Metrics to assess the condition of rangeland ecosystems in central Mongolia

S.J. Sinclair, C. Liu, K. Batpurev, O. Avirmed, B. Avirmed, M.D. White, B. Ricard, A. Erdengerel, E. Narmandakh, D. Odsuren, M. Kohout and K. Olson

August 2023

Arthur Rylah Institute for Environmental Research
Technical Report Series No. 365

Arthur Rylah Institute for environmental research

Wildlife Conservation Society

AVSF Agronomes & Vétérinaires Sans Frontières

Victoria State Government | Energy, Environment and Climate Action

Acknowledgment

We acknowledge and respect Victorian Traditional Owners as the original custodians of Victoria's land and waters, their unique ability to care for Country and deep spiritual connection to it. We honour Elders past and present whose knowledge and wisdom has ensured the continuation of culture and traditional practices.

We are committed to genuinely partner, and meaningfully engage, with Victoria's Traditional Owners and Aboriginal communities to support the protection of Country, the maintenance of spiritual and cultural practices and their broader aspirations in the 21st century and beyond.

Arthur Rylah Institute for Environmental Research
Department of Energy, Environment and Climate Action
PO Box 137
Heidelberg, Victoria 3084
Phone (03) 9450 8600
Website: www.ari.vic.gov.au

Citation: Sinclair, S.J., Liu, C., Batpurev, K., Avirmed, O., Avirmed, B., White, M.D., Ricard, B., Erdengerel, A., Narmandakh, E., Odsuren, D., Kohout, M. and Olson, K. (2023). Metrics to assess the condition of rangeland ecosystems in central Mongolia. Arthur Rylah Institute for Environmental Research Technical Report Series No. 365. Department of Energy, Environment and Climate Action, Heidelberg, Victoria.

Front cover photo: Heavily grazed High Mountain Steppe, Bayankhongor aimag (Steve Sinclair).

© The State of Victoria Department of Energy, Environment and Climate Action 2023

This work is licensed under a Creative Commons Attribution 3.0 Australia licence. You are free to re-use the work under that licence, on the condition that you credit the State of Victoria as author. The licence does not apply to any images, photographs or branding, including the Victorian Coat of Arms, the Victorian Government logo, the Department of Energy, Environment and Climate Action logo and the Arthur Rylah Institute logo. To view a copy of this licence, visit http://creativecommons.org/licenses/by/3.0/au/deed.en

Edited by David Meagher, Zymurgy Consulting, Meeniyan

ISSN 1835-3827 (print)
ISSN 1835-3835 (pdf))
ISBN 978-1-76136-382-5 (print)
ISBN 978-1-76136-383-2 (pdf/online)

Disclaimer
This publication may be of assistance to you but the State of Victoria and its employees do not guarantee that the publication is without flaw of any kind or is wholly appropriate for your particular purposes and therefore disclaims all liability for any error, loss or other consequence which may arise from you relying on any information in this publication.

Accessibility

If you would like to receive this publication in an alternative format, please telephone the DEECA Customer Service Centre on 136 186, email customer.service@delwp.vic.gov.au or contact us via the National Relay Service on 133 677 or www.relayservice.com.au. This document is also available on the internet at www.deeca.vic.gov.au

Metrics to assess the condition of rangeland ecosystems in central Mongolia

Steve J. Sinclair[1], Canran Liu[1], Khorloo Batpurev[1], Otgonsuren Avirmed[2], Batbold Avirmed[3], Matthew D. White[1], Bertrand Ricard[3], Ariunbold Erdengerel[4], Enkhriimaa Narmandakh[5], Davaasuren Odsuren[6], Michele Kohout[1] and Kirk Olson[2,7].

[1]Arthur Rylah Institute for Environmental Research
123 Brown Street, Heidelberg, Victoria 3084, Australia.

[2]Wildlife Conservation Society (WCS), Mongolia Country Program
POB 485, PO 38, Ulaanbaatar-15141, Mongolia.

[3]Agronomes et Vétérinaires Sans Frontières (AVSF)
Ulaanbaatar-14251, Mongolia.

[4] Mongolian Academy of Sciences, Institute of Geography and Geoecology
Ulaanbaatar-15170, Mongolia.

[5] Mongolian Academy of Sciences, Botanic Garden, Institute of history and Ethnology
Ulaanbaatar-133343, Mongolia.

[6]Bayalag-Eco, LLC, Mongolia.

[7]Smithsonian's National Zoo and Conservation Biology Institute, 3001 Connecticut Avenue, Washington DC 20008, USA.

Arthur Rylah Institute for Environmental Research
Technical Report Series No. 365

Arthur Rylah Institute for Environmental Research
Department of Energy, Environment and Climate Action
Heidelberg, Victoria

Acknowledgements

We acknowledge the contributions of all the stakeholders from Mongolia who gave their time and knowledge to this project. Many people travelled long distances away from their families to contribute, and we are grateful for their efforts. We thank Sergelenkhuu Jambal for her advice on the delineation of ecosystems. We thank Guillaume Touati, Manon Lelarge and Quentin Moreau (AVSF) for their support of this project. We also thank Fern Hames, Tim O'Brien and the admin team at ARI for their help in allowing ARI staff to travel overseas in safety.

Unofficial

Contents

Tables

Figures

Хураангуй (Mongolian Summary)

Нүүдэлч малчдын орлогын гол үүсвэр ямааны ноолуур байдаг ба нийгмийн нөхцөл өөрчлөгдөж, ноолуурын үнэ өссөнөөс малчид ямаагаа үржүүлэн өсгөж олшруулсан. Бэлчээрт үзүүлэх малын нөлөө дарамт ч үүнтэй хамт ихэссэн байдаг. Энэ сөрөг нөлөөний улмаас бэлчээрт ашиглагдаж байгаа экосистемийн биологийн олон янз байдаг алдагдах, хөрс элэгдэж эвдрэх аюул нөмөрч, мал сүрэг сульдаж цаг агаарын хүнд нөхцөлийн давах чадвар суларч байна.

Малчид орлогоо бууруулалгүйгээр малаа цөөлөхийг хөхүүлэн дэмжих зорилгоор “Хил Хязгааргүй Малын Эмч нар” байгууллага ноолуурын сертификат олгох тогтолцоо бий болгосон. Уг систем Баянхонгор аймгийн хэмжээнд хэрэгжиж байна. Бэлчээрээ зохистой ашиглаж буй малчдад уг сертификатыг олгодог бөгөөд сертификат авсан малчид зах зээлд өндөр үнээр ноолуураа борлуулах боломжтой болдог юм. Өөрөөр хэлбэл малын тоогоо бууруулсаны улмаас алдсан орлогоо ноолуураа өндөр үнээр зарсанаар нөхөх зарчимтай тогтолцоо юм. Ноолуурын энэ сертификатыг хүлээн зөвшөөрөгдөхүйц үр дүнтэй байлгахын тулд бэлчээрээ хамгаалж, сайжруулж буй малчид л уг сертификатыг авдаг байх ёстой. Бэлчээрийн экосистемийн нөхцөл байдлыг хянах арга зүй нь бэлчээрт орсон өөрчлөлтийг тодорхойлох боломжийг бий болгох ба энэ сертификатжуулах тогтолцоо бэлчээрийн экосистемийн үнэлгээний арга зүйгээс хамааралтай юм.

Баянхонгор аймагт тархсан бэлчээрийн таван өөр экосистемийн (Бореалын ой, Өндөр уулын хээр, Чийгсүү хээр, Хуурай хээр болон Хэт гандуу цөл) үнэлгээний арга зүйг бид энэ удаагийн тайланд багтаалаа. Уг арга зүй нь хэрэглэхэд хялбар, энгийн хээрийн хэмжилтийн мэдээг ашиглан өөрчлөлт ороогүй экосистем болон онгон байдлаас доройтож муудсан бүх нөхцөлийг ялган харуулах боломжтой тоон үнэлгээ байх ёстой.

Өмнө нь Австралийн экосистем болон Монголын говийн экосистемд зориулан боловсруулан, олон улсын сэтгүүлд хэвлүүлсэн арга зүйг бид ашиглан уг үнэлгээг боловсруулллаа. Экологийн нөхцөл байдал гэдэг ойлголт нь өөрөө субьектив шинжтэй учраас бид уг үнэлгээг хүмүүсийн мэдлэгт суурилсан бодол дээр үндэслэж боловсруулсан.

Энэ үнэлгээг боловсруулхад 151 хүний санаа бодлыг ашигласан ба эдгээрийн 81% нь малчид байсан. Тэд байгаль дээр байж болох талбайг ургамлын үзүүлэлтээр илэрхийлсэн картуудыг харж байгаад өөрийн тоон үнэлгээг өгсөн. Оролцогч бүр таван экосистемийн бүлэг картуудыг үнэлсэн ба ингэхдээ картан дээрхи талбай бүрийг 0-с (хамгийн муу нөхцөл) 100 (хамгийн сайн нөхцөл) онооны хооронд үнэлсэн. Эдгээр үнэлгээний мэдээгээр статистик загварыг сурган улмаар ургамлын үзүүлэлтийн мэдээлэл байхад л тухайн экосистемийн оноог гаргаж өгөх загвар бий болгосон. Мөн хээрийн талбайн ургамлын үзүүлэлтийг цуглуулах ажлын арга зүйг тайлбарлан оруулсан.

Үүний дараа майкрософт эксел программд уг боловсруулсан загварын томьёог бичиж оруулан *“эксэл хэрэглүүр”* үүсгэсэн. Талбайд бичиглэл хийсэн мэдээний үр дүнгээ Майкрософт эксел программын хүснэгтэд шивээд хэрэглүүрийг ашиглан экосистемийн оноог бодох боломжтой болж байгаа юм. Уг хэрэглүүрийн гаргасан оноо оролцогсодын өгөх байсан үнэлгээний дундажийг илэрхийлж байгаа бөгөөд хээрийн мэдээнээс өөр нэмэлт мэдээлэл цуглуулах шаардлагагүй.

Бид загвар сургахад ашиглагдаагүй мэдээгээр загвараа шалгахад таван экосистемийн загвар тус бүр илэрхий эерэг хамааралтай байсан (R^2 0.65 – 0.77). Таван экосистемийн гурав дээр нь мөн илүү нарийн шалгалт хийж үзсэн. Ингэхдээ хээр талбай дээр хүмүүсийн өгсөн оноог хэрэглүүрээр бодсон оноотой шууд харьцуулж үзэхэд өндөр уулын хээрийн загвар маш сайн хамааралтай байсан бол чийгсүү хээрийн загвар дунд зэрэг, хуурай хээрийн загвар сул эерэг хамааралтай байсан.

