Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

Continuous national Gross Domestic Product (GDP) time series for 195 countries: past observations (1850-2005) harmonized with future projections according to the Shared Socio-economic Pathways (2006-2100)

Tobias Geiger

Potsdam Institute for Climate Impacts Research, Telegraphenberg A 56, 14473 Potsdam, Germany

Correspondence to: Tobias Geiger (geiger@pik-potsdam.de)

Abstract. Gross Domestic Product (GDP) represents a widely used metric to compare economic development across time and space. GDP estimates have been routinely assembled only since the beginning of the second half of the 20th century, making comparisons with prior periods cumbersome or even impossible. In recent years various efforts have been put forward to re-estimate national GDP for specific years in the past centuries and even millennia, providing new insights of past economic development on a snapshot basis. In order to make this wealth of data utilizable across research disciplines, we here present a first continuous and consistent data set of GDP time series for 195 countries from 1850 to 2009, based mainly on data from the Maddison Project and other population and GDP sources. The GDP data is consistent with Penn World Tables v8.1 and future GDP projections from the Shared Socioeconomic Pathways (SSPs), and freely available at http://doi.org/10.5880/pik.2017.003. To ease usability, we additionally provide GPD per capita data and further supplementary and data description files in the online archive. We utilize various methods to handle missing data and discuss the advantages and limitations of our methodology. Despite known shortcomings this data set provides valuable input e.g. for climate impact research in order to consistently analyze economic impacts from pre-industrial times to the future.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

1 Introduction

The concept of measuring and comparing economic activity within and across countries using the Gross Domestic Product (GDP) is rather new in historic terms. Starting with some first attempts to quantify economic activity in the late 19th century for certain countries, comprehensive regular assessments of GDP have only been established in the second half of the 20th century. Since then GDP has become the standard indicator to assess nations' development, despite initial and more recent criticism concerning the incomplete representation of a nation's state via GDP and the potential problematic comparison across countries [*Speich Chasse*, 2013].

Nonetheless and because of lack of alternatives has GDP proven to be a useful measure to track the evolution of economic development within or across nations. Many other development proxies, e.g. the level of education, life expectancy, the population's health status, and many others, have been shown to correlate well with a nation's GDP, see e.g. [*Gennaioli and Porta*, 2013]. Similarly, a reduction in vulnerability (or an increase in resilience) to natural disasters has also been shown to correlate well with a nation's GDP [*Kousky*, 2013], resulting in less mortality and in fewer damages relative to GDP. Most research in this field focuses only on the last decades where sufficient coverage of economic activity exists for most countries of the world. However, many fields of research could benefit from more comprehensive economic data set that covers a larger time horizon and a larger number of nations to gain a better understanding of the drivers of long-term economic development.

Various global institutions (e.g. Worldbank, OECD, IMF) and research groups (e.g. Penn World Table) have therefore assembled global GDP data sets, most of which provide a comprehensive view across space but lack data prior to the 1960ies. In addition, there have been attempts to provide GDP or income (i.e. GDP per capita) estimates reaching further back in time based on proxies for specific periods, see e.g. [*Baier et al.*, 2002; *Mitchel*, 2003; *Maddison*, 2007; *Bolt and van Zanden*, 2014]. However, these estimates only provide snapshots of economic activities for specific periods without continuous global coverage across time.

Here, we contribute to increasing the usability of historical economic time series by creating a continuous and complete income and GDP time series from 1850 to the present. We do so by combination of various data sources and methods to interpolate and extrapolate missing data points in a pragmatic but most sensible way. The Maddison Project data base thereby constitutes the foundation of the period mostly before 1960 while Penn World Table (version 8.1) sets the basis for the more recent past. Our final GDP time series covers 195 countries (in their present constitution) from 1850 to 2009 and is consistent with GDP projections from the Shared Socioeconomic Pathways (SSPs) that extend the historical time series from 2010 to 2100.

The long record and complete coverage enhance the data set's usability. It is has already been assigned as input data for the current climate change impact model runs within the global Inter-sectoral Impact Model Inter-comparison Project (ISIMIP2b, www.isimip.org) [*Frieler et al.*, 2016], and has been used in a downscaling approach to provide spatially-explicit economic information on the grid level [*Murakami, D. and Yamagata*, 2017] that can e.g. be used to quantify economic

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

values exposed to climate extremes [*Geiger et al.*, 2017]. Despite various known shortcomings, that are discussed in detail below, this new data set has and will further broaden the applicability of historic estimates of economic activity and potentially feedback to foster a increased research interest in the field of economic history and the improvement of the current data set.

2 Data & Methods

In the following we present our methodology that is used to create a continuous and consistent GDP time series for 195 countries based on national accounts data from the Penn World Table (PWT), the Maddison project, World Development Indicators (WDI), the History Database of the Global Environment (HYDE), and projections from the Shared Socio-economic Pathways (SSPs). We will present the data sources first before we describe the consistent merging procedure across all sources.

