The first map of dominant crop sequences in the European Union over 2012-2018

Rémy Ballot[1], Nicolas Guilpart[1], Marie-Hélène Jeuffroy[1]

[1] Université Paris-Saclay, AgroParisTech, INRAE, UMR Agronomie, 78850, Thiverval-Grignon, France

5 *Correspondence to*: Rémy Ballot (remy.ballot@inrae.fr)

Abstract. Crop diversification is considered as an important linchpin of the agroecological transition, whereas current dominant cropping systems are known to rely only on a few crops species – like cereals in the European Union (EU). To assess the benefits of crop diversification at large scale, an accurate description of current crop sequences is required as a baseline. However, such a description is lacking at the scale of the EU. Here, we developed the first map of dominant crop sequences

10 in the EU for the period 2012-2018. We used the LUCAS dataset that provides temporally-incomplete land cover information from a stable grid of points covering the whole EU. Eight crop sequence types were identified using hierarchical clustering implemented on LUCAS data, and mapped over EU. We show, in France, that the relative importance of these eight crop sequence types (as estimated from LUCAS data) was highly consistent with those derived from an almost spatially-exhaustive temporally-complete national dataset (the French Land Parcel Identification System) for the same period, thus validating the

15 method and the typology for this country. Land use (i.e. crop production area) derived from our map of dominant crop sequences was also highly consistent with land use reported by official statistics, both at national and EU levels, validating the approach at the EU-scale. This first map of dominant crop sequences in the EU should be useful for future studies dealing with agricultural issues that are sensitive to crop rotations. The map of dominant crop sequences types in the EU derived from our work is available at https://doi.org/10.5281/zenodo.7016986 (Ballot et al., 2022).

20 **1 Introduction**

Crop diversification – the increase of crop diversity from field to national scale – is considered as an essential linchpin of the agroecological transition (Kremen et al., 2012; Lechenet et al., 2016; Renard and Tilman, 2021; Wezel et al., 2014), whereas current crop sequences are short and specialized in many parts of the world (Meynard et al., 2018; Salembier et al., 2016; Schott et al., 2010; Stein and Steinmann, 2018). Indeed, crop diversification supports a number of ecosystem services, with

25 positive effects on crop yield, soil fertility, nutrient cycling, carbon sequestration, climate and water regulation, pest control, biodiversity, and pollination (Beillouin et al., 2019, 2021a; Tamburini et al., 2020). Strategies of crop diversification practices include crop rotation, cultivar mixture, cover crops, intercropping and agroforestry. Among these strategies, diversifying the crop sequence is of particular interest in order to: (i) increase nutrient availability and limit fertilizer requirement (e.g. by the inclusion of legume crops), (ii) favor soil protection and conservation by enhancing soil cover, (iii) promote natural regulation

30   of pests and diseases by avoiding the presence of successive host crops, and (iv) reduce weed infestation (Wezel et al., 2014). Aiming at leveraging these benefits, several foresight studies have assessed scenarios for the future of agriculture in which crop sequences are modified to increase the services they provide (Billen et al., 2021; Poux and Aubert, 2018), or scenarios of organic farming expansion where significant differences in crop rotations between conventional and organic agriculture are acknowledged (Barbieri et al., 2017, 2019, 2021).

35   A robust assessment of benefits expected from crop diversification at large scale requires an accurate description of current crop sequences as a baseline. Indeed several ecosystem services provided by diversified farming systems depend on the pre-crop effect, thus needing to know precisely the nature of the crops successively grown on a same field (Bennett et al., 2012). However, such a description is still lacking for Europe, for at least two reasons: (i) existing datasets at the European scale, e.g. the LUCAS dataset, provide information about land use categories and crop species cultivated on agricultural land, but no

40   information about crop sequences because data are not available every year (e.g. the LUCAS dataset provides information every 3 years), (ii) even if national and sub-national datasets describing crop sequences are available within some EU-state members, e.g. the Land Parcel Identification Systems in France (Levavasseur et al., 2016), a lack of harmonization (e.g. spatial and temporal resolution) between these datasets makes them difficult to use in a cross-analysis. This situation has resulted in the production of regional to national studies of crop sequences (Levavasseur et al., 2016; Peltonen-Sainio and Jauhiainen,

45   2019; Stein and Steinmann, 2018; Xiao et al., 2014) but has hampered any analysis at the scale of the European Union (EU). Such an analysis would be useful, especially in the context of the EU's Farm to Fork strategy, in which the adoption of more diverse crop rotations is encouraged (EU, 2020). To overcome this problem, we developed an original method that combines European-level and national-level datasets to create a map of current dominant crop sequences at the European level.

2. Materials and Methods

50   **2.1 Method overview**

The purpose of this work is to map current dominant crop sequences from the European Land Use Cover Area frame statistical Survey (LUCAS, d'Andrimont et al. 2020). As this dataset is temporally incomplete (observations of land use on fixed points, only every three years), we proceeded in three steps (Figure 1) to assess how this incomplete information could be used to describe the diversity of crop sequences across the EU. First, we preprocessed the LUCAS data, by selecting points under non-

55   perennial agricultural land cover. For each of these points, we calculated frequencies of eight crops or groups of crops across 2012, 2015 and 2018 (the three last years of observation). Second, based on this set of eight variables, LUCAS points were classified into eight groups (hereafter referred to as crop sequence types), combining a principle component analysis (PCA) with hierarchical clustering. Third, we assessed the consistency of this classification against crop sequences derived with the quasi-exhaustive yearly French Land Parcel Identification System (LPIS), and with the Eurostat crop production data. All data

60   sources used in this study are detailed in Table 1.

2.2. Data sources

2.2.1. The LUCAS dataset

The Land Use Cover Area frame Survey (LUCAS) is an in-situ land-cover (i.e. physical cover observed at the earth's surface, e.g. cereals, root crops, fodder crops) and land-use (i.e. socio-economic function of the observed earth's surface, e.g.
65   agriculture, industry, residential) observation, carried out on approximately 300 000 points sampled from a stable grid of around 1 100 000 points across EU-28 from 2006, every three years. Last observation campaign was realized in 2018. The work presented here is based on the harmonized LUCAS database developed by d'Andrimont et al. (2020), which gathered, in a unique dataset, all the information collected from the beginning of this survey.

2.2.2. The French LPIS dataset

70   The French Land Parcel Identification System (LPIS) is based on yearly declarations made by farmers in compliance with Common Agricultural Policy subsidies. From 2008, it provides an annual almost spatially-exhaustive information of land use for agricultural land, detailing 28 crop categories until 2014, and approximately 300 from 2015. For our study, we used the "RPG Explorer Crop successions France version 2.0" dataset developed by Martin et al. (2021). This dataset compiles all annual LPIS data for France into a unique dataset of crop sequences for more than 18 million fields over the period 2012-2018,
75   and associated areas. This dataset represents the most exhaustive dataset of crop sequences in France, in terms of both spatial and temporal resolution.

2.2.3. The Eurostat dataset

The Eurostat crop production dataset provides harmonized information about annual crop-specific acreage per country from the year 2000 (EUROSTAT, 2022).

80   ### 2.3   Identification and mapping of dominant crop sequence types in the EU based on LUCAS data

2.3.1   Preprocessing of LUCAS data

From the harmonized LUCAS database (d'Andrimont et al., 2020), we selected points with observations for the three most recent campaigns (i.e. 2012, 2015, and 2018). We assumed the seven-year time period long enough to encompass the duration of main crop rotations, and thus we do not consider older campaigns (i.e. 2006 and 2009) which may be outdated to represent
85   current crop sequences. We discarded points under non-agricultural use or permanent agricultural use (e.g. orchards, vineyards) during the last observation (i.e. Land Cover (LC) not included in land cover classification B11 to B55). As an exception, we also conserved points identified as permanent grasslands in 2018 (i.e. LC E20), but identified as a non-permanent agricultural use in 2012 or 2015, as they may be the result of a confusion between temporary and permanent grassland during observation. This resulted in a dataset of 31 159 points, with information about crops cultivated in 2012, 2015 and 2018. If the whole 2006-

90   2018 period was considered, only 9 094 points would have remained with information about crop cultivated in 2006, 2009, 2021, 2015 and 2018. This choice would also have limited the analysis to 11 among the 27 EU countries (i.e. Belgium, Czechia, Germany, Spain, France, Hungary, Italy, Luxembourg, The Netherlands, Poland and Slovakia). This is another reason why we considered only the three most recent campaigns.

Depending on the purpose of analysis, crop sequences can be described in many different ways, including length and flexibility
95   of rotation (Castellazzi et al., 2008), nature and function of crops or crop groups (Barbieri et al., 2017; Videla-Mensegue et al., 2022), diversity of crops or crop groups (Beillouin et al., 2021b; Tamburini et al., 2020), diversity of sowing dates (Weisberger et al., 2019), order of succession of crops within the sequence (Peltonen-Sainio et al., 2019), and temporal frequency or return time of crops within the sequence (Nowak et al., 2022). Given the characteristics of the LUCAS dataset, the exact order of crop succession within the sequences could not be described, nor could be the variety of sowing dates (e.g. the LUCAS dataset
100   provides no information about whether the crop is sown in autumn or spring). Therefore, we chose to describe current dominant crop sequences by the temporal frequencies, over three years (i.e. 2012, 2015 and 2018), of eight crops or groups of crops described below. This choice is a compromise between addressing common issues related to crop rotations (e.g. nitrogen management and the role of legumes) and avoiding unnecessary complexity. These eight groups were defined based on two criteria: (i) their agronomic relevance, and (ii) their importance in terms of cultivated area in Europe. Groups considered were
105   (i) cereals (i.e. wheat, barley, oat, triticale, rye, corresponding to LC B11 to B19 in the LUCAS dataset nomenclature), (ii) corn (LC B16), (iii) rapeseed (LC B32), (iv) sunflower (LC B31), (v) pulses (i.e. dry pulses and soybean, corresponding to LC B33 and B41), (vi) root crops (i.e. beets and potatoes, corresponding to LC B21 to B23), (vii) forage legumes (i.e. alfalfa and clover, corresponding to LC B51 and B52) and (viii) temporary grassland (LC B53, B55, and E20).

