

A scalable workflow for ecosystem condition assessment from Earth observation data

Roozbeh Valavi^{1,*}, Kristen J Williams^{2,*}, Eric A Lehmann², Simon Collings³, Shaun Levick⁴, Katherine M Giljohann¹, Ning Liu², Stephanie Johnson⁵, Samantha EM Munroe⁶, Thomas G Van Niel⁷, Matthew Paget², Cassandra Malley⁸, Peter Lyon⁸, Thomas D Harwood⁹, and Simon Ferrier²

¹CSIRO, Clayton, VIC 3168, Australia

²CSIRO, GPO Box 1700, Canberra, ACT 2601, Australia

³CSIRO, Kensington, WA 6151, Australia

⁴CSIRO, Waite Campus, Urrbrae, SA 5064, Australia

⁵CSIRO, GPO Box 1538, Hobart, TAS 7001, Australia

⁶CSIRO, Townsville, QLD 4811, Australia

⁷CSIRO, Floreat, WA 6014, Australia

⁸Australian Government Department of Climate Change, Energy, the Environment and Water, Canberra, ACT 2601, Australia

⁹Environmental Change Institute, University of Oxford, Oxford, OX1 3QY, United Kingdom

*Corresponding authors: roozbeh.valavi@csiro.au; kristen.williams@csiro.au

October 2026

Abstract

1. Quantifying ecosystem condition consistently across space and time is a central challenge in biodiversity monitoring, ecosystem accounting, and ecological restoration. Earth observation (EO) data provide extensive coverage, but translating remotely sensed ecosystem characteristics into ecologically meaningful measures of condition requires approaches that account for ecological context, and natural variability across space and time.

2. The Habitat Condition Assessment System (HCAS) addresses this challenge through a reference-based benchmarking framework that evaluates ecosystem characteristics relative to ecologically comparable high-integrity reference ecosystems. However, the absence of a general, reproducible implementation has limited broader application of this approach across regions and datasets.

3. Here, we present a scalable workflow that formalises the HCAS framework as an analytical system, in which the **ClassicHCAS** R package implements core ecosystem condition calculations. The workflow integrates remotely sensed ecosystem condition variables, environmental covariates, and reference ecosystem data within a benchmarking framework. Ecological context is addressed

through reference ecosystem modelling, while ecosystem condition is inferred by benchmarking observations against multiple ecologically comparable reference ecosystems. The modular workflow supports alternative modelling approaches, analytical configurations and calibration strategies while maintaining consistency within a common reference-based framework.

4. We demonstrate the workflow using a continental-scale application for Australia (Supporting Information), generating spatially explicit ecosystem condition estimates and multi-decadal time series from Landsat data aggregated at 90 m resolution. Results capture broad gradients of ecosystem modification, fine-scale spatial structure and temporal dynamics relative to a consistent long-term reference framework. By providing open-source implementations of the HCAS workflow, including the `ClassicHCAS` R package and accompanying workflow code, we provide a reproducible and scalable approach to ecosystem condition assessment from EO data that can be readily adapted to different regions, datasets and applications.

KEYWORDS: biodiversity monitoring, ecosystem accounting, ecosystem characteristics, ecosystem integrity, large-scale analysis, open-source software, remote sensing, reproducible workflow

1- INTRODUCTION

Quantifying ecosystem condition consistently across space and time is central to biodiversity monitoring, ecosystem-based management, natural capital accounting, and restoration planning (Luxton *et al.* 2026). Earth observation (EO) data provide unprecedented spatial and temporal coverage for monitoring ecosystem characteristics (Ustin & Middleton 2021), yet translating these observations into ecologically meaningful measures of ecosystem condition remains a persistent challenge (Pettorelli *et al.* 2018). Interpretation of EO-derived ecosystem characteristics requires approaches that account for ecological context, natural variability and human-driven modification (Anderson 2018; Senf 2022).

The Habitat Condition Assessment System (HCAS) was developed to address this challenge by linking EO-derived ecosystem characteristics to ecologically comparable high-integrity reference ecosystems (Harwood *et al.* 2016). Within this framework, ecosystem condition is inferred through reference-based benchmarking, where observed ecosystem characteristics are evaluated relative to the range of characteristics exhibited by ecologically comparable high-integrity reference ecosystems. This approach explicitly recognises that ecosystem condition is context-dependent and that high-integrity ecosystems may exhibit multiple valid expressions of structure, function and composition across common environmental settings.

HCAS was originally proposed as a general conceptual framework rather than a fixed analytical method, integrating EO time-series, environmental covariates and reference ecosystem information within a reference-based benchmarking approach (Harwood *et al.* 2016). Despite its conceptual strengths and demonstrated applicability, broader application of HCAS has been constrained by the absence of a general, reproducible computational workflow. Early applications were tailored to specific datasets, modelling decisions, and processing workflows, limiting transparency, extensibility, and comparability across regions and analyses (Harwood *et al.* 2016; Williams *et al.* 2021). This reflects a broader challenge in computational ecology, where translation of

conceptual frameworks into accessible, reproducible software is critical for uptake and cumulative methodological development (Sandve *et al.* 2013; Lowndes *et al.* 2017).

Here we describe an extensible workflow that operationalises the HCAS framework as a modular and reproducible analytical system, with **ClassicHCAS**, an open-source R package, implementing the core benchmarking, time-series estimation and calibration functions (Figure 1). The workflow explicitly separates reference ecosystem modelling component from ecosystem condition benchmarking, enabling alternative modelling approaches, data inputs and calibration strategies to be implemented within a consistent reference-based framework.

By formalising HCAS into a transparent and scalable workflow, with **ClassicHCAS** at its core, we provide a generalisable foundation for integrating EO data, environmental covariates and reference ecosystem data within a consistent benchmarking framework. The modular workflow can be applied across whole regions or stratified ecosystem assessments, with ecosystem characteristics modelled individually or jointly, providing flexibility for different applications while maintaining conceptual consistency. Full implementation of the workflow enables repeated, spatially explicit estimation of ecosystem condition across regions and through time, supporting applications in biodiversity monitoring (e.g. Mokany *et al.* 2022; Ward *et al.* 2024), restoration planning (e.g. Murphy *et al.* 2025), and ecosystem accounting (e.g. Giljohann *et al.* 2025).

In presenting the workflow, we clarify which components are implemented within **ClassicHCAS** and which are external to the package (Figure 1). Throughout, we refer to the Supporting Information, which demonstrates application of the workflow across continental Australia.

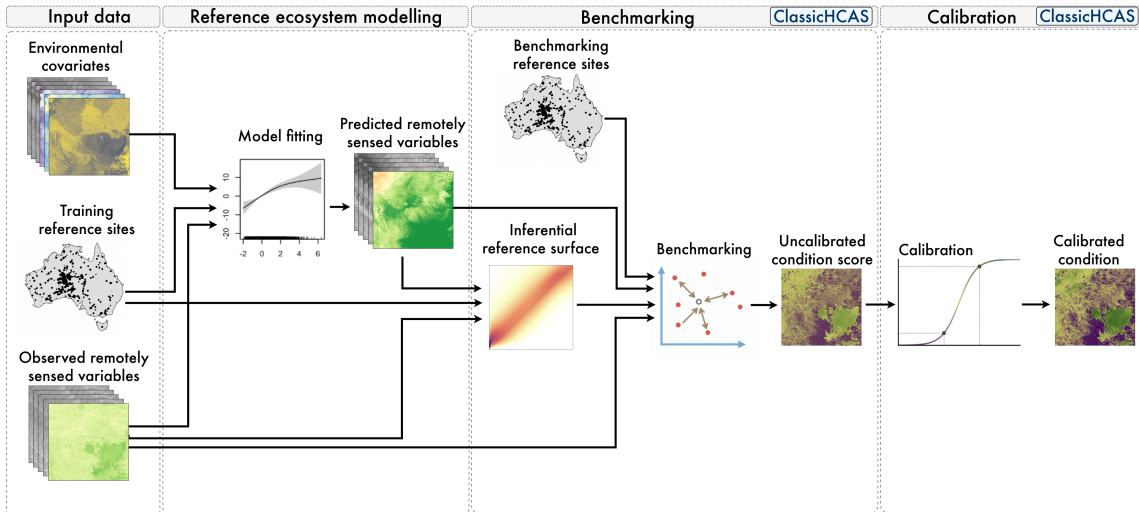


Figure 1: Diagrammatic workflow showing the main stages from input data and reference ecosystem modelling through benchmarking and calibration. Components implemented within the **ClassicHCAS** R package are indicated with blue labels.

2- WORKFLOW OVERVIEW AND DESIGN PRINCIPLES

The HCAS workflow is designed to address a set of ecological challenges inherent to estimating ecosystem condition from Earth observation data. These challenges shape the structure of

the workflow, the input data required, and the analytical procedures used to infer ecosystem condition. This section outlines the ecological rationale underpinning the workflow, focusing on their conceptual roles and relationships rather than their computational implementation. The workflow comprises a modular sequence of stages—data preparation, reference ecosystem modelling, condition benchmarking, time-series estimation, and calibration—each addressing a distinct aspect in ecosystem condition inference (Figure 1).

While the sequence and functional roles of these stages are fixed, the workflow is designed to be modular and analytically flexible (Figure 2). Alternative data inputs, reference ecosystem modelling approaches and condition-calibration strategies can be implemented within a consistent reference-based benchmarking framework. For example, analyses may be applied across entire regions or stratified by ecosystem type, and variables representing observed ecosystem characteristics may be modelled individually or jointly within a multivariate framework. This flexibility allows adaptation to different datasets and applications—such as ecosystem accounting (Czúcz *et al.* 2021b; United Nations *et al.* 2024)—without altering the underlying inferential logic of the workflow.

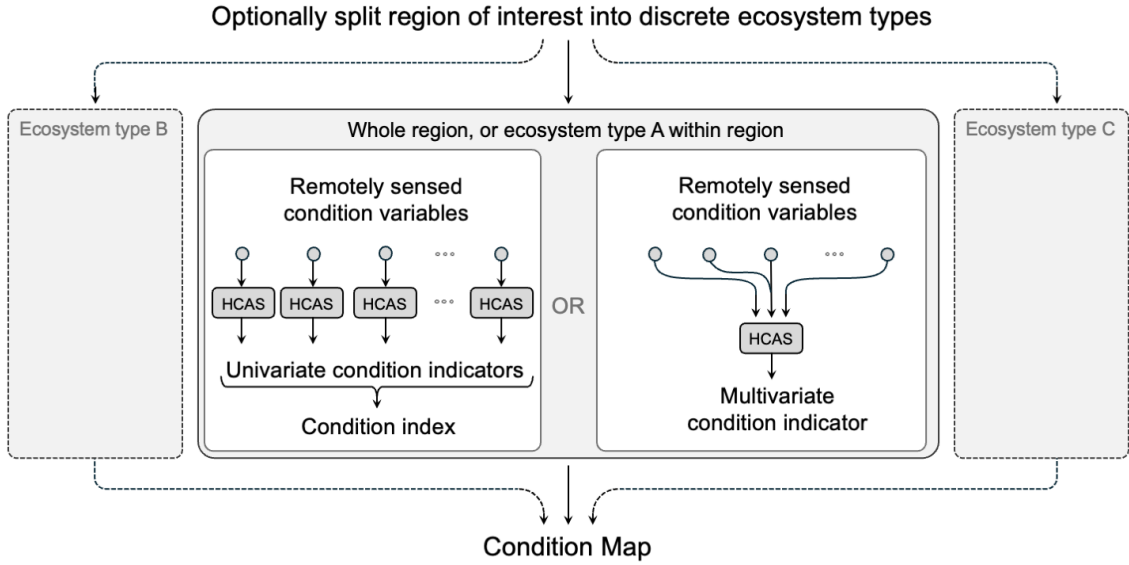


Figure 2: Flexible analytical pathways. The HCAS workflow can be applied across whole regions or stratified by ecosystem types, with condition variables modelled individually or jointly. This modularity enables adaptation to different data and contexts while maintaining a consistent benchmarking framework.

The following sections describe the ecological challenges that motivate the workflow (Section 2.1), the data required to implement it (Section 2.2), and the analytical architecture through which ecosystem condition is inferred. Computational procedures underpinning each stage are described in Section 3.

2.1 Ecological application challenges confronting ecosystem condition estimation using Earth observation data

HCAS estimates ecosystem condition by evaluating how closely the observed characteristics of an ecosystem align with those of ecologically comparable high-integrity reference ecosystems. Earth observation data provide measurements of ecosystem characteristics, representing aspects of ecosystem structure, function and composition, which are translated into condition variables (sensu [Keith *et al.* 2020](#)). Ecosystem condition is inferred within a reference-based benchmarking framework rather than measured directly. Implementing this framework requires addressing several interdependent ecological challenges ([Harwood *et al.* 2016](#)).

Ecosystem condition is inherently dependent on ecological context: similar observed ecosystem characteristics may indicate different levels of integrity in different environmental settings. For example, vegetation cover or productivity that is characteristic of a high-integrity ecosystem in a dry upland environment may indicate reduced condition in a more productive valley-bottom environment. Conversely, high-integrity ecosystems can exhibit a range of characteristics within comparable environmental settings, including those associated with alternative ecological states ([Beisner *et al.* 2003](#)). Condition assessments must therefore accommodate the range of characteristics exhibited by high-integrity ecosystems in an ecologically comparable setting. In addition, ecosystem characteristics vary through time in response to natural processes such as seasonal cycles, disturbance and recovery, requiring condition assessments to distinguish transient ecological dynamics from persistent ecological change. Finally, high-integrity reference ecosystems are often unevenly distributed or scarce across environmental space, creating challenges for the selection of representative benchmarks and increasing the risk of bias if limitations in reference data are not explicitly managed.

These ecological application challenges have shaped the structure of the HCAS workflow (Figure 1). Ecological context is addressed through reference ecosystem modelling, which models how remotely sensed characteristics of reference ecosystems are expected to vary as a function of spatial variation in environmental covariates (climate, soil, terrain) thereby informing the selection, and weighting, of reference ecosystems used to assess condition at any given location of interest. Alternative high-integrity ecosystem expressions and natural temporal variability are accommodated by benchmarking observations against multiple reference ecosystems within comparable environments, allowing condition to be evaluated relative to the range of characteristics exhibited by these ecosystems over both space and time. Limitations in the availability and representativeness of reference ecosystems are addressed by weighting benchmark reference sites according to their ecological comparability with the target site, ensuring that the most relevant available reference sites exert the greatest influence on condition estimates. Calibration then transforms benchmarked estimates onto a consistent scale that is interpretable and comparable across space and time.

2.2 Required input data

The HCAS workflow is underpinned by three core input data components that define the empirical basis for ecosystem condition estimation: (i) reference ecosystem and highly modified sites, (ii) remotely sensed ecosystem condition variables, and (iii) abiotic environmental covariates.

These components provide complementary information required for reference-based inference. Remotely sensed ecosystem condition variables describe observed ecosystem characteristics; environmental covariates define the abiotic settings in which these characteristics are expressed; and reference and highly modified sites provide the empirical bounds for interpreting ecosystem condition along a continuous scale.

2.2.1 Reference ecosystem sites and highly modified sites

Reference ecosystem sites are contemporary locations that approximate the dynamics of high-integrity ecosystem states across the environmental heterogeneity represented within the study domain. They provide the empirical basis for assessing ecosystem condition relative to high-integrity reference ecosystems and represent the upper range of ecosystem condition within the framework. Identification of reference sites typically integrates multiple data sources, including land-use and land-cover information, remnant vegetation mapping, on-ground observations and expert ecological knowledge (see Section S2.2 for the continental Australia implementation).

Reference ecosystems are intended to represent the range of high-integrity states observable under contemporary environments. When combined with Earth observation and environmental data, reference ecosystem sites provide the empirical basis for interpreting ecosystem characteristics relative to the range expressed by high-integrity ecosystems (Harwood *et al.* 2016).

Two complementary sets of reference sites are used within the workflow. Training reference sites are selected to represent the full range of environmental gradients, enabling robust prediction of ecosystem characteristics across the study domain. Benchmark reference sites are selected to represent variation in both environmental gradients and observed ecosystem characteristics, and provide the basis for benchmarking against multiple plausible reference states. The representativeness and distribution of reference sites strongly influence both reference ecosystem modelling and condition benchmarking.

Highly modified sites represent ecosystems substantially altered by sustained human activities such as intensive agriculture or urbanisation. These sites are used to anchor the lower end of the calibrated condition scale, support model evaluation, and improve interpretability of outputs.

2.2.2 Remotely sensed ecosystem condition variables

Remotely sensed ecosystem condition variables are proxies for ecosystem characteristics derived from Earth observation (EO) data, providing consistent measures of ecosystem structure, composition, and function across space and time (Graetz 1990; Czúcz *et al.* 2021a; Senf 2022). These variables form the observable component of ecosystem condition estimation within the

workflow.

Variables are selected to capture multiple dimensions of ecosystem characteristics while remaining ecologically meaningful, sensitive to ecosystem modification, robust to noise and data gaps, and consistently derivable at large spatial scales (Levick *et al.* 2025; Álvarez-Martínez *et al.* 2026). EO-derived variables may include spectral, structural, and functional metrics describing, for example, vegetation cover, productivity, phenology, and disturbance-response dynamics (Kooistra *et al.* 2024).

Long, temporally consistent EO time series are essential for distinguishing transient ecological dynamics from persistent ecosystem change. Multi-decadal satellite archives such as Landsat (Verbesselt *et al.* 2010; Kennedy *et al.* 2014; Hemati *et al.* 2021) provide the temporal depth required for this purpose. Although the HCAS framework is conceptually independent of any particular sensor, its application depends on the availability of EO data with sufficient temporal consistency and spatial coverage (Bathiany *et al.* 2025).

Two sets of ecosystem condition variables are derived from the same EO source data. One set represents long-term ecosystem characteristics by summarising observations across multi-decadal time series, while the other applies the same statistical summaries to shorter temporal periods to define assessment epochs. The long-term variables establish the condition inference process, while the short-term variables provide the basis for time-series condition estimation (Section 3.3). The Landsat-derived variables, long- and short-term aggregations, data-gap treatment and *z*-score standardisation used for continental Australia are described in Section S2.3.

2.2.3 Abiotic environmental covariates

Abiotic environmental covariates describe relatively stable gradients, such as climate, soils, landform, and hydrology, that influence the distribution and diversity of reference ecosystem characteristics over the region of interest. Their primary role is to characterise ecological context and support identification of ecologically comparable reference ecosystems.

Within the workflow, environmental covariates are used as predictors in the reference ecosystem model (Figure 1), enabling observed ecosystem characteristics to be interpreted relative to context-specific reference expectations. This reduces the risk that environmentally driven variation is misinterpreted as ecosystem modification and supports consistent comparison across heterogeneous landscapes.

Covariates are selected to capture major environmental gradients influencing reference ecosystem distribution and diversity while minimising direct representation of anthropogenic modification, thereby reducing circularity in condition inference. The environmental covariate selection process and final covariate sets used for continental Australia are described in Section S2.4 and Appendix A (in the Supporting Information). Environmental covariates also guide selection of reference site subsets for both model training and benchmarking, helping ensure adequate representation of environmental space (see Section S2.5 for the continental Australia sampling procedures).

2.2.4 Data preparation and alignment

Prior to analysis, remotely sensed ecosystem condition variables and environmental covariates are z -score standardised. The z -score statistics derived from the long-term remotely sensed variables are applied to all short-term variables so that both long-term and short-term variables can be represented within a consistent multivariate space during benchmarking. Standardising environmental covariates ensures stable reference ecosystem model fitting (Section 3.1).

All input datasets should be aligned within a common spatial and analytical framework prior to analysis. This includes harmonisation of coordinate reference systems, spatial resolution, and grid geometry, ensuring that reference and highly modified sites, remotely sensed ecosystem condition variables and environmental covariates correspond to the same spatial units of analysis. Appropriate quality control procedures should also be applied to ensure reliability of inputs, including treatment of missing data where appropriate. The analysis extent, grid resolution and spatial alignment used for continental Australia are documented in Section S2.1.

2.3 ClassicHCAS implementation architecture and workflow execution

Within the broader HCAS workflow, data preparation, reference ecosystem modelling, input-data selection and workflow orchestration are external to the **ClassicHCAS** R-package (Figure 1). This separation enables alternative approaches to these components while maintaining consistency in ecosystem condition inference through a common benchmarking framework.

The core computational functions of **ClassicHCAS**—construction of the inferential reference surface, condition benchmarking and calibration—are implemented in C++ via Rcpp ([Eddelbuettel & Francois 2011](#)) and parallelised using OpenMP ([Dagum & Menon 1998](#)), enabling efficient processing of large spatial datasets. The package supports both raster and tabular inputs, facilitating application across a range of spatial scales and use cases (Figure 2). This design allows individual workflow components to be configured, executed, and evaluated independently, supporting methodological flexibility and reproducibility.

Integration with geospatial processing frameworks such as the terra package ([Hijmans *et al.* 2026](#)) enables efficient handling of large raster datasets. In particular, out-of-memory processing allows data to be partitioned into manageable chunks rather than loaded entirely into memory, reducing computational constraints when working with high-resolution, continental-extent datasets.

Application of the HCAS workflow at large spatial scales requires execution strategies that extend beyond the core **ClassicHCAS** implementation. Large Earth observation datasets are typically partitioned into smaller spatial units (e.g. raster tiles) that can be processed independently, supporting execution within available memory and enabling parallel processing across multiple computing resources. The degree and configuration of parallelisation depend on the computational infrastructure within which the workflow is deployed. The tiled, distributed implementation used for continental Australia is described in Section S5.1.

Several workflow components may be computed once and subsequently reused across spatial subsets and temporal analyses. These include the reference ecosystem model (external to

ClassicHCAS), inferential reference surface, benchmarking parameters and calibration functions. Reuse of such invariant components reduces computational cost, promotes consistency among analyses, and enables **ClassicHCAS** to be embedded within a wide range of external processing pipelines as a transparent and repeatable foundation for ecosystem condition assessment. The complete workflow configuration and version-specific **ClassicHCAS** settings for the continental Australia implementation are summarised in Section S3.3.

3- CENTRAL ANALYTICAL WORKFLOW

This section describes the four analytical stages used within HCAS to infer ecosystem condition: reference ecosystem modelling, condition benchmarking, time-series condition estimation and calibration (Figure 1). Reference ecosystem modelling is undertaken externally to **ClassicHCAS**, whereas reference-surface construction, condition benchmarking and calibration are implemented within the package; time-series estimation applies the established long-term configuration to epoch-specific observations of ecosystem characteristics.

3.1 Reference ecosystem modelling

Reference ecosystem modelling is conducted externally, and the resulting spatial predictions applied to every location in the study region are provided as inputs to **ClassicHCAS** (Figure 1). The objective is to characterise relationships between abiotic environmental covariates and the summarised, long-term remotely sensed ecosystem condition variables using representative training observations from high-integrity reference ecosystems. These modelled relationships are used to predict the remote sensing distance expected between any given pair of sites—either between two training reference sites, or between a target location and a benchmark reference site—if both those sites were in reference condition. These predictions play three key roles when addressing ecological context in the benchmarking part of the HCAS workflow (Section 3.2). They contribute to construction of the inferential reference surface (Section 3.2.1), support identification of ecologically comparable benchmark reference sites for each target location, and provide the basis for weighting selected benchmark reference sites according to their expected ecological differences from the target location (Section 3.2.3).

The HCAS workflow is not tied to a specific modelling approach. Example implementations have employed flexible methods such as Generalised Additive Models (GAMs; via the **mgcv** R package; [Wood 2011](#)), which allow smooth responses to environmental predictors and can provide estimates of model uncertainty ([Pedersen *et al.* 2019](#)). The chosen modelling framework should accommodate large spatial datasets and complex, non-linear relationships between ecosystem characteristics and environmental gradients (see Section S2.6 for the continental Australia implementation).

Models are fitted using training data from reference ecosystem sites (Section 2.2.1). A separate model is fitted for each long-term remotely sensed ecosystem condition variable, enabling flexible representation of different ecological responses to environmental gradients, and supporting both

univariate and multivariate implementations of the workflow (Figure 2).

3.2 Condition benchmarking

Condition benchmarking is the core analytical component of the HCAS workflow and is implemented within **ClassicHCAS**. Ecosystem condition is inferred by comparing the observed characteristics of a target location with those of multiple ecologically comparable high-integrity reference ecosystems (Figure 3). Rather than benchmarking observations against a single expected value for each remotely sensed ecosystem characteristic, HCAS evaluates condition relative to the range of remotely sensed characteristics exhibited by comparable reference ecosystems. This enables ecosystem condition to be inferred while accommodating ecological context, alternative high-integrity ecosystem expressions and natural temporal variability.

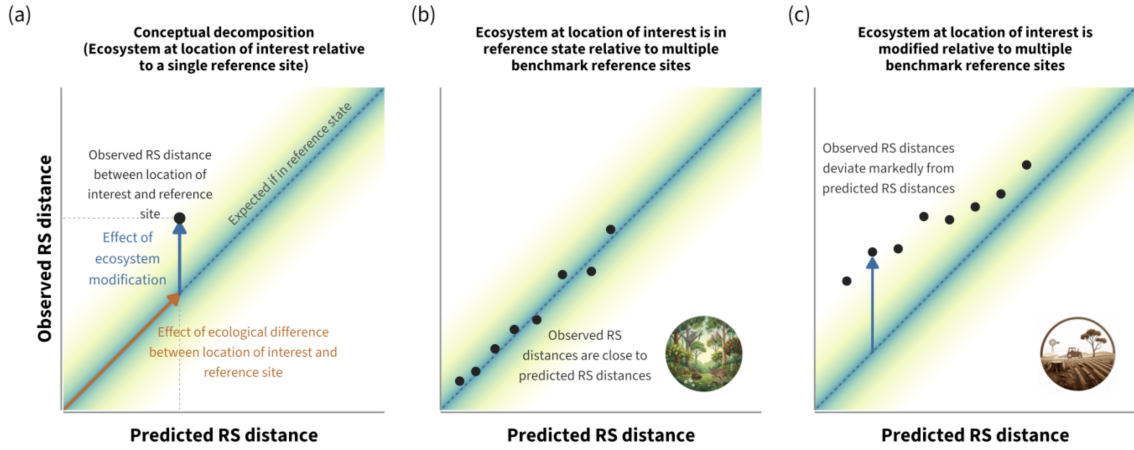


Figure 3: Conceptual interpretation of observed versus predicted remotely sensed (RS) distances in HCAS condition benchmarking. (a) Predicted RS distance is the difference in remotely sensed characteristics expected if both sites were in reference states, given their environmental differences. Departure of the observed RS distance from this prediction indicates ecosystem modification. (b) For ecosystem in reference state, observed RS distances remain close to their predicted distances from benchmark reference sites. (c) For modified ecosystems, observed RS distances depart from predictions, with greater departures indicating greater modification. Shading represents the expected reference relationship.

The inferential reference surface that supports this comparison is described in Section 3.2.1, followed by the procedure used to select benchmark reference sites and estimate condition for individual locations (Section 3.2.2).

3.2.1 Establishing the inferential reference surface

The inferential reference surface is constructed using observed remotely sensed ecosystem condition variables from training reference sites and corresponding predictions from the reference ecosystem model. Observed and predicted multivariate distances are calculated between pairs of training reference sites to characterise the remotely sensed empirical relationships among high-integrity reference ecosystems.

Let p_i and p_j denote two reference sites, with $\mathbf{O}_i$, $\mathbf{O}_j$, and $\mathbf{P}_i$, $\mathbf{P}_j$, their vectors of observed and predicted ecosystem condition variables respectively. The predicted and observed distances between the sites are:

$$d_{ij}^{\text{pred}} = \|\mathbf{P}_i - \mathbf{P}_j\|_1 \text{ as Eq. 1, and}$$

$$d_{ij}^{\text{obs}} = \|\mathbf{O}_i - \mathbf{O}_j\|_1 \text{ as Eq. 2.}$$

where $\|\bullet\|_1$ denotes the L_1 norm, or Manhattan distance. These distances are calculated for every unique pair of training reference sites within a specified geographic radius.

The resulting pairs $(d_{ij}^{\text{pred}}, d_{ij}^{\text{obs}})$ are processed to form a two-dimensional empirical probability density surface, $H(d^{\text{pred}}, d^{\text{obs}})$, representing the frequency distribution of predicted and observed distances among representative reference ecosystems. The resolution of this surface is controlled by two parameters: bin width and bin number. The surface is then trimmed to remove poorly supported regions, smoothed to fill gaps and reduce edge effects, and normalised with respect to predicted distance to ensure comparability across analyses, resulting in the *inferential reference surface*. The corresponding construction steps and **ClassicHCAS** settings for the continental Australia implementation are documented in Section S3.1.1.

3.2.2 Selecting benchmark reference sites

Benchmarking begins by identifying a candidate pool of benchmark reference sites within a specified geographic search radius of a target location. Ecological comparability is evaluated subsequently using predicted target–reference distances.

We will denote benchmarks that are known to be in reference condition as p_i for some index i , while sites with unspecified condition will be denoted as q_j . In particular, a site that is being assessed by the method (“the target location”) will be denoted as q_t . Note that q_t may itself be a known benchmark, but that does not affect the treatment.

Let $\mathbf{X}_t$, $\mathbf{X}_i$ denote the spatial coordinates of q_t and p_i , and $R_0 > 0$ be a specified spatial search radius (e.g. $R_0 = 200 \text{ km}$). Candidate reference sites are first restricted to the set:

$$\mathcal{C}_t^0 = \{p_i : \|\mathbf{X}_i - \mathbf{X}_t\|_2 \leq R_0 \text{ and } p_i \neq q_t\} \text{ as Eq. 3,}$$

where $\|\bullet\|_2$ denotes Euclidean geographic distance. Note that q_t is explicitly excluded from being its own candidate reference site.

