Ecosystems (2019) 22: 1902–1917
https://doi.org/10.1007/s10021-019-00380-y

ECOSYSTEMS

A Continental-Scale Validation of Ecosystem Service Models

Simon Willcock,[1,2]* Danny A. P. Hooftman,[3,4] Stefano Balbi,[5] Ryan Blanchard,[6] Terence P. Dawson,[7] Patrick J. O'Farrell,[6,8] Thomas Hickler,[9,10] Malcolm D. Hudson,[11] Mats Lindeskog,[12] Javier Martinez-Lopez,[5] Mark Mulligan,[7] Belinda Reyers,[13] Charlie Shackleton,[14] Nadia Sitas,[6] Ferdinando Villa,[5,15] Sophie M. Watts,[2] Felix Eigenbrod,[2,16] and James M. Bullock[4]

[1]*School of Natural Sciences, Bangor University, Bangor, UK;* [2]*Biological Sciences, University of Southampton, Southampton, UK;* [3]*Lactuca: Environmental Data Analyses and Modelling, Diemen, The Netherlands;* [4]*Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK;* [5]*Basque Centre of Climate Change, Leioa, Spain;* [6]*Council for Scientific and Industrial Research, Pretoria, South Africa;* [7]*Department of Geography, King's College London, London, UK;* [8]*Percy FitzPatrick Institute of African Ornithology, University of Cape Town, Cape Town, South Africa;* [9]*Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt, Germany;* [10]*Department of Physical Geography, Goethe University, Frankfurt, Germany;* [11]*Engineering and the Environment, University of Southampton, Southampton, UK;* [12]*Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden;* [13]*Department of Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa;* [14]*Department of Environmental Science, Rhodes University, Makhanda, South Africa;* [15]*IKERBASQUE, Basque Foundation for Science, Bilbao, Bizkaia, Spain;* [16]*Geography and Environment, University of Southampton, Southampton, UK*

Abstract

Faced with environmental degradation, governments worldwide are developing policies to safeguard ecosystem services (ES). Many ES models exist to support these policies, but they are generally poorly validated, especially at large scales, which undermines their credibility. To address this gap, we describe a study of multiple models of five ES, which we validate at an unprecedented scale against 1675 data points across sub-Saharan Africa. We find that potential ES (biophysical supply of carbon and water) are reasonably well predicted by the existing models. These potential ES models can also be used as inputs to new models for realised ES (use of charcoal, firewood, grazing resources and water), by adding information on human population density. We find that increasing model complexity can improve estimates of both potential and realised ES, suggesting that developing more detailed models of ES will be beneficial. Furthermore, in 85% of cases, human population density alone was as good or a better predictor of realised ES than ES models, suggesting that it is demand, rather than supply that is predominantly determining current patterns of ES use. Our study demonstrates the feasibility of ES model validation, even in data-deficient locations such as sub-Saharan Africa. Our work also shows the clear need for more work on

Received 3 February 2019; accepted 22 March 2019;
published online 22 April 2019

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s10021-019-00380-y) contains supplementary material, which is available to authorized users.

Authors Contribution FE, SW, DAPH, TPD, POF, MDH, BR, CS, FV and JMB conceived the project. SW, DAPH, RB, TPD, POF, MDH, MM, BR, CS, NS, FV, FE and JMB developed the investigation at a workshop on model complexity. SW and SMW collated and prepared the validation dataset. SW and MM ran the Co$ting Nature and WaterWorld models. DAPH ran the InVEST models and Scholes international and water supply models. SW, SB and JML ran LPJ-GUESS, ML provided the LPJ-GUESS code and, with TH, provided advice on the use of LPJ-GUESS. RB, POF and BR ran the local South African grazing and firewood models. DAPH, JMB, SW and FE analysed the results. SW, DAPH, FE and JMB wrote the manuscript, with comments and revisions from all other authors. DAPH collated the SI, with comments and revisions from all other authors.

Simon Willcock and Danny A. P. Hooftman are joint first authors and contributed equally to this work.
Corresponding author; e-mail: s.willcock@bangor.ac.uk

the demand side of ES models, and the importance of model validation in providing a stronger base to support policies which seek to achieve sustainable development in support of human well-being.

Key words: Africa; beneficiary; carbon; charcoal; complexity; firewood; grazing; natural capital; water.

Highlights

- We validate multiple ecosystem services (ES) models across sub-Saharan Africa (SSA)
- We find that more complex ES models sometimes provide more accurate estimates
- Realised use of ES is closely aligned with human population density (demand) in SSA

Introduction

Ecosystem services (ES)—nature's contributions to people (Pascual and others 2017)—are of global importance to human well-being, but are increasingly threatened by human activities (Steffen and others 2015). As a result, many governments are now moving to ES-based management of natural resources (Wong and others 2014) and 132 United Nation member states have signed up to the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services (IPBES; www.ipbes.net). This shift in policy requires accurate spatial modelling of ES (Malinga and others 2015), as managing ES requires an understanding of their spatial distribution and heterogeneity (Swetnam and others 2011; Spake and others 2017) and the ability to project and compare the outcomes of management scenarios (Willcock and others 2016). Models can provide credible information where empirical ES data are sparse, which is especially the case in many developing countries (Suich and others 2015).

To meet demand for an enhanced understanding of ES flows, many spatial modelling methods and tools for mapping ES have been developed, ranging from very simple land cover-based proxies to sophisticated process-based models (IPBES 2016). Whilst a growing literature is comparing the outputs and features of the different tools (Bagstad and others 2013; Turner and others 2016), validation of these models is challenging and thus rare in the literature (Bennett and others 2013). Few studies have validated single ES models against independent datasets and then only rarely at a larger, country scales (Mulligan and Burke 2005; Bruijnzeel and others 2011; Redhead and others 2016, 2018). Even more rare are studies that explicitly validate multiple ES models simultaneously, and these generally involve small areas at catchment scale (Sharps and others 2017). As a consequence, the uncertainties associated with most ES models and the datasets that underpin them remain largely unknown (Bryant and others 2018; van Soesbergen and Mulligan 2018). This is a particular issue as the results of local-scale validation are likely not to be transferable to new locations (Redhead and others 2016) or to the regional and national scales at which ES model outputs are most widely used (Willcock and others 2016). As a result, attempts at validation by those applying models in new settings are all the more important (Bryant and others 2018). Indeed, rescaling social–ecological patterns and processes to different spatial resolutions and extents can induce substantial systematic bias (Grêt-Regamey and others 2014), providing challenges to decision-making in situations where model results are the only source of information. Lack of proven credibility, salience and legitimacy are the major reasons for the 'implementation gap' between all ES research (not just ES models) and its incorporation into policy- and decision-making (Cash and others 2003; Voinov and others 2014; Wong and others 2014; Clark and others 2016).

Approaches to improve the reliability of model predictions in general include increasing model complexity [defined here as model structural complexity (Kolmogorov 1998), sometimes also referred to as model complicatedness (Sun and others 2016)]. Computational capacity has rapidly increased over time, enabling ES models to become more complex and multiple models to be run at higher resolutions across larger spatial ranges (Levin and others 2013). However, increasing the complexity of ecological models typically also increases the amount of data and expertise required for implementation and interpretation, with unclear consequences for the results (Merow and others 2014). In short, it is unclear whether an investment in increasing model complexity leads to more accurate information for policy- and decision-making on local and regional scales.

The unknown credibility of ES models (Voinov and others 2014) is most pronounced where they are arguably most needed—in many developing countries, where data collection and model development efforts are least advanced (Suich and others 2015). Such ES information is important

because the rural and urban poor are often the most dependent on ES (either directly or indirectly (Cumming and others 2014)), both for their livelihoods (Daw and others 2011; Suich and others 2015) and as a coping strategy for buffering shocks (Shackleton and Shackleton 2012). A major barrier to the understanding and management of these benefit flows to the poor is a lack of information on the potential supply and realised use of ES, particularly in the developing world (Wong and others 2014; Willcock and others 2016; Cruz-Garcia and others 2017). Indeed, the comparisons of ES models to primary data that do exist are all focused on potential and not realised ES (that is, biophysical supplies only and not actual use by beneficiaries) (Bagstad and others 2014). Analyses need to be disaggregated to focus on how people use ES, from which ecosystems, and how such benefits contribute to the people's well-being (Daw and others 2011; Bagstad and others 2014; Cruz-Garcia and others 2017).

