Global Inventory of Gas Geochemistry Data from Fossil Fuel, Microbial and Biomass Burning Sources, Version 2017

Owen A. Sherwood¹, Stefan Schwietzke²³, Victoria A. Arling³, Giuseppe Etiope⁴

¹ Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado, USA.
² Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.
³ NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, Colorado, USA.
⁴ Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy, and Faculty of Environmental Science and Engineering, Babes-Bolyai University, Cluj-Napoca, Romania
ABSTRACT

The concentration of atmospheric methane (CH$_4$) has more than doubled over the industrial era. To help constrain global and regional CH$_4$ budgets, inverse (top-down) models incorporate data on the concentration and stable carbon ($\delta^{13}$C) and hydrogen ($\delta^2$H) isotopic ratios of atmospheric CH$_4$. These models depend on accurate $\delta^{13}$C and $\delta^2$H end-member source signatures for each the main emissions categories. Compared with meticulous measurement and calibration of isotopic CH$_4$ in the atmosphere, there has been relatively little effort to characterize globally representative isotopic source signatures, particularly for fossil fuel sources, since the 1980s. Most global CH$_4$ budget models have so far relied on outdated source signature values derived from globally non-representative data. To correct this deficiency, we present a comprehensive, globally representative end-member database of the $\delta^{13}$C and $\delta^2$H of CH$_4$ from fossil fuel (conventional natural gas, shale gas and coal), modern microbial (wetlands, rice paddies, ruminants, termites, and landfills/waste) and biomass burning sources. Alkane and permanent gas molecular chemistry for fossil fuel categories are also included with the database. The database comprises 10,706 samples (8,734 fossil fuel, 1972 non-fossil) from 190 published references. Mean (unweighted) $\delta^{13}$C signatures for fossil fuel CH$_4$ are significantly lighter than values commonly used in CH$_4$ budget models, thus highlighting potential under-estimation of fossil fuel CH$_4$ emissions in previous CH$_4$ budget models. This living database will be updated every 2-3 years to provide the atmospheric modeling community with the most complete CH$_4$ source signature data possible. Database Digital Object Identifier (DOI):

https://doi.org/10.15138/G3201T.
1. INTRODUCTION:

Methane (CH$_4$) is a potent greenhouse gas that accounts for approximately 20% (0.48 W m$^{-2}$) of anthropogenic greenhouse gas radiative forcing in the lower atmosphere (Claas et al. 2013). Atmospheric CH$_4$ levels have more than doubled over the industrial era, increasing from about 700 ppb in the year 1750 to >1800 ppb today (Etheridge et al., 1992). Atmospheric CH$_4$ stabilized from 2000 to 2006 and increased again after 2007 (Nisbet et al., 2014, Dlugokencky et al., 2011). Specific contributions of natural and anthropogenic sources of CH$_4$ to this renewed increase, and to the global CH$_4$ budget in general, remain unclear (Kirschke et al. 2013; Saunois et al. 2016). Wetlands and agriculture have been suggested as dominant sources of renewed increases in CH$_4$ emissions (Dlugokencky et al., 2009; 2011; Bousquet et al. 2011; Bloom et al. 2010; Nisbet et al. 2014; Patra et al. 2016; Schaefer et al. 2016). The recent surge in unconventional oil and gas development in North America and growing awareness of CH$_4$ emissions from oil and gas infrastructure (Howarth et al. 2011; Karion et al. 2013; Brandt et al. 2014) informs alternative explanations for the increase in atmospheric CH$_4$ (Hausmann et al. 2016; Helmig et al. 2016; Rice et al. 2016). Increasing emissions of coal-related CH$_4$, particularly from China, is yet another possible source of increasing CH$_4$ emissions (Bergamaschi et al. 2013; Nisbet et al. 2014).

Current global CH$_4$ emission estimates rely, in part, on inverse (top-down) models that incorporate data on the concentration and stable carbon ($\delta^{13}$C) and hydrogen ($\delta^2$H) isotopic ratios of atmospheric CH$_4$ (e.g., Quay et al. 1991; Lowe et al. 1994; Bousquet et al. 2006; Whiticar and Schaefer, 2007; Neef et al. 2010; Monteil et al. 2011; Rice et al. 2016; Schwietzke et al. 2014a; Schwietzke et al. 2016; Schaefer et al. 2016). Ethane (C$_2$H$_6$) to CH$_4$ mixing ratios have also been used as an additional constraint on fossil fuel CH$_4$ emissions (Simpson et al. 2012; Schwietzke et al. 2014a; Hausmann et al. 2016; Helmig et al. 2016). These models are highly sensitive to the choice of $\delta^{13}$C$_{CH_4}$, $\delta^2$H$_{CH_4}$, and C$_2$H$_6$:CH$_4$ end-member signatures for each of the various emissions sources, broadly defined as microbial (wetlands, rice paddies, ruminants, termites, and landfills/waste), fossil fuel (coal, oil, natural gas and geological seepage), and biomass burning. For example, a 5% downward adjustment in the fossil fuel $\delta^{13}$C$_{CH_4}$ source signature increases anthropogenic fossil fuel emissions of CH$_4$ from approximately 100 to 150 Tg yr$^{-1}$ (Schwietzke et al. 2016).

Despite the critical importance of accurate source signature data, there has been no recent comprehensive effort to define CH$_4$ source signatures for the atmospheric modeling community (Table 1). Early studies from the 1980s and early 1990s provided tables of average values for each of the various CH$_4$ source categories, typically with little or no metadata on sample size or geographic origin (Deines, 1980; Quay et al. 1988; Stevens and Engelkemeir, 1988; Whiticar, 1989, 1993). Subsequent studies referred back to the original data tables with no accounting of sample size, error/range, or geographic and geological representation (e.g., Fung et al. 1991; Levin et al. 1994; Ferretti et al. 2005; Quay et al. 1999; Mikaloff Fletcher et al. 2004; Bousquet et al. 2006). Other top-down studies have often assumed a set of canonical end-member values used in previous modeling studies, without reference to the primary data, essentially a form of self-citation within the atmospheric modeling community (Gupta et al. 1996; Tyler et al. 1999; Houweling et al. 2000; Lassey et al. 2007; Neef et al. 2010; Dlugokencky et al. 2011; Monteil et al., 2014).
al. 2011). For the fossil fuel category of CH₄ source signatures, there have been virtually no original measurements since Levin et al. (1994). Moreover, model sensitivity to source signature values is rarely tested (e.g. Schwietzke et al. 2014a, 2016; Rice et al. 2016).