Let $\mathbf{P}_t$, $\mathbf{P}_i$ be the vectors of predicted remote sensing variables for sites q_t and p_i . For each candidate benchmark reference site $p_i \in \mathcal{C}_t^0$, the predicted remotely sensed distance from the target location is therefore:

$$d_{it}^{\text{pred}} = \|\mathbf{P}_i - \mathbf{P}_t\|_1 \text{ as Eq. 4.}$$

That is, d_{it}^{pred} is a target–reference pairwise distance in predicted reference ecosystem distance-space. This predicted distance establishes the level of ecological comparability between the target location and its available candidate benchmark reference sites.

Prior to candidate benchmark reference site selection, an optional spatial penalty can be applied, calculated as:

$$g_{it} = \|\hat{\mathbf{X}}_i - \hat{\mathbf{X}}_t\|_1 \text{ as Eq. 5,}$$

where $\hat{\mathbf{X}}_i$ and $\hat{\mathbf{X}}_t$ are the standardised geographic coordinates of p_i and q_t . This term (Eq. 5) complements Equation 6 by favouring geographically nearer benchmark reference site candidates, where available.

Candidate benchmark reference sites are ranked using the combined geographic selection parameters (from Eqs. 4 and 5):

$$a_{it} = d_{it}^{\text{pred}} + \mu g_{it} \text{ as Eq. 6,}$$

where $\mu \geq 0$ controls the influence of spatial separation. Higher values of μ penalise the selection of geographically distant reference sites. For a selected integer $k_1 > 0$, let T_{k_1} be the k_1^{th} smallest value of a_{it} . Then

$$\mathcal{C}_t^1 = \{p_i \in \mathcal{C}_t^0 : a_{it} \leq T_{k_1}\} \text{ as Eq. 7}$$

is the set of k_1 sites with the smallest values of a_{it} (for example $k_1 = 50$). This second stage restricts candidate benchmark reference sites to the most geographically proximal and ecologically comparable reference ecosystems relative to the target location.

For each retained candidate benchmark reference site in $\mathcal{C}_t^1$, the observed remotely sensed distance to its target location is:

$$d_{it}^{\text{obs}} = \|\mathbf{O}_i - \mathbf{O}_t\|_1 \text{ as Eq. 8,}$$

where $\mathbf{O}_t$, $\mathbf{O}_i$ is the observed remotely sensed variables at q_t and p_i .

Using d_{it}^{obs} and d_{it}^{pred} (from Eqs. 4 and 8) for target location q_t and candidate benchmark p_i , an estimate of the likelihood that q_t is drawn from the global population of sites in reference condition is assigned from the *inferential reference surface* using $H(d_{it}^{\text{pred}}, d_{it}^{\text{obs}})$. For each candidate benchmark reference site in $\mathcal{C}_t^1$, this measures how well the predicted remotely sensed distance from the target location matches the observed remotely sensed distance under the assumption that the target location represents a site in reference condition.

Next, sort and relabel the benchmark sites in $\mathcal{C}_t^1$ so that they are ranked by decreasing $H(d_{it}^{\text{pred}}, d_{it}^{\text{obs}})$, and so that now, $H(d_{1t}^{\text{pred}}, d_{1t}^{\text{obs}}) \geq H(d_{2t}^{\text{pred}}, d_{2t}^{\text{obs}}) \geq \dots \geq H(d_{k_1t}^{\text{pred}}, d_{k_1t}^{\text{obs}})$.

For a chosen k_2 with $0 < k_2 \leq k_1$, let:

$$\mathcal{C}_t^2 = \{p_1, p_2, \dots, p_{k_2}\} \text{ as Eq. 9}$$

That is, $\mathcal{C}_t^2$ is the set of k_2 candidate sites from $\mathcal{C}_t^1$ with the highest likelihood that the target location, q_t , is drawn from the global population of sites in reference condition (for example $k_2 = 10$).

The above describes the nested sets $\mathcal{C}_t^2 \subseteq \mathcal{C}_t^1 \subseteq \mathcal{C}_t^0$, representing the stages of benchmark selection, with $\mathcal{C}_t^2$ the final set of candidates to be used for benchmarking q_t .

3.2.3 Assigning pre-calibration condition scores

The likelihoods, $H(d_{it}^{\text{pred}}, d_{it}^{\text{obs}})$, from benchmark reference sites in $\mathcal{C}_t^2$ are weighted according to their predicted remotely sensed distance to the target location using a Gaussian kernel. For a benchmark reference site p_j , with predicted distance to the target location d_{jt}^{pred} and selected scale parameter $\lambda > 0$, the unnormalised kernel weight is:

$$\tilde{w}_{jt} = \exp \left[- \left(\frac{d_{jt}^{\text{pred}}}{\lambda} \right)^2 \right] \text{ as Eq. 10.}$$

The scale parameter λ controls how rapidly the benchmark reference site influence declines with decreasing ecological comparability to the target location.

This weighting scheme allows all selected k_2 benchmark reference sites to contribute to condition estimation while ensuring that ecologically more comparable reference ecosystems exert greater influence on the final condition estimate for the target location. **ClassicHCAS** also supports a Cauchy kernel as an alternative weighting function.

The benchmark reference site p_1 (with the largest likelihood $H(d_{1t}^{\text{pred}}, d_{1t}^{\text{obs}})$ from Eq. 9) is given additional influence through a boost parameter $\beta \geq 1$. This prevents the strongest empirically-supported target-reference analogue from being diluted by averaging across less comparable k_2 benchmark reference sites to target location, particularly where multiple plausible reference ecosystem expressions are present. The adjusted weights are defined as:

$$w_{jt} = \begin{cases} \beta \tilde{w}_{jt}, & j = 1, \\ \tilde{w}_{jt}, & j > 1. \end{cases} \text{ as Eq. 11.}$$

Values of $\beta > 1$ increase the influence of the highest reference-density benchmark, whereas $\beta = 1$ gives the ordinary kernel-weighted mean. By default, $\beta = k_2$. The weights of remaining $k_2 - 1$ benchmark reference sites are unchanged.

The weights, after boost adjustment, are then normalised so that their relative contributions sum to one:

$$\omega_{jt} = \frac{w_{jt}}{\sum_{\ell=1}^{k_2} w_{\ell t}} \text{ as Eq. 12.}$$

The pre-calibration condition index for target location q_t is the weighted mean of inferential reference surface values across its k_2 benchmark reference sites, calculated as:

$$\mathbf{h}(q_t) = \sum_{j=1}^{k_2} \omega_{jt} H(d_{jt}^{\text{pred}}, d_{jt}^{\text{obs}}) \text{ as Eq. 13.}$$

Larger values of $\mathbf{h}$ indicate stronger alignment between the observed ecosystem characteristics of the target location and the distribution of characteristics exhibited by comparable high-integrity reference ecosystems. The continental Australia implementation and software-version-specific settings for this benchmarking procedure are documented in Section S3.1.2.

The parameters settings for R_0 (Eq. 3), μ (Eq. 6), $k_1 = \text{size}(\mathcal{C}_t^1)$ (Eq. 7), $k_2 = \text{size}(\mathcal{C}_t^2)$ (Eq. 9), λ (Eq. 10), and β (Eq. 11) are chosen through the parameter-selection process described in Section 3.2.4.

3.2.4 Choosing benchmarking parameters

Parameter choices in the benchmarking stage influence how benchmark reference sites are selected, weighted and represented for each target location. Key parameters include the spatial search radius, the number of candidate and retained reference sites, and the weighting function used in condition estimation. Selection of these parameters is external to **ClassicHCAS** because it depends on the available data and application context.

In the absence of independent ecosystem condition data (e.g. ecological field surveys), parameter selection can be guided using mapped land-use classes representing increasing levels of ecosystem modification (e.g. [Hirsh-Pearson *et al.* 2022](#); [Zhou *et al.* 2025](#); [Venegas-Li *et al.* 2026](#)), together with reference and highly modified sites to approximate the upper and lower ends of the condition gradient (Section 2.2.1). Parameter tuning can be conducted using random or grid search approaches ([Bergstra & Bengio 2012](#)), with performance evaluated using the area under the receiver-operating characteristic curve (AUC; [Fawcett 2006](#)) based on the ability of the condition estimates (pre-calibration) to discriminate among classes along the modification gradient.

Under this approach, AUC is calculated for pairs of modification classes (e.g. reference versus moderately modified, moderately versus highly modified), and the optimal parameter set is selected by maximising mean discrimination across these pairs (see Section S3.1.3 for a worked example). The condition index is then assessed for its ability to preserve the expected ordering of ecosystem modification across these classes, with reference sites exhibiting the highest values and highly modified sites the lowest. The AUC quantifies the probability that a randomly

selected site from a less-modified class is assigned a higher condition value than a site from a more-modified class. To capture performance across the full modification gradient, AUC is calculated for adjacent class-pairs, and the optimal parameter set is defined as that which maximises mean sequential AUC across all pairs (see Section S3.1.3 for the continental Australia worked example).

3.3 Time-series condition estimation

Time-series condition estimates are derived by replacing the long-term observed remotely sensed ecosystem condition variables used in benchmarking with equivalent variables summarised over shorter temporal periods (e.g. annual or multi-year epochs), while retaining the long-term reference ecosystem model and inferential reference surface. Benchmarking is then applied as described in Section 3.2 for each temporal epoch, using the inferential reference surface and benchmarking parameters established for the long-term epoch. This results in epoch-specific pre-calibration condition estimates that can be interpreted consistently through time (see Section S3.4 for the continental Australia temporal implementation).

This approach ensures that temporal variation in condition estimates can be interpreted relative to a consistent long-term reference framework rather than artefacts arising from changing model expectations.

3.4 Calibration

Calibration in `ClassicHCAS` uses a monotonic Hermite spline (Fritsch & Carlson 1980; Dougherty *et al.* 1989), implemented with `splinefun` in the R `stats` package (R Core Team 2026), to transform the pre-calibration condition score into a standardised and interpretable measure of ecosystem condition on a continuous 0–1 scale. This non-parametric transformation preserves the relative ordering of sites. Calibrated values approaching 1 correspond to ecosystems consistent with high-integrity reference condition, and those approaching 0 correspond to highly modified ecosystems. Because pre-calibration values are aligned with ecologically comparable high-integrity reference ecosystems, they differ in range and distribution with variation in input data and model configurations. Therefore, calibration supports comparability among implementations.

Calibration coordinates are determined external to `ClassicHCAS`. The calibration coordinates used for continental Australia, including the provenance of the upper and lower internal coordinates, are documented in Section S3.2.

To describe the calibration process, let $\mathcal{R}$ be the set of pixels in reference condition and let $\mathcal{H}$ be the set of highly modified pixels. Recalling that $\mathbf{h}$ (Eq. 13) is the function that maps a point in the study area to its pre-calibrated score, let $H_{\mathcal{R}} = \{\mathbf{h}(q) : q \in \mathcal{R}\}$ be the set of all pre-calibrated scores over the set $\mathcal{R}$ and, similarly, $H_{\mathcal{H}} = \{\mathbf{h}(q) : q \in \mathcal{H}\}$ be the set of pre-calibrated scores over the set $\mathcal{H}$.

A corresponding ecosystem condition value on the 0–1 scale is assigned that represents an empirically-derived median value or expert estimate for each of these sets, say, $C_{\mathcal{R}}$ and $C_{\mathcal{H}}$.

Denote the median values of the pre-calibrated scores over these regions as $\text{med}(H_{\mathcal{R}})$ and $\text{med}(H_{\mathcal{H}})$. The calibration curve, α is a monotonic Hermite spline fitted to selected knots. Since α is monotonic, it will preserve the rank ordering of the pre-calibration scores and therefore preserves their median values. By including the two knot points $(\text{med}(H_{\mathcal{R}}), C_{\mathcal{R}})$ and $(\text{med}(H_{\mathcal{H}}), C_{\mathcal{H}})$ we can be assured that:

$$\text{med}(\alpha(H_{\mathcal{R}})) = \alpha(\text{med}(H_{\mathcal{R}})) = C_{\mathcal{R}} \text{ as \textbf{Eq. 14}}$$

and

$$\text{med}(\alpha(H_{\mathcal{H}})) = \alpha(\text{med}(H_{\mathcal{H}})) = C_{\mathcal{H}} \text{ as \textbf{Eq. 15.}}$$

The median values of reference and highly modified areas will therefore match the empirical or expert-determined estimates after calibration. To anchor the calibration at the lower end, the knot point $(0, 0)$ is included. At the upper end, the maximum of the pre-calibration scores should be calibrated to 1, but cannot be the sole tie point, because the maximum is sensitive to the exact choice of k_2 benchmark reference sites.

Let $h_{\max}$ be the maximum pre-calibrated score over the entire study area, and h_q be a very large quantile of the pre-calibrated score (for example the 99.99th percentile, with $q = 0.9999$). The two additional knots (h_q, q) and $(h_{\max}, 1)$ are included in the calibration curve to provide a robust transformation from pre-calibration scores to the interval $[0, 1]$. Table 1 provides a summary of the knot points used in the calibration curve and their purpose.

Table 1: Summary of the spline knots used in the calibration curve.

Pre-calibrated value	Calibrated value	Description
0	0	Lower anchor
$\text{med}(H_{\mathcal{H}})$	$C_{\mathcal{H}}$	Expert-assigned median highly modified condition
$\text{med}(H_{\mathcal{R}})$	$C_{\mathcal{R}}$	Expert-assigned median reference condition
h_q	0.9999	Robust upper anchor
$h_{\max}$	1	Upper anchor

Calibration coordinates derived from the long-term epoch used to establish the parameters for condition benchmarking are applied consistently across all long- and short-term epochs. To maintain temporal consistency, where required, gain scaling can be applied to short-term epoch pre-calibration condition estimates before calibration (see Section S3.4).

5- DISCUSSION

The HCAS workflow provides a scalable and reproducible implementation of a reference-based approach to ecosystem condition assessment from Earth observation (EO) data. By formalising the HCAS framework as a modular analytical workflow, the approach provides a transparent

and flexible foundation for large-scale ecosystem condition assessment. The continental-scale Australian application presented in the Supporting Information demonstrates that these concepts can be operationalised using multi-decadal EO datasets to generate spatially explicit and temporally consistent condition estimates.

Supporting evidence for the continental Australia application includes evaluation of the inferred reference-site dataset (Section S4.1) and comparison with an independently developed regional condition product (Section S4.2).

The HCAS workflow addresses several longstanding challenges associated with inferring ecosystem condition from EO data, including those relating to ecological context, alternative high-integrity ecosystem expressions, and benchmarking against ecologically comparable reference ecosystems. However, interpretation of condition estimates should still take into consideration ecological, implementation and data-related limitations, as well as uncertainty arising throughout the workflow. These considerations, and opportunities for future methodological development, are discussed below.

5.1 Limitations of the HCAS approach

Several limitations arise from the inherent challenges of defining and inferring ecosystem condition using remotely sensed data. As outlined in Section 2.1, ecosystem condition is context-dependent and must be interpreted relative to ecologically comparable high-integrity ecosystems. Although HCAS addresses ecological context through reference ecosystem modelling and benchmarking, no model can fully represent all abiotic and biotic drivers influencing ecosystem characteristics. Therefore, ecological context may be incompletely represented, particularly in complex or highly heterogeneous landscapes.

High-integrity ecosystems can also exhibit multiple valid expressions of structure, composition, and function within similar environmental settings. HCAS accommodates this challenge by benchmarking observations against multiple ecologically comparable reference ecosystems rather than a single reference state. Nevertheless, the range of high-integrity ecosystem expressions represented in any assessment is constrained by the available reference ecosystem sample and the ecosystem characteristics captured by the selected Earth observation variables. As a result, some valid reference ecosystem expressions may remain underrepresented.

Natural temporal variability presents a further challenge. Ecosystem characteristics vary through time in response to processes such as seasonal cycles, disturbance and recovery. While HCAS seeks to distinguish transient ecological dynamics from persistent ecological change, complete separation of these processes is inherently difficult. Temporal variation in condition estimates may therefore reflect some combination of natural variability and sustained ecological change, particularly where long-term reference information is limited.

The availability and representativeness of high-integrity reference ecosystems also constrain ecosystem condition inference. Reference ecosystems are often unevenly distributed across environmental space, and some environmental settings may have few suitable reference ecosystems. Although HCAS accommodates this challenge through benchmarking and ecological-

comparability weighting, uncertainty increases where benchmark support is limited or where available reference ecosystems incompletely represent the range of high-integrity ecosystem expressions.

More broadly, attribution of observed variation to ecosystem modification remains uncertain. Remotely sensed ecosystem characteristics integrate responses to both natural processes and human influences, and similar ecosystem responses may arise from multiple interacting drivers. Ecosystem condition estimates should therefore be interpreted as measures of alignment with the characteristics exhibited by ecologically comparable high-integrity reference ecosystems, rather than as direct evidence of specific ecological mechanisms or causes of change.

These challenges are not confined to HCAS alone but are instead shared by all efforts to assess ecosystem condition using Earth observation data. Condition estimates should therefore be interpreted as relative measures of ecosystem condition rather than absolute measures of ecosystem state.

5.2 Limitations of the ClassicHCAS implementation

Several limitations arise from the current implementation of **ClassicHCAS**. Benchmarking is based primarily on high-integrity reference ecosystems, with highly modified sites used to anchor the calibrated condition scale. While this ensures that condition is defined relative to ecologically meaningful reference states, it does not explicitly incorporate the full gradient of ecosystem modification during condition inference, which may limit discrimination among intermediate condition states. Yet [Harwood *et al.* \(2016\)](#) originally described the broader HCAS framework as integrally using both high-integrity reference ecosystems and highly modified ecosystems.

Calibration introduces an additional implementation dependency. The calibrated condition scale relies on assumptions regarding the ecological interpretation of reference and highly modified sites, and these assumptions may differ among regions. In landscapes where few high-integrity reference ecosystems remain, contemporary reference ecosystems may fall below the theoretical reference state, requiring expert-informed adjustment of upper calibration bounds.

Condition estimates are also influenced by the representativeness and density of reference site data. Under-representation of particular environmental settings or ecosystem expressions may reduce the performance of both reference ecosystem modelling and benchmarking. The continental Australia implementation incorporated model-based uncertainty quantification using bootstrap resampling of reference sites to generate confidence intervals around condition estimates (Section S3.5).

Parameterisation of the benchmarking process introduces an additional source of variability. Choices relating to neighbourhood definition, number of candidate and benchmark reference sites, weighting functions and associated tuning parameters influence condition estimates and may affect comparability among applications if not consistently selected and documented. Benchmarking parameter selection for the continental Australia implementation is described in Section S3.1.3.

The current implementation assumes that reference ecosystems retain high levels of integrity and

fluctuate within the range represented by the long-term reference framework. While this supports consistency in interpretation, it does not explicitly account for temporal variability in reference ecosystems. Reference ecosystems themselves may be subject to persistent environmental pressures (e.g. [Bergstrom *et al.* 2021](#); [Wu *et al.* 2026](#)), and such changes may not be represented within the benchmarking framework.

These limitations are specific to the current implementation of **ClassicHCAS**, rather than to the HCAS framework itself, and represent areas for continued methodological development and evaluation.

5.3 Limitations of Earth observation and environmental data

HCAS workflow implementations are necessarily constrained by the characteristics and quality of both Earth observation (EO) data and environmental inputs. Remotely sensed ecosystem condition variables provide an indirect representation of ecosystem state, capturing surface and near-surface properties but not local or fine-scale attributes of ecosystem integrity, such as species composition, understorey structure, soil condition, or faunal interactions. Condition estimates should therefore be interpreted as proxy-based indicators rather than comprehensive measures of ecosystem state.

Interpretation of EO-derived variables is also scale-dependent. Ecosystem condition is an emergent property that is most meaningful at spatial scales integrating ecosystem structure, function and composition ([Parkes *et al.* 2003](#); [Eyre *et al.* 2011](#)). At finer resolutions, EO variables may reflect local heterogeneity rather than system-level properties, affecting interpretability.

Temporal consistency of EO data is critical for reliable condition estimation. Data gaps, uneven observation density and persistent cloud cover can introduce bias, particularly in short-term condition estimates. Similarly, the quality and completeness of environmental covariates influence the performance of reference ecosystem modelling and the extent to which ecological context can be represented in condition assessments.

5.4 Quantifying uncertainty in condition estimation

Uncertainty in ecosystem condition estimates arises from multiple sources, including reference site selection, reference ecosystem modelling, benchmarking procedures and input data quality. Uncertainty is not uniform across space or time, with higher uncertainty generally expected in ecologically heterogeneous regions and areas with limited benchmark support. Model uncertainty propagates through benchmarking, while variability in the selection and weighting of benchmark reference ecosystems influences local condition estimates. Additional uncertainty may arise when estimating condition from short-term observations, particularly where Earth observation data quality is variable.

Quantitative approaches such as bootstrap resampling can be used to characterise uncertainty ([Lehmann *et al.* 2025](#)). The continental Australia implementation quantifies the model-based uncertainty in the estimated ecosystem condition due to reference site resampling using the

bootstrap analysis described in Section S3.5. However, interpretation of condition estimates should consider the combined effects of multiple sources of uncertainty (Williams *et al.* 2025).

5.5 Opportunities for future development

Several avenues exist to extend both the HCAS framework and its implementation in **ClassicHCAS**. Potential methodological developments include improved representation of the full condition gradient, enhanced treatment of alternative ecological states, and exploration of alternative benchmarking formulations and distance metrics. For example, a future temporal-benchmarking extension could compare contemporary target observations with the historical dynamics of selected contemporary high-integrity benchmark ecosystems over a defined baseline period, testing whether target location change is consistent with or departs from the temporal variation observed at those benchmarks. Future implementations may also revisit the incorporation of both high-integrity reference ecosystems and highly modified ecosystems within condition inference, consistent with the broader framework proposed by Harwood *et al.* (2016).

Data-related improvements include expansion of reference ecosystem datasets, integration of multiple Earth observation sensor sources to improve temporal coverage in cloud-dominated regions, and incorporation of complementary data sources such as field observations to support empirical calibration and accuracy assessment. Improved environmental covariates may further enhance representation of ecological context.

From a computational perspective, development of standardised parameterisation protocols, diagnostic-visualisation tools and workflow orchestration would support more repeatable and scalable application across regions and platforms.

6- CONCLUSIONS

The **ClassicHCAS** R package was developed to support reproducible and scalable HCAS workflows for ecosystem condition assessment using Earth observation data. By formalising the relationships between ecological context, reference ecosystems and observed ecosystem characteristics within a modular workflow, the approach enables consistent inference of ecosystem condition across space and time.

A key contribution of this work is the explicit separation of ecological application challenges and computational implementation. The workflow translates the core elements of HCAS—including those addressing ecological context, variation among high-integrity reference ecosystems, benchmarking, time-series estimation and calibration—into a transparent and flexible analytical system. This modular design enables alternative data sources, reference ecosystem modelling approaches, calibration strategies and analytical configurations to be implemented within a consistent benchmarking framework.

The repeated continental-scale applications across Australia (Supporting Information; e.g. Valavi *et al.* 2025; Williams *et al.* 2025) demonstrate the capacity of **ClassicHCAS**, when embedded in an operational workflow, to generate spatially explicit and temporally consistent ecosystem

condition estimates from multi-decadal EO data. These outputs capture broad gradients of ecosystem modification and temporal dynamics relative to a consistent long-term reference framework, supporting applications in biodiversity assessment, restoration planning, ecosystem accounting and environmental reporting. Because the required input datasets are increasingly available globally, the workflow is readily adaptable to other regions and jurisdictions.

By providing an open, modular, and extensible workflow, with **ClassicHCAS** at its core, we establish a foundation for continued methodological development and more consistent large-scale ecosystem condition assessment. The flexibility of the workflow enables adaptation to diverse ecological contexts and applications while maintaining a common inferential framework, supporting future implementation in emerging initiatives such as ecosystem accounting under the System of Environmental-Economic Accounting – Ecosystem Accounting framework (United Nations *et al.* 2024) and other large-scale environmental assessment and reporting programs.

Code and data availability

The **ClassicHCAS** R package is openly available on GitHub at <https://github.com/csiro/ClassicHCAS>, with a stable release planned for CRAN upon publication.

Published data collections and associated reports and application examples are listed on the project website (<https://research.csiro.au/biodiversity-knowledge/projects/hcas/>).

Acknowledgements

The work presented here was primarily funded by the Australian Government Department of Climate Change, Energy, the Environment and Water, through an ongoing partnership with CSIRO. The Terrestrial Ecosystem Research Network (TERN), an Australian Government NCRIS-enabled project, part-funded the work on environmental covariates in the Supporting Information. We thank Mike Birchall for review and feedback on C++ coding, Dr Sam Nicol and Rebecca Bartley for their thoughtful feedback on an early draft, and Dr Glenn Newnham for enabling opportunities in HCAS development. We acknowledge multiple past contributors to HCAS, as listed in technical reports on the project webpage. We recognise the enduring connections that Torres Strait Islanders and Aboriginal peoples have to the lands and seas of Australia, and pay our respects to their elders past and present. AI-assisted drafting tools (Microsoft 365 Copilot) were used for editorial support. Figures 4, 5 and 16 in the Supporting Information were prepared with OpenAI image-generation tools under author supervision.

9- REFERENCES

Álvarez-Martínez, J.M., Lugonja, T.N., Valdés, A., González Le Barbier, J., Suárez, M.P., Romero, G.H., Radulović, M., Knežević, M., Tarčak, S., Brkljač, B., Bokić, B., Radak, B., Andrić, A., Marković, M., Brdar, S., Lugonja, P., Simović, I., Giagnacovo, L. & Jiménez-Alfaro, B. (2026) Four decades of remote sensing for monitoring terrestrial ecosystems:

- a global review and future challenges. *Science of Remote Sensing*, 13, 100341. <https://doi.org/10.1016/j.srs.2025.100341>.
- Anderson, C.B. (2018) Biodiversity monitoring, earth observations and the ecology of scale. *Ecology Letters*, 21, 1572–1585. <https://doi.org/10.1111/ele.13106>.
- Bathiany, S., Bastiaansen, R., Bastos, A., Blaschke, L., Lever, J., Loriani, S., De Keersmaecker, W., Dorigo, W., Milenković, M., Senf, C., Smith, T., Verbesselt, J. & Boers, N. (2025) Ecosystem resilience monitoring and early warning using Earth observation data: Challenges and outlook. *Surveys in Geophysics*, 46, 265–301. <https://doi.org/10.1007/s10712-024-09833-z>.
- Beisner, B.E., Haydon, D.T. & Cuddington, K. (2003) Alternative stable states in ecology. *Frontiers in Ecology and the Environment*, 1, 376–382. [https://doi.org/10.1890/1540-9295\(2003\)001\[0376:ASSIE\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2003)001[0376:ASSIE]2.0.CO;2).
- Bergstra, J. & Bengio, Y. (2012) Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305. <https://jmlr.org/papers/v13/bergstra12a.html>.
- Bergstrom, D.M., Wienecke, B.C., van den Hoff, J., Hughes, L., Lindenmayer, D.B., Ainsworth, T.D., Baker, C.M., Bland, L., Bowman, D.M.J.S., Brooks, S.T., Canadell, J.G., Constable, A.J., Dafforn, K.A., Depledge, M.H., Dickson, C.R., Duke, N.C., Helmstedt, K.J., Holz, A., Johnson, C.R., McGeoch, M.A., Melbourne-Thomas, J., Morgain, R., Nicholson, E., Prober, S.M., Raymond, B., Ritchie, E.G., Robinson, S.A., Ruthrof, K.X., Setterfield, S.A., Sgrò, C.M., Stark, J.S., Travers, T., Trebilco, R., Ward, D.F.L., Wardle, G.M., Williams, K.J., Zylstra, P.J. & Shaw, J.D. (2021) Combating ecosystem collapse from the tropics to the Antarctic. *Global Change Biology*, 27, 1692–1703. <https://doi.org/10.1111/gcb.15539>.
- Czúcz, B., Keith, H., Driver, A., Jackson, B., Nicholson, E. & Maes, J. (2021a) A common typology for ecosystem characteristics and ecosystem condition variables. *One Ecosystem*, 6, e58218. <https://doi.org/10.3897/oneeco.6.e58218>.
- Czúcz, B., Keith, H., Maes, J., Driver, A., Jackson, B., Nicholson, E., Kiss, M. & Obst, C. (2021b) Selection criteria for ecosystem condition indicators. *Ecological Indicators*, 133, 108376. <https://doi.org/10.1016/j.ecolind.2021.108376>.
- Dagum, L. & Menon, R. (1998) OpenMP: an industry standard API for shared-memory programming. *IEEE Computational Science and Engineering*, 5, 46–55. <https://doi.org/10.1109/99.660313>.
- Dougherty, R.L., Edelman, A. & Hyman, J.M. (1989) Positivity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation. *Mathematics of Computation*, 52, 471–494. <https://doi.org/10.1090/S0025-5718-1989-0962209-1>.
- Eddelbuettel, D. & Francois, R. (2011) Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software*, 40, 1 – 18. <https://doi.org/10.18637/jss.v040.i08>.
- Eyre, T.J., Fisher, A., Hunt, L.P. & Kutt, A.S. (2011) Measure it to better manage it: a biodiversity monitoring framework for the Australian rangelands. *The Rangeland Journal*, 33, 239–253. <https://doi.org/10.1071/RJ10071>.
- Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- Fritsch, F.N. & Carlson, R.E. (1980) Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17, 238–246. <https://doi.org/10.1137/0717021>.

