Supplementary information

High-resolution national mapping of natural gas composition substantially updates methane leakage impacts

Philippine M. Burdeau, Evan D. Sherwin, Sébastien C. Biraud,
Elena S.F. Bermam, Adam R. Brandt

Table of contents

S1 Workflow and data processing details

S1.1 Consolidation of Enverus oil and gas production data at the well level

We detail here how we construct a national well-level dataset for all available years using Enverus DrillingInfo. Four datasets are available for download:

- `Producing Entity Header`: Contains information about the producing entity.
- `Well Header`: Contains information on individual wells.
- `Producing Entity Production`: Contains production data at the producing entity level.
- `Well Production`: Contains production data at the well level.

If all production data were reported at the well level, the process would be straightforward. However, production is sometimes reported at different levels.

Completion-level reporting

For wells with multiple completions, the `Well Header` dataset classifies the well as "MULTIPLE" in the Producing Entity Type field, with a corresponding Producing Entity ID listing completions. The `Production Header` dataset contains multiple rows corresponding to each completion (each Entity ID) and Entity Type is 'COM'. The `Producing Entity Production` dataset reports production at the completion level, with one row per completion. The `Well Production` dataset consolidates total production across completions into a single row per well.

For wells with a single completion, both datasets align, and production values match.

Lease-level reporting

In states where production is reported at the lease level, the `Well Header` dataset lists individual wells within a lease but does not provide producing entity details. The `Producing Entity Header` dataset provides coordinates for a representative well. The `Producing Entity Production` dataset includes an "API/UWI List" field, linking all wells within a lease. The `Well Production` dataset contains no data.

Resolving reporting inconsistencies

Some entities switch between WELL, UNIT, and LEASE types over time, leading to potential duplicate production records. To mitigate this, we do as follows.

When both WELL and LEASE production are reported for the same API/UWI and Entity ID, we retain only WELL-level production.

If only lease-level data are available, we allocate the production to the representative well. We could also allocate proportionally to each well in the lease, but in practice we computed that the average number of wells by lease is 5 and most wells are clustered spatially (only 1 case in Kansas with 2 wells separated by 188 miles).

We exclude entities such as SWD, waste plants, and drip points, as they are not directly related to individual wells.

S1.2 Gas composition prediction workflow summary

In Figure S1 we present the overall methodological workflow for estimating gas composition at specific production locations and times. It summarizes the integration of the different data sources, the neural network training and validation process, and the combination of predictions from spatio-temporal kriging and from using the neural network. The equation numbers shown in the diagram refer to the corresponding equations detailed in Section 4.2.

S2 Empirical support for study design

S2.1 Consistency of data reported to the GHGRP with proprietary data

Proprietary full gas composition data from inlet flows to three processing facilities were provided for this work by an anonymous industry partner. For a preliminary validation of GHGRP-reported methane (C_1) fractions, proprietary data were aggregated into annual averages to match the temporal resolution of the GHGRP dataset. Each dataset was grouped by year to facilitate comparison.

The Mean Absolute Error (MAE) was computed to quantify the difference between proprietary data and GHGRP reporting:

$$\mathrm{MAE} = \frac{1}{Y}\sum_{y=1}^{Y}\left|x^{\mathrm{prop}}_{\mathrm{C1,annual}}(y) - x^{\mathrm{GHGRP}}_{\mathrm{C1,annual}}(y)\right|,$$

where, Y denotes the number of years for which data are available, $x^{\mathrm{prop}}_{\mathrm{C1,annual}}(y)$ is the annual mean methane fraction in the proprietary dataset for year y, $x^{\mathrm{GHGRP}}_{\mathrm{C1,annual}}(y)$ is the annual mean methane fraction reported in the GHGRP dataset for year y.

For one facility (Facility A), proprietary methane composition was available from three separate inlet streams: main, low-pressure (LP), and high-pressure (HP).

To obtain a representative methane fraction for the entire facility, a molar-weighted average was computed to account for differences in gas volume and composition. The proprietary datasets were first merged by date to align corresponding measurements. The total inlet gas volume at each time step was computed as $V_{\mathrm{total}}(t) = V_{\mathrm{main}}(t) + V_{\mathrm{LP}}(t) + V_{\mathrm{HP}}(t)$, where $V_s(t)$ represents the gas volume (MCF) for stream $s \in \{\mathrm{main}, \mathrm{LP}, \mathrm{HP}\}$ at time t. Since methane fractions are expressed in molar terms, the total number of moles from each inlet was estimated as $n_s(t) = \frac{d_s(t)\cdot V_s(t)}{\sum_i M_i x_{i,s}(t)}$, where $d_s(t)$ is the specific gravity, M_i is the molar mass of component i, and $x_{i,s}(t)$ is the mole fraction of component i in stream s. The total molar inflow was then computed as $n_{\mathrm{total}}(t) = n_{\mathrm{main}}(t) + n_{\mathrm{LP}}(t) + n_{\mathrm{HP}}(t)$.

The overall methane fraction was obtained as follows:

$$x_{\mathrm{CH_4}}(t) = \frac{\sum_s n_s(t) x_{\mathrm{CH_4},s}(t)}{n_{\mathrm{total}}(t)}.$$

To determine which inlet best represents the GHGRP-reported methane fraction, the annual averages of each inlet were compared against GHGRP data using MAE. The main inlet exhibited the closest agreement with the reported values.

For another one (Facility B, see below), only high-pressure (HP) inlet data were available in the proprietary dataset. This limitation likely explains the greater discrepancy (MAE of 2%) between the proprietary values and the GHGRP values (see Fig. S2).

Overall, we see close agreement between data reported to the GHGRP and high-resolution operational data supplied by our industry partner. This agreement bolsters confidence in gas processing facility gas composition data reported to the GHGRP.

S2.2 Correlation analysis between Gas-to-Oil Ratio (GOR) and component fractions

Due to the limited availability of direct gas composition measurements and gaps in data coverage across key areas, we explored the gas-to-oil ratio (GOR) as an auxiliary variable to help infer gas composition. GOR is routinely reported and reflects underlying hydrocarbon production characteristics, making it a promising proxy. We computed Pearson correlation coefficients between the log-transformed GOR and individual gas component fractions using USGS data (all basins), observing a positive correlation for methane (C1) and negative correlations for heavier hydrocarbons (see Fig. S3 and Table S1). These correlations suggest that GOR contains information about gas composition, though the relationships are not strictly linear. This means that a model that incorporates gas and oil production volumes directly may well capture this potentially non-linear dependence.

S3 Intermediate steps in the gas composition prediction workflow

S3.1 Anisentropy parameter characterization

As described in Section 4.2.4, spatio-temporal kriging requires defining an appropriate distance metric, which is adjusted using an anisotropy factor α to account for differing rates of spatial and temporal variations in gas composition. We show the impact of different α values Fig. S4, where methane fraction differences exhibit a temporal pattern when $\alpha = 1.0$ and a spatial pattern when $\alpha = 40.0$.

The optimal anisotropy factor, $\alpha^* = 6.3$, is determined by minimizing the root mean square error (RMSE) in the spatial variogram fit, as shown in Fig. S5. This value results in approximately isotropic gas composition variations.

S3.2 Component fractions before and after quantile transformation

To ensure a more Gaussian-like distribution before applying kriging, a Quantile Transformer with $n_{\text{quantiles}} = 500$ was applied to the raw gas component data. This transformation preserves rank order while mapping the original data onto a nearly standard normal distribution. Fig. S6 presents Kernel Density Estimation (KDE) plots comparing the distributions of gas component

fractions before and after transformation for two datasets: USGS data and GHGRP production data.

S3.3 Variogram results

We present the estimated variograms used for kriging across basins and gas components.

Figure S7 shows example variograms for methane (C1) in three basins—Anadarko, Permian, and Gulf Coast—based on USGS (panel a) and GHGRP production data (panel b). Each plot displays the empirical variogram along with the fitted exponential model, including the corresponding sill and range parameters.

Figures S9 and S8 display the default variograms used in cases where basin-level estimation is not feasible due to limited data. These models were computed by concatenating data from all basins with sufficient sample size (at least 50 points), with variogram calculations restricted to within-basin distances.

Tables Table S2–Table S7 summarize the estimated variogram parameters (sill and range) for all basins and components. Entries marked as N/A indicate that a basin had too few samples for reliable variogram fitting. In such cases, the default variograms from Figures S9 and S8 are used instead.

S4 Assessment of methods robustness

S4.1 Ablation study and comparison to simple baselines

In order to validate the benefit of using gas and oil production as auxiliary variables, we performed an ablation study that consisted in removing the non-linear model and only relying on kriging. To confirm the importance of kriging, we compared our method to a much simpler interpolation method (nearest neighbor). We used 4-fold cross-validation on each basin to compute the accuracy of each method. Results are shown in Figure S10.

To illustrate the structure of the holdout test sets used in this evaluation, Figure S12 presents illustrative examples of holdout set selection for three basins — Anadarko, Permian, and Gulf Coast. For each basin, one of the four holdout regions is shown. As detailed in Section 4.5.1, this procedure is applied consistently across all basins, where holdout sets are defined as circular regions, and the centers of these regions are determined using k-means clustering ($k = 4$) to ensure a spatially structured and representative layout.

In absolute terms, the results indicate that our method achieves a Mean Absolute Error (MAE) of 5.8 mol% for C1 using USGS data, 5.0% using GHGRP data, and 2.3 mol% for C2 using USGS data.

In relative terms, our method reduces the Mean Absolute Error (MAE) by 15% compared to the kriging-only approach, when applying it to USGS C1 data. Compared to the nearest-neighbor method, this relative improvement reaches 39% on average. For USGS C2 data, while the improvement over kriging is limited to less than 1%, our method achieves a substantial 20% reduction in MAE compared to nearest-neighbor interpolation. For GHGRP C1, we achieve 1% improvment compared to kriging alone, and 10% compared to neares-neighbor.

Accuracy was also analyzed as a function of the distance to the nearest point in the dataset used for kriging. The relative accuracy of our method, compared to the kriging-only approach, increases with this distance At the farthest distances, the relative improvement reaches 30% for C1 and 5% for C2 using USGS data and 17% for C1 using GHGRP data. This highlights the value added by the auxiliary variable in data-sparse regions.

Beyond predictive accuracy, we evaluate the reliability of using gas and oil production as auxiliary variables by comparing the non-linear model's uncertainty — defined as its optimal weighting factor in the variance combination (see Section 4.2.6) — to the standard deviation from kriging. The results are shown in Figure S11. For C1 using USGS data, the two uncertainty estimates are closely aligned. In contrast, for C1 using GHGRP data and C2 using USGS data, the non-linear model tends to exhibit higher uncertainty on average.

S4.2 Sensitivity of gas composition to well selection criteria

To assess the sensitivity of our method to the definition of oil and gas wells, we apply it to a filtered version of the USGS dataset that excludes samples with high proportions of non-hydrocarbon components. This allows us to examine how gas composition estimates change when redefining what qualifies as a typical production well — e.g., one that predominantly produces hydrocarbons such as methane. Following this filtering, methane (C1) content decreases from 87.5% to 86.2%. Other hydrocarbons, including ethane (C2), propane (C3), butanes (N-C4, I-C4), pentanes (N-C5, I-C5), and heavier components (C6+), exhibit slight decreases. In contrast, non-hydrocarbon components such as H2S, nitrogen (N_2) and carbon dioxide (CO_2) increase significantly. Minor species like hydrogen (H_2) show inconsistent trends within the range of uncertainty.

