

Efficient Predictions of Global Free Gas and Gas Hydrate Formation using K-means Clustering

William Karl Eymold¹

Jennifer Mary Frederick¹, Michael Nole¹, Benjamin J Phrampus², Warren T Wood²

¹Sandia National Laboratories, ²Naval Research Laboratory

Abstract

Significant quantities of free gas and gas hydrate contained beneath the seafloor are crucial for climate modeling, carbon budget estimates, and determining acoustic velocity of seafloor sediment. While free gas and gas hydrate have been discovered at various global locations, the availability of geophysical data required for accurate predictions of their occurrence is sufficient in areas of active oil or gas production but remains scarce or absent in critical regions such as the Arctic. To compensate for the variations in data coverage, researchers at the Naval Research Laboratory used machine learning techniques to extend geophysical information from previously studied regions to poorly constrained areas to produce the Global Predictive Seabed Model (GPSM). We have developed a workflow that couples Dakota to PFLOTRAN to probabilistically predict free gas and gas hydrate occurrence. Dakota uses Latin hypercube sampling of the GPSM values and their uncertainties to determine distributions which are used as PFLOTRAN input parameters to simulate methanogenesis and predict hydrate and gas formation. We apply k-means clustering to the GPSM data from a study area of ~24,000 offshore locations between Svalbard and Norway (10°E - 30°E, 70°N - 80°N) to determine a subset of simplified clusters characterized by similarities in sedimentation rate, TOC, heat flux, temperature, and depth. Every region is described by a set of means and standard deviations for these parameters that are sampled on by Dakota to generate input decks for PFLOTRAN simulations. We ran 500 simulations for each cluster and map the probabilities of free gas and gas hydrate formation to their corresponding geographic regions. To verify the process, we also ran 50 simulations at all offshore locations in the study area and find strong agreement for free gas ($r=0.967$) and gas hydrate ($r=0.947$) formation rates from both the k-means and individual simulations. Both simulation methods predict elevated formation rates of free gas and gas hydrate in shallow regions between Svalbard and Norway. The k-means technique was then extended to the full GPSM dataset to make probabilistic predictions of global occurrence. This efficient technique provides preliminary predictions that identify important regions of gas and hydrate accumulation in seafloor sediment.

Introduction

- Presence of gas and hydrate in seafloor sediments critically affects economic, national security, and environmental predictions
 - 1-2% of gas dramatically reduces acoustic velocity of sediment (Sahoo et al., 2019)
- Global predictions of seafloor geophysical characteristics remain sparse but have been predicted as part of the Global Predictive Seabed Model (GPSM) (Lee et al., 2019)
- Coupling Dakota-PFLOTRAN has allowed us to probabilistically predict the occurrence of free gas and gas hydrate in Blake Ridge (Eymold et al., 2021)
 - By sampling on TOC, V, Q, depth, and temperature, we run a suite of methanogenesis simulations
 - At 5 arcmin spacing, 6,148,836 locations need to be simulated for the entire model space which is computationally intensive
- Clustering the study area allows us to simulate a dimensionally reduced representation in a much shorter time span
 - This process was tested in Svalbard and extended to the full global map
 - This novel approach can be applied to any simulation design and demonstrates good agreement with the full simulation technique

