Extending the Bureau's heatwave forecast to multi-week timescales

Hudson, D. and Marshall, A.G.

September 2016
Extending the Bureau's heatwave forecast to multi-week timescales

Hudson, D. and Marshall, A.G.

Bureau Research Report No. 016
September 2016
Enquiries should be addressed to:

Contact Name: Debra Hudson

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

Contact Email: D.Hudson@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. The POAMA forecast system ....................................................................................3
3. The Excess Heat Factor (EHF) and its application to POAMA data ..................4
4. Verification ................................................................................................................9
   4.1 Hindcasts ..............................................................................................................9
   4.2 Real-time case studies ......................................................................................19
5. Conclusions ..............................................................................................................27

Acknowledgements .......................................................................................................28

References ......................................................................................................................28
List of Tables

Table 1: Overview of the POAMA-2 system ................................................................. 3

Table 2: ROC areas of the probability of occurrence of a heatwave (\(EHF > 0\)) for forecasts initialised in December, January and February. Verification at different lead times is indicated (week 2, week 3, fortnight 1, fortnight 1.5, fortnight 2, month 1). Calculations are shown for all Australian land grid boxes (these ROC curves are shown in Figure 9), as well as for south-eastern and northern Australia. .......................... 16

List of Figures

Figure 1: Example of the Bureau heatwave forecast, showing the forecast for the three day period starting on the 1 January 2015, issued on the 1 January. 4

Figure 2: Average \(EHF\) (\(^\circ\)C)\(^2\) from the AWAP data in the period 1981-2010 for all a) heatwaves (i.e. positive \(EHF\) events), b) severe and extreme heatwaves (i.e. for \(EHF\) events which exceed the 85\(^{th}\) percentile of positive \(EHF\) events), and c) extreme heatwaves (i.e. for \(EHF\) events which exceed three times the value of the 85\(^{th}\) percentile of positive \(EHF\) events). 6

Figure 3: Average annual number of three-day periods (i.e. three-day periods per year) with a) positive \(EHF\), b) \(EHF\) above the severe threshold (\(EHF_{85}\)) and c) \(EHF\) above the extreme threshold, from the AWAP data in the period 1981-2010. 7

Figure 4: The average number of positive \(EHF\) events per year in October to March using the 1981-2010 period from the AWAP data. 8

Figure 5: Maximum percentage area of Australia in heatwave (positive \(EHF\)) within in a given fortnight from AWAP (black) and fortnight 1 forecasts (yellow) (i.e. weeks 1 and 2 of the forecast) in the 1981-2010 hindcast period. Data are available for forecasts initialised on the 1\(^{st}\), 11\(^{th}\) and 21\(^{st}\) of every month. The forecast probability threshold used to turn the probabilistic forecast into a deterministic forecast is 50\%, i.e. it is a “yes” for the occurrence of a heatwave if more than 50\% of the forecast ensemble members predict that a heatwave will occur within the fortnight period. The result is plotted against the first day of the period being forecast. 10

Figure 6: As for Figure 5, but for \(EHF\) above the severe threshold and the forecasts are shown in orange. 11

Figure 7: As for Figure 5, but for extreme heatwaves and the forecasts are shown in red. 12

Figure 8: ROC area (\(A\)) of the probability of the occurrence of a heatwave (\(EHF > 0\)) for all forecasts initialised in the months of December, January and February. The verification at different lead times is shown, i.e. a) week 2 (i.e. days 8-14 of the forecast), b) week 3, c) fortnight 1 (weeks 1 and 2), d) fortnight 1.5 (weeks 2 and 3), e) fortnight 2 (weeks 3 and 4), and f) month 1 (i.e. the first calendar month e.g. if the forecast starts on the 11\(^{th}\) November, then December is verified). ROC areas significant at the 5\(^{th}\) significance level are shown in colour. 14

Figure 9: ROC curves of the probability of the occurrence of a heatwave (\(EHF > 0\)) for all forecasts initialised in December, January and February (x-axis: false alarm rate; y-axis: hit rate). The verification at different lead times is shown, i.e. a) week 2, b)
EXTENDING THE BUREAU’S HEATWAVE FORECAST TO MULTI-WEEK TIMESCALES

week 3, c) fortnight 1, d) fortnight 1.5, e) fortnight 2, and f) month 1. All Australian land grid boxes are used in the calculations. The ROC area (A) is also shown.

Figure 10: Reliability diagrams of the probability of the occurrence of a heatwave (EHF > 0) for all forecasts initialised in December, January and February (x-axis: forecast probability; y-axis: observed relative frequency). The verification at different lead times is shown, i.e. a) week 2, b) week 3, c) fortnight 1, d) fortnight 1.5, e) fortnight 2, and f) month 1. All Australian land grid boxes are used in the calculations. The inset histograms represent the frequency of forecasts in each forecast probability bin.

Figure 11: As for Figure 10, but for south-eastern Australia (see map inset).

Figure 12: As for Figure 10, but for northern Australia (see map inset).

Figure 13: Forecast cases from the summer of 2012/13 for forecasts of fortnight 1 (i.e. weeks 1 and 2 of the forecast are shown). The graph at the top shows the percentage area of Australia in low-intensity (yellow), severe (red) and extreme (brown) heatwave from observations (solid) and from the forecast ensemble (dashed; based on a 50% ensemble probability threshold, as was done in Figures 5-7). The forecast products associated with cases A (verification period: 22/11/2012 - 5/12/2012), B (verification period: 3/01/2013 - 16/01/2013) and C (verification period: 14/02/2013 - 27/02/2013) are shown in the three rows below the graph. The products show the chance of a low-intensity (first column), severe (second column) or extreme (third column) heatwave occurring at any stage in the fortnight being forecast. Also shown is the observed outcome, i.e. the maximum severity level of heatwave observed within the fortnight (fourth column).

Figure 14: As for Figure 13, but for forecasts with a one week lead time (i.e. weeks 2 and 3 of the forecast). The verification periods of cases A, B and C are the same as for Figure 13 (therefore the observed lines on the graph and the observed maps in column 4 are the same in both figures).