S5 Supplementary basin-level results

S5.1 Spatial distribution of full gas composition within the Anadarko Basin

To illustrate the output of our gas composition mapping approach beyond methane alone (shown in Section 2), Figure S14 presents the estimated molar fractions of the 15 gas components in the Anadarko Basin, based on USGS data.

S5.2 Temporal variation in methane fraction

To illustrate how our method captures temporal changes in gas composition, we plot the production-weighted average molar fractions of methane (C1), heavier hydrocarbons (C2+), and non-hydrocarbon components in two example basins. As shown in Fig. S15, the Permian Basin exhibits an increasing share of heavier hydrocarbons over time, while the Uinta Basin displays the opposite trend. This contrast may reflect a shift in the type of production within the Permian toward more oil-rich reservoirs.

S5.3 Spatial distribution of methane fraction in multiple basins

To illustrate the basin-level outputs of our method beyond the Anadarko example shown in Section 2, Figure S16 presents our estimates for methane molar fractions for the 13 basins discussed in the Section 2 section, using GHGRP production data.

S6 Supplementary Figures

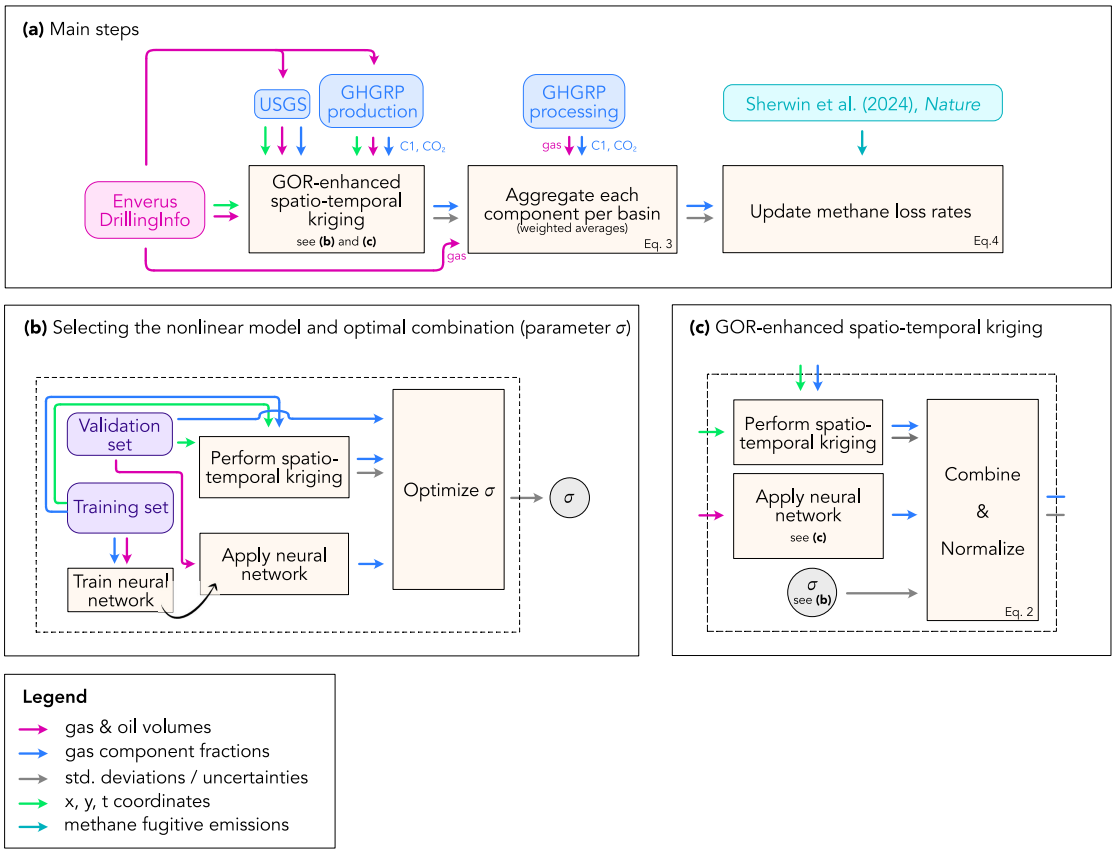

Fig. S1: **Workflow for estimating gas composition and updating production-normalized methane loss rates. (a)** Main steps: Integrates data from Enverus DrillingInfo, USGS, and GHGRP production into GOR-enhanced spatio-temporal kriging, aggregates components per basin, and updates fractional loss. **(b)** Details on training the neural network and finding σ: the initial set of gas composition samples is separated into a training set, test set and validation set. **(c)** GOR-enhanced spatio-temporal kriging: Combines kriging and neural network predictions, using the optimal weighting factor for the combination (σ). Arrows indicate data flows: gas & oil volumes (pink), gas component fractions (blue), uncertainties (grey), coordinates (green), and measured methane emissions (turquoise). The equation numbers refer to the equations presented in 4

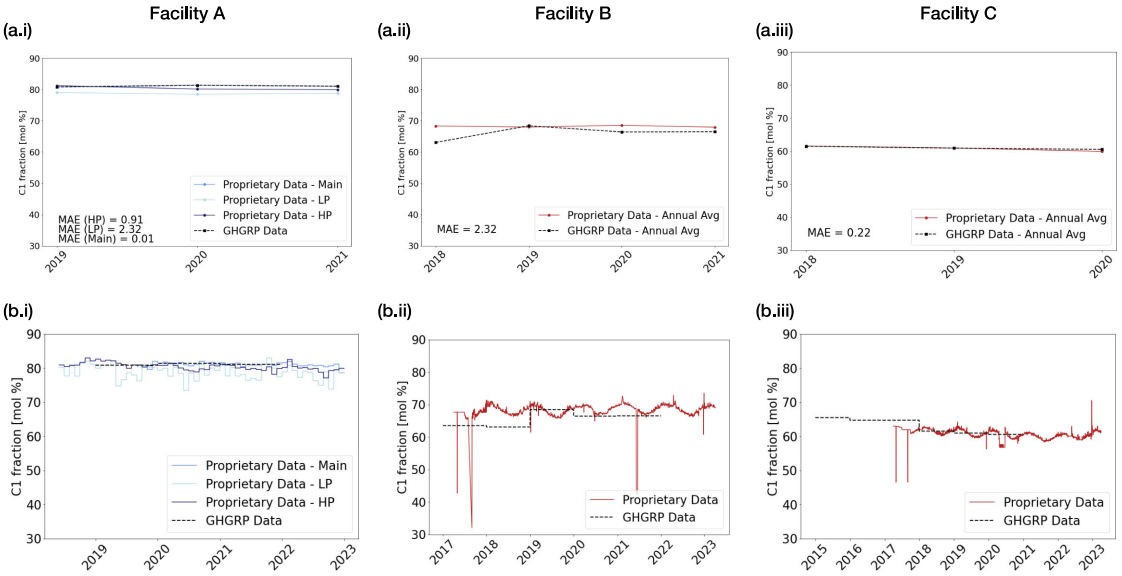

Fig. S2: **Comparison of proprietary and GHGRP-reported methane (C_1) fractions for three facilities. (a)** Annual average methane fractions for Facility A (**i**), Facility B (**ii**), and Facility C (**iii**). In **(a.i)**, methane fractions are reported separately for the main, low-pressure (LP), and high-pressure (HP) inlets, with the GHGRP data aligning most closely with the main inlet. **(a.ii)** and **(a.iii)** compare the proprietary and GHGRP annual averages for Facilities B and C, respectively, with mean absolute errors (MAE) indicated. **(b)** Time series of raw methane fraction data for Facility A (**i**), Facility B (**ii**), and Facility C (**iii**).

Component	C1	C2	C3	N-C4	I-C4	N-C5	I-C5	C6+	HE	CO2	H2	N2	H2S	AR	O2
Pearson r	0.42	-0.44	-0.38	-0.34	-0.34	-0.25	-0.22	-0.21	-0.14	-0.07	-0.04	-0.11	-0.31	-0.09	0.05

Table S1: Pearson correlation coefficients between the logarithm of the Gas-to-Oil Ratio (Log GOR) and hydrocarbon component fractions.

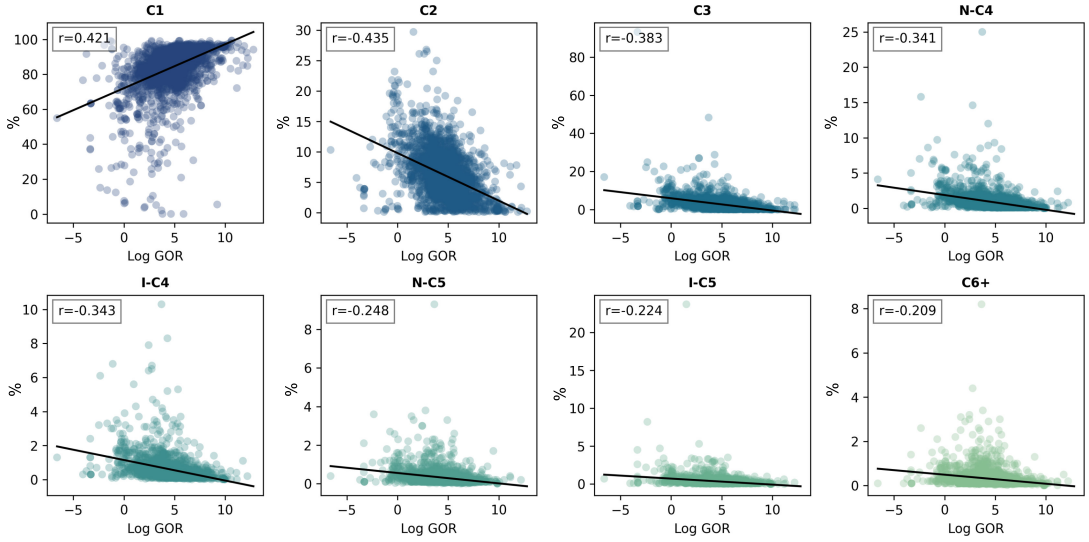

Fig. S3: **Scatter plots illustrating the correlation between Log GOR and hydrocarbon components.** The black lines represent linear regression fits, and the Pearson correlation coefficients (r) are displayed in each subplot.