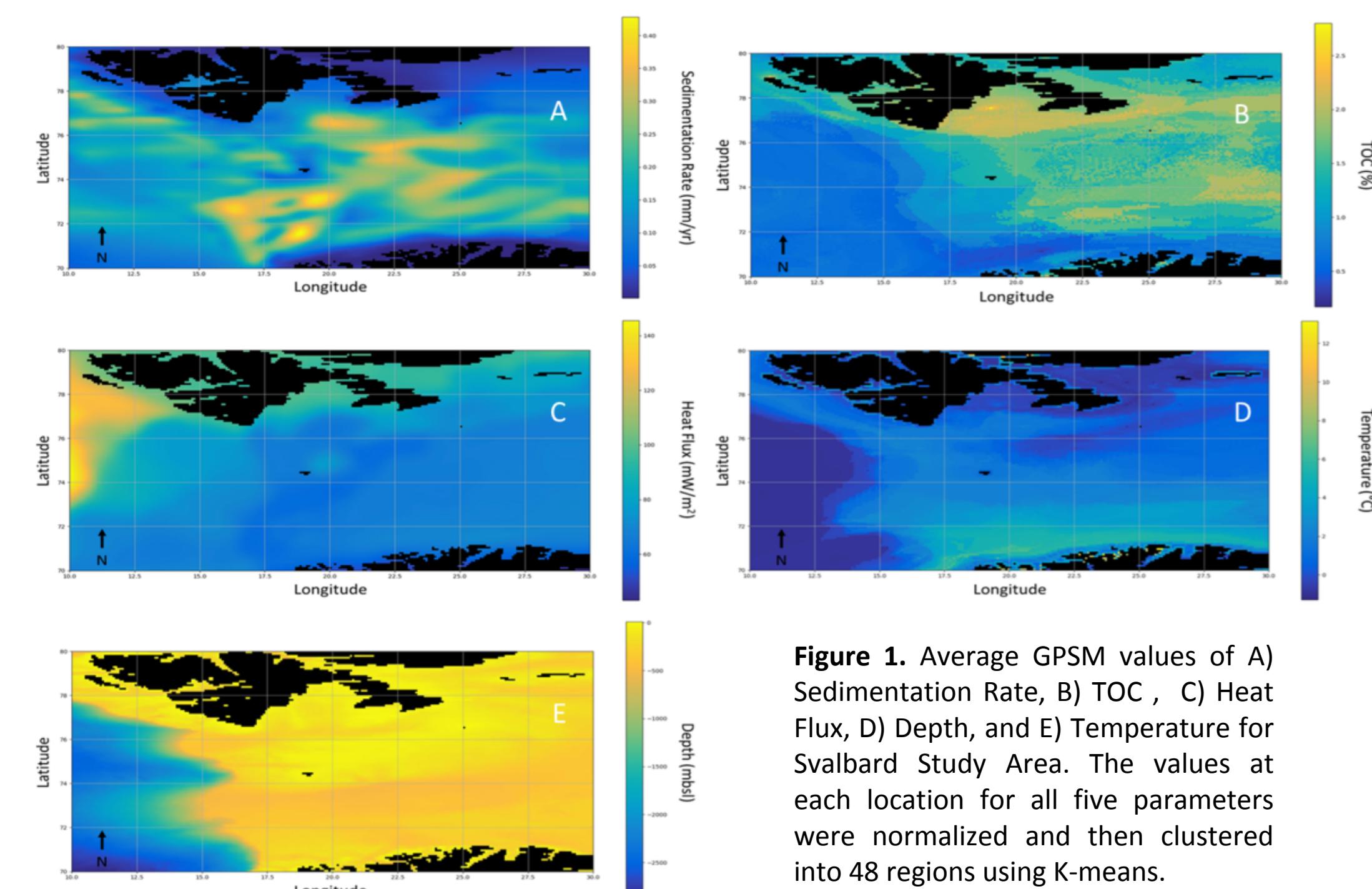


Figure 1. Average GPSM values of A) Sedimentation Rate, B) TOC, C) Heat Flux, D) Depth, and E) Temperature for Svalbard Study Area. The values at each location for all five parameters were normalized and then clustered into 48 regions using K-means.

References

- Etiope, G., G. Ciotoli, S. Schwietzke, and M. Schoell (2019), Gridded maps of geological methane emissions and their isotopic signature, *Earth System Science Data*, 11(1), 1-22.
- Eymold, W. K., J. M. Frederick, M. Nole, B. J. Phrampus, and W. T. Wood (2021), Prediction of Gas Hydrate Formation at Blake Ridge Using Machine Learning and Probabilistic Reservoir Simulation, *Geochemistry, Geophysics, Geosystems*, 22(4), e2020GC009574.
- Graves, C. A., R. H. James, C. J. Sapart, A. W. Stott, I. C. Wright, C. Berndt, G. K. Westbrook, and D. P. Connelly (2017), Methane in shallow subsurface sediments at the landward limit of the gas hydrate stability zone offshore western Svalbard, *Geochimica et Cosmochimica Acta*, 198, 419-438.
- Lee, T. R., W. T. Wood, and B. J. Phrampus (2019), A machine learning (KNN) approach to predicting global seafloor total organic carbon, *Global Biogeochemical Cycles*, 33(1), 37-46.
- Sahoo, S. K., L. J. North, H. Marín-Moreno, T. A. Minshull, and A. I. Best (2019), Laboratory observations of frequency-dependent ultrasonic P-wave velocity and attenuation during methane hydrate formation in Berea sandstone, *Geophysical Journal International*, 219(1), 713-723.

Methods

Clustered the full model (Svalbard or Global) of GPSM data into 48 regions

- All clustering was conducted via Python3 using Scikit-Learn
- Geophysical parameters were first normalized then used as features for k-means clustering
- Cluster output defined the mean (μ) and uncertainty (σ) for sedimentation rate, TOC, heat flux, depth, and temperature for each cluster's centroid

Dakota-PFLOTRAN workflow statistically sampled on the μ and σ for all five parameters to generate probabilistic simulations for each centroid or location

- Formation probability is defined as simulation in which ANY gas or hydrate formed within 1 Myr
- 500 simulations were run for each k-means centroid
- 50 simulations were run at each location in the Svalbard study area and for 4 Global regions