2.1 Gross Domestic Product (GDP)

2.1.1 Penn World Tables (PWT)

The PWT comprises national accounts data maintained by scholars at the University of California and the University of Groningen to measure real GDP across countries and over time. The data base is successively updated and extended, with the latest release being PWT 9.0 [*Feenstra et al.*, 2015], and provides the most extensive coverage for GDP reported in Purchasing Power Parity (PPP) across time. We here mostly rely on the PWT release 8.1 from 2015 for two reasons: First, PWT8.1 data is reported in 2005 PPP $ and thus being consistent with SSP projections. Second, PWT8.1 is in close agreement with the SSP initial data in 2010, thus reducing matching artefacts to a minimum. Moreover, PWT8.1 replaces the strongly criticized original PPPs for 2005, see e.g. [*Deaton and Heston*, 2010], by a modified version, see [*Inklaar and Rao*, 2017] for details. Missing countries in PWT8.1 are taken from PWT9.0 after rescaling from 2011 PPP $ to 2005 PPP $ dollars, see discussion below. PWT also provides national population estimates that we apply to generate income estimates based on national GDP.

2.1.2 World Development Indicators (WDI)

The WDI assembled by the Worldbank provide a vast resource of socio-economic data. Their present release of PPP-based GDP comes in 2011 PPP $ and is available from 1990 to 2015 [*Worldbank*, 2017]. As income estimates in 2005 PPP $ values are no longer available, we rescale from 2011 PPP $ to 2005 PPP $ to insert otherwise missing countries in the PWT data, see discussion below. WDI also provides national population estimates that we apply to generate income estimates based on national GDP.

2.1.3 Shared Socio-economic Pathways (SSPs)

The SSPs are story lines of plausible alternative evolutions of society at the global level that can be combined with assumptions about climate change and policy responses to evaluate climate change impacts, adaptation, and mitigation [*O'Neill et al.*, 2017]. Meanwhile different Integrated Assessment Models (IAMs) have generated GDP projections along the SSP story lines. The associated national time series all start in 2010. The historical data set we provide is designed to allow for a smooth transition from historical time series to the associated future projections. As such it can e.g. be used for transient climate impact model simulations covering the historical period and future projections under different socio-economic development pathways as described by the SSPs.

SSP projections are generally available at https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=about. We here use the OECD GDP data set [*Dellink et al.*, 2017] provided by the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), freely available at https://www.isimip.org/gettingstarted/availability-input-data-isimip2b/ upon registration. This data set was directly provided by the OECD and is advantageous as it contains data on seven additional countries.

2.1.4 Maddison Project

The Maddison project resembles a cooperative effort to assemble historical national accounts [*Bolt and van Zanden*, 2014], continuing the ground-breaking work by the late Angus Maddison. The Maddison project data set (MPD) is a freely available excel spread sheet [*Maddison*, 2016] that provides per capita income (in 1990 Geary-Khamis (G-K) dollars) for 184 countries and/or world regions between 1AD and 2010 in varying time intervals. Since 1800 data is provided annually, however, containing large fractions of missing values, in particular for the African continent, Western Asia and the former Soviet Union. The global missing value fraction is 61.8% between 1800 and 2010. It steadily decreases to 51.7% (36.9%) when analyzing the data from 1850 (1900) onwards. After 1950, when national accounts data was starting to be routinely collected, the missing value fraction drops to 8%. The data is nearly complete since 1990 (missing value fraction: 3.6%). See Fig. 1 and supplementary data available at http://doi.org/10.5880/pik.2017.003.

We selected 1850 as the start year for our present work for two reasons: first, coverage is somehow better than in 1800 and second, we can rely on available data points between 1800-1849 for the purpose of interpolation rather than extrapolation.

2.2 Population data

Part of the GDP data sets (PWT and WDI) also provide national population data to derive income. Whenever available we use the associated population estimates from the same source to derive income from GDP. However, to estimate national GDP from the income data generated within the MPD we use population estimates from the HYDE data set.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

2.2.1 History Database of the Global Environment (HYDE)

HYDE is developed under the authority of the Netherlands Environmental Assessment Agency and provides (gridded) time series of population and land use for the last 12,000 years [*Klein Goldewijk et al.*, 2010, 2011]. HYDE provides national population data decennially up to 2000 and annually up to 2015. Where required we linearly interpolate the data to derive annual distributions.

