High-resolution estimates of Australia's coastal population

Keping Chen[1] and John McAneney[1]

[1] Here we quantify Australian coastal population at a spatial resolution of 1 km and do this as a function of distance to shoreline and elevation above mean sea level. We also report comprehensive validations based on statistical and spatial relationships between very fine-resolution Australian data sets and the recent high-resolution global data sets on ambient population distribution (LandScan2003), shorelines (GSHHS) and elevation (SRTM). Estimates are heavily dependent upon the resolutions of all input data, as well as the resolution of analysis. Our results show about 50% of Australian addresses or population are located within 7 km of the shore, and that population decreases very rapidly with increasing distance from the shoreline. About 6.0% of Australian addresses are situated within 3 km of shorelines in areas with elevations below 5 m. **Citation:** Chen, K., and J. McAneney (2006), High-resolution estimates of Australia's coastal population, *Geophys. Res. Lett.*, *33*, L16601, doi:10.1029/2006GL026981.

1. Introduction

[2] The enormous death toll in coastal communities of the Indian Ocean caused by the mega-tsunami of the 26 December 2004 Sumatra-Andaman earthquake [*Marris*, 2005], the impact of destructive hurricanes and accompanying storm surges [*Travis*, 2005] and concerns about global warming and sea-level rise [*Church and White*, 2006] have all served to heighten awareness of the vulnerability of populations living in low-lying coastal areas. Living with natural disasters, especially low probability-high impact catastrophes [*Grossi and Kunreuther*, 2005], challenges the sustainability of such communities.

[3] Making risk-informed decisions about mitigation measures or policy requires consideration of all contributing agents to risk - hazard, exposure and their vulnerability [*Alexander*, 2000]; that is, *Risk = f (Hazard, Exposure, Vulnerability)*. (This functional relationship is shown in schematic form in Figure S1: see the auxiliary material[2], which includes Figures S1–S5). Here, we focus exclusively on the exposure component of this relationship. In particular, we seek to quantify the magnitude and spatial distribution of Australian addresses and population as a function of proximity to the shoreline (horizontal) and elevation above mean sea level (vertical). Previous attempts to do this have invariably been undertaken at coarse resolutions, but now thanks to the increasing availability of fine-resolution geospatial data such as digital terrain models, shore-lines and address-specific exposure data, more detailed estimations are possible.

2. Data and Analysis

[4] We have used very fine-resolution Australian data, including the Geocoded National Address File (G-NAF) of 2004, detailed shorelines (scales 1:4K at urban areas and 1:250K at remote areas), and 5m-resolution digital elevation models (DEM-5m, specific to ground level and with a vertical accuracy better than 1m). The latter covers three regions of New South Wales (Sydney Basin, South Coast, and Central Coast), and two regions of Queensland (Gold Coast and Mackay), comprising about 30% of all national addresses. Each Australian G-NAF address (residential, commercial or industrial) is represented by a latitude and longitude referring to the centroid of each land parcel. The detailed shorelines are specific to any coastal waters (e.g., rivers, lakes and estuaries) directly connecting to open ocean.

[5] For validation purposes, we also employ detailed and consistent global data sets on modelled ambient population (LandScan2003 at 30-arc-second resolution ($30''$) or about 1 km at the Equator, http://www.ornl.gov/landscan/), the Global Self-consistent, Hierarchical, High-resolution Shoreline (GSHHS) (at a 1:250K nominal scale, http://www.soest.hawaii.edu/wessel/gshhs/gshhs.html), and elevation ("finished" version of the Shuttle Radar Topography Mission – SRTM at 3-arc-second ($3''$) or about 90 m at the Equator, http://srtm.usgs.gov/). (See the auxiliary material for data availability, quality and a description of the pre-processing involved.) The distribution of physical G-NAF addresses (9.6 million unique addresses nationwide) represents an opportunity to "ground-truth" the population surface in LandScan2003 (19.7 million Australians represented in LandScan2003).

