AWAPer: An R package for area weighted catchment daily meteorological data anywhere within Australia

Peterson, Tim J.^{*1,2}, Wasko, Conrad², Saft, Margarita² and Peel, Murray²

tim.peterson@monash.edu

¹ Department of Civil Engineering, Monash University.

² Department of Infrastructure Engineering, The University of Melbourne.

Abstract

Meteorological time-series data are a fundamental input to hydrological investigations. But sourcing data is often laborious and plagued with difficulties. In an effort to improve efficiency and rigor we present an R-package, named *AWAPer* (<https://github.com/peterson-tim-j/AWAPer>), for the efficient estimation of daily area weighted catchment average and spatial variance of meteorological variables, including evapotranspiration. The package allows creation and updating of a data-cube of gridded daily data from 1900 onwards. Once created, point and area weighted estimates can be extracted at user-defined locations and time periods for anywhere within Australia. Examples of point and catchment average extraction are presented.

Keywords: Meteorological time-series data; hydrological modelling; R; Australia

1. Introduction

A primary issue facing climatological and hydrological assessments is the constant need for temperature, precipitation, and evaporation data at a location or catchment of interest. The starting point for any such analysis is to see what station data are available. Inevitably station data will not be available at the location of interest, there will be missing records, or a catchment average will be required, and the discussion will turn to which interpolation technique is best (e.g. Hwang et al, 2011; Wasko et al., 2013). Ambiguity and subjectiveness creeps into the analysis even before the data are used for the desired application.

To overcome these issues, gridded data sets are often used (e.g. Ledesma and Futter, 2017; Nathan et al., 2017; Peterson and Fulton, 2019) but often file formats make extraction of time-series data, catchment weighted averages, or otherwise, an onerous task, and one which is rarely undertaken outside the data set host's organisation. We believe, the inability of existing datasets to easily provide area weighted catchment estimates of daily meteorological data has unnecessarily

compromised the reliability of hydrological investigations (e.g. Chaubey et al, 1999; Faurès et al 1995). Furthermore, Zhao et al. (2013) showed that gridded meteorological data can contribute to understanding the significance of rainfall spatial variability on streamflow response, despite the grids being interpolated products. We believe that the difficulties in using gridded meteorological data has contributed to this issue remaining an open question (Emmanuel et al., 2017).

To increase the rigor and efficiency of hydrological investigations and to open new research directions, we present an R-package for the efficient estimation of daily point and area weighted catchment average meteorological variables, including evapotranspiration, for all the Australian Water Availability Project (AWAP) data from 1900 onwards (Jones et al., 2009). As Australia experiences very high temporal variability in precipitation and runoff for a given climate type (Peel et al., 2004), prolonged droughts (van Dijk et al., 2013), and complex multi-decade hydrologic variability (Peterson and Western, 2014; Peterson and Western; 2018; Peterson et al. 2018; Saft et al., 2015), this package will enable the international research community to explore such phenomena using reproducible data.

2. The R-Package

AWAPer is a GPL3+ open-source licensed R-package and is available at <https://github.com/peterson-tim-j/AWAPer>. The package includes documentation of each function and it is fully functional on both Windows and Linux operating systems.

2.1 Source Data

AWAPer uses AWAP gridded daily data of precipitation, minimum temperature, maximum temperature, vapour pressure deficit (calculated from dew point temperature) and solar radiation (Jones et al., 2009) from the Australian Bureau of Meteorology. AWAP has been extensively evaluated (Jones et al., 2009; Beesley et al. 2009) including for extreme precipitation (Nathan et al., 2016). The choice of using AWAP as the data source is a pragmatic choice. However, the package is sufficiently flexible that other gridded data sets, such as SILO (Jeffrey et al. 2001), can be easily used if they are in an ESRI-ASCII raster format.

The AWAP grids are $0.05^\circ \times 0.05^\circ$ (approximately 5 km $\times$ 5 km) and extend from 112°E to 156.25°E and -44.50°S to -10.00°S . Temperature and precipitation are available from 1900 (monthly) and 1911 (daily) respectively to the present day; vapour pressure from 1950 to the present day; and solar radiation from 1990 to the present day. For calculation of some evapotranspiration estimates, the Geoscience Australia 9 second ($\sim 250\text{m}$) DEM is also downloaded by the package. Eight different evapotranspiration calculations are implemented from Guo et al. (2016) including potential, reference crop, and actual evaporation.

