www.nature.com/scientificdata

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A global dataset of public agricultural R&D investment: 1960-2022

Michiel van Dijk[1,2] ✉, Keith Fuglie[3], Paul W. Heisey[4] & Haiyan Deng[5]

Investment in public agricultural R&D is a key driver of agricultural productivity, which is regarded as one of the major solutions to enhance global food security, reduce the environmental impact of agricultural production and enhance crop resilience to climate change. Detailed information on public agricultural R&D investment is required to inform national science and technology policies that aim to improve the performance of the agricultural sector and monitor whether countries are on track to reach (inter)national targets. Due to the long lead-time, it requires exceptionally long time-series to assess the impact of public agricultural R&D spending on present and future agricultural productivity. To address these issues, this paper presents GRAPE - the WUR-ERS Global Research on Agriculture: Personnel & Expenditures dataset, which contains data on the number of public agricultural researchers and R&D expenditures for 190 countries, broadly covering the period 1960–2022. To construct the dataset, information from a large number of sources was combined, different R&D classifications were harmonized and missing data were imputed.

Background & Summary

Increasing agricultural productivity is regarded as one of the main solutions to raise food production and reduce the environmental impact of agricultural production[1]. A large body of studies has shown that public investments in agricultural R&D are an important driver of agricultural productivity growth. A meta-analysis found internal rates of return between 34% and 42% for high- and low-income countries, respectively[2]. Due its high potential to contribute to global food security, increasing the investment in agricultural R&D is one of the Sustainable Development Goal (SDG) 2 targets, to achieve zero hunger in 2030. In addition to boosting productivity growth and agricultural production, investment in R&D is also essential to develop crops and farming practices that are resilient to climate change, particularly in low-income countries, where a large part of the population depends on agriculture for a living[3].

Detailed information on public agricultural R&D investment is essential to inform national science and technology policies that aim to improve the performance of the agricultural sector. International comparable data will help analysts and governments to compare agricultural research investments with those of other countries and support monitoring whether countries are on track to reach (inter)national targets.

Although institutions such as the CGIAR, EUROSTAT and the OECD, have systematically collected agricultural R&D statistics for low, middle and high-income countries, their efforts have a strong regional focus, resulting in fragmented datasets that only cover a limited number of countries and are not always directly comparable. Moreover, they do not always contain long-run time-series. Many studies point out that the period between the start of research on agricultural technologies and when resulting innovations are adopted and improve productivity may span between 35 to 50 years[4]. Hence, it requires equally long time-series to investigate the impact of investment in agricultural R&D on improvements in agricultural productivity.

Several researchers have attempted to overcome these issues by combining and harmonizing national statistics to construct a global dataset of public agricultural R&D investment (See[5] for a review). The first initiative to create

[1]Wageningen Social & Economic Research, Wageningen University & Research (WUR), the Hague, the Netherlands. [2]International Institute for Applied Systems Analysis (IIASA), Laxenbourg, Austria. [3]Economic Research Service (ERS), U.S. Department of Agriculture, Washington, DC, USA. [4]Agricultural economist, retired in 2019 after 21 years with the U.S. Department of Agriculture Economic Research Service and 13 years with the International Maize and Wheat Improvement Center (CIMMYT), Washington, USA. [5]School of Economics, Beijing Institute of Technology, 100081, Beijing, China. ✉e-mail: michiel.vandijk@wur.nl

such a dataset was undertaken by[6], who collected data on R&D spending and personnel for 30 countries from the 1950s to the early 1970s. Updates and revisions of this study are[7] and[8]. In the 1980s the International Service for National Agricultural Research (ISNAR) was launched, which aimed to systematically collect and standardize international information on public agricultural R&D expenditure and personnel. Key outputs were[9] and[10] that present internationally comparable data with public agricultural R&D statistics for 154 countries, spanning the period 1960-86. In the 1990s, ISNAR transitioned to the Agricultural Science and Technology Indicators (ASTI) initiative, led by the International Food Policy Research Institute (IFPRI) (but discontinued in 2023), which mainly focused on the collection of international comparable agricultural R&D statistics for low- and middle-income countries but regularly compiled global estimates[11,12]. Another initiative to compile a global dataset of international agricultural R&D statistics is the InSTePP International Innovation Accounts: Food and Agricultural R&D data series compiled by the International Science & Technology Practice & Policy (InSTePP) center at the University of Minnesota[13,14]. The latest version[5] includes private and public agricultural R&D expenditure data for 158 countries, spanning the years 1960 to 2011, although this dataset does not publish annual time series by country.

Despite the long history of creating global datasets and overviews of public agricultural R&D investment, the existing datasets are largely outdated now and often not available in the public domain, which limits their usefulness to inform evidence-based policy making. To address this gap, this paper presents GRAPE - the WUR-ERS Global Research on Agriculture: Personnel & Expenditures dataset that contains international comparable data on the number of public agricultural researchers and R&D expenditures for 190 countries, broadly covering the period 1960–2022.

The GRAPE dataset makes three contributions to the previous agricultural R&D data collection efforts. First, it extends the country coverage to 190 countries, covering 99.7% of global GDP and and 99.6% of agricultural GDP, and updates the time-series to the recent past. Second, it presents complementary time-series on the number of researchers, which is an indicator of the human resource commitment to public investment in agricultural research[9]. Although reported by international initiatives, such as ASTI, EUROSTAT and OECD, previous attempts to construct a global dataset on public investment in agricultural research[10] have not included an indicator of research staff. Finally, it tests and adopts a consistent approach to handle missing information, which is a major issue for any effort to construct a global resource with public agricultural R&D data that involves combining 'patchy' information from a large number of different sources.

Preliminary versions of the GRAPE dataset have been used in[15–17] and[18] to analyse global patterns of investment in public agricultural R&D and its impact on agricultural productivity, food security and climate adaptation. This paper presents an updated version that includes a larger number of countries and more recent information as well as detailed documentation on the construction of GRAPE, sources and the approach used to deal with gaps in the data.

Methods

Concepts and definitions. For our definition of 'public agricultural R&D', we relied on the OECD Frascati Manual[19], which provides guidelines for the collection of research and development statistics (Supplementary Section A lists the key concepts and definitions that are relevant for this study). Public R&D is the total of R&D undertaken within two performing sectors: government (e.g. local, state and national institutions) and higher education (e.g. universities and related research institutes).

Agricultural R&D refers to primary agriculture, fishery and forestry research. In practice, primary agriculture (i.e. crop and livestock) research makes up the largest share although there are notable exceptions such as Sweden and Iceland where forestry and fishery make up the largest part of agricultural GDP[20]. We followed[9] and[20] and attempted to define research by purpose i.e. the socio-economic objective of the research program or project (e.g. agricultural productivity in case of improving the fuel efficiency of agricultural equipment), rather than its content (improvements in energy use in the aforementioned example). This approach is preferred because it captures the R&D activities that potentially contribute to agricultural output and productivity growth, and is therefore the most relevant measure to inform agriculture sector R&D policies. The OECD socio-economic objectives (SEO) classification, which includes a category Agriculture, is a widely used system to categorize research by purpose.

Another option is to classify research by field of science, which refers to the nature or the knowledge domain of the activity. This is done by the OECD Fields of research and development (FORD) classification, which includes a category: Agricultural and veterinary sciences. Due to the nature of activities, most countries use socio-economic objective classifications to allocate government R&D, while field of science classifications, such as FORD, are preferred for the R&D performed by the higher education sector.

In practice, it is not straightforward to implement socio-economic and field of science classifications consistently as there are ambiguities within socio-economic sectors and fields of science classes, which means research activities could be allocated to multiple categories (see section 12.4 in[19]). Also countries tend to use variations of the SEO and FORD classifications and even might use different systems across performing sectors and indicators (i.e. expenditure and human resources). Where possible, we preferred the socio-economic objective over field of science classification when both types were available.

We included two indicators of public agricultural R&D. The first one is R&D spending (RD), expressed in 2017 PPP$ values (but we also provided values in local currency units (LCUs) and US$). This measure is commonly reported in inter(national) studies and reports that analyze investment in agricultural R&D. It is also a key statistic to assess the extent to which public investment contributes to an increase in agricultural (total factor) productivity[15,21,22]. In addition, we also compiled time-series on the number of researchers expressed in full time equivalent (HR) that are employed by public agricultural research entities. Alongside R&D spending, this indicator provides additional insights on the level of agricultural research capacity and commitment per country[10].

