CSIRO PUBLISHING

SPECIAL ISSUE | RESEARCH PAPER
https://doi.org/10.1071/RJ22041

THE RANGELAND JOURNAL Australian Rangeland Society

Can we benchmark annual ground cover maintenance?

Terrence S. Beutel[A,*] and F. Patrick Graz[A]

For full list of author affiliations and declarations see end of paper

***Correspondence to:**
Terrence S. Beutel
Department of Primary Industries and Fisheries, Rockhampton, Qld, Australia
Email: terry.beutel@daf.qld.gov.au

ABSTRACT

The capacity for rangeland stakeholders, including land managers, financiers and regulators, to regularly assess impacts of management practices on grazed landscapes has potential benefits. This paper describes the development of ground cover maintenance (GCM) spatial layers for a large study area in the catchment of the Great Barrier Reef in Queensland, Australia. GCM layers are an experimental product designed to benchmark the direction and strength of annual change in remotely sensed total ground cover (ΔTGC). This was achieved by predicting ΔTGC per pixel in a multivariate model, then using the quantile of the observed ΔTGC within its modelled prediction interval to benchmark observed ΔTGC. Under this approach, pixels with higher quantiles are those with a more positive annual observed ΔTGC after rainfall and other predictors in the multivariate model are taken into account. We then mapped these quantiles annually (2011–2021) across the study area and the annual spatial distribution of these quantiles is what we call the GCM layers. We identified two important issues to be addressed in future iterations of this work, namely, the potentially confounding impact of fire on GCM layers and their interpretation, and a need for more predictive skill in the underlying random forest model. Because management variables were not part of the underlying multivariate model but management practices can affect ΔTGC, we were interested in whether patterns in the mapped GCM values correlated with any known management practices or management-practice effects in the study area. We tested this idea on three datasets. In one, we compared GCM values from 12 well managed and 12 poorly managed grazing sites, finding no significant differences between the two groups. Another analysis looked at the relationship between grazing land condition and cumulative GCM values at two sets of sites (n = 110 and n = 189). Land condition and cumulative GCM values correlated significantly, although in only one of these data sets. Overall, we conclude that the developed GCM layers require further refinement to fit their desired purpose, but have potential to produce a number of benefits if current limitations can be addressed.

Keywords: grazing management, grazing pressure, landscape ecology, rangeland management, remote sensing.

Received: 10 October 2022
Accepted: 10 April 2023
Published: 27 April 2023

Cite this:
Beutel TS and Graz FP (2023)
The Rangeland Journal
44(5–6), 333–342. doi:10.1071/RJ22041

Introduction

Ground cover is an important feature of rangeland landscapes. It includes green (e.g. grasses, cryptogams) and non-green (e.g. leaf litter, logs, dung) components (DES 2014), provides ecoservices including primary production, carbon capture and soil protection, and varies in terms of both its composition and extent. In this work, we use the term 'total ground cover' (TGC) to define the horizontal coverage of these combined components of ground cover.

TGC varies naturally across rangelands due to factors including soil, topography and vegetation, but temporal changes are largely driven by two factors, namely, rainfall and management (often including fire). Temporal changes in species composition also play a role, but typically these changes follow some change in management and/or climate/rainfall (Bestelmeyer *et al.* 2003). In northern Australia, TGC typically correlates positively with rainfall across a range of time scales, and rainfall accounts for the majority of temporal variance in TGC at any location (Bastin *et al.* 2012). Management practices affect TGC, although often less dramatically and at a scale that is confounded or masked

by the larger fluctuations driven by rainfall. One key management impact is grazing pressure, with higher grazing pressure resulting in reduced TGC as pasture is removed and/or erosion areas expand. This relationship is not always straightforward; for example, changes in vegetation state (e.g. Stafford Smith *et al.* 2007) may delay or mask subsequent management impacts on TGC.

In Australia, a suite of remotely sensed TGC data and tools data are firmly embedded in the work of government agencies, natural resource management (NRM) groups and land managers to evaluate landscape health and pasture productivity, understand management impacts on the landscape and compare the value of public investments in rangeland health (e.g. Carroll *et al.* 2013; Zhang and Carter 2018; Beutel *et al.* 2019; Stone *et al.* 2019). It is worth noting that TGC data are used as a surrogate for such a wide range of outcomes in part because TGC correlates with these outcomes, but also because there are no more suitable surrogate data available at present. The foundational data behind these tools is the TGC time series (DES 2014), a seasonal (3 monthly, 1990–present), nationwide time series of Landsat-derived images of 30 m resolution that record percentage of TGC per pixel.

Benchmarking of any sort requires the availability of reference cases to which the subject is compared. A recurring theme in benchmarking satellite-derived cover indices, including TGC, is to derive the benchmark from parts of the surrounding area. A commonly used example of this approach is the regional comparison (RC; Zhang and Carter 2018; Beutel *et al.* 2019). RC benchmarks TGC at a site against TGC in parts of the surrounding region with similar soil, vegetation and topography. This process separates the effects of management and rainfall on TGC, and so highlights the timing and extent of management impacts. This approach is echoed in both the Compere (Donohue *et al.* 2022) and Dynamic Reference Cover Method (Bastin *et al.* 2012) and lends itself to outputs in either a graphed time-series format or mapped benchmark values, depending on the tools or study. Benchmarking against surrounding areas is a logical approach. However, it does require the choice of an appropriate buffer size; if the benchmark is too large, the reference and subject areas may differ too much climatically, and if too small, the reference area may be drawn from a small number of regionally atypical management units, which skews the benchmarking results. This type of benchmark also requires accurate stratification, so that the subject area is compared with a biologically similar reference area, and the availability of such data varies between jurisdictions in Australia (Beutel *et al.* 2019). An alternative benchmarking process would be modelling the reference area response via machine learning, and there is a range of machine learning tools suitable for such predictive tasks (Lesmeister 2017). This approach could draw on a large number of predictors (of possibly multiple scales) to develop a potentially more nuanced understanding of reference area response, would not require selection of a buffer area, or rely on a single or at least limited number of stratification layers. Statistical approaches to land degradation are common (e.g. Evans and Geerken 2004; Verbesselt *et al.* 2010; Burrell *et al.* 2017); however, these have largely focussed on assessment of a trend in an index value such as NDVI or its residuals once rainfall is taken into account. We are not aware of any published attempts to model either TGC or change in TGC as the basis for benchmarking observed values, but believe the method warrants consideration.