[6] Polygons for open ocean waters were rasterised at various spatial resolutions (i.e., 1 km, 500 m and 250 m), and then the shortest spherical distance [*Longley et al.*, 2001] - the great circle distance - between centroids of populated cells (at 1 km, then equally-split 500 m and 250 m) and shoreline cells calculated. To our knowledge, no mainstream GIS software can calculate the great circle distance in a sufficiently efficient manner for its implementation to a large region, and for this reason an optimized in-house tool was developed. The raster (or cell)-based approach has two obvious advantages: first, it avoids straightforward vector-based distance buffering with complex shorelines, a task that is extremely time-consuming and impracticable; and secondly, it takes advantage of readily-available population and elevation data. Distance calculations based on spherical

[1]Risk Frontiers–Natural Hazards Research Centre, Macquarie University, North Ryde, New South Wales, Australia.
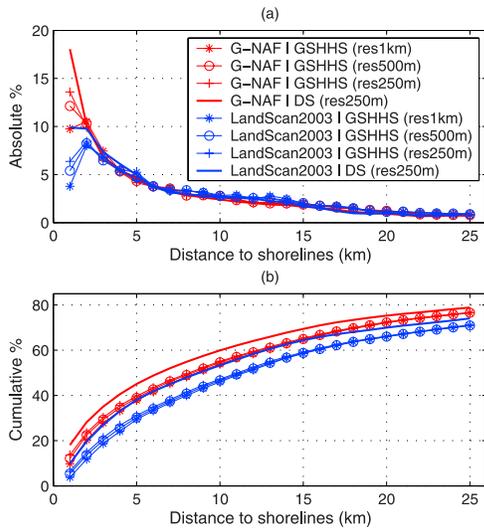
Figure 1. Percentages of G-NAF addresses and Land-Scan2003 population within the first 25 km from shorelines in Australia. Both Australia's detailed shorelines (DS) and global GSHHS shorelines were compared and all estimations conducted at 1 km, 500 m and 250 m resolution levels. (a) Absolute percentages. (b) Cumulative percentages.

trigonometry with explicit cell locations in terms of latitudes and longitude provide for accuracy and consistency.

3. Coastal Population in Australia

[7] Figures 1a and 1b compare absolute and cumulative percentages of G-NAF addresses and LandScan2003 population within the first 25 km of the coastline. Leaving aside for the moment obvious differences within the first 2 km, it is clear that absolute percentages of addresses decrease quickly with distance showing that more people choose to live closer to the coast. For distances beyond 2 km and for any resolution analysed - 1 km, 500 m and 250 m - absolute percentages of addresses and population are similar and corresponding cumulative curves all show remarkably consistent trends. This accord provides clear evidence for a strong spatial correlation between addresses and population.

[8] Large underestimates in the first 2 kms are related to the resolution of: exposure data; shorelines; and analysis. LandScan2003 shows how overall "populated areas" inferred from spatial contextures - road networks, land cover and night-time lights - are biased inland in comparison to the surface of G-NAF addresses (Figure S2) and thus cannot faithfully represent the extremely dense littoral population. Finer resolution shorelines delineate more land areas potentially exposed to ocean waters. Within the first 1 km, 18.1% and 13.6% of addresses are quantified with the most detailed shoreline and the GSHHS shoreline (both resampled at 250 m) respectively.

[9] Given more people are located closer to the shoreline, distance calculations based on cell centroids will include more population for smaller distance ranges as the population is resampled at finer resolutions (from 1 km to 500 to 250 m). This improves the accuracy of measurement for our results aggregated and reported at 1 km level. As resolution becomes increasingly finer (e.g., from 250 to 100 to 50 m),

however, there is a limit to this improvement as differences between estimates at two consecutive resolutions become negligible. We confirm this for a study area in eastern Australia (Figure S3). This validation suggests that for more objective estimates for the distance of the first pixel (1 km in our case), the resolution of analysis should be at least a half to a quarter of the intended reported resolution; otherwise, significant underestimates are likely and interpretations based on such results become inappropriate.

[10] From above, it is apparent that LandScan2003 only produces lower-bound estimates of coastal population for Australia. We conducted similar validations for each coastal state of Australia taking advantage of the wide spatial variability of addresses, population and sample numbers. The same overall trend and spatial patterns holds true for each state. Using the most detailed input and analysis, about 30% and 50% of Australian addresses are located within 2 km and 7 km from shorelines, reinforcing the observation that Australia is an 'island' country. In what follows, we examine how these coastal addresses or populations are distributed in relation to underlying elevation.

