Dataset of biomass and chemical quality of crop residues from European areas

P. Thiébeau, Lars Stoumann Jensen, Fabien Ferchaud, Sylvie Recous

Data in Brief 37 (2021) 107227

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Dataset of biomass and chemical quality of crop residues from European areas

Pascal Thiébeau[a,d,*], Lars Stoumann Jensen[b], Fabien Ferchaud[c], Sylvie Recous[d,a]

[a] INRAE, Université de Reims Champagne Ardenne, FARE, UMR A 614, 51100 Reims, France
[b] Plant and Soil Science Section, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, 1871 Frederiksberg, Denmark
[c] INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne, BioEcoAgro, 02000 Barenton-Bugny, France
[d] Université de Reims Champagne Ardenne, INRAE, FARE, UMR A 614, 51097 Reims, France

ARTICLE INFO

Article history:
Received 20 May 2021
Revised 10 June 2021
Accepted 11 June 2021
Available online 15 June 2021

Keywords:
Crop residues
Chemical characteristics
N content, C:N ratio
Literature survey

ABSTRACT

This dataset presents the chemical characteristics of plant biomass and crop residues from agrosystems in European areas (carbon and nitrogen contents and biochemical composition). These data have been collected from the scientific literature. The specific data and their origins are presented. The mean values from these data are also provided by major production type (main crops, forage and pasture crops, green manure and cover crops, vegetable crops and energy crops), species and litter type. These data were collected as part of the framework of the European project ResidueGas (ERA-GAS, 2017-2021), which aims to improve the estimation of greenhouse gas emissions associated with crop residues.

* Corresponding author at: INRAE, Université de Reims Champagne Ardenne, FARE, UMR A 614, 51100 Reims, France.
E-mail address: pascal.thiebeau@inrae.fr (P. Thiébeau).

https://doi.org/10.1016/j.dib.2021.107227

Specification Table

Subject	Agricultural science
Specific subject area	Agronomy and crop science: crop biomass and chemical characteristics
Type of data	Figure Table
How data were acquired	Data were collected from published and unpublished literature (PhD and master theses; technical reports) that provide plant litter/crop residue characteristics (secondary data). Plant biomass characteristics were obtained by elemental analysis of total carbon and total nitrogen. Biochemical composition was determined by proximate analysis using the Van Soest method [1,2].
Data format	Secondary data: "Literature data" files Raw data: aggregate data ("mean data" files) and figures of this article "List of terms and abbreviations" List of the primary data sources ("List of literature references") All files are presented as .xlsx extension file format
Parameters for data collection	Only experimental work carried out on the European areas and from a well-identified source were included in the dataset.
Description of data collection	Secondary data: We collected all the elementary data (e.g., biomass, % N, % C, and lignin content) that were found in the literature. Raw data: The elementary data were aggregated into five agricultural production types: main crops, pasture and forage crops, green manure and cover crops, vegetable crops and energy crops. Means and standard errors of the data were calculated.
Data source location	Raw data of figures are available as supplementary materials of *Data in Brief* site. Secondary data: a complete list of the primary data sources used, with their access link, is produced in the "List_of_literature_references_v1.tab" file in repository *Data INRAE* (see *Data accessibility* part). For any other file, the complete [dataset] [3] is securely stored in Data INRAE Center located in Jouy-en-Josas (FR) with a replication in the Data Center of Toulouse (FR).
Data accessibility	Repository name: Data INRAE Data identification number: 10.15454/LBI3U7 Direct URL to data: https://doi.org/10.15454/LBI3U7

Value of the Data

- This dataset is based on the collection of a large amount and diversity of data on the amount of plant biomass and its elemental and gross biochemical composition from crops at different stages of growth (young to full maturity) in the European areas;
- These data are primarily intended for researchers working on the agricultural and environmental impacts of agricultural activities (soil carbon storage, fertilization management, and greenhouse gas emissions) because the composition of residues remaining on the soil is one of the factors, in addition to climate, soil geophysical environment and management (e.g., tillage), that determines the consequences of crop residue management;
- This collection of data made it possible to generate average values by plant species and type of use, which should correspond to stakeholder needs, such as for national GHG inventory agencies in the European area and Intergovernmental Panel on Climate Change (IPCC) experts;
- These data are rarely available to users because they are expensive to acquire and mostly distributed across a wide range of scientific publications as well as unpublished data sources; this dataset is freely available to all users through the Data INRAE repository; and
- This dataset is intended to be updated regularly based on new data published in the literature.