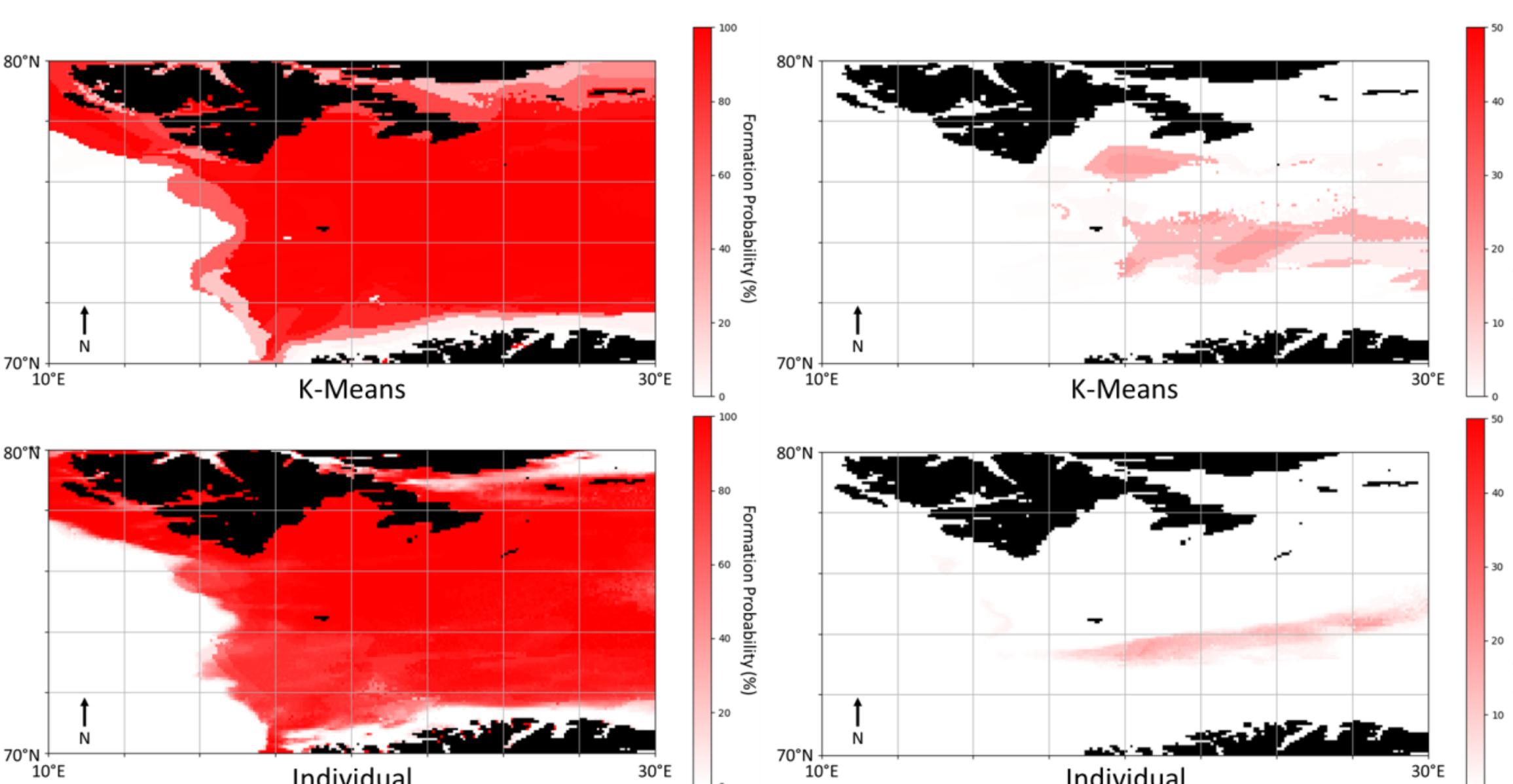


Figure 2. Maps of formation probability of gas (left) and hydrate (right) for Svalbard study area using K-means (top) and Individual simulations (bottom)

Figure 3. Individual formation probability versus K-means formation probability of A) gas and B) hydrate for Svalbard study area (gray circles) and four global clusters (squares) shown with colors matching those used in Figure 4). Dashed black lines represent $\pm 20\%$. Note that hydrate formation range was reduced to show many clusters do not form hydrate and the maximum formation probability was only 19.64%.