There is in fact a vast literature on the molecular and isotopic composition of natural and anthropogenic sources of CH₄, going back decades. The literature has grown significantly since the early studies of the 1980s from which most canonical source signature values originally were derived. This paper describes a comprehensive global database of δ¹³CCH₄, δ²HCH₄ and C₂H₆/C₄H₄ source signatures for fossil fuel, microbial and biomass burning sources of CH₄ compiled from public domain sources. Data distributions are discussed within the context of existing and evolving natural gas genetic origin frameworks (Schoell, 1983; Whiticar et al., 1986; Whiticar, 1989, 1999; Etiope 2015; Milkov et al., 2017). The database is intended primarily for use by atmospheric scientists working on top-down modeling of CH₄ emissions at regional to global scales. This “living” database will be updated every 2-3 years so that the modeling community has access to the most up-to-date and comprehensive collection of CH₄ source signature data available. The database may also prove useful for petroleum geoscientists interested in genetic characterization of natural gas across different basins and formations. Hydrogeochemists may use the database for analyzing the origin and fate of hydrocarbon gases in groundwater in specific oil and gas producing basins.

2. DATABASE METHODS AND DESCRIPTION
2.1 Database Version
The 2017 version of the source signature database is accessed from the NOAA Earth Systems Research Laboratory with this link: https://www.esrl.noaa.gov/gmd/ccgg/arc/?id=123. This version supersedes an earlier version (Sherwood et al., 2016) published as a complement to Schwietzke et al. (2016). Whereas the previous version reported values of δ¹³CCH₄ only, the 2017 version expands the range of geochemical parameters, as described in section 2.4 below. Other minor changes to the database are noted in the database “readme” file.

2.2 Types of Gas
The database is separated into fossil fuel and non-fossil fuel sources of CH₄. Fossil fuel sources comprise conventional natural gas, coal gas and shale gas. Shale gas is included as a separate category because of growing interest in CH₄ emissions associated with this form of unconventional gas production. Both conventional and shale gas include natural gas co-produced with oil. Coal gas includes both coal mine gases and coal bed methane (CBM). All three fossil fuel gas types are representative of reservoir gases measured from producing or previously producing oil or gas wells or coal mines. Data from exploratory wells were excluded, as these are not broadly representative of atmospheric emissions. The database does not currently distinguish between oil and non-oil associated gas or between different ranks of coal (i.e., lignite, bituminous and anthracite). However, the database includes the locations of each sample, which may be used to make this distinction based on activity data (e.g., production based on coal rank at a given coal mine). Pipeline (processed) distribution gases are not included in the database, primarily due to lack of data availability. Users of this database should be aware that, due to preferential stripping of alkane components, processed gases may have...
different molecular compositions than the reservoir gases represented herein. Also, the molecular composition of distribution gases in any region may change over time (Schwietzke et al., 2014b). By contrast, isotopic signatures are unaffected by gas processing except for mixing of 2 or more isotopic end-members (Schoell et al., 1993). Geological seepage gases, i.e. the natural source component of the “fossil fuel” category (Etiope et al. 2008; Etiope, 2009; 2015), are not included in this database. A global database of onshore seeps is discussed in Etiope (2015) and available by CGG (2015). The composition of seepage gases and their influence on the global CH₄ budget is the subject of ongoing research.

Non-fossil fuel sources of CH₄ in the database consist of modern microbial sources and biomass burning. Modern microbial data are from rice paddies, ruminants (C3- and C4-plant eating cattle, sheep, goats and their manure), termites, waste/landfills and wetlands (bogs/peat, deltas, estuaries, floodplains, lagoons, lakes, marshes, ponds, rivers, swamps and tundra). Biomass burning data are from brush, forests/woodlands, grasses, and pastures.

2.3 Data Gathering
Data were obtained from the peer-reviewed literature, conference proceedings, graduate theses, and government reports and databases. Government databases include the U.S. Geological Survey Energy Geochemistry Database (http://energy.usgs.gov/GEOchemistryGeophysics/GEOchemistryLaboratories/GEOchemistryLaboratories-GEOchemistryDatabase.aspx), the Geological Survey of the Netherlands “NLOG” database (available by request through http://nlog.nl/en/gas-properties), and the Geoscience Australia “ORGCHEM” database (available by request through http://www.ga.gov.au/search/index.html#/). Google Scholar, Web of Science, the American Association of Petroleum Geologists (AAPG; http://www.datapages.com/), and the Society of Petroleum Engineers (SPE; www.spe.org) were used to search for data. Use of English language search tools presented an unavoidable bias in data gathering. Searches focused on publications with gas isotopic data. Since gas compositional analysis is a pre-requisite for subsequent isotopic analysis in most laboratories, gas compositional data are included with δ¹³C₂H₄ and δ²H data if reported in the original source. Note that the literature contains far more publications with gas compositional data alone. All of the data can be traced back to original sources using the references provided. To maintain data transparency, industry proprietary data were excluded.

The database is separated into fossil fuel and non-fossil fuel (modern microbial and biomass burning sources) data tables for two practical reasons. First, the petroleum geochemistry literature tends to report analyses for discrete samples, for example, production gas analyses from individual wellheads or analyses from discrete stratigraphic horizons in a wellbore. By contrast, the literature on non-fossil fuel sources of CH₄ more commonly reports statistical summaries (e.g., multiple measurements at a given location/time) as opposed to discrete sample data. Second, fossil fuel data usually include gas composition of C₂+ alkanes and permanent gases and isotopic compositions of C₂+ alkanes. The non-fossil fuel literature rarely reports data on these additional parameters, even though microbial processes in fact produce C₂+, albeit in negligible quantities (< 0.1%) compared to CH₄ (Oremland, 1981; Ladygina et al.)
2006; Xie et al. 2013). Rather than trying to fit these two fundamentally different types of data into a common table format, they are presented separately.

### 2.4 Analytical Parameters

Table 2 lists analytical parameters included in the database. For fossil fuel gases, parameters include molar percent composition of permanent gases (N₂, O₂, CO₂, Ar, H₂, H₂S, He) and C₁ to C₅ alkanes (CH₄, C₂H₆, C₃H₈, iso-C₄H₁₀, n-C₄H₁₀, iso-C₅H₁₂, n-C₅H₁₂, C₆H₁₄), and δ¹³C and δ²H isotopic ratios of C₁ to C₅ alkanes. Though not commonly used in 3D inverse modeling studies of the global CH₄ budget, alkane compositions are important for source attribution in regional air quality/emissions studies (Karion et al. 2013; Pétron et al. 2014; Peischl et al., 2015; Townsend- Small et al. 2016; Kort et al. 2016). The δ¹³C and δ²H isotopic signatures of C₂+ alkanes may also prove useful as source tracers with future advances in analytical instrumentation. For non-fossil fuel samples, δ¹³C and δ²H of CH₄ are the only parameters provided in the database.