Benchmarking annual change in TGC has a number of potential benefits. First, the process of benchmarking (in this context evaluating something by comparison with a standard; Bruno 2014) the observed annual change in TGC to the range of potential change has potential for land managers and investors to robustly compare TGC maintenance across different sites. For example, is TGC better maintained on a site where it increases over 1 year from 60% to 70% or on a similar adjoining site where it increases from 95% to 98% over the same period? And is TGC at the former site equally well maintained in two separate years if that result occurs in a year of drought, then again some years later following above average rainfall? These kind of insights might allow rangeland stakeholders to objectively compare change (at least in terms of TGC) on sites with differing starting levels of ground cover, in different ecological contexts or at different times, and so better inform management and investment. A second advantage of annual benchmarking is potentially providing more immediate (annual) robust feedback to rangeland stakeholders about change in ground cover and its relationship to management. This is relevant to identifying early change and preventing entrenchment of longer-term problems. Annual benchmarks could also potentially be summed or stacked to give a longer-term picture of change. Benchmarking can, of course, occur over multiple time spans and other time spans have relevance to other possible goals, but those are not addressed in this work.

This work generated ground cover management (GCM) layers for three NRM regions in Queensland for the years 2011–2021. These layers benchmark annual ΔTGC and were built through a modelling approach. The modelling process predicted ΔTGC at a pixel scale by using a large number of contextual spatial layers, then used the quantiles of the model's prediction interval to benchmark observed ΔTGC per pixel. The values of a GCM layer are thus mapped quantiles and range between 0 and 100. Mathematically they indicate the proportion of the prediction interval for ΔTGC that falls below the observed ΔTGC in any pixel; however, in a more general sense they indicate how well ground cover has been maintained compared to modelled expectations. In this paper, we outline the development of these layers and discuss a number of observations and validation exercises designed to better describe the resulting layers and to provide initial insights into their capacity to benchmark TGC management, and to map land condition across the landscape.

Methods

Our methods are outlined below in two separate sections. The first details the development of the GCM layers, and the second outlines a number of validation exercises conducted to evaluate their potential utility as indices of management impact.

GCM development

GCM layers were generated for the combined Fitzroy Basin Association (Fitzroy), NQ Dry Tropics (Burdekin) and Burnett-Mary Regional Group (Burnett Mary) natural resource management regions (Fig. 1). These regions cover approximately 360 000 km^2, including approximately 280 000 km^2 of grazing land (QLUMP 2017). Rainfall in this area is summer dominant, with average annual rainfall across the region varying from above 1200 mm along sections of the eastern coast to 500 mm in the southern Burdekin and western Fitzroy regions.

The TGC time series used to build the GCM layers is a standard product consisting of seasonal composite images (1990–present; DES 2014). Annual ΔTGC in spring each year was calculated as:

$$\Delta\text{TGC} = \text{TGC}_t - \text{TGC}_{t-1},$$

where t = spring in the year of assessment and $t - 1$ = spring in the year prior to assessment. Spring data were used because ground cover varies most across the study area in spring (Beutel *et al.* 2019).

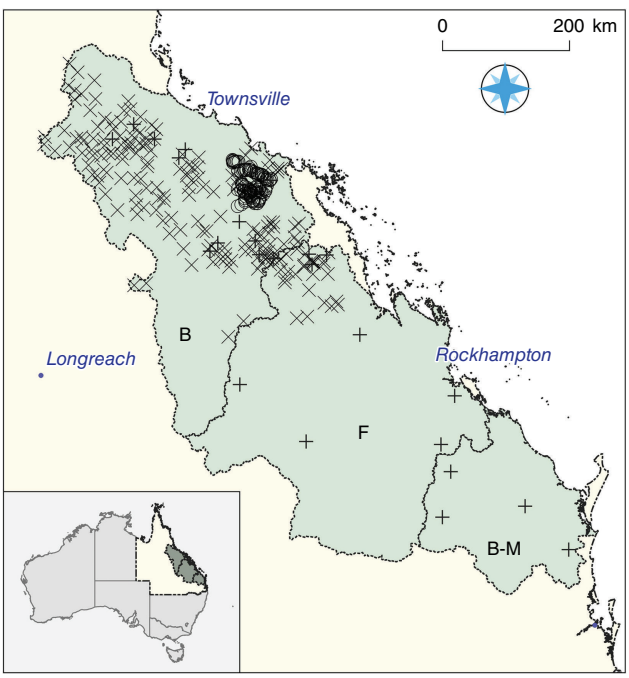

Fig. 1. Study area comprising the Burdekin (B), Fitzroy (F) and Burnett-Mary (B-M) natural resource management region. Validation site locations are shown for Datasets 1 (+), 2 (O) and 3 (X).

We modelled ΔTGC by using a starting pool of 35 predictors (Table 1) derived from mapped layers of the study area. These predictors included StartValue (TGC_{t-1}), a suite of rainfall and other climatic factors, and various landscape variables relating to vegetation cover and type, topography and soil. Predictors were sourced from a range of online and local datasets on the basis that they may relate to TGC across the region. Predictors were either static (unchanged through time, e.g. annual average rainfall, soil colour) or variant (changing through time, e.g. prior winter rainfall, woody plant cover).