4. Population Vulnerable to Significant Sea-Level Rise

[11] Since a very high-resolution (e.g., 5 m) DEM for the whole country does not exist, our quantification of the number of vulnerable coastal addresses at a national level is based on the SRTM data after its validation in various regions. As the first test, we selected all G-NAF addresses in the Sydney Basin within 3 km of the most detailed shoreline representation available, in order to see how they were distributed on two independent elevation surfaces: DEM-5m and SRTM-3″. Both surfaces display very similar distributions of height (Figure S4) and statistical results on
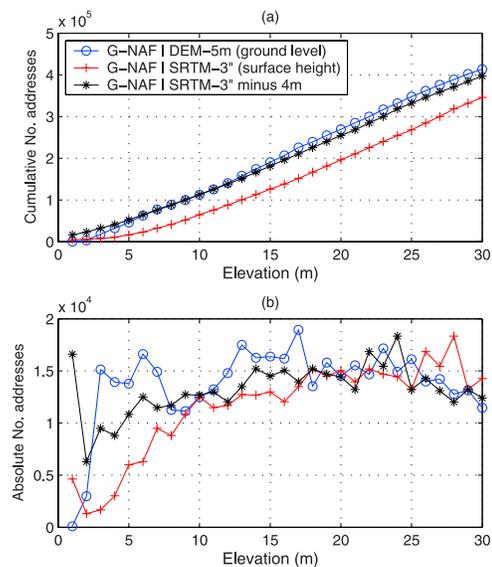

Figure 2. Validation of coastal addresses in relation to elevation in the Sydney Basin. All G-NAF addresses within 3 km of shorelines in the Sydney Basin are distributed at two elevation surfaces: DEM-5m and SRTM-3″. (a) Cumulative number. (b) Absolute number.
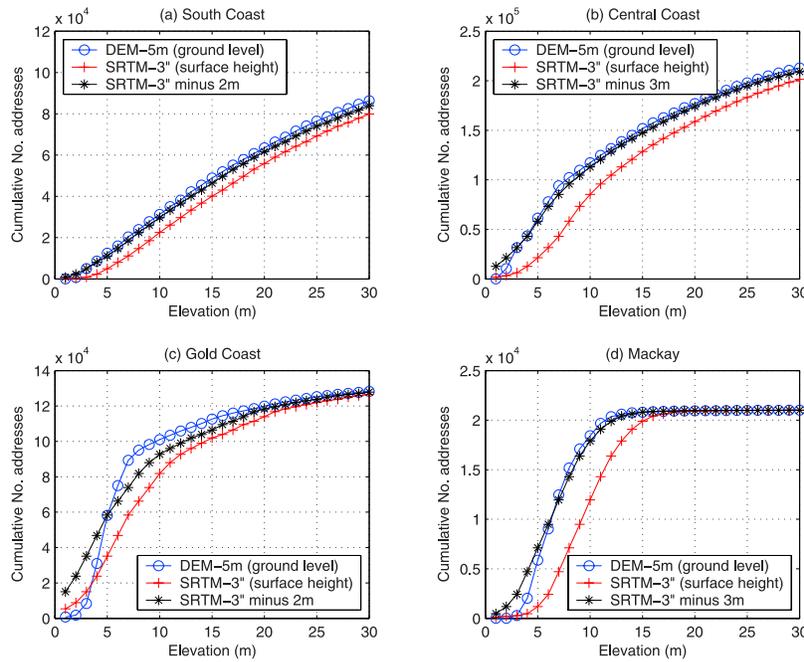
Figure 3. Validation of coastal addresses in relation to elevation in (a) South Coast, (b) Central Coast, (c) Gold Coast, and (d) Mackay. All G-NAF addresses within 3 km of shorelines in South Coast and Central Coast, and address sets in the Nerang River Catchment, Gold Coast and in Mackay were analysed.

cumulative numbers of addresses show consistent trends for both elevation surfaces (Figure 2a).

