# Synthesising multiple observations into annual environmental condition reports: the OzWALD system and Australia's Environment Explorer

**A.I.J.M. Van Dijk[a]** (ID) **and J. Rahman[b]**

[a] *ANU Centre for Water and Landscape Dynamics, Fenner School of Environment & Society, The Australian National University, ACT*
[b] *FlowMatter Pty. Ltd., Canberra, ACT*
*Email: albert.vandijk@anu.edu.au*

**Abstract:** Routine, nation-wide environmental data collection, analysis, and reporting remains a challenge in Australia. We wished to test what environmental data is already being collected and can be usefully synthesised, interpreted and reported upon, and to develop a deeper understanding of the scientific and technological opportunities and challenges. To that end, we developed a fully functional annual environmental reporting process, incorporating an automated workflow for data acquisition, integration and summary; a website to deliver the summary the data in a visually informative form; and an annual interpretation, reporting and communication process. Here we report mainly on the technological aspects.

The Australian Water and Landscape Dynamics (OzWALD) system is a data production workflow that can be run on demand. It includes the following components: (1) *Data acquisition* from several gridded climate and satellite data sources; (2) *Data reformatting*, e.g., spatial sub-setting, reprojection, temporal aggregation, and vector-grid transformation; (3) *Data fusion and enhancement*, referring to the blending of like data into a single best-estimate data set, and the improvement of desirable data characteristics, respectively. This results in a blended satellite-gauge precipitation data set, an internally-consistent set of dynamic water and land surface properties derived from satellite products (WALDMORF), and downscaled climate data at 500-m resolution; (4) *Model-data assimilation*, where a biophysical model of the same lineage as the AWRA Community Modelling System is infused with the observational data to estimate additional variables (soil moisture, streamflow generation and vegetation carbon uptake); and (5) *Statistical summary* by year, region and land use type, deriving temporal and regional statistics, used predominantly for web-based visualisation.

A website, *Australia's Environment Explorer* (AEE, www.ausenv.online), was developed to allow users visualise and explore environmental changes by region, location or land use type. The data can be queried in different ways and are visualized in charts. There have been four annual updates to the AEE since 2016, coinciding with the accompanying *Australia's Environment* report and public briefing. In addition to the AEE, the report also synthesises information from other sources and interprets specific events and temporal trends in global or non-spatial data.

Our experiences demonstrate that it is feasible to produce useful, observation-based annual environmental reports. Developing the experimental system and reporting process produced valuable insights, including: (1) the unstable ad continually evolving spatial environmental data services environment is a challenge for maintaining an operational workflow; (2) the rapid development of open source and cloud technologies provides major opportunities and efficiencies; (3) regular, detailed and accurate land cover and land use mapping will be required to achieve successful environmental accounting; (4) the abundance of past, current and future satellite mission observations provides many opportunities for environmental reporting; and (5) obtaining regular and reliable data on biodiversity remains a major challenge. Our goal is to continue, and as much as possible, improve the annual modelling and reporting process until a similar service is available from another source.

*Keywords: Data fusion, model-data assimilation, environmental data, spatial web services*

## 1.    INTRODUCTION

A 2014 review of national environmental reporting and accounting in Australia found that "*Despite strong demand for information [and] considerable effort and investment, nation-wide environmental data collection and analysis remains a substantially unmet challenge.*" (van Dijk et al., 2014). The authors identified a variety of reasons for the lack of progress, from institutional, societal and political to scientific and technical. They identified six challenges in science and technology, including (1) developing technologies that make better use of satellite and *in situ* data; (2) environmental modelling approaches to integrate the available observations; and (3) environmental accounting to promote consistency and isolate research and development priorities (van Dijk et al., 2014).