2.2 Data Structures

AWAPer builds netCDF data cubes from the source ARCMAP ASCII text files (one file per day and variable); one for precipitation, daily minimum temperature, daily maximum temperature, vapour pressure deficit and a second data cube for solar radiation. Both data cubes are in netCDF version 4 format, have fixed latitude and longitudinal coordinates and an unconstrained time dimension to allow efficient update of the data cube as new data become available. The time datum is 1st of January 1900 for all variables (including vapour pressure deficit), except solar radiation which has a time datum of 1st of January 1990.

2.3 Usage

AWAPer contains functions for building and updating netCDF data cubes, downloading and ingestion of the Geoscience Australia 9 second DEM, extracting area weighted catchment and point daily meteorological data, and subsequent calculation of evapotranspiration. The details of each function are presented in the package manual.

The netCDF files can be built for user defined meteorological variables and time range. Once built, they can be easily updated from the last date of stored data through simply running the default function again. The file size ranges from 5 GB for one variable for the entire time series (1900-2019) when maximum compression is used, to greater than 380 GB when the files contain all available variables and time points and no compression is used. Once built, the netCDFs files should be stored locally because the runtime for data extraction is often limited by the hard drive read speed and accessing the files remotely is likely to introduce significant delays. With no

compression, extraction of 119 years of daily area weighted precipitation for the two example catchments within the package required 30.2 minutes. With maximum compression, the extraction time increased by 8% to 32.7 minutes.

For the area weighted catchment extraction, the fraction of an AWAP grid cell within a catchment boundary is derived by dividing the grid cell into 100 sub-cells (i.e. approximately 500m x 500m) and assessing if the entire sub-cell is within the catchment. This results in an approximately 1% error along the boundary of each AWAP grid cell. The catchment average is the arithmetic average of each grid cell, partially or fully inside the catchment boundary, multiplied by the fraction of the cell inside the catchment boundary. The spatial variance is calculated from all grid cells partially or fully within the boundary, without the fractional weights. For the extraction of point data, bilinear interpolation is used, based on the four nearest grid cells. For point extraction, spatial variance is not estimated.

When solar radiation is extracted, or used within an ET estimate, the solar radiation prior to 1/1/1990 is estimated from the mean calendar day solar radiation from 1/1/1990 to the final date for which data is to be extracted. Days post 1/1/1990 without solar radiation data are estimated using linearly interpolated and gaps at the end of the record are infilled using the most recent prior observation. Estimates with and without interpolations are returned to the user.

3. Example Applications

3.1 Extraction at a point

Point data are extracted at the observation station for Bathurst (149.5559, -33.4289). This station is one of the few stations included within the high-quality Australian Climate Observations Reference Network (BOM, 2012) for both daily surface temperature (Trewin, 2013) and precipitation measurements (Lavery, 1992). Prior to December 1971 the screen did not strictly meet specifications and the station was moved in 1950, 1978 and 1996. All these dates are presented in the double mass curves in Figure 2 using data from 1911-2017 inclusive. Missing station data were omitted (there is no missing data in the gridded observations).

There is no great deviation from the one-to-one double mass curve for minimum temperature till the 1970's (Figure 1a). At this point there is a slight underestimation of the minimum temperature. However, Figure 1c shows that the difference between the estimates is modest (Figure 1c). The cause of the difference is most likely due to (1) AWAP estimates not being forced to honour observations and/or (2) *AWAPer* point estimates being derived from the bilinear interpolation of the $0.05^\circ \times 0.05^\circ$ (approximately 5 km $\times$ 5 km) AWAP grid cells.

The extracted AWAP precipitation shows a systematic overestimation compared to the station observation (Figure 1b), but the majority of data are zero days and matched well. The result is a slightly negatively skewed distribution of errors but in general these errors are small (Figure 1d). These results are not intended as an evaluation of AWAP, but confirmation that using *AWAPer* provides a simple means for generating an accurate time series of daily climatic data.

3.2 Extraction of catchment weighted

The primary benefit of *AWAPer* is extraction of catchment averages and variances. Figure 2a presents extracted catchment average precipitation for Bet Bet Creek at Norwood (407220, 351.7km²), Victoria, Australia, derived from 26 AWAP grid cells (14 full grid cells, 12 partial). For independent verification the catchment average precipitation is compared to the high quality recorded streamflow (Zhang et al., 2016) and 0-1m modelled soil moisture from the BOM Australian Water Resources Assessment Landscape (AWRA-L) model product (Viney et al., 2015; Frost et al., 2015; Holgate et al., 2016); which was derived from the same source data as used by *AWAPer* and so provides an independent assessment of *AWAPer* correctly extracting and processing data.

