On the optimal amplitude thresholds for tsunami warning

Diana J. M. Greenslade and Stewart C. R. Allen

October 2019
On the optimal amplitude thresholds for tsunami warning

Diana J. M. Greenslade and Stewart C. R. Allen

Bureau Research Report No. 039

October 2019
Enquiries should be addressed to:

Lead Author: Diana J.M. Greenslade

Bureau of Meteorology
GPO Box 1289, Melbourne
Victoria 3001, Australia

diana.greenslade@bom.gov.au:

Copyright and Disclaimer

© 2016 Bureau of Meteorology. To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of the Bureau of Meteorology.

The Bureau of Meteorology advise that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law and the Bureau of Meteorology (including each of its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.
Contents

Executive Summary ........................................................................................................... 1

1. Introduction .................................................................................................................... 2

2. Method ........................................................................................................................... 3
   2.1 New events ................................................................................................................. 4
   2.2 Scenario selection .................................................................................................... 6
   2.3 Alternative predictors ............................................................................................. 7
   2.4 New objective measures for threshold optimisation .................................................. 8

3. Results ............................................................................................................................. 10
   3.1 Australian mainland zones ....................................................................................... 13
   3.2 Offshore zones ......................................................................................................... 15
   3.3 Other predictors ........................................................................................................ 16

4. Discussion ....................................................................................................................... 19

5. Summary ......................................................................................................................... 20

6. Acknowledgements ........................................................................................................ 21

7. References ...................................................................................................................... 21
List of Figures

Figure 1: Australian coastal and offshore zones. Offshore zones drawn larger than to scale, in order to improve their visibility on the figure. ................................................................. 2

Figure 2: Desired warning schemes for the Chile 2010 (left) and Japan 2011 (right) events. Coastal zones for which a Marine Threat was deemed necessary are indicated in blue. No Land Threats were deemed necessary................................................................. 5

Figure 3: Example of a categorical performance diagram after Roebber (2009). The Success Ratio and Probability of Detection are plotted on the horizontal and vertical axes respectively, the Bias is given by the dashed lines and the Threat Score by the curved lines. Better forecasts are towards the top-right of the diagram. The symbols represent the scores for a range of different thresholds, as indicated by the colour bar. The thresholds with the highest Threat Score (20 cm and 20.5 cm) are indicated by the diamond. Note that some symbols may be overlaid, since successive thresholds may have the same score......................... 11

Figure 4: 95th percentile values of $H_{\text{max}}$ (i.e. P95) in each of the coastal zones for all events considered in AG10. Offshore zones are towards the right-hand side of the figure. For each event, Marine Threats were desired in the zones marked by blue symbols. The existing thresholds are marked by the dashed lines. ................................................................. 12

Figure 5: Threat Scores as a function of Marine Threat threshold for the data in Figure 4. The optimal threshold on the basis of maximum Threat Score is indicated by the dotted lines. 13

Figure 6: 95th percentile values of $H_{\text{max}}$ in each of the mainland coastal zones for all significant events considered in this study. The thresholds used in AG10 are shown by the solid lines and the lowest optimal threshold from this study based on maximum Threat Score is shown by the dashed line. ................................................................. 14

Figure 7: Threat Scores as a function of Marine Threat threshold for the data in Figure 6. The (lowest) optimal threshold, on the basis of maximum Threat Score, is indicated by the vertical dashed line. ................................................................. 14

Figure 8: Same as Figure 6 but for offshore zones ................................................................. 15

Figure 9: Same as Figure 7 but for offshore zones ................................................................. 16
EXECUTIVE SUMMARY

The Joint Australian Tsunami Warning Centre (JATWC) threat assessments are derived from a database consisting of more than two thousand pre-computed scenarios of tsunami propagation across the Indian and Pacific Ocean basins. Tsunami warnings are issued for pre-defined coastal warning zones. In the event of a potentially tsunamigenic earthquake, the 'closest' scenario is selected and predicted tsunami amplitudes within each zone are assessed to see if they exceed particular pre-determined warning thresholds. This guides the decision on whether a Land Threat, Marine Threat, or No Threat bulletin is issued. In the present work, the Marine Threat threshold is re-examined in the light of 8 recent tsunami events and their known impacts. In addition, a more robust and quantitative assessment technique is used to determine the optimal warning thresholds, and a number of alternative proxies for coastal impact are considered.
1. **INTRODUCTION**

The Joint Australian Tsunami Warning Centre (JATWC) was formally established in 2008 to provide Australia with the capability to warn the community of potential tsunami threats to Australian coastal locations and offshore territories. JATWC tsunami threat assessments are derived from the 'T2' scenario database consisting of more than two thousand pre-computed scenarios of tsunami propagation. Each scenario is associated with a specific hypothetical earthquake along subduction zones within the Indian, Pacific and South Atlantic Oceans (Greenslade et al. 2009; Greenslade et al. 2010; Simanjuntak et al. 2011).

