Ecosystem Services 66 (2024) 101599

Contents lists available at ScienceDirect

Ecosystem Services

journal homepage: www.elsevier.com/locate/ecoser

Full Length Article

'Uncertainty audit' for ecosystem accounting: Satellite-based ecosystem extent is biased without design-based area estimation and accuracy assessment

Zander S. Venter [*], Bálint Czúcz, Erik Stange, Megan S. Nowell, Trond Simensen, Bart Immerzeel, David N. Barton

Norwegian Institute for Nature Research – NINA, Sognsveien 68, 0855 Oslo, Norway

ARTICLE INFO

Keywords:
Uncertainty audit
Ecosystem service
Stratification
Remote sensing
Earth observation
Sampling

ABSTRACT

There are currently no guidelines in the System of Environmental-Economic Accounting Ecosystem Accounting (SEEA EA) for quantifying and disclosing uncertainty. However, without quantifying uncertainty, it is unclear whether or not accounting tables contain biased (erroneous) area estimates which do not reflect real land cover changes. We use Oslo municipality in Norway as a case study to illustrate best practices in quantifying unbiased area estimates using design-based statistical methods. As input for ecosystem extent accounts, we compared a custom Sentinel-2 land cover map with a globally available one called Dynamic World for 2015, 2018 and 2021. The design-based area estimation involved (i) generating a stratified probability sample of locations using the satellite-based maps to define strata, (ii) assigning ecosystem type labels to the samples using photointerpretation according to a response design protocol, and (iii) applying a stratified area estimator to produce 95% confidence intervals around opening, closing and change stocks in the extent accounting table. We found that pixel counting practices, currently adopted by the SEEA EA community, led to biased extent accounts, particularly for ecosystem conversions, with biases averaging 195% of the true change value derived from design-based methods. We found that the uncertainty inherent in state-of-the-art satellite-based maps exceeded the ability to detect real change in extent for some ecosystem types including water and bare/artificial surfaces. In general, uncertainty in extent accounts is higher for ecosystem type conversion classes compared to stable classes, and higher for 3-yr compared to 6-yr accounting periods. Custom, locally calibrated satellite-based maps of ecosystem extent changes were more accurate (81% overall accuracy) than globally available Dynamic World maps (75%). We suggest that rigorous accuracy assessment in SEEA EA will ensure that ecosystem extent (and consequently condition and service) accounts are credible. A standard for auditing uncertainty in ecosystem accounts is needed.

1. Introduction

Climate change and biodiversity loss are tightly intertwined, and both are related to land use. The Kunming-Montreal Global Biodiversity Framework aims to reduce the loss of areas of high biodiversity importance to near zero by 2030 through Target 1, which focuses on inclusive spatial planning and effective land use and land cover (LULC) policies. Quantifying, understanding, and communicating the dynamics of land change is a premise for knowledge-based land use policies, including strategies for degradation neutrality and no net loss of natural habitats (Cowie et al., 2018). So-called ecosystem extent accounts depict changes in the spatial extent or area of different ecosystems over a certain period within a focal area—i.e., a country, region, or a municipality (United Nations, 2021). Accounting for changes in ecosystem types (ET), their condition, and the services they provide for society is a challenging task with inherent uncertainty (Costanza et al., 1997; Foody, 2015), yet it is necessary for tracking progress towards achieving the sustainable development goals (Bebbington and Unerman, 2018). The UN Statistical Commission (UNSC) has adopted an internationals statistical standard for the System of Environmental-Economic Accounting Ecosystem Accounting (SEEA EA) (United Nations, 2021) and have developed guidelines for implementation (Edens et al., 2022).

* Corresponding author.
E-mail address: zander.venter@nina.no (Z.S. Venter).

https://doi.org/10.1016/j.ecoser.2024.101599
Received 1 August 2023; Received in revised form 16 November 2023; Accepted 17 January 2024
Available online 23 January 2024
2212-0416/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Work is now ongoing in national statistical offices to put the standard into national practice, and experimental ecosystem accounts have been carried out by the EU (e.g. Vysna et al., 2021). In addition, EUROSTAT is developing a regulation for common implementation by EU member states, with associated countries such as Norway following the same reporting standards under the SEEA agreement.

The current guidelines for biophysical accounting of changes in ecosystem extent, condition and services do not specify either minimum accuracy criteria nor methods for estimating and reporting uncertainty (United Nations, 2022). This is perhaps because SEEA EA emerged as a complement to the System for National Accounts (SNA) where there is no uncertainty: the accounts consist of transactions of exactly known amounts and there is no estimation, modelling, proxy indicators, or measurement errors. Nevertheless, GDP revisions can be quite large (e.g., Ghana 60 %, China 15 %, Netherlands 7 %) (Barton et al., 2019), indicating SNA are also victim to uncertainty even if it is not explicitly acknowledged. However, SEEA EA is fundamentally different to SNA because it builds on indicators and estimation, which are inherently uncertain. The lack of guidelines for reporting uncertainty is unfortunately consistent with recent state-of-the-art national ecosystem accounts where uncertainty is not quantified (Hein et al., 2020; Heris et al., 2021; Petersen et al., 2022). Business accounting research has demonstrated that different ways of accounting for and disclosing uncertainty can affect auditor and investor objectivity (Eilifsen et al., 2021; Mayhew et al., 2001). Therefore, communication of accuracy and uncertainty is considered vital for the credibility and utility of future ecosystem accounts (Bagstad et al., 2021; Schägner et al., 2013). Successful implementation of SEEA EA will depend on developing best practices for rigorous accuracy quantification, assessment and disclosure – an 'uncertainty audit' for ecosystem accounting.

Quantifying error in ecosystem extent maps and accounts is a priority because ecosystem extent is the foundation for biophysical modelling of ecosystem condition and services. Ecosystem extent accounts are essentially an area estimation of ETs and their changes over time, which often rely on existing LULC maps (United Nations, 2022). The classic approach to generating a statistical estimate for the area covered by various LULC categories relies on design-based (survey) statistics. First a representative sample is identified for the area of interest (the study area). The sampling units can be points (point-frame) or small polygons (area-frame samples), which are identified using a probability sampling design such as random, systematic or stratified random sampling (Gallego, 2004). Sampling units are then manually classified according to a target LULC typology with either ground surveys or visual interpretation of aerial imagery (orthophotos), creating a 'reference dataset' from which the total area for each category of the target typology is inferred.

Ecosystem accounting is inherently spatial and many steps in populating the ecosystem condition and service accounts requires a wall-to-wall spatial delineation of the ETs. Accordingly, the development of a comprehensive ET map was considered unavoidable to provide the necessary spatial detail. Perhaps the most commonly used wall-to-wall ET maps are those produced by national or international statistical or mapping authorities which use manual mapping approaches based on visual interpretation of aerial and satellite imagery. In Europe the Corine Land Cover (CLC) product (Büttner, 2014) includes single year ET maps, but also change maps which are mapped separately from the single year maps which can be used for ecosystem extent accounting purposes. Because such maps are manually digitized, they are considered highly accurate, but when subjected to external validation, they nevertheless reveal error rates of between 2 and 25 %, depending on the ET (Büttner, 2014). Apart from the error rates, they also have relatively restrictive minimum size thresholds (25 ha/100 m in the case of CLC products) and update frequencies (every three years for CLC). These constraints mean that CLC-like maps cannot meet all ecosystem accounting requirements especially when considering urban thematic accounts which require smaller minimum mapping units. Therefore, the ecosystem accounting community has increasingly relied on satellite-based remote sensing maps of land cover built with a combination of manual mapping and machine learning models for the purposes of quantifying ecosystem extent (Cord et al., 2017; de Araujo Barbosa et al., 2015). Such satellite-based maps provide wall-to-wall coverage of the globe, with flexibly-defined minimum mapping units, and can be continuously updated. State-of-the-art global LULC maps are based on Sentinel satellites from the European Space Agency which deliver result in 10 m resolution ET extent maps (Venter et al., 2022). Sentinel-based products offer improved spatial and temporal resolution compared to previous datasets such as CLC (Büttner, 2014) and Copernicus Global Land Cover (Buchhorn et al., 2020). Even though manually-digitized and satellite-based ET maps are easily complemented by design-based sampling to give better estimations for the total area of each ET, the extent to which this can improve the ecosystem accounting process has not been adequately explored.

Although satellite-based maps of ecosystem extent meet SEEA EA requirements, they can lead to biased area estimates (Foody, 2015; Gallego, 2004). The typical approach for estimating area size from satellite-based extent maps is 'pixel counting', whereby the number of pixels per ET are summed and multiplied by the pixel area (Fig. 1). However, pixel counting does not account for either classification errors caused by algorithms in artificial intelligence (AI) models that are used to convert satellite imagery into a categorical map (Olofsson et al., 2014), or errors in the data used to calibrate the AI models themselves (Foody et al., 2016). Such pixel counting bias can be particularly large for rare classes, including any "change classes" corresponding to a specific ET conversion (e.g. wetland loss) over a specific time period. ET conversions are the focus of ecosystem accounting for the purpose of change detection and are often of high policy interest. Depending on the length of accounting period and rate of landcover change, ET conversions nearly always cover only a very small proportion of an accounting area (Foody, 2013). In some cases, the bias in satellite-based maps may exceed the magnitude of actual change and therefore compromise the ability to detect statistically significant changes in ecosystem extent (Kleinewillinghöfer et al., 2022). Without quantifying area bias and uncertainty, one cannot conclude whether accounting tables reflect real changes or merely artifacts in the methods used to produce them. Errors in land cover classification can generate wildly disparate value estimates of the services ecosystems provide. For instance, after correcting for misclassification bias in a six-class national land cover map for the United States, Foody (2015) found that ecosystem services value changed from US$ 1118 billion yr^{-1} to US$ 600 billion yr^{-1}.

