A first-generation national storm surge forecast system

Stewart Allen, Diana Greenslade, Frank Colberg, Justin Freeman and Eric Schulz

January 2018
A first-generation national storm surge forecast system

Stewart Allen, Diana Greenslade, Frank Colberg, Justin Freeman and Eric Schulz

Bureau Research Report No. 028
January 2018
Enquiries should be addressed to:

Stewart Allen:

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

stewart.allen@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

Abstract ...................................................................................................................... 1

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

2 Other relevant systems ..................................................................................... 5

3 Hydrodynamic model ........................................................................................ 8
   3.1 Model forcing ............................................................................................................ 9
   3.2 Bathymetry ................................................................................................................ 9

4 Wave Setup ...................................................................................................... 11

5 Observations ................................................................................................... 13

6 Model grid design ........................................................................................... 16
   6.1 Offshore territories .................................................................................................. 21

7 Verification ...................................................................................................... 23
   7.1 Hindcast runs ........................................................................................................... 23
   7.2 Non-linear surge-tide interactions ........................................................................... 26
   7.3 Case studies ........................................................................................................... 30
      7.3.1 Case Study 1: Victoria, June 2014 .............................................................. 32
      7.3.2 Case Study 2: South Australia, May 2016 ............................................. 38
      7.3.3 Case Study 3: South Australia, September 2016 .................................. 44
      7.3.4 Case Study 4: Western Australia, September 2014 .......................... 51
      7.3.5 Case Study 5: Western Australia, May 2016 ....................................... 56
      7.3.6 Case Study 6: New South Wales, June 2016 ....................................... 60
      7.3.7 Case Study 7: Queensland, March 2017 ................................................... 66
      7.3.8 Summary analysis of all case studies ......................................................... 71

8 Conclusions and further work ........................................................................ 73

9 Acknowledgments .......................................................................................... 75

10 References ....................................................................................................... 76

Appendix A: Time series of model and observations for hindcast runs .............. 81

Appendix B: Formulations for statistical indices ................................................. 89
List of Figures

Figure 1-1: A schematic showing some of the components of variability that comprise storm tide. ................................................................................................................................................ 2

Figure 3-1: Map indicating the bathymetry over the region. The green markers and labels indicate the locations of the ABLSMP tide gauges used in this study. ............................................ 10

Figure 5-1: Location of the ABLSMP (green circles) and supplementary tide gauges (blue circles) in Victoria and Southern Australia from which observations were used in case studies 1, 2 and 3. The locations marked by black diamonds are additional sites used for forcing verification only (see Section 7.3). ......................................................... 14

Figure 5-2: Location of the ABLSMP (green circles) and supplementary tide gauges (blue circles) in Western Australia from which observations were used in case studies 4 and 5. 14

Figure 5-3: Location of the ABLSMP (green circles) and supplementary tide gauges (blue circles) in New South Wales from which observations were used in case study 6. .............. 15

Figure 5-4: Location of the supplementary tide gauges in Queensland from which observations were used in case study 7. ................................................................................................... 15

Figure 6-1: A representation of the regular grid, showing its coverage and extent. For clarity, only every 25th cell is plotted. Axes show the indicative (i.e. ‘zonal’ and ‘meridional’) orientation of the ξ and η model axes. .................................................................................. 16

Figure 6-2: A representation of the ribbon grid, showing its coverage and extent. For clarity, only every 25th cell is plotted. The position of the grid ‘seam’ is denoted by the bold line. Axes show the indicative (i.e. ‘along-shore’ and ‘cross-shore’) orientation of the ξ and η model axes. ............................................................................................................. 17

Figure 6-3: Time series of sea level elevation in metres for the year-long run for both the regular grid (orange) and ribbon grid (green). These time series are compared to observed residuals (teal) from the ABLSMP Tide gauges at Spring Bay, Tasmania (top), Cape Ferguson (Queensland) and Thevenard, South Australia (bottom). The model time series are taken from the model grid point nearest to the location of the tide gauges. .................. 19

Figure 7-1: a) mean absolute deviation of the non-linear interactions at coastal points. b) maximum absolute deviation of the non-linear interactions at coastal points. Units are in metres................................................................. 28

Figure 7-2: Mean absolute deviation and maximum absolute deviation of the non-linear interactions at the grid cell nearest to each of the tide gauges used in this study. Units on the vertical axis are in metres................................................................. 29

Figure 7-3: Official 0000UTC MSLP analyses for 22, 23 and 23 Jun 2014. ......................... 32

Figure 7-4: Time series of observed and ACCESS-R MSLP for three sites along the south west Victorian coastline. Time is in UTC. .................................................................................................................. 33

Figure 7-5: Hs observations from the Cape de Couedic Waverider buoy (blue markers) and AUSWAVE-R Hs forecasts (orange line). Time is in UTC. ................................................................................. 34

Figure 7-6: Time series of calculated surge and wave setup components and their sum (surge + setup) for the June 2014 case study. Also shown is the observed residual sea level, which has a 6-minute interval at all locations. Times are in UTC......................................................... 35

Figure 7-7: Official 0000UTC MSLP analyses for 8, 9 and 10 May 2016. ......................... 39

Figure 7-8: Time series of observed and ACCESS-R MSLP for three sites in the region of the South Australian Gulfs (Edithburgh is approximately 10km from Port Giles). Time is in UTC. ................................................................. 40
Figure 7-9: $H_s$ observations from the Cape de Couedic Waverider buoy (blue symbols) and AUSWAVE-R $H_s$ forecasts (orange line). Time is in UTC. .................................................. 40

Figure 7-10 Time series of calculated surge and wave setup components and their sum (surge + setup) for the May 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all sites except Thevenard, which has a 6-minute interval. Times are in UTC................................................................. 41

Figure 7-11: Official 0000UTC MSLP analyses for 28, 29 and 30 September 2016. ............... 45

Figure 7-12: Time series of observed and ACCESS-R MSLP for three sites in the region of the South Australian Gulfs (Edithburgh is approximately 10km from Port Giles). Time is in UTC. ........................................................................................................................... 45

Figure 7-13: $H_s$ observations from the Cape de Couedic Waverider buoy (blue markers) and AUSWAVE-R $H_s$ forecasts (orange line). Time is in UTC. .................................................. 46

Figure 7-14: Time series of calculated surge and wave setup components and their sum (surge + setup) for the September 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all sites except Thevenard, which has a 6-minute interval. Times are in UTC. ................................................................................................................ 47

Figure 7-14: Official 0000UTC MSLP analyses for 6, 7 and 8 Sept 2014. ......................... 52

Figure 7-16: Time series of calculated surge and wave setup components and their sum (surge + setup) for the September 2014 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all sites except Esperance and Hillarys, which have 6-minute intervals. Times are in UTC. ................................................................. 53

Figure 7-17: Official 0000UTC MSLP analyses for 20, 22 and 24 May 2016. .................. 56

Figure 7-18: Time series of calculated surge and wave setup components and their sum (surge + setup) for the late May 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at Fremantle and Geraldton and a 6-minute interval at Esperance and Hillarys. Times are in UTC. ................................................................. 58

Figure 7-19: Total rainfall between 4 and 7 June 2016 across eastern Australia (from Bureau of Meteorology, 2016). ................................................................. 60

Figure 7-20: Official 1200UTC MSLP analyses for 4, 5 and 6 June 2016. ......................... 61

Figure 7-21: Time series of calculated surge and wave setup components and their sum (surge + setup) for the June 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all locations. Times are in UTC. ................................................................. 63

Figure 7-22: Official 0000UTC MSLP analyses for 29,30 and 31 March 2017.................. 66

Figure 7-23: Time series of calculated surge and wave setup components and their sum (surge + setup) for the March 2017 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all locations. Times are in UTC. ................................................................. 68
ABSTRACT

A new national storm surge forecast system has been developed. This is the first system within the Bureau of Meteorology to provide numerical model guidance for forecasts of storm surge sea level around the entire Australian coast. The Regional Ocean Modeling System (ROMS) model was selected to calculate these forecasts. It is used in 2D, depth-averaged barotropic mode and the model grid is spaced uniformly in latitude and longitude with a spatial resolution of approximately 2.5 km around the Australian coastline. The model is forced by surface wind stresses and atmospheric pressures from the Bureau of Meteorology’s operational regional atmospheric forecast model ACCESS-R. In its operational configuration, the model is ‘warm run’ every 6 hours for 72-hour forecasts, aligned with the configuration of ACCESS-R. Estimates of wave setup are calculated from the Bureau of Meteorology’s AUSWAVE-R ocean wave forecast model and are linearly combined with the ROMS storm surge to provide forecasts of residual (i.e. non-tidal) sea level.

The model was evaluated in two ways. First, a 30-month hindcast run was created to assess the model’s long-term performance. When compared to observations from tide gauges from the Australian Baseline Sea Level Monitoring Project (ABLSMP) network, the average root-mean-square difference (RMSD) in sea level (with astronomical tides removed) was approximately 10 cm. This is comparable to, or better than, similar international systems. The skill was found to be better along the southern coastline than the northern coastline.

The model was then evaluated for several case studies of recent significant storm events occurring on Australia’s south, west and east coasts. The model was found to perform very well for five events, with modeled peak surges for these events within 16% of the observed peak surge amplitudes. It performed less well for two other events, with modeled peak surge within 37% and 41% of the observed peak surge. Timing of the surge was reasonably well simulated for the cases, although with some substantial outliers. With these outliers excluded, the model peak time is generally within 2 to 3 hours of the observed peak time. Forecasts with different lead-times were also examined and it was found that it is not always the case that short-range forecasts have better skill than the long-range forecasts.
1 INTRODUCTION

Storm surge is defined as an elevation of water level at the coast resulting from strong winds and reduced atmospheric pressure. Storm surges are very often associated with Tropical Cyclones as they come onshore but may also be generated in non-tropical areas, most typically by intense low-pressure systems. The definition of storm surge typically includes the effects of wave setup – an additional elevation of the water level at the coast due to wave-breaking. This is distinct from wave run-up, which relates to the impact of individual waves on shoreline structures or beaches and is often manifest as inundation or ‘over-topping’.

Storm tide is the combination of storm surge and astronomical tide – if a storm surge arrives at or close to the time of high tide, the impacts can be considerably more damaging than if it arrives at low tide. A depiction of these quantities and their contribution to a storm tide, as might be experienced at a coastal location, is shown in Figure 1-1.

Figure 1-1: A schematic showing some of the components of variability that comprise storm tide.

The Bureau of Meteorology (the Bureau) is undertaking a project to enhance its operational storm surge forecasting system.
The project consists of three key components:

1. An event-based Tropical Cyclone (TC) ensemble storm surge system,
2. A National Storm Surge Forecast System for forecasting anomalous sea levels due to mid-latitude storms and tropical lows, and
3. Operationalization of an existing aggregate sea level monitoring and alert system at tide gauge locations.

There are many scientific and technical commonalities to each of the three components of the project. This report describes the second component: The National Storm Surge Forecast System. It should also be noted that the systems in the first two components are traditionally referred to as ‘storm surge’ systems, despite the fact that they ultimately both include the effects of astronomical tides and are therefore can be considered ‘storm tide’ systems. This report only focuses on the non-tidal component of the National Storm Surge Forecast System.

As noted above, although storm surges are most often associated with TCs, there can also be significant impacts due to sub-tropical and mid-latitude low pressure systems. For example, there are numerous historical records of flooding of coastal areas and damage to coastal infrastructure in South Australia (SA) (Vecchio, 1980). More recently, the Port Germain jetty in Spencer Gulf has been damaged several times during severe weather events (ABC, 2016). The coastal topography and bathymetry of the South Australian Gulfs region make these parts of the South Australian coastline particularly susceptible (Vecchio 1980).

Historically, the Bureau has had no nationally consistent method or protocol to provide forecasts or warnings for non-TC related storm surges. The SA Regional Office provides a storm tide warning service for potentially high coastal sea levels based on simple relationships between atmospheric pressure, wind stress and sea level (Vecchio, 1980).

The capability now exists to run a numerical hydrodynamic model for coastal regions on a routine basis to provide guidance for these sorts of events. In this project component, a storm surge forecast system is developed using the Regional Ocean Modeling System (ROMS; Shchepetkin and McWilliams, 2005) that is forced by the Bureau’s operational regional atmospheric forecast model, ACCESS-R. The resulting storm surge forecasts from ROMS are combined with wave setup estimates to produce forecasts of storm-induced residual (i.e. non-tidal) sea level at the coast.
This report describes the configuration of the storm surge model and presents a series of verification results. The report is structured as follows: Section 2 places this work in context by describing some other relevant operational (non-TC) storm surge forecast systems and other Australian storm surge studies; Section 3 describes the configuration of ROMS, the hydrodynamic model; Section 4 describes the wave setup component of the system; Section 5 describes the observations used in the verifications; Section 6 presents some experiments undertaken to provide guidance on the choice of model grid; Section 7 presents an evaluation of the system for a 30 month hindcast and also examines several case studies of extreme events; and finally Section 8 presents the conclusions and some suggestions of further work to improve the system.
2 OTHER RELEVANT SYSTEMS

Many countries outside the tropics are vulnerable to the impacts of increased sea level due to intense storms, and as a response have developed storm surge models to provide forecasts and warnings. The most relevant operational forecast systems to this report are those for Canada, the United States (US) and the United Kingdom (UK) and are briefly described for context.

Since the late 1990’s, Environment Canada has been running an operational storm surge forecast system for its east coast based on the Princeton Ocean Model (POM; Mellor, 1998). There are currently two operational systems - a deterministic system at 1/30\textsuperscript{th} degree resolution and a 21-member ensemble system at 1/12\textsuperscript{th} degree resolution. The POM model is used for both systems and is run in depth-averaged barotropic mode with tides linearly superimposed on the surge forecasts. The model is forced using Environment Canada’s operational deterministic and ensemble atmospheric forecast systems respectively, and the forecast range is 10 days. Further details of the model configurations can be found in Bernier and Thompson (2015). Validation of the forecasts by Bernier and Thompson (2015) showed that the standard deviation of surge forecast error ranges from around 8 cm for short-range forecasts (up to 24 hours) to around 20 cm for the 10-day forecasts. They concluded that there is skill in the deterministic system for forecast ranges of up to 6 days. It is suggested that the relatively low resolution of the atmospheric forcing fields for the ensemble system (66 km) contributes to underestimation of large surge events.