Figure 15: As for Figure 13, but for the summer 2013/14 and with verification periods as follows: A: 2/01/2014 - 15/01/2014; B: 12/01/2014 - 25/01/2014; C: 6/03/2014 - 19/03/2014

Figure 16: As for Figure 15, but for forecasts with a one week lead time (i.e. weeks 2 and 3 of the forecast). The verification periods of cases A, B and C are the same as for Figure 15 (therefore the observed lines on the graph and the observed maps in column 4 are the same in both figures).

Figure 17: Heatwave forecasts: from weather to multi-week
ABSTRACT

Connecting weather and climate forecasts, especially for extreme events, is highly desirable for meeting societal needs. Here the extension of the Bureau of Meteorology's heatwave forecast from weather forecast timescales to multi-week timescales is described. The heatwave index has been applied to forecasts from the current operational multi-week to seasonal prediction system, POAMA-2, to provide information about the chance of low-intensity, severe and extreme heatwaves occurring over Australia in the upcoming weeks or month.

The heatwave forecasts have been verified using a hindcast set spanning a 30-year period. The ability of the forecast system to discriminate between the occurrence and non-occurrence of heatwaves, as determined by a ROC (Relative Operating Characteristic) analysis, is better than climatology at all lead times and regions across Australia. This forecast discrimination is particularly good over northern Australia, but less so over central and southern-central Australia. However, the issued probabilities exhibit a marked over-forecasting bias (the probabilities are too high i.e. too emphatic) over northern Australia, whereas the issued probabilities are more reliable over south-eastern Australia.

The results show that there is significant potential to augment traditional weather forecast warnings for extreme heat events to include guidance on longer timescales. The demand from the user-community for forecasts on these longer, multi-week timescales is growing rapidly. It is therefore desirable to expand on this research and development with the Bureau's next generation multi-week and seasonal forecasting system, ACCESS-S, which is currently being developed.
1. INTRODUCTION

In January 2014 the Bureau launched a Pilot Heatwave Program, issuing forecasts out to 7 days of the location of low-intensity, severe and extreme heatwaves over Australia (http://www.bom.gov.au/australia/heatwave/). There is substantial benefit to be gained by extending those forecasts beyond a weather forecast and into the multi-week or intra-seasonal timescale, potentially leading to a seamless service.

Heatwaves are a common feature of Australia’s climate and often result in severe impacts. They are responsible for more deaths in Australia than any other natural hazard including bushfires, floods and storms (Price Waterhouse Coopers 2011). Heatwave warnings are potentially useful across a range of sectors, such as health, agriculture, utilities, infrastructure, tourism and retail. Given that the frequency and intensity of heatwaves is increasing (Plummer et al. 1999, Collins et al. 2000, Griffiths et al. 2005, Alexander et al. 2007, IPCC 2013, Alexander 2016), and that this trend is projected to continue throughout the current century (IPCC 2013), the benefits of producing enhanced and actionable forecasts of heatwaves for climate exposed sectors of society are numerous and of substantial value for mitigating impacts.

The Bureau’s weather prediction heatwave warning is based on the Excess Heat Factor (EHF, Nairn and Fawcett 2013, 2015). A heatwave is defined as three or more consecutive days of high maximum and minimum temperatures that are unusual for a given location, based on two criteria. Firstly, a day is deemed “hot” if it exceeds the climatological 95th percentile threshold (calculated from a number of years of daily mean temperature data for the whole year for the location). Three or more consecutive days need to satisfy this criterion to be classified as a heatwave. The second criterion is a measure of how hot the three days are with respect to the recent past (the preceding 30 days). This criterion takes into account the idea that people can acclimatise to high temperatures, but may not be prepared for a sudden rise in temperature (Nairn and Fawcett 2013). The EHF index is designed to detect the most severe heatwaves (as it is based on raw temperature and not temperature anomalies) and these, by definition, will only occur in the summer half of the year (it does not pick up out-of-summer-season “warm waves”).

We have applied the EHF index to forecasts from the Bureau’s multi-week to seasonal prediction system, POAMA-2 (Hudson et al. 2013), to provide forecasts of the chance of low-intensity, severe and extreme heatwaves occurring over Australia in the upcoming weeks or month. This work forms part of a larger body of research that has investigated the capability of the prediction system to forecast heat extremes on multi-week and seasonal timescales, including understanding the role of key drivers of climate variability in producing heat extremes over Australia in both the observations and the model (Hudson et al. 2015a; Hudson et al. 2015b; Marshall et al. 2014; White et al. 2014).

The multi-week heatwave forecast product described in this report is available for trial use at http://poama.bom.gov.au/. Section 2 of the report provides a brief description of the POAMA forecast system, Section 3 describes the EHF index and its application to the POAMA forecasts, Section 4 assesses the performance of the heatwave forecasts and conclusions are presented in Section 5.
2. THE POAMA FORECAST SYSTEM

The heatwave forecast product is based on the current operational seasonal prediction system run at the Bureau of Meteorology, POAMA version 2. Full details of the model, data assimilation and ensemble generation are available in Hudson et al. (2013) and are summarised in Table 1. As noted in Table 1, hindcasts (retrospective forecasts) are generated on the 1st, 11th and 21st day of each month for the period 1981–2010. The hindcast set has since been expanded to having 6 forecast initialisations per month i.e. a forecast every 5 days, but the results in this report are based on 3-forecast starts per month. Real-time forecasts are generated twice per week (00z Sundays and Thursdays), although prior to January 2013 they were only generated once per week (00z Thursdays).