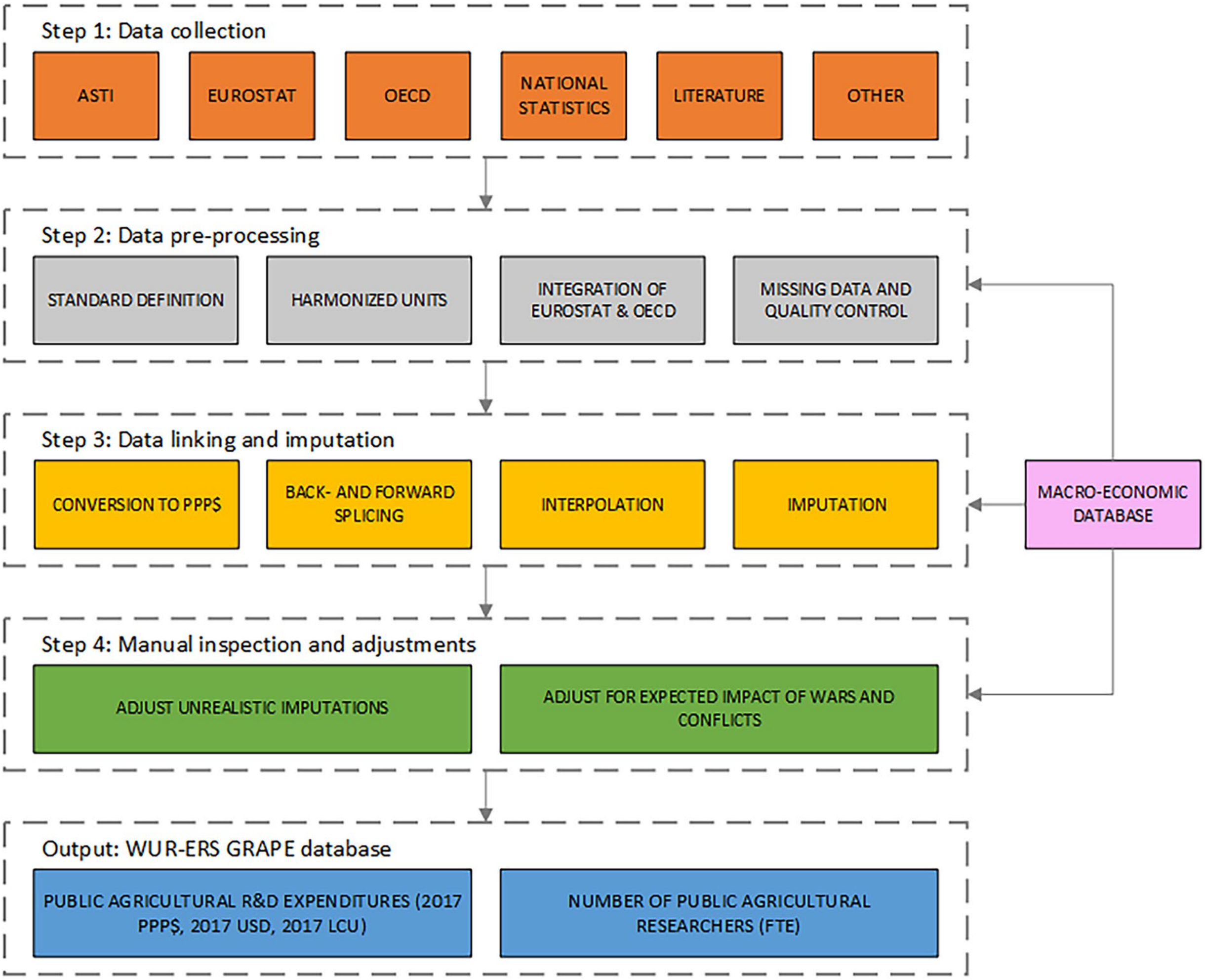

Fig. 1 Steps to construct the GRAPE dataset.

Construction of dataset. A schematic overview of the steps involved to construct the GRAPE dataset is presented in Fig. 1. It involves four steps: (1) Data collection; (2) Data pre-processing; (3) Data linking and imputation and (4) Manual inspection and adjustments, which are discussed in the sections below.

Step 1: Data collection. The first step in the creation of the GRAPE dataset was the systematic collection of data on RD and HR per country for the period 1960-2022. We started by gathering data from several 'core' initiatives, which we considered as the most detailed and reliable sources of national agricultural R&D statistics: ASTI (for low- and middle-income countries) and OECD and EUROSTAT (for high-income countries). In addition, we gathered data from other international and regional initiatives that present global (UNESCO) and regional (RICYT for middle- and Latin American countries) information. FAO (2024) also provides information on RD but only reports government budget allocations for agricultural R&D (GBARD), which reflects expenditures by government funding agencies, which may go to public or private research performers, and excludes non-government sources of funding for public research institutes[23]. An in-depth analysis showed that GBARD is a poor predictor of actual spending (see Supplementary Section B). We therefore decided to only use this type of data when no other sources of information were available and when trends seem in line with additional information on RD and/or HR for periods with overlapping information.

For data before 1980, which is not covered by most of the aforementioned sources, we took information from[6,9,10] and ISNAR country briefs (various). For countries, for which data was still missing or with limited coverage, we (a) consulted publications and websites from national statistical offices, building on the data presented in[20] and (b) collected information from scientific and grey literature, informed by earlier efforts of research team members to compile a global public agricultural research dataset[12,15]. Table 1 provides an overview of all sources used. Finally, for ten countries (See GRAPE documentation), we found data spanning the pre-1960 period that for some countries even goes back to the end of the 19th century, which we also added to the dataset. All literature sources for the dataset are listed in the data file and cited in this paper[6,9,20,23–81].

To support the harmonization of units and the imputation procedures (see below), we created a complementary global database with macro-economic information for the period 1960-2022. Key data series in the database include implicit GDP deflators, purchasing power parity (PPP) conversion factors, agricultural GDP

Source	Coverage	Time period	Indicator	Note
ASTI[56–59]	Low- and middle-income countries	1981–2021	HR & RD	Discontinued
EUROSTAT[77]	EU27 and several other countries	1980–2022	HR & RD	
FAO[23]	Global	2001–2022	GBARD	
OECD[75]	OECD member countries	1980–2022	HR & RD	
UNESCO[74]	Global	1996–2018	HR & RD	Discontinued
RICYT[73];	Central and Latin America	1996–2022	HR & RD	
ISNAR country briefs (various)	Low- and middle-income countries	1960–1992	HR & RD	Discontinued
Boyce & Evenson[6]	Global	1950–1974	HR & RD	Not always comparable
Pardey & Roseboom[9]	Global	1960–1986	HR & RD	
Pardey *et al.*[53]	Global	1961–1981	HR & RD	
Heisey and Fuglie[20]	High-income countries	1960–2013	RD	
National statistics (various)	Global	Varies	Varies	
Literature (various)	Global	Varies	Varies	

Table 1. Overview of GRAPE data sources. HR and RD refer to number of public agricultural researchers and R&D expenditures, respectively.

series and agricultural output statistics. The main source of the deflators was[82], which offers, depending on the country, macro-economic time-series from 1960 onward. For countries, with incomplete or missing data, we took complementary information from[83] (since 1970) and[84] (since 1980), which also provided GDP deflator time-series information. A comparison between countries, for which all three data sources include information, showed little if any difference. We used historical data from[9] to backcast the deflator series of some countries to 1960 using the annual rate of change. PPP conversion factors were taken from[82] and extended with similar information from[84] for a few countries that were still missing. Finally, we compiled times series on agricultural GDP, which were taken from[83], and agricultural output, which were obtained from[85].