In the US, the National Weather Service (NWS) currently runs two operational extratropical storm surge systems. The earlier of these (ET-SURGE; NOAA, 2013) is a variation on the NWS's Sea, Lake and Overland Surges from Hurricanes model (SLOSH; Jelesnianski et al., 1992) and is used for TC storm surge forecasting. ET-SURGE is run four times daily to produce numerical storm surge guidance for extratropical systems out to 96 hours. It is a barotropic ocean model and is forced by real time surface winds and pressure from the NCEP Global Forecast System (GFS). The second, more recently developed system, Extratropical Surge and Tide Operational Forecast System (ESTOFS; Feyen et al., 2013) provides a further set of forecast guidance in addition to the ET-SURGE model. It has a finer resolution (2.5 km vs. 5 km for ET-SURGE) and uses the Advanced Circulation model ADCIRC (Luettich and Westerink, 2004). The model is barotropic and includes tides dynamically. The ESTOFS model is run four times daily out to 180 hours. The model is also forced by real-time winds and pressure from the GFS system. When compared to observations of sea level from 48 tide gauges around the US Atlantic and Gulf coasts, forecasts of total sea level from ESTOFS were found to have an RMS error of approximately 20 cm (Feyen et al., 2013).
The operational storm surge system for the UK is similar to Canada’s and consists of a deterministic model and, as of 2017, two ensemble systems. The ensemble system was initially implemented in 2009, and upgraded and extended in 2011 (Flowerdew et al., 2013). The hydrodynamic model used is the CS3X model, based on that of Flather (2000), and is developed and maintained at the Proudman Oceanographic Laboratory, and run operationally at the UK Met Office. CS3X is a single layer barotropic model, and covers the northwest European continental shelf with horizontal resolution of 1/9th degree in latitude and 1/6th degree in longitude. Tides are dynamically included in the model, but in noting that harmonically analyzed tides are generally more accurate than dynamically modeled tides, a second model is run with tides only (no atmospheric forcing) and subtracted from the full model run. This creates storm surge forecasts that include the non-linear interactions between tides and surge. Harmonically analyzed tides are then linearly added to this surge forecast to provide total water level.

The two ensemble storm surge systems consist of 24 members forced with different versions of the Met Office Global and Regional Ensemble Predictions System (MOGREPS; Bowler et al., 2008). The first system is forced with the higher–resolution MOGREPS-R atmospheric ensemble out to a forecast range of 54 hours, and the second system is forced with the coarser resolution MOGREPS-15, allowing the storm surge ensemble to be run for forecast ranges up to 7 days. Flowerdew et al. (2013) provide further details of these implementations along with verification statistics.

Within the Bureau, a recently implemented aggregate sea level forecast system (Bureau National Operations Centre, 2016a; Taylor and Brassington, 2017) provides objective forecast guidance for coastal sea level at 116 tide gauge locations in the Australian region. The system is primarily a downstream product of OceanMAPS (Brassington et al., 2012), and incorporates the effects of sea level pressure from ACCESS-G forecasts (Puri et al., 2013), tidal harmonics and tide gauge observations. It produces 7-day forecasts every 24 hours. The skill of the current version is generally good along the southern shelves and within the Gulf of Carpentaria but relatively poor in northern NSW, Queensland and on the North West Shelf.

Some research in the Australian context relating to extratropical storm surges has been undertaken by Haigh et al. (2014). The Mike21-FM hydrodynamic model (DHI, 2010) was used to generate a 61-year hindcast of coastal water levels around Australia, forced by pressure and winds from the NCEP global reanalysis, and incorporating astronomical tides. The aim of this work was to generate exceedance probabilities of extreme water levels. The model was run in depth-averaged barotropic
mode on an unstructured grid with a spatial resolution of approximately 10 km around the Australian coastline. The model was found to reproduce the surge component very well, with a mean Root Mean Square Error of about 5 cm when compared to tide gauge observations around the Australian coastline during 1995.

Hetzel et al. (2018) examined a number of storm surge events both tropical and extra-tropical. The SCHISM hydrodynamic modeling system was used (Zhang et al., 2016), with a two-way coupled wind-wave model. This is one of few storm surge studies to use a full 3D hydrodynamic model, as opposed to a depth-averaged model. In this case, the model was run with 61 levels and an unstructured grid with spatial resolution of up to 100 m at the coast. It was found that the incorporation of wave setup could increase the total surge levels by up to 50%, depending on local water depth and coastline orientation.

McInnes and Hubbert (2003) undertook a numerical modeling study of two storm surges within Bass Strait. They found that for the two events studied, the wind stress forcing was significantly more important than the atmospheric pressure forcing, which resulted in considerably less sea level variability than would be expected through the inverse barometer effect. They also showed that in order to fully simulate sea level variability within Bass Strait, a model domain spanning much of the south coast of Australia is needed so that remotely generated coastal trapped waves are included.

Other extratropical storm surge research relating to the Australian coastline has focused on future climate scenarios. For example, Colberg and McInnes (2012) used ROMS (in barotropic mode) to investigate potential changes in sea level around the southern coast of Australia under several different future climate simulations. The model was run with a spatial resolution of approximately 5 km and astronomical tides were not included. It was found that oceanic response to the climate change signal in the future climate simulations is a reduction in daily maximum sea surface height along the south coast of Australia and an increase along the east coast.
3 HYDRODYNAMIC MODEL

The ROMS model was selected as the hydrodynamic model for this work. Colberg et al. (2013) completed a comprehensive review of potential numerical models, and concluded that there were several numerical models available that would be suitable for this application. ROMS was selected not only as the hydrodynamic model for the National Storm Surge Forecast System but also for the TC event-based Tropical Storm Surge Forecast System (Greenslade et al., 2018). It is also being employed in the eReefs project, where ROMS is being used to establish both an operational ocean forecasting system and a multi-year hindcast and reanalysis for currents, temperature, salinity, sea level and river tracer concentrations from major rivers for the Great Barrier Reef region. (Colberg et al, 2018; Sandery et al., 2018)

ROMS is an open-source model, based upon a free-surface, terrain-following formulation of the primitive equations. The equations are solved using a split-explicit time-stepping scheme. In essence, this means that barotropic (fast) and baroclinic (slow) modes are treated separately and with different time step lengths, but are carefully re-coupled at the end of each ‘slow’ time step. Any aliasing effects resulting from the different time-steps are removed. Time-stepping uses a stable, third-order predictor-corrector type scheme and is constrained to preserve volume conservation.

ROMS was configured to run in a 2D, depth-averaged barotropic mode, based on the shallow water equations only. This configuration allows for a rapid processing speed, while retaining the main modes of variability for long-period disturbances, such as storm surges. A secondary reason for this configuration is to retain similarity with the Tropical Storm Surge Forecast System, which has been developed in parallel.

By not including any vertical structure in the model, some modes of variability may not be included or well simulated. However, it has been argued by others that these modes do not contribute significantly to mid-latitude storm surge (Kauker and Langenberg, 2000; Colberg and McInnes, 2012). Moreover, this approach is consistent with current international practice, as discussed in Section 2.
Other specific configuration details of the model that were adopted include:

- Non-linear free-surface;
- Spherical coordinate system with curvilinear grid geometry;
- Open boundary conditions;
- Land/Sea masking with a minimum water depth of 5 m;
- Wetting and drying of coastal grid cells to improve model stability.

### 3.1 Model forcing

The ROMS model is forced by mean sea level pressure (MSLP) and surface wind stress from the Bureau’s ACCESS-R model (Puri et al., 2013). The ACCESS-R model is an operational regional scale numerical weather prediction model for Australia, whose domain covers from 65°S to 16.95°N in latitude and 65°E to 184.57°E in longitude. It has a horizontal spatial resolution of 0.11° (approximately 12 km) and 70 vertical levels. It has received several upgrades over time, particularly to its data assimilation scheme (Bureau National Operations Centre, 2016b). For this application, only surface fields of MSLP and wind stress are taken from ACCESS-R and used as forcing for the storm surge model.

Due to the higher surface roughness of land, the values of ACCESS-R wind stress will be higher over land than over the oceans. Given this and the coarseness of the ACCESS-R fields (compared to the spatial resolution of the storm surge model), some modification of the ACCESS-R wind stresses at the land/sea boundary is required: land values of ACCESS-R wind-stress were removed and then interpolated from ocean wind-stress values using a successive over-relaxation technique. This avoids the interpolation of large values of land-based ACCESS-R wind stress to ocean values.

The ACCESS-R model is run four times daily, at 0000, 0600, 1200 and 1800UTC, and provides 72-hour forecasts. The National Storm Surge Forecast System is run routinely, matching the ACCESS-R run schedule and duration.

### 3.2 Bathymetry

Two bathymetric data sets were used as the underlying grid bathymetry for this study. The majority of the region of interest is covered by the Geoscience Australia (GA) 250 m national bathymetry data set
The GA data set spans longitudes from 92°E to 172°E and latitudes from 60°S to 8°S.

For areas of north of 8°S that are not covered by the GA data set, the GEBCO_08 30” grid (version 20150318, http://www.gebco.net) was used. The two sets are blended together around 8°S and then down sampled to 500 m. A map of the bathymetry is shown in Figure 3-1. This blended bathymetry data set was sampled during the generation of the candidate grids discussed in Section 6.
4 WAVE SETUP

In addition to the contributions to sea level from atmospheric pressure and large-scale wind stress forcing, wind-wave breaking at the shoreline can add an additional elevation of the water level. Here, momentum from the wave field is transferred to the depth integrated water column in the surf zone. It can be a significant contribution to sea level, particularly in areas of steep coastal bathymetry, such as where the continental shelf is narrow, or at non-continental islands. However, it is debatable as to whether wave setup is always observable in sea level observations from tide gauges, given these gauges are rarely located in the surf zone.

Direct calculation of the contribution to sea level from wave setup is possible with a spectral wave model using radiation stress theory. However, the computational burden of this approach, as well as the relatively coarse (near-shore) resolution of existing operational wave forecast models mean that this is not presently possible in an operational setting. Therefore, a parameterised approach must be used, whereby the wave field some distance offshore is used to estimate the wave setup at the coastline.

As in Tropical Cyclone ensemble storm surge system (which was developed in parallel to this system; Greenslade et al., 2018), the wave setup is calculated using the parametric approach developed by CSIRO (O’Grady et al., 2015) and using the Bureau’s AUSWAVE-R operational forecasts (BNOC, 2016c).

The parameterisation of wave setup \( \eta_0 \) was derived from multiple simulations around the Australian coast made with the SWAN near-shore wave model and is calculated as follows:

\[
\eta_0 = 0.31 (0.8) H_S \left[ 0.325 \left( \frac{2\pi H_S}{g T_p^2} \right)^{\frac{1}{4}} \right]
\]

Here, \( H_S \) is Significant Wave Height (from AUSWAVE-R), \( T_p \) is peak wave period (from AUSWAVE-R), \( g \) is the acceleration due to gravity and \( S \) is the bathymetric slope. Estimates of bathymetric slopes were determined around the Australian coastline from the model grid (see Section 6), rather than using those contained in O’Grady et al. (2015).

This approach has minimal computational cost and the necessary inputs can easily be extracted from an existing Bureau operational system. However, it also has some limitations based on its formulation.
and assumptions. It does require parametric extrapolation from the deep-water wave-field, based on
the coarse wave model. O’Grady et al., (2015) discuss this wave setup calculation in more detail and
Greenslade (2016, unpublished manuscript) provides further discussion and presents alternative
approaches, each with their merits and shortcomings.

Note that wave setup is distinct from wave run-up, which relates to the impact of individual waves on
shoreline structures or beaches and is often manifest as inundation or ‘over-topping’. As for wave
setup, explicit calculation of wave run-up is possible but requires high-resolution phase-resolving
models, which, like the direct calculation of wave setup, has significant computational cost. In this
project, wave run-up is considered out of scope.
5 OBSERVATIONS

The primary observations used to verify the forecasts of the National Storm Surge Forecast System are the sea level values measured by the tide gauges that make up the Australian Baseline Sea Level Monitoring Project (ABLSMP) tide gauge network. The aim of the ABLSMP is to provide long-term, high-quality observations of sea level around Australia. The network is comprised of 16 tide gauges and began collecting observations in 1991. The Bureau owns and operates the majority of the gauges in the network, however port operators own two stations (Lorne and Stony Point). Of the ABLSMP network, one gauge (Port Stanvac) only has measurements until the end of 2010, another (Cocos Island) was outside the bounds of the grid, and a third (Stony Point) is in semi-enclosed waters that are not well represented by the grid. Therefore, the set of ABLSMP gauges from which observations have been used in this study comprises 13 gauges, the locations of which are shown in Figure 3-1.

Observations from a supplementary set of gauges, which provide extended coverage of the case study events, are used in Section 7.3. Some of these gauges are in the ABLSMP network, while others are owned by local port authorities. The locations of the non-ABLSMP gauges used in the case studies are shown from Figure 5-1 to Figure 5-4.

In its operational configuration, the forecast coastal sea level from the National Storm Surge Forecast System is obtained from linearly combining surge, astronomical tide and wave setup. In this report, we are mostly interested in the performance of the surge component alone (and wave setup, to a certain extent) so the astronomical tide component is removed from the observations. This was achieved by subtracting harmonic tidal forecasts from observations of sea level to produce data sets of observed residual sea level. The vertical datum of these observed residuals is the Australian Height Datum (AHD). Note that this datum metadata was not available for the Groote Eylandt gauge and so observed sea levels at this location cannot be normalized to AHD and may be erroneously offset.
Figure 5-1: Location of the ABLSMP (green circles) and supplementary tide gauges (blue circles) in Victoria and Southern Australia from which observations were used in case studies 1, 2 and 3. The locations marked by black diamonds are additional sites used for forcing verification only (see Section 7.3).

Figure 5-2: Location of the ABLSMP (green circles) and supplementary tide gauges (blue circles) in Western Australia from which observations were used in case studies 4 and 5.
Figure 5-3: Location of the ABLSMP (green circles) and supplementary tide gauges (blue circles) in New South Wales from which observations were used in case study 6.

Figure 5-4: Location of the supplementary tide gauges in Queensland from which observations were used in case study 7.
6 MODEL GRID DESIGN

Two different model grids were derived from the bathymetry data and considered as options for the National Storm Surge Forecast System. These two grids differed significantly in terms of their geometries. In this section, the grids are described and evaluated with respect to their computational performance and accuracy.

The first grid is a simple grid that is uniformly spaced in latitude and longitude. It will be referred to as the ‘regular grid’. Details are shown in Table 6-1 and a representation is shown in Figure 6-1. In the meridional direction (with respect to the ROMS coordinate system, this is the η-direction), the grid has 2350 cells and spans latitudes 47°S to 7°S. This gives a grid resolution of approximately 1 arc-minute or 1.89 km.

In the zonal direction (in the ROMS coordinate system, this is the ξ-direction), there are 2000 cells and the grid spans longitudes from 107.5°E to 157.5°E. This gives a resolution of 1.5 arc minutes, but
since lines of longitude converge towards the poles, the spatial resolution in this direction varies with latitude. At the northern edge of the grid the spatial resolution is approximately 2.75 km, while at the southern boundary of the grid the spatial resolution is approximately 1.85 km. At the most northern Australian coastline (Cape York in northern Queensland), the resolution is approximately 2.70 km, while at the most southern coastline (South East Cape in southern Tasmania) the resolution is approximately 2.00 km.

It can be seen from Figure 6-1 that the regular grid encompasses the entire Australian continental landmass. Grid cells in land areas are ‘masked’, that is, computations for these cells are performed but then set to a masked value. These areas represent areas of computational inefficiency.

