



Are remote sensing-based crop type classifications suitable for calculating a landscape heterogeneity metric? A data-fitness-for-purpose assessment

Jannes Säurich ^{a,g} , Marcel Schwieder ^{b,c}, Sebastian Preidl ^{d,e}, Florian Beyer ^a, Markus Möller ^{f,*}

^a Julius Kühn Institute (JKI), Federal Research Centre for Cultivated Plants, Institute for Crop and Soil Science, Bundesallee 58, 38116 Braunschweig, Germany

^b Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany

^c Earth Observation Lab, Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

^d Julius Kühn Institute (JKI), Federal Research Centre for Cultivated Plants, Institute of Forest Protection, Erwin-Baur-Straße 27, 06484 Quedlinburg, Germany

^e Helmholtz Centre for Environmental Research GmbH – UFZ, Department of Remote Sensing, Permoserstraße 15, 04318 Leipzig, Germany

^f Julius Kühn Institute (JKI), Federal Research Centre for Cultivated Plants, Department of Digitalisation and Artificial Intelligence, Stahnsdorfer Damm 81, 14532 Kleinmachnow, Germany

^g Thünen Institute of Climate-Smart Agriculture, Bundesallee 65, 38116 Braunschweig, Germany¹

ARTICLE INFO

Dataset link: <https://zenodo.org/records/17752848>, <https://zenodo.org/records/17753964>

Keywords:

Land use data
Integrated Administration and Control (IACS)
Crop type classification
Data quality metrics
Data-fitness-for-purpose
Accuracy
Uncertainty
Land use Heterogeneity Metrics

ABSTRACT

Remote sensing products are widely used to assess biodiversity in German agricultural landscapes. Evaluating their suitability for specific purposes is crucial for reliable decision-making. This study develops and applies a problem-oriented data-fitness-for-purpose (DFFP) perspective on satellite-based crop type classifications (CTC) to derive land use heterogeneity metrics (LHM) as proxies for biodiversity.

Two nationwide CTC (SWD, PRE) products were compared with detailed administrative reference data (IACS) for Lower Saxony and Brandenburg in Germany (2017–2019). Class schemes and geometries were harmonised, and the Shannon Evenness Index (SEI) was calculated on 1 km² hexagons as an illustrative LHM. Regional accuracy metrics differed substantially between classification products and between federal states, with systematically lower accuracies in Brandenburg. High overall accuracies masked pronounced class imbalances and regional variations in performance.

A simple, spatially explicit classification uncertainty metric was computed at the field level and aggregated to the hexagon scale, demonstrating that local uncertainty information can be carried through a multi-step workflow. The resulting patterns revealed systematic differences between federal states, classification products, and reference data.

Our findings indicate that nationally reported accuracy measures are insufficient to evaluate local or regional suitability of CTCs for biodiversity-relevant LHMs. Spatially explicit, domain-specific uncertainty layers should become standard components of remote sensing-based biodiversity indicators. Integrating such layers within ISO 19157-1 data quality concepts, including “spatial uncertainty” and provenance-rich, FAIR geospatial workflows, will enhance the robustness and interpretability of landscape heterogeneity assessments.

1. Introduction

The demands on agriculture have steadily increased in recent decades. This is due to population growth, changing eating habits, and an increasing demand for fuel (Tilman et al., 2002; von Braun, 2007). These challenges have led to an intensification of agricultural land use, facilitated by a high degree of specialisation, technological progress, and the use of fertilisers and pesticides (Bindi and Olesen, 2011; Smith et al., 2016; Jänicke et al., 2022). In addition, altered climatic conditions, such as extreme weather conditions (e.g., heat, drought, or

heavy rain events), make the cultivation of agricultural land even more challenging (Bindi and Olesen, 2011; Auerswald and Menzel, 2021; Riedesel et al., 2023, 2024). Agricultural expansion and intensification lead to major land cover changes, reduce landscape complexity, and subsequently contribute to a decline in biodiversity (Asam et al., 2022; Smith et al., 2016; Roilo et al., 2024). This is reflected in the loss of crop species and pollinators and has a negative impact on soil quality (Smith et al., 2016) and crop yield (Tilman et al., 2002). Agricultural landscapes, which are characterised by the cultivation of

* Corresponding author.

E-mail addresses: jannes.saeurich@thuenen.de (J. Säurich), markus.moeller@julius-kuehn.de (M. Möller).

¹ Present address.

Glossary

<i>H</i>	Shannon Index
<i>OA</i>	Overall Accuracy
<i>PA</i>	Producer Accuracy
<i>SEI</i>	Shannon Evenness Index
<i>UA</i>	User Accuracy
BB	Brandenburg
CTC	Crop Type Classification
DFFP	Data-Fitness-For-Purpose
EBV	Essential Biodiversity Variables
IACS	Integrated Administration and Control System
LHM	Land use Heterogeneity Metric
LS	Lower Saxony
PRE	Preidl
SWD	Schwieder

various crop types with broad crop rotation (Oppermann et al., 2005), on small and varying field sizes (Fahrig et al., 2015; Merckx et al., 2009), and the presence of landscape elements such as hedges, rows of trees, individual trees, and shrubs (Uthes et al., 2020; Oppermann et al., 2005), are generally considered to be conducive to biodiversity.

Monitoring the current state and changes in biodiversity is therefore crucial to support an effective management and improvement of agricultural landscapes (FAOSTAT, 2022; Asam et al., 2022; Wolff et al., 2021; Brown et al., 2022). In recent decades, numerous biodiversity metrics – such as species richness or land cover/use diversity – have been developed to quantitatively measure specific aspects of agrobiodiversity. These metrics provide the raw data that underpin biodiversity indicators, which are standardised tools designed to communicate trends and inform decision-making (Duelli and Obrist, 2003; Feld et al., 2010; Heink and Kowarik, 2010; Billeter et al., 2008; Riitters et al., 1995; Uemaa et al., 2009, 2013). Increasingly, indicators are based on Essential Biodiversity Variables (EBV), a standardised set of key metrics used to quantify the rate and direction of biodiversity change over time and space (Pereira et al., 2013). For example, land cover and land use data, types of EBVs, are crucial for understanding landscape structure and broader biodiversity trends.

Various studies have applied EBVs to assess biodiversity dynamics (e.g., Kissling et al., 2018; Jetz et al., 2019; Hoban et al., 2022; Schmeller et al., 2017), using datasets such as Integrated Administration and Control System (IACS), which tracks agricultural land use and crop types in Europe. This system, central to the Common Agricultural Policy (CAP), offers detailed annual information about land management in EU member states (Leonhardt et al., 2024). These data are already used region-specific, such as the federal state of Brandenburg (BB) in Germany, for landscape analysis on hexagonal grids (Wolff et al., 2021). These case studies demonstrate the potential of combining EBVs with IACS data to derive Land use Heterogeneity Metrics (LHM). They are designed to capture the diversity and spatial arrangement of land use types (Tonetti et al., 2023) and can be used to monitor and manage biodiversity in agricultural landscapes. However, IACS data might not always be available, and accessibility already varies considerably between EU member states. In Germany, for example, data are managed by the federal states, each applying different standards for data management and dissemination (Jänicke et al., 2025). As a result, data cannot serve as a comprehensive national basis for obtaining LHMs to assess variety and variability in agricultural ecosystems.

With the availability of satellite-based earth observation data (Drusch et al., 2012; Wulder et al., 2016; Claverie et al., 2018; Shang and Zhu, 2019; Defourny et al., 2019), the possibilities of deriving EBVs for biodiversity monitoring have expanded, leading to satellite

remote sensing-EBVs (Pettorelli et al., 2016; Pereira et al., 2013). In addition, a series of nationwide and European Crop Type Classifications (CTCs) products have emerged in recent years (Preidl et al., 2020b; Blickensdörfer et al., 2022; Griffiths et al., 2019; Orynbaiqyzy et al., 2020; Asam et al., 2022; Ghassemi et al., 2024), which are considered important for the frequent, cost-effective and extensive assessment of biodiversity of agricultural land use (Pettorelli et al., 2016; Asam et al., 2022).

Although satellite remote sensing-EBVs have advantages in terms of spatial coverage, interannual resolution, or cost efficiency, their specific adequacy must be assessed (Vihervaara et al., 2017; Petutschnig et al., 2024). Assessment of the suitability of geodata, especially spatial quality, can be achieved by a Data-Fitness-For-Use (DFFU) or Data-Fitness-For-Purpose (DFFP) evaluation (Pôças et al., 2014). DFFU criteria describes the general suitability of a dataset for further use and thus refer to the intrinsic quality of the data such as completeness, accuracy, consistency, and timeliness ready to be used according to non-functional requirements (Yang et al., 2013; Lacagnina et al., 2022). DFFP criteria are more context-dependent and emphasise the suitability of data for a specific application or purpose according to the suitability of the user's functional requirements (Lacagnina et al., 2022).

There are DFFP approaches that aim to enable users to select the appropriate geodata for their specific needs. However, these approaches differ significantly in their perspectives, methodologies, and outputs. They range from practical toolkits and quantitative frameworks to standardised rating systems (e.g., Fischer et al., 2023; Höck et al., 2020; Wentz and Shimizu, 2018; Petutschnig et al., 2024). All of these approaches acknowledge that DFFP is not an intrinsic property of data, but rather is defined by the context and requirements of a specific end user or application. The approaches thus aim at a retrospective evaluation of data from the user's perspective.

Against this background, this study simulates a typical geodata usage situation by using a LHM derivation as example. Considering both the user's and producer's perspectives, we focus on thematic accuracy as a key indicator of trustworthiness, examining the following questions:

1. What differences exist between CTC thematic accuracy metrics at the regional level?
2. How do the CTC/IACS differences affect LHMs?
3. Is sufficient CTC thematic accuracy information available to enrich LHMs based in remote sensing products with data quality information?

To address the questions, two harmonised CTC products are compared with IACS data for the federal states of Lower Saxony (LS) and Brandenburg (BB) for the years 2017 to 2019. In the next step, the Shannon Evenness Index (*SEI*), as a LHM, is calculated using the CTC products and IACS data. Hexagons of 1 km² are used as spatial units to assess how different input data affect the calculation of LHMs.

2. Materials and methods

2.1. Study area

The two federal states of Lower Saxony (LS) and Brandenburg (BB) are located in the north of Germany. LS has an area of 47,710 km², of which 58% is classified as arable land (Table 1; Destatis, 2024). The coastal region of LS in the north is predominantly characterised by grassland areas in the form of meadows and mowed pastures, while agricultural areas are mainly dominated by the cultivation of maize in the north and winter cereals, like winter wheat, winter rye or winter barley, in the south. As both the topographic gradient increases towards the southeast and the continental influence increases, greater differences in seasonal temperature and a decrease in precipitation are observed in the south (Brandt et al., 2024). The average annual temperature in 2017 to 2019 for all of LS ranges from 10.0 to 10.7°C and

Table 1

Details of agricultural area of Lower Saxony (LS) and Brandenburg (BB). The total area and total arable land is based on Destatis (2024) (marked by *). Crop area, number of crop fields as well as mean, standard deviation (sd) and maximal value (max) of crop field area is calculated based on Integrated Administration and Control System (IACS) 2019 after harmonisation to the Crop Type Classifications (CTCs) by Schwieder (SWD) or Preidl (PRE) (Section 2.2.1, 2.2.2, 2.3.1 and 3.1).

	Lower Saxony (LS)	Brandenburg (BB)
Total area [km ²]*	47,710	29,654
Arable land (total) [km ²]*	27,536 (58%)	14,036 (47%)
Crop area (SWD) [km ²]	18,362 (38.5%)	9617 (32.4%)
Crop area (PRE) [km ²]	18,465 (38.7%)	9682 (32.6%)
Number of crop fields (SWD)	519,852	86,189
Number of crop fields (PRE)	532,912	88,263
mean (crop field area) (SWD) [km ²]	0.04	0.11
mean (crop field area) (PRE) [km ²]	0.04	0.11
sd (crop field area) (SWD) [km ²]	0.04	0.16
sd (crop field area) (PRE) [km ²]	0.04	0.16
max (crop field area) (SWD) [km ²]	1.29	2.25
max (crop field area) (PRE) [km ²]	1.81	2.25

the average annual precipitation varies from 512 to 890 mm (Deutscher Wetterdienst, 2018, 2020a,b).

BB is characterised by a dry climate and sandy soils. In addition to large grasslands, most agricultural land is dominated by winter rye, maize, winter wheat, winter barley, sunflowers, and winter rape. In 2017 to 2019, the average annual temperature ranged from 9.9 to 11.1°C and the average annual rainfall varied between 390 and 719 mm (Deutscher Wetterdienst, 2018, 2020a,b).

In total, BB covers an area of 29,654 km², which corresponds to 62% of the area of LS (47,710 km²). In addition, the number of fields in BB (~87,000) is considerably lower than in LS (~525,000). This difference is partly explained by the substantially larger average field size in BB of 0.11 km² compared to LS with 0.04 km². The standard deviation of field size in LS is approximately equal to the mean, whereas in BB it exceeds the mean. This indicates a high degree of field size heterogeneity in both federal states, particularly in BB (Table 1 and Section 3.1). The proportion of crop area is similar in both federal states (~38% in LS and ~32% in BB). Thus, the main differences between the two regions, besides the overall size difference and the lower number of fields for BB, are the clearly larger field sizes in BB. This can be attributed to the fact that small fields were merged into large agricultural production communities within the framework of socialist agricultural production cooperatives in the former German Democratic Republic (Wolz, 2013).

2.2. Data base

2.2.1. Integrated Administration and Control System (IACS)

IACS is a tool established by the European Union to manage the payment of agricultural subsidies in its Common Agricultural Policy (CAP) (Tóth and Kučas, 2016; Schneider et al., 2022). As part of the IACS framework, farmers themselves register the boundaries of their fields and the crop types grown on them within the Land Parcel Identification System (LPIS). This results in an annual dataset that shows the geometry and crop types. Due to the high level of class detail data density at field level and the annual repetition frequency, this is considered a very detailed reference dataset for the comparison of remote sensing based CTC products (Section 2.2.2). Only fields that are actually reported by farmers are included in LPIS. However, since subsidy payments are tied to reporting, it can be assumed that the vast majority of agricultural land is registered in the system. Schneider et al. (2025) point out that the quality of IACS data is not backed by official EU quality assurance. They revealed that the state and federal authorities in Germany assume empirical correctness levels above 95%. In Germany, the data are only freely available for a few federal states, so in this why this article is focused on LS and BB.

2.2.2. Crop Type Classifications (CTCs)

The CTCs of Blickensdörfer et al. (2022) and Schwieder et al. (2022) are based on the preliminary work of Griffiths et al. (2019) and were continued by Schwieder et al. (2024) (from now on called SWD). The approach itself is based on Sentinel-2 and Landsat-8 time series in combination with monthly Sentinel-1 composites (backscatter) and additional environmental data using a random forest classification model. The nationwide agricultural land cover map includes 25 land cover classes, containing the dominant crop types of Germany.

In addition, there is another full-coverage CTC for Germany based on the approach called *adaptable pixel-based compositing and classification* (APiC) introduced by Preidl et al. (2020b) (from now on PRE). The work is based on a pure consideration of Sentinel-2 data time series and a regionally pixel-specific dynamic composition to address the problem of varying cloud cover. The highly variable temporal composite time series also required the training of numerous random forest models. A total of 19 agricultural and five non-agricultural classes were considered.

Both CTC products are available as raster datasets with a resolution of 10 m² (SWD) or 20 m² (PRE) and are characterised by accuracy metrics based on confusion matrices. These include generally recognised metrics based on accuracy assessment protocols, such as the crop type-specific User Accuracy (*UA*), Producer Accuracy (*PA*), *F*₁-scores, Macro-*F*₁, Weighted-*F*₁, and Overall Accuracy (*OA*) (Section 2.3.3; Stehman and Foody, 2019; Stehman and Wickham, 2020).