1. Data Description

This dataset is positioned in the open Data INRAE repository. It is described in English language. It is made up of twelve files: a file named "List_of_literature_references_v1.tab" produces the origin of the primary data and access link; a "List_of_Terms_and_Abbreviations_v1.tab" file produces the correspondences of the terms and abbreviations used in the dataset; and ten data files, described below, comprising five "Literature_data_xx_crops_v1.tab" files and five "Mean_data_xx_crops_v1.tab" files. The twelve files are displayed into two pages. Each file can be downloaded.

The dataset contains data collected in 177 documents, including 158 scientific articles, 3 books, 4 PhD theses, 2 Master theses, 7 reports and 3 conference presentations. This literature was published between January 1985 and January 2021. Their distribution of the publications over the years, except 2021, which is in progress, is presented in Fig. 1 (raw data in Supplementary Material SM 1).

The 158 scientific articles used were published in 44 journals. These studies were carried out in 17 European countries. The journals *Plant and Soil, Soil Biology and Biochemistry, Nutrient Cycling in Agroecosystems, Agriculture, Ecosystems and Environment,* and *European Journal of Agronomy* accounted for 56% of the data (raw data presented in SM 2).

The dataset includes approximately 2300 individual data records (see SM 3). It is structured into five categories according to the crop type and crop agricultural use: main crops (24% of occurrences), pasture and forage crops (21% of occurrences), green manure and cover crops (30% of occurrences), vegetable crops (5% of occurrences), and energy crops (21% of occurrences). While field crops are well represented, this is not the case for vegetable crops for which data are still scarce. On the other hand, many references are available for new crops, such as energy crops. The list of species, Latin names and botanical families of the five crop types identified in the dataset are presented in SM 3.

This effort allowed us to generate, for each category of crop type, a literature data file and a mean data file.

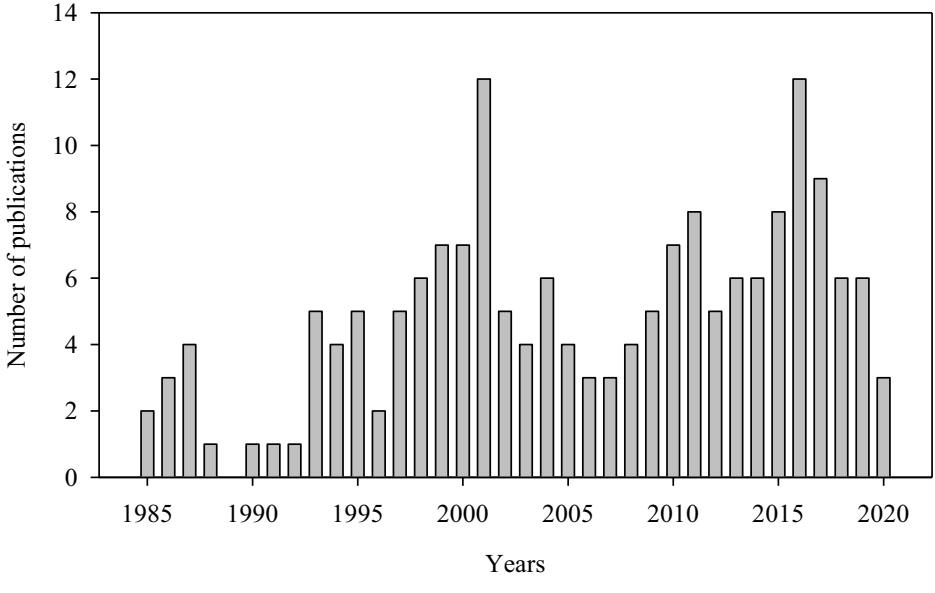

Fig. 1. Distribution and number of the publications over the years.