Хээрийн талбайн бичиглэл хийж, уг мэдээгээрээ экосистемийн нөхцөл байдлын оноо бодох энэ арга нь ноолуурын сертификат олгоход шаардлагатай бэлчээрийн нөхцөл байдлын хяналт болон тайлагнах ажилд тохирсон арга зүй болсон гэж үзэж байна.

Summary

Context:

The production of cashmere fibre (goat hair) is a major source of income for nomadic herders in Mongolia. Because of social changes and high cashmere prices, many herders have increased their goat herds substantially in recent decades. The resultant increase in grazing pressure has resulted in biodiversity loss, soil erosion, and increased the vulnerability of the goat herds to harsh environmental conditions.

To encourage herders to reduce herd sizes while maintaining a reasonable income, Agronomes et Vétérinaires Sans Frontières (AVSF) has developed a cashmere certification scheme. This scheme operates in the Bayankhongor aimag of central western Mongolia. It certifies herders who graze sustainably, and this certification provides access to premium markets with higher prices. The higher prices should offset any household income losses resulting from reduced herd sizes. To be credible, certification must be awarded only to herders whose grazing practices protect or improve the rangelands. The scheme therefore depends on methods to monitor the ecological conditions so that improvements can be demonstrated.

Aims:

We aimed to create user-friendly condition metrics for five important rangeland ecosystems in Bayankhongor aimag that are subject to grazing impacts: Boreal Coniferous Forest, High Mountain Steppe, Moderate Dry Steppe, Dry Steppe, and Extreme Arid Desert.

The metrics must be able to distinguish sites that vary in condition, along a spectrum from intact to degraded, using simple field-measured data.

Methods:

We used a method adapted from previously published work on ecosystems in Australia and the Mongolian Gobi Desert. The metrics are derived from informed opinion, because the approach maintains that the concept of ecological condition is subjective.

One hundred and fifty-one stakeholders contributed quantitative opinion data used to build the metrics. Most contributors were herders (81%). They evaluated a set of hypothetical rangeland sites (one set for each of the five ecosystems). They provided each site with a score between 100 (a desired state) and 0 (no ecological value remains). We used the evaluation data to train a statistical model for each ecosystem that is capable of predicting the score from the site variables.

We tested the metrics against the workshop-derived opinion dataset using cross validation (with test sites not used to train the model). For three of the five ecosystems, we also applied a more stringent test: we compared metric scores from field-measured test sites to scores assigned by a subset of stakeholders who examined the sites in the field. They were asked to give their opinion about site condition, using whatever observations and mental models of condition they felt appropriate.

We also developed a plot-based field method for measuring the relevant site variables in each ecosystem.

Results:

The metric tools are provided as fixed formulas in spreadsheets. The user assesses the site in the field using the plot-based method, then types the results into the spreadsheet. The spreadsheet automatically produces a condition score. To use the metrics, no further stakeholder opinion is required.

When tested against workshop-derived opinion, all metrics performed well, showing clear positive relationships (R^2 0.65 – 0.77). When tested against field-derived opinion data, the metric for High Mountain Steppe performed extremely well ($R^2 = 0.92$). The metric for Moderate Dry Steppe performed well ($R^2 = 0.50$). The metric for Dry Steppe did not perform well, but still showed a positive relationship ($R^2 = 0.28$).

Conclusions and implications:

The data collection method and the metrics for deriving condition scores are fit for monitoring and reporting on rangeland condition to support AVSF’s cashmere certification scheme.

1 Introduction

1.1 The socio-environmental context

The Eurasian Steppe is a grassland ecosystem spanning over 10 million km^2 of Europe and Asia, including most of Mongolia (Barbolini et al. 2020). Despite the extreme cold-arid climate, it has been inhabited by pastoralists for millennia (Fernández-Giménez and Allen-Diaz1999). The characters of the ecosystems of the region have developed under shifting climatic and land use patterns, so natural and anthropogenic influences on the environment are difficult to separate (Miehe et al. 2007).

In southern Mongolia, cashmere (goat hair) is a major income source, supplying high-end fashion products in Europe (Berger et al. 2013). Due to recent high cashmere prices, goat numbers in Mongolia have increased five-fold over the last 30 years (Bureau of Statistics, Ulaanbaatar; Tuvshintogtokh and Ariungerel 2013; Rao et al. 2015). Increased grazing pressure may cause ecological degradation and desertification, including the loss of palatable species, the increase of non-palatable species, the overall loss of vegetation cover, soil erosion, and the exclusion of some wild animals (Tserendash and Erdenebaatar 1993; Fernández-Giménez and Allen-Diaz 1999; Lkhagva et al. 2013, Jamiyansharav et al. 2018; Munkhzul et al. 2021). In years of '*dzud'*, when drought, low vegetation cover and severe winter weather combine to cause shortages of forage, livestock may suffer high mortality (Tachiiri and Shinoda 2012). Although the extent of ecological degradation and its impacts are much-debated (Jamsranjav et al. 2018), it is widely agreed that overgrazing is a serious issue in many parts of the Mongolian rangelands.

To encourage cashmere producers to appropriately manage grazing pressure, maintain sustainable pastoral production practices and ensure good animal welfare, Agronomes et Vétérinaires Sans Frontières (AVSF) co-designed a voluntary sustainable cashmere certification label with local herders. The Sustainable Cashmere Certification Committee (S3C) uses a 25-criteria checklist to guide herders towards eco-friendly practices at individual and collective levels (AVSF 2020). In 2021, S3C certified 505 herding families in the province (aimag) of Bayankhongor, Mongolia.

The S3C scheme certifies herders who graze their herds sustainably, and this certification provides access to premium markets with higher prices. The higher prices should offset any losses resulting from improved practice. To be credible, certification must be awarded only to herders whose grazing practices protect or improve the rangelands. The scheme therefore depends on methods to monitor the ecological condition of the rangeland so that improvements can be demonstrated.

1.2 Evaluation of ecological condition

We use the following definition of ecological condition:

> Ecological condition measures the retention (or loss) of the ecological attributes that characterise an ecosystem in its desired state.

Ecosystem condition cannot be measured directly (unlike length, weight, etc.), because it is a concept made up of several different attributes (Schlacher et al. 2014; Sinclair et al. 2015, 2018, Venables and Boon 2016). A metric for ecosystem condition is a formula for transforming multivariate information into a single number (Figure 1). In other words, it is an algorithm for reducing the dimensionality of data.

It is important to recognise that condition is a subjective idea. The 'desired state' represents what people want an ecosystem to be like. This cannot be inferred from data alone—it must come from people. Condition metrics should speak for a community of stakeholders who have an interest in the ecology of the system. Although it is subjective, most people tend to agree that high ecological condition relates to high species diversity, stable soils, and high levels of biomass or cover (given the productivity of the region) (Fleishman et al. 2006; Batpurev et al. 2022a).

This desired state represents a benchmark for the expectations of stakeholders. Those expectations will vary between ecosystems. For example, we would expect a forest ecosystem to have greater biomass (or total vegetation) than a desert ecosystem. Thus a vegetation cover of 20% in a desert might represent the desired state (a high condition score), but the same cover in a forest might be concerningly low (a low condition score).

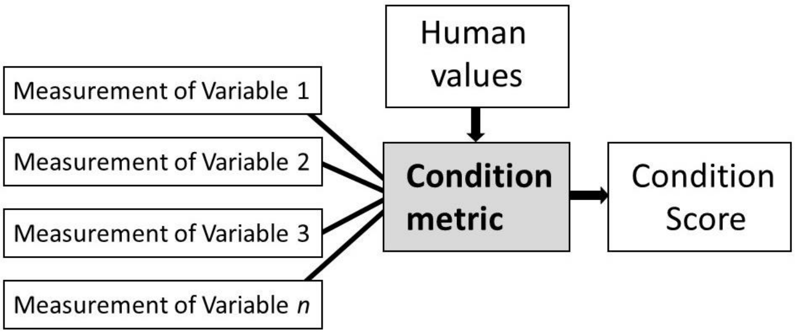


Figure 1. The conceptual structure of a condition metric. Metric is the algorithm, or general rule that determines relationship between the measured variables. Condition score is a specific number generated from measured variables using the algorithm (scenario specific result).

Goat herding in the Bayankhongor aimag occurs across a rainfall gradient, and different ecosystems occur across this gradient. These various systems have been defined and mapped by Gunin and Saandar (2019). Because these differences affect our expectations for vegetation cover across the region, ecological condition assessment within the S3C program requires different metrics for different places. A previous project created metrics for several of the ecosystems relevant to S3C (Desert Steppe, Semi Desert, True Desert, Saxaul, Elm Forest; Avirmed et al. 2018; Sinclair et al. 2021). These metrics can be used in S3C, but new metrics are required for several other ecosystems in Bayankhongor aimag. These are described below and their distribution in Bayankhongor aimag is shown in Figure 2.