Table 1: Overview of the POAMA-2 system

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmospheric model</td>
<td>BAM (Bureau atmospheric model, vn 3.0) \nT47 horizontal resolution, 17 vertical levels</td>
</tr>
<tr>
<td>Ocean model</td>
<td>ACOM2 (Australian Community Ocean Model, vn2.0) \n2° in the zonal direction; in the meridional direction it is 0.5° at the equator and gradually increases to 1.5° near the poles. \n25 vertical levels</td>
</tr>
<tr>
<td>Coupler</td>
<td>Ocean Atmosphere Sea Ice Soil (OASIS) coupling software (Valcke et al. 2000)</td>
</tr>
<tr>
<td>Atmosphere/land data assimilation</td>
<td>ALI (Atmosphere Land Initialisation Scheme; Hudson et al. 2011)</td>
</tr>
<tr>
<td>Ocean data assimilation</td>
<td>PEODAS (POAMA Ensemble Ocean Data Assimilation System; Yin et al. 2011)</td>
</tr>
<tr>
<td>Ensemble generation</td>
<td>33 members \nMulti-model (P2a,P2b,P2c) with 11 members from each Burst ensemble: Ocean and atmosphere perturbations from Coupled Ensemble Initialisation Scheme (CEIS) (Hudson et al. 2013)</td>
</tr>
<tr>
<td>Hindcast configuration</td>
<td>33 members on the 1st, 11th and 21st of the month out to 270 days (1981-2010)</td>
</tr>
<tr>
<td>Real-time configuration</td>
<td>33 members every 0z Sundays and Thursdays out to 270 days</td>
</tr>
</tbody>
</table>
3. THE EXCESS HEAT FACTOR (EHF) AND ITS APPLICATION TO POAMA DATA

The weather prediction heatwave service provides forecasts based on the Gridded Objective Consensus Forecast (GOFC) system, which allows forecasts of daily mean temperature out to 7 days (Fawcett and Nairn 2014). The heatwave forecasts provided to the public are for 5 consecutive overlapping three-day periods and the forecasts indicate the locations predicted to experience no heatwave, a low-intensity, severe or extreme heatwave (e.g. Figure 1; http://www.bom.gov.au/australia/heatwave/). The POAMA forecasts show the probability of having a low-intensity, severe or extreme heatwave for lead times beyond 7-days.

As described in Nairn and Fawcett (2013, 2015), there are two components of the EHF calculation. The first component defines the presence or absence of a heatwave and is referred to as the significance excess heat index (EHISig). For a heatwave to be present, the temperature averaged over three consecutive days has to be greater than the climatological 95th percentile (T95) of daily mean temperature for a given region. In other words, the period must be unusually hot with respect to the local climate and thus the value of EHISig, as shown in equation (1), must be positive. Daily mean temperature on a given day (T_i) is calculated as the average of the maximum temperature (Tmax) on day i and the minimum temperature (Tmin) on day i+1. T95 is calculated using all days in the year for the period 1971-2000.

\[ \text{EHISig} = \frac{1}{3} \left( T_i + T_{i+1} + T_{i+2} \right) - T_{95} \]  

(1)

The second component compares the same three-day averaged period to the average temperature of the preceding 30-days. This is the acclimatisation index (EHIAccl), shown in equation (2). If the three-day period is warmer than the recent past (i.e. EHIAccl is positive) then there is a potential lack of acclimatisation to the warmer temperatures, with possible detrimental impacts (Nairn and Fawcett 2013, 2015).

\[ \text{EHIAccl} = \frac{1}{3} \left( T_i + T_{i+1} + T_{i+2} \right) - \frac{1}{30} \left( T_{i-30} + \ldots + T_{i-1} \right) \]

(2)
The \( EHF \) index is the product of these two indices (equation 3), and has units of \((\text{°C})^2\). A heatwave is present, and \( EHF \) is positive, when \( EHI_{\text{sig}} \) is positive. However, if \( EHI_{\text{accl}} \) is also positive, then it will amplify the magnitude of the \( EHF \) value.

\[
EHF = EHI_{\text{sig}} \times \text{max}(1, EHI_{\text{accl}})
\]  

(3)

The location specific thresholds that determine whether a heatwave is severe or extreme are based on all positive \( EHF \) values within the period 1958-2011. The threshold that distinguishes low-intensity and severe heatwaves is the 85th percentile of all positive \( EHF \) values across the 1958-2011 period. A heatwave is deemed to be extreme if the \( EHF \) is greater than three times the value of this 85th percentile. Refer to Nairn and Fawcett (2013, 2015) for details of the rationale behind the selection of these specific thresholds.

As has been done for weather timescales (Nairn and Fawcett 2013, 2015) the \( EHF \) calculation is applied to daily gridded observations of \( T_{\text{max}} \) and \( T_{\text{min}} \) from the Australian Water Availability Project (AWAP) gridded dataset (Jones et al. 2009). However, in our application the AWAP data are first regridded to the POAMA T47 (~250 km resolution) grid. The reason for this, as will be described later, is that the regridded AWAP \( T_{\text{max}} \) and \( T_{\text{min}} \) climatology is used to bias-correct the POAMA forecasts.

The mean positive \( EHF \) value from the AWAP data, as represented on the POAMA grid for the period 1981-2010 (which corresponds to the POAMA hindcast period), is shown in Figure 2a. This is directly comparable to Figure 6 in Nairn and Fawcett (2015), although their \( EHF \) climatology is shown for the period 1958-2011 and is based on the 0.25° AWAP grid. Mean \( EHF \) values are highest in southern Australia and decrease towards northern Australia, indicating that less severe heatwaves occur at lower latitudes. As described in Nairn and Fawcett (2013, 2015), this is a consequence of the lower daily temperature variability in the tropical north, such that although summer temperatures are high, large \( EHI_{\text{accl}} \) values do not often occur. A similar north-south pattern is evident when considering the average magnitude of heatwaves exceeding the 85th percentile of positive \( EHF \) (i.e. all severe and extreme heatwaves, Figure 2b), implying that larger temperature variations are required to produce a severe heatwave in southern regions compared to northern regions (Nairn and Fawcett 2015). The most extreme heatwaves are located in southern South Australia and Victoria (Figure 2c).