The secondary data files (literature and unpublished data) provide the following information for each plant species:

Species: common name, Latin name, cultivar, and family name;
Agricultural conditions of growth: plant part, harvest stage, sowing date, harvest date, N fertilization ($kg.ha^{-1}$ yr^{-1}), country (region or town), and origin (field, glasshouse, lysimeter);
Dry matter: ton ha^{-1} and percentage of fresh matter (% FM);
Chemical composition: neutral detergent soluble (NDS), hemicellulose + cellulose, hemicellulose, cellulose, lignin, LCI (lignocellulose index = lignin/[hemicellulose + cellulose + lignin]) [4,5], water soluble C (WSC), total C, total N, and C:N ratio (carbon:nitrogen ratio = total C/total N);
Soil properties: texture characteristics, bulk density, layer, pH, total N, organic C, and soil classification according to authors; and
References: authors (year published), journal, issue (volume), pages or report number, and access link.

An example of extraction of the available data is produced by Fig. 2 (raw data in SM 4) for the total N concentration and C:N ratio of the species of the "main crop" type.

For the mean data files, they provide us with the following aggregated information:

Species: common name, Latin name, and family name;
Agricultural conditions of growth: plant part and harvest stage;
Dry matter: ton ha^{-1} and % FM;
Chemical composition: NDS, hemicellulose + cellulose, hemicellulose, cellulose, lignin, LCI, WSC, total C, total N, C:N ratio.

An example of data extraction is produced by Fig. 3 (raw data in SM 5) for the Van Soest chemical composition of species of the "main crop" type shown in Fig. 2. They allow us to compare species and reveal missing information (Fig. 3, raw data in SM 5).

The dataset includes a file that describes the list of terms and abbreviations used and a file with literature references: author name, year published, journal name and access link.

2. Experimental Design, Materials and Methods

The literature survey was conducted on studies carried out in the European area using the following keywords: "crop residue" OR "biomass" OR "aerial dry matter" OR "straw" OR "root" OR "root dry matter" OR "main crops" OR "forage crops" OR "meadow" OR "cover crop" OR "catch crop" OR "green manure" OR "legume", AND "N content" OR/AND "C content" OR/AND "biochemical composition". Several data retrieval directions were pursued and included in English and without limit of publication year and a query of several search engines: Web of Science, ResearchGate and Google Scholar; in addition, we systematically explored the reference list of the relevant articles, which allowed us to identify additional articles/reviews with which the initial keywords had not been associated; we also contacted the authors to obtain the raw data for their published works.

During this collection phase, we collected all available elementary data (e.g., biomass, % N, % C, and lignin content), even partial, and we calculated values where possible, based on the information given in the articles. For example, dry biomass per hectare and corresponding N quantity allow the N concentration of the residues to be calculated; the residue N concentration and C:N ratio made it possible to calculate the C concentration. However, many articles were not retained because they presented aggregated data and/or modelled plant growth adjustments without presenting sufficient information on the initial data.

We classified the data into five crop categories corresponding to the main agricultural production types: main crops, forage and pasture crops, green manure and cover crops (typical soil