Tsunami warnings are issued to the community within pre-defined Australian Bureau of Meteorology coastal warning zones. These cover the Australian mainland and offshore territories (Macquarie Island, Lord Howe Island, Norfolk Island, Willis Island, Christmas Island and Cocos (Keeling) Islands). Figure 1 shows the boundaries of the mainland and offshore zones. These were not designed specifically for tsunami warning, however the advantage of using these pre-existing zones is that forecasters, emergency managers and the general public are familiar with them as they are used for other marine forecast and warning products. The mainland coastal zones extend 60 nautical miles (approximately 110 km) offshore while the offshore zones extend 20 arc minutes offshore (approximately 35-40 km) around islands.

![Figure 1 Australian coastal and offshore zones. Offshore zones drawn larger than to scale, in order to improve their visibility on the figure.](image-url)
In event of a potentially tsunamigenic earthquake, the 'closest' T2 scenario is selected (see Section 2.2 for details) and predicted tsunami amplitudes within the zones are assessed to see if they exceed particular pre-determined warning thresholds. This guides the decision on whether a Land Threat, Marine Threat, or No Threat Bulletin is issued. A Marine Threat warns of potentially dangerous rips, waves and strong ocean currents in the marine environment and the possibility of only some localised overflow onto the immediate foreshore. A Land Threat is a warning for low-lying coastal areas of major land inundation, flooding, dangerous rips, waves and strong ocean currents.

The derivation of the warning thresholds was undertaken through analysis of observed impacts for previous tsunami events (up until 2009) and is described in Allen and Greenslade (2010, hereafter AG10). Since that work was undertaken, several significant tsunami events have occurred. In the present work, this analysis is revisited with recent events incorporated in order to assess the warning thresholds and determine if any changes are required. In addition, a number of different proxies for tsunami impact are investigated, and new techniques for objectively and quantitatively determining the optimal warning thresholds are introduced.

2. METHOD

The approach of AG10 for the determination of the optimal thresholds is briefly outlined here. For a number of historical events, evidence of impacts on the Australian coast was collated. This was generally through newspaper articles, personal communications with harbourmasters, anecdotal evidence etc. Based on these known impacts, the most appropriate warning level was allocated to each zone. For example, if an event was known to have caused some dangerous currents within a harbour but no inundation, then a Marine Threat was allocated to that zone. This produced a 'desired' warning scheme for each historical event. The closest T2 scenario (closest in space and magnitude) for each event was then identified and scaled as appropriate (see Section 2.2). Maximum amplitudes (over time) from that scenario at each grid point within each coastal zone were ranked and the 95th percentile value (P95) was extracted. These P95 values were then plotted for all events and all zones and used to select the threshold that most closely produced the desired warning schemes (see Figure 8 in AG10). This resulted in the threshold for Marine Threats being set to 20 cm. Since none of the events used produced any widespread inundation that would have required a Land Threat, the threshold for this level of warning was set to be just larger than the highest calculated P95 value of all the events, at 55 cm. This is because there was "no evidence to suggest that an event with amplitude above this threshold would not need a land warning" (AG10). In summary, P95 was used as the 'predictor' and the threshold levels for this predictor were set to be 20 cm for a Marine Threat and 55 cm for a Land Threat. These values were used for both mainland zones and offshore zones.
In addition to the inclusion of recent events, a number of other extensions to the method are explored. These are outlined in the following sections.

### 2.1 New events

In the present analysis, we are now able to consider a number of more recent events. The full set of events considered is listed in Table 1, with new events added since the last assessment indicated.