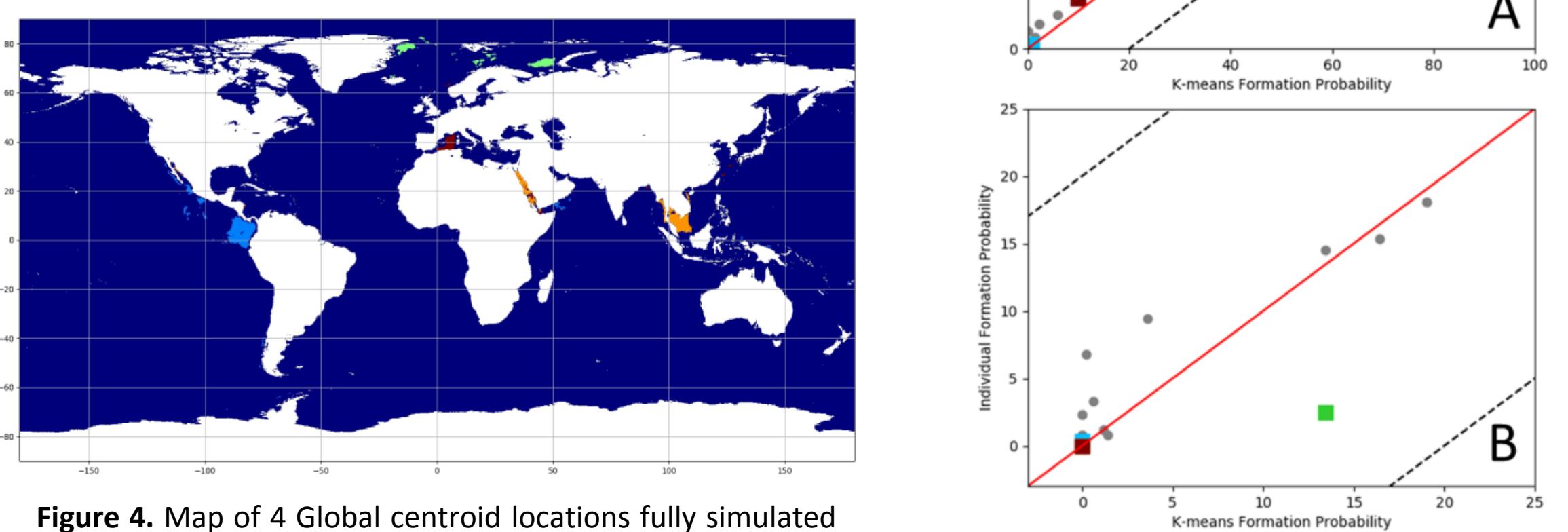


Figure 4. Map of 4 Global centroid locations fully simulated to compare to K-means predicted shown with color correlated symbols in Figure 3.

Results

General

- Clustering and simulation routine runs in ~8 hours on 48 processors compared to computational run time for all individual simulations of ~17 days on 48 processors
- Gas formation is less sensitive to generalizations of k-means centroids than hydrate formation

Svalbard

- Gas forms in almost all locations where depth <500m, higher TOC based on both k-means and individual simulations (Figure 2, left)
 - Landward limit of GHSZ is ~400 m west of Svalbard (Graves et al 2017)
- Hydrate is predicted to form in deeper areas between Svalbard and Norway at lower probabilities than gas formation (Figure 2, right)
- Strong agreement for gas ($r=0.96$) and hydrate formation probabilities ($r=0.94$) (Figure 3)

Global

- K-means consistently over-predicts gas formation in clusters simulated (Figure 4)
- Areas where gas probability is elevated (Figure 5) coincide with known seafloor seeps (c.f., Etiope et al. 2019)

