

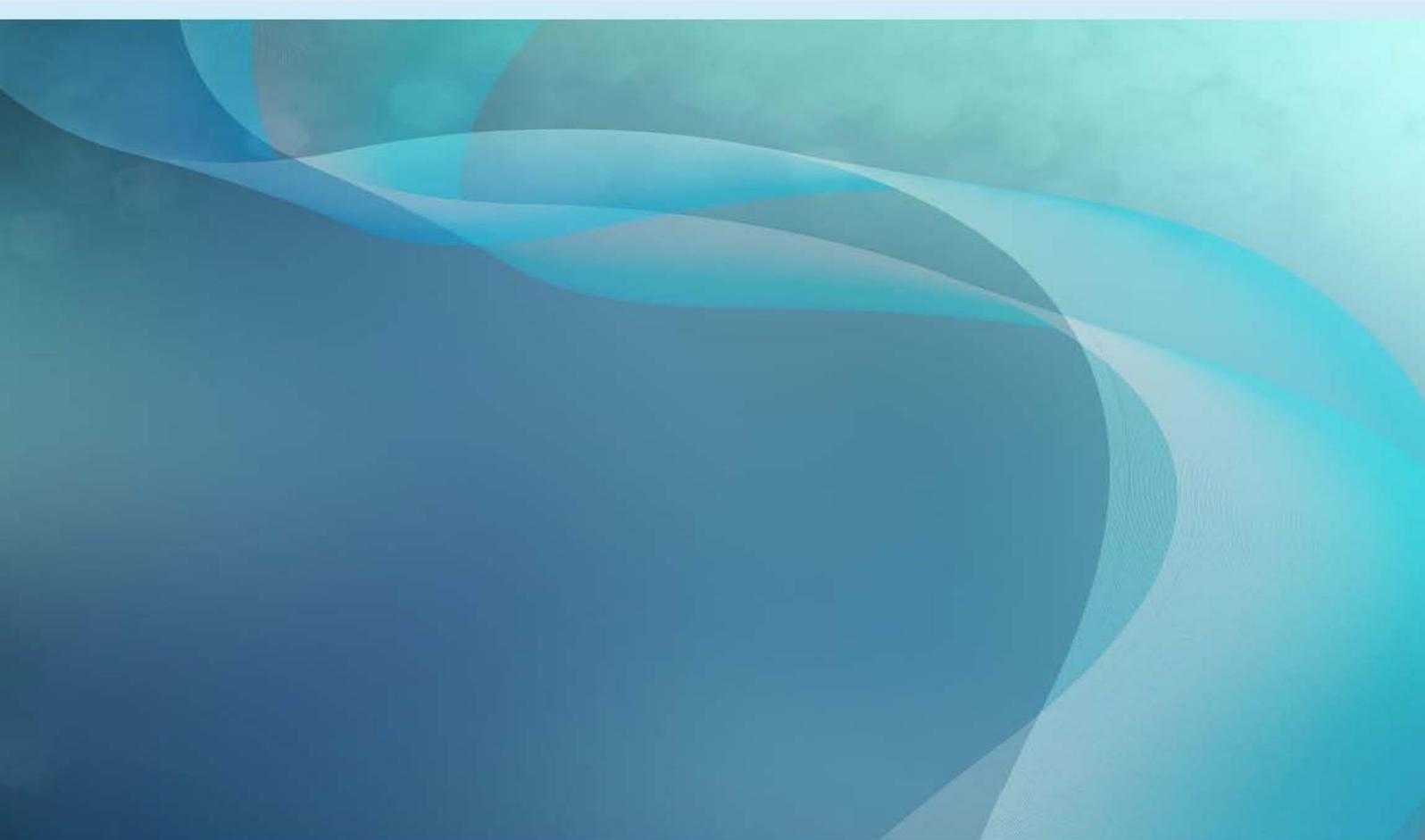


Australian Government
Bureau of Meteorology

A hybrid parametrisation for precipitation probability of exceedance data

Catherine de Burgh-Day, Francis Dillon

July 2021



A hybrid parametrisation for precipitation probability of exceedance data

Catherine de Burgh-Day, Francis Dillon

Bureau Research Report No. 052

July 2021

National Library of Australia Cataloguing-in-Publication entry

Authors: Catherine de Burgh-Day, Francis Dillon

Title: A hybrid parametrisation for precipitation probability of exceedance data

ISBN: 978-1-925738-30-8

ISSN: 2206-3366

Series: Bureau Research Report – BRR052

Enquiries should be addressed to:

Lead Author: Catherine de Burgh-Day

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

Catherine.deBurgh-Day@bom.gov.au

Copyright and Disclaimer

© 2021 Bureau of Meteorology. To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of the Bureau of Meteorology.