Figure 2a shows the catchment average precipitation differs greatly from an observational daily gauge located in the centre of the catchment demonstrating the importance of catchment average variables for hydrologic applications (Faurès et al., 1995). Over the four-month period, the timing of *AWAPer* derived precipitation events are consistent with soil moisture and streamflow response confirming sensible extraction of *AWAPer* area weighted catchment precipitation. To demonstrate the additional functionality daily (Jensen-Haise) and monthly (Morton CRAE) potential

evapotranspiration calculated using the Evapotranspiration R package (Guo et al., 2016) are also presented.

The catchment precipitation variance, indicative of storm variability across the catchment, is also presented using vertical black lines (Figure 2a). The event in mid-August was clearly not uniformly spread across the catchment. Assuming this spatial variability is real, and not an artefact of the interpolation of the rain gauges, is likely to affect the resultant hydrograph peak flow (Emmanuel et al., 2017), catchment water balance (Zhao et al., 2013) and internal fluxes (Sapriza-Azuri et al., 2015).

To further explore precipitation spatial variability, Figure 2b shows the monthly area weighted precipitation averaged over 1910-2017 inclusive and Figure 2c shows the monthly mean daily spatial variance. Comparing Figure 2b and 2c, June through August have the highest mean precipitation but low spatial variance. Conversely, February has almost the lowest mean precipitation but high spatial variability. These seasonal differences are consistent with precipitation during summer being primarily from thunderstorms while during winter precipitation primarily is from fronts (Kraus, 1954).

4. Conclusions

AWAPer is an R-package which allows the extraction of both point and catchment area weighted metrological data (including evaporation) across Australia. One of the most common issues plaguing hydrological and climatological analyses is data availability, and even when gridded data is available it is often in formats which are difficult to manipulate. This data note presents the *AWAPer* R-package which overcomes many of these limitations. *AWAPer* allows for construction of data cubes which can be easily updated. Daily area weighted time-series are easily extracted at any location in Australia and future developments will allow analysis of monthly, seasonal and annual data and additional area weighted analysis functions will soon be added. Although deciding which data set is best for any application is an arbitrary decision (Tozer et al., 2012), the use of standard climatological products, particularly in decision making, is likely to aid more consistent

design outcomes and this package represents a possibility for standardised hydrologic assessments.

Acknowledgements

This research, and the *AWAPer* R-package, were supported by the Victorian Water and Climate Initiative (Government of Victoria, Australia). The ACORN-SAT data set is freely available through the Bureau of Meteorology website (<http://www.bom.gov.au/climate/change/acorn-sat>). The high-quality precipitation can be found at <http://www.bom.gov.au/climate/change/datasets/datasets.shtml>. Additional catchment boundaries can be obtained from <http://www.bom.gov.au/water/geofabric/download.shtml>. Conrad Wasko receives funding from the University of Melbourne McKenzie Postdoctoral Fellowships Program.

Data Availability Statement

The source code that support the findings of this study is open-source and available on GitHub at <https://github.com/peterson-tim-j/AWAPer>. The input data to the software is available in the public domain: <http://www.bom.gov.au/jsp/awap/index.jsp>.

References

- Beesley, C. A., Frost A. J. & Zajackowski J. (2009). A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets. In Anderssen, R.S., R.D. Braddock and L.T.H. Newham (eds) 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, pp. 3886-3892. ISBN: 978-0-9758400-7-8. <https://www.mssanz.org.au/modsim09/I13/beesley.pdf>
- BOM. (2012). Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT): Station catalogue. Melbourne, Australia. Retrieved from <http://www.bom.gov.au/climate/change/acorn-sat/documents/ACORN-SAT-Station-Catalogue-2012-WEB.pdf>