In this paper, we validate ES models against measured ES data extending over 36 countries in sub-Saharan Africa, covering 16.7 million km^2—over half of the land area of Africa—and including some of the world's poorest regions (Handley and others 2009). We focus on five ES of high policy relevance in sub-Saharan Africa (Willcock and others 2016), and for which validation data exist in multiple locations. The potential supply of two ES (stored carbon and available water) is modelled using the existing models and a further three ES (firewood, charcoal and grazing resources) predominantly using new models generated from stored carbon outputs of the existing models. To assess ES use, we developed new standardised models for realised ES (that is, actual use by people) by weighting models of potential ES (biophysical supply) by human population density for the four measured ES where the location of beneficiaries is important (use of charcoal, firewood, grazing resources and water). We hypothesised that these new realised ES models have higher predictive power than potential ES models for these ES. We also assessed the performance of human population density alone as a predictor of ES use, as this represents the simplest possible globally available ES use model. Our rationale for doing so is that local population density is a straightforward indicator of the number of people making use of the ES, and such a simple approach for modelling realised services would be very useful if it proved to be accurate. We do not focus on comparing specific modelling platforms, as the identification of the best specific model for a particular use may shift as new models are developed and is likely be location specific: such site-specific comparisons have been done elsewhere (Bagstad and others 2013; Ochoa and Urbina-Cardona 2017). As such, our aims in this study are twofold: (1) to compare the general performance of models predicting ES supply (for stored carbon and available water) to realised ES (charcoal, firewood, grazing and water use); and (2) whether more complex ES models make better predictions.

Methods

Our approach to modelling and validation is summarised in Figure 1. We validated the existing and new—developed using outputs from the existing models (see below)—ES models against ES data, using 1675 data points from 16 independent datasets extending over sub-Saharan Africa (carbon: 214, water: 736, firewood: 285, charcoal: 59, grazing: 401; Table 1, Figure 2). We compared approaches for modelling ES ranging in complexity from simple land cover-driven production functions to process-based models (IPBES 2016). As our validation datasets vary in spatial extent and location, we accounted for the effects of spatial extent and context (Figure 1). We tested the hypotheses that ES models incorporating a more complex causal structure have higher predictive power. Since decision-makers in sub-Saharan Africa consider model complexity to mean more inputs being used to model more processes (Willcock and others 2016), we assessed model complexity in terms of the number of input variables, defining inputs as a coherent set of values covering the research area for a single feature, be it categorical or numerical (Merow and others 2014).

Description of Ecosystem Service Models

We selected ES models to test, focussing on: (1) ES models capable of estimating some of our selected potential ES (stored carbon, available water) and providing inputs to our new models of firewood, charcoal and grazing resources within our study area; (2) the subset of these models for which adequate validation data could be identified, allowing like-for-like comparisons between modelled outputs and validation data; and (3) models representing a range of complexities from simple production functions to process-based models. As such, we used six existing ES modelling frameworks that contain one or more models meeting these criteria (Table SI-1-1, SI-1-1). We selected InVEST (Kareiva 2011; McKenzie and others

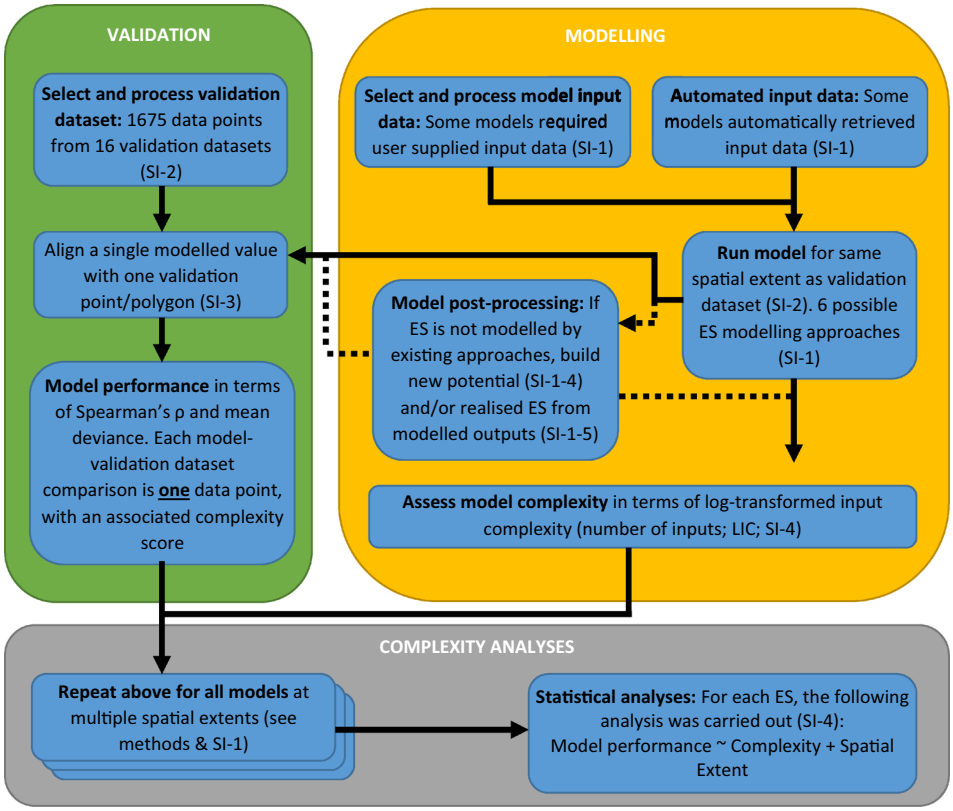

Figure 1. A summary of the analytical framework, divided into validation, modelling and analysis subsets.

2012), Co$ting Nature (Mulligan and others 2010; Mulligan 2015), WaterWorld (Mulligan 2013) and benefits transfer (based on coupling the Costanza and others (2014) values with GlobCover 2009 landcover categories; SI-1-2) due to their widespread use and global applicability (Bagstad and others 2013). We also included the well-known and partially validated (Pachzelt and others 2015) dynamic vegetation model LPJ-GUESS (Smith and others 2001, 2014). Although LPJ-GUESS is not traditionally considered an ES model and has a relatively coarse native resolution (0.5 × 0.5 degrees, but constrained mainly by the resolution of environmental input variables), it is increasingly used for ES modelling applications [including implementation within the ARIES technology (Villa and others 2014)] and it is a process-based model that gives outputs that effectively track the biophysical supply of many potential ES (Bagstad and others 2014; Lee and Lautenbach 2016). Furthermore, we included the Scholes models (comprising two grazing models and a rainfall surplus model) as it is the only large-scale ES models designed specifically for use in sub-Saharan Africa (Scholes 1998) (SI-1-3). Ideally, we would also have compared bespoke local models with local data. However, such models simply do not exist in sub-Saharan Africa in most places. Moreover, as the global models we compare run at fine spatial resolutions (except LPJ-GUESS), it is reasonable to investigate how well they perform in terms of accuracy against local data collected in many locations in many different ways (as is the case here).

At time of analysis (March 2017), InVEST, Co$ting Nature and LPJ-Guess did not have models that focus on firewood, charcoal or grazing resources, but they did explicitly output stored vegetation carbon. As the supply of these three ES is directly dependent on the amount of biomass present, which is what underpins estimates of stored vegetation carbon in all three models, we built eight new predictors using the outputs from these three existing carbon modules (to estimate the potential supply of these three additional ES (SI-1-4). These new models used spatial masks to estimate the biomass available on relevant land uses (SI-1). For example, we applied a 'grazing' spatial mask to derive grassland carbon from InVEST and Co$ting Nature standing carbon outputs. We excluded areas in which little to no grazing activity was expected (for example, protected areas) and

Table 1. Datasets Used to Validate the Ecosystem Service Models Included in This Study

	Source	Type	Period	Selected # data points	Unit of validation	Type of area	Service	Link to data source
1	Global Runoff Data Centre (GRDC)	Flow through Weirs	Running period of Weir	512	m^3 $year^{-1}$*	Catchments¶	Water supply	www.bafg.de/GRDC¶¶
2	South Africa, Department of Water and Sanitation (DWS)			188	m^3 $year^{-1}$*	Catchments¶	Water supply	www.dwa.gov.za/Hydrology
3	Aquastat	Water use per capita	1999–2011†	36	m^3 $year^{-1}$*	Countries	Water use	www.fao.org
4	Stanford Library	Ethiopia livestock Census	1999	70	LSU's*	Zones	Grazing	library.stanford.edu
5	Kenyan Government database	Kenya livestock Census	2009	46	LSU's*	Counties	Grazing	www.opendata.go.ke/
6	South-African Statistics department (Hamann and others 2015, 2016)	South-Africa per household census data	2011	234	Usage*	Municipalities	Grazing	Hamann and others (2015)
7				234	Usage*	Municipalities	Firewood use	
8	FAO databases	Animal stocks	2012	36	LSU's*	Countries	Grazing	www.fao.org/faostat
9		Carbon stock in living biomass in forest areas	2010	36	Tons above-ground Carbon	Countries	Stored woody carbon	www.fao.org/faostat
10		Total usage per year	2013	36	Tons biomass	Countries	Charcoal use	www.fao.org/faostat
11		Total usage per year	2013	36	m^3 wood fuel	Countries	Firewood use	www.fao.org/faostat
12	Carbon stocks via Laporte and others (2008)	Carbon stock in DRC in forest areas	2007	31	Tons above-ground Carbon	Districts	Stored woody carbon	whrc.org
13	ForestPlots.Net (Willcock and others 2014; Avitabile and others 2016)	Above-ground biomass using Chave and others (2005) approximation	Last available	147	Tons above-ground Carbon	Plots	Stored woody carbon	www.forestplots.net¶¶
14	CIFOR, Poverty Environment Network (PEN)	Fodder usage	2004–2008	23	Kg biomass $year^{-1}$‡	Villages‡	Grazing	www.cifor.org/pen¶¶
15		Charcoal production	2004–2008	23	Kg biomass $year^{-1}$‡	Villages‡	Charcoal use	www.cifor.org/pen¶¶
16		Firewood production	2004–2008	23	m^3 $year^{-1}$‡	Villages‡	Firewood use	www.cifor.org/pen¶¶