Step 2: Data pre-processing. After the data collection step, we processed the raw data to create international comparable series. In particular, we addressed the following issues:

- *Standard definition of public agricultural R&D*. As explained above, countries may classify public agricultural research activities by socio-economic objective, our preferred measure, or field of science. In case we collected primary information from national statistics (e.g. China and Canada), we tried to collect data by socio-economic objective. Similarly, in case data was taken from EUROSTAT and OECD datasets, which sometimes report data using different classifications systems for the same country (also see below), we selected the socio-economic objective series.
- *Harmonization of units*. First, we converted all current RD statistics in LCUs to 2017 PPP$ values, applying the recommended 'first deflate, then convert' approach[86]. A few, mainly historical sources, presented data in constant PPP$ values but with an older base year than 2017. We rebased these to 2017 PPP$ using relative PPP conversion factors if PPPs for the older base year were available. In other cases we used the US GDP deflator. Second, where needed, we converted HR in head count (HC) units to FTE. For the conversion, we used data from[74], which, for a selection of countries, presented data both in HC and FTE units. These conversion factors were subsequently applied to countries with similar levels of economic development to derive the number of researchers in FTE. In other cases we applied growth rates in HC series to backcast recent data in FTE.
- *Integration of EUROSTAT and OECD statistics*. Both the EUROSTAT and OECD datasets provide detailed information on research activities by performing sector for European countries using similar classifications systems. A close comparison showed that, although the statistics are nearly identical, longer time-series can be generated by combining data from both sources. For this reason, we integrated the statistics into one dataset (referred to as EUROSTAT & OECD in the dataset). In case of inconsistencies, we used EUROSTAT data for the European countries. We applied several simple imputation techniques to deal with missing data, such as using the growth rate of FORD statistics to extrapolate missing SEO data. In rare cases, we also used RD trends to extrapolate HR, and vice versa.
- *Handling of missing data and quality control*. Depending on the source and country, occasionally we found one to three year data gaps within the time series. We imputed missing values assuming a constant exponential growth rate (see next section). In rare cases, we omitted or adjusted (i.e. in case of obvious data entry errors) extreme values, which we considered as implausible. We applied exponential interpolation in case this resulted in data gaps.

Step 3: Data linking and imputation. In the final step, we linked the data from different sources to create consistent long-run time series of RD and HR, and imputed remaining missing values.

As mentioned above, dealing with missing information in public agricultural R&D statistics is a major issue that needs to be tackled in order to create long-run continuous time-series. Most previous studies have applied basic techniques, such as linear interpolation and back- and forward casting using the rate of change in agricultural GDP (also referred to as agricultural research intensity (ARI), which is the ratio of RD to agricultural GDP)[13]. To the best of our knowledge, there are no studies, which systematically investigated, the performance of these and other techniques to impute missing data in agricultural R&D statistics. To select the best-performing approaches, we conducted an in-depth assessment and comparison of techniques (see Supplementary Section B). The assessment established that an ensemble approach in which different imputation results are averaged provided the most accurate results. This is in line with the literature on forecasting, where ensemble approaches are also recommended[87]. This resulted in the following approach to handle missing data.

First, for each country, where possible, we spliced all data by using growth rates and information in overlapping years. As a rule, we started with the most recent statistics and worked backwards in time to splice all data together. The exception to this procedure was when ASTI and EUROSTAT-OECD, which we considered as the most reliable data sources, did not include the most recent information. In such cases, we spliced ASTI/EUROSTAT-OECD data forward using information from, for example FAO, RICYT and national statistics, which sometimes included more up to date statistics. Before merging the different data sources, we closely inspected all the available series for each country and discarded series which were deemed implausible or inconsistent.

Second, we applied several methods to deal with missing information that remained after splicing the data together. We distinguished between three different types of missing data, which were each treated differently (Fig. 2):

- *Data missing within time-series*, i.e. when both older and newer data were available, with a data gap in the middle. To impute this type of missing information, we applied simple exponential growth rate interpolation (EI) as this approach showed superior performance over ARI imputation and more advanced multiple imputation techniques[88], both of which require auxiliary data (e.g. agricultural GDP and other covariates) that are not always available for historical periods (Supplementary Section B).
- *Data missing for recent years*, i.e when data is missing between the year of the latest observation and 2022. For a large number of countries data for HR and RD were not available for the most recent one to five years. The main reason for this was that datasets such as EUROSTAT and OECD were not always fully up to date, whereas others, such as ASTI and UNESCO, were discontinued and therefore did not contain the most recent observations. To impute these values, we used an ensemble approach, which takes the mean of different imputation methods. Where possible, we combined the following imputation approaches: (a) Extrapolation of data using the growth rate of agricultural GDP (AGGDP); (b) Extrapolation of data using the growth rate of agricultural output (AGOUTPUT); (c) Extrapolation using the growth rate of RD or HR, when information on one of them was more complete than the other; (d) NAIVE forecasting, in which the last observation is carried forward; and two common time-series forecasting approaches: (e) simple exponential smoothing (SES), which is the preferred method for stationary data with no clear trend, giving more weight to recent observations; and (f) Damped Trend (DT), which works best for data with a (long-run) trend[89]. SES and DT require at least 3–4 times the length of forecast. We therefore applied these methods only in case the number of observations was 24 or larger and the number of missing years was six or smaller.
- *Data missing for distant years*, i.e. when data was missing between 1960 and the first year of observation. This problem mostly occurs for former USSR and several Eastern European countries, for which data collection started around 1990 and data for previous decades is frequently missing, as well as for a handful of countries that are not covered by[9] and ISNAR country reports. Unless, additional information on HR/RD was available, we decided not to impute data missing for distant years.

Step 4: Manual inspection and adjustments. After the data linking and imputation step, we manually inspected the results to see if it resulted in extreme values and implausible rates of change. We also took into account the expected negative impact of conflicts and wars on the investment in public agricultural R&D. For several countries, we decided to remove the AGGDP based imputation from the ensemble because of extreme fluctuations in the AGGDP series, resulting in unrealistic imputations. We also manually adjust the HR and RD series for Solomon Islands to very low levels following an episode of civil or communal violence that severely impaired the national agricultural R&D system (see GRAPE documentation).

Figure 3 provides HR and RD trends for Korea to illustrate the information in the GRAPE dataset. The figure shows that the RD series was constructed by splicing together data from various sources. Only the data point for 1960 was imputed using the ensemble approach (i_ens). For the HR series, the availability of information was more limited and consequently a large number of data points were imputed using the exponential growth interpolation (ei) and ensemble imputation approaches. The figure also shows the uncertainty range of the ensemble approach demarcated by the lower and upper bounds of the various underlying imputation methods.

Coverage. The core of the GRAPE dataset broadly covers the period 1960–2022. However, depending on data availability, actual intertemporal coverage per country might differ. Figure 4 depicts the number of countries per year for which information (including imputed) is available in GRAPE. For ten countries, we managed to add HR (only one country) and RD information for the pre-1960 period. For the majority of countries the first year of data is 1960 (74) or 1961 (83). Most of the remaining countries with shorter time-series are former USSR economies, located in the Europe & Central Asia region for which statistics are only available from 1990 onward. Another group of countries with incomplete data are small island nations, located in the East Asia & Pacific region.

Figure 5 provides an overview of the (weighted) share of observations per source for HR and RD for each decade. The figure shows that data coverage is very high in the period 1960-79, resulting in a relatively low share of imputed values. This is mainly due to the extensive data collection efforts of the, now defunct, International Service for National Agricultural Research (ISNAR)[9,10]. For the period 1980-2019, much less