2.2.3. Reference units

As the IACS data are only available on the federal state scale, the CTCs are tailored to the size of the federal states. The individual federal states LS and BB are considered separately and serve as a reference area for regional accuracy assessment.

The derived LHMs are reported using 1 km² hexagons (Perić et al., 2022). Hexagons allow for a complete coverage of a two-dimensional area to represent spatial continua. Compared to square cells, this offers advantages in the analysis of neighbourhood, movement, and connectivity and therefore serves as a typical reference unit in biodiversity analysis (Birch et al., 2007; Wolff et al., 2021). Hexagons with overlapping state boundaries were excluded to avoid reference units of less than 1 km². The individual hexagons represent the reference level for the local accuracy.

2.3. Methodology

Our analysis can be viewed as a two-stage process (Fig. 1). In deriving LHMs, operations are performed that affect the data quality of the final product:

1. Since the number and characteristics of the classes differ between IACS and the CTCs (Table 2), both the CTC and IACS subsets are subject to class harmonisation operations. The resulting raster information is then transferred to the field polygons (Section 2.3.1).
2. In the LHMs derivation procedure, the filtered and harmonised crop type classes are spatially aggregated to reference units (Section 2.3.2), which are represented here by hexagons (Section 2.2.3).

2.3.1. Field scale: filtering and harmonisation

The filtering and harmonisation process comprises three sequential steps: class selection, majority voting, and thematic aggregation.

Table 2

Crop type classes for the CTCs of SWD (left; Schwieder et al., 2022) and PRE (right; Preidl et al., 2020a) as well as the assignment of the IACS codes to the classes. All IACS classes that cannot be grouped into the classes are summarised as the extra class not classified (777). The grassland classes for both CTCs (below midrule) are removed during the processing. IACS codes with crop names can be taken from Table A.9 and Table A.10.

SWD			PRE		
Code	Name	IACS	Code	Name	IACS
1101	Winter wheat	112, 115	1	Winter wheat	112, 115, 156
1103	Winter rye	121	3	Winter rye	121
1102	Winter barley	131	4	Winter barley	131
			2	Spelt	114
1201	Spring barley	132	6	Spring barley	132
1202	Spring oat	143	7	Spring oat	143
			5	Spring wheat	113, 116, 157
1300	Maize	171, 172, 177, 410, 411, 412	8	Maize	411, 171, 172
1401	Potato	601, 602, 606	12	Potatoes	601, 602
1402	Sugar beet	603	13	Sugar beets	603
1501	Winter rapeseed	311	10	Rapeseed	311
1502	Sunflower	320			
1611	Peas	210, 240	9	Legumes	210, 211, 220, 230, 240
1612	Broad bean	220			
1613	Lupin	230			
1614	Soy	330			
1603	Vegetables	610, 613, 630, 632, 633, 634, 635, 637, 638, 707, 860	14	Strawberries	707
			11	Leeks	633
4003	Orchard	821, 824, 825, 826, 983	15	Stone fruits	821, 825, 826
4001	Grapevine	843, 844, 847	16	Vines	842, 843
4002	Hops	856	17	Hops	856
			18	Asparagus	860
777	not classified	all other IACS classes	777	not classified	all other IACS classes
200	Permanent grassland	441, 443, 451, 452, 453, 455, 459	19	Grassland	421, 422, 423, 424, 425, 441, 451, 452, 453, 454, 459
1602	Cultivated grassland	421, 422, 423, 424, 425, 433			

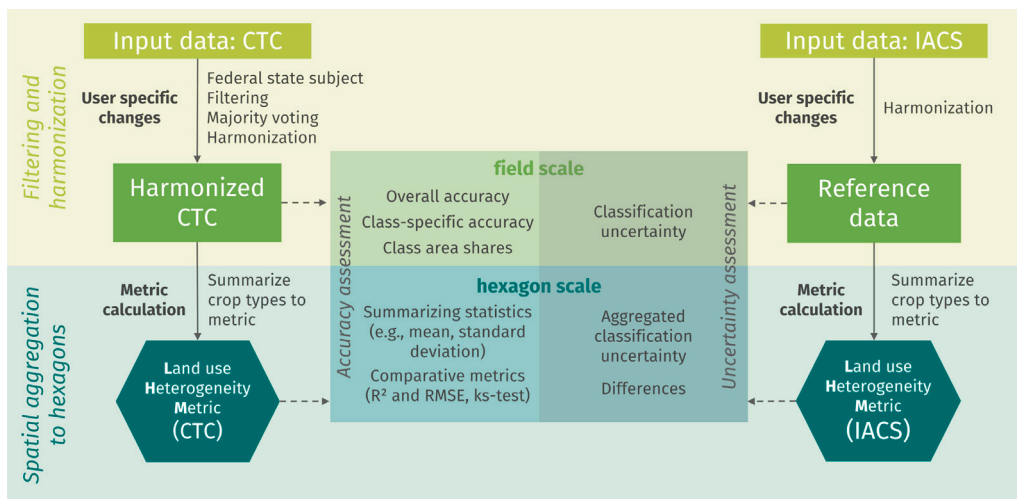


Fig. 1. Visualisation of the data workflow for deriving Land use Heterogeneity Metric (LHM) from Crop Type Classification (CTC) and Integrated Administration and Control System (IACS) as reference data as well as accuracy and uncertainty assessment. Changes to the data itself are represented by arrows. Dashed arrows show the quality assessment.

Class selection. Following Uthes et al. (2020), Hass et al. (2018), Asam et al. (2022), crop diversity is considered based on the use of arable land. Therefore, all grassland areas are excluded from this analysis (Table 2). For the dataset used for the calculation with PRE data, the geometries of the IACS classes 421, 422, 423, 424, 425, 441, 451, 452, 453, 454, 459 are removed (Table A.9). For SWD additional to the removed geometries form PRE also 433, 443 and 455 are removed while 454 remains included (Table A.11).

In addition, all other areas that do not represent a crop type are also removed. For PRE this contains unclassified pixels (pixel value = 0) as well as the classes *grassland* (19), *urban regions* (20), *water* (21), *other vegetations* (22) and *forest* (23). For SWD the classes *grassland*

(200), *cultivated grassland* (1602), *small woody features* (3001), *other agricultural areas* (3002), *fallow land* (3003), *other land* (3004) and *small woody features on other land* (3011) were excluded. In general, the focus on arable land leads to a reduction in approximately 20% of the Destatis (2024) crop area to the SWD and PRE crop area (Table 1).

Majority voting. In order to ensure comparability between the CTC and the IACS data, the classifications are transferred into the geometry of IACS data by a majority voting approach. The specific field is attributed the value of the crop type whose pixels occur most frequently in the geometry. If there are no crop types within the geometry, a *no data* value (NA) is assigned to the field polygon. That can be observed in places

where information has been excluded from the CTCs (e.g., classes *forest*, *urban areas*, or *water*). *No data* values have been ignored in the accuracy metric and LHMs calculation.

Thematic aggregation. After filtering all non-crop classes, only 18 crop type classes each remained for PRE and SWD. To ensure comparability, IACS classes are adapted to the classes of the respective CTC. The classes were defined by the authors of the respective classification themselves, resulting in different categories and different IACS mappings (Table 2). All IACS classes that are not covered by the specific CTC classes are grouped in the class *not classified* (777).

2.3.2. Hexagon scale: Spatial aggregation during LHM calculation

Since the geometries of the CTCs according to the majority voting and the reference data (IACS) are identical, only geometry-independent metrics are used in this study. As an example for a LHM the Shannon Evenness Index (*SEI*) will be calculated. The *SEI* is derived from the Shannon Index *H*. This index represents species occurrence and distribution in a single key value (Wolff et al., 2021; Uthes et al., 2020; Hass et al., 2018). The Shannon Index is well established in biodiversity assessment, making the results easily comprehensible and comparable (Klein et al., 2024). According to Breitschuh et al. (2004), a larger *H* value represents a greater diversity of species. *H* was calculated using the R package *vegan* (Oksanen et al., 2025) and is based on Eq. (1).

$$H = - \sum_i p_i \times \ln(p_i) \quad \text{with } p_i = \frac{n_i}{A} \quad (1)$$

p_i describes the proportional share of the crop type *i* that is calculated by the specific area of the crop type. n_i is divided by the entire arable area *A* within a single hexagon.

In this study, the *H* variant *SEI* is used, which normalises *H* to continuous values between 0 and 1. According to Eq. (2), *H* is divided by its potential maximum H_{max} , which depends on the number of different crops (*NoC*) within the hexagon.

$$SEI = \frac{H}{H_{max}} \quad \text{with } H_{max} = \ln(NoC) \quad (2)$$

If only one crop type is represented per hexagon, *SEI* is 0. If no crop type is specified, *no data* value (*NA*) is returned. Small values can occur if several crop types have only small proportions. Large values indicate a higher biodiversity that is generally produced by several crop types with medium field sizes (Schiller et al., 2024). For the calculation of *SEI*, only the number and shares within a hexagon are taken into account, and therefore the areas are cut by the hexagon borders.

2.3.3. Accuracy and uncertainty assessment

The terms “accuracy” and “uncertainty” in earth observation represent two related perceptions of data quality. While accuracy refers to closeness to truth (validation data), spatial uncertainty quantifies the lack of certainty or degree of “doubt” in modelling or classification products caused by all factors affecting data reliability (Yang et al., 2013). This means that the uncertainty goes beyond the pure measurement or classification error and also includes aspects of vagueness and ambiguity, such as unclear class definitions.

ISO 19157-1 (2023) provides a taxonomy of geodata quality categories. Trustworthiness is strongly related to the element “Thematic Quality” and its subelements “Thematic Classification Correctness” and “Quantitative Attribute Accuracy”. “Thematic Classification Correctness” refers to classified geodata products. The associated key metrics are based on the confusion matrix, which is an established core element for the accuracy assessment of land cover classifications used in this study (Stehman and Czaplewski, 1998; Stehman and Wickham, 2020; Stehman and Foody, 2019; Foody, 2002; Kaur et al., 2023). In contrast, the metrics of “Quantitative Attribute Accuracy” quantify “closeness of the value of a quantitative attribute to a value accepted as or known to be true” (ISO19157-1, 2023). A prerequisite is the availability of reference values that are accepted as true (Section 2.3.3).

For many applications, information on the spatial distribution of crop type classification uncertainty is of particular interest (Foody, 2002; Comber et al., 2012), which is also true for the LHM derivation applied in this study. In this study, we derived a simplified spatial uncertainty metric, as no metrics have been published for either classification.

After calculating the LHM with spatial aggregation on a hexagon scale in phase 2, the resulting distribution variants are compared. In doing so, we calculated the attribute accuracy metrics (Kaur et al., 2023; ISO 19157-1, 2023), performed a goodness-of-fit test (Thas, 2010), and evaluated the differences of the spatial uncertainty metric variants.

Regional accuracy. Both CTCs have already been described by their authors using general uncertainty metrics orientated to production, which refers to the total product and thus represents a global scale (Section 2.2.2). In order to determine metrics for the cropped, filtered and harmonised CTC and IACS subsets, we are using established confusion matrix-based metrics like the Overall Accuracy (*OA*) (Foody, 2002), the class-specific F_1 -score (Sokolova et al., 2006), the Macro- F_1 and the Weighted- F_1 (Farhadpour et al., 2024) (Table A.15).

The Overall Accuracy (*OA*) is defined as the number of correctly classified classes C_{ii} compared to the total number of samples *N* (Eq. (3)).

$$OA = \frac{\sum_{i=1}^N C_{ii}}{N} \quad (3)$$

According to Eq. (4), the class-specific F_1 -score of a class *i* results from the harmonic mean of the metrics *User Accuracy* (*UA*) and *Producer Accuracy* (*PA*).

$$F_1^i\text{-score} = 2 \times \frac{(UA_i \times PA_i)}{(UA_i + PA_i)} \quad (4)$$

While *UA* measures how often positive predictions of a model are correct (Eq. (5)), *PA* expresses how well a model finds all positive instances in a dataset (Eq. (6)).

$$UA = \frac{\text{Correct classifications (per class)}}{\text{Total predicted instances of that class}} \quad (5)$$

$$PA = \frac{\text{Correct classifications (per class)}}{\text{Reference instances of that class}} \quad (6)$$

The Macro- F_1 calculates the unweighted average of the class-specific F_1 -score across all *N* classes. It thus describes how evenly the model performs across all classes (Eq. (7)). Since all classes contribute equally, regardless of their frequency, the Macro- F_1 is sensitive to class imbalance.

$$\text{Macro-}F_1 = \frac{1}{N} \sum_{i=1}^N F_1^i \quad (7)$$

The Weighted- F_1 averages the class-specific F_1 -scores according to the relative frequency of each class in the dataset (Eq. (8)). Therefore, it reflects the overall performance of the model while adjusting for class imbalance.

$$\text{Weighted-}F_1 = \frac{\sum_{i=1}^N n_i F_1^i}{\sum_{i=1}^N n_i} \quad (8)$$

n_i refers to the number of actual instances of class *i* and *N* is the total number of instances. Large discrepancies between Macro- F_1 and Weighted- F_1 indicate that the model performs unevenly across classes.

All of these metrics provide different perspectives on the classification accuracy of the CTC. In addition, the proportions of area per crop type in a federal state are calculated.

Spatial classification uncertainty. After filtering and harmonising the different classes between the CTC and the reference data, the spatial classification uncertainty *P* (Table A.16) is determined by comparing the crop types *C* of the CTC with those of the IACS reference data. If the crop type classes are identical, *P* is set to 1, and if they differ, *P* is

set to 0 (Eq. (9)). If there is no crop type in a field for one of the input datasets, P is displayed with *no data* value (*NA*).

$$P = \begin{cases} 0 & \text{if } C_{CTC} \neq C_{IACS} \\ 1 & \text{if } C_{CTC} = C_{IACS} \end{cases} \quad (9)$$

As a result, each field is characterised by a P value, which can be summarised or aggregated to any reference unit.

Comparing distributions. After calculating the LHMs with spatial aggregation on a hexagon scale, the results based on CTC are compared with those based on IACS on a regional scale for each of the two federal states LS and BB. Apart from calculating the descriptive statistical metrics *mean* (μ), *standard deviation* (σ) and *difference of means* ($\Delta\mu$), the non-parametric *Kolmogorov–Smirnov goodness-of-fit test* (*KS test*) is used to test whether the empirical cumulative distributions (*ECDF*) of the calculated LHM distributions are significantly different from each other (Thas, 2010). The degree of difference is determined by the maximum absolute difference D_{KS} between the cumulative distributions of IACS- and CTC-based metrics (Eq. (10)).

$$D_{KS} = \max |ECDF_{IACS} - ECDF_{CTC}| \quad (10)$$

In addition, the accuracy metrics *Coefficient of determination* (R^2) and the *Root Mean Square Error* (*RMSE*) are calculated (Strahler et al., 2006) (Table A.17).

By subtracting the Shannon Evenness Index (*SEI*) based on IACS (SEI_{IACS}) by the metric value based on the CTC (SEI_{CTC}), a local difference per hexagon D_{HX} is calculated (Eq. (11)).

$$D_{HX} = SEI_{IACS} - SEI_{CTC} \quad (11)$$

The difference between the *SEI* values in a hexagon indicates differences between the input datasets. This may indicate a different number or distribution, or a combination of both, between the two input data sets (Breitschuh et al., 2004). A positive D_{HX} value indicates a higher *SEI* based on IACS. In contrast, a negative D_{HX} represents a higher CTC-based *SEI* (Table A.18).

The uncertainty values from field scale P are summarised by an area-weighted mean to present the uncertainty on hexagon scale P_{HX} . The aggregation transforms the integer values of P into floating numbers of P_{HX} . Areas with *no data* values (*NA*) are not taken into account (Eq. (12)).

$$P_{HX} = \frac{\sum_{i=1}^n (A_i \times P_i)}{\sum_{i=1}^n A_i} \quad (12)$$

n is the number of fields within the hexagon, A_i is the area of the i th fields within the hexagon, and P_i is the classification uncertainty metric value for the i th field.