**Unit converted from original source, see SI-2; ‡ quarterly data summed into years and all families summed per village, ¶Catchments bespoke DEM-delineated to weir points. †Latest values as was available per country in the database in July 2015. ¶¶Under permission of use for this project only. Each separate model-validation dataset comparison is a single independent data point within our analysis.*

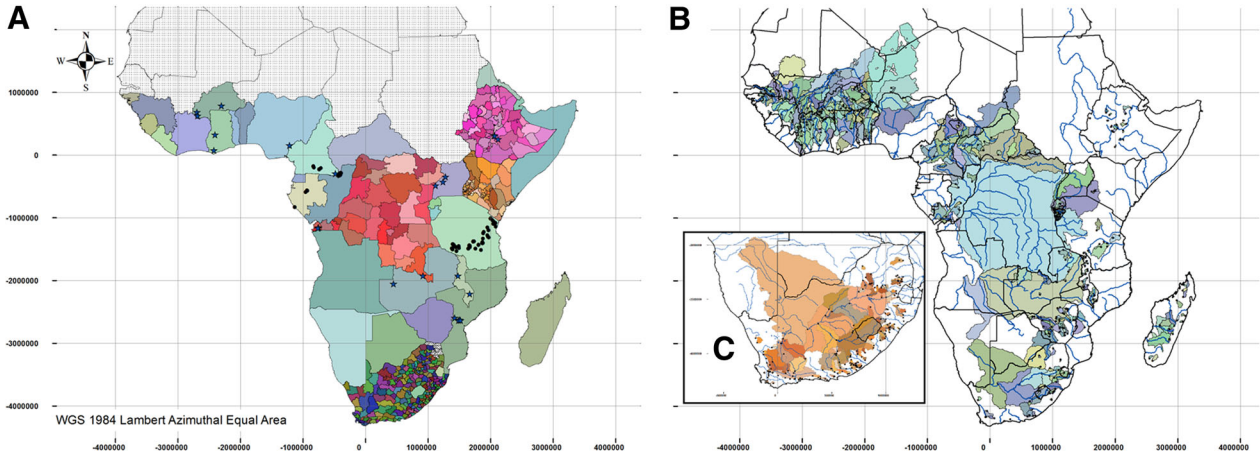

Figure 2. Locations at which validation datasets were gathered (SI-2). **A** Coloured countries show our study area and our validation data at the country scale; dots represent standing carbon plots; stars represent PEN sites used for charcoal, firewood and grazing; districts in the Democratic Republic of the Congo are used for standing carbon; counties in Ethiopia and Kenya for grazing; and municipalities in South Africa for charcoal, firewood and grazing. **B** Catchments used through the Global Runoff Data Centre managed weir dataset. **C** Catchments through the South African weir data managed by the Department for Water and Sanitation. Colours in all figures are present only to allow distinction among different units within datasets.

included areas in which most of the above-ground carbon is assumed to be available for livestock grazing (Figure SI-1-1A; Table SI-1-5). For LPJ-Guess, we used C3/C4 carbon outputs as estimate for grazing resources. Thereafter, we converted grazing biomass to FAO livestock units for sub-Saharan Africa using the conversion factors from Houerou and Hoste (1977). Henceforth, we refer to these carbon-based predictors as ES models. Finally, we created models of realised ES by weighting models of potential ES (models of biophysical supply only; for example, the Scholes models, WaterWorld and our new models of firewood, charcoal and grazing resources) by human population (Stevens and others 2015). We also conducted like-for-like comparisons of these new models for realised use of water, firewood, charcoal and grazing resources with relative rural population data alone—the simplest possible model of ES use. We also assessed whether these new realised ES models have higher predictive power than potential ES models when compared to ES use data. We excluded urban populations for all analyses except the Poverty Environment Network usage data and water use (Table 1).

Validation Datasets

As we considered the performance of ES models separately for each ES, we did not require locations that provided primary data on all ES together. This enabled us to access 1675 data points from 16 separate validation datasets—the maximum number available to us that were suitable for the purposes of this study (that is, independent of the model calibration data; Figure 2, Table 1, SI-2). These data are diverse, being collected using a range of methods of varying reliability, including expert opinion (for example, country-level statistics from the FAO), census data (for example, district level for Kenya and Ethiopia, household level for South Africa) and biophysical measurement (for example, tree inventory plots, and weir data on water flow [both from across sub-Saharan Africa]) (Table 1). As such, each dataset has associated uncertainties (Grainger 2008) but, because the 'true value' can never be absolutely determined, provides acceptable reference values for validation. Given that the datasets cover a wide range of independent methods and our focus is on ranked correlative relationships between models and data, there is unlikely to be systematic bias and so data quality issues should impact our results minimally. In our analyses, some of the validation data required processing to ensure like-for-like comparison with modelled outputs. All ES models were either run at 1 × 1 km or resampled from their minimum native resolution to an exact 1 × 1 km resolution (that is, for the Scholes Firewood model [native resolution: 5 × 5 km] and for LPJ-Guess [native resolution: 55.6 × 55.6 km]). We then extracted a single summary value per polygon to align model outputs with polygon validation data (for example, each catchment for the Global Runoff

Data Centre [GRDC]; each district for Kenya; each country for FAO data; see SI-2). For forest plot point validation data (the ForestPlots.net data), we compared the point data to the 1 × 1 km grid cell it was in. For the PEN data (fodder, charcoal and firewood use), we buffered the point estimate of each village location by 10 km (to align with walking distances for firewood or water (Agarwal 1983; Sewell and others 2016)) and calculated the summary value for each model for each polygon. Hence, we extracted model data to be as comparable as possible to the validation data points. This means that single values as similar in area and units as possible were extracted from each model to be compared to the single validation values as provided by the datasets listed in Table 1 (see SI-2 for full details of these methods). All data were normalised following Verhagen and others (2017) to equalise any unit differences (SI-3-3).

Statistical Analyses

Calculation of Model Performance

There is no single comprehensive measure of model performance (Bennett and others 2013). Criteria commonly considered are: (1) trueness—the closeness of the agreement between the reference value and the average model value, largely affected by systematic error or bias within the model); (2) precision—the closeness of agreement between repeated model runs, largely affected by random variables or distributions that feature within the model code; and (3) accuracy—an overall summary of precision and trueness that describes the closeness of the agreement between the reference value and the values obtained from the model run(s) (IOS 1994). We focussed on accuracy and trueness here, as we only considered a single output dataset from each model (derived from a single set of parameters) and assessed these using two metrics. The first metric was the rank correlation between modelled and validation values (Spearman's ρ)—a measure of accuracy ranging from − 1 to 1, with 1 indicating a perfect positive correlation, 0 no correlation and − 1 a perfect negative correlation. Thus, ρ is a useful metric as in many cases policy-makers want to rank locations by their relative ES values (Willcock and others 2016). The second metric was the average absolute deviance of modelled values from the 1-to-1 line representing a perfect fit of normalised model values to the normalised validation values—a measure of accuracy and trueness, as it reflects the degree to which models consistently reflect validation values (SI-2). In our normalised setting (with values inverted for consistency with ρ), deviance ranged from 0 (poor fit) to 1 (perfect fit). For interpretation, we follow the generally used criteria employed in AUC (area under the curve) in which a result below 0.7 should be considered as likely random (Swets and others 1979; Marmion and others 2009; Hooftman and others 2016) and a value of at least 0.7 shows a close fit between the modelled value and the validation data. It is entirely possible for a model to have a high rank correlation value, but also have high deviance from the 1–1 line and vice versa, so the two metrics are complementary (Table 2, Figure 3). We calculated both metrics separately for every ES model for each relevant validation dataset, with the ES models run at 1 × 1 km spatial resolution in most instances, giving 100 comparisons (carbon: 12; water supply: 21; water use: 18; charcoal use: 8; firewood use: 15; grazing use: 26 Figure 1).