2.3.2   Identification of crop sequence types with hierarchical clustering

110   A principle component analysis (PCA) was performed with the PCA() function of the R package FactoMineR v2.1 (Lê et al., 2008) on all the eight variables describing temporal frequencies of cereals, rapeseed, sunflower, pulses, corn, root crops, forage legumes, and temporary grassland. Seven components displayed eigenvalues higher than or close to 1, cumulating more than 99% of variance (Figure 2a). These seven components were used as inputs for a hierarchical clustering, performed with the HCPC() function from the same R package. Inertia gain displayed a clear break from eight to nine (Figure 2b), thus eight
115   groups of crop sequences (crop sequence types) were considered. To analyze the geographical distribution of dominant crop sequences, observed locations of these eight crop sequence types were mapped across the whole EU.

2.3.3   Quality assessment

The crop sequence types derived from incomplete temporal sequences (LUCAS dataset) were compared to crop sequence types derived from complete temporal sequences (French LPIS dataset). To this aim, a three-step procedure was followed.
120   First, a random forest (RF) model was trained on LUCAS data to predict crop sequence type from crop (or crop group) frequencies (i.e. a total of eight predictors). The RF model was fitted using the randomForest() function of the R package

randomForest v4.6.14 (Liaw and Wiener, 2002), with default settings. The RF model showed good performances, as indicated by an out-of-bag (OOB) error rate of only 0.12% (Table 3). Second, the RF model was applied on French LPIS data to classify observed crop sequences into the eight crop sequence types. Third, the crop sequence type distributions derived from the
125 LUCAS and from French LPIS datasets were compared at the national and regional levels in France. This allowed to check if the lower temporal and spatial resolutions of the LUCAS dataset led to some bias in the identification of crop sequence types and their geographical distribution.

To ensure the map of crop sequence types presented in this paper could be used as a baseline in studies assessing scenarios for the future of agriculture in the EU, it appears essential to assess its consistency with observed land use. To this aim, average
130 crop or crop group harvested areas between 2012 and 2018 were calculated from (i) their relative frequencies within the crop sequence types derived from the LUCAS dataset, (ii) relative importance of crop sequence types, and (iii) total arable land area. This calculation was applied for each country in the EU, at the national level, and compared to crop- or crop group harvested areas reported by Eurostat (2022).

3 Results

3.1 Eight types of crop sequences identified based on the LUCAS dataset

Eight different types of crop sequences were identified based on the LUCAS dataset (Table 2). Two crop sequence types largely dominate in terms of relative importance within the EU: (i) the first one accounts for 40% of LUCAS points, and is dominated by temporary grasslands (90%), followed by cereals (4%) and corn (2%); (ii) the second one accounts for 21% of LUCAS points, and is dominated by cereals (87%), followed by corn (6%) and grasslands (4%). Together, these two crop
140 sequence types represent 61% of LUCAS points. Then comes a group of three crop sequence types of moderate relative importance, each one representing 7% to 9% of LUCAS points. This group includes a crop sequence type dominated by rapeseed (43%) and cereals (43%), a second one dominated by corn (81%) and cereals (15%), and a third one dominated by root crops (43%) and cereals (35%). The remaining three crop sequence types can be considered as minor as they individually account for only 5% of LUCAS points. This includes a crop sequence type dominated by forage legumes (47%) and cereals
145 (23%), a second one dominated by sunflower (42%) and cereals (39%), and a third one dominated by pulses (37%) and cereals (36%).

Three (groups of) crops are present in all crop sequence types: cereals, corn and temporary grassland (Table 2). This highlights the central role of these crops in crop rotations in the EU. On the other hand, pulses, and to a lesser extent sunflower, root crops and forage legumes, appear only in few (i.e. 1 to 3) crop sequence types. Some crop sequence types show a high degree of diversity, like the "pulses and cereals" in which all (groups of) crops are represented, while some others are much less
150 diverse, like the "cereals" and "corn and cereals" types. These two crop sequence types account for almost one third (29%) of all LUCAS points and are composed of cereals and corn by more than 90%.

If all LUCAS campaigns available were considered (i.e. 2006 to 2018), the classification would have resulted in the same eight types of crop sequences. The eight types would have ranked in the same order, regarding their relative importance. Their

155   relative importance would have changed slightly: 29 instead of 40% for temporary grassland dominated crop sequence type, 21 instead of 22% for cereals one, 14 instead of 8%, for corn and cereals, 13 instead of 9% for rapeseed and cereals, 7 instead of 5% for forage legumes and cereals, 6 instead of 5% for pulses and cereals, 5% for sunflower and cereals (relative importance unchanged), 5 instead of 7% for root crops and cereals. These differences were consistent as 16 among 27 countries were discarded if we considered only points with observations for all the 2006 to 2018 LUCAS campaigns.

160   **3.2   Quality assessment**

3.2.1   Incomplete temporal crop sequences are a good proxy of complete temporal crop sequences at national and regional scale in France

On average at the national scale in France, the relative importance of crop sequence types estimated from the LUCAS dataset is in good agreement (R^2=0.93, RMSE=0.04) with estimates based on the French LPIS dataset, with no systematic bias (Figure

165   3, Figure S1). However, estimates based on the LUCAS dataset overestimate the proportion of the "grasslands" crop sequence type, and slightly underestimate the proportion of the "rapeseed and cereals", "root crops and cereals", and "forage legumes and cereals" ones.

At the regional scale, the agreement between the relative importance of crop sequence types derived from the LUCAS or French LPIS datasets remains good, but varies by crop group (Figure 4). The crop sequence types "cereals", "corn and cereals",

170   "forage legumes and cereals" and "sunflower and cereals" show the best agreement, with high R^2 (higher than 0.8), low RMSE (lower than 0.05), and no systematic bias, whereas the crop sequence type "pulses and cereals" shows the lowest agreement (R^2=0.26, RMSE=0.02). Overestimation of the crop sequence type "grasslands" for almost all regions, and underestimation of the crop sequence types "rapeseed and cereals" and "root crops and cereals for the regional highest values are confirmed.

Overall, regional differences in crop sequence types are well captured with the LUCAS dataset, as shown by two conclusions.

175   First, any given crop sequence type may have a low relative importance in one region and a high one in another. This range of relative importance across regions is well captured by estimates derived from the LUCAS dataset for all crop sequence types (Figure 4, Figure S2). Second, regional specificities in terms of dominant crop sequences types are also well captured by LUCAS data (Figure S2). For example, the region Picardie is dominated by three crop sequence types: "root crops and cereals", "cereals", and "rapeseed and cereals"; while the region Aquitaine is dominated by the crop sequence types: "grasslands", "corn

180   and cereals", and "sunflower and cereals".

3.2.2   Land use derived from crop sequence types is in good agreement with land use reported by official statistics at national and EU levels

Comparison of crop harvested areas derived either from crop sequence types or from official statistics (EUROSTAT, 2022) shows good agreement at the EU scale (Figure 5), with R^2 higher than 0.98 and no bias. Comparison for each country of the

185 EU (Figure 6) reveals good levels of correlation between areas derived from crop sequence types and areas reported by Eurostat with R^2 ranging from 0.87 for forage legumes, to 0.99 for grasslands. Nevertheless, some crop-country combinations display bad agreement between estimated and observed harvested areas (Figure S3). Harvested area of forage legumes are overestimated in Spain and France, whereas they are underestimated in Germany. Regarding corn, harvested area is overestimated for Spain. For root crops, harvested areas is also overestimated for Spain, but underestimated for Cyprus and

190 Montenegro. Sunflower harvested area is overestimated in Spain and underestimated for Romania. Cereals harvested area is overestimated for Spain.

3.3   Spatial distribution of crop sequence types in Europe

The maps of crop sequence types presented in Figure 7 and Figure 8 show a strong spatial pattern of crop sequence types distribution in the EU. The crop sequence types "grasslands", "corn and cereals", "cereals", and, to a lesser extent, "pulses and

195 cereals" and "forage legumes and cereals" are present in all EU countries, whereas the other crop sequence types are concentrated in specific parts of the EU. Indeed, "rapeseed and cereals" and "root crops and cereals" are mainly found in the north of France, Belgium, The Netherlands, Germany, Czech Republic, and Poland, whereas "sunflower and cereals" is mainly found in the south of France, Spain, Hungary, Bulgaria, and Romania. These differences are best highlighted when looking at the latitudinal distribution of crop sequence types shown in Figure 9: Indeed "grasslands", "pulses and cereals", and "cereals"

200 are found in a wide range of latitudes (from 40°N to 60°N) while other crop sequence types are limited to narrower ranges of latitudes with some in the north, like "root crops and cereals" that is concentrated between 50°N and 55°N, and some in the south, like "sunflower and cereals" that is concentrated between 40°N and 50°N. Most crop sequence types display regions in which they are more concentrated, like the north of France and Belgium for "root crops and cereals" or central Italy for "forage legumes and cereals". However, the "pulses and cereals" crop sequence type seems more regularly distributed from 40°N to

205 60°N. Several crop sequence types coexist in most regions. For example, all crop sequence types except "sunflower and cereals" and "forage legumes and cereals" are found in the very north of France (Figure 8). This demonstrates the importance of considering crop sequences in addition to land use, which do not reflect this diversity.