![Figure 6-2: A representation of the ribbon grid, showing its coverage and extent. For clarity, only every 25th cell is plotted. The position of the grid ‘seam’ is denoted by the bold line. Axes show the indicative (i.e. ‘along-shore’ and ‘cross-shore’) orientation of the ξ and η model axes.](image)
With this inefficiency in mind, the other candidate grid considered is oriented as a ‘ribbon’ encircling the coastline of mainland Australia and Tasmania. A representation of this ribbon grid is shown in Figure 6-2. The intent behind this design is to reduce the land area encompassed by the grid and potentially increase computational efficiency. Given its geometry, the ξ- and η-directions for each of the ROMS axes should not be considered as the zonal and meridional directions, but rather the ξ-axis should be thought of as being oriented along-shore (clockwise around the continent) and the η-axis cross-shore (outwards from the coastline).

Two boundaries of the ribbon grid meet at a ‘seam’. The location of the seam was placed in a region where it could be designed to have low curvature, ensuring adjacent grid points on either side of the seam are closely located to each other. This allows correct transfer of information across the seam via periodic boundary conditions.

<table>
<thead>
<tr>
<th>Table 6-1: Parameters of the regular and ribbon grids.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cells in ξ direction</td>
</tr>
<tr>
<td>Cells in η direction</td>
</tr>
<tr>
<td>Total grid cells</td>
</tr>
<tr>
<td>Proportion of masked nodes</td>
</tr>
<tr>
<td>Spatial resolution range (km, along each grid axis)</td>
</tr>
<tr>
<td>Median spatial resolution (km, along each grid axis)</td>
</tr>
<tr>
<td>Time step (seconds)</td>
</tr>
<tr>
<td>Courant number</td>
</tr>
</tbody>
</table>

Some parameters of the ribbon grid are shown in Table 6-1. In particular, it can be seen from Table 6-1 that not only does the ribbon grid have approximately 36% fewer grid cells than the regular grid, but a smaller proportion of its cells are masked (i.e. land).
One feature of the ribbon grid is that the local curvature results in a more variable grid resolution, compared to the regular grid. The resolution in either direction varies from as fine as approximately 0.6 km to as coarse as 7.7 km. However, over 95% of grid cells have a resolution of less than 3 km.

Most of the grid is internally concave, resulting in the desired quality of increased resolution near the coast (the coastline being the region of greatest interest). However, there are some coastal regions where the grid is internally convex and therefore relatively coarse in its resolution, namely along much of the coastline of South Australia. The southern coastline of Tasmania is also located in a region of relatively coarse resolution. This presents an undesirable quality of the ribbon grid.

To compare the computational performance of both grids, a model run of 1-year duration (2014) was performed using each grid. An important value to consider when configuring a model is the Courant number. To ensure numerical stability this value should be below 1. The finer spatial resolution in the ribbon grid requires a shorter time step than the regular grid in order to ensure a Courant number less than unity and maintain numerical stability. Therefore, the total number of computational steps required by the ribbon grid is greater than for the regular grid for an equivalent run duration. When run on the National Computing Infrastructure’s Raijin supercomputer using 256 CPUs, the ribbon grid took approximately 53.5 hours to complete the 1-year run, while the regular grid took approximately 51.5 hours (see Table 6-2). Note that these run times are merely indicative - variations are expected across repeat runs.

Figure 6-3: Time series of sea level elevation in metres for the year-long run for both the regular grid (orange) and ribbon grid (green). These time series are compared to observed residuals (teal) from the ABLSMP Tide gauges at Spring Bay, Tasmania (top), Cape Ferguson (Queensland) and Thevenard, South Australia (bottom). The model time series are taken from the model grid point nearest to the location of the tide gauges.
simulations due to external factors, such as system load at the time of running.

From a numerical performance perspective, the gains made with the ribbon grid by reducing the number of grid cells and the proportion of masked cells are effectively lost by the need for a shorter time step. The difference between the two grids with respect to computational performance is therefore negligible.

Although there is little to separate the computational performance of the two grids, the ribbon grid has several features such as fewer land-masked points and increased coastal resolution (in most areas) that may provide advantages, such as improved accuracy, over the regular grid. To evaluate this, the results from the runs for each grid are compared to observations from the ABSLMP tide gauges.

Examples from three of the 13 tide gauge locations are shown in Figure 6-3 and the performance of both grids was assessed quantitatively by determining the bias, mean absolute difference (MAD) and root mean square difference (RMSD) for each grid (Table 6-2). These statistics were calculated from time series taken from the model grid points nearest to each the 13 ABLSMP Tide Gauges.

Table 6-2: Computation time and statistics for the year-long run for both the regular grid and ribbon grid. These statistics were determined by comparing a time series for each of the grids taken from the model grid point nearest to the location of the 13 ABLSMP Tide Gauges.

<table>
<thead>
<tr>
<th></th>
<th>Regular grid</th>
<th>Ribbon grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computation time (hours)</td>
<td>51.5</td>
<td>53.5</td>
</tr>
<tr>
<td>Bias (cm)</td>
<td>-0.01</td>
<td>1.17</td>
</tr>
<tr>
<td>MAD (cm)</td>
<td>6.91</td>
<td>6.89</td>
</tr>
<tr>
<td>RMSD (cm)</td>
<td>8.65</td>
<td>8.59</td>
</tr>
</tbody>
</table>

Given that there is little difference between the two grids in terms of both their performance and accuracy, there does not appear to be any compelling reason to use the more complex ribbon grid over the simpler the regular grid. Furthermore, other reasons related to possible future development (discussed in Section 8) make the regular grid a better choice over the ribbon grid. Therefore, it was decided to base the National Storm Surge Forecast System on the regular grid, henceforth referred to only as ‘the grid’ or ‘the model grid’.
6.1 Offshore territories

In addition to the model grid outlines in the previous section, additional and separate model grids were considered for specific Australian offshore territories, specifically Cocos Islands, Christmas Island, Lord Howe Island, Norfolk Island and Macquarie Island. Each of these territories is remote from the Australian continent and is an island atop an underwater seamount. The topography of each differs, particularly with respect to their coastlines. Some may be particularly prone to ocean impacts from storms, while others have natural geographic coastal defences from storms. However even these natural defensible islands contain a few key locations that house coastal infrastructure vital to the island.

Grids for each of the five islands were designed approximately 1000km square, ensuring a length scale large enough to resolve synoptic scale weather systems and their forcing (i.e. with sufficient ‘fetch’). This is based on analysis performed for the Tropical System, as described in Greenslade et al. (2018). The bathymetry used to make the grids for Cocos and Christmas Islands was taken from the Geoscience Australia (GA) 250 m national bathymetry data set (Whiteway, 2009), except for areas north of 8°S, where the GEBCO_08 30” grid (version 20150318, http://www.gebco.net) was used instead. However, for Lord Howe, Norfolk and Macquarie Islands, it was found that the GA 250m dataset doesn’t resolve these islands. Therefore, grids for these islands were created solely from GEBCO_08 30” grid. For all islands, the source bathymetry was down-sampled to a resolution of approximately 2 km.

The grids were tested by running ROMS with ACCESS-R forcing as described above for a 12-month simulation spanning 2014. Contributions of wave setup, calculated using the method described in section 4 were added to the ROMS output at coastal grid points to produce simulations of total residual (i.e. non-tidal) sea level.

The models were verified by comparing model time series to observations of residual sea level from tide gauges at each of the locations (with the exception of Macquarie Island, for which observations were not available). The model time series were taken from the coastal grid point that was either nearest to, or representative of the tide gauge location. Full details are contained in Allen and Greenslade (2017, unpublished manuscript).

The comparisons (not shown) show that the storm surge system performs poorly for the offshore territories, in terms of predicting total sea level as the sum of surge and wave setup. In general, the
The sum of modelled surge and wave set-up is considerably higher than the observed signal, and this is predominantly due to high values of wave set-up, which are typically at least 99% of the total signal.

The storm surge component of the system is, in comparison, generally negligible in magnitude and does not show any variability at the scales seen in the observations. The dominance of wave setup over the effect of wind- and pressure-driven storm surge is expected and is well understood, particularly in the context of impacts on offshore islands, where the coastal bathymetric profile is very steep compared to continental coastal profiles (Kennedy et al. 2012).

The coarse resolution of the model grid means that in most cases, the model grid does not resolve the island coastline or topography very well, and complex, fine-scale coastal features are not sufficiently resolved. Past studies have suggested that modelling on the scale of 10s of metres is required to accurately simulate flows across fringing reefs and through channels, the discharge and recharge of atoll lagoons and other morphological features (Hoeke et al., 2011).

Given these results, it is not recommended that the model configuration documented in this section be applied to offshore territories. While subsequent sections of this report will show that the National Storm Surge Forecast System is effective in simulating wind and wave surge impacts at Australian continental shores, it cannot accurately do the same for offshore islands. Full details of this study for the offshore territories are provided in Allen and Greenslade (2017, unpublished manuscript).
7 VERIFICATION

The forecasts for the mainland Australian coast were evaluated in two ways. First, a relatively long ‘hindcast’ run was performed to assess the model’s long-term performance. This provides some information about the ability of the model’s underlying physics to correctly simulate the sea level variability arising from pressure and wind stress forcing. These runs are described in Section 7.1.

Second, the model was evaluated for seven case studies of past significant storm events. These runs for were for shorter periods (i.e. the duration of the events) and the model was run in its operational configuration, such as using a recurring warm-start mode and adding the effects of parameterized wave setup (see Section 4). In this case, the model was evaluated using observations from an extended network of tide gauges. Details and discussion of these runs are in Section 7.3.

7.1 Hindcast runs

The hindcast model runs were forced using the ACCESS-R data, taking an analysis field and the subsequent forecast for the ensuing hours until the next analysis is available. This results in an hourly sequence of forcing fields consisting of analyses at 6-hourly intervals, with 5 hours of short-range forecasts between each analysis. The model was initialized with a 5-day spin up period, starting from an ocean at rest, and then run for a 30-month period from 1 January 2014 to 30 June 2016.

As noted in section 4, it is not clear whether wave setup is detectable in observations from the tide gauges used in this study. Given this, the effects of wave setup are not added to the hindcast runs in this section, although they will be included in section 7.3, where the model is evaluated under operational conditions. Evaluation of whether wave setup is discernible in observations from tide gauges is an area of for further investigation (see section 8).

Using this long period ensures that many modes of variability are included in the observations and forcing. In particular, there were some 10 Tropical Cyclones making landfall on to the mainland Australian coastline during this period. It is known that ACCESS-R poorly resolves the intensity and strong spatial gradients associated with Tropical Cyclones. Therefore, the resultant storm surge would be expected to be underestimated by the National Storm Surge Forecast System.

In addition, there are some low-frequency modes of ocean variability that may be apparent in the observations but are not well simulated or resolved by the model. These occur on seasonal to
interannual time-scales and result from phenomena such as the El-Niño and Southern Oscillation (ENSO). They may be manifest as thermosteric and halosteric sea level variations, and since the National Storm Surge Forecast System does not model ocean temperature and salinity, these variations will not be resolved in the model. Furthermore, it is known that these low-frequency modes can modulate the near-shore ocean circulation and, consequently, the near-shore sea level. For example, steric sea level variations have been linked to seasonal variations in the strength of the Leeuwin Current that flows southward along the West Australian coastline (Godfrey and Ridgway, 1985). This may be a factor in simulations of sea level in this region.

Time series of sea level elevations from the model were compared to observed sea level residuals from the ABLSMP tide gauges described in Section 5. Both the model and observed time series have a temporal frequency of one hour. The model time series were taken from the grid point closest to each gauge and plots are shown in Appendix A. Over the 30-month study period, data from the 14 tide gauges was 98.53% complete, with the least complete gauge having 93.33% of observations.

A number of different performance metrics were calculated in order to undertake quantitative assessment of the model. These metrics are bias, Mean Absolute Difference (MAD), Root Mean Square Difference (RMSD), Index of agreement (IOA), Index of efficiency ($E_1$) and two skill indices, skill$_V$ and skill$_R$. Details on how these are calculated are provided in Appendix B. Several authors have discussed how to interpret these indices and how they can best be used to assess model performance. Pielke (1984) argues that for a system to show skill, RSMD should be less than observed standard deviation (skill$_R$ should be less than 1) and the forecast standard deviation should be close to the observed standard deviation (skill$_V$ should be close to 1). Legates and McCabe (2013) argue that a model should ideally have an $E_1$ value at least greater than zero, with a score of 1 representing a perfect score. A score greater than zero means the model is forecasting better than the mean or climatology. Hurley (2000) suggests that the Index of Agreement (IOA) “provides a more consistent measure of performance than the correlation coefficient”, although the updated version of the IOA (Willmott et al., 2011) is used here and it is bound between the values of -1 and 1, with 1 being a perfect prediction.

The calculated indices for each of the time series shown in Appendix A and the means of each index averaged across all sites are listed in Table 7-1. It can be seen that model bias is less than 3 cm, except at Cape Ferguson, Rosslyn Bay, Broome, Darwin and Groote Eylandt, where it is on the order of +5 to +10 cm. Similarly, MADs are less than 10 cm at all locations except Cape Ferguson, Broome, Darwin and Groote Eylandt. The
RMSDs range from less than 7 cm at Port Kembla to around 15 cm at Darwin and Groote Eylandt. The IOA is quite consistent across most sites, with values between 0.6 and 0.7. However, it is around 0.4 at Cape Ferguson, Rosslyn Bay and Groote Eylandt, and at Broome and Darwin the IOA is less than 0.2. There are similar patterns for both the $E_1$ index and $\text{skill}_R$ scores: the sites where $E_1$ is less than zero are Cape Ferguson, Rosslyn Bay, Broome, Darwin and Groote Eylandt. There is no set goal for $\text{skill}_V$ scores but if we arbitrarily aim for forecast variability to be within $\frac{1}{3}$ of the observed variability, then only three sites do not have a $\text{skill}_V$ score in the range 0.67 to 1.33: Port Kembla, Broome and Darwin. These measures of skill are comparable to those of similar international systems and other Australian studies (see Section 2).

Table 7-1: Statistics of model performance for the hindcast runs. All indices are dimensionless except where indicated.

