

# How thunderstorms appear in weather forecasts

B. Price and M. Foley

*Science and Innovation Group, Bureau of Meteorology, Australia  
Email: michael.foley@bom.gov.au*

**Abstract:** Thunderstorms present a range of dangers to the community. Therefore, information about thunderstorms in the Australian Bureau of Meteorology's 7-day forecasts for the public can be important for peoples' decision making. In these forecasts, thunderstorms are described with likelihood-related text phrases such as 'possible thunderstorm' and 'thunderstorm likely'. There is a complex series of steps that lead to the placement of thunderstorms in the forecast. The process, which has been shaped by several historical factors, is linked to the probability of precipitation and involves both human-driven and automated aspects.

The thunderstorm forecasts lack specific explicit quantitative definitions, which makes it hard to assess the quality of the service. However, by measuring the characteristics of the forecasts, informed by knowledge of the production process, we can start to understand the nature of the service currently being provided. We have done this using a complete set of national thunderstorm forecasts issued during Summer (December to February) 2020/2021, comparing these forecasts to thunderstorm observations derived from lightning strike data. This has given a quantitative understanding of what forecasts currently represent, which can inform future decisions around service definitions and improvements to the production process.

From these results, we highlight notable differences that exist in the current service that depend on climatic zones, forecast horizons, and particular offices issuing the forecast. We observe an overall trend whereby the same forecast terms represent increasingly larger storm likelihoods as we move from temperate parts of the country to the tropics. This is driven by climatological factors and local forecast practices. Additionally, we observe a convergence of forecast terms towards the climatological frequency with increasing forecast horizon.

We suggest what a nationally consistent and clearly defined service, driven by thunderstorm probability information, might look like. The findings indicate that moving towards such a service, consistent at the national level, will result in service changes over different parts of the country and different forecast horizons. This poses questions around how the service should be best designed to most effectively support decision making.

Once a thunderstorm definition has been established, forecast skill and accuracy can be measured. This facilitates objective improvement of the forecast production system including determination of where human input does or does not add value to the service, over automated approaches.

**Keywords:** *Thunderstorm forecast process, probabilistic forecasts, automation, forecast verification*

## 1. INTRODUCTION

Thunderstorms are significant weather phenomena which pose numerous risks and hazards to the community. Lightning strikes from thunderstorms lead to several deaths and injuries in Australia each year, as well as damage to property (Coates et al. 1993). 'Severe' thunderstorms bring additional threats such as large hail, flash flooding, damaging wind gusts and even tornadoes (Allen et al. 2011). Therefore, the risk of thunderstorms is a significant aspect of the information presented in the weather forecast.

Forecasting thunderstorms and the potential for severity is a specialised activity with many decision points (Colquhoun 1987, Doswell 2001). In this paper, we will describe a subset of this activity: the process by which the thunderstorm forecast feeds into the Australian Bureau of Meteorology's (BoM) 7-day public weather forecasts. This process includes automated steps and allows for manual involvement, and its design shapes the characteristics of the BoM's current thunderstorm forecasts.

The performance of the thunderstorm forecasts within the BoM 7-day forecast service has not been measured previously. We will present the initial results of a comparison of these thunderstorm forecasts to observations of lightning, for Summer (December to February) 2020/2021. We note differences in the nature of these forecasts for different climatic zones, different forecast horizons and also for different forecast offices. We will discuss how manual approaches to the process contribute to these variations and the implications of the differences for the BoM's current and future forecast services.

The BoM forecast production processes are undergoing significant transformation, including the use of greater levels of automation (Just and Foley 2020). By better understanding the way thunderstorms are represented within the current 7-day forecast service, and by establishing a more clearly defined service, we will be better placed to improve the service and forecast process, including how manual and automated elements feed into decisions about the level of thunderstorm risk. We will discuss the implications of our results, which relate to these questions. The ultimate outcome of such investigations should be an improved forecast service, which will better enable the Australian community to decide how they conduct thunderstorm-sensitive activities.