[12] We also observe the wide disparity of the absolute numbers of addresses in low SRTM elevations (e.g., less than 9 m, Figure 2b). Since SRTM data represent the average height of reflecting surfaces (mainly rooftops in this area), we first adjusted for this by removing the measured 4.0 m average difference between surface height (SRTM-3″) and ground level (DEM-5m) for address locations. The result is to shift the original SRTM-based curve leftward to closely match the DEM-5m curve (Figure 2a). This suggests that for gross, cumulative estimations of vulnerable addresses at larger elevation levels (e.g., ≥5 m), reliable results can be effectively derived using the SRTM data set and with considerable cost savings over the DEM-5m in production or purchase.

[13] This general finding is supported by empirical tests for address sets at other distance ranges (e.g., 1 km or 5 km) and in the other regions - South Coast, Central Coast, Gold Coast, and Mackay – used for validation (Figure 3). For reporting cumulative results at 5 m level, we find reasonable SRTM elevation adjustments to be about 2−4 m. At all validation sites, the number of addresses with elevations less than 1 m is virtually zero and is minimal less than 2 m. Given local tidal oscillations and storm surges, these features are perhaps unsurprising.

[14] Beyond 2 m, the number of addresses begins to increase at a rate depending upon the locality. For estimates at low elevations, e.g., 3−5 m, we suggest taking advantage of the linearity of the cumulative curve and extrapolating back from estimates at higher elevations. While it is apparent that detailed elevation adjustments vary regionally, the above findings suggest if no elevation adjustment is undertaken for the direct SRTM-based estimate at higher

elevations, the exposure result can be regarded as a plausible lower bound; in other words less than the number of addresses impacted by ground-level flooding, for example.

[15] Based on the above regional validations, we now report national results. All G-NAF addresses within 3 km of the shorelines were first selected. Using either the full resolution SRTM-3″ or aggregated SRTM-30″ elevation surface, similar cumulative percentages of G-NAF at various elevation levels were estimated (Figure 4). This comes as no surprise given the spatial aggregation was done on continuous, as opposed to categorical, surfaces such as elevation, and with the large number of samples (addresses), errors are averaged out. If we apply the SRTM height adjustments of 2−4 m, the percentages of vulnerable addresses at elevation less than 5 m range between 4.6%
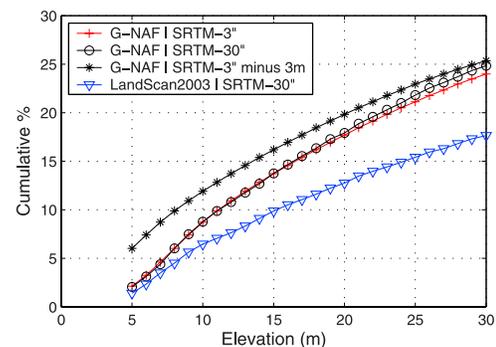

Figure 4. Cumulative percentages of G-NAF addresses and LandScan2003 population within 3 km of detailed shorelines in Australia in relation to two elevation surfaces: SRTM-3″ and SRTM-30″. Only estimates at larger elevation levels (≥5 m) are shown.

and 7.4%. However, we do not regard the height adjustments of 2 m observed from very low-lying areas such as Gold Coast and 4 m from the densely populated Sydney Basin as being typical for the entire country. An adjustment of 3 m results in 6.0% of national addresses located within 3 km of the shoreline and with elevation below 5 m.

[16] As far as the spatial distribution of such addresses is concerned, the majority are adjacent to sea-connected coastal waters - alongside lakes or lagoons, river banks and estuaries, rather than directly facing the open ocean. This finding is useful in directing mitigation measures and looking at the vulnerability of critical assets in relation to significant sea-level rise.

[17] For all LandScan2003 population within 3 km of the shorelines, the estimated percentages of vulnerable population are smaller than those of G-NAF addresses, reflecting again the lower-bound estimation of coastal population obtained with LandScan2003. The underestimation becomes larger as elevation gets higher (e.g., >10 m) (Figure 4). Our validations in the Australian region suggest that the combination of LandScan2003, GSHHS shorelines and SRTM at 30″-resolution only provides lower-bound estimates of vulnerable coastal population.