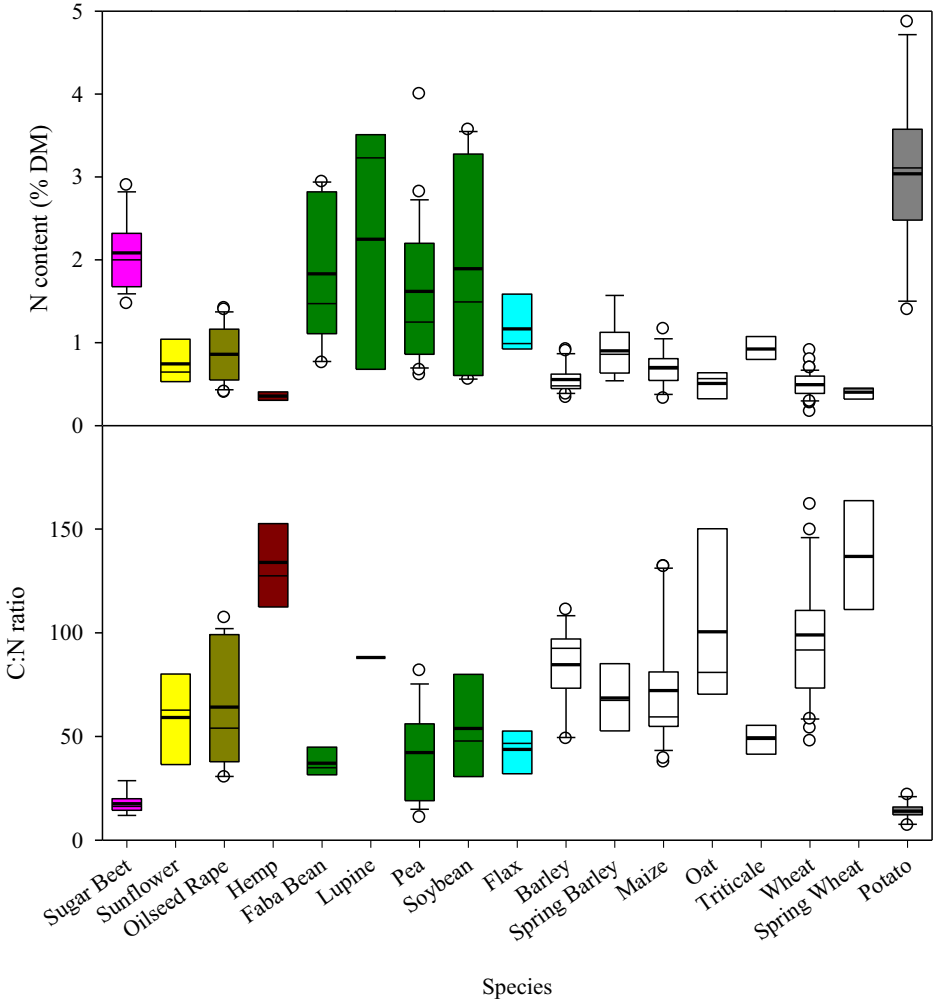

Fig. 2. Distribution of the C:N ratio and N contents (% Dry Matter) of aboveground residues for the main crops categories available in the « Raw_data » file. Data represented: points (data dispersion), error bars (10th –90th percentile), box (25th –75th percentile), mean (thick line) and median (thin line).
The color legend correspondance (family species) is : pink = amaranthaceae, yellow = asteraceae, dark green = brassicaceae, brown = cannabaceae, green = fabaceae, blue = linaceae, white = poaceae, and grey = solanaceae. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

incorporated in an immature state rather than harvested), vegetable crops, and energy crops. We retained input variables presenting enough occurrences to justify their presence in the dataset. For example, we chose the biochemical composition determined by proximate analysis according to the Van Soest method [1,2] and did not keep the data on lignin obtained by the Klason method [6] because they were very scarce. We also did not retain the data on sugar concentrations (glucose, sucrose, fructose, etc.), which were rarely available in the literature, and we did not retain data on lipids, resins and starch contents.

