with a metadata statement. This dataset extends over four small areas; Sullivans Cove, Tasman Bridge Risdon Wharf and Selfs Point (Figure 5). Depth information required conversion from LAT (Lowest Astronomical Tide and depths in positive values) to AHD (bathymetry elevations in negative values). The data appears to be a good representation of reality at the time of acquisition but it is important to note that subsequent dredging has occurred at the overseas wharf (Sullivans Cove) that cannot be accounted for by this data.

FIGURE 4 Examples of probable bathymetric artefacts on the continental shelf and slope (two of many are highlighted in yellow) used on publicly available imagery (e.g. Google Earth).



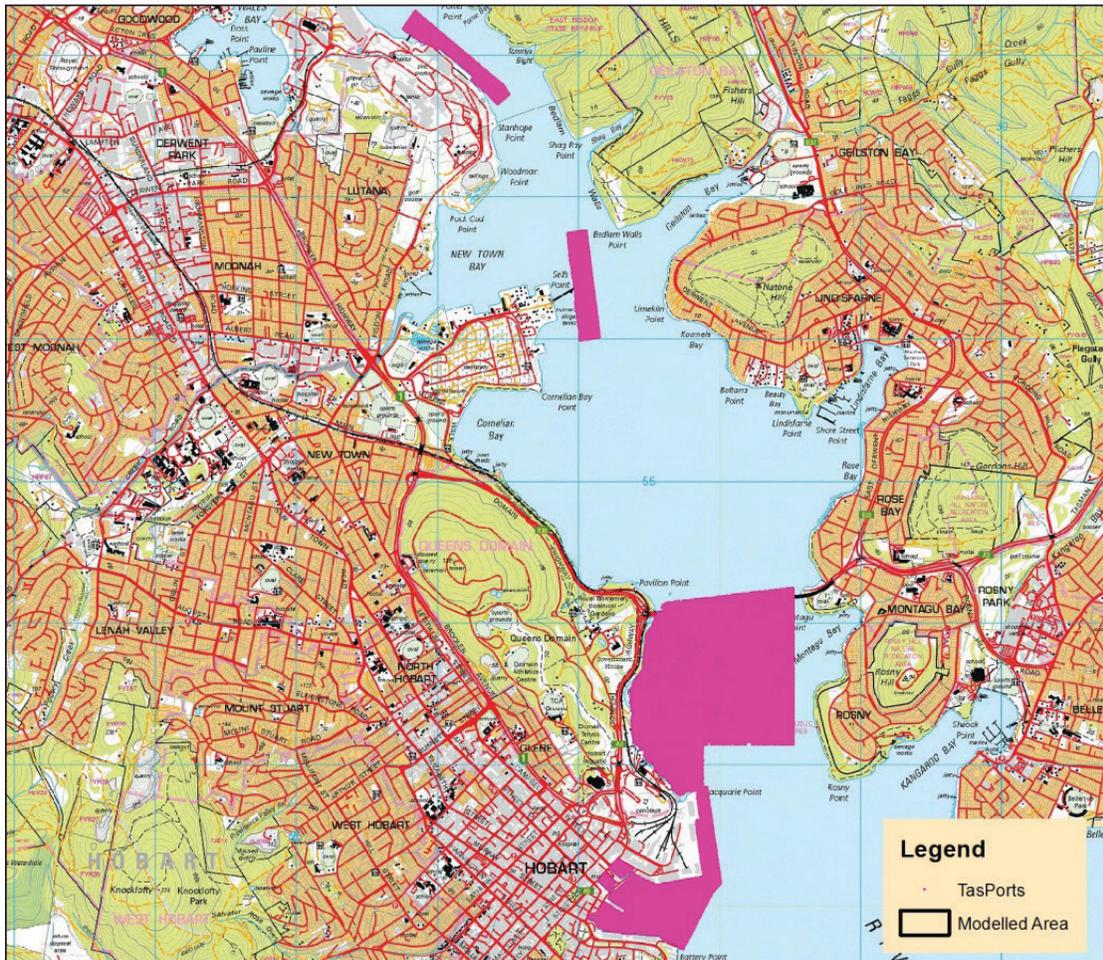


FIGURE 5 TasPorts swath data

2. SeaMap sonar data (formerly TAFI (Tasmanian Aquaculture and Fisheries Institute) now IMAS (Institute of Marine and Antarctic Science, University of Tasmania))

Several point datasets were obtained including a dataset used in the 2009 modelling by Geoscience Australia (the Bruny Bioregion described below) (Figure 6). The data have been collected by a small vessel traversing the coastal and estuarine waterways using a sonar device.

a. Bruny Bioregion dataset (acquisition date: 2001)

This coastal dataset extends from South East Cape through to Marion Bay but excludes the inner estuaries, such as the Derwent and Huon. A report by Barrett et al. (2001) describes the purpose and main findings of the study. A point file of elevations was kindly supplied to MRT on request, delivered

as a single shapefile but without a projection file, inadequate metadata and with many spatial problems. A published contour dataset available on the LIST web viewer (Southeastern Tasmania marine contour map 1:25 000) is presumably derived from a clean version of the point file and provided a means of identifying and fixing the data problems.