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

Contents

Executive Summary	1
1. Introduction	1
2. Methods	3
2.1 Data	3
2.1.1 AWAP Climatological data	3
2.1.2 ACCESS-S1 Forecast data.....	4
2.2 Computing a PoE from discrete data	4
2.3 Motivation for the choice of fitting function	5
2.4 Application of the fitting function	6
3. Results	7
4. Discussion and Conclusion	11
Acknowledgements	12
References	13

List of Tables

Table 1: Timescales, time averaging details and lead times that curves were fit to for each start date	3
Table 2: Fit Mean Absolute Error (MAE) statistics for the fits to the 556 stations for the 5 th of February 2021. The units of the MAE statistics are the same as the y-axis of Figure 1 and Figure 2 (i.e., percent probability).	10
Table 3: Fit Mean Absolute Error (MAE) statistics for the fits to the 23 regions for the 5 th of February 2021. The units of the MAE statistics are the same as the y-axis of Figure 1 and Figure 2 (i.e., percent probability).	10

List of Figures

Figure 1: An example of raw (dashed lines) and fitted (solid lines) forecast (black lines) and climatological (grey lines) PoEs for the Kimberley Downs station from the 1 st of April 2021 for the week beginning on the 8 th of April 2021.	8
Figure 2: An example of raw (dashed lines) and fitted (solid lines) forecast (black lines) and climatological (grey lines) PoEs for the Shipton's Flat station from the 1 st of April 2021 for the month of May 2021.	9

EXECUTIVE SUMMARY

Precipitation probability of exceedance (PoE) curves enable users to understand the distribution of possible rainfall amounts in their location in a coming period.

Forecast PoE curves can be constructed from ensemble data by ranking the ensemble members, and climatological PoE curves can be constructed by ranking the years in the climatological period. However, when the ensemble size or climatological period is small, the resulting PoE often exhibits spurious features resulting from sampling error, model uncertainty, measurement error, or model error. There is a risk that these features will be overinterpreted by users, and so it is common to fit smooth curves to the PoEs prior to visualisation to avoid this outcome. However, for locations or accumulation periods in which there are instances of zero precipitation, fitting curves to precipitation data becomes challenging due to zero-inflation.

Here, the Bureau of Meteorology's multi-week and seasonal prediction system, ACCESS-S1 (Hudson et al. 2017; Bureau of Meteorology, 2018) has been used to construct precipitation PoEs by ranking the 99-member ensembles it produces each day (Bureau of Meteorology, 2019; de Burgh-Day et al. 2020). Climatological PoEs were also created using AWAP data from 1980-2017.

It is then shown that a hybrid complementary gamma cumulative distribution function with a zero-inflation term can be fit to these PoEs even in the presence of zero inflation. This approach is shown to be highly effective for a range of stations (556) and key agricultural regions (23) around Australia, from timescales of weeks to seasons.

1. INTRODUCTION

Probability of Exceedance (PoE) curves such as those shown in Figure 1 and Figure 2 are used in weather, climate and hydrology forecasts (e.g. Seasonal PoE maps and curves produced by the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center¹). They are a useful tool to aid in forecast interpretation and decision making because they allow the user to choose the threshold of interest in a given parameter and read off the associated probability of exceeding that value. Thus a single probability of exceedance curve can be used by a range of users from different industries and sectors, with different needs and thresholds of interest. For example, a seasonal precipitation PoE curve could help horticulturalists plan for irrigation water purchases for the coming season and could also inform water management authorities on expected water storage capacities. A weekly or fortnightly precipitation PoE curve could provide planning authorities with warning of any upcoming flood conditions, and could also give advance warning of extended wet events, which can have both positive and negative impacts on graziers.

The Rural R&D For Profit: Forewarned is Forearmed² (FWFA) project includes the development of a suite of trial forecast products on a range of timescales for over 500 locations

¹https://www.cpc.ncep.noaa.gov/products/predictions/long_range/poe_index.php?lead=1&var=p

²<https://www.agriculture.gov.au/ag-farm-food/innovation/rural-research-development-for-profit/approved-projects-round3>

throughout Australia. Already a large number of experimental research products have been developed and are being trialled with users through FWFA and other projects (Bureau of Meteorology, 2019; de Burgh-Day et al. 2020). One of the newer products developed is a precipitation PoE plot. However, these PoEs are constructed from a limited sample of data, and exhibit features that are likely to be due to the small sample size, uncertainty in the model, measurement error, or model error. There is a risk that users will overinterpret these features, so it is desirable to fit a smooth curve to the data to avoid this situation.

A great deal of attention has been given to daily precipitation probability distribution characterisation in the context of stochastic precipitation models, frequency analysis of precipitation and precipitation trends related to global climate change (Ye et al. 2018 and references therein). The majority of this research has focussed on distributions for wet-day data (i.e., precipitation series where only wet days are considered). This is suitable for analyses using “two-part” stochastic daily precipitation models, where a Probability Distribution Function (PDF; often a gamma distribution) is used to describe precipitation amounts on wet days, and a probabilistic representation of precipitation occurrences is applied separately (e.g. Geng et al., 1986; Watterson, 2005). A full wet *and* dry days probability distribution would be zero-inflated (i.e., it would feature an accumulation of counts at zero alongside the wet days precipitation distribution) and would be discontinuous.