## 2. THUNDERSTORMS IN THE FORECAST PRODUCTION PROCESS

### 2.1. Representation of Thunderstorms in Public Weather Forecasts

The BoM's public weather forecasts, available at <http://www.bom.gov.au/> and on the 'BOM Weather' app, are expressed in several different ways. These include: precis forecasts which describe each day's weather for a location in a few words; corresponding daily weather icons; and more detailed forecasts which describe the weather for a location or across an area (e.g. Sydney metropolitan area or Central Victorian forecast district) in several sentences. Gridded forecasts are displayed as 'MetEye' graphics at <http://www.bom.gov.au/australia/meteye/> and are available as data in the Australian Digital Forecast Database (Bureau of Meteorology 2020). The information is available at a daily timescale for the rest of today and the next seven days, which we will describe as 'lead days' numbered from 0 to 7. There is 3-hourly timescale information available in the Australian Digital Forecast Database and some sub-daily information can be given in the detailed text. The risk of thunderstorms is described in the forecasts in various ways using qualitative terms as indicated in Table 1. The terms are linked to 5 categories related to the production process to be described in the next section.

**Table 1.** How thunderstorms are represented in forecast products, relationship to thunderstorm categories and time periods that the information covers.

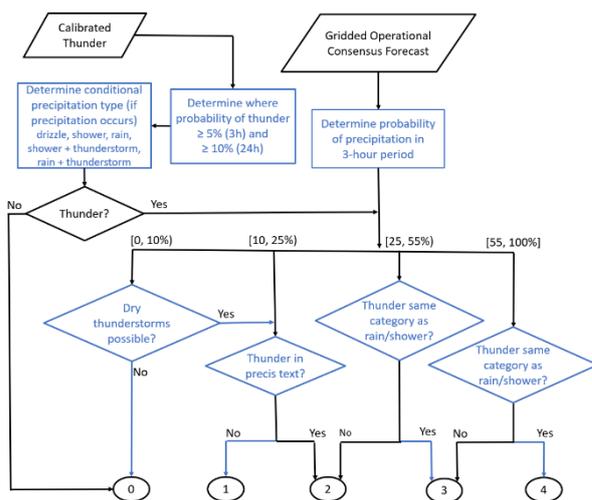
Cat. (3h)	Precis text/icons (24h)	Detailed text (24h)	MetEye/Digital Forecast Database (3h)
0	No mention	No mention	No mention
1	No mention	Chance of thunderstorm	Low probability
2	Possible thunderstorm	Chance of thunderstorm	Low probability
3	Thunderstorm likely	Thunderstorm likely	Low probability
4	Thunderstorms	Thunderstorms	High probability

## 2.2. The Grid-Based Forecast Production Process

From 2014 the BoM's 7-day public weather forecasts for all of Australia have been produced using the Graphical Forecast Editor (Hart 2019). In this system, forecasters curate gridded forecasts of weather conditions and produce text and icon forecasts with the help of natural language generation software. Forecast production has been conducted separately in 7 Regional Forecast Centres responsible for each of the Australian states and the Northern Territory, with the regional grids being combined into a national grid at around 6-km resolution. During 2021, as part of the transformation of production, the process is moving to a national one, which is operated using a Graphical Forecast Editor instance covering the whole country.

The gridded forecasts include information about near-surface conditions for elements of the forecast such as temperature, humidity, wind, and chance and amount of precipitation. There is also indication of the presence of notable weather phenomena within 3-hourly periods conforming to a nominal standard 1500 UTC 'midnight' applied across all Australian time zones, for the rest of today and the next 7 days. This includes a representation of the thunderstorm forecast in one of five categories for each 3-hour period.

For historical reasons the thunderstorm forecast categories in the Graphical Forecast Editor are labelled 'None', 'Chance', 'Isolated', 'Scattered' and 'Widespread'. What is being represented by these terms is not precisely defined. The categories include a mix of probabilistic terms ('Chance') and spatial distribution terms which come from aviation forecasting parlance ('Isolated', 'Scattered' and 'Widespread' in order of increasing spatial prevalence). The presence of these terms in the software reflects the way weather forecasts were written by weather forecasters before the move to grid-based forecast production, and the terms were used directly to describe thunderstorms and showers in text produced by early versions of the software. However, as can be seen in Table 1, currently the categories are mapped exclusively to qualitative probabilistic terms. Given the lack of a clear definition, we will refer to the thunderstorm forecast categories with labels 0 to 4.