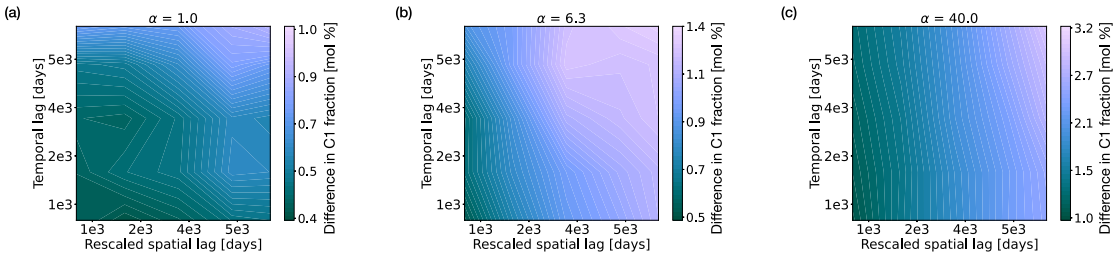

Fig. S4: **Contour maps of mean difference in methane fraction (C1) as a function of binned temporal and rescaled spatial lags for different anisotropy factors α.** The analysis was made using C1 fraction data from USGS. The spatial lag is rescaled using the factor α (indicated in each subplot). In **(a)**, where $\alpha = 1.0$, the difference primarily varies with time, showing horizontal contour lines, indicating minimal spatial dependence. In **(b)**, with $\alpha = 6.3$, the difference exhibits approximately isotropic behavior in space and time, as seen from diagonal contour lines. In **(c)**, where $\alpha = 40.0$, the difference primarily varies with spatial distance, leading to vertical contour lines, implying minimal temporal dependence.

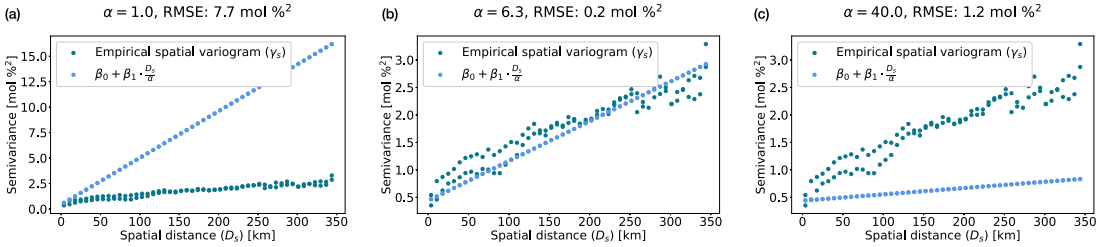

Fig. S5: **Empirical spatial variogram and temporal model fit for different anisotropy factors α.** The spatial semivariance $\gamma_S(h)$—defined as half the average squared difference in gas composition between pairs of observations separated by spatial distance h—is shown for various values of the anisotropy factor α. The curves are compared to the temporal variogram model $\beta_0 + \beta_1 \cdot \frac{D_S}{\alpha}$, where β_0 and β_1 are coefficients obtained by fitting a linear model to the temporal variogram at short lags (see Section 4.2.4). This model expresses how temporal variation is rescaled onto space using α. **(b)** shows the optimal fit obtained for $\alpha^* = 6.3$ m/day by minimizing the root mean square error (RMSE) between the empirical spatial variogram and the temporal model (Eq. (2)). Panels **(a)** and **(c)** illustrate less optimal fits for $\alpha = 1.0$ and $\alpha = 40.0$, which yield larger RMSE values.

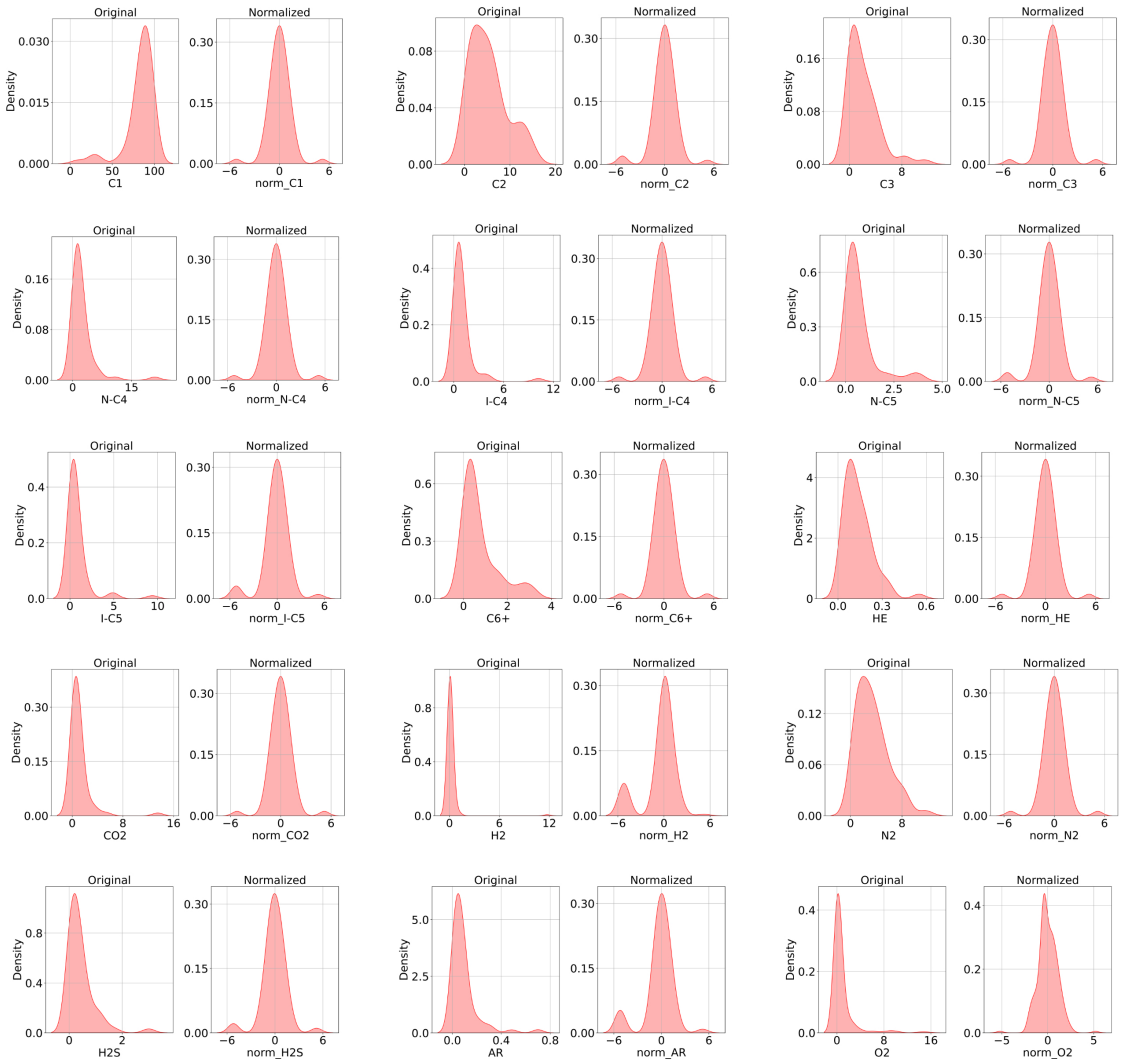

Fig. S6: **Effect of quantile transformation on gas component distributions.** Kernel Density Estimation (KDE) plots show the distribution of each gas component before (left panel of each pair) and after (right panel) quantile transformation. Each row corresponds to a different set of components, with 5 rows and 3 columns covering the 15 components: C1, C2, C3, N-C4, I-C4, N-C5, I-C5, C6+, CO2, H2, N2, H2S, AR, O2, and HE.

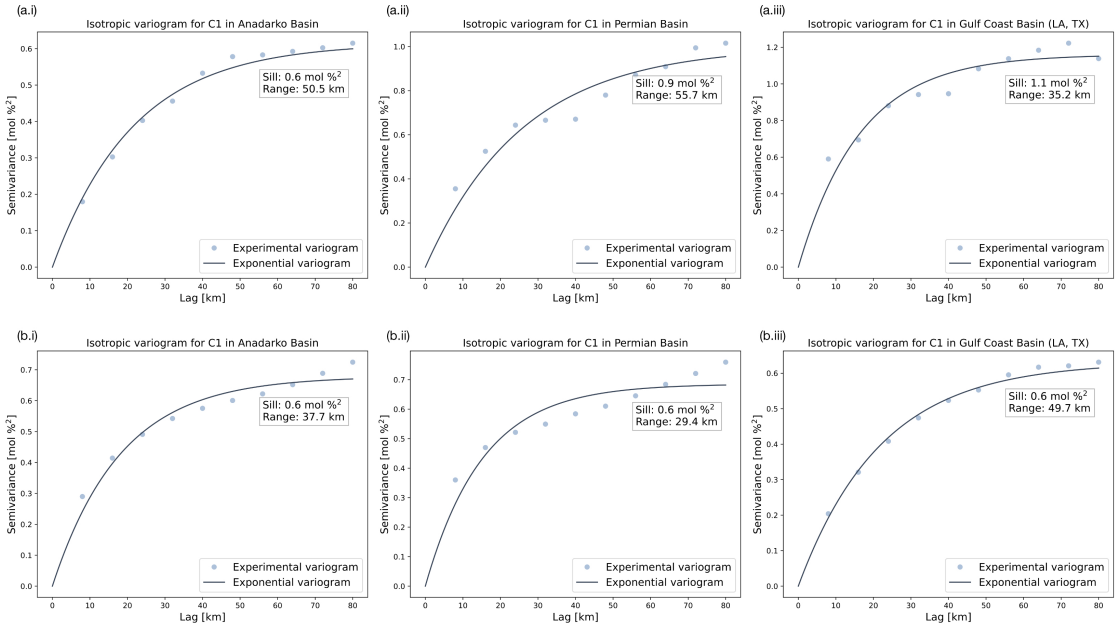

Fig. S7: **Isotropic variogram analysis for methane (C1) in three selected basins.** Panel **(a)** presents variograms derived from USGS data for the Anadarko Basin **(a.i)**, Permian Basin **(a.ii)**, and Gulf Coast Basin (LA, TX) **(a.iii)**. Panel **(b)** shows corresponding results using GHGRP production data: Anadarko **(b.i)**, Permian **(b.ii)**, and Gulf Coast **(b.iii)**. Each plot compares the experimental variogram (blue points) with the fitted exponential model (black curve), and reports the estimated sill and range values. While only these three basins are shown, accompanying tables (Table S2, Table S3, Table S4, Table S5, Table S6, Table S7) provide variogram parameters for all studied basins.

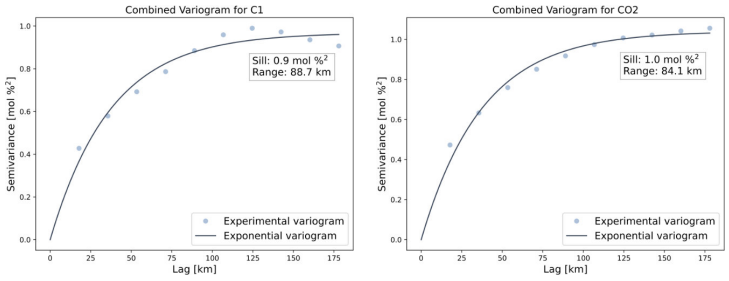

Fig. S8: **Default isotropic variograms for GHGRP production data**. Combined experimental and fitted exponential variograms for C1 and CO_2 using GHGRP production data across selected basins: Appalachian Basin, Appalachian Basin (Eastern Overthrust Area), Permian Basin, Arkla Basin, Anadarko Basin, Denver Basin, Green River Basin, Arkoma Basin, Gulf Coast Basin (LA, TX), and East Texas Basin. The empirical variogram (scatter points) and the fitted exponential model (solid line) illustrate the spatial correlation structure, with sill and range values provided for each component.