In contrast, less attention in the literature has been given to the best approach for characterising the distribution of precipitation in an operational forecasting context for precipitation accumulations over greater than daily accumulation periods³. In this setting, computational overheads are a limiting factor and an approach which is robust across a wide variety of precipitation distributions is necessary. Thus, in this case a single parameterisation (as opposed to the abovementioned two-part approach) is desirable to reduce computational cost and complexity.

One approach to presenting a smooth precipitation PoE curve for a given accumulation period (i.e. over a month or season) in a forecasting context is to assume all the samples in the distribution are “wet samples” (i.e. wet months or wet seasons), and simply fit a PDF to the precipitation distribution before computing the associated PoE curve. An alternative method is to again assume all samples are wet samples, transform the data to an approximated normal prior to fitting, then transform back to the original distribution before computing the PoE curve (as is done at the NOAA Climate Prediction Centre).

These approaches work well in cases where the precipitation distribution covers a longer time period (i.e. months up to seasons and beyond), or is for a region with larger precipitation amounts, since in these cases the assumption of wet samples is often a fair one. However, when considering shorter timescales (i.e. weeks up to months), or regions which are often dry (e.g. much of inland Australia), these methods become less effective because of the high likelihood of instances of no precipitation within the period. This introduces zero-inflation, making it impossible to perform a robust fit with a single smooth function since the distribution is discontinuous.

³ The only example the authors could find was from the NOAA Climate Prediction Centre: <https://www.cpc.ncep.noaa.gov/pacdir/NFORdir/INTR.html>.

Here we present a precipitation PoE fitting methodology which overcomes these challenges and is able to robustly fit PoE curves for a large number of locations around Australia on timescales from weeks to seasons. This methodology is intended for an applied real-time setting where computational cost and complexity are considerations.

In Section 2 we describe the data used, motivate our choice of fitting function, and present an overview of our method. In Section 3 we present the results of an assessment of our method. In Section 4 we discuss these results and provide some concluding remarks.

2. METHODS

2.1 Data

When presenting forecast PoEs, it is desirable to have a reference PoE curve to compare to. Shifts in the forecast PoE curve relative to the observed climatological PoE curve provide information about whether there is any shift to larger or smaller precipitation totals compared to the historical odds. Both observed climatological and model forecast PoE curves were considered in this analysis. Since both the forecast and climatological PoEs are representing the same quantity (i.e., precipitation), one might assume that if the PoE curve fitting methodology described here is effective for one type of curve it will be effective for the other. However, the number of samples contributing to the forecast and climatological PoE curves are different, and are sourced from different types of systems (a numerical prediction system and an observation analysis respectively). It is not possible therefore to say with any confidence that the PoE fitting routine described here will be effective for one of these types of PoE curve because it performs well for the other, and so both are investigated.

2.1.1 AWAP Climatological data

The climatological precipitation PoE data was based on Australian Water Availability Project (AWAP) region and station data from 1980 to 2017 (i.e., 38 years), derived from a 5km grid (Jones et al., 2009). The stations were obtained from the 5km grids by extracting the grid cell closest to the latitude, longitude point of each station. Regions were obtained by taking the area average of the grid cells inside each region boundary. Weekly, fortnightly, 4-weekly, monthly and seasonal totals were obtained by accumulating along the time axis of the data as detailed in Table 1.

Table 1: Timescales, time averaging details and lead times that curves were fit to for each start date

Timescale	Time averaging method	Forecast lead times
Weekly	Daily data accumulated to weeks in serial ⁴ . Starting from the forecast start date for ACCESS-S1 data.	0, 1, 2, 3
Fortnightly	Daily data accumulated to weeks in serial, then accumulated to fortnights as a rolling accumulation.	0, 1, 2, 3

⁴ I.e., week 0 is days 0 to 6, week 1 is days 7 to 13, etc.

	Starting from the forecast start date for ACCESS-S1 data.	
4-weekly	Daily data accumulated to weeks in serial, then accumulated to 4-week means as a rolling accumulation. Starting from the forecast start date for ACCESS-S1 data.	0, 1
Monthly	Daily data accumulated to calendar months in serial. Starting from the first full month for ACCESS-S1 data.	0, 1, 2, 3
Seasonal	Daily data accumulated to calendar months in serial, then accumulated to 3-month means as a rolling accumulation. Starting from the first full month for ACCESS-S1 data.	0, 1, 2

2.1.2 ACCESS-S1 Forecast data

The forecast precipitation PoE data was drawn from the ACCESS-S1 real time post processing pipeline (Bureau of Meteorology, 2019; de Burgh-Day et al. 2020). This pipeline uses operational data produced by the Bureau’s seasonal prediction system ACCESS-S1 (Hudson et al., 2017; Bureau of Meteorology, 2018). The postprocessing pipeline performs a number of operations on the raw model outputs. Those of interest to this work are:

1. Re-gridding the spatial precipitation data from a 60km to a 5km grid
2. Bias-correcting and calibrating the re-gridded precipitation forecast data
3. Constructing time lagged 99-member weekly and seasonal ensembles from the 33 members produced each day
4. Computing precipitation totals across a range to timespans from weekly to seasonal
5. Extracting stations (i.e. individual grid cells) and computing regional averages from the spatial, calibrated data for each timespan

For a detailed description of the postprocessing steps applied to ACCESS-S1 see Bureau of Meteorology (2019) and de Burgh-Day et al. (2020). For more information on ACCESS-S1 see Hudson et al. (2017) and Bureau of Meteorology (2018).