**Figure 1.** How thunderstorms are mapped to forecast categories for 3-hour periods. Black indicates a standard process and blue indicates manual input.

The Graphical Forecast Editor receives forecast information from a range of Australian and international numerical weather prediction models, as well as from a statistically postprocessed blend of such models known as the Gridded Operational Consensus Forecast (Engel and Ebert 2012). The forecaster can take this information and use it to populate the gridded forecast. A variety of graphical editing tools are provided to enable adjustments to be made to the incoming information. These range from simple 'paint box' tools to tools which use science-based relationships between parameters to adjust the grids.

A key design principle which enables the forecaster to manage the forecasting process is that as many elements of the forecast as possible can be built up from a smaller number of base grids that they oversee. As far as possible, a standard process is defined, which could produce the forecast process in an automated fashion, but which the forecaster

can adjust if needed based on their expert judgement. In the case of thunderstorms, such a process is represented in Figure 1. Precipitating types of weather are controlled by two base grids to capture the decisions: how likely is precipitation (3-hourly probability) and if precipitation occurs, what type is expected?

The Gridded Operational Consensus Forecast provides the main input for the precipitation probability grids. Thunderstorms can be included in the conditional precipitation type grids in several ways. The standard process indicated in Figure 1 uses forecasts of thunderstorm probability within a 10 km radius ('Calibrated Thunder') following Bright et al. (2005), which have recently been implemented within BoM based on Australian model inputs. However, forecasters are accustomed to using other approaches to set thunderstorms in the conditional precipitation type grids. This includes the use of editing tools that diagnose deep atmospheric instability, a necessary ingredient for thunderstorm formation, as discussed for example by Colquhoun (1987). (A comprehensive discussion of approaches to thunderstorm forecasting is beyond the scope of the current paper.)

The probability of precipitation is mapped to four categories based on ranges as shown in Figure 1. In earlier versions of the text generation software, these were used to define phrases such as 'scattered showers'. If thunderstorms are present in the precipitation type grid, they typically occur with a lower frequency than the

main precipitation type and thus will be described with a lower category, for instance, 'scattered showers and isolated thunderstorms'. Although the descriptive terms are no longer used, and the text for non-thunderstorm precipitation types is now based on other information (the 24-hour probability of precipitation), mapping to a lower category for thunderstorms remains in the process, at the discretion of the forecaster. The forecaster can also choose the lower category 1 for circumstances where they feel that the thunderstorm risk does not warrant indication on the less-detailed forecast products. (The thunderstorm categories are not yet based directly on the thunderstorm probabilities, because this forecast information has only recently become available.)

Conversely, there can be situations (where precipitation evaporates before hitting the ground) where thunderstorms can be indicated in the forecast as 'dry thunderstorms' even when the precipitation probabilities are negligible. The forecaster has tools to add this to the weather phenomena forecast. This is a less common occurrence, and the analysis in this paper does not separate out this category.

Following preparation of the forecast grids, the forecasters run software tools to generate natural language text forecasts from the information in the gridded forecasts. This involves much summarisation. When there are different thunderstorm categories across the time period being described, the highest numbered category is used. Therefore, for the products in Table 1 that show daily forecasts, the input is the maximum thunderstorm category from the eight 3-hourly periods covering the day. It is possible for forecasters to edit the machine-generated text. However, because of the large number of public weather forecasts, this is only done in occasional high-profile cases such as for metropolitan areas or during significant community events where forecasters take particular care to refine messaging to support community decision making.