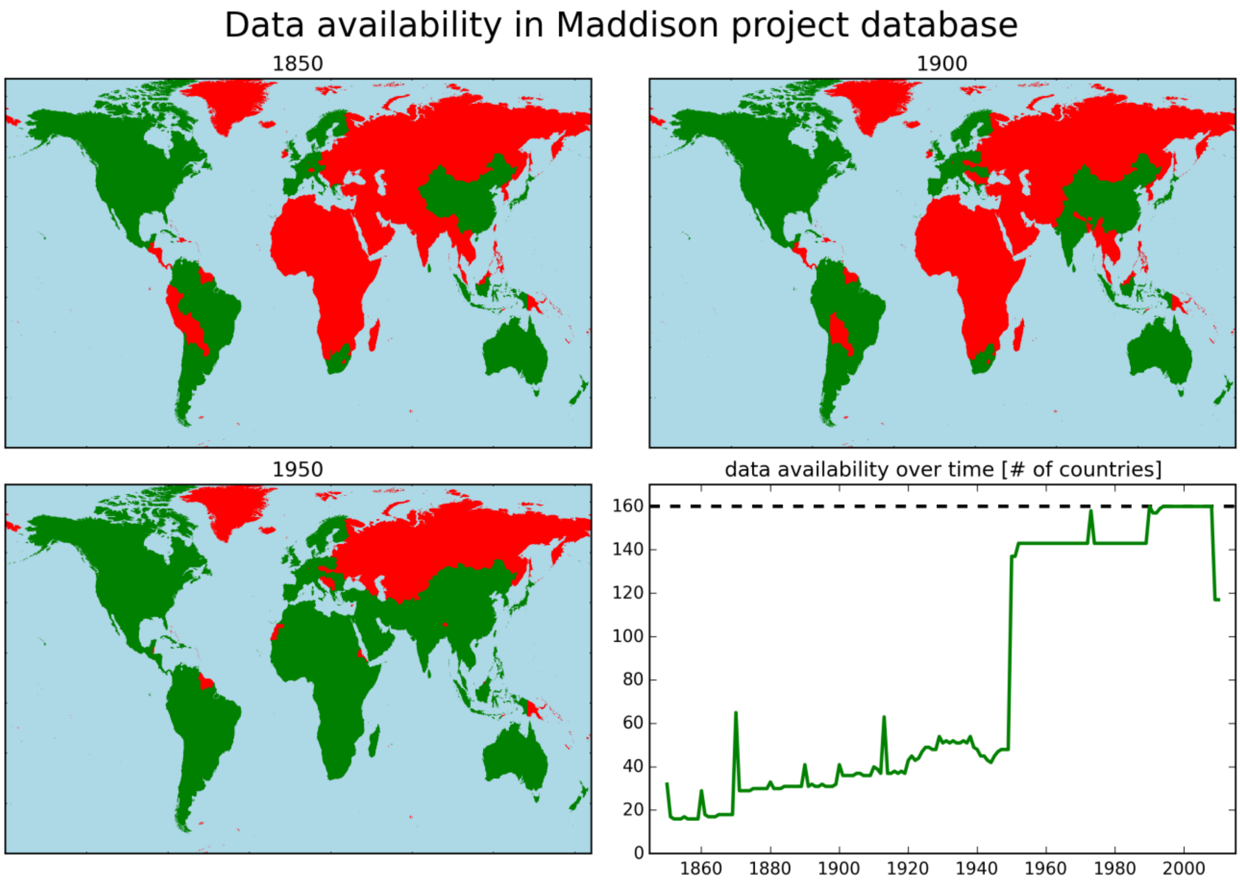

Figure 1: Illustration of data availability in the Maddison project data base over time. Maps show geographical distribution of available data points for three selected years (1850, 1900, 1950), while the plot displays the actual number of available countries over the full period. Countries are displayed using their current borders. We only show data for those 160 countries that are explicitly named in the Maddison project data base; country group estimates lacking national resolution are not shown (see Table 1 for details).

2.3 Completing the Maddison Project Data

MDP provides income data on national and supranational level, where the supranational data represents population-weighted averages of national values (see Table 1). In addition, there exist three world region-specific groups of small countries (group of 14 small European countries, 21 Caribbean countries, 24 South-East Asian countries) which lack national

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

resolution, i.e. member countries are prescribed identical growth paths (see Table 1). In the following, each of the three country groups is treated like an individual country with respect to replacing of missing data.

Generally, gap filling is first done on the supranational level and later gaps in national time series are filled by accounting for growth rates of either neighboring countries, the associated supranational level, or the associated world region the country belongs to.

In the following we present our gap-filling methodology, first general steps and then for each world region in detail, starting with regions with least missing values.

World region	Supranational group 1	Supranational group 2	Small country groups
Europe	12 Western European (WE)	7 Eastern European (EE)	14 small European countries
Latin America	8 Latin American (LA)	-	21 Caribbean countries
Asia	16 Eastern Asian (EA)	Western Asian (WA)	24 South-East Asian countries
Africa	African total (AT)	-	-

Table 1: Overview of the world region-specific supranational country groups and the small country groups, that lack national resolution, as used within the Maddison project data base.

2.3.1 Preparatory steps

As a first step we populate all missing data points in 1850, the initial year of our data product, by linear interpolation between the last available data point before 1850 and the first one after 1850, ensuring that it is not more distant in time than 1870. Next and if available, we generate annual data by linear interpolation of data points between 1850, 1860, and 1870. These preparatory steps reduce the missing value fraction from 51.7% to 48.5%.

2.3.2 Europe

The preparatory steps completed the country-level data for most countries in Western Europe. These individual country-level data is then used to complete the time series for the supranational group of 12 Western European (WE) countries by population-weighted mean income using HYDE's national population data, as growth rates from this time series are used to approximate missing values for the country group of small 14 European countries. Similarly, United Kingdom's growth path is used to complete Ireland's time series.

Next, gaps in the supranational group of 7 Eastern Europe (EE) countries time series are linearly interpolated (including the time of World War 2) and then, starting in 1870, used to extrapolate individual country growth paths in Eastern Europe back to 1850.