On assigning an AGD66 datum most of the data lined up with coastal features. However, one day of data collection transgressed onto land in parts. It was determined that the data on this day was in GDA94 MGA projection and once reprojected, the data fitted the coastal constraints well. All of the data were subsequently reprojected to the common datum.

There were also obvious problems close to sea cliffs, particularly on the Tasman Peninsula, where data transgresses onto land, probably due to poor GPS signals, and this data was selectively removed.

There were multiple instances where streams of data had fixed elevations or clearly were not realistic for their setting. A considerable amount of manual editing was undertaken to remove anomalous points. Problematic points were identified through the construction of contours using Delaney triangulation in an ESRI Terrain Dataset. Problem areas where the depth value was zero or showed a fixed value over large distances and showed as linear ridge lines. Another common problem was unrealistic depth values in intertidal zones or spurious isolated values (too deep or too shallow) that showed a bullseye pattern on the plotted contours. Once the problematic points were removed the contours closely matched those published by TAFI.

The data are transect-based with highly variable distances between each transect; exceeding 1 km in places and with individual data points tens of metres apart.

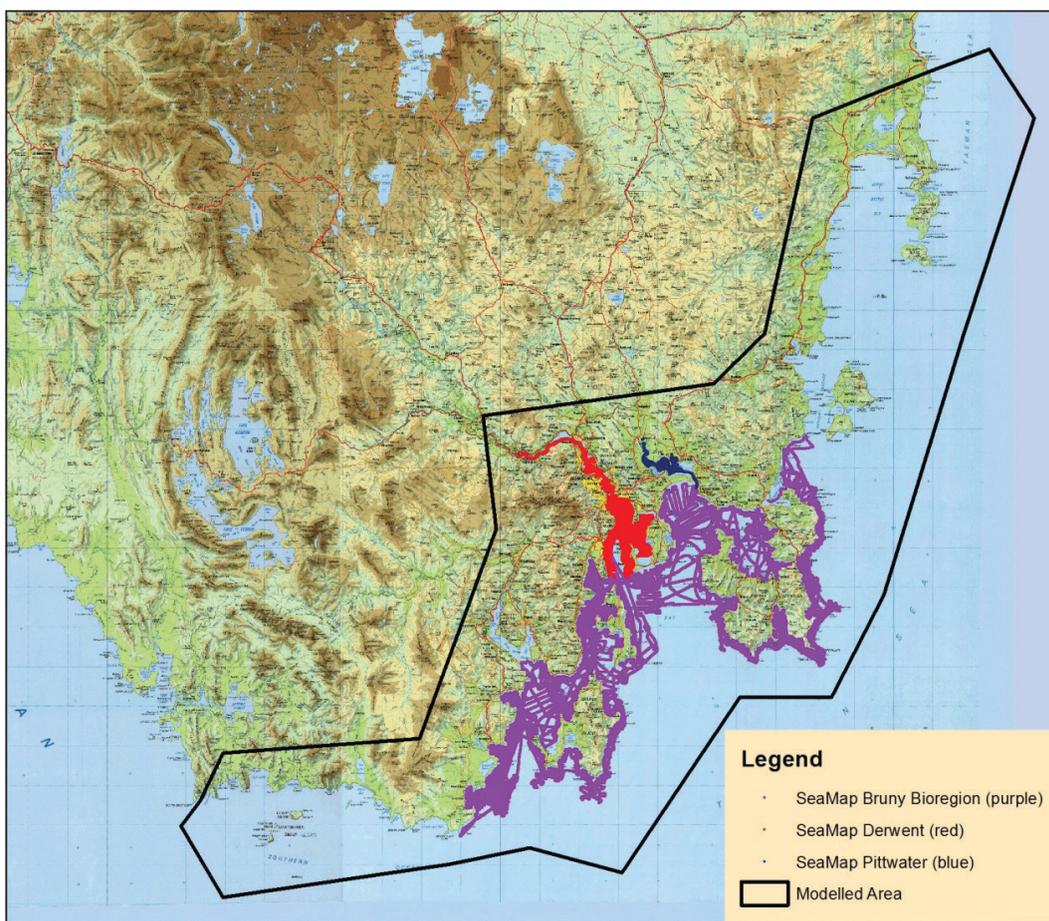
b. Derwent Estuary Program (acquisition date: 2007)