5. Discussion and Conclusion

[18] Detailed geospatial data and comprehensive validations undertaken in this study provide evidence for reliable and high-resolution estimates of the coastal population. The results are useful for benchmarking and monitoring decadal changes in coastal populations, and similar analyses can be pursued for any coastal region of interest that could be or has been impacted by natural catastrophes such as tsunamis or storm surges [e.g., *Ward and Day*, 2001; *Travis*, 2005]. Our findings on coastal population down to 1 km horizontal resolution and elevation down to 5 m level start to provide concrete material for improved understanding of the potential impact of, say, changes in the global mean sea level [*Kerr*, 2006; *Overpeck et al.*, 2006]. More importantly, the validation methodology reported here directly contributes toward the effort of finding high-resolution estimates at the global level [e.g., *Small et al.*, 2000].

[19] On a more general level, there is often a need to quantify regional or global environmental attributes using data sets with inherent resolution limitations. In such cases, spatial scaling through empirical validation tests shown here offers the only cost-effective solution. Validation methodology should reveal the process of the scaling employed. In this study, we established statistical and spatial correlations between the most detailed representation of address-level exposure data and population surfaces depicted by LandScan2003. By delineating socially important objects (e.g., individual dwellings or addresses, population density and land use), up-scaling can be performed on "true" spatial units of analysis (SUAs), rather than "artificial" SUAs such as administrative or census units (Figure S5). It is spatial

contextures that should be explicitly coupled during the prescription of valid SUAs and then the nature of spatial data that exist at multiple resolutions can be respected in aggregation and scaling [*Chen*, 2003]. For estimates at gross levels, it is not always necessary to increase the spatial resolution of the underlying data. The conceptual emphasis on "objects" promotes best scientific practice of spatial scale analysis and is of significance to other fields such as biology, ecology, cosmology and physics.

[20] Sufficiently understanding the human dimension of natural disasters requires detailed accounting of the number and spatial distribution of lives and assets at risk. This increasingly important task is exactly complementary to scientific endeavours to improve our knowledge about the physics of hazard agents. Much more effort should be expanded to overcome limitations of global observational data, including limited coverage and access, coarse resolution and inconsistency across countries. Our study reported here takes advantage of regional high-resolution data sets and develops efficient scaling approaches to derive gross estimates of vulnerable coastal populations. This information is invaluable for the public and high-level policy makers confronting environmental challenges.

References

Alexander, D. (2000), *Confronting Catastrophe: New Perspectives on Natural Disasters*, Oxford Univ. Press, New York.

Chen, K. (2003), Scale analysis for remotely sensed images and areal census data: Stressing objects, paper presented at 30th International Symposium on Remote Sensing of Environment, Int. Comm. Remote Sens. Environ., Honolulu, Hawaii, 10–14 Nov.

Church, J. A., and N. J. White (2006), A 20th century acceleration in global sea-level rise, *Geophys. Res. Lett.*, *33*, L01602, doi:10.1029/2005GL024826.

Grossi, P., and H. Kunreuther (Eds.) (2005), *Catastrophe Modeling: A New Approach to Managing Risk*, Springer, New York.

Kerr, R. A. (2006), Climate change: A worrying trend of less ice, higher seas, *Science*, *311*, 1698–1701.

Longley, P., M. F. Goodchild, D. J. Maguire, and D. W. Rhind (2001), *Geographical Information Systems and Science*, John Wiley, Hoboken, N. J.

Marris, E. (2005), Inadequate warning system left Asia at the mercy of tsunami, *Nature*, *433*, 3–5.

Overpeck, J. T., B. L. Otto-Bliesner, G. H. Miller, D. R. Muhs, R. B. Alley, and J. T. Kiehl (2006), Paleoclimatic evidence for future ice-sheet instability and rapid sea-level rise, *Science*, *311*, 1747–1750.

Small, C., V. Gornitz, and J. E. Cohen (2000), Coastal hazards and the global distribution of population, *Environ. Geosci.*, *7*, 3–12.

Travis, J. (2005), Scientists' fears come true as hurricane floods New Orleans, *Science*, *309*, 1656–1659.

Ward, S. N., and S. Day (2001), Cumbre Vieja Volcano—Potential collapse and tsunami at La Palma, Canary Islands, *Geophys. Res. Lett.*, *28*, 3397–3400.

————————

K. Chen and J. McAneney, Risk Frontiers–Natural Hazards Research Centre, Macquarie University, Building E7A, North Ryde, NSW 2109, Australia. (kchen@els.mq.edu.au)