We generated a random forest model by using the randomForest package (Liaw and Wiener 2002) in the R software (https://r-project.org) to predict ΔTGC per pixel from the data for spring 1991 to spring 2010 inclusive. We used this model to provide prediction intervals for the ΔTGC quantiles of subsequent years (2011–2021). The underlying assumption here is that 1991–2010 is climatically representative of 2011–2021. Our investigation indicated that total rainfall in the latter period was approximately 10% higher across the entire region, but we were limited in choice of modelling period since the model building period starts at the earliest possible date and had to conclude sufficiently early to allow creation of what we thought was a suitable number of GCM layers. The methodology for this process was as follows:

1. Generate 125 000 random sample points in the grazing land (QLUMP 2017) of the study area, then assign each point randomly to a year (1991–2010).
2. The ΔTGC and all predictor values at each point were extracted from their respective spatial layers, ensuring that variant layer values corresponded to their assigned year (Step 1).
3. These data were divided randomly into a training (a) and independent test (b) data sets at a ratio of 4:1.
4. We then ran a VSURF variable selection process (Genuer *et al.* 2015) on the training data (Step 3a) to identify a subset of predictors with a high predictive power. Variables with poor predictive power were discarded from the training data. Note that although VSURF generates random forest models, its purpose was variable selection and not to generate a final predictive random forest.
5. Build a final random forest model from the simplified training data (Step 4) to predict ΔTGC, tuning the *ntree* and *mtry* parameters (Probst *et al.* 2019) before measuring the fit of the model to the independent test data.
6. We then collated the predictor layers for the study area that match those from Step 4 for each year of 2011–2021.
7. Use the random forest model (Step 5) and the predictor layers (Step 6) to generate layers of predicted ΔTGC for each year of 2011–2021, then calculate the quantile of each observed ΔTGC by using the out-of-bag prediction interval method (Zhang *et al.* 2019) to produce the GCM layers.

Table 1. Variables and their sources used for modelling ΔTGC.

Variable	X/Y	Definition and source information
ΔTGC	Y	EndValue − StartValue
StartValue	X	Ground cover in spring prior to assessment year
Rain1	X	Total November rainfall in assessed year[A]
Rain2	X	Total October rainfall in assessed year[A]
Rain3	X	Total September rainfall in assessed year[A]
Rain4	X	Total winter rainfall (June–August) prior to the assessed season[A]
Rain5	X	Total autumn rainfall (March–May) prior to the assessed season[A]
Rain6	X	Total summer rainfall (December–February) prior to the assessed season[A]
Rain7	X	Total rainfall June–November in year prior to the assessed season[A]
Rain8	X	Total rainfall December–May in year prior to the assessed season[A]
Rain9	X	Total rainfall December–Nov 2 years prior to the assessed season[A]
Djaa	X	Woody vegetation cover index[B]
RainAnn	X	Mean annual rainfall (BoM product: IDCJCM004)
RainVar	X	Annual rainfall variability index (BoM product: IDCJCM0009)
RelHum09	X	Mean relative humidity at 9:00 am (BoM product: IDCJCM0014)
RelHum15	X	Mean relative humidity at 3:00 pm (BoM product: IDCJCM0014)
TempMax	X	Mean annual max temperature (BoM product: IDCJCM0005)
TempMean	X	Mean annual mean temperature (BoM product: IDCJCM0005)
TempMin	X	Mean annual minimum temperature (BoM product: IDCJCM0005)
BareBlue	X	Blue reflectance of bare earth component[C]
BareGreen	X	Green reflectance of bare earth component[C]
BareNIR	X	Near-infrared (NIR) reflectance of bare earth component[C]
BareRed	X	Red reflectance of bare earth component[C]
BareSWIR1	X	Short-wave IR reflectance of bare earth component[C]
BareSWIR2	X	Short-wave IR reflectance of bare earth component[C]
TopoInt	X	Topographic position Index – intermediate topography[D]
TopoLoc	X	Topographic position Index – local topography[D]
TopoReg	X	Topographic position Index – regional topography[D]
Weather	X	Surface weathering intensity index[E]
Slope	X	Percent slope[F]
Landform2F	X	$\bar{X}$ encoded regional ecosystem land zones[G]
Landform2sd	X	S_X encoded regional ecosystem land zones[G]
Glm60F	X	$\bar{X}$ encoded GLM land types (simplified)[H]
Glm60sd	X	S_X encoded GLM land types (simplified)[H]
SubIbraF	X	$\bar{X}$ encoded subIBRA[I]
SubIbrasd	X	S_X encoded subIBRA[I]

[A]Derived from BoM monthly rainfall rasters (Product code: IDCK200A00).
[B]Gill *et al.* (2017).
[C]Roberts *et al.* (2019).
[D]Wilford *et al.* (2020).
[E]Wilford (2011).
[F]Grundy *et al.* (2015).
[G]Neldner *et al.* (2019).
[H]https://www.data.qld.gov.au/dataset/grazing-land-management-land-types-series/resource/c951ebf8-c3b8-4247-b324-9d469c0e7d7e.
[I]https://www.environment.gov.au/land/nrs/science/ibra.

The resulting GCM layers were intended to benchmark ΔTGC across the study area for each year of 2011–2021. Layers were generated only for years outside the model period (1991–2010). The mapped GCM values are rounded quantiles ranging from 0 to 100, with higher values indicating that the actual ΔTGC was higher than the modelled ΔTGC. For example, a GCM value of 95 suggests a level of observed ΔTGC in the 95th percentile of the prediction interval and indicates a high level of TGC maintenance in that year.

Results include a number of visual and statistical summaries of the GCM time series. They highlight some of the more obvious features in these data and are included in part because of the novelty of the GCM layers and to assist in their potential use and interpretation.