The SEEA EA biophysical accounts guidelines suggest reporting classification accuracy statistics for the AI model or the ET map using an error matrix, with the assumption that this is a reliable reflection of the accuracy of area estimates reported in the extent accounts (United Nations, 2022). However, class-specific, overall model or map accuracy may have little bearing on the bias and uncertainty in area estimates reported in an extent account (Radoux and Bogaert, 2020). The divergence between map accuracy and extent account accuracy is especially large, with potentially high bias for class-specific accuracies, when ground truth data used to quantify accuracy are not based on a probability sample or when accuracy estimates are not calculated at the appropriate scale. For example, reporting map accuracies from continental/global maps when conducting (sub)national accounts is problematic because map accuracy varies with spatial scale and the corresponding frequency and distribution of land classes. Furthermore, we cannot assume that accuracies of either a LULC time series product or the AI model used to produce it will be representative of the error in the LULC change maps when comparing annual maps over an accounting period.

We can overcome the limitations of pixel-counting and produce unbiased ecosystem extent accounts by using a combination of satellite-based ecosystem extent maps and sampling-based reference data. In a *design-based area estimator* (Fig. 1), we use an algorithm's ET classifications of satellite-based ecosystem extent maps as a basis for stratified

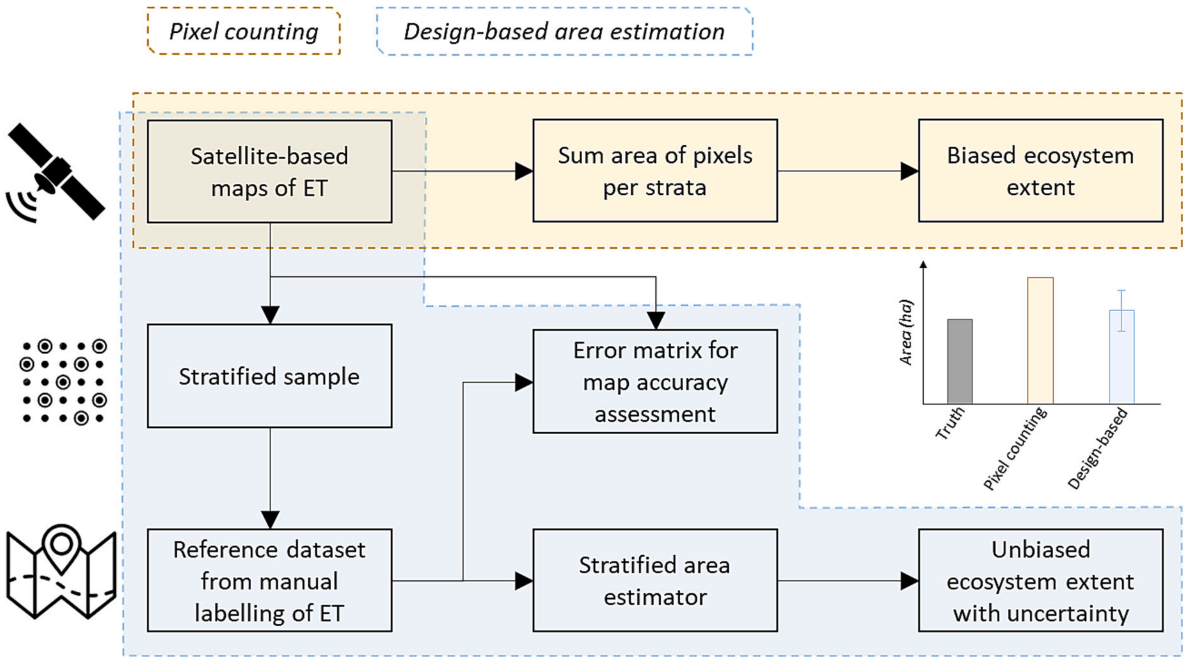

Fig. 1. Workflow for deriving ecosystem extents from pixel counting methods and design-based area estimation methods.

random sampling, and manually label each sampling unit through ground surveys or visual (photo)interpretation to generate a reference dataset (Olofsson et al., 2014). We then use the reference dataset to calculate the area covered by each ET (i.e., the ecosystem extents) and quantify the uncertainty of each ET extent estimate—often reported as 95 % confidence intervals. Such design-based area estimation has been adopted by the remote sensing community to provide statistics on agriculture, forestry, and other similar domains (e.g. Arévalo et al., 2020; Gallego, 2004).

The application of design-based area estimation to ecosystem accounting has not yet been adequately explored, however, and several research gaps need to be addressed to inform guidelines for implementing design-based area estimation in satellite-based ecosystem extent accounts. One, we need better understanding of how much bias is introduced through pixel counting methods as opposed to design-based area estimation. Two, we need to explore the accuracy requirements for extent accounting at varying administrative levels. The accuracy needs of ecosystem accounting to support national versus local government policy and planning needs are different (Grammatikopoulou et al., 2023). For example, are the global high resolution (10 m) LULC maps (e.g., Venter et al., 2022) accurate enough for municipal-level extent accounting to support landuse planning, or are locally-calibrated LULC maps necessary? Three, we need to assess whether uncertainty inherent in satellite-based maps is presently too large to detect real changes over shorter accounting periods (i.e. annual change as in national accounts), or if longer accounting periods or more generalized ET typologies would be more appropriate.

To address these research gaps, we investigated an ecosystem accounting case study in Oslo, Norway. We explored how bias and uncertainty in ecosystem extent estimates vary with input data type and accounting period. We define bias as the difference between pixel counts of algorithm ET classification and the statistical estimator based on the reference dataset, while uncertainty is defined as the likely limits to the aforementioned bias for each area estimate. Our study investigates how bias and uncertainty in ecosystem extent differ between 1) a custom Sentinel-2 land cover map vs the globally available Dynamic World land cover map, and 2) between 3- and 6-year accounting periods. By seeking answers to these research questions, we attempt to illustrate best practices in accounting for ecosystem extent.

2. Methods and materials

2.1. Study area

Oslo municipality (59′55 N, 10′45 E) contains the capital of Norway and had a population of 699,827 in 2021, or 13 % of the country's population. The 454 km^2 municipality is largely covered by forest, with a built-up zone interspersed with grass, trees and water bodies (Fig. 2). Oslo is one of Europe's fastest growing cities, which has resulted in large scale urban restructuring since the early 1990′s to accommodate this rapid growth (Kjærås, 2023), focused on urban densification and redevelopment as the predominant processes. Oslo has been a focus area for testing urban ecosystem service mapping and assessment, a lab for experimental urban ecosystem accounting and forerunner for testing downscaling of ecosystem accounting standards to the municipal level (e.g. Barton et al., 2015; Cimburova and Barton, 2020; Garnåsjordet et al., 2021; Hanssen et al., 2021; Venter et al., 2021; 2020). Oslo was the first municipality in Norway to implement urban vegetation cover extent accounts (Oslo kommune, 2018). As a city situated between a coastal fjord, boreal forests and agricultural landscapes, Oslo comprises a wide range of different ETs within a relatively small area.

2.2. Satellite-based ecosystem extent maps

To map ecosystem extent and its changes, we use two 10 m resolution satellite-based maps which both provide annual status maps and do not directly map changes between two years. There is a well-known trade-off in satellite-based LULC maps where global maps generalize well over large spatial extents, but have regional/local inaccuracies that might limit their utility at municipal scales. Therefore, we wanted to compare a global LULC map with a local one. For a global LULC map we chose a continuously updated map called Dynamic World produced by Google (Brown et al., 2022), which is the only existing LULC map derived from 10 m resolution satellite data with a time series dating back to 2015. Dynamic World is produced using a deep learning model which classifies each Sentinel-2 satellite image into a 9-class LULC map and provides class-specific probability scores defining the likelihood that each pixel belongs to a given LULC class. The model has been applied to the entire Sentinel-2 image archive (2015 to present) and is

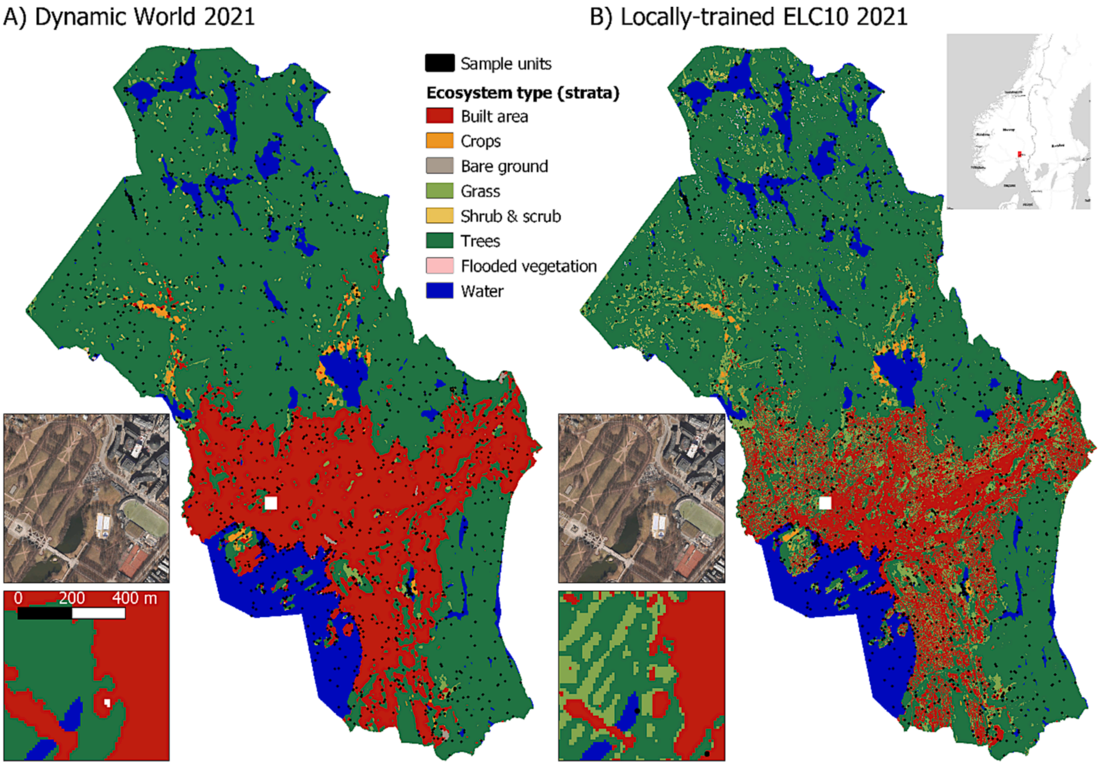

Fig. 2. Maps of ecosystem extent in 2021 for Dynamic World and custom map which is a version of ELC10 trained on local reference data. Inset map shows Frogner park and is illustrated in white. The location of sampling units (10x10m quadrats) used for design-based area estimation and map accuracy assessment are shown in black.

operationally delivering near real-time LULC maps as new Sentinel-2 scenes become available (every 5 days). Although Dynamic World has inferior accuracy over Europe compared to the European Space Agency's World Cover (Venter et al., 2022), it is the only global LULC dataset to span the period 2015 to present. We created annual LULC composites for 2015, 2018 and 2021 by calculating the median probability scores for all LULC classes across all Dynamic World images during each year's growing season (April to September) and then classifying by taking the class with the highest probability score per pixel.

