Challenges and recommendations for improved identification of low ILUC-risk agricultural biomass

Supplementary Material

Cato Sandford, Chris Malins, Calliope Panoutsou

Annex A Data filtering and cleaning

The data source is Eurostat's apro_cpshr[1], which spans the period 2000-2021. Variables of interest for this study are productive area (AR, units of ha) and the production (YI_HU_EU, units of t) for each crop and region in the dataset. Crop yield is calculated as the production divided by area for each region; Eurostat does give data for yield, but this dataset is sparsely populated.

Crops: From the available crops, we filter down to those relevant to BIKE. Each crop or crop category is associated with a standard code (e.g., barley is C1300 – see Annex C).

Areas: Countries in the dataset include all the EU plus some other European and peripheral countries like Armenia and Turkey; we restrict focus to larger EU countries plus the UK.

Data processing: The three elements of data processing are: cleaning raw data; calculating variables of interest; and visualisation of the results. Scripts for these are written in Python, with some graphing of additional yields done in Excel.

Cleaning pipeline: Cleaning consists of transforming the data into a format more useful for subsequent calculations, eliminating datapoints which appear to be erroneous to prevent them from skewing results, and dealing with gaps in the record. Specifically, the steps are: pivoting the panel dataset into an annual format; dropping duplicated entries[2]; flagging values that appear to be rounded to a lower level of precision (for manual follow-up); automatically eliminating erroneous data-points identified from jumps in time-series; eliminating erroneous data-points which have been identified by manually checking peculiar results; merging and filling records whose NUTS2 identifier codes have changed over time; and "filling" missing yield data with appropriate proxy values.

Data filling: The "filling" routine just mentioned identifies where gaps in the yield data for one crop may be filled by values for similar crop categories, to arrive at a reasonable approximation. For example, a country may not report the required data for the crop "winter wheat"; however, if it does report data for the more general "wheat" category, then we use this to infer the "winter wheat" trends. As can be verified by looking at countries which report both crop categories, this provides a useful and decently accurate proxy. The virtue of this whole procedure is to allow yield trends to be compared between countries on a common (and limited) set of crops, with maximum specificity where specific data exist, but without throwing away useful data from the parent categories. However, by using one crop category as a stand-in for another, we naturally sacrifice accuracy, and so any detailed investigation of a particular country or sub-national district should be done on the un-filled dataset.

[1] https://ec.europa.eu/eurostat/databrowser/view/APRO_CPSHR/default/table?lang=en

[2] I.e., cases in which two or more regional labels are associated with identical data, including cases where the regional label is also repeated.

Validation: As a check, we compare the post-processing EUROSTAT data with national data from FAOSTAT. This serves a dual purpose of validating our data cleaning and calculations and flagging any inherent discrepancies between the datasets (including difference in crop category terminology in different countries). We find that the results from analysing both datasets are extremely similar, which increases confidence in our analysis pipeline when applied to sub-national data.

Annex B Cleaned dataset

A zip file is included as supplementary material. This has two major components:

1. A csv file containing the yields for selected crops (identified by their five-character Eurostat ID) at the national, NUTS1, and NUTS2 levels (identified by their two-, three-, and four-character NUTS IDs respectively).

2. A Python script which cleans and transforms the raw Eurostat datafile, and outputs the csv file mentioned above. More documentation can be found in the file itself. Associated with the Python script are external dictionary files, which are included in the zip.

Annex C FAOstat crop yield slopes

The global trend in yield growth for selected crops is provided in Table S1 for reference. This is used for calculating dynamic yield baselines for certifying low ILUC-risk production.

Table S1. Yield slopes for selected crops

Crop	Crop ID	Crop Type	Yield Slope (t/ha/yr²)	Source
Barley (Summer)	C1320	Cereal	0.035	Implementing Regulation (barley)
Barley (Winter)	C1310	Cereal	0.035	Implementing Regulation (barley)
Durum Wheat	C1120	Cereal	0.040	Implementing Regulation (wheat)
Maize	C1500	Cereal	0.074	Implementing Regulation
Oats	C1410	Cereal	0.026	Calculated FAOstat
Palm	--	Oil	0.200	Implementing Regulation
Rapeseed	I1111	Oil	0.036 0.034	Implementing Regulation Calculated FAOstat
Rye	C1210	Cereal	0.050	Calculated FAOstat
Sorghum	C1700	Cereal	0.006	Calculated FAOstat
Soybean	I1130	Oil	0.028	Implementing Regulation
Sugar Beet	R2000	Sugar	1.276	Implementing Regulation
Sugarcane	--	Sugar	0.379	Implementing Regulation
Sunflower Seed	I1120	Oil	0.035 0.034	Implementing Regulation Calculated FAOstat
Triticale	C1600	Cereal	0.011	Calculated FAOstat
Wheat (Summer)	C1112	Cereal	0.040	Implementing Regulation (wheat)
Wheat (Winter)	C1111	Cereal	0.040	Implementing Regulation (wheat)

The slopes for some crops are indicated to be taken from the Implementing Regulation; these were calculated by the Commission by fitting a straight line to 20 years of FAOstat data from the period 1998-2017. For crops not included in the Implementing Regulation, we calculate the yield slopes following exactly the same procedure, with the caveat that the FAOstat record for the period has been updated and amended since the Implementing Regulation was drafted, so we are working from a slightly different dataset.

