FRONTIER SI

IMPACT OF SATELLITE DATA QUALITY ON EO-DERIVED INSIGHTS FOR DECISION MAKING

Dr Fang Yuan

Dr Claire Fisk

September 2024

TABLE OF CONTENTS

ACKNOWLEDGEMENTS OF COUNTRY

We respectfully acknowledge the Aboriginal and Torres Strait Islander people of Australia, first custodians of the lands, air and waters that sustain the places we live, work and play. These first peoples have had a vibrant, living culture that has remained in sustainable synergy with the natural environment for tens of thousands of years, and continues to do so. Remotely sensed data about the natural world is a powerful tool to create shared community vision accessible to many groups about sustainable ways of living in harmony with Country. We believe that through capturing high-quality remotely sensed data of the real world, we are doing our best to respect Country by accurately representing this real world in a way that ensures the integrity, trust and quality of data used for decision for the good of people and planet into the future. We recognise that the lands of the Aboriginal and Torres Strait Islander people of Australia were never ceded, and coexist with the Commonwealth of Australia.

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Earth Observation (EO) data is crucial in Australia, underpinning a wide range of applications and informing decision-making processes in both government and industry. This study aims to quantify and assess the potential impact of EO data calibration quality on decision-making in Australia.

Our report begins with a review of existing literature covering topics on the importance of EO data quality, the potential downstream effects of calibration errors, quantifications of sensor degradations, and validations of EO data and derived products in Australia. Despite the general recognition of the importance of well-calibrated and validated data, we found limited literature on how data quality issues actually affect the derived information products and the subsequent decisions made based on these products. Depending on the use case, the impact on derived information can be significant.

We then present a case study of the Habitat Condition Assessment System (HCAS), a representative EO analysis supporting the Commonwealth government's national nature-positive policies. We designed two assessments based on the HCAS modelling approach and the input remote sensing variables used. Various types of perturbations were introduced to baseline Landsat surface reflectance data to simulate errors resulting from calibration and data processing issues, with the amount determined based on validation assessments found in our literature review.

Our first assessment explores the impact of sensor differences on input remote sensing variables to the HCAS model. We find that simple spectral indices are highly sensitive to potential errors in surface reflectance data. In HCAS, the use of remote sensing distance calculated from normalized variables appears to mitigate some impacts of input data errors. Our second assessment attempts to evaluate how errors in surface reflectance may change the condition assessments using a simplified remote sensing distance-based condition assessment model. Although a complete replication of the HCAS system was not feasible in our study, these analyses provide useful insights into the extent of the impact due to data calibration issues. Our analysis indicates that different types of error distributions in input data result in different biases in the output. In some scenarios, when perturbed input data is used, the distribution of condition similarity metrics changed significantly, with more than 5% of assessment results changing in our simplified model.

We conclude by discussing the implications of our findings for both EO data providers and users. It is imperative for data providers to communicate data quality assessments and limitations clearly to users to prevent data misinterpretation or misuse. Regular and ongoing calibration and validation of sensors and data products are essential to maintain data integrity. Users of EO data should be aware of the inherent uncertainties associated with EO data and consider these uncertainties when evaluating derived information. Such considerations are important to enable the effective use of EO-based solutions for decision support.

1. INTRODUCTION

Earth Observation (EO) data is crucial in Australia, underpinning a wide range of applications and informing decision-making processes in both government and industry. This study aims to quantify and assess the potential impact of Earth observation data calibration quality on decision-making in Australia through reviewing existing relevant work locally and internationally and conducting analysis for a high-profile use case in Australia, namely the CSIRO's Habitat Condition Assessment System (HCAS).

Findings from this study will highlight the importance of data quality to remote sensing data users and providers as well as provide quantitative evidence to inform data quality discussions.

2. LITERATURE REVIEW

2.1 Background

Earth Observation (EO) data plays an important role across a wide variety of applications from environmental, agriculture, land use and climate monitoring, mineral exploration, mining impact assessment, emergency response and preparedness as well as various defence purposes. It is critical therefore that users have access to well calibrated and validated data to ensure that it is fit for purpose and that limitations of the data are made clear as they may influence how the information can be used.

Historically, EO data has been acquired by national space agencies via large expensive satellites such as Landsat, Sentinel and Moderate Resolution Imaging Spectroradiometer (MODIS), but in recent years the introduction of nanosatellites or cubesats has meant more commercial companies are able to launch low cost EO missions while also enabling greater coverage and an increase in the volume of data that is captured (Baber, 2021). Prior to launch, satellites typically undergo pre-launch calibration but over time they begin to degrade no matter how sophisticated these instruments are designed to be. Therefore, post-launch calibration is critical to ensure the quality of information collected from these sensors post-launch continues to play a fundamental role in the basis of reliable remote sensing (Müller, 2014).

The aim of this literature review is to explore examples of work that has been completed relating to the calibration and validation of EO data, and, specifically, to examine the potential flow on effects from poor calibration along with the impacts of sensor degradation.

2.2 Calibration and effects of calibration errors

The Committee on Earth Observation Satellites (CEOS) defines calibration as "the process of quantitatively defining a system's response to known and controlled signal inputs" (CEOS, 2023). Calibration of a sensor typically occurs both before and after a sensor is launched. Pre-launch calibration helps to ensure that the sensor meets its original design specifications while post-launch calibration is critical to ensure the quality of the data is maintained with appropriate correction applied as the sensor starts to degrade. Post-launch calibration of sensors, also known as vicarious calibration, is a method of correction that utilises 'invariant' natural targets of the earth for calibration (Müller, 2014). Calibration can refer to spectral calibration, radiometric, or geometric corrections and it is important to appreciate how calibration errors may affect data products produced.

Baber (2021) examined the downstream effects of radiometric calibration error across two application areas: crop classification and harmful algal bloom detection. The first study emulated the conversion of Top of Atmosphere (TOA) digital numbers to TOA reflectance for Landsat 8 imagery by introducing errors in the gain factors for each band before the radiometric calibration was performed. Errors applied to each band ranged from -1% to 1%. The resulting TOA reflectance products were then used to conduct a crop classification and based on those results quantified the effect that the errors in gain values for a sensor can have on the resulting crop classification maps. The second study into harmful algal bloom detection examined the influences of atmospheric correction and identified several key takeaways. Firstly, the influence of radiance errors carries over linearly into the final data products meaning that, depending on the algorithm used, radiance errors of ±1% can produce broad false positives or negatives of algal blooms. Secondly, they found that band errors are correlated with each other due to atmospheric correction models using various bands as inputs; therefore, errors in one band can go on to affect the accuracy of other bands. Overall, both studies showed that radiometric error from a single band is traceable but at the same time correlated with radiometric errors from other bands once atmospheric corrections are applied.

Others have looked to demonstrate the potential downstream effects errors can have on earth observation outputs and decision making. Siddiqi et al. (2020a) used the Normalised Difference Vegetation Index (NDVI) as an example and applied a data value-chain approach finding that a 3% error in the reflectance ratio of the red and near infrared (NIR) bands can translate into a difference of 60% for category 5 high vegetation (value ≥ 0.4) value in an NDVI classification. This 60% reduction in the count of image pixels containing high vegetation will lead to an underestimation of dense / healthy vegetation within their region of interest. While class 5 was significantly affected, this study showed that not all classes are affected equally for example class 3 which represented low vegetation ($0.1 \leq NDVI < 0.25$) appeared to be unaffected by the same degradation effects.

2.3 Sensor Degradation

Sensor degradation can be caused by a range of factors, some examples include mechanical issues, exposure of the sensor to UV as well as thermal and electrical effects. While degradation of any sensor is inevitable, these effects can be offset by calibration. Large satellites are able to accommodate on-board calibrators and may have more resources available to conduct on-board calibration. In comparison, nanosatellite or cubesat that overall have less resources available may not have the space required to enable on-board calibration. Therefore, alternative calibration options are required which can include cross-calibration with reference sensors, vicarious calibration campaigns and dedicated calibration sites (Baber, 2021).

The first MODIS instrument aboard the Terra satellite was launched in December 1999 and became operational in 2000, collecting imagery up until October 2022. The second MODIS instrument was launched aboard the Aqua satellite in 2002. Both MODIS sensors operated continuously for more than 18 years. These sensors were developed to help support NASA's Earth Observing System to meet the needs of both the operational and scientific community by extending and enhancing the data records being collected of the Earth. On a regular basis more than 30 products were generated that are distributed globally and are openly available to the global community. These and numerous other products and insights that have been derived from MODIS imagery has supported a range of applications and studies of the Earth's system including monitoring various geophysical and environmental parameters of land, oceans and the atmosphere as well as impacts due to natural and human effects on the Earth's climate (Xiong and Butler, 2020). MODIS underwent a thorough prelaunch calibration process and on-going on-board calibration post launch but as the sensor aged and got closer to its end-of-life, growing concerns were raised regarding sensor degradation and the flow on impacts to derived products.

Wang et al. (2012) examined the impacts of sensor degradation on the MODIS NDVI time series. Their study found that sensor degradation varied by spectral band, view angle and mirror side. Between 2002 and 2010, bands 1 to 3 on the Aqua MODIS sensor only showed a <2% change in TOA reflectance while data collect by the Terra MODIS sensor exhibited substantial degradation, particularly at shorter wavelengths. Specifically, the Terra MODIS blue band had the most pronounced degradation with a decrease of nearly 7% in TOA reflectance at near-nadir view angles over the decade. They expected that by 2013 this degradation would exceed 10% since launch. The study recognised that while this band isn't directly used to calculated NDVI it would impact the calculation of surface reflectance for other spectral bands and eventually other higher-level products including the NDVI product. Next, they examined the NDVI product and found a decreasing trend in NDVI values had occurred between 2000 and 2010 due to the degradation of the blue band. Based on these learnings' corrections could be made to the MODIS collection 6 product that would eliminate the negative bias detected in this study. This demonstrates how important it is that calibration errors be continually assessed throughout the lifespan of a sensor to detect and correct for calibration errors or bias.