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Date and Time</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Moment Magnitude</th>
<th>Warnings?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chile 1960</td>
<td>22 May 1960 19:11</td>
<td>39° 30' S</td>
<td>74° 30' W</td>
<td>9.5</td>
<td>Yes</td>
</tr>
<tr>
<td>Sumatra 2004</td>
<td>26 Dec 2004 00:59</td>
<td>3° 18' N</td>
<td>95° 47' E</td>
<td>9.1</td>
<td>Yes</td>
</tr>
<tr>
<td>Tonga 2006</td>
<td>3 May 2006 15:27</td>
<td>19° 54' S</td>
<td>174° 12' E</td>
<td>7.9</td>
<td>No</td>
</tr>
<tr>
<td>Java 2006</td>
<td>17 Jul 2006 08:19</td>
<td>9° 18' S</td>
<td>107° 18' E</td>
<td>7.7</td>
<td>Yes</td>
</tr>
<tr>
<td>Solomons 2007</td>
<td>1 Apr 2007 20:40</td>
<td>8° 29' S</td>
<td>156° 59' E</td>
<td>8.1</td>
<td>Yes</td>
</tr>
<tr>
<td>Sumatra 2007</td>
<td>12 Sep 2007 11:10</td>
<td>4° 31' S</td>
<td>101° 23' E</td>
<td>8.4</td>
<td>Yes</td>
</tr>
<tr>
<td>Puysegur 2007</td>
<td>30 Sep 2007 05:23</td>
<td>49° 23' S</td>
<td>134° 01' E</td>
<td>7.4</td>
<td>No</td>
</tr>
<tr>
<td>Puysegur 2009</td>
<td>15 Jul 2009 09:22</td>
<td>45° 58' S</td>
<td>166° 28' E</td>
<td>7.8</td>
<td>No</td>
</tr>
<tr>
<td>Samoa 2009</td>
<td>29 Sep 2009 17:48</td>
<td>15° 34' S</td>
<td>172° 4' W</td>
<td>8.1</td>
<td>No</td>
</tr>
<tr>
<td>Chile 2010</td>
<td>27 Feb 2010 06:34</td>
<td>35° 51' S</td>
<td>72° 43' W</td>
<td>8.8</td>
<td>Yes</td>
</tr>
<tr>
<td>Japan 2011</td>
<td>11 Mar 2011 05:46</td>
<td>38° 19' N</td>
<td>142° 22' E</td>
<td>9.1</td>
<td>Yes</td>
</tr>
<tr>
<td>Solomons 2013</td>
<td>6 Feb 2013 01:12</td>
<td>11° 13' S</td>
<td>164° 44' E</td>
<td>7.9</td>
<td>No</td>
</tr>
<tr>
<td>Chile 2014</td>
<td>1 Apr 2014 23:46</td>
<td>19° 38' S</td>
<td>70° 49' W</td>
<td>8.2</td>
<td>No</td>
</tr>
<tr>
<td>Chile 2015</td>
<td>16 Sep 2015 22:54</td>
<td>31° 34' S</td>
<td>71° 39' W</td>
<td>8.3</td>
<td>No</td>
</tr>
<tr>
<td>Solomons 2016</td>
<td>8 Dec 2016 17:39</td>
<td>10° 41' S</td>
<td>161° 20' E</td>
<td>7.8</td>
<td>No</td>
</tr>
<tr>
<td>Mexico 2017</td>
<td>8 Sep 2017 04:49</td>
<td>15° 4' N</td>
<td>93° 43' W</td>
<td>8.1</td>
<td>No</td>
</tr>
</tbody>
</table>
Note that all of the new events have occurred in the Pacific Ocean basin - there were no suitable Indian Ocean events in this period. Moreover, the Sumatra event of 2005 that was originally included in AG10 has been excluded from this analysis. The rationale for this relates to the tsunamigenic potential of that particular earthquake. Most of the slip for that event occurred under nearby islands or in shallow water, so the earthquake was not able to uplift a large amount of water (Gahalaut and Catherine, 2006). This is quite different to the characteristics of the relevant ‘closest’ T2 scenario, for which most of the slip occurs in deep water and a large tsunami is predicted. We do not wish to tune our thresholds to include events for which there is no appropriate analogy in the T2 database, so the Sumatra 2005 event is excluded from this analysis. See Section 4 for further discussion of this issue.

Desired warning schemes for the first eight events in Table 1 are described in AG10 and are not reproduced here. Of the eight events since 2009, there are only two for which, in hindsight, warnings are deemed necessary for the Australian coastline: the Chile 2010 and Japan 2011 events. The desired warning schemes for these events are shown in Figure 2. To reiterate, these schemes are hindcasts which reflect what warnings (if any) should have been issued for the Australian coastal zones. These were determined in consultation with staff from the Joint Australian Tsunami Warning Centre, by considering all the

Figure 2: Desired warning schemes for the Chile 2010 (left) and Japan 2011 (right) events. Coastal zones for which a Marine Threat was deemed necessary are indicated in blue. No Land Threats were deemed necessary.
known impacts that arose from the events. A summary of these impacts has been collated by the JATWC (Bureau of Meteorology, 2019). As an example, a report from Merimbula, on the far south coast of NSW states that: "On Saturday 12th March 2011 at 2:00pm, coinciding with the high tide, a significant surge of water, 0.6 to 1 metre above normal, was experienced over a two minute period causing several swimmers to be washed into the lagoon 500 metres away. The current generated chop of approximately 30cm in height." Based on this report, a marine threat is assigned to this zone for this event.

### 2.2 Scenario selection

The T2 scenarios are computed with fixed Mw values of 7.0, 7.5, 8.0, 8.5 and 9.0. When a tsunamigenic earthquake occurs, linear scaling is applied to the amplitudes from the relevant T2 scenario (or scenarios – see below) to match the Mw of the event. In operations, the scaling is limited to no more than ± 0.2 Mw away from the magnitude of the scenario, i.e. to obtain a forecast for a Mw = 8.8 event, the standard operating procedure is to scale down the closest Mw = 9.0 scenario, whereas a Mw = 8.7 event would be scaled up from the closest Mw = 8.5 scenario. Furthermore, for Mw ≥ 7.8, the operational procedure does not use the single T2 scenario with the closest epicentre to the event, but uses an 'envelope' of scenarios, to take account of the uncertainty in rupture direction in real-time. In order to maintain conservatism, warnings are issued according to the worst case of all of the possible scenarios in the envelope. Given the structure of the T2 scenarios (see Greenslade et al., 2009), this means that an envelope of 2 adjacent scenarios is used for a Mw = 8.0 event and also for events scaled from 8.0, i.e. for events with 7.8 ≤ Mw ≤ 8.2; up to 4 adjacent scenarios are used for events with 8.3 ≤ Mw ≤ 8.7 and up to 10 adjacent scenarios for Mw ≥ 8.8. Single scenarios are used for Mw ≤ 7.7.