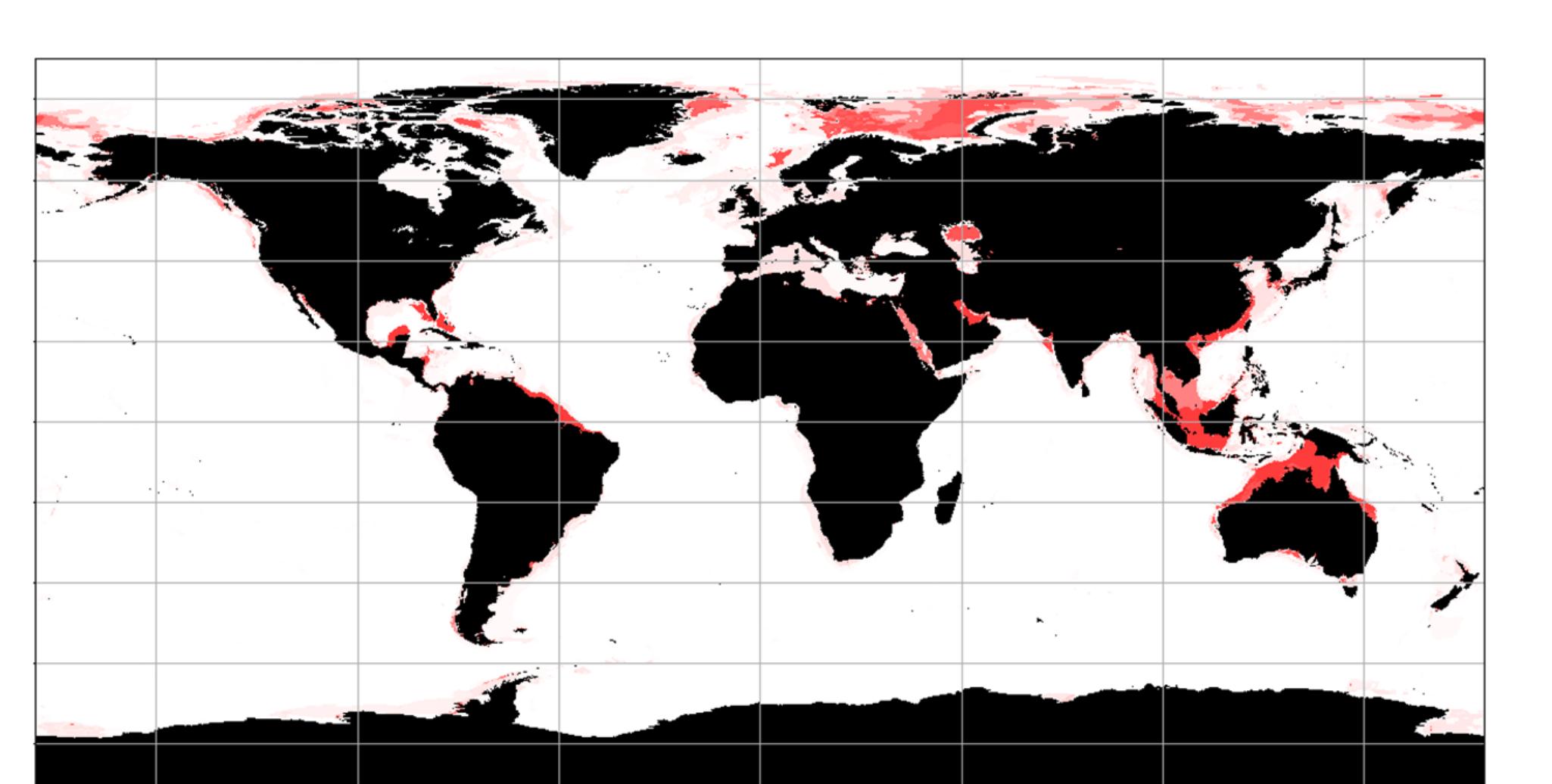


Figure 5. Global predictions for formation probability of gas using the K-means clustered simulations.

Discussion/Conclusions

- Comparison between the results from full simulations of all 24,092 locations in the Svalbard study to those predicted using k-means clustering driven simulations serve as a proof of concept for applying the approach to larger models
- K-means clustering represents an efficient technique to design simulation strategies
 - Individual simulations at thousands or millions of locations requires weeks to simulate even using multiple processors compared to less than a single day for k-means
 - Many clusters do not form any gas or hydrate and this method can identify locations that do not require simulation to expedite the individual simulation process
- Simpler gas system allows for the generalized clusters to adequately represent expected findings based on full simulations
- Hydrate system is more complex, predictions are generally in agreement but k-means predictions are not as successful at capturing nuances associated with individual locations
- Results can be used to identify locations which warrant further investigation and avoid unnecessary simulations for areas unsuitable to gas or hydrate formation
- Number of simulations where formation occurs (formation probability) can be extracted to represent likelihood of gas presence in sediments
- Newly constructed maps can inform predictions of acoustic velocity alterations that need to be considered for accurate bathymetric studies