### 3. MEASURING PROPERTIES OF THUNDERSTORM FORECASTS

#### 3.1. Methodology

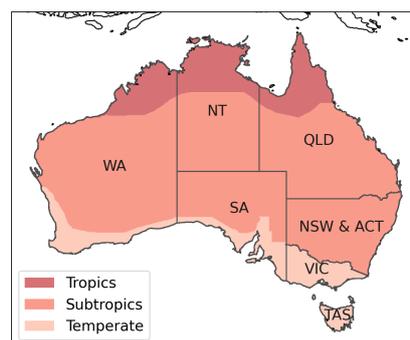
To gain insights into the nature of the thunderstorms forecasts from the production process described above, we have analysed one summer (December 2020 to February 2021) of gridded 3-hourly forecasts of thunderstorm category (0 to 4), as produced in the Graphical Forecast Editor. Forecasts spanning lead day 1 (tomorrow) to lead day 7 have been examined.

As, by definition, a thunderstorm must have associated lighting, we have compared forecasts to observed lightning from the Weatherzone Total Lightning Detection Network, which is part of the Earth Networks Total Lightning Network (Rudlosky 2015). Gridded lightning strike counts were derived from the number of lightning strikes which occurred within a 10 km radius in 3-hour periods corresponding to the forecasts. The radius was chosen to match thunderstorm definitions currently in use within the Bureau. An observed thunderstorm event for a period at a grid cell was recorded if there was a non-zero lightning count. It is estimated that less than 1% of strikes in the Weatherzone dataset are false positives (M. Palmer, personal communication, April 29, 2020). Therefore, the occurrence of spurious observations is not expected to affect the results significantly.

Data have been aggregated across each of the State and Territory areas of responsibility of the 7 offices issuing the forecasts, as well as 3 climatic zones: 'tropics', 'subtropics', and 'temperate' and all of Australia. Only results over land have been included, as we are focusing on land-based forecasts. The analysis areas are indicated in Figure 2.

As well as analysing the 3 hour forecast periods, we have combined the data into whole day results which are comparable to the way the information appears in the BoM's text forecasts. Consistent with the natural language generation rules, such 'maximum-in-day' results use the maximum numeric forecast thunderstorm category for each grid cell in each day aligned with the 1500 UTC standard midnight, and an observed thunderstorm event is recorded if there was any lightning observed in that day.

Where confidence intervals are shown on results, these have been generated using a bootstrapping technique (Efron and Tibshirani 1994). The data for the season is divided into 90 daily samples which are then randomly resampled with replacement to build up a distribution of possible outcomes, using 10,000 trials in total. Confidence intervals are then derived from this distribution. There are various complex spatial and temporal



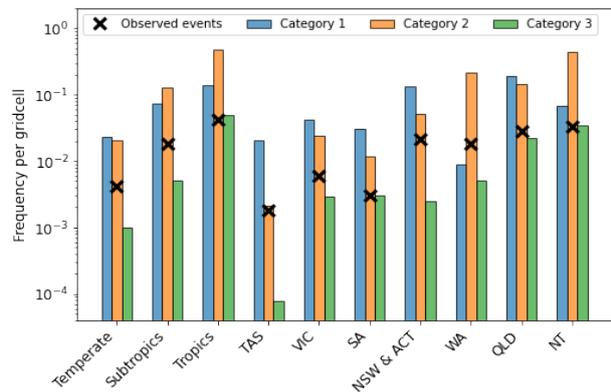
**Figure 2.** The administrative and climatological zones analysed in this study.

dependencies in the thunderstorm data which mean that this will not be a rigorous assessment, but it gives a rough indication of the comparative certainty of different results.

### 3.2. Results

Figure 3 displays separately the frequencies of observed thunderstorm events and frequencies of forecasts of thunderstorms for all the 3-hourly periods for all lead day 1 forecasts.