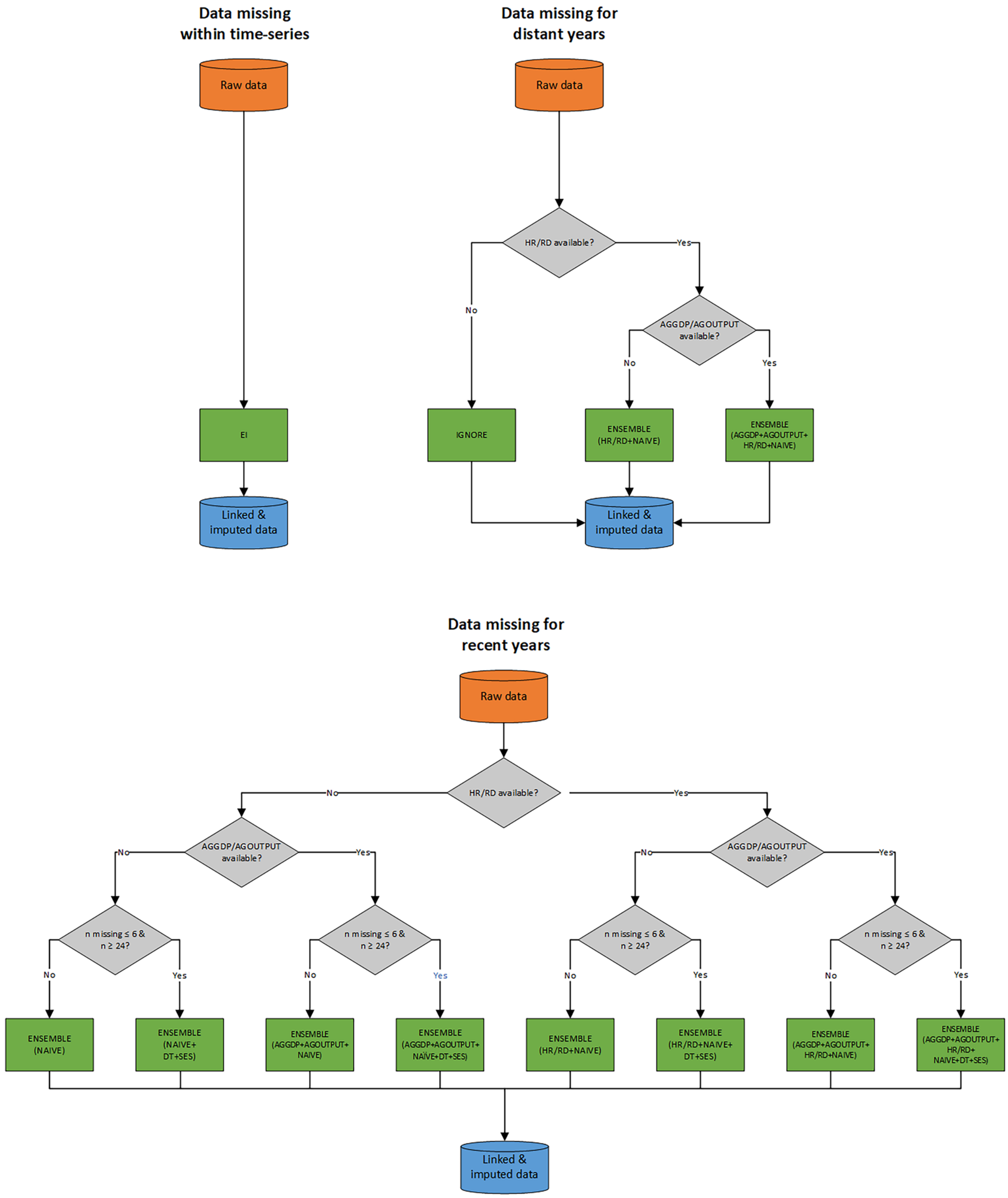

Fig. 2 Decision tree for imputation of missing data. Note that for a few, mostly small-island economies AGGDP or AGOUTPUT data was not available and therefore only one of the related imputation techniques was applied.

observations are available as indicated by a share of 33–57% imputed values. However, the share of imputed values weighted by R&D spending is much lower because information is predominantly missing for small countries. This is not the case for HR, where data for India and China were often not available, resulting in a high imputation share. For the most recent period (2020-22), data is largely incomplete because of the discontinuation (e.g. ASTI and UNESCO) and limited updating (e.g. EUROSTAT, OECD and FAO) of international data collection efforts.

There are also stark regional differences in data availability and the share of observations that was imputed (Fig. 6). Data coverage for both HR and RD is generally high for North and South-America, and a selection of

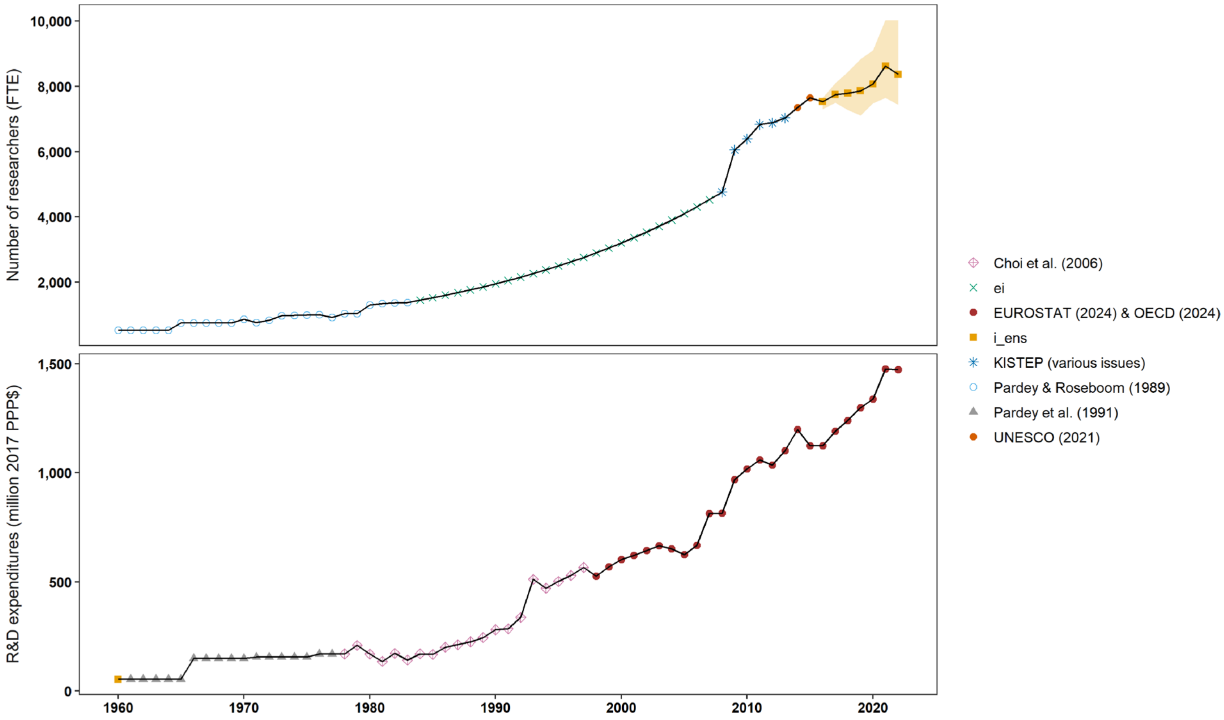

Fig. 3 Example of GRAPE country information. The figure shows the HR and RD trends for Korea for the period 1960-2022.The yellow area demarcates the lower and upper bound of the ensemble imputation approach (i_ens).