Relative changes in the former Yugoslavia time series are used to extrapolate their temporary constituents back to 1850, and to extrapolate Kosovo's income between 1991 and 2010. The same procedure is conducted for the constituents of former Czechoslovakia.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

Relative EE changes prior to 1885 are used to complete the former USSR time series. The relative changes of the USSR time series are used to extrapolate income for all member countries of the USSR up to 1973, and to interpolate between 1973 and 1990. The fraction of missing values in the entire data set thus reduces to 34.4%.

2.3.3 North America, Australia, and New Zealand

The preparatory steps were sufficient to fill the respective time series completely.

2.3.4 Latin America

Following the country grouping in the MPD, we generate a complete population-weighted average income time series for a group of 8 large Latin American (LA) countries. To do so, Peru's and Mexico's income is respectively extrapolated before 1870 and interpolated after 1870 based on the mean growth of the 7 remaining countries. Relative changes in the LA time series are applied to estimate the income of the remaining South American countries. The time series for the group of 21 Caribbean countries is linearly interpolated and then used to fill gaps in the remaining Caribbean and Central American countries, except for Jamaica and Cuba, whose time series is sufficiently dense to be interpolated individually.
Finally, the fraction of missing values is reduced to 29.7%.

2.3.5 Asia

Asia is separated into East and West Asia, reduced by the fraction that belonged to the former USSR.
Where required relative income changes based on the linearly interpolated time series for a group of 16 Eastern Asian (EA) countries are used to extrapolate individual country data before 1870. Furthermore, data gaps for selected countries with quite dense coverage (India, Japan, Indonesia (Java before 1880), Sri Lanka, and China) are filled by linear interpolation on an individual basis. Using a step-wise procedure, we fill the EA time series after 1870 with population-weighted average income of all countries for which original data is available at any given time step. Relative changes of the EA time series are then used to fill gaps in all remaining East and South-East Asian countries, in particular to complete the time series for the joint group of 24 South-East Asian countries. However, missing values for Bangladesh and Pakistan are estimated based on India's relative changes. Thereby, the fraction of missing values is reduced to 24.3%.
In a second step, Western Asian countries including the Gulf States are treated, where data availability is very limited before 1950. Except for Turkey (former Ottoman Empire, complete data since 1923), relevant data point exist only for 1820, 1870, and 1913. To also include countries with completely missing data before 1950, we assume that the income of Israel is equal to West Bank/Gaza's income in 1870 and 1913, and that the income of some Gulf States is equal to Iran's income. For each country we linearly interpolate the data until 1913, and again until 1950, except for Kuwait, Qatar, and UAE. For those three countries we estimate income growth to 1939 based on average income growth for all Western Asian (WA) countries and then linearly interpolate each country individually between 1940 and 1950, thereby ensuring that sharp income rises only occur after the discovery of oil.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

Upon completing Asian time series the fraction of missing values is reduced to 19.5%.

2.3.6 Africa

The MPD contains only 6 countries with income data prior to 1950: Egypt, Tunisia, Morocco, Algeria, and South Africa/Cape Colony (all since 1820), and Ghana (since 1870). Therefore, the African total (AT) population-weighted average income prior to 1950 is only defined by 6 countries. Due to historic and geographic reasons, we assume that those countries define the upper income limit when extrapolating the remaining countries back in time. So if a country's 1950 income is smaller than the 6 country population-weighted mean, the income fraction is used to define the scaled income in 1913; otherwise, it is set to equal the 6 country mean in 1913. Missing data between 1913 and 1950 is then interpolated according to scaled AT relative changes, while the bare AT relative changes are used to extrapolate missing values from 1913 to 1850.

2.3.7 Summary

Using all steps described above all missing values are filled and a complete time series (1850-2010) is obtained for all countries listed within the MPD. This time series, reported in 1990 G-K dollars and matched with IS03 country codes, can be found at http://doi.org/10.5880/pik.2017.003. Note that for some countries the reporting period ends in 2008; no extrapolation was done thereafter as more reliable data from other sources is available for this period.

2.4 Matching and time series conversion

Our final data base is reported in PPP dollars referenced to 2005, the original unit of PWT8.1 and the SSPs. Consequently, data from the MPD, the WDI, and PWT9.0 requires conversion. As for MPD no official conversion factors are available, time series of historic income from the MPD are scaled to match PWT8.1 income data at the earliest reporting year of PWT8.1 data. To test the robustness of the conversion factors they are calculated not only for the earliest reporting year of PWT8.1 but for each of the first five years. If the five individual conversion factors significantly vary (fraction of standard deviation of the first five conversion factors and the first conversion factor larger than the iteratively derived threshold of 4 %), we use the five-year mean conversion factor instead. All conversion factors and the base year of matching are available at http://doi.org/10.5880/pik.2017.003. Figure 2 provides an illustration of the conversion factors, here shown as conversion rates in per-cent to highlight the symmetrical scattering around the cross-country mean, which is larger than unity, as expected, due to the transformation forward in time between 1990 and 2005. Most countries' conversion factors scatter closely around the median (Fig. 2b) while BRN (Brunei) and SOM (Somalia) are clear outliers, introducing rather large uncertainty in their respective converted time series before 1970 and 2010. Time series for countries with no individual resolution in the MPD are matched to PWT8.1 data using the respective country group data, see Table 1.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