For the purpose of deriving the optimal warning thresholds, there are therefore a number of possible approaches to selecting the appropriate T2 scenario(s) for each event. These are:

1) Use the worst-case of the envelope of scenarios that would have been used in real-time - this is the approach that is closest to operational usage;
2) Use the single best scenario that would have been selected according to operational scaling protocols, with hindsight knowledge of the event by using available seismic data and sea-level observations; or
3) Use the best matching T2 scenario, even if this is outside operational scaling protocols. For example, Allen and Greenslade (2016) showed that the T2 scenario that best matched the Mw = 9.1 Japan 2011 event was a Mw = 8.5 scenario scaled up by 0.6 Mw to 9.1, rather than a Mw = 9.0 scenario scaled up by 0.1 Mw.
The advantage of approach (3) is that it provides the very best possible estimate of amplitudes in the coastal zone. However, it results in amplitudes that are not actually possible to obtain with the operational system because the operational system has constraints on the scaling that can be applied. So the thresholds that would be derived in this manner would not be applicable to the operational system.

Approach (1) would ensure that the thresholds are tuned according to the actual amplitudes in the coastal zone that would be used operationally. This would therefore find the best tuning and best threshold for the current operational system. The disadvantage of this is that it develops a tuning that incorporates all the errors in the system and is not ‘future-proof’. These errors may well evolve and reduce in time, which would require the thresholds to be re-evaluated. For example, the JATWC is currently developing a scenario selection technique that incorporates the earthquake centroid in addition to the epicentre. This provides some information on the rupture direction and would likely reduce the number of scenarios selected in the ‘envelope’ and thus change the value of the amplitudes in the coastal zone. This means that the derived thresholds would no longer be relevant.

So, we want to use the best possible amplitudes in the coastal zone that can be obtained with the current operational system. This means we should use the best matching T2 scenario to the event, but only consider the scaled-scenario options that are within current operational constraints. This will find the threshold that should be applied to the operational scenarios given no other errors in the system, such as magnitude errors, rupture direction uncertainty etc. Therefore, in this study, we use approach (2) above, i.e., the best matching scenario within operational constraints.

### 2.3 Alternative predictors

A key aspect of this work is the use of tsunami amplitudes within the coastal zones as a proxy for coastal impacts. Ideally, forecasts of coastal impacts and resulting warnings would be based on inundation modelling, but this is not currently incorporated in the operational system. In the initial development of this threshold technique, Allen and Greenslade (2008) used the single maximum amplitude (over both time and space) within each coastal zone as the proxy for coastal impact, i.e., as the predictor. In AG10, this was re-assessed based on the fact that the single maximum value within a coastal zone was often an outlier and not necessarily representative of all the amplitudes within the coastal zone. AG10 considered the maximum amplitude ($H_{\text{max}}$) at each model grid point (where maximum is determined over the 24 hours of the T2 scenario), ranked these values within the coastal zone and selected the 95th percentile value from the resulting distribution, i.e., $P_{95}$, was the predictor. 90th percentile values were also considered but did not produce results that were substantially different.
In the present study, the choice of predictor is revisited, and a number of other possible predictors are examined. These include:

- Other percentiles of $H_{\text{max}}$: $100^{\text{th}}$ (i.e. the maximum), $90^{\text{th}}$, $50^{\text{th}}$ (median)
- Only using $H_{\text{max}}$ from grid points in the coastal zone with depths less than 100m, in order to ensure that the amplitudes more closely represent coastal values, noting that the T2 scenarios are limited to depths greater than 20m.
- Applying a normalisation factor to the $H_{\text{max}}$ values to account for tsunami wavelength, where the normalisation factor is a function of earthquake rupture width. This is an attempt to address the fact that waves of different wavelength but the same amplitude are expected to have different coastal impacts (e.g. Charvet et al., 2013).
- Various combinations of these options

### 2.4 New objective measures for threshold optimisation

In AG10, the threshold for Marine Threats was determined somewhat subjectively, by inspection. In this study we take a more objective and quantitative approach through the use of contingency tables and derived metrics. Contingency tables provide a framework for describing whether an event is forecast or not, and whether it eventuated or not. In the context of our study, the 'events' are the Threat levels for each zone and each historical tsunami, and they are either 'forecast' (i.e. Threat level was predicted according to the range of predictors described in Section 2.3) or they were 'desired' (i.e. the Threat level was desired according to knowledge of the impacts as described in Section 2.1). Consequently, a contingency table can be devised that allows us to assess the overall forecast for each predictor:

<table>
<thead>
<tr>
<th>Desired warning?</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast warning?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Hit</td>
<td>False Alarm</td>
<td>Forecast yes</td>
</tr>
<tr>
<td>No</td>
<td>Miss</td>
<td>Correct Negative</td>
<td>Forecast no</td>
</tr>
<tr>
<td>Total</td>
<td>Observed yes</td>
<td>Observed no</td>
<td>Total</td>
</tr>
</tbody>
</table>

We can evaluate the quality of the forecast by counting the scores in each category of the table and using the counts to calculate various measures.