The second LULC map included in our study was a local one was customized to Norwegian conditions. Given that Norway does not have any time series national land cover maps, we conducted a custom implementation of a 10 m resolution European land cover map called ELC10 which is built on Sentinel-1 and −2 imagery (Venter and Sydenham, 2021). Here we used the same workflow as used to produce ELC10, except that we substituted the model's training data with local calibration data collected over Oslo municipality. The calibration dataset consisted of 2500 randomly distributed 10x10 m sampling quadrats aligned to the Sentinel-2-pixel grid. The samples were labelled using visual interpretation of very high resolution orthophotos with a reference year of 2018 according to the Dynamic World LULC typology (Table 1). A random forest model was then trained on the reference sample with Spectral-temporal metrics for all Sentinel-1 and −2 images during 2018. The model was then used to classify LULC for 2015, 2018 and 2021 over the entire municipality. All processing of satellite imagery and LULC data was performed in Google Earth Engine (Gorelick et al., 2017).

For the purposes of addressing our research aim, we employed a simplified 4-class ecosystem typology which has a cross-walk to the Eurostat level 1 typology, recommended by SEEA EA and based on the IUCN ecosystem typology (Keith et al., 2022). We considered three accounting periods (2015–2018; 2018–2021; and 2015–2021), which gave two instances of 3-year intervals and one instance of a 6-year interval. Three years is the frequency proposed for ecosystem extent accounting in the EU (European Commission, 2022). The temporal extent was determined by the limited availability of Sentinel-2 imagery, and also the 3-year update cycle of commonly used European statistical datasets such as Corine land cover (Büttner, 2014) and LUCAS (d'Andrimont et al., 2020).

Table 1
Ecosystem extent typology used with descriptions from Dynamic World. Cross-walk to Eurostat level 1 typology is provided along with the simplified 4-class typology used in the paper.

Dynamic World	Eurostat level 1	Custom 4-class
Built area	Settlements and other artificial areas	Bare
Bare ground	Sparsely vegetated ecosystems	
Crops	Cropland	Vegetated - short
Grass	Grassland	
Flooded vegetation	Inland wetlands	
Shrub & scrub	Heathland and shrub	Vegetated - tall
Trees	Forest and woodland	
Water	Rivers and canals	Water
	Lakes and reservoirs	
	Marine inlets and transitional waters	

2.3. Pixel counting

The most common approach to calculate ecosystem extent accounts is pixel counting. For each ET, the number of pixels classified as that ET are counted and then multiplied by the pixel area. We conducted pixel counting over Oslo municipality a total of six times for each unique combination of accounting period (2015–2018, 2018–2021, and 2015–2021) and satellite data source (Dynamic World and custom ELC10). Each map consists of a 16-class typology with 4 stable, and 12 conversion classes (Fig. 3). Pixel counting areas were used to quantify opening, change and closing stocks for each ET for each accounting period.

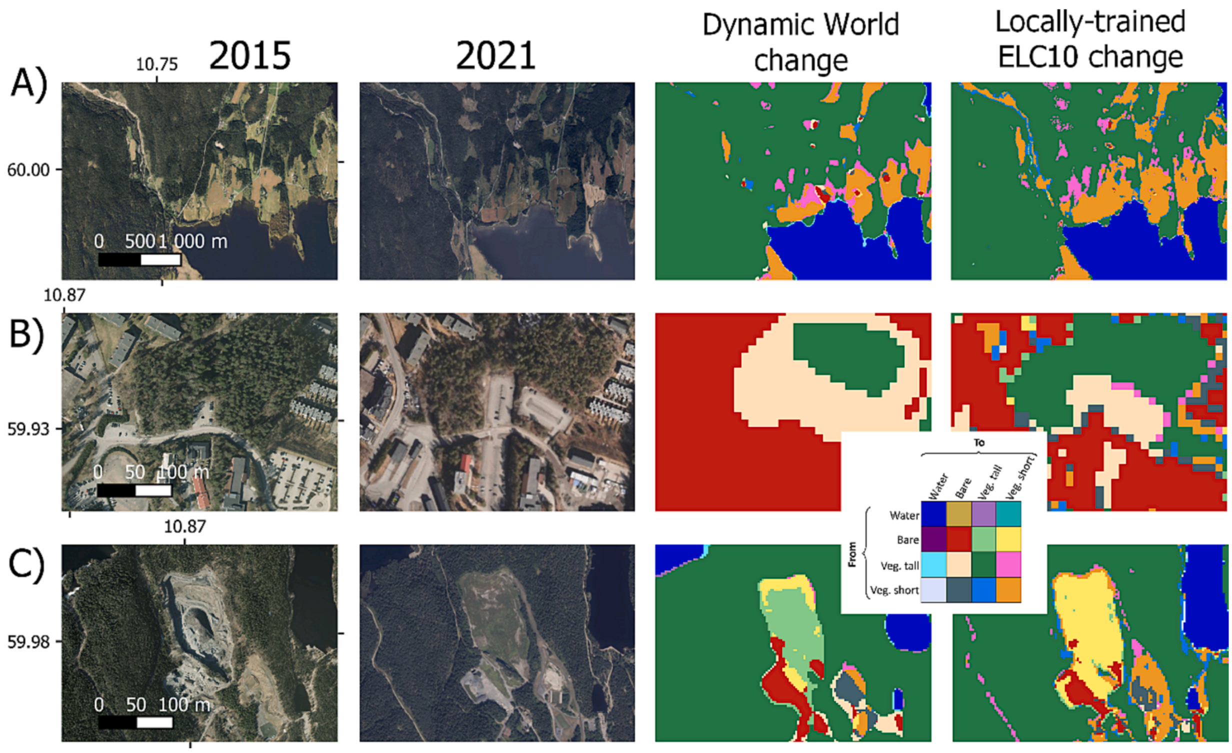

Fig. 3. Maps of ecosystem extent change between 2015 and 2021 according to a 4-class typology resulting in a 16-class change typology. The maps highlight certain types of change and are illustrative of the differences between Dynamic World and the ELC10 approach. The color legend is a transition matrix and should be read from left to right in terms of direction of change.

2.4. Design-based area estimator

The approach we adopt for design-based area estimation is the stratified area estimation described in detail in Olofsson et al. (2014). We summarise the approach as applied to ecosystem extent accounting here (see Fig. 1), however, for a comprehensive description of the statistical methodology and formulae, we refer the reader to Olofsson et al. (2014). As with the pixel counting, the design-based estimation is repeated for each unique combination of accounting period (2015–2018, 2018–2021, and 2015–2021) and satellite data source (Dynamic World and custom ELC10). This results in six unique sets of stratified samples, area estimates with associated uncertainty intervals. It is important to note that this process is completely distinct from the collection of calibration data for the ELC10 AI model described in section 2.2.

i) *Stratified sample:* The satellite-derived LULC change maps were used to generate a series of stratified random samples of 10x10m quadrats, aligned with the Sentinel-2 pixel grid, over Oslo municipality. The mapped ET and their change classes define the strata and the number of samples allocated per strata are proportional to the mapped strata areas. The total sample size was determined using the stratified variance estimator described in Olofsson et al. (2014). This requires one to define a target standard error of 2 % for the anticipated overall accuracy estimates. In addition, one needs to speculate an expected error matrix based on previous experience with mapping ET stable and transition strata. Based on previous experience in Venter and Sydenham (2021), class-specific accuracies were anticipated as 60 % for ET conversion classes and 80 % for stable classes. As recommended good practice, for very rare conversion classes, we set a minimum sample size of 50. The resulting sample allocations are presented in Table 2. The AREA2 tool (https://area2.readthedocs.io/en/latest/overview.html) in Google Earth Engine was used to randomly select the sampling units from each stratum.

ii) *Manual labelling of ecosystem types:* We developed a response design protocol to label the sampling units according to the ecosystem typology defined in Table 1 for the years 2015, 2018 and 2021. To facilitate the labelling of samples, we developed a Google Earth Engine web app that allowed for visual interpretation of multi-temporal orthophotos provided by the Norwegian Mapping Authority, Sentinel-2 mosaics, and very high resolution satellite images provided by the Copernicus Contributing Missions. The web app allowed interpreters to easily navigate between samples, inspect the reference imagery interactively, and assign ET labels with the click of a button. The samples were randomly allocated among five interpreters who were trained and provided with a sampling manual to achieve a common understanding of the typology and what possible conversion classes look like in satellite/aerial imagery. A random sub-sample of 100 units were labelled by all interpreters in order to estimate variability and potential error in the reference data collection. The group of interpreters consisted of researchers in landscape ecology and ecosystem accounting and were thus familiar with ET definitions and orthophoto interpretation methods.

iii) *Error matrix for map accuracy assessment:* The reference data were used to produce an error matrix in terms of proportion of area from mapped vs reference samples. We quantified an estimate of the overall accuracy for each map, and we also estimated the user's accuracy (or commission error rate), and producer's accuracy (or omission error rate) for each ET category in each map. The sampling variability of the different accuracy metrics was quantified as their standard error.

iv) *Stratified area estimator:* Area was estimated from the reference dataset using a simple stratified estimator (Olofsson et al., 2014). Uncertainty in area estimates was quantified as 95 % confidence intervals. Steps iii and iv were conducted in R using the

Table 2
Sample allocation for the stratified random sample. Number of sampling units per ecosystem type change strata used for quantifying map accuracy, and bias and uncertainty in ecosystem extent estimates. Class abbreviations are shown in parentheses for reference in later figures.