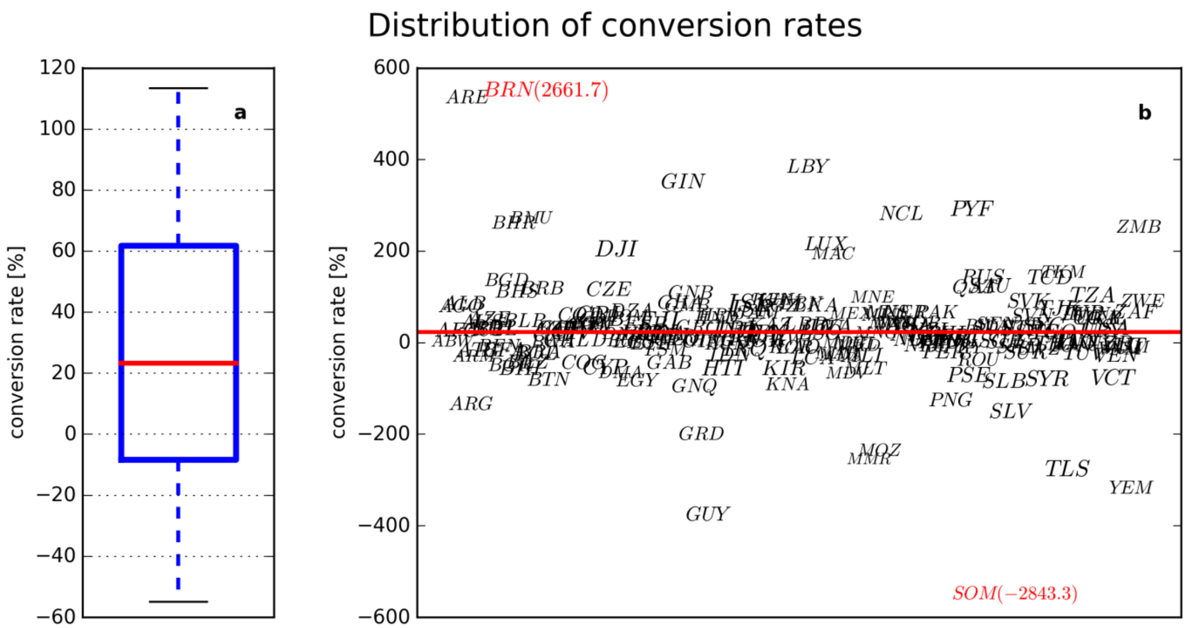

Figure 2: Statistical (a) and country-specific (b, using ISO3 country codes) distribution of conversion rates used to convert GK 1990 $ to PPP 2005 $. Most countries are close to the cross-country median conversion rate (+23.3%, red line in both panels) while SOM (Somalia) and BRN (Brunei) (shown in red in b) are clear outliers with respective conversion rates shown in parentheses. Whiskers in (a) display the 10% to 90% percentile range.

Additional countries that are reported in the SSPs but not in PWT8.1 are included based on data first from PWT9.0 and second from WDI, except for El Salvador and Zimbabwe where PWT9.0 data is chosen in place of available PWT8.1 due to known issues in the data set, see http://www.rug.nl/ggdc/docs/what_is_new_in_pwt_81.pdf for details.

Conversion of PWT9.0 data between 2011 PPP $ and 2005 PPP $ values is conducted by using the PWT-provided exchange rates (PWT abbreviation: xr) and price levels of GDP (PWT abbreviation: pl_gdpo). For WDI data we use the provided PPP conversion table [*Worldbank*, 2017] and corrections for inflation based on the USA GDP deflator [*Worldbank*, 2017]. North Korea (ISO3: PRK) is only available in MPD and therefore excluded from the analysis.

In order to be consistent with SSP projections starting in 2010, we truncate the observation-based time series in 2005 and linearly interpolate from 2006 to 2009. An illustrative example of the matching procedure for four selected countries is displayed in Fig. 3.

In a next step we multiply income and population time series to obtain GDP time series. As for the income matching between historical observations and SSP projections, population estimates are truncated in 2005 and linearly interpolated to match SSP population projections in 2010. For almost all countries the initial value of the SSP time series in 2010 and actually observed GDP values in 2010 only slightly deviate, see Fig. 3. Even though the deviation illustrated in Fig. 3 corresponds to income deviations only, it reflects also GDP differences well as the SSP income projections are the main

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

source of deviations, mostly due to the fact that SSP simulations were completed before the financial crisis in 2008. However, two countries show larger discrepancies that are caused by difference both in income and population estimates. The difference in 2010 GDP from the reconstructed time series and the associated SSP value in 2010 for Zimbabwe (ZWE) and Aruba (ABW) is about 70% and 600%, respectively. When using these two time series for historical analysis only, we recommend truncation of the time series in 2005. Furthermore, for following countries (Antigua and Barbuda (ATG), Bermuda (BMU), Dominica (DMA), Federated States of Micronesia (FSM), Grenada (GRD), Kiribati (KIR), Saint Kitts and Nevis (KNA), Marshall Islands (MHL), Nauru (NRU), Seychelles (SYC), Tuvalu (TUV)) no GDP projections exist in the OECD data base. We therefore used the original data from our reconstruction up to 2009.