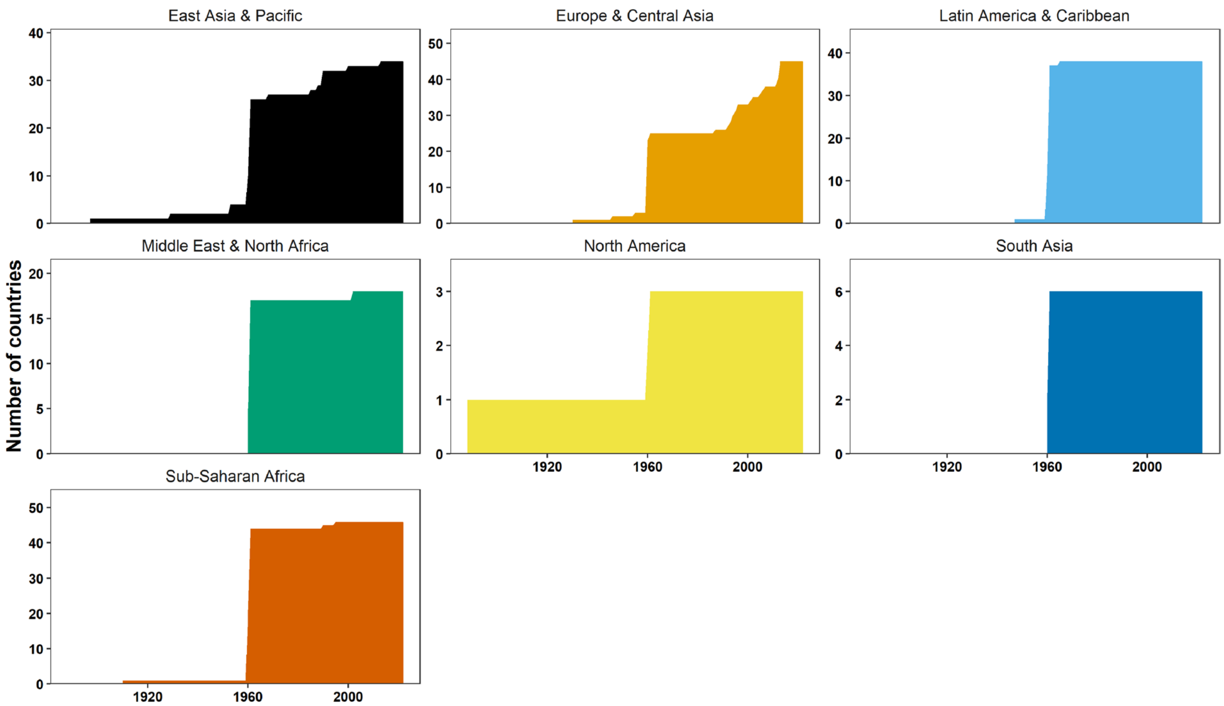

Fig. 4 Coverage of GRAPE by region and over time. The figure shows the number countries for which RD is included between 1888 and 2022. HR is only included after 1960 for each country and year, except for South Africa for which RD and HR are available from 1910 to 2022.

Asian and African countries, while for Europe, parts of Asia and Oceania only RD information was identified. In contrast, for a number of African countries and most of the former USSR economies, only a limited number of data points are available for both HR and RD, resulting in a high share of imputed values.

Fig. 5 Summary of data sources by decade and indicator. Share of imputed (a) RD and (b) HR values. Imputed refers to all observations that were imputed using one of the approaches described in the text and indicated by one or more of the following codes: ei, i_aux, i_naive and, i_ens and i_man (Tables 2 and 3).

Data Records

The GRAPE dataset and supporting information is available on Zenodo at the following repository: https://zenodo.org/records/15507361[90]. The repository includes nine files. The version number of the GRAPE dataset decribed in this paper is v1.0.0.

- grape_v1.0.0.xlsx is the main data file, which contains the HR (FTE) and RD (2017 PPP$) data series from the first year data was available to 2022. The file includes (a) two columns to identify the country (country name and the corresponding ISO three character code); (b) year, variable (HR or RD) and unit columns; (c) a column with a short reference to the data source; (d) a column with a code to describe the adjustments to the raw data (e.g. treatment of extreme values and missing data - see Usage notes); (e) a column with a code to describe the data linking and imputation approaches (see Usage notes) and (f) three columns with data series, including the HR/RD value and the lower and upper bound in case of ensemble imputation.
- rd_v1.0.0.xlsx, presents the RD series in 2017 PPP$ values (identical to the main data file), constant 2017 LCU and constant 2017 US$.
- hr_pre_imp_v1.0.0.rds, contains the HR (FTE) series before imputation in the R[91] data format.
- rd_pre_imp_v1.0.0.rds, contains the RD (2017 PPP$) series before imputation in the R[91] data format.
- ppp_2017_db_v1.0.0.rds, contains the 2017 PPP$ conversion factors in the R[91] data format.
- xr_2017_db_v1.0.0.rds, contains the 2017 LCU-USD exchange rates in the R[91] data format.
- macro_db_v1.0.0.xlsx, contains the macro-economic database with time series for PPP$, GDP deflator, population, share of rural population, share of agricultural GDP, agricultural GDP, GDP per capita, agricultural output and agricultural GDP.
- grape_documentation_v1.0.0.pdf, provides detailed country level documentation, explanation of codes and the full reference list of data sources that can be linked to the short reference column in the main data file.
- grape_version_history.pdf, contains the version history of the GRAPE dataset, including a summary of changes made to the data.

Technical Validation

Manual validation and adjustments. During various steps to construct the GRAPE dataset (Fig. 1), we undertook actions to inspect, validate and, where needed, correct the data. In step 2, we manually inspected the raw data for outliers, implausible patterns and missing information. Examples of such errors include zero values where a missing value would be most logical, the extrapolation of period averages to individual years, and obvious data entry mistakes (e.g. misplacement of the decimal separator). Depending on the problem, we omitted the observation and replaced it using our imputation procedure or corrected the error.

When splicing the data from different sources (step 3), all available time-series were visually inspected country by country. In case of unexpected or contradicting trends between different data sources, the country series

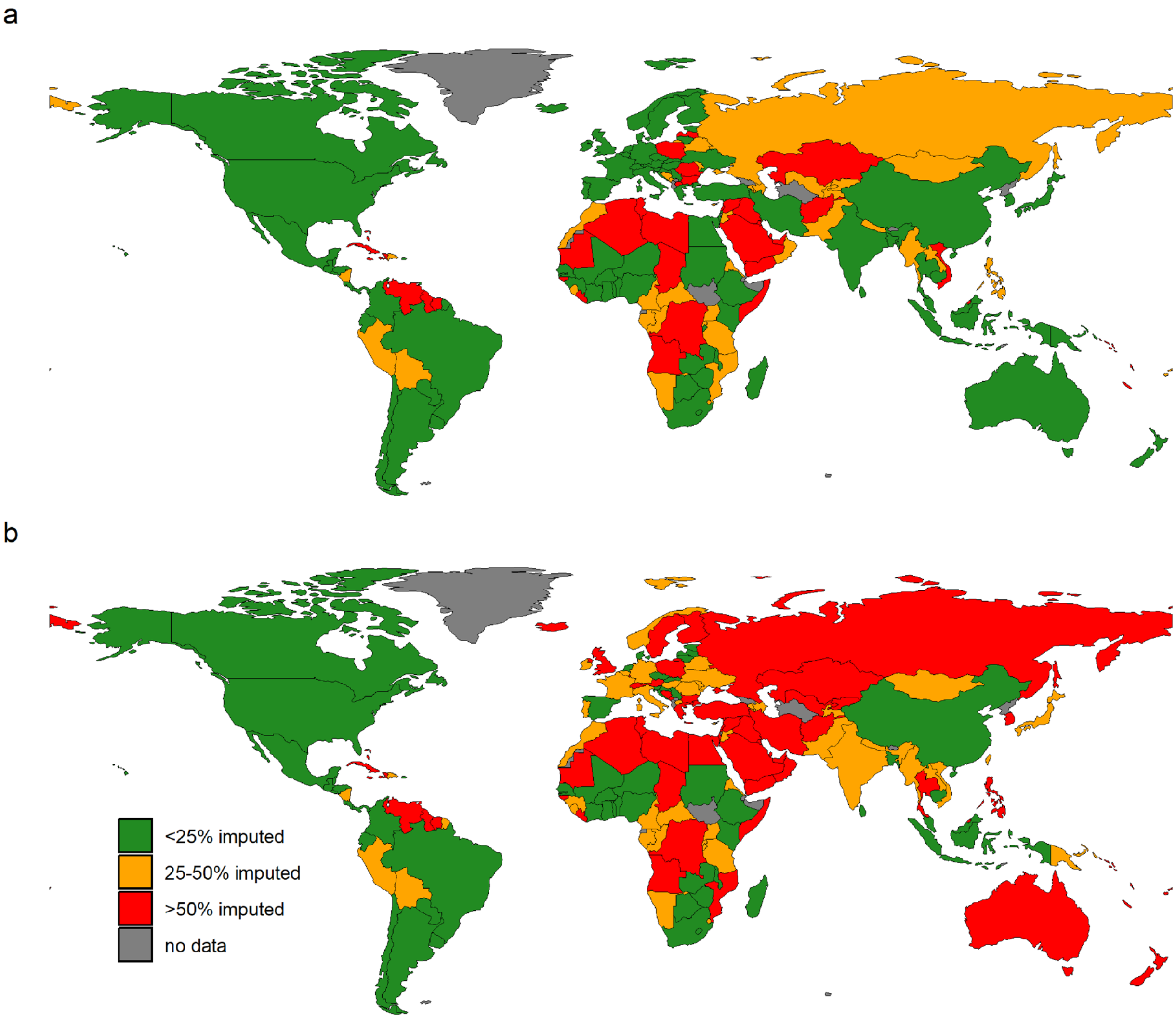

Fig. 6 Geographical distribution of the share of imputed observations for (a) RD and (b) HR per country. The share of imputed observations is calculated by dividing the number of imputed observations by the sum of number of observed and imputed observations, as such accounting for the relative short time-series of certain countries (e.g. former USSR).

were discussed with the research team and a joint decision was taken on which sources to use and how to process and combine the available information. For several countries, we decided not to use the[23] GBARD data because the pattern was regarded as too volatile and not in line with ASTI and EUROSTAT-OECD datasets, which were regarded as the most consistent and reliable.