EXTRA IMAGES AND TABLES

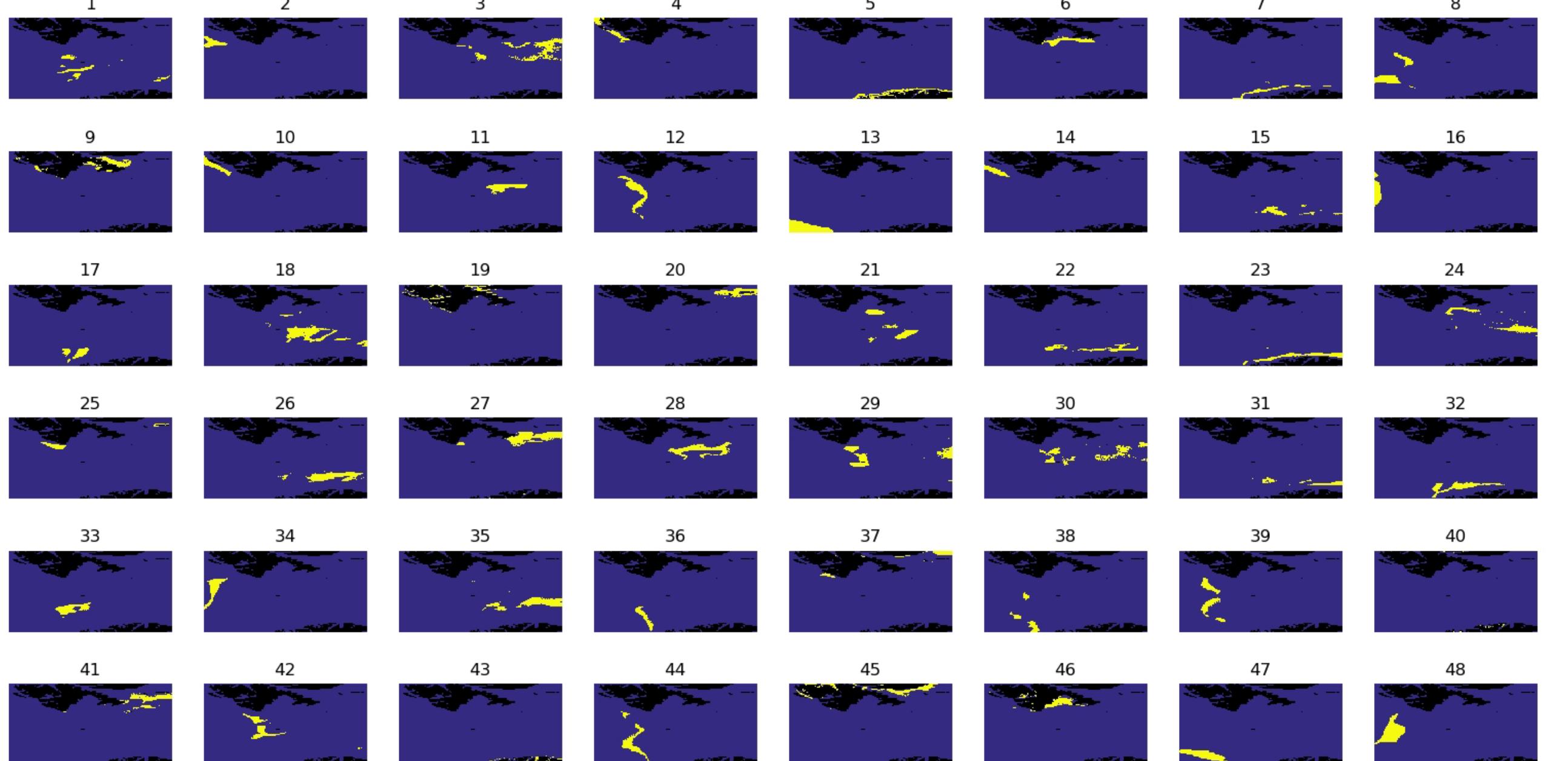


Figure 6. Map of cluster areas for Svalbard study area. Cluster number is indicated above each panel and region is indicated in bright yellow. Corresponding formation probabilities are provided in **Table 1**.