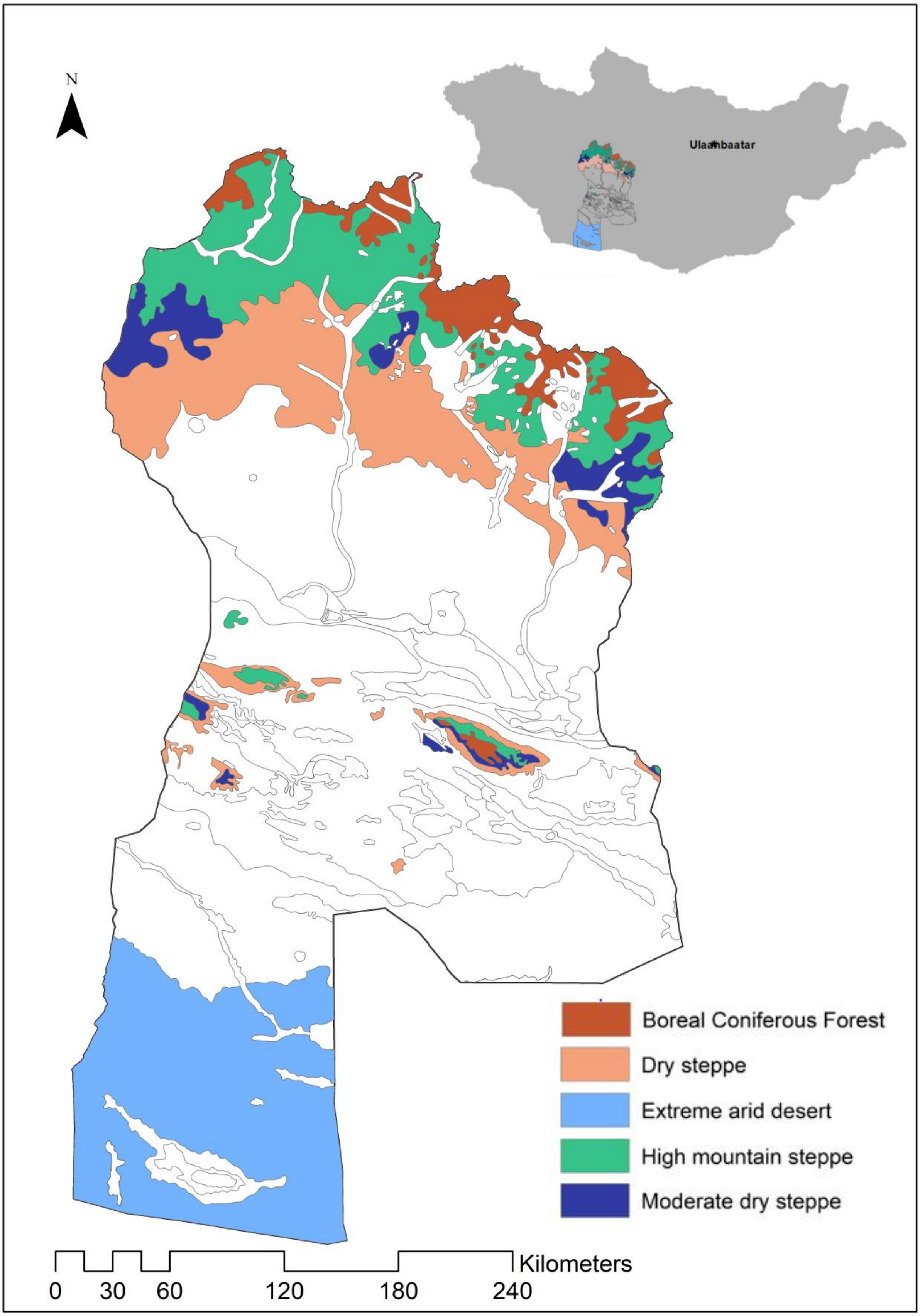


Figure 2. Bayankhongor aimag and the distribution of ecosystems within it (only showing relevant five that are described in this report). The inset shows Bayankhongor aimag in relation to the whole of Mongolia, with capital Ulaanbaatar for spatial context.

1.3 The target ecosystems

1.3.1 Boreal Coniferous Forest

This ecosystem is characterised by the presence of coniferous trees, most commonly the deciduous Siberian Larch (*Larix sibirica*). Other coniferous trees that may be present include the deciduous Dahurian Larch (*Larix gmelinii*) and evergreen species such as Siberian Pine (*Pinus sibirica*) and Siberian Fir (*Abies sibirica*). Total vegetation cover is high in this system, often approaching 100%.

Boreal Coniferous Forest occurs in regions of relatively high annual rainfall, excess of 275 mm approximately. It tends to occur at high elevations, on hilltops and high slopes, sometimes with rock outcrops, and often on warmer south-facing aspects. The soils are often relatively high in organic matter, sometimes peaty, and frequently frozen. The ground may be littered with logs, twigs and leaf litter.

The ground layer is rich in forbs, including berries such as *Fragaria orientalis*, and grasses including *Poa attenuata* and *Calamagrostis* species. Rocky patches often support moss-beds and small succulents such as *Orostachys spicata*. There is sometimes a middle-layer of small trees and shrubs, including species of *Betula*, *Vaccinium* and *Bergenia*, along with *Rosa acicularis* and *Dasiphora fruticosa*.

In Bayankhongor aimag, Boreal Coniferous Forest is at its southern-most limit, and generally occurs as small patches in the hills and ranges of the northern part of the province, often surrounded by High Mountain Steppe (below). Areas further south are too dry. Boreal Coniferous Forest is shown in Figure 3.

Figure 3. An example of Boreal Coniferous Forest (Gorkhi Terelj National Park, outside study area; Photo Steve Sinclair).

1.3.2 High Mountain Steppe

High Mountain Steppe occurs in regions of moderate to high rainfall (150–275 mm), on hilltops, slopes and rocky outcrops, mostly at elevations between 2300 and 2600 m altitude. The soils may be relatively high in organic content but are often shallow and stony, or the surface is covered with scree. Vegetation cover is generally high, sometimes approaching 100%.

The vegetation is dominated by grasses, such as *Agropyrum cristatum*, *Festuca lenensis*, *Koeleria macrantha* and *Helictotrichon mongolicum*. It also includes a range of sedges (*Carex duriuscula*, *Carex stenopheloides*, *Kobresia* spp.), numerous forbs (*Aster hispidus*, *Saussurea* spp., *Potentilla gelida*, *Eritrichium pauciflorum*). Sometimes, shrubs occur in patches or as scattered individuals. These include *Cotoneaster melanocarpus*, *Rosa pimpinellifolia*, *Berberis sibirica* and *Juniperus pseudosabina*.

In Bayankhongor aimag this ecosystem occurs mostly on the ranges in the northern half of the province. It is shown in Figure 4.

Figure 4. An example of High Mountain Steppe, Bayankhongor aimag (Photo Steve Sinclair).

1.3.3 Moderate Dry Steppe

Moderate Dry Steppe occurs on slopes and valley floors (usually 550–1600 m altitude, higher than Dry Steppe described below), in regions with moderate and variable annual rainfall. It generally occupies coarse, brown soils, which are often calcareous, and contain more organic content than Dry Steppe. Vegetation cover is relatively high, often between 40 and 75%, but rarely approaches 100%.

The vegetation is strongly dominated by grasses, including several *Stipa* species (*S. glareosa*, *S. grandis*, *S. krylovii*, *S. baicalensis*), *Agropyron cristatum*, *Festuca lenensis*, *Cleistogenes squarrosa*, *Poa pratensis*, *Poa attenuata*, *Elymus chinensis*, *Koeleria macrantha*, and *Kobresia* spp. It also supports numerous forbs such as *Potentilla* spp., *Saussurea salicifolia*, *Cymbaria dahurica*, *Eremogone capillaris*, *Eurotia ceratoides*, *Kochia prostrata* and *Pedicularis flava*.

Moderate Dry Steppe is extensive and common throughout Bayankhongor aimag and is important animal pasture. It is shown in Figure 5.

Figure 5. An example of Moderate Dry Steppe, Bayankhongor aimag (Photo Steve Sinclair).

1.3.4 Dry Steppe

Dry Steppe occurs on low-relief slopes and pediments in regions with an annual rainfall of 180–250 mm. It often occupies soils that are stony, or often calcareous, and sometimes subject to wind erosion. It is very similar to moderate dry steppe but occupies drier sites at lower altitudes. Vegetation cover is moderate, often between 30 and 70%.

The vegetation is strongly dominated by grasses, including *Stipa. krylovii*, *Stipa gobica*, *Agropyron cristatum*, *Cleistogenes squarrosa* and *Koelaria cristata*. It also supports a range of forbs such as *Allium* spp., *Stellaria dichotoma*, *Ptilotrichum tenuifolium* and *Astragalus brevifolium*. Sometimes, Dry Steppe has scattered shrubs at low cover (e.g., *Caragana* spp.).

Dry Steppe is extensive and common throughout Bayankhongor aimag and is important animal pasture. It is shown in Figure 6.

Figure 6. An example of Dry Steppe, Bayankhongor aimag (Photo Steve Sinclair).

1.3.5 Extreme Arid Desert

Extreme Arid Desert occurs on sandy and stony plains, sometimes with salinisation, in regions where annual rainfall is less than 50 mm. Vegetation cover is very low (often < 10%), and this ecosystem is often subject to wind erosion.

Most of the vegetation is composed of small chenopod shrubs (*Sympegma regelli*, *Iljinia regellii*, *Anabasis brevifolia, Kallidium* spp.) along with *Ephedra przewalskii*, *Nitraria sphaerocarpa*, *Reaumuria soongorica*. Grasses and forbs are rare. After rains, several annual species (such as *Bassia dasyphylla* and *Peganum nigellastrum*) erupt from the seed bank and, for a short time, have high cover.

Extreme arid desert only occurs in the far southern portion of Bayankhongor aimag. It is rarely used as pasture, most grazing there occurring opportunistically when other pastures are exhausted or there have been good rains to encourage fresh growth. This ecosystem is shown in Figure 7.

Figure 7. An example of Extreme Arid Desert, Bayankhongor aimag (Photo Wildlife Conservation Society, Mongolia).

1.4 Specific aims of the current project

The Arthur Rylah Institute for Environmental Research was contracted by AVSF to work with Wildlife Conservation Society Mongolia (WCS) and AVSF to develop and test metrics for the five target ecosystems.

The following specific aims guided the work presented here. They were formulated within the context described above and the inherent limitations on the creation of condition metrics.

The work aimed to produce robust quality metrics for the target ecosystems that:

- can distinguish sites of different condition, including sites at the extreme ends of the condition spectrum,
- are based on data that are easily derived from field plots, which can be completed by any moderately skilled botanist within 1 hour, without follow-up laboratory analysis,
- can detect changes related to land-use change over multi-year periods,
- are not unduly influenced by natural and short-term fluctuations,
- are supported, tested and justified by easily measurable and reliable data,
- are explicitly linked to the views of stakeholders,
- facilitate comparisons of condition both within and between ecosystems.

The metrics are *not* designed to:

- explicitly evaluate habitat for any species of plant or animal (although habitat quality for wildlife does contribute to the conception of condition),
- consider the area, spatial arrangement or context of sites,
- be calculable from remote sensed data.

2 Methods

2.1 Overview

The method used here closely follows our previously published work (Avirmed et al. 2018; Sinclair et al. 2015; 2021; White et al. 2023). It can be summarised as follows:

- We maintain that condition is a subjective quantity, and that a condition metric should represent the judgements of an appropriate stakeholder group.
- We assume that cashmere producers and ecological scientists are appropriate stakeholder groups, because they understand the dynamics of the ecosystems (Fernández-Giménez and Allen-Diaz 2000).
- We gather judgements in a workshop context. Stakeholders are shown hypothetical sites, described by a set of variables that summarise the vegetation at the site.
- The stakeholders are asked to judge the condition of the hypothetical sites on an intuitive scale (here, 0–100).
- We take the resultant dataset (site variables, with related scores) and create models that predict score from the suite of site variables (here, regression trees).
- We test the models, both against the assessments of hypothetical sites, and against assessments of the condition of real field sites.
- The models are converted into user-friendly metric tools (one for each system) in a spreadsheet.
- To assess a site in the field, an assessor physically measures the relevant variables in the field. They enter these measurements into the tool, and a condition score is produced.
- No further expert consultation is required to use the metrics after the spreadsheet has been built.

2.2 Variable selection

We sought a set of measurable site values that capture important information about ecological condition, relevant to each ecosystem. Consistent with the previous work, condition relates largely to the reliably observed characteristics of sites i.e., the vegetation and soil (not the animal community).