The 2014 review directly motivated the research reported on in this paper. We wished to demonstrate which environmental data are already being collected and can be usefully synthesised, interpreted and reported; and develop a deeper understanding of the scientific and technological opportunities and challenges towards comprehensive environmental reports and accounts. The review and initial research made it clear that these goals could not be achieved through academic study alone, as many challenges are operational and only manifest themselves when attempting regular information production. Hence, our research objective became to develop a functional annual environmental reporting process, incorporating (1) technology for routine data acquisition, integration and summary; (2) a web application to communicate the data in an intuitive and informative manner; and (3) an associated annual interpretation, reporting and communication process. In this paper, our primary focus is on the technological aspects, which include the data processing and modelling components, but also the web application and communication products.

## 2.    WORKFLOW

### 2.1.    Introduction

The Australian Water and Landscape Dynamics (OzWALD) system is a modular data acquisition and processing workflow that has been developed over the past four years to support the *Australia's Environment* report released annually by the Australian National University's Centre for Water and Landscape Dynamics (ANU-WALD). The workflow is implemented using a combination of MatLab and Python scripts and libraries and executed on the National Computational Infrastructure[1]. The workflow is largely automated but does require some manual operations in data acquisition. The processing requirements amount to an equivalent of approximately 30 thousand processor-hours. The computationally most demanding steps are deployed as distributed processes on NCI's 'Raijin' High-Performance Computing infrastructure. The sequential workflow components include data acquisition; data reformatting; data fusion and enhancement; model-data assimilation; and the calculation of summary statistics. Each are described in turn below.

### 2.2.    Data acquisition

Most input data are sourced from NCI's file system. They include locally updated copies of (1) satellite-derived surface reflectance and albedo (product code MCD43A4.006 and MCD43A3.006) (Schaaf and Wang, 2015), vegetation leaf area index (LAI, MCD15A3H.006) (Myneni et al., 2015) and land surface temperature (LST, MYD11A1.006) (Wan et al., 2015) produced by NASA; (2) estimates of the fractions of bare soil, photosynthetic vegetation and non-photosynthetic vegetation derived from MCD43A4.006 (Guerschman and Hill, 2018) and produced by CSIRO; (3) Landsat-based water presence mapping data (Water Observations from Space) (Mueller et al., 2016) available from the Digital Earth Australia database maintained by Geoscience Australia (GA); (4) precipitation and short-wave incoming radiation produced as part of the Australian gridded climate data (AGDC) (Jones et al., 2009) by Bureau of Meteorology (BoM); (5) hourly temperature, humidity, pressure and wind speed produced by the European Centre for Medium-range Weather Forecasting (ERA5 , ECMWF) (Copernicus Climate Change Service, 2017). These data are available and accessible to NCI users, while most can also be accessed via the public-facing NCI THREDDS[2]. In addition, we have been downloading the following data as a scheduled daily process or on demand: (6) satellite-derived rainfall estimates (IMERG, Huffman et al., 2015) from NASA[3]; (7) satellite-detected fire occurrence and intensity from GA's Sentinel Hotspots system (Geoscience Australia, 2014)[4]; and (8) fire carbon emission

---

[1] www.nci.org.au
[2] http://dapds00.nci.org.au/thredds/
[3] https://pmm.nasa.gov/data-access/downloads/gpm
[4] https://sentinel.ga.gov.au/

estimates (Kaiser et al., 2012) from ECMWF[5]. Data acquisition is the least robust component of the workflow and has consistently been failing every year in one or more ways. Supervision and ad hoc modifications are required to adapt to discontinued services, superseded datasets, and changes in access protocols, files, databases or data services.

## 2.3. Data reformatting

All original data are resampled and reformatted to a single standard in geographic coordinates and NetCDF file format, a self-described spatio-temporal gridded data format. Depending on the dataset, there may be spatial sub-setting, changes in spatial projection, temporal aggregation, and mapping vector data to a grid. The minimum temporal resolution is daily, but otherwise the spatial and temporal resolution are kept equivalent to the original data. Workflow failures in this processing step have occurred on occasion where the file format or conceptual structure of the data were changed.