4. Discussion

The maps of crop sequence types derived from our study fill an important void in our knowledge of the spatial distribution of

210 crop sequences in the EU. Despite they were based on uncomplete data, both in time and space (the LUCAS dataset), these maps have been shown to represent well the distribution of crop sequence types in France at both national and regional scales when compared to an almost spatially-exhaustive dataset of temporally complete crop sequences in France. This consistency shows that the temporal incomplete information from LUCAS (i.e. only three crops known – 2012, 2015, and 2018 – from a seven-year crop sequence) allows to distinguish crop sequence types. While the quality assessment of this map of crop

215 sequence types in the EU would benefit from comparisons with other datasets at national or subnational levels in other

countries, we nevertheless highlight that all crop sequence types are present in France allowing for a complete quality assessment over the eight crop sequence types.

Yet, the comparison with the French LPIS dataset reveals an overestimation of grassland crop sequences, which leads to an underestimation of other crop sequence types. A possible explanation for this overestimation could be the inability of the three
220   sample years available in the LUCAS dataset (i.e. 2012, 2015, 2018) to capture the full diversity of longer crop rotations, such as rotations including temporary grassland. For example, let's consider a cyclical crop rotation starting in 2012 with 3 years of consecutive grassland followed by wheat, maize and barley. Then, observation in 2012, 2015 and 2018 would be grassland, wheat, and grassland respectively, yielding a proportion of grassland of two-third instead of half. Another possible explanation could be that temporary grassland may sometimes be confused with permanent grassland, during LUCAS observations, as
225   well as during farmers' declaration in the French LPIS.

However, the comparison between land use derived from our map of crop sequence types and land use derived from official statistics revealed good consistency at EU level, except for some crop-country combinations. This could be partly explained by the spatial sampling effort of LUCAS data, which can be quantified as the number of LUCAS points per unit area of cropland by country. Analysis of this variable reveals important variation across countries, with the spatial sampling effort
230   ranging from 0.13 (Hungary) to 1.77 (Montenegro) points per thousand hectares of cropland, with a median of 0.31 (Table 4). Accuracy of the map of crop sequence types should therefore be analyzed in the light of this varying spatial sampling effort. For example, our results pointed harvested areas overestimated for most (groups of) crops for Spain, which displays a sampling effort lower than the median (0.23). On the other hand, Montenegro displays the highest sampling effort (1.77), but also an important discrepancy between estimated and observed harvested areas. Indeed, this high rate hides a low number of
235   observations (15) for a small arable land. Thus the map accuracy may be limited either by a low sampling effort or by a low number of observations in small-sized countries.

We assume our maps of crop sequences to be useful for future studies dealing with agricultural issues that are sensitive to crop rotations, including nature-based pest control (Lechenet et al., 2016), pesticide use intensity (Jacquet et al., 2011), nitrogen management and cycling (De Notaris et al., 2018), biodiversity (Li et al., 2021), and crop yield (Bennett et al., 2012). We also
240   highlight that the methodology presented in this paper that leverages temporally incomplete information about crop sequences based on the LUCAS dataset can be adapted to specific objectives of future studies. For example, looking at the specific place of soybean in crop sequences in the EU would probably require to consider this crop specifically, instead of part of the group "pulses". Here we chose the eight crops or groups of crops for their relevance to major agronomical issues (see methods, section 2.3.1). Cereals (wheat, barley, oat, triticale, rye) are all species of the Poaceae family, sharing similar features regarding
245   their effects on crop sequences. For example, they all have straws with a relatively high C:N ratio, which contributes to limit nitrogen availability for the subsequent crop, when not exported from the field (Justes et al., 2009). Pulses (e.g. pea, fababean, soybean) are all from the Fabaceae family, and have the ability to fix nitrogen from the atmosphere, so that they are usually not fertilized with nitrogen, and contribute to increase nitrogen availability to subsequent crops thanks to the relatively low C:N ratio of their crop residues (Guinet et al., 2020). Root crops (beets and potatoes) may have a specific effect on topsoil and

Earth System
Science
Data

250 subsoil structure as belowground organs are harvested (Gabarron-Galeote et al., 2019). Forage legumes and temporary grasslands are commonly cultivated for at least two or three years, and present numerous benefits, including controlling weeds (Martin et al., 2020). Corn, sunflower and rapeseed are three widely grown crops, and have specific characteristics including low pesticide use for sunflower, high insecticide use for rapeseed (Urruty et al., 2016), and high irrigation requirements for maize, especially in the south of Europe (Senthilkumar et al., 2015).

255 Based on these agronomic knowledge, we assume that the crop sequence types identified may also go with differentiated cultivation practices, i.e. winter wheat within a "rapeseed and cereals" crop sequence may be cultivated differently as within a "pulses and cereals" one. Thus, these crop sequence types may display contrasted environmental, economic and social performances. The relevance of this crop sequence typology to differentiate cropping systems with contrasted performances could be assessed by combining a dataset including crop sequence with a one gathering crop management information.

260 We hope the methodology developed here will stimulate a more detailed description of crop sequences, both in time and space, and their effect on agricultural production and sustainability in future studies.

Data availability

Data described in this manuscript can be accessed at https://doi.org/10.5281/zenodo.7016986 (Ballot et al., 2022).

Code availability

265 Code used to produce data described in this manuscript, as well as to create figures and tables, can be accessed at https://doi.org/10.5281/zenodo.7018283

Author contribution

RB and MHJ initiated research. RB, MHJ and NG designed research. RB performed the analysis with help of NG. RB and NG wrote the manuscript with contributions from MHJ.

270 **Competing interests**

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by the LegValue project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement N°727672. Access to some confidential data, on which is based this work, has been made
275 possible within a secure environment offered by CASD – Centre d'accès sécurisé aux données (Ref. 10.34724/CASD)

References

Ballot, R., Guilpart, N., & Jeuffroy, M.H.. (2022). Harmonised LUCAS database classified by crop sequence type (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7016987

Barbieri, P., Pellerin, S., Nesme, T., 2017. Comparing crop rotations between organic and conventional farming. Scientific
280 Reports 7, 1–10. https://doi.org/10.1038/s41598-017-14271-6

Barbieri, P., Pellerin, S., Seufert, V., Nesme, T., 2019. Changes in crop rotations would impact food production in an organically farmed world. Nat. Sustain. 2, 378–385. https://doi.org/10.1038/s41893-019-0259-5

Barbieri, P., Pellerin, S., Seufert, V., Smith, L., Ramankutty, N., Nesme, T., 2021. Global option space for organic agriculture is delimited by nitrogen availability. Nat Food 2, 363–372. https://doi.org/10.1038/s43016-021-00276-y

285 Beillouin, D., Ben-ari, T., Makowski, D., 2019. Evidence map of crop diversification strategies at the global scale. Environmental Research Letters 14, 123001. https://doi.org/10.1088/1748-9326/ab4449

Beillouin, D., Ben-Ari, T., Malézieux, E., Seufert, V., Makowski, D., 2021a. Positive but variable effects of crop diversification on biodiversity and ecosystem services. Global Change Biology 27, 4697–4710. https://doi.org/10.1111/gcb.15747

Beillouin, D., Pelzer, E., Baranger, E., Carrouée, B., Cernay, C., de Chezelles, E., Schneider, A., Jeuffroy, M.H., 2021b.
290 Diversifying cropping sequence reduces nitrogen leaching risks. Field Crops Research 272. https://doi.org/10.1016/j.fcr.2021.108268

Bennett, A.J., Bending, G.D., Chandler, D., Hilton, S., Mills, P., 2012. Meeting the demand for crop production: the challenge of yield decline in crops grown in short rotations. Biol. Rev. 87, 52–71. https://doi.org/10.1111/j.1469-185X.2011.00184.x

Billen, G., Aguilera, E., Einarsson, R., Garnier, J., Gingrich, S., Grizzetti, B., Lassaletta, L., Le Noë, J., Sanz-Cobena, A.,
295 2021. Reshaping the European agro-food system and closing its nitrogen cycle: The potential of combining dietary change, agroecology, and circularity. One Earth 4, 839–850. https://doi.org/10.1016/j.oneear.2021.05.008

Castellazzi, M.S., Wood, G.A., Burgess, P.J., Morris, J., Conrad, K.F., Perry, J.N., 2008. A systematic representation of crop rotations 97, 26–33. https://doi.org/10.1016/j.agsy.2007.10.006

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H.I., Joebges,
300 C., Lemoine, G., van der Velde, M., 2020. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. Scientific Data 7, 1–15. https://doi.org/10.1038/s41597-020-00675-z

De Notaris, C., Rasmussen, J., Sørensen, P., Olesen, J.E., 2018. Nitrogen leaching: A crop rotation perspective on the effect of N surplus, field management and use of catch crops. Agriculture, Ecosystems & Environment 255, 1–11. https://doi.org/10.1016/j.agee.2017.12.009

305 EU, 2020. Farm to fork strategy: for a fair, healthy and environmentally-friendly food system. DG SANTE/Unit 'Food information and composition, food waste''.