Strata (ET change classes)	2015—2018	2018—2021	2015—2021	Total
Water - Water (W to W)	706	834	792	2332
Water - Bare (W to B)	79	60	75	214
Water - Vegetation tall (W to VT)	75	63	69	207
Water - Vegetation short (W to VS)	137	65	61	263
Bare - Water (B to W)	95	61	84	240
Bare - Bare (B to B)	925	1001	935	2861
Bare - Vegetation tall (B to VT)	57	66	68	191
Bare - Vegetation short (B to VS)	121	91	111	323
Vegetation tall - Water (VT to W)	103	85	120	308
Vegetation tall - Bare (VT to B)	105	61	112	278
Vegetation tall - Vegetation tall (VT to VT)	2759	2742	2710	8211
Vegetation tall - Vegetation short (VT to VS)	143	95	168	406
Vegetation short - Water (VS to W)	118	139	123	380
Vegetation short - Bare (VS to B)	110	92	92	294
Vegetation short - Vegetation tall (VS to VT)	92	98	122	312
Vegetation short - Vegetation short (VS to VS)	487	559	470	1516
Total	6112	6112	6112	18,336

'mapaccuracy' package (Costa, 2022) which implements the formulae in Olofsson et al. (2014).

v) *Unbiased ecosystem extent with uncertainty:* Area estimates with 95 % confidence intervals were reported in accounting tables showing opening, change and closing stocks for each ET and each accounting period.

2.5. Effects of input data type, accounting period, and sampling effort

We explored the effect of input data type (Dynamic World vs custom ELC10 maps) and accounting period (3- vs 6-year) on pixel counting relative bias and estimator uncertainty. Relative bias (B) was quantified as the percentage difference between the area proportion estimate derived from pixel counting (A_{pc}), and that derived from the design-based approach (A_{db}):

$$B = \frac{A_{pc} - A_{db}}{A_{db}}$$

Uncertainty for design-based area estimates was quantified as 95 % confidence intervals (CI) based on the reference sample using equations in Olofsson et al. (2014). For purposes of comparing across ETs, we calculated estimator relative uncertainty (U) as a percentage of the class-specific area estimate (A_{db}) as:

$$U = \frac{CI}{A_{db}}$$

3. Results

3.1. Bias in pixel counting approach

For "stable" ecosystem change classes (i.e. the ones corresponding to no change, Fig. 4A), pixel counting led to area estimates that were proportional to the area estimates derived from design-based methods for water (W to W) and tall vegetation (VT to VT). However, there was substantial bias introduced by the pixel counting for stable bare (B to B) and short vegetation (VS to VS), particularly for estimates based on Dynamic World. For ET conversion classes (Fig. 4B), pixel counting produced area estimates that were greater or smaller than those from design-based methods. In most cases, the ecosystem extents from pixel counting exceeded the 95 % confidence intervals around the design-based extent estimates (Fig. 4). In extreme cases, such as conversions from short vegetation (VS) to tall vegetation (VT) and water (W), the relative bias in area estimates resulting from pixel counting exceeded 500 % (Fig. 5). Comparatively, pixel counting of stable classes such as tall vegetation to tall vegetation led to smaller relative biases. When averaged across data source and accounting period, pixel counting led to a relative bias value of 15 % for stable classes, compared to 195 % for conversion classes.

Pixel counting from locally-trained ELC10 maps produced greater bias (168 %) compared to Dynamic World maps (71 %). The pixel counting bias in ELC10 maps was greatest for ecosystem conversion classes. This is despite the fact that the map accuracy assessment (Table 4) showed that ELC10 was more accurate than Dynamic World, highlighting the fact that map accuracy and pixel counting bias are not necessarily related. Pixel counting in the 3-year accounting period produced greater bias (99 %) compared to the 6-year accounting period (81 %; Fig. 5B). The difference in pixel counting bias between 3-yr and 6-yr accounting period was greatest for conversions between tall and short vegetation (Fig. 5A). Such conversions represent a gradient in vegetation structure changes that are typically gradual (e.g. tree growth), and are therefore more accurately captured by longer accounting periods.

The pixel counting bias propagated through to the extent accounting tables (Table 3) which quantify opening and closing extents along with net change for a given accounting period. In some cases, the net change in ecosystem extents estimated from pixel counting differed in both magnitude and direction compared to the design-based estimates. For instance, pixel counting from the ELC10 map revealed a net gain in tall vegetation (+189 ha) between 2015 and 2021, yet the unbiased estimate from design-based methods revealed a statistically significant net loss of tall vegetation (-252 ± 212 ha).

3.2. Uncertainty in design-based estimation

The relative uncertainty in design-based area estimates was on average much greater for ET conversion classes (97 %) compared to stable classes (6 %; Fig. 6A). In some cases, such as the conversion of short vegetation to water, the width of the 95 % confidence intervals was nearly double the magnitude of the area estimate. ET conversion classes that were estimated with the lowest relative uncertainty included tall vegetation to short vegetation (characteristic of forest clear-cutting), bare to water (characteristic of rising reservoir levels), and short vegetation to bare (characteristic of vegetation clearing for infrastructure development).

Uncertainty in area estimates from locally-trained ELC10 maps were on average lower (69 %) than the uncertainty from Dynamic World maps (84 %). On a class-by-class basis, the difference in map accuracies between ELC10 and Dynamic World (Table 4) corresponded to the differences in relative uncertainty in their area estimates. For instance, ELC10 produced a higher producer's accuracy for tall vegetation to short vegetation class compared to Dynamic World, and this corresponded to a lower uncertainty value in the resulting area estimate for this class.

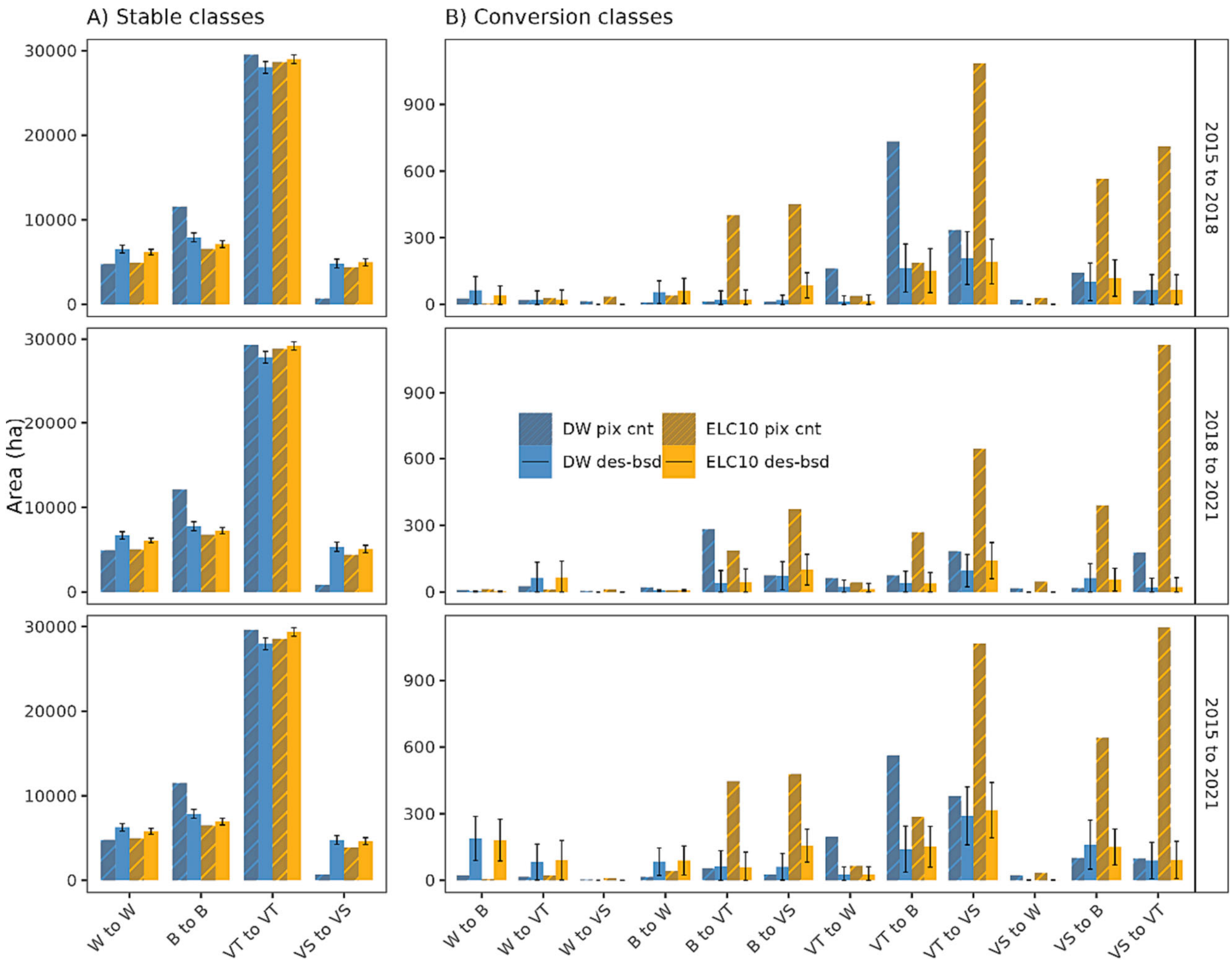

Fig. 4. Area estimates for stable (A) and conversion (B) ecosystem extent classes. Extents for satellite-based maps (DW = Dynamic World; ELC10 = locally trained ELC10 map) and area estimation method (pix cnt = pixel counting; des-bsd = design-based area estimation) are differentiated by color and pattern. Design-based area estimates are provided with 95 % confidence interval error bars. The accounting period and opening, change and closing stocks are presented as facetted panels.

The 3-year accounting period produced greater uncertainty (95 %) compared to the 6-year accounting period (65 %; Fig. 6B).