This dataset is a transect based acquisition along the Derwent Estuary, extending upstream of New Norfolk down to the Iron Pot near Kingston and including Ralphs Bay (Lucieer, et al., 2007). Transects are spaced approximately 200 m apart and individual points are at approximately 2 m separation. The data were in good condition and no anomalies were detected. However, the data were trimmed to remove points that intersected the TasPorts dataset in order to avoid introduction of artefacts.

c. Pittwater Estuary (acquisition date: 2002)

Data were obtained for the Pittwater Estuary in shapefile format based on a study by Davies et al. (2002). Unfortunately the data had issues in common with the Bruny Bioregion dataset, containing mixed projections and anomalous elevations requiring significant manual editing to produce a realistic model. The data were trimmed so as not to overlap with the TasPorts swath data.

FIGURE 6 Distribution of SeaMap datasets



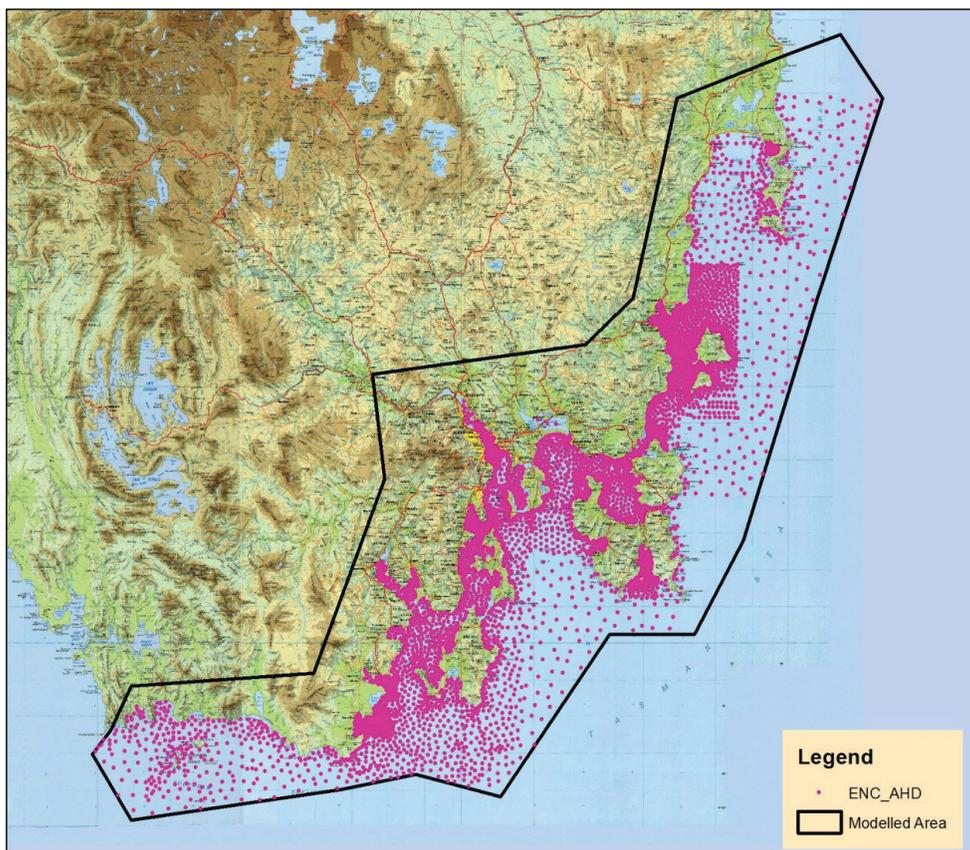


FIGURE 7 Electronic Nautical Chart data from the Australian Hydrographic Office used in the study

3. Australian Hydrographic Office Electronic Nautical Charts

Areas covered by the ENC (Electronic Nautical Charts) series were purchased under a licence agreement with the Australian Hydrographic Office (Figure 7). The data were in an uncommon vector format but which can be viewed and exported to other GIS formats utilising a free S57 viewer extension in the ArcGIS environment. The ENC data were derived from a variety of sources with varying accuracies. It is supplied in files corresponding to each nautical chart and where overlaps occur, data are duplicated. However, in viewing duplicate data from adjacent charts, it is observed that some of the data points are in slightly different locations. I suspect that this shifting of points resulted from a cartographic exercise to prevent overlaps with other features on the chart. While not an ideal situation, all duplicates were left in the elevation dataset and the process of creating the mesh should not have been seriously compromised.

4. CSIRO Bathymetry

A small dataset derived from the CSIRO was incorporated in to the model in the Huon area (Figure 8). These data have been smoothed by unknown parties into a regular grid of 50 m cell size.