EUROSTAT, 2022. Crop production in EU standard humidity. URL https://ec.europa.eu/eurostat/web/main/data/database?p_p_id=NavTreeportletprod_WAR_NavTreeportletprod_INSTANCE_ nPqeVbPXRmWQ&p_p_lifecycle=0&p_p_state=pop_up&p_p_mode=view&_NavTreeportletprod_WAR_NavTreeportletpr 310 od_INSTANCE_nPqeVbPXRmWQ_nodeInfoService=true&nodeId=1001945

Gabarron-Galeote, M.A., Hannam, J.A., Mayr, T., Jarvis, P.J., 2019. BEETSOIL: a decision support tool for forecasting the impact of soil conditions on sugar beet harvest. Soil and Tillage Research 191, 131–141. https://doi.org/10.1016/j.still.2019.04.001

Guinet, M., Nicolardot, B., Voisin, A.-S., 2020. Provision of contrasted nitrogen-related ecosystem services among grain 315 legumes. Agron. Sustain. Dev. 40, 33. https://doi.org/10.1007/s13593-020-00637-0

Jacquet, F., Butault, J.-P., Guichard, L., 2011. An economic analysis of the possibility of reducing pesticides in French field crops. Ecological Economics 70, 1638–1648. https://doi.org/10.1016/j.ecolecon.2011.04.003

Justes, E., Mary, B., Nicolardot, B., 2009. Quantifying and modelling C and N mineralization kinetics of catch crop residues in soil: parameterization of the residue decomposition module of STICS model for mature and non mature residues. Plant Soil 320 325, 171–185. https://doi.org/10.1007/s11104-009-9966-4

Kremen, C., Iles, A., Bacon, C., 2012. Diversified Farming Systems: An Agroecological, Systems-based Alternative to Modern Industrial Agriculture. E&S 17, art44. https://doi.org/10.5751/ES-05103-170444

Lê, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. Journal of statistical software 25, 1–18.

325 Lechenet, M., Makowski, D., Py, G., Munier-Jolain, N., 2016. Profiling farming management strategies with contrasting pesticide use in France. Agric. Syst. 149, 40–53. https://doi.org/10.1016/j.agsy.2016.08.005

Levavasseur, F., Martin, P., Bouty, C., Barbottin, A., Bretagnolle, V., Thérond, O., Scheurer, O., Piskiewicz, N., 2016. RPG Explorer: A new tool to ease the analysis of agricultural landscape dynamics with the Land Parcel Identification System. Computers and Electronics in Agriculture 127, 541–552. https://doi.org/10.1016/j.compag.2016.07.015

330 Li, M., Guo, J., Ren, T., Luo, G., Shen, Q., Lu, J., Guo, S., Ling, N., 2021. Crop rotation history constrains soil biodiversity and multifunctionality relationships. Agriculture, Ecosystems & Environment 319, 107550. https://doi.org/10.1016/j.agee.2021.107550

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18–22.

Martin, G., Durand, J.-L., Duru, M., Gastal, F., Julier, B., Litrico, I., Louarn, G., Médiène, S., Moreau, D., Valentin-Morison,

335   M., Novak, S., Parnaudeau, V., Paschalidou, F., Vertès, F., Voisin, A.-S., Cellier, P., Jeuffroy, M.-H., 2020. Role of ley pastures in tomorrow's cropping systems. A review. Agron. Sustain. Dev. 40, 17. https://doi.org/10.1007/s13593-020-00620-9

Martin, P., Rabenandrasana, N., Poméon, T., Serard, P., 2021. RPG Explorer Crop successions France 2007-2014, 2007-2019, 2015-2019 v2. Portail Data INRAE.

Meynard, J.-M., Charrier, F., Fares, M., Le Bail, M., Magrini, M.-B., Charlier, A., Messean, A., 2018. Socio-technical lock-in

340   hinders crop diversification in France. Agron. Sustain. Dev. 38, 54. https://doi.org/10.1007/s13593-018-0535-1

Nowak, B., Michaud, A., Marliac, G., 2022. Assessment of the diversity of crop rotations based on network analysis indicators. Agricultural Systems 199, 103402. https://doi.org/10.1016/j.agsy.2022.103402

Peltonen-Sainio, P., Jauhiainen, L., 2019. Unexploited potential to diversify monotonous crop sequencing at high latitudes. Agricultural Systems 174, 73–82. https://doi.org/10.1016/j.agsy.2019.04.011

345   Peltonen-Sainio, P., Jauhiainen, L., Honkavaara, E., Wittke, S., Karjalainen, M., Puttonen, E., 2019. Pre-crop values from satellite images for various previous and subsequent crop combinations. Frontiers in Plant Science 10, 1–13. https://doi.org/10.3389/fpls.2019.00462

Poux, X., Aubert, P.-M., 2018. An agroecological Europe in 2050: multifunctional agriculture for healthy eating. Findings from the Ten Years For Agroecology (TYFA) modelling exercise, Study N°09/18. Paris.

350   Renard, D., Tilman, D., 2021. Cultivate biodiversity to harvest food security and sustainability. Current Biology 31, R1154–R1158. https://doi.org/10.1016/j.cub.2021.06.082

Salembier, C., Elverdin, J.H., Meynard, J.-M., 2016. Tracking on-farm innovations to unearth alternatives to the dominant soybean-based system in the Argentinean Pampa. Agron. Sustain. Dev. 36, 1. https://doi.org/10.1007/s13593-015-0343-9

Schott, C., Mignolet, C., Meynard, J.-M., 2010. Les oléoprotéagineux dans les systèmes de culture : évolution des assolements

355   et des successions culturales depuis les années 1970 dans le bassin de la Seine. OCL 17, 276–291. https://doi.org/10.1051/ocl.2010.0334

Senthilkumar, K., Bergez, J.-E., Leenhardt, D., 2015. Can farmers use maize earliness choice and sowing dates to cope with future water scarcity? A modelling approach applied to south-western France. Agricultural Water Management 152, 125–134. https://doi.org/10.1016/j.agwat.2015.01.004

360   Stein, S., Steinmann, H.-H., 2018. Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems – A case study from Central Europe. European Journal of Agronomy 92, 30–40. https://doi.org/10.1016/j.eja.2017.09.010

Tamburini, G., Bommarco, R., Wanger, T.C., Kremen, C., van der Heijden, M.G.A., Liebman, M., Hallin, S., 2020. Agricultural diversification promotes multiple ecosystem services without compromising yield. Science Advances 6.

365   https://doi.org/10.1126/SCIADV.ABA1715

Urruty, N., Deveaud, T., Guyomard, H., Boiffin, J., 2016. Impacts of agricultural land use changes on pesticide use in French agriculture. European Journal of Agronomy 80, 113–123. https://doi.org/10.1016/j.eja.2016.07.004

Videla-Mensegue, H., Caviglia, O.P., Sadras, V.O., 2022. Functional crop types are more important than diversity for the productivity, profit and risk of crop sequences in the inner Argentinean Pampas. Agricultural Systems 196, 103333. https://doi.org/10.1016/j.agsy.2021.103333

Weisberger, D., Nichols, V., Liebman, M., 2019. Does diversifying crop rotations suppress weeds? A meta-analysis. PLoS ONE 14, 1–12. https://doi.org/10.1371/journal.pone.0219847

Wezel, A., Casagrande, M., Celette, F., Vian, J.-F., Ferrer, A., Peigné, J., 2014. Agroecological practices for sustainable agriculture. A review. Agron. Sustain. Dev. 34, 1–20. https://doi.org/10.1007/s13593-013-0180-7

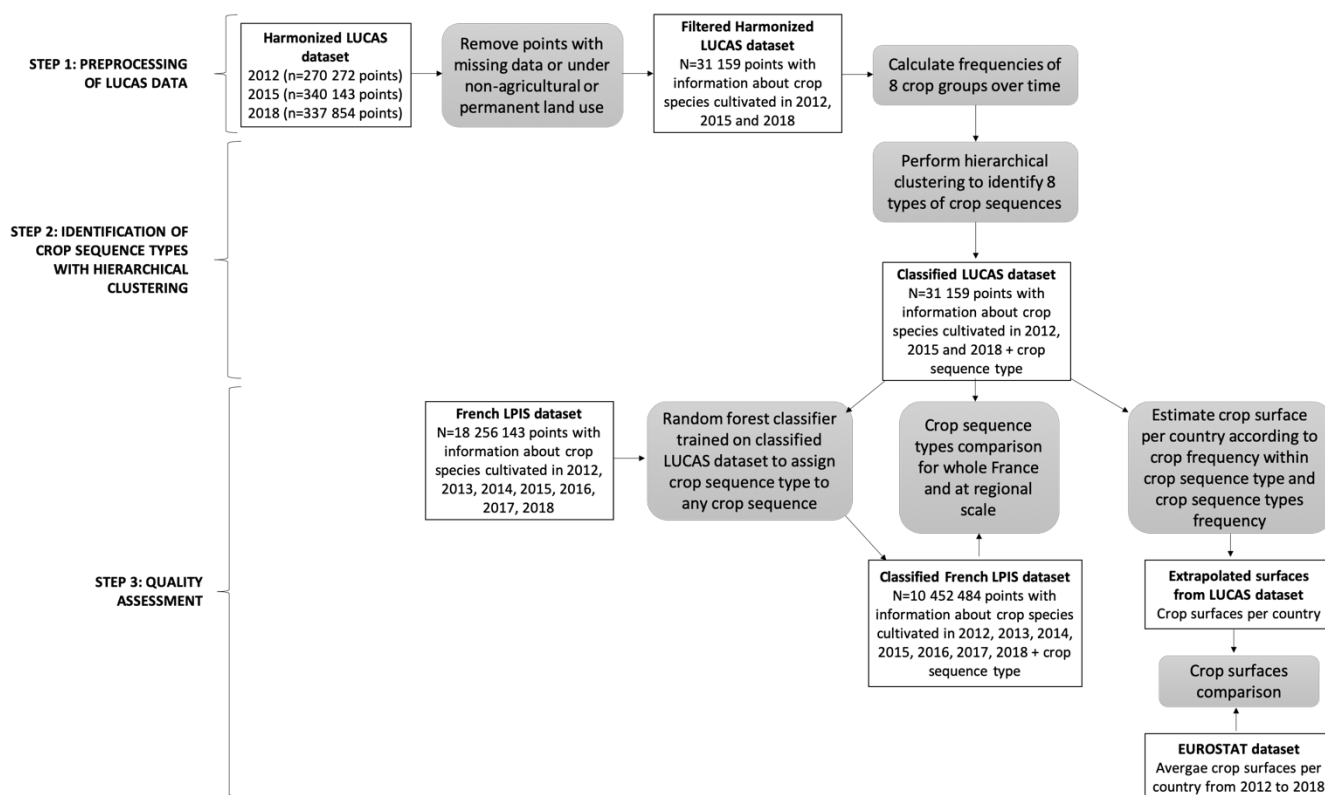
Xiao, Y., Mignolet, C., Mari, J.F., Benoît, M., 2014. Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France. Computers and Electronics in Agriculture 102, 51–63. https://doi.org/10.1016/j.compag.2014.01.010

380

Figure 1: Workflow diagram for developing spatially-explicit maps of dominant crop sequences across EU.