<table>
<thead>
<tr>
<th>Location</th>
<th>Bias (cm)</th>
<th>MAD (cm)</th>
<th>RMSD (cm)</th>
<th>IOA</th>
<th>$E_1$</th>
<th>$\text{skill}_V$</th>
<th>$\text{skill}_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cape Ferguson</td>
<td>9.70</td>
<td>10.24</td>
<td>12.07</td>
<td>0.33</td>
<td>-0.34</td>
<td>0.74</td>
<td>1.26</td>
</tr>
<tr>
<td>Rosslyn Bay</td>
<td>5.84</td>
<td>8.49</td>
<td>10.53</td>
<td>0.40</td>
<td>-0.19</td>
<td>1.03</td>
<td>1.16</td>
</tr>
<tr>
<td>Port Kembla</td>
<td>-1.43</td>
<td>5.20</td>
<td>6.88</td>
<td>0.62</td>
<td>0.24</td>
<td>0.63</td>
<td>0.79</td>
</tr>
<tr>
<td>Spring Bay</td>
<td>2.20</td>
<td>5.76</td>
<td>7.34</td>
<td>0.71</td>
<td>0.41</td>
<td>0.85</td>
<td>0.60</td>
</tr>
<tr>
<td>Burnie</td>
<td>-1.08</td>
<td>6.60</td>
<td>8.52</td>
<td>0.62</td>
<td>0.25</td>
<td>0.83</td>
<td>0.74</td>
</tr>
<tr>
<td>Lorne</td>
<td>0.60</td>
<td>6.18</td>
<td>7.88</td>
<td>0.71</td>
<td>0.43</td>
<td>0.86</td>
<td>0.56</td>
</tr>
<tr>
<td>Portland</td>
<td>-0.99</td>
<td>5.95</td>
<td>7.47</td>
<td>0.69</td>
<td>0.38</td>
<td>0.81</td>
<td>0.61</td>
</tr>
<tr>
<td>Thevenard</td>
<td>-1.92</td>
<td>8.56</td>
<td>10.68</td>
<td>0.70</td>
<td>0.41</td>
<td>0.92</td>
<td>0.56</td>
</tr>
<tr>
<td>Esperance</td>
<td>-0.43</td>
<td>6.33</td>
<td>7.97</td>
<td>0.68</td>
<td>0.37</td>
<td>0.74</td>
<td>0.62</td>
</tr>
<tr>
<td>Hillarys</td>
<td>-0.72</td>
<td>6.40</td>
<td>7.92</td>
<td>0.66</td>
<td>0.32</td>
<td>0.69</td>
<td>0.66</td>
</tr>
<tr>
<td>Broome</td>
<td>10.15</td>
<td>11.44</td>
<td>13.67</td>
<td>0.17</td>
<td>-0.66</td>
<td>0.64</td>
<td>1.55</td>
</tr>
<tr>
<td>Darwin</td>
<td>10.67</td>
<td>12.10</td>
<td>14.43</td>
<td>0.17</td>
<td>-0.67</td>
<td>0.59</td>
<td>1.53</td>
</tr>
<tr>
<td>Groote Eylandt</td>
<td>6.87</td>
<td>12.42</td>
<td>15.19</td>
<td>0.42</td>
<td>-0.15</td>
<td>0.97</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>3.04</strong></td>
<td><strong>8.13</strong></td>
<td><strong>10.04</strong></td>
<td><strong>0.65</strong></td>
<td><strong>0.06</strong></td>
<td><strong>0.79</strong></td>
<td><strong>0.90</strong></td>
</tr>
</tbody>
</table>

Although values of the indices vary from one another and across the sites, some clear patterns emerge from Table 7-1. There is, generally, good model performance at most sites located in the south of Australia, but only fair performance at most northern sites (specifically, Broome, Darwin, Cape Ferguson, Groote Eylandt and Rosslyn Bay. Of all sites, the one where the model appears to perform best is Spring Bay, which is within the top three scores for five of the seven indices. The difference in
performance along the southern coastline compared to the northern coastline of Australia can also be seen in the Bureau’s tide gauge aggregate sea level system discussed in Section 2.

The poorer performance of the model in the tropical region could be attributed to different types of sea level variability there, which are either not or not well represented in the model. As stated earlier, Tropical Cyclones (TCs) in particular are not well represented in the atmospheric forcing (ACCESS-R) and therefore the resulting TC storm surge is unlikely to be well modeled.

There is also a seasonal steric component in tropical waters (Miles et al., 2014) and potentially along the West Australian coastline (Godfrey and Ridgway, 1985), which may not be captured by the modeled sea levels. This is likely to be a contributor to the relatively large bias at the northern sites. A large bias will also contribute to the value of some other indices, such as MAD and RMSD.

In addition, river outflow, particularly after heavy rainfall, may also contribute to variability in the observed sea levels, which is not captured by the model. Each of these factors would contribute to some of the differences seen with the observed signal.

The US National Ocean Service (NOS) recommend that the RMSD of a coastal forecast model should be less than 15 cm to be suitable for operational forecasting (Hess et al., 2003). The results above show that this system is well within that criterion for the mean value and also at all individual sites, with the exception of Groote Eylandt, which is just outside. It is worth noting that the vertical datum adjustment at Groote Eylandt was not available and this could influence the statistics presented for this site. It should also be noted that the US NOS recommendation is based on total sea level (including tides), and so this assessment of the National Storm Surge Forecast System is made under the assumption that the harmonic tide predictions used to determine the observed residual sea level are reasonably accurate.

### 7.2 Non-linear surge-tide interactions

As noted in Section 3, the model has been configured to simulate sea level arising from atmospheric pressure and wind stress forcing. No other sources of sea level variability are simulated by the model. In particular, to compare these model forecasts to observations, the observations have had the tidal component removed by subtracting a harmonic tidal prediction (resulting in a residual, as described in Section 5). However, this assumes that the sea level signals due to tidal forcing and atmospheric forcing do not interact in a non-linear manner.
There is a considerable body of work examining the interactions between tides and storm surges. For example, Horsburgh and Wilson (2007) studied sea levels at 5 tide gauges along the UK coast in the North Sea and found that surges most frequently occurred on rising tides. The key physical mechanism for this is through small phase shifts in the tide and/or surge as they interact and modify the water depth and thus the phase speeds. Similar results were found within the Bay of Bengal by Antony and Unnikrishnan (2013), and on the French coast of the English Channel by Idier et al. (2012), who also found that the amplitude of the tide-surge interactions can reach up to 70 cm at some locations within the Channel. Olbert et al. (2013) investigated the effects of tide-surge interactions on extreme water levels in Irish coastal waters. They found that where interactions exist, they lead to lower total water levels in comparison to the case where they are assumed independent. They also found a strong dependence between surge and phase of the tide, although the specific phase of the tide at which surges tended to cluster varied with location.

Similar studies have also been undertaken for the Australian coastline. For example, Tang et al. (1996) examined the interactions between tides and TC-induced storm surges on the North Queensland coast and found that including the interactions between tide and surges generally led to lower sea levels. In their study of storm surges within Bass Strait, McInnes and Hubbert (2003) found that a significant component of the observed sea level variability could be attributed to a phase delay in the tides, resulting from the interaction between tides and surge. More recently, Mawdsley and Haigh (2016) analyzed a global dataset of 220 tide gauges and undertook a global assessment of the spatial and temporal variability in storm surges. Through an examination of the differences between non-tidal residual and skew surge (the difference between the maximum observed sea level and the maximum predicted tidal level, regardless of timing) they found that the tide-surge interaction was strongest in regions of shallow bathymetry including the north-west shelf of Australia.

To investigate the importance of non-linear interactions between the tide and surge at the locations of interest for the present system, three different configurations of the model were run and evaluated:

- A run that included boundary forcing from a number of calculated tidal constituents in addition to the atmospheric forcing used in previous sections. This was deemed run 1.
- A run that included only the boundary forcing from the tidal constituents. No atmospheric forcing was included. This was deemed run 2.
- A run that included only the atmospheric forcing, identical to the runs in the previous section. This was deemed run 3.
From these, the difference in sea level between run 1 (which incorporates full non-linear interactions between surge and tide) and the sum of runs 2 and 3 (where the tide and surge are assumed to interact in a linear manner) was calculated. This difference provides some guidance as to the expected magnitude of the non-linear surge-tide interactions, hereafter referred to as the ‘non-linear interactions’.

The non-linear interactions were calculated using the method above over the entire model grid for one year, for the period 1-Jan 2014 to 31-Dec 2014 (with a 5-day spin up). Since we are specifically interested in the non-linear interactions at coastal locations, subsequent analysis is only conducted at grid points that are ‘coastal’, i.e. adjacent to one or more land grid points, including islands. There are 39787 of these coastal points on the model grid, which constitute about 1% of unmasked grid points.

The non-linear interactions at the coastal points were found to be mostly small over the entire study period, compared to the sea level variations due to tidal and surge forcing. Figure 7-1a shows the mean absolute deviation of the non-linear interactions over the study period at the coastal points around Australia. The largest mean absolute deviation was less than 11 cm and 98% of coastal points were less than 5 cm. The areas where the absolute deviations were largest are the gulfs of South Australia, the Gulf of Carpentaria, some parts of the northwest coast and Bass Strait and other small bays.

Figure 7-1: a) mean absolute deviation of the non-linear interactions at coastal points. b) maximum absolute deviation of the non-linear interactions at coastal points. Units are in metres.
On a few occasions during the study period, the non-linear interactions were large. Figure 7-1b shows the maximum absolute deviation for the coastal points around Australia. The largest of these maximum absolute deviations at any coastal point was 1.00 m. This occurred at the upper reaches of a bay on the central Queensland coast. However, more generally across the grid, 75% of the coastal points have a maximum absolute deviation less than 20 cm and 99% of points are less than 50 cm. The locations of the largest maximum absolute deviations show close correspondence the locations of the largest mean absolute deviations, with largest values again in enclosed coastal waters, particularly the South Australian gulfs and Bass Strait.

Since much of the model verification is conducted against sea level observations from tide gauges, the non-linear effects at the locations of these tide gauges are examined in more detail. Recall that Figure 3-1 and Figure 5-1 to Figure 5-4 show the locations of the tide gauges used in this study. The mean and maximum absolute deviations of the non-linear interactions at the location of the tide gauges are shown in Figure 7-2. From this figure, we can see:

- The mean absolute deviations are relatively small, around 5 cm or less.
- On occasion, the absolute deviations can be large, with maxima as much as 42 cm.
- The largest values occur at gauges in the South Australian gulfs and Bass Strait, although values are also large in Northern Australia (from Broome to the central Queensland coast).

![Figure 7-2: Mean absolute deviation and maximum absolute deviation of the non-linear interactions at the grid cell nearest to each of the tide gauges used in this study. Units on the vertical axis are in metres.](image-url)
The magnitudes of the non-linear interactions seen here are comparable to those reported by other studies, for example Idier et al. (2012). Given that the non-linear interactions can occasionally be large, they need to be taken into consideration when interpreting the model results. In particular, they could partially explain differences between the model and the observations, especially since the locations where non-linear effects are important are near the coastline and in enclosed areas where some tide gauges are located. Implications for the National Storm Surge Forecast System will be discussed further in Section 8.

7.3 Case studies

In order to test the model on shorter time-scales and in a context closer to its intended use, its performance was evaluated using case studies of past surge events. For this test, the model was run in its operational configuration. After a five-day spin-up, the model was forced using a single ACCESS-R forecast run for 72 hours. The 6-hour forecast from this run was then used to initialize the next forecast, which is forced by the next ACCESS-R forecast (recall that ACCESS-R has a 6-hourly forecast cycle). This method of using the previous run to initialize a given forecast run is known as the recurring warm-start method.

From this method, the first 6 hours can be extracted from each run and concatenated to create what will be deemed the 0-hour forecast. Similarly, the 24 to 29 hour forecasts for each run can be used to create a 24-hour forecast, and so on. In principle, the model should show decreased skill as the forecast range increases.

In addition to calculating the pressure and wind stress-induced surge, the contribution to coastal sea level from wave setup is included, following the method described by O’Grady et al. (2015). AUSWAVE-R forecasts of significant wave height ($H_s$) and peak wave period ($T_p$) are interpolated to the model grid, and then combined with the near-shore coastal slope to produce estimates of wave setup, as described in Section 4. The contributions from surge and wave setup, after being calculated separately, are added together and compared to observed residual sea level. This summed quantity is hereafter referred to as ‘surge + setup’.

Seven events were chosen that were all associated with a significant mid-latitude cyclone, East Coast Low or ex-Tropical Cyclone. In each case, the events were observed on a number of tide gauges that are not part of the ABLSMP network. The locations of these gauges are shown from Figure 5-1 to Figure 5-4.
As described in Section 5, harmonic tidal predictions were subtracted from the observed sea level to produce time series of residual sea level. Furthermore, the observations used in this section have 6- and 10-minute observation intervals. Given this higher frequency of observations (compared to the hourly observations from the ABLSMP network used in Section 7.1), the model was set to output every 10 minutes. The reason for this increased frequency is so that the model is run in a manner close to how it would be used operationally. In particular, accurate determination of the observed and simulated peak storm surge water levels is of prime interest. There are more likely to be discerned using observations and model predictions taken at a higher temporal frequency.
7.3.1 Case Study 1: Victoria, June 2014

During the second last week of June 2014, a mid-latitude cyclone and associated cold front passed to the south of south-eastern Australia. The event was associated with severe and damaging weather, such as high winds and rainfall. The winds and low pressure also impacted marine conditions along parts of the Victorian coastline. Large waves, a tide higher than the predicted astronomical tide and subsequent flooding were observed at many sites, including open coastal locations, enclosed waters such as Port Phillip Bay and estuaries such as the lower Yarra River. (Note that weekly rainfall totals in the catchment area of this river were modest, on the order of 15-25 mm. Therefore, flooding was most likely due to storm surge, rather than from rainfall).

Figure 7-3: Official 0000UTC MSLP analyses for 22, 23 and 23 Jun 2014.
Figure 7.3 shows the official MSLP manual analyses that show the passage of the system across the Australian coastline and includes the peak of the event. In these analyses, a cold front crossed southeastern Australia ahead of a south-westerly airstream that was directed on to the Victorian coastline. The large pressure gradient across this south-westerly airstream suggests high wind speeds.

As discussed in Section 3.1, the model is forced by ACCESS-R pressure and wind stress. These forcing fields were verified against observations to provide confidence that the storm surge model is being forced correctly. Figure 7.4 shows time series of observed MSLP from sites along the south west Victorian coastline and compares them to time series of 0-hour forecast ACCESS-R MSLP from the grid point nearest the observation location. It can be seen that the time series show a high degree of correspondence, with correlation coefficients above 0.99. However, it is important to note that these observations are assimilated into ACCESS-R and so this does not represent an independent comparison. The largest difference here is at Cape Otway, which shows a deviation of 3.8 hPa at 2300UTC on 24/6. Note that, using basic hydrostatic arguments, a rule of thumb is that the inverse barometer effect contributes 1 cm of increased sea level for every 1 hPa of decreased atmospheric pressure (Weisse, 2010).

![Figure 7-4: Time series of observed and ACCESS-R MSLP for three sites along the south west Victorian coastline. Time is in UTC.](image)

It is more difficult to verify the wave setup component directly. Given the wave setup calculation uses AUSWAVE-R forecasts of Hs and Tp, Waverider buoy observations of Hs can be used to verify the AUSWAVE-R forecasts, as an indirect verification of the wave setup. This also has the advantage of being an indirect verification of the surface winds, as it is generally assumed that the dominant source of error in a wave forecast is due to error in the forcing wind fields. Moreover, the wave field at any particular point is a combination of both locally and remotely generated waves (swell), hence the Waverider observations of Hs can be considered an integrated assessment of the forcing wind field.
Figure 7-5 shows a time series of $H_s$ observed at 10-min frequency by the Cape de Couedic Waverider buoy (its location is marked in Figure 5-1), along with the 0-hour AUSWAVE-R forecast from this location. As for MSLP, there appears to be a high degree of correspondence between the two series. The correlation between the series of $H_s$ values at the buoy (taken on the hour) and the AUSWAVE-R forecast has a coefficient of 0.94. Waverider buoy observations from just off Cape Sorrell on the west Tasmanian coast were also examined (not shown) and showed a similar degree of correspondence.

Considering the verifications of both the MSLP and $H_s$, we can be relatively confident that the surface forcing fields during this time period are not a major source of error in the storm surge forecasts.