Figure 3 shows the average annual number of heatwaves (all positive \( EHF \) three-day periods), severe heatwaves and extreme heatwaves in 1981-2010. There are comparable figures in Nairn and Fawcett (2015) for the period 1958-2011. In general, heatwaves occur more frequently in northern than southern parts of the country (Figure 3a), although, as shown previously they tend to be less severe in the north (Figure 2). The pattern of the frequency of severe and extreme heatwaves (Figures 3b and c) is noisier than when considering all heatwaves, probably because of the reduced sample size. Nairn and Fawcett's (2015) calculation, based on the longer time period, indicates that the severe heatwaves also tend to occur more frequently in the tropical north and that lowest rates are around the southern coastline.

Figure 4 shows the seasonal cycle of the average number of positive \( EHF \) events (all heatwaves). By design, the \( EHF \) index mainly detects heatwaves in the summer half of the year, when temperatures are higher in an absolute sense (the 95th percentile threshold is calculated using daily data from all days in the year). The \( EHF \) index will not detect out-of-summer-season “warm spells”, when temperatures may be high in a relative sense (i.e. high temperature anomalies). Figure 4 shows that January is the peak month for heatwaves, although for northern
In Australia, heatwaves are more common in November and December before the peak of the northern wet season (Figure 4).

Figure 2: Average EHF (°C)² from the AWAP data in the period 1981-2010 for all a) heatwaves (i.e. positive EHF events), b) severe and extreme heatwaves (i.e. for EHF events which exceed the 85th percentile of positive EHF events), and c) extreme heatwaves (i.e. for EHF events which exceed three times the value of the 85th percentile of positive EHF events).
Figure 3: Average annual number of three-day periods (i.e. three-day periods per year) with a) positive EHF, b) EHF above the severe threshold (EHF85) and c) EHF above the extreme threshold, from the AWAP data in the period 1981-2010.
For multi-week and seasonal forecasts it is common practice to bias-correct the forecasts to account for model drift (e.g. Stockdale 1997). This is done by forming anomalies relative to the ensemble mean climatology from the hindcasts. This climatology is a function of the start date, lead time and grid box, and thus a first-order linear correction for model mean bias is made to the forecasts. As shown in section 2, the POAMA climatology period is from 1981-2010 and the hindcasts comprise forecasts starting on the 1st, 11th and 21st of each month in that period. To
calibrate the forecasts for the calculation of EHF values, the model’s forecast Tmax and Tmin anomaly is added to the observed, regridded AWAP Tmax and Tmin climatology respectively (using the period 1981-2010). \(EH_{\text{acc}}\) requires the calculation of the temperature of the 30 days immediately prior to a three-day period in question (see equation 2). For POAMA forecasts at a lead time of \(n\) days into the forecast, \(EH_{\text{acc}}\) is calculated using the mean of the previous \(n\) days of the forecast plus the previous \(30-n\) days of observed data up to the start of the forecast. Since the AWAP data is not readily available in real-time, to calculate the observed data up to the start of the real-time forecast the ALI reanalysis (Hudson et al. 2011) Tmax and Tmin anomalies are added to the AWAP climatology.

For the POAMA forecasts we calculate the probability of having a low-intensity, severe or extreme heatwave at any time within week 2, week 3, fortnight 1 (i.e. weeks 1 & 2), fortnight 1.5 (i.e. weeks 2 & 3), fortnight 2 (i.e. weeks 3 and 4) and in the next full calendar month. The probability of a low-intensity heatwave occurring at each grid point is the percentage of ensemble members for which the \(EH\) is greater than zero at any stage within the forecast period of interest, and similarly greater than the respective thresholds for the probability of a severe or extreme heatwave.

4. VERIFICATION

4.1 Hindcasts

In this section, the heatwave forecasts in the hindcast period (1981-2010) are compared to heatwaves calculated from the AWAP observations.

The heatwave forecasts are probabilistic, but as a first look at verification, we convert the probabilistic forecast to a dichotomous (yes/no) forecast (e.g. is a severe heatwave being forecast or not at each grid box) using a forecast probability threshold, above which the forecast will be ‘yes’ and below which it will be ‘no’. Figure 5 shows the maximum percentage area of Australia in heatwave (positive \(EH\)) within a 2-week period from observations and from the forecast of fortnight 1 (i.e. weeks 1 and 2 of the forecast). Figures 6 and 7 show equivalent graphs, but for severe and extreme heatwaves respectively. Forecasts are available on the 1\textsuperscript{st}, 11\textsuperscript{th} and 21\textsuperscript{st} of every month and the result is plotted against the first day of the period being forecast. A forecast probability threshold of 50\% is used to turn the probabilistic forecast into a deterministic forecast, i.e. it is a ‘yes’ for the occurrence of a heatwave at a particular grid box if more than 50\% of the forecast ensemble predicts that a heatwave will occur within the fortnight period. Note that to some extent this conversion of the probabilistic forecast to a deterministic forecast is arbitrary since in reality the threshold that is chosen depends on the user of the forecast and the decision being made (Wilks 2006). In addition, if the forecast probability is above or below 50\%, irrespective of whether or not the heatwave occurs, the probabilistic forecast cannot be described as “incorrect”. However, as mentioned by Mason (2014), interest in measuring the “correctness” (i.e. was the forecast “correct”) of probabilistic forecasts remains widespread, both amongst end-users and forecasters. More appropriate measures of the performance of probabilistic forecasts are shown later, but this “correctness” of a deterministic forecast does provide some insight into the forecast. The 50\% probability threshold is also the threshold above which we use colour shading in the forecast product (e.g. see Figure 13). This means that users will inadvertently focus on regions with probabilities of heatwave occurrence greater than this threshold.
Figure 5: Maximum percentage area of Australia in heatwave (positive EHF) within a given fortnight from AWAP (black) and fortnight 1 forecasts (yellow) (i.e. weeks 1 and 2 of the forecast) in the 1981-2010 hindcast period. Data are available for forecasts initialised on the 1st, 11th and 21st of every month. The forecast probability threshold used to turn the probabilistic forecast into a deterministic forecast is 50%, i.e. it is a “yes” for the occurrence of a heatwave if more than 50% of the forecast ensemble members predict that a heatwave will occur within the fortnight period. The result is plotted against the first day of the period being forecast.
Figure 6: As for Figure 5, but for EHF above the severe threshold and the forecasts are shown in orange.
Figure 7: As for Figure 5, but for extreme heatwaves and the forecasts are shown in red.
In general, there is good correspondence between the forecast and observations for the percentage area of Australia in heatwave for the first fortnight of the forecast (Figure 5). There are, however, a number of occurrences where the percentage area is overestimated in the forecast, most noticeably in the summers of 1983/84, 1985/86, 1988/89, 1994/95 and 2001/02. There are fewer occurrences where the percentage area is underestimated, with the summer of 2006/07 being a noticeable exception (although, Figure 6 shows that the percentage area of Australia experiencing a severe heatwave in 2006/07 was relatively accurately forecast).