GCM validation

Once the GCM layers were completed, they were assessed against three datasets to evaluate their potential utility. In the first analysis, we assessed GCM values across 24 grazed-land parcels (Dataset 1) from 2017 to 2021 (Fig. 1). These parcels were identified by local extension and research professionals in the Queensland Government, each with at least 10 years of knowledge of grazing management practices in their targeted area and knowledge of longer-term management practices on the identified parcels. Experts classified parcels as having either consistently good (n = 12) or consistently poor (n = 12) pasture management over at least the previous 5 years based on their understanding of contemporaneous grazing pressure on each parcel. We compared the mean GCM values between the consistently good and consistently poor parcels over this period (2017–1021), on the assumption that better pasture managers generally maintain TGC better and should thus demonstrate generally higher GCM values over time.

A second analysis examined the correlation between cumulative GCM and grazing land condition class (Chilcott *et al.* 2003), which classes grazing land condition on an ordinal four-point scale from Good (A) to Very Poor (D). Grazing land condition changes more slowly than ground cover and is often driven by cumulative management practices such as ongoing heavy grazing or regular wet season spelling (Chilcott *et al.* 2003; McIvor 2012). If annual GCM values reflect annual management skill, then cumulative GCM values might predict grazing land condition. We tested this assumption on two separate datasets (Datasets 2 and 3). Dataset 2 documents land condition ratings collected in the field at 110 × 1 ha sites in and around the Bowen catchment in 2018, and Dataset 3 details land condition ratings on 189 land parcels in the Burdekin and Fitzroy regions (Fig. 1). The land condition of Dataset 3 parcels was determined by consensus at a workshop of regional research and extension staff regarding regional land condition in 2018 (Beutel *et al.* 2021). For these analyses, we calculated the temporal mean GCM value per pixel (2014–2018) across the region (we ignored 2019–2021 data because they were collected after the 2018 land condition assessments and 2011–2013 data due to significant missing data (discussed below)). We then averaged pixel values from this 5-year mean image within each site polygon to extract a 5-year mean GCM for each site. Mean GCM was used in preference to a 5-year cumulative sum because only the former accounts for missing values in any pixel time series. We then tested the correlation between land condition ratings and the corresponding 5-year mean GCM in each dataset.

Results

The results of the GCM layer development are discussed below in two main sections focussing on GCM development and validation. GCM development describes the GCM random forest model and its resulting 2011–2021 layers. GCM validation outlines the statistical testing of these layers at sites identified in Datasets 1, 2 and 3.

GCM development

The random forest model for ΔTGC fitted independent test data relatively well (Fig. 2, R^2 = 0.52). This model included 10 predictors selected through the VSURF process, and these are shown in the variable importance plot for the model (Fig. 3). The most influential predictor in the model was StartValue (TGC at the start of the assessment year). This makes intuitive sense; TGC at the start of the year is likely to generally reflect TGC at the end of the year. In addition, sites with higher starting TGC have less scope to increase over the year (e.g. a site starting on 100% cannot increase at all), sites with low starting TGC have less scope to decline in cover, overall making ΔTGC easier to predict. Rainfall

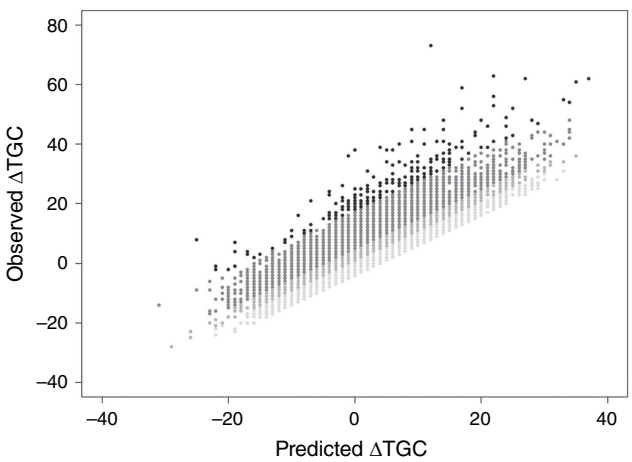

Fig. 2. Observed vs predicted ΔTGC values for 25 000 independent test cases (R^2 = 0.52). The four greyscale gradations indicate the quartiles of the observed ΔTGC values and also the relationship among observed ΔTGC, predicted ΔTGC and GCM value. The palest points have GCM values 1–25 through to darkest (76–100).

variables were also important predictors, particularly rainfall in the summer (Rain6), winter (Rain4) and autumn (Rain5) immediately before the assessment season. Soil colour (BareRed) and woody vegetation coverage (Djaa) contributed least to predictive power of the predictors selected in the VSURF process.

Fig. 4 shows the relationship between observed and predicted ΔTGC and their corresponding GCM quantiles. Quantiles are higher where the observed ΔTGC exceeds the predicted ΔTGC (circle) and lower where predicted ΔTGC exceeds the observed ΔTGC (rectangle) in the given year.

The impact of fire is often clearly visible in the GCM layers. Fire obviously has a dramatic impact on TGC, and visual inspections of imagery at known and suspected fire sites generally indicated very low GCM values in the year of the fire. Less commonly, the same site had very high GCM values in the year following the fire because cover increased from the very low base (Fig. 5). However, this post-fire 'bounce' was not as common as was the initial drop in GCM values immediately after a fire.

A second feature noted in the GCM imagery was a large number of missing pixel values in layers for 2011–2013 (Fig. 6). Missing data can result from factors including cloud, water and woody cover masks, but the extra missing data in these years have three main sources. Landsat 5 was decommissioned in mid-2013, reducing available data, and this along with a known Landsat 7 fault (Andrefouet *et al.* 2003) flowed through to TGC images and the GCM layers. These problems were relieved by the launch of Landsat 8 in February 2013, which produced more complete imagery post-2013. Missing data in the GCM layers were additionally caused by needing both TGC_t and TGC_{t-1} to calculate ΔTGC and, ultimately, the GCM quantile (Fig. 7). This means that any missing TGC_t value produces a corresponding missing GCM value in two consecutive years (t and $t + 1$). As a result, missing values in the GCM layers were still relatively high in spring 2013 when Landsat 8 had been operational for approximately 6 months.