Earth System Discussions
Open Access Science
Data

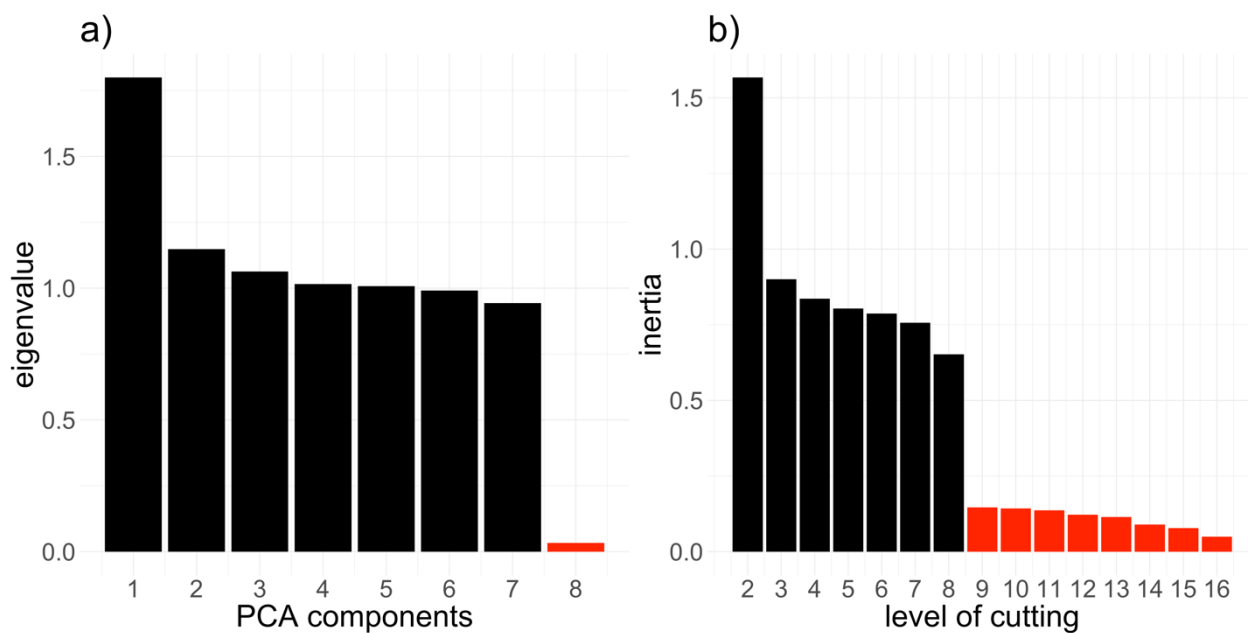



Figure 2. Identification of crop sequence types based on LUCAS dataset. (A) Eigenvalues of the components from the Principle
Components Analysis. (B) Inertia gain according to the level of cutting. The hierarchical clustering is performed on 43 291 points from the
385   harmonized LUCAS dataset for the years 2012, 2015 and 2018, based on the frequency of 8 crop groups (see text for details).

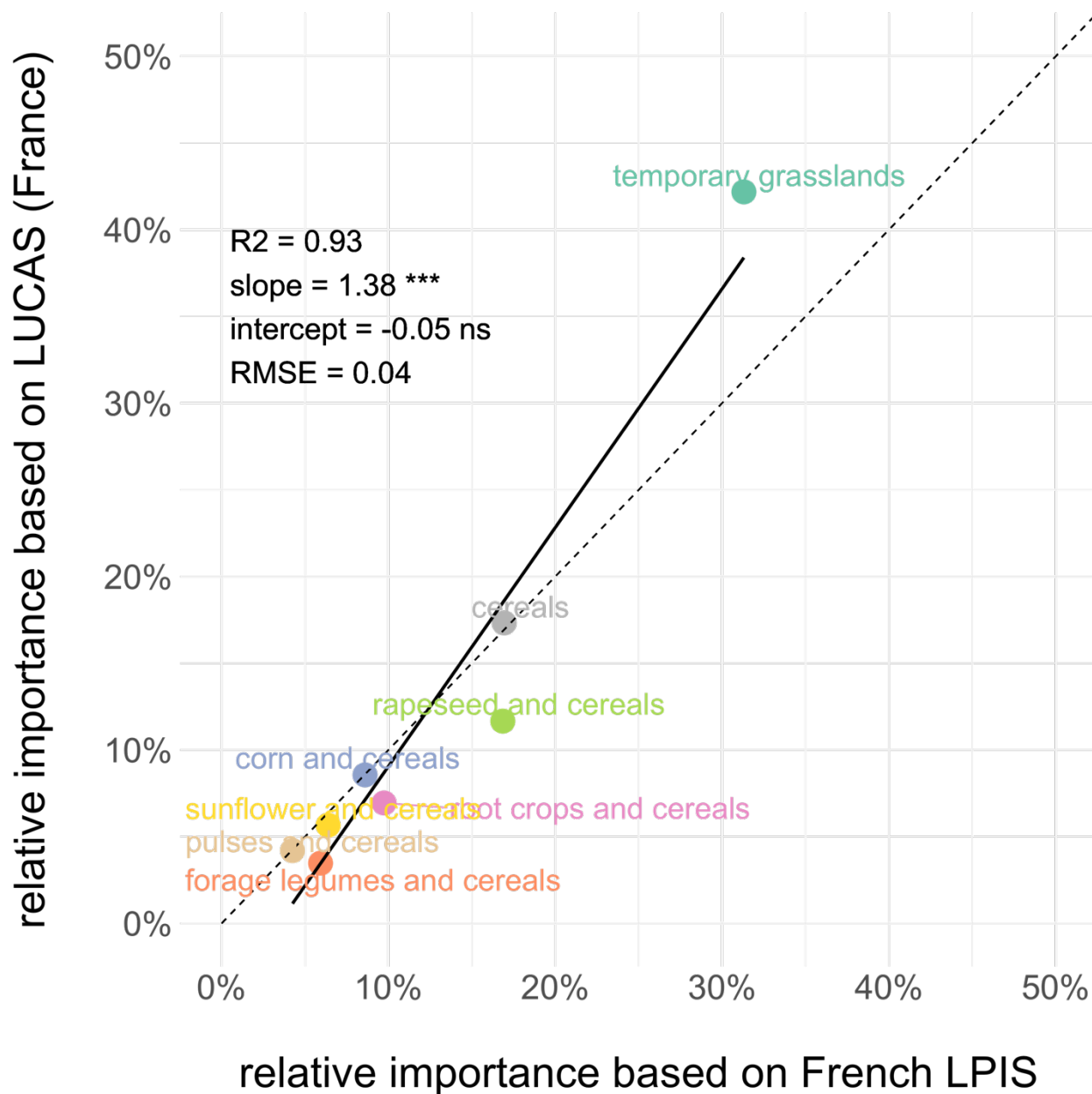



Figure 3. Comparison of the relative importance of crop sequence types derived from the LUCAS and French LPIS datasets at the national scale in France. The dotted line denotes the 1:1 line. Also shown are the fitted linear-regression model and the associated slope and intercept. RMSE: Root Mean Square Error.