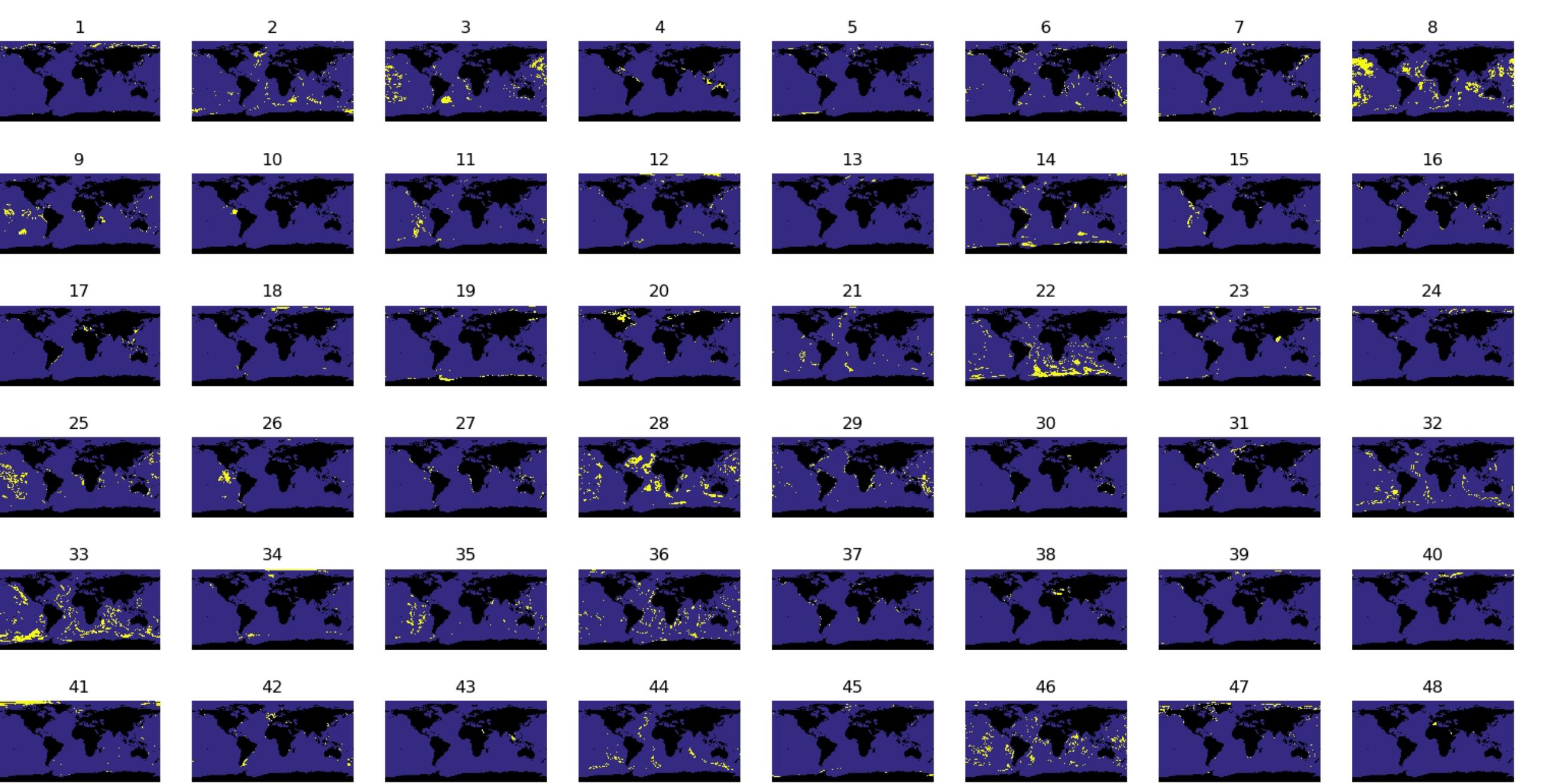


Figure 7. Map of cluster areas for Global study area. Cluster number is indicated above each panel and region is indicated in bright yellow. Note that each region is not limited to a coherent geographical area but instead can include areas across the globe which have similar geophysical characteristics.

Cluster	K-means		Full Simulation		Cluster	K-means		Full Simulation	
	Gas Formation	Hydrate Formation	Gas Formation	Hydrate Formation		Gas Formation	Hydrate Formation	Gas Formation	Hydrate Formation
1	99.40%	0.60%	85.00%	3.28%	25	96.20%	0.00%	89.24%	0.00%
2	1.40%	0.00%	3.04%	0.00%	26	99.60%	0.00%	91.61%	0.19%
3	99.80%	1.40%	98.02%	0.98%	27	99.60%	0.00%	98.67%	0.04%
4	83.60%	0.00%	77.93%	0.00%	28	100.00%	0.00%	98.52%	0.00%
5	2.20%	0.00%	2.57%	0.00%	29	98.60%	1.20%	86.03%	1.51%
6	100.00%	0.00%	99.95%	0.00%	30	100.00%	0.00%	96.39%	0.41%
7	44.20%	0.00%	46.98%	0.00%	31	93.20%	0.00%	72.60%	0.00%
8	0.00%	0.00%	0.00%	0.00%	32	84.80%	0.00%	71.54%	0.00%
9	72.00%	0.00%	64.81%	0.00%	33	98.60%	0.20%	80.11%	6.75%
10	99.60%	0.00%	97.68%	0.00%	34	0.00%	0.00%	0.00%	0.00%
11	100.00%	0.00%	98.87%	0.00%	35	99.80%	3.60%	95.81%	9.42%
12	62.40%	0.00%	50.98%	0.58%	36	22.60%	0.00%	27.95%	0.00%
13	0.00%	0.00%	0.00%	0.00%	37	44.40%	0.00%	15.06%	0.00%
14	94.00%	0.00%	85.64%	0.00%	38	0.00%	0.00%	0.00%	0.00%
15	99.20%	0.00%	86.57%	2.33%	39	0.00%	0.00%	0.00%	0.00%
16	0.00%	0.00%	0.00%	0.00%	40	13.40%	0.00%	1.75%	0.00%
17	96.80%	0.00%	82.29%	0.00%	41	94.40%	0.00%	91.22%	0.00%
18	99.80%	13.40%	95.03%	14.71%	42	98.80%	0.00%	83.95%	0.10%
19	75.80%	0.00%	49.94%	0.00%	43	0.00%	0.00%	0.00%	0.00%
20	57.60%	0.00%	48.47%	0.00%	44	0.20%	0.00%	1.06%	0.81%
21	99.60%	19.00%	96.64%	19.64%	45	26.40%	0.00%	29.93%	0.00%
22	98.60%	0.00%	83.79%	0.00%	46	100.00%	0.00%	98.72%	0.00%
23	5.80%	0.00%	8.54%	0.00%	47	0.00%	0.00%	0.00%	0.00%
24	100.00%	16.40%	99.29%	17.40%	48	0.00%	0.00%	0.00%	0.00%

Table 1. Formation probability of gas and hydrate for Svalbard study area based on k-means centroids (left) and individual simulations (right).