The Effect of Model Complexity, Spatial Extent and Adding Beneficiaries on ES Model Performance

For each ES, we assessed model complexity in terms of the number of input variables (input complexity [IC]). We considered an input to include a coherent set of values covering a geographic region for a single feature (for example, land use or elevation). GIS processing without changing the feature parameter was not considered an additional input, and neither were layers created by combining inputs, although the parameter values of an equation could be independent, single-value datasets (see SI-4 for full details). Thus, our new models (developed via GIS processing of the outputs from the existing models of ES potential) retained the complexity score of the associated existing model (Figure 1; SI-4). As such, our complexity metric captures the generalisation that models with large numbers of equations generally require more inputs (Sun and others 2016). From a user experience perspective, this complexity often relates to the sourcing and processing of these required input datasets (Willcock and others 2016). Our continuous complexity metric is more subtle and precise than simple categorisation of models, for example, process-based vs production function. Therefore, it allowed us to advance from previous model comparisons, which often focus on identification of the best model specific to a location, by identifying generalisable conclusions relating model complexity to model accuracy. We log-transformed the IC value (LIC) in all instances as the data were skewed by extreme values.

Importantly, we considered each separate model *vs* validation dataset comparison a *single* independent data point (for example, InVEST stored carbon

Table 2. Comparison of Individual Ecosystem Service Model Performance (Rank Correlation [ρ] or Mean Deviance) with That of Human Population Density

Ecosystem service	Response variable	Least squares mean value for the best-fitting individual model	Is human population density a significantly better predictor than all individual models?
Stored carbon (biophysical supply)*	ρ	0.677	n/a
	Deviance	0.790	n/a
Water (biophysical supply)*	ρ	0.695	n/a
	Deviance	0.855	n/a
Water (use)†	ρ	0.137	Worse$^{0.0341}$
	Deviance	0.800	Worse$^{0.0014}$
Charcoal (use)†	ρ	0.185	Equal
	Deviance	0.660	Equal
Firewood (use)†	ρ	0.407	Equal
	Deviance	0.731	Equal
Grazing (use)†	ρ	0.463	Equal
	Deviance	0.728	Equal

*Modelled using the existing approaches (SI-1-1).
†Modelled using our new approaches (SI-1-4, SI-1-5).
For both measures, a value closer to 1 indicates better fit. n/a not applicable. p is given in superscript for significant effects (otherwise $p > 0.05$).

validated against carbon storage per unit area derived from tree inventory plots was a single data point) (Figure 1). This is because we were interested in assessing how well models performed in general in sub-Saharan Africa for different types of validation data collected in different locations. This approach also enabled us to use very different types of validation datasets, thereby overcoming the issue of there not being consistent validation data for all ES across most of sub-Saharan Africa.

By considering each model vs validation dataset comparison (in terms of rank correlation or deviance) a single data point, we were able to build separate generalised linear models (GLMs) for each of the two model performance measures (the response variable y; rank correlation or deviance), and for each ES. In each case, the GLM was: $y \sim Complexity\ Measure + Spatial\ Extent$. Thus, LIC was chosen as the complexity metric, with spatial extent (local, regional, country) modelled as a fixed factor. This allowed us to test if more complex models better predict the biophysical supply or realised use of each of ES, whilst controlling for any effects of spatial extent. Where the validation data were of realised ES, we compared models of potential ES, ES demand and realised ES.

Results

Model Validation

In general, at least one model for each ES produced outputs that represented the validation data well, calculated in terms of their deviance measure and Spearman's ρ, with deviance showing better fits (mean of the least squares mean [LSM] values for best-fit model: $\rho = 0.43$, deviance = 0.76; Table 2; Figure 3; SI-2). *Potential ES:* The LSM value of the response variable for the best-fit model showed that the best of the existing models of potential ES (carbon and water supply) matched the validation data well (mean LSM value for best-fit potential ES models: $\rho = 0.69$, deviance = 0.82; Table 2; Figure 3A). *Realised ES:* Whilst still producing reasonable fit to validation data, the new models of realised ES did not show a fit as good as the models of potential ES to their respective validation data (mean LSM value for best-fit realised ES models: $\rho = 0.30$, deviance = 0.73; Table 2; Figure 3B). When compared to realised ES data (Table 3), some (3 of 8 [38%]) of the simple models of realised ES performed better than models of ES potential, and none performed worse. However, for our models of realised charcoal, firewood and grazing services, a majority (45 of 47 [96%]) were predicted as well by human population density alone as by our models, and in two cases (4%) population density was a better predictor than our models (p values < 0.05; Table 2). By contrast, the comparison of realised water with the water use data showed population density to be a worse predictor than our new realised ES models (6 of 6 [100%]; Table 2).

Model Complexity

Our comparisons showed either no (1 of 4 [25%] potential ES; 6 of 8 [75%] realised ES) or a positive

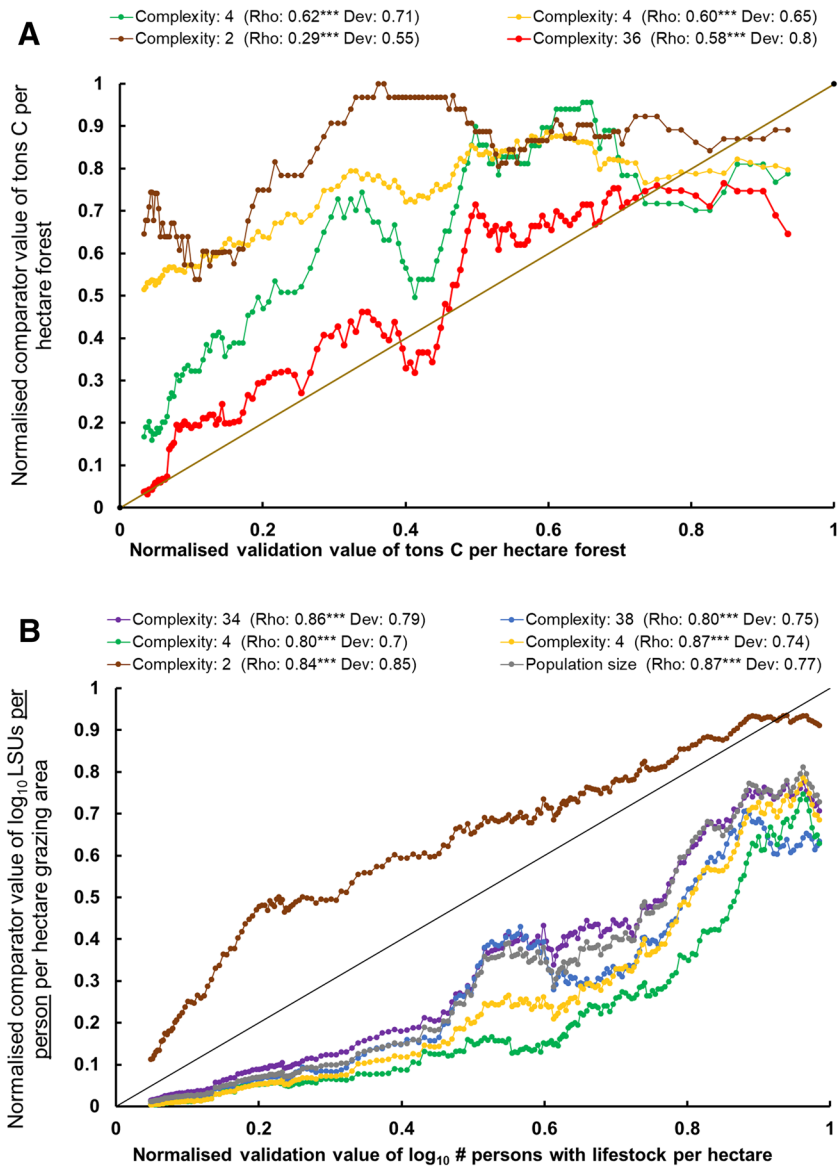

Figure 3. Examples of ecosystem service model validation for **A** potential biophysical carbon supply and **B** realised grazing use. X-axis is (**A**) tons carbon per hectare forest in ForestPlot.net (Willcock and others 2014; Avitabile and others 2016) and **B** the validation set of South African data (Hamann and others 2015), being the normalised $\log_{10}$ number of people with livestock per hectare. *Y* axis is the normalised modelled value. Different lines are different models, characterised by their complexity score. The lines are added to the graphs for visual clarity only, to allow the reader to see trends; we smoothed the lines with a 10% running average.

(3 of 4 [75%] potential ES; 2 of 8 [25%] realised ES) effect of complexity on model fit, with no cases of a negative effect (Table 3). Responses to model complexity were not consistent among the two model accuracy metrics (ρ and deviance), reflecting their different properties. Notably, we found positive effects across both metrics for stored carbon, but complexity was more rarely a significant predictor of model fit for firewood use, charcoal use and water availability, and in these cases was only detected for one of the two accuracy metrics. Grazing use and water use showed no effect of complexity for either metric (Table 3).