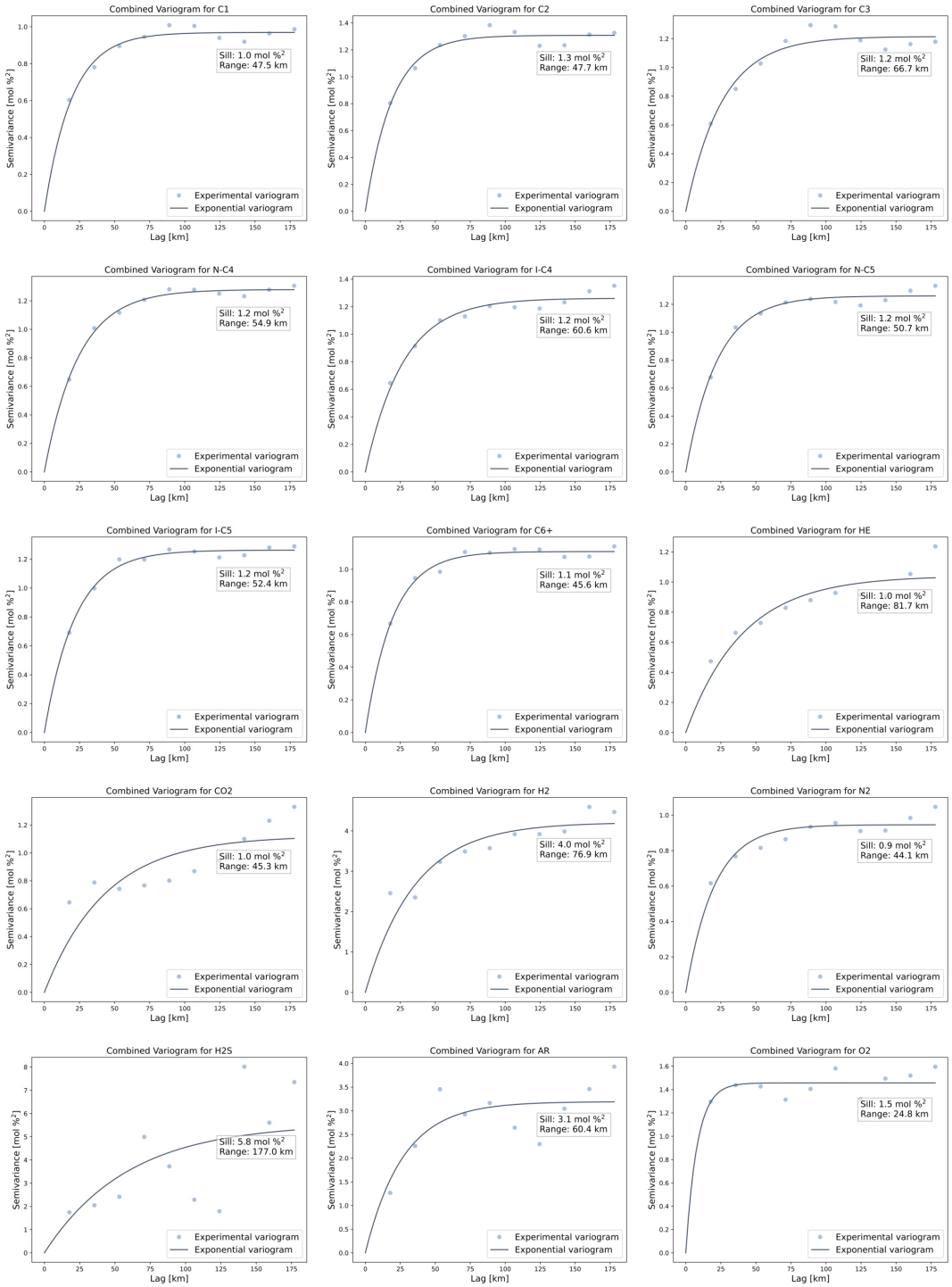

Fig. S9: **Default isotropic variograms for USGS data**. Combined experimental and fitted exponential variograms for gas component fractions using USGS data across selected basins: Appalachian Basin, Appalachian Basin (Eastern Overthrust Area), Permian Basin, Arkla Basin, Anadarko Basin, Denver Basin, Green River Basin, Arkoma Basin, Gulf Coast Basin (LA, TX), and East Texas Basin. Each subplot represents a different gas component, showing the empirical variogram (scatter points) and the fitted exponential model (solid line). The sill and range values for each component are indicated within the plots.

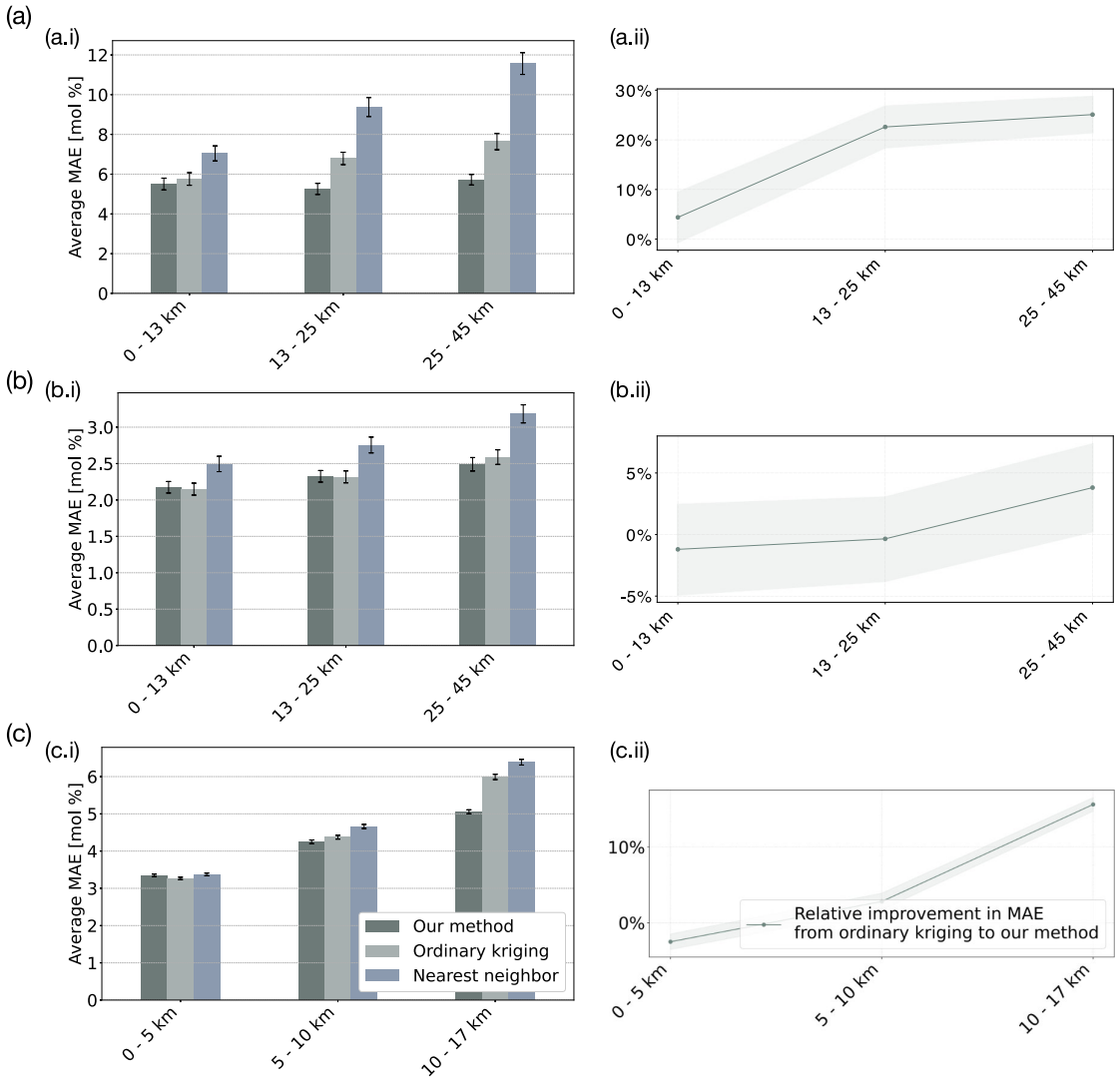

Fig. S10: **Comparison of interpolation methods for predicting C1 and C2 fractions as a function of distance.** Panels (a), (b), and (c) show results using USGS C1 data, USGS C2 data, and GHGRP C1 data, respectively. In each panel, subfigure (i) presents the absolute difference in mean absolute error (MAE) between methods, while subfigure (ii) shows the relative improvement in MAE of our method compared to ordinary kriging. Each distance bin contains 729 points for USGS (panels a and b) and 23470 points for GHGRP (panel c); error bars indicate the standard error of the mean.

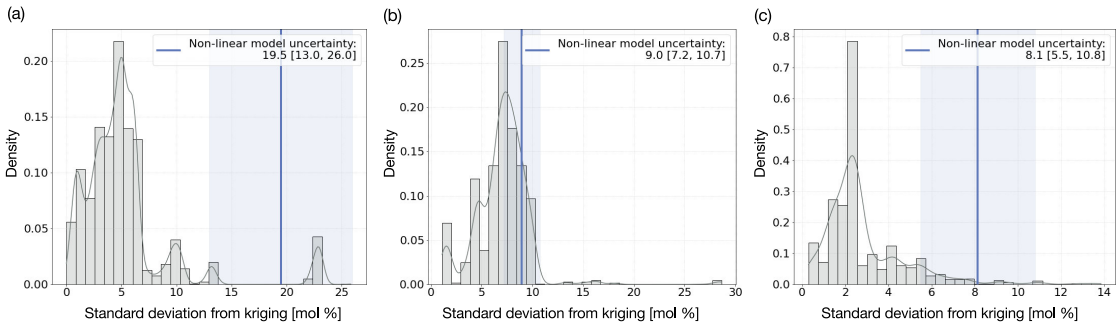

Fig. S11: **Comparison of uncertainty estimates from kriging and the non-linear model.** Panels (a), (b), and (c) show the distribution of kriging standard deviations for GHGRP C1, USGS C1, and USGS C2 data, respectively. Vertical lines indicate the average uncertainty of the non-linear model, defined as its optimal weighting factor in the combined variance (see Section 4.2.6).