### 2.5 Stable Isotope Notation and Standardization

Stable isotopic data are reported in conventional delta notation: 

\[ \Delta X = \left( \frac{R_{\text{Sample}}}{R_{\text{Standard}}} - 1 \right) \times 1000, \text{ where } \Delta X = \delta_{13}C \text{ or } \delta_{2}H \text{ and } R = \frac{^{13}C}{^{12}C} \text{ or } \frac{^2H}{^1H}, \text{ respectively.} \]

δ¹³C data are reported on the PeeDee Belemnite/Vienna PeeDee Belemnite (PDB/VPDB) scale and δ²H data are reported on the Vienna Standard Mean Ocean Water (VSMOW) scale. The “Vienna” version of the PDB scale, signifying that the original PDB reference material used to define the scale ran out and was replaced with the NBS-19 reference material, is nominally identical to the predecessor PDB scale (Gröning, 2004). For references in which the scales were not stated explicitly, we assume the use of PDB/VPDB and VSMOW scales, based on the fact that the use of PDB to define the δ¹³C scale and VSMOW to define the δ²H scale goes back to the 1950s and early 1960s (Craig, 1953; 1961) and that the oldest reference in the database (Dubrova and Nesmelova, 1968), postdates formal recognition of these scales.

### 2.6 Data Screening

Data screening for the fossil fuel data consisted of the following steps. 1) Location metadata (Country, State/Region, Basin, Formation) were checked for logical compatibility. 2) To aid searching for basin-specific data, wherever possible fossil fuel data were assigned to a corresponding sedimentary basin in the Robertson Tellus Sedimentary Basins of the World (available at: http://www.datapages.com/gis-map-publishing-program/gis-open-files/global-framework/robertson-tellus-sedimentary-basins-of-the-world-map). 3) Data duplicates were merged. This step was particularly important for the USGS Energy Geochemistry Database as it includes data from several other sources including the Gas Research Institute report on U.S. natural gas analyses (Jenden & Kaplan 1989), peer-reviewed papers, and other U.S.G.S. data reports. For merged duplicates, references to both sources are provided. 4) Obvious outliers, such as individual gas concentrations greater than 100%, O₂ concentrations greater than 21%, total gas compositions summing to greater than 100% (plus 10% to allow for analytical and rounding errors), and positive values of δ¹³C and δ²H were omitted. For the non-fossil fuel data, no data-screening steps were taken; data are provided as originally reported in the respective sources.
2.7 Data Quality

This database was not subject to a data quality assessment. The data were generated from countless laboratories in different countries over a span of five decades. Source publications also span a wide range in academic rigor, from conference proceedings to peer-reviewed journals. Milkov (2010) analyzed natural gas data from the West Siberia Basin and found that Soviet Russia era papers from the 1970s reported $\delta^{13}$C$_{CH4}$ values that were too negative by $\sim 7$% compared to data generated in the late 1990s by U.S., German and Russian labs, while Soviet era papers from the 1980s reported values that were too positive by $\sim 4.5$%. We make no attempts to correct for these systematic errors; rather we caution users of this database to evaluate and use the data appropriately. By sheer number of samples ($n = 10,706$) and data sources, systematic errors inherent to any single dataset average out over the whole database while random errors have negligible impact on measures of central tendency.

3. RESULTS AND DISCUSSION

3.1 Data Summary

Fossil fuel sources comprise 8,734 data records from 149 published sources. Table 3 provides a summary of the number of countries, basins, fields, formations, and published source by gas type (conventional gas, coal gas, shale gas) and specified analytical parameter ($\delta^{13}$C$_{CH4}$, $\delta^2$H$_{CH4}$, C$_2$H$_6$:CH$_4$). Non-fossil fuel sources comprise 1,972 data records from 41 published sources. Table 4 provides a summary of the number of countries, regions and published source by CH$_4$ source (rice paddies, ruminants, termites, waste, wetlands, biomass burning) and parameter ($\delta^{13}$C$_{CH4}$, $\delta^2$H$_{CH4}$). Finally, Table 5 provides unweighted statistical summaries by CH$_4$ source and parameter.

3.2 Data Representativeness

Fig. 1 shows global maps of the number of samples in each country by gas type. These maps are further broken down by parameter ($\delta^{13}$C$_{CH4}$, $\delta^2$H$_{CH4}$, C$_2$H$_6$:CH$_4$) in Fig. 2. Representativeness of the fossil fuel data is assessed by comparison of sample counts from each country to that country’s coal and natural gas production volumes from the BP Statistical Review of World Energy, 2016 (http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html) (Fig. 3). This was done at the level of individual countries owing to difficulty in obtaining production statistics at the sub-country level for all the countries in the database. We note that reservoir gases vary compositionally and isotopically within individual countries, basins, fields and formations (Fig. 4). Within an individual formation, for example, natural gas can range from microbial gas in shallow/thermally immature areas, to oil-associated gas in deeper/thermally mature areas, to non-associated dry gas in thermally over-mature areas. Similarly, the type (i.e. rank) of coal gas data presented for any specific country may not be representative of the dominant coal type produced in that country.

Despite isotopic and compositional variability within countries, country-level analysis is the finest practical spatial resolution that can be assessed for the global dataset. Shale gas was excluded from this analysis of representativeness, since shale gas production is limited mostly to Canada and USA. For the parameter $\delta^{13}$C$_{CH4}$, the database is representative of 84% of global
natural gas production and 80% of global coal production for the time period 2000-2015. For conventional gas, the countries with the highest numbers of samples with δ^{13}C_{CH4} are USA (n = 2,042), China (834), Russia (556), Canada (402) and Australia (400) (Fig. 3). Countries with no conventional gas data include Algeria, Malaysia, Turkmenistan, United Arab Emirates and Venezuela, which together account for 12.2% of global natural gas production. For coal gas, the countries with the largest sample sizes include USA (722), China (196), Australia (110) and Poland (105) (Fig. 3). Countries with no coal gas data representation include India, Indonesia, Kazakhstan, Ukraine and Columbia, which together account for 14.5% of global coal production.

For the parameter δ^{3}H_{CH4}, the database is representative of 73% of global natural gas production and 74% of global coal production. For C_{2}H_{6}:CH_{4} ratio data, the database is representative of 76% of global natural gas production and 31% of global coal production.

Sample biases can be mitigated by weighting values by each country’s fraction of global gas or coal production (Schwietzke et al. 2016) or by other methods suited to the specific data use.

Representativeness is generally poorer for the non-fossil data, owing in part to the smaller total sample sizes and the lack of data for several key areas. For example, there are few microbial or biomass burning data from Southeast Asia and Africa, two areas of significant wetlands, termite and biomass burning CH_{4} emissions. These areas constitute important data gaps that should be targeted for more intensive data mining and/or future field studies.

### 3.3 Genetic Characterization

Fig. 5 shows a natural gas genetic characterization plot of δ^{13}C_{CH4} versus δ^{3}H_{CH4}, first presented in Whiticar et al. (1986), and modified in Whiticar (1989, 1999). The characterization framework in Fig. 5 and in other plots of δ^{13}C_{CH4} versus alkane molecular compositions (Bernard 1978; Schoell 1983; Faber and Stahl, 1984) originally were developed by researchers at the German Federal Institute for Geosciences and Natural Resources (BGR) in the 1970s and early 1980s.