Variables selected for the previously published work were shown to be appropriate for systems similar to those targeted here (Avirmed et al. 2018; Sinclair et al. 2021). Thus, we retained the same variables for the desert and steppe ecosystems. We modified the variable set to deal with the different structure and species composition of Boreal Coniferous Forest. These modified variables were developed in consultation with ecological experts at WCS, using the principles of variable selection described by Avirmed et al. (2018). Table 1 shows the variables selected and how they apply to the previous project and each of the ecosystems covered here. Appendix 1 provides definitions of the terms used to define the variables.

Table 1. The variables used to assess condition for each ecosystem.*

Variable	Previous work	BCF	HMS	MDS	DS	EAD
Total vegetation cover	✓	✓	✓	✓	✓	✓
Cover of all shrubs	✓	✓	✓	✓	✓	✓
Richness of all shrubs	✓	✓	✓	✓	✓	✓
Cover all perennial grasses and sedges	✓	✓	✓	✓	✓	✓
Richness of all grasses and sedges	✓	✓	✓	✓	✓	✓
Cover of perennial forbs	✓	✓	✓	✓	✓	✓
Richness of all forbs	✓	✓	✓	✓	✓	✓
Cover of organic litter	✓	✓	✓	✓	✓	✓
Max. height of exposed root pedestals	✓		✓	✓	✓	✓
Cover of annual forbs	✓	✓	✓	✓	✓	✓
Cover of all annual grasses and sedges	✓	✓	✓	✓	✓	✓
Cover of all succulent shrubs	✓		✓	✓	✓	✓
Cover of fabaceous shrubs	✓		✓	✓	✓	✓
Cover of *Artemisia* species	✓		✓	✓	✓	✓
Cover of coniferous trees		✓				
Cover of broadleaved trees		✓				
Density of large conifers		✓				
Density of conifer saplings		✓				
Cover of berry plants		✓				
Cover of dead wood		✓				
Total number of variables	**14**	**16**	**14**	**14**	**14**	**14**

* Appendix 1 gives definitions of all terms used. The ecosystem names are abbreviated: BCF (Boreal Coniferous Forest), HMS (High Mountain Steppe), MDS (Moderate Dry Steppe), DS (Dry Steppe) and EAD (Extreme Arid Desert).

2.3 Consultation scenarios

We required evaluations of numerous sites, spanning a wide range of variation in each ecosystem. Given the unavoidable logistical constraints associated with field assessments in the vast steppe region, most site evaluations were undertaken in a workshop context, using sites represented on paper cards. The cards included the random number identifying the site, the values of each variable (with Mongolian text for the variable names), and a space for the stakeholder to write the score. Each card described a single site for a single ecosystem, using the variables in Table 1. An example of a card is shown in Figure 8.

Each expert assessed a set of 16 site cards (or 18 in the final workshop; see below). Each set was composed of cards from the following types. The experts were not aware of these different card types.

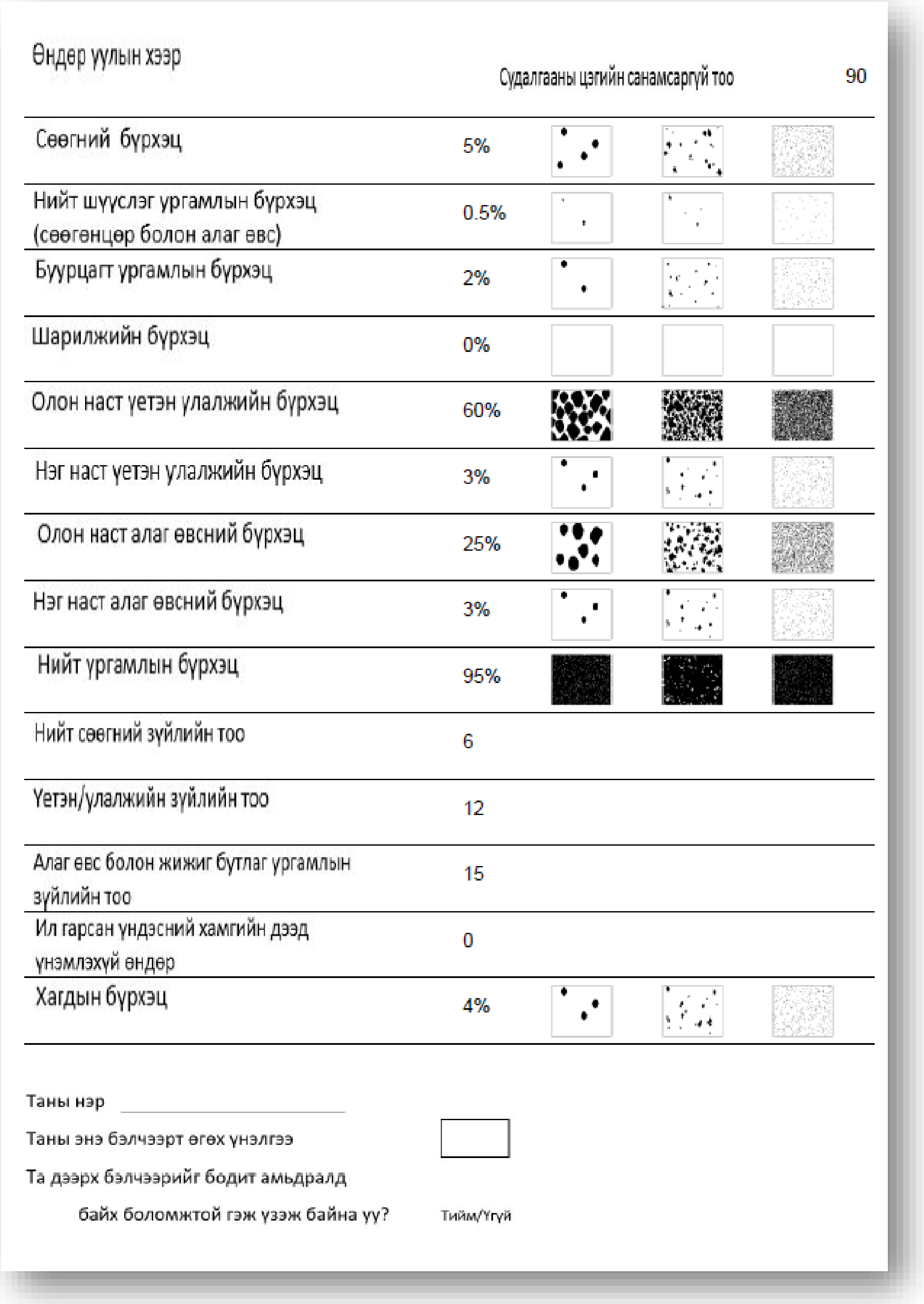
- **Card type 1: 'Common' sites** (*n* = 10 per ecosystem in total; every expert given 1 or 2 at random). These sites represent typical or widespread vegetation conditions for each ecosystem type. They represent the environmental space over which the metrics are most likely to be operating, and they were given to participants at a relatively high rate to ensure good model training in this important domain. They were constructed by estimating the variables from photographs of each ecosystem representing vegetation known to be widespread and unexceptional, as well as from the quantitative data in Fernández-Giménez and Allen-Diaz (2001).
- **Card type 2: General synthetic sites** (*n* = 203 per ecosystem in total; every expert given 11 or 12 at random). These sites were designed to ensure that the dataset included sites covering the widest conceivable range of variation within each ecosystem and included a wide range of permutations of values for each variable. They were created manually by examining photographs and depicting all types of variation that could be found, and by systematically varying each variable against all others.
- **Card type 3: Calibration sites** (*n* = 3 per ecosystem; every expert given all 3). It is important that all stakeholders are evaluating sites on a common scale. To encourage this we included a common set of pre-judged 'high' (2 cards) and 'low' (1 card) calibration sites. The 'high' calibration cards were made by hand, subjectively, in consultation with WCS. They represented the most intact, highest condition sites that were considered possible for each system (generally very high species richness combined with high cover). The 'low' calibration card represented a site with no vegetation cover (all cover and richness variables set to 0), and some erosion (the maximum height of roots exposed by erosion set to 20 cm).

An additional card type, based on our field work and therefore not available earlier, was added for the final two workshops (Bayankhongor town centre and Ulaanbaatar):

- **Card type 4: Real field sites** (*n* = 9 in total for High Mountain Steppe, 7 for Dry Steppe, 7 for Moderate Dry Steppe; none for other systems. Each expert was given 2 cards at random, in addition to the 16 cards described above). These cards directly represented the field sites we had sampled in the first half of our field work, converted into cards in time for the final two workshops.

We automated the production of the cards (as PDF files ready for use) from the data describing the sites. We did this by creating images of each Mongolian phrase and creating a script to call up the correct image and place it in the correct position on the card, based on the data describing the sites. This was processed with three packages jpeg (Urbanek 2014), plotrix (Lemon 2006) and grDevices in R (R Core Team 2016).

We also presented cover diagrams to help stakeholders who were not familiar with visualising covers from numerical values. Each cover value was represented by three cover images accurately representing that cover: one strongly clumped, one more dispersed, and one randomly dispersed. These were produced by colouring cells on a grid—black for cover and white for no cover—using ArcGIS 10.3 (ESRI). The pixel counts confirm that the cover represented in each image is correct.



Өндөр уулын хээр

Судалгааны цэгийн санамсаргүй тоо 90

Сөөгний бүрхэц	5%
Нийт шүүслэг ургамлын бүрхэц (сөөгөнцөр болон алаг өвс)	0.5%
Бүүрцагт ургамлын бүрхэц	2%
Шарилжийн бүрхэц	0%
Олон наст үетэн улалжийн бүрхэц	60%
Нэг наст үетэн улалжийн бүрхэц	3%
Олон наст алаг өвсний бүрхэц	25%
Нэг наст алаг өвсний бүрхэц	3%
Нийт ургамлын бүрхэц	95%
Нийт сөөгний зүйлийн тоо	6
Үетэн/улалжийн зүйлийн тоо	12
Алаг өвс болон жижиг бутлаг ургамлын зүйлийн тоо	15
Ил гарсан үндэсний хамгийн дээд үнэмлэхүй өндөр	0
Хагдын бүрхэц	4%

Таны нэр ______________

Таны энэ бэлчээрт өгөх үнэлгээ

Та дээрх бэлчээрийг бодит амьдралд байх боломжтой гэж үзэж байна уу? Тийм/Үгүй

Figure 8. An example of a site card for consultation. The label at the top right identifies the ecosystem (in this case, High Mountain Steppe). The number at the top right is a random site identifier. The cover variables are shown with example cover diagrams. At the bottom, a space is provided for the stakeholder to provide their name (optional), their score, and their assessment of whether the site was plausible.