Figure S1 shows an example of global yield and yield growth for two oil crops. The annual average yield data are shown as connected points, and the 20-year trend as straight solid lines. The slopes of the trendlines are indicated in parentheses in the figure legend; it so happens that, in the period considered, these slopes are identical to two significant figures.

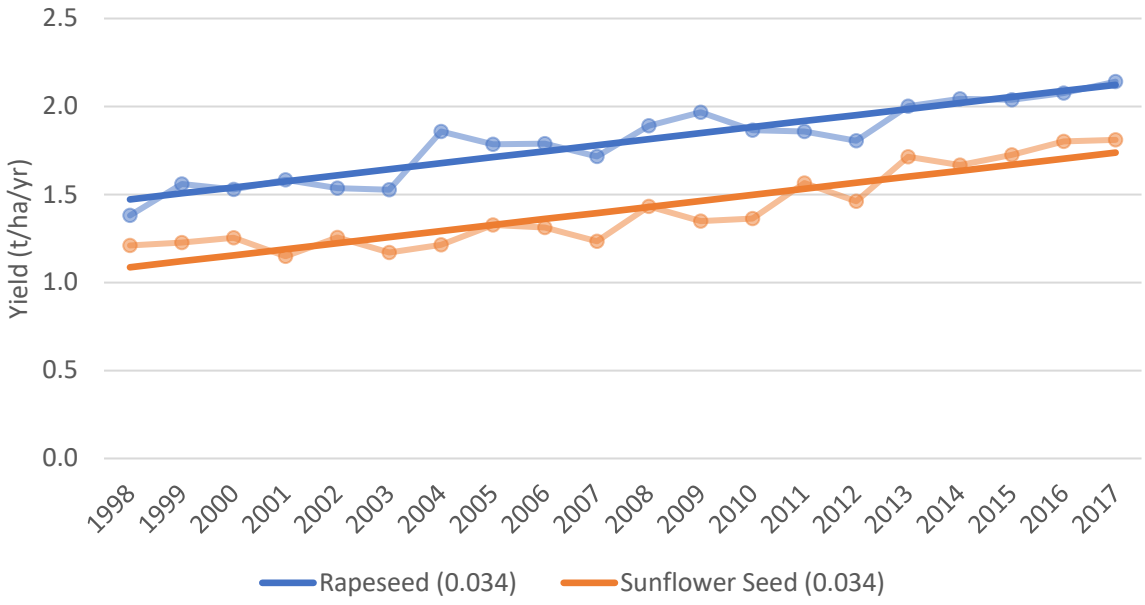

Figure S1. Global average yields for selected oil crops in the period 1998-2017, with best-fit lines and slopes

Note: The time-period is chosen to follow the Implementing Regulation.

Annex D Country codes

For reference, the country code abbreviations for EU countries and the UK are provided in Table S2.

Table S2. Country codes

Country ID	*Country Name*
AT	Austria
BE	Belgium
BG	Bulgaria
HR	Croatia
CY	Cyprus
CZ	Czechia
DK	Denmark
EE	Estonia
FI	Finland
FR	France
DE	Germany
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
LV	Latvia
LT	Lithuania
LU	Luxembourg
MT	Malta
NL	Netherlands
PL	Poland
PT	Portugal
RO	Romania
SK	Slovakia
SI	Slovenia
ES	Spain
SE	Sweden
UK	United Kingdom

Annex E Above-baseline yield statistics

Figure S2 counts the number of districts which could hypothetically have declared spurious above-baseline yield, taking winter barley as an example. Each panel shows a different year, from 2015 to 2020. The bars at zero above-baseline yield dominate the plots, signifying that, in a given year, the majority of districts are producing at or below the baseline – as we would expect, because no additionality measure has been taken. Nevertheless, it is evident that a significant number of districts could have declared above-baseline material.