## 2.4. Data fusion and enhancement

Data fusion refers to the blending of like data into a single consistent best-estimate data set, whereas data enhancement creates more desirable data characteristics; in this case, reduced noise or increased spatial resolution. Data fusion (or blending) is applied in the OzWALD system to produce observation-based estimates of precipitation and surface water extent, respectively. Estimates of daily accumulated rainfall are derived by merging the AGCD and IMERG grids. The AGCD are derived by interpolation of station data and are more accurate than the IMERG estimates where gauge density is >1 per ~500 km$^2$ but become gradually less accurate at greater distance (Renzullo et al., 2011). An inverse-distance weighting scheme blends the two datasets. The Water And Landscape Dynamics Multi-Observation Reanalysis and Filtering (WALDMORF) algorithm is a workflow module that produces a surface property synthesis by merging the fractions of surface water, snow, bare soil and living and dead vegetation, as well as LAI and albedo, into an internally-consistent data set at 500-m and 8-day resolution. A pixel-based algorithm interpolates the more accurate but frequently missing Landsat-based estimates of surface water fraction to train a machine-learning algorithm using the MODIS reflectance data. Estimates of fraction snow cover are produced with a simple pixel-based algorithm based on MODIS albedo and LST. Missing values in all time series are interpolated using the other series, as well as reflectances and derived indices, where they are available and a statistical relationships exist. As a final step, heuristic rules are applied to ensure the sum of fractions equals unity and a temporal filter that detects and interpolates outliers and missing values are passed over the respective time series.

Another form of data enhancement is the downscaling of gridded climate data to 500-m resolution. Key to the downscaling is the 1-second Hydrologically Enforced Digital Elevation Model (DEM-H) derived from observations by NASA's Shuttle Radar Topography Mission in 2000 (Gallant et al., 2011). This DEM formed the basis for derived products including the estimated ratio between average monthly terrain illumination from incoming shortwave radiation and that for a hypothetical horizontal plane, produced by CSIRO using the SRAD algorithm (Wilson and Gallant, 2000)[6]. Both datasets were resampled to 500-m resolution to support downscaling. AGDC short-wave radiation was downscaled using the SRAD data. To account for the effect of elevation on air temperature, screen level air temperature was downscaled by dynamically calculating the vertical temperature lapse rate from the multiple ERA5 pressure levels and using the lapse rate in combination with the DEM (Figure 1). Vapour pressure was downscaled by assuming constant specific humidity within the ERA5 grid cells and over elevation range using standard meteorological equations. Wind speed, down-welling longwave radiation and precipitation were not downscaled so far, but have been trialled and may be included in future releases.
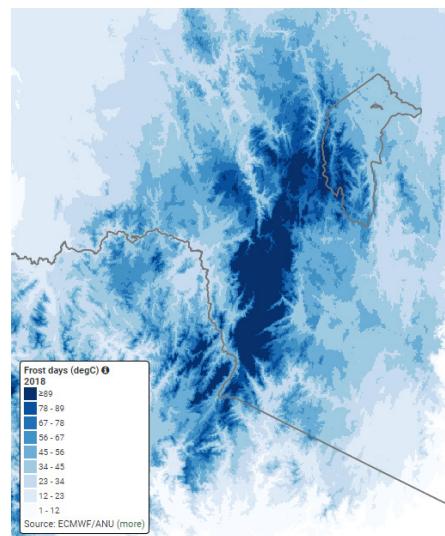


**Figure 1**. Number of frost days in 2018 in the Snowy Mountains region, illustrating downscaled air temperature (source: www.ausenv.online).

---

### 2.5. Model-data assimilation

In the model-data fusion step, the eponymous OzWALD biophysical model is used to estimate environmental variables for which spatial data are not directly available. Reported output variables include root zone soil moisture, streamflow generation, and carbon uptake by the vegetation. (The model produces numerous other energy, water and carbon balance variables, some of which are stored on disk, but they are not currently used in reporting.) The OzWALD model evolved from the original AWRA-L model v0.5 (Van Dijk, 2010), which formed the core of BoM's operational daily landscape water balance modelling system, the associated spatial information service[7], and the AWRA Community Modelling System[8]. The operational model at the time of writing (version 6) is largely identical to the original model (version 0.5) in terms of model theory and equations, with subsequent updates focusing on enhanced spatial parameter fields and model parameter calibration (Frost et al., 2016). Most of those changes have been imported into OzWALD. Key differences are that OzWALD runs at higher spatial resolution (~500 m *vs.* ~5 km); includes simulation of snow hydrology, vegetation carbon cycle processes, and grid-based river flow routing; and assimilates dynamic satellite observations (specifically, satellite precipitation and WALDMORF surface properties).