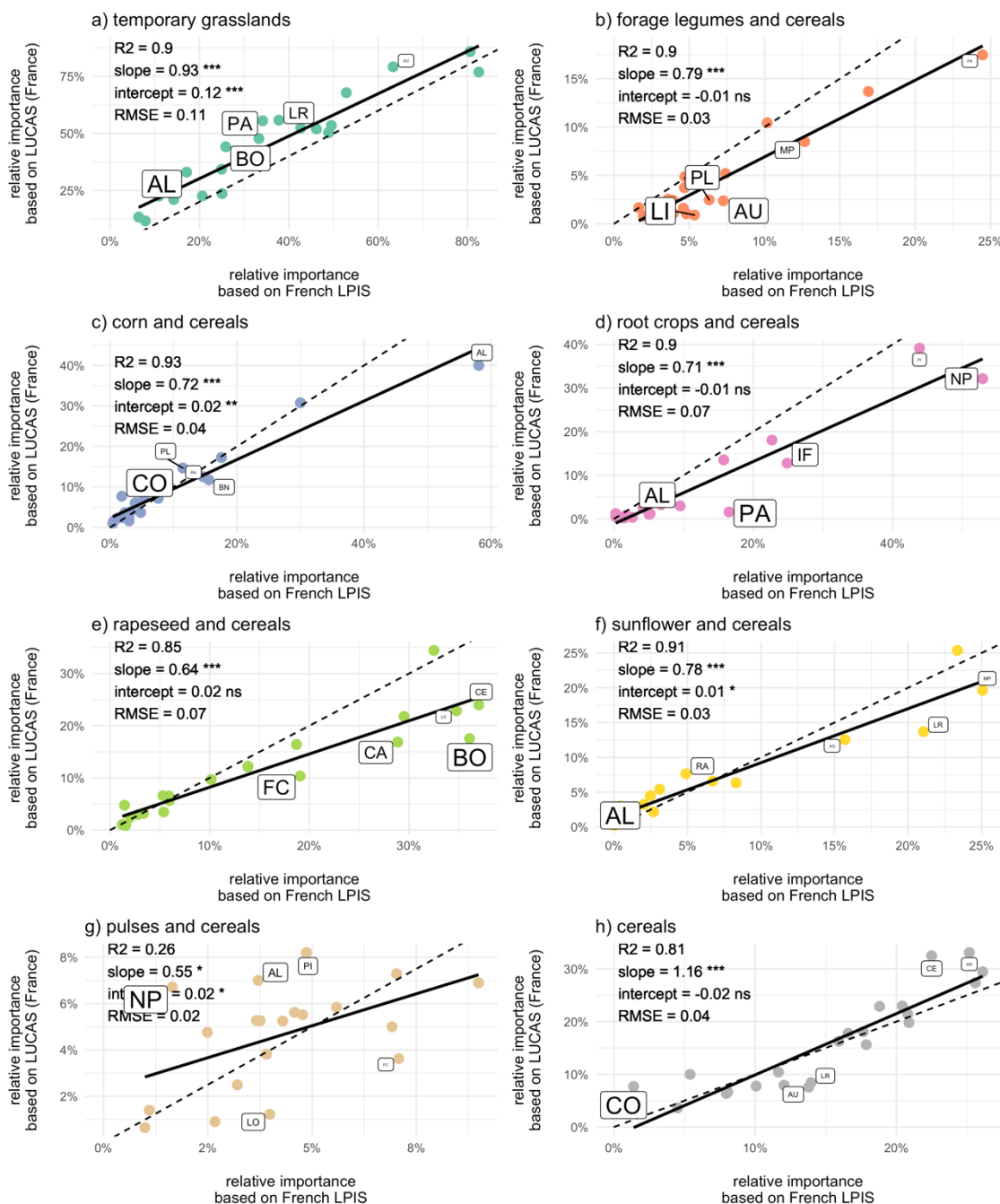



Figure 4. Comparison of the relative importance of crop sequence types derived from the LUCAS and French LPIS datasets for the 22 regions in France. The dotted line denotes the 1:1 line. Also shown are the fitted linear regression models and the associated slopes and intercepts. RMSE: Root Mean Square Error. Labels identify the five regions with the highest difference between predicted and observed area (font size is proportional to area difference relative to observed area). AL: Alsace, AQ: Aquitaine, AU: Auvergne, BN: Basse-Normandie, BO: Bourgogne, BR: Bretagne, CE: Centre, CA: Champagne-Ardenne, CO: Corse, FC: Franche-Comté, HN: Haute-Normandie, IF: Île-de-France, LR: Languedoc-Roussillon, LI: Limousin, LO: Lorraine, MP: Midi-Pyrénées, NP: Nord-Pas-de-Calais, PL: Pays de la Loire, PI: Picardie, PC: Poitou-Charentes, PA: Provence-Alpes-Côte d'Azur, RA: Rhône-Alpes

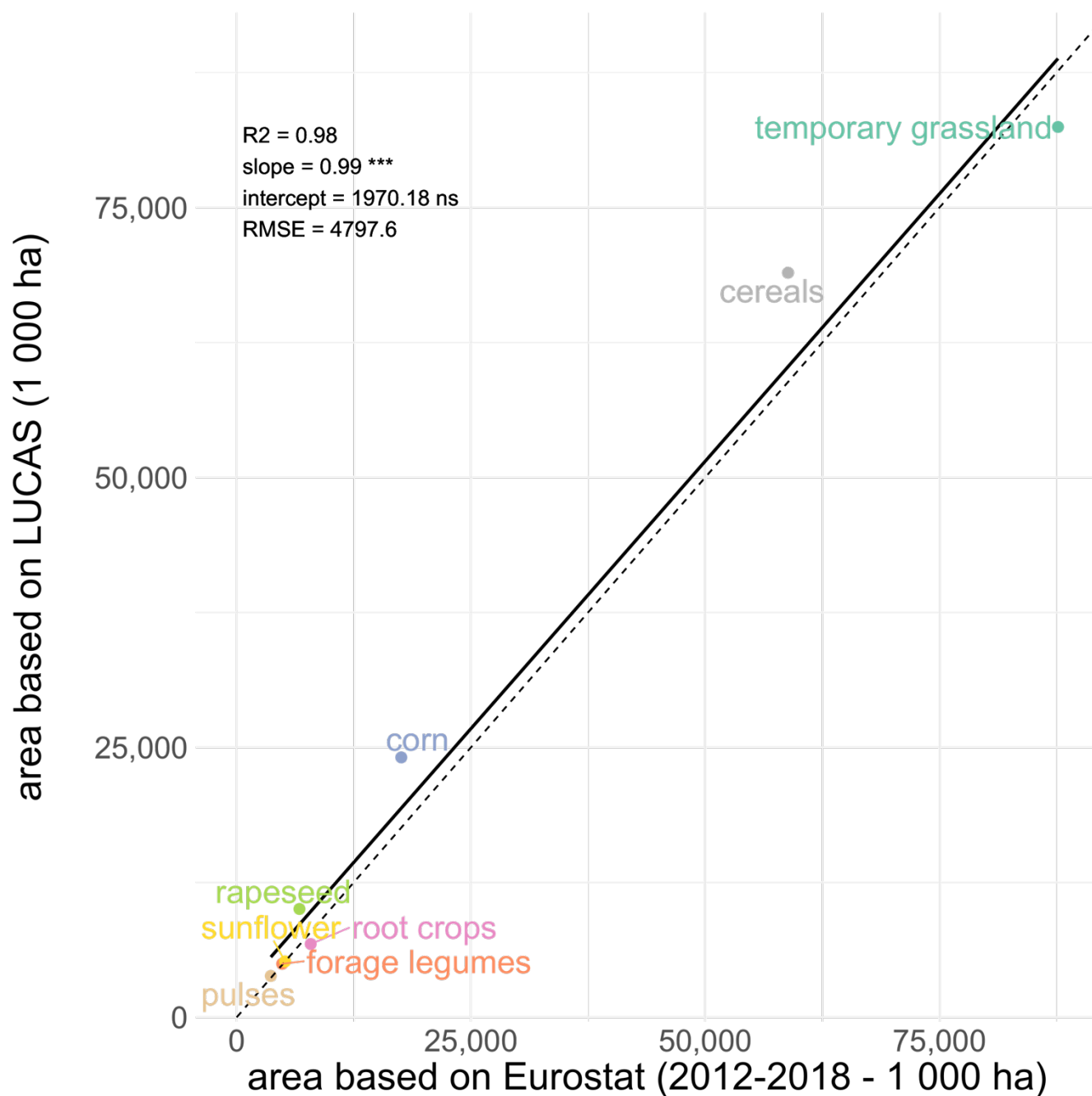



Figure 5. Comparison of harvested areas reported by Eurostat and derived from crop sequence types based on the LUCAS dataset at the whole EU scale. The dotted line denotes the 1:1 line. Also shown is the fitted linear-regression model and the associated slope and intercept. RMSE: Root Mean Square Error.