2.4 Landsat degradation

With sensor degradation being an inevitable process it is therefore important to understand and where possible model the impacts of degradation so users can understand the limitations of the data or make corrections. As discussed in section 2.2, Baber (2021) explores the effects of degradation through errors in Landsat gain values. Siddiqi et al. (2020b) examines another form of degradation and presents a method of modelling the value of multi-spacecraft earth observation missions with globally dispersed calibration systems for frequent radiometric calibration of earth imaging sensors. The model is called the Calibration integration Value Analysis Framework (CalVIN) and aims to enable integrated analysis of calibration and the value of EO systems. The framework consists of a set of computational modules that obtain simulations of orbits and coverage and instrument performance and combines this information with models of data quality degradation and calibration, and system costs and revenue to compute value. The case study they applied this method to was a hypothetical cubesat constellation of

12 spacecraft with passive imaging sensors viewing the continental United States for a 60-day simulation. As an initial proof of concept, they modelled the Adjusted Effective Data Acquired (AEDA) Value for the constellation which at the beginning of life post initial calibration (no degradation) was 2,912.80 terabytes (TB) of data. After 60 days with 0.25% degradation the AEDA declined to 2,386.54 TB whereas with 1% degradation at 60 days AEDA declined to 1,636.73 TB. In this paper degradation encompassed a combination of factors including thermal, optical, sensor response, and electronics that can all lead to degradation. While there are limitations to this model which requires further research, this is a useful tool for projects in pre-phase A development and clearly highlights the reduction in effective data acquired by sensors as degradation occurs over time with and without calibration.

2.5 Validation

Users' trust in satellite imagery or satellite-derived products is critical for most applications when using EO data, whether it is for analysis by researchers or industry groups, as a visual aid, or as an input in decision making by governments. No matter the application, it is important to ensure that what is reflected in a satellite image or in a derived product is representative of the object or phenomenon being observed. Validation and calibration can also provide useful insights into whether a product is appropriate for a specific use case and can assist in ensuring that data is not being misused or misrepresented. CEOS defines validation as "the process of assessing, by independent means, the quality of the data products derived from those system outputs" (CEOS, 2023). To enable this process of assessment and verification, conventional ground-based observations are required ideally to be collected by calibrated and traceable field instrument and other associated field methodologies. Around the world, great efforts are being made to collect trusted validation data. AERONET, the Terrestrial Ecosystem Research Network (TERN) and IMOS Satellite Altimetry Calibration and Validation Facility are just a few examples of groups or networks that are working to collect high quality reference measurements for calibration or validation.

In Australia, Digital Earth Australia (DEA) produces analysis ready products such as the Landsat and Sentinel-2 surface reflectance products. CEOS defines analysis ready data (ARD) for land as "satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets"(CEOS, 2021). To assist in ensuring DEA ARD surface reflectance products meet the CEOS standard, Geoscience Australia (GA) in 2017 commenced a comprehensive validation program of their Landsat-8 and Sentinel 2 ARD surface reflectance products. Byrne et al. (2021) reports on the phase 1 validation project that ran from January 2018 to June 2019 which focused on capturing ground spectral data on cloud free days coincident with Landsat-8 or Sentinel 2A/B overpasses. Field surveys were conducted over 12 sites distributed across Australia by a network of field teams from various institutions and agencies. When the field site surface reflectance measurements were compared to the Landsat 8 and Sentinel ARD products, both were found to be highly correlated with the field measurements (Landsat 8 R^2 = 0.99, Sentinel 2A = 0.992 and Sentinel 2B = 0.991). They concluded that while some of the individual data points may lie away from the line of best fit that the field data does validate the satellite data. Overall, they concluded that this project was able to validate the DEA's ARD Landsat and Sentinel products to within 10%.

Further validation work by Byrne et al. (2024) conducted a comparison of Landsat-9 and Landsat-8 during an underpass event where field data was collected across five sites with seven validation datasets in total being collected. This work enabled a comparison of the radiometric performance between Landsat-8 and -9 as well as a comparison between ARD processing models from the USGS and DEA. As part of this work CSIRO published field methods used as a technical handbook outlining a community approach to the standardised validation of surface reflectance products (Malthus et al., 2019) to assist in ensuring that these ARD products are meeting the CEOS standard.

Byrne et al. (2021) study found that both the USGS and DEA ARD products had a similar level of accuracy, but that overall the DEA products showed a slightly better level of agreement with the field measurements compared to the USGS products. They also found that at the time of the underpass event, Landsat-8 and -9 were recording near identical EO data. This is backed up by Eon et al. (2023) study who validated Landsat-8 and Landsat-9 during the underpass event as part of the commissioning stage of the newly launched Landsat-9 satellite that aimed to support cross-comparison studies of the two sensors. Using field spectral data collected in North Carolina, USA, they found Landsat-8 and -9 during the underpass event collected consistent data, and both surface reflectance ARD-produced products were in good agreement—only deviating 2% from the ground survey data for the VNIR and 5-8% in the short-wave infrared (SWIR) bands.

3. CASE STUDY - HABITAT CONDITION ASSESSMENT SYSTEM (HCAS)

3.1 Background

HCAS is a system designed by CSIRO for nationally consistent site-scale habitat condition assessments, focused on habitat's capacity to support native biodiversity (Williams et al., 2021). It is trusted and used by commonwealth government to support nature positive policies, including as a performance measure (EN01) in annual reporting on environment and heritage (Department of Climate Change, Energy, the Environment and Water, DCCEEW). We selected HCAS as the use case for our study because:

- it uses several popular public EO datasets as input
- the modelling process represents a time series analysis that combines multiple remote sensing inputs with environmental variables to examine changes over time
- the system is continuously improved to incorporate new EO datasets and generate results at a finer spatial and temporal resolution
- the accuracy and reliability of HCAS' outputs potentially influence policy and decision-making.

The most recent publicly released HCAS version 2 (Harwood et al., 2023; Williams et al., 2023Williams et al., 2021) reports the current best estimate of habitat condition using an 18-year epoch (2001 to 2018) and a time-series of condition change using 18 annual epochs, which are used to derive trends in condition. The 250 m grided products were generated using MODIS data. Currently, the CSIRO HCAS team is working on improving the spatial resolution of the system to 90 m by using Digital Earth Australia's Landsat derived products, to be published as version 3.

To evaluate the potential impact of input EO data quality on HCAS model outputs, we have designed the substitute analysis based on an interpretation of the modelling approaches detailed in the technical reports for HCAS version 2 (Lehmann et al., 2018; Williams et al., 2020, 2021) and through direct discussions with the HCAS team.

3.2 HCAS modelling approach

HCAS uses satellite EO data to estimate habitat condition resulting from anthropogenic disturbance, to provide a consistent estimate for national-level applications. Figure 1 shows a schematic diagram of the approach. The model defines reference conditions using at least 100,000 sites selected across the country that represent the least modified natural areas and ecosystem diversity from primarily terrestrial zones. Environmental variables such as soil properties, climate variability, and landforms are used as covariates to predict remote sensing signatures of the natural landscape. For a given site of interest (i.e. test site), actual remote sensing observations are compared using 20 benchmarks representing local ecosystem reference state variability, selected based on their similarity in modelled reference signatures to the site of interest. Deviations from reference conditions indicate modified ecosystems and reduced habitat quality for indigenous biodiversity.

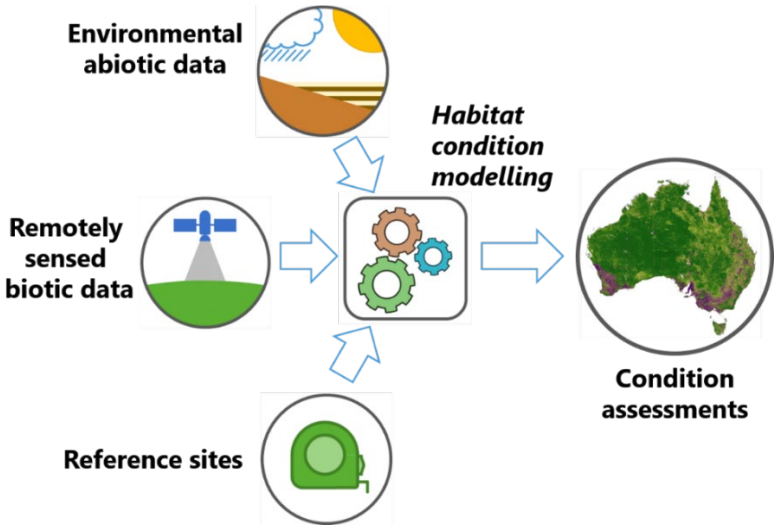

Figure 1: HCAS inputs and modelling approach, reproduced from Lehmann et al. (2021).

The HCAS method relies on two major assumptions about remote sensing data. First is that the type of remote sensing data used can measure key characteristics, such as composition, structure and function dynamics, of a wide range of ecosystems. Due to data availability, previous and current versions of HCAS rely on time series information derived from optical satellites such as MODIS and Landsat. These remote sensing datasets map vegetation covers and their seasonal changes but provides limited information on the structure of the vegetation. This limitation can't be addressed by improving the quality of current input data alone. Incorporation of additional datasets such as lidar-derived canopy height models can significantly enhance the understanding of vegetation structure and lead to a more comprehensive habitat condition assessment, for example.