At the highest level, the OzWALD system consists of a (1) HPC job generation module that deploys the workflow for specified model domains; (2) a model workflow management module that proceeds through the various modelling steps on the HPC node; and (3) a post-processing module that collates the domain outputs into annual national grids and stores them in NetCDF format. At the next level down, the model workflow proceeds through the following steps: (2.1) setting the system environment; (2.2) initialising the static model parameters and model states stored on disk; (2.3) an optional spin-up model run for a specified period to produce 'hot' initial states if needed; and (2.4) a final model run that produces and stores the requested variables. The spin-up and final model run use the same workflow, which consists of retrieving the relevant (2.4.1) climate forcing data and (2.4.2) satellite-derived surface properties from disk and transforming those to the model-expected format; (2.4.3) the assimilation step, which updates several model parameters and states using the satellite observations using a simple nudging scheme; (2.4.4) the daily time step model, which evolves the states one day forward; and (2.4.5) a module that concatenates the requested data and writes them to temporary storage.

### 2.6. Statistical summary by year and region

The daily or 8-daily, continental, 500-m resolution data files produced in the previous steps are published through NCI THREDDS. However, they are likely to be unattractive or challenging to work with for many users. Therefore, temporal and regional summary statistics are calculated in a final post-processing step, and these summary data are also what users are most likely to visualise through web applications. The following summary data are computed:

1. *Annual summary maps* at the original spatial data resolution, for the purpose of map visualisation. In most cases, minimum, maximum and average and/or accumulated daily values are calculated for the year, and these are stored in gridded NETCDF format. In some cases, slightly more elaborate summary statistics are calculated. They are stored as multi-year files and published through NCI THREDDS.
2. *Regional-average annual time series*. The annual summary maps are intersected with different region definitions derived from thematic vector maps (see Section 3). Because of their small data volume, these summary data are stored as CSV files on an external web server.
3. *Regional-average sub-annual time series.* These are calculated following the same approach, but at the original temporal resolution (i.e. daily or 8-daily). They are also stored as NetCDF files on NCI THREDDS. The two spatial dimensions are replaced by the region identifier.
4. *Annual regional summaries by broad land use type.* These are computed similar to (2), but further stratifying by broad land-use type, and as cumulative spatial sums rather than averages where appropriate. The land-use classes are an amalgamation of the many classes in the 50-m resolution Catchment-scale Land Use Data (ABARES, 2017) into 20 broad classes. The resulting data are quite comparable to that used to construct environmental accounts.