Discussion

This study—the first multi-country validation of multiple ES models (to the best of our knowledge)—suggests that the existing ES models provide good predictions across two potential ES of high policy relevance (Willcock and others 2016). But, for the ES models we investigated, models of

Table 3. The Effects of Variables on Ecosystem Service Model Performance, Derived From Generalised Linear Models as Follows: Model Performance (Rank Correlation [ρ] or Mean Deviance) ~ Complexity Measure + Spatial Extent

Ecosystem service	Response variable	Complexity (LIC)		Spatial extent	Realised versus potential ES model performance
		Direction of effect	Coefficient		
Stored carbon (biophysical supply)*	ρ	$+^{0.0055}$	0.2089	0	n/a
	Deviance	$+^{0.0030}$	0.0979	0	n/a
Water (biophysical supply)*	ρ	$+^{0.0332}$	0.2010	n/a	n/a
	Deviance	0	n/a	n/a	n/a
Water (use)†	ρ	0	n/a	0	0
	Deviance	0	n/a	0	0
Charcoal (use)†	ρ	0	n/a	0	0
	Deviance	$+^{0.0139}$	0.0732	$L > C^{<0.0001}$	$V2 > V1^{<0.0001}$
Firewood (use)†	ρ	0	n/a	$R > L,C^{0.0389}$	0
	Deviance	$+^{0.0508}$	0.0457	$R,L > C^{0.00331}$	$V2 > V1^{<0.0001}$
Grazing (use)†	ρ	0	n/a	$C,R > L^{<0.0001}$	0
	Deviance	0	n/a	$C,R > L^{<0.0001}$	$V2 > V1^{0.0243}$

Log-transformed input complexity is the measure of complexity, and spatial extent (local [L], regional [R], country [C]) are fixed effects. Additionally, whether potential ES (V1) or realised ES (V2) models types performed better is indicated for ES use. 0 indicates no significant effect ($p > 0.05$) and n/a not applicable. Significant effects are shown with p in superscript and the direction of effect as ± for continuous variables (that is, complexity) or in terms of differences among factor levels (determined using the lsmeans statement in SAS 9.3).
**Modelled using the existing approaches (SI-1-1)*
†Modelled using our new approaches (SI-1-4, SI-1-5)

potential ES (biophysical supply) were more accurate than our new models of realised ES (use; Table 2). This difference can be explained by the facts that: a) building models for realised ES is more challenging; and b) there is a research bias towards the biophysical supply of a few provisioning and regulating services—food supply, water availability and stored carbon (Egoh and others 2012; Martínez-Harms and Balvanera 2012; Wong and others 2014).

The Importance of Social Systems

Decision makers require information on a wide range of ES and across a variety of temporal and spatial scales (Scholes and others 2013; McKenzie and others 2014; Willcock and others 2016). Meeting these needs will require a shift in the focus of most models towards understanding the beneficiaries of ES and quantifying their demand, access to and utilisation of services, as well as the consequences for well-being (Bagstad and others 2014; Poppy and others 2014). Whilst some studies (Hamann and others 2016) and models (Mulligan 2015; Suwarno and others 2018; Martínez-López and others 2019) do include the demand and use of ES, our new models of realised ES (created by weighting outputs of models of ES potential by human population) generally showed lower predictive power when compared with the ability of the existing models of ES potential to predict biophysical supply. Indeed, many of our new models were unable to predict ES more accurately than human population density alone (Table 2, Table 3). This suggests that rural human population density is a good proxy for ES demand, and realised use of ES is closely aligned with demand in sub-Saharan Africa. The only exception is water use, where our new models were better predictors of realised water use than human population density (Table 2). Further combining social science theory and data to explain the social–ecological processes of ES co-production, use and well-being consequences will likely result in substantial improvements in our understanding and estimates of ES use (Bagstad and others 2014; Díaz and others 2015; Suich and others 2015; Pascual and others 2017). This is an area of active research, and some modelling frameworks are beginning to address this deficiency. Socio-economic data on ES use, perceptions and well-being contributions collected over large regions can and has been incorporated into models to address questions about the impacts of ecosystem change on the well-being of regional and socio-economic groups (Díaz and others 2015; Hamann and others 2016; Egarter Vigl and others 2017). Spatial multi-criteria analyses can be used to model how consistent available potential ES are

with local demand, highlighting trade-offs between beneficiary groups where demand varies (Martínez-López and others 2019). Furthermore, the impact of individual decision-making on ES use can be captured through the integration of agent-based models depicting human behaviour with biophysical models (Villa and others 2017; Suwarno and others 2018)). Coupling models of potential ES with models of demand to estimate realised ES will likely result in models that are more complex than the existing models (Zhang and others 2017), which our findings suggest could improve accuracy, and new modelling techniques (for example, machine learning (Willcock and others 2018)) may be needed to enable this (Bryant and others 2018).

The Impact of Model Complexity

The effect of model complexity on the accuracy of ES results has not been investigated in detail previously. For example, a Web of Science search (20 June 2018) for 'model' and 'complexity' and 'accuracy' and 'ecosystem service' resulted in only 19 studies, few of which actually assess how ES model complexity affects accuracy. Our results suggested a tendency for ES model complexity to be correlated with increased performance (particularly for potential ES), and strong evidence that increased model complexity does not lead to worse predictions (for both potential and realised ES). However, each successive increase in complexity brings diminishing returns. For example, for each unit increase in LIC for models of stored carbon, ρ increased by 0.25 and deviance by 0.10 (Table 3). Since LIC is log-transformed, each unit increase is achieved by a tenfold increase in inputs. Furthermore, a trade-off with benefits of additional complexity may be the feasibility of running and interpreting such models (Willcock and others 2016). Results from elsewhere in the literature are mixed, often dependent on the specific context of the comparison. For example, Villarino and others (2014) compare simple 'Tier 1' carbon accounting methods with more complex 'Tier 2' methods, reporting increased accuracy with model complexity. However, studies that extend this analysis to the most complex 'Tier 3' methods report limits in the gains in accuracy, with intermediate 'Tier 2' and complex 'Tier 3' models producing similar predictions (Hill and others 2013; Willcock and others 2014). Furthermore, increasing model complexity does not necessarily lead to better model performance when predicting ground water recharge rates (von Freyberg and others 2015) nor agricultural yield (Quiroz and others 2017). Nevertheless, model performance should not be the only variable considered when selecting between models of differing complexity. From an ecological perspective, simple functional forms (for example, linear or nonlinear regression equations having a sufficiently high explanatory power) can be easier to interpret and translate into applications (that is, from science to policy). However, they may lack predictive power in novel locations and/or future time points if they insufficiently represent spatial heterogeneity in form and process (Syfert and others 2013). A certain level of complexity may be required before sufficiently reliable results can be obtained (Merow and others 2014; Salmina and others 2016), such as our observation that human population is poor predictor of water use, likely as it completely fails to capture the behaviour of the presence and flow of water. Simpler models may accurately represent more basic aspects of a system (for example, estimating natural capital), but incorporation of additional complexity may be necessary to describe the underlying processes accurately (for example, the interactions and feedbacks between people and ecosystems) (Merow and others 2014; Willcock and others 2014; Dunham and Grand 2016), and how different trade-offs and benefit flows can be understood and managed. Thus, model complexity should be considered in terms of how complex the ES being modelled are, what objectives need to be met, and to what end.

Limitations

Our analysis comes with several important caveats with respect to validation. Here, we highlight these in part to act as a 'call to arms' for ES scientists concerning areas demanding further development.

Primary data collection, particularly at large scales, should be a priority for ES scientists. As validation of modelled outputs must involve like-for-like comparisons (that is, comparing potential ES outputs to biophysical supply and realised ES outputs to observed ES use), we were unable to validate all models and we would have liked. For example, we were unable to include the models of realised ES produced by Co$ting Nature (Mulligan and others 2010; Mulligan 2015) due to the lack of corresponding validation data.

Another priority for future work is to link better different types of ES models to bespoke validation data to understand their performance fully. For instance, the different carbon models we used to some extent model different constructs. Co$ting Nature's stored carbon model includes both below- and above-ground carbon, whilst other models predict

only above-ground carbon (see SI-1), and we validated these with above-ground carbon data. As below-ground carbon can exceed that of above ground, these data are not ideal to validate the Co$ting Nature model. As below-ground carbon in forests is often consistent across forests or proportional to above-ground carbon storage (Lewis and others 2013), it is unlikely that this particular issue affected our findings. However, similar issues arise when linking the Costanza and others (2014) benefit transfer models with validation data, as the former estimate value but are validated against either biophysical or use data (SI-1). Because benefit transfer models are derived by combining global values with land cover data, one might expect the values to be more indicative of the biophysical supply of services, but be poorly matched to actual ES use. To reduce these issues and enable like-for-like comparisons, we generated new models (for example, for firewood, charcoal and grazing resources) from stored carbon outputs of the existing models (SI-1) in our analyses. These new models used spatial masks to estimate the biomass available on relevant land uses (SI-1). The outputs from these new models are likely to be overestimates as, for example, not all grassland vegetation will be grazed and not all grazed land will stock at maximum capacity (Fetzel and others 2017). However, since our statistical analyses focused on relative ranking (see methods), it is unlikely that these uncertainties impacted our findings greatly (that is, sites with the highest maximum capacity are likely to be the sites with highest potential and/or realised grazing).