A final feature noted in the image time series was the tendency for significant areas to fluctuate annually across a wide range of GCM values, such that it was possible for a large area, for example, to have generally low GCM values in 1 year and generally high values in the next and *vice versa*. Within these areas, property- and subproperty-scale variation was also visible, but the broader fluctuation suggests either that TGC maintenance is temporally correlated among large numbers of properties, or that the model did not adequately capture variation in ΔTGC among years (Fig. 8).

GCM validation

Fig. 9 shows the relationship between mean GCM value and pasture management on 12 parcels with 'good' and 12 parcels with 'poor' management, as assessed by regional experts (Dataset 1). We analysed the repeated measures by using residual maximum likelihood (REML) and modelled the correlation structure induced by the repeated measures. Various models were investigated, and the best (based on Schwarz information criterion) was an autoregressive process of order 2. This showed no interaction between management type and year ($F_{4,62} = 0.17$, $P > 0.05$) or effect of management type ($F_{1,27} = 2.20$, $P > 0.05$), but a significant effect of year ($F_{4,65} = 5.27$, $P < 0.001$). Fisher's protected least significant difference (l.s.d.) tests ($P = 0.05$) showed that mean GCM for 2017, 2018 and 2021 were significantly higher than that for 2019, with that for 2020 not being different from any year.

Fig. 10 shows the relationship between grazing land condition ratings and 5-year mean GCM values for Datasets 2 and 3. In both cases, sites with higher land condition ratings tended to have higher GCM values, although the strength of this correlation (using Kendall's Tau_B) was significant only for Dataset 3 ($T_B = -0.26$, $P < 0.001$).

Discussion

Our work developed a set of spatial raster layers designed to benchmark ΔTGC across the landscape and provide insight into TGC management. Each layer benchmarks a separate year of remotely sensed ΔTGC (spring to spring) at a pixel scale. The underlying random forest model predicted ΔTGC on independent data reasonably well and relied heavily on TGC_{t-1} and rainfall to explain annual ΔTGC.

The layers showed much higher proportions of missing data prior to 2014. As noted, this was largely driven by a Landsat 7 error, and exacerbated by the need for StartValue and EndValue in any year to calculate the GCM quantile. However, this problem is largely historical now and coverage since 2014 has been very good (> 90%), with most of the gaps due to masking of TGC by high levels of woody cover.

The impact of fire on GCM quantiles has important implications for how the layers might be used. Fire generally causes a sharp GCM decline in the year of the burn. Although this pattern makes sense, different burning and post-fire management strategies can produce very different outcomes (e.g. McIvor 2012; Hunt *et al.* 2014), and this makes it difficult to equate this cause of low GCM with poor land management. Fire clearly affects GCM values and validation of GCM layers should take this into account because mapped GCM values in a burn year may possibly mask other management impacts intended to protect the landscape in the longer term. Future development and testing may require masking burn sites from analyses or at least noting when fire affected an analysis.

The tendency noted for broader areas to fluctuate significantly over time suggests either that temporal variation in TGC management is correlated among large numbers of grazing properties or that a larger scale driver of GCM was not captured by the random forest model. The former seems unlikely, but at least two scenarios are plausible for the

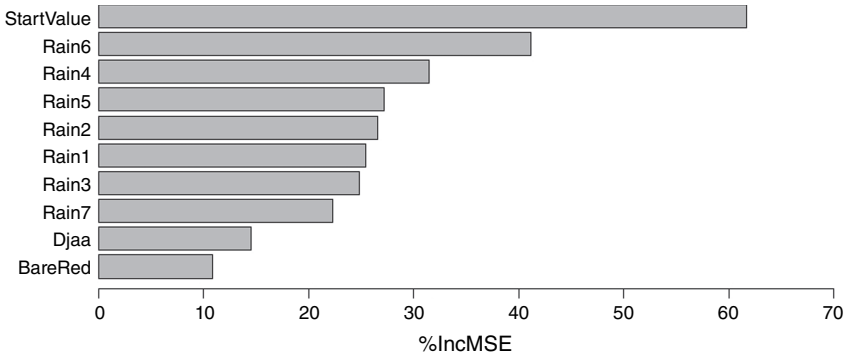

Fig. 3. Variable importance plot for the random forest model predicting ΔTGC. Variables are defined in Table 1.

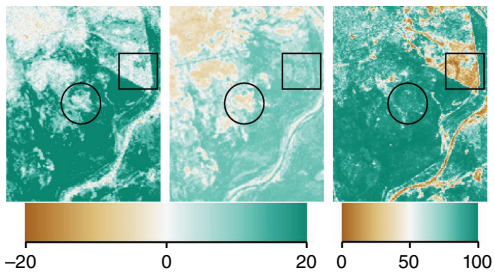

Fig. 4. Actual ΔTGC (left), predicted ΔTGC (middle) and corresponding GCM values (right) for a subset of the study area in 2014. In the circled area, the actual ΔTGC exceeds the predicted ΔTGC and, consequently, the GCM quantile for the circled area is high. Conversely, the rectangle highlights a location where the actual ΔTGC is below the predicted ΔTGC and in this area GCM values are relatively low.