---

[7] http://www.bom.gov.au/water/landscape/
[8] https://github.com/awracms/awra_cms
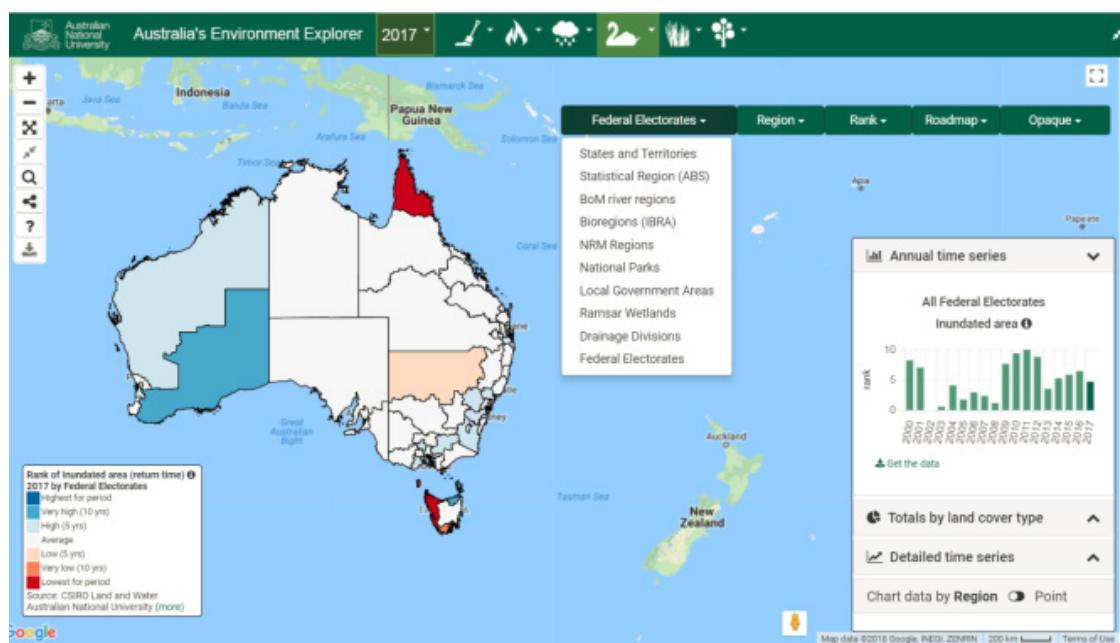
**Figure 2**. Screen shot of Australia's Environment Explorer ([www.ausenv.online](www.ausenv.online))

## 3.    DATA DELIVERY AND COMMUNICATION

*Australia's Environment Explorer* (AEE, Fig. 2) is a web application that was designed to make it easier for non-specialist users to visualise and investigate environmental changes by region, location or land use type. The AEE aims to provide direct visual access to the summary data described in Section 2.6 in the form of maps and charts. The AEE currently offers data on 13 indicators in 6 themes. All are produced at some stage in the OzWALD workflow; some are only reformatted (e.g., fire-related data), whereas others are the result of downstream model-data assimilation (soil moisture, runoff and vegetation carbon uptake).

The data are visualized as bar charts of annual values for the entire period (typically 2000 onwards), as pie charts attributing variables to main land-use classes, and as sub-annual time series for the selected years. Data can be queried for a particular location, or as the average across a predefined region. Available region definitions currently include political and administrative units (states and territories, local government areas, federal electorates, natural resource management regions, national parks, Ramsar wetlands), biophysical units (bioregions, river regions and drainage divisions) and statistical accounting regions. The user can download the regional summary data directly via the website or edit them online using the plotly ([http://plot.ly/](http://plot.ly/)) visualization app. The original and summary gridded time series can be accessed and downloaded from the NCI Data Collection, which provides an API to the data. All data is provided under a Creative Commons license (CC BY 3.0 AU).

The sub-annual and annual data produced by OzWALD are also made available through the ANU-TERN Landscape Data Visualiser ([http://maps.tern.org.au](http://maps.tern.org.au)), a web app designed to help researchers and other interested users to visualise and explore data collected or produced by the Terrestrial Ecosystem Research Network or ANU-WALD. Unlike the AEE, this website provides tools to visualise sub-annual gridded data, and to compare different datasets (e.g. flux tower measurements and gridded data) for a location of interest.

Since the first release in 2016, there have been four updates to the AEE, coinciding with the release of the annual *Australia's Environment* report and briefing by ANU-WALD. This release has been occurring between February and April after the year reported on, as a function of data availability and latency, required workflow modifications or innovation tests, and available resources.