More generally, it may be good practice to validate models against more than one dataset, as validation data have their own intrinsic inaccuracies. For example, in this study we used more than one validation dataset for each ES (Table 1). More work is required to understand how best to validate ES models, allowing model validation to become standard practice within the ES community, increasing confidence and helping to reduce the implementation gap between ES models and policy- and decision-making (Cash and others 2003; Voinov and others 2014; Wong and others 2014; Clark and others 2016). However, there will always be financial and practical limits to model validation, especially at large scales. Collection of high-quality data is challenging and expensive, and as such would require further investments; indeed, the reason ES models are used is often because of the lack of primary data.

Finally, more work is required to develop and test more complex use models. Whilst we highlight that ES models need to move beyond biophysical production to realised use by beneficiaries, our very simple ES use models require substantial improvement, for example, by incorporating flows of ES (Bagstad and others 2014; Villa and others 2014). Similarly, none of the models we consider here adequately represent temporal dynamics (that is, when are ES being used?) (Scholes and others 2013; Willcock and others 2016), nor can they disaggregate between beneficiary groups (that is, who is using which services?) (García-Nieto and others 2013; Bagstad and others 2014), nor estimate if such use is sustainable. All three points are highly relevant to understanding if the Sustainable Development Goals (https://sustainabledevelopment.un.org/) are being achieved, and hence represent a critical and hugely challenging frontier in both ES modelling and validation. This is further complicated by the fact that model reliability may differ across spatial scales (Scholes and others 2013). For example, because the focus of decision-makers across sub-Saharan Africa predominantly ranges from local to national scales, they require ES information at different gridcell sizes (Willcock and others 2016), and so it is necessary to understand better how the accuracy of ES models varies with spatial resolution.

Conclusions

Our study demonstrates the feasibility of ES model validation, even in data-deficient locations such as sub-Saharan Africa (Suich and others 2015; Willcock and others 2016). Although this demonstration has been long overdue, the lack of such large-scale, multi-model validations is perhaps reflective of the complications involved. In partnership with decision-makers, the advances suggested here could help to ensure ES research continues to inform ongoing policy processes (Voinov and others 2014) (such as the IPBES, the Sustainable Development Goals and CBD Aichi targets). Our findings are of particular relevance to sub-Saharan Africa. Whilst the continent is perceived as relatively data-deficient (Suich and others 2015; Willcock and others 2016), we have shown that adequate data exist to run and validate multiple models for ES of high policy relevance (Willcock and others 2016), particularly related to supplies of ES. Thus, ES models could help to meet the information demand from policy-makers in sub-Saharan Africa (Willcock and others 2016).

Acknowledgements

This work took place under the 'WISER: Which Ecosystem Service Models Best Capture the Needs of the Rural Poor?' project (NE/L001322/1), funded

by the UK Ecosystem Services for Poverty Alleviation program (ESPA; www.espa.ac.uk). ESPA receives its funding from the UK Department for International Development, the Economic and Social Research Council and the Natural Environment Research Council. We thank Kate Schreckenberg and the anonymous reviewers, whose input substantially improved the manuscript.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

OPEN ACCESS

REFERENCES

Agarwal B. 1983. Diffusion of rural innovations: some analytical issues and the case of wood-burning stoves. World Dev 11:359–76.

Avitabile V, Herold M, Heuvelink GBM, Lewis SL, Phillips OL, Asner GP, Armston J, Ashton PS, Banin L, Bayol N, Berry NJ, Boeckx P, de Jong BHJ, Devries B, Girardin CAJ, Kearsley E, Lindsell JA, Lopez-Gonzalez G, Lucas R, Malhi Y, Morel A, Mitchard ETA, Nagy L, Qie L, Quinones MJ, Ryan CM, Ferry SJW, Sunderland T, Laurin GV, Gatti RC, Valentini R, Verbeeck H, Wijaya A, Willcock S. 2016. An integrated pan-tropical biomass map using multiple reference datasets. Glob Chang Biol 22:1406–20.

Bagstad KJ, Semmens DJ, Waage S, Winthrop R. 2013. A comparative assessment of decision-support tools for ecosystem services quantification and valuation. Ecosyst Serv 5:27–39.

Bagstad KJ, Villa F, Batker D, Harrison-Cox J, Voigt B, Johnson GW. 2014. From theoretical to actual ecosystem services: mapping beneficiaries and spatial flows in ecosystem service assessments. Ecol Soc 19:art64.

Bennett ND, Croke BFW, Guariso G, Guillaume JHA, Hamilton SH, Jakeman AJ, Marsili-Libelli S, Newham LTH, Norton JP, Perrin C, Pierce SA, Robson B, Seppelt R, Voinov AA, Fath BD, Andreassian V. 2013. Characterising performance of environmental models. Environ Model Softw 40:1–20.

Bruijnzeel LA, Mulligan M, Scatena FN. 2011. Hydrometeorology of tropical montane cloud forests: emerging patterns. Hydrol Process 25:465–98.

Bryant BP, Borsuk M, Hamela P, Olesond KLL, Schulpe CJE, Willcock S. 2018. Transparent and feasible uncertainty assessment can add value to applied ecosystem services modeling. Ecosyst Serv 33:103–9.

Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, Jäger J, Mitchell RB. 2003. Knowledge systems for sustainable development. Proc Natl Acad Sci USA 100:8086–91.

Chave J, Andalo C, Brown S, Cairns MA, Chambers JQ, Eamus D, Folster H, Fromard F, Higuchi N, Kira T, Lescure J-PP, Nelson BW, Ogawa H, Puig H, Riera B, Yamakura T, Fölster H, Riéra B. 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145:87–99.

Clark WC, Tomich TP, van Noordwijk M, Guston D, Catacutan D, Dickson NM, McNie E. 2016. Boundary work for sustainable development: Natural resource management at the Consultative Group on International Agricultural Research (CGIAR). Proc Natl Acad Sci USA 113:4615–22.

Costanza R, de Groot R, Sutton P, van der Ploeg S, Anderson SJ, Kubiszewski I, Farber S, Turner RK. 2014. Changes in the global value of ecosystem services. Glob Environ Change 26:152–8.

Cruz-Garcia GS, Sachet E, Blundo-Canto G, Vanegas M, Quintero M. 2017. To what extent have the links between ecosystem services and human well-being been researched in Africa, Asia, and Latin America? Ecosyst Serv 25:201–12.

Cumming GS, Buerkert A, Hoffmann EM, Schlecht E, von Cramon-Taubadel S, Tscharntke T. 2014. Implications of agricultural transitions and urbanization for ecosystem services. Nature 515:50–7.

Daw T, Brown K, Rosendo S, Pomeroy R. 2011. Applying the ecosystem services concept to poverty alleviation: the need to disaggregate human well-being. Environ Conserv 38:370–9.

Díaz S, Demissew S, Carabias J, Joly C, Lonsdale M, Ash N, Larigauderie A, Adhikari JR, Arico S, Báldi A, Bartuska A, Baste IA, Bilgin A, Brondizio E, Chan KM, Figueroa VE, Duraiappah A, Fischer M, Hill R, Koetz T, Leadley P, Lyver P, Mace GM, Martin-Lopez B, Okumura M, Pacheco D, Pascual U, Pérez ES, Reyers B, Roth E, Saito O, Scholes RJ, Sharma N, Tallis H, Thaman R, Watson R, Yahara T, Hamid ZA, Akosim C, Al-Hafedh Y, Allahverdiyev R, Amankwah E, Asah ST, Asfaw Z, Bartus G, Brooks LA, Caillaux J, Dalle G, Darnaedi D, Driver A, Erpul G, Escobar-Eyzaguirre P, Failler P, Fouda AMM, Fu B, Gundimeda H, Hashimoto S, Homer F, Lavorel S, Lichtenstein G, Mala WA, Mandivenyi W, Matczak P, Mbizvo C, Mehrdadi M, Metzger JP, Mikissa JB, Moller H, Mooney HA, Mumby P, Nagendra H, Nesshover C, Oteng-Yeboah AA, Pataki G, Roué M, Rubis J, Schultz M, Smith P, Sumaila R, Takeuchi K, Thomas S, Verma M, Yeo-Chang Y, Zlatanova D. 2015. The IPBES conceptual framework—connecting nature and people. Curr Opin Environ Sustain 14:1–16.