- Giljohann, K.M., Mokany, K., Ferrier, S., Harwood, T.D., Ware, C. & Williams, K.J. (2025) Accounting for Australia's threatened species: estimating historical and recent change in terrestrial habitat Ecological Indicators, 170, 1–10. <https://doi.org/10.1016/j.ecolind.2024.112978>.
- Graetz, R.D. (1990) Remote sensing of terrestrial ecosystem structure: An ecologist's pragmatic view. Remote Sensing of Biosphere Functioning (eds R.J. Hobbs & H.A. Mooney), pp. 5–30. Springer New York. https://doi.org/10.1007/978-1-4612-3302-2_2.
- Harwood, T.D., Donohue, R.J., Williams, K.J., Ferrier, S., McVicar, T.R., Newell, G. & White, M. (2016) Habitat Condition Assessment System: A new way to assess the condition of natural habitats for terrestrial biodiversity across whole regions using remote sensing data. Methods in Ecology and Evolution, 7, 1050–1059. <https://doi.org/10.1111/2041-210X.12579>.
- Hemati, M., Hasanlou, M., Mahdianpari, M. & Mohammadimanesh, F. (2021) A Systematic Review of Landsat Data for Change Detection Applications: 50 Years of Monitoring the Earth. Remote Sensing, 13, 2869. <https://www.mdpi.com/2072-4292/13/15/2869>
- Hijmans, R.J., Brown, A., Barbosa, M., Dyba, K., Bivand, R., Chirico, M., Cordano, E., Pebesma, E., Rowlingson, B. & Sumner, M.D. (2026) Package 'terra': Spatial Data Analysis, version 1.9-27. CranR Project. <https://rspatial.github.io/terra/>.
- Hirsh-Pearson, K., Johnson, C.J., Schuster, R., Wheate, R.D. & Venter, O. (2022) Canada's human footprint reveals large intact areas juxtaposed against areas under immense anthropogenic pressure. FACETS, 7, 398–419. <https://doi.org/10.1139/facets-2021-0063>.
- Keith, H., Czúcz, B., Jackson, B., Driver, A., Nicholson, E. & Maes, J. (2020) A conceptual framework and practical structure for implementing ecosystem condition accounts. One Ecosystem, 5. <https://doi.org/10.3897/oneeco.5.e58216>.
- Kennedy, R.E., Andréfouët, S., Cohen, W.B., Gómez, C., Griffiths, P., Hais, M., Healey, S.P., Helmer, E.H., Hostert, P., Lyons, M.B., Meigs, G.W., Pflugmacher, D., Phinn, S.R., Powell, S.L., Scarth, P., Sen, S., Schroeder, T.A., Schneider, A., Sonnenschein, R., Vogelmann, J.E., Wulder, M.A. & Zhu, Z. (2014) Bringing an ecological view of change to Landsat-based remote sensing. Frontiers in Ecology and the Environment, 12, 339–346. <https://doi.org/10.1890/130066>.
- Kooistra, L., Berger, K., Brede, B., Graf, L.V., Aasen, H., Roujean, J.L., Machwitz, M., Schlerf, M., Atzberger, C., Prikaziuk, E., Ganeva, D., Tomelleri, E., Croft, H., Reyes Muñoz, P., Garcia Millan, V., Darvishzadeh, R., Koren, G., Herrmann, I., Rozenstein, O., Belda, S., Rautiainen, M., Rune Karlsen, S., Figueira Silva, C., Cerasoli, S., Pierre, J., Tanır Kayıkçı, E., Halabuk, A., Tunc Gormus, E., Fluit, F., Cai, Z., Kycko, M., Udelhoven, T. & Verrelst, J. (2024) Reviews and syntheses: Remotely sensed optical time series for monitoring vegetation productivity. Biogeosciences, 21, 473–511. <https://doi.org/10.5194/bg-21-473-2024>.
- Lehmann, E.A., Collings, S., Valavi, R. & Williams, K.J. (2025) Model-based uncertainty quantification for the Habitat Condition Assessment System (HCAS). Publication EP2025-2296. CSIRO. <https://doi.org/10.25919/7czb-an54>.
- Levick, S.R., Johnson, S., Williams, K.J., van Niel, T.G., Valavi, R., Paget, M., Lehmann, E.A., Collings, S. & Newnham, G. (2025) Remotely sensed ecosystem characteristics for condition estimation in the Habitat Condition Assessment System (HCAS): transitioning to Landsat.

- Technical Report EP2025-2350. CSIRO. <https://publications.csiro.au/publications/publication/PIcsi:EP2025-2350>.
- Lowndes, J.S.S., Best, B.D., Scarborough, C., Afflerbach, J.C., Frazier, M.R., O'Hara, C.C., Jiang, N. & Halpern, B.S. (2017) Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*, 1, 0160. <https://doi.org/10.1038/s41559-017-0160>.
- Luxton, S., Avila-Ortega, D.I. & Dunnett, S. (2026) Integrating biodiversity and business: Emerging methods driving the nature economy. *Methods in Ecology and Evolution*, 17, 1326–1335. <https://doi.org/10.1111/2041-210x.70307>.
- Mokany, K., Ware, C., Harwood, T.D., Schmidt, R.K., Tetreault-Campbell, S. & Ferrier, S. (2022) Habitat-based biodiversity assessment for ecosystem accounting in the Murray-Darling Basin. *Conservation Biology*, 36, e13915. <https://doi.org/10.1111/cobi.13915>.
- Murphy, H., Ferrier, S., Fitch, P., Hoskins, A., Lyons, P., Moran, C., Munroe, S., Richards, A., Bugnot, A., Butler, D., Duncan, P., Good, M., Hosack, G., Lemon, D., McEvoy, J., Prober, S., Saunders, M., Terasaki Hart, D. & Williams, K.J. (2025) The Ecological Knowledge System for the Nature Repair Market. Technical report EP2024-6312. CSIRO. <https://doi.org/10.25919/fyzz-kb05>.
- Parkes, D., Newell, G. & Cheal, D. (2003) Assessing the quality of native vegetation: The 'habitat hectares' approach. *Ecological Management & Restoration*, 4, S29–S38. <https://doi.org/10.1046/j.1442-8903.4.s.4.x>.
- Pedersen, E.J., Miller, D.L., Simpson, G.L. & Ross, N. (2019) Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*, 7, e6876. <https://doi.org/10.7717/peerj.6876>.
- Pettorelli, N., Schulte to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A.M., Keith, D.A., Wegmann, M., Schrod, F., Stellmes, M., Sonnenschein, R., Geller, G.N., Roy, S., Somers, B., Murray, N., Bland, L., Geijzenborffer, I., Kerr, J.T., Broszeit, S., Leitão, P.J., Duncan, C., El Serafy, G., He, K.S., Blanchard, J.L., Lucas, R., Mairota, P., Webb, T.J. & Nicholson, E. (2018) Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation*, 4, 71–93. <https://doi.org/10.1002/rse2.59>.
- R Core Team (2026) R documentation for splinefun in package stats: Interpolating Splines. R Foundation for Statistical Computing. Viewed 10 September 2026, <<https://search.r-project.org/R/refmans/stats/html/splinefun.html>>.
- Sandve, G.K., Nekrutenko, A., Taylor, J. & Hovig, E. (2013) Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9, e1003285. <https://doi.org/10.1371/journal.pcbi.1003285>.
- Senf, C. (2022) Seeing the system from above: The use and potential of remote sensing for studying ecosystem dynamics. *Ecosystems*, 25, 1719–1737. <https://doi.org/10.1007/s10021-022-00777-2>.
- United Nations, European Union, Food and Agriculture Organization of the United Nations, International Monetary Fund, Organisation for Economic Co-operation and Development & The International Bank for Reconstruction and Development/The World Bank (2024) System of Environmental-Economic Accounting – Ecosystem Accounting. United Nations

- Department of Economic and Social Affairs, Statistics Division. https://seea.un.org/sites/seea.un.org/files/documents/EA/seea_ea_f124_web_12dec24.pdf
- Ustin, S.L. & Middleton, E.M. (2021) Current and near-term advances in Earth observation for ecological applications. *Ecological Processes*, 10, 1. <https://doi.org/10.1186/s13717-020-00255-4>.
- Valavi, R., Levick, S.R., Lehmann, E.A., Liu, N., Giljohann, K.M., Williams, K.J., Collings, S., Johnson, S., Botha, E.J., Munroe, S.E.M., Van Niel, T.G., Newnham, G., Paget, M., Malley, C., Carlile, P., Gunawardana, D., Lyon, P., Richards, A.E., Tetreault Campbell, S. & Ferrier, S. (2025) HCAS 3.3 (1988-2024) base model estimate of habitat condition (90m grid), National Connectivity Index 2.0 (NCI) and annual time series for continental Australia. Data collection 65549. CSIRO. <https://doi.org/10.25919/hngh-v675>.
- Venegas-Li, R., Atkinson, S., de Andrade Junior, M.A.U., Fletcher, R., Owen, P., Morales-Barquero, L., Aska, B., Arias-Patino, M., Grantham, H.S., Possingham, H., Venter, O., Ward, M. & Watson, J.E.M. (2026) Australia's terrestrial industrial footprint and ecological intactness. *Earth Syst. Sci. Data*, 18, 2179–2201. <https://doi.org/10.5194/essd-18-2179-2026>.
- Verbesselt, J., Hyndman, R., Newnham, G. & Culvenor, D. (2010) Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114, 106–115. <https://doi.org/10.1016/j.rse.2009.08.014>.
- Ward, M., Ashman, K., Lindenmayer, D.B., Legge, S., Kindler, G., Cadman, T., Fletcher, R., Whiterod, N., Lintermans, M., Zylstra, P., Stewart, R., Thomas, H., Blanch, S. & Watson, J.E.M. (2024) Shifting baselines clarify the impact of contemporary logging on forest-dependent threatened species. *Conservation Science and Practice*, 6, e13185. <https://doi.org/10.1111/csp2.13185>.
- Williams, K.J., Harwood, T.D., Lehmann, E.A., Ware, C., Lyon, P., Bakar, S., Pinner, L., Schmidt, R.K., Mokany, K., Van Niel, T.G., Richards, A.E., Dickson, F., McVicar, T.R. & Ferrier, S. (2021) Habitat Condition Assessment System (HCAS version 2.1): Enhanced method for mapping habitat condition and change across Australia. CSIRO. <https://doi.org/10.25919/n3c6-7w60>.
- Williams, K.J., Liu, N., Valavi, R., Levick, S.R., Lehmann, E.A., Collings, S., Giljohann, K.M., Johnson, S., Botha, E.J., Munroe, S.E.M., Van Niel, T.G., Newnham, G., Paget, M., Malley, C., Carlile, P., Gunawardana, D., Lyon, P., Richards, A.E., Tetreault-Campbell, S. & Ferrier, S. (2025) Habitat Condition Assessment System (HCAS) version 3.3: Guide to the 90-metre data collection. Edition 2. Technical report EP2025-2979. CSIRO. <https://doi.org/10.25919/jz55-5d44>.
- Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.
- Wu, T., Liu, Y., Wang, Z., Yao, Y., Cheng, X. & Zhao, Y. (2026) A systematic review of terrestrial ecosystem resilience: Trends and adaptations over the past 50 years. *Geography and Sustainability*, 7, 100410. <https://doi.org/10.1016/j.geosus.2026.100410>.

Zhou, J., Wu, N. & Lu, T. (2025) Mapping human footprint changes over Qingzang Plateau. *Geography and Sustainability*, 6, 100249. <https://doi.org/10.1016/j.geosus.2024.09.010>.

Supporting information

Supporting Information: Continental-scale example of ecosystem condition workflow for Australia

Supporting Information: continental-scale example of ecosystem condition workflow for Australia

Roosbeh Valavi^{1,*}, Kristen J Williams^{2,*}, Eric A Lehmann², Simon Collings³, Shaun Levick⁴, Katherine M Giljohann¹, Ning Liu², Stephanie Johnson⁵, Samantha EM Munroe⁶, Thomas G Van Niel³, Matthew Paget², Cassandra Malley⁷, Peter Lyon⁷, Thomas D Harwood⁸, Simon Ferrier²

* corresponding authors: roosbeh.valavi@csiro.au, kristen.williams@csiro.au

Contents

Supporting Information: continental-scale example of ecosystem condition workflow for Australia	1
Summary statement	3
S1.0 Introduction	3
S1.1 Scope of the workflow demonstrated	3
S1.2 Summary of the approach to the worked example	4
S1.3 Continental Australia	5
S2 Data preparation	6
S2.1 Analysis extent	6
S2.2 Reference and highly modified sites	6
S2.3 Remotely sensed ecosystem condition variables	8
S2.4 Environmental covariates	10
S2.5 Sampling of training and benchmark reference sites	11
S2.6 Reference ecosystem modelling (REM)	11
S3 ClassicHCAS implementation	13
S3.1 Condition benchmarking with ClassicHCAS	13
S3.2 Calibration with ClassicHCAS	21
S3.3 Summary of HCAS v3.3 workflow configuration and ClassicHCAS settings	26
S3.4 Temporal condition estimation	31
S3.5 Bootstrap uncertainty from reference-site resampling	37
S4 Evaluation of the Australian HCAS v3.3 application	39
S4.1 Evaluation of inferred reference sites	39

S4.2 Comparison with Queensland Spatial BioCondition	40
S5 Implementation considerations.....	45
S6 Conclusions	46
Appendix A – Environmental covariates selected for use in HCAS v3.3.....	46
References	53

Summary statement

This Supporting Information provides a worked example of the HCAS workflow applied at continental scale in Australia using HCAS v3.3. It documents the preparation of input datasets, reference ecosystem modelling, implementation of condition benchmarking and calibration using **ClassicHCAS**, and generation of spatially explicit and temporal ecosystem condition estimates. The worked example provides the application-specific data, configuration and parameterisation needed to understand and reproduce the Australian implementation, complementing the conceptual and methodological description provided in the main manuscript.

S1.0 Introduction

This worked example documents the continental Australian application of the Habitat Condition Assessment System (HCAS) workflow used to generate HCAS v3.3. It describes the application-specific data preparation, reference ecosystem modelling, condition benchmarking, calibration, temporal estimation and uncertainty analysis used to generate spatially explicit ecosystem condition estimates from Earth observation (EO) data. **ClassicHCAS** provides the core computational functions for condition benchmarking and calibration within this broader workflow. The conceptual and methodological basis of HCAS is described in the accompanying manuscript (Valavi *et al.* 2026); this Supporting Information focuses on its implementation for continental Australia.

S1.1 Scope of the workflow demonstrated

The Australian application comprises data preparation, reference ecosystem modelling (REM), condition benchmarking, calibration, temporal condition estimation and uncertainty quantification. **ClassicHCAS** implements the core condition benchmarking and calibration functions; data preparation, reference ecosystem modelling, derivation of calibration coordinates and uncertainty analysis are undertaken externally (Table 1). This distinction is important for interpreting the role of **ClassicHCAS** within the broader HCAS workflow.

Table 1 Components of the HCAS workflow implemented externally and within **ClassicHCAS** for the continental Australia HCAS v3.3 application.

Workflow component	Process	Implementation
Data preparation	Preparation of remotely sensed ecosystem condition variables	External to ClassicHCAS
	Preparation of environmental covariates	External to ClassicHCAS
Reference system	Identification of inferred reference and highly modified sites	External to ClassicHCAS
	Sampling of reference sites for training and benchmarking	External to ClassicHCAS
Reference ecosystem modelling	Development of reference ecosystem models	External to ClassicHCAS

Workflow component	Process	Implementation
Calibration preparation	Derivation of calibration coordinates	External to ClassicHCAS
Inferential reference framework	Derivation of the inferential reference surface	ClassicHCAS
Condition benchmarking	Benchmarking of target locations against ecologically comparable reference ecosystems	ClassicHCAS
Calibration	Calibration of pre-calibration condition estimates to a standardised 0–1 scale	ClassicHCAS

A schematic overview of the broader HCAS workflow and the functions implemented within **ClassicHCAS** is provided in Figure 1 of the main text, with the conceptual and methodological basis described in Sections 2–3.

Uncertainty quantification for HCAS v3.3 was undertaken partly externally and partly internally to **ClassicHCAS**, using bootstrap resampling of training and benchmark reference-site samples. The associated **ClassicHCAS** reference ecosystem modelling, condition estimation and calibration steps were repeated to quantify variability in ecosystem condition estimates arising from reference-site sampling (Section S3.5).

S1.2 Summary of the approach to the worked example

The Australian implementation presented here follows the general HCAS workflow, adapted to a continental-scale dataset:

- remotely sensed ecosystem condition variables are derived from Landsat observations
- environmental covariates are used to model variation in ecosystem characteristics among inferred reference ecosystems
- inferred reference sites support reference ecosystem modelling and benchmarking, while independently derived ecosystem modification classes inform parameter tuning and calibration
- benchmarking and calibration are implemented using **ClassicHCAS**, with calibration informed by empirical interpretation of the condition gradient
- temporal ecosystem condition estimates were generated by applying the same reference ecosystem models, reference data and condition benchmarking configuration across all temporal epochs.

The resulting outputs comprise the HCAS v3.3 data collection for continental Australia ([Valavi *et al.* 2025b](#); [Williams *et al.* 2025b](#)). Section S2 describes data preparation and reference ecosystem modelling, Section S3 documents the **ClassicHCAS** implementation and temporal and uncertainty analyses, Section S4 presents evaluation of the Australian HCAS v3.3 application, and Section S5 describes additional implementation considerations.

S1.3 Continental Australia

Australia is an island continent in the Southern Hemisphere and the world's smallest continent, covering approximately 7.69 million km². Across the continent, climates range from humid tropical and monsoonal conditions in the north and along parts of the coast, through subtropical and temperate regions, to winter-rainfall climates in the south-west and an extensive arid and semi-arid interior. Strong gradients in rainfall amount and seasonality, temperature and water availability occur with latitude, distance from the coast and topography (Stern *et al.* 2000).

Vegetation includes tropical rainforests, tropical savannas, eucalypt forests and woodlands, temperate grasslands, shrublands and extensive arid grasslands and shrublands. At continental scale, the distribution and expression of these vegetation types reflect interactions among water and nutrient availability, temperature, soils, geomorphology and disturbance, particularly fire and flooding (Keith 2017). These patterns occur across predominantly old and generally subdued landscapes, although geomorphic histories and rates of landscape change vary substantially among regions (Bishop & Pillans 2010). Hydrological settings range from large internally draining basins to drainage systems flowing to the coast (Stein *et al.* 2014; Bureau of Meteorology 2023). Variation in these interacting environmental attributes creates markedly different ecological contexts that must be represented in continental-scale ecosystem condition assessment.

Human land use adds a further spatial gradient. Higher-intensity agriculture—including cropping, horticulture and grazing on sown or otherwise modified pastures—is concentrated mainly in the southern, eastern and south-western farming regions and in smaller irrigated districts inland (Williams *et al.* 2021b; ABARES 2024). By contrast, extensive livestock grazing occurs across much of the interior and northern rangelands, often within native vegetation, while large areas are managed primarily for conservation or are subject to other relatively low-intensity uses (Foran *et al.* 2019). Australia's population is concentrated in a small number of coastal metropolitan centres, and continuing expansion at their urban fringes creates pronounced gradients of urban and peri-urban land-use modification (McFarland 2015; Dühr 2024). Land-use intensity and ecosystem modification therefore vary within as well as among climatic and biogeographic regions.

Fire regimes also vary widely. Frequent, predominantly grass-fuelled fires characterise much of monsoonal northern Australia, whereas very high-intensity, low-frequency fire regimes occur particularly in some high-biomass forests of southern and eastern Australia (Murphy *et al.* 2013). In arid and semi-arid systems, fire activity is often limited by discontinuous fuels but can increase episodically after above-average rainfall promotes herbaceous biomass and fuel continuity (Bradstock 2010). Other disturbances, including drought and flooding, further influence ecosystem dynamics and remotely sensed ecosystem characteristics. Interactions among climate, water and nutrient availability, land use and disturbance create a complex mosaic of natural variability and ecosystem modification, making persistent anthropogenic change difficult to distinguish from natural ecological dynamics.

S2 Data preparation

The continental Australian HCAS v3.3 application required three primary data inputs: (i) environmental covariates describing the abiotic characteristics that define ecological context; (ii) remotely sensed ecosystem condition variables representing observed ecosystem characteristics; and (iii) inferred reference-site data used for reference ecosystem modelling and condition benchmarking. Additional data representing ecosystem modification were used for parameter tuning and calibration. This section describes the preparation and organisation of these inputs for the Australian application.

Input datasets were also required to support open and unrestricted reuse, with licensing compatible with release of HCAS outputs under a Creative Commons Attribution 4.0 licence.¹

S2.1 Analysis extent

The analytical extent comprised continental Australia with a 3 km buffer to ensure complete processing coverage of coastal and intertidal areas (Liu *et al.* 2025b). At the 90 m analysis resolution, this extent comprised approximately 967 million pixels. Final outputs were clipped to the authoritative national coastline.

A spatial resolution of 90 m was selected to align with the national Soil and Landscape Grid of Australia, which provided many of the environmental covariates used in the workflow (Malone *et al.* 2025).

All spatial data were processed using the Australian Albers equal-area projection (GDA94; EPSG:3577)² and aligned to a common analysis grid. Extent masks defined valid data regions and maintained spatial alignment among workflow components. Source datasets and preprocessing procedures are described in Liu *et al.* (2025b), with the resulting HCAS v3.3 datasets documented in Williams *et al.* (2025b).

S2.2 Reference and highly modified sites

Inferred reference sites and highly modified areas were identified through separate rule-based workflows integrating multiple lines of spatial evidence (Giljohann *et al.* 2025) (Figure 1 – left). The reference site dataset provided candidate locations for reference ecosystem modelling and condition benchmarking (Sections S2.5–S3.1). Because direct field observations were not available consistently across continental Australia, these locations are designated as *inferred reference sites*. Particular attention was given to improving representation of coastal–intertidal and inland water ecosystems, for which suitable reference locations were comparatively sparse (Williams *et al.* 2025b).

Ecological evidence was also used to refine the inferred reference-site dataset where mapped locations were unlikely to represent high-integrity ecosystems. For example, recent fires affected

¹<https://creativecommons.org/licenses/by/4.0/>

²<https://epsg.io/3577>

fire-sensitive Gondwanan rainforests in Lamington National Park that are characterised by very long fire-return intervals (e.g. [Godfree *et al.* 2021](#)). Recently burnt rainforest areas were therefore excluded from the inferred reference-site pool ([Giljohann *et al.* 2025](#)). This example provided an ecologically interpretable test of reference site selection and subsequent condition estimation ([Williams *et al.* 2025a](#)), although comparable ecological evidence was not available consistently across all Australian ecosystems.

Reference and highly modified areas were delineated using source data at their highest available spatial resolution and subsequently aligned to the 90 m HCAS v3.3 analysis grid. A temporal dimension to reference-site identification was investigated but was not implemented because available source datasets were insufficiently consistent through time for continental application ([Giljohann *et al.* 2025](#)).

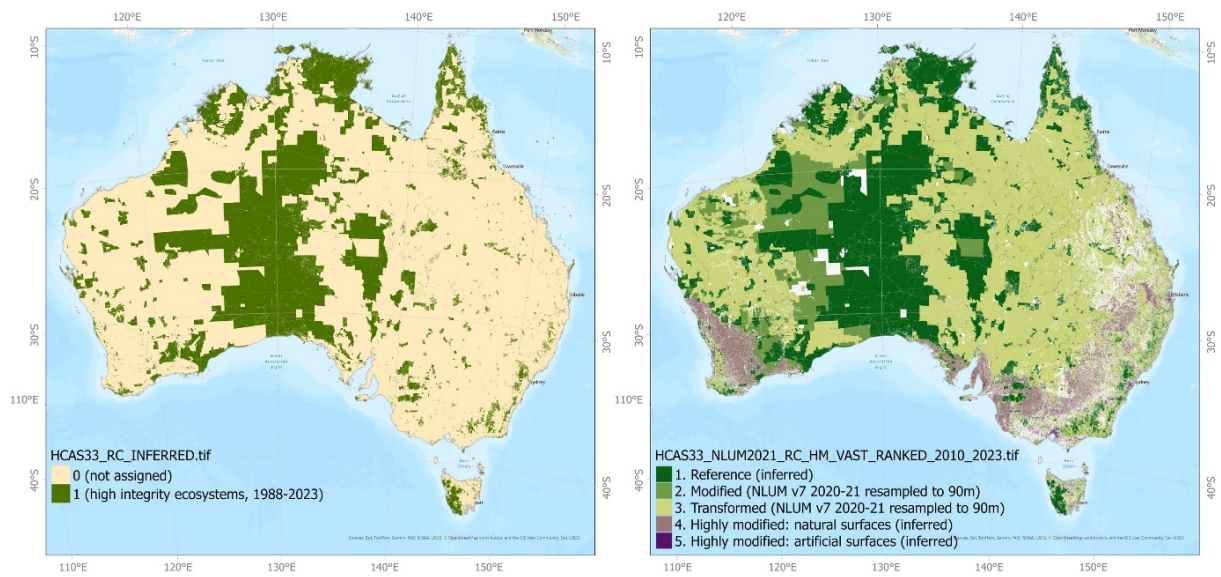


Figure 1 Reference areas and ecosystem modification classes used in the continental Australian HCAS v3.3 application. Left: conservatively mapped inferred reference ecosystems used to identify structured samples of reference sites for reference ecosystem modelling and condition benchmarking ([Giljohann *et al.* 2025](#)). Right: ordinal ecosystem modification classes derived from inferred reference and highly modified areas ([Giljohann *et al.* 2025](#)) with a background of 2020–21 National Land Use Map categories (2020–21 NLUM version 7.0; [ABARES 2024](#)) assigned to ‘Modified’ or ‘Transformed’ classes of the Vegetation Assets, States and Transitions (VAST) framework ([Thackway & Lesslie 2006](#); [Thackway & Lesslie 2008](#)) by [Giljohann *et al.* \(2024\)](#). Data and methods are described in [Giljohann *et al.* \(2024\)](#); [Giljohann *et al.* \(2025\)](#); [Valavi *et al.* \(2025b\)](#); [Williams *et al.* \(2025b\)](#). The modification classes subsequently used for HCAS v3.3 benchmarking parameter tuning and calibration used only the inferred reference sites with the 2020–21 National Land Use Map categories (2020–21 NLUM version 7.0; [ABARES 2024](#)) assigned to the VAST framework by [Giljohann *et al.* \(2024\)](#).

Reference site delineation was intentionally conservative and reflects the sensitivity of the HCAS workflow to the validity of reference sites. Locations likely to contain mixtures of high-integrity and modified ecosystem states were excluded, reducing the likelihood that modified locations were included in the reference pool but also reducing reference site coverage in environments where high-integrity ecosystems are scarce. At 90 m resolution, the resulting dataset comprised approximately 256 million inferred reference pixels.

Highly modified areas were identified separately using contemporary land-use, infrastructure and

other evidence of ecosystem modification (Giljohann *et al.* 2025). At 90 m resolution, this dataset comprised approximately 53 million pixels. The combined pattern of conservatively identified inferred reference and highly modified sites with a background of other modification classes is shown in Figure 1 (left). This inferred highly modified dataset was distinct from the ecosystem modification classes subsequently used for HCAS v3.3 parameter tuning and calibration.

For parameter tuning and calibration, national land-use mapping for 2020–21 (ABARES 2024) was cross-walked to the Vegetation Assets, States and Transitions (VAST) framework (Thackway & Lesslie 2006; Thackway & Lesslie 2008), following Giljohann *et al.* (2024), to derive ordinal classes representing increasing ecosystem modification. These classes, combined with the inferred reference site dataset, were used in benchmarking parameter selection (Section S3.1.3) and calibration of the ecosystem condition scale (Section S3.2).

S2.3 Remotely sensed ecosystem condition variables

Remotely sensed ecosystem condition variables provide spatially explicit proxies for ecosystem characteristics associated primarily with structure and function. For HCAS v3.3, these variables were derived from Landsat Collection 3 Analysis Ready Data (Commonwealth of Australia 2021) Derivative Products version 4.0.0, including GeoMedians, GeoMADs and Fractional Cover (Geoscience Australia 2024; Lymburner 2024; Mueller *et al.* 2024). The original annual data at 30 m resolution were aggregated to 90 m to align with the HCAS v3.3 analysis grid and reduce sensitivity to fine-scale sensor artefacts and spatial variability (Levick *et al.* 2025).

A set of 36 candidate variables was initially derived from these datasets. Fourteen ecosystem condition variables were subsequently selected based on ecological relevance and exploratory analysis to reduce dimensionality while retaining complementary information about ecosystem characteristics (Table 2). The selected variables comprise fractional-cover metrics for photosynthetic vegetation, non-photosynthetic vegetation and bare soil; spectral indices associated with vegetation productivity, fire response, canopy moisture, surface water and soil exposure; and measures of intra-annual spectral and fractional-cover variability.

Two temporal aggregations were generated for each variable. Long-term summaries were calculated over 1988–2024 and provided the ecosystem characteristic data used to establish the long-term HCAS v3.3 condition inference configuration. Short-term summaries were calculated using overlapping 3-year windows to represent temporal variation in the same ecosystem characteristics. Spatial and temporal aggregation also reduced sensitivity to observation gaps and residual contamination (e.g. due to cloud, smoke, or smog) in the underlying satellite data.

Some spatial gaps remained where persistent cloud cover resulted in insufficient observations for the intra-annual fractional-cover percentile variables. These gaps were filled using local neighbourhood averaging followed by global mean imputation to provide spatially complete inputs (Levick *et al.* 2025). A parallel workflow retained information on the location of gap-filled observations to support evaluation of data quality and interpretation of HCAS v3.3 outputs.

Each long-term ecosystem condition variable was mean-centred and scaled by its standard deviation (z -score standardisation). The same variable-specific means and standard deviations

were then applied to the corresponding short-term variables, ensuring that long- and short-term observations were represented within the same standardised multivariate feature space.

These variables provide the observed ecosystem characteristics used in condition inference. Reference ecosystem models generate corresponding predictions of these characteristics for inferred reference ecosystems based on environmental covariates (Section S2.6). The observed and reference ecosystem model predictions are subsequently used to construct the reference information required for condition benchmarking in **ClassicHCAS** (Section S3.1).

Table 2 Remotely sensed ecosystem condition variables derived from Landsat observations and used to represent ecosystem characteristics in HCAS v3.3. Ecosystem Condition Typology classes follow [Czúcz et al. \(2021\)](#) and Table 5.1 of [United Nations et al. \(2024\)](#). Data derivation and variable selection are described in [Levick et al. \(2025\)](#) and [Williams et al. \(2025b\)](#).