2.4 Stakeholder workshops

Our metrics had to represent the judgements of a wide and diverse range of stakeholders. To ensure representation among nomadic herders in remote areas, five workshops were convened at local centres across Bayankhongor aimag (Bayankhongor, Bumbugur, Shinejinst, Bayantsagaan, Bogd (Figure 9).

Figure 9. Workshop locations in Bayankhongor aimag. The inset shows the location of the aimag in Mongolia.

We also convened a workshop targeting scientists and policymakers in Ulaanbaatar. All workshops took place in August 2022. Unlike previous work (Avirmed et al. 2018), we did not gather any additional assessments from stakeholders via online means. AVSF recruited stakeholders from the S3C scheme and their existing networks. Each participant was offered about 2 days minimum wage. In total we consulted 151 participants (see Results).

Each workshop began with an introduction to the context and aims of the project, an explanation of how we intended to collect and use data, an understanding that all data provided by the stakeholders would be treated anonymously, and an invitation to participate in further discussion, data analysis and report writing after the workshops. We allowed time for questions. All written and verbal communication was in Mongolian. Consultation aligned with the principles of 'free, prior, informed consent' (FAO 2016), and all stakeholders consented to their answers being used in the manner described above, by signing a consent form.

We then defined what we meant by 'ecological condition' to the workshop participants, consistent with previous work in Mongolia (Avirmed et al. 2018), that is:

> Ecological condition measures the retention (or loss) of the ecological attributes that characterise an ecosystem in its desired state.

We explained that condition may include elements of 'quality', 'intactness', 'health' or 'conservation value', including any or all of the following factors (to any degree):

- the value of the site in providing key ecological functions,
- the provision of habitat for the wildlife of the ecosystem,
- the provision of habitat for the plants of the ecosystem,
- the stabilisation of the soil,
- the value of the site as an example of its type,
- the abundance of particularly important species or life-forms,
- how important the site should be for conservation / protection,
- the degree to which the site resembles a site that has suffered no loss of condition,
- how much a well-informed (expert) stakeholder 'likes' the site.

We explained that the following considerations were not to be included in the conception of condition (although their importance in other contexts is acknowledged):

- the personal wealth that could be derived from the site (livestock or money),
- the value of the site for any other purpose other than as an example of its ecosystem type,
- the likely future for the site (whether good or bad),
- the cost of rehabilitating the site.

Stakeholders were instructed to first look over their sites, imagine them, and rank them from highest to lowest condition (Figure 10). Ties in rank were permitted. They were then asked to quantitatively evaluate site condition by writing a score on each card reflecting the quality of the sites. Again, ties were permitted. It was explained that scores need not be evenly distributed across the range of possible scores.

Figure 10. Stakeholders participating in the elicitation exercise, Shinejinst local hall (Photo Steve Sinclair)

All participants were required to mark one of their cards 0 and one card 100. Based on previous work, we expected that most stakeholders would perceive the 'low' and 'high' calibration cards to represent 0 and 100 respectively. Most did this, as expected. If, however, any stakeholders felt that their concept of 0 or 100 was not represented in their card set, they were invited to create their own new card that represented 0 and/or 100. Only a small minority chose to do this, resulting in 3 new cards created for Boreal Coniferous Forest, 1 for High Mountain Steppe, 6 for Moderate Dry Steppe, 7 for Dry Steppe and 7 for Extreme Arid Desert. These cards were added to the final dataset. All people that elected to create their own 0 or 100 cards still gave a score below 20 for the low calibration card, or above 80 for the 'high' calibration card, so we decided not to rescale any data.

We also asked for a judgement as to whether each site was plausible in reality. In previous work we used these data to correct metrics which did not perform well initially (Sinclair et al. 2020), so we collected the equivalent information here in case it was required for a similar correction process. As discussed below, the models performed well, so no correction was necessary and these data were not used.

Clarifying questions were allowed throughout, but no communication was permitted between participants during the quantitative elicitation exercises.

We used a paper questionnaire to collect additional information about the participants' expertise and affiliations. This information was used to characterise the pool of people we consulted, and hence the pool of judgement represented by the resultant metric tool.

2.5 Data treatment, model testing and metric creation

Our approach to the treatment of data and the testing of models differed slightly from our previous approach and was based on subsequent unpublished work testing and evaluating similar models. The modelling approach was identical for all five ecosystems treated here.

2.5.6 Data treatment

Since the stakeholder scores are bounded by 0 and 100, the stakeholder scores were transformed using arcsine square root transformation before the analysis, to improve model performance. All models described below used these transformed values as the dependent variable. The final metrics, and all reported results below, use back-transformed values on the original scale (0–100).

Because the calibration sites were assessed by every stakeholder, and most of the 'low' sites were scored 0, we discarded most of this data and used one representative 'low' calibration row only.

For each dataset (with n values), we used a 'leave-one-out' cross-validation approach for outlier detection. We used $n - 1$ data points to build a random forest model, using the R randomForest package (Liaw and Wiener 2002). We applied that model to the $n - 1$ training points and the one remaining test point. We repeated this n times so that every data point served as the test point once.

For each point, we calculated the difference between the actual score and the predicted score from when that point was left out as a test. We calculated the standard deviation (sd) of the absolute prediction errors. We identified outliers as any observations with an absolute prediction error larger than 2 sd. We removed these outliers from all further analyses.

2.5.7 Test models

We divided the full range of scores (0–100) evenly into 10 segments. We randomly selected 10 points from each group as test data. From the remaining data, we randomly selected 20 points from each group (with replacement if less than 30 in the group) as training data. We built a regression tree model with these training data points using the R rpart package (Therneau and Atkinson 2022). We applied this model to the test data to obtain predicted values. We repeated the above selection of training data and modelling 30 times. Finally, the predictions for the 10 test data points from the 30 models were aggregated, using both the mean and the median, to see which aggregation method was best.

2.5.8 Final models

To create the final metrics, we randomly selected 30 points from each stratum (that is, each 10-point range, as described above) to use as training data, and built a regression tree model. We repeated this 30 times and obtained 30 models for each ecosystem type. Theoretically these final models should be better than the tested models, above, since more training points were used here.

These 30 models were each transformed into IF/THEN statements using syntax suitable for Microsoft Excel. These are incorporated into a spreadsheet which forms the metric tool. The final tool aggregated these 30 models using their median value. We also explored aggregation using the mean but found that performance was essentially the same whether median or mean aggregation was used. We used the median for consistency with previous work (Avirmed et al. 2018).

2.6 Field assessment method

The field sampling method we used for the previous work was shown to be effective and practical (Avirmed et al. 2018; Sinclair et al. 2021). We retain that method here, unchanged, and describe it briefly below.

Condition is always assessed at a discrete site: a square plot that measures 30 x 30 m (900 m^2).

Within this plot, 4 parallel tape measures are laid out, crossing the plot edges at 6 m, 12 m, 18 m and 24 m. Each of these tape measures defines a point intercept sampling line. 120 sampling points are distributed evenly along each line, spaced every 0.25 m (commencing at 0.25, ending at 30.0), totalling 480 points per plot (Figure 11).

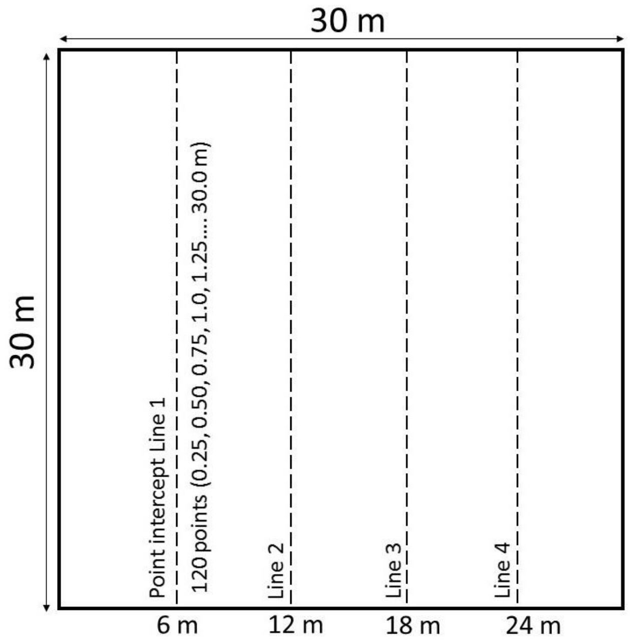


Figure 11. The plot method used to sample vegetation in the field. Left: Diagram of the plot used to sample vegetation in the field. Right: Observers quantifying vegetation cover in a plot, using tape measures to define the point intercept lines. This plot is a degraded example of Dry Steppe, with low cover as a result of heavy grazing.

At each point a narrow steel pin is held vertically, and any plant species or organic litter in contact with the pin is recorded. Multiple species (and litter) can be recorded at a single point, but each species can only be recorded once per point (i.e., we do not quantify overlapping cover). The cover of each species (and litter) is calculated using the following formula:

Percentage cover of species = (number of points species recorded / 480) x 100

These species-specific cover data can be used to calculate all of the cover-based variables required in the metric tool (such as cover of all shrubs, cover of berry plants), by summing the covers of all species in each lifeform category.

To sample species richness, a single experienced botanist examines the plot for 10 minutes, recording all vascular plant species, regardless of their cover. Richness values for each of the lifeforms are calculated by simply counting the number of species in each lifeform.

To quantify the maximum height of roots exposed by soil loss, a single observer quickly checks the root systems of all shrubs in the plot. For shrubs where some of the root system was exposed by soil loss, the vertical distance between the root–shoot junction and the point of contact with the current soil level is measured (Figure 12). The maximum distance found on any shrub in the plot is recorded. This process is easily completed within the 10-minute search time allotted to the botanical observer.

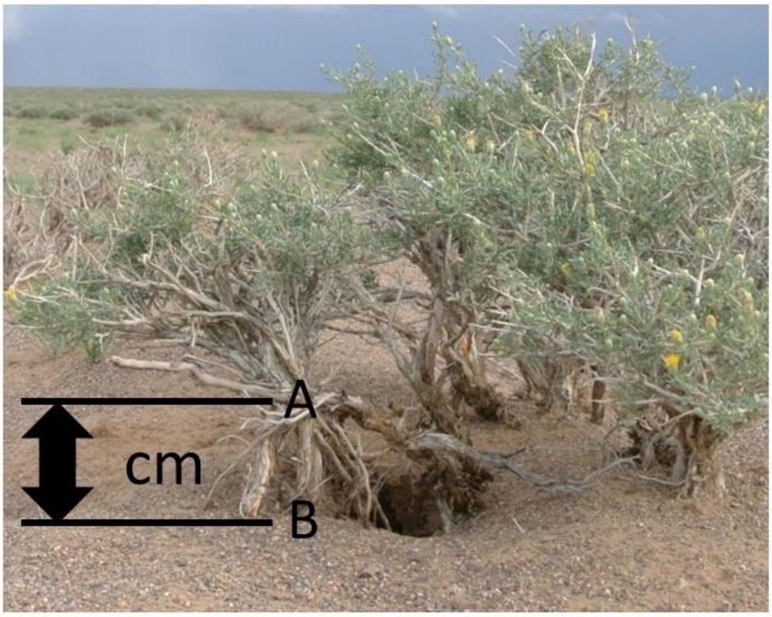


Figure 12. Measurement of roots exposed by soil loss. Maximum height of exposed root is the vertical distance between the root–shoot junction visible on a plant (A) and the junction between the plant's root system and the soil level (B), in centimetres. The example shown uses *Brachanthemum gobicum* (Asteraceae) (Photo Steve Sinclair).