The uncertainty in area estimates for opening and closing extents were small enough to draw conclusions about differences in ecosystem extents between the four ETs (Table 3). By this we mean that the difference in extents between classes exceeds the width of the 95 % confidence intervals and are therefore significantly different. However, the uncertainty in area estimates for net changes were too large to conclude whether there had been significant changes in most cases (Table 3; Figure S1). The few cases where uncertainty was low enough to identify significant changes were net increases in bare cover (including built land), and losses in tall vegetation.

4. Discussion

Pixel counting is a fast, and straightforward method to provide a rough overview of land cover changes and is therefore widely adopted by the ecosystem accounting community. However, our results indicate that the practice of pixel counting for area estimation may lead to biased extent accounts irrespective of ecosystem type, input data source, or accounting period length. In the Oslo case study, the extent accounts resulting from pixel counting were in some cases biased both in magnitude and direction of change. In other words, pixel counting can lead to conclusions about ecosystem loss or gain that are inaccurate. There are several policy domains including agricultural, forestry, and climate policy that require area estimations for rare LULC types (i.e. extremely small proportional cover), such as ET conversions, in their reporting workflows. In many of these domains the application of unbiased area estimators based on design-based methods has already become a best practice or even an official requirement (e.g. Kleinewillinghöfer et al., 2022). When remote sensing is used to map such rare LULC classes (e.g. deforestation for REDD+, Mitchell et al., 2017) design-based methods involve a (post-)stratified sampling to (1) remove the bias introduced by the satellite-based classification in mapped areas derived from pixel-counting, and (2) produce 95 % confidence intervals for area estimates derived from the reference sample. In some cases, this can be achieved even by reusing the same dataset that was used by the producers of the map to validate their product (Stehman, 2013). In our study, we apply this methodology to ecosystem extent accounts in Oslo and provide evidence that confirms the risks of relying on pixel counting for area estimation. To the best of our knowledge, no national ecosystem extent accounts have adopted design-based approaches to area estimation. We suggest that design-based area estimation has a lot to offer for the SEEA EA community, and we recommend that this approach be explored and integrated in the relevant guideline documents. Nevertheless, we acknowledge that pixel counting may offer advantages compared to design-based methods particularly in contexts without human resource to perform photointerpretation. Further research is required to better elucidate the trade-offs between pixel counting and design-based approaches in the

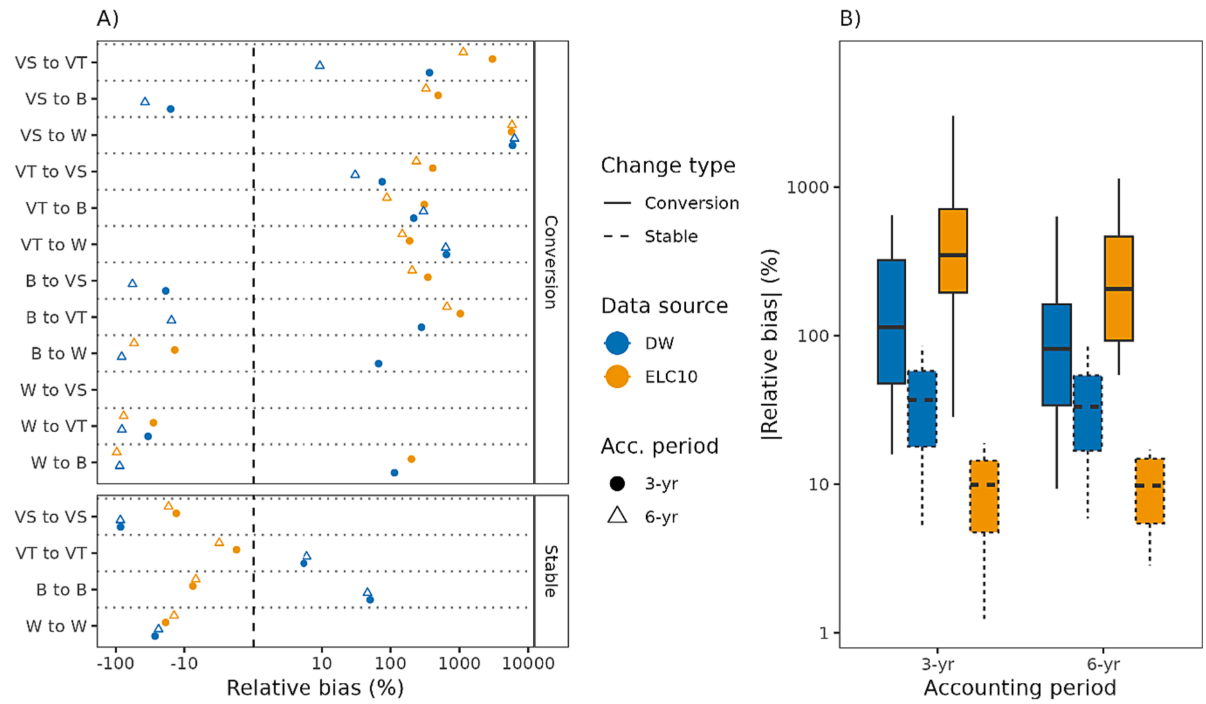

Fig. 5. Variation in pixel counting bias between ecosystem type change classes (A) for two accounting periods including 3-year (average of 2015–2018 and 2018–2021) and 6-year (2015–2021), and satellite-based maps (DW = Dynamic World; ELC10 = locally trained ELC10 map). Relative bias is defined as the percentage difference in area estimated between pixel counting and design-based area estimation. A pseudo-log transformation is applied to the x-axis mapping numbers to a signed logarithmic scale with a smooth transition to linear scale around 0. The distribution of absolute values for relative bias are shown in B with box-and-whisker plots.

Table 3
Ecosystem extent accounting table for a 4-class ecosystem typology over three accounting periods. Values are expressed in hectares. Extents derived from traditional pixel counting (PC) are adjacent estimates from design-based (DB) methods which are accompanied by ± 95 % CI. * PC areas which are significantly biased by exceeding the 95 % CI of the DB estimates. †DB conversion estimates which reflect significant changes in ET extent (also highlighted in bold).

		Water		Bare		Vegetation tall		Vegetation short	
		PC	DB	PC	DB	PC	DB	PC	DB
		Dynamic World							
2015 to 2018	Opening	4818*	6628 ± 565	11599*	8025 ± 640	30755*	28416 ± 954	903*	5006 ± 678
	Change	132*	−15 ± 94	868*	**234 ± 166** †	−1137*	**−281 ± 186** †	137	62 ± 163
	Closing	4951*	6613 ± 541	12467*	8259 ± 781	29618*	28135 ± 852	1040*	5069 ± 663
2018 to 2021	Opening	4951*	6749 ± 513	12467*	7895 ± 650	29618*	28014 ± 849	1040*	5417 ± 654
	Change	60*	−36 ± 78	−276*	−16 ± 120	165*	−34 ± 138	51	86 ± 124
	Closing	5011*	6713 ± 474	12190*	7879 ± 648	29783*	27981 ± 860	1091*	5503 ± 684
2015 to 2021	Opening	4818*	6547 ± 624	11599*	8074 ± 723	30755*	28431 ± 967	903*	5024 ± 720
	Change	193*	**−160 ± 146** †	592*	**282 ± 212** †	−972*	**−223 ± 217** †	188	101 ± 198
	Closing	5011*	6387 ± 540	12190*	8356 ± 843	29783*	28208 ± 932	1091*	5125 ± 718
		Locally-trained ELC10							
2015 to 2018	Opening	5007*	6249 ± 418	7445	7299 ± 573	29,961	29357 ± 741	5662	5171 ± 579
	Change	40	13 ± 88	−136*	142 ± 163	−168	**−250 ± 171** †	264*	94 ± 157
	Closing	5048*	6262 ± 419	7309	7441 ± 640	29792*	29108 ± 667	5926*	5265 ± 585
2018 to 2021	Opening	5048*	6151 ± 351	7309	7392 ± 508	29,792	29387 ± 644	5926*	5146 ± 536
	Change	62*	−47 ± 78	104*	−55 ± 116	357*	−62 ± 143	−523*	**164 ± 126** †
	Closing	5109*	6104 ± 304	7413	7337 ± 477	30150*	29325 ± 665	5403	5310 ± 593
2015 to 2021	Opening	5007*	6064 ± 539	7445	7253 ± 610	29,961	29872 ± 754	5662*	4886 ± 577
	Change	102*	**−156 ± 149** †	−32*	178 ± 195	189*	**−252 ± 212** †	−259*	**230 ± 186** †
	Closing	5109*	5908 ± 459	7413	7432 ± 668	30,150	29619 ± 742	5403	5116 ± 610

context of ecosystem accounting.

4.1. Factors influencing bias and uncertainty in extent accounts

Bias and uncertainty in ecosystem extent accounts are particularly large for ET conversion classes. Given that the purpose of extent accounts is to track changes in ecosystem extent over time, it is vital to ensure that the area estimates of change are credible. The higher uncertainties associated with ET conversion classes, relative to stable classes, in our study are consistent with other studies described in the broader remote sensing literature (Gallego, 2004; Kleinewillinghöfer et al., 2022). The uncertainty in changes in vegetation classes – in our case short to tall vegetation – can be particularly high. This may be a result of the spectral similarity between vegetation classes. This is