Dunham K, Grand JB. 2016. Effects of model complexity and priors on estimation using sequential importance sampling/resampling for species conservation. Ecol Model 340:28–36.

Egarter Vigl L, Depellegrin D, Pereira P, de Groot R, Tappeiner U. 2017. Mapping the ecosystem service delivery chain: capacity, flow, and demand pertaining to aesthetic experiences in mountain landscapes. Sci Total Environ 574:422–36.

Egoh B, Drakou EG, Dunbar MB, Maes J, Willemen L. 2012. Indicators for mapping ecosystem services: a review. Report EUR 25456 EN. Luxembourg, Luxembourg

Fetzel T, Havlik P, Herrero M, Kaplan JO, Kastner T, Kroisleitner C, Rolinski S, Searchinger T, Van Bodegom PM, Wirsenius S, Erb K-H. 2017. Quantification of uncertainties in global grazing systems assessment. Global Biogeochem Cycles 31:1089–102.

García-Nieto AP, García-Llorente M, Iniesta-Arandia I, Martín-López B. 2013. Mapping forest ecosystem services: from providing units to beneficiaries. Ecosyst Serv 4:126–38.

Grainger A. 2008. Difficulties in tracking the long-term global trend in tropical forest area. Proc Natl Acad Sci 105:818–23. http://www.pnas.org/content/105/2/818.abstract

Grêt-Regamey A, Weibel B, Bagstad KJ, Ferrari M, Geneletti D, Klug H, Schirpke U, Tappeiner U. 2014. On the effects of scale for ecosystem services mapping. PLoS ONE 9:e112601.

Hamann M, Biggs R, Reyers B. 2015. Mapping social–ecological systems: Identifying 'green-loop' and 'red-loop' dynamics based on characteristic bundles of ecosystem service use. Glob Environ Change 34:218–26.

Hamann M, Biggs R, Reyers B, Pomeroy R, Abunge C, Galafassi D. 2016. An exploration of human well-being bundles as identifiers of ecosystem service use patterns. PLoS ONE 11:e0163476.

Handley G, Higgins K, Sharma B, Bird K, Cammack D. 2009. Poverty and poverty reduction in sub-Saharan Africa: an overview of the issues. London, UK

Hill TC, Williams M, Bloom AA, Mitchard ETA, Ryan CM. 2013. Are inventory based and remotely sensed above-ground biomass estimates consistent? PLoS ONE 8:e74170.

Hooftman DAP, Edwards B, Bullock JM. 2016. Reductions in connectivity and habitat quality drive local extinctions in a plant diversity hotspot. Ecography (Cop) 39:583–92.

Houerou HNL, Hoste CH. 1977. Rangeland production and annual rainfall relations in the Mediterranean Basin and in the African Sahelo-Sudanian Zone. J Range Manag 30:181.

IOS. 1994. ISO 5725-1:1994 Accuracy (trueness and precision) of measurement methods and results—part 1: general principles and definitions.

IPBES. 2016. Methodological assessment of scenarios and models of biodiversity and ecosystem services. (Ferrier S, Ninan KN, Leadley P, Alkemade R, Acosta LA, Akçakaya HR, Brotons L, Cheung WWL, V.Christensen, Harhash KA, Kabubo-Mariara J, Lundquist C, Obersteiner M, Pereira H, Peterson G, Pichs-Madruga R, Ravindranath N, Rondinini C, Wintle BA, editors.). Bonn: Secretariat of the Intergovernmental Platform for Biodiversity and Ecosystem Services

Kareiva PM. 2011. Natural capital: theory and practice of mapping ecosystem services. Oxford: Oxford University Press.

Kolmogorov AN. 1998. On tables of random numbers. Theor Comput Sci 207:387–95.

Laporte N, Merry F, Baccini A, Goetz S, Stabach J, Bowman M. 2008. Réduire les emissions de CO2 du déboisement et de la degradation dans la République Démocratique du Congo: Un premier aperçu. Falmouth, MA, USA

Lee H, Lautenbach S. 2016. A quantitative review of relationships between ecosystem services. Ecol Indic 66:340–51.

Levin S, Xepapadeas T, Crépin A-S, Norberg J, de Zeeuw A, Folke C, Hughes T, Arrow K, Barrett S, Daily G, Ehrlich P, Kautsky N, Mäler K-G, Polasky S, Troell M, Vincent J, Walker B. 2013. Social-ecological systems as complex adaptive systems: modeling and policy implications. Environ Dev Econ 18:111–32.

Lewis SL, Sonké B, Sunderland T, Begne SK, Lopez-Gonzalez G, van der Heijden GMF, Phillips OL, Affum-Baffoe K, Baker TR, Banin L, Bastin J-F, Beeckman H, Boeckx P, Bogaert J, De Cannière C, Chezeaux E, Clark CJ, Collins M, Djagbletey G, Djuikouo MNK, Droissart V, Doucet J-L, Ewango CEN, Fauset S, Feldpausch TR, Foli EG, Gillet J-F, Hamilton AC, Harris DJ, Hart TB, de Haulleville T, Hladik A, Hufkens K, Huygens D, Jeanmart P, Jeffery KJ, Kearsley E, Leal ME, Lloyd J, Lovett JC, Makana J-R, Malhi Y, Marshall AR, Ojo L, Peh KS-H, Pickavance G, Poulsen JR, Reitsma JM, Sheil D, Simo M, Steppe K, Taedoumg HE, Talbot J, Taplin JRD, Taylor D, Thomas SC, Toirambe B, Verbeeck H, Vleminckx J, White LJT, Willcock S, Woell H, Zemagho L. 2013. Above-ground biomass and structure of 260 African tropical forests. Philos Trans R Soc B Biol Sci 368:20120295.

Malinga R, Gordon LJ, Jewitt G, Lindborg R. 2015. Mapping ecosystem services across scales and continents—a review. Ecosyst Serv 13:57–63.

Marmion M, Parviainen M, Luoto M. 2009. Evaluation of consensus methods in predictive species distribution modelling. Divers Distrib 15:59–69.

Martínez-Harms MJ, Balvanera P. 2012. Methods for mapping ecosystem service supply: a review. Int J Biodivers Sci Ecosyst Serv Manag 8:17–25.

Martínez-López J, Bagstad KJ, Balbi S, Magrach A, Voigt B, Athanasiadis I, Pascual M, Willcock S, Villa F. 2019. Towards globally customizable ecosystem service models. Sci Total Environ 650:2325–36.

McKenzie E, Posner S, Tillmann P, Bernhardt JR, Howard K, Rosenthal A. 2014. Understanding the use of ecosystem service knowledge in decision making: lessons from international experiences of spatial planning. Environ Plan C Gov Policy 32:320–40.

McKenzie E, Rosenthal A, Bernhardt J, Girvetz E, Kovacs K, Olwero N, Tof J. 2012. Guidance and Case Studies for InVEST Users. (WWF, editor.). Washington, USA: World Wildlife Fund

Merow C, Smith MJ, Edwards TC, Guisan A, McMahon SM, Normand S, Thuiller W, Wüest RO, Zimmermann NE, Elith J. 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography (Cop) 37:1267–81.

Mulligan M. 2013. WaterWorld: a self-parameterising, physically based model for application in data-poor but problem-rich environments globally. Hydrol Res 44:748–69.

Mulligan M. 2015. Trading off agriculture with nature's other benefits, spatially. In: Zolin C, de Rodrigues RA, Eds. Impact of Climate Change on Water Resources in Agriculture. Boca Raton: CRC Press.

Mulligan M, Burke S. 2005. Global cloud forests and environmental change in a hydrological context. DFID FRP Project ZF0216 Final Technical Report

Mulligan M, Guerry A, Arkema K, Bagstad K, Villa F. 2010. Capturing and quantifying the flow of ecosystem services. In: Silvestri S, Kershaw F, Eds. Framing the flow: innovative approaches to understand, protect and value ecosystem services across linked habitats. Cambridge: UNEP World Conservation Monitoring Centre. p 26–33.

Ochoa V, Urbina-Cardona N. 2017. Tools for spatially modeling ecosystem services: publication trends, conceptual reflections and future challenges. Ecosyst Serv 26:155–69.

Pachzelt A, Forrest M, Rammig A, Higgins SI, Hickler T. 2015. Potential impact of large ungulate grazers on African vegetation, carbon storage and fire regimes. Glob Ecol Biogeogr 24:991–1002.

Pascual U, Balvanera P, Díaz S, Pataki G, Roth E, Stenseke M, Watson RT, Başak Dessane E, Islar M, Kelemen E, Maris V, Quaas M, Subramanian SM, Wittmer H, Adlan A, Ahn S, Al-Hafedh YS, Amankwah E, Asah ST, Berry P, Bilgin A, Breslow SJ, Bullock C, Cáceres D, Daly-Hassen H, Figueroa E, Golden

CD, Gómez-Baggethun E, González-Jiménez D, Houdet J, Keune H, Kumar R, Ma K, May PH, Mead A, O'Farrell P, Pandit R, Pengue W, Pichis-Madruga R, Popa F, Preston S, Pacheco-Balanza D, Saarikoski H, Strassburg BB, van den Belt M, Verma M, Wickson F, Yagi N. 2017. Valuing nature's contributions to people: the IPBES approach. Curr Opin Environ Sustain 26–27:7–16.