2.4. DFFP assessment

In this study, we use the term DFFP in a problem-oriented sense. Specifically, we understand DFFP as the degree to which a remote sensing-based land-cover product provides the spatial, thematic, and temporal information required for robust calculation and interpretation of LHMs. Our objective is not to propose a complete, operational DFFP evaluation framework, but to define and structure the problem space for a specific domain: we identify which characteristics of mapping products are critical for different classes of landscape metrics and how these characteristics are only partially captured by conventional accuracy assessments. Accordingly, this paper should be understood as a conceptual and diagnostic contribution that motivates and delineates requirements for future, domain-specific DFFP frameworks for landscape heterogeneity analysis, rather than as a final framework in itself.

Table 3

Mean value for the years 2017 to 2019 of regional Overall Accuracy (*OA*), Macro- F_1 and Weighted- F_1 for the federal states Lower Saxony (LS) and Brandenburg (BB) and the sources Preidl (PRE) and Schwieder (SWD). All annual data for 2017 to 2019 can be found in Table A.13 and Table A.14.

Federal state	Source	<i>OA</i>	Macro- F_1	Weighted- F_1
LS	PRE	0.73 ± 0.02	0.56 ± 0.01	0.67 ± 0.02
BB	PRE	0.53 ± 0.04	0.42 ± 0.02	0.46 ± 0.06
LS	SWD	0.76 ± 0.01	0.59 ± 0.02	0.70 ± 0.01
BB	SWD	0.62 ± 0.05	0.51 ± 0.02	0.56 ± 0.05

3. Results

3.1. Field scale

Spatial intersection and class-specific filtering and harmonisation operations on the field scale result in two IACS and two CTC datasets that differ from the original datasets (see Table 1). This concerns the spatial coverage, which is the smallest area common to both the IACS and each CTC variant. Above this, classes and therefore areas have been removed or summarised (Section 2.3.1). As a result, the adjusted agricultural crop area without grassland for LS in 2019 for PRE is 18,465 km² and for SWD 18,362 km², while for BB 9682 km² (PRE) and 9617 km² (SWD) are classified (see Table 1).

A detailed comparison of the two harmonised datasets reveals slightly larger agricultural areas for PRE (LS: 38.7% (PRE) vs. 38.5% (SWD), BB: 32.6% (PRE) vs. 32.4% (SWD)) and a higher number of fields in both federal states for PRE. However, this difference does not affect the mean field size or its standard deviation, which remain identical in both datasets. Only the maximum field size in LS differs, with PRE containing a larger field (1.81 km²) than SWD (1.29 km²). In addition, it can be observed that for both datasets the crop area after harmonisation with the input data (SWD and PRE) is approximately 20% smaller than arable land. This results from the exclusion of temporary grassland, short-term fallow, catch crops, and minor/specialty crops that are part of arable land according to Destatis (2024) but are not included in the harmonised crop-type class schemes (see Table 1). The slight differences between CTCs are affected by the data processing steps explained in Section 2.3.1.

3.1.1. Overall accuracies

Table 3 show the mean regional accuracy metrics calculated for the CTCs provided by Preidl et al. (2020b) and Schwieder et al. (2024) using IACS as validation data. The classification results differ significantly between the two regions (LS, BB) and between the two input datasets (PRE, SWD). Overall, higher accuracies were achieved for LS than for BB. Based on PRE, the Overall Accuracy (*OA*) in LS was 0.73 ± 0.02, while in BB it was only 0.53 ± 0.04. A similar pattern can be seen for the Macro- F_1 (LS: 0.56 ± 0.01, BB: 0.42 ± 0.02) and the Weighted- F_1 (LS: 0.67 ± 0.02, BB: 0.46 ± 0.06).

The use of SWD leads to higher values in both federal states compared to PRE. For SWD in LS, an *OA* of 0.76 ± 0.01 was achieved, with a Macro- F_1 of 0.59 ± 0.02 and a Weighted- F_1 of 0.70 ± 0.01. The values also increase for BB (*OA*: 0.62 ± 0.05, Macro- F_1 : 0.51 ± 0.02, Weighted- F_1 : 0.56 ± 0.05), but remain significantly below the values for LS.

The partly high values of the overall accuracy (*OA* > 0.70) indicate a good overall recognition rate. The fact that Macro- F_1 and Weighted- F_1 have lower values than the *OA* indicates that the classification accuracy varies between classes. The lower values of the Macro- F_1 metric compared to the Weighted- F_1 metric suggest that dominant classes are classified more frequently correctly, while less common classes tend to be recognised less accurately.

Table 4

Proportion of area (A), User Accuracy (UA), Producer Accuracy (PA) and F_1 -scores (F_1), per crop type based on Preidl (PRE) for Lower Saxony (LS) and Brandenburg (BB) 2019 sorted by the F_1 -scores. No data values (NA) indicates that no F_1 -scores could be calculated. The four largest area percentages and all F_1 -score > 0.8 are highlighted in bold.

Crop classes	Code	A	UA	PA	F_1	A	UA	PA	F_1
Maize	8	34%	0.87	0.99	0.93	28%	0.59	0.98	0.74
Winter barley	4	8%	0.91	0.88	0.89	9%	0.86	0.74	0.80
Winter wheat	1	25%	0.79	0.92	0.85	27%	0.61	0.84	0.71
Rapeseed	10	4%	0.75	0.98	0.85	7%	0.88	0.95	0.91
Sugar beets	13	6%	0.76	0.97	0.85	1%	0.33	0.93	0.49
Winter rye	3	8%	0.73	0.86	0.79	16%	0.76	0.67	0.71
Potatoes	12	7%	0.66	0.87	0.75	2%	0.27	0.47	0.35
Stone fruits	15	1%	0.59	0.87	0.70	0.3%	0	NA	NA
Spring barley	6	3%	0.51	0.76	0.61	0.2%	0.23	0.09	0.13
Strawberries	14	0.5%	0.22	0.67	0.33	0	NA	NA	NA
Spring oat	7	1%	0.22	0.37	0.27	4%	0.25	0.66	0.37
Legumes	9	1%	0.12	0.66	0.20	3%	0.34	0.61	0.44
Asparagus	18	1%	0.10	0.70	0.18	0	NA	NA	NA
Spelt	2	0.1%	0.11	0.15	0.13	0.1%	0.12	0.11	0.12
Spring wheat	5	0.4%	0.11	0.15	0.13	0.2%	0.03	0.01	0.01
Leeks	11	0.001%	0	0	0	0.1%	0	0	0
Vines	16	0.1%	0	0	0	1%	0	0.38	0
not classified	NA	1%	NA	0	NA	1%	NA	0	NA
Federal state		LS				BB			

Table 5

Proportion of area (A), User Accuracy (UA), Producer Accuracy (PA) and F_1 -scores (F_1) per crop type based on Schwieder (SWD) for Lower Saxony (LS) and Brandenburg (BB) 2019 sorted by the F_1 -scores. No data values (NA) indicates that no F_1 -scores could be calculated. The four largest area percentages and all F_1 -score > 0.8 are highlighted in bold.

Crop classes	Code	A	UA	PA	F_1	A	UA	PA	F_1
Maize	1300	33%	0.88	0.97	0.92	20%	0.83	0.76	0.79
Winter rapeseed	1501	4%	0.80	0.98	0.88	7%	0.83	0.95	0.89
Sugar beet	1402	6%	0.80	0.97	0.87	1%	0.66	0.83	0.73
Winter wheat	1101	23%	0.77	0.96	0.86	15%	0.83	0.69	0.75
Winter barley	1102	9%	0.78	0.91	0.84	11%	0.77	0.78	0.78
Potato	1401	7%	0.80	0.89	0.84	1%	0.64	0.25	0.36
Orchards	4003	1%	0.85	0.81	0.83	0.02%	0.06	0.02	0.03
Spring barley	1201	2%	0.79	0.69	0.74	0.1%	0.66	0.06	0.10
Winter rye	1103	10%	0.59	0.93	0.72	29%	0.54	0.96	0.69
Vegetables	1603	2%	0.38	0.81	0.51	1%	0.25	0.40	0.31
Peas	1611	0.2%	0.36	0.64	0.46	1%	0.80	0.74	0.77
Broad bean	1612	1%	0.29	0.87	0.43	0.04%	0.58	0.24	0.34
Spring oat	1202	1%	0.30	0.67	0.42	3%	0.34	0.66	0.45
Soy	1614	0.1%	0.30	0.41	0.35	0.04%	0.46	0.38	0.42
Sunflower	1502	0.02%	0.12	0.18	0.14	6%	0.13	0.94	0.23
Lupin	1613	0.2%	0.04	0.36	0.07	3%	0.20	0.84	0.33
Grapevine	4001	NA	NA	0	NA	0.001%	0	NA	NA
not classified	NA	2%	NA	0	NA	2%	NA	0	NA
Federal state		LS				BB			

3.1.2. Class-specific accuracies

Tables 4 and 5 list for both classifications the accuracy metrics UA, PA and F_1 -score as well as their corresponding proportions for LS and BB for the year 2019. The accuracy metric for the years 2017 and 2018 can be found in the appendix (Table A.19, Table A.20, Table A.21, Table A.22). For both classifications from 2019, higher F_1 -scores can be observed for LS than in comparison to BB. In BB, there are two crop types with F_1 -score > 0.8 for PRE, while there are five crop types for LS. For SWD, there is one crop type with F_1 -score > 0.8 for BB and seven for LS. The highest F_1 -score for PRE in LS show the classes *maize* (0.93), *winter barley* (0.89) and *winter wheat*, *sugar beet*, and *rapeseed* (0.85 each). For BB, the highest F_1 -scores are observed for the classes *rapeseed* (0.91), *winter barley* (0.80), *maize* (0.74), *winter wheat* and *winter rye* (0.71). Overall, the same crop type classes achieve the highest F_1 -scores across both federal states and both classification approaches.

The largest total areas for PRE in both LS and BB are associated with the class *maize* (34% LS and 28% BB), followed by *winter wheat* (LS 25% and BB 27%), *winter rye* (LS 8% and BB 16%) and *winter barley* (LS 8% and BB 9%). All these classes show class accuracies of

F_1 -score > 0.7. Crop types with smaller total proportions (< 3%) also tend to have smaller F_1 -scores (< 0.5). Exceptions are the classes *stone fruits* in LS, which have class accuracies of F_1 -score = 0.70 with only 1% of the total area or the class *sugar beet* in BB, which have a class accuracy of F_1 -score = 0.49 with 1% area. In LS, the classes *vines* and *leeks* and in BB *stone fruits* and *leek*, are never correctly classified, so the F_1 -scores is determined as 0. In addition, there are no 777-values in the classification of the crop types, so that no F_1 -scores can be determined for them either. Additional, for BB, the classes *strawberry* and *asparagus* are not present, so they are described by *no data* (NA) values.

As with PRE, the main crops of SWD are *maize* (33% LS and 20% BB), *winter wheat* (23% LS and 15% BB), *winter rye* (10% LS and 29% BB) and *winter barley* (9% LS and 11% BB) account for the largest share of the total area. The corresponding class accuracies are all characterised by F_1 -score > 0.69. The class *grapevine* does not exist in LS and is not correctly classified in BB.

3.1.3. Classification uncertainty

Table 6 summarises the proportions of the spatial classification uncertainty categories P defined in Eq. (9) for both classifications and

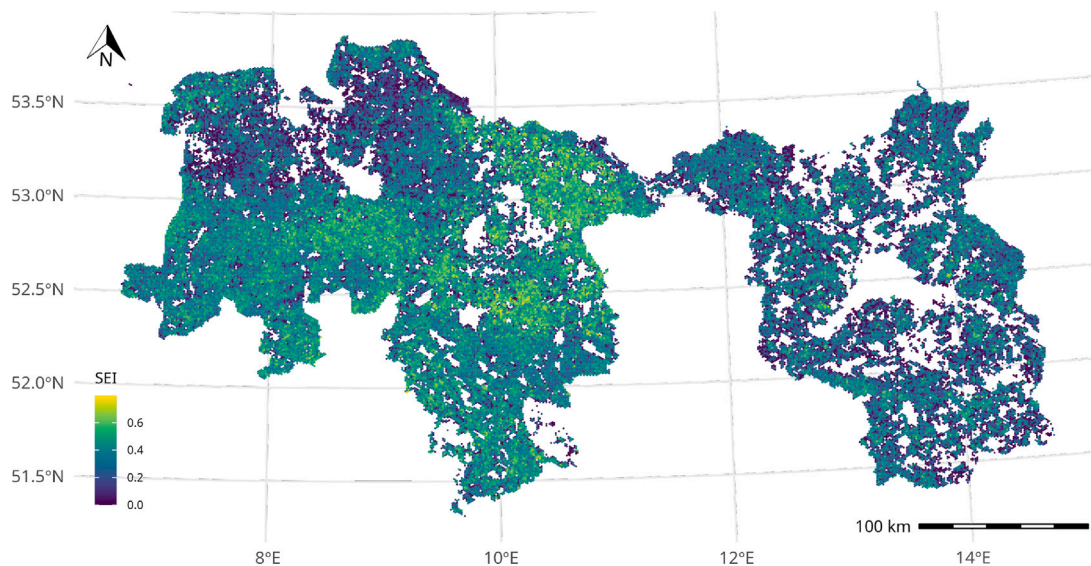


Fig. 2. Shannon Evenness Index (*SEI*) for Lower Saxony (LS) (left, $n = 46543$) and Brandenburg (BB) (right, $n = 28859$) for 2019 based on Preidl (PRE) on hexagon level. Colours shifting from blue (low values) to yellow (high values) with violet for $SEI = 0$.

Table 6

Proportions [%] of classification uncertainty categories P (see Eq. (9), *NA*: no value in datasets) on field-scale in Lower Saxony (LS) and Brandenburg (BB) for 2017, 2018 and 2019 and CTCs by Schwieder (SWD) and Preidl (PRE). The number of fields varies by federal state and CTC (see Table A.12).

Federal state	Source	P	2017	2018	2019
LS	SWD	0	22.9	24.4	21.3
LS	SWD	1	72.0	69.7	72.7
LS	SWD	NA	5.0	5.8	5.9
LS	PRE	0	22.5	27.0	21.6
LS	PRE	1	73.8	69.7	75.2
LS	PRE	NA	3.7	3.4	3.2
BB	SWD	0	28.8	37.9	37.0
BB	SWD	1	63.0	51.4	52.9
BB	SWD	NA	8.2	10.7	10.1
BB	PRE	0	44.7	36.6	35.9
BB	PRE	1	49.4	56.9	60.4
BB	PRE	NA	5.9	6.5	3.7

for all years. It shows that in both classifications clearly more fields are correctly classified ($P = 1$) in LS than in BB. But LS also has a significantly larger number of fields than BB. Differences between both CTCs in the two federal states can be observed primarily in BB. In particular, the year 2017 shows major deviations. The higher proportion of *no data* values (*NA*) for SWD is also noticeable for BB over the years.

3.2. Hexagon scale

3.2.1. Land use heterogeneity metrics

Fig. 2 illustrates the Shannon Evenness Index (*SEI*) as LHMs for agricultural areas in the two federal states of LS and BB for the year 2019. Urban regions (cities such as Bremen, Hanover, Berlin, Potsdam, etc.) and areas with other landscape types such as mountains (Harz mountains, Fläming) that do not contain any crop types are shown as empty. In LS, regions with higher *SEI* (>0.6) can be found in the centre, east and northeast. The region in the north is characterised by low crop diversity. For BB values are significantly lower compared to LS.