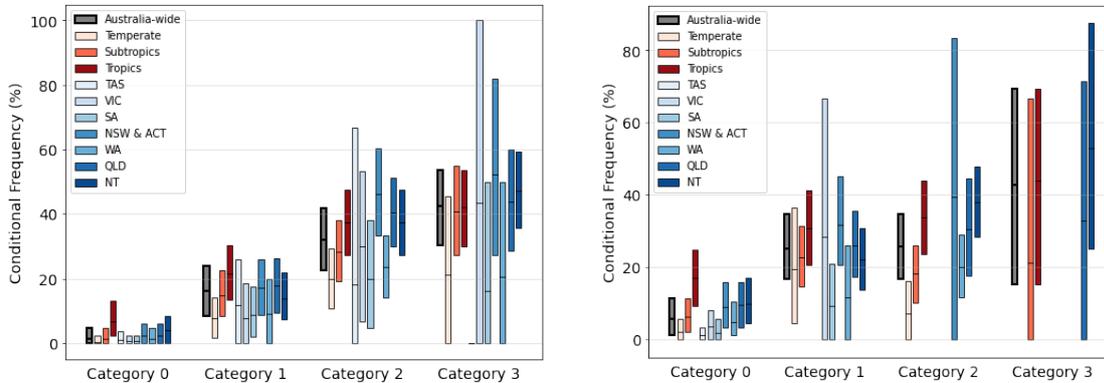
There is a clear trend moving from temperate zones to the tropics of thunderstorms being forecast more frequently, corresponding to their more frequent occurrence. There is also a trend towards issuing category 2 more frequently than category 1.



**Figure 3.** Frequency of category 1, 2 and 3 forecasts and observed events for each administrative and climatological zone for lead day 1, on a log scale.

Greater variation is evident between forecast offices. Tasmania issues eleven times as many category 1 forecasts as category 2 while Western Australia issues twelve times as many category 2 forecasts as category 1. Tasmania issues an order of magnitude fewer category 3 forecasts than any other office.

We estimate the performance characteristics of the text forecast products outlined in Table 1 by examining the observed frequencies of a thunderstorm event in a day, binned based on maximum-in-day forecast category from 0 to 3. Results are shown in Figure 4 for lead days 1 (left) and 7 (right) to indicate how the forecast performance changes between early and late in the forecast range. Category 4 results are not shown, as forecasts for this category were only issued at lead day 1 for a small area on a single day within the study period. For some of the aggregation areas and lead days there are no results shown due to similar lack of data (e.g. Tasmanian category 3 forecasts at lead day 1). Mean results are shown, along with 95% confidence intervals.



**Figure 4.** Observed frequencies conditional on maximum-in-day category 0 to 3 forecasts for lead day 1 (left) and 7 (right). The bars represent 95% confidence intervals, with the mean indicated by a horizontal line.

Maximum-in-day lead day 1 category 1 forecasts ('chance of thunderstorm' in detailed forecasts but no mention in precis forecasts) on average across Australia correspond to around a 16% likelihood of a thunderstorm occurring during that day. Results differ between forecast offices; however, the most significant difference is between the temperate and the tropics zones (around 8% versus around 21%, respectively). On average, category 2 forecasts ('possible thunderstorm' precis forecasts) correspond to around a 32% likelihood of a storm on a given day. As with category 1 forecasts, differences between climate zones are again observed for category 2, ranging from 20% for the temperate zone to 37% for the tropics. Differences emerge between forecast offices for this category. In particular, New South Wales, Northern Territory and Queensland-issued forecasts all represent a significantly larger likelihood of thunderstorms than South and Western Australia. Category 3 maximum-in-day forecasts ('thunderstorm likely' text forecasts) results are generally noisy, with far fewer forecasts issued in this category (see Figure 3), and it is not clear if there is a significant difference in thunderstorm likelihood compared with category 2. Finally, for category 0 forecasts (no mention of a

thunderstorm in the text forecasts), significant differences are observed between the temperate zone and the tropics (around 0% and 7% likelihoods, respectively).

In general, results for lead day 7 forecasts exhibit much greater uncertainty than results for lead day 1, with a wider confidence interval for temperate zone forecasts. Nationally, lead day 7 category 1 forecasts represent around a 25% storm likelihood. Interestingly, there is no clear progression of increasing likelihood with category across 1, 2 and 3, with mixed signals for different areas. In the tropics, lead day 7 no-storm forecasts (category 0) represent notably larger storm likelihoods (17%) compared to lead day 1.