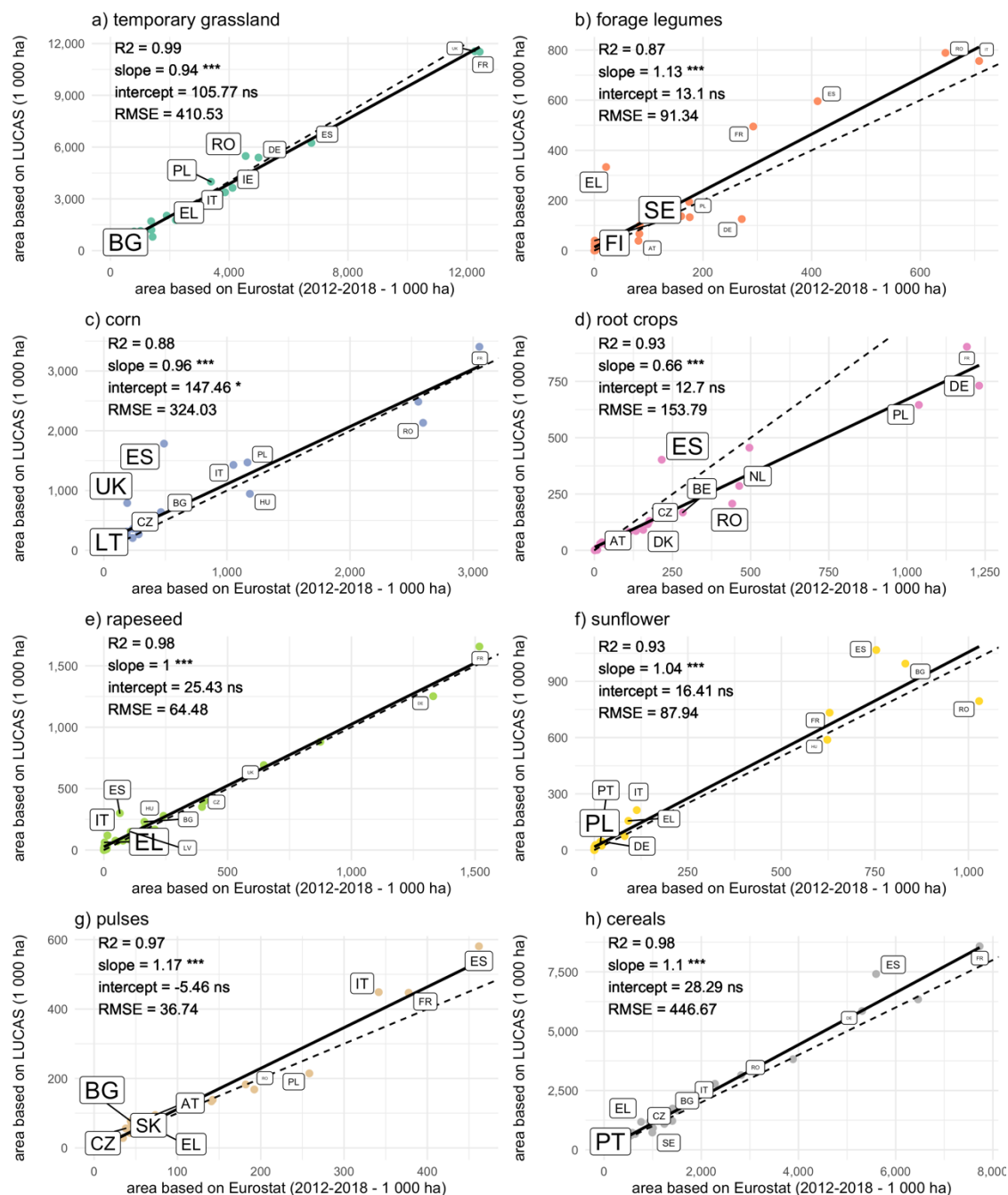



Figure 6. Comparison of harvested areas reported by Eurostat and derived from the crop sequence types based on the LUCAS dataset, for the various countries in the EU. The dotted line denotes the 1:1 line. Also shown are the fitted linear-regression models and the associated slopes and intercepts. RMSE: Root Mean Square Error. Labels identify the 10 countries with the highest difference between predicted and observed area (font size is proportional to area difference relative to observed area). AT: Austria, BE: Belgium, BG: Bulgaria, CY: Cyprus, CZ: Czechia, DE: Germany, DK: Denmark, EE: Estonia, EL: Greece, ES: Spain, FI: Finland, FR: France, HU: Hungary, IE: Ireland, IT: Italy, LT: Lithuania, LU: Luxembourg, LV: Latvia, MT: Montenegro, NL: Netherlands, PL: Poland, PT: Portugal, RO: Romania, SE: Sweden, SI: Slovenia, SK: Slovakia, UK: United Kingdom.

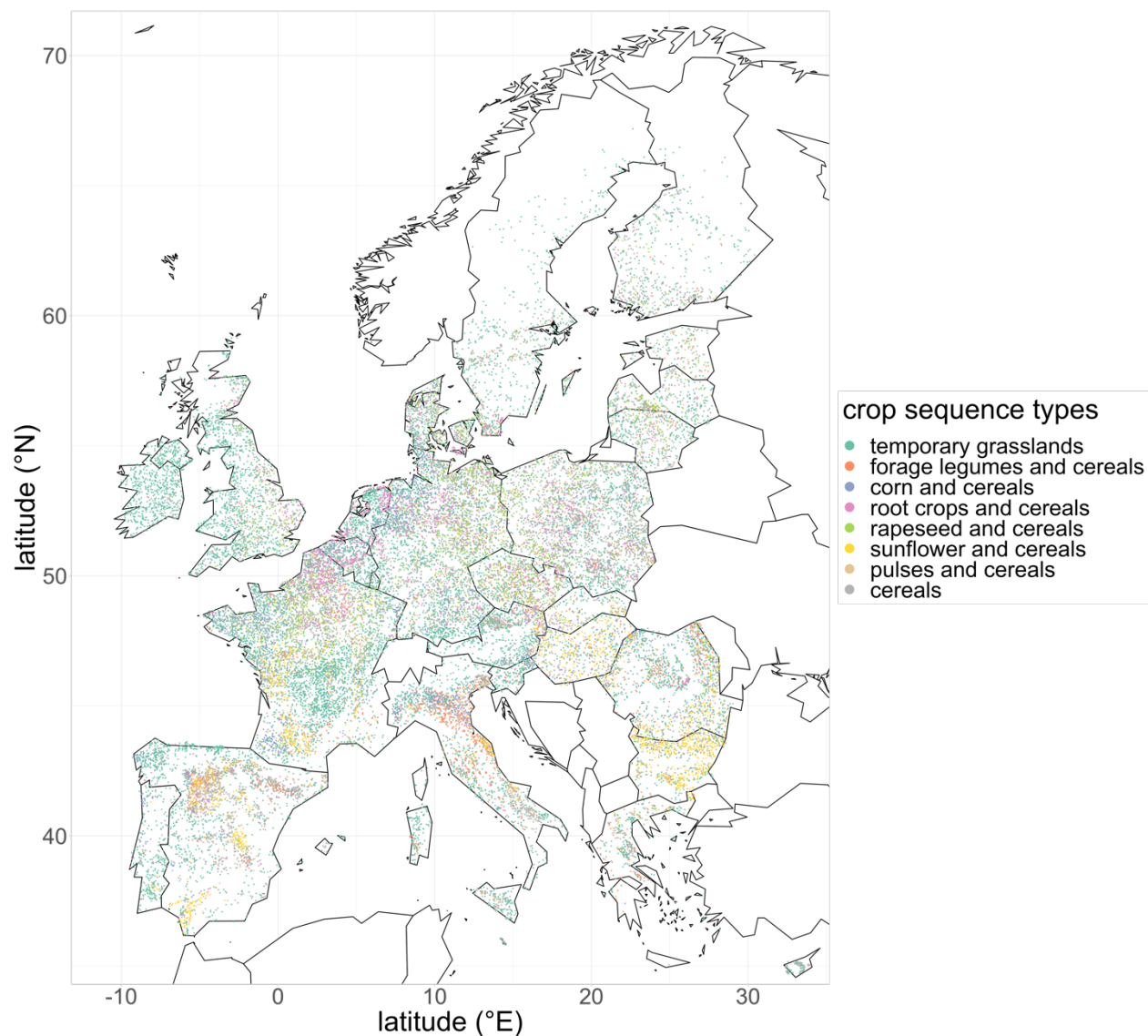
405

410

Figure 7. Map of crop sequence types in the EU based on the LUCAS dataset. Note that crop sequence types have been given short names that reflect dominant crops or crop groups within the sequence. For example, the crop sequence type "grasslands" also includes cereals and corn, but these crops are less frequent in the sequence. Composition of each crop sequence type in terms of temporal frequencies of (groups of) crops is presented in Table 2.

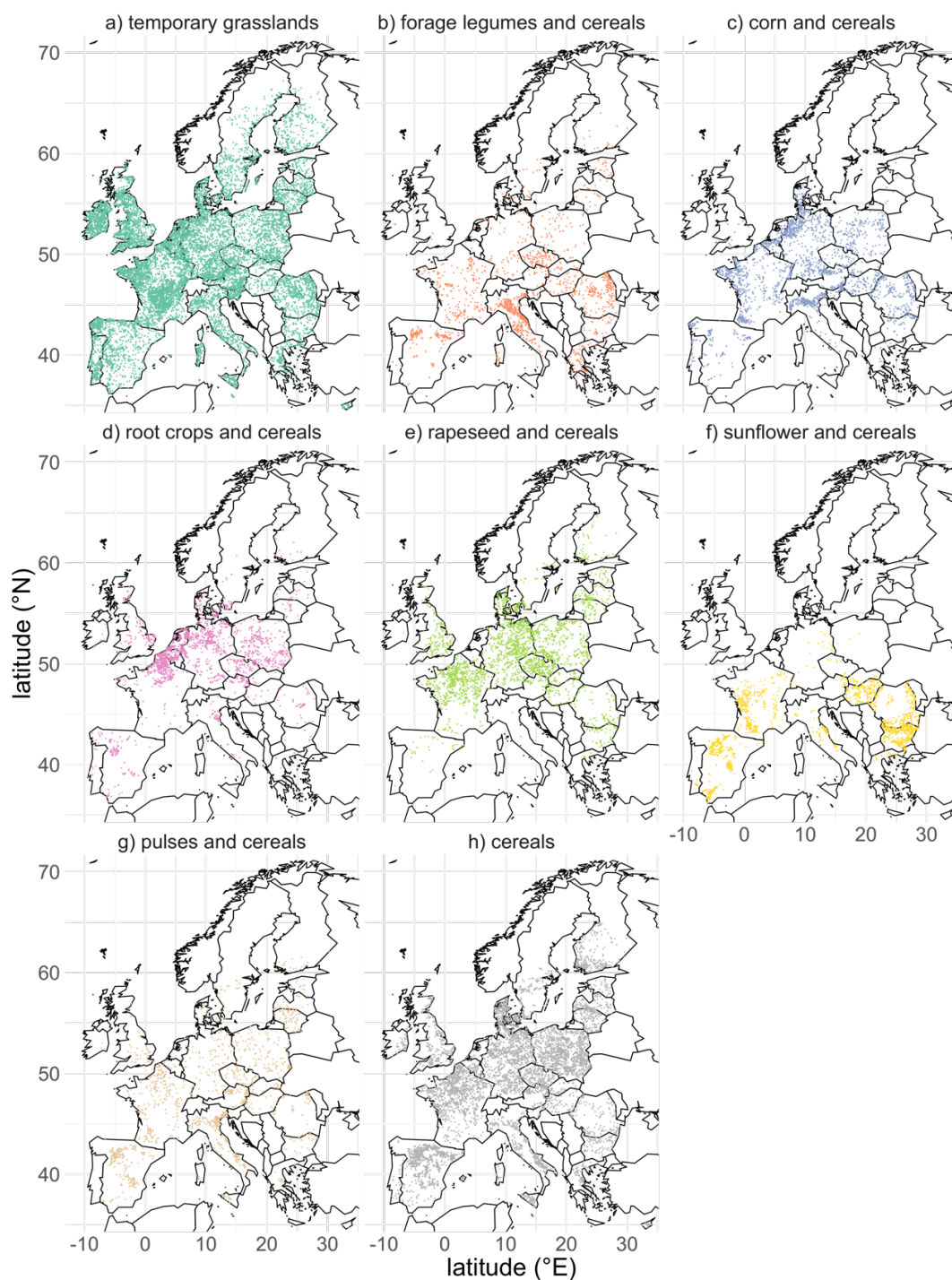



Figure 8. Individual maps of crop sequence types in the EU based on LUCAS dataset. Note that crop sequence types have been given short names that reflect dominant crops or group of crops within the sequence. For example, the crop sequence type "grasslands" also includes cereals and corn, but these crops are less frequent in the sequence. Composition of each crop sequence type in terms of temporal frequencies of (groups of) crops is presented in Table 2.

