| | |
|---|---|
| **Name of dataset or data source:** | Wetland Vegetation of the Lachlan - Great Cumbung Swamp 2023 |
| **Custodian of the dataset or data source:** | ED Science (E&H) |
| **Description:** | This wetland vegetation map of the Great Cumbung Swamp was produced using a machine learning-based classification framework that integrates multi-source satellite and terrain with a cluster-guided training approach (Wen et al., 2025). Inputs and training data Inputs included Sentinel-1 synthetic aperture radar (SAR) time series, Sentinel-2 optical time series, and hydro-morphological variables derived from a gap-filled 5 m LiDAR digital elevation model (DEM) and hydrologically enforced shuttle radar topography mission (SRTM) DEM. To capture the high spatial and seasonal variability of wetland vegetation, K-means clustering was used to guide sample selection. Clusters were reviewed by an expert vegetation ecologist against high-resolution aerial and drone imagery, topographic context, and existing field data, and then assigned to plant community types (PCTs) where appropriate. The verified clusters formed the basis of the training dataset for a Random Forest classifier which used 48 predictors (spectral, temporal, structural, terrain). Model outputs were produced at three hierarchical class levels: NSW Vegetation Formations (L1: 9 classes), Functional (L2: 14 classes) and PCTs (L3: 23 classes). Post-processing and manual edits Following classification, model outputs were post-processed to enhance spatial coherence while preserving hydrologically meaningful patches. Steps included edge-aware smoothing and progressive gap-filling/merging with class-specific minimum mapping units (MMU): < 0.1 ha for non-woody wetland PCTs and < 0.2 ha for woody wetland PCTs. Outputs were then manually edited by an expert vegetation ecologist to resolve any residual artifacts and boundary issues. Model accuracy assessment The following metrics are the raw model output (before post-processing and editing) performance for each class level in Wen et al. (2025) (reported on internal independent test set). Metrics include Overall Accuracy (OA), Cohens Kappa (κ) and Matthews Correlation Coefficient (MCC): - NSW |

Vegetation Formations (L1): OA ≈ 97 %, κ ≈ 0.96, MCC ≈ 0.96; - Functional (L2): OA ≈ 94 %, κ ≈ 0.93, MCC ≈ 0.93; - PCTs (L3): OA ≈ 93 %, κ ≈ 0.91, MCC ≈ 0.89 Class hierarchy Labels were assigned at PCT level using the NSW BioNet Vegetation Classification (https://vegetation.bionet.nsw.gov.au/) and then aligned to the NSW framework's Vegetation Class and Formation levels (https://www.environment.nsw.gov.au/topics/animals-and-plants/biodiversity/nsw-bionet/the-nsw-vegetation-classification-framework). For water management reporting, each wetland PCT was aligned to a Monitoring, Evaluation and Reporting (MER) Functional Group consistent with the Lachlan Long-Term Water Plan (LTWP) (https://www.environment.nsw.gov.au/sites/default/files/lachlan-long-term-water-plan). Key fields dictionary 'PCT_ID' (PCT Code); 'PCT_Desc' (PCT Name); 'Veg_Class' (NSW Vegetation Class); 'Veg_Format' (NSW Vegetation Formation); 'MER_FG' (MER Functional Group for LTWP reporting); 'Hectares' (polygon area); 'DN' (classifier code) and 'Functional' (model specific functional group per Wen et al., 2025). Context classes ('Bare ground', 'Cleared/Disturbed', 'Open water', 'Dam') are included for completeness and accuracy assessment. Intended use Baseline for environmental water planning, MER reporting under the LTWP, conservation management, and long-term monitoring at landscape and site scales. Not intended for statutory site assessment without targeted field verification. Input data limitations Cloud, inundation state and sensor geometry may influence satellite image quality and contribute to classification error; LiDAR and ancillary datasets may differ in acquisition date from satellite inputs. Localised errors in source DEMs/orthophoto errors can propagate to terrain-derived predictors. Validation scope The above model accuracy metrics are from internal hold-out testing (80/20 train-test split) and repeated cross-validation of the expert-labelled dataset in Wen et al. (2025). A withheld, ground-based validation dataset collected independent of model training will be used to validate the final post-processed and edited map product; those results will be provided in future versions to supplement the raw model accuracy values for reporting purposes. Users requiring statutory-grade evidence should conduct targeted field verification. Versioning This version is v1.0 (release date: 2025-10-30). Results

| | |
|---|---|
| | are versioned; Identified errors will be corrected in subsequent releases with an accompanying changelog. Acknowledgements This mapping project was funded by the NSW Water for the Environment Program. Related publication Wen, L., Ryan, S., Powell, M., and Ling, J.E. (2025). From Clusters to Communities: Enhancing Wetland Vegetation Mapping Using Unsupervised and Supervised Synergy. Remote Sensing, 17(13): 2279. https://doi.org/10.3390/rs17132279 |
| **Data quality rating:** | Institutional Environment - 5<br>Accuracy - 4<br>Coherence - 3<br>Interpretability - 5<br>Accessibility - 4 |

## INSTITUTIONAL ENVIRONMENT — Excellent

Does the information have the potential to enhance services or service delivery?

The data aligns with the Data Quality Framework, including:

- Legislation
- Policies
- Information Asset Governance
- Standards
- Data Management Plans

The following governance roles and responsibilities for this asset are clearly assigned:

- Information Asset Owner
- Information Asset Custodian
- Information Steward

Data collection is authorised by law, regulation or agreement

The Custodial agency has no commercial interest or conflict of interest in the data

## ACCURACY — Very Good

Data has been subject to a data assurance process (for example: Checking for errors at each stage of data collection and processing, or verifying data entry and making corrections if necessary.)