Table 7

Summary of all calculated comparative metrics, mean (μ), standard deviation (σ), Coefficient of determination (R^2), Root Mean Square Error (*RMSE*) and Kolmogorov–Smirnov goodness-of-fit test (D_{KS} , p -values = 0 for all D -values) for the variants of Shannon Evenness Index (*SEI*) based on Preidl (PRE) and Schwieder (SWD) as well as Integrated Administration and Control System (IACS) for Lower Saxony (LS) ($n = 46543$) and Brandenburg (BB) ($n = 28859$) as mean for 2017 to 2019.

FS	Source	μ_{IACS}	μ_{CTC}	σ_{IACS}	σ_{CTC}	R^2	<i>RMSE</i>	D_{KS}
LS	SWD	0.36	0.35	0.17	0.18	0.84	0.07	0.03
LS	PRE	0.36	0.35	0.17	0.17	0.81	0.07	0.03
BB	SWD	0.25	0.23	0.17	0.16	0.71	0.09	0.06
BB	PRE	0.25	0.25	0.16	0.17	0.69	0.09	0.04

3.2.2. Comparative metrics

The results of the comparison of the IACS- and CTC-based *SEI* distributions at the federal state level are listed in Table 7. Consequently, the mean values (μ) for both CTC variants are significantly lower in BB than in LS, which might be an expression of the reduced crop diversity in BB. Furthermore, SWD in BB provides slightly lower means than PRE.

The KS test results with $p = 0$ show that the null hypothesis of identical *SEI* distributions based IACS and CTC is rejected. However, the low D -values ($D = 0.03 - 0.06$), which represents the difference between the distributions, are very small. This indicates that while the differences are statistically detectable due to the large sample size, the overall agreement between the two datasets is high and the discrepancies are only minimal. For BB the D -values are generally higher than for LS, which indicates that the differences between the distributions are larger for BB. Also, the R^2 is lower in BB ($R^2 = 0.69 - 0.71$) compared to LS ($R^2 = 0.81 - 0.84$). That the discrepancies between the classifications (CTC) and the reference data (IACS) are larger for BB is already evident in the accuracy assessment of the classification themselves (Table 3). For both federal states, the SWD-based *SEI* show higher R^2 .

The differences between *SEI* variants based on IACS and CTC are illustrated as a map in Fig. 3. Negative values (blue) indicate that the CTC-based *SEI* is greater than the IACS-based *SEI* and vice versa (red). The differences here are the result of differences in the input data.

For the *SEI* difference based on PRE relative to IACS, both positive and negative deviations are observed in LS. Positive differences occur primarily in the north-east of LS, in the transition zone towards the north-west of BB, as well as in the northern parts of LS. In contrast,

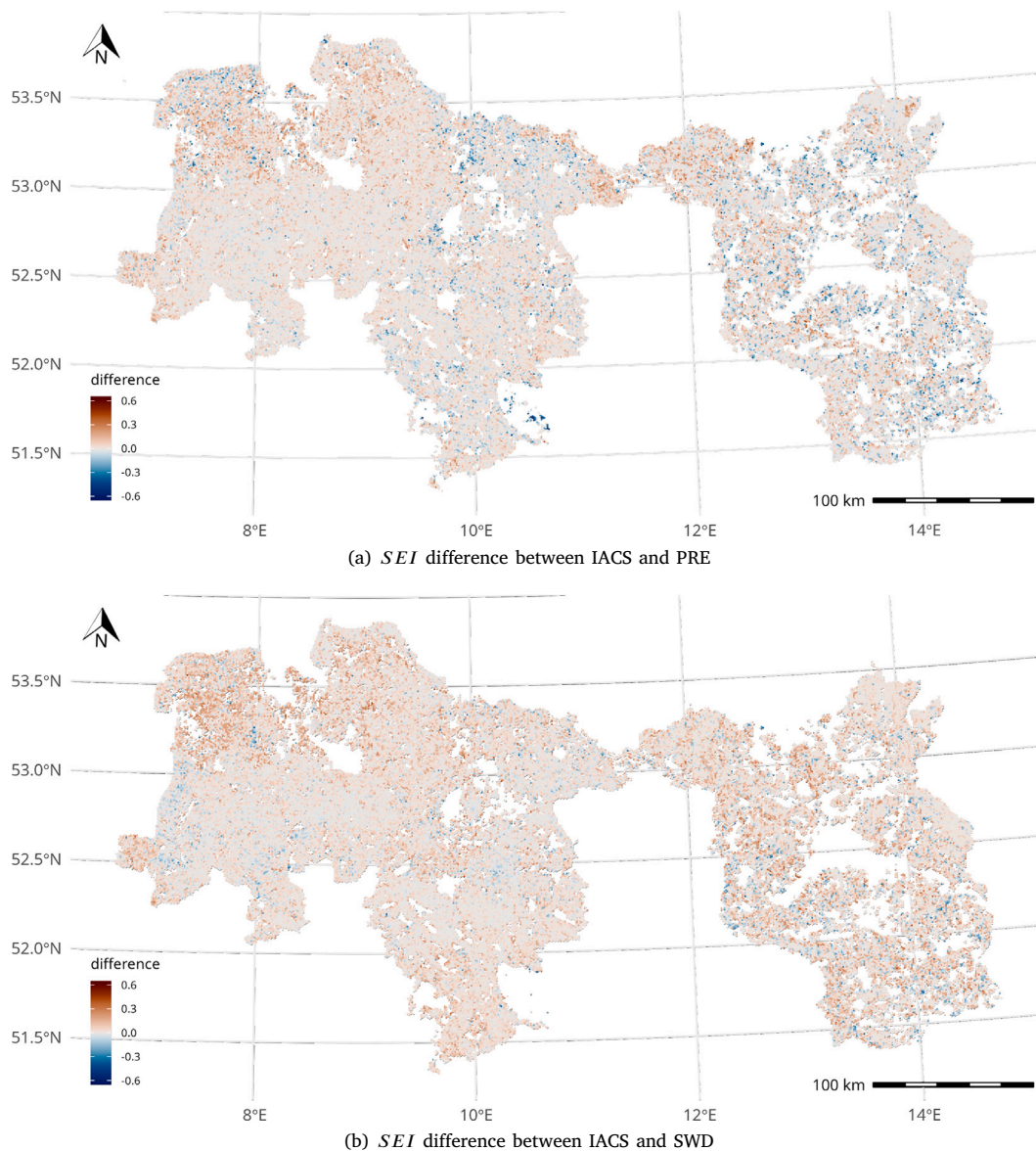


Fig. 3. Shannon Evenness Index (*SEI*) differences between IACS and CTC variants Preidl (PRE) (top) and Schwieder (SWD) (bottom) for Lower Saxony (LS) (left, $n = 46543$) and Brandenburg (BB) (right, $n = 28859$) in 2019 on hexagon level. Negative values (blue) indicate that the CTC-based values are higher than the IACS-based ones and vice versa (positive values in red). Colour saturation is related to the difference value. Grey corresponds to a difference of 0, i.e. identical values between IACS and CTC.

the coastal region itself is characterised by negative differences. In addition, several areas in eastern LS exhibit negative deviations, with a particularly pronounced cluster of strongly negative values in the south-east of LS. In BB, regions with positive and negative deviations are distributed across the entire federal state.

For the *SEI* difference based on SWD compared to IACS, both LS and BB are predominantly characterised by areas with positive differences, indicating that the reference data (IACS) provide larger values for the calculated metric than the crop type classification (SWD). Only in a few isolated regions do negative differences dominate, for example, north-east of Hanover or along the western border of LS. The observed differences often arise when one or more fields are assigned the IACS class 777. This class serves as a catch-all category for all IACS classes that are not represented in the CTC, thereby reflecting the major advantage of the high thematic resolution provided by IACS. Positive differences frequently occur when only a single field is classified differently. In contrast, if multiple fields within a hexagon are assigned the

777 class in IACS, the resulting *SEI* decreases because an apparently lower crop diversity is present. In this case, a negative difference is produced. In example, the comparison of IACS with PRE shows a region in the south-east of Lower Saxony with notably negative differences. Here, all fields in IACS are classified as 777, resulting in a very low *SEI*, as no diversity is recorded.

Fields with discrepancies in their assigned crop type may also extend across several hexagons, thereby giving the impression of a broader spatial influence, even though the underlying issue concerns only one field or a small group of fields. For this reason, larger differences are observed in BB, where field sizes are generally larger (Table 1).

3.2.3. Aggregated classification uncertainty

The aggregated classification uncertainty P_{HX} for both CTC variants in 2019 is shown in Fig. 4. LS is dominated by predominantly green areas ($P_{HX} > 0.5$). The few red areas ($P_{HX} < 0.5$) are mainly aggregated

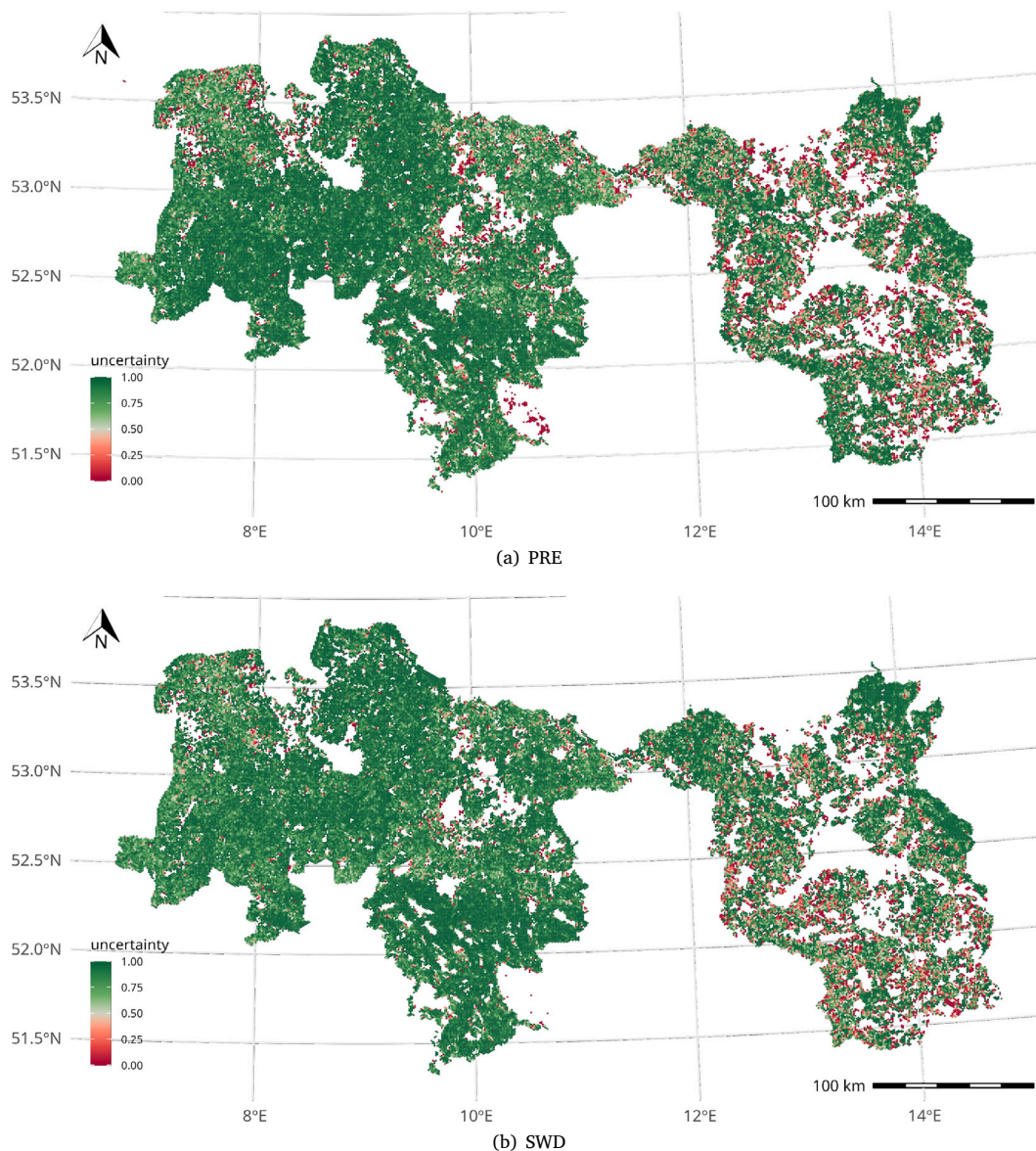


Fig. 4. Aggregated classification uncertainty P_{HX} for Lower Saxony (LS) (left, $n = 46543$) and Brandenburg (BB) (right, $n = 28859$) based on Preidl (PRE) (top) and Schwieder (SWD) (bottom) for 2019. Red colours ($P_{HX} = 0$) indicate no alignment between the crop type classes between IACS and the CTC variants, while green colours indicate a complete alignment ($P_{HX} = 1$).

and occur at the edges of the agricultural areas. There are red areas for PRE that do not occur at all in SWD (e.g. in the south-east of LS). In contrast, BB is characterised by more red areas compared to LS. Also here red areas are observed in the PRE dataset where *no data* values are presented in the data based on SWD. Except of differences for single hexagons, a similar distribution can be observed between the two CTCs on a federal state scale.

A closer examination of the map reveals that hexagons with extreme values (0 or 1) are typically defined by the presence of only a single field, meaning the hexagon’s value is largely determined by that field. This effect is particularly pronounced in BB, where large fields often span multiple hexagons, amplifying their influence.

Table 8 presents the proportions of the spatial aggregated uncertainty P to hexagon scale. By summarising the P -values of each individual polygon for hexagon level, the original discrete field level values of 0 and 1 are transformed into continuous values between 0 and 1. As a result of this spatial aggregation, the share of values that are exactly 0 or exactly 1 naturally decreases. In Figure A.5

corresponding density plots as a summary of the uncertainty P across the federal state on a hexagonal scale are presented, which also displays the continuous values. While the field-level calculation considered only polygons classified as agricultural land, the hexagon level computation is performed across the entire area. This also includes hexagons without any agricultural land, which therefore receive *no data* values. Consequently, the proportion of *no data* values increases substantially when moving from the field (Table 6) to the hexagon scale (Table 8).

A higher amount of *no data* values can be observed in BB than in LS (e.g., 26.6% in BB versus 12.1% in LS for PRE in 2019). In addition, BB contains more hexagons where all fields in the hexagon are misclassified ($P_{HX} = 0$; e.g., 5.8% in BB vs. 1.3% in LS for SWD in 2019). However, BB also shows a larger proportion of hexagons whose fields were completely correctly classified ($P_{HX} = 1$) with 14.1% PRE and 15.7% for SWD in BB compared to 11.3% for PRE and 13.0% for SWD in LS for 2019.

In general, both classifications show a similar trend for the two federal states. Compared to PRE, SWD shows a slightly higher number

Table 8

Proportions [%] of spatial aggregated classification uncertainty categories $P = 0$, $P = 1$, $0 < P < 1$ and $P = NA$ (see Eq. (9), NA: no value in one dataset) on hexagon scale in Lower Saxony (LS) and Brandenburg (BB) for 2017, 2018 and 2019 and Crop Type Classification (CTC) by Schwieder (SWD) and Preidl (PRE). The percentage values refer to the total number of hexagons in the federal states (LS: 46543, BB: 28859).

Federal state	Source	P	2017	2018	2019
LS	SWD	0	1.5	1.6	1.3
LS	SWD	1	11.1	10.1	13.0
LS	SWD	$0 < P < 1$	74.4	75.0	72.3
LS	SWD	NA	13.0	13.3	13.4
LS	PRE	0	2.4	2.7	2.5
LS	PRE	1	9.9	8.6	11.3
LS	PRE	$0 < P < 1$	75.4	76.5	74.1
LS	PRE	NA	12.3	12.2	12.1
BB	SWD	0	3.8	5.7	5.8
BB	SWD	1	24.0	14.5	15.7
BB	SWD	$0 < P < 1$	43.7	52.0	51.2
BB	SWD	NA	28.5	27.8	27.3
BB	PRE	0	12.2	8.0	7.0
BB	PRE	1	8.7	12.2	14.1
BB	PRE	$0 < P < 1$	52.5	53.2	52.3
BB	PRE	NA	26.6	26.6	26.6

of hexagons with completely correctly ($P_{HX} = 1$) fields and a lower proportion of hexagons with completely misclassified fields ($P_{HX} = 0$) for both federal states. However, SWD has a higher number of no data values.