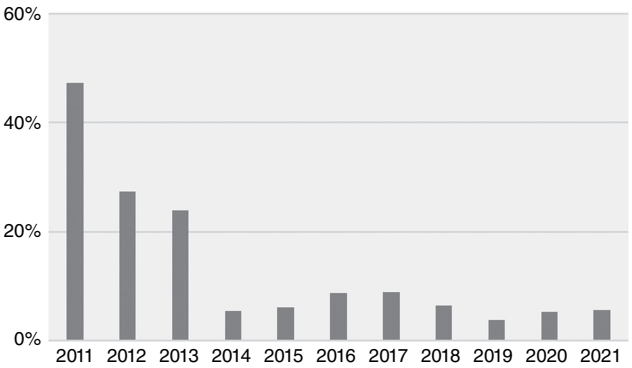

Fig. 6. Percentage of pixels with missing data in GCM images for 2011–2021. Values are lower post-2013 due to availability of Landsat 8 imagery and subsequently, lower reliance on Landsat 7 data from early 2013.

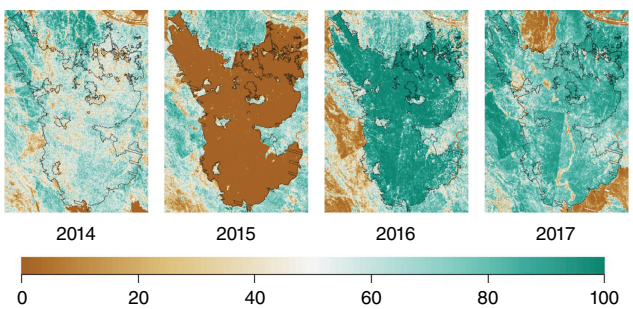

Fig. 5. GCM layers for a fire site (2014–2017). The burned area (black outline) was identified from fire-scar mapping (Goodwin and Collett 2014) and occurred in autumn 2015. Note the typically low quantile values in the first year post-burn (2015) and relatively high quantile values in the subsequent year (2016) as cover increased from a low base. Observable impacts in the 2017 image are more limited.

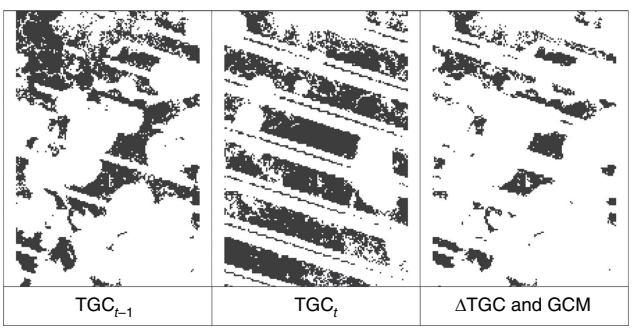

Fig. 7. Coverage of corresponding TGC$_{t-1}$, TGC$_t$, ΔTGC and GCM images. White space represents missing pixel data. ΔTGC and GCM values can be calculated only where both TGC$_{t-1}$ and TGC$_t$ are available and so inherit gaps from both the parent layers.

latter. One possibility is that rainfall impacts on TGC may not have been adequately captured in the model; so, while rainfall did provide a significant part of the model's predictive power, it may not have been sufficient to fully describe annual fluctuations. We note that in preliminary testing we trialled a variety of rainfall indices, including the standardised precipitation index, and selected the configuration of rainfall variables in Table 1 on the basis of their better predictive power. Another potential explanation derives from the fact that the modelling data were collected over 20 years (1991–2010), and a wider range of environmental conditions than would be expected in any single year. This could result in predictions for a given year occupying only a subset of the prediction space and thus producing only a subset (sometimes high or low) of potential GCM values. In either case, these broader fluctuations affect interpretation of the GCM layers and are likely to have contributed to

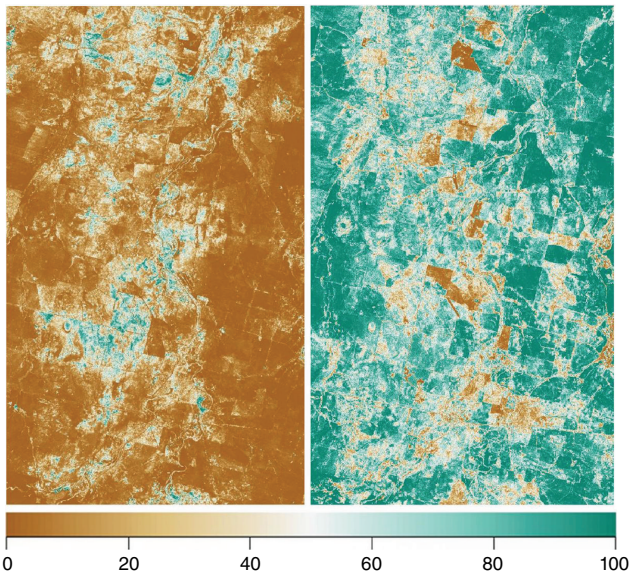

Fig. 8. GCM image subsets for corresponding area of Fitzroy Basin for 2020 (left) and 2021 (right). Property- and paddock-scale contrasts are visible, but there is also a distinct overall difference between the two images, with the 2020 image having generally much lower GCM values.

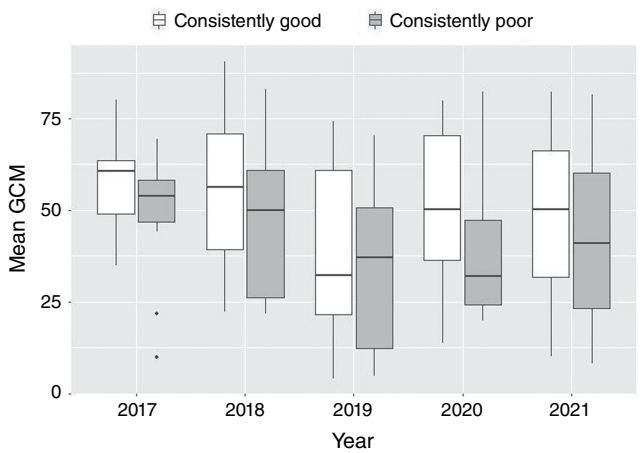

Fig. 9. Boxplot of annual average GCM values on parcels with consistently good and consistently poor pasture management (2017–2021) in the Dataset 1 parcels.

limited performance in the validation testing. This issue would need further investigation in future iterations of the GCM layers, and might, for example, include separately modelling for wet and dry parts of the climate cycle or trialling alternative rainfall indices.