Sea level observations for the event were obtained from three gauges. These observations are shown in Figure 7-6. In addition to the observed residual sea level, the calculated surge and wave setup for the 0-hour forecasts are shown, as well as the sum of the surge and setup. The verification statistics for the peak surge amplitude and timing are shown in Table 7-2.

At each of the locations, the observations show an increase from 23 June to a peak around 24 June. The modeled surge and setup shows its peak around the same time. The verification statistics in Table
7-2 show that the model is capable of predicting the peak surge within 2 hours of the observed peak, although it is early at two locations and late at the other one.

In general, the total modeled sea level (surge + setup) overestimates the observed surge. This is reflected in the verification statistics which show a mean overestimate of 38 cm (45% of the mean observed surge of 0.85 m). Given that we are relatively confident of the quality of the atmospheric forcing, we should consider other factors to explain the discrepancies.

![Figure 7-6: Time series of calculated surge and wave setup components and their sum (surge + setup) for the June 2014 case study. Also shown is the observed residual sea level, which has a 6-minute interval at all locations. Times are in UTC.](image)

In this case, it can be seen that all three sites, the wave setup component is relatively large, reaching 0.7 m at Portland. While the bathymetric profile along the south west Victorian coastline is relatively steep, (see Figure 7 of O’Grady et al., 2015) and would therefore contribute to high values of wave
setup, a value of 0.7 m here is unlikely to be realistic. This suggests that the wave setup calculation and/or input data should be reviewed and possibly revised.

Another possibility for the discrepancy seen between the observed and modeled surge could be related to the astronomical tides. It can be seen that the amplitude of both the setup and surge components decays immediately following the peak of the event. However, in this period following the peak, there are some oscillations in the observations apparent at Lorne and Point Lonsdale with amplitudes of around 40 cm and 20 cm respectively. The period of these oscillations is approximately 12 hours in each case. This suggests that these oscillations may be tidal in nature, but it is unclear if this is due to an imperfect tidal prediction (and thus imperfect removal from the observations) or an example of surge-tide interaction. The results of Section 7.2 suggest that the amplitude of any non-linear interactions between tide and surge could potentially reach around 25 cm at Lorne and around 30 cm at Point Lonsdale, so even if these interactions are important here, they would not explain the full discrepancies seen in the time series. This requires further investigation.

Table 7-2: Validation of the model (surge + setup) for the June 2014 event (0-hr forecast). Negative time differences imply a model peak that is earlier than the observed peak.

<table>
<thead>
<tr>
<th>Location</th>
<th>Observed max (m)</th>
<th>Model max (m)</th>
<th>Diff. (m)</th>
<th>abs(Diff.) (m)</th>
<th>Time of obs. peak</th>
<th>Time of model peak</th>
<th>Diff. (hrs)</th>
<th>abs(Diff.) (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portland</td>
<td>0.74</td>
<td>1.02</td>
<td>0.33</td>
<td>0.33</td>
<td>23/06/14 23:10</td>
<td>23/06/14 21:20</td>
<td>-1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>Lorne</td>
<td>0.96</td>
<td>1.25</td>
<td>0.29</td>
<td>0.29</td>
<td>24/06/14 00:18</td>
<td>24/06/14 01:30</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Point Lonsdale</td>
<td>0.86</td>
<td>1.38</td>
<td>0.52</td>
<td>0.52</td>
<td>24/06/14 02:36</td>
<td>24/06/14 02:10</td>
<td>-0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>mean</td>
<td>0.85</td>
<td>1.23</td>
<td>0.38</td>
<td>0.38</td>
<td>n/a</td>
<td>n/a</td>
<td>-0.3</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Another question of interest is whether model forecasts improve with decreased lead-time. In principle, forecasts close to the time of the event should be more accurate than forecasts with longer lead-time. For this event, 0-, 24-, 48- and 66-hour forecasts of surge + setup were compared to the observations (note that the 66-hour forecast represents the longest analysis/forecast forcing cycle in a 72-hour forecast). The biases of the mean forecast peak amplitudes (averaged over all four locations) and their times are shown in Table 7-3. Interestingly, the forecasts of the peak amplitude do not improve with decreasing lead-time; indeed, the shortest-range forecast (0-hour) actually performs the worst here. However, the timing of the peak does have a tendency to improve with decreasing lead-time, with the longest-range
forecasts being incorrect by more than 4 hours. However, given the long lead time, this still provides very useful forecast guidance.

Table 7-3: Mean biases in amplitude and timing of peak of surge + setup for increasing forecast lead-time across all sites for the June 2014 event.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Amplitude (m)</th>
<th>Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>abs(Bias)</td>
</tr>
<tr>
<td>0-hour</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>24-hour</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>48-hour</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>66-hour</td>
<td>0.15</td>
<td>0.15</td>
</tr>
</tbody>
</table>
7.3.2 Case Study 2: South Australia, May 2016

This case study occurred in May 2016, affecting the South Australian coastline. A low-pressure system formed and passed through the Great Australian Bight. The low-pressure system was intense, with a lowest central pressure around 978 hPa. The steep pressure gradient associated with the system produced high wind speeds, particularly on its northern and western flanks. The system passed over the South Australian Gulfs and onto the coastline on 9 May and the peak of the surge occurred at high tide. This high tide was, coincidentally, close in amplitude to Highest Astronomical Tide. Consequently, the event produced very high sea levels and coastal flooding. The tide gauge at Port Adelaide observed its highest ever measurement of 3.81m, some 1.09m above the predicted tide of 2.72m. The peak residual of 1.17m had occurred approximately one hour prior.

A series of official MSLP analyses from this time is shown in Figure 7-7. The series of figures show the low-pressure system moving eastward along the South Australian coast, before moving south towards the west of Tasmania. The strong pressure gradient on the rear flank of the system indicates a region of strong winds that are oriented onshore to the South Australia coastline.
The pressure forcing for this case was checked against local observations. Figure 7-8 shows, for three locations in the vicinity of the event, time series of observed MSLP and the ACCESS-R MSLP from the grid cell nearest each location. There is a very high degree of correspondence between the observed and ACCESS-R series, with correlation coefficients above 0.99. This provides confidence in the pressure forcing. As in case study 1, it should be noted that these observations are likely assimilated into the ACCESS-R model so may not provide an independent source of data.
A FIRST-GENERATION NATIONAL STORM SURGE FORECAST SYSTEM

Figure 7-8: Time series of observed and ACCESS-R MSLP for three sites in the region of the South Australian Gulfs (Edithburgh is approximately 10km from Port Giles). Time is in UTC.

As in case study 1, the winds and wave setup are verified indirectly by comparing observations of $H_s$ from the Cape de Couedic Waverider buoy to AUSWAVE-R forecasts for that location. These are shown in Figure 7-9, where it can be seen that observations and forecast show close agreement, with a correlation coefficient of 0.88. Waverider buoy observations from just off Cape Sorrell on the west Tasmanian coast were also examined (not shown) and showed a similar high degree of correspondence.

Figure 7-9: $H_s$ observations from the Cape de Couedic Waverider buoy (blue symbols) and AUSWAVE-R $H_s$ forecasts (orange line). Time is in UTC.

There were seven locations for which suitable tide gauge observations of this event were available. Time series of observed sea level residual are shown in Figure 7-10 along with the forecast surge, wave setup and their sum. The graphs in this figure show that for this event, the model performs extremely well in terms of forecasting the peak amplitudes. Indeed, at each location, the “surge + setup” matches the observations very well in terms of the general variability over the entire time period. The timing of the peak surge is also generally very good, although the model is 3 to 4 hours late at Whyalla and Port Pirie. These results are quantified by the statistics listed in Table 7-4. It can be seen from
Table 7-4 that there is negligible systematic bias, and a very small absolute difference in the peak amplitude for this event. The model peak surge is on average within 4% of the observed peak amplitude.

Figure 7-10 Time series of calculated surge and wave setup components and their sum (surge + setup) for the May 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all sites except Thevenard, which has a 6-minute interval. Times are in UTC.
Although it is noted above that the model peak is somewhat late at some locations, the relevant statistics are actually not reflective of the quality of the forecast. This is due to the model forecast at Thevenard, which shows a large difference (more than 6 hours) between modeled and observed peak time. However, inspection of the top panel in Figure 7-10 shows that the later peak is only marginally higher than the earlier peak (which arrives at the correct time) and is predominantly due to a small increase in the wave setup. If Thevenard is excluded from the statistics, then the mean difference in
timing of the other sites is less than 2 hours although it can be noted that the model is consistently late.

Table 7-4: Validation of the model for the May 2016 event. Negative time differences imply a forecast peak that is earlier than the observed peak.

<table>
<thead>
<tr>
<th>Location</th>
<th>Observed max (m)</th>
<th>Forecast max (m)</th>
<th>Diff. (m)</th>
<th>abs(Diff.) (m)</th>
<th>Time of obs. peak</th>
<th>Time of f’cast peak</th>
<th>Diff. (hrs)</th>
<th>abs(Diff.) (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thevenard</td>
<td>1.25</td>
<td>1.26</td>
<td>0.01</td>
<td>0.01</td>
<td>8/05/2016 17:54</td>
<td>9/05/2016 0:10</td>
<td>6.3</td>
<td>6.3</td>
</tr>
<tr>
<td>Whyalla</td>
<td>1.50</td>
<td>1.50</td>
<td>0.00</td>
<td>0.00</td>
<td>9/05/2016 6:00</td>
<td>9/05/2016 9:50</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>Port Pirie</td>
<td>1.72</td>
<td>1.69</td>
<td>-0.03</td>
<td>0.03</td>
<td>9/05/2016 6:30</td>
<td>9/05/2016 10:10</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>Wallaroo</td>
<td>1.44</td>
<td>1.39</td>
<td>-0.05</td>
<td>0.05</td>
<td>9/05/2016 6:50</td>
<td>9/05/2016 8:10</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Port Giles</td>
<td>0.98</td>
<td>1.02</td>
<td>0.04</td>
<td>0.04</td>
<td>9/05/2016 8:00</td>
<td>9/05/2016 8:50</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Port Adelaide</td>
<td>1.17</td>
<td>1.26</td>
<td>0.09</td>
<td>0.09</td>
<td>9/05/2016 6:50</td>
<td>9/05/2016 7:00</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Victor Harbor</td>
<td>1.03</td>
<td>0.87</td>
<td>-0.16</td>
<td>0.16</td>
<td>9/05/2016 9:00</td>
<td>9/05/2016 9:10</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Mean</td>
<td>1.30</td>
<td>1.28</td>
<td>-0.02</td>
<td>0.06</td>
<td>n/a</td>
<td>n/a</td>
<td>2.3</td>
<td>2.3</td>
</tr>
</tbody>
</table>

We consider again how the quality of the forecasts varies with forecast range in Table 7-5. In this case, there is a slight tendency towards better forecasts for both peak amplitude and peak timing for shorter lead times, with the exception of the 66-hour forecasts, which show good skill in both.

Table 7-5: Mean biases in amplitude and timing of peak for increasing forecast lead-time across all sites assessed for the May 2016 event.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Amplitude (m)</th>
<th>Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-hour</td>
<td>-0.01</td>
<td>2.3</td>
</tr>
<tr>
<td>24-hour</td>
<td>-0.04</td>
<td>1.3</td>
</tr>
<tr>
<td>48-hour</td>
<td>-0.10</td>
<td>3.1</td>
</tr>
<tr>
<td>66-hour</td>
<td>0.00</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

A FIRST-GENERATION NATIONAL STORM SURGE FORECAST SYSTEM
7.3.3 Case Study 3: South Australia, September 2016

This event occurred just 5 months after the case study in Section 7.3.2, affecting the same region of Australia. It, too, arose from the passage of a maturing low-pressure system, which evolved in the Great Australian Bight and intensified before making landfall in south-east Australia near the South Australian Gulfs. This event also peaked near high-tide, producing rough seas and coastal flooding.

Figure 7-11 shows a sequence of official MSLP analyses during the event. It is immediately apparent that this case bears a close resemblance to the equivalent MSLP charts presented for case study 2 (Figure 7-7). This was a deep, mature frontal system that was beginning to occlude. There is a strong pressure gradient on the western and northern flanks of the system, indicating strong onshore winds. Unlike case study 2, as the system moved away from South Australia, it moved into Bass Strait and began to weaken. At this time, the pressure gradient on the rear flank of the system was less intense than that seen in case study 2.

The MSLP and wind forcing from ACCESS-R that was used with the storm surge model was once again verified against observations. Since this event occurred in the same region as case study 2, observed MSLP from the same sites was also used here. Figure 7-12 shows the MSLP verification and Figure 7-13 shows the verification of the winds, again done indirectly by comparing the observed Hs from the Cape de Couedic Waverider buoy to that forecast by AUSWAVE-R at that location. Both the MSLP and Hs forecasts show very close agreement with the observations. The MSLP have correlation coefficients above 0.99 and the Hs forecasts correlate with the observations at 0.97.

However, the AUSWAVE-R forecasts do slightly over estimate Hs at the peak of the event. Waverider buoy observations from just off Cape Sorrell on the west Tasmanian coast were also examined (not shown) and showed a similar degree of correspondence. Given the high agreement between the observations and the forcing fields, this verification analysis is not included for further case studies.
Figure 7-11: Official 0000UTC MSLP analyses for 28, 29 and 30 September 2016.

Figure 7-12: Time series of observed and ACCESS-R MSLP for three sites in the region of the South Australian Gulfs (Edithburgh is approximately 10km from Port Giles). Time is in UTC.
Observations were obtained from eight of the locations shown in Figure 5-1. A harmonic tidal prediction is subtracted from the observations to produce a time series of residual sea level. These are compared to time series of surge, wave setup and their sum in Figure 7-14.

It can be seen that there are some issues with the observations at Port Pirie and Wallaroo. There is a large data gap in the Port Pirie observations, just before what is presumed to be the peak. At Wallaroo, there appears to be significant instrument error, approximately one day after the observed peak. These errors could be sufficient justification for discarding these observations; however, we are interested in just the timing and amplitude of the peak. Comparison with the nearby site of Whyalla suggests that the peak of the surge was recorded correctly at both Port Pirie and Wallaroo. Therefore, we can take the records at both sites as being complete enough to allow comparison with the model.
Figure 7-14: Time series of calculated surge and wave setup components and their sum (surge + setup) for the September 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all sites except Thevenard, which has a 6-minute interval. Times are in UTC.
The observations of residual sea level show this event to have experienced a very large surge. At Whyalla, Port Pirie and Wallaroo, three sites located nearby to each other in the upper reaches of Spencer Gulf, the peak of the surge was in excess of 2 m. All of the observation sites recorded a peak surge amplitude in excess of 1 m.
It can be seen in Figure 7-14 that as in case study 2, the wave setup component is relatively small compared to the surge component, and relatively invariant over time at all locations except Victor Harbor.

As for case study 2, the model in general does a very good job at simulating this event. At most locations, the model simulates the peak surge very well in terms of both the amplitude and the timing. The exceptions to this are an underestimation of the peak amplitude at Whyalla and Port Pirie and to a lesser extent at Port Giles and Port Adelaide. This is reflected in the validation statistics for this event, shown in Table 7-6.