2.7 Metric evaluation using field data

We tested the real-world performance of the metric tools by trialling them in the field. We did this for High Mountain Steppe, Moderate Dry Steppe and Dry Steppe. The other two ecosystems were not tested in the field due to limited accessibility during our field schedule.

For test sites we selected a range of sites across the landscape that clearly differed in their vegetation characteristics. They included sites with widely different terrain, geomorphology and vegetation cover, along with sites adjacent to camps and roads, and sites very distant from camps and roads. Our aim was to stretch our test as far across the condition spectrum as we could.

At these sites we implemented the sampling method described above, so that we were able to derive a condition score for each site using the final metric tools.

We took a group of stakeholders to all these sites. We asked them to examine each plot (approx. 10–30 minutes was allowed) and evaluate the condition of the site, using a score between 0 (no ecological value) and 100 (the highest ecological condition they could imagine for vegetation of this kind, at this site, in August, following a year of normal rainfall).

The evaluations were carried out independently, with no communication between participants. The participants were not provided with any instruction on which variables to consider, nor how to interpret or weight them. The stakeholders were aware that vegetation cover, species richness and soil loss were variables under consideration, given that these were measured at each plot.

The stakeholders were ecological experts (n = 8), a 'policymaker and land administrator' working for AVSF (n = 1), and one person with a herding background. Most of the authors of this report were among the stakeholders. Occasionally, we opportunistically enlisted local herders to provide their judgements, and occasionally one or more stakeholders were unavailable for a given site. All sites were assessed by 8 to 11 people.

4 Results

4.1 The characteristics of the stakeholders

It is important to describe the characteristics of the stakeholders who participated in the workshops, because the final metric tools represent their collective judgements of ecological condition. Put another way, the metrics are intended to speak for a defined stakeholder group.

Figure 13 shows the spread of expertise within our stakeholder group. The data are based on self-assessment surveys undertaken by stakeholders. It is notable that the expertise here is far more focussed on herding practice than in our previous work, which had a higher representation of scientists and land administrators.

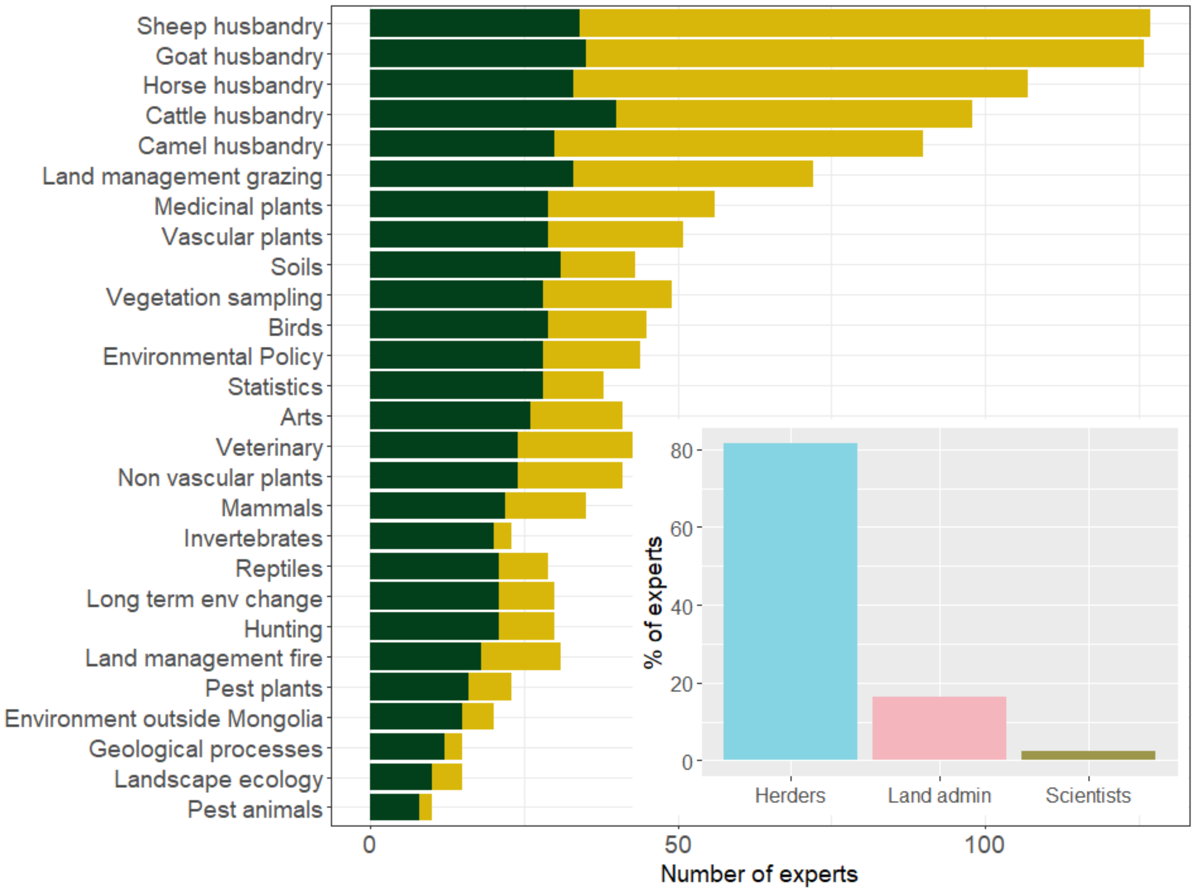


Figure 13. A summary of the expertise within the stakeholder set.

The dark green bar represents the number of participants who indicated they were ‘fairly knowledgeable’ in the topic area; the yellow bar represents those who indicated they were ‘expert’. Inset: ‘Herder’ means anyone who derives their primary income from herding, regardless of their other affiliations or knowledge. ‘Scientist’ means anyone employed by a university or government agency with a role in data collection or analysis (WCS, university researchers). ‘Land administrator’ means anyone involved in the organisation, regulation, logistics or social support of the herding industry and or rangeland management (staff of pasture user groups, local administrators, AVSF, etc.).

4.2 Evaluation of metric performance

4.2.9 Metric cross validation, tested against workshop-derived data

We tested the metrics for all five ecosystems against the workshop-derived opinion dataset. We used a cross-validation approach, where each test point was not used to train the model being tested (see Methods). All metrics performed well, showing clear positive relationships (R^2 = 0.67 – 0.86; Figure 14). This result is comparable to the metrics for Mongolian ecosystems we have previously published (R^2 = 0.68 – 0.82; Avirmed et al. 2018; Sinclair et al. 2021). It demonstrates that the metrics can represent the middle ground of the stakeholder opinion.

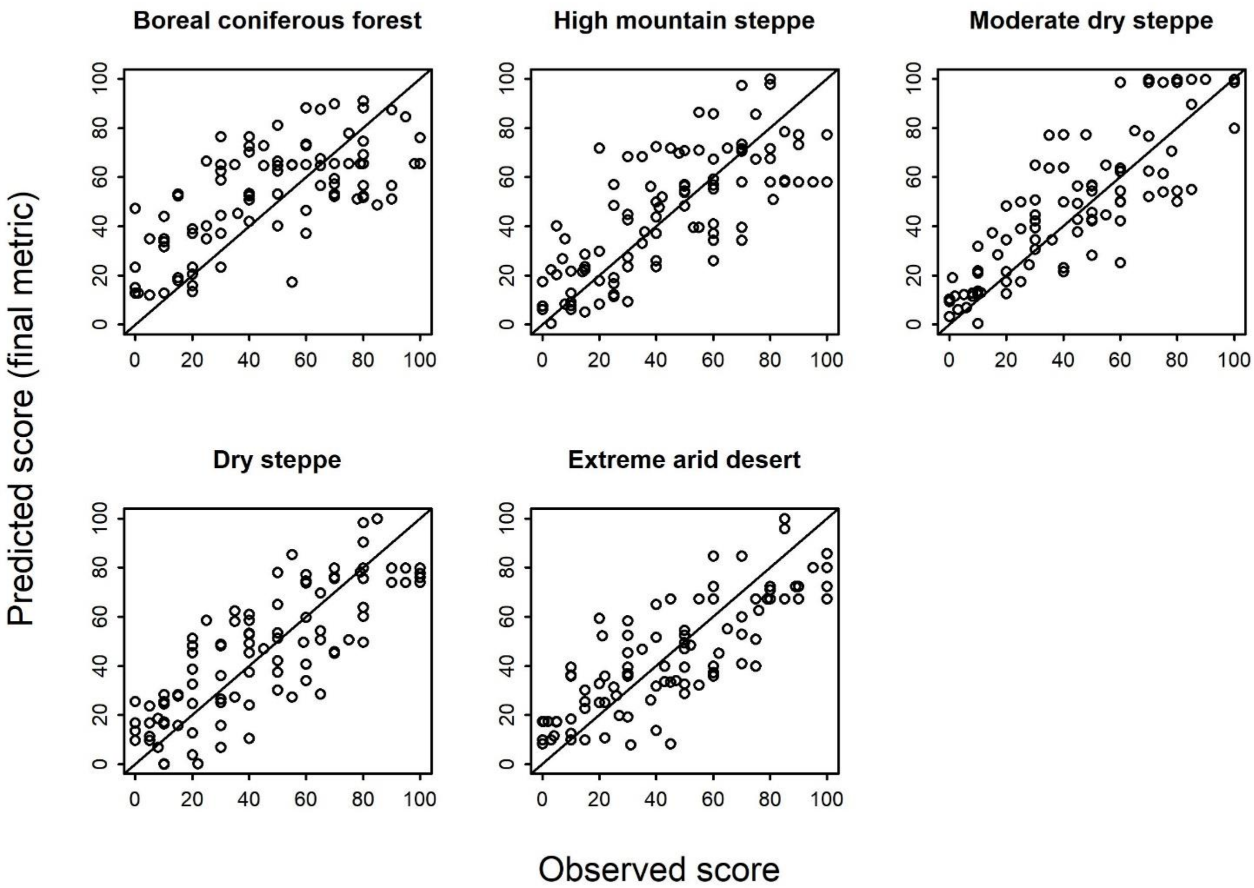


Figure 14. The predictive capacity of the metrics, for each ecosystem. R^2 values: Boreal Coniferous Forest 0.67, High Mountain Steppe 0.74; Moderate Dry Steppe 0.86, Dry Steppe 0.80, Extreme Arid Desert 0.81.