We have encountered several situations where the common names differed for the same species, while the Latin name assigned by the authors was identical. In these cases, we have chosen to group them under the same Latin name item (e.g., *Raphanus sativus* L. correspond

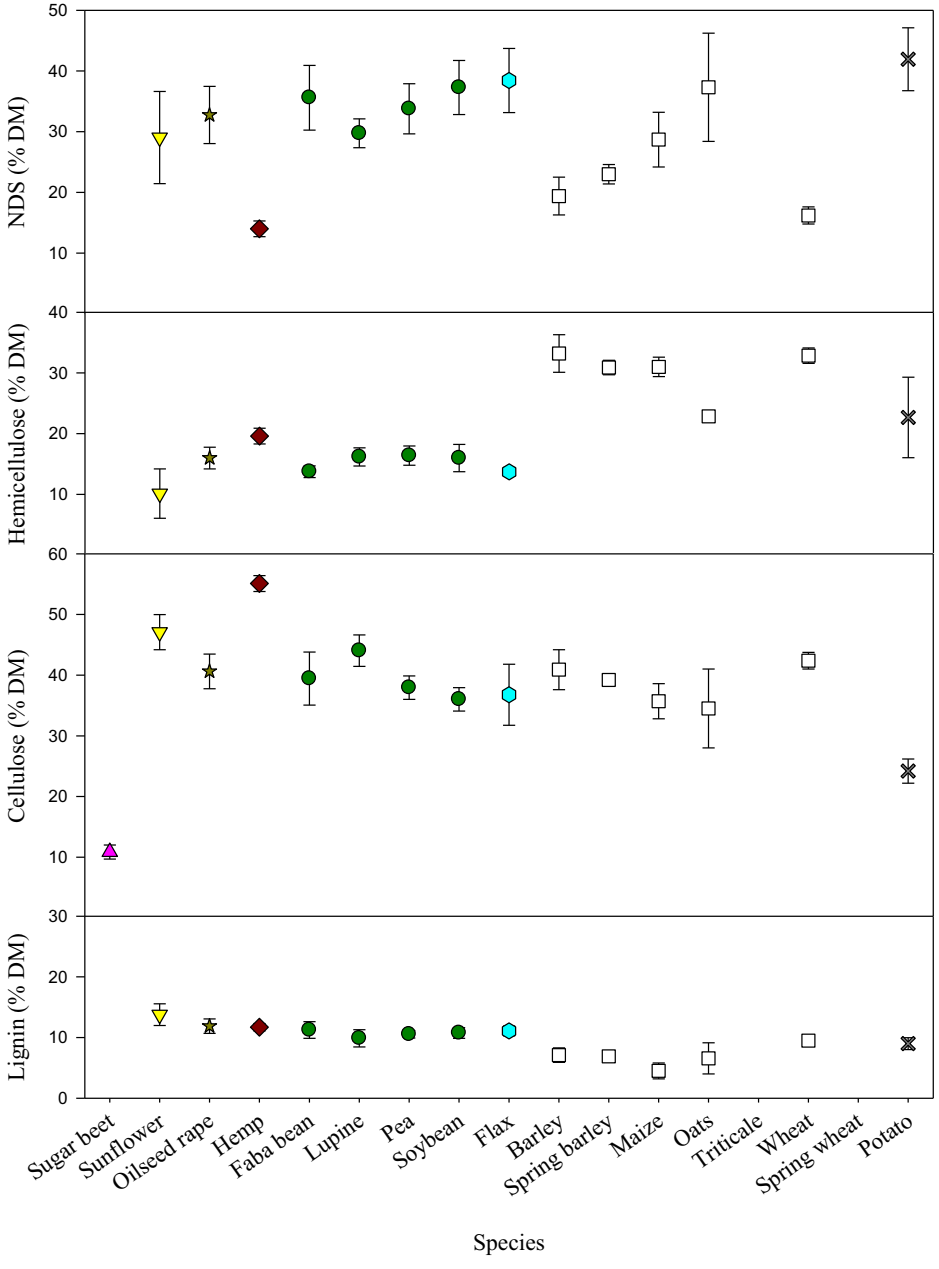

Fig. 3. Biochemical composition (mean and standard error) of aboveground residues for the main crops categories, available in the « Mean_data » file.
NDS : Neutral Detergent Soluble ; %DM : percent of Dry Matter. The symbol shapes and color legend correspondance (family species) are : triangle up pink = amaranthaceae, triangle down yellow = asteraceae, star dark green = brassicaceae, diamond brown = cannabaceae, circle green = fabaceae, hex blue = linaceae, square white = poaceae, x grey = solanaceae. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to fodder radish, oil radish and daikon radish; *Vicia faba* L. correspond to broad bean and faba bean). For the situations where two Latin names have been met for the same common names, we have chosen to refer to the crop sowing dates to classify them correctly (e.g., *Hordeum vulgare* L. and *Hordeum distichum* L. were both used for barley and spring barley).