Label	Description	Relevance	Ecosystem Condition Typology
kNDVI	Kernel Normalised Difference Vegetation Index	Vegetation productivity and photosynthetic activity – ecosystem function	B3
NBR	Normalised Burn Ratio	Vegetation recovery-response to fire and disturbance – ecosystem function	A1/B3
NDMI	Normalised Difference Moisture Index	Vegetation and canopy moisture status – ecosystem function	B3
NDWI	Normalised Difference Water Index	Surface-water and vegetation-water status – ecosystem function	A1/B3
BSI	Bare Soil Index	Bare-ground and soil exposure – ecosystem structure	A1/B2
SDEV	Spectral distance median absolute deviation (SMAD)	Multivariate intra-annual spectral variability – ecosystem diversity	B3
EDEV	Euclidean distance median absolute deviation (EMAD)	Multivariate intra-annual spectral variability – ecosystem diversity	B3
BCDEV	Bray-Curtis distance median absolute deviation (BCMAD)	Multivariate intra-annual spectral variability – ecosystem diversity	B3
PV10	10th percentile of intra-annual photosynthetic vegetation cover fractions	Lower percentile of photosynthetic vegetation cover – ecosystem structure	B2
PVrange	Inter-percentile range (10 th – 90 th percentile difference) of intra-annual photosynthetic vegetation cover fractions	Intra-annual range in photosynthetic vegetation cover – ecosystem dynamics	B3
NPV10	10th percentile of intra-annual non-photosynthetic cover fractions	Lower percentile of non-photosynthetic vegetation cover – ecosystem structure	B2
NPVrange	Inter-percentile range (10 th – 90 th percentile difference) of intra-annual non-photosynthetic vegetation cover fractions	Intra-annual range in non-photosynthetic vegetation cover – ecosystem dynamics	B3
BS10	10th percentile of intra-annual bare soil cover fractions	Lower percentile of bare-soil cover – ecosystem structure	B2

BSrange	Inter-percentile range (10 th – 90 th percentile difference) of intra-annual bare soil cover fractions	Intra-annual range in bare-soil cover – ecosystem dynamics	B3
---------	--	--	----

S2.4 Environmental covariates

Environmental covariates represent abiotic characteristics used to describe ecological context for reference ecosystem modelling (REM). For HCAS v3.3, more than 125 candidate covariates were generated at 90 m spatial resolution or compiled from national datasets (Liu *et al.* 2025b). Principal sources included the Soil and Landscape Grid of Australia (Grundy *et al.* 2015; Malone *et al.* 2025) and associated environmental covariate stacks developed for digital soil mapping (Searle *et al.* 2022), from which candidate covariates relevant to major continental environmental gradients were identified.

Candidate covariates were selected to represent relatively stable abiotic characteristics rather than contemporary disturbance or land use. These included long-term climate variables based on 30-year climatologies centred on 1990 (Xu & Hutchinson 2011; Xu & Hutchinson 2013), together with soil, lithological, topographic and hydrological variables that collectively influence ecosystem distribution and variability; consistent with established approaches in correlative ecological modelling (Guisan & Zimmermann 2000; Williams *et al.* 2012).

An important consideration was minimising dependence between environmental covariates used to represent ecological context and the remotely sensed ecosystem condition variables subsequently used in condition inference. Some modelled soil, lithological and geophysical datasets incorporate remotely sensed predictors that may also contain information represented in the remotely sensed ecosystem condition variables (Grundy *et al.* 2020; Kidd *et al.* 2020; Wilford & Roberts 2021; Wilford *et al.* 2023a; Wilford *et al.* 2023b). Where necessary, selected soil attribute and geophysical covariates were rederived without remotely sensed predictors (Searle 2023; Liu *et al.* 2025b).

Additional environmental covariates were developed for coastal–intertidal systems and episodic inland-water environments, where standard continental datasets provided limited representation of relevant environmental gradients (Liu *et al.* 2025b). All covariates were aligned to the common analysis grid, small data gaps were filled, and selected datasets were extended across the buffered coastal analysis extent. The resulting environmental covariate datasets were published at 250 m and 90 m spatial resolutions (Liu *et al.* 2025a; Liu *et al.* 2025b).

Exploratory analysis and expert review were used to reduce redundancy while retaining complementary covariates representing climate, soil properties, lithology, topography and surface hydrology (Liu *et al.* 2025b). This process resulted in 58 covariates for HCAS v3.3, of which 46 were retained for final reference ecosystem modelling following exploratory modelling (Section S2.6). Four additional covariates were retained where required for sampling benchmark reference sites (Section S2.5). The membership of each covariate set is identified in Appendix A (Table 6). All covariates used within each analytical component were *z*-score standardised before analysis.

S2.5 Sampling of training and benchmark reference sites

The inferred reference site dataset (Section S2.2) is unevenly distributed across geographic and environmental space and is too large to use directly in reference ecosystem modelling and condition benchmarking. HCAS v3.3 therefore used two independently sampled subsets of inferred reference sites (Section S2.2): *training reference sites* for reference ecosystem modelling (REM) and *benchmark reference sites* for condition benchmarking.

Training reference sites were sampled to represent the environmental gradients occurring across inferred reference ecosystems and thereby support REM prediction across the range of environments represented in the study domain. For HCAS v3.3, training sites were selected using a feature-space coverage sampling approach (Brus 2019). Given the approximately 256 million inferred reference pixels at 90 m resolution, a staged implementation based on k -means clustering was used to approximate optimal coverage of the environmental feature space represented by 48 environmental covariates (Appendix A, Table 6). This resulted in 488,619 training reference sites. The sampling methodology and testing framework are described in Valavi *et al.* (2025a).

Benchmark reference sites were sampled separately because condition benchmarking requires representation of variation in both environmental covariates and remotely sensed ecosystem characteristics. A principal components analysis (PCA) was applied to a combined feature set comprising the 48 environmental covariates used for reference site sampling and the 14 long-term ecosystem condition variables (Section S2.3; Table 2). The first 20 principal components, accounting for 90% of the variance in the combined feature set, were used for benchmark reference site sampling. Four supplementary environmental covariates were included to improve representation of complex terrain, riparian zones, coastal–intertidal environments and areas affected by seasonal snow (Appendix A, Table 6).

A staged clustering procedure was used to generate the initial benchmark reference site sample. Preliminary inspection of condition estimates subsequently identified areas where low condition estimates coincided with limited representation of local reference ecosystems. The benchmark sample was therefore augmented to improve representation of wetland ecosystems (Macfarlane *et al.* 2024b; 2024a) and protected areas (DCCEEW 2025a; 2025b). The final benchmark reference dataset for HCAS v3.3 comprised 590,342 sites, improving representation of environmental gradients and variation in ecosystem characteristics across continental Australia (Williams *et al.* 2025b).

S2.6 Reference ecosystem modelling (REM)

Reference ecosystem modelling (REM) models relationships between environmental covariates and ecosystem characteristics observed at training reference sites, enabling those ecosystem characteristics to be predicted across the study domain. The resulting predictions provide an input to condition benchmarking (Section S3.1).

For HCAS v3.3, the REM was fitted using 488,619 training reference sites, each attributed with 46 environmental covariates (Section S2.4; Appendix A, Table 6) and 14 long-term remotely sensed

ecosystem condition variables (Section S2.3; 2). These data represent ecosystem characteristics observed at inferred reference ecosystems and the environmental contexts in which they occur. The z -score-standardised environmental covariates and ecosystem condition variables were used for REM fitting and subsequent condition inference analyses.

The REM was implemented using the `mgcv` R package (Wood 2011), with one generalised additive model (GAM) fitted for each of the 14 ecosystem condition variables. GAMs provide a flexible framework for representing potentially nonlinear ecological relationships between response variables and environmental predictors (Pedersen *et al.* 2019). Each model related an observed ecosystem characteristic at the training reference sites to the environmental covariates, and the fitted models were subsequently used to predict that characteristic across the study domain. Models were fitted using the `bam()` function with fast restricted maximum likelihood estimation, discrete approximation and parallel processing to accommodate the large training dataset (Marra & Wood 2011; Li & Wood 2020; Valavi *et al.* 2025a).

All 46 environmental covariates were included in each of the 14 models. Smoothness selection and penalisation allowed flexible non-linear relationships while controlling model complexity. Model predictive performance was evaluated using five-fold cross-validation (Valavi *et al.* 2025a).

Predictive performance varied among ecosystem condition variables (Figure 2). Several spectral indices, including kNDVI, NBR, NDMI and NDWI, showed relatively high predictive performance, with R^2 values generally exceeding 0.80. Fractional-cover and bare-ground variables also showed strong to moderate performance: PV10, BSI and BS10 had relatively high R^2 , whereas NPV10 was more moderately predicted. Prediction error showed a somewhat different pattern, with NDWI having particularly low RMSE, while some variables with high R^2 , including NBR and NDMI, retained comparatively larger errors in standardised units.

Variables representing intra-annual variability generally showed lower explanatory power. The inter-percentile range metrics PVrange, NPVrange and BSRrange, together with the multivariate spectral variability metrics EDEV and SDEV, had comparatively low R^2 values, with BSRrange among the poorest-performing variables. BCDEV was an exception among the variability metrics, showing substantially higher explanatory performance. RMSE did not follow the same ranking as R^2 across all variables, although BSRrange combined relatively low explanatory power with comparatively high prediction error. The weaker performance of several variability metrics may reflect their greater sensitivity to seasonal conditions, weather and episodic disturbance, which are not explicitly represented by the relatively stable abiotic covariates used in the REM. In contrast, the stronger performance of many spectral indices and fractional-cover variables is consistent with these variables responding more directly to broad environmental gradients represented by the REM covariates.

The fitted REMs provide spatial predictions of the 14 ecosystem characteristics conditional on the environmental relationships represented by the training reference sites. These predictions are subsequently used with observed ecosystem characteristics in condition benchmarking (Section S3.1).

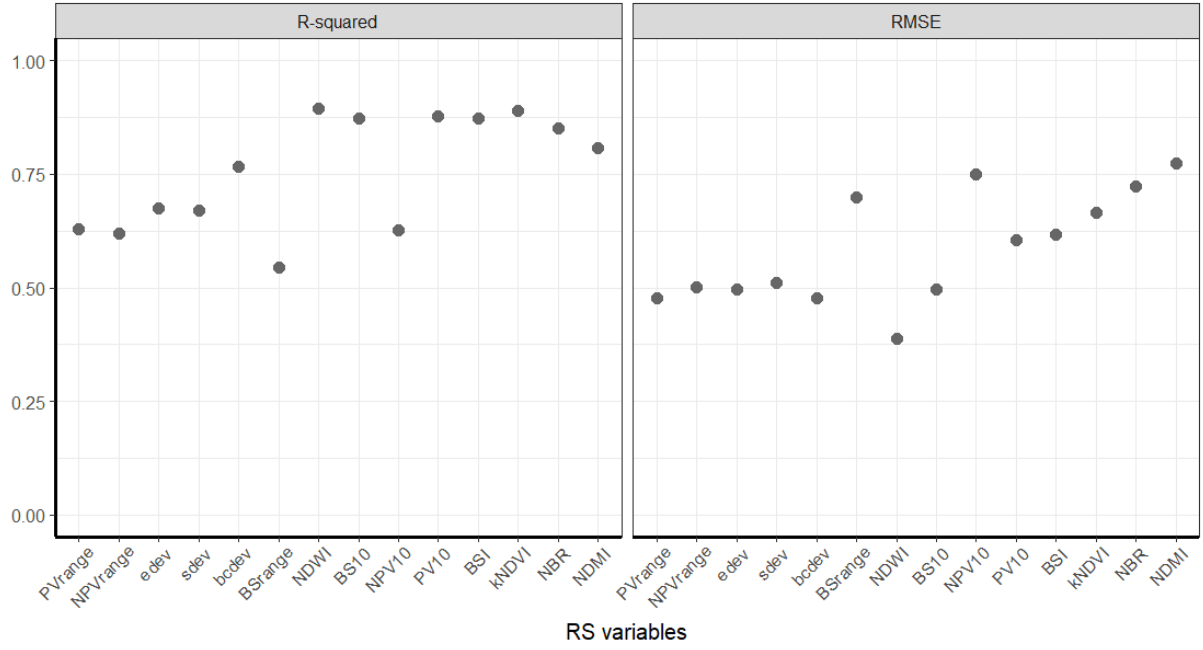


Figure 2 Predictive performance of the 14 reference ecosystem models fitted for HCAS v3.3, evaluated using five-fold cross-validation. Panels show the coefficient of determination (R^2) and root mean square error (RMSE) for each remotely sensed ecosystem condition variable (Table 2). RMSE values are reported in standardised units following z-score standardisation of the response variables. Source: [Valavi et al. \(2025a\)](#).

S3 ClassicHCAS implementation

HCAS v3.3 was generated using **ClassicHCAS** version 0.2.0; this Supporting Information reports functions and arguments using the current **ClassicHCAS** API (version 2.0.0), with correspondence to the original production terminology given in the workflow/settings table (Table 5, Section S3.3), and does not imply that HCAS v3.3 was regenerated with the current software.

S3.1 Condition benchmarking with **ClassicHCAS**

Following data preparation and reference ecosystem modelling (Section S2), the reference surfaces and condition estimates for HCAS v3.3 were generated using **ClassicHCAS**. The conceptual and mathematical basis of condition benchmarking is described in Section 3.2 of the manuscript; this section documents the inputs, software functions and settings used for the HCAS v3.3 continental Australia application.

Condition benchmarking for HCAS v3.3 used four inputs:

1. Training reference sites: 488,619 sampled inferred reference sites (Section S2.5) used to construct the reference density surface and inferential reference surface.
2. Benchmark reference sites: 590,342 independently sampled inferred reference sites (Section S2.5) used to select ecologically comparable reference ecosystems for each target location during condition benchmarking.

3. Observed ecosystem condition variables: aligned raster layers containing the 14 long-term remotely sensed ecosystem condition variables used to represent observed ecosystem characteristics (Section S2.3; Table 2).
4. Reference ecosystem model predictions: aligned raster layers containing REM predictions corresponding to the same 14 long-term remotely sensed ecosystem condition variables (Section S2.6). These predictions were used to calculate predicted pairwise distances between locations, representing differences associated with ecological context.

Observed and predicted remotely sensed ecosystem condition variables were ordered consistently and represented within the same standardised multivariate feature space. Training and benchmark reference sites were attributed with their spatial coordinates and the corresponding observed and REM-predicted values for the 14 long-term ecosystem condition variables.

3.1.1 Construction of the reference surfaces

Two processing steps generated the reference surfaces used during HCAS v3.3 condition benchmarking.

1. Reference density surface

The reference density surface was constructed using the `ref_density()` function in `ClassicHCAS`. Pairwise comparisons were calculated among the 488,619 training reference sites within a spatial search radius of 1,000 km (`radius_km = 1000` for HCAS v3.3). For each site pair, Manhattan (L_1) distances were calculated separately across the 14 REM-predicted variables and the 14 observed variables.

The resulting predicted and observed pairwise distances were aggregated into a two-dimensional reference density surface. The horizontal axis represents predicted remote-sensing distance, the vertical axis represents observed remote-sensing distance, and each cell records the number of training-reference pairs within that distance combination. For HCAS v3.3, `bin_width = 0.05` and `bin_num = 650` were selected following visual comparison of alternative bin widths (Figure 3).

2. Inferential reference surface

The reference density surface was processed using `normalise()`. For HCAS v3.3, Gaussian smoothing was applied across neighbouring cells, the surface was restricted using `trim_size = 400` and `offset = 0`, and cell values were normalised within each predicted-distance class (Figure 3). The resulting inferential reference surface represents the empirical distribution of observed pairwise distances associated with each predicted pairwise distance among the training reference ecosystems

The inferential reference surface was subsequently used during condition benchmarking to assign a relative surface value to each retained target–benchmark reference-site pair. Construction of

the surface is distinct from the process of selecting and weighting of benchmark reference sites described in Section S3.1.2.

3.1.2 Benchmarking against reference ecosystems

For each target location, `ClassicHCAS benchmark()` selects and weights a target-specific set of benchmark reference sites and then combines the associated *inferential reference surface* values to produce a pre-calibration condition estimate (Figure 4). The conceptual and mathematical basis of condition benchmarking is described in Section 3.2 of the manuscript; the subsections below document the process and settings used for HCAS v3.3.

A. Selection and weighting of benchmark reference sites

The benchmark-reference sample used for each target location was derived through three sequential filters.

1. *Spatial candidate selection.* Candidate benchmark reference sites were drawn from the independently sampled set of 590,342 benchmark reference sites (Section S2.5). For HCAS v3.3, sites within a 200 km benchmark-candidate search radius of the target location were considered (`radius_km = 200` in `benchmark()`). This setting differs from the 1,000 km training-reference pair search radius used to construct the reference density surface (Section S3.1.1).
2. *First-stage selection using predicted pairwise distance.* Predicted pairwise distances were calculated between the target location and the spatial candidate reference sites using the 14 reference ecosystem model predictions. Up to 70 sites with the smallest adjusted predicted distances were retained (`k1 = 70`). A geographic-distance penalty (`xy_penalty = 30`) was applied during this first-stage ranking to favour geographically nearer reference sites among sites with otherwise similar predicted distances.
3. *Second-stage selection using the inferential reference surface.* For each of the retained target-reference pairs, the predicted and observed pairwise distances identified a position on the inferential reference surface. The surface value at that position quantified how strongly the pair was represented in the empirical distribution derived from training reference ecosystems. Up to 10 pairs with the highest inferential-reference-surface values were retained as the final target-specific benchmark reference site sample (`k2 = 10`).

The final benchmark reference sites were then weighted according to their predicted pairwise distance from the target location. HCAS v3.3 applied a Cauchy-type weighting function with `lambda = 2`, assigning greater influence to retained reference sites with smaller predicted distances.

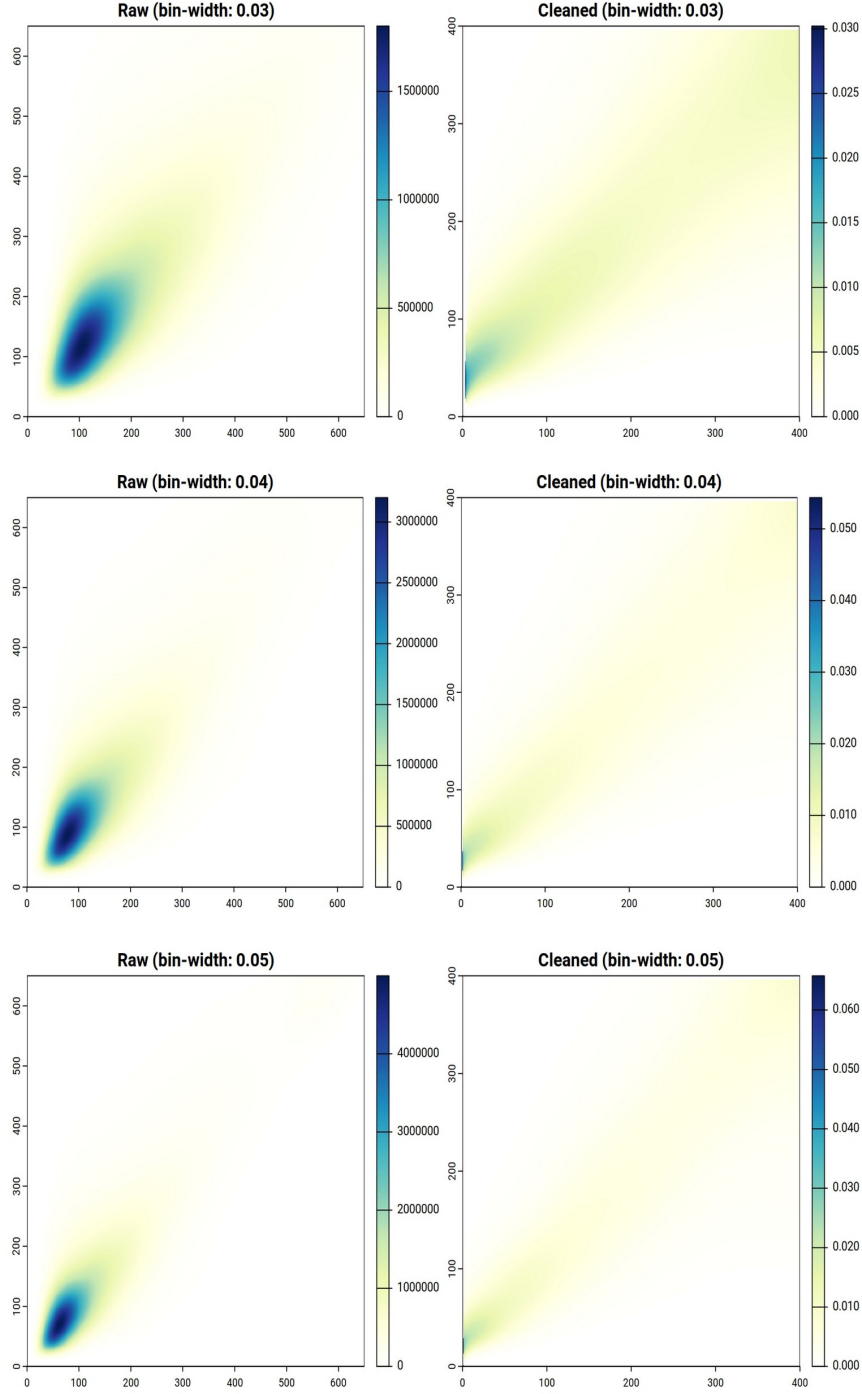


Figure 3 Reference surfaces generated using candidate bin widths of 0.03, 0.04 and 0.05 for the HCAS v3.3 continental Australian application. The left panels show the reference density surfaces constructed from predicted and observed pairwise remote-sensing distances among training reference sites. The right panels show the corresponding inferential reference surfaces after Gaussian smoothing, trimming and normalisation within predicted-distance classes. A bin width of 0.05 was selected for HCAS v3.3 following visual assessment of surface stability and consistency with earlier HCAS implementations (e.g. [Williams *et al.* 2021a](#); [Williams *et al.* 2025a](#)).

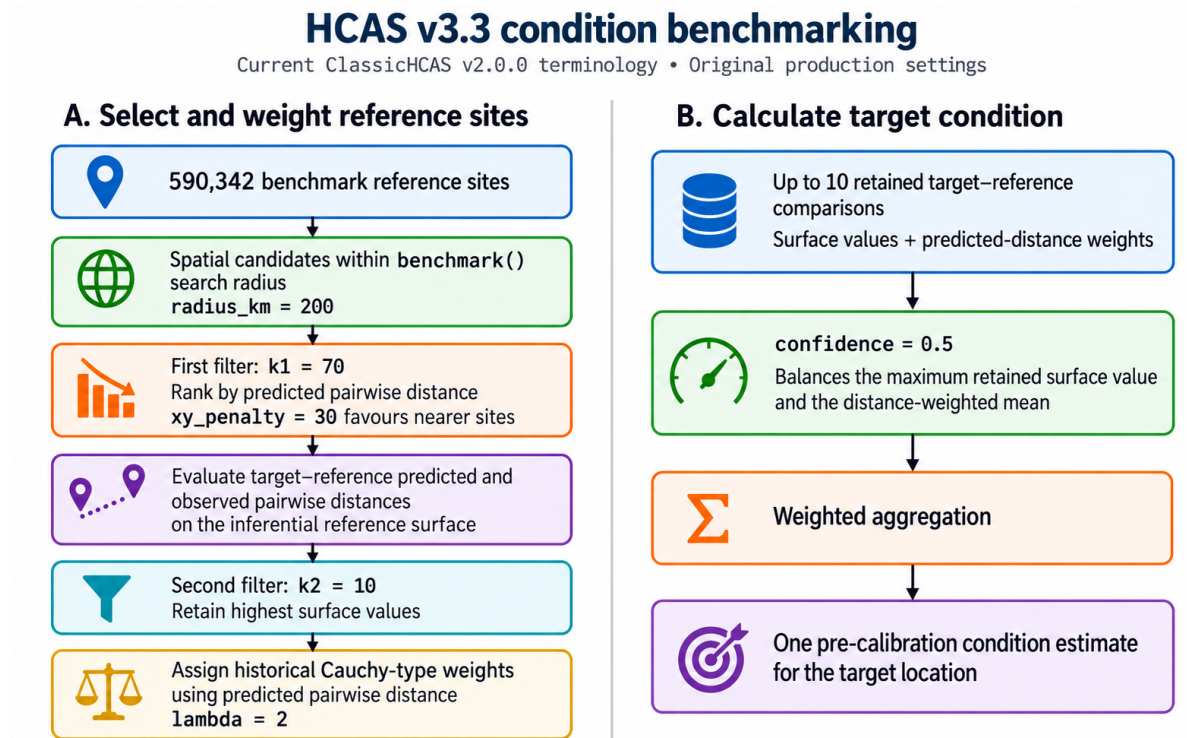


Figure 4 Sequencing of HCAS v3.3 benchmarking starts with benchmark reference site selection and weighting (left panel), followed by calculation of a single pre-calibration condition estimate for each target location (right panel). Current **ClassicHCAS** parameter names are shown with the original HCAS v3.3 production settings. Current API: **confidence** = 0.5 remains supported with **boost** = NULL (or NA). Default **boost** = **k2** is an alternative; HCAS v3.3 did not use **boost**. Current **kernel** = "Cauchy" differs from the historical Cauchy-type weighting. HCAS v3.3 was generated with **ClassicHCAS** v0.2.0. Current API labels do not imply regeneration. AI-assisted schematic prepared under author supervision using OpenAI image-generation tools.

B. Calculation of the pre-calibration condition estimate

Each target–reference pair in the final benchmark-reference sample contributed an inferential-reference-surface value and an associated predicted-distance weight. **ClassicHCAS** combined these values across the retained reference sites to produce one pre-calibration condition estimate for the target location. For HCAS v3.3, **confidence** = 0.5 controlled the relative contribution of the maximum retained inferential-reference-surface value and the predicted-distance-weighted mean of the retained values to this calculation.

The resulting pre-calibration condition estimate represents the target location’s relative alignment with ecologically comparable high-integrity reference ecosystems. Higher values indicate greater alignment with the empirical relationships between predicted and observed pairwise distances represented by the training reference ecosystems. The pre-calibration estimates were subsequently transformed to the 0–1 ecosystem condition scale described in Section S3.2.

3.1.3 Parameter selection, tuning and diagnostic inspection

HCAS v3.3 production settings were established through sequential AUC-based parameter tuning supported by diagnostic inspection (Figure 5).

A subset of benchmarking parameters for HCAS v3.3 were selected using a sequential AUC-based search across alternative combinations of **ClassicHCAS** arguments (Figure 6) (Valavi *et al.* 2025a). The tuning analysis varied the benchmark-reference candidate and retention settings (**k1** and **k2**), geographic-distance penalty (**xy_penalty**) and **confidence** setting. The benchmark-candidate search radius was fixed at **radius_km** = 200 for the analysis shown in Figure 6. The Cauchy distance-weighting setting was also held constant at **lambda** = 2.0 and was not included in the tuning search.

The tuning data comprised inferred reference sites and 2020–21 National Land Use Map classes cross-walked to ordinal ecosystem modification classes based on the VAST framework (Section S2.2). The two most modified classes (‘removed’ and ‘replaced-managed’) were combined, resulting in four classes used in the tuning analysis.

Area under the receiver-operating characteristic curve (AUC) was used to evaluate separation among the ordinal ecosystem modification classes for each parameter combination considered during the sequential search. Candidate settings were preferred where pre-calibration condition estimates showed clearer separation among classes and generally decreased with increasing ecosystem modification. The modification classes provided an ordinal tuning criterion rather than independent observations of ecosystem condition; the analysis therefore supported selection of operational benchmarking settings but did not constitute validation of HCAS condition estimates.

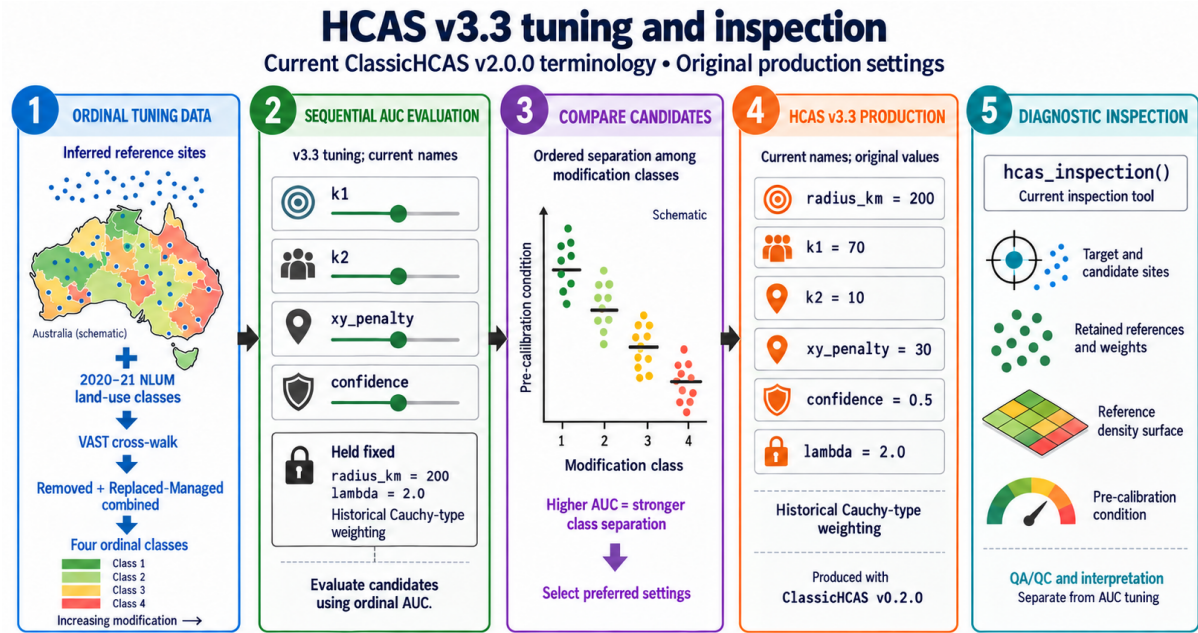


Figure 5 Schematic distinguishing the HCAS v3.3 sequential AUC-based tuning process, the production-run settings selected, the constant weighting setting, and the separate diagnostic role of the inspection tool. Current API: `confidence = 0.5` remains supported with `boost = NULL` (or NA). Default `boost = k2` is an alternative; HCAS v3.3 did not use `boost`. Current `kernel = "Gaussian"` is the default; `kernel = "Cauchy"` is optional and differs from v3.3 Cauchy-type weighting. HCAS v3.3 was generated with ClassicHCAS v0.2.0. Current API labels do not imply regeneration. AI-assisted schematic prepared under author supervision using OpenAI image-generation tools.