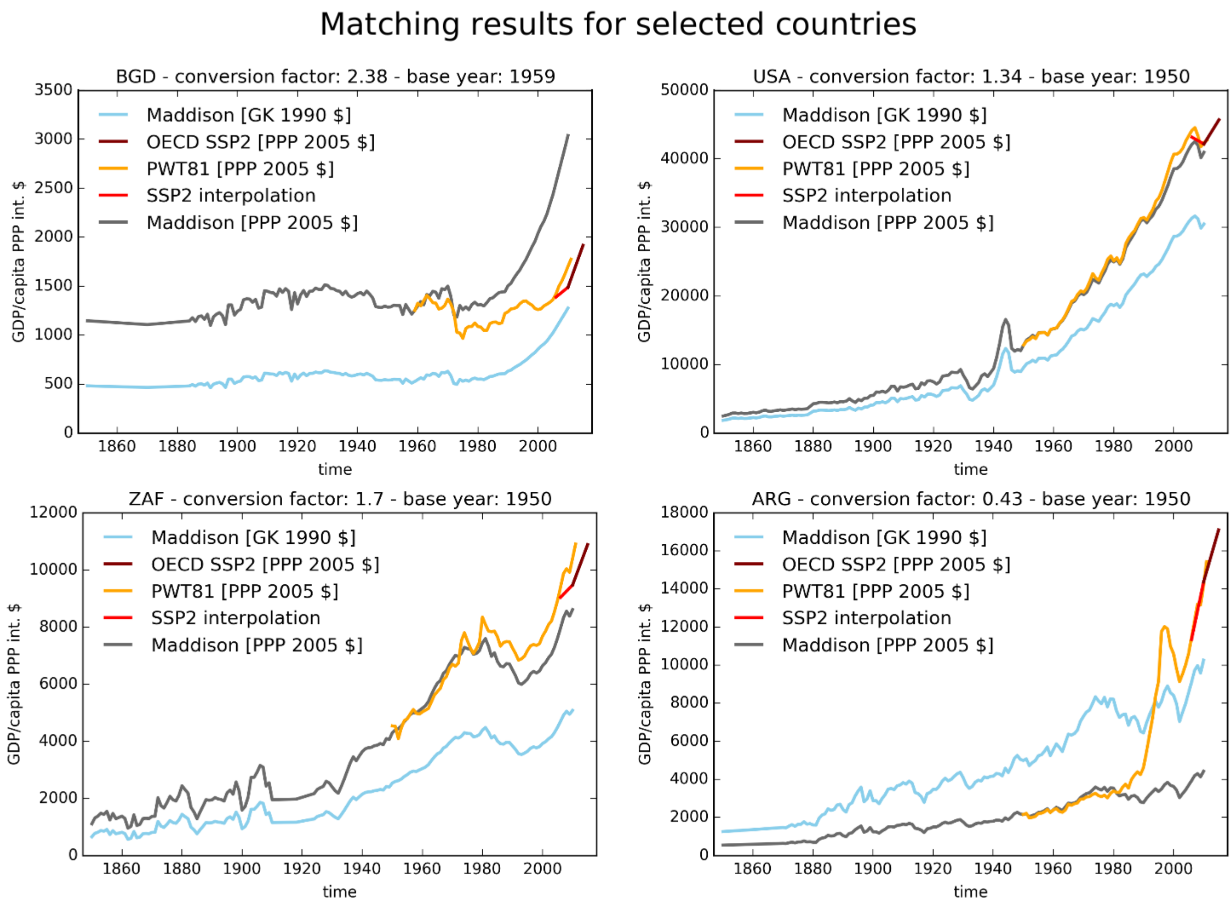

Figure 3: Results of matched GDP per capita time series for Maddison, PWT, and OECD SSP2 data for selected countries, ARG: Argentina; BDG: Bangladesh; USA: United States of America; ZAF: South Africa. Original Maddison data (blue, in 1990 GK $) is matched to PWT data (yellow, in 2005 PPP $) using a country-specific base year and conversion factor, to obtain converted Maddison data (gray, in 2005 PPP $). Note that for ZAF an overlap of 5 years is used to determine the mean conversion factor. Between 2006 and 2009, PWT data is interpolated (red) to match OECD SSP2 projections (maroon) starting in 2010.

For some small countries HYDE data has missing population values prior to certain years (Bermuda (BMU), Macao (MAC), Maldives (MDV) < 1970; Federated States of Micronesia (FSM), Kiribati (KIR), Marshall Islands (MHL), Nauru (NRU), French Polynesia (PYF), Seychelles (SYC), Tuvalu (TUV) < 1960). Therefore, the corresponding GDP time series contains missing values for this period.

3 Overview of the data set and data availability

We provide three different primary data sets, a data description file, and two supplementary data sets in the online archive at http://doi.org/10.5880/pik.2017.003:

1) A continuous table of global income data (in original 1990 G-K $) based on the MPD for 160 individual countries and 3 groups of countries from 1850 to 2010.

2) A continuous and consistent table of global income data (in 2005 PPP $) for 195 countries based on the merged MPD and PWT8.1 data and extended using PWT9.0 and WDI data from 1850 to 2009, and consistent with OECD SSP2 income projections starting in 2010.