Table 7-6: Validation of the model for the September 2016 event. Negative time differences imply a forecast peak that is earlier than the observed peak.

<table>
<thead>
<tr>
<th>Location</th>
<th>Observed max (m)</th>
<th>Forecast max (m)</th>
<th>Diff. (m)</th>
<th>abs(Diff.) (m)</th>
<th>Time of obs. peak</th>
<th>Time of f'cast peak</th>
<th>Diff. (hrs)</th>
<th>abs(Diff.) (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thevenard</td>
<td>1.74</td>
<td>1.66</td>
<td>-0.08</td>
<td>0.08</td>
<td>28/09/2016 16:30</td>
<td>28/09/2016 17:10</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Port Lincoln</td>
<td>1.19</td>
<td>1.05</td>
<td>-0.14</td>
<td>0.14</td>
<td>28/09/2016 23:10</td>
<td>28/09/2016 23:20</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Whyalla</td>
<td>2.34</td>
<td>1.87</td>
<td>-0.47</td>
<td>0.47</td>
<td>29/09/2016 4:30</td>
<td>29/09/2016 7:20</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>Port Pirie</td>
<td>2.64</td>
<td>2.13</td>
<td>-0.51</td>
<td>0.51</td>
<td>29/09/2016 5:20</td>
<td>29/09/2016 6:10</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Wallaroo</td>
<td>2.02</td>
<td>1.87</td>
<td>-0.15</td>
<td>0.15</td>
<td>29/09/2016 6:10</td>
<td>29/09/2016 6:30</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Port Giles</td>
<td>1.49</td>
<td>1.15</td>
<td>-0.34</td>
<td>0.34</td>
<td>29/09/2016 2:40</td>
<td>29/09/2016 3:40</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Port Adelaide</td>
<td>1.72</td>
<td>1.43</td>
<td>-0.29</td>
<td>0.29</td>
<td>29/09/2016 1:10</td>
<td>29/09/2016 7:20</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>Victor Harbor</td>
<td>1.01</td>
<td>1.15</td>
<td>0.14</td>
<td>0.14</td>
<td>29/09/2016 11:40</td>
<td>29/09/2016 11:30</td>
<td>-0.2</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>1.77</strong></td>
<td><strong>1.54</strong></td>
<td><strong>-0.23</strong></td>
<td><strong>0.27</strong></td>
<td>n/a</td>
<td>n/a</td>
<td><strong>1.48</strong></td>
<td><strong>1.52</strong></td>
</tr>
</tbody>
</table>

It can be seen that on average, there is an underestimation of the peak amplitude of 23 cm, and the mean absolute difference here is 27 cm, which is less than 15% of the observed peak amplitude, suggesting a very good simulation. The timing of the peak amplitude is also very good here, with the
model on average forecasting the peak to occur within approximately 1.5 hours of the observed time. Again, as with the previous case study there is one significant outlier here, and in this case, it is at Port Adelaide, where the modeled peak is more than 6 hours later than the observed peak. However, again, this is not an accurate reflection of the quality of the forecast. In this case, the observations at Port Adelaide show a broad surge, incorporating two peaks of similar magnitude. The model overall depicts the broad surge very well, but does not simulate the two-phased nature of the peak, placing instead, a third shorter period but higher amplitude peak in between them.

An evaluation of the accuracy of the forecasts as a function of lead time in Table 7-7 shows that, as expected, there is a slight tendency for decreasing accuracy for longer lead times for peak surge amplitude and timing, although as for the other case studies, it is fairly noisy.

Table 7-7: Mean biases in amplitude and timing of peak for increasing forecast lead-time across all sites in Figure 7-14.
7.3.4 Case Study 4: Western Australia, September 2014

In September 2014, a mature mid-latitude cyclone passed to the south of Western Australia. Although this system occurred well to the south of the continent, a series of associated cold fronts which extended northward from the system towards the southern West Australian coast, led to a surge event for this region.

The meteorological aspects of this event are depicted in Figure 7-15, which shows a series of MSLP analyses. The first cold front approached the West Australian coast on 5 September. Behind the front was a strong pressure gradient, implying a band of intense south-westerly winds. As it crossed the coast around 0600UTC on 6 September, the front matured and began to decay, and the trailing winds weakened.

A second cold front crossed the coast around 1800UTC on 7 September, again associated with a parent mid-latitude cyclone well to the south. The passage of this front saw the return of an intense south westerly air stream on to the coastline, which persisted until 9 September.

The winds from these two cold fronts caused widespread but relatively minor damage in parts of southern Western Australia. Wind gusts in excess of 90 km/h were recorded at several locations, including a gust of 119 km/h at Cape Leeuwin (Bureau of Meteorology, 2014).

Tide gauge observations for 5 locations were obtained for a period that spans this event. The locations of these gauges are shown in Figure 5-2 and span a length of coastline from Geraldton in the north, around to Esperance in the east, a distance in excess of 1500 km.

Time series of observations from these gauges were obtained and a harmonic tidal prediction was removed to produce time series of residual sea level; these are shown in Figure 7-16. All of the gauges, except perhaps Fremantle, display relatively large-amplitude high-frequency oscillations. They are largest at Bunbury and Esperance, where they are on the order of 20 cm, but occasionally larger. Superimposed on this high-frequency component are low-frequency oscillations due to the storm surge associated with the meteorological events described above.
At all sites, except Esperance, there is a peak in the low-frequency oscillations early on 6 September. The peak is earlier at Geraldton, occurring late on 5 September. The magnitude of the peak is on the order of 50 cm, although at Bunbury, there is a single spike in the residual sea level that extends to 76 cm. This low-frequency peak coincides with the transit across the coast of the first cold front.

From this time, the observed residual sea level decreases until the latter part of 7 September. Another low-frequency peak is observed at all sites at this time, except Esperance, where it occurs late on 8 September. This peak is on the order of 70 cm to 80 cm at all sites except Geraldton, where it peaks at...
60 cm. This second peak coincides with the passage of the second cold front and associated burst of south-westerly winds.

Figure 7-16: Time series of calculated surge and wave setup components and their sum (surge + setup) for the September 2014 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all sites except Esperance and Hillarys, which have 6-minute intervals. Times are in UTC.
Also shown in Figure 7-16 are the modeled surge, the wave setup and their sum. Like the observed residual sea level, the surge component contains oscillations that are relatively high-frequency, superimposed over the lower-frequency surge. Considered alone, the surge underestimates the observations. However, when the wave setup is added to the surge, their sum correlates well with the time series of observations at all five sites. The main difference is that the high-frequency component of the model time series is lower in frequency and amplitude than the high-frequency component of the observations. The low-frequency components (i.e. the surge) of the model and observations match well and this is the component we are most interested in.

Another feature that can be observed are what appear to be discontinuities in the wave setup time series. They are most evident at Geraldton, but closer inspection reveals that similar, smaller jumps are present at the other sites. This is due to noisiness inherent in the peak wave period used in the calculation of wave setup. The wave period spectrum is known to have several peaks, with one peak usually (but not always) larger than the rest. Which peak is largest may change from forecast run to forecast run, and this results in the jumps apparent in the wave setup time series. Normally these jumps are small, but as seen in Figure 7-16, they can occasionally be large.

Table 7-8: Validation of the model for the September 2014 case study. Negative time differences imply a forecast peak that is earlier than the observed peak.

<table>
<thead>
<tr>
<th>Location</th>
<th>Observed max (m)</th>
<th>Forecast max (m)</th>
<th>Diff. (m)</th>
<th>abs(Diff.) (m)</th>
<th>Time of obs. peak</th>
<th>Time of f’cast peak</th>
<th>Diff. (hrs)</th>
<th>abs(Diff.) (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunbury</td>
<td>0.80</td>
<td>0.84</td>
<td>0.04</td>
<td>0.04</td>
<td>7/9/14 14:50</td>
<td>8/9/14 09:00</td>
<td>18.2</td>
<td>18.2</td>
</tr>
<tr>
<td>Fremantle</td>
<td>0.71</td>
<td>0.78</td>
<td>0.07</td>
<td>0.07</td>
<td>7/9/14 15:40</td>
<td>7/9/14 19:00</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Hillarys</td>
<td>0.82</td>
<td>0.88</td>
<td>0.06</td>
<td>0.06</td>
<td>7/9/14 15:38</td>
<td>7/9/14 18:20</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Geraldton</td>
<td>0.60</td>
<td>0.71</td>
<td>0.11</td>
<td>0.11</td>
<td>7/9/14 18:20</td>
<td>7/9/14 17:50</td>
<td>-0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Esperance</td>
<td>0.74</td>
<td>0.76</td>
<td>0.02</td>
<td>0.02</td>
<td>8/9/14 19:42</td>
<td>8/9/14 18:20</td>
<td>-1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Mean</td>
<td>0.73</td>
<td>0.79</td>
<td>0.06</td>
<td>0.06</td>
<td>n/a</td>
<td>n/a</td>
<td>4.5</td>
<td>5.2</td>
</tr>
</tbody>
</table>

With respect to forecasting the peak of the surge, statistics are shown in
Table 7-9. The table shows that with respect to the amplitude of the peak, the mean difference across all five sites is 6 cm, which is approximately 8% of the mean observed peak. The model overestimates at all sites, with the largest difference being 11 cm at Geraldton. Estimates of the timing of the peak are reasonable, with Geraldton and Esperance predicting approximately 1 hour too early, while Fremantle and Hillarys predict the peak approximately 3 hours too late. Bunbury is an outlier, as it predicts a maximum surge of 84 cm on 8 September, while the observed maximum occurred on 7 September. However, the model predicted a secondary maximum 83 cm, at 1320 UTC on 7 September, 3 cm larger and only 1.5 hours earlier than the observed peak.

When considering changes in accuracy with increasing forecast lead time, there is no clear pattern. There is little change in the amplitude bias, and, if anything, a tendency for reduced bias in the timing of the peak. In this case, the 66-hour forecast is arguably the most accurate while the 0-hour forecast is the least accurate. However, the differences between the forecasts are small. This unclear result is similar to that of previous case studies.

Table 7-9: Mean biases in amplitude and timing of peak for increasing forecast lead-time across all sites for the September 2014 case study.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Amplitude (m)</th>
<th>Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>abs(Bias)</td>
</tr>
<tr>
<td>0-hour</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>24-hour</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>48-hour</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>66-hour</td>
<td>0.02</td>
<td>0.04</td>
</tr>
</tbody>
</table>
7.3.5 Case Study 5: Western Australia, May 2016

In the latter half of May 2016, a surge event was observed in southern Western Australia. Like case study 4, it was linked to the passage of two cold fronts that were connected to parent low-pressure systems well to the south of the continent.

The evolution of this event is shown in Figure 7-17. The front approached the coastline during 20 May. As it approached, the front weakened slightly and a small, secondary low-pressure system and front formed in its wake. Associated with this frontal passage was the usual strong pressure gradient and associated high winds, with some wind gust observations in excess of 100 km/h.

Figure 7-17: Official 0000UTC MSLP analyses for 20, 22 and 24 May 2016.
After crossing the coast around 0000UTC on 21 May, the strong winds persisted until late on 22 May, until a high-pressure system moved onshore, and the winds weakened. However, the respite was short-lived, as a second front passed onshore early on 24 May. Although a period of strong winds occurred with this front, they only persisted for approximately 12 hours. During this event, a number of impacts were reported. The passage of the first front coincided with a very high tide, specifically a spring tide. In addition to reports of very high water readings, some erosion was also reported at several beaches in the Perth region.

Tide gauge observations for the same 5 locations used in Section 7.3.4 were obtained for a period spanning this event. Time series of sea level observations had a harmonic tidal prediction removed to produce time series of residual sea level; and are shown in Figure 7-18.

It was previously mentioned that the time of landfall of the first cold front coincided with a spring tide. The resulting high-water level means that the gauge at Bunbury appears to have been ‘overtopped’ around the time of when the maximum surge is suspected to occurred (based on analysis of the time series from Bunbury and nearby Fremantle and Hillarys). Given that it is quite possible that the maximum observed residual at Bunbury was not recorded correctly, the observations from cannot be used in this case study and were excluded.

All of the gauges show the same high-frequency oscillations described in case study 4. They are of similar amplitude and again are largest at Esperance (Despite the missing data, it is suspected that the surge may also have been relatively large at Bunbury). This suggests that there may be location-based, rather than event-based. Superimposed on this high-frequency component are low-frequency ‘bulges’, the first around 0000UTC on 21 May and the second in the latter part of 23 May. These represent the storm surges associated with the meteorological events described above, with their timing coinciding with the passage of the two cold fronts described above. The only site that differs substantially is Esperance, where the first surge occurred in the later on 21 May, and the second surge is significantly smaller.
Figure 7-18: Time series of calculated surge and wave setup components and their sum (surge + setup) for the late May 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at Fremantle and Geraldton and a 6-minute interval at Esperance and Hillarys. Times are in UTC.

Also shown in Figure 7-18 are the time series of calculated surge, wave setup and their sum. At all locations, the surge component shows very good agreement with the observations. However, after adding the wave setup component, their sum overestimates the observed residual sea level (assuming an accurate tidal prediction). The wave setup appears to increase from around 20 cm to 45 cm just after the first surge, before slowly decreasing until the time of the second surge. This increase means
the time series of surge + setup does not decay from the peak of the first surge as quickly as the observations.

Comparisons of the peak amplitude between the model and observations, and their timings are shown in Table 7-10. The model predicts the amplitude well, with the mean of the differences being 10 cm or 15% of the mean observed peak. It should be noted that at all sites, the peaks used to calculate these statistics relate to the one that occurs following the passage of the first front around 21 May, rather than the peak associated with the passage of the second front on 24 May.

Table 7-10: Validation of the model for the late May 2016 case study. Negative time differences imply a forecast peak that is earlier than the observed peak.

<table>
<thead>
<tr>
<th>Location</th>
<th>Observed max (m)</th>
<th>Forecast max (m)</th>
<th>Diff. (m)</th>
<th>abs(Diff.) (m)</th>
<th>Time of obs. peak</th>
<th>Time of f'cast peak</th>
<th>Diff. (hrs)</th>
<th>abs(Diff.) (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fremantle</td>
<td>0.65</td>
<td>0.76</td>
<td>0.11</td>
<td>0.11</td>
<td>20/5/16 22:45</td>
<td>21/5/16 02:40</td>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td>Hillarys</td>
<td>0.74</td>
<td>0.82</td>
<td>0.08</td>
<td>0.08</td>
<td>20/5/16 22:57</td>
<td>21/5/16 05:20</td>
<td>6.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Geraldton</td>
<td>0.51</td>
<td>0.67</td>
<td>0.16</td>
<td>0.16</td>
<td>21/5/16 02:20</td>
<td>21/5/16 03:40</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Esperance</td>
<td>0.66</td>
<td>0.70</td>
<td>0.04</td>
<td>0.04</td>
<td>21/5/16 08:28</td>
<td>21/5/16 16:20</td>
<td>7.9</td>
<td>7.9</td>
</tr>
<tr>
<td>Mean</td>
<td>0.65</td>
<td>0.76</td>
<td>0.10</td>
<td>0.10</td>
<td>n/a</td>
<td>n/a</td>
<td>4.9</td>
<td>4.9</td>
</tr>
</tbody>
</table>

The timing of the peak is not as well estimated as the amplitude. The peak at Geraldton is predicted within 2 hours of the observed peak, but all other sites differ by 4 hours or more.