2.2 Computing a PoE from discrete data

Before a smooth curve was fitted to it, the forecast and climatological PoEs were obtained for each station and region. The method to do this is different for the forecast ensemble and the AWAP climatology:

- The AWAP climatological PoE for a given date (i.e., a given week of the year) was obtained by ranking each *year* of data for that date and scaling the ranking count between 0 and 100.

- The forecast PoE for a given forecast start time, lead time combination was obtained by ranking each *ensemble member* in the time-lagged 99-member ensemble (Bureau of Meteorology, 2019) and scaling the ranking count between 0 and 100.

In both cases the inverse of this scaled ranking is the PoE (i.e. $100 - \text{ranking} = 100 * \text{PoE}$). For precipitation all ranked values will have ≥ 0 mm of precipitation for a given date (i.e., the probability of exceeding any value less than 0 is 100%), and that there will be some upper precipitation amount which no value in the ranking will exceed (the probability of exceeding this value is 0%). The ranking will provide information about the probabilities between these two extremes, based on precipitation amounts from each year in the climatological period for the climatological PoE and each ensemble member in the 99-member ensemble for the forecast PoE.

2.3 Motivation for the choice of fitting function

As described in the previous section, when constructing a PoE from discrete data the samples are ranked, scaled, and the inverse is then computed. The equivalent PoE curve would be defined as the graph of $1 - F$, where F is the Cumulative Distribution Function (CDF). Thus one strategy to obtain a smooth PoE curve can be to fit a gamma distribution to the precipitation PDF, compute the F from the PDF, and then compute the PoE curve from this. Another approach is to convert the precipitation PDF to a normal distribution using a power transformation, fit a Gaussian function to this, and then obtain the PoE from this in a similar manner⁵. However, transforming to a normal distribution requires estimation of the skewness of the data, necessitating an extra model parameter and increasing the risk of overfitting.

Locations with low rainfall amounts can be highly skewed or zero-inflated (i.e. an accumulation of counts at $x = 0$, resulting in a discontinuous distribution) because even over longer timespans there may be little or no precipitation, and similarly when considering shorter timespans such as weeks or fortnights the chance that there is little or no precipitation in the period of interest increases.

Fitting using a two-parameter gamma or a normal distribution is not effective for zero-inflated data because the distribution being modelled is not smooth. This can be overcome with a hybrid gamma distribution which includes a zero-inflation term, resulting in a three-parameter fitting model. This is appropriate when the aim is to fit the precipitation PDF, however for the purposes of this work we are interested in precipitation PoE curves specifically, and thus for simplicity we can fit a zero-inflated model directly to PoE data.

Therefore, the method investigated here is to compute the PoE directly from the discrete data, and to fit to that. As mentioned above, the PoE and CDF of any precipitation distribution are complementary, so that the PoE can be obtained by calculating $1 - F$ (hereafter referred to as the *complementary gamma CDF*), where F denotes the CDF of the distribution. Assuming that the precipitation PDF is well represented by a gamma distribution, the CDF will be defined as

⁵ As is done by the NOAA Climate Prediction Center (<https://www.cpc.ncep.noaa.gov/pacdir/NFORdir/INTR.html>)

$$F_{gamma}(x; \alpha) = \frac{\gamma(\alpha, x)}{\Gamma(\alpha)} \quad (1)$$

where

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (2)$$

is the upper complete gamma function and

$$\gamma(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt \quad (3)$$

is the lower incomplete gamma function.

One of the things that makes fitting a gamma function to a precipitation PDF challenging is that the mode (i.e., the peak of the distribution) is not fixed at $x = 0$, making the addition of a zero-inflation term more complex since the value of the PDF at $x = 0$ must be fit for. Since the complementary gamma CDF is always maximum at $x = 0$ (where the probability of exceedance is 100%) and is monotonically decreasing for $x \geq 0$, zero-inflation can be more easily incorporated by allowing the fitting function to translate in x , and fixing all values of $x \leq 0$ to 1. In this formulation, the free parameter introduced by zero-inflation is the translation term. The hybrid fitting function used in this work thus includes three free parameters; a skewness term α , a scaling parameter, β , and a translation term, δ :

$$F_{hybrid}(x; \alpha, \beta, \delta) = \begin{cases} 1, & x \leq 0 \\ \left(1 - F_{gamma}\left(\frac{x - \delta}{\beta}; \alpha\right)\right), & x > 0 \end{cases} \quad (4)$$

where

$$\begin{aligned} -\infty &\leq \alpha \leq \infty \\ 0 &< \beta \leq \infty \\ -\infty &\leq \delta \leq \infty \end{aligned} \quad (5)$$

This hybrid complementary Gamma CDF, F_{hybrid} , is able to robustly fit a large variety of PoE curves from timescales of weeks to seasons, across a wide range of climatological zones in Australia.