4. Discussion

IACS has become a critical resource for scientific research, particularly in agricultural and environmental studies enabling precise analyses of agricultural landscapes (Pe'er et al., 2014). Main applications are tracking crop rotations and land use changes over time, calibrating ecological and agronomic models or validating remote sensing products. The limitations of using IACS include spatial coverage gaps due to the exclusion of non-subsidised agricultural land, restrictions on data access due to country-specific data protection regulations, and delays in data provision (Leonhardt et al., 2024).

Although remote sensing enables the scalable monitoring of changes in land use, IACS is superior due to its direct link to farmers' data, particularly in terms of classification depth and granularity. Remote sensing products can be used for areas not covered by IACS data. However, major limitations of remote sensing products are misclassification and class omission, which is expressed by different quality metrics. Thus, the systematic use of remote sensing data products, such as CTCs, requires the definition of prerequisites, including thematic accuracy metrics and spatial uncertainty that allow limitations to be communicated across different processing steps and scales (Höck et al., 2020).

In the following sections, we discuss different perspectives on CTCs quality and its implications for LHM calculation. We use SEI as a widely applied and conceptually transparent example of an LHM. Thus, the detailed quantitative results formally apply to SEI and closely related diversity-type indices. However, the uncertainty concept and computational workflow we employ are independent of the choice of metric and can, in principle, be applied to any LHM.

4.1. What differences exist between the accuracy metrics at the regional level?

The results of the overall accuracy assessment at regional level indicates that all accuracy metrics such as OA , UA , PA and F_1 -scores (see Tables 3, 4 and 5) exhibit differences within the input datasets

and, separately, within the regions. The different metrics do not show a complete agreement between CTC and IACS.

For LS, higher accuracy values are achieved across all metrics compared to BB (Table 3), which can partially be attributed to the different environmental structures. While BB is characterised by large fields due to its post-socialist history, Lower Saxony is characterised by smaller field sizes (Table 1, Uthes et al., 2020). The higher overall OA and F_1 -scores suggest also that the data variability within the classes is lower and the classes are more easily separable in LS. In addition, the two federal states differ in terms of different climate and soil conditions (Preidl et al., 2020b; Blickensdörfer et al., 2022) as well as differences in crop rotations and agricultural practices. Blickensdörfer et al. (2022) indicate that crop rotations in LS are typically dominated by spring cereals, whereas in BB, they are predominantly composed of winter cereals. Uthes et al. (2020) also highlight the different agricultural management systems and agricultural structures in the regions, which in their analyses have also led to lower biodiversity values in BB. Our analyses also show lower values for land use diversity for BB (Fig. 3), which results from the larger field sizes (Schiller et al., 2024). In calculations based on hexagons, which are commonly applied in biodiversity assessments (Birch et al., 2007; Wolff et al., 2021), the larger field sizes in BB result in fewer individual plots and crop types per reference unit. Consequently, fewer crop categories contribute to the computation of the SEI , potentially leading to lower diversity values irrespective of the actual agricultural diversity within the region. This highlights that the choice of spatial reference unit can substantially influence the resulting metrics. For example, Wolff et al. (2021) used hexagons with an area of 10 km², to capture large field sizes in their biodiversity assessment. Other reference units would be administrative areas (Griffith et al., 2000), but these vary in size and shape and are therefore not suitable for large-scale comparisons. Therefore, the reference units applied in such analyses should be critically evaluated and, if possible, made adaptable to the specific landscape context.

Across the federal states and input data, the Macro- F_1 is always below the Weighted- F_1 and the OA . The comparatively high Weighted- F_1 shows that the model achieves good results overall for frequent classes, while the lower Macro- F_1 illustrates that performance is less balanced across all classes. More specifically, we observed that classifications correctly recognise dominant classes more often, while rare classes are classified less accurately. A comparison of the F_1 -scores on a federal state level (Tables 4 and 5) clearly shows that fewer classes achieve high F_1 -scores in BB for 2019 (e.g., F_1 -scores > 0.8: PRE in BB/LS: 2/5 crop types, SWD in BB/LS: 1/7 crop types). We observed that in both CTCs less dominant classes tend to have lower F_1 -score, which might be due to their small area shares and related constraints (e.g., more mixed pixel at the edges) that make them more challenging to classify accurately (Orynbaikyzy et al., 2020).

Comparison of SEI differences derived from SWD and PRE relative to IACS shows that PRE exhibits greater variation in both positive and negative directions, whereas SWD produces almost exclusively positive values (Fig. 3). Positive values indicate that the SEI for IACS is higher than that for the respective CTC in a given hexagon. Translated to crop types, this means that crop diversity in the IACS data is higher. This may be due to a generally larger number of crop types or, even with the same number of crop types, to a more even distribution of proportions in the IACS data or a combination of both (Breitschuh et al., 2004).

In our case, it is well known that IACS offers substantially greater class diversity than the CTCs. To harmonise the datasets, all IACS classes not represented in the CTCs were aggregated into an additional class (777). Permanent grassland was removed entirely. Removing all unmatched classes would have eliminated the key advantage of IACSs, namely its high thematic resolution. Nevertheless, this creates a considerable imbalance between the input datasets, although harmonisation has mitigated this to some extent. But, without harmonisation, a comparison would not have been possible at all. Despite these adjustments, the additional class is responsible for most of the observed differences.

When a single field is classified differently, this commonly results in a positive difference, which describes a higher *SEI* for IACS. When multiple fields are assigned to class 777, this often reduces the *SEI* in IACS, leading to a negative difference between the SWD and IACS. It is notable that in many cases only a single field within a hexagon is classified differently. Nevertheless, the *SEI* is sensitive enough to reflect such changes, indicating its high responsiveness. At the same time, the strong influence of the additional class illustrates the considerable difference in classification depth between IACS and CTC, which is, one of the major advantages of IACS.

There are also differences in accuracy metrics between the CTC variants, with SWD classification tending to be characterised by slightly higher R^2 values. Apart from methodology, these differences may be due to various factors, in particular classification input data. For example, Asam et al. (2022), Blickensdörfer et al. (2022), Orynbaikyzy et al. (2020) show that combining optical (Sentinel-2) and SAR (Sentinel-1) data increases *OA* by 6 to 10% compared to approaches using only data from one sensor type. This could explain the slightly lower accuracy for PRE using optical (Sentinel-2) data.

Both CTCs have already been described by their authors in terms of global *OA*, which is given for PRE by 0.88 to 0.89. and for SWD by 0.84 to 0.85. However, the values for *OA* after purpose-specific adjustment are significantly lower than the national values reported by the authors (Table 3). In particular, removing of grassland classes from the original datasets to focus on the arable lands use in our use case seems to have a strong influence on *OA*. The excluding of grasslands, follows the approach of Uthes et al. (2020), who also only use arable land for diversity considerations, as grassland use is said to be inadequately represented in IACS. Hass et al. (2018) also uses pure crop diversity in the form of the Shannon Index based on crop type richness and cover for their compositional heterogeneity analyses. Due to the high classification accuracy of grasslands, excluding them leads to a decrease in *OA* values, which is also shown by Asam et al. (2022). During the harmonisation of crop types, they also removed the grassland class from their classification in their work, leading to a Germany-wide *OA* of 0.75, which is comparable to our results for LS (see Table 3). Orynbaikyzy et al. (2020) achieved a map accuracy of 75.5% in their classification for Germany in 2018, but including grassland classes.

4.2. How do the CTC/IACS differences affect LHMs?

The analysis of differences between CTCs and IACS as well as additional comparison metrics, such as R^2 and the KS test, revealed variations in the *SEI* resulting from the different input datasets (see Table 7). This shows that spatial patterns of differences between IACS and CTCs remain even after LHM derivation and after the associated spatial aggregation from the field level to the hexagon level.

To compare and evaluate the effects of different input datasets on the derivation of the *SEI*, both CTCs and IACS were harmonised semantically and geometrically:

- Semantic harmonisation was carried out according to the principle of the lowest common denominator. This means that the IACS dataset in particular, due to its granularity and greater class depth (e.g., 226 classes for LS), was subjected to an expert-based semantic aggregation. While this approach enables basic interoperability, there are several limitations like information loss, risk of over-simplification or semantic misalignment. Balancing standardisation with flexibility, such as preserving granularity through tiered harmonisation or supplementary metadata, could mitigate these trade-offs (Van Den Brink et al., 2017; Cheng et al., 2024). Future efforts might integrate machine learning to semi-automate semantic alignment while retaining critical details (Pang et al., 2024).

- The IACS geometries served as a first reference units, whereby existing geometric differences between CTCs and IACS were neglected. However, they also represent a potential source of data uncertainty, which can be quantified using geometric accuracy measures (Möller et al., 2013; Ye et al., 2018). During the metric calculation the data is finally summarised to the hexagon scale. Furthermore, the spatial reference units used (e.g. quadratic or hexagonal) can exert a substantial influence on LHMs (Roilo et al., 2024), which was not investigated in the study.

The largest discrepancies between IACS and CTC are probably caused by misclassification. These might arise from both spatial resolution (Griffiths et al., 2019) and, more importantly, from spectral limitations (Preidl et al., 2020b; Blickensdörfer et al., 2022; Asam et al., 2022; Orynbaikyzy et al., 2020). At the current 10 m resolution of Sentinel-2 as primary input for crop-type classification, the representation of real-world field conditions is inherently constrained. In addition, the spectral signatures and phenological patterns of many crops are highly similar. As a result, remote sensing approaches typically aggregate crop classes, improving classification performance but reducing class diversity (Preidl et al., 2020b; Blickensdörfer et al., 2022; Orynbaikyzy et al., 2020).

The probability P analysis (Section 3.2.3) showed that the largest differences at the hexagon level occur when only a single field is present. When multiple fields are included, discrepancies tend to average out and thus have a weaker effect on the final metric. Consequently, both spectral ambiguity and the number and spatial distribution of fields influence the observed differences. Nevertheless, on a larger spatial scale, consistent interregional patterns were observed for the *SEI*, which was derived from both CTC and IACS data. Hence, while local inconsistencies exist (hexagon scale), large-scale spatial patterns (federal state level) can still be reliably represented at the regional level. In addition, the main crop types can already be mapped very well using remote sensing products, as demonstrated by the high F_1 -values (Tables 4 and 5). So, depending on the purpose, CTCs could be an alternative to IACS. In particular, for countries with limited or no access to datasets such as IACS, remote sensing products could be a cost-effective and efficient alternative. However, accurate and quality information must also be provided with remote sensing products to reflect the credibility of the data.

4.3. Is sufficient CTC thematic accuracy information available to enrich LHM based on remote sensing-based products with data quality information?

The datasets from PRE and SWD, as well as other CTCs like Orynbaikyzy et al. (2020) and Griffiths et al. (2019), are described in their respective articles using common evaluation metrics such as *OA*. Class-specific measures like UA , PA , and F_1 -score, Macro- F_1 , Weighted- F_1 enable a more nuanced evaluation, considering that imbalanced class distributions reveal the semantic limitations and strengths of a dataset (Strahler et al., 2006). However, these metrics are reported only on a global scale, i.e., for Germany as a whole, whereas Preidl et al. (2020b) and Asam et al. (2022) also present their results divided into six regions.

The workflow presented in this study is divided into two production phases, each phase running at a specific spatial level. In the first phase, the original input CTCs datasets contained Germany-wide thematic accuracy metrics, which are commonly used in the accuracy assessment of crop type maps and correspond to existing standards such as ISO19157-1 (2023). However, they describe the quality of geodata from the perspective of producers and, therefore, may not be suitable for providing users with appropriate information (Fischer et al., 2023).

For a potential user who is interested in local and regional LHM quality, there are the following options:

1. If additional validation data are available, regional and thematic accuracy metrics can be derived.

2. If no validation data are available, LHM variants based on different CTC can be calculated.
3. Additional domain-specific data quality metrics are provided by the data producers or requested from them.

The third option is supported by Foody (2022), who suggested deriving additional information on classification uncertainty, often a by-product of classifications, to provide local information and reveal spatial variations in classification quality (e.g., Cheng et al., 2021; Valle et al., 2023). Common approaches include probability-based metrics that quantify pixel-level uncertainty using class probability distributions or ensemble-based uncertainty measures that quantify uncertainty through inter-model variance and disagreement among multiple classifier predictions (Lark et al., 2022; Dou et al., 2024). Such metrics are not part of existing standards such as ISO19157-1 (2023), but they can be considered an example of key metrics that address DFFP. Both, standardised and DFFP metrics, enable the evaluation of data quality from producer and user perspectives, as discussed in Leyk et al. (2005) or Brodeur et al. (2019).

As comprehensive DFFP descriptions often lack clear metadata, Petutschnig et al. (2024) propose a framework for assessing datasets based on their constraints. Applied to data producers, this would imply that, in addition to conventional metadata, information on dataset performance and suitable application domains should be included upon publication. To simplify the hurdles, it would be helpful to integrate such metadata fields of limits and benefits into existing ISO standards.

Therefore, in phase two, we used a simplified classification uncertainty metric P at the field scale. This allowed us to demonstrate that local uncertainty measures can be carried throughout the process chain and can be scaled to different spatial scales, such as from field scale (Table 6) to hexagon scale (Table 8). As a result, in addition to the actual analysis product “land use heterogeneity metric”, a spatial assessment of the local uncertainty on the hexagon scale is possible due to the available reference data. However, it has also become clear that the number of fields in the hexagon can have a significant impact on the quality of the data. For future policy applications, reporting should include not only accuracy values but also the number of observations on which these values are based. This information could be used, for example, as a filtering criterion to improve the reliability of decision-making. In general, especially in political decision-making, it is crucial to question credibility. Lokers et al. (2016) describes that, in the agro-environmental domain, “veracity” is considered the most important data characteristic. It should reflect the integrity and accuracy of data and data sources; therefore, it provides confidence for critical decision-making. Furthermore, particular attention should be paid to ensuring that the information provided is accurate and relevant to the specific purpose (Failing and Gregory, 2003).

It is important to clarify that established metrics such as OA , UA , PA , and F_1 -scores are mathematically sound and widely accepted in accuracy assessment protocols (Stehman and Foody, 2019; Stehman and Wickham, 2020; Foody, 2002). In principle, these metrics could reflect local and regional quality if validation were conducted with adequate sample sizes, appropriate spatial distribution, and stratified sampling designs at relevant regional scales (Stehman and Czaplewski, 1998). However, from a DFFP perspective, the practical limitation is that data producers typically do not provide such detailed spatial accuracy information. As demonstrated by our results, the CTC products used in this study, like many remote sensing classification products, report accuracy only at the national scale. This study simulates a scenario in which a potential user is primarily interested in the final product and its documented quality, rather than re-conducting the accuracy assessment. Therefore, our focus is on the description of the product’s data quality as communicated to users, enabling them to assess the product’s suitability for their specific purposes. This does not diminish the importance of documenting the full data lifecycle, including provenance, but recognises that users require accessible, spatially

explicit quality information with the published product (Fischer et al., 2023). Even well-designed accuracy assessments based on probability sampling cannot inform end users about local quality if the results are only reported globally (Foody, 2005; Comber et al., 2012).