Analogous to 'headline indicators' on other topics, regional and national Environmental Condition Scores (ECS) or 'report cards' are produced. The ECS is calculated from seven variables (Fig. 3). First, the actual annual values of each variable are replaced by their relative ranking in the time series for 2000 onwards and rescaled to the range 0–10 to produce a component score, with 10 indicating the 'most desirable' score. The ECS is subsequently calculated as the mean of the component scores.
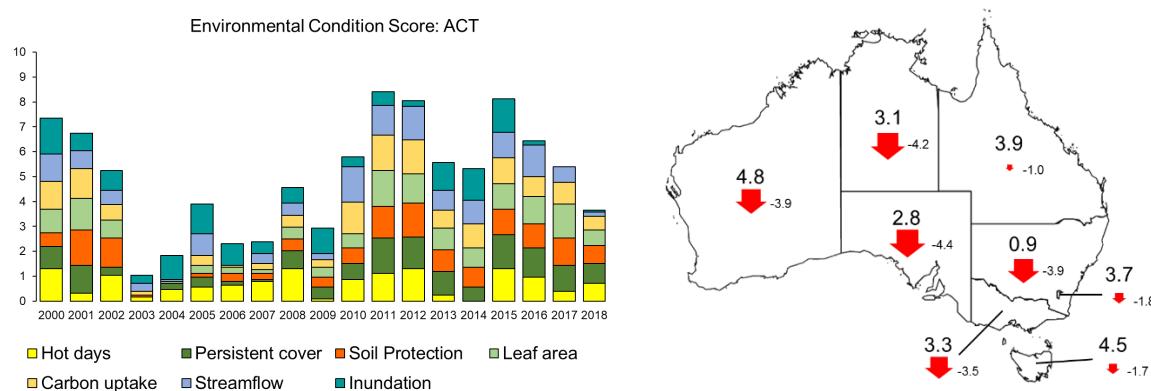
**Figure 3**. Environmental Condition Scores for 2018 shown (left) for the ACT as a time series along with its components, and (right) by State and Territory as a map indicating movement from the previous year.

However, interpretation of the ECS component scores confirms the scientific consensus that environmental condition is directly dependent on water availability for most of Australia, which partially mitigates its limitations. The reports emphasize the subjectivity and limitations of the score. Perhaps most limiting is the lack of national-scale information and hence inclusion of changes in important environmental aspects, such as biodiversity, land clearing and soil health.

Beyond reporting on the AEE data, the *Australia's Environment* report also synthesises new and up-to-date information on environmental conditions from other credible sources within and outside Australia (e.g., the Australian Bureau of Statistics, BoM, Integrated Marine Observing System, Terrestrial Ecosystem Research Network, and Atlas of Living Australia, World Meteorological Organisation). This provides additional opportunities to interpret information and report on time series representing smaller or unspecified areas within Australia (e.g., bio-indicators and citizen science), specific events and their impacts (e.g., cyclones), different parts of the environment (e.g., the marine environment), and the global context (e.g., greenhouse gas concentrations, global warming and sea level rise).

*Australia's Environment* report is presented as (a) a four-page factsheet with key dot points and figures, (b) the slide pack used in the annual briefing, providing a more comprehensive visual summary, and (c) an article in The Conversation, containing a short summary and interpretation in accessible language. Reports for previous years can be found on the *Australia's Environment* website (http://wald.anu.edu.au/australias-environment/).

## 4. DISCUSSION AND CONCLUSIONS

Our stated goal was to gain further insights into the possibilities and barriers in developing a system that can routinely provide information on environmental conditions in Australia in a more comprehensive way than available. We described the OzWALD workflow system, which combines and synthesises several sources of spatial information into a biophysical model. Next, we described the AEE website as part of a broader data delivery and communication strategy, targeting different audiences.

Building an operational workflow has provided insights that would have been unlikely to have been obtained by other means. This includes first-hand insights into the challenges of maintaining a (pseudo-) operational workflow in the face of rapidly changing data services. The stability, reliability, latency and continuity of data services are all important factors that constrain what can be achieved and the associated effort. A case in point was the replacement of one data collection (MODIS Collection 5) with another (Collection 6), which on the face of it was a minor upgrade but caused a series of challenges. Conversely, the rapid development of open source technologies provided unexpected convenience, opportunities and efficiencies. This includes the establishment and popularity of common data formats (e.g., NetCDF), data service technologies (e.g., OpenDAP THREDDS and WMS/WCS), cloud technologies, and open libraries and frameworks for mapping and data visualisation.