These plots were derived largely from proprietary industry data. Because the data could not be publicized, the characterization plots were published without showing the underlying data used in their development. These characterization schemes are still widely used to this day, despite that fact that the literature data on gas isotope ratios and compositions has expanded by orders of magnitude since the 1980s. Fig. 5 shows the distribution of conventional gas, coal gas and shale gas in relation the major genetic fields: thermogenic, microbial CO_{2} reduction and microbial fermentation. It also shows the field for gases from geothermal, hydrothermal and crystalline rocks. Overall, the low percentage of samples falling outside any of the principal genetic fields in Fig. 5 indicates that this original classification scheme captures essentially the full range of isotopic variability in natural gases; however, the breakdown of sample counts by genetic origin changes with revision to the classic characterization scheme. For example, while the canonical thermogenic field assumes a δ^{13}C value of -50 % or -55‰ as the limit between thermogenic and microbial CH_{4} (Stahl et al. 1974; Schoell 1983; Whiticar et al. 1986), recent work extends the thermogenic field to isotopically lighter values; see below.

Fig. 6 shows a more recent version of the δ^{13}C_{CH4} versus δ^{3}H_{CH4} plot, updated in Etiope (2015) based on a previous, unpublished version of a fossil fuel reservoir dataset. This diagram
distinguishes more types of thermogenic gas, following Etiope et al. (2013) and Hunt (1996), and reports an updated genetic field for abiotic gas, i.e., gas formed by chemical reactions of inorganically derived gases such as carbon dioxide (CO₂), and hydrogen (H₂), and not from degradation of organic matter (Etiopé and Schoell, 2014). The thermogenic field in Fig. 6 extends to δ¹³C = -67 ‰ due to the existence of low-maturity thermogenic gas (Rowe and Muehlenbachs 1999; Milkov & Dzou 2007) and secondary alterations (biodegradation; Milkov, 2010; 2011) that would otherwise be mistaken for primary microbial gas.

Of the 8,734 fossil fuel samples in the database, a subset of (n = 2861) have both δ¹³C and δ²H data and thus are represented on the plot. For conventional gas (n = 1951 δ¹³C − δ²H data pairs), a majority (78%) of the samples plot within the thermogenic field. A smaller percentage of samples plot within the microbial field (17%) or the abiotic field (5%). For coal gas (n = 511), data are more evenly distributed between thermogenic (56%) and microbial (39%) fields, with a smaller percentage falling within the abiotic (2%) field. Because of overlapping genetic fields, percentages sum to > 100%. Additionally, it is important to outline that conventional or coal gases falling within the “abiotic” field, even where it is not overlapping with the thermogenic field, actually have a dominant thermogenic origin. These δ¹³C-enriched gases are, in fact, mainly from over-mature (late stage catagenesis) source rocks from NW Germany (Rotliegend) and China (Songliao and Tarim basins). Further refinement of the genetic characterization plot should therefore account for these late stage thermogenic gases. Shale gas data (n = 396) fall almost entirely within the thermogenic field (91%), with the majority of the data clustered toward the dry gas (“Td” in Fig. 6) end of the thermogenic maturity spectrum. Non-fossil source data (rice paddies, ruminants, waste, wetlands, termites) plot entirely within the microbial fermentation field. Biomass burning has a characteristically enriched isotopic signature, falling within the abiogenic field despite a fundamentally different generation pathway compared to abiogenic natural gas. A revision of the genetic diagram is in fact in progress (Milkov et al. 2017), and statistics of our database will be re-adjusted taking into account this new re-assessment of microbial versus thermogenic isotopic genetic characterization.

3.4 Importance of Isotopically Light Natural Gas and Coal Gas

A long-standing view in the petroleum geochemical literature held that “more than 20% of the world’s discovered gas reserves are of biogenic origin” (Rice & Claypool 1981). This “biogenic gas” was loosely defined by cutoffs of δ¹³C(CH₄) < -55 ‰ and < 2% C₂+ alkanes (C₃H₆ through pentane (C₅H₁₂)). For conventional natural gas in the current database, 14% of the samples have δ¹³C(CH₄) < -55 ‰ and 23% have a C₂H₆:CH₄ ratio < 0.02. These percentages envelope the original Rice & Claypool (1981) estimate. However, it is now known that natural gas within the δ¹³C(CH₄) and % C₂+ cutoffs encompass primary microbial gas (i.e. “biogenic” gas in Rice & Claypool, 1981; formed from microbial CO₂ reduction and methyl fermentation in shallow sediments), secondary microbial gas (formed from biodegradation of thermogenic hydrocarbons; Zengler et al. 1999; Head et al. 2003; Jones et al. 2008), and low maturity thermogenic gas (Rowe and Muehlenbachs 1999; Milkov & Dzou 2007). Analysis of the δ¹³C and molecular ratios of C₂+ alkanes and CO₂ is often the only means of distinguishing between these three types of gas (Milkov, 2011).
At the global level, primary and secondary microbial gas are thought to account for ~3-4% and ~5-11%, respectively, of conventional recoverable natural gas reserves (Milkov, 2011).

Secondary microbial gas accounts for a larger share of global conventional gas production: giant Cenomanian gas pools of secondary microbial CH₄ (mean δ¹³C(CH₄) = -51.8 ‰) found at depths < 1500 m in the West Siberia Basin alone account for ~17 % of the global conventional gas production (Milkov, 2010).

Microbial methanogenesis is even more significant for coals (Rice, 1993), with an approximately even distribution between thermogenic and microbial genetic origins (Fig. 5, 6). The two largest coal mines in the world (North Antelope Rochelle and Black Thunder mines) are located in the Powder River Basin, Wyoming, USA. Coal gas from these formations is microbial (fermentation) in origin (mean δ¹³C(CH₄) = -59.1 ‰, n = 267; mean δ²H(CH₄) = -309.6 ‰, n = 118). However, as discussed above, we note that some gas, traditionally considered microbial because of its low δ¹³C values, may actually have a thermogenic origin. Coals can also generate secondary microbial gas (Scott et al. 1994).

3.5 Data Distributions

Figs. 7 and 8 show normalized probability distributions of δ¹³C(CH₄) and δ²H(CH₄) for fossil fuel and modern microbial processes (with their respective sub-categories) and biomass burning sources of CH₄. The distributions show wide overlap between different CH₄ source categories, thus highlighting the critical need for robust weighting schemes that result in globally or regionally representative measures of central tendency (discussed below).

Data distributions for modern microbial processes have relatively normal distributions with tight overlap between the different sub-categories. The distributions for biomass burning show characteristic bimodality, caused by differences between isotopically lighter C₃ and isotopically heavier C₄ vegetation. Fossil fuel δ¹³C and δ²H exhibit left-skewed (conventional and shale gas) or bimodal (coal) distributions arising from the presence of microbial and low-maturity thermogenic gas as described above. This also leads to relatively wider data ranges than the non-fossil categories.