Over the 30-year period there is a tendency for the percentage area in severe heatwave to be overestimated (Figure 6) and this is also particularly the case when considering extreme heatwaves (Figure 7). The observations in Figure 7 underscore the rarity of extreme heatwaves, such that in many summers there are no detectible extreme heatwaves on the scale of the graphs shown and when they do occur, they generally cover less than 5% of the area of Australia. The largest observed area in extreme heatwave (about 10%) occurred in the first 2 weeks of January 1994 (Figure 7) and was focused over eastern Queensland and northern New South Wales (not shown). The forecasts were able to warn of this heatwave, although in terms of the diagnostic calculated for Figure 7, over-forecast the area expected to be in extreme heatwave by a factor of two. However, the region of higher forecast probabilities (>70%) very closely matched the area that was observed to experience an extreme heatwave (not shown). Apart from the general over-forecasting of extreme heatwaves, including false alarms, there are also some missed extreme heatwave events, in particular, in October 1987, November 1990, February 2004 and November 2009 (Figure 7).

As mentioned previously, verification of this nature can, however, be misleading, since the probabilistic nature of the forecasts is being removed. Also, there is nothing in this particular analysis that considers the location of the heatwaves. Subsequent evaluation will take into account the probabilistic nature of the forecasts through appropriate verification metrics and will also consider other forecast lead times.

The rarities of the severe and extreme heatwaves pose challenges for verification due to the small sample size. As such, we have chosen to look at the severe and extreme heatwaves in case studies in Section 4.2. Note: for the experimental forecast product (http://poama.bom.gov.au/bom/ehf.shtml) we also show all the available heatwave forecast cases from the entire 30-year hindcast period and these are accompanied by a map showing what was actually observed. This facilitates an easy case-by-case assessment of the quality of the forecasts.

Assessment of the probabilistic forecasts in the remainder of this section deals with all heatwaves (i.e., $EHF > 0$) and are based on all forecast start dates within the months of December, January and February (i.e. 270 forecast cases comprising 30 years, 3 months, 3 start-dates per month), the key months for heatwaves (Figure 4). The metrics used are the Relative Operating Characteristic (ROC) score, ROC curve and reliability diagram (e.g. Mason and Graham 2002; Wilks 2006). These metrics are used for verifying the performance of dichotomous (yes/no) forecasts, in this case, whether there is a heatwave ($EHF > 0$) or not. They are based on contingency tables of the number of observed occurrences and non-occurrences of a heatwave falling into predefined forecast probability bins (we use 10 equally-sized forecast probability bins).

The ROC score (also referred to as the ROC area, $A$) measures the ability of the forecasting system to discriminate between events and non-events, thereby providing information on
forecast resolution (discrimination). The ROC curve is produced by plotting the hit rate (fraction of observed events that were correctly forecast) against the false alarm rate (fraction of non-events that were incorrectly forecast as events) calculated for each probability bin. The no-skill line on an ROC curve is the diagonal, where hit rates equal false alarm rates. A forecast system with positive skill has a curve which lies above the diagonal and bends towards the top left corner, such that hit rates exceed false alarm rates. The area under the ROC curve is often used to summarize the skill. It is normalized such that a perfect forecast system has an area of 1 and a curve lying on the diagonal (no resolution) has an area of 0.5, the threshold for positive skill. ROC areas greater than 0.5 imply that the forecast system has some ability to discriminate between the occurrence and non-occurrence of a heatwave. Figure 8 shows the ROC area of the probability of the occurrence of a heatwave (\(EHF > 0\)) for summer forecast start months at different forecast lead times, i.e. week 2, week 3, fortnight 1 (weeks 1 and 2), fortnight 1.5 (weeks 2 and 3), fortnight 2 (weeks 3 and 4), and month 1 (the first calendar month i.e. if the forecast starts on the 11th November, then December is verified). Statistical significance of the ROC area is determined using the Mann–Whitney \(U\) statistic (Mason and Graham, 2002; Wilks, 2006). Virtually all of Australia exhibits ROC scores significantly greater than 0.5 (at the 95% confidence level) for all forecast lead times, indicating generally good forecast resolution. ROC areas are highest for the forecast of the first fortnight (Figure 8c). Most of this skill comes from the first week of the forecast (not shown) and then the skill rapidly degrades going to week 2 and then week 3 (Figures 8a, b). There is not much difference in forecast discrimination between fortnight 1.5 (weeks 2 and 3) and fortnight 2 (weeks 3 and 4). In general, for all lead times, poorest resolution or discrimination is found over central and southern-central Australia. There is particularly good forecast discrimination over northern tropical Australia at all lead times.