Furthermore, the reputation of data providers and the accuracy of geodata and data sources are critical factors for the trustworthiness of environmental metrics and indicators that support policy-making (Yang et al., 2013; Lokers et al., 2016; Rowland et al., 2021). This applies to all steps in a workflow, where the challenge lies in transferring quality information from one step to the next. Höck et al. (2020) introduced the Quality Maturity Matrix (QMM) framework, which defines different levels of data production phases and is closely related to ISO 19157-1 (2023). Each level is characterised by quality metrics that fit into a hierarchical framework of overarching quality criteria and a set of sub-criteria or aspects. In turn, quality metrics are assigned to each aspect, which may differ depending on the (intermediate) data product. This means that modelling products are described by different accuracy metrics depending on the method used (e.g., classification, interpolation) (Kaur et al., 2023). From a methodological perspective, our study examines the propagation of uncertainties through a multi-stage geospatial workflow, demonstrating how input data quality affects the reliability of derived products at various spatial scales (Heuvelink, 2018).

In the context of biodiversity assessment, it is necessary to integrate additional metrics from multiple domains to enable a truly comprehensive evaluation (Klein et al., 2024). This is a fundamental prerequisite for allowing policy-makers to make well-informed, evidence-based decisions regarding biodiversity. For this reason, the this report includes 41 indicator profiles that provide a robust scientific basis for policy advice (Klein et al., 2024). Nevertheless, it remains essential that reliable and transparent quality and uncertainty information is provided for each individual component.

An example of considering uncertainty information in the context of the derivation of metrics and indicators is given by Möller et al. (2019). They present an uncertainty-aware framework for the standardised calculation of weather indices by applying statistical operations on the start dates of the annual and crop-specific phenological development stages. Taking into account additional phenological uncertainty layers during the calculation allows for the representation of uncertainty ranges in the final weather index products.

The key property of this approach is that the uncertainty propagation is independent of the specific metric or index. Once spatially explicit uncertainty layers are available for the input variables, any calculated product can be evaluated repeatedly on a set of metrics. Consequently, while we demonstrate this concept using the SEI as a concrete example, the same procedure can be applied to any landscape heterogeneity metric.

5. Outlook: Towards a dynamic DFFP framework

Several recent contributions provide general concepts and tools for assessing DFFP and geospatial data adequacy (e.g., Wentz and Shimizu, 2018; Höck et al., 2020; Fischer et al., 2023; Petutschnig et al., 2024). Building on these works, a logical next step is to develop a domain-specific DFFP framework for landscape heterogeneity analysis. Such a framework would explicitly link

1. the analytical requirements of different metric families (e.g. patch-based, edge-based, connectivity metrics),
2. relevant data quality dimensions (e.g. spatial resolution, thematic consistency, temporal comparability, provenance), and
3. decision-oriented criteria and thresholds for accepting or rejecting particular data products for specific landscape-ecological questions.

Our study contributes to this agenda by clarifying and structuring the problem: we show where standard accuracy assessment fails to reveal critical limitations for landscape metrics and which data characteristics appear most influential. Developing and validating a comprehensive evaluation uncertainty-aware schema tailored to LHM will therefore be the focus of future work, resulting in a multi-metric uncertainty propagation framework.

As this study further illustrates, the direct quantification of the effects of spatial uncertainties in the classification of crop types like misclassification and class omission on specific biodiversity indices has not yet been sufficiently investigated (Sandler and Rashford, 2018; Aramburu Merlos and Hijmans, 2020). A dedicated sensitivity analysis, stratified by P_{HX} and landscape complexity, would clarify where the heterogeneity signal remains robust despite classification noise and class omission. This highlights a critical knowledge gap for monitoring biodiversity in agriculture and planning conservation measures (Cánibe et al., 2022). The derivation of suitable spatial uncertainty metrics would be supported by the existence of a standardised framework and associated guidelines and best practice examples. For geodata, the ISO 19157-1 (2023) standard already provides a framework that offers guidance for assessing data quality and describes: “a data producer and a data user can potentially view data quality from different perspectives”. Thus, ISO 19157-1 (2023) is created as an extensible “framework for defining the quality of geographic data [...] with rules for how to add additional data quality measures”. Against this background, the approach introduced in this study can be considered as an expert-based process identifying relevant domain-specific DFFP metrics. Here, a potential domain-specific DFFP extension could be the element “Spatial Uncertainty” and the subelements “Class Spatial Uncertainty” and “Numerical Spatial Uncertainty” with associated techniques and metrics.

To automate the extension of ISO19157-1 (2023) with domain-specific DFFP elements such as “Spatial Uncertainty”, the next steps should involve leveraging LLM-driven text mining (e.g., Lawley et al., 2023; Zhang et al., 2024) to extract relevant metadata categories from the scientific literature (e.g., Farag et al., 2025; Piikki et al., 2021), formalising them as new quality elements using ontology and metadata standards, and integrating automated metric calculations into geospatial tool-use chains (Zhao et al., 2024; Zhang et al., 2025).

6. Conclusion

The study evaluated the suitability of satellite-based CTCs for calculating LHMs in agricultural landscapes, focussing on two German federal states, LS and BB, from 2017 to 2019. Two CTC products (SWD and PRE) were compared against high-quality reference data from the IACS.

As a result, significant regional and local differences in classification accuracy were observed, with consistently higher accuracies in LS compared to BB. Consequently, the LHM (*SEI*) derived from CTCs consistently differed from those based on IACS data, whereas CTC-derived metrics generally showed slightly lower values than IACS-derived metrics. Spatial aggregation from field to hexagon scale did not eliminate these discrepancies, indicating persistent spatial uncertainties in satellite-derived input data and associated LHMs. Nevertheless, similar large-scale patterns were observed at the regional level, indicating that, under certain conditions, results based on CTC data can effectively substitute for IACS data. This is particularly relevant for areas where IACS data are unavailable or cannot be accessed.

Although the data quality of the CTC products used followed standardised and established procedures, the thematic accuracy metrics provided enabled quality assessment primarily from the producer’s perspective, but not from the user’s perspective. This applies in particular to local and regional accuracy assessments. Thus, to ensure trustworthiness of LHM used in agricultural management and policy making, we recommend the following:

- extending standardised quality frameworks for geospatial products with additional DFFP elements like “spatial uncertainty”,
- providing detailed spatial uncertainty metrics with classification products enabling local and regional accuracy assessments.

In conclusion, while satellite-based CTCs offer advantages in spatial coverage and cost-efficiency, their inherent uncertainties must be explicitly accounted for when calculating LHMs. Future research should focus on developing standardised quality frameworks and providing detailed spatial uncertainty information to ensure reliable products like LHM as biodiversity proxies for agricultural management and policy making. In addition, transparency in geospatial workflows should be enhanced by aligning with FAIR (Findable, Accessible, Interoperable, Reusable) principles and embedding data quality information throughout the processing chain (Specka et al., 2023; García Brizuela et al., 2024).

CRediT authorship contribution statement

Jannes Säurich: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Marcel Schwieder:** Writing – review & editing, Methodology, Data curation. **Sebastian Preidl:** Writing – review & editing, Data curation. **Florian Beyer:** Writing – original draft, Software, Methodology, Conceptualization. **Markus Möller:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization.

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.

Acknowledgements

The data presented are from the collaborative project “Monitoring of Biodiversity in Agricultural Landscapes” (MonViA, funded by the German Federal Ministry of Agriculture, Food and Regional Identity). The analysis was carried out as part of the FAIRagro project within the German initiative Nationale Forschungsdateninfrastruktur funded by the German Research Foundation (DFG) – project number 501899475. Special thanks for ideas during the development of the DFFP analysis to Emilia Sophia Klaußner and Daniel Martini.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2026.103660>.

Data availability

Data generated and analysed in the study, are available here: <https://zenodo.org/records/17752848> (CC BY 4.0). R code for data processing and analysis is published here: <https://zenodo.org/records/17753964> (Apache 2.0).

References

- Aramburu Merlos, F., Hijmans, R.J., 2020. The scale dependency of spatial crop species diversity and its relation to temporal diversity. *Proc. Natl. Acad. Sci.* 117 (42), 26176–26182. <http://dx.doi.org/10.1073/pnas.2011702117>, URL: <https://pnas.org/doi/full/10.1073/pnas.2011702117>.
- Asam, S., Gessner, U., Almengor González, R., Wenzl, M., Kriese, J., Kuenzer, C., 2022. Mapping crop types of Germany by combining temporal statistical metrics of sentinel-1 and sentinel-2 time series with LPIS data. *Remote. Sens.* 14 (13), 2981. <http://dx.doi.org/10.3390/rs14132981>.
- Auerswald, K., Menzel, A., 2021. Change in erosion potential of crops due to climate change. *Agric. Ecol. Forest. Meteorol.* 300, 108338. <http://dx.doi.org/10.1016/j.agrformet.2021.108338>.
- Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., de Blust, G., de Cock, R., Diekötter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J.P., Opdam, P., Roubalova, M., Schermann, A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M., Speelmans, M., Simova, P., Verboom, J., van Wingerden, W., Zobel, M., Edwards, P.J., 2008. Indicators for biodiversity in agricultural landscapes: A pan-European study. *J. Appl. Ecol.* 45 (1), 141–150. <http://dx.doi.org/10.1111/j.1365-2664.2007.01393.x>.
- Bindi, M., Olesen, J.E., 2011. The responses of agriculture in Europe to climate change. *Reg. Environ. Chang.* 11 (S1), 151–158. <http://dx.doi.org/10.1007/s10113-010-0173-x>.
- Birch, C.P., Oom, S.P., Beecham, J.A., 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecol. Model.* 206 (3–4), 347–359. <http://dx.doi.org/10.1016/j.ecolmodel.2007.03.041>.
- Blickensdorfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmí, S., Hostert, P., 2022. Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sens. Environ.* 269, 112831. <http://dx.doi.org/10.1016/j.rse.2021.112831>.
- Brandt, P., Beyer, F., Borrmann, P., Möller, M., Gerighausen, H., 2024. Ensemble learning-based crop yield estimation: a scalable approach for supporting agricultural statistics. *GIScience Remote Sensing* 61, 2367808. <http://dx.doi.org/10.1080/15481603.2024.2367808>, URL: <https://www.tandfonline.com/doi/full/10.1080/15481603.2024.2367808>.
- Breitschuh, G., Eckert, H., Feige, H., Germand, U., Sauerbeck, D., 2004. Entwicklung eines Umweltcontrolling-/Umweltoptimierungssystems in der Landwirtschaft. *Umweltforschungsplan Des Bundesmin. Für Umw. Naturschutz Und Reakt.*
- Brodeur, Coetzee, Danko, Garcia, Hjelmager, 2019. Geographic information metadata – An outlook from the international standardization perspective. *ISPRS Int. J. Geo-Information* 8 (6), 280. <http://dx.doi.org/10.3390/ijgi8060280>, URL: <https://www.mdpi.com/2220-9964/8/6/280>.
- 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 (1), <http://dx.doi.org/10.1038/s41597-022-01307-4>.
- Cánibe, M., Titeux, N., Domínguez, J., Regos, A., 2022. Assessing the uncertainty arising from standard land-cover mapping procedures when modelling species distributions. In: Zhang, Z. (Ed.), *Diversity and Distributions* 28 (4), 636–648. <http://dx.doi.org/10.1111/ddi.13456>, URL: <https://onlinelibrary.wiley.com/doi/10.1111/ddi.13456>.
- Cheng, K.-S., Ling, J.-Y., Lin, T.-W., Liu, Y.-T., Shen, Y.-C., Kono, Y., 2021. Quantifying uncertainty in land-use/land-cover classification accuracy: A stochastic simulation approach. *Front. Environ. Sci.* 9, 628214. <http://dx.doi.org/10.3389/fenvs.2021.628214>, URL: <https://www.frontiersin.org/articles/10.3389/fenvs.2021.628214/full>.
- Cheng, C., Messerschmidt, L., Bravo, I., Waldbauer, M., Bhavikatti, R., Schenk, C., Grujic, V., Model, T., Kubinec, R., Barceló, J., 2024. A general primer for data harmonization. *Sci. Data* 11 (1), 152. <http://dx.doi.org/10.1038/s41597-024-02956-3>, URL: <https://www.nature.com/articles/s41597-024-02956-3>.
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The harmonized landsat and sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161. <http://dx.doi.org/10.1016/j.rse.2018.09.002>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425718304139>.
- Comber, A., Fisher, P., Brunson, C., Khmag, A., 2012. Spatial analysis of remote sensing image classification accuracy. *Remote Sens. Environ.* 127, 237–246. <http://dx.doi.org/10.1016/j.rse.2012.09.005>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425712003598>.
- Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroui, C., Valero, S., Bégué, A., Dejoux, J.-F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* 221, 551–568. <http://dx.doi.org/10.1016/j.rse.2018.11.007>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425718305145>.
- Destatis, 2024. Bodenfläche nach Nutzungsarten und Bundesländern. URL: <https://tinyurl.com/3jfr6cks>.
- Deutscher Wetterdienst, 2018. In: Deutscher Wetterdienst (Ed.), *Klimastatusbericht Deutschland Jahr 2017*. Deutscher Wetterdienst, Geschäftsbereich Klima und Umwelt, Offenbach, URL: <https://tinyurl.com/yc5p3yez>.
- Deutscher Wetterdienst, 2020a. *Klimastatusbericht Deutschland Jahr 2018*. Selbstverlag des Deutschen Wetterdienstes, Offenbach, URL: <https://tinyurl.com/ysnd735p>.
- Deutscher Wetterdienst, 2020b. *Klimastatusbericht Deutschland Jahr 2019*. Selbstverlag des Deutschen Wetterdienstes, Offenbach, URL: <https://tinyurl.com/yp8jp6zj>.
- Dou, P., Huang, C., Han, W., Hou, J., Zhang, Y., Gu, J., 2024. Remote sensing image classification using an ensemble framework without multiple classifiers. *ISPRS J. Photogramm. Remote Sens.* 208, 190–209. <http://dx.doi.org/10.1016/j.isprsjprs.2023.12.012>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0924271624000145>.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36. <http://dx.doi.org/10.1016/j.rse.2011.11.026>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425712000636>.
- Duelli, P., Obrist, M.K., 2003. Biodiversity indicators: the choice of values and measures. *Agric. Ecosyst. Environ.* 98 (1–3), 87–98. [http://dx.doi.org/10.1016/S0167-8809\(03\)00072-0](http://dx.doi.org/10.1016/S0167-8809(03)00072-0), URL: <https://linkinghub.elsevier.com/retrieve/pii/S0167880903000720>.
- Fahrig, L., Girard, J., Duro, D., Pasher, J., Smith, A., Javorek, S., King, D., Lindsay, K.F., Mitchell, S., Tischendorf, L., 2015. Farmlands with smaller crop fields have higher within-field biodiversity. *Agric. Ecosyst. Environ.* 200, 219–234. <http://dx.doi.org/10.1016/j.agee.2014.11.018>.
- Failing, L., Gregory, R., 2003. Ten common mistakes in designing biodiversity indicators for forest policy. *J. Environ. Manag.* 68 (2), 121–132. [http://dx.doi.org/10.1016/S0301-4797\(03\)00014-8](http://dx.doi.org/10.1016/S0301-4797(03)00014-8), URL: <https://www.sciencedirect.com/science/article/pii/S0301479703000148>.
- FAOSTAT, 2022. Data. URL: <https://www.fao.org/faostat/en/#data>.
- Farang, M., Emam, A., Leonhardt, J., Roscher, R., 2025. Enhancing decision support in crop production: Analyzing conformal prediction for uncertainty quantification. *Comput. Electron. Agric.* 237, 110559. <http://dx.doi.org/10.1016/j.compag.2025.110559>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0168169925006659>.
- Farhadpour, S., Warner, T.A., Maxwell, A.E., 2024. Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices. *Remote Sens.* 16 (3), 533. <http://dx.doi.org/10.3390/rs16030533>.
- Feld, C.K., Sousa, J.P., Da Silva, P.M., Dawson, T.P., 2010. Indicators for biodiversity and ecosystem services: towards an improved framework for ecosystems assessment. *Biodivers. Conserv.* 19 (10), 2895–2919. <http://dx.doi.org/10.1007/s10531-010-9875-0>, URL: <http://link.springer.com/10.1007/s10531-010-9875-0>.
- Fischer, J., Egli, L., Groth, J., Barrasso, C., Ehrmann, S., Figgemeier, H., Henzen, C., Meyer, C., Müller-Pfefferkorn, R., Rümmler, A., Wagner, M., Bernard, L., Seppelt, R., 2023. Approaches and tools for user-driven provenance and data quality information in spatial data infrastructures. *Int. J. Digit. Earth* 16 (1), 1510–1529. <http://dx.doi.org/10.1080/17538947.2023.2198778>, URL: <https://www.tandfonline.com/doi/full/10.1080/17538947.2023.2198778>.
- Footy, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* 80 (1), 185–201. [http://dx.doi.org/10.1016/S0034-4257\(01\)00295-4](http://dx.doi.org/10.1016/S0034-4257(01)00295-4), URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425701002954>.
- Footy, G.M., 2005. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *Int. J. Remote Sens.* 26 (6), 1217–1228. <http://dx.doi.org/10.1080/01431160512331326521>.
- Footy, G.M., 2022. Global and local assessment of image classification quality on an overall and per-class basis without ground reference data. *Remote Sens.* 14 (21), 5380. <http://dx.doi.org/10.3390/rs14215380>, URL: <https://www.mdpi.com/2072-4292/14/21/5380>.
- García Brizuela, J., Scharfenberg, C., Scheuner, C., Hoedt, F., König, P., Kranz, A., Leidel, A., Martini, D., Schneider, G., Schneider, J., Singson, L.S., Von Waldow, H., Wehrmeyer, N., Usadel, B., Lesch, S., Specka, X., Lange, M., Arend, D., 2024. A roadmap for a middleware as a federation service for integrative data retrieval of agricultural data. *J. Integr. Bioinform.* 21 (3), 20240027. <http://dx.doi.org/10.1515/jib-2024-0027>, URL: <https://www.degruyter.com/document/doi/10.1515/jib-2024-0027/html>.
- Ghassemi, B., Izquierdo-Verdiguier, E., Verhegghen, A., Yordanov, M., Lemoine, G., Moreno Martínez, A., De Marchi, D., Van Der Velde, M., Vuolo, F., d'Andrion, R., 2024. European Union crop map 2022: Earth observation's 10-meter dive into Europe's crop tapestry. *Sci. Data* 11 (1), 1048. <http://dx.doi.org/10.1038/s41597-024-03884-y>, URL: <https://www.nature.com/articles/s41597-024-03884-y>.
- Griffith, J.A., Martinko, E.A., Price, K.P., 2000. Landscape structure analysis of Kansas at three scales. *Landsc. Urban Plan.* 52 (1), 45–61. [http://dx.doi.org/10.1016/S0169-2046\(00\)00112-2](http://dx.doi.org/10.1016/S0169-2046(00)00112-2).
- Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sens. Environ.* 220, 135–151. <http://dx.doi.org/10.1016/j.rse.2018.10.031>.