The second assumption is that remote sensing signatures of natural habitats in reference condition can be predicted by environmental covariates and variation in remote sensing signatures can serve as indicators of habitat condition changes from reference states. Both habitat conditions and remote sensing signatures exhibit natural variability. Statistical models aim to identify trends linking environmental variables and remote sensing data. However, inherent variability in these factors introduces uncertainty into the models. Quality issues with remote sensing data can exacerbate this variability, increasing model uncertainty. Consequently, higher uncertainty reduces the model's ability to differentiate between actual habitat degradation and natural variability. In addition, inaccuracies in remote sensing data can lead to model bias and erroneous evaluations of habitat conditions. This assumption is sensitive to remote sensing data quality issues and therefore, these issues will be the focus of our study.

A key HCAS concept is observed remote sensing distance, defined as the Manhattan distance (sum of pairwise distances between two points in n-dimensional space) calculated within the remote sensing variable space between a test site and a reference site or among reference sites. This multi-dimensional distance measure allows HCAS to summarise differences across multiple remote sensing characteristics. Using data over reference condition sites, the dependence of each remote sensing characteristic, measured by an input remote sensing variable, on environmental covariates is modelled through a multivariate regression. This model enables the prediction of remote sensing variables and subsequent calculation of predicted remote sensing distances for any location with defined environmental covariates. Between a test site and a reference site, the predicted remote sensing distance can be seen as a proxy for similarity of ecosystem types, and the observed remote sensing distance is a measure of similarity of condition characteristics. Intrinsic variability resulting from ecosystem dynamics means that even with natural drivers of disturbance, the predicted remote sensing distances will not have a one-to-one relationship with the observed remote sensing distances.

The HCAS modelling framework accounts for the intrinsic variability through the construction of a reference condition probability density surface (Figure 2, reproduced from Figure 31 in Williams et al. (2021), a two-step process to select not a single but a set of reference sites, known as benchmark sites, for each test site (Figure 3), and the estimation of test site condition as a distance-weighted average of its reference condition probabilities from the comparison to each benchmark site.

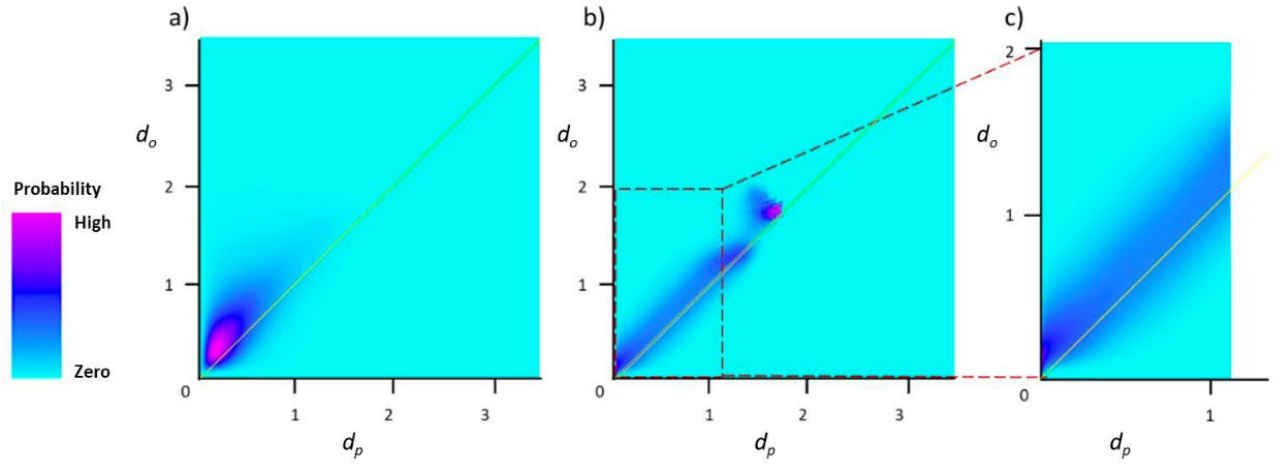

Figure 2: Construction of reference condition probability density surface in HCAS version 2, reproduced from Figure 31 in Williams et al. (2021). a) shows the original 2d histogram of observed (do) vs predicted (dp) remote sensing distances between reference site pairs. b) shows the histogram from a) normalised within bins of predicted distance (dp) so that within a given predicted distance (dp) bin, probabilities over all possible observed distances (do) sum up to 1. c) shows the final probability density surface, generated from removing the part of surface where normalisation reveals unstable behaviour (outside of the red boundary) from b).

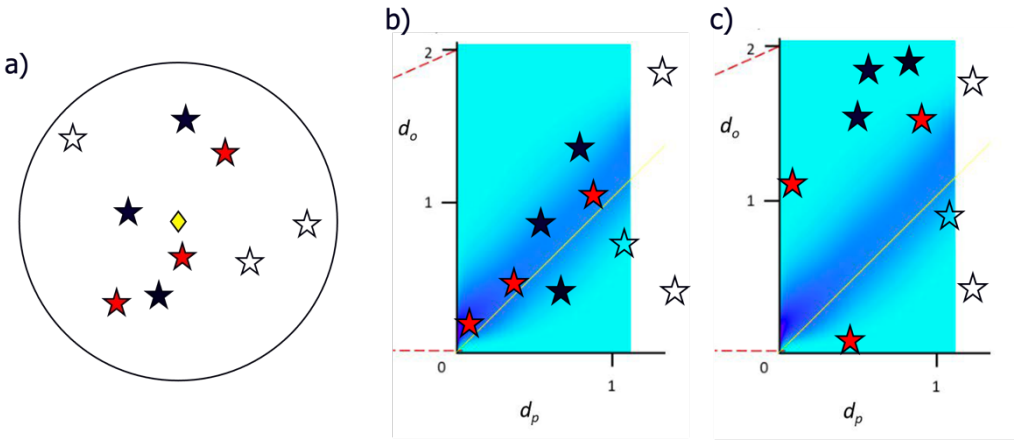

Figure 3: Schematic examples of HCAS condition assessment. a) shows the spatial location of a test site (yellow diamond), with 9 reference condition sites found within a defined distance (R). Out of the nine reference condition sites, six sites (np, filled stars) are first selected as they have the smallest predicted distances to the test site. Sites with large, predicted distances (unfilled stars) are considered more likely to have different ecosystem types to the test site. Subsequently, three sites (nref, filled red stars) are selected as benchmark sites as they have the highest probability values on the reference condition probability density surface, to best represent the dynamics of the ecosystem of the test site as if it were in its reference state. In HCAS version 2, selection radius R is 200km, np is 50 and nref is 20. b) shows the observed and predicted distances between the test site and reference condition sites, overlayed on the probability surface from Figure 2c, for a test site with a high probability of being in reference condition and hence a high condition score (the red filled stars are close to the 1:1 position on the reference condition probability density surface). c) shows the observed and predicted distances for a test site with a low probability of being in reference condition and hence a low condition score (the red filled stars are distant from the 1:1 position on the reference condition probability density surface).

Due to the complexity of the HCAS modelling framework, it is not feasible to emulate the entire workflow in this study to trace the impact of input data errors throughout the model construction and condition assessment process. Instead, we examine selected components of HCAS and conduct analyses to assess the potential scale of change in the outputs caused by varied quality of remote sensing input data.

3.3 Remote sensing data quality considerations

3.3.1 Relevance for the HCAS input components

The environmental covariates used by HCAS (version 3 series) include the Soil and Landscape Grid of Australia (SLGA) products that are optimised to minimise the effects of anthropogenic influences (Searle, 2023) and utilise other TERN and GA products aligned or adjusted to remove anthropogenic signatures. Remote sensing covariates used to generate the original SLGA are not used in the HCAS optimised SLGA model. Therefore, the environmental co-variates are not impacted by remote sensing data quality issues.

Our study will focus on the remote sensing input variables that are used in HCAS. HCAS version 2 uses MODIS collection 6 spanning from 2001 to 2018 as input. Sourced from NASA, this dataset is well-calibrated and offers a high temporal resolution with frequent revisits that capture dynamic changes. However, MODIS data is provided at a low spatial resolution of 250 or 500 m, and the sensor reached the end of its operational lifespan in 2023. Therefore, it can't be relied on for future condition assessment or at finer scale. In addition, the MODIS data product has been affected by sensor degradation throughout its lifespan as found by Wang et al. (2012). The decreasing NDVI, if not corrected, may bias trend analyses.

HCAS version 3 uses data from the Landsat series to provide condition assessment at a higher spatial resolution of 90 m and over a longer time period. Landsat is unique as it has collected high-resolution imagery with a consistent spectral coverage since the mid-1980s and will continue into the foreseeable future. However, the sensor technology within the series has evolved significantly over time, starting from Landsat-5's Thematic Mapper (TM) operational from 1984 to 2013, to Landsat-7's Enhanced Thematic Mapper Plus (ETM+) operational from 1999 to present, and then to Landsat-8's Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) operational from 2013 to present, and the most recent, Landsat-9 OLI-2 and TIRS-2, operational from 2021 to present. Radiometric resolution has seen significant improvement across sensor generations. The other major difference between sensors is the spectral responses. The Landsat 7 sensor has similar spectral responses to

Landsat 5 in the visible, near infrared, and short-wave infrared wavelengths. Landsat 8 and 9 are equipped with similar sensors, with different spectral responses to Landsat 7 and 5. Notably, Landsat 7 experienced a well-documented issue known as the "scan line corrector (SLC) failure" on May 2003. This failure resulted in data gaps in Landsat 7 images since then, which appear as stripes of missing data. Differences in sensors lead to variations in reflectance measurements. These inconsistencies are not due to calibration errors; rather, as data calibration and processing become more precise, these differences become more pronounced. When combining data from different sensors, it may not always be possible or necessary to distinguish differences attributable to sensor variability from those due to calibration procedures. It is, however, important to be aware of the possible sources of difference, so cross-calibration can be carried out properly. When not all differences can be reconciled, the potential impact should be considered. Such consideration becomes more important if additional remote sensing datasets, covering different periods of time or offering measurements at different spatial resolution, are incorporated to complement Landsat in HCAS. We will use Landsat data products as a baseline in our impact analysis in sections 3.4 and 3.5.