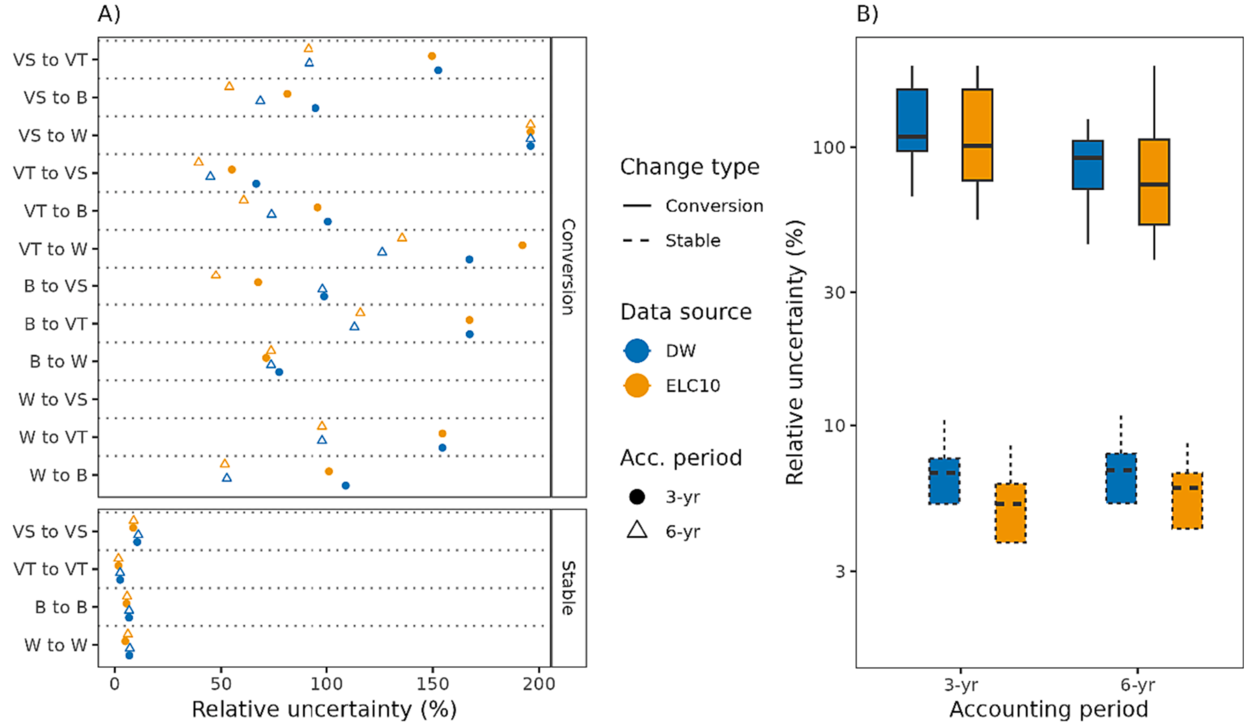

Fig. 6. Variation in uncertainty from design-based area estimation between ecosystem type change classes (A) for two accounting periods including 3-year (average of 2015–2018 and 2018–2021) and 6-year (2015–2021), and satellite-based maps (DW = Dynamic World; ELC10 = locally trained ELC10 map). Relative uncertainty was calculated defined by 95 % confidence intervals as a percentage of the class-specific area estimate. The distribution of relative uncertainty are shown in B with box-and-whisker plots.

Table 4
Validation and accuracy assessment results for Dynamic World (DW) and custom ELC10 maps of ecosystem extent changes between 2015 and 2018, 2018–2021, and 2015–2021. Accuracies are averaged over the accounting periods and presented with 95% confidence intervals in parentheses. The overall accuracies for DW and ELC10 were 75.2% (1.6%) and 81.2% (1.2%), respectively.

	User's accuracy (%)		Producer's accuracy (%)	
	DW	ELC10	DW	ELC10
W to W	83.9 (3.6)	92.4 (2.4)	62.2 (4.1)	76.3 (3.8)
W to B	6.7 (6)	7.6 (7.5)	1.8 (2.2)	0.5 (0.6)
W to VT	–	–	–	–
W to VS	–	–	–	–
B to W	0.9 (1.7)	1.6 (3.1)	2.8 (5.3)	1.7 (3.3)
B to B	57.6 (4)	76.4 (4.4)	85.8 (2.7)	70.9 (3)
B to VT	–	–	–	–
B to VS	22.8 (13.5)	8.9 (6.9)	13.2 (15.6)	32.3 (24.2)
VT to W	0.5 (0.9)	–	3.3 (7.1)	–
VT to B	0.5 (0.9)	4.1 (7)	1.9 (3.8)	11 (20)
VT to VT	83.9 (1.9)	92.6 (1.4)	88.5 (1.3)	90.9 (0.9)
VT to VS	13.7 (8.9)	14.9 (8.1)	21.7 (17)	65.1 (25.4)
VS to W	1.1 (2.1)	1.1 (2.2)	100 (0)	100 (0)
VS to B	2.2 (3.6)	1.9 (2.9)	2.4 (4)	8.2 (12.1)
VS to VT	1.3 (2.6)	0.9 (1.7)	1.2 (2.7)	10.2 (19.4)
VS to VS	78 (7.9)	62.8 (4.8)	11.4 (1.6)	53.7 (4.5)

reflected in the relatively low user's and producer's accuracies in agreement between interpreters of orthophoto imagery during our post-stratified sampling (Table S1). Additionally, the spectral signatures can be further complicated by inter-annual variability due to changes in rainfall and temperature which influence vegetation structure and vigour. For instance, the spectral signature of a forest during a drought year may converge with the spectral signature of shrubs or grass during a normal year. Therefore, AI models that classify satellite imagery may struggle to differentiate changes in spectrally similar classes between years. This highlights the importance of over-sampling ET conversion classes in extent maps during the stratified sampling process (Olofsson et al., 2013). By allocating more samples to strata which are expected to have low map accuracies, one can reduce the standard error and thus 95 % confidence intervals around the resulting area estimates up to a point.

Uncertainty in area estimates is largely a product of the accuracy of the input satellite-based map – in addition to the sampling effort applied during the post-stratified area estimation, and human classification error in the stratified sample. We found that custom satellite-based maps tailored to local conditions are more accurate than global maps and therefore reduced uncertainty in extent accounts. It is possible that alternative global satellite-based maps such as the European Space Agency's WorldCover map (Zanaga et al., 2021) produce extent account uncertainties that are comparable to our custom ELC10 map. WorldCover has the highest map accuracy amongst global 10 m resolution land cover maps when considering a 100 m^2 minimum mapping unit (Venter et al., 2022). However, we could not test its performance in extent accounts given that it is only available since 2020. Nevertheless, regardless of the input data source, design-based methods can be used to account for inaccuracies in satellite-based maps and, given enough post-stratified sampling, one can still use global maps to produce unbiased area estimates. In principle, the less accurate the extent map, the more sampling effort needs to be applied to constrain uncertainty in area estimates. However, the confidence intervals will never be smaller than the inherent error in the measurement, regardless of how many samples are assigned to a given ET.

The length of the accounting period is another factor determining the uncertainty in satellite-based extent accounts. We found that shorter accounting periods (3-yr vs 6-yr) produced greater bias and uncertainty in extent accounts. Satellite spectral signatures compared over shorter periods are more vulnerable to the effects of inter-annual changes in weather conditions on vegetation, which may produce artifacts in spectral signatures. Therefore, AI models are more likely to mis-classify ecosystem change over shorter periods. In addition, some ecosystem change processes take several years to develop (e.g. the succession from low to tall vegetation) or are less likely to occur (e.g. bare to tall vegetation). Therefore, shorter accounting periods may result in artifacts or misclassification of ecosystem change types. For instance, gradual ecosystem changes such as succession from grassland into shrubland ("woody plant encroachment") occur over longer time periods and produce spectral time series that are not easily classified by AI models compared to, for example, abrupt deforestation. Our results confirm the expectation that shorter accounting periods require greater sampling effort in order to maintain similar 95 % confidence intervals (i.e. uncertainty) around extent change estimates.

Irrespective of input data type or accounting period length, our design-based area estimation for net ecosystem extent changes were mostly too uncertain to conclude anything about the direction of magnitude of change. The main exceptions were net increases in bare cover, and losses in tall vegetation. The former is probably driven by urban development and the latter by forest clear cutting. Both cases introduce sudden and large changes to the spectral signature detected by satellites and are thus more confidently mapped as change. Nevertheless, the fact that the majority of extent changes were undetectable even for a simple ecosystem typology with four ETs highlights the difficulty of performing extent accounts with satellite-derived maps. This finding points to a trade-off between the thematic resolution of extent accounts and the certainty with which one can quantify changes in extent. The SEEA EA guidelines currently identify the IUCN Global Ecosystem Typology level 3 units (containing 108 distinct classes) as the desired ecosystem typology for extent accounts. Our results, based on a 4-class typology, question how realistic it is to expect countries to accurately account for changes in ecosystem extent at level 3 of the typology. Results from field-based ET mapping have identified fine-grained and complex typologies with a high number of classes as one of three major sources of error in the assignment of land-cover types (Naas et al., 2023). Therefore, although simplified ET typologies may not provide enough detail for some use-cases of ecosystem accounts (e.g. municipal managers needing to account for carbon accumulation in wetlands vs heathlands, or forest managers needing to distinguish deciduous vs coniferous forests), more complex typologies may result in extent accounts with levels of uncertainty that are too high for reliable decision making.

In order to address the challenge of class inseparability using remote sensing alone, practitioners may need to consider simplified ecosystem typologies. For instance, one may be to report baseline ecosystem extents for the full IUCN level 3 ecosystem typology, but report extent changes only at a higher level in the classification hierarchy. Since many ecosystem properties and processes are not detectable by remote sensing, there is a trade-off between typologies tailor-made to describe ecological relevant properties, and typologies tailor-made to be consistently detected and classified by remote sensing methods. The simplification of change typologies has knock-on implications for estimating changes in ecosystem service accounts, where some ES models purposed for accounting may use more disaggregated landcover (Buchhorn et al., 2022) than the number of ET for which we can afford to calculate design-based uncertainty estimates. A key challenge will be understanding the strengths and limitations of remote sensing-based indices when making ecological interpretation of changes in ecosystem extent, and for the purpose of estimating changes in ecosystem services.

4.2. Areas for future research

Perhaps the greatest challenge with applying design-based methods to ecosystem accounting is that the resulting ecosystem extent tables violates an important accounting identity: the closing stocks of a "T1-T2" account should match the opening stock of a "T2-T3" account. While pixel counting conforms to this identity (see Table 3), design-based methods do not because each opening and closing stock is estimated with a unique survey with different uncertainty estimates. This begs the question of whether the accounting identities should be enforced with some form of post-processing of the design-based estimates, or whether a violation of accounting identities should acceptable given that SEEA EA necessitates modelling, estimation and measurement – all of which have inherent uncertainty. Further research and guidance in this direction is needed.