- Hass, A.L., Kormann, U.G., Tschardt, T., Clough, Y., Baillo, A.B., Sirami, C., Fahrig, L., Martin, J.-L., Baudry, J., Bertrand, C., Bosch, J., Brotons, L., Burel, F., Georges, R., Giral, D., Marcos-García, M.Á., Ricarte, A., Siriwardena, G., Batáry, P., 2018. Landscape configurational heterogeneity by small-scale agriculture, not crop diversity, maintains pollinators and plant reproduction in western Europe. *Proc. Biological Sci.* 285 (1872), <http://dx.doi.org/10.1098/rspb.2017.2242>.
- Heink, U., Kowarik, I., 2010. What criteria should be used to select biodiversity indicators? *Biodivers. Conserv.* 19 (13), 3769–3797. <http://dx.doi.org/10.1007/s10531-010-9926-6>, URL: <http://link.springer.com/10.1007/s10531-010-9926-6>.
- Heuvelink, G.B.M., 2018. Uncertainty and uncertainty propagation in soil mapping and modelling. In: McBratney, A.B., Minasny, B., Stockmann, U. (Eds.), *Pedometrics*. Springer, Cham, pp. 439–461. http://dx.doi.org/10.1007/978-3-319-63439-5_14.
- Hoban, S., Archer, F.I., Bertola, L.D., Bragg, J.G., Breed, M.F., Bruford, M.W., Coleman, M.A., Ekblom, R., Funk, W.C., Grueber, C.E., Hand, B.K., Jaffé, R., Jensen, E., Johnson, J.S., Kershaw, F., Liggins, L., MacDonald, A.J., Mergeay, J., Miller, J.M., Muller-Karger, F., O'Brien, D., Paz-Vinas, I., Potter, K.M., Razgour, O., Vernesi, C., Hunter, M.E., 2022. Global genetic diversity status and trends: towards a suite of essential biodiversity variables (EBVs) for genetic composition. *Biological Rev.* 97 (4), 1511–1538. <http://dx.doi.org/10.1111/brv.12852>, URL: <https://onlinelibrary.wiley.com/doi/10.1111/brv.12852>.
- Höck, H., Toussaint, F., Thiemann, H., 2020. Fitness for use of data objects described with quality maturity matrix at different phases of data production. *Data Sci. J.* 19, <http://dx.doi.org/10.5334/dsj-2020-045>.
- ISO 19157-1, 2023. Geographic Information: Data Quality. Technical Report, International Organization for Standardization, URL: <https://www.iso.org/standard/78900.html>.
- ISO19157-1, 2023. Geographic Information: Data Quality. Technical Report, International Organization for Standardization, Geneva, Switzerland, URL: <https://www.iso.org/standard/78900.html>.
- Jänicke, C., Goddard, A., Stein, S., Steinmann, H.-H., Lakes, T., Nendel, C., Müller, D., 2022. Field-level land-use data reveal heterogeneous crop sequences with distinct regional differences in Germany. *Eur. J. Agron.* 141, 126632. <http://dx.doi.org/10.1016/j.eja.2022.126632>, URL: <https://www.sciencedirect.com/science/article/pii/S1161030122001800>.
- Jänicke, C., Petersen, K.A., Schmidts, P., Müller, D., Jepsen, M.R., 2025. Field and farm-level data on agricultural land use for the European Union. *Sci. Data* 12 (1), 1050. <http://dx.doi.org/10.1038/s41597-025-05210-6>, URL: <https://www.nature.com/articles/s41597-025-05210-6>.
- Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M., Geller, G.N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F.E., Pereira, H.M., Regan, E.C., Schmeller, D.S., Turak, E., 2019. Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* 3 (4), 539–551. <http://dx.doi.org/10.1038/s41559-019-0826-1>, URL: <https://www.nature.com/articles/s41559-019-0826-1>.
- Kaur, M., Singh, D., Jabarulla, M.Y., Kumar, V., Kang, J., Lee, H.-N., 2023. Computational deep air quality prediction techniques: a systematic review. *Artif. Intell. Rev.* 56 (S2), 2053–2098. <http://dx.doi.org/10.1007/s10462-023-10570-9>, URL: <https://link.springer.com/10.1007/s10462-023-10570-9>.
- Kissling, W.D., Walls, R., Bowser, A., Jones, M.O., Kattge, J., Agosti, D., Amengual, J., Basset, A., Van Bodegom, P.M., Cornelissen, J.H.C., Denny, E.G., Deudero, S., Eglhoff, W., Elmendorf, S.C., Alonso García, E., Jones, K.D., Jones, O.R., Lavorel, S., Lear, D., Navarro, L.M., Pawar, S., Pirzl, R., Rüger, N., Sal, S., Salguero-Gómez, R., Schigel, D., Schulz, K.-S., Skidmore, A., Guralnick, R.P., 2018. Towards global data products of essential biodiversity variables on species traits. *Nat. Ecol. Evol.* 2 (10), 1531–1540. <http://dx.doi.org/10.1038/s41559-018-0667-3>, URL: <https://www.nature.com/articles/s41559-018-0667-3>.
- Klein, K., Ogan, S., Rottstock, T., Tönnschhof, C., Ackermann, A., Alkassab, A., Balzar, L., Beer, H., Beyer, Böhner, H., Briem, F., Dauber, J., Dieker, P., Erasmi, S., Finn, D., Früchtenicht, E., Gerighausen, H., Gocht, A., Göderz, H., Golla, Grabener, S., Greil, H., Gummert, A., Hamm, Hartmann, M., Hellwig, N., Hendriksma, H.P., Herwig, Herz, A., Hoffmann, C., Hommel, B., Kaczmarek, M., Kämper, W., Kasiske, T., Kassel, P., Kehlenbeck, H., Klimek, S., Krahnert, A., Krenzel-Horney, S., Krüger, L., Lakemann, L., Lehmsius, J., Lettow, N., Levers, C., Lindermann, L., Lodenkemper, R., Lorenz, S., Lüken, D.J., Meinikmann, K., Mohr, K.A., Möller, M., Perić, Z., Pingel, M., Pistorius, J., Richter, A., Riedel, T., Röder, N., Ruf, L.C., Schröder, S., Schwieder, M., Sensen, S., Sichel, W., Sittinger, M., Sommerlandt, F., Stahl, J., Strassmeyer, J., Tebbe, C., Uhler, J., Uhlott, J., Ulber, L., Urso, L.-M., Vaupel, A., von Redwitz, C., von Hörsten, D., Wang, H., Wegener, J.K., Wider, J., Winkler, M., Yang, J., 2024. MonViA Indikatorenbericht 2024: Bundesweites Monitoring der biologischen Vielfalt in Agrarlandschaften. URL: https://www.openagrar.de/receive/openagrar_mods_00097695.
- Lacagnina, C., David, R., Nikiforova, A., Kuusniemi, M.-E., Cappiello, C., Biehlmaier, O., Wright, L., Schubert, C., Bertino, A., Thiemann, H., Dennis, R., 2022. Towards a Data Quality Framework for EOSC Authorship Community. Technical Report, EOSC Association, <http://dx.doi.org/10.5281/zenodo.7515816>, URL: <https://hal.science/hal-04017152>.
- Lark, R., Chagumaira, C., Milne, A., 2022. Decisions, uncertainty and spatial information. *Spat. Stat.* 50, 100619. <http://dx.doi.org/10.1016/j.spasta.2022.100619>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S2211675322000161>.
- Lawley, C.J.M., Gadd, M.G., Parsa, M., Lederer, G.W., Graham, G.E., Ford, A., 2023. Applications of natural language processing to geoscience text data and prospectivity modeling. *Nat. Resour. Res.* 32 (4), 1503–1527. <http://dx.doi.org/10.1007/s11053-023-10216-1>, URL: <https://link.springer.com/10.1007/s11053-023-10216-1>.
- Leonhardt, H., Wesemeyer, M., Eder, A., Hüttel, S., Lakes, T., Schaak, H., Seifert, S., Wolff, S., 2024. Use cases and scientific potential of land use data from the EU's integrated administration and control system: A systematic mapping review. *Ecol. Indic.* 167, 112709. <http://dx.doi.org/10.1016/j.ecolind.2024.112709>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S1470160X2401166X>.
- Leyk, S., Boesch, R., Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Trans. GIS* 9 (3), 291–322. <http://dx.doi.org/10.1111/j.1467-9671.2005.00220.x>, URL: <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9671.2005.00220.x>.
- Lokers, R., Knapen, R., Janssen, S., van Randen, Y., Jansen, J., 2016. Analysis of Big Data technologies for use in agro-environmental science. *Environ. Model. Softw.* 84, 494–504. <http://dx.doi.org/10.1016/j.envsoft.2016.07.017>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S1364815216304194>.
- Merckx, T., Feber, R.E., Dulieu, R.L., Townsend, M.C., Parsons, M.S., Bourn, N.A., Riordan, P., Macdonald, D.W., 2009. Effect of field margins on moths depends on species mobility: Field-based evidence for landscape-scale conservation. *Agric. Ecosyst. Environ.* 129 (1–3), 302–309. <http://dx.doi.org/10.1016/j.agee.2008.10.004>.
- Möller, M., Birger, J., Gidudu, A., Gläßer, C., 2013. A framework for the geometric accuracy assessment of classified objects. *Int. J. Remote Sens.* 34 (24), 8685–8698. <http://dx.doi.org/10.1080/01431161.2013.845319>, URL: <https://www.tandfonline.com/doi/full/10.1080/01431161.2013.845319>.
- Möller, M., Doms, J., Gerstmann, H., Feike, T., 2019. A framework for standardized calculation of weather indices in Germany. *Theor. Appl. Climatol.* 136 (1–2), 377–390. <http://dx.doi.org/10.1007/s00704-018-2473-x>, URL: <http://link.springer.com/10.1007/s00704-018-2473-x>.
- Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R., Solymos, P., Stevens, M.H.H., Szocs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., Martino, C., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J., Weedon, J., 2025. Community ecology package: Package 'vegan'. URL: <https://github.com/vegan/vegan>.
- Oppermann, R., Braband, D., Haack, S., 2005. Nature indicators for agriculture. *Berichte Über Landwirtschaft.* (83), 76–102.
- Orynbaikyzy, A., Gessner, U., Mack, B., Conrad, C., 2020. Crop type classification using fusion of sentinel-1 and sentinel-2 data: Assessing the impact of feature selection, optical data availability, and parcel sizes on the accuracies. *Remote Sens.* 12 (17), 2779. <http://dx.doi.org/10.3390/rs12172779>.
- Pang, K., Martínez, L., Li, N., Liu, J., Zou, L., Lu, M., 2024. A concept lattice-based expert opinion aggregation method for multi-attribute group decision-making with linguistic information. *Expert Syst.* Appl. 237, 121485. <http://dx.doi.org/10.1016/j.eswa.2023.121485>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0957417423019875>.
- Pe'er, G., Dicks, L.V., Visconti, P., Arlettaz, R., Báladi, A., Benton, T.G., Collins, S., Dieterich, M., Gregory, R.D., Hartig, F., Henle, K., Hobson, P.R., Kleijn, D., Neumann, R.K., Robijns, T., Schmidt, J., Schwartz, A., Sutherland, W.J., Turbé, A., Wulf, P., Scott, A.V., 2014. EU agricultural reform fails on biodiversity. *Science* 344 (6188), 1090–1092. <http://dx.doi.org/10.1126/science.1253425>, URL: <https://www.science.org/doi/10.1126/science.1253425>.
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurr, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. *Science* 339 (6117), 277–278. <http://dx.doi.org/10.1126/science.1229931>, URL: <https://www.science.org/doi/10.1126/science.1229931>.
- Perić, Z., Neukampf, R., Sinn, C., 2022. Geografisches hexagonales Gitter mit 1 Quadratkilometer Zellengröße für Deutschland – Geographical hexagonal grid with one square kilometer cell size for Germany. <http://dx.doi.org/10.5281/ZENODO.6623511>, Zenodo.
- Pettorelli, N., Wegmann, M., Skidmore, A., Muecher, S., Dawson, T.P., Fernandez, M., Lucas, R., Schapman, M.E., Wang, T., O'Connor, B., Jongman, R.H., Kempeneers, P., Sonnenschein, R., Leidner, A.K., Böhm, M., He, K.S., Nagendra, H., Dubois, G., Fatoyinbo, T., Hansen, M.C., Paganini, M., de Klerk, H.M., Asner, G.P., Kerr, J.T., Estes, A.B., Schmeller, D.S., Heiden, U., Rocchini, D., Pereira, H.M., Turak, E., Fernandez, N., Lausch, A., Cho, M.A., Alcaraz-Segura, D., McGeoch, M.A., Turner, W., Mueller, A., St-Louis, V., Penner, J., Vihervaara, P., Belward, A., Reyers, B., Geller, G.N., 2016. Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. In: Boyd, D. (Ed.), *Remote Sens. Ecol. Conserv.* 2 (3), 122–131. <http://dx.doi.org/10.1002/rse2.15>, URL: <https://onlinelibrary.wiley.com/doi/10.1002/rse2.15>.