As in case study 4, there is no clear pattern when considering changes in accuracy with increasing forecast lead time. There is no change in the amplitude bias, however the timing of the forecast peak appears worse for the 48 and 66 hour forecasts.

Table 7-11: Mean biases in amplitude and timing of peak for increasing forecast lead-time across all sites for the late May 2016 case study.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Amplitude (m)</th>
<th>Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>abs(Bias)</td>
</tr>
<tr>
<td>0-hour</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>24-hour</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>48-hour</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>66-hour</td>
<td>0.10</td>
<td>0.10</td>
</tr>
</tbody>
</table>
7.3.6 Case Study 6: New South Wales, June 2016

In the early part of June 2016, the New South Wales coastline, as well as parts of southeast Queensland, eastern Victoria and northern Tasmania, were impacted by an East Coast Low (ECL) event that began around 4 June and persisted until 7 June.

The broad synoptic situation was the interaction of a complex surface low-pressure system with an upper-level trough. The evolution of the event is depicted in the MSLP charts in Figure 7-20. The event produced substantial rainfall (Figure 7-19) that resulted in a number of new June rainfall records for some NSW river catchment areas (Bureau of Meteorology, 2016). In addition to the river flooding that this rainfall produced, there were also high winds and large waves. The event also coincided with the winter solstice spring tide, meaning tidal levels near Highest Astronomical Tide.

![Figure 7-19: Total rainfall between 4 and 7 June 2016 across eastern Australia (from Bureau of Meteorology, 2016).](image)
Figure 7-20: Official 1200UTC MSLP analyses for 4, 5 and 6 June 2016.

Waves during this event caused significant coastal erosion in many areas. The scale of the erosion was considerable at some Sydney beaches, in some cases eroding some 50 m of beach inland. The erosion substantially damaged many coastal properties, roads and other infrastructure.

However, despite these significant coastal impacts, it was suspected at the time that this event did not contain a large storm surge component. In a study conducted shortly after the event, Louis et al. (2016) noted that most of the impact resulted from ‘massive ocean waves’ coinciding with the winter solstice spring tide. They estimated tidal residuals were on the order of 35 cm, which although not insignificant, is certainly not extreme. Nor would one expect such a surge to wreak damage on the scale documented. Indeed, despite the events coincidence with the King Tide, Louis et al. (2016) state
“the tide levels observed along the NSW coastline were not altogether unusual for an ECL event” but that “The most unique aspect of the June ECL from a tidal perspective was the scale and duration of the event.”

Time series of observed sea level were obtained from eight gauges along the NSW coastline (see Figure 5-3). The locations span Brunswick heads in the North to Bermagui on the south coast, a distance over 1000 km. A tidal prediction was subtracted to produce time series of residuals; these are shown in Figure 7-21. At all sites, a low-frequency ‘bulge’ is apparent from around 4 June to the end of the series on 8 June, although it is slightly earlier in northern sites, compared to southern ones (i.e. the bulge begins around 1200UTC on 3 June at Brunswick Heads but around 0600UTC on 4 June at Bermagui Boat Harbour). There is also a high-frequency component that is comparatively small in amplitude, being largest at Coffs Harbour.

Another mode of variability is apparent, as several of the sites show oscillations with periods on the order of 12 hours. The overall duration of these oscillations varies between sites, as do their amplitude. For example, they are evident at Hastings River Breakwater from 4 June through to the end of the series at 8 June and range from 20 to 40 cm. In contrast, the oscillations at Brunswick Heads are much larger, at times 40 to 80 cm, but are only apparent from approximately 4 June until 6 June. It is suspected that these peaks may be due to downstream-flowing river discharge, which increased following the heavy rains during this event (Figure 7-19). This is corroborated by river volume data, which show large increases at some gauges, particularly those in rivers such as Brunswick and Hastings Rivers (these two catchments were among those that received record June rainfall during this event.). Therefore, some of these large oscillations, such as at Brunswick Heads and Hastings River Breakwater, are likely due to water moving ‘downstream’ rather than ‘upstream’, i.e., from a storm surge.

Figure 7-21 shows the time series of the calculated surge wave setup and their sum. It is apparent that for this case study, the wave setup is substantially larger than the surge component, which, for this event is small. The range of the surge component is, at most, approximately 20 cm, while the setup component is approximately 40 to 50 cm. Having the wave setup substantially larger than the surge component makes this case study different to the other case studies examined so far. However, this agrees with the findings of Louis et al. (2016) described above.
Figure 7-21: Time series of calculated surge and wave setup components and their sum (surge + setup) for the June 2016 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all locations. Times are in UTC.
Figure 7-22: Continued.

Despite the unusually small surge component, the sum of the surge and wave setup matches the observations moderately well. Table 7-12 compares the predicted peak surge to the observed peak surge. The difference in peak surge is less than 15 cm at four of the sites, but larger at the other four. Overall, mean errors are around 38% of the observed peak. The model substantially under-predicts the peak at Brunswick heads, while over-predicting at Port Kembla. However, as noted above, some of the under-predictions may be due to a large river discharge component in the observations. The over-prediction at Port Kembla and other sites may be due to the gauges being in sheltered locations where wave setup is small.

Despite the broad nature of this event the model does a reasonable job of predicting the timing of the peak surge, except for two outliers. Time differences are 4 hours or less, except at Yamba and Coffs Harbour, which are over 12 hours too early and late respectively. The mean timing error listed in Table 7-12 is approximately 5 hours, but if Yamba and Coffs Harbour are excluded, this drops to 2.5 hours. The particularly broad nature of this event means that the errors in estimating the timing of the
peak are sensitive to high-frequency oscillations. Moreover, the influence of high river discharge on the observations should also be borne in mind.

Table 7-12: Validation of the model for the June 2016 case study. Negative time differences imply a forecast peak that is earlier than the observed peak.

<table>
<thead>
<tr>
<th>Location</th>
<th>Observed max (m)</th>
<th>Forecast max (m)</th>
<th>Diff. (m)</th>
<th>abs(Diff.) (m)</th>
<th>Time of obs. peak</th>
<th>Time of f’cast peak</th>
<th>Diff. (hrs)</th>
<th>abs(Diff.) (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brunswick Heads</td>
<td>1.15</td>
<td>0.58</td>
<td>-0.57</td>
<td>0.57</td>
<td>4/6/16 16:10</td>
<td>4/6/16 12:50</td>
<td>-3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Yamba</td>
<td>0.42</td>
<td>0.55</td>
<td>0.13</td>
<td>0.13</td>
<td>5/6/16 01:40</td>
<td>4/6/16 13:00</td>
<td>-12.7</td>
<td>12.7</td>
</tr>
<tr>
<td>Coffs Harbour (Inner)</td>
<td>0.50</td>
<td>0.56</td>
<td>0.06</td>
<td>0.06</td>
<td>4/6/16 16:00</td>
<td>5/6/16 04:50</td>
<td>12.8</td>
<td>12.8</td>
</tr>
<tr>
<td>Hastings River Breakwater</td>
<td>0.78</td>
<td>0.66</td>
<td>-0.12</td>
<td>0.12</td>
<td>5/6/16 03:40</td>
<td>5/6/16 02:30</td>
<td>-1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Swansea Channel</td>
<td>0.56</td>
<td>0.64</td>
<td>0.08</td>
<td>0.08</td>
<td>5/6/16 05:20</td>
<td>5/6/16 07:40</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Port Kembla</td>
<td>0.29</td>
<td>0.73</td>
<td>0.44</td>
<td>0.44</td>
<td>5/6/16 17:20</td>
<td>5/6/16 14:20</td>
<td>-3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Crookhaven</td>
<td>1.09</td>
<td>0.71</td>
<td>-0.38</td>
<td>0.38</td>
<td>5/6/16 16:50</td>
<td>5/6/16 15:30</td>
<td>-1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Bermagui Boat Harbour</td>
<td>0.58</td>
<td>0.83</td>
<td>0.25</td>
<td>0.25</td>
<td>5/6/16 17:10</td>
<td>5/6/16 21:10</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Mean</td>
<td>0.67</td>
<td>0.66</td>
<td>-0.01</td>
<td>0.25</td>
<td>n/a</td>
<td>n/a</td>
<td>-0.3</td>
<td>5.1</td>
</tr>
</tbody>
</table>

When considering the effect of increased lead-time (Table 7-13), this event does show a slight tendency for errors in predicting peak amplitude to increase with longer forecasts. Moreover, longer forecasts show consistently smaller peak amplitudes. With respect to predicting the time of the peak, errors grow more significantly with longer forecasts.

Table 7-13: Mean biases in amplitude and timing of peak for increasing forecast lead-time across all sites for the June 2016 case study.

<table>
<thead>
<tr>
<th></th>
<th>Amplitude (m)</th>
<th>Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast</td>
<td>Bias</td>
</tr>
<tr>
<td>0-hour</td>
<td>-0.01</td>
<td>0.25</td>
</tr>
<tr>
<td>24-hour</td>
<td>-0.08</td>
<td>0.26</td>
</tr>
<tr>
<td>48-hour</td>
<td>-0.09</td>
<td>0.26</td>
</tr>
<tr>
<td>66-hour</td>
<td>-0.11</td>
<td>0.28</td>
</tr>
</tbody>
</table>
7.3.7 Case Study 7: Queensland, March 2017

Severe Tropical Cyclone Debbie was a late season system that made landfall on the mid-Queensland coast during late March 2017. The rainfall, winds and associated storm surge during this landfall period were substantial. Although it is intended that this event would normally be forecast using the Bureau’s Tropical Storm Surge System (Greenslade et al., 2018), the event can also be used to assess the quality of the National Storm Surge Forecast System. This case focuses on the surge that occurred after landfall.

Figure 7-22: Official 0000UTC MSLP analyses for 29, 30 and 31 March 2017.
Landfall occurred around 0000UTC on 28 March when Severe TC Debbie was a category-4 storm with a minimum central pressure estimated to be 943 hPa and sustained 1-minute winds to 195 km/h. After making landfall, Severe TC Debbie rapidly weakened into a tropical depression. By 0000UTC on 29 March, its minimum central pressure was analyzed at 996 hPa (Figure 7-22).

The system continued to track inland in a south-westerly direction, before turning towards the east and moving offshore near the South-East Queensland coast. During this post-landfall period, the system continued to produce torrential rainfall, resulting in significant flooding. The Tweed River catchment recorded a record daily rainfall of 374 mm on 31 March, nearly 50 mm above the previous record (Bureau of Meteorology, 2017). Strong winds were also observed with winds gusts around 100 km/h and warnings for dangerous surf conditions issued.

Time series of sea level observations were available from five tide gauges in the area of interest (Figure 5-4). All are located in Queensland, except Tweed Heads, which is located about 200 m across the border in New South Wales. Tidal predictions were subtracted to produce time series of residual sea level (Figure 7-23).

The system moved offshore around 1200UTC on 30 March in the vicinity of the Sunshine coast (of the sites in Figure 7-23, this is nearest to Mooloolaba). The observations show large surges to the north of this region, which are all located north of Fraser Island. They are approximately 0.5m at Gladstone and Burnett Heads, increasing in height to 0.9m at Urangan in Hervey Bay. After peaking, the surge rapidly decreases at all three sites.

At Mooloolaba and Tweed Heads, the observations suggest any surge, if present, was substantially smaller. In particular, Mooloolaba has a spike in the residual sea level around 0700UTC on 30 March, with an amplitude of 30 cm and a duration of just 60 to 90 minutes. At Tweed Heads, there is a broader signal on 30 June that spans many hours and may be storm surge related. Given the record rainfall on 31 March mentioned above, subsequent larger oscillations at this site are suspected to be due to river discharge.

Also shown in Figure 7-23 are time series of the calculated surge, wave setup and their sum. The model appears to simulate the surge at Gladstone, Burnett Heads and Urangan well. It accurately simulates the onset and decay of the surge and matches the amplitude of the peak in the observed
residual sea level, with the exception of underestimating the maximum amplitude at Urangan by 16 cm.

Figure 7-23: Time series of calculated surge and wave setup components and their sum (surge + setup) for the March 2017 case study. Also shown is the observed residual sea level, which has a 10-minute interval at all locations. Times are in UTC.
At the two southern sites, Mooloolaba and Tweed Heads, the modeled surge component is small and on the order of 10 cm, while the wave setup component is larger and on the scale of 20 to 30 cm. The total predicted surge appears to over-estimate the observations by around 10 cm, but as mentioned, the observations on these gauges may be influenced by river flooding from heavy rainfall upstream.

For example, at Tweed Heads, it could arguably be more realistic to compare the model to the peak of 36 cm in the observations that occurs at 1150UTC on 30 March. However, across all five sites, the model predicts the peak surge to within 7 cm of the mean observed peak of 56 cm or to within 13%. The predicted timings of the peak for this case study are also very good, with the exception of one outlier, Tweed Heads. The mean error is 4 hours, but if Tweed Heads is excluded, this drops to less than 0.75 hours.

Table 7-14: Validation of the model for the March 2017 case study. Negative time differences imply a forecast peak that is earlier than the observed peak.

<table>
<thead>
<tr>
<th>Location</th>
<th>Observed max (m)</th>
<th>Forecast max (m)</th>
<th>Diff. (m)</th>
<th>abs(Diff.) (m)</th>
<th>Time of obs. peak</th>
<th>Time of f’cast peak</th>
<th>Diff. (hrs)</th>
<th>abs(Diff.) (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweed Heads</td>
<td>0.57</td>
<td>0.46</td>
<td>-0.11</td>
<td>0.11</td>
<td>31/3/2017 7:00</td>
<td>30/3/17 13:50</td>
<td>-17.2</td>
<td>17.2</td>
</tr>
<tr>
<td>Mooloolaba</td>
<td>0.30</td>
<td>0.40</td>
<td>0.10</td>
<td>0.10</td>
<td>30/3/17 07:50</td>
<td>30/3/17 07:10</td>
<td>-0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Urangan</td>
<td>0.89</td>
<td>0.73</td>
<td>-0.16</td>
<td>0.16</td>
<td>30/3/17 06:00</td>
<td>30/3/17 06:30</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Burnett Heads</td>
<td>0.57</td>
<td>0.57</td>
<td>0.00</td>
<td>0.00</td>
<td>30/3/17 03:30</td>
<td>30/3/17 03:40</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Gladstone</td>
<td>0.48</td>
<td>0.48</td>
<td>0.00</td>
<td>0.00</td>
<td>30/3/17 00:00</td>
<td>29/3/17 22:30</td>
<td>-1.5</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>0.56</strong></td>
<td><strong>0.53</strong></td>
<td><strong>-0.03</strong></td>
<td><strong>0.07</strong></td>
<td>n/a</td>
<td>n/a</td>
<td><strong>-3.7</strong></td>
<td><strong>4.0</strong></td>
</tr>
</tbody>
</table>

As with most of the other case studies, there is no clear pattern of forecasts degrading with increased lead-time. Table 7-15 shows that estimates of the amplitude of the peak surge do not vary much for longer forecasts. Predictions of the timing of the peak slow a slight increase in error with longer lead-times, with the 66-hour forecast showing a notably larger error than the 48-hour forecast.
Table 7-15: Mean biases in amplitude and timing of peak for increasing forecast lead-time across all sites for the March 2017 case study.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Amplitude (m)</th>
<th>Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>abs(Bias)</td>
</tr>
<tr>
<td>0-hour</td>
<td>-0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>24-hour</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>48-hour</td>
<td>0.02</td>
<td>0.11</td>
</tr>
<tr>
<td>66-hour</td>
<td>-0.01</td>
<td>0.09</td>
</tr>
</tbody>
</table>
7.3.8 Summary analysis of all case studies

Section 7.3 has shown for seven case studies that the model is usually capable of accurately forecasting storm surge events. Over all the case studies, the mean accuracy of the maximum surge amplitude was within 16% of the observed peak for all but two cases. For those two exceptions, the mean model accuracy was around 45% (case study 1 and 6).