415

420

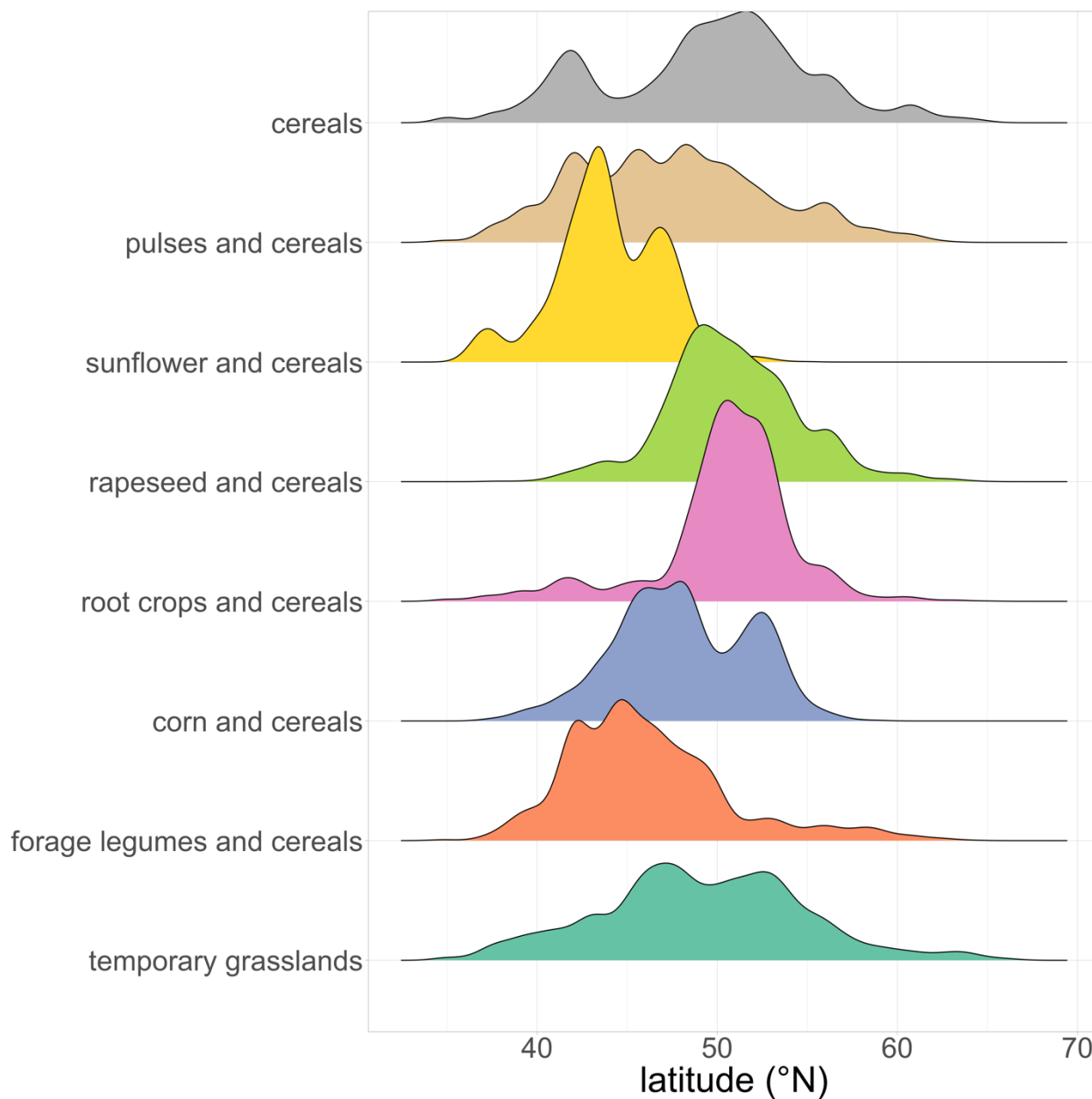



Figure 9. Latitudinal distribution of crop sequence types based on the LUCAS dataset. Note that crop sequence types have been given short names that reflect dominant crops or group of crops within the sequence. For example, the crop sequence type "grasslands" also includes cereals and corn, but these crops are less frequent in the sequence. Composition of each crop sequence type in terms of temporal frequencies of (groups of) crops is presented in Table 2.

425

Table 1. Summary of data sources used in this study.

Name	Dataset description	References
Harmonised LUCAS dataset	Harmonised LUCAS *in-situ* land cover and use database for field surveys from 2006 to 2018 in the European Union	(d'Andrimont et al., 2020) Available for download at: https://data.jrc.ec.europa.eu/dataset/f85907ae-d123-471f-a44a-8cca993485a2
French LPIS dataset	RPG Explorer Crop successions France 2007-2014, 2007-2019, 2015-2019	Martin et al. (2021) Available for download at: https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/XH84QB
Eurostat dataset	Annual crop acreage at national scale for EU countries from 2000	Available for download at: https://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?file=data/apro_cpsh1.tsv.gz

430 **Table 2. Average temporal frequency of crops or crop groups within each crop sequence type.** Values in bold indicate the dominant (or two dominants) crop(s) or group(s) of crops in the crop sequence type. Values in brackets indicate standard deviation.

Crop sequence type (short name)	Temporal frequency of crop or crop group over the 2012-2018 period								Relative importance of crop sequence types across EU
	Cereals	Corn	Rapeseed	Sunflower	Pulses	Root crops	Grassland	Forage legumes	
1- Temporary grassland	0.04	0.02	0	0	0	0	**0.90**	0	40%
	(0.11)	(0.08)	(0)	(0)	(0)	(0)	(0.24)	(0)	
2- Forage legumes and cereals	**0.3**	0.11	0.02	0.02	0	0	0.13	**0.47**	5%
	(0.25)	(0.20)	(0.09)	(0.07)	(0)	(0.02)	(0.22)	(0.20)	
3- Corn and cereals	**0.15**	**0.81**	0	0	0	0	0.03	0	8%
	(0.17)	(0.16)	(0)	(0)	(0)	(0)	(0.10)	(0)	
4- Root crops and cereals	**0.35**	0.09	0.05	0	0	**0.43**	0.04	0.01	7%
	(0.26)	(0.18)	(0.13)	(0)	(0)	(0.19)	(0.13)	(9.05)	
5- Rapeseed and cereals	**0.43**	0.09	**0.43**	0	0	0	0.03	0	9%
	(0.25)	(0.18)	(0.19)	(0)	(0)	(0)	(0.12)	(0)	
6- Sunflower and cereals	**0.39**	0.11	0.03	**0.42**	0	0.01	0.02	0	5%
	(0.25)	(0.20)	(0.11)	(0.16)	(0)	(0.05)	((0.10)	(0.02)	
7- Pulses and cereals	**0.36**	0.09	0.04	0.03	**0.37**	0.03	0.04	0.01	5%
	(0.26)	(0.19)	(0.13)	(0.10)	(0.12)	(0.10)	(0.12)	(0.07)	
8- Cereals	**0.87**	0.06	0	0	0	0	0.04	0	21%
	(0.16)	(0.13)	(0)	(0)	(0)	(0)	(0.11)	(0)	

Table 3. Confusion matrix of the Random Forest model

Observed class	Predicted class								Error rate (%)
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
1- Temporary grassland	4082	0	0	0	0	0	0	0	0
2- Forage legumes and cereals	0	3385	0	0	1	0	0	0	0.03
3- Corn (and cereals)	0	0	4815	0	1	0	0	0	0.02
4- Root crops and cereals	0	0	0	1481	0	0	0	0	0
5- Rapeseed and cereals	0	0	0	0	2235	0	0	0	0
6- Sunflower and cereals	0	0	0	0	0	1700	0	0	0
7- Pulses and cereals	0	0	28	0	0	0	4124	0	0.67
8- Cereals	0	0	0	0	0	0	0	2598	0

435

Table 4: Spatial sampling effort of LUCAS data by country. The sampling effort is quantified as the number of LUCAS points for which observations are available in 2012, 2015, and 2018 per unit area of arable land in 2018 by country as reported by (EUROSTAT, 2022)

Country	Number of LUCAS points	Arable land area (x1000ha)	Number of LUCAS points / agricultural area
Montenegro	15	8	1.77
Ireland	544	476	1.14
Slovenia	189	167	1.13
Cyprus	104	94	1.11
Luxemburg	47	62	0.75
Austria	862	1408	0.61
Belgium	500	976	0.51
The Netherlands	622	1269	0.49
Czech Republic	887	2069	0.43
Latvia	445	1075	0.41
Italy	2263	6142	0.37
Germany	4210	11772	0.36
Sweden	812	2461	0.33
France	5934	19092	0.31
Denmark	712	2347	0.30
Finland	644	2171	0.30
Poland	3133	10908	0.29
Portugal	324	1158	0.28
United Kingdom	1817	7023	0.26
Estonia	165	661	0.25
Bulgaria	775	3289	0.24
Lithuania	516	2209	0.23
Spain	2841	12396	0.23
Greece	403	1972	0.20
Slovakia	248	1358	0.18
Romania	1615	8946	0.18
Hungary	532	4228	0.13