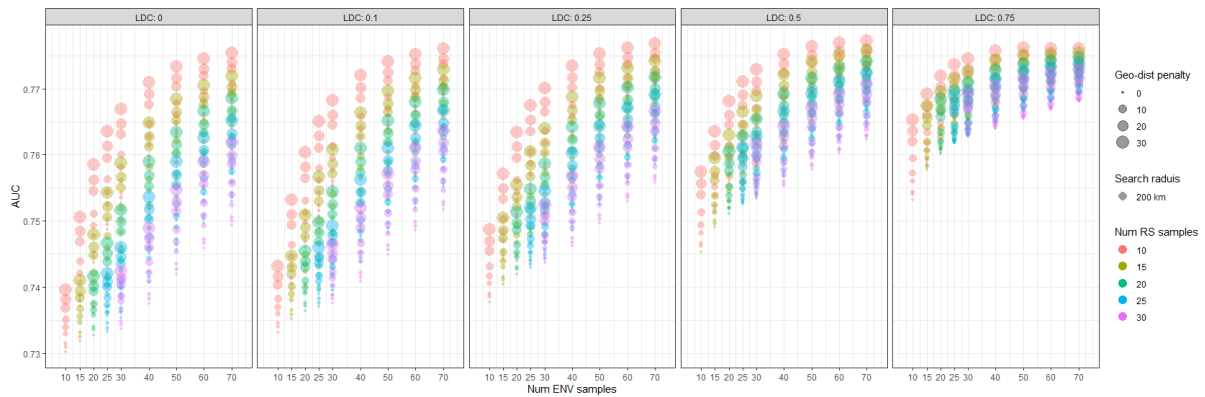


Figure 6 Sequential AUC-based evaluation of ClassicHCAS version 0.2.0 benchmarking settings for the HCAS v3.3 continental Australia application. Points show AUC values for parameter combinations evaluated during the sequential search using four ordinal ecosystem modification classes derived from inferred reference sites and 2020–21 National Land Use Map classes cross-walked to the VAST framework. The horizontal axis shows `k1` (denoted ‘Num ENV samples’), point size represents `xy_penalty` (denoted ‘Geo-dist penalty’), point colour represents `k2` (denoted ‘Num RS samples’), and panels represent values of `confidence` (denoted ‘LDC’). The benchmark-candidate search radius was fixed at `radius_km = 200` for the analysis shown, and the Cauchy distance-weighting setting was held constant at `lambda = 2.0`. Source: Valavi *et al.* (2025a).

The selected HCAS v3.3 production settings (summarised in Table 5) were:

- Benchmark-candidate search radius: `radius_km` = 200
- first-stage benchmark-reference retention: `k1` = 70
- final benchmark-reference retention: `k2` = 10
- geographic distance penalty: `xy_penalty` = 30
- Cauchy distance-weighting setting: `lambda` = 2.0
- Condition-calculation setting: `confidence` = 0.5.

An interactive inspection tool was also used during development of HCAS v3.3 to examine how alternative benchmarking settings affected target-specific reference-site selection and pre-calibration condition estimates (Figure 7). The tool displayed the target location, spatial candidate sites, retained benchmark reference sites and the positions of target–reference pairs on the inferential reference surface. It provided a complementary diagnostic for inspecting individual locations but was separate from the sequential AUC-based parameter-selection analysis (Valavi *et al.* 2025a).

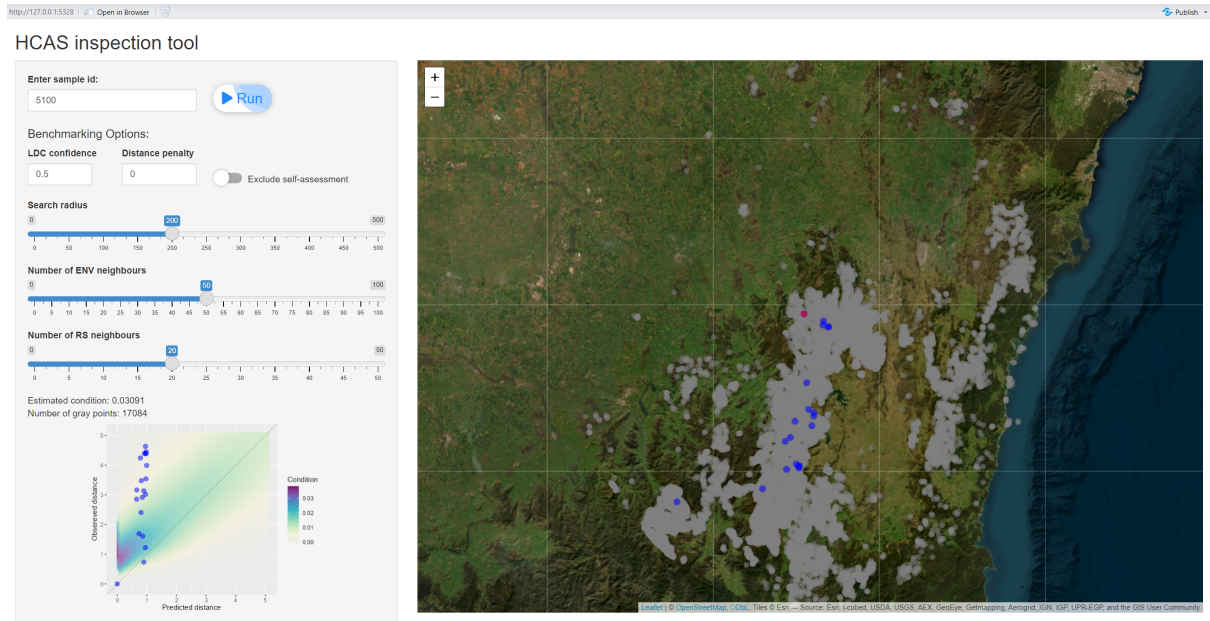


Figure 7 Interactive inspection of condition benchmarking for an example target location during development of HCAS v3.3. The map shows the target location, benchmark reference sites within the selected benchmark-candidate search radius and the final target-specific benchmark-reference sample retained using `k2`. The inferential-reference-surface panel shows the positions of the retained target–reference pairs in predicted and observed pairwise-distance space and the resulting pre-calibration condition estimate. In the example shown, `xy_penalty` = 0 (denoted ‘Distance penalty’), and self-assessment was enabled for diagnostic illustration, placing the target–self pair at the origin of the inferential reference surface. Self-assessment was excluded from HCAS v3.3 production condition estimation (`exclude_slef` = TRUE). Historical parameter labels are retained, as for Figure 6. Source: Valavi *et al.* (2025a).

S3.2 Calibration with ClassicHCAS

Calibration transformed the HCAS v3.3 pre-calibration condition estimates to a continuous 0–1 ecosystem condition scale. The transformation was applied using the `ClassicHCAS calibrate()` function and a monotone Hermite spline fitted through four calibration coordinates: minimum and maximum endpoints and two internal coordinates. The spline was implemented using `splinefun (method = "monoH.FC")` in R (R Core Team 2026), following the monotonic interpolation method of Fritsch and Carlson (1980). The calibration curve was derived from the HCAS v3.3 long-term epoch for 1988–2024 and subsequently applied to the HCAS v3.3 condition estimates.

Software-version note: To improve calibration robustness in future implementations of `ClassicHCAS`, an upper limit x - y coordinate (99.99th percentile, 0.9999) was introduced to support the end point (absolute maximum, 1), as outlined in the manuscript.

Each calibration coordinate comprised:

- an x -value representing an HCAS v3.3 pre-calibration condition estimate; and
- a corresponding y -value representing its assigned ecosystem condition score on the 0–1 scale.

The two internal x -values were derived from the distributions of pre-calibration condition estimates within mapped extents representing relatively high and low levels of ecosystem modification (Figure 8; Figure 9). The corresponding y -values were derived separately from expert-assigned ecosystem condition scores associated with conceptually aligned VAST classes (Table 3). Table 4 summarises the four coordinates used to calibrate HCAS v3.3.

Upper internal coordinate

The upper internal x -value was the median HCAS v3.3 pre-calibration condition estimate across the inferred reference-site extent compiled by Giljohann *et al.* (2025). This produced an x -value of 0.0207154.

The corresponding y -value of 0.8827 was derived from the expert-assigned condition score for the VAST Residual class reported by Giljohann *et al.* (2024). The Residual class was used because its description of unmodified natural regenerative capacity provided the closest conceptual alignment with the inferred high-integrity reference ecosystems used in HCAS. The x - and y -values therefore have related ecological interpretations but were derived from different evidence sources.

Lower internal coordinate

The lower internal x -value was derived from the pre-HMOD calibration dataset used in the HCAS v3.3 production workflow. This dataset combined inferred reference sites with categories from the 2020–21 National Land Use Map (ABARES 2024) cross-walked to ordinal ecosystem modification classes based on the VAST framework (Thackway & Lesslie 2006; Thackway & Lesslie 2008;

Giljohann *et al.* 2024). It was distinct from the separately derived highly modified-site dataset described by Giljohann *et al.* (2025).

Selected NLUM categories assigned to the VAST Replaced-Managed and Removed classes defined the lower calibration extent. The median HCAS v3.3 pre-calibration condition estimate within this combined extent produced an x -value of 0.0016073.

The corresponding y -value of 0.0641 was calculated as an area-weighted combination of the expert-derived scores for the VAST Replaced-Managed and Removed classes. The weights were derived from the respective proportional mapped extents of these classes in the pre-HMOD 2020–21 NLUM dataset cross-walked to VAST by Giljohann *et al.* (2024).

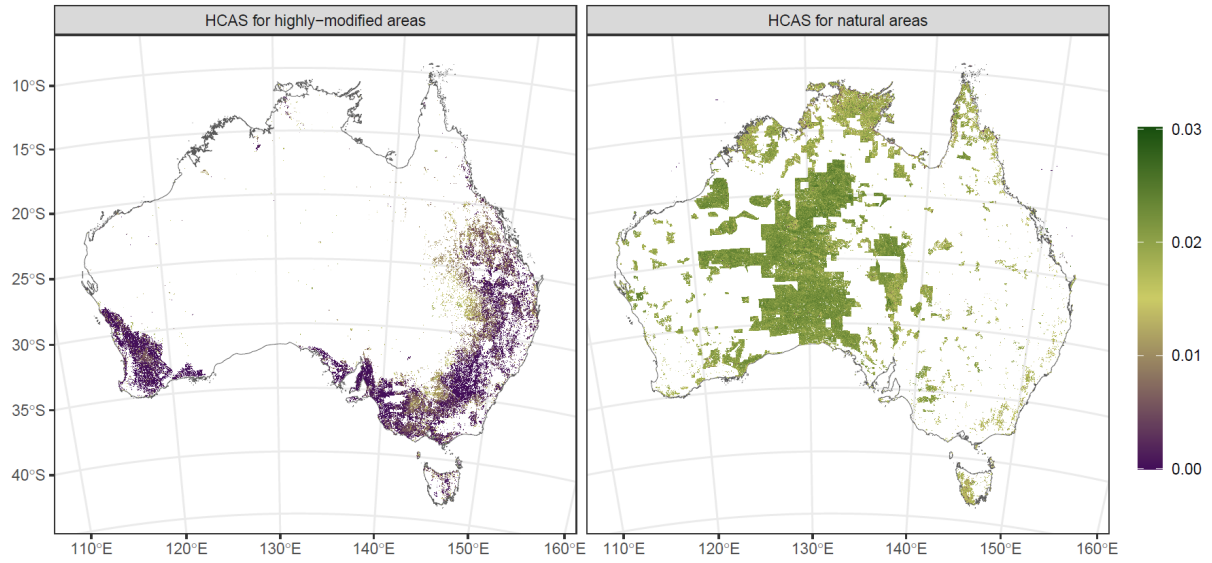


Figure 8 Mapped extents used to derive the two internal x -values for the HCAS v3.3 calibration. Left: selected categories from the 2020–21 National Land Use Map (ABARES 2024) cross-walked to the VAST Replaced-Managed and Removed classes following Giljohann *et al.* (2024). This pre-HMOD dataset defined the extent used to derive the lower internal x -value and was distinct from the separately inferred highly modified-site dataset (Section S2.2). Right: inferred reference sites compiled by Giljohann *et al.* (2025), used to derive the upper internal x -value. Map colours represent HCAS v3.3 long-term pre-calibration condition estimates and are linearly scaled from 0 to the maximum value of 0.0301316 using the HCAS colour palette (Williams *et al.* 2025b).

Endpoints and calibration function

The minimum endpoint mapped a pre-calibration value of 0 to an ecosystem condition score of 0. The maximum endpoint mapped the maximum HCAS v3.3 long-term pre-calibration value of 0.0301316 to an ecosystem condition score of 1.

The four coordinates—minimum endpoint, lower internal coordinate, upper internal coordinate and maximum endpoint (Table 4)—were ordered monotonically in both x and y . The fitted monotone Hermite spline therefore provided a continuous, monotonically increasing transformation between the HCAS v3.3 pre-calibration values and the specified 0–1 ecosystem condition scale (Figure 10). Higher calibrated values indicate greater alignment with ecologically comparable high-integrity reference ecosystems.

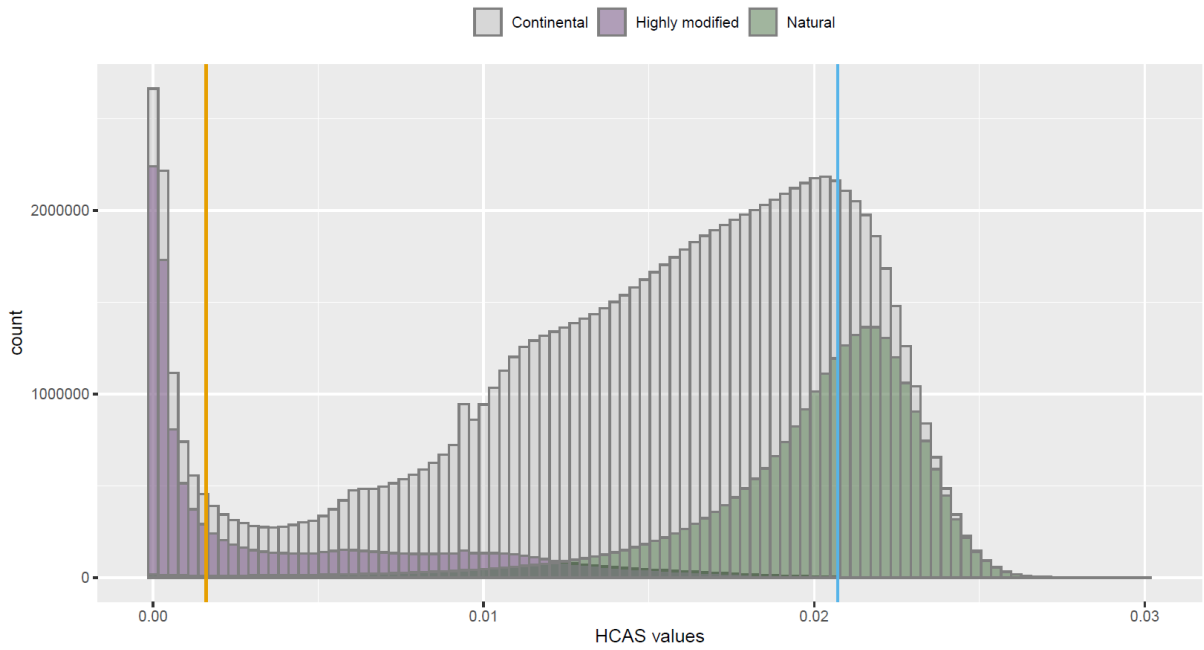


Figure 9 Distributions of HCAS v3.3 long-term pre-calibration condition estimates within the mapped extents used to derive the internal calibration x -values. Mauve bars show estimates within selected 2020–21 NLUM categories cross-walked to the VAST Replaced-Managed and Removed classes; green bars show estimates within inferred reference sites; and grey bars show the background distribution for the continental sample. Vertical lines identify the median estimate for the lower calibration extent (0.0016073) and inferred reference-site extent (0.0207154). Histograms were calculated from a 10% random sample of valid data pixels ($n = 96,728,199$) using 100 equal-width bins over the observed pre-calibration range of 0–0.0301316.

Table 3 Expert-derived ecosystem condition scores and class boundaries for VAST ecosystem modification classes. Scores are from the class-as-factor analysis reported by [Giljohann *et al.* \(2024\)](#). The HCAS v3.3 calibration used the Residual class score for the upper internal y -value and an area-weighted combination of the Replaced-Managed and Removed scores for the lower internal y -value. The remaining classes are shown to provide context for the expert-derived condition scale. VAST class descriptions follow [Thackway and Lesslie \(2006\)](#).

VAST CLASS	VAST DESCRIPTION: CURRENT REGENERATIVE CAPACITY (Thackway & Lesslie 2006; 2008)	CLASS AS FACTOR ANALYSIS		
		LOWER CLASS BOUNDARY	EXPERT-DERIVED CLASS SCORE	UPPER CLASS BOUNDARY
Residual	Natural regenerative capacity unmodified (includes naturally bare, ephemerals and lower plants)	≥ 0.74	0.88	≤ 1.0
Modified	Natural regeneration tolerates or endures under past and or current land management practices	≥ 0.49	0.60	< 0.74
Transformed	Natural regenerative capacity limited or at risk under past and or current land use or land management practices. Rehabilitation and restoration possible through modified land management practice	≥ 0.31	0.38	< 0.49
Replaced-Adventive	Regeneration of native vegetation community has been suppressed by ongoing disturbances of the natural regenerative capacity; limited potential for restoration	≥ 0.18	0.23	< 0.31
Replaced-Managed	Regeneration of native vegetation community lost or suppressed by intensive land management; limited potential for restoration	≥ 0.09	0.14	< 0.18
Removed	Native vegetation community removed	≥ 0.00	0.04	< 0.09

Table 4 Calibration coordinates used to transform HCAS v3.3 pre-calibration condition estimates to the 0–1 ecosystem condition scale.

COORDINATE POSITION	PRE-CALIBRATION VALUE (x)	CALIBRATED CONDITION SCORE (y)	PROVENANCE
Minimum endpoint	0.0000000	0.0000	Maximum HCAS v3.3 long-term pre-calibration estimate mapped to condition 1
Lower internal coordinate	0.0016073	0.0641	<i>x</i> : median pre-calibration estimate across selected 2020–21 NLUM categories cross-walked to VAST Replaced-Managed and Removed classes. <i>y</i> : area-weighted combination of expert-derived Replaced-Managed and Removed scores
Upper internal coordinate	0.0207154	0.8827	<i>x</i> : median pre-calibration estimate across inferred reference sites. <i>y</i> : expert-derived VAST Residual class score
Maximum endpoint	0.0301316	1.0000	Minimum of the pre-calibration scale mapped to zero condition

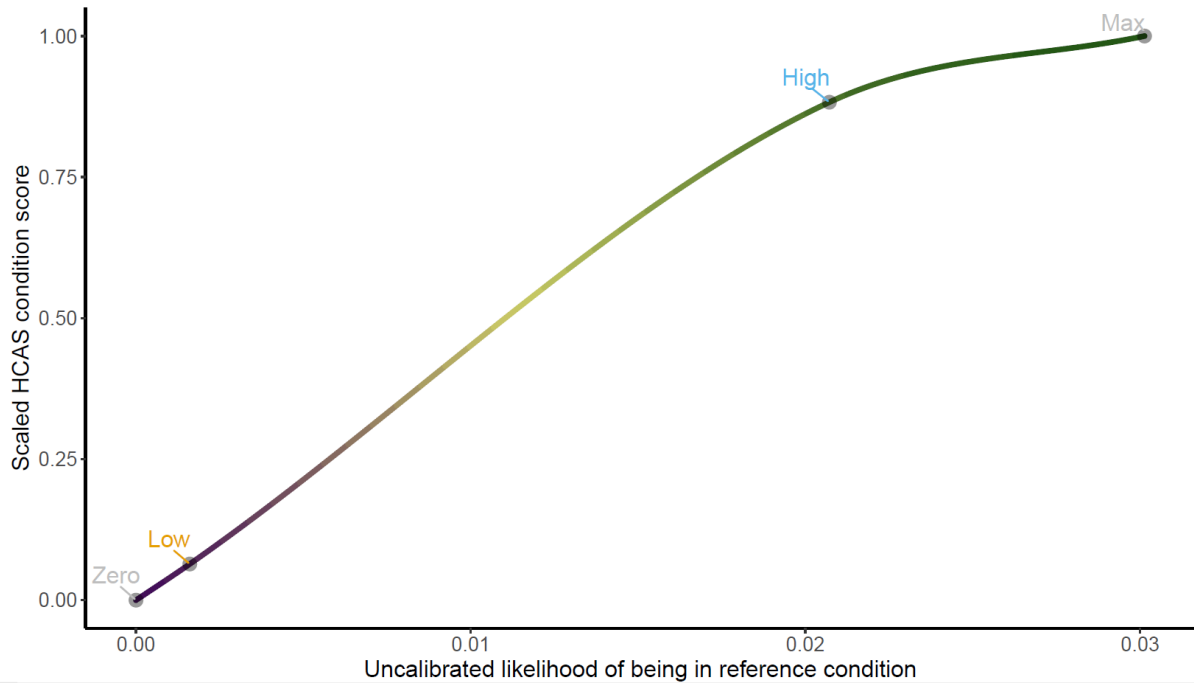


Figure 10 Monotonic calibration function used to transform HCAS v3.3 long-term pre-calibration estimates to the 0–1 ecosystem condition scale. The curve is a monotone Hermite spline fitted through the minimum endpoint (0,0), lower internal coordinate (0.0016073,0.0641), upper internal coordinate (0.0207154,0.8827) and maximum endpoint (0.0301316,1). The spline was implemented using `splinefun(method = 'monoH.FC')` following [Fritsch and Carlson \(1980\)](#). Curve colours follow the HCAS colour palette ([Williams *et al.* 2025b](#)). Colours are consistent with related phenomena in Figure 8 and Figure 9.

S3.3 Summary of HCAS v3.3 workflow configuration and ClassicHCAS settings

The external workflow configuration, inputs and ClassicHCAS function settings used to generate the HCAS v3.3 long-term condition estimate are summarised in Table 5.

The resulting spatial pattern of calibrated ecosystem condition for the 1988–2024 long-term epoch is shown in Figure 11. Temporal estimation is described in Section S3.4, and bootstrap uncertainty associated with reference-site sampling is described in Section S3.5.

Table 5 External workflow configuration and ClassicHCAS function settings used to generate HCAS v3.3, expressed using the current API (version 2.0.0). Original production values and corresponding version 0.2.0 terminology are retained, with semantic differences noted. Settings or inputs determined outside ClassicHCAS are identified explicitly.

WORKFLOW COMPONENT	CURRENT CLASSICHCAS FUNCTION / ARGUMENT	HCAS v3.3 PRODUCTION VALUE	ROLE IN HCAS V3.3	V0.2.0 TERMINOLOGY / VERSION NOTES
Production metadata	External to ClassicHCAS: software version	ClassicHCAS 0.2.0	Package version used for the production condition calculation	Current repository version: 2.0.0; not directly output-equivalent because several algorithms and interfaces have changed
Production metadata	External to ClassicHCAS: production identifier	HCAS v3.3; PROD3	Internal identifier for the published production run, combines as HCAS v3.3	External workflow metadata
Analysis grid	External to ClassicHCAS: grid resolution	90 × 90 m	Resolution of the common raster analysis grid	External workflow configuration
Training reference data; Reference ecosystem modelling	External to ClassicHCAS: training reference sites	488,619	Representative sample of training reference sites attributed with observed remotely sensed variables and environmental covariates for use in reference ecosystem models (REM)	User-supplied reference data; current full tables require x, y, predicted variables, then observed variables
Reference ecosystem modelling	External to ClassicHCAS: environmental covariates	46	Environmental covariates to attribute values in training reference sites used in reference ecosystem models (REM)	REM remains external to ClassicHCAS

WORKFLOW COMPONENT	CURRENT CLASSICHCAS FUNCTION / ARGUMENT	HCAS v3.3 PRODUCTION VALUE	ROLE IN HCAS V3.3	V0.2.0 TERMINOLOGY / VERSION NOTES
Benchmark reference data	External to ClassicHCAS: benchmark reference sites	590,342	Representative sample of candidate benchmark reference sites used during target-location-specific condition benchmarking	User-supplied benchmark data
Ecosystem characteristics	External to ClassicHCAS: observed variables	14	Spatially explicit observed remotely sensed variables, used to attribute values in training reference sites used in reference ecosystem models (REM), and to calculate observed pairwise distances	User-supplied observed layers; current rasters require predicted layers before observed layers
Ecosystem characteristics	External to ClassicHCAS: REM-predicted variables	14	Spatially explicit predicted remotely sensed variables, used to calculate predicted pairwise distances	User-supplied predicted layers
Reference density surface	<code>ref_density(radius_km)</code>	1,000 km	Maximum training reference site-pair search radius used for reference density surface construction	<code>histogram(radius_km)</code> ; current code counts each pair once rather than twice; this scaling does not affect final result due to normalisation
Reference density surface	<code>ref_density(bin_num)</code>	650	The maximum dimension of the reference-density surface specified by the number of bins used to classify the observed and predicted distances for training reference site-pairs	<code>histogram(bin_num)</code>
Reference density surface	<code>ref_density(bin_width)</code>	0.05	The equidistant width of each class used to bin the predicted and observed distances for training reference site-pairs	<code>histogram(bin_width)</code>
Reference density surface	<code>ref_density(drop_features)</code>	NULL	All observed and predicted remotely sensed variables retained during reference density surface construction	<code>histogram(drop_features)</code> ; excluded variables are now removed from the active feature set rather than zeroed
Inferential reference surface	<code>normalise(trim_size)</code>	400	Restricted construction of the inferential reference surface to the selected 400 × 400-bin extent of the input reference density surface	<code>normalise(trim_size)</code> ; unchanged argument

WORKFLOW COMPONENT	CURRENT CLASSICHCAS FUNCTION / ARGUMENT	HCAS v3.3 PRODUCTION VALUE	ROLE IN HCAS V3.3	V0.2.0 TERMINOLOGY / VERSION NOTES
Inferential reference surface	<code>normalise(offset)</code>	0	No additional near-origin bins of the input reference density surface excluded using this setting	<code>normalise(offset)</code> ; unchanged argument
Inferential reference surface	<code>normalise(legacy)</code>	FALSE	Applied the revised normalisation method introduced in version 0.2.0 to reduce edge effects during inferential reference surface construction	<code>normalise(legacy)</code> ; FALSE still selects the R normalisation method
Inferential-surface lookup	<code>benchmark(interpolate)</code>	TRUE	Interpolated between inferential reference-surface cells during benchmarking	<code>benchmark(interpolate)</code> ; still supported for bilinear surface lookup
Benchmark candidate selection	<code>benchmark(radius_km)</code>	200 km	Maximum target-to-reference search radius used to identify spatial benchmark reference site candidates	<code>benchmark(radius_km)</code> ; geographic-distance calculations have changed
Benchmark candidate selection	<code>benchmark(k1)</code>	70	Retained up to 70 candidate benchmark reference sites with the smallest adjusted predicted pairwise distances	<code>benchmark(k_pred)</code> ; same first-stage filter role
Benchmark candidate selection	<code>benchmark(xy_penalty)</code>	30	Geographic-distance penalty applied during first stage ranking of candidate benchmark reference sites	<code>benchmark(xy_penalty)</code> ; unchanged argument
Benchmark candidate selection	<code>benchmark(k2)</code>	10	Retained up to 10 benchmark reference sites to pair with the specific target location	<code>benchmark(k_obs)</code> ; same second-stage filter role
Condition calculation	<code>benchmark(lambda, kernel)</code>	2.0	Scale setting for the Cauchy-type predicted-distance weighting applied to the retained benchmark reference sites for the target	<code>benchmark(lambda)</code> ; no selectable kernel. Current <code>kernel</code> = "Cauchy" omits the production weighting's inverse-distance factor; it is not equivalent. <code>kernel</code> = "Gaussian" is the current default