Cluster	Sedimentation Rate (mm/yr)	TOC (%)	Heat Flux (mW/m ²)	Depth (mbsl)	Temperature (°C)	K-means		Full Simulation	
	(mm/yr)	(%)	(mW/m ²)	(mbsl)	(°C)	Gas Formation	Hydrate Formation	Gas Formation	Hydrate Formation
1	0.251 ± 0.022	1.074 ± 0.009	65.367 ± 0.504	230 ± 10.5	2.45 ± 0.41	66.446 ± 10.404	51.2 ± 6.764	-0.13 ± 1.21	-
2	0.233 ± 0.027	0.706 ± 0.155	122.50 ± 7.183	1548 ± 94.4	0.26 ± 0.64	14.9%	0.00%	3.04%	0.60%
3	0.156 ± 0.023	1.720 ± 0.095	68.846 ± 2.808	183.5 ± 94.4	0.81 ± 0.40	99.93%	1.40%	98.42%	0.78%
4	0.053 ± 0.019	1.174 ± 0.258	109.71 ± 4.737	125.6 ± 109.9	0.87 ± 0.53	93.40%	0.00%	88.62%	0.00%
5	0.120 ± 0.027	0.920 ± 0.059	122.50 ± 7.183	1548 ± 94.4	0.26 ± 0.64	14.9%	0.00%	3.04%	0.60%
6	0.134 ± 0.028	2.097 ± 0.065	62.319 ± 4.249	90.5 ± 55.9	0.29 ± 0.63	100.00%	0.00%	99.95%	0.00%
7	0.110 ± 0.027	0.534 ± 0.126	65.518 ± 3.268	279.1 ± 139.3	5.54 ± 0.43	44.20%	0.00%	48.07%	0.00%
8	0.178 ± 0.018	0.640 ± 0.088	75.481 ± 4.257	139.9 ± 76.9	-0.78 ± 0.28	0.00%	0.00%	0.00%	0.00%
9	0.098 ± 0.020	0.640 ± 0.088	65.518 ± 3.268	279.1 ± 139.3	5.54 ± 0.43	44.20%	0.00%	48.07%	0.00%
10	0.117 ± 0.024	1.321 ± 0.146	124.316 ± 3.784	162.7 ± 77.1	1.33 ± 0.37	99.69%	0.00%	98.18%	0.00%
11	0.306 ± 0.024	1.501 ± 0.166	62.823 ± 2.632	79.6 ± 43.8	0.26 ± 0.46	100.00%	0.00%	99.64%	0.00%
12	0.205 ± 0.026	0.706 ± 0.107	76.022 ± 4.387	449.2 ± 181.6	1.80 ± 0.42	62.40%	0.00%	50.98%	0.38%
13	0.120 ± 0.027	0.640 ± 0.088	65.518 ± 3.268	279.1 ± 139.3	5.54 ± 0.43	44.20%	0.00%	48.07%	0.00%
14	0.323 ± 0.024	0.916 ± 0.164	126.437 ± 5.048	106.2 ± 27.9	1.62 ± 0.32	98.60%	0.00%	93.79%	0.00%
15	0.163 ± 0.027	1.574 ± 0.099	65.667 ± 1.348	341.7 ± 69.9	3.30 ± 0.43	99.20%	0.00%	86.31%	2.32%
16	0.139 ± 0.027	0.660 ± 0.070	132.025 ± 4.697	232.7 ± 166.7	-0.91 ± 0.10	0.00%	0.00%	0.00%	0.00%
17	0.120 ± 0.027	0.640 ± 0.088	65.518 ± 3.268	279.1 ± 139.3	5.54 ± 0.43	44.20%	0.00%	48.07%	0.00%
18	0.265 ± 0.019	1.581 ± 0.122	45.422 ± 4.211	331.7 ± 127.0	2.10 ± 0.32	99.99%	13.40%	95.08%	14.54%
19	0.035 ± 0.036	1.492 ± 0.215	39.009 ± 3.901	105.3 ± 9.7	2.38 ± 0.58	70.62%	0.00%	-	-
20	0.037 ± 0.011	1.224 ± 0.096	33.847 ± 2.630	161.5 ± 77.7	-0.46 ± 0.35	57.69%	0.00%	58.24%	0.00%
21	0.120 ± 0.027	0.640 ± 0.088	65.518 ± 3.268	279.1 ± 139.3	5.54 ± 0.43	44.20%	0.00%	48.07%	0.00%
22	0.200 ± 0.022	1.085 ± 0.124	46.869 ± 2.947	394.5 ± 43.2	3.33 ± 0.43	98.60%	0.00%	83.79%	0.00%
23	0.039 ± 0.019	0.662 ± 0.138	65.609 ± 2.910	300.9 ± 75.7	5.10 ± 0.59	5.80%	0.00%	8.94%	0.00%
24	0.209 ± 0.024	1.924 ± 0.141	69.266 ± 3.317	277.3 ± 132.4	1.34 ± 0.29	100.00%	0.00%	99.39%	14.95%
25	0.120 ± 0.027	0.640 ± 0.088	65.518 ± 3.268	279.1 ± 139.3	5.54 ± 0.43	44.20%	0.00%	48.07%	0.00%
26	0.344 ± 0.022	1.528 ± 0.155	67.855 ± 4.507	291.5 ± 155.9	3.41 ± 0.26	99.69%	0.00%	91.61%	0.19%
27	0.079 ± 0.021	1.917 ± 0.109	74.095 ± 2.642	136.3 ± 2.764	0.17 ± 0.49	99.60%	0.00%	98.69%	0.02%
28	0.235 ± 0.019	1.527 ± 0.143	67.725 ± 2.266	94.2 ± 44.8	0.38 ± 0.43	99.34%	0.00%	99.34%	0.00%
29	0.120 ± 0.027	0.640 ± 0.088	65.518 ± 3.268	279.1 ± 139.3	5.54 ± 0.43	44.20%	0.00%	48.07%	0.00%
30	0.183 ± 0.021	1.429 ± 0.098	65.713 ± 1.838	167.7 ± 74.6	1.				