In this section, the case studies are combined into a single data set to give an overall assessment of the model’s performance. Of the 41 observations across the seven case studies, the mean difference of the forecast peak was 17 cm or 17% of the observed peak. Just over 75% of sites had a modeled peak surge that was within 25% of the observed peak. These statistics supplement those presented in Section 7.1 for the hindcast study.

It should be noted that these summary statistics contain some sites that could arguably be dismissed, such as when the observations include variability that the model is not expected to include (e.g. high river discharge due to heavy rainfall at Brunswick Heads and Hastings River Breakwater in Sections 7.3.6 and at Tweed Heads in Section 7.3.7).
Table 7-16: Summary of model performance across all case studies. Sites that have an absolute difference that is 25% or less than the observed peak surge are shaded in green, and those greater than 25% are shaded in red.

<table>
<thead>
<tr>
<th>Location, Date</th>
<th>observed peak (m)</th>
<th>model peak (m)</th>
<th>abs(model-obs) (m)</th>
<th>percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Victoria, June 2014</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portland</td>
<td>0.74</td>
<td>1.07</td>
<td>0.33</td>
<td>45%</td>
</tr>
<tr>
<td>Lorne</td>
<td>0.96</td>
<td>1.25</td>
<td>0.29</td>
<td>30%</td>
</tr>
<tr>
<td>Point Lonsdale</td>
<td>0.86</td>
<td>1.38</td>
<td>0.52</td>
<td>60%</td>
</tr>
<tr>
<td><strong>South Australia, May 2016</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thevenard</td>
<td>1.25</td>
<td>1.26</td>
<td>0.01</td>
<td>1%</td>
</tr>
<tr>
<td>Whyalla</td>
<td>1.50</td>
<td>1.50</td>
<td>0.00</td>
<td>0%</td>
</tr>
<tr>
<td>Port Pirie</td>
<td>1.72</td>
<td>1.69</td>
<td>0.03</td>
<td>2%</td>
</tr>
<tr>
<td>Wallaroo</td>
<td>1.44</td>
<td>1.39</td>
<td>0.05</td>
<td>3%</td>
</tr>
<tr>
<td>Port Giles</td>
<td>0.98</td>
<td>1.02</td>
<td>0.04</td>
<td>4%</td>
</tr>
<tr>
<td>Port Adelaide</td>
<td>1.17</td>
<td>1.26</td>
<td>0.09</td>
<td>8%</td>
</tr>
<tr>
<td>Victor Harbor</td>
<td>1.03</td>
<td>0.87</td>
<td>0.16</td>
<td>16%</td>
</tr>
<tr>
<td><strong>South Australia, September 2016</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thevenard</td>
<td>1.74</td>
<td>1.66</td>
<td>0.08</td>
<td>5%</td>
</tr>
<tr>
<td>Port Lincoln</td>
<td>1.19</td>
<td>1.05</td>
<td>0.14</td>
<td>12%</td>
</tr>
<tr>
<td>Whyalla</td>
<td>2.34</td>
<td>1.87</td>
<td>0.47</td>
<td>20%</td>
</tr>
<tr>
<td>Port Pirie</td>
<td>2.64</td>
<td>2.13</td>
<td>0.51</td>
<td>19%</td>
</tr>
<tr>
<td>Wallaroo</td>
<td>2.02</td>
<td>1.87</td>
<td>0.15</td>
<td>7%</td>
</tr>
<tr>
<td>Port Giles</td>
<td>1.49</td>
<td>1.15</td>
<td>0.34</td>
<td>23%</td>
</tr>
<tr>
<td>Port Adelaide</td>
<td>1.72</td>
<td>1.43</td>
<td>0.29</td>
<td>17%</td>
</tr>
<tr>
<td>Victor Harbor</td>
<td>1.01</td>
<td>1.15</td>
<td>0.14</td>
<td>14%</td>
</tr>
<tr>
<td><strong>Western Australia, September 2014</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bunbury</td>
<td>0.80</td>
<td>0.84</td>
<td>0.04</td>
<td>5%</td>
</tr>
<tr>
<td>Fremantle</td>
<td>0.71</td>
<td>0.78</td>
<td>0.07</td>
<td>10%</td>
</tr>
<tr>
<td>Hillarys</td>
<td>0.82</td>
<td>0.88</td>
<td>0.06</td>
<td>7%</td>
</tr>
<tr>
<td>Geraldton</td>
<td>0.60</td>
<td>0.71</td>
<td>0.11</td>
<td>18%</td>
</tr>
<tr>
<td>Esperance</td>
<td>0.74</td>
<td>0.76</td>
<td>0.02</td>
<td>3%</td>
</tr>
<tr>
<td><strong>Western Australia, May 2016</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fremantle</td>
<td>0.65</td>
<td>0.76</td>
<td>0.11</td>
<td>17%</td>
</tr>
<tr>
<td>Hillarys</td>
<td>0.74</td>
<td>0.82</td>
<td>0.08</td>
<td>11%</td>
</tr>
<tr>
<td>Geraldton</td>
<td>0.51</td>
<td>0.67</td>
<td>0.16</td>
<td>31%</td>
</tr>
<tr>
<td>Esperance</td>
<td>0.66</td>
<td>0.70</td>
<td>0.04</td>
<td>6%</td>
</tr>
</tbody>
</table>
This report has outlined the design and development of the National Storm Surge Forecast System and described a series of experiments undertaken to validate the model.

Using a 30-month hindcast, it was shown that the model is capable of realistically simulating sea level variations arising from atmospheric pressure and wind stress forcing at most mid-latitude locations of Australia. The model was assessed quantitatively using several different metrics and was found to generally perform well in southern Australia but with reduced performance in northern Australia. This reduced performance in the north can be attributed to some physical limitations of the model and its forcing. However, RMSD was approximately 10 cm and within the 15 cm threshold at all observation sites bar one, as recommended by the US National Ocean Service (NOS) (Hess et al., 2003).

The system was evaluated for 7 case studies. The model predicted the magnitude of the peak surge well for 5 of the cases, to within 16% of the observed peak surges. In case studies 1 and 6, the model predicted the peak surge less accurately, to within 41% and 31% of the observed peaks respectively. The prediction of the timing of the maximum surge was reasonable, although in all cases, there were some outliers. If these are excluded, the timing of the peak is predicted within 2-3 hours of the observed peak.
In each of the cases, the forecast accuracy was assessed with longer lead times. It might ordinarily be expected that forecast skill would decrease as forecast lead-time increases, however this was not always the case. In some cases, the longer forecasts were as accurate as the short-term forecasts. This does suggest that we can have confidence with longer forecasts from the model.

The importance of non-linear interactions between the surge and tides was examined. It was found that the non-linear interactions are typically small and on the order of less than 5 cm. However, at isolated times and locations, they could be much larger, on the scale of 10’s of centimetres, with the maximum absolute deviation calculated to be 1.00 m. However, 75% of the largest absolute deviations at coastal points were less than 20 cm. The locations where these non-linear effects appeared to be most important were typically in semi-enclosed waters, notably the South Australian Gulfs and Bass Strait, northwest coast and the Gulf of Carpentaria. This suggests that a potential area of future work could be to include a suitable tidal model that could allow for explicit calculation of any non-linear interactions with the tide, for example, in a manner similar to that used in the UK operational model (Flowerdew et al, 2013). This might result in the improvement of surge forecasts in regions and at times when these non-linear interactions are significant. However, correct determination of the non-linear interactions relies on the assumption that the tidal calculations are accurate.

There are several other aspects of future work that may lead to improvements to the model. Middleton and Wright (1988) suggest that a barotropic, depth-averaged model will erroneously backscatter coastally trapped short waves in response to changes in bathymetry, particularly in locations where there is an abrupt long-shore change in bathymetry. This scattering would potentially reduce the modeled surge. In contrast, a stratified model (i.e. with multiple vertical levels) would allow these short waves to propagate past the topography and possibly better represent the surface layer. Adding this complexity to the National Storm Surge Forecast System would require further development and increased computational costs, however it may be worth investigating where there are measurable gains from adding even a modest number of vertical levels.

Verifications of this system have been undertaken using observations of coastal sea level at tide gauges. In section 7.1, modelled wave setup was not added to the ROMS storm surge for the comparison while in section 7.3, modelled wave setup was included. While this is somewhat inconsistent, it is not obvious which approach is the most appropriate. Wave setup is expected to be observed at open coasts and beaches, whereas most tide gauges whose data was used in this study are
located in enclosed harbours. Therefore, while at least some wave setup would be observed at some of the tide gauges, the extent to which it would be observed is unknown. Further work should include an examination of Australian tide gauge sites to investigate this issue. One way to do this could be to use a high-resolution wave model which is able to explicitly resolve wave setup, such as SWAN (Booij et al., 1999).

Another area of work that could be undertaken is the nesting of the model within the Bureau’s OceanMAPS ocean forecast model. This would allow OceanMAPS to provide boundary forcing of realistic open-ocean variability to the model, which may be important in some coastal regions.

9 ACKNOWLEDGMENTS

The authors acknowledge the following people for their contributions to this work. Holly Sims and Pavel Sakov provided some computing assistance. Andy Taylor in the Bureau’s National Operations Centre and Thomas Boeck, Kylie Egan, Mike Davis and others in the South Australian Regional Office and National Tides Unit provided observational data. Stefan Zieger provided assistance with the wave verifications. Paul Sandery, Gary Brassington and Tony Hirst provided useful scientific advice. Elaine Miles, Mikhail Entel and Tony Hirst provided helpful reviews of the manuscript.
10 REFERENCES


Figure A-1: Time series of modeled sea level from the grid point nearest the Cape Ferguson tide gauge (orange) and observed sea level residuals from the Cape Ferguson tide gauge (teal). Sea level is in metres and the study period is from 1 January 2014 to 30 June 2016.

Figure A-2: Same as Figure A-1 but for Rosslyn bay
Figure A-3: Same as Figure A-1 but for Port Kembla.

Figure A-4: Same as Figure A-1 but for Spring Bay.
Figure A-5: Same as Figure A-1 but for Burnie.

Figure A-6: Same as Figure A-1 but for Lorne.
Figure A-7: Same as Figure A-1 but for Portland.

Figure A-8: Same as Figure A-1 but for Thevenard.
Figure A-9: Same as Figure A-1 but for Esperance.

Figure A-10: Same as Figure A-1 but for Hillarys.
Figure A-11: Same as Figure A-1 but for Broome.

Figure A-12: Same as Figure A-1 but for Darwin.
Figure A-13: Same as Figure A-1 but for Groote Eylandt.
APPENDIX B: FORMULATIONS FOR STATISTICAL INDICES

The following indices are described in terms of time series, one consisting of observations, $O$, and another of predictions, $P$.

The series of observations, $O$, consists of $n$ time-ordered elements $\{O_1, O_2, \ldots, O_n\}$. Similarly, the series of predictions, $P$, will also consist of $n$ time-ordered elements $\{P_1, P_2, \ldots, P_n\}$. Each element in $P$ corresponds in time to an element in $O$. These series have means, $\bar{O}$ and $\bar{P}$ respectively and standard deviations $\sigma_O$ and $\sigma_P$ respectively.

- **Bias and Mean Absolute Difference (MAD)**

  The bias is a measure of the mean difference between the observations and the predictions, giving an average forecast difference:

  $\text{Bias} = \bar{P} - \bar{O}$

  A positive bias suggests the forecast is, on average, an over-prediction, while a negative bias suggests the forecast is under-predicting. The MAD gives an indication of how close the predictions are to the observations, with no information on whether they are too high or too low:

  $$\text{MAD} = \frac{1}{n} \sum_{i=1}^{n} |P_i - O_i|$$

- **Index of agreement, IOA**

  Willmott (1981) devised an index that aims to specify “the degree to which the observed deviations about the observed mean correspond, both in magnitude and sign, to the forecasted deviations about the observed mean” (Willmott, 1981). This index has been commonly applied in a meteorological context on a wide range scales (from mesoscale to climatological).

  Willmott (2011) sought to make further refinements to the index of agreement, so that it more readily represents a measure of the potential error that is independent of the model. Also, the formulation was adjusted so that the index remains bounded, but between -1 and 1, with perfect agreement indicated by a value of 1:

  $$\text{IOA} = \begin{cases} 
  1 - \frac{\sum_{i=1}^{n} |P_i - O_i|}{2 \sum_{i=1}^{n} |O_i - \bar{O}|} & \text{when } \sum_{i=1}^{n} |P_i - O_i| \leq 2 \sum_{i=1}^{n} |O_i - \bar{O}| \\
  -1 & \text{when } \sum_{i=1}^{n} |P_i - O_i| > 2 \sum_{i=1}^{n} |O_i - \bar{O}| 
  \end{cases}$$
• **Index of efficiency, \( E_1 \)**

Legates and McCabe (1999) developed their index for use in a hydrological context and have termed it the ‘index of efficiency’. It is a measure that compares deviations between the two series to deviations of the observed series from its mean.

\[
E_1 = 1 - \frac{\sum_{i=1}^{n} |P_i - O_i|}{\sum_{i=1}^{n} |O_i - \bar{O}|}
\]

The index has a maximum value of 1, indicating perfect agreement, but is unbounded at its lower end. However, a score less than zero indicates the predictions provide no better forecast of the observations than does the observed mean or climatology (Legates and McCabe, 2013).

• **RMSD and skill\(_V\) and skill\(_R\) indices**

The root mean square difference or deviation is commonly formulated as:

\[
\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2}
\]

It is also often termed root mean square error (RMSE) when the observations are considered ‘truth’ and the deviations of the predictions from this truth therefore represent an error. However, given the observations used in this report are not an exact measure of storm surge, but rather non-tidal residual sea level, the term RMSD is preferred over RMSE.

Given this, the two indices of skill are formulated as:

\[
\text{skill}_V = \frac{\sigma_P}{\sigma_O}
\]

\[
\text{skill}_R = \frac{\text{RMSD}}{\sigma_O}
\]

Pielke (1984) argued that a model shows skill when skill\(_V\) is close to one and skill\(_R\) is less than one.