Producing summaries by region and land use class demonstrated the technical feasibility of producing environmental accounts. However, there remain data challenges that would need to be addressed. Important among these is the production of reliable, high-resolution land cover and land use mapping products that are suited for annual change analysis. The developing GA Digital Earth Australia infrastructure and derived products bodes well in this regard. More generally, the rapidly increasing abundance and reduced cost of

environmental data from past, current and new satellite missions and derived modelling products also continues to provide new opportunities for environmental reporting. More challenging is the lack of regular and reliable observations on ecological health and biodiversity across the nation, and more broadly, the continued collection of field data to evaluate satellite- and model-derived spatial information. Citizen science, e.g., as harnessed by the ALA, and high-throughput sample analysis methods may provide the best opportunities but will require new approaches to deal with different nature and reduced accuracy of the data.

Overall, our research demonstrates that it is technically feasible to produce annual environmental reports based on a solid foundation of on-ground and satellite observations. Our intent is to continue to maintain and where possible further improve the reporting process and technology until a similar operational service has become available from another source.

## ACKNOWLEDGMENTS

## REFERENCES

ABARES, 2017. Catchment Scale Land Use of Australia - Update September 2017. ABARES, Canberra.

Copernicus Climate Change Service, 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. In: ECMWF (Editor). Copernicus Climate Change Service Climate Data Store (CDS), https://cds.climate.copernicus.eu/cdsapp#!/home.

Frost, A.J., Ramchurn, A. and Smith, A.B., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model, Melbourne.

Gallant, J.C. et al., 2011. 1 second SRTM Derived Digital Elevation Models User Guide, Geoscience Australia (available at www.ga.gov.au/topographic-mapping/digital-elevation-data.html), Canberra.

Geoscience Australia, 2014. Sentinel Hotspots Product Description Document V1.2, code D2014-145826, Geocat Reference 70869.

Guerschman, J.P. and Hill, M.J., 2018. Calibration and validation of the Australian fractional cover product for MODIS collection 6. Remote Sensing Letters, 9(7): 696-705.

Huffman, G.J. et al., 2015. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm theoretical basis document, version, 4: 30.

Jones, D., Wang, W. and Fawcett, R., 2009. High-quality spatial climate data-sets for Australia. Australian Meteorological and Oceanographic Journal, 58: 233-248.

Kaiser, J.W. et al., 2012. Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. Biogeosciences, 9(1): 527-554.

Mueller, N. et al., 2016. Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. Remote Sensing of Environment, 174: 341-352.

Myneni, R., Knyazikhin, Y. and Park, T., 2015. MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006. 2015. NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MCD15A3H.006.

Renzullo, L.J. et al., 2011. An Assessment of Statistically Blended Satellite-Gauge Precipitation Data for Daily Rainfall Analysis in Australia, ISRSE, Sydney.

Schaaf, C. and Wang, Z., 2015. MCD43A4 MODIS/Terra+ Aqua BRDF/Albedo Nadir BRDF Adjusted RefDaily L3 Global-500m V006. NASA EOSDIS Land Processes DAAC.

van Dijk, A., Mount, R., Gibbons, P., Vardon, M. and Canadell, P., 2014. Environmental reporting and accounting in Australia: Progress, prospects and research priorities. Science of The Total Environment, 473–474(0): 338-349.

Van Dijk, A.I.J.M., 2010. AWRA Technical Report 3. Landscape Model (version 0.5) Technical Description, WIRADA / CSIRO Water for a Healthy Country Flagship, Canberra.

Wan, Z., Hook, S. and Hulley, G., 2015. MYD11A1 MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006. NASA EOSDIS Land Processes DAAC.

Wilson, J.P. and Gallant, J.C., 2000. Secondary topographic attributes. Terrain analysis: Principles and applications: 87-131.