5. Private kayak soundings on the Huon

A kayak based dataset of soundings in the Huon area was provided to Entura from a private individual (Figure 8). Not much is known about this information other than that the XY locations were probably collected using a hand held GPS device (~5 m accuracy) but no information is available regarding the vertical accuracy and whether it is tidally corrected. Given that other data sources were absent upstream of Port Huon, the inclusion of the poorly constrained information significantly improves the model in this area.



FIGURE 8 Data sources in the Huon area

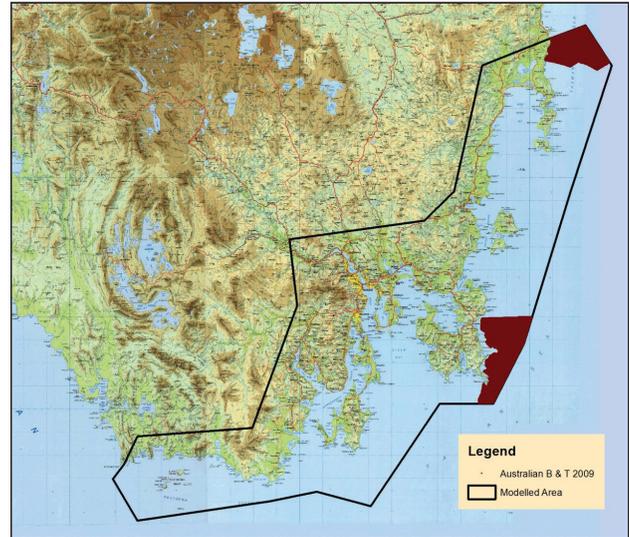


FIGURE 9 Selected data from the Geoscience Australia national bathymetry and topography dataset

6. Australian Bathymetry and Topography Grid 2009 (Whiteway, 2009)

This is a national bathymetric gridded dataset (raster) with approximately 200 m cell spacing. The grid was converted to point elevations and is used in a limited manner (Figure 9). On comparison with other datasets, it appears that there are many interpolation errors in this dataset, especially in shallow water where it conflicts with nautical charts. For this reason it has been used as a last resort to complete the elevation model.

Terrestrial Data

Public LiDAR datasets

A number of publicly available terrestrial LiDAR datasets were used in this study, all of which have been acquired since 2009. The Climate Futures of Tasmania LiDAR dataset used by Geoscience Australia in 2009 was not used as it is known to have sub-standard elevation control and it has been superseded by subsequent surveys. All of the datasets listed below have index files available on the LIST website so only the names and acquisition dates are provided in this report.

1. Mt Wellington 2011
2. Geoscience Australia Greater Hobart 2013
3. Geoscience Australia Huon 2013
4. DPAC Coastal 2014
5. Coal Mines (Tasman Peninsula) 2015

For the first three datasets listed, the data were trimmed to extract the points below 15 m AHD to minimise the file size. 15 m AHD was considered to be comfortably above the highest run-up value. The remaining datasets were relatively small and were included without modification.

In working with this data it was discovered that the Geoscience Australia datasets listed above contained many discernible classification errors, where points clearly overlying water have been classified as ground returns (Figure 10). In this case, the points were manually deleted.

A similar, albeit more subtle, problem occurred with the Mt Wellington LiDAR dataset. In the area adjacent to the TasPorts bathymetry datasets (Figure 5), there were areas of LiDAR ground returns in water that conflicted with the bathymetric data. In these areas LiDAR points were carefully deleted. This was considered most important wherever there were wharves in the vicinity to ensure that an accurate model was created. The extent of the combined LiDAR datasets is shown on Figure 11.