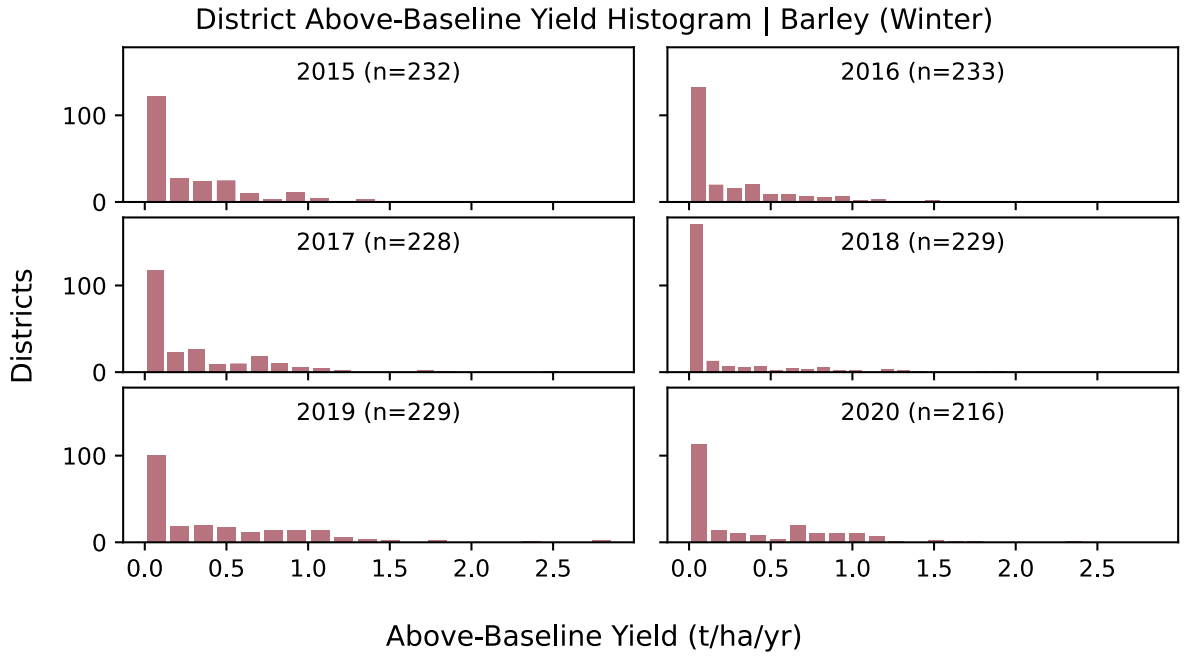

Figure S2. Above-baseline yield histogram encompassing all NUTS2 districts, for winter barley

Note: For this analysis, the yield baseline begins in 2015. This certification year differs from the one used in Figures 10 and 11. Previously, we were interested in how the production of above-baseline material plays out over a longer time-span, so we chose an earlier certification year. Now, we are looking at annual snapshots rather than time evolution, and moreover we and to maximise the number of NUTS2 regions represented in the dataset; this favours a later certification year.

Annex F Country-specific yield slopes

Yield slopes for selected crops are presented, country by country, in Table S3. These slopes are calculated on national data on the period 2000-21.

Table S3. Yield slopes for selected crops in the EU

Yield Slope (t/ha/yr^2)	Wheat (Winter)	Durum Wheat	Rye	Barley (Winter)	Rapeseed (Winter)	Sunflower Seed	Sugar Beet
Austria	0.038	0.047	0.052	0.088	0.032	0.007	0.685
Belgium	0.015	--	--	0.026	--	--	1.730
Bulgaria	0.125	0.085	0.022	0.113	0.085	0.040	--
Czechia	0.083	--	0.074	0.106	0.039	0.022	0.996
Germany	0.014	0.004	0.007	0.042	0.001	-0.004	1.081
Estonia	0.127	--	0.079	0.143	0.066	--	--
Spain	0.032	0.034	0.035	0.013	0.045	0.013	1.411
Finland	0.049	--	0.086	--	--	--	0.237
France	0.001	0.048	-0.004	0.000	0.010	-0.003	0.468
Croatia	0.083	0.024	0.057	0.104	0.035	0.052	1.711
Hungary	0.099	0.081	0.068	0.143	0.069	0.053	1.048
Ireland	0.028	--	--	0.088	0.071	--	--
Italy	0.049	0.054	0.037	0.039	0.082	0.019	0.986
Lithuania	0.087	--	0.015	0.082	0.058	--	1.620
Latvia	0.118	--	0.116	0.156	0.059	--	--
Netherlands	0.028	--	-0.065	0.114	-0.005	--	1.424
Poland	0.068	--	0.045	0.063	0.023	0.025	1.301
Portugal	0.050	0.054	0.008	0.050	--	0.062	-2.383
Romania	0.109	0.069	0.055	0.107	0.079	0.079	1.029
Sweden	0.051	--	0.049	0.049	0.014	--	1.197
Slovenia	--	--	0.065	--	0.013	0.074	0.149
Slovakia	0.097	0.091	0.058	0.128	0.073	0.047	1.196
United Kingdom	0.017	--	-0.199	0.037	-0.004	--	0.960