The analysis of Dataset 1 examined a set of sites where local experts viewed pasture management as either consistently good or consistently poor over the period 2017–2021, and we expected to see consistently higher GCM values on good parcels over that period. While good sites had higher average GCM values, no significant difference was detected

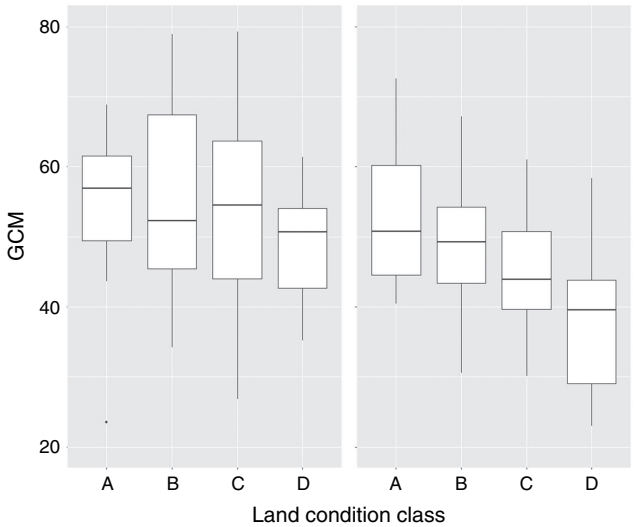

Fig. 10. Boxplots of 5-year mean GCM values on 110 × 1 ha land condition assessment sites (Dataset 2) in the Bowen catchment (left) and 189 larger grazing property parcels (Dataset 3) across the Burdekin and Fitzroy NRM areas (right).

between the two groups of parcels; so GCM did not discriminate the two management groups. This suggests that in their current form, the GCM layers are not a reliable index of TGC maintenance skill since the comparison was made between two groups of parcels that should have separated reasonably well, given their relatively large difference in management outcomes.

The land condition analyses (Datasets 2 and 3) showed that cumulative GCM values discriminated land condition classes well in one of the two case-study data sets, although, in both data sets, better-condition sites tended to have higher GCM values than did poor-condition sites. GCMs are intended to reflect TGC management, and although TGC is a significant component of grazing land condition, condition is multidimensional and is not explained by TGC alone (e.g. sites with heavy cover of weedy grasses have good cover but poor grazing land condition). Given this, it is encouraging that cumulative GCMs did correlate with land condition as well as they did, and the potential use of these values in mapping or modelling land condition spatially may be worth further investigation.

Conclusion

The capacity to assess management practices annually has a number of potential uses for rangeland stakeholders, including managers, financiers and regulators. This kind of product has potential uses in understanding and informing management decisions on individual sites, assessing the success of extension interventions and possibly even in mapping and modelling grazing land condition across the landscape. Consequently, new and/or improved approaches to

the goals of this work are warranted. Overall, the validation work in this paper suggests that the GCM layers we developed require further refinement to fit their desired purpose. The underlying model of ΔTGC had reasonably good predictive skill, but also substantial room for improvement. We also identified a number of factors that complicate the use and interpretation of the layers. Of particular note is the impact of fires on sequential GCM values and the tendency for broader-scale fluctuations in GCM values among years. If these issues can be addressed and the predictive power of the underlying model improved, then the GCM approach we proposed here could provide more reliable results and potentially fit its intended purpose more reliably.

References

Andrefouet S, Bindschadler R, Brown de Colstoun E, Choate M, Chomentowski W, Christopherson J, Doorn B, Hall DK, Holifield C, Howard S, Kranenburg C (2003) 'Preliminary assessment of the value of Landsat-7 ETM+ data following scan line corrector malfunction.' (US Geological Survey, EROS Data Center: Sioux Falls, SD, USA) doi:10.1016/j.jenvman.2017.12.070

Bastin G, Scarth P, Chewings V, Sparrow A, Denham R, Schmidt M, O'Reagain P, Shepherd R, Abbott B (2012) Separating grazing and rainfall effects at regional scale using remote sensing imagery: a dynamic reference-cover method. *Remote Sensing of Environment* **121**, 443–457. doi:10.1016/j.rse.2012.02.021

Bestelmeyer BT, Brown JR, Havstad KM, Alexander R, Chavez G, Herrick JE (2003) Development and use of state-and-transition models for rangelands. *Journal of Range Management* **56**, 114–126. doi:10.2307/4003894

Beutel TS, Trevithick R, Scarth P, Tindall D (2019) VegMachine.net. online land cover analysis for the Australian rangelands. *The Rangeland Journal* **41**, 355–362. doi:10.1071/RJ19013

Beutel TS, Shepherd R, Karfs RA, Abbott BN, Eyre T, Hall TJ, Barbi E (2021) Is ground cover a useful indicator of grazing land condition. *The Rangeland Journal* **43**, 55–64. doi:10.1071/RJ21018

Bruno I (2014) Benchmarking. In 'Encyclopedia of quality of life and well-being research'. (Ed. AC Michalos) (Springer: Dordrecht, Netherlands) doi:10.1007/978-94-007-0753-5_170

Burrell AL, Evans JP, Liu Y (2017) Detecting dryland degradation using time series segmentation and residual trend analysis (TSS-RESTREND). *Remote Sensing of Environment* **197**, 43–57. doi:10.1016/j.rse.2017.05.018

Carroll C, Waters D, Ellis R, McCosker K, Gongora M, Chinn C, Gale K (2013) Great Barrier reef paddock to reef monitoring and modelling program. In 'MODSIM2013, 20th international congress on modelling and simulation', Modelling and Simulation Society of Australia and New Zealand, Adelaide, SA, Australia.