There are several limitations to our study which also serve as avenues for further research. Firstly, we did not consider the effect of spatial scale on bias and uncertainty. The size (scale) of the basic spatial units (also referred to as minimum mapping units) has significant effects on the area estimates derived from design-based estimation (Gallego, 2004). For instance, an accounting unit of 100 m^2 might identify clumps of trees in a city as forest, whereas the same clump of trees might be considered as urban in an accounting unit of 10000 m^2 which considers a mosaicked approach to ET definition. Like the accounting unit, the scale of the accounting area over which ecosystem extents are estimated can affect the bias and uncertainty in extent accounts. Here we used a single accounting area, Oslo municipality, however further research is needed to quantify the effects of spatial scale on uncertainty, and whether uncertainty estimates from one accounting area can be generalized to another. The expectation is that municipalities with similar land use mosaics/ ET proportions will have similar uncertainty estimates. Secondly, our study took place at municipal level and it is unclear whether our findings are applicable to the national level – the level at which Eurostat requires ecosystem accounts to be reported. Thirdly, we considered a 3-yr and 6-yr accounting period, the latter recommended by the draft EUROSTAT regulation. Given our results, we expect uncertainties in change estimates to be even higher for 1-yr accounting periods used in national accounts. Fourthly, we did not account for the uncertainty in sample photointerpretation (Table S1) and how this error propagates to area estimates in extent accounts. Nor did we compare validate our photointerpretation ET labels with field-based ground truths which are known to exhibit discrepancies (Naas et al., 2023). Fifthly, because our focus is on satellite-based maps derived using machine learning, we did not consider including manually-mapped ET products such as CLC, which have single year status maps, but also 3-year change maps. We expect that using the CLC change product would produce more certain extent accounts than using the difference (intersection) of the two consecutive status layers. However, the error rates of CLC change maps exceed 15 % in many cases (Büttner, 2014) and may be as uncertain as the satellite-based maps presented here. The comparison of manually-mapped ET changes versus satellite-derived machine learning change maps in terms of accuracy and resource requirements for producing such maps requires further research.

4.3. Implications for ecosystem service accounting

Our analysis was restricted in scope to ecosystem extent and further research is needed on the wider implications for ecosystem accounting as a standard for generating national statistics. Since SEEA EA is specifically designed as a cascade, extent accounts outputs serve as inputs to condition accounts, services accounts and asset accounts, but the implications of this for the reliability of ecosystem accounting as a standard practice for supporting policy are unclear (United Nations, 2021). Urgency is needed now that SEEA EA is an international statistical standard and is being applied to support regulatory frameworks such as the proposed Nature Restoration Law (Maes et al., 2023) and EU Common Agricultural Policy (Grondard et al., 2021).

SEEA EA accounting standards define extent accounts as "stock" accounts with opening and closing extents, additions and reductions (see e.g. Figure and Table 2.2 (United Nations, 2021), pp.31–32). Our design-based approach highlights the uncertainty in using this standardised approach. The condition accounts are also based on opening and closing entries, with net change computed as the difference. The biophysical and monetary ecosystem services supply-use accounts are flow accounts (Table 2.4 (United Nations, 2021). For ecosystem services where flows are directly recorded (e.g. crop production), any change in cropland in the accounting period will be implicitly integrated in the annually recorded crop production. However, the majority of ecosystem services are modelled based on units of ecosystem extent (sometimes including condition variables), with a convention being to use the opening state as the basis for the calculation (e.g. 6.111 "The carbon retention component of the service is quantified by recording the stock of carbon retained in ecosystems at the beginning of the accounting period (i.e., the opening stock)."). A change in ecosystem service flow in this case is detected between two accounting periods, rather than within the accounting period as with extent and condition. Given the uncertainty in the underlying extent accounts documented in this paper, change in ES supply will be harder to detect.

Valuing total ecosystem service flows versus incremental changes in flows raises some methodological questions that are addressed in the SEEA EA standard in general, but without operational recommendations. "Monetary values are of most applicability in analysing changes that are marginal, i.e., concerning the effects of relatively small changes in stocks or flows of a particular asset, good or service. For example, analysing the changes in agricultural production associated with changes in soil fertility. When there is a requirement to analyse large, non-marginal changes, such as the permanent loss of a water resource, analysis should incorporate the assessment of physical changes in stocks in relation to appropriate thresholds" (p.178 (United Nations, 2021)). An operational recommendation from our paper could therefore be that when statistically significant changes in ecosystem extent are identified in the extent account, an assessment of the validity of accounting prices used to value ecosystem services based on the opening extent should be explicitly assessed and reported with the monetary ecosystem service supply-use tables.

5. Conclusion and recommendations

Although satellite-based maps of ecosystem extent may be well-suited to indicative ecosystem accounting, we argue that rigorous accuracy assessment is needed to maintain the credibility of SEEA EA. Without addressing uncertainty, decision-makers risk mismanaging valuable and irreplaceable natural assets. Using Oslo municipality as a case study, we found that without design-based area estimation, satellite-based extent maps lead to biases in extent accounts, even with a simple ecosystem typology with four classes. In general, uncertainties and limitations in ecosystem accounts should be communicated to users and policy makers. Specifically, we suggest some recommendations for the SEEA EA community below:

- Pixel counting from satellite-based maps for generating extent accounts can lead to biased area estimates and should therefore be discouraged unless there are insufficient resources to support design-based area estimation from a photointerpretation survey.
- Instead of a dedicated survey (stratified with the satellite-based ET map), it is also possible to re-use an existing reference data set – typically a survey-based dataset, that was created for a similar purpose using a compatible typology (i.e. one with a reliable cross-walk towards the target ecosystem typology). In many countries there are national reference datasets for statistical area estimation (e.g. the LUCAS survey in the EU – d'Andrimont et al., 2020), and often the validation dataset of the ET map itself can be considered given some caveats (Stehman and Foody, 2019).
- In cases where pixel counting takes place, this should be explicitly communicated and the potential for biased area estimates acknowledged. Practitioners should not use the classification accuracy of the AI model or the resulting map to inspire confidence in the resulting ecosystem extent estimates. Map accuracy metrics like those recommended in the SEEA EA guidelines (United Nations, 2022) including users, producers, and overall accuracy do not necessarily mean the area estimates derived from the map will be unbiased or certain.
- Longer accounting periods offer more precise extent change estimates and may be preferred depending on the accuracy requirement of the ecosystem account end-users.
- Simplified ecosystem typologies which are possible to classify reliably with remote sensing, should be considered when estimating ecosystem changes with satellite-based extent maps.
- Satellite-based extent maps generated from locally-trained AI models should be evaluated as alternatives to global land cover maps because they may produce less uncertainty in the resulting extent accounts.
- The compounding effect of uncertainty in extent accounts down the ecosystem accounting cascade is a useful and urgent avenue for further research. For example, if uncertainty analysis is to be applied to ecosystem service models being proposed for ecosystem accounting, they may need to be simplified to match the limitation in number of ecosystem types that can be subjected to an ecosystem account 'uncertainty audit'.
- Standards for quantifying, assessing and disclosing uncertainty in ecosystem accounting are needed to complement the existing guidance in the SEEA EA statistical standard. An 'uncertainty audit' for ecosystem accounting could play an analogous role to that of an 'information systems audit' vis a vis corporate financial accounts.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data and code to support this paper will be made available here: https://github.com/zanderVenter/ecosystemExtentUncertainty

Acknowledgements

We acknowledge funding from the Research Council of Norway through the projects MAREA (contract #326976), SPARE (contract #326641), and EcoGaps (contract #320042). Thanks to Anders Kolstad for assisting with the photo interpretation of stratified samples. Thanks to Stephen Stehman for advice and guidance on design-based area estimation methods.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoser.2024.101599.

References

Arévalo, P., Olofsson, P., Woodcock, C.E., 2020. Continuous monitoring of land change activities and post-disturbance dynamics from Landsat time series: A test methodology for REDD+ reporting. Remote Sens. Environ. Time Series Analysis with High Spatial Resolution Imagery 238, 111051. https://doi.org/10.1016/j.rse.2019.01.013.

Bagstad, K.J., Ingram, J.C., Shapiro, C.D., La Notte, A., Maes, J., Vallecillo, S., Casey, C. F., Glynn, P.D., Heris, M.P., Johnson, J.A., Lauer, C., Matuszak, J., Oleson, K.L.L., Posner, S.M., Rhodes, C., Voigt, B., 2021. Lessons learned from development of natural capital accounts in the United States and European Union. Ecosyst. Serv. 52, 101359 https://doi.org/10.1016/j.ecoser.2021.101359.

Barton, D.N., Stange, E., Blumentrath, S., Vågnes Traaholt, N., 2015. Economic valuation of ecosystem services for policy. A pilot study on green infrastructure in Oslo.

Barton D.N., Obst C., Day B., Caparrós A., Dadvand P., Fenichel E., Havinga I., Hein L., McPhearson T., Randrup T., Zulian G. (2019). Discussion paper 10: Recreation services from ecosystems. Paper submitted to the Expert Meeting on Advancing the Measurement of Ecosystem Services for Ecosystem Accounting, New York, 22-24 January 2019 and subsequently revised. Version of 25 March 2019. Available at: https://seea.un.org/events/expert-meeting-advancing-measurement-ecosystemservices-ecosystem-accounting.

Bebbington, J., Unerman, J., 2018. Achieving the United Nations Sustainable Development Goals: An enabling role for accounting research. Account. Audit. Account. J. 31, 2–24. https://doi.org/10.1108/AAAJ-05-2017-2929.

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Sci. Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4.

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., Smets, B., 2020. Copernicus global land cover layers—collection 2. Remote Sens. (Basel) 12, 1044. https://doi.org/10.3390/rs12061044.

Buchhorn, M., Smets, B., Danckaert, T., van Loo, M., Broekx, S., Peelaerts, W., 2022. Establishing a reference tool for ecosystem accounting in Europe, based on the INCA methodology. One Ecosystem 7, e85389. https://doi.org/10.3897/oneeco.7.e85389.

Büttner, G., 2014. CORINE land cover and land cover change products, in: Land Use and Land Cover Mapping in Europe. Springer, pp. 55–74.

Cimburova, Z., Barton, D.N., 2020. The potential of geospatial analysis and Bayesian networks to enable i-Tree Eco assessment of existing tree inventories. Urban For. Urban Green. 55, 126801.