2.4 Application of the fitting function

Hybrid complementary gamma CDF curves were fit to ranked PoE data from 556 stations and 23 regions around Australia. Here station data are drawn from a single grid cell for each station, while region data are area averages of the grid cells within a defined domain. Fits were

performed for both climatological data and forecast data for each station and region, for the timescales shown in Table 1. This amounts to a total of $(556 + 23) \times (4 + 4 + 2 + 4 + 3) \times 2 = 19,686$ individual fits of F_{hybrid} to raw ranked PoE data for a single forecast start date.

In order to simplify the implementation of the fitting routine, the raw data was linearly-interpolated to 300 points before fitting. This was performed using the Python routine `scipy.interpolate.interp1d`⁶ where the lower and upper bounds of the interpolation were fixed at 100 and 0 respectively. Without this step, the default configuration of the Python fitting routine used frequently rejected the fit due to the sparsity of data points. Since the number of free parameters in the model is fixed and we have made the assumption that the data follow this model, this interpolation does not increase the risk of overfitting. However, since the model being fit is not linear between the samples, this interpolation may degrade the quality of the fit. This is not visually apparent on the plots this methodology was developed for (e.g. Figure 1 and Figure 2), but it underscores that this approach should not be employed for applications where the fitted parameterisation is used for any further analysis.

The fits to the interpolated data were then performed using the Python routine `scipy.optimize.curve_fit`⁷ with a maximum number of function evaluations of 10,000 and a least squares optimisation approach where the fitting tolerance for termination was specified as where the rate of change of the independent fitting variables was $<10^{-8}$. The initial guess for each fit was $(\alpha, \beta, \delta) = (1, 1, 0)$, and the gamma CDF was estimated using the Python function `scipy.stats.gamma.cdf`⁸.

The Mean Absolute Error (MAE) between the interpolated data and the fitted curve was calculated for each fit as a second measure of the robustness of the fitting method, and the mean and median of the MAE distribution for each timespan, lead time combination were calculated for the stations and the regions.

When implemented in a real-time setting the MAE between the interpolated data and fitted curve is used as a criterion for acceptance or rejection of the fit (i.e., if the MAE of a given fit exceeds a specified value, the fit is rejected even if it was successful according to the least squares measure used). By fitting with a least squares minimisation method and assessing the robustness of the fit using MAE, we have imposed a tighter constraint on the goodness of fit than the least squares minimisation would alone. The MAE statistics presented here include values for all fits, whether or not each fit was accepted or rejected.

3. RESULTS

Fits were performed for all stations and regions for the lead times specified in Table 1 for forecasts starting on the 1st, 5th and 10th of February, and the 1st of April 2021. Examples of raw

⁶ <https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.interp1d.html>

⁷ https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html

⁸ <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gamma.html>

ranked PoEs and successfully fitted curves for the climatological data and forecast data are shown in Figure 1 and Figure 2 for forecasts starting on the 1st of April.

Figure 1 shows raw and fitted PoEs for the Kimberley Downs station (in Western Australia) from the 1st of April 2021 for the week beginning on the 8th of April 2021. The fitted climatology and forecast curves demonstrate the effect of the zero-inflation term (meeting the y-axis at ~55%). The Forecast and Climatology statistics in the figure are

Precipitation Probability of Exceedance Start: 01-Apr-2021
 Region: Kimberley Downs Period: Week 08-Apr-2021 to 14-Apr-2021