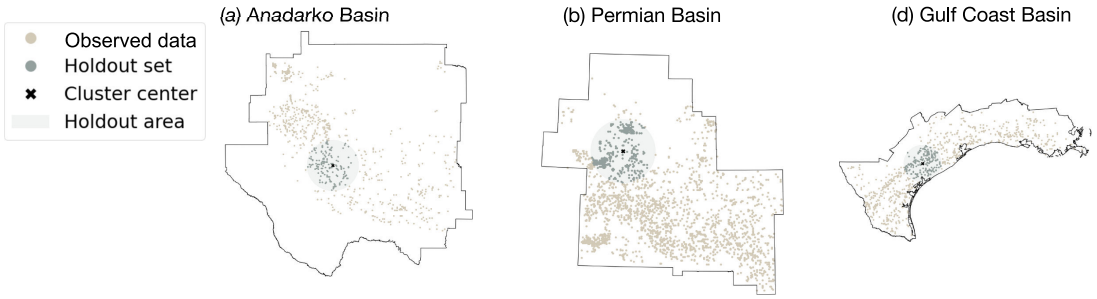

Fig. S12: **Spatial distribution of observed data and holdout sets in three basins.** The figure shows the locations of observed gas composition data in the USGS database (blue) used to fit the kriging model and train the neural network. Orange points represent locations held out from model fitting in three example basins: Anadarko, Permian, and Gulf Coast. These holdout sets are used to estimate the optimal parameter σ that combines kriging and neural network predictions. An analogous holdout procedure is applied to separate test sets used to evaluate final model performance.

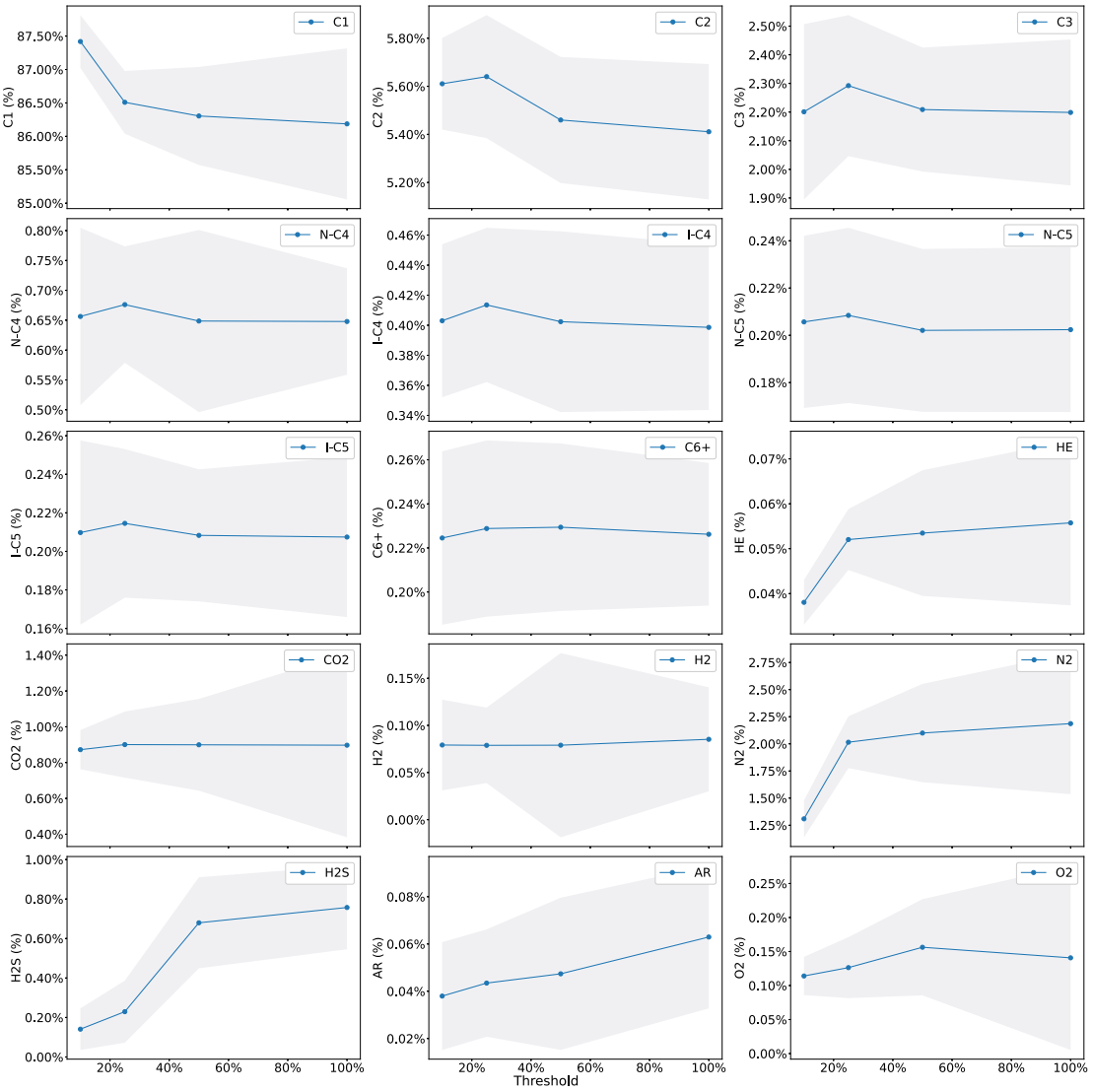

Fig. S13: **Effect of filtering thresholds on interpolated average gas composition (USGS data).** Shown are average component fractions predicted by our spatio-temporal interpolation model, weighted by gas production and grouped by filtering threshold. Thresholds correspond to the maximum allowed proportion of non-hydrocarbon gases in the original USGS samples (HE, CO_2, H_2, N_2, H_2S, AR, O_2). A threshold of 100% indicates no filtering. Shaded areas show 95% confidence intervals, estimated using standard errors from the simulation procedure described in Section 4.4.1.

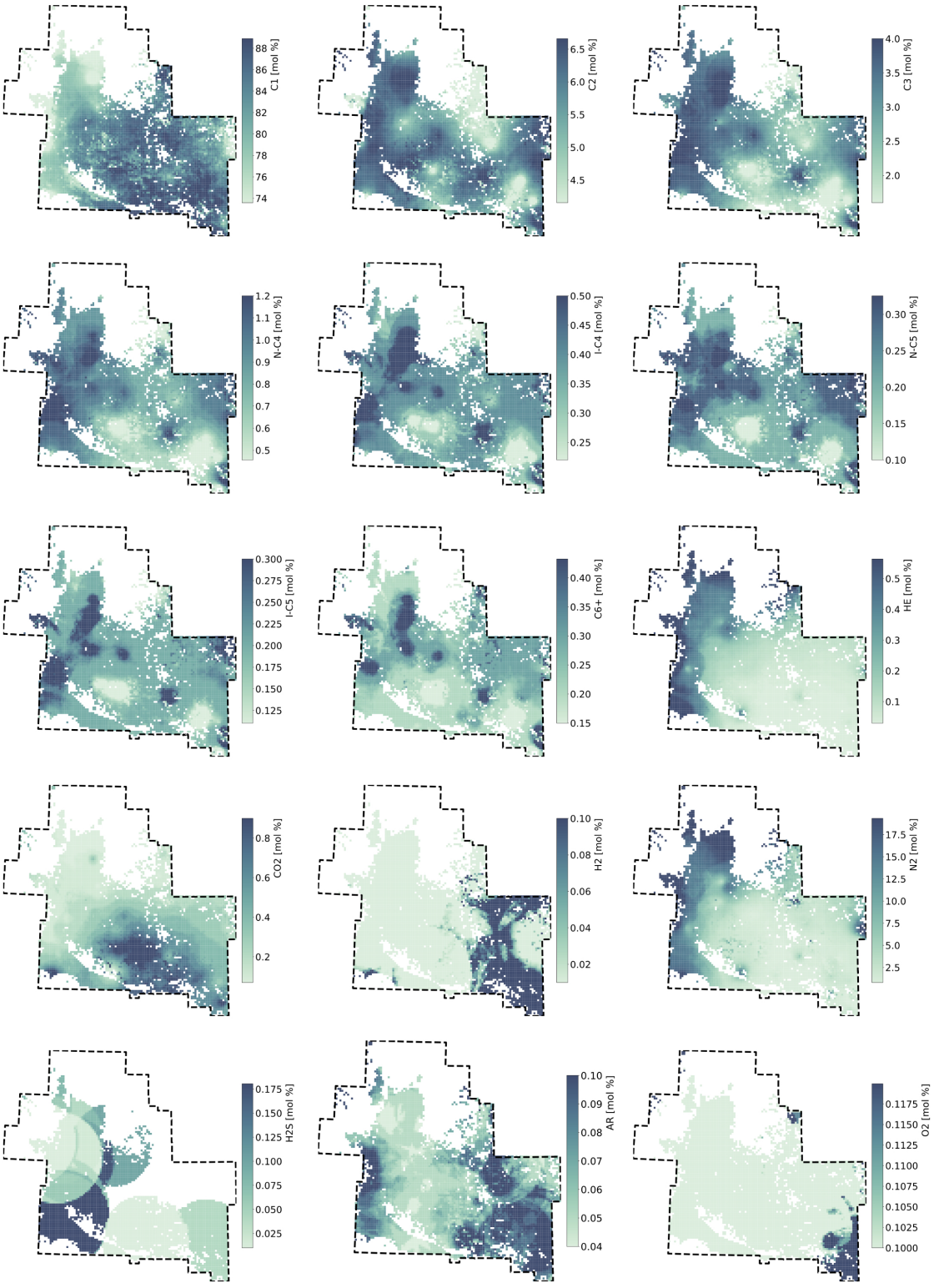

Fig. S14: **Molar fractions of 15 gas components in the Anadarko Basin from USGS data.** The figure shows the estimated molar fractions of major gas components, including methane, ethane, propane, and heavier hydrocarbons, as well as non-hydrocarbon species. These values are obtained using our spatio-temporal interpolation approach applied to USGS well sample data.

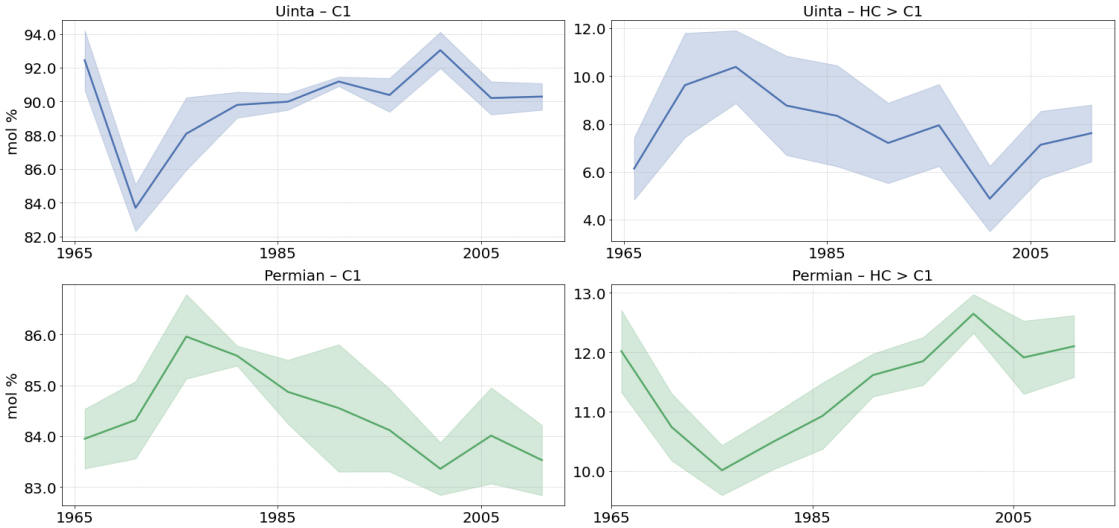

Fig. S15: **Temporal evolution of methane (C1) and heavier hydrocarbon (HC ¿ C1) molar fractions in natural gas from the Uinta and Permian Basins.** Each panel shows production-weighted mean values with associated standard errors. In the Permian Basin, methane content has declined over time, while heavier hydrocarbons increased—potentially reflecting a shift toward more liquids-rich production. In contrast, the Uinta Basin shows the opposite trend, with increasing methane and decreasing heavier hydrocarbons.