This "Literature_data" file also specifies the plant parts considered by the authors: aerial dry matter (ADM = aboveground dry matter – grain yield), leaves, stems, pods, stubble (= ADM – straw exported from the field), straw (= leaves + stems at harvest stage, without the chaff), roots, etc. In situations where plants were able to be harvested/destroyed during different periods (e.g., green manure/cover crops and energy crops), we distinguished "autumn destruction" (AuD) or "autumn harvest" (AuH) versus "winter harvest" (WiH) or "spring destruction" (SpD) or "at harvest time (AtHT)" versus "at destruction time (AtDT)" for the forage crops. The harvest height was rarely specified. When it was, we noted it (e.g., "aerial DM" versus "aerial DM (>5 cm)"). We carried out the same process for the root measurement depth (e.g., "roots" versus "roots (0-20 cm)"), which was rarely indicated.

We performed the mean and standard error calculations whenever possible, specifying the number of observations. When the amount of data collected was too low or the range of the values was too large, we did not calculate means, leaving this action to the users.

This dataset is intended to be updated regularly based on new data published in the literature.

CRediT Author Statement

Pascal Thiébeau: Conceptualization, Methodology, Investigation, Validation, Formal analysis, Data curation, Writing - original draft & editing, Visualization; **Lars Stoumann Jensen:** Resources, Writing - review & editing; **Fabien Ferchaud:** Resources, Writing-review & editing; **Sylvie Recous:** Conceptualization, Methodology, Validation, Writing - review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgements

The present study was supported by the FACCE ERA-GAS Cofund (European Research Area – for mitigation of Greenhouse gases from Agri- and Silvi-culture; id number: 696356), project ResidueGas coordinated by Professor Jørgen Eivind Olesen (https://projects.au.dk/residuegas/). The French contribution to ResidueGas was funded by the Agence Nationale de la Recherche (ANR-17-EGAS-0003) and the National Institute for Food, Agriculture and the Environment (INRAE).

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107227.

References

[1] P.J. Van Soest, Use of detergents in analysis of fibrous feeds. 2. A rapid method for determination of fiber and lignin, J. Assoc. Off. Agric. Chem. 46 (5) (1963) 825–835, doi:10.1093/jaoac/73.4.487.
[2] P.J. Van Soest, Development of a comprehensive system of feed analyses and its application to forages, J. Anim. Sci. 26 (1) (1967) 119–128, doi:10.2527/jas1967.261119x.
[3] [dataset] P. Thiébeau, L.S. Jensen, F. Ferchaud, S. Recous, Biomass and Chemical Quality of Crop Residues From European areas, Data INRAE, v1, 2021, doi:10.15454/LBI3U7.
[4] J.M. Melillo, J.D. Aber, A.E. Linkins, A. Ricca, B. Fry, K.J. Nadelhoffer, Carbon and nitrogen dynamics along the decay continuum: plant litter to soil organic matter, Plant Soil 115 (2) (1989) 189–198, doi:10.1007/BF02202587.
[5] J. Herman, D. Moorhead, B. Berg, The relationship between rates of lignin and cellolose decay in aboveground forest litter, Soil Biol. Biochem. 40 (10) (2008) 2620–2626, doi:10.1016/j.soilbio.2008.07.003.
[6] C.W. Dence, The determination of lignin, in: S.Y. Lin, C.W. Dence (Eds.), Methods in Lignin Chemistry, Springer series in Wood Science, Springer, Berlin, 1992, pp. 33–61, doi:10.1007/978-3-642-74065-7_3.