- Petutschnig, L., Clemen, T., Klaußner, E.S., Clemen, U., Lang, S., 2024. Evaluating geospatial data adequacy for integrated risk assessments: A malaria risk use case. *ISPRS Int. J. Geo-Information* 13 (2), 33. <http://dx.doi.org/10.3390/ijgi13020033>, URL: <https://www.mdpi.com/2220-9964/13/2/33>.
- Piikki, K., Wetterlind, J., Söderström, M., Stenberg, B., 2021. Perspectives on validation in digital soil mapping of continuous attributes—A review. *Soil Use Manag.* 37 (1), 7–21. <http://dx.doi.org/10.1111/sum.12694>, URL: <https://onlinelibrary.wiley.com/doi/10.1111/sum.12694>.
- Pôças, I., Gonçalves, J., Marcos, B., Alonso, J., Castro, P., Honrado, J.P., 2014. Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring. *Int. J. Geogr. Inf. Sci.* 28 (11), 2356–2371. <http://dx.doi.org/10.1080/13658816.2014.924627>, URL: <http://www.tandfonline.com/doi/full/10.1080/13658816.2014.924627>.
- Preidl, S., Lange, M., Doktor, D., 2020a. Introducing APIC for regionalised land cover mapping on the national scale using sentinel-2A imagery. *Remote Sens. Environ.* 240, 111673. <http://dx.doi.org/10.1016/j.rse.2020.111673>.
- Preidl, S., Lange, M., Doktor, D., 2020b. Land cover classification map of Germany's agricultural area based on sentinel-2A data from 2016. <http://dx.doi.org/10.1594/PANGAEA.910837>.
- Riedesel, L., Ma, D., Piepho, H.-P., Laidig, F., Möller, M., Golla, B., Kautz, T., Feike, T., 2024. Climate change induced heat and drought stress hamper climate change mitigation in German cereal production. *Field Crop. Res.* 317, 109551. <http://dx.doi.org/10.1016/j.fcr.2024.109551>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0378429024003046>.
- Riedesel, L., Möller, M., Horney, P., Golla, B., Piepho, H.-P., Kautz, T., Feike, T., 2021. Timing and intensity of heat and drought stress determine wheat yield losses in Germany. In: Hamed, M.M. (Ed.), *PLoS One* 18 (7), e0288202. <http://dx.doi.org/10.1371/journal.pone.0288202>, URL: <https://dx.plos.org/10.1371/journal.pone.0288202>.
- Ritters, K.H., O'Neill, R.V., Hunsaker, C.T., Wickham, J.D., Yankee, D.H., Timmins, S.P., Jones, K.B., Jackson, B.L., 1995. A factor analysis of landscape pattern and structure metrics. *Landsc. Ecol.* 10 (1), 23–39. <http://dx.doi.org/10.1007/BF00158551>.
- Roilo, S., Paulus, A., Alarcón-Segura, V., Kock, L., Beckmann, M., Klein, N., Cord, A.F., 2024. Quantifying agricultural land-use intensity for spatial biodiversity modelling: implications of different metrics and spatial aggregation methods. *Landsc. Ecol.* 39 (3), 55. <http://dx.doi.org/10.1007/s10980-024-01853-9>, URL: <https://link.springer.com/10.1007/s10980-024-01853-9>.
- Rowland, J.A., Bland, L.M., James, S., Nicholson, E., 2021. A guide to representing variability and uncertainty in biodiversity indicators. *Conserv. Biol.* 35 (5), 1669–1682. <http://dx.doi.org/10.1111/cobi.13699>, URL: <https://onlinelibrary.wiley.com/doi/10.1111/cobi.13699>.
- Sandler, A.M., Rashford, B.S., 2018. Misclassification error in satellite imagery data: Implications for empirical land-use models. *Land Use Policy* 75, 530–537. <http://dx.doi.org/10.1016/j.landusepol.2018.04.008>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0264837717314242>.
- Schiller, J., Jänicke, C., Reckling, M., Ryo, M., 2024. Higher crop rotational diversity in more simplified agricultural landscapes in Northeastern Germany. *Landsc. Ecol.* 39 (4), 90. <http://dx.doi.org/10.1007/s10980-024-01889-x>, URL: <https://link.springer.com/10.1007/s10980-024-01889-x>.
- Schmeller, D.S., Mihoub, J.-B., Bowser, A., Arvanitidis, C., Costello, M.J., Fernandez, M., Geller, G.N., Hobern, D., Kissling, W.D., Regan, E., Saarenmaa, H., Turak, E., Isaac, N.J.B., 2017. An operational definition of essential biodiversity variables. *Biodivers. Conserv.* 26 (12), 2967–2972. <http://dx.doi.org/10.1007/s10531-017-1386-9>, URL: <http://link.springer.com/10.1007/s10531-017-1386-9>.
- Schneider, M., Gackstetter, D., Prexl, J., Meyer, S.T., Körner, M., 2025. Advancing transnational assessments of biodiversity drivers in European agriculture with an updated hierarchical crop and agriculture taxonomy (HCAT). *Npj Sustain. Agric.* 3 (1), 3. <http://dx.doi.org/10.1038/s44264-024-00037-x>, URL: <https://www.nature.com/articles/s44264-024-00037-x>.
- Schneider, M., Marchington, C., Körner, M., 2022. Challenges and opportunities of large transnational datasets: A case study on European administrative crop data. <http://dx.doi.org/10.48550/ARXIV.2210.07178>, ArXiv, URL: <https://arxiv.org/abs/2210.07178>.
- Schwieder, M., Erasmi, S., Nendel, C., Hostert, P., 2022. National-scale crop type maps for Germany from combined time series of sentinel-1, sentinel-2 and landsat 8 data (2020). <http://dx.doi.org/10.5281/ZENODO.6451848>, Zenodo.
- Schwieder, M., Tetteh, G.O., Blickensdörfer, L., Gocht, A., Erasmi, S., 2024. Agricultural land use (raster) : National-scale crop type maps for Germany from combined time series of sentinel-1, sentinel-2 and landsat data (2017 to 2021). <http://dx.doi.org/10.5281/zenodo.10640528>, URL: <https://zenodo.org/records/10640528>.
- Shang, R., Zhu, Z., 2019. Harmonizing landsat 8 and sentinel-2: A time-series-based reflectance adjustment approach. *Remote Sens. Environ.* 235, 111439. <http://dx.doi.org/10.1016/j.rse.2019.111439>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425719304584>.
- Smith, P., House, J.I., Bustamante, M., Sobocká, J., Harper, R., Pan, G., West, P.C., Clark, J.M., Adhya, T., Rumpel, C., Paustian, K., Kuikman, P., Cotrufo, M.F., Elliott, J.A., McDowell, R., Griffiths, R.I., Asakawa, S., Bondeau, A., Jain, A.K., Meersmans, J., Pugh, T.A.M., 2016. Global change pressures on soils from land use and management. *Global Change Biol.* 22 (3), 1008–1028. <http://dx.doi.org/10.1111/gcb.13068>.
- Sokolova, M., Japkowicz, N., Szpakowicz, S., 2006. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In: Sattar, A., Kang, B.-h. (Eds.), *AI 2006: Advances in Artificial Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1015–1021.
- Specka, X., Martini, D., Weiland, C., Arend, D., Asseng, S., Boehm, F., Feike, T., Fluck, J., Gackstetter, D., Gonzales-Mellido, A., Hartmann, T., Haunert, J.-H., Hoedt, F., Hoffmann, C., König, P., Lange, M., Lesch, S., Lindstädt, B., Lischeid, G., Möller, M., Rascher, U., Reif, J.C., Schmalzl, M., Senft, M., Stahl, U., Svoboda, N., Usadel, B., Webber, H., Ewert, F., 2023. FAIRagro: Ein Konsortium in der Nationalen Forschungsdateninfrastruktur (NFDI) für Forschungsdaten in der Agrosystemforschung: Herausforderungen und Lösungsansätze für den Aufbau einer FAIRen Forschungsdateninfrastruktur. *Inform. Spektrum* 46 (1), 24–35. <http://dx.doi.org/10.1007/s00287-022-01520-w>, URL: <https://link.springer.com/10.1007/s00287-022-01520-w>.
- Stehman, S.V., Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment. *Remote Sens. Environ.* 64 (3), 331–344. [http://dx.doi.org/10.1016/S0034-4257\(98\)00010-8](http://dx.doi.org/10.1016/S0034-4257(98)00010-8), URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425798000108>.
- Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. *Remote Sens. Environ.* 231, 111199. <http://dx.doi.org/10.1016/j.rse.2019.05.018>.
- Stehman, S.V., Wickham, J., 2020. A guide for evaluating and reporting map data quality: Affirming Shao et al. "Overselling overall map accuracy misinforms about research reliability". *Landsc. Ecol.* 35 (6), 1263–1267. <http://dx.doi.org/10.1007/s10980-020-01029-1>, URL: <https://link.springer.com/10.1007/s10980-020-01029-1>.
- Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morissette, J.T., Stehman, S.V., Woodcock, C.E., 2006. Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *Eur. Communities, Luxemb.* 51 (4), 1–60.
- Thas, O., 2010. Comparing Distributions. Springer series in statistics, Springer, New York ; London, <http://dx.doi.org/10.1007/978-0-387-92710-7>, oCLC:ocn401153891.
- Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. *Nature* 418 (6898), 671–677. <http://dx.doi.org/10.1038/nature01014>.
- Tonetti, V., Pena, J.C., Scarpelli, M.D., Sugai, L.S., Barros, F.M., Anunciação, P.R., Santos, P.M., Tavares, A.L., Ribeiro, M.C., 2023. Landscape heterogeneity: concepts, quantification, challenges and future perspectives. *Environ. Conserv.* 50 (2), 83–92. <http://dx.doi.org/10.1017/S0376892923000097>, URL: https://www.cambridge.org/core/product/identifier/S0376892923000097/type/journal_article.
- Tóth, K., Kučas, A., 2016. Spatial information in European agricultural data management: requirements and interoperability supported by a domain model. *Land Use Policy* 57, 64–79. <http://dx.doi.org/10.1016/j.landusepol.2016.05.023>.
- Uthes, S., Kelly, E., König, H.J., 2020. Farm-level indicators for crop and landscape diversity derived from agricultural beneficiaries data. *Ecol. Indic.* 108, 105725. <http://dx.doi.org/10.1016/j.ecolind.2019.105725>.
- Uuemaa, E., Antrop, M., Roosare, J., Marja, R., Mander, Ü., 2009. Landscape metrics and indices: An overview of their use in landscape research. *Living Rev. Landsc. Res.* (3), URL: <http://lrlr.landscapeonline.de/Articles/lrlr-2009-1/download/lrlr-2009-1Color.pdf>.
- Uuemaa, E., Mander, Ü., Marja, R., 2013. Trends in the use of landscape spatial metrics as landscape indicators: A review. *Ecol. Indic.* 28, 100–106. <http://dx.doi.org/10.1016/j.ecolind.2012.07.018>.
- Valle, D., Izbicki, R., Leite, R.V., 2023. Quantifying uncertainty in land-use land-cover classification using conformal statistics. *Remote Sens. Environ.* 295, 113682. <http://dx.doi.org/10.1016/j.rse.2023.113682>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S003442572300233X>.
- Van Den Brink, L., Janssen, P., Quak, W., Stoter, J., 2017. Towards a high level of semantic harmonisation in the geospatial domain. *Comput. Environ. Urban Syst.* 62, 233–242. <http://dx.doi.org/10.1016/j.compenurbysys.2016.12.002>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0198971516304136>.
- Vihervaara, P., Auvinen, A.-P., Mononen, L., Törmä, M., Ahlroth, P., Anttila, S., Böttcher, K., Forsius, M., Heino, J., Heliölä, J., Koskelainen, M., Kuussaari, M., Meissner, K., Ojala, O., Tuominen, S., Viitasalo, M., Virkkala, R., 2017. How essential biodiversity variables and remote sensing can help national biodiversity monitoring. *Glob. Ecol. Conserv.* 10, 43–59. <http://dx.doi.org/10.1016/j.gecco.2017.01.007>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S2351989416301482>.
- von Braun, J., 2007. The world food situation: New driving forces and required actions. *Internat. Food Policy Research Inst, International Food Policy Research Institute, Washington, DC*.
- Wentz, E., Shimizu, M., 2018. Measuring spatial data fitness-for-use through multiple criteria decision making. *Ann. Am. Assoc. Geogr.* 108 (4), 1150–1167. <http://dx.doi.org/10.1080/24694452.2017.1411246>.
- Wolff, S., Hüttel, S., Nendel, C., Lakes, T., 2021. Agricultural landscapes in Brandenburg, Germany: An analysis of characteristics and spatial patterns. *Int. J. Environ. Res.* 15 (3), 487–507. <http://dx.doi.org/10.1007/s41742-021-00328-y>.

- Wolz, A., 2013. The Organisation of Agricultural Production in East Germany Since World War II: Historical Roots and Present Situation; Organisation Der Agrarproduktion in Ostdeutschland Seit Dem 2. Weltkrieg: Historische Wurzeln Und Aktuelle Situation. Discussion Paper 139, Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO), Halle (Saale), URL: <https://hdl.handle.net/10419/83981>, urn:nbn:de:gbv:3:2-27321.
- Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P., 2016. The global Landsat archive: Status, consolidation, and direction. *Remote Sens. Environ.* 185, 271–283. <http://dx.doi.org/10.1016/j.rse.2015.11.032>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425715302194>.
- Yang, X., Blower, J.D., Bastin, L., Lush, V., Zabala, A., Masó, J., Cornford, D., Díaz, P., Lumsden, J., 2013. An integrated view of data quality in earth observation. *Philos. Trans. Ser. A, Math. Phys. Eng. Sci.* 371 (1983), 20120072. <http://dx.doi.org/10.1098/rsta.2012.0072>.
- Ye, S., Pontius, R.G., Rakshit, R., 2018. A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS J. Photogramm. Remote Sens.* 141, 137–147. <http://dx.doi.org/10.1016/j.isprsjprs.2018.04.002>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0924271618300947>.
- Zhang, Y., Li, J., Wang, Z., He, Z., Guan, Q., Lin, J., Yu, W., 2025. Geospatial large language model trained with a simulated environment for generating tool-use chains autonomously. *Int. J. Appl. Earth Obs. Geoinf.* 136, 104312. <http://dx.doi.org/10.1016/j.jag.2024.104312>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S1569843224006708>.
- Zhang, Y., Wei, C., He, Z., Yu, W., 2024. GeoGPT: An assistant for understanding and processing geospatial tasks. *Int. J. Appl. Earth Obs. Geoinf.* 131, 103976. <http://dx.doi.org/10.1016/j.jag.2024.103976>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S1569843224003303>.
- Zhao, T., Wang, S., Ouyang, C., Chen, M., Liu, C., Zhang, J., Yu, L., Wang, F., Xie, Y., Li, J., Wang, F., Grunwald, S., Wong, B.M., Zhang, F., Qian, Z., Xu, Y., Yu, C., Han, W., Sun, T., Shao, Z., Qian, T., Chen, Z., Zeng, J., Zhang, H., Letu, H., Zhang, B., Wang, L., Luo, L., Shi, C., Su, H., Zhang, H., Yin, S., Huang, N., Zhao, W., Li, N., Zheng, C., Zhou, Y., Huang, C., Feng, D., Xu, Q., Wu, Y., Hong, D., Wang, Z., Lin, Y., Zhang, T., Kumar, P., Plaza, A., Chanussot, J., Zhang, J., Shi, J., Wang, L., 2024. Artificial intelligence for geoscience: Progress, challenges, and perspectives. *Innov.* 5 (5), 100691. <http://dx.doi.org/10.1016/j.xinn.2024.100691>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S2666675824001292>.