3) A continuous and consistent table of global GDP data (in 2005 PPP $) for 195 countries based on the merged MPD and PWT8.1 data, extended using PWT9.1 and WDI data, and consistent with OECD SSP2 GDP projections starting in 2010.

The second data set is illustrated in Fig. 4, showing the population-weighted income distribution across continents and percentiles.

Furthermore, we provide two supplementary data sets:

1) A mask table complementing the first primary data set indicating which of the data points are original values and which are estimated based on our present methodology, and used to generate Fig. 1.

2) A table of conversion factors (1990 GK $ to 2005 PPP $) also indicating the data source used for matching and the methodology used for conversion, and used to generate Fig. 2.

All data sets are provided in "csv" format and are freely available at http://doi.org/10.5880/pik.2017.003. Please also consult the data description file for additional information on the data set.

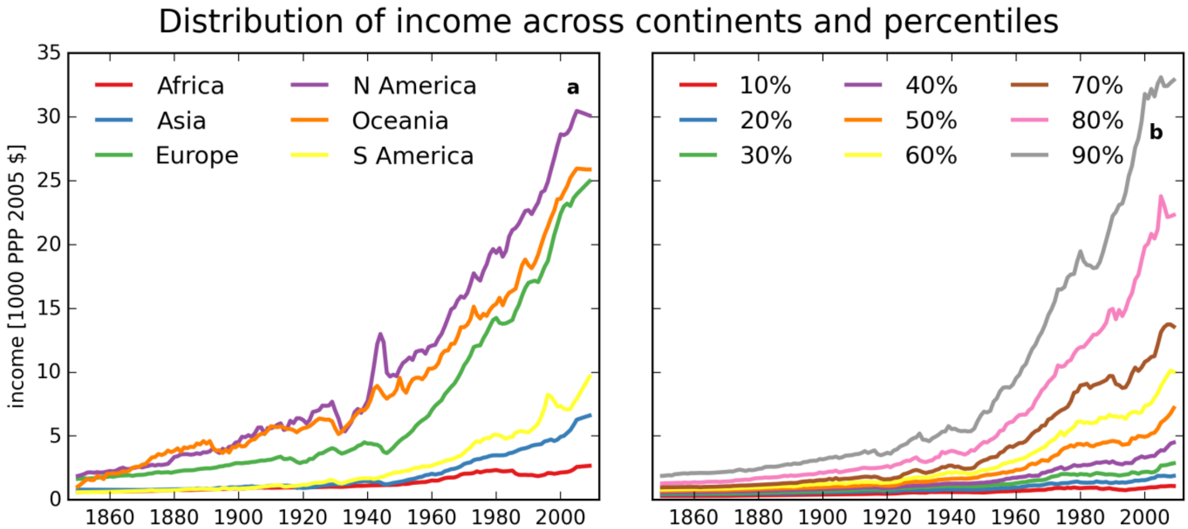

Figure 4: Distribution of population-weighted income over time across continents (a) and percentiles (b) that can be analysed for e.g. income inequality or other distributional effects. Percentiles rank the number of countries by average national income.

4 Discussion & Conclusion

In conclusion, we provide a continuous time series for per capita income and GDP between 1850 and 2009 for 195 countries. Foundation of our work is the income data set maintained by the Maddison project, that is completed using various interpolation and extrapolation techniques and harmonized with economic data from the Penn World Tables. The main objective is to provide a continuous economic time series that is readily applicable across disciplines and by non-experts in the field. The methodology applied is rather simple and comes with several limitations and caveats.

As we do not devise new historic economic figures ourselves, we are bound to existing data only that is combined with various techniques to replace missing values. Our main assumption is that geographically close countries have observed similar growth paths over the last 150-200 years, such that we can use neighboring (groups of) countries to estimate missing data. This assumption strongly depends on many political, economical, and societal aspects (e.g. the economic system, the membership in alliances, the occurrence of wars, the colonial background, and many more) that, however, where not considered in our work. While rather exhaustive data exists for Western European countries, these limitations might be less of a problem than for most African countries. As a consequence, one should treat the data with care and allow for uncertainties, in particular were data coverage is limited or almost non-existing.

Another limitation arises due to different units: The original MPD is measured in 1990 GK $ while all other data sets use PPP equivalents either in 2005 or 2011 $. The required PPP-transformation can lead to under- or overestimation of economic figures, in particular further back in time. As mentioned above for Brunei (discovery of fossil fuels) and Somalia (ethnic conflict and civil war), large conversion factors are due to rapid changes in a country's income over a short period of time that can deteriorate the PPP conversion. However, as currently there exists no reliable solution to circumvent this conversion

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

problem one has to interpret transformed data with caution. For example, countries that mainly depend on fossil-fuel exports had rapid income jumps in the not so distant past that can overestimate their income before oil discovery. For this reason we also provide the interpolated MPD in original 1990 GK $.