Poppy GM, Chiotha S, Eigenbrod F, Harvey CA, Honzák M, Hudson MD, Jarvis A, Madise NJ, Schreckenberg K, Shackleton CM, Villa F, Dawson TP. 2014. Food security in a perfect storm: using the ecosystem services framework to increase understanding. Philos Trans R Soc Lond B Biol Sci 369:20120288.

Quiroz R, Loayza H, Barreda C, Gavilán C, Posadas A, Ramírez DA. 2017. Linking process-based potato models with light reflectance data: Does model complexity enhance yield prediction accuracy? Eur J Agron 82:104–12.

Redhead JW, May L, Oliver TH, Hamel P, Sharp R, Bullock JM. 2018. National scale evaluation of the InVEST nutrient retention model in the United Kingdom. Sci Total Environ 610–611:666–77.

Redhead JW, Stratford C, Sharps K, Jones L, Ziv G, Clarke D, Oliver TH, Bullock JM. 2016. Empirical validation of the InVEST water yield ecosystem service model at a national scale. Sci Total Environ 569:1–9.

Salmina ES, Wondrousch D, Kühne R, Potemkin VA, Schüürmann G. 2016. Variation in predicted internal concentrations in relation to PBPK model complexity for rainbow trout. Sci Total Environ 550:586–97.

Scholes R, Reyers B, Biggs R, Spierenburg M, Duriappah A. 2013. Multi-scale and cross-scale assessments of social–ecological systems and their ecosystem services. Curr Opin Environ Sustain 5:16–25.

Scholes RJ. 1998. The South African 1: 250 000 maps of areas of homogeneous grazing potential. CSIR, South Africa, Report ENV-P-C 98190

Sewell SJ, Sewell SJ, Desai S. 2016. The impacts of undeveloped roads on the livelihoods of rural women. Rev Soc Sci 1:15–29.

Shackleton SE, Shackleton CM. 2012. Linking poverty, HIV/AIDS and climate change to human and ecosystem vulnerability in southern Africa: consequences for livelihoods and sustainable ecosystem management. Int J Sustain Dev World Ecol 19:275–86.

Sharps K, Masante D, Thomas A, Jackson B, Redhead J, May L, Prosser H, Cosby B, Emmett B, Jones L. 2017. Comparing strengths and weaknesses of three ecosystem services modelling tools in a diverse UK river catchment. Sci Total Environ 584:118–30.

Smith B, Prentice IC, Sykes MT. 2001. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. Glob Ecol Biogeogr 10:621–37.

Smith B, Wårlind D, Arneth A, Hickler T, Leadley P, Siltberg J, Zaehle S. 2014. Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences 11:2027–54.

Spake R, Lasseur R, Crouzat E, Bullock JM, Lavorel S, Parks KE, Schaafsma M, Bennett EM, Maes J, Mulligan M, Mouchet M, Peterson GD, Schulp CJE, Thuiller W, Turner MG, Verburg PH, Eigenbrod F. 2017. Unpacking ecosystem service bundles: Towards predictive mapping of synergies and trade-offs between ecosystem services. Glob Environ Chang 47:37–50.

Steffen W, Richardson K, Rockström J, Cornell SE, Fetzer I, Bennett EM, Biggs R, Carpenter SR, de Vries W, de Wit CA, Folke C, Gerten D, Heinke J, Mace GM, Persson LM, Ramanathan V, Reyers B, Sörlin S. 2015. Planetary boundaries: Guiding human development on a changing planet. Science 80:347.

Stevens FR, Gaughan AE, Linard C, Tatem AJ, Jarvis A, Hashimoto H. 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS ONE 10:e0107042.

Suich H, Howe C, Mace G. 2015. Ecosystem services and poverty alleviation: a review of the empirical links. Ecosyst Serv 12:137–47.

Sun Z, Lorscheid I, Millington JD, Lauf S, Magliocca NR, Groeneveld J, Balbi S, Nolzen H, Müller B, Schulze J, Buchmann CM. 2016. Simple or complicated agent-based models? A complicated issue. Environ Model Softw 86:56–67.

Suwarno A, van Noordwijk M, Weikard H-P, Suyamto D. 2018. Indonesia's forest conversion moratorium assessed with an agent-based model of land-use change and ecosystem services (LUCES). Mitig Adapt Strateg Glob Chang 23:211–29.

Swetnam RD, Fisher B, Mbilinyi BP, Munishi PKT, Willcock S, Ricketts T, Mwakalila S, Balmford A, Burgess ND, Marshall AR, Lewis SL. 2011. Mapping socio-economic scenarios of land cover change: a GIS method to enable ecosystem service modelling. J Environ Manage 92:563–74.

Swets J, Pickett R, Whitehead S, Getty D, Schnur J, Swets J, Freeman B. 1979. Assessment of diagnostic technologies. Science (80-) 205:753–9.

Syfert MM, Smith MJ, Coomes DA, Meagher T, Roberts D. 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PLoS ONE 8:e55158.

Turner KG, Anderson S, Gonzales-Chang M, Costanza R, Courville S, Dalgaard T, Dominati E, Kubiszewski I, Ogilvy S, Porfirio L, Ratna N, Sandhu H, Sutton PC, Svenning J-C, Turner GM, Varennes Y-D, Voinov A, Wratten S. 2016. A review of methods, data, and models to assess changes in the value of ecosystem services from land degradation and restoration. Ecol Modell 319:190–207.

van Soesbergen A, Mulligan M. 2018. Uncertainty in data for hydrological ecosystem services modelling: Potential implications for estimating services and beneficiaries for the CAZ Madagascar. Ecosyst Serv 33:175–86.

Verhagen W, Kukkala AS, Moilanen A, van Teeffelen AJA, Verburg PH. 2017. Use of demand for and spatial flow of ecosystem services to identify priority areas. Conserv Biol 31:860–71.

Villa F, Bagstad KJ, Voigt B, Johnson GW, Portela R, Honzák M, Batker D. 2014. A methodology for adaptable and robust ecosystem services assessment. PLoS ONE 9:e91001.

Villa F, Balbi S, Athanasiadis IN, Caracciolo C. 2017. Semantics for interoperability of distributed data and models: foundations for better-connected information. F1000Research 6:686.

Villarino SH, Studdert GA, Laterra P, Cendoya MG. 2014. Agricultural impact on soil organic carbon content: testing the IPCC carbon accounting method for evaluations at county scale. Agric Ecosyst Environ 185:118–32.

Voinov A, Seppelt R, Reis S, Nabel JEMS, Shokravi S. 2014. Values in socio-environmental modelling: persuasion for action or excuse for inaction. Environ Model Softw 53:207–12.

von Freyberg J, Moeck C, Schirmer M. 2015. Estimation of groundwater recharge and drought severity with varying model complexity. J Hydrol 527:844–57.

Willcock S, Hooftman D, Sitas N, O'Farrell P, Hudson MD, Reyers B, Eigenbrod F, Bullock JM. 2016. Do ecosystem service maps and models meet stakeholders' needs? A preliminary survey across sub-Saharan Africa. Ecosyst Serv 18:110–17.

Willcock S, Martínez-López J, Hooftman DAP, Bagstad KJ, Balbi S, Marzo A, Prato C, Sciandrello S, Signorello G, Voigt B, Villa F, Bullock JM, Athanasiadis IN. 2018. Machine learning for ecosystem services. Ecosyst Serv 33:165–74.

Willcock S, Phillips OL, Platts PJ, Balmford A, Burgess ND, Lovett JC, Ahrends A, Bayliss J, Doggart N, Doody K, Fanning E, Green JM, Hall J, Howell KL, Marchant R, Marshall AR, Mbilinyi B, Munishi PK, Owen N, Swetnam RD, Topp-Jorgensen EJ, Lewis SL. 2014. Quantifying and understanding carbon storage and sequestration within the Eastern Arc Mountains of Tanzania, a tropical biodiversity hotspot. Carbon Balance Manag 9:2.

Wong CP, Jiang B, Kinzig AP, Lee KN, Ouyang Z. 2014. Linking ecosystem characteristics to final ecosystem services for public policy. Ecol Lett 18:108–18.

Zhang L, Peng J, Liu Y, Wu J. 2017. Coupling ecosystem services supply and human ecological demand to identify landscape ecological security pattern: A case study in Beijing–Tianjin–Hebei region, China. Urban Ecosyst 20:701–14. http://link.springer.com/10.1007/s11252-016-0629-y. Last accessed 29/01/2019