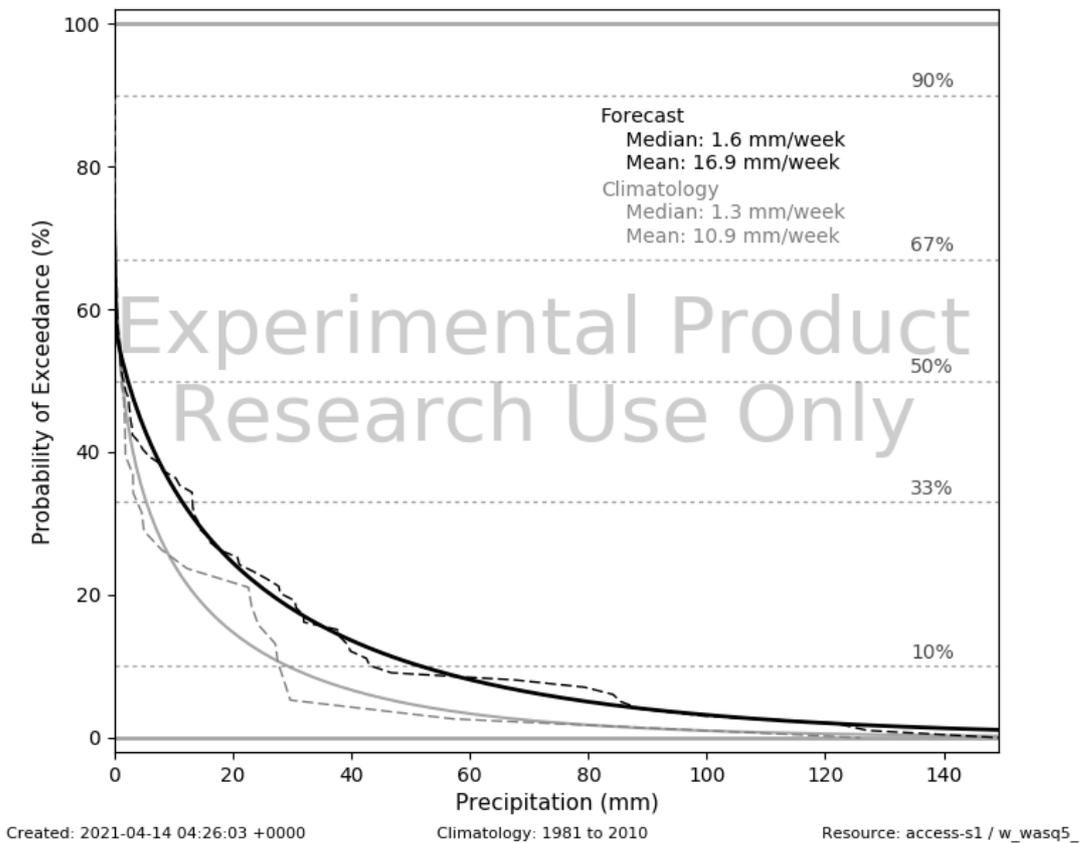


Figure 1: An example of raw (dashed lines) and fitted (solid lines) forecast (black lines) and climatological (grey lines) PoEs for the Kimberley Downs station from the 1st of April 2021 for the week beginning on the 8th of April 2021.

calculated directly from the raw data. This figure also provides a typical example of the lesser degree of smoothness in the raw climatological data due to the smaller sample size available (38 years vs the 99 ensemble members for the forecast data). The prediction for this fortnight is that the odds are shifted to wetter than normal conditions compared to the historical reference period.

Figure 2 shows raw and fitted PoEs for the Shipton’s Flat station (in northern Queensland) from the 1st of April 2021 for the month of May 2021. This is an example of a case where both

the forecast and climatological raw PoE data was relatively smooth and was fit well by F_{hybrid} . This figure is also a good example of an interesting forecast; ACCESS-S1 has predicted that the month of May will be very dry (the climatological mean exceeds the forecast mean by more than a factor of 3) compared to the historical reference at a lead time of 1 month.

Precipitation Probability of Exceedance Start: 01-Apr-2021
 Region: Shiptons Flat Period: Month 01-May-2021 to 31-May-2021

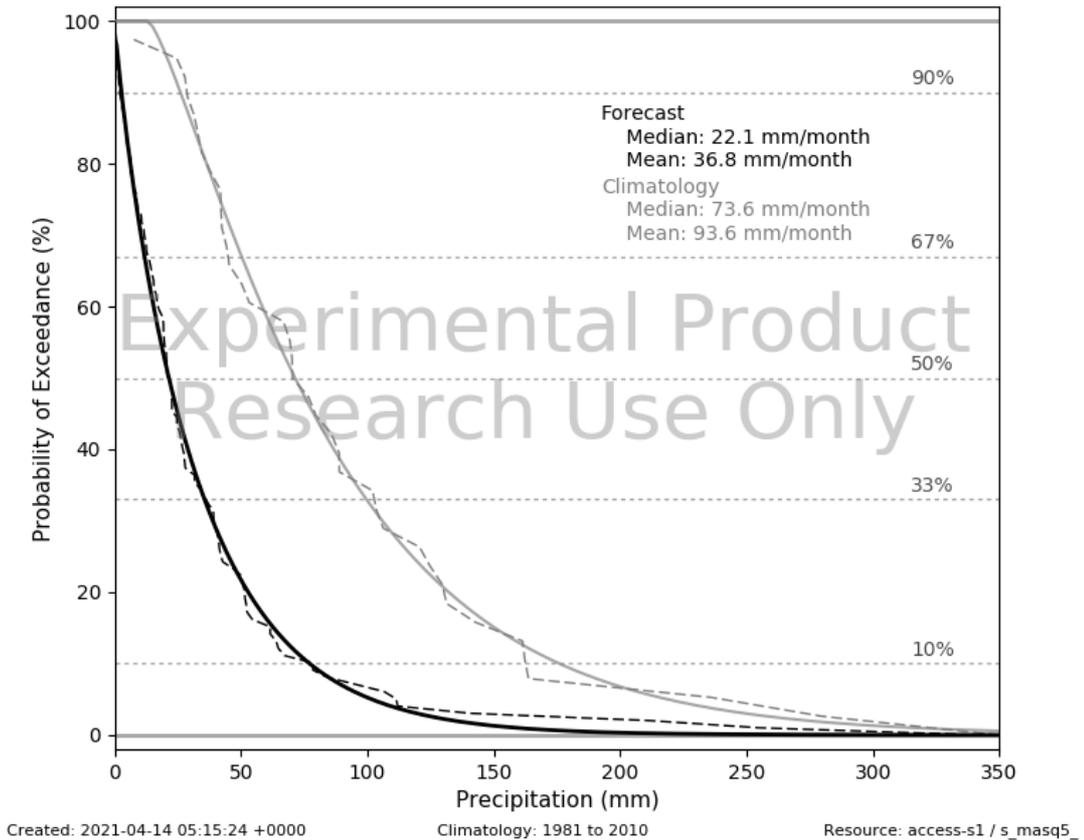


Figure 2: An example of raw (dashed lines) and fitted (solid lines) forecast (black lines) and climatological (grey lines) PoEs for the Shipton’s Flat station from the 1st of April 2021 for the month of May 2021.