Another problem arises due to shifting national borders, and the formation of new and disappearance of old nations: the data sets only reflect the current political map. To circumvent this problem, we also provide per capita income time series that can be used in combination with historical population data to generate additional GDP estimates. The provided GDP time series is generated by multiplying national population and income data. For consistency reasons population and income estimates are selected from identical sources, except for the MPD where HYDE data is used. However, when merging different time series inconsistencies arise due to different country definitions, e.g. as is the case for countries in former Yugoslavia or for the transition between historical data and SSP projections. Existing discrepancies were harmonized and an interval between 2005 and 2010 was used to allow for a smooth transition to SSP projections.

Furthermore and as mentioned above, historical population estimates are unavailable for some small countries making the GDP time series incomplete.

Despite these shortcomings and uncertainties, this new data set will broaden the applicability of historic estimates of economic activity, e.g. in the field of climate impact research in order to facilitate impact simulations on centennial time scales [*Frieler et al.*, 2016]. It further provides the opportunity to generate gridded GDP distributions for the past based on recent downscaling initiatives [*Murakami, D. and Yamagata*, 2017]. Moreover, the increased research interest in past GDP data will provide valuable feedback to the historians and economists working in this field and might stimulate further advances. These advances are expected to improve this current data set further.

5 Author contribution

TG created and analyzed the data set and wrote the paper.

6 Competing Interests

The author declares that he has no conflict of interest.

7 Acknowledgements

We wish to thank Johannes Gütschow for his valuable comments on the generation of this data set and Katja Frieler for her helpful remarks to improve the manuscript. We also thank Jutta Bolt for advice on the Maddison project data base.

This work was funded through the framework of the Leibniz Competition (SAW-2013-PIK-5 and SAW-2016-PIK-1).

8 References

Baier, S. L., G. P. Dwyer, and R. Tamura (2002), How Important Are Capital and Total Factor Productivity for Economic Growth?, *SSRN Electron. J.*, doi:10.2139/ssrn.301213.

Bolt, J., and J. L. van Zanden (2014), The Maddison Project: collaborative research on historical national accounts, *Econ. Hist. Rev.*, *67*(3), n/a–n/a, doi:10.1111/1468-0289.12032.

Deaton, A., and A. Heston (2010), Understanding PPPs and PPP-based National Accounts, *Am. Econ. J. Macroecon.*, *2*(4), 1–35, doi:10.1257/mac.2.4.1.

Dellink, R., J. Chateau, E. Lanzi, and B. Magné (2017), Long-term economic growth projections in the Shared Socioeconomic Pathways, *Glob. Environ. Chang.*, *42*, 200–214, doi:10.1016/j.gloenvcha.2015.06.004.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015), The Next Generation of the Penn World Table, *Am. Econ. Rev.*, *105*(10), 3150–3182, available for download at www.ggdc.net/, doi:10.1257/aer.20130954.

Frieler, K. et al. (2016), Assessing the impacts of 1.5°C global warming - simulation protocol of the Inter-Sectoral Model Intercomparison Project (ISIMIP2b), *Geosci. Model Dev.*, (October), submitted, doi:10.5194/gmd-2016-229.

Geiger, T., K. Frieler, and D. N. Bresch (2017), A global data set of tropical cyclone exposure (TCE-DAT), Available from: http://doi.org/10.5880/pik.2017.005

Gennaioli, N., and R. La Porta (2013), Human capital and regional development, *Q. J. …*, *128*(1), 105–164, doi:10.1093/qje/qjs050.Advance.

Inklaar, R., and D. S. P. Rao (2017), Cross-Country Income Levels over Time: Did the Developing World Suddenly Become Much Richer?, *Am. Econ. J. Macroecon.*, *9*(1), 265–290, doi:10.1257/mac.20150155.

Klein Goldewijk, K., A. Beusen, and P. Janssen (2010), Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1, *The Holocene*, *20*(4), 565–573, doi:10.1177/0959683609356587.

Klein Goldewijk, K., A. Beusen, G. Van Drecht, and M. De Vos (2011), The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years, *Glob. Ecol. Biogeogr.*, *20*(1), 73–86, doi:10.1111/j.1466-8238.2010.00587.x.

Kousky, C. (2013), Informing climate adaptation: A review of the economic costs of natural disasters, *Energy Econ.*, doi:10.1016/j.eneco.2013.09.029.

Maddison (2016), Maddison Project, Available from: http://www.ggdc.net/maddison/maddison-project/data.htm (Accessed 25 October 2016)

Maddison, A. (2007), *Contours of the world economy, 1-2030 AD : essays in macro-economic history*, Oxford University Press.

Mitchel, B. R. (2003), *International Historical Statistics*, Palgrave Macmillan UK, London.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-80
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 24 August 2017

Murakami, D. and Yamagata, Y. (2017), Estimation of gridded population and GDP scenarios with spatially explicit statistical downscaling., *Environ. Res. Lett.*, (under review), available at https://arxiv.org/abs/1610.09041.

O'Neill, B. C. et al. (2017), The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century, *Glob. Environ. Chang.*, *42*, 169–180, doi:10.1016/j.gloenvcha.2015.01.004.

Speich Chasse, D. (2013), *Die Erfindung des Bruttosozialprodukts : globale Ungleichheit in der Wissensgeschichte der Økonomie*, Vandenhock & Ruprecht.

Worldbank (2017), World Development Indicators, Available from: http://data.worldbank.org/ (Accessed 11 January 2017)