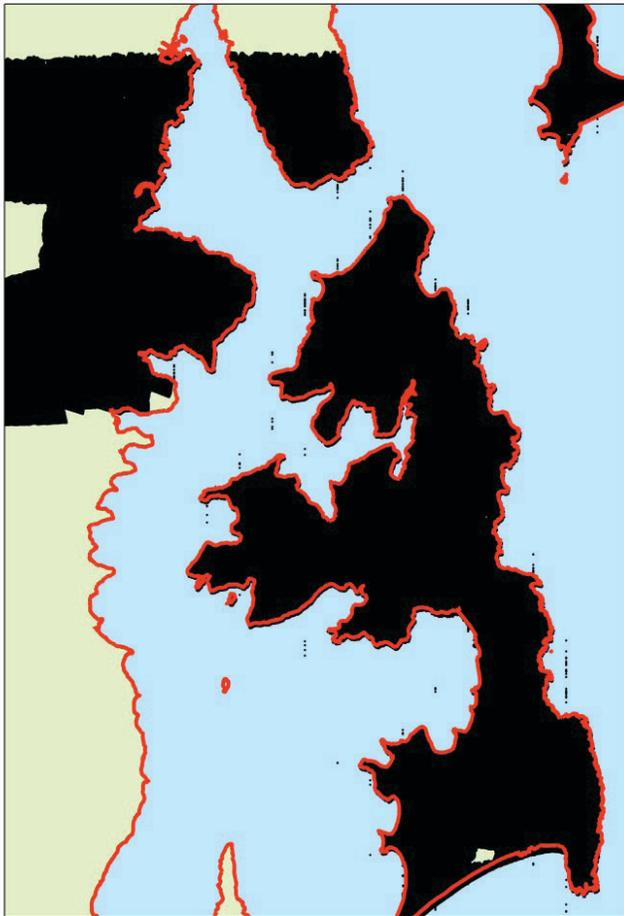


FIGURE 10: Example of LiDAR classification errors adjacent to North Bruny Island. These are all classified as ground returns.

A second dataset of points classified as buildings was extracted from the datasets listed above. This dataset was added to the elevation model as a trial of detailed inundation modelling in selected areas and as an alternative to the Mannings-N approach at a detailed level.

Photogrammetric derived topographic DEM

XYZ points were extracted from the Statewide 25m DEM to populate the remaining terrestrial areas. This dataset originates from Land Tasmania, DPIPW and has been created from photogrammetric contours at 10 m spacing.

A seamless elevation model for South East Tasmania

The *ESRI Terrain* model provides a powerful data type with associated tools to manage large volumes of diverse elevation data effectively. A particular strength of the terrain model is its ability to rapidly visualise the elevation model and identify potential errors before the data is used by the ANUGA software. A considerable amount of time was spent in the QA process ensuring that there were no serious errors or join artefacts visible. Two visualisations are provided (figures 12 and 13) that compare results between the GA 2009 and the MRT elevation models.

The final stage in the compilation process involved exporting the completed terrain model (Figure 14) as a single point file for importing into ANUGA.

Surface Roughness Model

A surface roughness model of the entire area was compiled from publically available datasets into the form of a raster grid with assigned Mannings-N values (Figure 15). This information is an important control on wave and run-up attenuation over the modelled area. For instance, some objects, such as buildings have high attenuation effects on run-up whereas smooth surfaces such as road pavement have low values. This project has used the Mannings-N coefficients listed below based on commonly used values and experience within the team. It is important to note that the modelling undertaken by Geoscience Australia in 2009 used a single Mannings-N value over the entire area and therefore did not account for variation in roughness.

Mannings Value	Surface Type
0.5	Solid buildings
0.071	Built up areas
0.055	Vegetated areas
0.035	Land (default)
0.03	Bare ground
0.025	Water courses
0.018	Roads
0.01	Oceans and estuaries



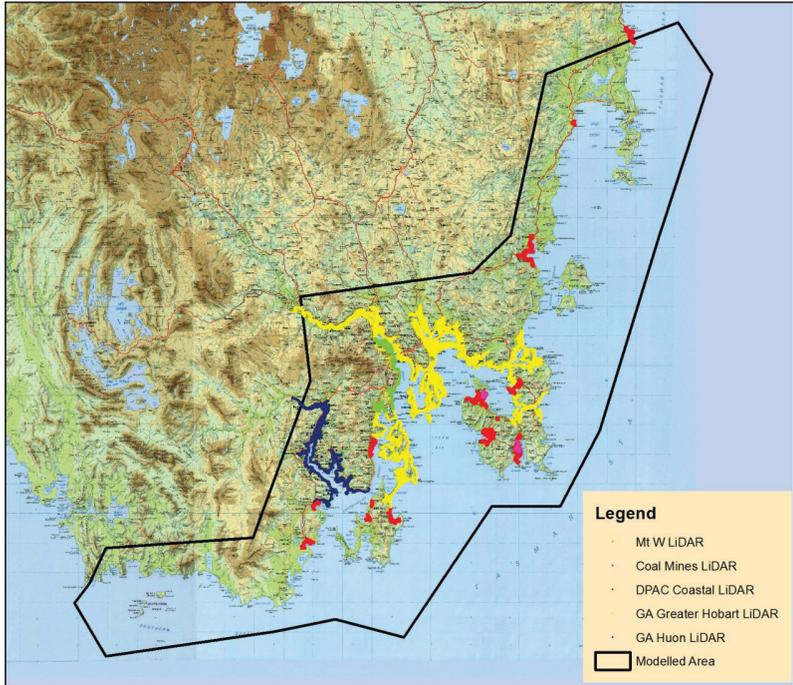


FIGURE 11 Ground classified LiDAR used in elevation model. Most of the datasets were trimmed to a maximum of 15m elevation.