The most obvious measure is to determine what fraction of forecasts are correct. This is the *Accuracy*, i.e. how well do the corresponding ‘yes’ and ‘no’ categories match:
ON THE OPTIMAL AMPLITUDE THRESHOLDS FOR TSUNAMI WARNING

\[
Accuracy = \frac{\text{Hits} + \text{Correct Negatives}}{\text{Total}}
\]

This measure ranges from zero to one, with a score of one representing a perfect forecast, that is, there are no misses or false alarms.

This measure, though simple and intuitive is often biased by the most common category, which is usually the ‘correct negative’ category. This is particularly the case in the present context, because all of the events considered tend to impact a relatively short portion of Australia's long coastline, if at all. This means that there are many zones where no warning is required and no warning is predicted. An alternative approach excludes ‘correct negatives’ from our calculation of forecast accuracy, giving a measure known as the Threat Score:

\[
\text{Threat Score} = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{false alarms}}
\]

In effect, the Threat Score only considers forecasts that are important. It is more conservative, being sensitive to variations in the number of hits, misses and false alarms. Like the Accuracy, it has a range from zero to one, with a perfect forecast given by a value of one. Note that sometimes the Threat Score is also known as the ‘critical success index’.

Another useful measure is the Bias:

\[
\text{Bias} = \frac{\text{hits + false alarms}}{\text{hits + misses}}
\]

This is the ratio of the frequency of forecast events to the frequency of observed events and indicates whether the forecast system has a tendency to under-forecast (Bias < 1) or over-forecast (Bias > 1) events. In contrast to the Threat Score, the Bias does not measure how well the forecast corresponds to the observations, it only measures relative frequencies. A score of one is ideal.

Other similar measures commonly used include the Probability of Detection (POD):

\[
\text{POD} = \frac{\text{hits}}{\text{hits + misses}}
\]

which measures fraction of the observed "yes" events that were correctly forecast, and the Success Ratio (SR):
\[ SR = \frac{\text{hits}}{\text{hits} + \text{false alarms}} \]

which measures the fraction of the forecast "yes" events that were correctly observed. In both cases, the measures have a range from zero to one, with one representing a perfect forecast. The SR can be expressed in terms of the more familiar False Alarm Ratio (FAR) as \( SR = 1 - \text{FAR} \).

Both the Threat Score and the Bias can be written in terms of the POD and the SR as:

\[
\text{Bias} = \frac{\text{POD}}{SR} \\
\text{Threat Score} = \frac{1}{\frac{1}{SR} + \frac{1}{\text{POD}} - 1}
\]

Given this relationship, these four measures can all be plotted on a diagram known as the categorical performance diagram (Roebber, 2009), an example of which is shown in Figure 3. In this diagram, the SR and POD are plotted on the horizontal and vertical axes respectively, the Bias is given by the dashed lines and the Threat Score by the curved contour lines. The Bias = 1 line (representing an unbiased forecast) is at 45° (from the bottom-left to top-right) and values of the Threat Score indicating a better forecast are closer to the top-right.

**3. RESULTS**

In this section, we focus specifically on the Marine Threat threshold, predominantly because, as in AG10, there are no recent events for which a Land Threat was either issued or desired for the Australian region. The results of AG10 were first reproduced in order to check the appropriateness of the existing Marine Threat threshold using this more objective technique. Figure 4 shows a reproduction of Figure 8 in AG10. Each zone is shown along the horizontal axis, and the vertical axis designates the P95 value. For each event shown in the legend, the P95 value of the relevant T2 scenario (see Section 2.2) is plotted for each zone, and coloured blue if a Marine Threat was desired, or black if No Threat was desired (according to Section 2.1). Note that this figure is not quite identical to Figure 8 in AG10 due to some minor changes to the number and extent of the coastal zones in the intervening period.

The various metrics described in Section 2.4 were calculated for warning thresholds ranging from 10 cm to 30 cm, at 0.5 cm increments. The Categorical Performance Diagram based on these metrics is the one shown in Figure 3.
ON THE OPTIMAL AMPLITUDE THRESHOLDS FOR TSUNAMI WARNING

Figure 3: Example of a categorical performance diagram after Roebber (2009). The Success Ratio and Probability of Detection are plotted on the horizontal and vertical axes respectively, the Bias is given by the dashed lines and the Threat Score by the curved lines. Better forecasts are towards the top-right of the diagram. The symbols represent the scores for a range of different thresholds, as indicated by the colour bar. The thresholds with the highest Threat Score (20 cm and 20.5 cm) are indicated by the diamond. Note that some symbols may be overlaid, since successive thresholds may have the same score.