![ROC area](image)

Figure 8: ROC area \((A)\) of the probability of the occurrence of a heatwave \((EHF > 0)\) for all forecasts initialised in the months of December, January and February. The verification at different lead times is shown, i.e. a) week 2 (i.e. days 8-14 of the forecast), b) week 3, c) fortnight 1 (weeks 1 and 2), d) fortnight 1.5 (weeks 2 and 3), e) fortnight 2 (weeks 3 and 4), and f) month 1 (i.e. the first calendar month e.g. if the forecast starts on the 11th November, then December is verified). ROC areas significant at the 5% significance level are shown in colour.
These results are further summarised in the ROC curves in Figure 9 and are based on all land grid boxes over Australia contributing to the contingency tables. As mentioned previously, the forecast lead time with the best resolution is fortnight 1 (i.e., weeks 1 and 2; $A=0.76$) and the lead time with the lowest resolution is week 3 ($A=0.64$). This is related to both the decline in skill with lead time, as well as the ability to compensate by aggregating periods (e.g. weeks into fortnights). These ROC areas are also provided in Table 2, together with the ROC areas for the northern Australian region and the south-eastern Australian region. Forecast resolution over northern Australia is high and consistently better than all-Australia and south-eastern Australia at all lead times.

Figure 9: ROC curves of the probability of the occurrence of a heatwave ($EHF > 0$) for all forecasts initialised in December, January and February (x-axis: false alarm rate; y-axis: hit rate). The verification at different lead times is shown, i.e. a) week 2, b) week 3, c) fortnight 1, d) fortnight 1.5, e) fortnight 2, and f) month 1. All Australian land grid boxes are used in the calculations. The ROC area ($A$) is also shown.
Table 2: ROC areas of the probability of occurrence of a heatwave ($EHF > 0$) for forecasts initialised in December, January and February. Verification at different lead times is indicated (week 2, week 3, fortnight 1, fortnight 1.5, fortnight 2, month 1). Calculations are shown for all Australian land grid boxes (these ROC curves are shown in Figure 9), as well as for south-eastern and northern Australia.

<table>
<thead>
<tr>
<th></th>
<th>All Australia</th>
<th>South-Eastern Australia</th>
<th>Northern Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2</td>
<td>0.69</td>
<td>0.68</td>
<td>0.72</td>
</tr>
<tr>
<td>Week 3</td>
<td>0.64</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td>Fortnight 1 (weeks 1+2)</td>
<td>0.76</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>Fortnight 1.5 (weeks 2+3)</td>
<td>0.69</td>
<td>0.67</td>
<td>0.73</td>
</tr>
<tr>
<td>Fortnight 2 (weeks 3+4)</td>
<td>0.68</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td>Month 1</td>
<td>0.70</td>
<td>0.66</td>
<td>0.75</td>
</tr>
</tbody>
</table>

The above assessment says nothing of the reliability of the issued probabilities for the occurrence of a heatwave. Reliability diagrams, as shown in Figure 10, show the conditional relative frequency of occurrence of an event (observed relative frequency) as a function of the forecast probability, thereby providing information on forecast reliability. The diagonal line on the diagrams represents perfect reliability and any deviation from that line gives the conditional bias. The relative sample size in each forecast probability bin is shown as a histogram inset on each graph. Most of the reliability diagrams for the lead times shown in Figure 10 show a tendency towards either over-confidence (points lie below the diagonal for high probabilities and above the diagonal for low probabilities) or over-forecasting (points lie below the diagonal, thus forecast probabilities are all generally too high). The latter is the case for forecasts of fortnight 2 (Figure 10e). Reliability is worst for forecasts of week 3 (Figure 10b) and best for forecasts of month 1 (Figure 10f).
EXTENDING THE BUREAU’S HEATWAVE FORECASTS TO MULTI-WEEK TIMESCALES

Figure 10: Reliability diagrams of the probability of the occurrence of a heatwave (EHF > 0) for all forecasts initialised in December, January and February (x-axis: forecast probability; y-axis: observed relative frequency). The verification at different lead times is shown, i.e. a) week 2, b) week 3, c) fortnight 1, d) fortnight 1.5, e) fortnight 2, and f) month 1. All Australian land grid boxes are used in the calculations. The inset histograms represent the frequency of forecasts in each forecast probability bin.

We have also examined the reliability over the two areas mentioned previously, south-eastern (Figure 11) and northern (Figure 12) Australia. Forecasts over south-eastern Australia tend to be over-confident (Figure 11), which is a common situation for multi-week and seasonal forecasts (Hudson et al. 2013, Mason and Stephenson 2008). The fortnightly forecasts (Figure 11c, d, e) exhibit fairly good reliability and are more reliable than the weekly forecasts (Figure 11a, b). The month 1 forecasts (Figure 11f) appear very unreliable for the lower forecast probabilities, but the results should be viewed in the context of the frequency histogram which shows that the sample size in those bins is small. Deviations from perfect reliability in the diagram can be due to sampling limitations rather than necessarily true deviations from reliability (e.g., Toth et al. 2003), and that may be the case here. The reliability diagram for month 1 (Figure 11f) indicates that the higher probabilities, which are being issued more frequently (see the histogram), are fairly reliable.

The reliability curves for northern Australia (Figure 12) are noticeably different from those for south-eastern Australia. Interestingly, although there is generally good forecast discrimination over northern Australia at all lead times (Figure 8 and Table 2), the issued probabilities exhibit an over-forecasting bias for all lead times i.e., the issued probabilities are too high (Figure 12).
The over-forecasting bias is worse for the higher forecast probabilities. For example, the worst reliability is seen for week 3 forecasts (Figure 12b), where an 80% issued forecast probability is associated with a 35% relative observed frequency, a far cry from a one-to-one relationship.