Another aspect highlighted by the HCAS use case is that application development using satellite data often leverage expertise that differ markedly from the skills needed to calibrate and preprocess satellite data. In the case of HCAS, statistical modelling and domain expertise such as ecological knowledge are critical. It is neither efficient nor feasible for most satellite data users to undertake sensor calibration and preprocessing tasks such as atmospheric correction themselves. Recognizing this, CEOS launched the Analysis Ready Data Initiative. The goal of this initiative is to provide data that has been pre-processed to meet specific standards, thereby enabling wider adoption and significantly increasing the data's impact.

The MODIS datasets used in HCAS version 2 were pre-processed by NASA and released as 16-day NDVI (MOD13Q1) or surface reflectance (MCD43A4) composites. During the pre-processing, sensor calibration, atmospheric correction and Bidirectional Reflectance Distribution Function (BRDF) corrections have been applied (Schaaf and Wang, 2015). HCAS version 3 uses GA DEA's Landsat-derived annual summary products, generated from GA's Landsat surface reflectance ARD. The pre-processing to ARD includes sensor calibration carried out by USGS and atmospheric, BRDF and terrain illumination corrections implemented by GA (Fuqin et al., 2019).

Some preprocessing steps, such as atmospheric correction, utilize information measured across the spectrum. This can propagate calibration errors non-linearly into the processed data product, introducing correlated errors among spectral bands. The manifestation of these errors depends on the algorithms used. In our study, we aim to focus on the processes that generate the information products supporting decision-making. Therefore, we will not attempt to model error propagation in preprocessing but acknowledge that the typical starting point for analysis is the ARD. We will investigate how inaccuracies in surface reflectance products influence the analyses that assess land surface conditions.

Finally, while the selection of HCAS reference sites does not depend on remote sensing data, it is important to note that a large number of sites spread across Australia are selected to represent diverse ecosystem types. These sites exhibit a wide range of observable characteristics, reflected in the extensive range of values and spectral signatures detected by sensors. Factors such as a site's longitude, latitude, and topographic settings can influence observing conditions—like sun angle and geometry—which in turn affect the quality of reflectance measurements due to the precision achievable or the absence of relevant corrections. In our study, we will utilize subsets of condition reference sites from HCAS version 2 to extract Landsat data. This will help us construct baseline distributions of remote sensing variables and assess the spatial and spectral type dependence of any biases arising from input data issues.

3.3.2 Contributing factors to data quality issues

Satellite data quality issues can broadly be categorized into geometric and radiometric concerns. Geometric issues, such as distortions and geolocation accuracy can arise from factors including sensor optic errors, platform instability, and inaccuracies during data processing. These errors may result in misplacement of measurements, causing observed features to be inaccurately represented at their actual locations on Earth's surface. For typical MODIS and Landsat observations, geolocation errors are usually confined to within one pixel (Dwyer et al., 2018; Wolfe and Nishihama, 2009). While these errors should be acknowledged and minimized, they are not expected to significantly affect the overall accuracy of HCAS assessments at 250 m or 90 m grids.

Radiometric errors directly impact reflectance measurements and may originate from issues with sensor radiometric calibration and stability. Inaccurate sensor calibration can lead to incorrect measurements of radiance or reflectance, while stability concerns over time may affect longitudinal studies. Sensor calibration is typically managed by satellite mission operators. Furthermore, errors in data preprocessing—such as incorrect atmospheric correction, geometric correction (including BRDF

and terrain), and uncorrected effects of clouds and their shadows—can introduce artifacts in reflectance measurements. Composite products may also encounter issues from integrating data from multiple sources, sensors, or varying observing conditions without adequate corrections.

When higher-level products derived from ARD are used, additional modelling errors may be introduced. These can result from the incorrect choice of algorithms, misinterpretation of input data, and overfitting of models to data. For instance, HCAS uses fractional cover derived from MODIS and Landsat in versions 2 and 3, respectively. For a fractional cover product, an unmixing algorithm is applied to convert each multispectral reflectance measurement into proportions of green (leaves, grass, and growing crops), brown (branches, dry grass or hay, and dead leaf litter), and bare ground (soil or rock) components (Scarth 2010, Lymburner 2021). The accuracy of these fractional measures depends on the representativeness of the spectral members used in the algorithm for the actual spectral signatures of the cover components at a specific site. The algorithm may perform unexpectedly in environments that differ from where it was trained. A recent study by Sutton et al. (2022) focused on accessing the accuracy of Landsat Vegetation Fraction Cover products in Australia Drylands. The study found that public products currently provided in Australia are generally consistent and have reasonable accuracy when compared to available field data. When the DEA fractional cover product, used in HCAS, was compared to the State-wide Landcover and Tree Study (SLATS) field dataset which consists of 4,207 unique observations, it produced an overall root mean square (RMSE) error of 14.39. Of the three fractions assessed the green fraction had the lowest RSME of 9.80, followed by the bare soil fraction (14.28 RMSE) and the brown fraction (17.93 RMSE). Studies like Sutton et al. (2022) can help to provide context for error and uncertainties that may be introduced in the HCAS algorithm via the input datasets.

3.4 Impact of sensor difference and data issues on HCAS input variables

3.4.1 HCAS version 3 input variables

HCAS version 3 uses DEA annual Geometric Median and Median Absolute Deviation (GeoMAD) (Geoscience Australia, 2022) and Fractional Cover Percentiles (Lymburner, 2021) products from 1988 to 2022 to derive remote sensing variables (Levick et al., 2023). Following protocols described in Lehmann et al. (2018), a total of 14 remote sensing variables are preliminarily selected, which include five spectral indices calculated from average annual geometric median spectral bands, three averaged median absolute deviations (MADs), three averaged fractional cover percentiles and three averaged fractional cover percentile ranges (Table 1).

Table 1: Preliminary selection of remote sensing variables for HCAS version 3, extracted from Table 5 in Levick et al. (2023).

Variable	Description
kNDVI	Kernal Normalised Difference Vegetation Index, 35-year average of annual geometric medians
NBR	Normalised Burn Ratio, 35-year average of annual geometric medians
NDMI	Normalised Difference Moisture Index, 35-year average of annual geometric medians
NDWI	Normalised Difference Water Index, 35-year average of annual geometric medians
BSI	Bare Soil Index, 35-year average of annual geometric medians
Sdev	The cosine (spectral) distance from the median absolute deviation of the geometric median of the multi-band pixel composite mosaic (i.e. SMAD)
Edev	The Euclidean distance from the median absolute deviation of the geometric median of the multi-band pixel composite mosaic (i.e. EMAD)
Bcdev	The Bray-Curtis distance from the median absolute deviation of the geometric median of the multi-band pixel composite mosaic (i.e. BCMAD)
PV10	10^{th} percentile of intra-annual photosynthetic cover fractions, averaged over 35 years

Variable	Description
PVrange	Inter-percentile range (10^{th} – 90^{th}) of intra-annual photosynthetic cover fractions, averaged over 35 years
NPV10	10^{th} percentile of intra-annual non-photosynthetic cover fractions, averaged over 35 years
NPVrange	Inter-percentile range (10^{th} – 90^{th}) of intra-annual non-photosynthetic cover fractions, averaged over 35 years
BS10	10^{th} percentile of intra-annual bare soil cover fractions, averaged over 35 years
BSrange	Inter-percentile range (10^{th} – 90^{th}) of intra-annual bare soil cover fractions, averaged over 35 years

All variables listed are likely to be affected by radiometric issues in the Landsat surface reflectance products. The GeoMAD products from Landsat 5, 7, and 8 have been combined without accounting for their differences. In Section 3.4.2, we will compare these variables as measured by individual Landsat sensors to assess the extent of these differences. However, a comparison of the fractional cover variables will not be conducted, as annual products from individual sensors are not readily available.

As previously mentioned, when different datasets are merged in the analysis, distinguishing inconsistencies due to sensor differences from those due to calibration or processing errors is not always possible. In Section 3.4.3, we will introduce perturbations into the baseline measurements from Landsat 8 to evaluate the impact of various types of errors. This assessment will focus solely on the effects on spectral indices, thanks to the straightforward nature of their calculation.

3.4.2 Remote sensing variables measured by different sensors

For this assessment, we selected around 12,000 sites from a random selection of 200,000 reference condition sites provided by the HCAS team. The selection employed a spatial clustering algorithm (DBSCAN) first and then picked one random site from each cluster. Through this process, spatial clustering is reduced. The sites are further split into two groups, East and West with roughly 6000 sites each (Figure 4).

For each site location, Landsat 7 GeoMAD and Landsat 8 GeoMAD measurements within a 90 by 90 m spatial window, for 2014 to 2018 (inclusive), are extracted and averaged. Linear regression between Landsat 7 and Landsat 8 variables is then conducted for the two groups separately (examples in Figure 5). In addition to comparing the individual remote sensing variables, we also compare remote sensing distances calculated using normalized (to mean of 0 and standard deviation of 1) input variables over pairs of reference sites. For the remote sensing distance comparison, only the eight GeoMAD derived variables are used. Out of all pairs of reference sites, 10% are used in the linear regression to reduce computation requirement.