It can be seen from Figure 3 that the thresholds that give the best POD (towards the top of the diagram) are the lowest values of the thresholds – this is not surprising as a lower threshold will naturally result in more warnings being issued and thus a higher POD. Conversely, the best SR (towards the right of the diagram) is given by the higher thresholds – as the threshold increases, this will reduce the number of false alarms and so
the SR increases. The best Bias (=1.0) is given by a threshold of 21cm but this value does not provide the best Threat Score. The best Threat Score is given by thresholds of 20cm and 20.5cm. We note here that a Bias > 1 is preferred to a Bias < 1, e.g. Bias = 1.1 is preferred to Bias = 0.9 because if it is a choice between issuing a warning when not required or missing a warning when it was required, then from a perspective of minimising damage and lives lost, false alarms are preferred to misses.

In the remainder of this work, the threshold will be optimised with respect to Threat Score while also considering the Bias. This is based on the fact that Threat Score represents a balance between ensuring events are not missed, while not issuing an excessive number of false alarms. As discussed above, it would also be possible to optimise with respect to POD or SR, however, this would merely result in raising the threshold until POD is maximised (with SR reduced and many subsequent false alarms) or lowering the threshold until SR is maximised (with POD reduced and many subsequent missed events).

A diagram showing only Threat Score as a function of threshold is shown in Figure 5. It can be seen from this figure that on the basis of Threat Score, and using P95 as the predictor, the optimal Marine Threat threshold for this set of events is 20 cm or 20.5 cm. A further decision made here is that if a range of thresholds provide the same Threat
Figure 5: Threat Scores as a function of Marine Threat threshold for the data in Figure 4. The optimal threshold on the basis of maximum Threat Score is indicated by the dotted lines.

Score, we use the lowest value as it represents the conservative option. So here, we choose 20cm as the optimal warning threshold. This is exactly the same value that was previously determined by AG10 using a more subjective technique. Note that this analysis includes both mainland and offshore zones – this is discussed in more detail below.

3.1 Australian mainland zones

The analysis is now extended to include all events listed in Table 1. A further change is that the offshore zones are excluded and will be analysed separately.

A comparison of the desired warnings and predicted warnings for the same predictor, i.e. P95 is shown in Figure 6. This includes all events in Table 1 but with very small events omitted. The Threat Score as a function of threshold for this case is shown in Figure 7 and indicates that any threshold from 18 cm to 20.5 cm is optimal. The previous Marine Threat threshold of 20 cm is indicated on Figure 6, as well as the lowest of the new optimal thresholds (18 cm). This is encouraging and demonstrates that even in the light of these recent events the selected threshold of 20 cm (using P95) is appropriate.
Figure 6: 95th percentile values of $H_{\text{max}}$ in each of the mainland coastal zones for all significant events considered in this study. The thresholds used in AG10 are shown by the solid lines and the lowest optimal threshold from this study based on maximum Threat Score is shown by the dashed line.

Figure 7: Threat Scores as a function of Marine Threat threshold for the data in Figure 6. The (lowest) optimal threshold, on the basis of maximum Threat Score, is indicated by the vertical dashed line.
It should be noted that the Threat Scores here are lower than in Figure 5. This is because there are an increased number of 'misses' in the present dataset. Specifically, there are forecasts for Chile 2010 and Japan 2011 which were deemed to have required a Marine Threat, but which have very low values of P95, so they result in 'No threats' for all thresholds > 10cm. These 'misses' are potentially a concern as it means that there would not be a warning for impacts such as described in Section 2.1 at Merimbula. This is discussed further in Section 3.3 below.

### 3.2 Offshore zones

The analysis for the offshore zones was conducted separately. This is because they generally have quite different bathymetric profiles to the mainland zones, leading to different tsunami dynamics in the coastal zone. Furthermore, because of this different bathymetric profile, they have different coastal zone designs, with a shorter offshore extent, as noted in Section 1. Results are shown in Figure 8 and Figure 9. It can be seen that using the existing threshold of 20 cm for the offshore territories will result in a number of misses, which could be avoided with a lower threshold. The optimal threshold for offshore zones is found to be between 9.5 cm and 10.5 cm, inclusive.

![Figure 8: Same as Figure 6 but for offshore zones](image-url)
3.3 Other predictors

The results discussed in the previous sections have used P95 the predictor. As noted in Section 3.1, the Threat Score using this predictor and the full set of events is lower than the Threat Score with the initial set of 8 events. It is therefore worth considering whether there are alternate predictors that might produce a higher Threat Score and improved tsunami warnings.