Figure 11: As for Figure 10, but for south-eastern Australia (see map inset).
4.2 Real-time case studies

In this section, example cases from the real-time forecasts are shown. Figures 13 and 14 show forecast cases from the summer of 2012/13 for forecasting a fortnightly period at zero lead (i.e. fortnight comprising weeks 1 and 2, Figure 13) and at one week lead (i.e. fortnight comprising weeks 2 and 3, Figure 14). The graph at the top of Figure 13 shows the percentage of Australia in low-intensity, severe and extreme heatwave from observations and from the forecast ensemble (based on a 50% ensemble probability threshold, as was done in Figures 5-7) over the summer period. The graph was used to select three forecast cases for examination. The selected cases, indicated on the graph (A, B and C), correspond to peaks in the observed heatwave coverage. All three cases are associated with peaks for the observed low-intensity heatwaves, cases A and B are also associated with peaks for the severe heatwaves and case B for extreme heatwaves. The forecast products associated with each of these cases are shown in the maps on Figure 13 (a row of maps for each forecast case). The forecasts show the chance of a low-intensity (first column), severe (second column) or extreme (third column) heatwave occurring at any stage in the fortnight being forecast. Also shown is the observed outcome, i.e. the maximum severity level of heatwave observed within the fortnight (fourth column).
Figure 13: Forecast cases from the summer of 2012/13 for forecasts of fortnight 1 (i.e. weeks 1 and 2 of the forecast are shown). The graph at the top shows the percentage area of Australia in low-intensity (yellow), severe (red) and extreme (brown) heatwave from observations (solid) and from the forecast ensemble (dashed; based on a 50% ensemble probability threshold, as was done in Figures 5-7). The forecast products associated with cases A (verification period: 22/11/2012 - 5/12/2012), B (verification period: 3/01/2013 - 16/01/2013) and C (verification period: 14/02/2013 - 27/02/2013) are shown in the three rows below the graph. The products show the chance of a low-intensity (first column), severe (second column) or extreme (third column) heatwave occurring at any stage in the fortnight being forecast. Also shown is the observed outcome, i.e. the maximum severity level of heatwave observed within the fortnight (fourth column).
Case A shows forecasts initialised on the 22 November for the fortnight of 22 November-5 December (Figure 13). Virtually all of eastern Australia experienced a heatwave within that fortnight, with south-eastern and far northern Australia experiencing severe heatwaves. The forecasts for this period performed well, with the areas forecast to have more than a 50% chance of a low-intensity heatwave closely matching the observed outcome. The areas showing more
than a 50% chance of having a severe heatwave correspond to the observed outcome (i.e. severe heatwaves eventuated) over parts of the south-eastern and northern Australia. Forecasts of this fortnight one week in advance (initialised on 15 November, Case A on Figure 14) had forecast probabilities that were not as emphatic, but still effectively warned of the occurrence of low-intensity and to some extent, the severe heatwaves (particularly over far northern Australia), in that fortnight.

Case B was associated with the most extreme heatwaves for the summer of 2012/13. January 2013 was Australia’s hottest January and month on record and there were major bushfires in south-eastern Australia and Tasmania (Bureau of Meteorology 2013). The heatwave at the beginning of January was notable in terms of the spatial extent of high temperatures (Bureau of Meteorology 2013 and Figure 13) and the 7th January holds the record for being the hottest day for Australia as a whole. The area-averaged maximum temperature for Australia exceeded 39 °C on seven consecutive days from 2-8 January (the previous record was 4 days in December 1972) (Bureau of Meteorology 2013). Case B includes forecasts for the fortnight of the 3-16 January at zero lead time (Figure 13, forecast initialised on the 3 January) and at one week lead time (Figure 14, forecast initialised on the 27 December). POAMA strongly warned (probabilities exceeding 90% over most of the country) of the upcoming heatwave conditions over the whole of Australia at both lead times. A possible exception was the forecast for Tasmania, where the forecast probabilities for the chance of a heatwave were low (Figure 13, 14: Case B column 1). The peak of the January heatwave over Tasmania fell within this fortnight, on the 4th January (Bureau of Meteorology, 2013). The horizontal resolution of the POAMA model is ~250 km (T47 grid) and this, together with the grid configuration, means that Tasmania is not resolved as land. The forecast is therefore reflecting ocean surface temperatures from the model. This highlights a need to improve model resolution in future versions of this model.

A large proportion of Australia, apart from northern regions, experienced severe heatwaves within the verifying fortnight (Figure 13, Case B, 4th column). At zero lead time POAMA strongly warned of upcoming severe heatwaves over most of the areas where severe heatwaves eventuated (Figure 13, Case B, 2nd column). Exceptions include far northern Australia, where the model indicated high probabilities of there being severe and even extreme heatwaves (Figure 13, Case B, 3rd column), where in fact only low intensity heatwaves occurred; and over far Western Australia where the model showed no strong indication of severe or extreme heatwaves occurring, and severe and extreme heatwaves were observed. At one week lead time (Figure 14) the model strongly warned of the chance of severe heatwaves over south-eastern Australia, as did occur.

For Case C (14-27 February), POAMA correctly warned of the upcoming heatwave conditions over western and far southern Australia, but the predictions over northern Australia indicate the potential for severe and even extreme heatwaves, which did not eventuate (Figure 13). The latter is particularly the case for the forecasts at one week lead time (Figure 14).

Figures 15 and 16 show forecast cases from the summer of 2013/14 for forecasting a fortnightly period at zero lead and at one week lead respectively. Case A was selected as it is associated with a peak in observed heatwave coverage and severity (Figure 15). Case B was selected to look in more detail at the potential over-forecast of severe heatwave conditions, and Case C for the over-forecast of low-intensity heatwave conditions (Figure 15). For Case A, virtually all of Australia experienced a heatwave during the fortnight of 2-15 January 2014 (Figure 15, Case A, 4th column). POAMA effectively warned of the upcoming low-intensity heatwave conditions at
both zero- (Figure 15, Case A, 1st column) and one-week lead times (Figure 16, Case A, 1st column). At both lead times, POAMA also warned of the severe and extreme heatwave conditions that eventuated over north-eastern Australia (although the forecast was less emphatic at one-week lead). In contrast, the model did not warn of the severe conditions over much of southern Australia and the extreme conditions over the south-east (Figures 15 and 16, Case A, 2nd and 3rd columns).

For Case B (12-25 January 2014), it is clear on the graph that the model over-estimated the area in severe heatwave (using a 50% probability threshold). The forecast maps show that there is a very good match between the high forecast probabilities for a severe heatwave (probabilities >90%) and the observed outcome over the southern part of Australia and the model correctly warns of the extreme heatwave conditions over parts of south-eastern Australia (Figure 15). However, the model forecast over northern Australia does not verify well, with the model forecasting the chance of extreme heatwaves in the fortnight, when no heatwaves (of any severity) eventuated (Figure 15). At one week lead time, the forecast for severe heatwave conditions over southern Australia are less emphatic and the forecast for extreme heatwaves is located over the coast of New South Wales rather than over coastal Victoria.