WORKFLOW COMPONENT	CURRENT CLASSICHCAS FUNCTION / ARGUMENT	HCAS v3.3 PRODUCTION VALUE	ROLE IN HCAS V3.3	V0.2.0 TERMINOLOGY / VERSION NOTES
Condition calculation	<code>benchmark(confidence)</code>	0.5	Controlled the balance between the maximum retained inferential reference surface value and predicted distance-weighted mean in the condition calculation for the observed and predicted target location distances to retained benchmark reference sites	<code>benchmark(confidence)</code> ; current <code>boost = NULL</code> or <code>NA</code> activates this blend. Default <code>boost = k2</code> instead multiplies the highest-probability site's kernel weight; no fixed one-to-one mapping
Self-assessment control	<code>benchmark(exclude_slef)</code>	TRUE	Prevents a benchmark reference site also being a target location from benchmarking against itself	<code>benchmark(exclude_slef)</code> ; this API spelling is unchanged from v0.2.0. Excludes predicted distances below the effective <code>bin_width</code> , not only identical sites
Alternative feature sets	<code>benchmark(drop_features)</code>	NULL	All 14 remotely sensed variables retained during benchmarking; non-NULL values allow specified variables to be excluded	<code>benchmark(drop_features)</code> ; excluded variables are now removed from the active feature set rather than zeroed
Calibration-coordinate derivation	External to ClassicHCAS: lower internal x -value	0.0016073	Median pre-calibration estimate across selected 2020–21 NLUM categories cross-walked to VAST Replaced-Managed and Removed classes	Externally derived and supplied to <code>calibrate()</code>
Calibration-coordinate derivation	External to ClassicHCAS: lower internal y -value	0.0641	Area-weighted combination of expert-derived scores for VAST Replaced-Managed and Removed, using their mapped 2020–21 NLUM categories cross-walked to VAST Replaced-Managed and Removed classes	Externally derived and supplied to <code>calibrate()</code>
Calibration-coordinate derivation	External to ClassicHCAS: upper internal x -value	0.0207154	Median pre-calibration estimate across the mapped extent of inferred reference sites equated with the VAST-Residual class	Externally derived and supplied to <code>calibrate()</code>
Calibration-coordinate derivation	External to ClassicHCAS: upper internal y -value	0.8827	Expert-derived condition score aligned with the VAST Residual class	Externally derived and supplied to <code>calibrate()</code>

WORKFLOW COMPONENT	CURRENT CLASSICHCAS FUNCTION / ARGUMENT	HCAS v3.3 PRODUCTION VALUE	ROLE IN HCAS V3.3	V0.2.0 TERMINOLOGY / VERSION NOTES
Calibration operation	<code>calibrate(x_values)</code>	0; 0.0016073; 0.0207154; 0.0301316	Supplied pre-calibration coordinates defining the minimum endpoint, two internal coordinates and maximum endpoint	<code>calibrate(x_values)</code> ; unchanged argument
Calibration operation	<code>calibrate(y_values)</code>	0; 0.0641; 0.8827; 1	Supplied condition scores corresponding to the four x -coordinates	<code>calibrate(y_values)</code> ; unchanged argument
Calibration operation	<code>calibrate()</code> monotone spline	Monotone Hermite spline	Applied the supplied coordinates to transform pre-calibration estimates to the 0–1 condition scale; ClassicHCAS did not derive the coordinates	<code>calibrate()</code> monotone spline; v0.2.0 interpolated a sampled spline curve, whereas the current function evaluates the fitted spline directly. <code>interpolate</code> remains a <code>benchmark()</code> argument
Computational execution	<code>tiling()</code> / external tiled HPC workflow	Distributed tiled processing	Partitioned and executed continental benchmarking through the external production workflow	No <code>tiling()</code> in v0.2.0; added in version 1.0.0. Production tiling and distributed execution were external; current <code>tiling()</code> creates tiles but does not dispatch HPC jobs

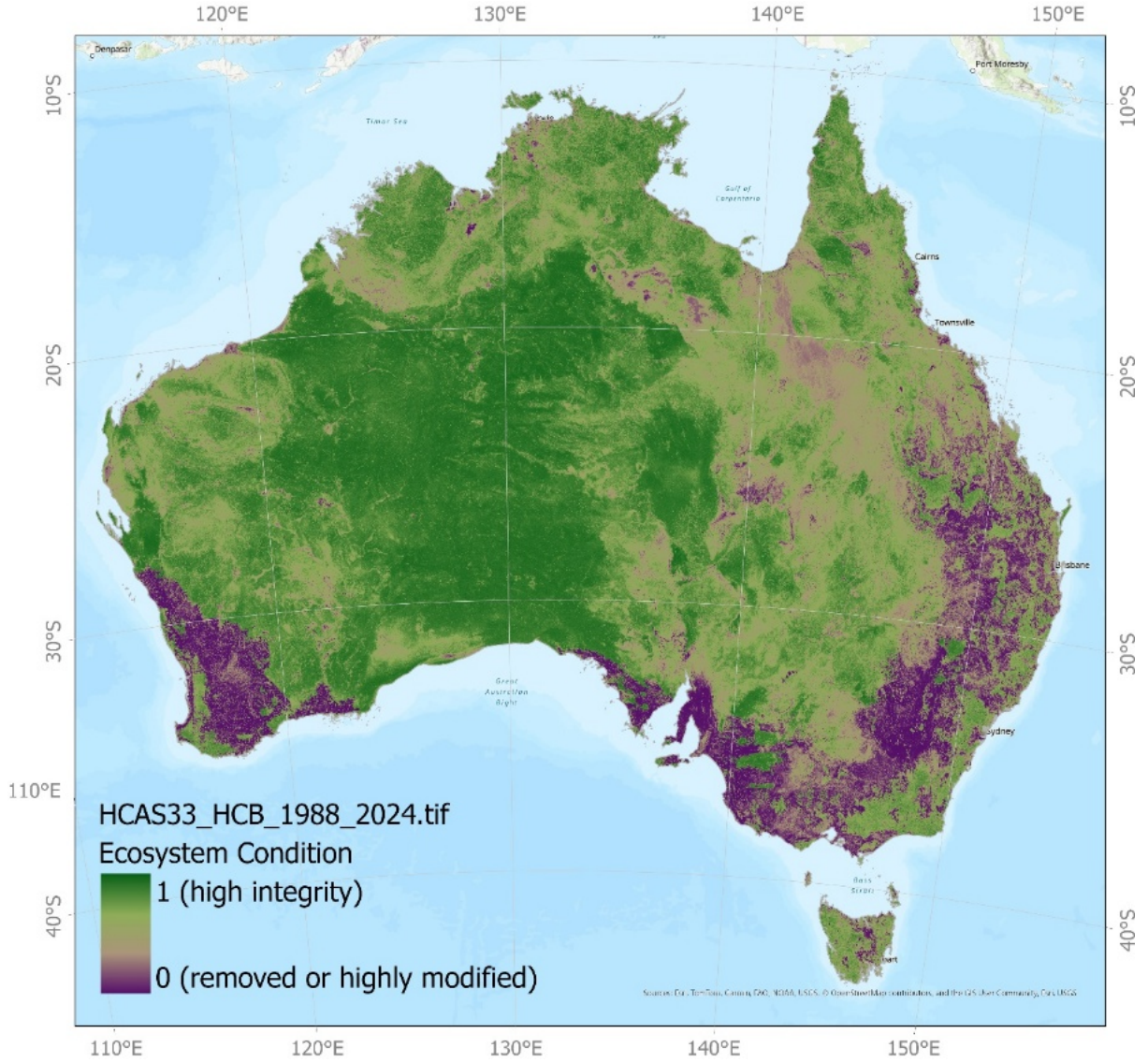


Figure 11 Ecosystem condition estimates for continental Australia generated from the HCAS v3.3 long-term epoch for 1988–2024. Condition scores are calibrated to the 0–1 scale described in Section S3.2. Source: HCAS33_HCB_1988_2024.tif in the HCAS v3.3 data collection (Valavi *et al.* 2025b).

S3.4 Temporal condition estimation

The model configuration and parameters derived from the long-term epoch—including the reference ecosystem model, inferential reference surface, benchmarking settings and long-term calibration curve—are held constant (Table 5) and applied across all short-term temporal epochs. The only part that varies is the observed remotely sensed ecosystem condition variables. These influence the benchmark reference sites retained for each target location and their observed target-reference distances when reading likelihood values from the inferential reference surface for condition estimation (as illustrated in Figure 12 to Figure 15).

S3.4.1 Gain rescaling of pre-calibration condition estimates

Prior to calibration, and external to **ClassicHCAS** functionality, pre-calibration condition estimates for each short-term epoch are adjusted using an epoch-specific gain (multiplier). Gain rescaling adjusts the magnitude of pre-calibration condition estimates across temporal epochs. The need for this adjustment arises when the observed remotely sensed variables of reference ecosystems summarised over shorter time windows (e.g. 3-year epochs) differ in their statistical properties from the long-term summaries used to establish the reference framework for condition benchmarking. Gain rescaling is intended to reduce these systematic differences by adjusting the amplitude of pre-calibration condition values for each short-term epoch relative to the long-term epoch, while preserving the relative ranking of sites within each short-term epoch in a temporal series.

This process is expressed mathematically as follows. For temporal short-term epoch j , let $H_{\mathcal{R},j} = \{h_j(p) : p \in \mathcal{R}\}$, where here h_j is the pre-calibration condition estimate for that epoch. Here, $\mathcal{R}$ is a common random sample of 200,000 VAST class-1 locations (reference sites) used across temporal epochs. Note that $\mathcal{R}$ is presumed to be in reference condition for the entire temporal series. Let $H_{\mathcal{R}}$ denote the corresponding long-term pre-calibration condition estimates, and let C denote the fixed long-term calibration function, applied to each estimate. Because of sensor and other data-related effects, the mean of the calibrated condition estimate of a short-term epoch often does not match the mean of the calibrated condition estimate for the long-term estimate, $\overline{C(H_{\mathcal{R}})}$. That is, we frequently observe that $\overline{C(H_{\mathcal{R},j})} \neq \overline{C(H_{\mathcal{R}})}$. To minimise this difference, we select a gain to apply to the pre-calibration condition estimates of each short-term epoch:

$$\gamma_j = \arg \min_{0.5 \leq \gamma \leq 1.5} \left| \overline{C(\gamma H_{\mathcal{R},j})} - \overline{C(H_{\mathcal{R}})} \right|$$

The gain is estimated using `stats::optimise()` and applied to all pre-calibration condition estimates for that epoch as $\tilde{h}_j(p) = \gamma_j h_j(p)$. The means are calculated over available sample values, excluding missing values.

This aligns, as closely as possible, the mean calibrated condition score over the common sample of inferred reference sites (from Section S2.2) between each short-term epoch and the long-term epoch used to establish the condition benchmarking framework, using a common calibration curve. This adjustment was applied to the HCAS v3.3 pre-calibration condition estimates across all short-term temporal epochs ([Valavi et al. 2025a](#); [Williams et al. 2025b](#)).

S3.4.2 Calibration and temporal comparability

Following gain rescaling, the calibration curve derived from the long-term epoch (Section S3.2; Table 5) is applied consistently to all short-term epochs. Applying the same calibration curve maintains a common 0–1 reporting scale across epochs, allowing temporal differences in condition estimates to be interpreted relative to a consistent calibration framework. This improves comparability among epochs but does not remove residual effects associated with sensor characteristics,

seasonal variability or other temporal artefacts in the underlying observations.

S3.4.3 Temporal patterns of condition estimates

The HCAS v3.3 short-term epochs provide repeated estimates of ecosystem condition generated using the fixed long-term reference ecosystem models, reference surfaces, benchmarking configuration and calibration curve described in Sections S3.1–S3.4.2. Variation among epochs arises from changes in the remotely sensed ecosystem characteristics summarised within each overlapping 3-year period. This variation may include the residual effects of responses to weather, disturbance and recovery, as well as observational artefacts, and does not by itself demonstrate persistent or statistically supported change (Williams *et al.* 2025b).

Figure 12 shows variation in the continental mean HCAS v3.3 condition estimate across the overlapping 3-year epochs, together with the mean of the long-term 1988–2024 condition product. Variation among the short-term continental means is relatively small. This continental summary is descriptive and may mask stronger, spatially restricted changes or offsetting patterns among regions. Persistent directional change can be evaluated separately using methods designed for serially correlated time series (Lehmann *et al.* 2023), while statistical comparison of selected epoch means can be undertaken using the method described by (Lehmann & Williams 2025). Broader limitations affecting temporal interpretation are discussed in Section 5 of the manuscript.

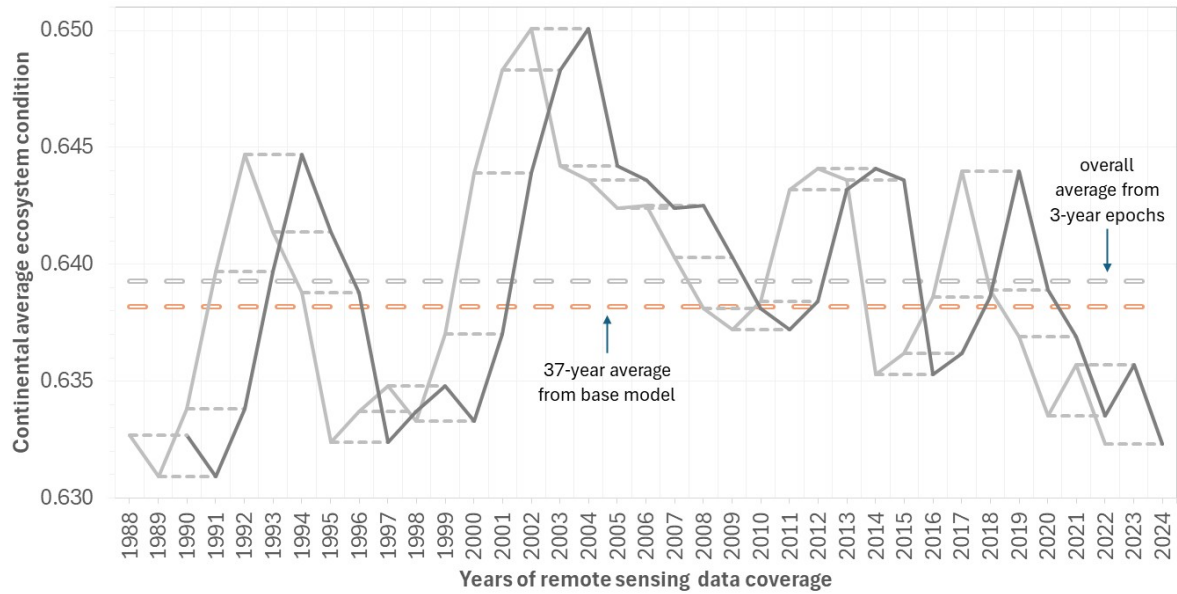


Figure 12 Continental mean ecosystem condition estimated from overlapping 3-year HCAS v3.3 epochs, compared with the mean of the long-term 1988–2024 condition product. Each short-term epoch is commonly referred to by the final year of its 3-year period. The truncated vertical axis highlights the relatively small variation among the plotted continental means and should be considered when interpreting the apparent magnitude of change. The figure provides a descriptive continental summary and does not show the results of formal trend or epoch-comparison testing. Sources: HCAS v3.3 data collection and Product Guide (Valavi *et al.* 2025b; Williams *et al.* 2025b).

Individual short-term epochs retain spatial variation in condition at the 90 m analysis resolution, as illustrated for 2022–2024 in Figure 13. Fine-scale spatial patterns are also evident in the long-

HCAS33_AHC_2022_2024.tif
Ecosystem Condition

1 (high integrity)

0 (removed or highly modified)

Sources: Cori, Tim-Tam, Cairns, FMO, IGMA, USCS, © OpenStreetMap contributors, and The-GIS-Lovers Community, Cori, USGS

34

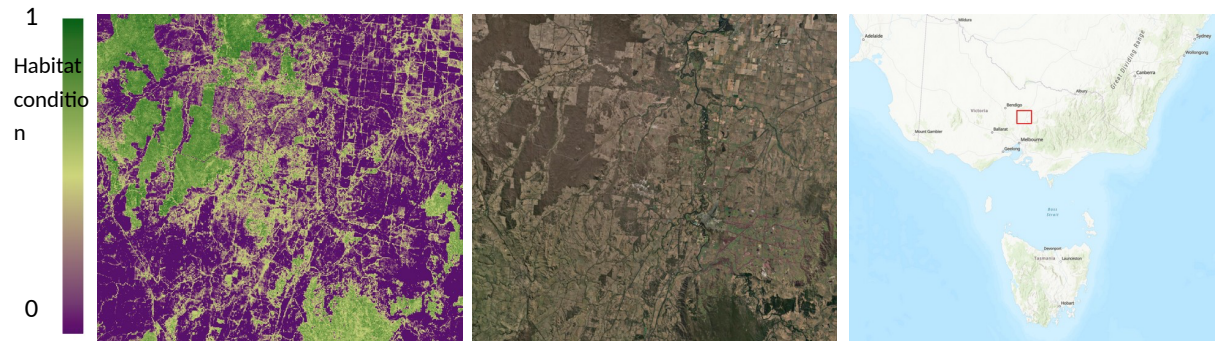


Figure 14 Fine-scale spatial pattern in the HCAS v3.3 long-term ecosystem condition product for an example landscape in central Victoria, Australia. Left: ecosystem condition estimates for the 1988–2024 long-term epoch, with lower values shown in dark purple and higher values in green. Middle: corresponding Landsat imagery. Right: location of the example landscape within Australia. Sources: HCAS condition data, HCAS33_HCB_1988_2024.tif, in [Valavi *et al.* \(2025b\)](#); Landsat imagery: ; location-map basemap: .

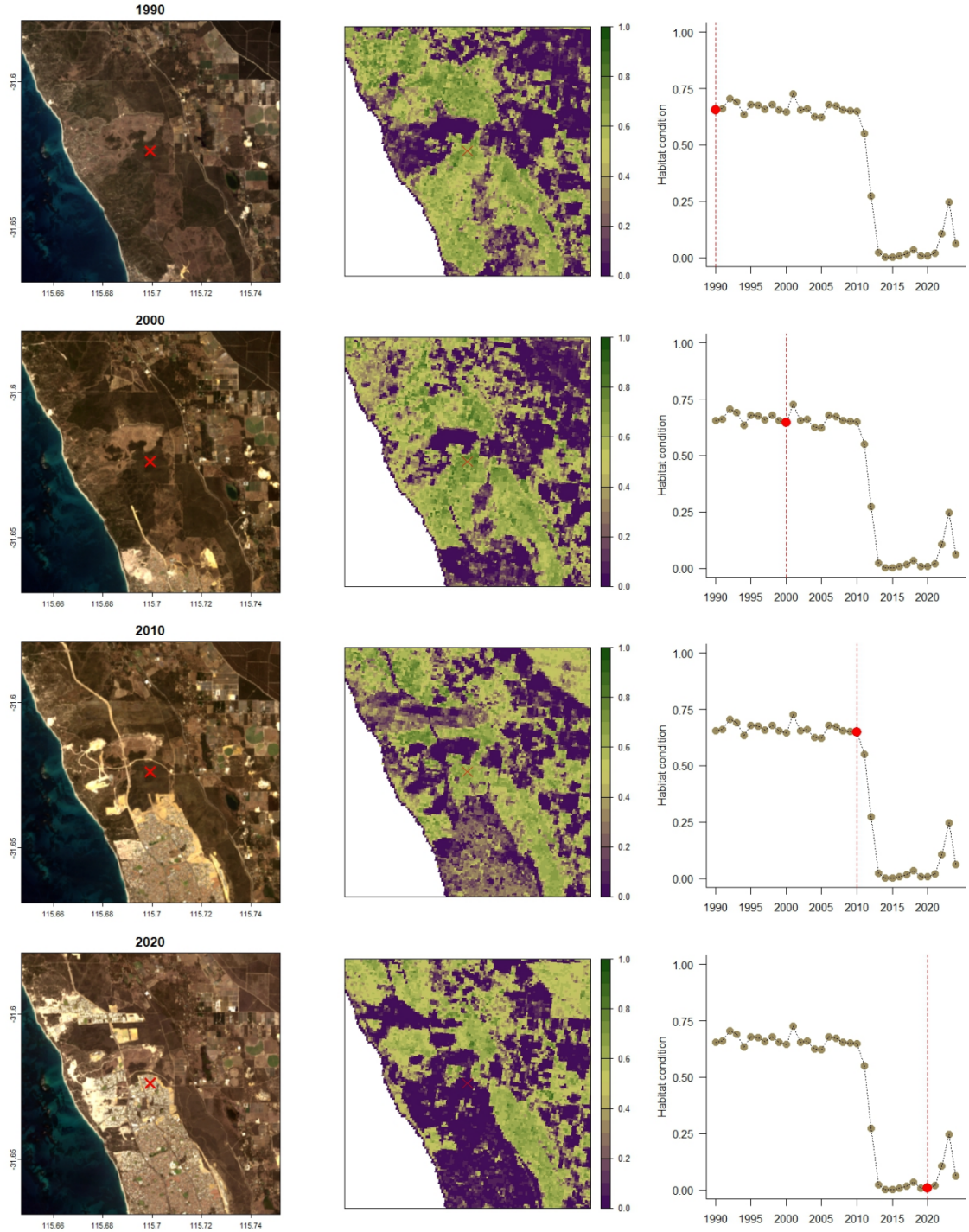


Figure 15 Temporal ecosystem condition estimates for a rapidly urbanising example landscape north of Perth, Western Australia, from 1990 to 2024. HCAS condition estimates are derived from overlapping 3-year epochs and show a sustained decline corresponding spatially and temporally with urban expansion visible in the accompanying imagery. Because adjacent epochs share two years of observations, abrupt changes are smoothed across successive estimates and their full expression may occur after the initial landscape change. Sources: HCAS v3.3 short-term ecosystem condition rasters, HCAS33_HC_1988_1990.tif, HCAS33_HC_1989_1991.tif, ... , HCAS33_HC_2022_2024.tif in [Valavi et al. \(2025b\)](#); imagery: .

S3.5 Bootstrap uncertainty from reference-site resampling

We evaluated the uncertainty in HCAS v3.3 ecosystem condition estimates due to the reference-site samples used in reference ecosystem modelling and condition benchmarking using a bootstrap resampling approach (Efron & Tibshirani 1994). One hundred bootstrap replicates were generated by resampling, with replacement, the training reference sites used to fit the reference ecosystem model and the benchmark reference sites used for target-specific condition estimation (Lehmann *et al.* 2025). The two reference site sets were resampled from their respective production samples (Section S2.5). These bootstrap processes were statistically independent of each other (Lehmann *et al.* 2025). For an N -out-of- N bootstrap sample, approximately 63% of the sampled reference site locations are expected to be unique, with the remaining draws comprising repeated selections of those locations.

For each replicate, the reference ecosystem model was refitted using the resampled training reference sites. The reference-density distribution and inferential reference surface were then reconstructed, condition benchmarking was rerun using the resampled benchmark reference sites, and the resulting condition estimates were calibrated using a replicate-specific calibration curve. The HCAS v3.3 input rasters (remotely sensed variables and environmental covariates), variable definitions, calibration method and other production settings were otherwise treated as fixed. This procedure generated one calibrated ecosystem condition estimate at each grid cell for each bootstrap replicate (Figure 16).

At each grid cell, the 2.5th and 97.5th percentiles of the 100 calibrated bootstrap estimates were used as the lower and upper limits of a 95% bootstrap percentile interval. Bootstrap interval width was calculated as the upper limit minus the lower limit. Figure 17 presents the long-term calibrated ecosystem condition estimate from the full HCAS v3.3 production run and the corresponding bootstrap interval width. The production run estimate is not the mean of the bootstrap replicates, but rather the HCAS output corresponding to the full (non-resampled) set of training and benchmark site data. Bootstrap resampling was also applied to selected temporal epochs and included in the HCAS v3.3 data collection (Valavi *et al.* 2025b; Williams *et al.* 2025b).

The bootstrap distribution quantifies the model-based uncertainty in the estimated ecosystem condition due to resampling the training and benchmark reference sites and to the effects of that resampling propagated through the refitted or recalculated workflow components. It should not be interpreted as a comprehensive estimate of uncertainty in ecosystem condition. In particular, the procedure does not independently propagate Earth observation measurement error, uncertainty in environmental covariates, model structural or functional-form uncertainty, assumptions embodied in the calibration method, residual temporal or sensor-related artefacts, or uncertainty in other fixed inputs and production settings. These sources may affect the condition estimates but are not represented as distinct uncertainty components unless their effects vary as a consequence of reference site resampling.

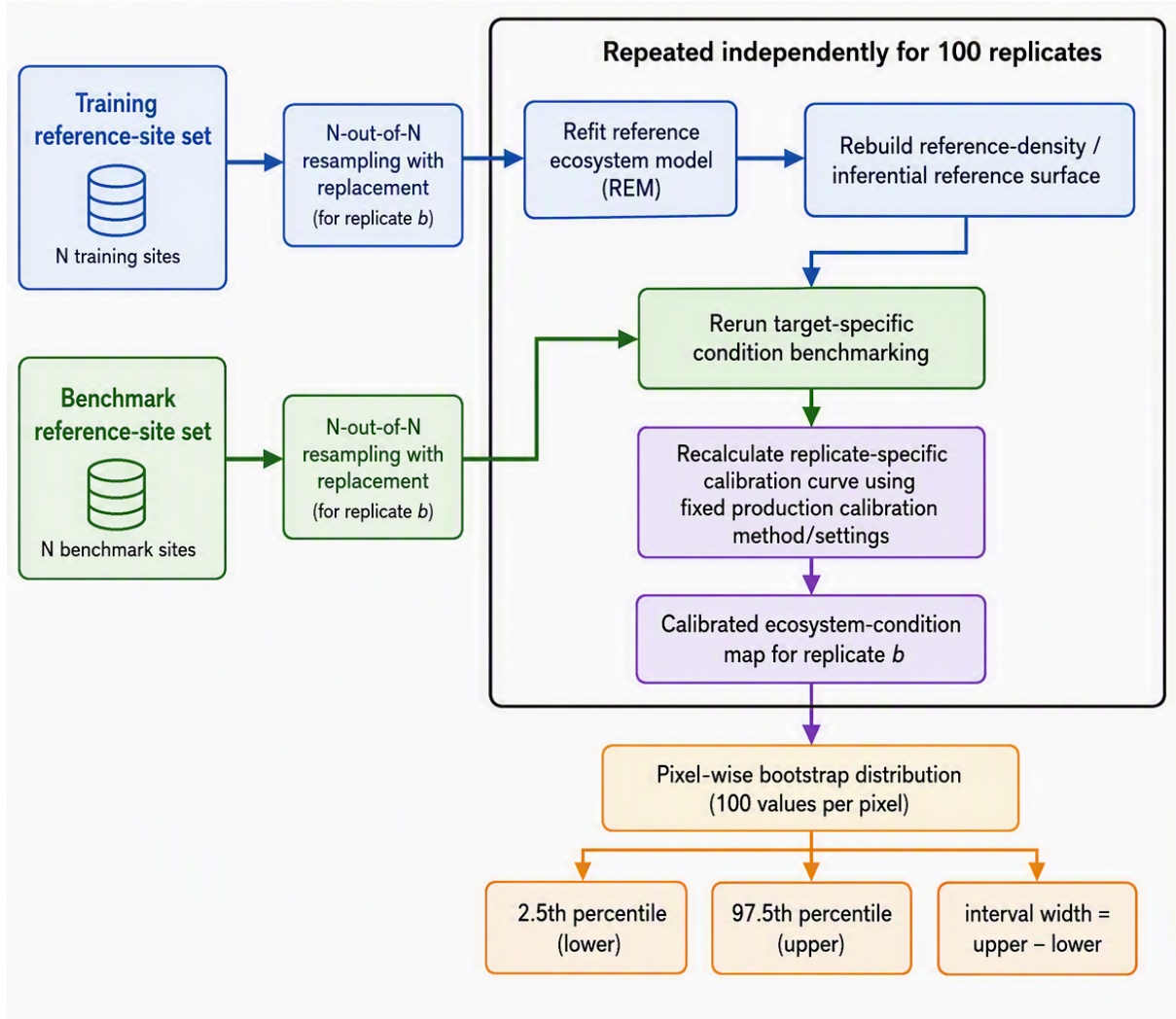


Figure 16 Bootstrap workflow for HCAS v3.3 uncertainty products. The 95% confidence interval width summarises uncertainty due to resampling the training and benchmark reference sites through the rerun of the HCAS workflow; it is not a comprehensive propagation of all uncertainty sources. AI-assisted schematic prepared under author supervision using OpenAI image-generation tools.

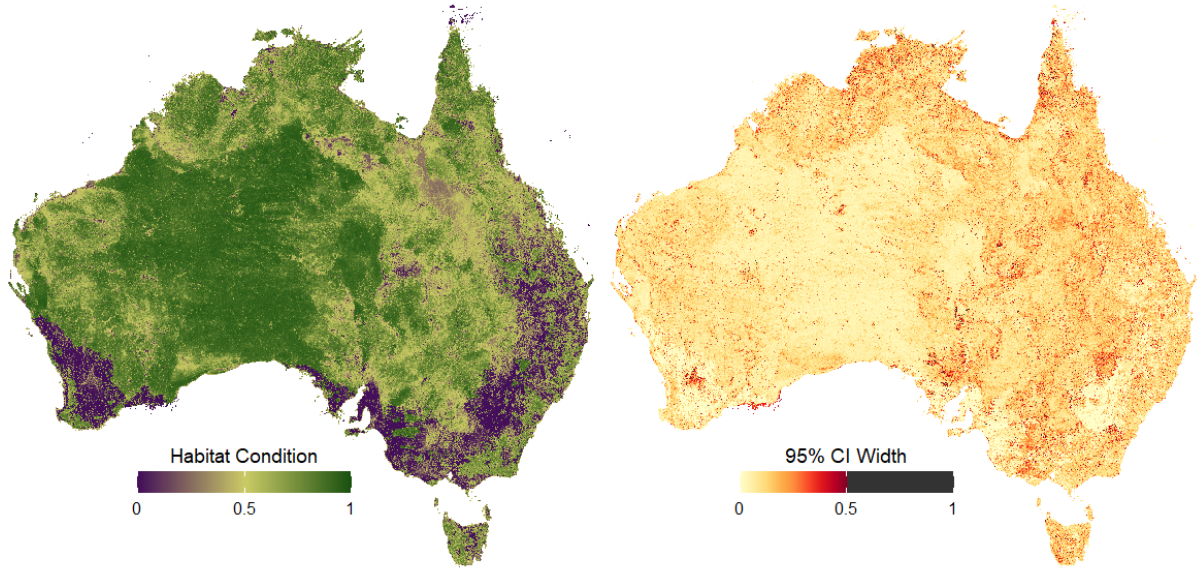


Figure 17 The long-term calibrated ecosystem condition estimates (left), and the 95% confidence interval (CI) width derived from 100 rounds of bootstrap resampling of training and benchmark reference sites. Data: HCAS33_HCB_1988_2024.tif (left) and HCAS33_HCB_CIW_1988_2024.tif in [Valavi et al. \(2025b\)](#)

S3.5 Benchmarking support diagnostic

When requested, `ClassicHCAS` benchmarking also returns the support diagnostic: $SU_t = \log \left(\sum_{j=1}^{m_t} \tilde{w}_{jt} \right)$, which is calculated from the original, unboosted kernel weights resulting from the condition estimation (see Eq. 10 in the manuscript). This diagnostic summarises the availability and ecological proximity provided by the k_2 benchmark reference sites for each target location but is not a calibrated confidence interval (see Section S3.5). It is referred to as an index of benchmark reference site coverage in the HCAS v3.3 Product Guide ([Williams et al. 2025b](#)) and provided as a dataset in the data collection ([Valavi et al. 2025b](#)).