In step 4, we used two simple metrics to validate the consistency of the HR and RD series. First, we inspected the distribution of R&D spending per researcher (i.e. RD divided by HR) across broad regions. Although, differences across countries and regions were expected, 2017 PPP$ per FTE should be broadly comparable between countries in the same stage of economic development. Second, we calculated the ARI for each country and year. Also here, it was expected that the share of R&D expenditures in agricultural GDP was within a comparable range. We closely inspected that country series with extreme values and, where appropriate adjusted outliers or discarded raw data sources with extreme values.

Finally, the HR and RD data series for each country were manually inspected by the research team, some of which have considerable experience with constructing and analyzing long-run public agricultural datasets. As described above, for a few countries, the imputations were considered as unrealistic and adjusted accordingly. All actions to adjust the data were given a code in the main data file and were summarized in the GRAPE documentation.

Comparison with alternative dataset. We validated the GRAPE dataset by comparing it with a comparable dataset presented in[14], which analysed global patterns of public and private agricultural R&D spending using the latest version of the InSTePP International Innovation Accounts: Food and Agricultural R&D database[5]. Unfortunately, the annex of the paper includes only a subset of the global InSTePP dataset, which has not been made public and therefore a detailed comparison was not possible. Nonetheless, the available data can be regarded as a good proxy because it covers six years spanning the period 1980–2011 and 20 countries, including various low-, middle and high-income economies.

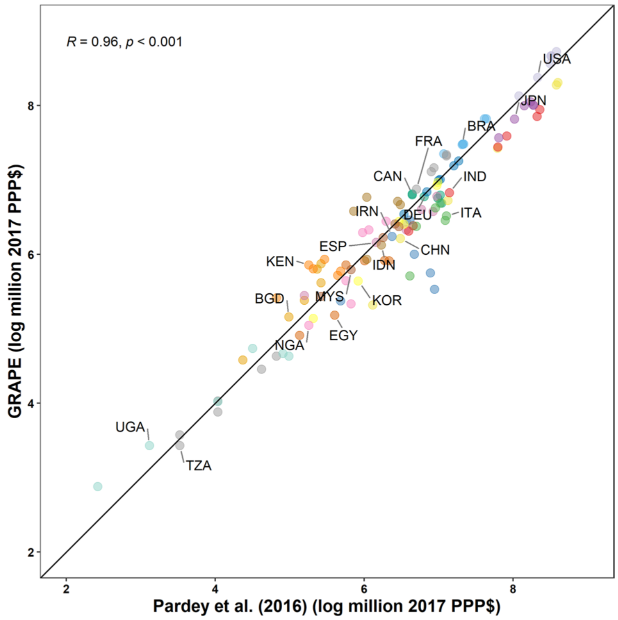

Fig. 7 Comparison between GRAPE agricultural R&D expenditures and InSTePP dataset presented in[14] (20 countries, 6 periods). For comparison, InSTePP data expressed in 2009 PPP$ values were converted to 2017 PPP$ values using the US GDP deflator from[82].

Code	Description
none	No imputation, interpolation or data changes, excluding conversion to standard units (i.e. 2017 PPP$ and FTE).
pa	Use of period averages for each individual year.
i_aux	Government and/or higher education series were imputed using the growth rate of an auxilliary data series (e.g. SEO-FORD or HR-RD series). EUROSTAT & OECD only.
ei	Interpolation of missing values using a constant exponential growth rate between two years.

Table 2. Codes for quality control adjustments and treatment of missing data (step 2 in Fig. 1).

Code	Description
none	No linking of data.
splice_new	Data was spliced together with more recent data series applying the growth rate for overlapping years.
splice_old	Data was spliced together with older data series applying the growth rate for overlapping years.
ei	Interpolation of missing values using a constant exponential growth rate between two years.
i_naive	Imputation using naive forecasting, i.e. carry the last observation forward.
i_ens	Imputation using an ensemble of imputed values - see GRAPE documentation for details.
i_man	Manual adjustment of imputed values - see GRAPE country documentation for details.

Table 3. Codes for data linking and imputation approaches (step 3 in Fig. 1).

We found a very high correlation (0.96) with the InSTePP dataset (Fig. 7). This is not surprising as for a large number of countries, both the datasets relied on the same international efforts to collect agricultural R&D expenditure information (e.g. ASTI, EUROSTAT and OECD) although methods to process, harmonize and impute the data might be different. Only for Iran (IRN), the GRAPE dataset includes several observations that are somewhat lower than those in the InSTePP dataset. The most likely reason for this is that we integrated new data from[61], which were not used in the InSTePP dataset.

Usage notes

Despite our efforts to create an internationally comparable dataset, the GRAPE dataset is not without limitations. First, countries use a variety of classifications systems to measure the size of investment in public agricultural R&D, including socio-economic objective and field of science approaches. We made an effort to harmonize the data series, use consistent definitions and validate results but complete harmonization was not always possible.

Second, in order to create long-run series, covering multiple decades, a relatively large-number of data points had to be imputed. Data availability differs considerably between the two indicators and in time and space. While there is good coverage for high-income, Latin American and several African and central Asian countries, data availability is limited for other regions, resulting in a large share of imputed values. Further, in particular for the most recent period 2020-22, the number of imputed values is large. Overall, the coverage for the HR series is much lower than that for RD information. In one case (Solomon Islands), we adjusted the imputed data to account for the impact of civil conflict on public agricultural R&D investment. But such cases are rare. In counties where agricultural innovation systems face large year-to-year fluctuations in funding (due to a government budget crisis or to a large infusion of funds for, say capital expansion), our imputation approach would tend toward predicting funding in a 'normal' year. Users of GRAPE should be mindful of these limitations and be careful when using the dataset to study international public agricultural R&D trends.

The main data file includes two columns with codes that refer to the approaches that were used to handle extreme values and missing information in the raw data and the methods used to link information from various data sources, such as back- and forward splicing, and imputation. Tables 2 and 3 present a brief explanation of the codes. The main data file also includes the lower and upper bound in case ensemble imputation was applied. By using these codes and bounds, users of GRAPE have the option to easily identify modifications to the raw data, remove adjusted observations or take into account uncertainty when conducting statistical analysis and apply alternative imputation techniques to deal with missing information. We aim to update the GRAPE dataset in the coming years with more recent data that will replace currently imputed values.

Code availability

Code was stored in a GitHub repository: https://github.com/michielvandijk/GRAPE. This includes R code to apply the imputation approaches to the pre-processed R&D statistics and convert R&D expenditure data in 2017 PPP$, LCU and USD values. It also includes code to reproduce all the figures in this paper and the Supplementary Information.