The linear regression results are listed in Table 2. All spectral indices measured by Landsat 7 and Landsat 8 show strong correlations, with R-squared values close to 1. The largest difference is observed in the NDWI of Group 1, with a best-fit slope of 0.9 between Landsat 7 and Landsat 8. Intra-annual variability measurements, such as MADs, are less well correlated, with the lowest R-squared value being 0.74 for Sdev (or SMAD) of Group 2. This discrepancy may be due to various factors, including the use of all spectral bands in the calculation and differences in observation times. When only a small number of observations are available, the MADs may be biased, so the difference cannot be attributed solely to sensor differences.

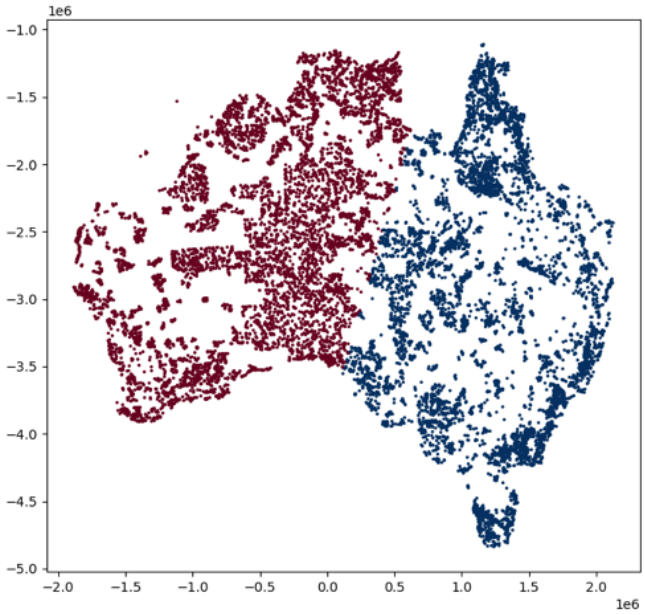

Figure 4: Locations of 12,736 sites selected from HCAS reference condition sites, split into two spatial groups, Group 1 (red) and Group 2 (blue).

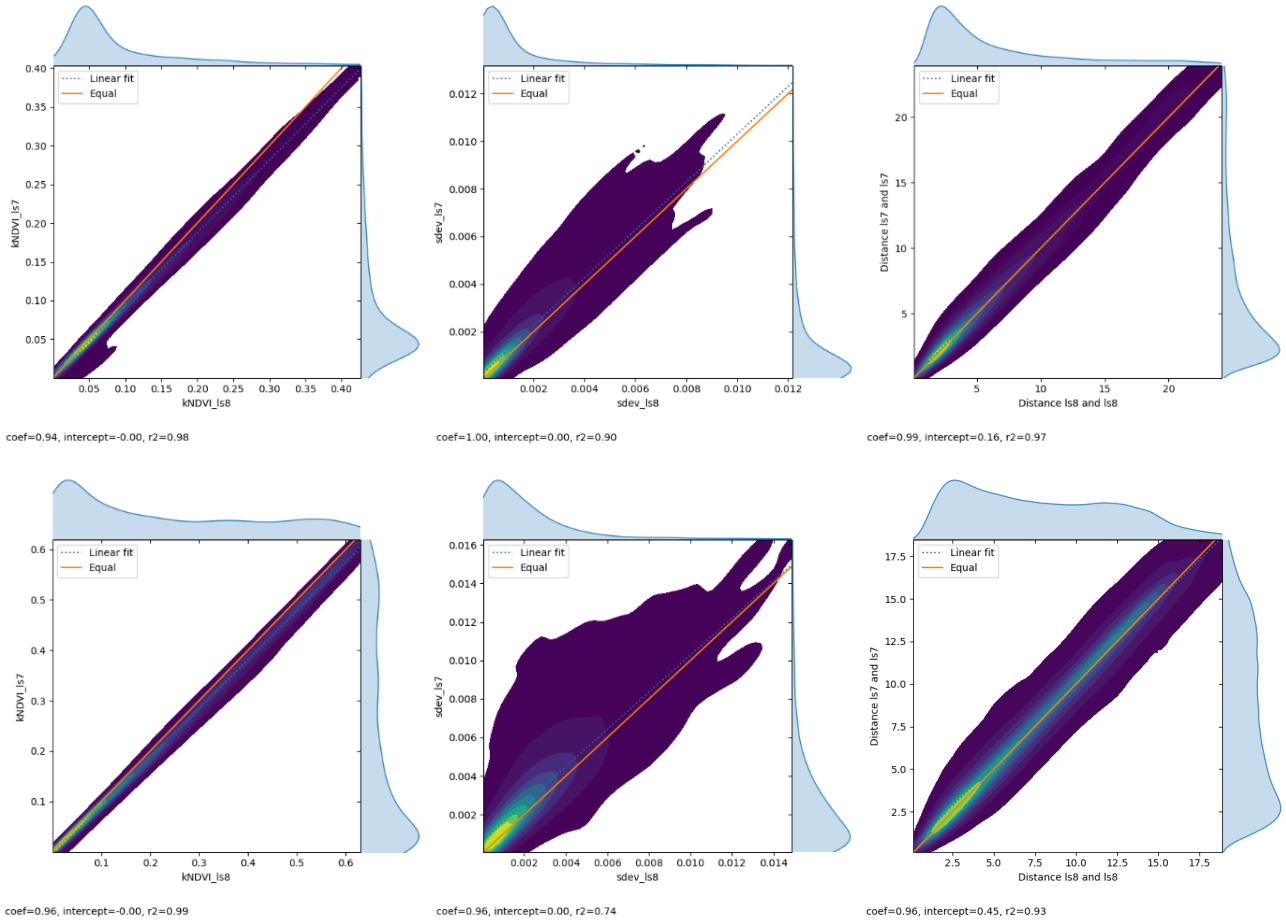

Figure 5: Linear regressions between selected variables measured by Landsat 7 (y-axis) and Landsat 8 (x-axis) over reference sites. The variables shown from left to right are kNDVI, Sdev, and remote sensing distance. The top and bottom rows show the comparisons for Group 1 and Group 2 respectively.

In contrast to individual remote sensing variables, the remote sensing distances calculated from normalized Landsat 7 and Landsat 8 variables are highly correlated. It is expected that scale differences in individual variables are minimized through the normalisation process, making remote sensing distance a more robust measure compared to individual variables.

The differences between Landsat 7 and Landsat 5 measured variables are at a similar level to differences between Landsat 7 and Landsat 8, therefore we won't include the comparison here.

We note that the arbitrary split into west and east groups is simply designed to show that the correlation between sensors is quite sensitive to the spectral distribution. Different correlations are likely to be observed for other types of groupings, e.g. between low and high reflectance areas.

*Table 2: Linear regression coefficients between variables measured by Landsat 7 (Var_ls7) and Landsat 8 (Var_ls8), where Var_ls7 = a*Var_ls8 +b. The intercept (b) values are not listed when they are close to 0. RS dist = Remote Sensing Distance.*

Site group		kNDVI	NBR	NDMI	NDWI	BSI	Sdev (SMAD)	Edev (EMAD)	Bcdev (BCMAD)	RS dist
Group (west)	1	**a=0.94, r2=0.98**	a=1.00, r2=0.99	a=1.00, r2=0.98, b=0.01	**a=0.90, r2=0.99, b=-0.05**	a=0.95, r2=0.99, b=0.02	a=1.00, r2=0.90	**a=0.95, r2=0.88**	a=0.99, r2=0.94	a=0.99, r2=0.97
Group (east)	2	a=0.96, r2=0.99	a=1.00, r2=0.99	a=0.99, r2=0.99	a=0.93, r2=0.99, b=-0.04	a=0.97, r2=0.99, b=0.01	**a=0.96, r2=0.74**	**a=0.92, r2=0.87**	a=0.98, r2=0.86, b=0.01	a=0.96, r2=0.93

3.4.3 Impact of data perturbation

As discussed earlier, we use surface reflectance as the starting point for our analysis since it is provided as analysis-ready data according to CEOS standards. In this section, we introduce perturbations into the surface reflectance products and perform linear regression between remote sensing variables derived from perturbed and unperturbed data. A low correlation would indicate that the perturbation has significantly altered the measured remote sensing signal, rendering it unreliable. The threshold at which remote sensing measurements become unusable depends on the accuracy required by the application.

When introducing perturbations, we consider two different types of offsets added to reflectance values. The first is a 10% relative offset, equivalent to scaling individual reflectance measurements by 1.1. This type of error might result from an incorrect gain value used during the conversion of digital numbers to radiometric measurements. The scale of this error is chosen to be slightly larger than the uncertainty measured in Byrne et al. (2021), representing the best quality ARD available.

The second type of offset involves adding 0.01 to reflectance values. This results in a larger relative change when reflectance is low and a smaller relative change when reflectance is high. For reference, the mean reflectance in the blue, green, and red spectral bands across all reference sites is around 0.1, while the mean reflectance in the NIR, SWIR1, and SWIR2 bands is around 0.02. This type of error might be introduced by an incorrect atmospheric correction.

In reality, atmospheric correction is likely to introduce correlated errors between spectral bands. For any individual surface reflectance measurement, several factors contribute to its uncertainty. This analysis aims to illustrate how different types of perturbations can result in varying biases in remote sensing variables. It's important to note that the error distribution in real datasets will be much more complex.