- 190 Chaubey, I., Haan, C. T., Grunwald, S., & Salisbury, J. M. (1999). Uncertainty in the model
191 parameters due to spatial variability of rainfall. *Journal of Hydrology*, 220(1-2), 48-61.
- 192 Emmanuel I., Payraastre, O., Andrieu, H., Zuber, F. (2017), A method for assessing the influence of
193 rainfall spatial variability on hydrograph modeling. First case study in the Cevennes Region,
194 southern France, *Journal of Hydrology*, 555, pp. 314-322, DOI: 10.1016/j.jhydrol.2017.10.011.
- 195 Faurès, J. M., Goodrich, D. C., Woolhiser, D. A., & Sorooshian, S. (1995). Impact of small-scale
196 spatial rainfall variability on runoff modeling. *Journal of hydrology*, 173(1-4), 309-326.
- 197 Frost, A. J., Ramchurn, A., & Hafeez, M. (2015). Evaluation of AWRA-L for national drought and
198 soil moisture monitoring. *The Art and Science of Water - 36th Hydrology and Water Resources*
199 *Symposium, HWRS 2015*, 1496–1505. Retrieved from
200 [https://www.scopus.com/inward/record.uri?eid=2-s2.0-](https://www.scopus.com/inward/record.uri?eid=2-s2.0-84974663165&partnerID=40&md5=9f5a08ad6fef1e6f416fd1be80290dfc)
201 [84974663165&partnerID=40&md5=9f5a08ad6fef1e6f416fd1be80290dfc](https://www.scopus.com/inward/record.uri?eid=2-s2.0-84974663165&partnerID=40&md5=9f5a08ad6fef1e6f416fd1be80290dfc)
- 202 Guo, D., Westra S., Maier H. R. (2016), An R package for modelling actual, potential and
203 reference evapotranspiration, *Environmental Modelling & Software*, 78, DOI:
204 10.1016/j.envsoft.2015.12.019.
- 205 Holgate, C., De Jeu, R. A. M., van Dijk, A. I. J. ., Liu, Y. ., Renzullo, L. J., Vinodkumar, et al.
206 (2016). Comparison of remotely sensed and modelled soil moisture data sets across Australia.
207 *Remote Sensing of Environment*, 186, 479–500. <https://doi.org/10.1016/j.rse.2016.09.015>
- 208 Hwang, Y., M. Clark, B. Rajagopalan, and G. Leavesley (2011), Spatial interpolation schemes of
209 daily precipitation for hydrologic modeling, *Stochastic Environ. Res. Risk Assess.*, 26(2), 295–320,
210 doi:10.1007/ s00477-011-0509-1
- 211 Jeffrey, S. J., J. O. Carter, K. B. Moodie, and A. R. Beswick. (2001). Using Spatial Interpolation to
212 Construct a Comprehensive Archive of Australian Climate Data. *Environmental Modelling &*
213 *Software* 16: 309–330.
- 214 Jones, D., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia.
215 *Australian Meteorological and Oceanographic Journal*, 58, 233–248.
- 216 Kraus, E. B. (1954). Secular changes in the rainfall regime of south-eastern Australia. *Quarterly*
217 *Journal of the Royal Meteorological Society*, 80, 591-601.

- 218 Lavery, B., Kariko, A., & Nicholls, N. (1992). A historical rainfall data set for Australia. Australian
 219 Meteorological Magazine, 40(1), 33-39
- 220 Ledesma, J. L. J., & Futter, M. N. (2017). Gridded climate data products are an alternative to
 221 instrumental measurements as inputs to rainfall–runoff models. Hydrological Processes, 31(18),
 222 3283–3293. <https://doi.org/10.1002/hyp.11269>
- 223 Nathan, R., P. Jordan, M. Scorch, S. Lang, G. Kuczera, M. Schaefer, and E. Weinmann. (2016).
 224 Estimating the Exceedance Probability of Extreme Rainfalls up to the Probable Maximum
 225 Precipitation. Journal of Hydrology 543: 706–720.
- 226 Nathan, R. J., & McMahon, T. A. (2017). Recommended practice for hydrologic investigations and
 227 reporting. Australian Journal of Water Resources, 21(1), 3–19.
 228 <https://doi.org/10.1080/13241583.2017.1362136>
- 229 Peel, M. C., McMahon, T. A. & Finlayson, B. L. (2004). Continental differences in the variability of
 230 annual runoff-update and reassessment. Journal of Hydrology, 295,
 231 <https://doi.org/10.1016/j.jhydrol.2004.03.004>
- 232 Peterson, T. J. and Fulton, S. (2019), Joint estimation of gross recharge, groundwater usage and
 233 hydraulic properties within HydroSight. Groundwater. Accepted Author Manuscript.
 234 doi:10.1111/gwat.12946
- 235 Peterson, T. J. & Western, A. W. (2014). Nonlinear time-series modeling of unconfined
 236 groundwater head. Water Resources Research, 50, 8330–8355.
 237 <https://doi.org/10.1002/2013WR014800>
- 238 Peterson, T. J., & Western, A. W. (2018). Statistical interpolation of groundwater hydrographs.
 239 Water Resources Research, 54. <https://doi.org/10.1029/2017WR021838>
- 240 Peterson, T. J., Western, A. W. & Cheng, X. (2018). The good, the bad and the outliers:
 241 automated detection of errors and outliers from groundwater hydrographs. Hydrogeology Journal,
 242 26, <https://doi.org/10.1007/s10040-017-1660-7>
- 243 Saft, M., Western, A. W., Zhang, L., Peel, M. C. & Potter, N. J. (2015). The influence of multiyear
 244 drought on the annual rainfall-runoff relationship: An Australian perspective. Water Resources
 245 Research, 51, <https://doi.org/10.1002/2014WR015348>