Received: 22 April 2025; Accepted: 3 June 2025;
Published online: 03 July 2025

References

1. Foley, J. A. *et al*. Solutions for a cultivated planet. *Nature*. https://doi.org/10.1038/nature10452 (2011).
2. Rao, X., Hurley, T. M. & Pardey, P. G. Are agricultural R&D returns declining and development dependent? *World Development* **122**, 27–37, https://doi.org/10.1016/j.worlddev.2019.05.009 (2019).
3. Jayne, T. S. & Sanchez, P. A. Agricultural productivity must improve in sub-Saharan Africa. *Science* **372**(6546), 1045–1047, https://doi.org/10.1126/science.abf5413 (2021).
4. Alston, J. M., Pardey, P. G., Serfas, D. & Wang, S. Slow Magic: Agricultural Versus Industrial R&D Lag Models. *Annual Review of Resource Economics* **15**(1), 471–493, https://doi.org/10.1146/annurev-resource-111820-034312 (2023).
5. Pardey, P. G., Chan-Kang, C., Beddow, J. M. & Dehmer, S. P. *In STePP International Innovation Accounts Research and Development Spending, version 3.5*. (2016).
6. Boyce, J. K. & Evenson, R. E. *National and international agricultural research and extension programs*. Agricultural Development Council. (1975).
7. Oram, P. A. & Bindlish, V. *Resource Allocations to National Agricultural Research: Trends in the 1970's*. International Service for National Agricultural Research; International Food Policy Research Institute. (1981).
8. Judd, M. A., Boyce, J. K. & Evenson, R. E. Investing in Agricultural Supply: The Determinants of Agricultural Research and Extension. *Economic Development and Cultural Change1* **35**(1), 77–113 (1986).
9. Pardey, P. G. & Roseboom, J. *ISNAR sgricultural research indicator series: A global data base on national agricultural research systems*. Cambridge University Press. (1989).
10. Pardey, P. G., Roseboom, J. & Anderson, J. R. Regional perspectives on National Agricultural Research. In Pardey, P. G., Roseboom, J. & Anderson, J. R. (Eds.), *Agricultural research policy: International quantitative perspectives* (pp. 197–264). Cambridge University Press. (1991).
11. Pardey, P. G. & Beintema, N. M. *Slow Magic: Agricultural R&D a Century After Mendel, IFPRI Food Policy Report*. International Food Policy Research Institute. (2001).
12. Beintema, N., Stads, G.-J., Fuglie, K. O. & Heisey, P. *ASTI global assessment of agricultural R&D spending*. IFPRI & GFAR. https://doi.org/10.2499/9780896298026 (2012).
13. Pardey, P. G., Alston, J. M. & Chan-Kang, C. Public agricultural R&D over the past half century: An emerging new world order. *Agricultural Economics (United Kingdom)* **44**(SUPPL1), 103–113, https://doi.org/10.1111/agec.12055 (2013).
14. Pardey, P. G., Chan-Kang, C., Dehmer, S. P. & Beddow, J. M. Agricultural R&D is on the move. *Nature* **537**(7620), 301–303, https://doi.org/10.1038/537301a (2016).
15. Fuglie, K. R&D Capital, R&D Spillovers, and Productivity Growth in World Agriculture. *Applied Economic Perspectives and Policy* **40**(3), 421–444, https://doi.org/10.1093/aepp/ppx045 (2018).
16. Baldos, U. L. C., Fuglie, K. O. & Hertel, T. W. The research cost of adapting agriculture to climate change: A global analysis to 2050. *Agricultural Economics* **51**(2), 207–220, https://doi.org/10.1111/agec.12550 (2020).
17. Hertel, B. T. W., Baldos, U. L. C. & Fuglie, K. O. Trade in technology: A potential solution to the food security challenges of the 21st century. *European Economic Review* **127**, 103479, https://doi.org/10.1016/j.euroecorev.2020.103479 (2020).
18. Jin, Y., Kristkova, Z. S., Kardung, M. & Wesseler, J. Impacts of accelerating agricultural R&D transfer on global food security. *GM Crops & Food* **15**(1), 1–12, https://doi.org/10.1080/21645698.2024.2438419 (2024).
19. OECD. *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*. OECD. https://doi.org/10.1787/9789264239012-en (2015).
20. Heisey, P. W. & Fuglie, K. O. *Research Investment and Policy Reform in High-Income Countries*. U.S. Department of Agriculture, Economic Research Service. (2018).
21. Alston, J. M., Andersen, M. A., James, J. S. & Pardey, P. G. The economic returns to U.S. Public agricultural research. *American Journal of Agricultural Economics* **93**(5), 1257–1277, https://doi.org/10.1093/ajae/aar044 (2011).
22. Baldos, U. L. C., Viens, F. G., Hertel, T. W. & Fuglie, K. O. R&D Spending, Knowledge Capital, and Agricultural Productivity Growth: A Bayesian Approach. *American Journal of Agricultural Economics* **101**(1), 291–310, https://doi.org/10.1093/ajae/aay039 (2019).
23. FAO. *FAOSTAT: Government expenditure*. https://www.fao.org/faostat/en/%7B/#%7Ddata/IG (2024).

24. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Botswana*. International Service for National Agricultural Research. (1994).
25. Mazzucato, V. *Statistical Brief on the National Agricultural Research System of Burkina Faso*. International Service for National Agricultural Research. (1994).
26. Beintema, N. M., Pardey, P. G. & Roseboom, J. *Statistical Brief on the National Agricultural Research System of Cape Verde*. International Service for National Agricultural Research. (1994).
27. Falconi, C. A. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Colombia*. International Service for National Agricultural Research. (1993).
28. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Cote d'Ivoire*. International Service for National Agricultural Research. (1994).
29. Roseboom, J., Beintema, N. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Ethiopia*. International Service for National Agricultural Research. (1994).
30. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Ghana*. International Service for National Agricultural Research. (1994).
31. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Kenya*. International Service for National Agricultural Research. (1994).
32. Beintema, N. M., Pardey, P. G. & Roseboom, J. *Statistical Brief on the National Agricultural Research System of Lesotho*. International Service for National Agricultural Research. (1995).
33. Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Madagascar*. International Service for National Agricultural Research. (1994).
34. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Malawi*. International Service for National Agricultural Research. (1993).
35. Mazzucato, V. *Statistical Brief on the National Agricultural Research System of Mali*. International Service for National Agricultural Research. (1994).
36. Beintema, N. M., Pardey, P. G. & Roseboom, J. *Statistical Brief on the National Agricultural Research System of Mauritius*. International Service for National Agricultural Research. (1995).
37. Beintema, N. M., Pardey, P. G. & Roseboom, J. *Statistical brief on the national agricultural research system of Namibia*. International Service for National Agricultural Research. (1994).
38. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Niger*. International Service for National Agricultural Research. (1993).
39. Roseboom, J., Beintema, N. M., Pardey, P. G. & Oyedipe, E. O. *Statistical Brief on the National Agricultural Research System of Nigeria*. International Service for National Agricultural Research. (1994).
40. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Rwanda*. International Service for National Agricultural Research (ISNAR). (1993).
41. Mazzucato, V. & El Habib LY, M. *Statistical Brief on the National Agricultural Research System of Senegal*. International Service for National Agricultural Research. (1994).
42. Roseboom, J., Pardey, P. G., Bach, H. S. V. & Zyl, J. V. *Statistical Brief on the National Agricultural Research System of South Africa*. International Service for National Agricultural Research. (1995).
43. Beintema, N. M., Pardey, P. G. & Roseboom, J. *Statistical Brief on the National Agricultural Research System of Sudan*. International Service for National Agricultural Research. (1995).
44. Beintema, N. M., Pardey, P. G. & Roseboom, J. *Statistical Brief on the National Agricultural Research System of Swaziland*. International Service for National Agricultural Research. (1995).
45. Beintema, N. M., Pardey, P. G. & Rose. *Statistical Brief on the National Agricultural Research System of Togo*. International Service for National Agricultural Research. (1996).
46. Roseboom, J. & Pardey, P. G. *Statistical Brief on the National Agricultural Research System of Zambia*. International Service for National Agricultural Research. (1995).
47. Roseboom, J., Pardey, P. G., Beintema, N. M. & Mudimu, G. D. *Statistical Brief on the National Agricultural Research System of Zimbabwe*. International Service for National Agricultural Research. (1995).
48. Hayami, Y. *A Century of Agricultural Growth in Japan*. University of Minnesota Press. (1975).
49. Childress, M. D. *Agrarian Research Institutes and Civil Society in Kazakhstan and Kyrgyzstan - In Search of Linkages*. United Nations Research Institute for Social Development. (2004).
50. Fuglie, K. O. & Nelson, K. P. *Agricultural and Food Research and Development Expenditures in the United States. ERS Public Data Product. April 2025 update*. U.S. Department of Agriculture. https://www.ers.usda.gov/data-products/agricultural-and-food-research-and-development-expenditures-in-the-united-states (2025).
51. Van der Eng, P. *Agricultural Growth in Indonesia: Productivity Change and Policy Impact Since 1880*. St. Martin's Press. (1996).
52. Chancellor, W. *Agricultural research and development investment in Australia*. Australian Bureau of Agricultural; Resource Economics; Sciences (ABARES). (2023).
53. Pardey, P. G., Roseboom, J. & Anderson, J. R. (Eds.). *Agricultural research policy international quantitative perspectives*. Cambridge University Press. (1991).
54. Agroscope. (n.d.). *Annual report - various issues*. Agroscope – Agricultural Research for Switzerland.
55. Lemarié, S. *et al*. Assessing the long-term impact of agricultural research on productivity: evidence from France. *European Review of Agricultural Economics* **47**(4), 1559–1586, https://doi.org/10.1093/eurrag/jbz051 (2020).
56. ASTI. *ASTI global R&D database*. https://www.asti.cgiar.org (2017).
57. ASTI. *ASTI global R&D database*. https://www.asti.cgiar.org (2014).
58. ASTI. *ASTI global R&D database*. https://www.asti.cgiar.org (2024).
59. ASTI. *ASTI West Africa update*. International Food Policy Research Institute. (2024).
60. Ardito-Barletta, N. *Costs and Benefits of Agricultural Research in Mexico* [PhD thesis]. University of Chicago. (1971).
61. Shamsadini, S., Yazdani, S. & Moghaddasi, R. Effects of R&D spending and Spill-Over on Agricultural TFP in Iran. *International Journal of Agricultural Science and Research*, *3*. (2014).
62. Echeverria, R., Trigo, E. J. & Byerlee, D. Financing agricultural research in Latin America. In Tabor, S. R., Janssen, W. & Bruneau, H. (Eds.), *Financing agricultural research: AS ourcebook*. (pp. 323–339). International Service for National Agricultural Research. (1998).
63. Pal, S. & Byerlee, D. India: the funding and organization of agricultural R&D - evolution and emerging policy issues. In Pardey, P. G., Alston, J. M. & Piggott, R. R. (Eds.), *Agricultural r&d in the developing world: Too little, too late?* International Food Policy Research Institute. (2006).
64. National Science and Technology Council Taiwan. *Indicators of Science and Technology*. https://wsts.nstc.gov.tw/stsweb/technology/TechnologyDataIndex.aspx?language=E (2024).
65. Fuglie, K. O. & Piggott, R. R. Indonesia: Coping with Economic and Political Instability. In Pardey, P. G., Alston, J. M. & Piggott, R. R. (Eds.), *Agricultual r&d in the developing world, too little, too late?* International Food Policy Research Institute. (2006).