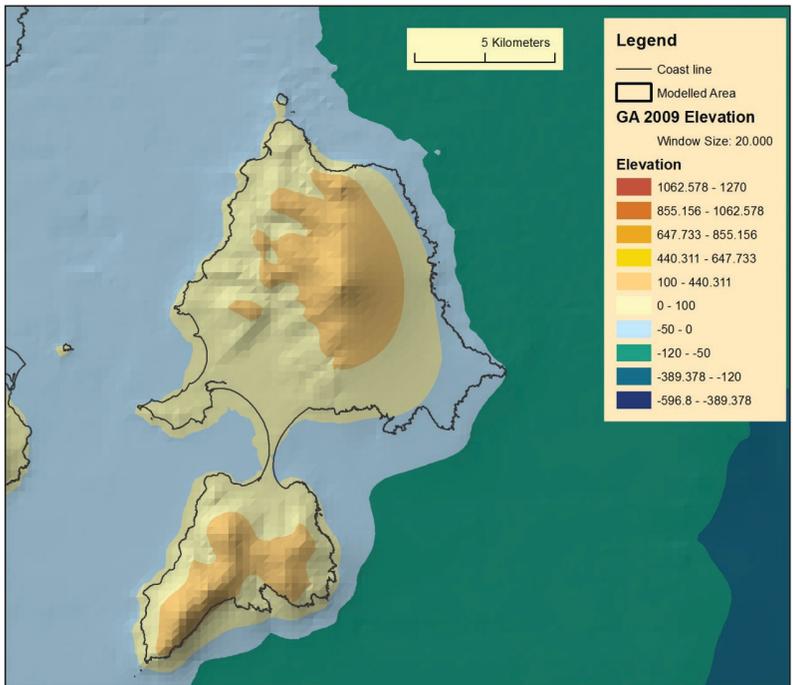


FIGURE 12 GA 2009 elevation model in detail at Maria Island. Note the mismatch of the model with the official coastline

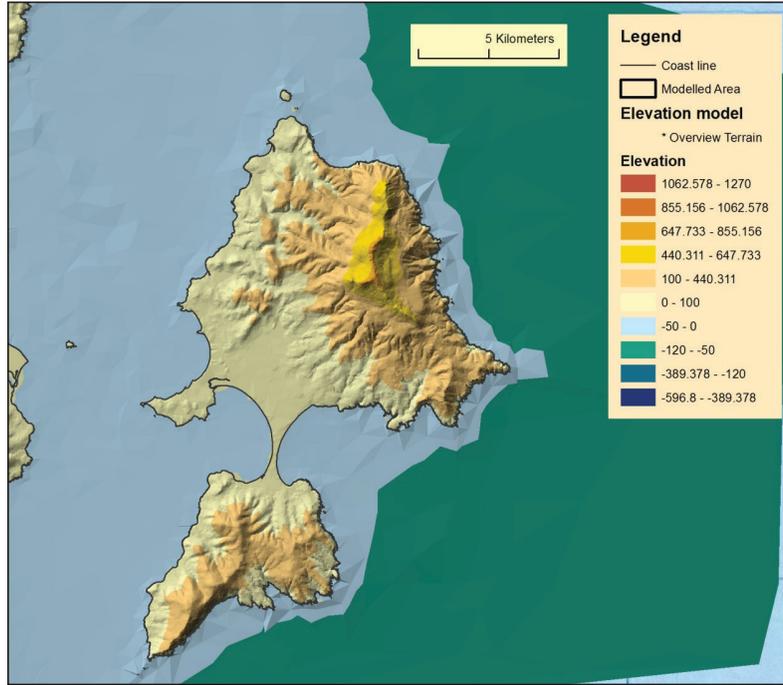


FIGURE 13 Detail at Maria Island for the 2016 elevation model to compare with the 2009 model (Figure 12)