Optimal threshold values (i.e. those which maximised the Threat Score) were therefore determined for a number of the other possible predictors described in Section 2.3. The results for mainland and offshore zones are shown in Table 2. The predictors marked "depth limited" are those for which only grid points in water depths less than 100m were used. "Normalised by width" indicates that the amplitudes were multiplied by a scaling factor based on the width of the earthquake rupture, such that amplitudes for tsunamis generated by earthquakes with narrower ruptures were increased relative to those with wider ruptures. "Reverse normalised by width" indicates the inverse of this, that is, amplitudes for tsunamis generated by earthquakes with narrower ruptures were decreased relative to those with wider ruptures. In most cases, it can be seen that there was a range of possible threshold values which produced the same Threat Score.

For mainland zones, the Threat Scores for most of the predictors are very close to, or slightly better than for P95. The best predictor is provided by the 100th percentile value,
i.e. the maximum value, which increases the Threat Score from 0.65 to 0.69. In this case, the Bias is also increased from 0.7 to 0.8. However, using 100th percentile for the offshore zones results in a degradation in both the Threat Score (0.73 to 0.56) and in the Bias (1.11 to 0.56).

For the offshore zones, the largest improvement in the Threat Score can be obtained by considering the 50th percentile (0.73 to 0.82), but this degrades the Bias (1.11 to 1.22). The best predictor for the offshore zones is the 90th percentile normalised by width, which increases the Threat Score to 0.8 while giving a Bias of 1.0. However, for mainland zones this does not provide any improvement to either the Threat Score or the Bias.

There is no single best option that improves both the Threat Score and the Bias for both the mainland zones and the offshore zones. However, as noted in the previous paragraph, the 90th percentile normalised by width, provides an improvement for the offshore zones while giving the same results for mainland zones.

For simplicity of operational implementation, there are benefits to having the same technique and even the same thresholds for offshore and for mainland zones. However, it seems reasonable to also consider different predictors for mainland and offshore zones particularly if gains in skill can be made. In that case, the 100th percentile could be considered for the mainland zones and the 90th percentile normalised by width for the offshore zones.

This work has concentrated on establishing the best option for the Marine Threat thresholds. As in AG10, no information is available to guide the magnitude of the Land Threat threshold as none of the events to date have produced any impacts that would need to be classified as a Land Threat. Note that this can be addressed through the application of inundation modelling (see Greenslade et al., submitted). For the current work, we apply the principle of AG10, which is that the Land Threat threshold is set just above the highest Marine Threat values, in order to maintain conservatism.
Table 2: Results for mainland and offshore zones for a range of proposed predictors. Optimal thresholds were determined through maximising the Threat Score.

<table>
<thead>
<tr>
<th></th>
<th>Mainland zones</th>
<th>Offshore zones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Optimal threshold (cm)</td>
<td>Threat Score</td>
</tr>
<tr>
<td>95th percentile (P95: control)</td>
<td>18 – 20.5</td>
<td>0.65</td>
</tr>
<tr>
<td>90th percentile</td>
<td>16 – 17.5</td>
<td>0.65</td>
</tr>
<tr>
<td>50th percentile</td>
<td>7.5</td>
<td>0.57</td>
</tr>
<tr>
<td>100th percentile (i.e. maximum value)</td>
<td>31.5 - 32</td>
<td>0.69</td>
</tr>
<tr>
<td>95th percentile [depth limited]</td>
<td>25.5 – 26.5</td>
<td>0.67</td>
</tr>
<tr>
<td>95th percentile [normalised by width]</td>
<td>18 – 20.5</td>
<td>0.65</td>
</tr>
<tr>
<td>95th percentile [reverse normalised by width]</td>
<td>36 – 41.5</td>
<td>0.65</td>
</tr>
<tr>
<td>90th percentile [normalised by width]</td>
<td>16 – 17.5</td>
<td>0.65</td>
</tr>
<tr>
<td>100th percentile [normalised by width]</td>
<td>31.5 - 32</td>
<td>0.67</td>
</tr>
<tr>
<td>95th percentile [normalised by width] and [depth limited]</td>
<td>25.5 – 26.5</td>
<td>0.67</td>
</tr>
</tbody>
</table>
In summary, there are a few options that could be considered. Values in square brackets are the Threat Scores for mainland zones and offshore zones respectively.

- [0.65, 0.73] Continue to use the 95\textsuperscript{th} percentile for both mainland and offshore zones, with thresholds of 20 cm and 55 cm for Marine and Land Threats respectively for the mainland zones and codify the offshore zone thresholds to 10 cm and 50 cm. The benefit of this is that there is minimal change in operational application.

- [0.65, 0.8] For both mainland and offshore zones use the 90\textsuperscript{th} percentile normalised by width as the predictor, with thresholds of 16 cm and 42 cm for Marine and Land Threats respectively for the mainland zones, and 10.5 cm and 40 cm for the offshore zones. In this case, the same predictor is used for all zones, but improvements in the warnings are only seen in the offshore zones. This technique is slightly more complicated to implement operationally as there is a need to incorporate the width of the rupture, and the gains in the skill are perhaps not significant enough to justify this effort. Indeed, this involves a change in technique for the mainland zones, but no change in skill.