66. Choi, J.-S., Sumner, D. A. & Lee, H. Korea: Growth, Consolidation and Prospects for Realignment. In Pardey, P. G., Alston, J. M. & Piggott, R. R. (Eds.), *Agricultural r&d in the developing world: Too little, too late?* (pp. 105–128). International Food Policy Research Institute. (2016).
67. Ministere de l'Enseignement Superieur de la Recherche et de l'Innovation. *La recherche dans les administrations par grandes finalite*. http://www.enseignementsup-recherche.gouv.fr/reperes/public/chiffres/france/adm.htm (2022).
68. Thirtle, C., Piesse, J. & Schimmelpfennig, D. Modeling the length and shape of the R&D lag: an application to UK agricultural productivity. *Agricultural Economics*, **39**(1), 73–85. https://doi.org/10.1111/J.1574-0862.2008.00316.X (2008).
69. Morales, C. *National Agricultural Research Systems in Latin America and the Caribbean: Changes and Challenges*. United Nations. (1998).
70. Rutten, H. *Productivity growth of Dutch agriculture, 1949-1989*. Landbouw Economisch Instituut; LEI-DLO. (1992).
71. Poapongsakorn, N. *R&D and Performance of the Thai Agriculture and Food Processing Industry: The Role of Government, Agribusiness Firms, and Farmers* (pp. 401–475). Economic Research Institute for ASEAN & East Asia. (2009).
72. Ratinger, T. & Kristkova, Z. R&D investments, technology spillovers and agricultural productivity, case of the Czech Republic. *Agricultural Economics (Czech Republic)*, **61**(7), 297–313. https://doi.org/10.17221/148/2014-AGRICECON (2015).
73. RICYT. *Red de indicatores de ciencia y technologia*. www.ricyt.org (2024).
74. UNESCO. *Research and Development (R&D) Archive March 2021*. https://uis.unesco.org/bdds (2021).
75. OECD. *Research and Development Statistics (RDS) Data*. https://www.oecd.org/en/data/datasets/research-and-development-statistics.html (2024).
76. Huffman, W. E. & Evenson, R. E. *Science for agriculture: A long-run perspective* (2nd Ed.). Blackwell Publishing. (2006).
77. EUROSTAT. *Science, technology and innovation*. https://ec.europa.eu/eurostat/web/science-technology-innovation/database (2024).
78. Liebenberg, F. *South African Agricultural Production, Productivity and Research Performance in the 20th Century* [PhD thesis]. University of Pretoria. (2013).
79. KISTEP. (n.d.). *Survey of research and development in Korea (various issues)*. Ministry of Science, ICT; Future Planning, Korea Institute of S&T Evaluation; Planning.
80. Mullen, J. Trends in investment in agricultural R&D in Australia and its potential contribution to productivity. *Australasian Agribusiness Review* **18**, 18–29 (2010).
81. World Bank. *Venezuela Agricultural Research, Extension and Education Subsector Review*. World Bank. (1991).
82. World Bank. *World Development Indicators*. https://databank.worldbank.org/source/world-development-indicators (2025).
83. United Nations. *National Accounts - Analysis of Main Aggregates (AMA)*. https://unstats.un.org/unsd/snaama/ (2024).
84. IMF. *World Economic Outlook Database: October 2024 edition*. https://www.imf.org/en/Publications/WEO/weo-database/2024/October (2024).
85. USDA-ERS. *Machine-readable and long format file of TFP indices and components for countries, regions, countries grouped by income level, and the world, 1961–2022*. https://www.ers.usda.gov/data-products/international-agricultural-productivity (2025).
86. Craig, B. J., Pardey, P. G. & Roseboom, J. Internationally Comparable Growth, Development, and Research Measures. In Pardey, P. G., Roseboom, J. & Anderson, J. R. (Eds.), *Agricultural research policy: International quantitative perspectives* (pp. 131–172). Cambridge University Press. (1991).
87. Wang, X., Hyndman, R. J., Li, F. & Kang, Y. Forecast combinations: An over 50-year review. *International Journal of Forecasting* **39**(4), 1518–1547, https://doi.org/10.1016/j.ijforecast.2022.11.005 (2023).
88. Honaker, J. & King, G. What to Do about Missing Values in Time-Series Cross-Section Data. *American Journal of Political Science* **54**(2), 561–581, https://doi.org/10.1111/j.1540-5907.2010.00447.x (2010).
89. Hyndman, R. J. & Athanasopoulos, G. *Forecasting: Principles and Practice* (3rd editio). OTexts. https://otexts.com/fpp3/ (2021).
90. Van Dijk, M., Fuglie, K. O., Heisey, P. W. & Deng, H. *GRAPE database (v.1.0.0)*. https://doi.org/10.5281/zenodo.15507361 (2025).
91. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org/ (2023).

Acknowledgements
This work was supported by the U.S. Department of Agriculture, Economic Research Service. Michiel van Dijk acknowledges funding from the European Union's Horizon Europe Programme through the BrightSpace project (Grant Agreement No 101060075). The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. The authors would like to express their gratitude to all data providers, statisticians and experts that shared data, knowledge and expertise to support the construction of GRAPE.

Author contributions
M.v.D. and K.F. led the research; M.v.D., P.H. and K.F. proposed the generation of GRAPE and collected, harmonized and processed the raw data, partly building on earlier data collection efforts. H.D. collected and processed additional information for China. M.v.D. and K.F. curated the data and performed technical validation; M.v.D. prepared code to harmonize, process and combine the data; M.v.D. wrote the manuscript and dataset documentation with inputs from H.D., K.F. and P.H.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-05331-y.

Correspondence and requests for materials should be addressed to M.v.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.