The MAE statistics were similar for all forecast start dates. For brevity, only the statistics for the 5th of February 2021 are shown here. The rate at which forecast (climatology) fits failed to converge was below 0.37% (1.51%) across the 556 stations for all timespans and lead times. The rate at which both forecast and climatology fits failed to converge for the 23 regions was 0% for all lead times for weekly, fortnightly, four-weekly and monthly timespans, and was 0% and 2.9% for forecast and climatology fits respectively for the seasonal timespan.

Table 2 shows the mean and median fit MAE by timespan and lead time across the 556 stations, and Table 3 shows the mean and median fit MAE by timespan and lead time across the 23 regions. For both region and station fits the mean and median values are comparable, indicating that there is little skew in the distribution of fit errors. The climatology fits have

generally higher MAEs than the forecast fits, which can be attributed to the fact that the underlying raw ranked PoEs are constructed from 38 and 99 samples for the climatology and forecasts respectively. This results in the raw climatology PoE curves being less smooth and harder to fit a curve to. The difference in smoothness between the forecast and climatology raw PoEs is apparent in Figure 1.

Table 2: Fit Mean Absolute Error (MAE) statistics for the fits to the 556 stations for the 5th of February 2021. The units of the MAE statistics are the same as the y-axis of Figure 1 and Figure 2 (i.e., percent probability).

Lead times/timesteps into the period		0	1	2	3	0	1	2	3
Timescale	Data Type	Median MAE (%)				Mean MAE (%)			
Weekly	Forecast	1.03	0.83	0.78	0.77	1.16	0.88	0.83	0.82
	Climatology	1.07	1.28	1.18	1.11	1.18	1.23	1.25	1.22
Fortnightly	Forecast	0.89	0.9	0.8	0.84	1.01	0.93	0.86	0.88
	Climatology	1.18	1.24	1.24	1.3	1.3	1.33	1.31	1.39
4-weekly	Forecast	0.83	0.87	-	-	0.88	0.92	-	-
	Climatology	1.39	1.38	-	-	1.45	1.46	-	-
Monthly	Forecast	0.95	0.86	0.83	0.74	0.99	0.9	0.86	0.78
	Climatology	1.41	1.3	1.31	1.27	1.45	1.4	1.39	1.34
Seasonal	Forecast	0.91	0.79	0.76	-	0.96	0.82	0.79	-
	Climatology	1.38	1.38	1.31	-	1.47	1.46	1.37	-

Table 3: Fit Mean Absolute Error (MAE) statistics for the fits to the 23 regions for the 5th of February 2021. The units of the MAE statistics are the same as the y-axis of Figure 1 and Figure 2 (i.e., percent probability).

Lead times/timesteps into the period		0	1	2	3	0	1	2	3
Timescale	Data Type	Median MAE (%)				Mean MAE (%)			
Weekly	Forecast	0.88	0.83	0.9	0.77	1.08	0.99	1.06	0.86
	Climatology	1.13	1.24	1.71	1.3	1.23	1.31	1.26	1.32
Fortnightly	Forecast	0.81	0.99	0.99	0.86	0.87	1.08	0.98	1.01
	Climatology	1.1	1.22	1.63	1.7	1.26	1.32	1.51	1.64
4-weekly	Forecast	0.91	0.89	-	-	0.87	0.96	-	-
	Climatology	1.34	1.26	-	-	1.43	1.38	-	-
Monthly	Forecast	1	0.92	0.69	0.8	1.03	0.95	0.75	0.83
	Climatology	1.34	1.34	1.42	1.31	1.49	1.54	1.48	1.39
Seasonal	Forecast	0.82	0.8	0.75	-	0.81	0.78	0.71	-
	Climatology	1.49	1.2	1.36	-	1.42	1.35	1.31	-

4. DISCUSSION AND CONCLUSION

While fitted curves of forecast and climatological PoEs are provided by other centres, these focus on seasonal timescales and larger regions so that there are fewer zero precipitation values, eliminating the challenge of accounting for zero-inflation. In contrast, the aim of this work is to be able to fit smooth curves to precipitation PoEs in the presence of zero-inflation, while minimizing the number of fitting parameters and therefore minimizing the risk of overfitting. Here we have shown that a hybrid complementary gamma CDF with a zero-inflation term is a good candidate for directly fitting to a wide variety of raw PoE curves, with forecast and climatology data on all timescales assessed having a fit failure rate below 1.51% for 556 stations and 2.91% for 23 regions. The average fit MAE across all stations (regions) for all timescales and lead times was 0.90% (0.91%) for forecast data and 1.37% (1.39%) for climatology data.