Case C, encompassing the fortnight of 6-19 March 2014, was a period of very little heatwave activity. Low-intensity heatwaves were observed in the Carnarvon region of Western Australia, Tasmania and north of Brisbane. The model strongly indicated the chance of low-intensity heatwaves over the former two regions, but not for the Brisbane region (with low probabilities < 30% for that region). The model also indicated a relatively high chance (>50%) of there being low-intensity and severe heatwaves over regions where they did not eventuate, leading to the over-forecast of the area in heatwave that can be seen in the graph.

A key feature of this multi-week heatwave product is that it is an extension of the Bureau’s weather forecast heatwave product and there is therefore the potential for use as a seamless forecasting tool, across weather and climate timescales. Connecting weather and climate forecasts, especially for extreme events, is highly desirable and has been flagged as a priority among users. Figure 17 (also shown in Hudson et al. 2015a) shows examples of the multi-week heatwave forecasts for January 2014, presented in a “Ready-Set-Go” framework along with weather forecasts. The “Ready-Set-Go” concept was proposed by the Red Cross Climate Centre and the IRI (International Research Institute for Climate and Society) in the context of humanitarian aid and disaster preparedness and demonstrates possible actions that could be taken with seasonal, multi-week and weather forecasts (Hellmuth et al. 2011). For example, with the seasonal forecasts (“Ready”) one may update contingency plans, enable early warning systems, train volunteers; with the multi-week forecasts (“Set”) one may alert volunteers, mobilise assessment teams, undergo local preparation activities; and then with the short-range weather forecasts (“Go”) one may deploy the assessment team and activate the volunteers and response plan. This “Ready-Set-Go” concept could be useful in other sectors, such as health, for framing the utility of the forecasts on these different timescales.
Figure 15: As for Figure 13, but for the summer 2013/14 and with verification periods as follows: A: 2/01/2014 - 15/01/2014; B: 12/01/2014 - 25/01/2014; C: 6/03/2014 - 19/03/2014
Figure 16: As for Figure 15, but for forecasts with a one week lead time (i.e. weeks 2 and 3 of the forecast). The verification periods of cases A, B and C are the same as for Figure 15 (therefore the observed lines on the graph and the observed maps in column 4 are the same in both figures).
Example: January 2014

One of the most significant multi-day heatwaves on record affected southeast Australia over the period from 13 to 18 January 2014.

**POAMA Forecasts (chance of a heatwave occurring in the period)**

Forecast start date on **29 December 2013** for the month of **January**

Forecast start date **5 January 2014** for **12 to 25 January** (i.e. weeks 2 & 3)

**Weather (NWP) Forecasts for 13 to 15 January**

Forecast start date **8 January 2014**
Forecast start date **12 January 2014**

Figure 17: Heatwave forecasts: from weather to multi-week
5. CONCLUSIONS

There is significant potential to extend the Bureau’s heatwave warning system to multi-week timescales. The weather forecast product (http://www.bom.gov.au/australia/heatwave/) is based on the Excess Heat Factor index (EHF, Nairn and Fawcett 2013), where a heatwave is defined as three or more consecutive days of high maximum and minimum temperatures that are unusual for a given location. Heatwaves are classified as low-intensity, severe or extreme. Here the EHF index is applied to forecasts generated by the Bureau’s coupled model multi-week to seasonal prediction system, POAMA-2. Whilst the weather forecast product is deterministic, the multi-week product is probabilistic. It shows the chance of a low-intensity, severe or extreme heatwave occurring at any stage in the upcoming weeks, fortnights or month (http://poama.bom.gov.au/bom/ehf.shtml).

Verification of the multi-week forecast product, using the 30-year hindcast set, has shown that POAMA-2 has generally good skill in detecting heatwaves at all lead times over Australia (i.e. week 2, week 3, fortnight 1, fortnight 1.5, fortnight 2 and month 1 of the forecast), with ROC areas significantly better than climatology. This ability to discriminate between the occurrence and non-occurrence of a heatwave is particularly good over northern tropical Australia, but less so over central and southern-central Australia.

However, there is a strong over-forecasting bias of the issued probabilities at all lead times over northern Australia (i.e. the probabilities are too high). In contrast, the issued probabilities are more reliable over south-eastern Australia. Analysis of case studies from the summers of 2012/3 and 2013/14 demonstrate the potential utility of the heatwave forecasts, showing that there were many instances of the system strongly warning of upcoming heatwaves that did eventuate. However, there does seem to be a tendency to overestimate the area of Australia in heatwave, especially for severe and extreme heatwaves. There is scope to further improve the utility and skill of this forecast product by additional calibration and by altering the display by changing the threshold above which colour shading is used (users will inadvertently focus on regions with probabilities of heatwave occurrence greater than this threshold).

This multi-week heatwave product shows the potential for seamless forecasting across weather and climate timescales, and is in the process of being trialled by specific industry clients. In the meantime, the Bureau is developing a new multi-week and seasonal forecasting system, ACCESS-S (the seasonal prediction version of the Australian Community Climate and Earth-System Simulator), which will have significant enhancements compared to POAMA-2. These enhancements should lead to an increased capability to forecast extremes, primarily due to an improved simulation of the climate, very much enhanced spatial resolution, and state-of-the-art physics parameterisation schemes. It is highly desirable to continue this research and development of heatwave (and other extreme events) forecasting with ACCESS-S, including developing the capability for seamless forecasting across the weather-climate interface. Connecting weather and climate predictions, especially for extreme events, is highly desirable for a range of sectors of society.
ACKNOWLEDGEMENTS

This work was partially supported by the Managing Climate Variability Program (MCV00031). We would like to thank John Nairn and David Jones for their useful reviews of this manuscript.

REFERENCES