Table 3 and Table 4 summarise the linear regression results for perturbed reflectance in individual spectral bands. The 10% relative perturbations tend to result in an offset (intercept of up to 0.06) in the variables, but the perturbed variables remain highly correlated with the original values (R-squared close to 1). In contrast, the offset of 0.01 reflectance values lead to lower

correlation (e.g. in NDWI). Since the NIR band is used in the calculation of all spectral indices, perturbation in NIR leads to larger changes in remote sensing distances compared to perturbation in other bands.

The remote sensing variables that are less sensitive to perturbation in reflectance values may be considered more robust against input errors. Our analysis shows that the remote sensing distance is a relatively robust measurement as the distances calculated from perturbed reflectance remain highly correlated with the unperturbed distances.

Table 3: Linear regression coefficients between perturbed (Var_perturbed) and original (Var_original) variables, where Var_perturbed = a*Var_original+b. In each row, 10% relative offset is added to surface reflectance of one spectral band leaving other bands unchanged. Regression results for indices that show relatively large change are highlighted in bold. All R-squared values are close to 1, therefore not listed.

Perturbed band	kNDVI	NBR	NDMI	NDWI	BSI	RS Distance
Blue	-	-	-	-	a=1.00, b=-0.01	a=1.00, b=0.01
Green	-	-	-	**a=1.01, b=0.04**	-	a=1.00, b=0.00
Red	a=0.98, b=-0.02	-	-	-	a=1.00, b=0.01	**a=1.01, b=-0.05**
NIR	a=1.01, b=0.02	**a=0.97, b=0.05**	**a=0.99, b=0.05**	**a=0.98, b=-0.04**	**a=0.99, b=-0.04**	**a=0.99, b=0.06**
SWIR1	-	-	**a=1.01, b=-0.05**	-	a=1.00, b=0.03	a=1.00, b=-0.02
SWIR2	-	**a=1.03, b=-0.05**	-	-	-	a=1.00, b=-0.01

Table 4: Linear regression coefficients between perturbed (Var_perturbed) and original (Var_original) variables, where Var_perturbed = a*Var_original+b. In each row, an offset of 0.01 is added to surface reflectance of one spectral band leaving other bands unchanged. Regression results for indices that show relatively large change are highlighted in bold. R-squared values that are close to 1 are not listed.

Perturbed band	kNDVI	NBR	NDMI	NDWI	BSI	RS Distance
Blue	-	-	-	-	a=1.02, b=-0.02	a=1.00, b=0.00
Green	-	-	-	**a=0.99, b=0.04**	-	a=1.00, b=-0.02
Red	**a=0.85, b=-0.00**	-	-	-	a=0.92, b=0.03	a=1.01, b=-0.04
NIR	a=1.00, b=0.01	a=1.00, b=0.03, r2=0.97	a=1.03, b=0.03, r2=0.95	**a=0.87, b=-0.08**	a=1.02, b=-0.02	**a=1.00, b=0.07**
SWIR1	-	-	**a=0.90, b=-0.03, r2=0.90**	-	a=0.92, b=0.03	a=0.99, b=0.02
SWIR2	-	**a=0.93, b=-0.03, r2=0.93**	-	-	-	a=1.00, b=0.00

3.4.4 Summary of findings

Our analyses above indicate that simple spectral indices are highly sensitive to potential errors in surface reflectance data. Different types of errors in surface reflectance can lead to varying changes in spectral indices or remote sensing distance calculations. The impact on remote sensing variables depends on the spectral shape of the target, suggesting that the spectral shape should be considered in uncertainty estimates whenever possible. Furthermore, in the HCAS case, using remote sensing distance calculated from normalized variables appears to mitigate some impacts of input data errors.

3.5 Impact of data issues on condition assessment

The ultimate goal of HCAS is to provide reliable habitat condition assessments, making it crucial to understand how input data impacts the assessment results. However, replicating the complete workflow of HCAS to conduct thorough error propagation is not feasible in our study. To evaluate how errors in surface reflectance might affect condition assessments, we developed a significantly simplified model focusing on the key metric of remote sensing distance. This evaluation should not be seen as a substitute for comprehensive error propagation but rather as a demonstration that input errors can nonlinearly affect condition assessments. By illustrating these potential impacts, we emphasize the importance of accurate input data in achieving reliable habitat condition assessments.

3.5.1 A simplified condition assessment model

As described in section 3.2, HCAS uses remote sensing distance as a key metric for condition assessment. Between a test and a reference site, the predicted remote sensing distance is a proxy for similarity of ecosystem types, and the observed remote sensing distance is a measure of similarity of condition characteristics. To account for the intrinsic variability of natural ecosystems, a multi-step process is used to select reference sites for each test location. Errors in input remote sensing variables would affect the modelled reference condition probability surface, the selection of reference sites, and the measured remote sensing distances, all leading to changes in condition assessment scores.

For our simplified model, we assume most sites and their spatial neighbours have the same ecosystem type. We used the Landsat 8 GeoMAD spectral measurements retrieved in section 3.4.2 over the 12,000 reference sites as our baseline and calculated the observed remote sensing distances between all neighbouring reference site pairs (within 20km, Figure 6a). We used only spectral indices, i.e. kNDVI, NBR, NDMI, NDWI, and BSI, in the remote sensing distance calculation, so that when we introduce errors into the spectral measurements later we can calculate the perturbed remote sensing distances. The measured distance distribution (Figure 6b) peaks at a small positive value and has a long tail extending into large distances. The peak value of non-zero reflects the intrinsic variability of ecosystems, and the large values can be understood as the result of a small fraction of site pairs being distinct, in terms of both ecosystem type and condition. This distribution is similar to the vertical profile (along the y-axis) at a small, predicted distance (close to 0 on the x-axis) in the condition probability density surface shown in Figure 2c.

In the next section, we examine how changes in input variables lead to changes in this distance distribution. To measure the change in condition assessment, we assign 90 percentile baseline distance as a threshold (Figure 6b) whereby site pairs with observed distances smaller than this threshold are assumed in similar condition and site pairs with larger observed distances are assessed to have different conditions.

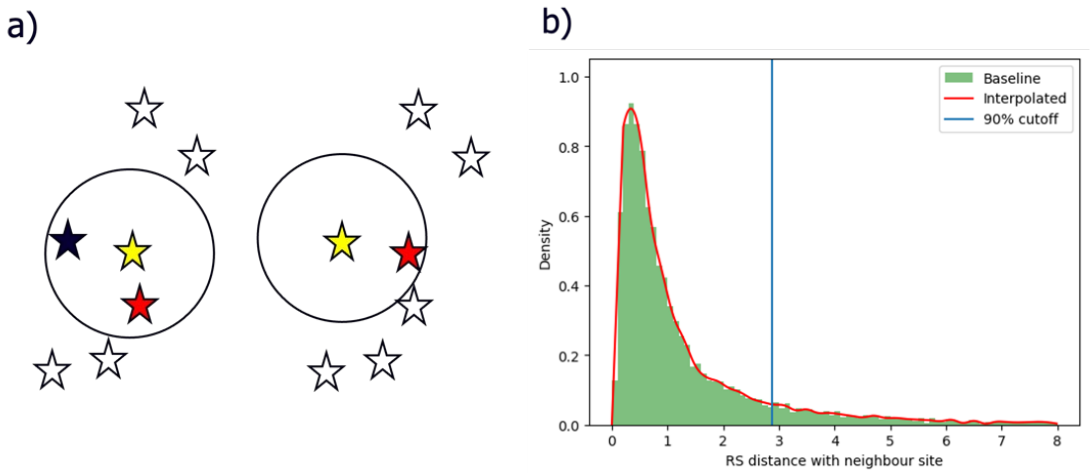

Figure 6: Simplified condition assessment model. a) Schematic example of site pair selections. For each site (yellow star), a single comparison site (red star) is selected as the closest site within a 20km radius. b) Density distribution of the observed remote sensing distances measured for all site pairs. 90 percentile threshold is marked by the blue vertical line.

3.5.2 Impact of data perturbation

Two types of perturbations are introduced into surface reflectance data and their impact on the distance distribution and condition similarity assessments are measured.

For the first perturbation experiment, randomly distributed, with mean of 0 and standard deviation of 5% relative errors are added to the surface reflectance values for all six spectral bands. The perturbed surface reflectance values are used to derive spectral indices and subsequently, remote sensing distance measurement between neighbouring reference site pairs. This altered remote sensing distance distribution is shown in comparison to the baseline distribution in Figure 7. Significant changes including a shift towards larger distance values are observed. While most of the site pairs estimated to have similar conditions in the baseline model remain to be in similar condition, i.e. with distance measures lower than the threshold, if we take the distance to the threshold as a measure of confidence, the shift towards large distance means a significant number of condition assessments have reduced confidence or higher uncertainty. Moreover, the binary condition assessment result for about 2% of the sites changed from similar to their neighbour to different to their neighbour and about 1% of the site assessments changed in the other direction.

When calculating the altered remote sensing distances, since the value distribution of input indices is different to the distribution of the original indices, we have the choice to either apply the original normalisation factors to the indices (Figure 7a) or generate new normalisation factors (Figure 7b). For the first perturbation experiment, the normalisation procedure choice doesn't appear to affect the change trend observed.

For the second perturbation experiment, randomly distributed, with a mean of 0 and standard deviation of 0.01, reflectance errors are added to surface reflectance values of all six spectral bands. This time, the result is sensitive to normalisation procedure chosen (Figure 8). When the original normalisation factors are applied on the altered variables, the measured remote sensing distances increased significantly (Figure 8a). About 6% of the sites changed from similar to their neighbour to different to their neighbour and about 1% of the site assessment changed in the other direction. When a new set of normalisation factors are calculated using perturbed input (Figure 8b), the altered remote sensing distance distribution appears to be more similar to the original distribution. However, change in condition assessment remains significant. About 1% of the sites changed from similar to their neighbour to different to their neighbour and about 6% of the site assessment changed in the other direction.