Cord, A.F., Brauman, K.A., Chaplin-Kramer, R., Huth, A., Ziv, G., Seppelt, R., 2017. Priorities to advance monitoring of ecosystem services using earth observation. Trends Ecol. Evol. 32, 416–428. https://doi.org/10.1016/j.tree.2017.03.003.

Costa, H., 2022. mapaccuracy: Unbiased Thematic Map Accuracy and Area.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260. https://doi.org/10.1038/387253a0.

Cowie, A.L., Orr, B.J., Sanchez, V.M.C., Chasek, P., Crossman, N.D., Erlewein, A., Louwagie, G., Maron, M., Metternicht, G.I., Minelli, S., 2018. Land in balance: The scientific conceptual framework for Land Degradation Neutrality. Environ. Sci. Policy 79, 25–35.

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H.I., Joebges, C., Lemoine, G., 2020. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. Sci. Data 7, 352.

de Araujo Barbosa, C.C., Atkinson, P.M., Dearing, J.A., 2015. Remote sensing of ecosystem services: A systematic review. Ecol. Ind. 52, 430–443. https://doi.org/10.1016/j.ecolind.2015.01.007.

Edens, B., Maes, J., Hein, L., Obst, C., Siikamaki, J., Schenau, S., Javorsek, M., Chow, J., Chan, J.Y., Steurer, A., Alfieri, A., 2022. Establishing the SEEA Ecosystem Accounting as a global standard. Ecosyst. Serv. 54, 101413 https://doi.org/10.1016/j.ecoser.2022.101413.

Eilifsen, A., Hamilton, E.L., Messier Jr, W.F., 2021. The importance of quantifying uncertainty: Examining the effects of quantitative sensitivity analysis and audit materiality disclosures on investors' judgments and decisions. Acc. Organ. Soc. 90, 101169.

European Commission, 2022. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Regulation (EU) No 691/2011 as regards introducing new environmental economic accounts modules. Brussels.

Foody, G.M., 2013. Ground reference data error and the mis-estimation of the area of land cover change as a function of its abundance. Remote Sens. Lett. 4, 783–792. https://doi.org/10.1080/2150704X.2013.798708.

Foody, G.M., 2015. Valuing map validation: The need for rigorous land cover map accuracy assessment in economic valuations of ecosystem services. Ecol. Econ. 111, 23–28. https://doi.org/10.1016/j.ecolecon.2015.01.003.

Foody, G.M., Pal, M., Rocchini, D., Garzon-Lopez, C.X., Bastin, L., 2016. The sensitivity of mapping methods to reference data quality: training supervised image classifications with imperfect reference data. ISPRS Int. J. Geo Inf. 5, 199. https://doi.org/10.3390/ijgi5110199.

Gallego, F.J., 2004. Remote sensing and land cover area estimation. Int. J. Remote Sens. 25, 3019–3047. https://doi.org/10.1080/01431160310001619607.

Garnåsjordet, P.A., Steinnes, M., Cimburova, Z., Nowell, M., Barton, D.N., Aslaksen, I., 2021. Urban green. Integrating ecosystem extent and condition data in urban ecosystem accounts. Examples from the Oslo region. Stat. J. IAOS 37, 1247–1274.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. Big Remotely Sensed Data: Tools Appl. Experiences 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Grammatikopoulou, I., Badura, T., Johnston, R.J., Barton, D.N., Ferrini, S., Schaafsma, M., La Notte, A., 2023. Value transfer in ecosystem accounting applications. J. Environ. Manage. 326, 116784 https://doi.org/10.1016/j.jenvman.2022.116784.

Grondard, N., Hein, L., Van Bussel, L.G.J., 2021. Ecosystem accounting to support the Common Agricultural Policy. Ecol. Ind. 131, 108157 https://doi.org/10.1016/j.ecolind.2021.108157.

Hanssen, F., Barton, D.N., Venter, Z.S., Nowell, M.S., Cimburova, Z., 2021. Utilizing LiDAR data to map tree canopy for urban ecosystem extent and condition accounts in Oslo. Ecol. Ind. 130, 108007.

Hein, L., Remme, R.P., Schenau, S., Bogaart, P.W., Lof, M.E., Horlings, E., 2020. Ecosystem accounting in the Netherlands. Ecosyst. Serv. 44, 101118 https://doi.org/10.1016/j.ecoser.2020.101118.

Heris, M., Bagstad, K.J., Rhodes, C., Troy, A., Middel, A., Hopkins, K.G., Matuszak, J., 2021. Piloting urban ecosystem accounting for the United States. Ecosyst. Serv. 48, 101226 https://doi.org/10.1016/j.ecoser.2020.101226.

Keith, D.A., Ferrer-Paris, J.R., Nicholson, E., Bishop, M.J., Polidoro, B.A., Ramirez-Llodra, E., Tozer, M.G., Nel, J.L., Mac Nally, R., Gregr, E.J., Watermeyer, K.E., Essl, F., Faber-Langendoen, D., Franklin, J., Lehmann, C.E.R., Etter, A., Roux, D.J., Stark, J.S., Rowland, J.A., Brummitt, N.A., Fernandez-Arcaya, U.C., Suthers, I.M., Wiser, S.K., Donohue, I., Jackson, L.J., Pennington, R.T., Iliffe, T.M., Gerovasileiou, V., Giller, P., Robson, B.J., Pettorelli, N., Andrade, A., Lindgaard, A., Tahvanainen, T., Terauds, A., Chadwick, M.A., Murray, N.J., Moat, J., Pliscoff, P., Zager, I., Kingsford, R.T., 2022. A function-based typology for Earth's ecosystems. Nature 610, 513–518. https://doi.org/10.1038/s41586-022-05318-4.

Kjærås, K., 2023. The politics of urban densification in Oslo. Urban Stud. https://doi.org/10.1177/00420980231178190.

Kleinewillinghöfer, L., Olofsson, P., Pebesma, E., Meyer, H., Buck, O., Haub, C., Eiselt, B., 2022. Unbiased area estimation using Copernicus high resolution layers and reference data. Remote Sens. (Basel) 14, 4903. https://doi.org/10.3390/rs14194903.

Maes, J., Bruzón, A.G., Barredo, J.I., Vallecillo, S., Vogt, P., Rivero, I.M., Santos-Martín, F., 2023. Accounting for forest condition in Europe based on an international statistical standard. Nat. Commun. 14, 3723. https://doi.org/10.1038/s41467-023-39434-0.

Mayhew, B.W., Schatzberg, J.W., Sevcik, G.R., 2001. The effect of accounting uncertainty and auditor reputation on auditor objectivity. Audit. J. Pract. Theory 20, 49–70.

Mitchell, A.L., Rosenqvist, A., Mora, B., 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. Carbon Balance Manag. 12, 1–22.

Naas, A.E., Halvorsen, R., Horvath, P., Wollan, A.K., Bratli, H., Brynildsrud, K., Finne, E. A., Keetz, L.T., Lieungh, E., Olson, C., Simensen, T., Skarpaas, O., Tandstad, H.R., Torma, M., Værland, E.S., Bryn, A., 2023. What explains inconsistencies in field-based ecosystem mapping? Appl. Veg. Sci. 26, e12715. https://doi.org/10.1111/avsc.12715.

Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sens. Environ. 129, 122–131.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015.

Oslo kommune, 2018. Grøntregnskap: En Måling Av grønnstruktur i Oslos Byggesone. Plan Og Bygningsetaten.

Petersen, J.-E., Mancosu, E., King, S., 2022. Ecosystem extent accounts for Europe. Ecosyst. Serv. 57, 101457 https://doi.org/10.1016/j.ecoser.2022.101457.

Radoux, J., Bogaert, P., 2020. About the pitfall of erroneous validation data in the estimation of confusion matrices. Remote Sens. (Basel) 12, 4128. https://doi.org/10.3390/rs12244128.

Schägner, J.P., Brander, L., Maes, J., Hartje, V., 2013. Mapping ecosystem services' values: Current practice and future prospects. Ecosyst. Services, Special Issue Mapping Model. Ecosyst. Serv. 4, 33–46. https://doi.org/10.1016/j.ecoser.2013.02.003.

Stehman, S.V., 2013. Estimating area from an accuracy assessment error matrix. Remote Sens. Environ. 132, 202–211. https://doi.org/10.1016/j.rse.2013.01.016.

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199 https://doi.org/10.1016/j.rse.2019.05.018.

United Nations, 2022. Guidelines on Biophysical Modelling for Ecosystem Accounting. United Nations Department of Economic and Social Affairs Statistics Division, New York.

United Nations, 2021. System of Environmental-Economic Accounting—Ecosystem Accounting (SEEA EA). White cover publication.

Venter, Z.S., Krog, N.H., Barton, D.N., 2020. Linking green infrastructure to urban heat and human health risk mitigation in Oslo, Norway. Sci. Total Environ. 709, 136193.

Venter, Z.S., Barton, D.N., Martinez-Izquierdo, L., Langemeyer, J., Baró, F., McPhearson, T., 2021. Interactive spatial planning of urban green infrastructure–Retrofitting green roofs where ecosystem services are most needed in Oslo. Ecosyst. Serv. 50, 101314.

Venter, Z.S., Barton, D.N., Chakraborty, T., Simensen, T., Singh, G., 2022. Global 10 m land use land cover datasets: A comparison of dynamic world, world cover and Esri land cover. Remote Sens. (Basel) 14, 4101. https://doi.org/10.3390/rs14164101.

Venter, Z.S., Sydenham, M.A.K., 2021. Continental-scale land cover mapping at 10 m resolution over Europe (ELC10). Remote Sens. (Basel) 13, 2301. https://doi.org/10.3390/rs13122301.

Vysna, V., Maes, J., Petersen, J.-E., La Notte, A., Vallecillo, S., Aizpurua, N., Ivits-Wasser, E., Teller, A., 2021. Accounting for Ecosystems and Their Services in the European Union (INCA): Final Report from Phase II of the INCA Project Aiming to Develop a Pilot for an Integrated System of Ecosystem Accounts for the EU. Publications office of the European Union.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., 2021. ESA WorldCover 10 m 2020 v100. Zenodo.