- [0.69, 0.8] Use the 100\textsuperscript{th} percentile as the predictor for mainland zones with thresholds of 31.5 cm and 86 cm for Marine and Land Threats respectively and use the 90\textsuperscript{th} percentile normalised by width for the offshore zones with thresholds of 10.5 cm and 40 cm for Marine and Land Threats respectively. With this technique, improvements would be seen for all zones, but using different techniques for different zones means that more effort is required to implement and maintain the system.

4. **DISCUSSION**

The options presented above should be considered in the context of a number of caveats associated with this work. One issue in particular that should be borne in mind is that the analysis is very sensitive to the choices made in the desired warning schemes, which are somewhat subjective. For example, for the Chile 2010 event, there were only two mainland coastal zones that were deemed to have needed warnings. If it were to be decided that additional zones should have had warnings, or that the indicated warnings were not really necessary, then the optimal thresholds and the resulting Threat Scores would change. The sensitivity of the results to the underlying data is further demonstrated by the fact that the Threat Score lines in Figure 5, Figure 7 and Figure 9 are not smooth.

On a related point, the desired warnings have been defined according to known impacts. There may well be other locations within the coastal zone, or in other coastal zones where impacts occurred, but we did not know about them. Therefore, the selected warning schemes, as used in this study, arguably represent the ‘minimum’ desired warnings. This implies that the suggested marine thresholds are perhaps at a ‘maximum’.
In section 2.1, the exclusion of the Sumatra 2005 event from the present analysis was discussed, because it is not well represented in the T2 scenario database. Of course, there may well be future events which are not well described in the T2 scenario database, and for which this technique is not able to provide accurate warnings. One pathway to address this is to have a dynamic forecast system in which a tsunami forecast model is initiated with appropriate earthquake rupture characteristics in real-time. The technique of using amplitudes in the coastal zones as a proxy for coastal impact should in principle still be valid.

In the present work, we have used the Threat Score as the key metric for the skill of the system, but have also considered the Bias, which provides guidance on whether the system under- or over-forecasts. Another way to incorporate this could be to treat misses and false alarms differently, i.e., give them a different weighting in the Threat Score, under the assumption that false alarms are preferable to misses. This is not done here, but left for further work.

Further work could also involve consideration of a wider range of predictors. In the present work, we have considered 4 different percentile values (100th, 95th, 90th, 50th) and two adjustment factors (limiting by depth, and scaling according to wavelength, or inverse wavelength). These adjustment factors were not applied to all percentiles. For example, they were not applied to the 50th percentile because the Threat Scores for the 50th percentile were significantly lower for mainland zones than the other percentiles. It was assumed that adjustment factors would therefore not provide significant improvement. This assumption could be tested more robustly through the consideration of the adjustment factors (or potentially other factors) applied to all percentiles.

The incorporation of a factor to scale the amplitudes by wavelength (or inverse wavelength) is an attempt to account for the fact that waves of different wavelengths are likely to cause different levels of impact, in particular different amounts of run-up. The results found here do not in general support this idea. However, it may be that any differences are more likely to be seen with larger events, where there is in fact some inundation expected. Given the lack of tsunami inundation events for Australia, this would require the analysis to be undertaken for a different region, or undertaken using inundation modelling studies, as in Greenslade et al. (submitted).

5. SUMMARY

The Marine Threat threshold for the Australian Tsunami Warning System has been re-examined in the light of recent tsunami events. This is the first time this threshold has been assessed since the work of AG10. In addition to including 8 recent tsunami events,
a more robust and quantitative assessment technique has been used to determine the optimal thresholds, and alternative 'predictors' to the 95\textsuperscript{th} percentile amplitude value within coastal zones have been considered.

On the basis of the results, three possible options for the predictors are presented. Related optimal thresholds and Threat Scores for each option are provided in Section 3.3.

1) Continue to use the 95\textsuperscript{th} percentile for both mainland and offshore zones.
2) For both mainland and offshore zones use the 90\textsuperscript{th} percentile normalised by width as the predictor.
3) Use the 100\textsuperscript{th} percentile as the predictor for mainland zones and use the 90\textsuperscript{th} percentile normalised by width for the offshore zones.

In considering these various options, the various caveats in the technique discussed in Section 4 should be noted, and it should also be noted that the improvements in Threat Score are arguably marginal. This report presents the results but does not make a formal recommendation as to which technique should be incorporated into the operational system.

6. ACKNOWLEDGEMENTS

The authors would like to thank Deryn Griffiths and Simon Allen for their useful comments on early drafts of the manuscript.

7. REFERENCES

Charvet, I., Eames, I. and Rossetto, T. (2013) New tsunami runup experiments based on long wave experiments, Ocean Modelling, 69, 79-92, doi.org/10.1016/j.ocemod.2013.05.009