Since a random set of errors was introduced, we ran the experiments a few times to confirm the order of change remains similar. We have also tested with a larger number (~20,000) of sites to confirm that the result is not sensitive to the selection of neighbouring sites in our sample.

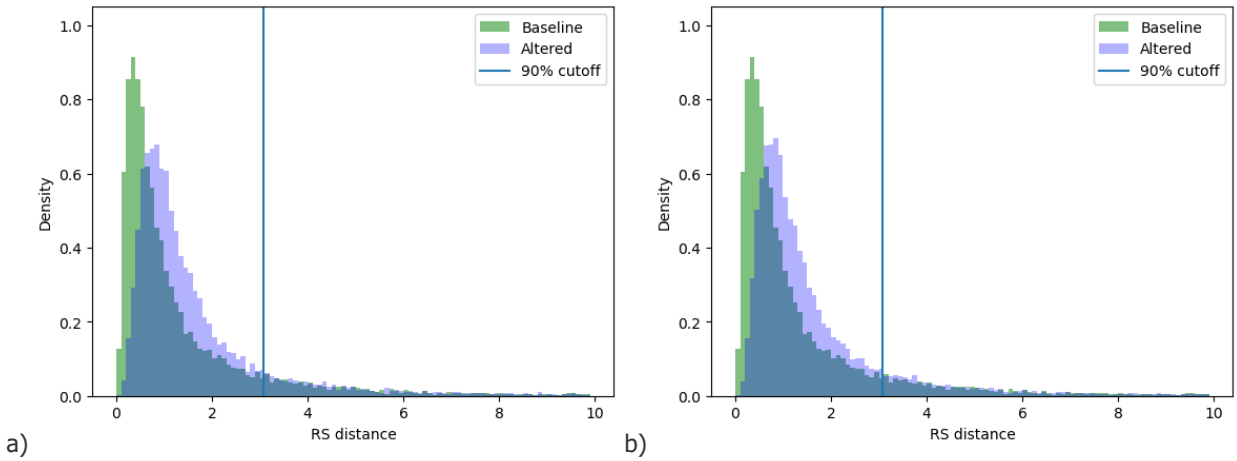

Figure 7: Distribution of baseline and altered remote sensing distances between reference site pairs, when altered distribution is generated by adding random relative errors with mean of 0 and standard deviation of 5% to reflectance values in all spectral bands. 90 percentile of the baseline distribution is marked by vertical blue lines. a) Normalisation factors derived from the original distributions are used to normalise altered variables. b) Normalisation factors are derived from the altered variables and applied.

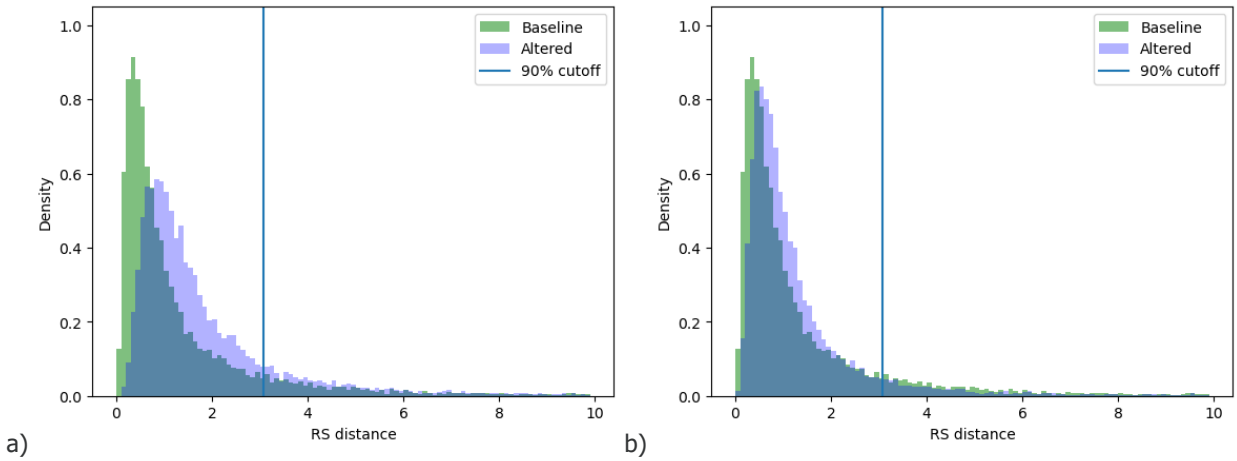

Figure 8: Distribution of baseline and altered remote sensing distances between reference site pairs, when altered distribution is generated by adding random reflectance errors with mean of 0 and standard deviation of 0.01 to reflectance values in all spectral bands. 90 percentile of the baseline distribution is marked by vertical blue lines. a) Normalisation factors derived from the original distributions are used to normalise altered variables. b) Normalisation factors are derived from the altered variables and applied.

3.5.3 Summary of findings

Our simplified condition assessment model, derived from elements of the HCAS framework, allowed us to explore how errors in surface reflectance data might affect condition assessments through changes in remote sensing distances measured between reference sites. This result should not be interpreted as a comprehensive uncertainty assessment for HCAS output, as our process does not account for all remote sensing input variables, the condition probability surface modeling involving environmental variables, or the sophisticated reference site selection process employed in HCAS. However, we have demonstrated that condition assessments and their associated uncertainties are likely impacted by errors in input remote sensing data. At a large scale, changes in condition assessments across a few percent of sites are significant.

Our assessment applies to scenarios where a reference condition probability surface is developed using variables measured by one sensor (e.g., Landsat 8), and condition assessments are conducted using variables measured by a different sensor. This approach might be necessary when the second sensor lacks sufficient spatial or temporal coverage for building a reference condition model but provides observations for assessing conditions at specific locations or time periods of interest. Any bias or increased scatter in data from the second sensor may lead to errors or reduced confidence in the condition assessment.

Our analysis indicates that different types of error distributions in input data will result in different biases in the output. In practice, numerous factors contribute to errors in remote sensing measurements, and errors in spectral bands may be correlated. It is worth noting that annual statistical products are used as input variables in HCAS. This is because intra-annual and inter-annual variabilities are important characteristics of land covers responding to seasonal and year-to-year changes in climate, including variations in temperature, precipitation, and sunlight availability. When combining observations across a year into an annual product, sensor noise and degradation, changes in observation conditions like solar angle, or uncorrected effects of clouds and their shadows all contribute to increased noise in data. The uncertainties from remote sensing input variables may also interact with other sources of uncertainty from additional input datasets and the modeling process in HCAS, complicating the influence on condition assessment results.

We have not considered remote sensing input variables such as the MADs (Median Absolute Deviations) and fractional covers, for which error propagation is more complex than for spectral indices. For fractional cover, the accuracy of the model is sensitive to how well the endmembers represent the environments being mapped, therefore model errors contribute to the overall error budget.

Based on our analysis, it is recommended that a comprehensive uncertainty assessment of HCAS should include uncertainties in the input remote sensing variables. The error distribution and subsequent impact on HCAS results may depend on the spectral shape and the ecosystem type of the observation target.

4. DISCUSSION

We have used HCAS as a case study to assess the impact of input remote sensing data quality on derived insights. While it is expected that input data quality will impact the reliability of a derived product, the exact level of impact can vary significantly based on the model used and is often unknown. For remote sensing applications sharing relevant analysis features with HCAS, several factors may be considered:

- When time series input is used, temporal and spatial consistency across time is essential. Any type of change over time, including sensor degradation, calibration shift, increased or decreased noise will impact comparisons made between time periods.
- When multiple remote sensing datasets are combined in analysis, it is important to understand, and correct for when possible, their intrinsic differences resulting from sensor difference and the preprocessing algorithms applied.
- Many factors contribute to uncertainties in remote sensing data, from sensor characteristics to preprocessing algorithms used. For preprocessed data such as surface reflectance, errors are often correlated between spectral bands and may depend on the spectral shape of observation targets. This complicates error estimation especially for users who don't have access to details of sensor specifications and the preprocessing methods applied.
- Uncertainty estimation of derived products is further complicated when sensing data is combined with other type of datasets, such as environmental and climate variables in analysis. This complexity is best addressed as part of the modeling process to ensure that all sources of uncertainty are accounted for, and their interactions are understood.

While some validation campaigns are periodically conducted (such as DEA surface reflectance validation), remote sensing data providers typically do not offer comprehensive error assessments of their data, even when data is provided as analysis ready. For large public missions, information on sensor degradations and issues is published (e.g., Landsat and Sentinel-2), but these details can be difficult for non-experts to interpret. For commercial missions, such information is often not available.

Users developing solutions using remote sensing data may not be aware of possible data issues or may be left to empirically determine data uncertainty themselves, which is not always feasible. When input data quality issues are not considered during information product generation, it becomes challenging to interpret the derived products. At a minimum, the reliability of the derived information may be under or overestimated, potentially leading to incorrect assessments.

In conclusion, it is recommended that a comprehensive uncertainty assessment include input data quality considerations. The reliability of model outputs depends significantly on the accuracy and consistency of input data, and addressing these issues is crucial for reliable habitat condition assessments and other remote sensing applications.

In our study, we haven't explored the follow-on implication of incorrect habitat assessment, partly because this will depend on how the information will be used in decision making. Understanding the level of uncertainty in information products is crucial for correctly interpreting and using the information in decision-making. In scenarios where additional in situ data is collected to verify remote sensing-based assessments, high uncertainty in remote sensing derived information can lead to increased effort and costs for verification. Incorrect decisions may also result in negative consequences, such as inefficient allocation of funds to address perceived problems and, more seriously, a loss of trust in the decisions made and the information used.