Fig. 7 also indicates the δ¹³C of atmospheric CH₄ (~52.6 ‰) before fractionation by photodegradation, calculated as measured atmospheric δ¹³C(CH₄) (mean -47.3 ‰ in the year 2016; White et al., 2017) plus an average fractionation factor ε = -6.3 ± 0.8 ‰ (Schwietzke et al., 2016). This value represents the hinge point upon which CH₄ emissions fluxes are estimated by isotopic mass balance (e.g. Whiticar and Schaefer, 2007). Modern microbial processes have δ¹³C(CH₄) signatures falling to the left of the hinge point, thus lower δ¹³C(CH₄) requires lower emissions to isotopically balance fossil fuel and biomass burning sources; higher δ¹³C(CH₄) requires higher emissions. Conversely, fossil fuel and biomass burning source categories have δ¹³C(CH₄) signatures falling to the right of the hinge point, thus lower δ¹³C(CH₄) requires higher emissions; higher δ¹³C(CH₄) requires lower emissions. Biomass burning falls furthest from the
hinge point (mean δ^{13}C_{CH4} = -26.2 ± 4.8 ‰, unweighted by proportion of C3 and C4 vegetation), therefore it has the most leverage on the isotopic mass balance.

In Fig. 8 the pre-fractionation hinge point is more poorly constrained, owing to greater variability in measured atmospheric δ^{13}H_{CH4} (-95 ± 5 ‰) and, more importantly, uncertainty in the estimated fractionation factor ε = -235 ± 80 ‰ (Gierczak et al., 1997). Modern microbial δ^{13}H_{CH4} signatures are within range of the estimated pre-fractionation atmosphere. Biomass burning and fossil fuel signatures fall to the right of the hinge “point”, hence lower δ^{2}H_{CH4} requires higher emissions and higher δ^{3}H_{CH4} requires lower emissions for both these categories.

Unweighted mean δ^{13}C_{CH4} and δ^{2}H_{CH4} for modern microbial processes and biomass burning (Table 5) are generally within about 2 ‰ of typical values used in published CH4 budget models (Schwietzke et al., 2016). By contrast, fossil fuel δ^{13}C_{CH4} and δ^{2}H_{CH4} summary statistics (Table 5) show wider disparity with source signature values used in published CH4 budget models (Table 1). Remarkably, unweighted mean δ^{13}C_{CH4} for conventional natural gas (-44.0 ± 10.7 ‰) is identical to the value (-44 ‰) originally indicated by Craig et al. (1988), Stevens and Engelkemier (1988; thermogenic gas), Quay et al. (1991; oil-associated gas) and Whiticar (1989; 1993). However, this value is 4-6 ‰ lighter than the range of -38 to -40 ‰ typically used in more recently published models (Gupta 1996; Lassey et al. 2000; 2007; Tyler et al., 2007; Neef et al., 2010; Montiel et al., 2011; Rigby et al., 2012; Ghosh et al., 2015; Schaefer et al., 2016).

For coal gas, unweighted mean δ^{13}C_{CH4} (-49.5 ‰) is even more significantly depleted compared to typical values of -35 to -37 ‰ assumed in virtually all previous studies. These canonical values likely were derived from bituminous and anthracite coal, which is isotopically heavier than lignite and sub-bituminous coal (Rice, 1993; Zazzeri et al. 2016), yet lignite and sub-bituminous coal account for more than half of world coal production (World Energy Council, 2007). Similarly, mean δ^{1}H_{CH4} for natural gas (-194 ‰) and coal (-232 ‰) is 10-15 ‰ and 60-120 ‰, respectively, lighter than literature values. Shale gas, which tends to be isotopically heavier than oil-associated gas because of its higher thermal maturity (Zumberge et al. 2012; Tilley & Muehlenbachs 2013), also exhibits lower mean δ^{13}C_{CH4} (-42.5 ‰) and δ^{2}H_{CH4} (-167 ‰) than indicated in the CH4 budget literature. For all fossil fuel data (uncategorized) unweighted mean δ^{13}C_{CH4} = -44.8 ± 11 ‰ (1 st. dev.) and δ^{2}H_{CH4} = -197 ± 50 ‰.

These results highlight the possibility that widespread use of too-heavy δ^{13}C_{CH4} and δ^{2}H_{CH4} fossil fuel source signatures could have led to systematic underestimation of fossil fuel emissions in the CH4 budget literature. Indeed, Schwietzke et al. (2016) re-analyzed the global CH4 budget using weighted source signature data calculated from an earlier version of this database (Sherwood et al., 2016) and showed that total fossil fuel emissions (excluding geological seepage) is about 50% higher than previously estimated.

Database users are encouraged to adopt appropriate weighting criteria for estimating spatially averaged source signatures. For instance, at the global level, Schwietzke et al. (2016) developed a method to weight fossil fuel δ^{13}C_{CH4} data at the country-level and non-fossil fuel δ^{13}C_{CH4} data at the emission subcategory-level. Weighting fossil fuel δ^{13}C_{CH4} data at the basin-level may be
practical for some countries with sufficient sample size. Basin-level gas production statistics may be used in the weighting procedure as a proxy for basin-level CH$_4$ emissions. However, note that basin-level CH$_4$ emissions may be correlated with basin-level δ$^{13}$CH$_4$. A basin with mature dry gas and no associated oil production (and thus relatively heavy δ$^{13}$C$_{CH_4}$ typically employs less gas processing infrastructure than a basin with associated gas production (and thus relatively light δ$^{13}$C$_{CH_4}$). The former is therefore likely to emit less CH$_4$ per unit of gas production than the latter. This is substantiated by CH$_4$ emission estimates from multiple U.S. oil and gas basins. For example, the dry gas basins Marcellus Shale and Fayetteville are estimated to emit on average 0.3% and 1.9%, respectively, per unit of gas produced (Peischl et al., 2015), whereas the wet gas Denver and Uinta Basins emit on average 4.1% and 8.9%, respectively, per unit of gas produced (Karion et al., 2013; Petron et al., 2014). Thus, using gas production statistics to weight individual basins without knowledge of the respective CH$_4$ emissions may introduce biases.

4. CONCLUSIONS

The database described here is the most comprehensive CH$_4$ source signature database ever compiled. For the fossil fuel category (conventional gas, shale gas and coal gas), the data comprise 8,734 unique records representing 84%, 73% and 76% (respectively for δ$^{13}$C$_{CH_4}$, δ$^2$H$_{CH_4}$ and C$_2$H$_6$:CH$_4$ ratio) of global conventional natural gas production and 80%, 74% and 31% (respectively for δ$^{13}$C$_{CH_4}$, δ$^2$H$_{CH_4}$ and C$_2$H$_6$:CH$_4$ ratio) of global coal production at the country level. For the non-fossil category (rice paddies, ruminants, termites, landfills/waste, wetlands and biomass burning), the data comprise 1,972 records from 15 countries on 5 continents. While this constitutes the most comprehensive global data compilation, additional data may help further reduce uncertainty in the global CH$_4$ budget, especially for regionally distinct CH$_4$ source attribution. In particular, additional wetland and ruminant δ$^{13}$C$_{CH_4}$ data are needed given their large contribution to the global CH$_4$ budget. Database users are encouraged to adopt appropriate weighting criteria to account for variability in emissions specific to each source category.