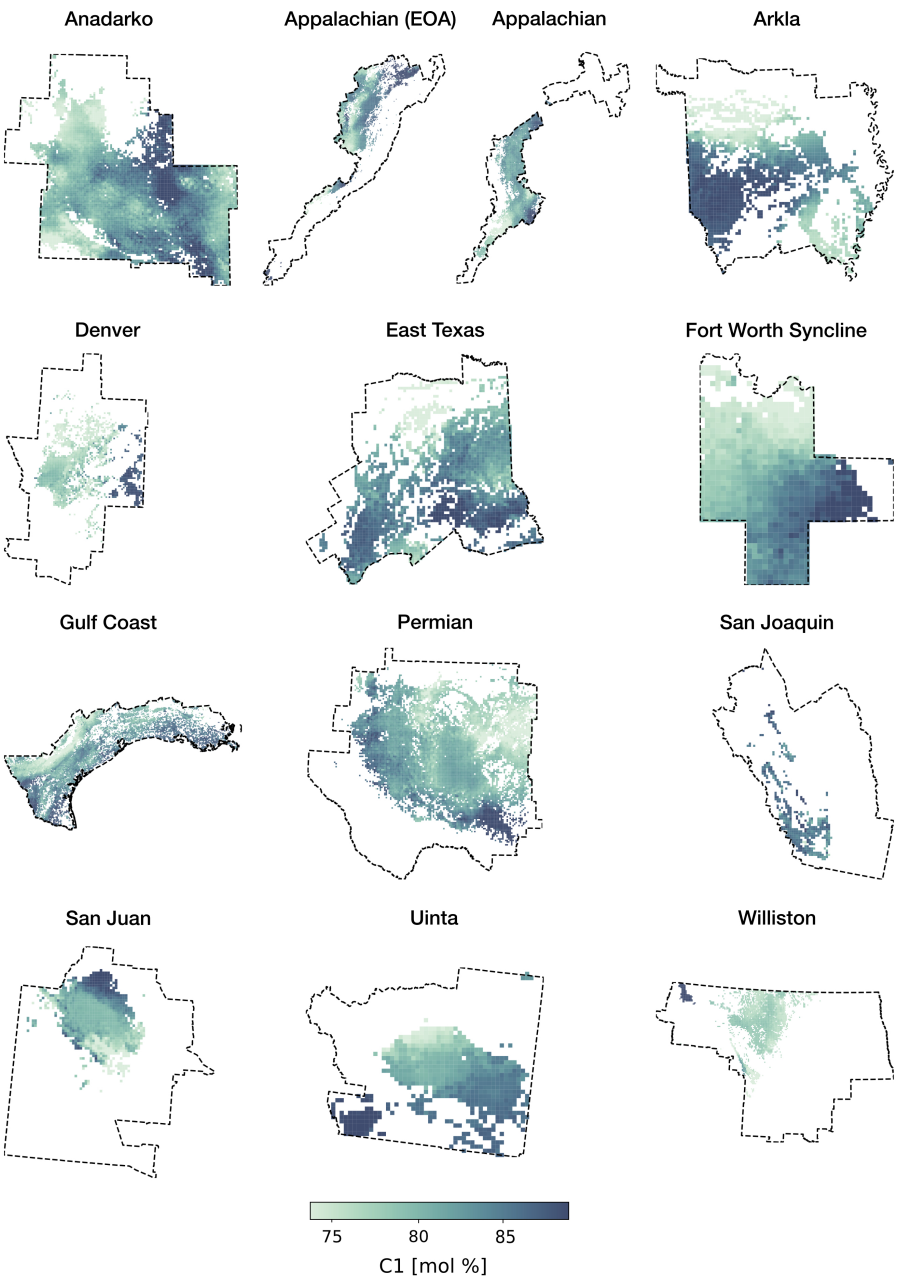

Fig. S16: **Methane molar fractions estimated using GHGRP production data.** Results are shown for 13 basins, with values predicted using our interpolation method.

Component	Metric	East Texas	Gulf Coast (LA, TX)	Denver	South Oklahoma Folded Belt	Permian	Uinta	Appalachian (Eastern Overthrust Area)	Bend Arch
AR	Sill	N/A	5.4	N/A	N/A	3.8	N/A	N/A	N/A
	Range [km]	N/A	14.8	N/A	N/A	28.3	N/A	N/A	N/A
C1	Sill	0.9	1.2	0.4	1.7	1.0	0.8	1.1	1.5
	Range [km]	54.0	49.6	80.0	80.0	78.2	24.5	80.0	59.7
C2	Sill	1.2	1.0	0.5	2.2	1.0	1.1	1.0	1.6
	Range [km]	80.0	61.8	66.9	80.0	80.0	80.0	80.0	80.0
C3	Sill	2.3	1.1	0.6	2.2	0.9	1.0	1.3	1.6
	Range [km]	39.6	71.4	0.0	79.6	80.0	14.6	80.0	51.7
C6+	Sill	1.1	0.9	0.8	1.8	1.1	1.4	N/A	4.8
	Range [km]	64.9	33.6	0.1	0.4	78.0	45.6	N/A	29.6
CO2	Sill	1.1	1.0	1.1	1.9	0.8	0.9	2.0	4.6
	Range [km]	29.7	80.0	67.5	79.6	10.0	15.8	80.0	44.1
H2	Sill	0.8	0.8	4.5	N/A	4.3	N/A	N/A	N/A
	Range [km]	79.8	0.1	16.7	N/A	0.4	N/A	N/A	N/A
H2S	Sill	N/A	N/A	N/A	N/A	1.0	N/A	N/A	N/A
	Range [km]	N/A	N/A	N/A	N/A	52.0	N/A	N/A	N/A
HE	Sill	0.9	0.8	0.6	3.1	0.5	1.0	2.9	1.4
	Range [km]	76.6	80.0	54.3	0.4	67.5	18.0	80.0	49.1
I-C4	Sill	3.5	1.0	0.8	2.1	1.1	1.0	1.2	2.2
	Range [km]	66.2	38.4	0.2	79.6	70.1	21.5	0.8	34.5
I-C5	Sill	1.3	1.0	0.9	2.3	1.1	1.5	N/A	1.6
	Range [km]	80.0	20.8	65.9	0.8	75.9	80.0	N/A	56.2
N-C4	Sill	2.9	1.0	0.8	2.0	1.0	0.9	1.6	1.5
	Range [km]	35.0	48.2	56.5	80.0	80.0	15.9	80.0	39.6
N-C5	Sill	5.6	0.8	1.3	N/A	1.2	1.2	N/A	2.4
	Range [km]	10.0	25.1	80.0	N/A	80.0	75.2	N/A	32.4
N2	Sill	1.3	1.0	0.7	1.6	0.9	1.2	1.0	1.3
	Range [km]	60.0	20.6	80.0	50.6	63.3	27.4	74.4	76.0
O2	Sill	1.6	1.4	N/A	N/A	0.8	N/A	N/A	N/A
	Range [km]	23.3	21.0	N/A	N/A	13.5	N/A	N/A	N/A

Table S2: Variogram parameters (Sill [sq. mol %] and Range [km]) per component and basin [USGS data, Part 1/3]

Component	Metric	Green River	Chautauqua Platform	Palo Duro	San Juan	Fort Worth Syncline	Las Animas Arch	Anadarko	Arkoma
AR	Sill	2.9	N/A	N/A	N/A	N/A	N/A	1.8	3.9
	Range [km]	79.9	N/A	N/A	N/A	N/A	N/A	80.0	9.9
C1	Sill	1.2	0.9	1.7	1.1	1.4	1.4	0.6	1.1
	Range [km]	52.7	0.0	73.0	48.3	80.0	80.0	65.1	9.9
C2	Sill	2.3	1.0	1.2	1.1	1.3	2.0	1.1	1.0
	Range [km]	18.6	39.0	63.7	32.9	21.2	63.0	59.4	0.1
C3	Sill	1.3	1.1	1.1	1.2	1.3	2.0	0.9	0.8
	Range [km]	55.0	45.9	51.7	34.1	61.1	24.2	57.6	14.3
C6+	Sill	1.3	1.4	1.2	1.2	1.5	2.2	0.9	N/A
	Range [km]	12.0	21.0	0.1	12.7	8.7	44.5	58.3	N/A
CO2	Sill	1.1	1.1	1.4	1.2	1.0	2.2	0.6	0.9
	Range [km]	22.4	80.0	23.8	61.9	80.0	0.0	56.1	31.9
H2	Sill	N/A	N/A	N/A	5.1	N/A	N/A	3.0	1.9
	Range [km]	N/A	N/A	N/A	80.0	N/A	N/A	80.0	33.6
H2S	Sill	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Range [km]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HE	Sill	1.8	1.1	1.7	1.1	0.9	1.6	0.5	1.1
	Range [km]	80.0	55.6	41.4	46.4	0.0	19.3	61.2	72.5
I-C4	Sill	1.3	1.2	1.4	1.0	1.4	2.1	1.0	N/A
	Range [km]	49.9	27.8	77.0	21.9	15.0	23.3	57.1	N/A
I-C5	Sill	1.4	1.2	1.2	1.1	1.5	3.2	1.0	1.5
	Range [km]	21.2	19.1	38.7	24.4	0.3	21.5	52.9	16.4
N-C4	Sill	1.2	1.1	1.3	1.1	1.4	2.0	0.9	N/A
	Range [km]	27.1	15.7	75.0	24.4	0.0	33.6	59.5	N/A
N-C5	Sill	1.7	1.2	1.0	1.3	2.3	2.4	1.0	N/A
	Range [km]	38.7	12.7	0.9	9.7	26.7	30.5	57.9	N/A
N2	Sill	1.4	1.2	1.9	1.0	1.5	1.4	0.5	1.1
	Range [km]	80.0	19.7	0.4	49.2	80.0	80.0	66.8	22.2
O2	Sill	1.9	2.4	N/A	1.9	N/A	N/A	1.4	1.3
	Range [km]	29.7	0.3	N/A	53.4	N/A	N/A	15.1	16.1

Table S3: Variogram parameters (Sill [sq. mol %] and Range [km]) per component and basin [USGS data, Part 2/3]