S4 Evaluation of the Australian HCAS v3.3 application

This section presents two complementary evaluations relevant to the continental Australian HCAS v3.3 application. First, the inferred reference-site dataset used in reference ecosystem modelling and condition benchmarking is assessed against independent field-based floristic observations. Second, HCAS v3.3 ecosystem condition estimates are compared with Queensland Spatial BioCondition v2, an independently developed regional vegetation-condition product. General limitations and sources of uncertainty affecting the HCAS workflow are discussed in Section 5 of the main manuscript (further details are provided in [Williams et al. 2025b](#)).

S4.1 Evaluation of inferred reference sites

The suitability of inferred reference sites affects both reference ecosystem modelling and condition benchmarking. [Giljohann et al. \(2025\)](#) evaluated the Australian inferred reference-site dataset

against independent field-based floristic observations from the Harmonised Australian Vegetation Plot dataset (HAVPlot). HAVPlot is a continental collation of more than 200,000 vegetation survey plots compiled from major Australian vegetation-data custodians (Mokany *et al.* 2022a; Mokany *et al.* 2022b).

Regression analyses considered native-species proportion, survey year, plot size and species richness. Inferred reference sites were preferentially associated with vegetation plots exhibiting higher relative naturalness. This result provides independent supporting evidence that the site-inference method preferentially identifies relatively natural vegetation. It does not demonstrate that every inferred reference site is in reference condition, establish equivalent coverage of all Australian ecosystems, or independently validate the HCAS v3.3 ecosystem condition estimates.

National evaluation of HCAS condition estimates remains constrained by the limited availability and comparability of field-based ecosystem condition observations. Earlier HCAS implementations were compared with compiled field observations and structured expert assessments (Williams *et al.* 2021a; Williams *et al.* 2023), but differences in variables, reference concepts, assessment methods and spatial support among datasets restrict their consistent application at continental scale. Structured expert assessments collected using the Habitat Condition Assessment Tool provide a potential source of site-level evidence for future evaluation of HCAS condition estimates and inferred reference ecosystems (White *et al.* 2019; White *et al.* 2023; Munroe *et al.* 2025; Munroe *et al.* 2026). These data were not used to evaluate the HCAS v3.3 application reported here.

S4.2 Comparison with Queensland Spatial BioCondition

Queensland Spatial BioCondition version 2 is an independently developed modelling and mapping framework that predicts vegetation condition for biodiversity relative to Regional Ecosystem specific reference or ‘best-on-offer’ states (DES 2021). The version used here provides predictions for 2021 across the Brigalow Belt, Central Queensland Coast and Southeast Queensland product regions on a 0–100 scale (DESI 2024a; DESI 2024d; DESI 2024b). For comparison with HCAS v3.3, the Spatial BioCondition predictions were rescaled to 0–1 and paired with the HCAS v3.3 short-term epoch summarising 2019–2021.

The comparison covered the common mapped extent of Brigalow Belt North, Brigalow Belt South, South East Queensland and Central Queensland Coast (Figure 18). The two Brigalow Belt IBRA bioregions were represented within the single Queensland Spatial BioCondition Brigalow Belt product. Locations masked or lacking estimates in either product were omitted from the paired comparison.

Correspondence between paired pixel values was evaluated using Type II major-axis regression implemented in R with `lmodel2` (Legendre & Oksanen 2026). Type II regression was used because both axes contain modelled estimates and neither product provides an error-free independent observation of ecosystem condition (Figure 19; major-axis slope = 0.88, intercept = 0.12, RMSOE = 0.13). The major-axis relationship was treated as the primary product-comparison model; an ordinary least-squares line is shown in Figure 19 only as descriptive context.

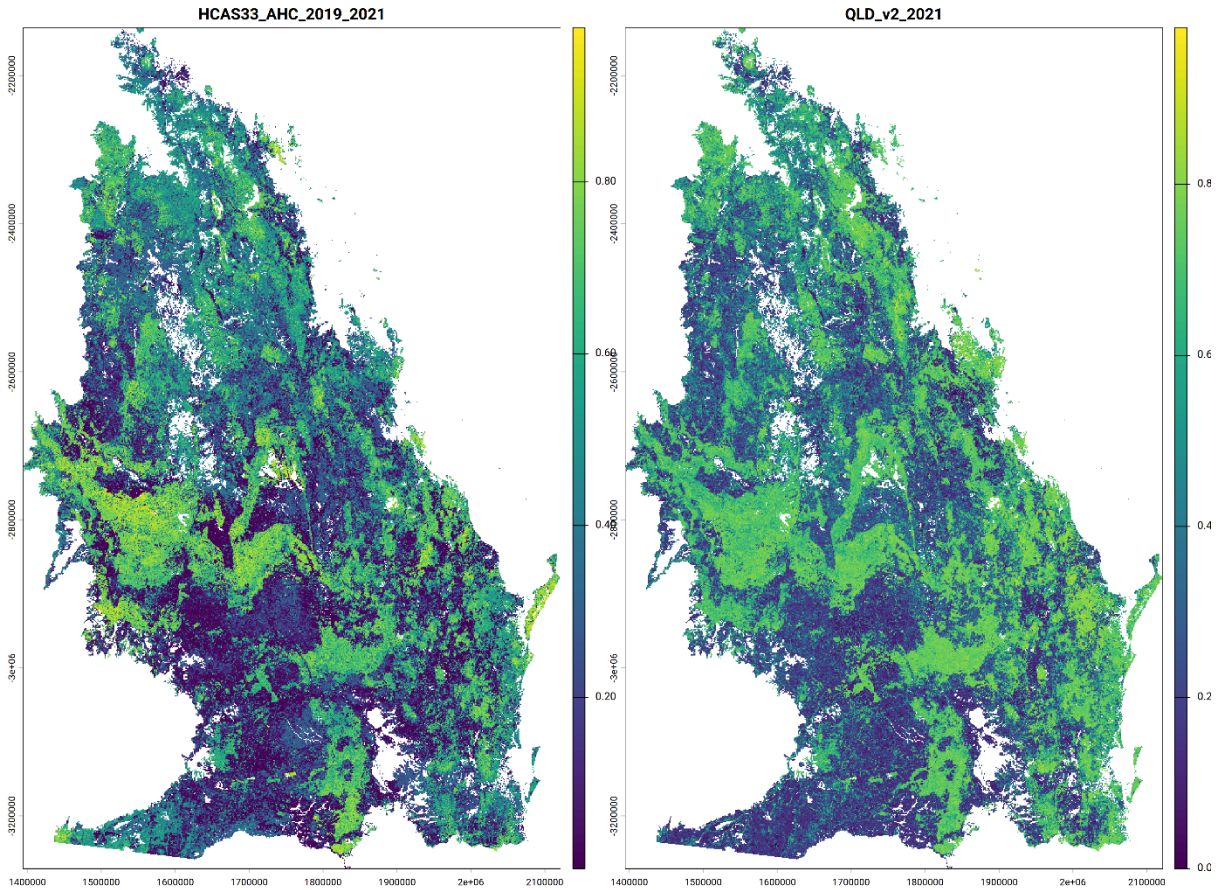


Figure 18 Comparison of mapped ecosystem condition estimates from HCAS v3.3 and Queensland Spatial BioCondition v2. Left: HCAS v3.3 ecosystem condition estimates for the 2019–2021 short-term epoch. Right: Queensland Spatial BioCondition v2 predictions representing vegetation condition for biodiversity in 2021 (mean predicted values per pixel). The comparison covers Brigalow Belt North, Brigalow Belt South, South East Queensland and Central Queensland Coast; the two Brigalow Belt IBRA bioregions are represented within the Queensland Spatial BioCondition Brigalow Belt product. Spatial BioCondition scores were rescaled from 0–100 to 0–1. White areas indicate locations for which Spatial BioCondition estimates were masked or unavailable. Sources: HCAS33_AHC_2019_2021.tif in [Valavi et al. \(2025b\)](#); Queensland Spatial BioCondition v2 datasets ([DESI 2024a](#); [DESI 2024d](#); [DESI 2024b](#)).

The products showed closer correspondence at higher condition values and greater divergence towards the lower end of the condition scale, where Queensland Spatial BioCondition generally assigned higher values than HCAS v3.3 (Figure 19). With HCAS v3.3 plotted on the horizontal axis and Queensland Spatial BioCondition on the vertical axis, the fitted major-axis relationship had a slope of 0.88 and an intercept of 0.12. These results indicate broad correspondence between two independently developed condition products while also identifying a systematic difference in their representation of the lower condition range. Differences may reflect contrasts in input data, spatial resolution, model structure, reference frameworks, mapped exclusions and calibration. The comparison does not establish that either product provides the more accurate estimate of ecosystem condition.

Queensland Spatial BioCondition was also evaluated independently against field-based BioCondition assessments from 270 sites ([DESI 2024c](#)); reproduced in Figure 20. The Queensland evaluation reported $R^2 = 0.70$ and a mean absolute error of 13.4 BioCondition score points on the original 0–100 scale. This Type I ordinary least-squares regression treated all field-assessment scores as the predictor and therefore as measured without error. This field-based evaluation provides context for the predictive performance and limitations of Spatial BioCondition. It provides useful context for interpreting the divergence between HCAS v3.3 and Spatial BioCondition.

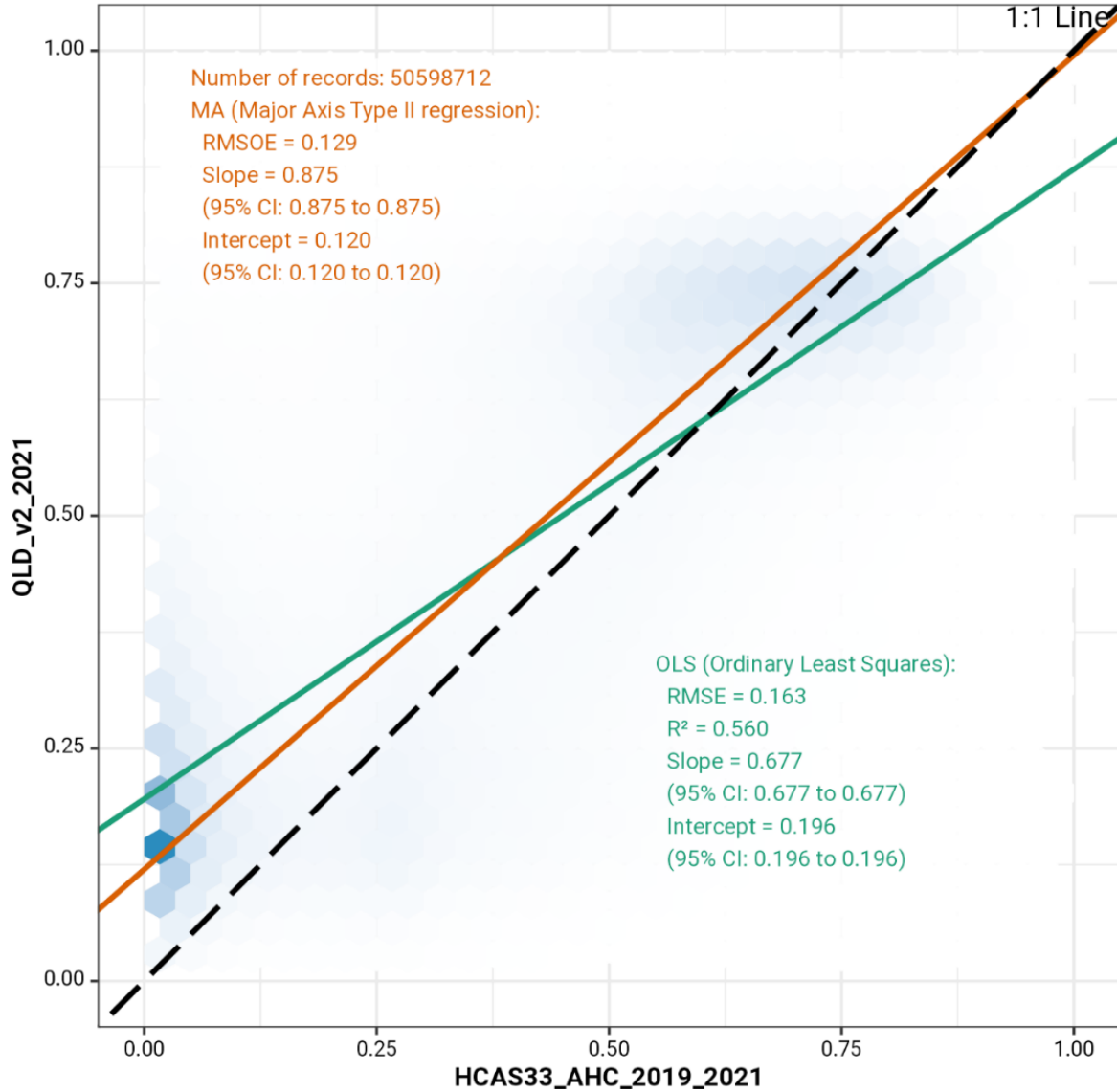


Figure 19 Pixel-level comparison of HCAS v3.3 and Queensland Spatial BioCondition v2 over their common mapped extent. HCAS v3.3 ecosystem condition estimates for the 2019–2021 short-term epoch are plotted on the horizontal axis and Queensland Spatial BioCondition v2 predictions, rescaled from 0–100 to 0–1, are plotted on the vertical axis. Blue hexagons show the density of paired pixel values, with darker shading indicating greater density. The solid line shows the Type II major-axis regression and the dashed line shows the 1:1 relationship. The ordinary least-squares regression (Type I) is shown in green for descriptive context and was not the primary product-comparison model (for purposes of the ordinary least-squares model, Queensland Spatial BioCondition v2 scores are treated as the predictor and therefore as measured without error). Sources: HCAS33_AHC_2019_2021.tif in [Valavi et al. \(2025b\)](#); Queensland Spatial BioCondition v2 datasets – mean predicted values per pixel ([DESI 2024a](#); [DESI 2024d](#); [DESI 2024b](#)).

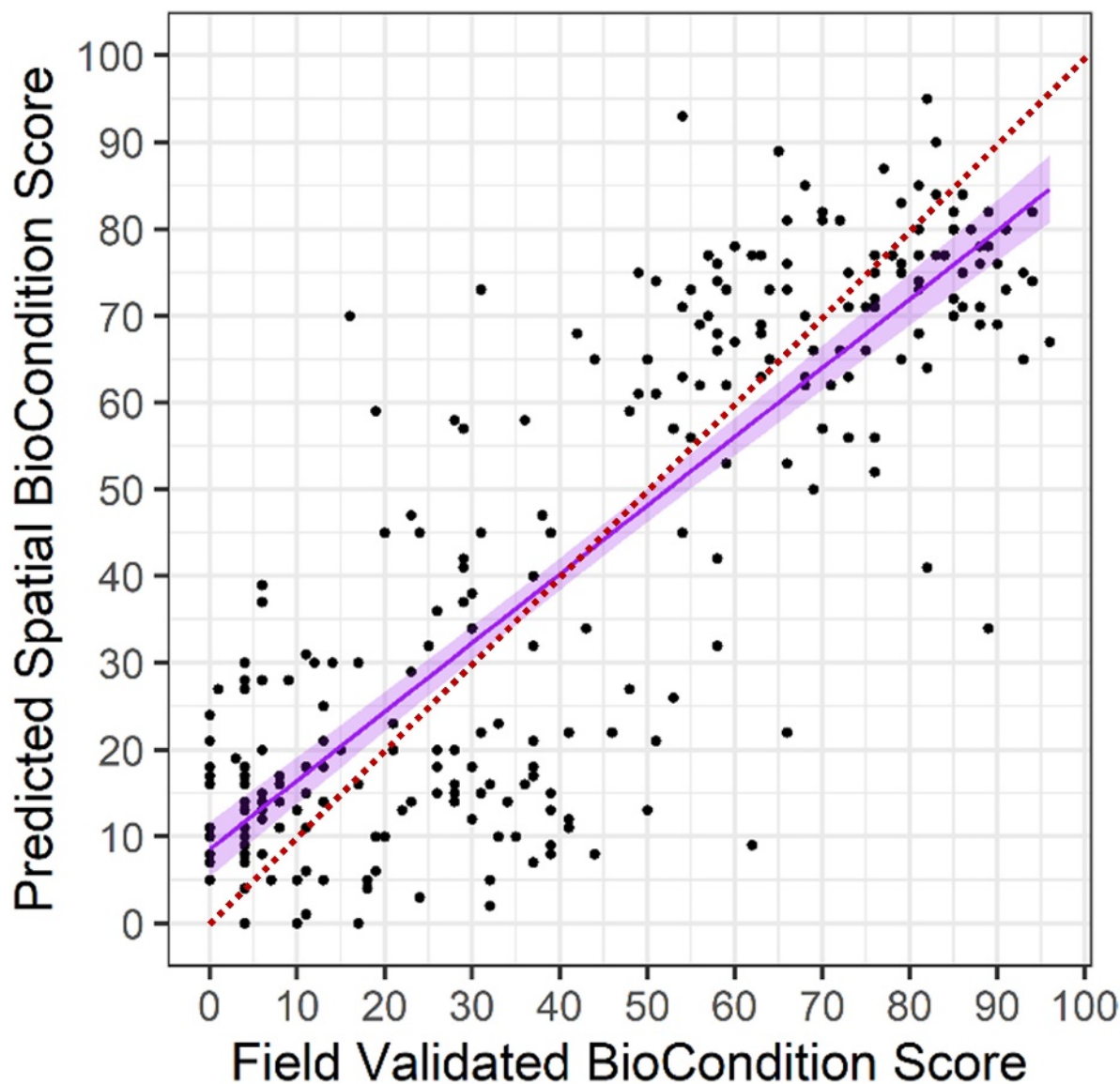


Figure 20 Independent Queensland evaluation of Spatial BioCondition v2 predictions against field-based BioCondition assessments from 270 sites. The published evaluation reported $R^2 = 0.70$ and a mean absolute error of 13.4 BioCondition score points on the original 0–100 scale. The fitted line is the ordinary least-squares regression (Type I) and the dotted line is the 1:1 relationship. For purposes of the ordinary least-squares model, field-assessment scores are treated as the predictor and therefore as measured without error; this is a statistical model assumption and does not imply that the field observations were collected without measurement error. This Queensland analysis is reproduced to provide context for interpreting Spatial BioCondition and was not used to evaluate HCAS v3.3 directly. Source: [DESI \(2024c\)](#).

S5 Implementation considerations

The relationship between the broader HCAS workflow and the functions implemented within **ClassicHCAS** is described in Section 2.3 and Figure 1 of the manuscript. This section documents computational and software-configuration considerations relevant to the continental Australian HCAS v3.3 application.

S5.1 Computational implementation at continental scale

The HCAS v3.3 production workflow partitioned the continental analysis grid into spatial tiles and distributed condition benchmarking across high-performance computing resources. Tiling and parallel execution were managed by the external production workflow rather than by **ClassicHCAS** version 0.2.0. Tiling enables large spatial datasets to be processed as independent subsets and supported workload distribution across the continental domain (requiring between 1 and ~10 minutes to complete; [Valavi *et al.* 2025a](#))

The tiled workflow allowed target locations to be processed as separate spatial subsets using the same reference surfaces, benchmarking settings and calibration curve. For temporal estimation, the reference ecosystem model predictions, training- and benchmark-reference data, reference surfaces, benchmarking configuration and calibration curve established for the long-term epoch were retained across short-term epochs. Target-specific condition inference was repeated for each spatial tile and temporal epoch using the observed ecosystem characteristics for that epoch. Outputs from the independently processed tiles were subsequently assembled into spatially continuous continental products. Further optimisation of tiling strategies could improve load balancing and reduce overall computational cost.

S5.2 Alternative feature-set configurations

ClassicHCAS can construct reference surfaces and estimate condition using specified subsets of the remotely sensed ecosystem characteristic variables, including individual variables where appropriate. The selected variables define the multivariate feature space in which predicted and observed pairwise distances are calculated. The same feature set must therefore be used consistently during reference surface construction and condition benchmarking.

HCAS v3.3 used all 14 ecosystem characteristics described in Section S2.3; no alternative feature subset or univariate condition products were generated as part of this application. Condition estimates derived using different feature sets may differ because the underlying distance space has changed and should not be assumed to be directly comparable with the HCAS v3.3 production estimates. Alternative feature set configurations may also require separate evaluation of benchmarking settings and calibration.

S6 Conclusions

This worked example documents implementation of the HCAS workflow at continental scale using multi-decadal Earth observation, environmental and inferred reference-ecosystem data. The Australian application generated spatially explicit ecosystem condition estimates and repeated temporal estimates on a common 0–1 reporting scale, capturing broad gradients of ecosystem modification, fine-scale spatial pattern and temporal variation across a large and environmentally heterogeneous domain.

The application also documents reference ecosystem modelling, condition benchmarking and calibration, independent supporting evidence for the inferred reference-site dataset, comparison with an independently developed regional condition product, and bootstrap sensitivity to resampling training and benchmark reference sites. Broader interpretation, limitations and implications of the HCAS workflow are discussed in Section 5 of the manuscript.

Appendix A – Environmental covariates selected for use in HCAS v3.3

All environmental covariates listed in Table 6 are 90 m Cloud Optimised GeoTIFFs (COGs) in Australian Albers projection (EPSG:3577, GDA 1994). The full list of 127 environmental covariates published in [Liu *et al.* \(2025a\)](#) and selection process can be found in [Liu *et al.* \(2025b\)](#).

Table 6 The 58 environmental covariates selected to represent long-term, stable abiotic gradients across Australia, used in developing HCAS v3.3 (Liu *et al.* 2025a; Liu *et al.* 2025b). Covariates subsequently excluded from use in the reference ecosystem modelling step are indicated by an asterisk ‘*’, and covariates additionally included to support benchmark reference site sampling are indicated by a hash ‘#’.

COVARIATE GROUP	DATASET NAME	DESCRIPTION	UNITS	CITATION
Climate	Clim_EAAS	Annual total actual evapotranspiration modelled using terrain-scaled water holding capacity, using Budyko framework: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	mm	Harwood <i>et al.</i> (2018)
Climate	Clim_EPA	Total annual potential evaporation: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	mm	Harwood <i>et al.</i> (2018)
Climate	Clim_EPRX	Rate of seasonal change in potential evaporation - maximum difference between successive months: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	mm/day	Harwood <i>et al.</i> (2018)
Climate	Clim_EPX	Maximum monthly potential evaporation: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	mm	Harwood <i>et al.</i> (2018)
Climate	Clim_PTI	Minimum monthly precipitation: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	mm	Harwood <i>et al.</i> (2018)
Climate	Clim_PTRX	Rate of seasonal change in precipitation - maximum of differences between successive months: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	mm/day	Harwood <i>et al.</i> (2018)
Climate	Clim_RSI	Short-wave solar radiation - monthly minimum (SRAD data): 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	MJ/m ² /day	Harwood <i>et al.</i> (2018)
Climate	Clim_RSRI	Rate of seasonal change in short-wave solar radiation - min difference between successive months (SRAD data): 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant <i>et al.</i> 2009)	MJ/m ² /day	Harwood <i>et al.</i> (2018)

COVARIATE GROUP	DATASET NAME	DESCRIPTION	UNITS	CITATION
Climate	Clim_RSRX	Rate of seasonal change in short-wave solar radiation - max difference between successive months (SRAD data): 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	MJ/m ² /day	Harwood et al. (2018)
Climate	Clim_RSX	Short-wave solar radiation - monthly maximum (SRAD data): 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	MJ/m ² /day	Harwood et al. (2018)
Climate	Clim_TNRI	Rate of seasonal change in minimum temperature - minimum difference between successive months: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	°C/day	Harwood et al. (2018)
Climate	Clim_TNRX	Rate of seasonal change in minimum temperature - maximum difference between successive months: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	°C/day	Harwood et al. (2018)
Climate	Clim_TRI	Minimum monthly mean diurnal temperature range: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	°C	Harwood et al. (2018)
Climate	Clim_TRX	Maximum monthly mean diurnal temperature range: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	°C	Harwood et al. (2018)
Climate	Clim_TXRX	Rate of seasonal change in maximum temperature - maximum difference between successive months: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	°C/day	Harwood et al. (2018)
Climate	Clim_TXX	Maximum temperature - monthly maximum: 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	°C	Harwood et al. (2018)
Climate	Clim_WDI	Minimum monthly atmospheric water deficit (PT-EP): 1975-2025 (ANUCLIM version 6.1 derived - Xu & Hutchinson 2011) using 3-second SRTM DEM (Gallant et al. 2009)	mm	Harwood et al. (2018)

COVARIATE GROUP	DATASET NAME	DESCRIPTION	UNITS	CITATION
Climate	Snow_days_avg_1975_2022 [#]	Mean number of days with snow calculated from ANUClimate version 2.0 daily mean temperature and rainfall from 1975 to 2022 (Hutchinson et al. 2022) using the 'Snow Cover' tool in SAGA version 9.2.0 (Conrad et al. 2015)	Days	This report
Parent Material	PM_gravity_2019_B4_CBA*	Gravity – B series terrain corrected complete Bouger anomaly grid 2019	µm/s ²	Lane et al. (2020b) ; Lane et al. (2020a)
Parent Material	PM_magmap_v7_2019_1VD*	Airborne-derived total magnetic intensity, 7 th Edition, 2019 – enhanced grid, 1VD	nT/m	Minty and Poudjom Djomani (2019)
Parent Material	PM_radmap_v4_2019_filtered_pctk_GAPFilled	Radiometric grid of Australia (Radmap) v4 2019 – predictive infilled filtered percent natural radioactive decay of Potassium (derived from Poudjom Djomani & Minty 2019a)	%	Wilford and Kroll (2020)
Parent Material	PM_radmap_v4_2019_filtered_ppmt_GAPFilled	Radiometric grid of Australia (Radmap) v4 2019 – predictive infilled filtered natural radioactive decay of Thorium (derived from Poudjom Djomani & Minty 2019b)	ppm	Wilford and Kroll (2020)
Parent Material	PM_radmap_v4_2019_ratio_tk_GAPFilled	Radiometric grid of Australia (Radmap) v4 2019 - predictive infilled filtered natural radioactive decay of Thorium-Potassium ratio (derived from Poudjom Djomani & Minty 2019c)	Ratio (unitless)	Wilford and Kroll (2020)
Parent Material	PM_radmap_v4_2019_ratio_uk_GAPFilled*	Radiometric grid of Australia (Radmap) v4 2019 - predictive infilled filtered natural radioactive decay of Uranium-Potassium ratio (derived from Poudjom Djomani & Minty 2019d)	Ratio (unitless)	Wilford and Kroll (2020)
Parent Material	PM_radmap_v4_2019_ratio_ut_GAPFilled*	Radiometric grid of Australia (Radmap) v4 2019 - predictive infilled filtered natural radioactive decay of Uranium-Thorium ratio (derived from Poudjom Djomani & Minty 2019e)	Ratio (unitless)	Wilford and Kroll (2020)
Relief	Relief_dems_3s_mosaic1	Elevation derived from the 3” SRTM Digital Elevation Model (DEM) Version 1.0	metres	Gallant et al. (2009)
Relief	Relief_profile_curvature_3 [#]	Profile Curvature derived from 1" SRTM DEM-S and aggregated to 3”	Radians/m	Gallant and Austin (2012)
Relief	Relief_roughness	Landscape roughness quantifies topographic heterogeneity within a neighbourhood – difference in elevation from a centre cell and 8 directly surrounding cells using the method of Riley et al. (1999)	metres	Searle et al. (2022)

COVARIATE GROUP	DATASET NAME	DESCRIPTION	UNITS	CITATION
Relief	Relief_SWI_mosaic [#]	SAGA Wetness Index derived from 3 second SRTM Derived Digital Elevation Model (DEM) Version 1.0 (Gallant <i>et al.</i> 2009). This is similar to the 'Topographic Wetness Index' (Relief_twi_3s), but it is based on a modified	index	Searle <i>et al.</i> (2022)
Relief	Exposed_rate	The annual percentage exposure rate of land between 1988 to 2016 was reclassified from the intertidal extent data (Sagar & Bala 2017)	Fraction	Liu <i>et al.</i> (2025b)
Relief	AccFlw_MFD_3s_log10 [#]	Accumulated flow using the multiple low directions (MFD) method with the 3s DEM resampled using 'bilinear' method from 1s DEM-H (Gallant <i>et al.</i> 2009), log10(accFlow+1)	m ³	Liu <i>et al.</i> (2025b)
Soil	SubSoil_AvailP	Predicted inherent plant available soil phosphorus (P), depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	mg/kg	Searle (2023)
Soil	SubSoil_BDW	Predicted whole soil bulk density (including coarse fragments in mass per unit volume by a method equivalent to the core method) depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	g/cm ³	Searle (2023)
Soil	SubSoil_Clay	Predicted clay soil texture (< 2 um mass fraction of the < 2 mm soil material determined using the pipette method) depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	%	Searle (2023)
Soil	SubSoil_DUL	Predicted soil hydraulic properties – volumetric drained upper limit (one-third bar) depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	mm/mm	Searle (2023)
Soil	SubSoil_pHWater	Soil pH (1:5 water; method 4A1 in Rayment & Lyons 2011) depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	Index	Searle (2023)
Soil	SubSoil_SOC	Predicted total soil organic carbon content (mass fraction by weight in the <2 mm soil material as determined by dry combustion at 900 °C) depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	%	Searle (2023)
Soil	SubSoil_TotalN [*]	Predicted total soil nitrogen (N) content (mass fraction in the soil by weight) depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	%	Searle (2023)