- 246 Sapriza-Azuri, G., Jódar, J., Navarro, L. J. Slooten, J. Carrera, and H. V. Gupta (2015), Impacts of
 247 rainfall spatial variability on hydrogeological response, *Water Resour. Res.*, 51, pp. 1300-1314,
 248 doi:10.1002/2014WR016168.
- 249 Tozer, C. R., A. S. Kiem, and D. C. Verdon-Kidd. (2012). On the Uncertainties Associated with
 250 Using Gridded Rainfall Data as a Proxy for Observed. *Hydrology and Earth System Sciences* 16:
 251 1481–1499.
- 252 Trewin, B. (2013). A daily homogenized temperature data set for Australia. *International Journal of*
 253 *Climatology*, 33(6), 1510–1529. <https://doi.org/10.1002/joc.3530>
- 254 van Dijk, A., Beck, H., Crosbie, R., de Jeu, R., Liu, Y., Podger, G., Timbal, B. & Viney, N. (2013).
 255 The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and
 256 implications for water resources, ecosystems, economy, and society. *Water Resources Research*,
 257 49, <https://doi.org/10.1002/wrcr.20123>
- 258 Viney, N., Vaze, J., Crosbie, R., Wang, B., Dawes, W., & Frost, A. (2015). AWRA-L v5.0: technical
 259 description of model algorithms and inputs. <https://doi.org/10.4225/08/58518bc790ff7>
- 260 Wasko, C., & Nathan, R. (2019). Influence of changes in rainfall and soil moisture on trends in
 261 flooding. *Journal of Hydrology*, 575, 432-441.
- 262 Wasko, C., Sharma, A., & Rasmussen, P. (2013). Improved spatial prediction: A combinatorial
 263 approach. *Water Resources Research*, 49(7), 3927–3935. <https://doi.org/10.1002/wrcr.20290>
- 264 Zhang, X. S., Amirthanathan, G. E., Bari, M. A., Laugesen, R. M., Shin, D., Kent, D. M., et al.
 265 (2016). How streamflow has changed across Australia since the 1950s: evidence from the network
 266 of hydrologic reference stations. *Hydrology and Earth System Sciences*, 20(9), 3947–3965.
 267 <https://doi.org/10.5194/hess-20-3947-2016>
- 268 Zhao F., Zhang L., Chiew F.H.S., Vaze J., Cheng L., (2013), The effect of spatial rainfall variability
 269 on water balance modelling for south-eastern Australian catchments, *Journal of Hydrology*, 493,
 270 pp. 16-29, DOI: 10.1016/j.jhydrol.2013.04.028.

273 **Figure 1.** Data extracted at Bathurst (063005) compared to station observation. Double mass curves for (a) minimum
274 daily temperature (b) daily precipitation. The difference between station observations and extracted data for (c) minimum
275 daily temperature and (d) daily precipitation.

276

277 **Figure 2.** (a) Time series of daily area weighted catchment average precipitation, streamflow, AWRA-L soil moisture, and
278 calculated potential evapotranspiration (PET) for the 351.7km² Bet Bet Creek catchment at Norwood (407220). The
279 vertical dark grey bars present +/- one standard deviation of the observed area weighted catchment precipitation. Spider
280 plots of (b) the area weighted mean monthly precipitation and (c) the mean monthly spatial variance at Bet Bet Creek at
281 Norwood (407220) which show that the months of highest mean precipitation have the lowest spatial variance and vice
282 versa.

hyp_13637_fig1.eps

hyp_13637_fig2.eps

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Peterson, TJ;Wasko, C;Saft, M;Peel, MC

Title:

AWAPer: An R package for area weighted catchment daily meteorological data anywhere within Australia

Date:

2020-02-28

Citation:

Peterson, T. J., Wasko, C., Saft, M. & Peel, M. C. (2020). AWAPer: An R package for area weighted catchment daily meteorological data anywhere within Australia. *Hydrological Processes*, 34 (5), pp.1301-1306. <https://doi.org/10.1002/hyp.13637>.

Persistent Link:

<http://hdl.handle.net/11343/286684>