### 3.3. Discussion

The existing process for placing thunderstorms in the forecast is suboptimal, as it is hampered by lack of clear service definitions, has been shaped by historical service considerations which are no longer relevant, and does not make full use of emerging probabilistic thunderstorm forecast information such as is provided by Calibrated Thunder. The link of the thunderstorm forecasts to the overall probability of precipitation in the current process is problematic because the chance of precipitation is defined on the scale of a rain gauge, while a thunderstorm is experienced if thunder and lightning occurs within some wider radius, such as the 10km definition used by Calibrated Thunder. As the current service is expressed in probabilistic terms (Table 1), it would be simpler to have a process whereby the text was generated using grids of thunderstorm probability directly, with appropriate thresholds. The place of probabilities over 3-hourly and 24-hourly periods in supporting daily and sub-daily aspects of forecast products would need to be determined. The results of this study both act to inform decisions on this future service definition and facilitate a better understanding of possible implications of changing the existing service to one driven by such definitions.

A future service definition based on daily storm probabilities (%) for categories 0 to 3, could for instance use ranges [0-10), [10-25), [25-40), and [40-55) respectively (with category 4 above). This definition is roughly based on the current Australia-wide results for lead day 1 from Figure 4. By adopting consistent meanings, the service would offer a simple and consistent set of quantitative definitions for users around the country.

Such a definition would have notable differences to the existing service for different climate zones. Consider, for example, category 2 forecasts (which result in a 'possible thunderstorm' forecast in the precis text). After implementation of this new fixed definition, this forecast would represent a roughly 10% greater likelihood of storms over temperate zone locations than it currently does. Conversely, this forecast would represent a decrease of almost 10% relative to the current service in the tropics. Existing differences in observed frequencies between these climate zones are likely due to differing treatments of category 1 and 2 forecasts by individual forecast offices. This is presumed to be at least in part driven by local climatology (with thunderstorms more prevalent in the tropics as evident in Figure 3). Additional to this, as mentioned in Section 2.2, category 1 is represented with the probabilistic term 'chance' and category 2 is represented with the spatial term 'isolated', and the latter might be chosen by forecasters as more appropriate to a situation where thunderstorms are expected in the wider area, even though the chance of a person experiencing one is low. In southern Australia, whether thunderstorms will occur at all over an area can be much less certain, which leads to a preference for the less-prominent category 1. Such observations prompt the question of whether the thunderstorm service should be defined differently in different parts of the country due to different local climatology. The 'chance' versus 'isolated' discussion raises the more subtle question of whether public weather services should be based simply on the risk of phenomena in the vicinity of a person, or whether a likelihood of phenomena on broader length scales should be taken into account.

In addition to climatological differences, we also see unique treatment of categories by specific forecasting offices. For instance, New South Wales demonstrated consistently high observed frequencies relative to other offices. This is particularly apparent for category 2 forecasts. Combined with the fact that New South Wales issued relatively few category 2 forecasts, this suggests that they were reserving this category for events with much higher confidence than required by other forecast offices. It is unsatisfactory if the meaning of forecasts change just because a person is on one side or the other of a state border.

Differences in results across lead days also raise interesting questions regarding whether the thunderstorm service should be defined differently for different forecast horizons. From Figure 4, we note that in many instances observed frequencies for a given category took on distinctly different values for lead day 7 as compared to lead day 1. This is particularly seen in category 0 observed frequencies for forecasts over the

tropics. This is presumably in large part due to the fact that, as forecast skill decreases with increasing lead time, observed frequencies of any given category should converge towards the unconditional observed frequency. The ability of the forecast system to distinguish between events, based on lead time, should also be considered in deciding how the service is defined at this forecast horizon.