COVARIATE GROUP	DATASET NAME	DESCRIPTION	UNITS	CITATION
Soil	SubSoil_TotalP	Predicted total soil phosphorus (P) content (mass fraction in the soil by weight) depth-weighted average for subsoil depths greater than 30 cm – hole-filled non-anthropogenic version	%	Searle (2023)
Soil	Topsoil_AvailP	Predicted inherent plant available soil phosphorus (P) depth-weighted average for topsoil depths from 0 to 30 cm – hole-filled non-anthropogenic version	mg/kg	Searle (2023)
Soil	Topsoil_Clay	Predicted clay soil texture (< 2 um mass fraction of the < 2 mm soil material determined using the pipette method) depth-weighted average for topsoil depths from 0 to 30 cm – hole-filled non-anthropogenic version	%	Searle (2023)
Soil	Topsoil_DUL	Predicted soil hydraulic properties – volumetric soil moisture at drained upper limit (one-third bar) depth-weighted average for topsoil depths from 0 to 30 cm – hole-filled non-anthropogenic version	mm/mm	Searle (2023)
Soil	Topsoil_pHWater	Soil pH (1:5 water; method 4A1 in Rayment & Lyons 2011) depth-weighted average for topsoil depths from 0 to 30 cm – hole-filled non-anthropogenic version	index	Searle (2023)
Soil	Topsoil_SOC	Predicted total soil organic carbon content (mass fraction by weight in the <2 mm soil material as determined by dry combustion at 900 °C) depth-weighted average for topsoil depths from 0 to 30 cm – hole-filled non-anthropogenic version	%	Searle (2023)
Soil	Topsoil_TotalP	Predicted total soil phosphorus (P) content (mass fraction in the soil by weight) depth-weighted average for topsoil depths from 0 to 30 cm – hole-filled non-anthropogenic version	%	Searle (2023)
Soil	Depth_mean	Depth of soil profile – combining A and B horizons (best estimate including > 2m) – hole-filled non-anthropogenic version	metres	Searle (2023)
Soil	DepthR_mean	Depth to hard rock - inclusive of all regolith – hole-filled non-anthropogenic version	metres	Searle (2023)
Soil	Cond_prediction_0_4m	Median predicted surface conductivity (up to 4 metres depth) from airborne electromagnetic (AEM) conductivity surveys (Ley-Cooper et al., 2020) – hole-filled non-anthropogenic version	S/m	Wilford <i>et al.</i> (2023b)
Soil	Al_prediction*	Median predicted oxide of aluminium (Al ₂ O ₃) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)

COVARIATE GROUP	DATASET NAME	DESCRIPTION	UNITS	CITATION
Soil	Ca_prediction	Median predicted oxide of calcium (CaO) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	Fe_prediction*	Median predicted oxide of iron (Fe ₂ O ₃ tot) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	K_prediction	Median predicted oxide of potassium (K ₂ O) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	Mg_prediction*	Median predicted oxide of magnesium (MgO) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	Mn_prediction*	Median predicted oxide of manganese (MnO) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	Na_prediction*	Median predicted oxide of sodium (Na ₂ O) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	P_prediction*	Median predicted oxide of phosphorus (P ₂ O ₅) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	Si_prediction	Median predicted oxide of silicon (SiO ₂) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)
Soil	Ti_prediction*	Median predicted oxide of titanium (TiO ₂) concentration in surface rock and regolith – hole-filled non-anthropogenic version	Log ₁₀ weight %	Wilford <i>et al.</i> (2023a)

References

- ABARES (2024) Land use of Australia 2010–11 to 2020–21. Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES), Australian Government Department of Agriculture, Fisheries and Forestry. <https://doi.org/10.25814/w175-xh85>.
- Bishop, P. & Pillans, B. (2010) Australian Landscapes. Geological Society, London, Special Publications. Lyell Collection. <https://doi.org/10.1144/SP346.0>.
- Bradstock, R.A. (2010) A biogeographic model of fire regimes in Australia: current and future implications. *Global Ecology and Biogeography*, 19, 145–158. <https://doi.org/10.1111/j.1466-8238.2009.00512.x>.
- Brus, D.J. (2019) Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma*, 338, 464–480. <https://doi.org/10.1016/j.geoderma.2018.07.036>.
- Bureau of Meteorology (2023) Australian Hydrological Geospatial Fabric (Geofabric) Product Suite, National V3.3. The Australian Bureau of Meteorology (BOM). <https://portal.wsapl.cloud.bom.gov.au/arcgis/apps/sites/#/australian-water-data-service/maps/35719064c4ea4ad79faa82f5c9c22068/about>
- Commonwealth of Australia (2021) Geoscience Australia Landsat Analysis Ready Data Collection 3. National Computing Infrastructure. <https://doi.org/10.25914/6099413995ed0>.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. & Böhner, J. (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.*, 8, 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>.
- Czúcz, B., Keith, H., Driver, A., Jackson, B., Nicholson, E. & Maes, J. (2021) A common typology for ecosystem characteristics and ecosystem condition variables. *One Ecosystem*, 6, e58218. <https://doi.org/10.3897/oneeco.6.e58218>.
- DCCEEW (2025a) Collaborative Australian Protected Areas Database (CAPAD) 2024 - marine. Australian Government Department of Climate Change, Energy, the Environment and Water. <https://www.dcceew.gov.au/environment/land/nrs/science/capad/2024>
- DCCEEW (2025b) Collaborative Australian Protected Areas Database (CAPAD) 2024 - terrestrial. Australian Government Department of Climate Change, Energy, the Environment and Water. <https://www.dcceew.gov.au/environment/land/nrs/science/capad/2024>
- DES (2021) Spatial BioCondition: Vegetation condition map for Queensland. Queensland Government Department of Environment and Science (DES). https://www.qld.gov.au/_data/assets/pdf_file/0015/230019/spatial-biocondition-vegetation-condition-map-for-queensland.pdf
- DESI (2024a) Brigalow Belt Bioregion Spatial BioCondition - 2021 - version 2.0. Queensland Department of Environment, Science and Innovation. <https://qldspatial.information.qld.gov.au/catalogue/custom/search.page?q=spatial+biocondition#>
- DESI (2024b) Central Queensland Coast Bioregion Spatial BioCondition - 2021 - version 2.0. Queensland Department of Environment, Science and Innovation. <https://qldspatial.information.qld.gov.au/catalogue/custom/search.page?q=spatial+biocondition#>

- DESI (2024c) Queensland Spatial BioCondition Framework (SBC) Factsheet (version 2). Queensland Department of Environment, Science and Innovation. https://www.qld.gov.au/__data/assets/pdf_file/0016/404161/spatial-biocondition-fact-sheet.pdf
- DESI (2024d) South East Queensland Bioregion Spatial BioCondition - 2021 - version 2.0. Queensland Department of Environment, Science and Innovation. <https://qldspatial.information.qld.gov.au/catalogue/custom/search.page?q=spatial+biocondition#>
- Dühr, S. (2024) Population Growth and Distribution in Australia: Policy and Governance Challenges for a More Balanced Settlement Structure. *Urban Policy and Research*, 42, 109–123. <https://doi.org/10.1080/08111146.2024.2320651>.
- Efron, B. & Tibshirani, R.J. (1994) *An Introduction to the Bootstrap* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9780429246593>.
- Foran, B., Smith, M.S., Burnside, D., Andrew, M., Blesing, D., Forrest, K. & Taylor, J. (2019) Australian rangeland futures: time now for systemic responses to interconnected challenges. *The Rangeland Journal*, 41, 271–292. <https://doi.org/10.1071/rj18105>.
- Fritsch, F.N. & Carlson, R.E. (1980) Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17, 238–246. <https://doi.org/10.1137/0717021>.
- Gallant, J.C. & Austin, J. (2012) Profile Curvature derived from 1" SRTM DEM-S (included 3" product). CSIRO. Data Collection. <https://doi.org/10.4225/08/56E9DEBF65706>.
- Gallant, J.C., Wilson, N., Tickle, P.K., Dowling, T. & Read, A. (2009) 3 second SRTM Derived Digital Elevation Model (DEM) Version 1.0. Record 1.0. Geoscience Australia. <http://pid.geoscience.gov.au/dataset/ga/69888>.
- Geoscience Australia (2024) DEA Geometric Median and Median Absolute Deviation (Landsat) version 4.0.0. Geoscience Australia Landsat Collection 3. Digital Earth Australia, Geoscience Australia. <https://doi.org/10.26186/146261>.
- Giljohann, K.M., Munroe, S., Botha, E., Liu, N., Williams, K.J., Ferrier, S. & Joehnk, K. (2025) Identifying reference and highly modified sites for use in the Habitat Condition Assessment System (HCAS). Technical Report EP2025-2017. CSIRO. <https://doi.org/10.25919/nh6j-8c14>.
- Giljohann, K.M., Williams, K.J., Hossack, G.R., Lehmann, E.A., Liu, N. & Richards, A.E. (2024) Using expert opinion to reclassify ecosystem condition – development of methods. Technical Report. Publication EP2024-6358. CSIRO. <https://publications.csiro.au/publications/publication/PIcsi:EP2024-6358>
- Godfree, R.C., Knerr, N., Encinas-Viso, F., Albrecht, D., Bush, D., Christine Cargill, D., Clements, M., Gueidan, C., Guja, L.K., Harwood, T., Joseph, L., Lepschi, B., Nargar, K., Schmidt-Lebuhn, A. & Broadhurst, L.M. (2021) Implications of the 2019–2020 megafires for the biogeography and conservation of Australian vegetation. *Nature Communications*, 12, 1023. <https://doi.org/10.1038/s41467-021-21266-5>.
- Grundy, M.J., Searle, R., Meier, E.A., Ringrose-Voase, A.J., Kidd, D., Orton, T.G., Triantafyllis, J., Philip, S., Liddicoat, C., Malone, B., Thomas, M., Gray, J. & Bennett, J.M. (2020) Digital soil assessment delivers impact across scales in Australia and the Philippines. *Geoderma Regional*, 22, e00314. <https://doi.org/10.1016/j.geodrs.2020.e00314>.

- Grundy, M.J., Viscarra Rossel, R.A., Searle, R.D., Wilson, P.L., Chen, C. & Gregory, L.J. (2015) Soil and Landscape Grid of Australia. *Soil Research*, 53, 835–844. <https://doi.org/10.1071/SR15191>.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9).
- Harwood, T.D., King, D., Nolan, M., Gallant, J., Ware, C., Austin, J. & Williams, K.J. (2018) 3 arcsecond climatology for continental Australia 1976-2005: Summary variables with elevation and radiative adjustment for modelling biodiversity patterns. CSIRO. Data Collection. <https://doi.org/10.25919/gayv-1e32>.
- Hutchinson, M., Kesteven, J. & Xu, T. (2022) ANUClimate 2.0. Australian National University. https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f2576_7854_4065_1457
- Keith, D.A. (2017) *Australian Vegetation*, Third Edition. pp. 747. Cambridge University Press. <https://doi.org/10.1017/9781316339701>.
- Kidd, D., Searle, R., Grundy, M., McBratney, A., Robinson, N., O'Brien, L., Zund, P., Arrouays, D., Thomas, M., Padarian, J., Jones, E., Bennett, J.M., Minasny, B., Holmes, K., Malone, B.P., Liddicoat, C., Meier, E.A., Stockmann, U., Wilson, P., Wilford, J., Payne, J., Ringrose-Voase, A., Slater, B., Odgers, N., Gray, J., van Gool, D., Andrews, K., Harms, B., Stower, L. & Triantafyllis, J. (2020) Operationalising digital soil mapping – Lessons from Australia. *Geoderma Regional*, 23, e00335. <https://doi.org/10.1016/j.geodrs.2020.e00335>.
- Lane, R.J.L., Wynne, P.E., Poudjom Djomani, Y., Stratford, W.R., Barretto, J.A. & Caratori Tontini, F. (2020a) 2019 Australian National Gravity Grids Explanatory Notes. Geoscience Australia. <https://doi.org/10.11636/Record.2020.022>.
- Lane, R.J.L., Wynne, P.E., Poudjom Djomani, Y., Stratford, W.R., Barretto, J.A. & Caratori Tontini, F. (2020b) 2019 Australian National Gravity Grids: Free Air Anomaly, Complete Bouguer Anomaly, De-trended Global Isostatic Residual, 400 m cell size (includes Point Located Data). Geoscience Australia. <https://pid.geoscience.gov.au/dataset/ga/133023>.
- Legendre, P. & Oksanen, J. (2026) *lmodel2*: Model II Regression. CRAN-R. <https://doi.org/10.32614/CRAN.package.lmodel2>.
- Lehmann, E. & Williams, K.J. (2025) Assessing change between Habitat Condition Assessment System (HCAS) epoch means: test of statistical significance. Technical report EP2025-4799. CSIRO. <https://doi.org/10.25919/dchz-j724>.
- Lehmann, E.A., Collings, S., Valavi, R. & Williams, K.J. (2025) Model-based uncertainty quantification for the Habitat Condition Assessment System (HCAS). Publication EP2025-2296. CSIRO. <https://doi.org/10.25919/7czb-an54>.
- Lehmann, E.A., Ferrier, S. & Williams, K.J. (2023) Trend test for serially correlated time series data. A technical report from the Regional Ecosystem Accounting Pilot projects. Publication number: EP2023-3255. CSIRO. <https://doi.org/10.25919/e4gd-4f88>.
- Levick, S.R., Johnson, S., Williams, K.J., van Niel, T.G., Valavi, R., Paget, M., Lehmann, E.A., Collings, S. & Newnham, G. (2025) Remotely sensed ecosystem characteristics for condition estimation in the Habitat Condition Assessment System (HCAS): transitioning to Landsat.

- Technical Report EP2025-2350. CSIRO. <https://publications.csiro.au/publications/publication/PIcsiro:EP2025-2350>.
- Li, Z. & Wood, S.N. (2020) Faster model matrix crossproducts for large generalized linear models with discretized covariates. *Statistics and Computing*, 30, 19–25. <https://doi.org/10.1007/s11222-019-09864-2>.
- Liu, N., Williams, K., Searle, R., Wilford, J., Botha, H., Johnson, S., Read, A., Valavi, R., Lehmann, E., Giljohann, K. & Joehnk, K. (2025a) Environmental covariates stack for biodiversity and ecosystem modelling at 90m and 250m for continental Australia. Data Collection 66355. CSIRO. <https://doi.org/10.25919/73vy-sc68>.
- Liu, N., Williams, K.J., Searle, R., Wilford, J., Botha, E., Johnson, S., Read, A., Valavi, R., Lehmann, E., Giljohann, K.M. & Joehnk, K. (2025b) Environmental covariates for predicting the reference state of ecosystems in the Habitat Condition Assessment System (HCAS). Technical Report EP2025-2352. CSIRO. <https://doi.org/10.25919/wt8y-dr64>.
- Lymburner, L. (2024) Geoscience Australia Landsat Analysis Ready Data Collection 3 Fractional Cover Percentiles v4.0.0. Geoscience Australia. <https://pid.geoscience.gov.au/dataset/ga/145501>.
- Macfarlane, C., Cox, S., Francis, R.J., Jordon, R., Keith, D.A., Kingsford, R.T., Liu, N., Newnham, G.J., Nicholson, E., Prober, S.M., Richards, A.E., Tetreault-Campbell, S., Trebilco, R. & Schmidt, R.K. (2024a) Methods for developing spatial extent of IUCN ecosystem functional groups at 250 m resolution for freshwater, terrestrial, marine and transitional realms for Australia: pre-1750, 2010-11, 2015-16 and 2020-21. Publication Number 2024-5047. CSIRO. <https://doi.org/10.25919/13nv-pm70>.
- Macfarlane, C., Cox, S., Francis, R.J., Jordon, R., Keith, D.A., Kingsford, R.T., Liu, N., Newnham, G.J., Nicholson, E., Prober, S.M., Richards, A.E., Tetreault-Campbell, S., Trebilco, R. & Schmidt, R.K. (2024b) Spatial extent of IUCN ecosystem functional groups at 250 m resolution for freshwater, terrestrial, marine and transitional realms for Australia: pre-1750, 2010-11, 2015-16 and 2020-21. CSIRO. <https://doi.org/10.25919/t8fc-er82>.
- Malone, B.P., Searle, R., Stenson, M., McJannet, D., Zund, P., Román Dobarco, M., Wadoux, A.M.J.C., Minasny, B., McBratney, A. & Grundy, M. (2025) Update and expansion of the soil and landscape grid of Australia. *Geoderma*, 455, 117226. <https://doi.org/10.1016/j.geoderma.2025.117226>.
- Marra, G. & Wood, S.N. (2011) Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55, 2372–2387. <https://doi.org/10.1016/j.csda.2011.02.004>.
- McFarland, P. (2015) The Peri-urban Land-Use Planning Tangle: An Australian Perspective. *International Planning Studies*, 20, 161–179. <https://doi.org/10.1080/13563475.2014.965250>.
- Minty, B.R.S. & Poudjom Djomani, Y. (2019) Total Magnetic Intensity (TMI) Grid of Australia 2019 - seventh edition - 40 m cell size. Geoscience Australia. <https://doi.org/10.26186/5dd5fb557fffb>
- Mokany, K., McCarthy, J.K., Falster, D.S., Gallagher, R., Harwood, T.D., Kooyman, R. & Westoby, M. (2022a) Harmonised Australian Vegetation Plot dataset (HAVPlot). v1. Data Collection. CSIRO. <https://doi.org/10.25919/5cex-4s70>.

- Mokany, K., McCarthy, J.K., Falster, D.S., Gallagher, R.V., Harwood, T.D., Kooyman, R. & Westoby, M. (2022b) Patterns and drivers of plant diversity across Australia. *Ecography*, 2022, e06426. <https://doi.org/10.1111/ecog.06426>.
- Mueller, N., Roberts, D., Emma, A. & Miller, J. (2024) DEA Geometric Median and Median Absolute Deviation (Landsat). Geoscience Australia. <https://doi.org/10.26186/149305>.
- Munroe, S., Brinkman, J., Giljohann, K., Williams, K.J., Hosack, G., Brenton, P., Raisbeck-Brown, N., Warnick, A., Dickson, F., Pirzl, R., Sinclair, S.J. & White, M. (2026) Habitat Condition Assessment Tool (HCAT) v3.0: Expert-derived ecosystem condition assessments for Australia (data collected July 2025–April 2026). Data Collection. CSIRO. <https://doi.org/10.25919/jacx-e840>.
- Munroe, S., Giljohann, K.M., Brenton, P., Williams, K.J., Hosack, G., Richards, A., Brinkman, J., Checksfield, S., Salomon, J., Raisbeck-Brown, N., Warnick, A., Dickson, F., Pirzl, R., Sinclair, S. & White, M. (2025) The Habitat Condition Assessment Tool: Instructions for participants, version 3.0. Publication EP2025-2633. CSIRO. <https://doi.org/10.25919/gkr9-cx03>.
- Murphy, B.P., Bradstock, R.A., Boer, M.M., Carter, J., Cary, G.J., Cochrane, M.A., Fensham, R.J., Russell-Smith, J., Williamson, G.J. & Bowman, D.M.J.S. (2013) Fire regimes of Australia: a pyrogeographic model system. *Journal of Biogeography*, 40, 1048–1058. <https://doi.org/10.1111/jbi.12065>.
- Pedersen, E.J., Miller, D.L., Simpson, G.L. & Ross, N. (2019) Hierarchical generalized additive models in ecology: an introduction with *mgcv*. *PeerJ*, 7, e6876. <https://doi.org/10.7717/peerj.6876>.
- Poudjom Djomani, Y. & Minty, B.R.S. (2019a) Radiometric Grid of Australia (Radmap) v4 2019 filtered pct potassium. Geoscience Australia. <https://doi.org/10.26186/5dd48d628f4f6>.
- Poudjom Djomani, Y. & Minty, B.R.S. (2019b) Radiometric Grid of Australia (Radmap) v4 2019 filtered ppm thorium. Geoscience Australia. <https://doi.org/10.26186/5dd48e3eb6367>.
- Poudjom Djomani, Y. & Minty, B.R.S. (2019c) Radiometric Grid of Australia (Radmap) v4 2019 ratio thorium over potassium. Geoscience Australia. <https://doi.org/10.26186/5dd48f8e542f2>.
- Poudjom Djomani, Y. & Minty, B.R.S. (2019d) Radiometric Grid of Australia (Radmap) v4 2019 ratio uranium over potassium. Geoscience Australia. <https://doi.org/10.26186/5dd4a5864e8dd>.
- Poudjom Djomani, Y. & Minty, B.R.S. (2019e) Radiometric Grid of Australia (Radmap) v4 2019 ratio uranium over thorium. Geoscience Australia. <https://doi.org/10.26186/5dd4a63603704>.
- R Core Team (2026) R documentation for *splinefun* in package *stats*: Interpolating Splines. R Foundation for Statistical Computing. Viewed 10 September 2026, <<https://search.r-project.org/R/refmans/stats/html/splinefun.html>>.
- Rayment, G.E. & Lyons, D.J. (2011) Soil Chemical Methods - Australasia. CSIRO Publishing. <https://doi.org/10.1071/9780643101364>.
- Riley, S.J., DeGloria, S.D. & Elliot, R. (1999) A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5, 23–27. <https://arc.lib.montana.edu/ojs/index.php/IJS/article/view/2914>.

- Sagar, S. & Bala, B. (2017) Intertidal Extents Model 25m v. 2.0.0 (Geoscience Australia Landsat Collection 2). Geoscience Australia. <https://doi.org/10.4225/25/5a602cc9eb358>.
- Searle, R. (2023) HCAS Optimised SLGA Products. Terrestrial Ecosystem Research Network. <https://aussoilsdsm.esoil.io/other/hcas-optimised-slga-products>
- Searle, R., Malone, B., Wilford, J., Austin, J., Ware, C., Webb, M., Roman Dobarco, M. & Van Niel, T. (2022) TERN Digital Soil Mapping Raster Covariate Stacks. CSIRO. Data Collection. <https://doi.org/10.25919/jr32-yq58>.
- Stein, J.L., Hutchinson, M.F. & Stein, J.A. (2014) A new stream and nested catchment framework for Australia. *Hydrology and Earth System Sciences*, 18, 1917–1933. <https://doi.org/10.5194/hess-18-1917-2014>.
- Stern, H., Hoedt, G.d. & Ernst, J. (2000) Objective classification of Australian climates. *Australian Meteorological Magazine*, 49, 87–96. <https://doi.org/10.1071/es00009>.
- Thackway, R. & Lesslie, R. (2006) Reporting vegetation condition using the Vegetation Assets, States and Transitions (VAST) framework. *Ecological Management & Restoration*, 7, S53–S62. <https://doi.org/10.1111/j.1442-8903.2006.00292.x>.
- Thackway, R. & Lesslie, R. (2008) Describing and mapping human-induced vegetation change in the Australian landscape. *Environmental Management*, 42, 572–590. <https://doi.org/10.1007/s00267-008-9131-5>.
- United Nations, European Union, Food and Agriculture Organization of the United Nations, International Monetary Fund, Organisation for Economic Co-operation and Development & The International Bank for Reconstruction and Development/The World Bank (2024) System of Environmental-Economic Accounting – Ecosystem Accounting. United Nations Department of Economic and Social Affairs, Statistics Division. https://seea.un.org/sites/seea.un.org/files/documents/EA/seea_ea_f124_web_12dec24.pdf
- Valavi, R., Collings, S., Williams, K.J., Lehmann, E.A., Liu, N., Niel, T.G.V. & Ferrier, S. (2025a) Innovation and improvements in the Habitat Condition Assessment System (HCAS): Developing version 3. Technical Report EP2025-2353. CSIRO. <https://publications.csiro.au/publications/publication/PIcsi:EP2025-2353>.
- Valavi, R., Levick, S.R., Lehmann, E.A., Liu, N., Giljohann, K.M., Williams, K.J., Collings, S., Johnson, S., Botha, E.J., Munroe, S.E.M., Van Niel, T.G., Newnham, G., Paget, M., Malley, C., Carlile, P., Gunawardana, D., Lyon, P., Richards, A.E., Tetreault Campbell, S. & Ferrier, S. (2025b) HCAS 3.3 (1988-2024) base model estimate of habitat condition (90m grid), National Connectivity Index 2.0 (NCI) and annual time series for continental Australia. Data collection 65549. CSIRO. <https://doi.org/10.25919/hngh-v675>.
- Valavi, R., Williams, K.J., Lehmann, E.A., Collings, S., Levick, S.R., Giljohann, K.M., Liu, N., Johnson, S., Munroe, S.E.M., Van Niel, T.G., Paget, M., Malley, C., Lyon, P., Harwood, T.D. & Ferrier, S. (2026) A scalable workflow for ecosystem condition assessment from Earth observation data. *EcoevoRxiv*, <https://publications.csiro.au/publications/publication/PIcsi:EP2026-2108>.
- White, M.D., Hollings, T., Williams, K.J., Dickson, F., Brenton, P., Raisbeck-Brown, N., Warnick, A., Lyon, P., Mokany, K., Liu, C., Sinclair, S.J. & Pirzl, R. (2023) Towards a continent-wide ecological site-condition database using calibrated expert evaluations. *Ecological Applications*, 33, e2729. <https://doi.org/10.1002/eap.2729>.

- White, M.D., Raisbeck-Brown, N., Williams, K.J., Warnick, A., Mokany, K., Brenton, P., Sathya Moorthy, S. & Pirzl, R. (2019) Habitat condition data for Australia from expert elicitation. Data Collection 38438. CSIRO. <https://doi.org/10.25919/5c7da9661dbcc>.
- Wilford, J., de Caritat, P. & Basak, S. (2023a) Predictive grids of major oxide concentrations in surface rock and regolith over the Australian continent. Geoscience Australia. <https://doi.org/10.26186/148587>.
- Wilford, J., LeyCooper, Y. & Basak, S. (2023b) National surface and near-surface conductivity grids. Geoscience Australia. <https://doi.org/10.26186/148588>.
- Wilford, J. & Roberts, D. (2021) Enhanced Bare Earth Covariates for Soil and Lithological Modelling. Terrestrial Ecosystem Research Network (TERN). <https://researchdata.edu.au/enhanced-bare-earth-lithological-modelling/1885434>.
- Wilford, J.R. & Kroll, A. (2020) Complete Radiometric Grid of Australia (Radmap) v4 2019 with modelled infill. Geoscience Australia. <https://pid.geoscience.gov.au/dataset/ga/144413>.
- Williams, K.J., Belbin, L., Austin, M.P., Stein, J. & Ferrier, S. (2012) Which environmental variables should I use in my biodiversity model? International Journal of Geographic Information Sciences, 26, 2009–2047. <https://doi.org/10.1080/13658816.2012.698015>.
- Williams, K.J., Giljohann, K.M., Richards, A.E., Harwood, T.D., Lehmann, E.A., Liu, N., Ferrier, S., Murphy, H.T., Prober, S.M., Hosack, G.R., Tetreault-Campbell, S. & Schmidt, R.K. (2023) Extended methods used in developing the Habitat Condition Assessment System (HCAS) version 2.3, ecosystem condition account-ready data and experimental accounts for two mixed-use landscapes. A technical report from the Regional Ecosystem Accounting Pilot projects. Publication number: EP2023-1426. CSIRO. <https://doi.org/10.25919/r12v-ya32>.
- Williams, K.J., Harwood, T.D., Lehmann, E.A., Ware, C., Lyon, P., Bakar, S., Pinner, L., Schmidt, R.K., Mokany, K., Van Niel, T.G., Richards, A.E., Dickson, F., McVicar, T.R. & Ferrier, S. (2021a) Habitat Condition Assessment System (HCAS version 2.1): Enhanced method for mapping habitat condition and change across Australia. CSIRO. <https://doi.org/10.25919/n3c6-7w60>.
- Williams, K.J., Hunter, B., Schmidt, R.K., Woodward, E. & Cresswell, I. (2021b) Australia State of the Environment 2021 - Land. <https://doi.org/10.26194/6EAM-6G50>.
- Williams, K.J., Liu, N., Valavi, R., Lehmann, E.A., Levick, S.R., Giljohann, K.M., Johnson, S., Botha, E.J., Munroe, S.E., Collings, S., Searle, R., Van Niel, T.G., Newnham, G., Paget, M., Joehnk, K., Hosack, G.R., Harwood, T.D., Malley, C., Gunawardana, D., Sivanandam, P., Richards, A., Tetreault Campbell, S. & Ferrier, S. (2025a) Habitat Condition Assessment System (HCAS) version 3.1: Guide to the 90-metre data collection with uncertainty. Technical report EP2025-1549, Version 2 edn. CSIRO. <https://doi.org/10.25919/nn6k-hn53>.
- Williams, K.J., Liu, N., Valavi, R., Levick, S.R., Lehmann, E.A., Collings, S., Giljohann, K.M., Johnson, S., Botha, E.J., Munroe, S.E.M., Van Niel, T.G., Newnham, G., Paget, M., Malley, C., Carlile, P., Gunawardana, D., Lyon, P., Richards, A.E., Tetreault-Campbell, S. & Ferrier, S. (2025b) Habitat Condition Assessment System (HCAS) version 3.3: Guide to the 90-metre data collection. Edition 2. Technical report EP2025-2979. CSIRO. <https://doi.org/10.25919/jz55-5d44>.

- Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.
- Xu, T. & Hutchinson, M. (2011) ANUCLIM Version 6.1 User Guide. The Australian National University, Fenner School of Environment and Society. <https://fennerschool.anu.edu.au/files/anuclim61.pdf>
- Xu, T. & Hutchinson, M.F. (2013) New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package. *Environmental Modelling & Software*, 40, 267–279. <https://doi.org/10.1016/j.envsoft.2012.10.003>.