There are no known gaps in the data or if there are gaps (for example: non-responses, missing records, data not collected), they have been identified in caveats attached to the dataset.

No changes have been made or other factors identified (for example: weighting, rounding, de-identification of data, changes or flaws in data collection or verification methods) that could affect the validity of the data; or any changes/factors have been identified in caveats attached to the asset.

The data collection met the objectives of the primary user. The data correctly represents what it was designed to measure, monitor or report.

Data is revised and the revision is published if errors are identified

i Find out more about the quality assurance processes from the NSW Government Standard for Data Quality Reporting. https://www.finance.nsw.gov.au/ict/resources/data-quality-standard

**COHERENCE** **Good**

Standard definitions, common concepts, classifications and data recording practices have been used.

Elements within the data can be meaningfully compared.

This data is generally consistent with similar or related data sources from the same discipline

The data can be analysed over time (for example, there have not been any significant changes in the way items are defined, classified or counted over time).

The data does not form part of a collection or, if it is the latest in a series of data releases, there have not been any changes in methodology or external impacts since the last data release.

**INTERPRETABILITY** **Excellent**

A data dictionary is available to explain the meaning of data elements, their origin, format and relationships

Information is available about the primary data sources and methods of data collection (e.g. instruments, forms, instructions).

Information is available to help users evaluate the accuracy of the data and any level of error

Information is available to explain concepts, help users correctly interpret the data and understand how it can be used

Information is available to explain ambiguous or technical terms used in the data

i Find out more about the data dictionary from the Custodian (contact details below).

i Find out more about the primary data sources and methods of data collection from the Custodian (contact details below).

i Find out more about concepts used in this dataset and how to understand or interpret the data from the Custodian (contact details below).

i Find out more about ambiguous or technical terms used in the data from the Custodian (contact details below).

**ACCESSIBILITY** **Very Good**

Data is available online with an open licence

Data is available in machine-processable, structured form (e.g. CSV format instead of an image scan of a table)

Data is available in a non-proprietary format (e.g. CSV, XML)

Data is linked to other data, to provide context (e.g. employee ID is linked to employee name or species name is linked to genus)

Data is described using open standards (e.g. RDF, SPARQL) and persistent identifiers (URIs or DOIs)

**DATA DISCLAIMER**

You must check and comply with the licensing conditions for the information you wish to use. This may require you to contact the Department, or other custodial agency, or the third party copyright owner for permission to use the material. You may also use any material in accordance with rights you may have under the Copyright Act 1968 (Cth), for example under the fair dealing provisions or statutory licences. Use

of material in a way not permitted by this copyright notice may be an infringement of copyright. Infringing copyright may expose you to legal action by, and liability to, the copyright owner. Wherever a third party holds copyright in material, the copyright remains with that party. Their permission may be required to use the material and you should contact that party directly. As far as practicable, material for which the copyright is owned by a third party will be clearly labelled. Excluded material can only be used under the specific terms of use attached to that material. If you want to use this material in a manner that is not covered by those specific terms of use, you must request permission from the copyright owner of the material.

The Department endeavours to make sure that information provided is correct at the time of its publication. However, as necessary you should obtain independent advice before making any decision based on the information. The information is made available on the understanding that custodial agencies and the State of NSW accept no responsibility for any damage, cost, loss or expense incurred by you as a result of:

- any error, omission or misrepresentation in the information provided
- without limiting the above, any delay, failure or error in recording, displaying or updating information, including but not limited to, data relating to credit holdings.

Custodial agencies and the State of New South Wales disclaim all responsibility and all liability (including without limitation, liability in negligence) for all expenses, losses, damages and costs you might incur as a result of the information being inaccurate or incomplete in any way, and for any reason.

| | |
|---|---|
| **For more information about this dataset or data source, contact:** | NSW Department of Climate Change, Energy, the Environment and Water |
| **Data Broker email:** | data.broker@environment.nsw.gov.au |
| **Data Broker phone:** | 131555 |

## Understanding the Data Quality Statement

The data quality statement aims to help you understand how a particular dataset could be used and whether it can be compared with other, similar datasets. It provides a description of the characteristics of the data to help you decide whether the data will be fit for your specific purpose.

**About the quality rating:**

The reporting questionnaire asks five questions for each of these data quality dimensions:

- Institutional Environment
- Accuracy
- Coherence
- Interpretability
- Accessibility

For each question: “yes” = 1 point; “no” = 0 points

The number of points determines the Quality Level for each dimension (high, medium, low).

Only dimensions with four or five points receive a star.

| Points | Quality Level | Star / No Star |
|---|---|---|

| | | |
|---|---|---|
| 0 | Poor | No Star |
| 1 | Poor | No Star |
| 2 | Fair | No Star |
| 3 | Good | No Star |
| 4 | Very Good | Star |
| 5 | Excellent | Star |

## Evaluating data quality

Quality relates to the data's "fitness for purpose". Users can make different assessments about the dataquality of the same data, depending on their "purpose" or the way they plan to use the data. The following questions may help you evaluate data quality for your requirements. This list is not exhaustive.Generate your own questions to assess data quality according to your specific needs and environment.

- What was the primary purpose or aim for collecting the data?
- How well does the coverage (and exclusions) match your needs?
- How useful are these data at small levels of geography?
- Does the population presented by the data match your needs?
- To what extent does the method of data collection seem appropriate for the information being gathered?
- Have standard classifications (eg industry or occupation classifications) been used in the collection of the data?If not, why? Does this affect the ability to compare or bring together data from different sources?
- Have rates and percentages been calculated consistently throughout the data?
- Is there a time difference between your reference period, and the reference period of the data?
- What is the gap of time between the reference period (when the data were collected) and the release date of thedata?
- Will there be subsequent surveys or data collection exercises for this topic?
- Are there likely to be updates or revisions to the data after official release?