The horizontal axis shows the median score provided by stakeholders for each site. The vertical axis shows the final metric score (median of 30 regression trees). The results are from cross-validation, where the test sites are not used to train the models that predict their score.

4.2.10 Field testing of metrics

We also tested three of the metrics in the field. We compared the metric scores for real sites measured in the field, to judgements made about those same sites by stakeholders who inspected them in the field.

All metrics showed a positive relationship with the test data (Figure 15). However, metric performance differed substantially among the ecosystems. High Mountain Steppe performed exceptionally well (R^2 = 0.92), Moderate Dry Steppe moderately well (R^2 = 0.5), and Dry Steppe far less well, with a weaker positive relationship (R^2 = 0.28). This field-based test is much more stringent than the cross-validation test presented above (see Discussion below).

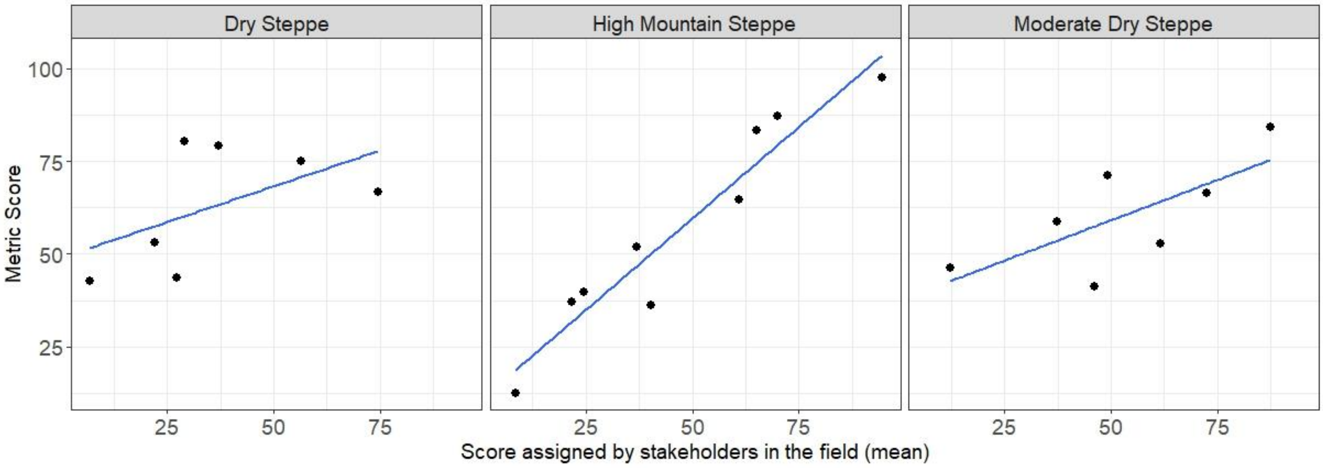


Figure 15. The performance of the metrics, tested with field assessments.

The horizontal axis shows the mean score provided by stakeholders for each site, assessed in the field. The vertical axis shows the metric score for that site.

5 Discussion

By implementing the field assessment described here, and entering the results into the relevant metric, a score can be produced that represents the condition of the assessed site. This score ranges between 0 and 100, where 100 represents the best possible condition for the given ecosystem.

Clearly, these are useful tools for the S3C program. The pastures of participating herders can be assessed, and degradation or improvement in pasture conditions can be detected. This should allow the scheme to certify those herders who are implementing grazing practices which result in the maintenance or improvement of condition in their pastures. It can also identify those herders whose pastures are suffering degradation, so that the S3C scheme can work with those herders to help them achieve certification.

The metrics presented here are recommended for use, but there are several conditions that should be kept in mind when using and interpreting them. We describe the details of these conditions in the ‘Caveats and conditions’ section below.

One of the strengths of the metrics presented here is that they ‘speak for’ the stakeholders who provided judgement. In this case, these stakeholders were mostly nomadic herders. This seems appropriate for S3C. Some certification schemes have been criticised for their top-down colonial attitudes, and lack of engagement with producers (Vandergeest and Unno 2012). Direct, explicit representation of the opinions of producers means that herders can have a collective voice in the way that certification is awarded.

Batpurev et al. (2022a) showed that the judgments of herders, scientists and policymakers were extremely similar. Each group of stakeholders produced opinions that mapped closely onto other groups of stakeholders. For this reason, we expect that the metrics produced here will be suitable for use alongside the previous metrics (Avirmed et al. 2018; Sinclair et al. 2021), even though those previous metrics were based on a different stakeholder set (fewer herders, more scientists and policymakers).

This potential difference in stakeholder groups (e.g., herders, vs scientists) is also relevant to the field tests that we implemented here. While the metrics were created from a dataset that was largely composed of herder judgements, the stakeholders who made the field assessments were largely scientists and policymakers. While it is possible that this difference explains some of the reduced performance in the field test, we don’t believe it is the major factor, for two reasons: firstly, because the same stakeholder sets produced a wide variety of test results (see Figure 15), and secondly because the work of Batpurev et al. (2022a) suggests that we should not expect too much difference between stakeholder groups.

5.1 Caveats and conditions for using the metric

Condition 1 The metrics are intended for use in summer (August), when vegetation cover is high. Measurement of pastures in other seasons may result in scores that are suppressed by lower vegetation cover.

Condition 2 Condition scores will fluctuate between years, depending on local patterns of rainfall. In wetter years, vegetation cover may be higher, and many species may be easier to detect, resulting in higher scores. These changes may be transient and may not represent meaningful shifts in pasture condition. Any assessment of grazing outcomes should therefore be considered in context: over multiple years, and in comparison, to other assessments at different locations taken in the same year.

Condition 3 The metrics produce results with some variation, noise and error that is not related to long-term condition changes. For example, a small drop in condition score does not necessarily mean that a pasture is degraded. Larger changes, and consistent longer-term changes are of far more concern.

Condition 4 Pastures under the influence of any herding family will differ markedly across space. Pastures adjacent to a regular winter camp are likely to be degraded, with low perennial grass cover and high cover of opportunistic annuals, whereas pastures on the margins of herder influence are likely to retain higher grass cover.

For these reasons we recommend a certification process which includes regular August assessments and uses multi-year data to make certification decisions. We also recommend that sampling is carefully and consistently positioned within the pastures of each herder, so that fair comparisons are made. One way to achieve this consistency is to position plots near to the winter camp (1-3km radius), where it is known that the herds return year after year. These issues were discussed in detail in our review of monitoring data in the Gobi region (Batpurev et al. 2022b).

5.2 Model performance

The field test we used to evaluate the metric is very stringent, and simultaneously tests a number of underlying factors. For a strong positive correlation to be found, when we compare stakeholders' judgements in the field with metric scores applied to those same field sites, ALL of the following conditions must be met:

- The variable set must adequately capture those attributes that stakeholders use to understand condition.
- The field measurement method must adequately quantify those variables.
- The hypothetical scenarios used in the workshops must present enough realistic variation to train a model that is useful in the field.
- The stakeholder set that carries out field assessment must have consistent judgements compared with the stakeholders who provided their judgements in the workshop context.
- Stakeholders in both workshop and field settings should share a common understanding of each ecosystem and its potential expression under conservative management in an 'average rainfall' year.
- The model fits the data well, such that the final metrics have good prediction capacity.

A breakdown in any one of these factors will result in reduced correlation, and reduced performance on the field test. Small errors in several of these factors will compound to erode test performance.

When this test was employed in previous work (Avirmed et al. 2018), all metrics performed very well (R^2 0.78 – 0.82;). In the present study the ecosystems were more divergent; High Mountain Steppe performed exceptionally well, but Dry Steppe performed poorly.

There are several reasons why metric performance varies between systems. We believe the primary reason is that some systems are tightly defined so that most people have a similar understanding of their dynamics and a consistent search image for good quality exemplars (Batpurev et al. 2022a). For these systems there is considerable consensus, and the metrics perform well. High Mountain Steppe is one such system; it has a clear niche that is relatively narrowly conceived and a clear set of species that characterise the system. On the other hand, Moderate Dry Steppe is widespread across many different landforms, and stakeholders may have differing views of what it is and how it should look, leading to more variation in judgements and a lower-performing metric.

Another reason for differing metric performance is simply that with such small sample sizes (field assessed sites ranged from only 5–7 per ecosystem), much of the variation in correlation is simply 'noise'. In other words, a few unusual sites or a few non-representative assessments from stakeholders may be enough to cause a large reduction in R^2 (Liu et al. 2018).

Overall, despite some variation in performance, the metrics described here are clearly capable of calculating condition scores that relate meaningfully to stakeholder perceptions of condition. They represent stakeholders, but they also bring consistency and transparency: while any individual stakeholder can provide a judgement of condition, the metrics bring the benefit of being strictly repeatable and transparent. These attributes add greatly to the credibility of assessment, and consequently add to the credibility of the S3C scheme. We hope that the metrics produced here can strengthen the S3C scheme, and ultimately lead to better management of Mongolian rangelands.

References

Avirmed, O., White, M.D., Batpurev, K., Griffioen, P.A., Liu, C., Jambal, S., Sime H, Olson, K. and Sinclair, S.J. (2018). Rangeland condition metrics for the Gobi Desert, derived from stakeholder evaluations *Arthur Rylah Institute for Environmental Research Technical Report* 289. Arthur Rylah Institute and Wildlife Conservation Society Mongolia Country Program, report for Oyu Tolgoi.

AVSF (2020). *Certification AVSF Sustainable Cashmere*. Agronomes et Vétérinaires Sans Frontières, Ulaanbaatar, 2020.

Barbolini, N., Woutersen, A., Dupont-Nivet, G., Silvestro, D., Tardif, D., Coster, P.M.C., Meijer, N., Chang, C., Zhang, H.-X., Licht, A., Rydin, C., Koutsodendris, A., Han, F., Rohtmann, A., Liu, X.-J., Zhang, Y., Donnadieu, Y., Fluteau, F., Ladant, J.-B., LeHir, G. and Hoorn, C. (2020). Cenozoic evolution of the steppe-desert biome in Central Asia. *Science Advances* **6,** eabb8227.

Batpurev, K., Sinclair, S. J., Avirmed, O., Scroggie, M. P., Olson, K., and White, M. D. (2022a). Stakeholders from diverse backgrounds make similar judgments about ecological condition and collapse in Mongolian rangelands. *Conservation Science and Practice* **4,** e574.