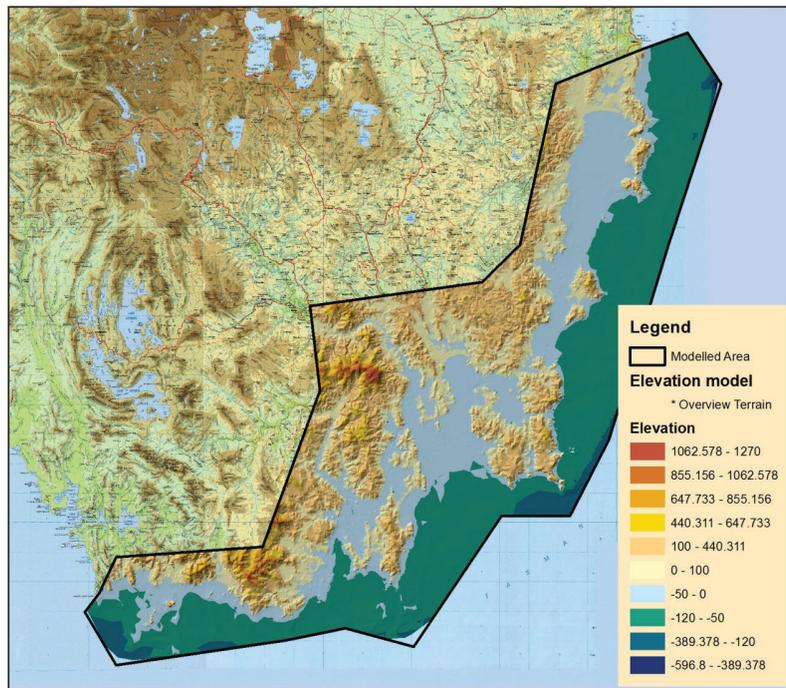


FIGURE 14 Elevation model adopted for the 2016 modelling (this study).

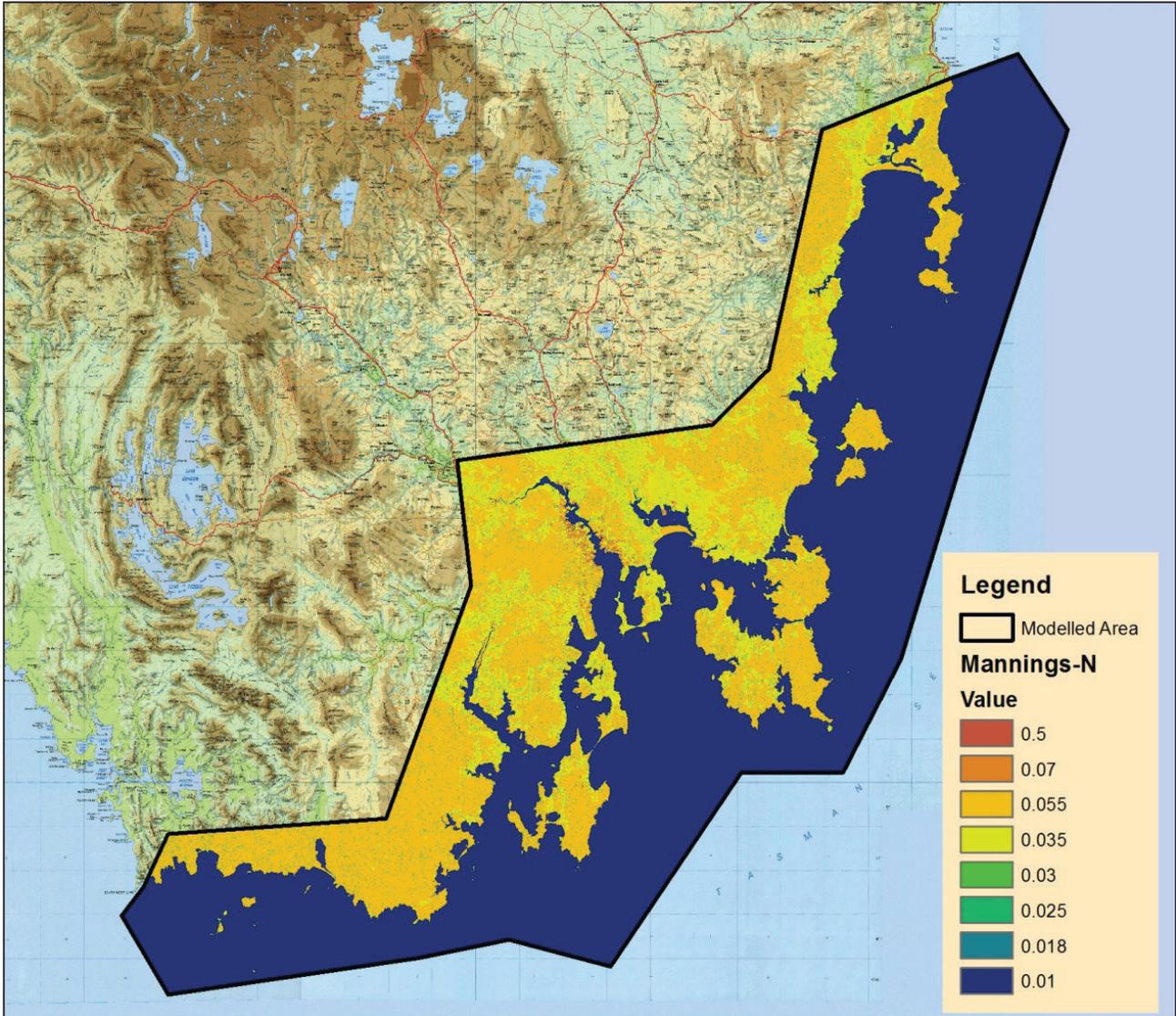


FIGURE 15 Mannings roughness map at a regional scale

The process of compiling this layer was based on a method originally developed by M. Hannon (Department of State Growth). The method he used was entirely a polygon vector-based geoprocessing operation using a third party extension to avoid using functionality only available at the “Advanced” licensing level of ArcGIS.