Chilcott CR, Paton CJ, Quirk MF, McCallum BS (2003) 'Grazing land management education package workshop notes – Burnett.' (Meat and Livestock Australia Limited: Sydney, NSW, Australia)

DES (2014) Seasonal ground cover - Landsat, JRSRP algorithm, Australia Coverage. (Version 1.0.0.) Department of Environment and Science, Queensland Government, Terrestrial Ecosystem Research Network (TERN) dataset. Available at https://portal.tern.org.au/seasonal-ground-cover-australia-coverage/22022

Donohue RJ, Mokany K, McVicar TR, O'Grady AP (2022) Identifying management-driven dynamics in vegetation cover: applying the Compere framework to Cooper Creek, Australia. *Ecosphere* **13**(3), e4006. doi:10.1002/ecs2.4006

Evans J, Geerken R (2004) Discrimination between climate and human-induced dryland degradation. *Journal of Arid Environments* **57**, 535–554. doi:10.1016/S0140-1963(03)00121-6

Genuer R, Poggi J-M, Tuleau-Malot C (2015) VSURF: An R package for variable selection using random forests. *The R Journal* **7**, 19–33. doi:10.32614/RJ-2015-018

Gill T, Johansen K, Phinn S, Trevithick R, Scarth P, Armston J (2017) A method for mapping Australian woody vegetation cover by linking continental-scale field data and long-term Landsat time series. *International Journal of Remote Sensing* **38**, 679–705. doi:10.1080/01431161.2016.1266112

Goodwin NR, Collett LJ (2014) Development of an automated method for mapping fire history captured in Landsat TM and ETM+ time series across Queensland, Australia. *Remote Sensing of Environment* **148**, 206–221. doi:10.1016/j.rse.2014.03.021

Grundy MJ, Rossel RAV, Searle RD, Wilson PL, Chen C, Gregory LJ (2015) Soil and landscape grid of Australia. *Soil Research* **53**, 835–844. doi:10.1071/SR15191

Hunt LP, McIvor JG, Grice AC, Bray SG (2014) Principles and guidelines for managing cattle grazing in the grazing lands of northern Australia: stocking rates, pasture resting, prescribed fire, paddock size and water points – a review. *The Rangeland Journal* **36**, 105–119. doi:10.1071/RJ13070

Lesmeister C (2017) 'Mastering machine learning with R.' (Packt Publishing: Birmingham, UK)

Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* **2**, 18–22.

McIvor J (2012) 'Sustainable management of the Burdekin grazing lands – a technical guide of options for stocking rate management, pasture spelling, infrastructure development and prescribed burning to optimise animal production, profitability, land condition and water quality outcomes.' (State of Queensland, Australia)

Neldner VJ, Butler DW, Guymer GP (2019) 'Queensland's regional ecosystems: building a maintaining a biodiversity inventory, planning framework and information system for Queensland, Version 2.0.' (Queensland Herbarium, Queensland Department of Environment and Science: Brisbane, Qld, Australia)

Probst P, Wright MN, Boulesteix A-L (2019) Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery* **9**, e1301. doi:10.1002/widm.1301

QLUMP (2017) 'Queensland land use mapping program.' (Remote Sensing Centre, Department of Environment and Science, Queensland Government: Qld, Australia)

Roberts D, Wilford J, Ghattas O (2019) Exposed soil and mineral map of the Australian continent revealing the land at its barest. *Nature Communications* **10**, 5297. doi:10.1038/s41467-019-13276-1

Stafford Smith DM, McKeon GM, Watson IW, Henry BK, Stone GS, Hall WB, Howden SM (2007) Learning from episodes of degradation and recovery in variable Australian rangelands. *Proceedings of the National Academy of Sciences* **104**, 20690–20695. doi:10.1073/pnas.0704837104

Stone G, Pozza RD, Carter J, McKeon G (2019) Long Paddock: climate risk and grazing information for Australian rangelands and grazing communities. *The Rangeland Journal* **41**, 225–232. doi:10.1071/RJ18036

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. doi:10.1016/j.rse.2009.08.014

Wilford J (2011) Weathering intensity map of the Australian continent. *AusGeo News* **101**, 1–5.

Wilford JR, Lindsay J, Basak S (2020) Multiscale topographic position image of the Australia continent. In 'Exploring for the future: Extended Abstracts'. (Eds K Czarnota, I Roach, S Abbott, M Haynes, N Kositcin, A Ray, E Slatter) pp. 1–4. (Geoscience Australia)

Zhang B, Carter J (2018) FORAGE – An online system for generating and delivering property-scale decision support information for grazing land and environmental management. *Computers and Electronics in Agriculture* **150**, 302–311. doi:10.1016/j.compag.2018.05.010

Zhang H, Zimmerman J, Nettleton D, Nordman DJ (2019) Random forest prediction intervals. *The American Statistician* **74**, 392–406. doi:10.1080/00031305.2019.1585288

Data availability. The data used in this study are available upon reasonable request to the corresponding author.

Conflicts of interest. The authors of this paper declare no conflicts of interest.

Declaration of funding. This work was funded by the Queensland Reef Water Quality Science Program through RP234.

Acknowledgements. Jim Fletcher, Karl McKellar and Rhonda Melzer, Roger Sneath, Mick Sullivan and Giselle Whish evaluated the Dataset 1 land parcels. David Reid performed the statistical analysis of the Dataset 1. Bec Trevithick and Emily Barbi reviewed drafts of this article. Thanks go to all.

Author affiliation
[A]Department of Primary Industries and Fisheries, Rockhampton, Qld, Australia.