Batpurev, K., Liu, C., Sinclair S.J., Avirmed, O. and Olson, K. (2022b). Monitoring the condition of rangelands in the Gobi Desert: Monitoring strategies to detect change. *Arthur Rylah Institute for Environmental Research Technical Report Series* 337. Department of Environment, Land, Water and Planning, Heidelberg, Victoria.

Bureau of Statistics (n.d.). *National Statistics Office*. Mongolian Statistical Information Service. https://www.1212.mn/en

Gunin PD and Saandar M (eds) (2019). *Ecosystems of Mongolia Atlas*. A.N. Severtsov Institute of Ecology and Evolution of Russian Academy of Sciences. Admon. Ulaanbaatar.

Berger, J., Buuveibaatar, B. and Mishra, C. (2013). Globalization of the cashmere market and the decline of large mammals in Central Asia. *Conservation Biology* **27,** 679–689.

FAO (2016). *Free, Prior and Informed Consent. An indigenous peoples' right and a good practice for local communities. Manual for Project Practitioners*. Food and Agricultural Organization (FAO) of the United Nations, 2016. 52 pages.

Fernández-Giménez, M.E. and Allen-Diaz, B. (1999), Testing a non-equilibrium model of rangeland vegetation dynamics in Mongolia. Journal of Applied Ecology **36**, 871–885.

Fernández-Giménez, M.E. and Allen-Diaz, B. (2000). The role of Mongolian nomadic pastoralists' ecological knowledge in rangeland management. *Ecological Applications* **10,** 1318–1326.

Fernández-Giménez, M.E. and Allen-Diaz, B. (2001). Vegetation change along gradients from water sources in three grazed Mongolian ecosystems. *Plant Ecology* **157,** 101–118.

Fleishman, E., Noss, R. F., and Noon, B. R. (2006). Utility and limitations of species richness metrics for conservation planning. *Ecological Indicators* **6,** 543–553.

Jamiyansharav, K., Fernández-Giménez, M.E., Angerer, J.P., Yadamsuren, B. and Dash, Z. (2018). Plant community change in three Mongolian steppe ecosystems 1994–2013: applications to state-and-transition models. *Ecosphere* **9,** 1–26.

Jamsranjav, C., Reid, R.S., Fernández-Giménez, M.E., Tsevlee, A., Yadamsuren, B. and Heiner, M. (2018). Applying a dryland degradation framework for rangelands: the case of Mongolia. *Ecological Applications* **28,** 622–642.

Lemon, J. (2006). Plotrix: a package in the red light district of R. *R News* **6,** 8–12.

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

Liu, C, White, M, Newell, G. (2018). Detecting outliers in species distribution data. *Journal of Biogeography* **45**:, 164–176.

Lkhagva, A., Boldgiv, B. Goulden, C. Yadamsuren, O. and Lauenroth, W. (2013). Effects of grazing on plant community structure and aboveground net primary production of semiarid boreal steppe of northern Mongolia. Grassland Science **59**, 135-145.

Miehe, G., Schlütz, F., Miehe, S., Opgenoorth, L., Cermak, J., Samiya, R., Jager, E. and Wesche, K. (2007). Mountain forest islands and Holocene environmental changes in Central Asia: A case study from the southern Gobi Altay, Mongolia. *Palaeogeography, Palaeoclimatology, Palaeoecology, 250*(1), 150-166.

Munkhzul O, Oyundelger K, Narantuya N, Tuvshintogtokh I, Oyuntsetseg B, Wesche K. and Jäschke Y. 2021. Grazing Effects on Mongolian Steppe Vegetation—A Systematic Review of Local Literature. *Frontiers in Ecology and Evolution* **9**, article 703220, doi 10.3389/fevo.2021.703220.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rao, M.P., Davi, N.K., D'Arrigo, R.D., Skees, J., Nachin, B., Leland, C., Lyon, B., Wang, S-Y. and Byambasuren, O. (2015). Dzuds, droughts, and livestock mortality in Mongolia. *Environmental Research Letters* **10,** article 0740012. doi:10.1088/1748-9326/10/7/074012

Schlacher, T.A., Schoeman, D.S., Jones, A.R., Dugan, J.E., Hubbard, D.M., Defeo, O., Peterson, C.H. Weston, M.A., Maslo, B., Olds, A.D., Scapini, F., Nel, R., Harrise, L.R., Lucrez, S., Lastra, M., Huijbers, C.M. and Connolly, R.M. (2014). Metrics to assess ecological condition, change, and impacts in sandy beach ecosystems. Journal of Environmental Management **144**, 322–335.

Sinclair, S. J., Griffioen, P., Duncan, D. H., Millett-Riley, J. E. and White, M. D. (2015). Quantifying ecosystem quality by modeling multi-attribute expert opinion. *Ecological Applications* **25,** 1463–1477.

Sinclair, S. J., Bruce, M. J., Griffioen, P., Dodd, A. and White, M. D. (2018). A condition metric for Eucalyptus woodland derived from expert evaluations. *Conservation Biology* **32,** 195–204.

Sinclair, S.J., Avirmed, O., Batpurev, K., Griffioen, P.A., White, M.D. and Olson, K. (2020). Condition metrics for Elm Forest and Saxaul ecosystems in the Gobi Desert. *Arthur Rylah Institute for Environmental Research Technical Report Series* 310. Department of Environment, Land, Water and Planning, Heidelberg, Victoria.

Sinclair SJ, Avirmed O, White MD, Batpurev K, Griffioen PA, Liu C, Jambal S, Sime H and Olson KA. 2021. Rangeland condition assessment in the Gobi Desert: A quantitative approach that places stakeholder evaluations front and Centre. *Ecological Economics* **181,** article 106891.

Tachiiri, K. and M. Shinoda (2012). "Quantitative risk assessment for future meteorological disasters." Climatic Change **113**, 867–882.

Therneau T, Atkinson B (2022). rpart: Recursive Partitioning and Regression Trees. R package version 4.1.19. https://CRAN.R-project.org/package=rpart.

Tserendash, S., & Erdenebaatar, B. (1993). Performance and management of natural pasture in Mongolia. *Nomadic Peoples* **33**, 9–15.

Tuvshintogtokh, I., Ariungerel, D. (2013). Degradation of Mongolian Grassland Vegetation Under Overgrazing by Livestock and Its Recovery by Protection from Livestock Grazing. In: Yamamura, N., Fujita, N., Maekawa, A. (eds) The Mongolian Ecosystem Network. Ecological Research Monographs. Springer, Tokyo.

Urbanek, S. (2014). jpeg: Read and write JPEG images. R package version 0.1-8. (https://CRAN.R-project.org/package=jpeg).

Vandergeest P and Unno A. (2012). A new extraterritoriality? Aquaculture certification, sovereignty, and empire. *Political Geography* **31,** 358–367.

Venables, A. and Boon, P.I. (2016). What environmental, social or economic factors identify high-value wetlands? Data-mining a wetlands database from south-eastern Australia. Pacific Conservation Biology 22, 312-337.

White, M.D., Hollings, T., Sinclair, S.J., Williams, K.J., Dickson, F., Brenton, P., Raisbeck-Brown, N., Warnick, A., Lyon, P., Mokany, K. and Liu, C. (2023). Towards a continent-wide ecological site condition database using calibrated expert evaluations. *Ecological Applications* **33,** p.e2729.

Appendix 1: Definitions of variables

The terms used to define the site variables are explained below. It is important to note that these groups are not mutually exclusive (some species belong to multiple groups), and that some groups are nested within others.

Cover Projective foliage cover, i.e., the shadow cast by the species (including all leaves, branches, trunk, etc., but not double counting overlapping cover).

Density Density refers to the number of the items per 900 m^2 plot.

Richness The count of species within the 900 m^2 plot.

Annual Any species which obligately completes its life cycle in a single year. A common annual forb is *Corispermum mongolicus*. A common annual grass is *Eragrostis minor*.

Perennial (forbs / grasses and sedges) Any species which is not annual. This group includes biennials and species which may be facultatively annual under harsh conditions.

Shrub A dicotyledonous plant (of any family) which forms perennial, above-ground woody stems. Such stems have secondary thickening and can be snapped. Common examples include *Artemisia xerophytica*, *Caragana leucophylla*, *Haloxylon ammodendron*, *Kallidium gracile, Nitraria sibirica, Oxytropis aciphylla* and *Reaumuria soongorica.*

Forbs Any species of angiosperm (monocot or dicot) that is not a shrub, and not a member of the Poaceae or Cyperaceae. Common examples include *Asparagus gobicus*, *Corispermum mongolicus*, *Ptilotrichum canescens*, *Rheum nanum* and *Scorzonera divaricata*. This group also includes sub-shrubs (or semi-shrubs) such as *Anabasis brevifolia*, *Peganum nigellastrum* and *Salsola collina*, and onions and their relatives (*Allium* spp.).

Grass / sedge Any species in the families Poaceae (grasses) or Cyperaceae (sedges). Common examples include *Achnatherum splendens*, *Aristida heymannii*, *Carex duriuscula*, *Carex pediformis*, *Cleistogenes soongorica*, *Cleistogenes squarrosa*, *Stipa glareosa* and *Stipa gobica*.

Coniferous tree Any species of conifer capable of growing to 4m or more in height. Most often *Larix sibirica*, *Larix gmelinii* and *Pinus sibirica.* May be evergreen or deciduous. Includes all examples of these species, including their young saplings.

Broadleaved tree Any non-coniferous species that can grow to 4 m or more in height. Deciduous. Usually *Betula* species, occasionally others such as *Alnus* and *Cotoneaster*. Includes all life stages of these species, including their young saplings.

Berry plant Any species with juicy fruits. Often Rosaceae or Ericaceae. May be shrubs, subshrubs, scramblers, brambles or forbs.

Succulent species Any species of dicot (shrub of forb) which has thickened, fleshy foliage that is ‘juicy’. Examples include several extremely common species such as *Anabasis brevifolia* and *Haloxylon ammodendron*.

Fabaceous shrub Any shrub in the family Fabaceae (Peas). Prominent genera include *Caragana* and *Oxytropis*.

***Artemisia* species** Any species in the genus *Artemisia*, whether a shrub (such as *A. xerophytica*) or a forb (such as *A. frigida*).

Litter Any plant material that is detached from the plant on which it formed, such as discarded leaves and twigs, and branches etc. with a diameter < 10 cm.

Dead wood Logs, stumps or standing dead trees, where the timber is > 10 cm in diameter.

Exposed roots/pedestals Roots which formed below ground but have been exposed by the erosion of soil. The height is measured vertically, from the root–trunk boundary to the point at which the lowest root is concealed by soil. The variable measures the highest example that can be found in the plot (not the mean).

www.deeca.vic.gov.au

www.ari.vic.gov.au