Unweighted mean δ$^{13}$C$_{CH_4}$ and δ$^2$H$_{CH_4}$ signatures for the non-fossil sub-categories are generally within range of a few per mil of typical values used in the CH$_4$ budget modeling literature. Unweighted mean δ$^{13}$C$_{CH_4}$ and δ$^2$H$_{CH_4}$ signatures for the fossil category, by contrast, are significantly lighter than the canonical values, particularly for coal gas. The origin of this bias is unknown, but may be caused in part, by a tendency among CH$_4$ budget modelers to reference other modeling studies instead of the primary literature on isotopic characterization of natural gas. In addition, an evolving understanding of natural gas genetic origins blurs the traditional cutoffs between microbial or “biogenic” and thermogenic natural gas: fossil fuel CH$_4$ is not exclusively thermogenic and the δ$^{13}$C$_{CH_4}$ of thermogenic CH$_4$ can be < -55 ‰.

Finally, the database includes a relatively new category of fossil fuel CH$_4$, shale gas; these data will become more useful as this resource assumes increasing share of global natural gas production. The availability of gas molecular concentrations will provide additional end-member constraints on fossil fuel emissions at global and regional scales. This “living” database
will be updated every 2-3 years to provide a comprehensive and up-to-date resource for the CH₄ modeling community.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Cooperative Institute for Research in Environmental Sciences (CIRES) and funding from the National Oceanographic and Atmospheric Administration (NOAA). We thank Martin Schoell (Gas Consult International) for compiling an initial version of the fossil fuel database and John Miller and Pieter Tans (NOAA) for initial discussions.
REFERENCES

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A.,
Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen, H.,
modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, J.

University, College Station, Texas, 144 pp., 1978.

Bloom, A. A., Palmer, P. I., Fraser, A., Reay, D. S., and Frankenberg, C.: Large-scale controls of
methanogenesis inferred from methane and gravity spaceborne data. Science, 327(5963), 322–

Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., Prigent, C., Van der
Werg, G. R., Peylin, P., Brunke, E. G., Carouge, C., Langenfelds, R. L., Lathiere, J., Papa, F.,
Ramonet, M., Schmidt, M., Steele, L. P., Tyler, S. C., and White, J.: Contribution of
anthropogenic and natural sources to atmospheric methane variability, Nature, 443, 439–443,
2006.

Bousquet, P., Ringleval, B., Pison, I., Dlugokencky, E. J., Brunke, E.-G., Carouge, C., Chevallier, F.,
Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D. A., Krummel, P. B., Langenfelds, R. L.,
attribution of the changes in atmospheric methane for 2006–2008, Atmos. Chem. Phys., 11,

Harriss, R.: Methane Leaks from North American Natural Gas Systems, Science, 343, 733–735,

Bréas, O., Guillou, C., Reniero, F., and Wada, E.: The Global Methane Cycle: Isotopes and Mixing
Ratios, Sources and Sinks, Isotopes in Environmental and Health Studies, 37(4), 257–379, 2001,
doi:10.1080/10256010108033302.

CGG: Organic Geochemistry Data from FRogi and the Fluid Features Database.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R.,
Contribution of Working Group I to the Fifth Assessment Report of IPCC, edited by: Stocker, T.
F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and


Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J.G., Dlugokencky, E.J.,

Etiop, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F.N., Castaldi, S.,

Jackson, R.B., Alexe, M., Arora, V.K., Beerling, D.J., Bergamaschi, P., Blake, D.R., Brailsford, G.,

Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney,

N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krummel, P.,

Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K.C.,


Peng, C., Peng, S., Peters, G.P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W.J., Saito,

M., Santini, M., Schroeder, R., Simpson, I.J., Spahni, R., Steele, P., Takizawa, A., Thornton, B.F.,

Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., van Weele, M., van der Werf, G.R., Weiss, R.,


Schaefer, H., Fletcher, S. E. M., Veidt, C., Lassey, K. R., Brailsford, G. W., Bromley, T. M.,

Dlugokencky, E. J., Michel, S. E., Miller, J. B., Levin, I., Lowe, D. C., Martin, R. J., Vaughn, B.H.,

and White, J. W. C.: A 21st century shift from fossil-fuel to biogenic methane emissions


Schoell, M.: The hydrogen and carbon isotopic composition of methane from natural gases of


Schwitzke, S., Griffin, W. M., Matthews, H. S., and Bruhwiler, L. M. P.: Natural gas fugitive

emissions rates constrained by global atmospheric methane and ethane, Environ. Sci. Technol.,


Schwitzke, S., Griffin, W. M., Matthews, H. S., and Bruhwiler, L. M. P.: Global bottom-up fossil

fuel fugitive methane and ethane emissions inventory for atmospheric modeling, ACS


Schwitzke, S., Sherwood, O. A., Bruhwiler, L. M. P., Miller, J. B., Etiop, G., Dlugokencky, E. J.,


global fossil fuel methane emissions based on isotope database, Nature, 538, 88–91,


Juan Basin, Colorado and New Mexico – Implications for coalbed gas producibility, AAPG Bull.,


### Table 1: Representative list of atmospheric modeling studies in which isotopic ratios were used to constrain emissions from fossil fuel sources of CH$_4$, showing values of $\delta^{13}$C$_{CH_4}$ and $\delta^2$H$_{CH_4}$ used and the source of those values.