The method used here requires only the “basic” licensing level of ArcGIS but with *Spatial Analyst* enabled. It differs in that all vector layers are converted to raster and a very efficient raster algebra operation is performed assigning values based on a priority system at a resolution of 10 m.

The vector layers used are listed below in decreasing order of priority:

1. Roads derived from the LIST transport layer (polyline) with bridges removed
2. Buildings extracted from all LiDAR layers discussed above
3. Water courses from the LIST water course polyline layer
4. Vegetation derived from the TASVEG polygon layer
5. Land use derived from the LIST cadastre polygon layer
6. Oceans and estuaries based on the LIST coastal polygon



The geoprocessing operation was constructed using *Model Builder*, a graphical user interface for coding within ArcGIS. Through experimentation a set of queries and geoprocessing operations were developed to prepare the data for assembly.

For the buildings layer, the LiDAR points classified as buildings and under 15 m elevation were converted to raster format (cell size 5 m) using the *point to raster* geoprocessing tool and excluding a count < 3 (less than 3 points in the cell). These cells were assigned a Mannings-N value of 0.5. A buildings polygon layer is also available from the LIST, which is used in the Tasmanian Street Atlas product. Unfortunately, the completeness of this layer is highly variable within local government areas and individual buildings are not sufficiently accurate for this purpose.

For the roads layer, the query consisted of selecting the following from the TRAN_CLASS field in the LIST transport layer and assigning a Mannings-N value of 0.071 (Built Areas): 'Arterial Road', 'Access Road', 'Local Road', 'National/State Highway', 'Sub Arterial Road', 'Collector Road'.

For the landuse layer, entries in the CAD_TYPE2 field were assigned the following values:

'Private Parcel' = 0.071 #Built Areas

'Aurora Energy Pty Ltd' = 0.030 #Open ground

'Department of Health and Human Servic' = 0.071 #Built Areas

'Housing Tasmania' = 0.071 #Built Areas

For the vegetation layer, records with the following entries in the VEGCODE field were excluded: 'FUR', 'FPE', 'OAO', 'FAG', 'OSM', 'FUM', 'FMG', 'FRG', 'ASS', 'AHS', 'AHF', 'ASF', 'AUS', 'GHC'. The remaining records were assigned a Mannings-N value of 0.055.

REFERENCES

Barrett, N.; Sanderson, J.c.; Lawler, M.; Halley, V.; Jordan, A. 2001. Mapping of inshore marine habitats in south-eastern Tasmania for marine protected area planning and marine management, vol. 7. Tasmanian Aquaculture and Fisheries Institute Technical Report Series: Hobart. 73p.

Burbidge, D.; Mleczko, R.; Thomas, C.; Cummins, P.; Nielsen, O.; Dhu, T. 2008. A probabilistic tsunami hazard assessment for Australia, Geoscience Australia Professional Opinion 2008/04: Canberra.

Davies, P.E.; Crawford, C.M.; Wells, F.A.; Dunstan, P.; Mitchell, I.M. 2002. Environmental Flow Assessment of the Lower Coal River and Pitt Water Estuary. Freshwater Systems and the Tasmanian Aquaculture and Fisheries Institute.

Lucieer, V.L.; Lawler, M.; Morffew, M.; Pender, A. 2007. Estuarine Habitat Mapping in the Derwent - 2007. A Resurvey of Marine Habitats by SeaMap Tasmania. Tasmanian Aquaculture and Fisheries Institute, University of Tasmania: Hobart.

Maspataud, A.; Biscara, L.; Hébert, H.; Schmitt, T.; Créach, R. 2015. Coastal Digital Elevation Models (DEMs) for tsunami hazard assessment on the French coasts (Poster). EGU 2015 General Assembly Vienna, Austria, 13-17 April 2015.

Schäfer, A.M.; Daniell, J.E.; Mazengarb, C.; Rigby, E.H.; Wenzel, F. 2016. The Tsunami Risk of South-East Tasmania, Australia – the Impact of the Puysegur Trench on greater Hobart (Poster). European Geoscience Union, Vienna.

Van Putten, K.; Fountain, L.; Griffin, J.; Sexton, J.; Nielsen, O.; Wilson, R. 2009. Capacity building for tsunami planning and preparation: inundation models for south east Tasmania. Geoscience Australia Professional Opinion. No.2009/07. Geoscience Australia: Canberra.

Whiteway, T. 2009. *Australian Bathymetry and Topography Grid, June 2009. Scale 1:5 000 000*. Geoscience Australia, Canberra.