There are some important considerations that must be kept in mind when applying this fitting methodology. Firstly, unusually shaped PoE curves, or PoE curves which are not sufficiently smooth, will not be fit successfully. In this work we have made the underlying assumption that the wet precipitation data follows a gamma distribution, and that a hybrid model can be used to account for the likelihood of dry conditions. Any features in the data which deviate from these assumptions will not be fit well.

The simple approach employed here to construct empirical PoEs (i.e., ranking the data and rescaling between 0% and 100%) can result in a poor representation of the underlying distribution where the sample size is small, because clustering of the data is not accounted for. This is apparent in the results here, where the raw climatological PoEs are constructed from a smaller sample of data (38 years) compared to the forecast PoEs (99 ensemble members), and thus have larger MAEs on average, and a higher rate of fit failures. Conversely, a longer climatological period or larger forecast ensemble would increase the robustness of this methodology and with a sufficiently large sample a fitted curve could even become unnecessary, since the risk of overinterpretation of the data due to sampling issues would be largely eliminated.

Alternative methods exist for obtaining a parameterisation for the PoE curve which don't require ranking the data and would thus be less sensitive to sample size. One example of this is a Maximum Likelihood approach, in which the data is related directly to the parameterisation. However, implementation of these alternative methods would be more complex than what is described in this Research Report, and is left to future work.

Secondly, this method is intended only to provide a more user-friendly smoothed curve and avoid overinterpretation of the underlying raw PoE. Any model parameters obtained from this method should not be used for anything other than visualisation purposes (e.g. they should not be used for verification purposes, to compute derived statistics, etc.). A more detailed analysis of the statistics of the fitted distributions would be required to establish whether they exhibit the correct statistical features for precipitation distributions.

Here we have described an approach to fitting to precipitation PoEs in which the PoE is computed from the raw data, and a curve is fitted directly to that. We have proposed a hybrid

complementary gamma CDF with a zero-inflation term as a suitable curve for fitting directly to precipitation PoEs. We have demonstrated that this parameterisation results in robust fits in the majority of cases, on a variety of accumulation timescales from weeks to seasons, and in a variety of climatological zones across Australia. Future work to build on this approach would include investigating more robust fitting approaches such as Maximum Likelihood parameter estimation, or developing a hybrid gamma precipitation distribution model with a zero-inflation term, so that the precipitation distribution could be fit to directly. While this is not the most efficient approach for the purposes of this work, it would result in a more versatile result which could be used to produce a range of derived products, including PoE curves.

ACKNOWLEDGEMENTS

We would like to thank all the team members who have been involved with the design and development of the ACCESS-S1 post-processing pipeline which enables the generation of the PoE products discussed here. Special thanks to Robert Taggart, Belinda Trotta and Martin Schweitzer for reviews of early versions of this manuscript. This work is supported by the Forewarned is Forearmed project, which is funded by the Australian Government Department of Agriculture, Water and the Environment as part of its Rural Research and Development for Profit programme.

REFERENCES

- Bureau of Meteorology, 2018. Operational Implementation of ACCESS-S1 Seasonal Prediction System. Bureau National Operations Centre Operations Bulletin Number 120.
- Bureau of Meteorology, 2019. Operational Implementation of ACCESS-S1 Forecast Post-Processing. Bureau National Operations Centre Operations Bulletin Number 124.
- de Burgh-Day, C., Griffiths, M., Yan, H., Young, G., & Hudson, D., 2020. An adaptable framework for development and real time production of experimental sub-seasonal to seasonal forecast products. *Bureau of Meteorology Research Report Number 42*. ISBN: 978-1-925738-15-5
- Geng, S., de Vries, F.W.P. and Supit, I., 1986. A simple method for generating daily rainfall data. *Agricultural and Forest meteorology*, 36(4), pp.363-376.
- Hudson, D., Alves, O., Hendon, H. H., Lim, E., Liu, G., Luo, J-J., MacLachlan, C., Marshall, A. G., Shi, L., Wang, G., Wedd, R., Young, G., Zhao, M., Zhou, X., 2017. ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system. *Journal of Southern Hemisphere Earth Systems Science*. 67: 132-159. doi: 10.22499/3.6703.001
- Jones, D.A., Wang, W. and Fawcett, R., 2009. High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, 58(4), p.233. doi: 10.1.1.222.6311
- Watterson, I.G., 2005. Simulated changes due to global warming in the variability of precipitation, and their interpretation using a gamma-distributed stochastic model. *Advances in Water Resources*, 28(12), pp.1368-1381.
- Ye, L., Hanson, L.S., Ding, P., Wang, D. and Vogel, R.M., 2018. The probability distribution of daily precipitation at the point and catchment scales in the United States. *Hydrology and Earth System Sciences*, 22(12), pp.6519-6531.