<table>
<thead>
<tr>
<th>Study</th>
<th>Source$^a$</th>
<th>$\delta^{13}$C$_{CH_4}$ (‰)</th>
<th>$\delta^2$H$_{CH_4}$ (‰)</th>
<th>Data Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig et al. 1988</td>
<td>NG</td>
<td>-44</td>
<td>na</td>
<td>Schoell 1980; Rice &amp; Claypool 1981</td>
</tr>
<tr>
<td>Stevens &amp; Engelkeimer 1988</td>
<td>NG (thermo.)</td>
<td>-44 (80 to -25)</td>
<td>na</td>
<td>Schoell 1980; Rice &amp; Claypool 1981</td>
</tr>
<tr>
<td></td>
<td>NG (oil-assoc.)</td>
<td>-40 (30 to -50)</td>
<td>na</td>
<td>Deines, 1980</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37 (44 to -60)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Quay et al. 1988</td>
<td>NG (thermo.)</td>
<td>-42 (76 to -21)</td>
<td>na</td>
<td>Deines, 1980; Schoell 1980</td>
</tr>
<tr>
<td></td>
<td>NG (oil-assoc.)</td>
<td>-41 (60 to -30)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37 (70 to -15)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Whiticar 1989, 1993</td>
<td>NG</td>
<td>-44</td>
<td>-180</td>
<td>Unspecified</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>-130</td>
<td></td>
</tr>
<tr>
<td>Fung et al. 1991</td>
<td>NG</td>
<td>-70 to -41</td>
<td>na</td>
<td>Quay et al. 1991</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-70 to -15</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Quay et al. 1991</td>
<td>NG (thermo.)</td>
<td>-41 (41 to -76)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NG (oil-assoc.)</td>
<td>-44 (60 to -30)</td>
<td>na</td>
<td>Deines, 1980; Schoell 1980</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-35 (70 to -15)</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Levin et al. 1994</td>
<td>NG</td>
<td>-40.5 ± 6.2</td>
<td>185 ± 29</td>
<td>Original measurements</td>
</tr>
<tr>
<td>Stevens 1993</td>
<td>NG</td>
<td>-43 ± 4</td>
<td>na</td>
<td>Schoell 1980; Rice &amp; Claypool 1981</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37 ± 4</td>
<td>na</td>
<td>Deines, 1980</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-35 ± 3</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Gupta et al. 1996</td>
<td>NG</td>
<td>-38</td>
<td>na</td>
<td>Unspecified</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Froney et al. 1999</td>
<td>NG</td>
<td>-40</td>
<td>na</td>
<td>Unspecified</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-35</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Quay et al. 1999</td>
<td>NG</td>
<td>-43 ± 7</td>
<td>185 ± 20</td>
<td>Stevens &amp; Engelkeimer 1988; Quay et al. 1991; Levin et al. 1994; Stevens 1993; Gupta et al. 1996</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-36 ± 7</td>
<td>140 ± 20</td>
<td></td>
</tr>
<tr>
<td>Tyler et al. 1999</td>
<td>NG</td>
<td>-38</td>
<td>na</td>
<td>Fung et al. 1991</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Houweling et al. 2000</td>
<td>NG (thermo.)</td>
<td>-40</td>
<td>na</td>
<td>Levin et al. 1994; Quay et al. 1999</td>
</tr>
<tr>
<td>Lassey et al. 2000</td>
<td>FF</td>
<td>-40</td>
<td>na</td>
<td>Unspecified</td>
</tr>
<tr>
<td>Mikaloff Fletcher et al. 2004</td>
<td>NG</td>
<td>-44</td>
<td>na</td>
<td>Whiticar, 1993</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Feronetti et al. 2005</td>
<td>FF</td>
<td>-40</td>
<td>na</td>
<td>Unspecified</td>
</tr>
<tr>
<td>Bouquet et al. 2006</td>
<td>NG</td>
<td>-44</td>
<td>na</td>
<td>Mikaloff Fletcher et al. 2004</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Lassey et al. 2007</td>
<td>NG</td>
<td>-35 ± 5</td>
<td>na</td>
<td>not specified</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-40 ± 5</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Tyler et al. 2007</td>
<td>NG</td>
<td>-38</td>
<td>-175</td>
<td>Gupta et al. 1996</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>-175</td>
<td></td>
</tr>
<tr>
<td>Whiticar &amp; Schaefer 2007</td>
<td>NG</td>
<td>-44</td>
<td>-180</td>
<td>Unspecified</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>-140</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geol</td>
<td>-41.8</td>
<td>-200</td>
<td></td>
</tr>
<tr>
<td>Neef et al. 2010</td>
<td>NG</td>
<td>-35</td>
<td>na</td>
<td>Lassey et al. 2007</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-40</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Dlugokencky et al. 2011</td>
<td>NG</td>
<td>-34 to -50 (± 3)</td>
<td>175 ± 10</td>
<td>Unspecified</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-35 ± 3</td>
<td>175 ± 10</td>
<td></td>
</tr>
<tr>
<td>Montiel et al. 2011</td>
<td>NG</td>
<td>-40</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-35</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Kirschke et al. 2013</td>
<td>FF</td>
<td>-25 to -55</td>
<td>na</td>
<td>Montiel et al. 2011; Neef et al. 2010</td>
</tr>
<tr>
<td>Ghosh et al. 2014</td>
<td>NG</td>
<td>-40</td>
<td>na</td>
<td>Unspecified</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-35</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>Rice et al. 2016</td>
<td>NG</td>
<td>-44</td>
<td>-175</td>
<td>Whiticar 1993, Tyler et al. 2007</td>
</tr>
<tr>
<td></td>
<td>Coal</td>
<td>-37</td>
<td>-175</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ NG, natural gas; FF, fossil fuels; Geol, geological seepage
Table 2: List of geochemical parameters by gas type included in the database.

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Parameters</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fossil Fuel</td>
<td>Permanent Gases: N₂, O₂, CO₂, Ar, H₂, H₂S, He</td>
<td>Mol. %</td>
</tr>
<tr>
<td></td>
<td>Alkanes: CnH₂₄, CnH₄₀, iso-C₄H₁₀, n-C₄H₁₀, iso-C₅H₁₂, n-C₅H₁₂, C₆H₁₄</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Isotopes: δ¹³C CH₄, δ¹³C C₂H₆, δ¹³C C₃H₈, δ¹³C i-C₄H₁₀, δ¹³C n-C₄H₁₀, δ¹³C i-C₅H₁₂, δ¹³C n-C₅H₁₂, δ¹³C C₆H₁₄</td>
<td>‰</td>
</tr>
<tr>
<td>Non-Fossil Fuel</td>
<td>Isotopes: δ¹³C CH₄ and δ²H CH₄</td>
<td>‰</td>
</tr>
</tbody>
</table>
Table 3: Fossil fuel data: number of countries, basins, fields, formations, and references by gas type and specified chemical parameter.

<table>
<thead>
<tr>
<th></th>
<th>Conventional Gas</th>
<th>Coal</th>
<th>Shale Gas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\delta^{13}$C</td>
<td>$\delta^{2}$H</td>
<td>$\delta^{2}$H</td>
</tr>
<tr>
<td>Countries</td>
<td>43</td>
<td>27</td>
<td>36</td>
</tr>
<tr>
<td>Basins*</td>
<td>151</td>
<td>70</td>
<td>188</td>
</tr>
<tr>
<td>Fields*</td>
<td>1238</td>
<td>424</td>
<td>989</td>
</tr>
<tr>
<td>Formations*</td>
<td>723</td>
<td>108</td>
<td>187</td>
</tr>
<tr>
<td>References</td>
<td>112</td>
<td>56</td>
<td>83</td>
</tr>
</tbody>
</table>

*Does not account for unknown/unspecified basins, fields, or formations.
Table 4: Non-fossil data: Number of countries, regions and references by source and specified chemical parameter.