Component	Metric	Paradox	Arkla	Piceance	Williston	Appalachian	Powder River	Wind River	San Joaquin
AR	Sill	3.2	N/A	5.7	N/A	N/A	N/A	N/A	N/A
	Range [km]	46.2	N/A	80.0	N/A	N/A	N/A	N/A	N/A
C1	Sill	1.6	1.1	1.1	1.1	2.4	N/A	N/A	N/A
	Range [km]	79.9	63.9	80.0	79.8	79.4	N/A	N/A	N/A
C2	Sill	1.5	3.0	1.2	0.8	1.9	N/A	N/A	N/A
	Range [km]	79.9	80.0	62.9	79.8	58.3	N/A	N/A	N/A
C3	Sill	0.9	1.2	1.2	1.5	1.5	N/A	N/A	N/A
	Range [km]	79.9	39.9	70.1	76.8	79.4	N/A	N/A	N/A
C6+	Sill	1.0	1.4	1.3	N/A	N/A	N/A	N/A	N/A
	Range [km]	19.2	35.2	65.2	N/A	N/A	N/A	N/A	N/A
CO2	Sill	1.6	2.4	1.3	0.8	N/A	N/A	N/A	N/A
	Range [km]	79.9	40.7	80.0	54.6	N/A	N/A	N/A	N/A
H2	Sill	8.4	N/A	8.8	N/A	N/A	N/A	N/A	N/A
	Range [km]	58.8	N/A	57.3	N/A	N/A	N/A	N/A	N/A
H2S	Sill	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Range [km]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HE	Sill	2.7	1.4	1.2	1.5	2.1	N/A	N/A	N/A
	Range [km]	79.9	40.7	80.0	79.8	79.4	N/A	N/A	N/A
I-C4	Sill	1.1	1.3	1.1	N/A	N/A	N/A	N/A	N/A
	Range [km]	35.3	46.2	33.2	N/A	N/A	N/A	N/A	N/A
I-C5	Sill	1.0	2.0	1.0	N/A	N/A	N/A	N/A	N/A
	Range [km]	21.3	52.2	32.3	N/A	N/A	N/A	N/A	N/A
N-C4	Sill	0.7	1.7	1.5	N/A	N/A	N/A	N/A	N/A
	Range [km]	13.0	80.0	79.0	N/A	N/A	N/A	N/A	N/A
N-C5	Sill	1.0	1.8	1.5	N/A	N/A	N/A	N/A	N/A
	Range [km]	17.8	70.2	52.8	N/A	N/A	N/A	N/A	N/A
N2	Sill	1.2	2.1	1.0	2.2	2.4	N/A	N/A	N/A
	Range [km]	0.2	28.6	80.0	79.8	79.4	N/A	N/A	N/A
O2	Sill	N/A	N/A	3.5	N/A	N/A	N/A	N/A	N/A
	Range [km]	N/A	N/A	0.0	N/A	N/A	N/A	N/A	N/A

Table S4: Variogram parameters (Sill [sq. mol %] and Range [km]) per component and basin [USGS data, Part 3/3]

Component	Metric	Las Animas Arch	Paradox	Green River	Arkoma	San Joaquin	Powder River	East Texas	Williston
C1	Sill	21.7	10.2	0.9	0.6	1.2	1.1	1.1	0.6
	Range [km]	66.5	23.8	80.0	80.0	33.3	40.6	80.0	58.0
CO2	Sill	21.7	10.2	0.6	0.9	1.0	1.2	1.0	0.5
	Range [km]	66.5	23.9	55.7	76.7	30.9	49.1	51.9	44.8

Table S5: Variogram parameters (Sill [sq. mol %] and Range [km]) per component and basin [GHGRP Production data, Part 1/3]

Component	Metric	Fort Worth Syncline	Wind River	Gulf Coast (LA, TX)	San Juan	Palo Duro	Appalachian (Eastern Overthrust Area)	Chautauqua Platform	Appalachian
C1	Sill	1.3	2.0	0.6	1.1	1.7	0.7	1.0	1.0
	Range [km]	80.0	80.0	66.0	80.0	61.2	80.0	80.0	80.0
CO2	Sill	1.6	3.0	0.7	1.0	7.6	0.7	1.2	0.9
	Range [km]	80.0	80.0	55.3	80.0	37.5	58.2	0.4	52.5

Table S6: Variogram parameters (Sill [sq. mol %] and Range [km]) per component and basin [GHGRP Production data, Part 2/3]

Component	Metric	Permian	Uinta	Denver	Anadarko	South Oklahoma Folded Belt	Piceance	Arkla	Bend Arch
C1	Sill	0.7	1.4	0.9	0.7	1.1	1.2	1.0	N/A
	Range [km]	45.8	80.0	44.8	54.6	58.2	80.0	80.0	N/A
CO2	Sill	0.9	1.0	0.8	0.6	1.2	1.3	1.0	N/A
	Range [km]	54.6	70.0	80.0	71.2	46.0	80.0	80.0	N/A

Table S7: Variogram parameters (Sill [sq. mol %] and Range [km]) per component and basin [GHGRP Production data, Part 3/3]

References

[1] Intergovernmental Panel on Climate Change. (2021). Global warming potential values for methane. In *Climate Change 2021: The Physical Science Basis*. Retrieved from https://www.ipcc.ch/report/ar6/wg1/

[2] Peischl, J., Ryerson, T., Aikin, K., De Gouw, J., Gilman, J., Holloway, J., Lerner, B., Nadkarni, R., Neuman, J., Nowak, J. & Others Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions. *Journal Of Geophysical Research: Atmospheres*. **120**, 2119-2139 (2015)

[3] Karion, A., Sweeney, C., Kort, E., Shepson, P., Brewer, A., Cambaliza, M., Conley, S., Davis, K., Deng, A., Hardesty, M. & Others Aircraft-based estimate of total methane emissions from the Barnett Shale region. *Environmental Science & Technology*. **49**, 8124-8131 (2015)

[4] Barkley, Z., Lauvaux, T., Davis, K., Deng, A., Miles, N., Richardson, S., Cao, Y., Sweeney, C., Karion, A., Smith, M. & Others Quantifying methane emissions from natural gas production in north-eastern Pennsylvania. *Atmospheric Chemistry And Physics*. **17**, 13941-13966 (2017)

[5] Omara, M., Zavala-Araiza, D., Lyon, D., Hmiel, B., Roberts, K. & Hamburg, S. Methane emissions from US low production oil and natural gas well sites. *Nature Communications*. **13**, 2085 (2022)

[6] Alvarez, R., Zavala-Araiza, D., Lyon, D., Allen, D., Barkley, Z., Brandt, A., Davis, K., Herndon, S., Jacob, D., Karion, A. & Others Assessment of methane emissions from the US oil and gas supply chain. *Science*. **361**, 186-188 (2018)

[7] Sherwin, E., Rutherford, J., Zhang, Z., Chen, Y., Wetherley, E., Yakovlev, P., Berman, E., Jones, B., Cusworth, D., Thorpe, A. & Others US oil and gas system emissions from nearly one million aerial site measurements. *Nature*. **627**, 328-334 (2024)

[8] Zavala-Araiza, D., Lyon, D., Alvarez, R., Palacios, V., Harriss, R., Lan, X., Talbot, R. & Hamburg, S. Toward a functional definition of methane super-emitters: Application to natural gas production sites. *Environmental Science & Technology*. **49**, 8167-8174 (2015)

[9] Chen, Y., Sherwin, E., Berman, E., Jones, B., Gordon, M., Wetherley, E., Kort, E. & Brandt, A. Quantifying regional methane emissions in the New Mexico Permian Basin with a comprehensive aerial survey. *Environmental Science & Technology*. **56**, 4317-4323 (2022)

[10] Cardoso-Saldaña, F., Pierce, K., Chen, Q., Kimura, Y. & Allen, D. A searchable database for prediction of emission compositions from upstream oil and gas sources. *Environmental Science & Technology*. **55**, 3210-3218 (2021)

[11] Smith, M., Gvakharia, A., Kort, E., Sweeney, C., Conley, S., Faloona, I., Newberger, T., Schnell, R., Schwietzke, S. & Wolter, S. Airborne quantification of methane emissions over the four corners region. *Environmental Science & Technology*. **51**, 5832-5837 (2017)

[12] Schwietzke, S., Pétron, G., Conley, S., Pickering, C., Mielke-Maday, I., Dlugokencky, E., Tans, P., Vaughn, T., Bell, C., Zimmerle, D. & Others Improved mechanistic understanding of

natural gas methane emissions from spatially resolved aircraft measurements. *Environmental Science & Technology*. **51**, 7286-7294 (2017)

[13] Peischl, J., Karion, A., Sweeney, C., Kort, E., Smith, M., Brandt, A., Yeskoo, T., Aikin, K., Conley, S., Gvakharia, A. & Others Quantifying atmospheric methane emissions from oil and natural gas production in the Bakken shale region of North Dakota. *Journal Of Geophysical Research: Atmospheres*. **121**, 6101-6111 (2016)

[14] Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B., Newberger, T., Wolter, S., Banta, R. & Others Methane emissions estimate from airborne measurements over a western United States natural gas field. *Geophysical Research Letters*. **40**, 4393-4397 (2013)

[15] Pétron, G., Karion, A., Sweeney, C., Miller, B., Montzka, S., Frost, G., Trainer, M., Tans, P., Andrews, A., Kofler, J. & Others A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin. *Journal Of Geophysical Research: Atmospheres*. **119**, 6836-6852 (2014)

[16] Zumberge, J., Ferworn, K. & Brown, S. Isotopic reversal ('rollover') in shale gases produced from the Mississippian Barnett and Fayetteville formations. *Marine And Petroleum Geology*. **31**, 43-52 (2012)

[17] Jackson, R., Vengosh, A., Darrah, T., Warner, N., Down, A., Poreda, R., Osborn, S., Zhao, K. & Karr, J. Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. *Proceedings Of The National Academy Of Sciences*. **110**, 11250-11255 (2013)

[18] Zavala-Araiza, D., Sullivan, D. & Allen, D. Atmospheric hydrocarbon emissions and concentrations in the Barnett Shale natural gas production region. *Environmental Science & Technology*. **48**, 5314-5321 (2014)

[19] Allen, D., Torres, V., Thomas, J., Sullivan, D., Harrison, M. & Others Measurements of Methane Emissions at Natural Gas Production Sites: Study Appendices and Database. (University of Texas at Autin: Austin, TX. http://dept. ceer. utexas. edu, 2013)

[20] Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Petron, G., Jordaan, S. M., Tans, P., Wilcox, J., Gopstein, A. M., Arent, D., Wofsy, S., Brown, N. J., Bradley, R., Stucky, G. D., Eardley, D., Harriss, R. (2014). Methane Leaks from North American Natural Gas Systems. *Science*, 343(6172), 733-735.