#### 4. CONCLUSIONS

In the absence of a strict quantitative definition of the Bureau's categorical thunderstorm forecast service, differing 'implicit' definitions and interpretations of this service have emerged based on the forecast office issuing the forecast. At the national level, this results in spatially inconsistent forecast products, whereby text forecast terms such as 'chance of a thunderstorm' and 'possible thunderstorms' take on different meanings in terms of thunderstorm likelihood from north to south and between different forecast offices. Although there is more work to be done (e.g. examining more seasons of data), by examining the nature of historical forecasts this study offers a starting point from which to establish a nationally consistent definition of the thunderstorm forecast service.

Clear service definition will also allow the forecast production process to be better directed. Rather than linking the categories to precipitation probabilities with convoluted adjustments, a simpler process in the Graphical Forecast Editor could revolve around explicit forecasts of thunderstorm probability, guided by inputs such as Calibrated Thunder which give direct probabilities of lightning within a 10 km radius, adjusted as necessary by the forecaster then converted directly to categories using probability thresholds.

Establishing a service definition also introduces the opportunity to measure forecast skill and accuracy. This in turn enables continuous monitoring of forecast quality, and systematic improvement of forecast processes by objectively comparing alternative systems and measuring where human input does or does not add value compared to fully automated approaches.

A clearly defined service means that the BoM can tell the public exactly what is meant by the forecast and how good it is. The insights in this paper provide a step towards improved decision-making with regard to thunderstorm risk for the community at large.

#### ACKNOWLEDGEMENTS

The authors would like to thank Deryn Griffiths, Alistair McKelvie, and Nicholas Loveday who have contributed to aspects of the technical and scientific work referred to in the paper.

#### REFERENCES

- Allen, J.T., Karoly, D.J., Mills, G.A., 2011. A severe thunderstorm climatology for Australia and associated thunderstorm environments. *Aust. Meteorol. Oceanogr. J.* 61, 143–158.
- Bright, D.R., Wandishin, M.S., Jewell, R.E., Weiss, S.J., 2005. A physically based parameter for lightning prediction and its calibration in ensemble forecasts, in: 85th AMS Annual Meeting, American Meteorological Society - Combined Preprints. pp. 5699–5709.
- Bureau of Meteorology, 2020. Australian Digital Forecast Database (ADFD) User Guide. Available at <http://www.bom.gov.au/catalogue/adfdUserGuide.pdf>
- Coates, L., Blong, R., Siciliano, F., 1993. Lightning fatalities in Australia, 1824–1991. *Nat. Hazards* 1993 83 8, 217–233. <https://doi.org/10.1007/BF00690909>
- Colquhoun, J.R., 1987. A Decision Tree Method of Forecasting Thunderstorms, Severe Thunderstorms and Tornadoes. *Weather Forecast.* 2, 337–345. [https://doi.org/10.1175/1520-0434\(1987\)002<0337:adtmof>2.0.co;2](https://doi.org/10.1175/1520-0434(1987)002<0337:adtmof>2.0.co;2)
- Doswell, C. ed., 2001. Severe convective storms. Springer.
- Efron, B., Tibshirani, R.J., 1994. An introduction to the bootstrap. CRC press.
- Engel, C., Ebert, E., 2012. Gridded Operational Consensus Forecasts of 2-m Temperature over Australia. *Weather Forecast.* 27, 301–322. <https://doi.org/10.1175/WAF-D-11-00069.1>
- Hart, T., 2019. The Road Taken: An account, with lessons learned, of a 15 year journey in translating advances in science and technology into streamlining the weather forecast process and enhancing Bureau services to the Australian community. Bureau of Meteorology, Melbourne, Australia. Available at [http://www.bom.gov.au/research/publications/otherreports/FSEP\\_to\\_NexGenFWS.pdf](http://www.bom.gov.au/research/publications/otherreports/FSEP_to_NexGenFWS.pdf)
- Just, A., Foley, M., 2020. Streamlining the graphical forecast process. *J. South. Hemisph. Earth Syst. Sci.* 70, 108–113. <https://doi.org/10.1071/ES19047>
- Rudlosky, S., 2015. Evaluating ENTLN performance relative to TRMM/LIS. *J. Operational Meteor.* 3, 11-20. <http://dx.doi.org/10.15191/nwajom.2015.0302>