<table>
<thead>
<tr>
<th>Source</th>
<th>Rice Paddies</th>
<th>Ruminants</th>
<th>Termites</th>
<th>Waste</th>
<th>Wetlands</th>
<th>Biomass Burning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\delta^{13}$C</td>
<td>$\delta^{13}$C</td>
<td>$\delta^{13}$C</td>
<td>$\delta^{2}$H</td>
<td>$\delta^{13}$C</td>
<td>$\delta^{2}$H</td>
</tr>
<tr>
<td>Countries*</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Regions</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>References</td>
<td>11</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

*Does not account for unknown/unspecified countries or regions
Table 5: Database summary statistics (unweighted) by gas type and parameter.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Statistic</th>
<th>Conventional Oil &amp; Gas</th>
<th>Fossil Fuel</th>
<th>Modern Microbial</th>
<th>Biomass Burning*</th>
<th>Waste</th>
<th>Wetlands</th>
<th>All Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Coal</td>
<td>Shale Gas</td>
<td>Rice paddies</td>
<td>Ruminant&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Termites</td>
<td>Waste</td>
<td>Wetlands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All Sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ&lt;sup&gt;13&lt;/sup&gt;C&lt;sub&gt;13&lt;/sub&gt;H&lt;sub&gt;2&lt;/sub&gt;&lt;sup&gt;6&lt;/sup&gt;:CH&lt;sub&gt;4&lt;/sub&gt; (‰)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Mean</td>
<td>-44.0</td>
<td>-40.5</td>
<td>-42.5</td>
<td>-44.8</td>
<td>-6.2</td>
<td>-45.4</td>
<td>-6.4</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>-42.2</td>
<td>-49.8</td>
<td>-41.1</td>
<td>-42.9</td>
<td>-6.3</td>
<td>-47.1</td>
<td>-6.3</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>-87.0</td>
<td>-85.6</td>
<td>-69.7</td>
<td>-87</td>
<td>-6.7</td>
<td>-74.4</td>
<td>-72.8</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>-14.8</td>
<td>-16.8</td>
<td>-24.4</td>
<td>-14.8</td>
<td>-54</td>
<td>-50.3</td>
<td>-55.7</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>10.7</td>
<td>11.2</td>
<td>6.7</td>
<td>10.7</td>
<td>3.9</td>
<td>6.7</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>St.br.</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>1.2</td>
</tr>
<tr>
<td>δ&lt;sup&gt;15&lt;/sup&gt;N&lt;sub&gt;2&lt;/sub&gt; (%)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Mean</td>
<td>-1.94</td>
<td>-2.32</td>
<td>-1.67</td>
<td>-1.97</td>
<td>-3.23</td>
<td>-3.56</td>
<td>-3.43</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>-1.86</td>
<td>-2.35</td>
<td>-1.46</td>
<td>-1.92</td>
<td>-3.28</td>
<td>-3.05</td>
<td>-3.43</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>-3.93</td>
<td>-4.05</td>
<td>-3.15</td>
<td>-4.15</td>
<td>-3.96</td>
<td>-3.58</td>
<td>-3.43</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>-6.2</td>
<td>-7.5</td>
<td>-1.01</td>
<td>-6.2</td>
<td>-3.01</td>
<td>-2.95</td>
<td>-3.43</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>47</td>
<td>52</td>
<td>4.4</td>
<td>51</td>
<td>36</td>
<td>29</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>St.br.</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>n/a</td>
</tr>
<tr>
<td>C&lt;sub&gt;4&lt;/sub&gt;H&lt;sub&gt;2&lt;/sub&gt;:CH&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Mean</td>
<td>0.0740</td>
<td>0.0166</td>
<td>0.0140</td>
<td>0.0359</td>
<td>0.059</td>
<td>0.036</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0890</td>
<td>0.0180</td>
<td>0.0159</td>
<td>0.0396</td>
<td>0.057</td>
<td>0.036</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>3.966</td>
<td>1.3277</td>
<td>0.3911</td>
<td>2.966</td>
<td>0.018</td>
<td>0.018</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>St.Dev.</td>
<td>0.1380</td>
<td>0.0581</td>
<td>0.0197</td>
<td>0.138</td>
<td>0.005</td>
<td>0.005</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>St.br.</td>
<td>0.0017</td>
<td>0.0028</td>
<td>0.0024</td>
<td>0.0034</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<sup>a</sup> Raw values, not weighted by proportion of C3- versus C4-eating ruminants.

<sup>b</sup> Raw values, not weighted by proportion of C3 versus C4 vegetation.
Figure 1: Global maps of country-specific sample counts for conventional gas, coal gas, shale gas and non-fossil CH$_4$. 

Sample Count

- 1001–5000
- 501–1000
- 101–500
- 51–100
- 11–50
- 1–10
Figure 2: Global maps of country-specific sample counts for conventional gas, coal gas, shale gas and non-fossil methane by geochemical parameter.
Figure 3: Tornado plots of $\delta^{13}$C$_{\text{CH}_4}$ sample counts versus production statistics for a) conventional natural gas and b) coal. Shale is gas not included because it is primarily from USA and Canada. “Other Asia Pacific” represents Afghanistan, Japan, New Zealand, Taiwan; “Other Africa” represents Angola and Tunisia; “Other Europe & Eurasia” represents Austria, France, Hungary, Lithuania, and Turkey; “Other South and Central America” represents Barbados; “Other Middle East” represents Israel.
Figure 4: Stripchart of conventional gas $\delta^{13}$C$_{CH_4}$ by continent and sedimentary basin, demonstrating high levels of variability within individual basins.
Figure 5: Genetic characterization plot of $\delta^{13}$C$_{\text{CH}_4}$ versus $\delta^{2}$H$_{\text{CH}_4}$ showing data distributions with respect to genetic domains, as traced from Whiticar (1999). Atmospheric value represents global average atmospheric CH$_4$ in the year 2015.
Figure 6: Genetic characterization plot of $\delta^{2}H_{\text{CH}_{4}}$ versus $\delta^{13}C_{\text{CH}_{4}}$ redrew from Etiophe (2015), based on thermogenic fields by Hunt (1996) and Milkov (2011) and abiotic gas from Etiophe and Sherwood Lollar (2013) and Etiophe and Schoell (2014). The reversed vertical and horizontal axes as compared to Fig. 5 follows conventions established previously to emphasize abiotic fields. M, microbial; T, thermogenic; A, abiotic; MCR, microbial CO$_{2}$ reduction; MAF, microbial acetate fermentation; ME, microbial in evaporitic environment; TO, thermogenic with oil; TC, thermogenic with condensate; TD, dry thermogenic; TH, thermogenic with high-temperature CO$_{2}$–CH$_{4}$ equilibration; and TLM, thermogenic low maturity; GV, geothermal-volcanic systems; S, serpentinitized ultramafic rocks; PC, Precambrian crystalline shields.
Figure 7: Normalized probability density distributions for the $\delta^{13}C_{CH_4}$ of microbial, fossil and biomass burning sources of $CH_4$. The flux-weighted average of all sources produces a mean atmospheric $\delta^{13}C_{CH_4}$ of $\sim$-52.5‰, as inferred from measured atmospheric $\delta^{13}C_{CH_4}$ and isotopic fractionation associated with photochemical $CH_4$ destruction (see text).
Figure 8: Normalized probability density distributions for the $\delta^2$H$_{CH_4}$ of microbial, fossil and biomass burning sources of CH$_4$. The flux-weighted average of all sources produces a mean atmospheric $\delta^2$H$_{CH_4}$ of between -245 and -415 ‰, as inferred from measured atmospheric $\delta^2$H$_{CH_4}$ and isotopic fractionation associated with photochemical CH$_4$ destruction (see text).