[21] Brandt, A., Yeskoo, T., McNally, S., Vafi, K., Cai, H. & Wang, M. Energy Intensity and Greenhouse Gas Emissions from Crude Oil Production in the Bakken Formation: Input Data and Analysis Methods. (Department of Energy, Argonne National Laboratory. https://greet.anl.gov/publication-bakken-oil,2015)

[22] Pétron, G., Frost, G., Miller, B., Hirsch, A., Montzka, S., Karion, A., Trainer, M., Sweeney, C., Andrews, A., Miller, L. & Others Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study. *Journal Of Geophysical Research: Atmospheres*. **117** (2012)

[23] Duren, R., Thorpe, A., Foster, K., Rafiq, T., Hopkins, F., Yadav, V., Bue, B., Thompson, D., Conley, S., Colombi, N., Frankenberg, C., McCubbin, I., Eastwood, M., Falk, M., Herner, J., Croes, B., Green, R. & Miller, C. California's methane super-emitters. *Nature*. **575**, 180-184 (2019,11), https://doi.org/10.1038/s41586-019-1720-3

[24] Peischl, J., Eilerman, S., Neuman, J., Aikin, K., De Gouw, J., Gilman, J., Herndon, S., Nadkarni, R., Trainer, M., Warneke, C. & Others Quantifying methane and ethane emissions to the atmosphere from central and western US oil and natural gas production regions. *Journal Of Geophysical Research: Atmospheres*. **123**, 7725-7740 (2018)

[25] Cusworth, D., Duren, R., Thorpe, A., Olson-Duvall, W., Heckler, J., Chapman, J., Eastwood, M., Helmlinger, M., Green, R., Asner, G. & Others Intermittency of large methane emitters in the Permian Basin. *Environmental Science & Technology Letters*. **8**, 567-573 (2021)

[26] Schneising, O., Buchwitz, M., Reuter, M., Vanselow, S., Bovensmann, H. & Burrows, J. Remote sensing of methane leakage from natural gas and petroleum systems revisited. *Atmospheric Chemistry And Physics*. **20**, 9169-9182 (2020)

[27] Rodriguez, N. & Philp, R. Geochemical characterization of gases from the Mississippian Barnett shale, Fort Worth basin, Texas. *AAPG Bulletin*. **94**, 1641-1656 (2010)

[28] Hill, R., Jarvie, D., Zumberge, J., Henry, M. & Pollastro, R. Oil and gas geochemistry and petroleum systems of the Fort Worth Basin. *AAPG Bulletin*. **91**, 445-473 (2007)

[29] Sherwood, O., Schwietzke, S., Arling, V. & Etiope, G. Global inventory of gas geochemistry data from fossil fuel, microbial and burning sources, version 2017. *Earth System Science Data*. **9**, 639-656 (2017)

[30] Zhang, T., Sun, X., Milliken, K., Ruppel, S. & Enriquez, D. Empirical relationship between gas composition and thermal maturity in Eagle Ford Shale, south Texas. *AAPG Bulletin*. **101**, 1277-1307 (2017)

[31] Bowles Jr, E. Effect of shale plays on US natural gas composition. *Pipeline & Gas Journal*. **241**, 38-41 (2014)

[32] George, D. & Bowles, E. Shale gas measurement and associated issues. *Pipeline & Gas Journal*. **238**, 38-41 (2011)

[33] Pring, M. Condensate Tank Oil and Gas Activities. *Final Report*. (2012)

[34] Lange, B., Pring, M. & Treimel, S. Specified oil & gas well activities emissions inventory update. (Eastern Research Group, Inc., 2014), https://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/ei/5821199776FY1426-20140801-erg-oil_gas_ei_update.pdf, Retrieved from Texas Commission on Environmental Quality website.

[35] Conder, M. & Lawlor, K. Production characteristics of liquids-rich resource plays challenge facility design. *The American Oil & Gas Reporter*. (2014), http://www.aogr.com/magazine/sneak-peek-preview/production-characteristics-of-liquids-rich-resource-plays-facilitydesign, Accessed online

[36] Barkley, Z., Davis, K., Feng, S., Balashov, N., Fried, A., DiGangi, J., Choi, Y. & Halliday, H. Forward modeling and optimization of methane emissions in the South Central United States using aircraft transects across frontal boundaries. *Geophysical Research Letters*. **46**, 13564-13573 (2019)

[37] Barkley, Z., Davis, K., Feng, S., Cui, Y., Fried, A., Weibring, P., Richter, D., Walega, J., Miller, S., Eckl, M. & Others Analysis of oil and gas ethane and methane emissions in the southcentral and eastern United States using four seasons of continuous aircraft ethane measurements. *Journal Of Geophysical Research: Atmospheres*. **126**, e2020JD034194 (2021)

[38] Barkley, Z., Lauvaux, T., Davis, K., Deng, A., Fried, A., Weibring, P., Richter, D., Walega, J., DiGangi, J., Ehrman, S. & Others Estimating methane emissions from underground coal and natural gas production in southwestern Pennsylvania. *Geophysical Research Letters*. **46**, 4531-4540 (2019)

[39] Caulton, D., Lu, J., Lane, H., Buchholz, B., Fitts, J., Golston, L., Guo, X., Li, Q., McSpiritt, J., Pan, D. & Others Importance of superemitter natural gas well pads in the Marcellus shale. *Environmental Science & Technology*. **53**, 4747-4754 (2019)

[40] Barkley, Z., Davis, K., Miles, N., Richardson, S., Deng, A., Hmiel, B., Lyon, D. & Lauvaux, T. Quantification of oil and gas methane emissions in the Delaware and Marcellus basins using a network of continuous tower-based measurements. *Atmospheric Chemistry And Physics Discussions*. **2022** pp. 1-25 (2022)

[41] Omara, M., Sullivan, M., Li, X., Subramanian, R., Robinson, A. & Presto, A. Methane emissions from conventional and unconventional natural gas production sites in the Marcellus Shale Basin. *Environmental Science & Technology*. **50**, 2099-2107 (2016)

[42] Brennan, S., Rivera, J., Creitz, R., Varela, B. & Park, A. Natural Gas Compositional Analyses Dataset of Gases from United States Wells. (U.S. Geological Survey data release,2021), https://doi.org/10.5066/P9TR93E3

[43] Environmental Protection Agency Greenhouse Gas Reporting Program. (https://www.epa.gov/ghgreporting, 2023), Accessed: 2023-05-08

[44] Enverus Enverus Homepage. (https://www.enverus.com, 2023), Accessed: 2023-05-08

[45] Hodnebrog, Ø., Dalsøren, S. & Myhre, G. Lifetimes, direct and indirect radiative forcing, and global warming potentials of ethane (C2H6), propane (C3H8), and butane (C4H10). *Atmospheric Science Letters*. **19**, e804 (2018)

[46] MacKie, E., Field, M., Wang, L., Yin, Z., Schoedl, N., Hibbs, M. & Zhang, A. GStatSim V1. 0: a Python package for geostatistical interpolation and conditional simulation. *Geoscientific Model Development*. **16**, 3765-3783 (2023)

[47] West Virginia Department of Environmental Protection API Well Number Explanation. (https://dep.wv.gov/oil-and-gas/GI/API_Explanation/Pages/default.aspx), Accessed: 2024-05-15

[48] International Energy Agency. (2024). *Global Methane Tracker 2024*. Retrieved from https://www.iea.org/reports/global-methane-tracker-2024

[49] World Bank, *Methane Explained*, 2024. Available: https://www.worldbank.org/en/programs/gasflaringreduction/methane-explained

[50] Edzer Pebesma and Benedikt Graeler. *gstat: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation*. R package version 2.1-2, 2024. Available at: https://github.com/r-spatial/gstat/.

[51] U.S. Energy Information Administration. *Natural Gas Processing Plants*. Available at: https://atlas.eia.gov/datasets/eia::natural-gas-processing-plants-2/explore.

[52] U.S. Environmental Protection Agency. *Greenhouse Gas Reporting Program (GHGRP) Query Builder: Subpart W - Petroleum and Natural Gas Systems (Reporting Years 2015-2023)*. Available at: https://enviro.epa.gov/query-builder/ghg/SUBPART%20W%20-%20PETROLEUM%20AND%20NATURAL%20GAS%20SYSTEMS%20(RY%202015-2023)/.

[53] Enverus. *DrillingInfo Gallery*. Available at: https://app.drillinginfo.com/gallery/.

[54] U.S. Environmental Protection Agency. *GHGRP Facility Information Query Builder*. Available at: https://enviro.epa.gov/query-builder/ghg/FACILITY%20INFORMATION/RLPS_GHG_EMITTER_FACILITIES.

[55] U.S. Environmental Protection Agency. *Geologic Basin Boundaries (GHGRP GIS Layer)*. Available at: https://catalog.data.gov/dataset/geologic-basin-boundaries-basins_ghgrp-gis-layer9.

[56] U.S. Census Bureau. *Cartographic Boundary Files*. Available at: https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html.

[57] Smedley, E.M., Hackley, P.C., Warwick, P.D., and Baldassare, F.J. (2022). *Natural gas compositional analyses dataset for gases from United States wells*. U.S. Geological Survey data release. Available at: https://www.usgs.gov/data/natural-gas-compositional-analyses-dataset-gases-united-states-wells.

[58] Doe, J., & Smith, A. (2023). Gas–oil ratio correlations with hydrocarbon composition in natural gas reservoirs. *Journal of Marine Science and Engineering*, *11*(10), 1891. Available: https://www.mdpi.com/2077-1312/11/10/1891

[59] U.S. Environmental Protection Agency (EPA). Volatile Organic Compounds' Impact on Indoor Air Quality. Available at: https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality. Accessed 2023.

[60] European Environment Agency (EEA). Emissions of Air Pollutants in the UK: Non-Methane Volatile Organic Compounds (NMVOCs). Available at: https://www.gov.uk/government/statistics/emissions-of-air-pollutants/emissions-of-air-pollutants-in-the-uk-non-methane-volatile-organic-compounds-nmvocs.

Accessed 2023.

[61] Collins, W. J., Derwent, R. G., Johnson, C. E., & Stevenson, D. S. (2002). The oxidation of organic compounds in the troposphere and their global warming potentials. *Climate Dynamics*, *18*(1), 71–84. Available at: https://doi.org/10.1007/s00382-001-0205-3.

[62] Fry, M. M., Shindell, D. T., & Wagner, L. E. (2014). The influence of ozone precursor emissions from four world regions on tropospheric composition and radiative climate forcing. *Journal of Geophysical Research: Atmospheres*, *119*(2), 1390–1412. Available at: https://doi.org/10.1002/2013JD020641.

[63] Hodnebrog, Ø., Myhre, G., & Samset, B. H. (2018). How shorter-lived climate forcers affect climate projections. *Nature Climate Change*, *8*(2), 95–100. Available at: https://doi.org/10.1038/s41558-017-0057-2.

[64] Texas Railroad Commission, "Oil and Gas Well Records," Available at: https://rrcsearch3.neubus.com/esd3-rrc/index.php?_module_=esd&_action_=keysearch&profile=17, accessed March 4, 2025.

[65] Intergovernmental Panel on Climate Change (IPCC). (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report. Available at: https://www.ipcc.ch/report/ar6/wg1/.

[66] Klaus Radunsky and Michael Gillenwater. Chapter 7: Precursors and Indirect Emissions. In *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Volume 1: General Guidance and Reporting*, Intergovernmental Panel on Climate Change (IPCC), 2019. https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/1_Volume1/19R_V1_Ch07_Precursors_Indirect.pdf.

[67] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019.

[68] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015. [Online]. Available: https://arxiv.org/abs/1412.6980

[69] U.S. Geological Survey. *AAPG Geologic Provinces*. Available at: https://ngmdb.usgs.gov/Geolex/stratres/provinces. Accessed April 2025.