One approach that can be adopted to model the economic impact of a remote sensing solution was proposed by Siddiqi et al. (2020b). The study used a decision tree to evaluate the cost implication of remote sensing-based harmful algal bloom monitoring. Outcomes were compared between scenarios when no monitoring is implemented and when monitoring is implemented and may lead to correct or false assessment. Changes in input data quality will result in changes in the rate of false positives and false negatives, and subsequently the expected outcome of each branch and impact the evaluation of whether a monitoring capability is worthwhile.

4.1 Implications and insights for remote sensing data providers

We believe it is the responsibility of the data providers to provide data quality assessment and information to their users. If it is not feasible to provide an uncertainty estimate for every measurement taken, it's important to communicate the level of uncertainty and limitations of data under different conditions. When data is misinterpreted and misused, it leads to a loss of trust in specific data sources and potentially the use of EO data in general.

To provide the most up-to-date data quality assessment, data providers must conduct ongoing and regular calibration and validation of their sensor and data products. Validation at a single point in time is not sufficient to support time series analysis and change monitoring applications.

A network of resources shared between data providers could potentially help reduce the cost and technical barriers to implementing continued calibration and validation, especially for small missions. A consistent approach to communicating calibration and validation and data quality issues will also help raise user awareness and ensure trust in data.

4.2 Implications and insights for remote sensing data users

Users of remote sensing data should understand that all data comes with inherent uncertainty and many factors, such as satellite operation, observing conditions, and data processing methods applied, influence the quality of remote sensing data. While new algorithms and increased computational power are enabling more accurate analysis-ready data, some artifacts, atmospheric effects, and observational variations cannot be entirely corrected. These data quality issues impact the reliability of modelled results and should be considered when assessing the uncertainty of output products. Application developers need to communicate the uncertainty of information products to end users, making it essential to consider the impact of input data uncertainty in their work.

In many cases, additional independent validation of products derived from remote sensing data is necessary. This involves collecting validation data, such as biophysical quantities, in ways that are comparable to satellite observations. Ongoing monitoring campaigns are also beneficial to ensure that changes in input data over time do not lead to discrepancies in derived products.

5. ACKNOWLEDGEMENTS

We are grateful to the generous support provided by the HCAS team during our study, in particular, Dr Kristen Williams, providing reference sites data and sharing insights on the HCAS modelling approach and current work to improve the system. We would like to thank Geoscience Australia DEA team for discussing and providing information on DEA datasets. Finally, we would like to thank Dr Matt Garthwaite and Andy Allen for reviewing and providing feedback to improve our report. This study was funded by the CSIRO Centre for Earth Observation.

6. REFERENCES

Baber S. 2021. The Impact of Radiometric Calibration Error on Earth Observation-supported Decision Making. Bachelor's Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

Byrne G, Walsh A, Thankappan M, Broomhall M, Hay E. 2021. DEA Analysis Ready Data Phase 1 Validation Project: Data Summary. Geoscience Australia. https://doi.org/10.26186/145101

Byrne G, Broomhall M, Walsh A.J, Thankappan M, Hay E, Li F, McAtee B, Garcia, R, Anstee J, Kerrisk G et al. Validating Digital Earth Australia NBART for the Landsat 9 Underfly of Landsat 8. Remote Sensing. 2024; 16(7):1233. https://doi.org/10.3390/rs16071233

CEOS. 2023. Working Group on Calibration & Validation (WGCV). Accessed February 4, 2024. https://ceos.org/ourwork/workinggroups/wgcv/

CEOS. 2021. CEOS Analysis Ready Data Governance Framework. Accessed November 13, 2023. https://ceos.org/ard/files/CEOS_ARD_Governance_Framework_18-October-2021.pdf

Dwyer J.L, Roy D.P, Sauer B, Jenkerson C.B, Zhang H.K, Lymburner L. Analysis Ready Data: Enabling Analysis of the Landsat Archive. Remote Sensing. 2018; 10(9):1363. https://doi.org/10.3390/rs10091363

Eon R, Gerace A, Falcon L, Poole E, Kleynhans T, Raqueño N, Bauch T. Validation of Landsat-9 and Landsat-8 Surface Temperature and Reflectance during the Underfly Event. *Remote Sensing*. 2023; 15(13):3370. https://doi.org/10.3390/rs15133370

Fuqin L, Jupp D.L.B, Sixsmith J, Wang L, Dorj P, Vincent A, Alam I, Hooke J, Oliver S, Thankappan M. 2019. GA Landsat 8 OLI/TIRS Analysis Ready Data Collection 3. Dataset. Geoscience Australia, Canberra. Accessed May 15, 2024. https://pid.geoscience.gov.au/dataset/ga/132317

Geoscience Australia. 2022. DEA Geometric Median and Median Absolute Deviation (Landsat). Geoscience Australia, Canberra. Accessed May 14, 2024. https://dx.doi.org/10.26186/146261

Harwood T.D, Lehmann E.A, Giljohann K.M, Williams K.J, Liu N, Ferrier S, Ware C, Donohue R.J, Sivanandam P, Malley C and Schmidt R.K. 2023. 9-arcsecond gridded HCAS 2.3 (2001-2018) base model estimation of habitat condition and general connectivity for terrestrial biodiversity, ecosystem site condition, annual epochs and 18-year trends for continental Australia. Data Collection. CSIRO, Canberra, Australia. https://doi.org/10.25919/arew-q819.

Lehmann E, Harwood T, Williams K, Ferrier S. 2018. HCAS Activity 3a(1): Environmental Space Modelling – Methods Document. Canberra.

Lehmann E, Williams K, Harwood T, Ferrier S, 2021. A not-too-technical introduction to the HCAS v2.x mechanics A revised method for mapping habitat condition across Australia. Slide pack Version 3.4. https://doi.org/10.25919/ek91-wj41

Levick S, Joehnk K, Botha E, Johnson S, van Niel T, Newnham G, Paget M, Lehmann E, Collings S, McVicar T, Williams K.J. 2023. HCAS-3: Remote sensing variable selection. Activity report for Progress Report 1 (a preliminary draft report prepared for DCCEEW). Canberra.

Lymburner L. 2021. Geoscience Australia Landsat Fractional Cover Percentiles Collection 3. Geoscience Australia, Canberra. Accessed May 14, 2024. https://pid.geoscience.gov.au/dataset/ga/145501

Malthus T.J, Ong C, Lau I, Fearns P, Byrne G, Thankappan M, Chisholm L, Suarez Barranco L, Clarke K, Scarth P. 2019. A community approach to the standardised validation of surface reflectance data. A technical handbook to support the collection

of field reflectance data. Brisbane. CSIRO Centre for Earth Observation. Release version 1.0. https://doi.org/10.25919/5c9d0ba9e9c12

Müller R. Calibration and Verification of Remote Sensing Instruments and Observations. Remote Sensing. 2014; 6(6):5692-5695. https://doi.org/10.3390/rs6065692

Schaaf C, Wang Z. 2015. MCD43A4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006. Dataset. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed May 14, 2024. https://doi.org/10.5067/MODIS/MCD43A4.006

Searle R. 2023. HCAS Optimised SLGA Products. Terrestrial Ecosystem Research Network, Canberra. Accessed May 14, 2024. https://aussoilsdsm.esoil.io/other/hcas-optimised-slga-products

Siddiqi A, Baber S, de Weck O, Durell C. 2020a. Error and Uncertainty in Earth Observation Value Chains, in: IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. Waikoloa, HI, USA. pp. 3158–3161. 10.1109/IGARSS39084.2020.9323463

Siddiqi A, Baber S, de Weck O.L, Durell C, Russell B, Holt J. 2020b. Integrating Globally Dispersed Calibration in Small Satellites Mission Value. 34th Annual Small Satellite Conference. Accessed November 5, 2024. https://digitalcommons.usu.edu/smallsat/2020/all2020/25/

Sutton A, Fisher A, Metternicht G. 2022. Assessing the Accuracy of Landsat Vegetation Fractional Cover for Monitoring Australian Drylands. Remote Sensing. 14(24):6322. https://doi.org/10.3390/rs14246322

Wang D, Morton D, Masek J, Wu A, Nagol J, Xion, X, Levy R, Vermote E, Wolfe R. 2012. Impact of sensor degradation on the MODIS NDVI time series. Remote Sens Environ 119, 55–61. https://doi.org/10.1016/j.rse.2011.12.001

Williams K, Donohue R, Harwood T, Lehmann E, Lyon P, Dickson F, Ware C, Richards A, Gallant J, Pinner L, Ozolins M, Austin J, White M, McVicar T, Ferrier S. 2020. Habitat Condition Assessment System: developing HCAS version 2.0 (beta): A revised method for mapping habitat condition across Australia. Canberra, Australia. https://doi.org/10.25919/85f4-1k65

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. 2021. Habitat Condition Assessment System (HCAS version 2.1): Enhanced method for mapping habitat condition and change across Australia. Canberra, Australia. https://doi.org/10.25919/n3c6-7w60

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, Canberra, Australia. https://doi.org/10.25919/r12v-ya3

Wolfe R.E, Nishihama M. 2009. Trends in MODIS geolocation error analysis, in: Earth Observing Systems XIV. SPIE, pp. 177–183. https://doi.org/10.1117/12.826598

Xiong X, Butler J.J. MODIS and VIIRS Calibration History and Future Outlook. Remote Sensing. 2020; 12(16):2523. https://doi.org/10.3390/rs12162523