SANDIA REPORT
SAND2021-10675
Printed September 2021

Sandia National Laboratories

Forecasting Marine Sediment Properties with Geospatial Machine Learning

Frederick, Jennifer M.[1], William K. Eymold[1], Michael A. Nole[1], Benjamin J. Phrampus[2], Taylor Lee[2], Warren T. Wood[2], David Fukuyama[3], Olin Carty[3], Hugh Daigle[3], Hongkyu Yoon[1], and Ethan Conley[1]

[1] Sandia National Laboratories, Albuquerque, NM.

[2] U.S. Naval Research Laboratory, Stennis Space Center, MS.

[3] University of Texas at Austin, Austin, TX.

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

Using a combination of geospatial machine learning prediction and sediment thermodynamic/physical modeling, we have developed a novel software workflow to create probabilistic maps of geoacoustic and geomechanical sediment properties of the global seabed. This new technique for producing reliable estimates of seafloor properties can better support Naval operations relying on sonar performance and seabed strength, can constrain models of shallow tomographic structure important for nuclear treaty compliance monitoring/detection, and can provide constraints on the distribution and inventory of shallow methane gas and gas hydrate accumulations on the continental shelves.

ACKNOWLEDGEMENTS

We are grateful for the support of the Laboratory Directed Research and Development program at Sandia National Laboratories and the Earth Science Investment Area for providing funding and supporting the national security vision of this project.

The Academic Alliance program at Sandia National Laboratories allowed us to work with faculty and students the University of Texas at Austin. This collaboration was mutually beneficial, and we hope to continue the collaboration started through this project.

Throughout this project, we had incredible support from our management team, including Lori Parrott, Kyle Jones, Emily Stein, Erik Webb, and Ben Cook. We are also very appreciative of the business professionals on the team that kept us on budget and helped with contracting paperwork, including Eric DeBlassie, Tracy Woolever, Matthew Roybal, and Aaron Levenson.

We thank Alex Sinclair for her interest in the project and her work towards investigating water column and seafloor acoustics on the Alaskan North slope.

Many times during this project, we laughed at the fact that our team was physically spread out across four U.S. time zones. Sometimes it made scheduling virtual meetings difficult but made us appreciate our annual face-to-face group meetings that much more! We are grateful to have hosted meetings on the UT Austin campus, at the Stennis Space Center in Mississippi, and at Sandia National Laboratory in Albuquerque, NM.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
ANNs	artificial neural networks
ASW	anti-submarine warfare
CNNs	convolutional neural networks
GIRT	global implicit reactive transport
GHSZ	gas hydrate stability zone
GML	geospatial machine learning
GMRT	global multi-resolution topography
IODP	International Ocean Discovery Program
LHS	Latin hypercube sampling
MBSF	meters below seafloor
MIW	mine warfare
PFLOTRAN	parallel flow and transport (simulator)
TOC	total organic content

1. MOTIVATION

The Arctic Ocean is one of Earth's last frontiers. As climate change continues to expose this once frozen marine expanse, accessibility and global interest in seafloor sovereignty has been steadily increasing. The United States, one of the eight Arctic Nations, faces new security challenges including the defense of its Arctic borders, protection of economic interests such as petroleum and fisheries, maintaining freedom of navigation, and ensuring maritime domain control.

The U.S. Navy has operated submarines under the Arctic since the passage of the USS Nautilis in 1958 (Allen et al., 2017). Recently, a new Naval Arctic strategy has been released that now includes "blue-water Arctic" operations (Eckstein, 2018). Such Naval operations, including navigation, anti-submarine warfare (ASW), and mine warfare (MIW) depend heavily on sonar performance and seabed strength.

The Navy relies on sound (acoustic waves) in the ocean to perform a variety of tasks critical to its national security mission. Sound interacts with everything in the water column, including the sea surface and seafloor. Sound can also be altered significantly as it penetrates the seafloor and interacts with geologic structures before returning to the water column to propagate further. Over long distances there can be many interactions with the sea surface and seafloor. Therefore, knowing the seabed geology constrains its acoustic properties, specifically the sound speed, density, and attenuation. Knowing the acoustic properties allows us to forecast the sound interaction with the seabed, and subsequently the performance of any given sonar system.

Reliable estimates of key seafloor properties (e.g. porosity, sediment type, and phase fractions of solid, liquid, and gas) are required to predict geoacoustic and geomechanical properties. In particular, shallow methane accumulations in seafloor sediments have enormous effects on sonar performance and seabed strength, however, only in the gas phase. Sub-seafloor thermodynamics, such as the temperature and pressure, control gas solubility and the gas hydrate stability boundary, and subsequently, where gas-phase methane occurs.

The Arctic Ocean is a new, challenging theater of operations for the military, where direct seafloor observations (data) are far sparser than in previous areas of Naval interest, leaving operational activities such as ASW, MIW, and others unsupported. Large campaigns of seafloor data acquisition (the current solution to this problem elsewhere) are unlikely in the Arctic for the foreseeable future. Instead, a means of probabilistically estimating seafloor properties based on existing sparse observations with predictive numerical simulation, can form the optimal environmental input to support ASW and MIW tactical decision aids. Using a combination of geospatial machine learning prediction and sediment thermodynamic/physical modeling, we have created probabilistic maps of key seafloor characteristics that control geoacoustic and geomechanical sediment properties. This new technique for producing reliable estimates of Arctic seafloor properties will better support Naval operations relying on sonar performance and seabed strength.

This report summarizes the software workflow we have developed and used to produce probabilistic maps of any seafloor quantity of interest. It includes descriptions of the model validation exercises, and demonstrations of the workflow at several seafloor locations. We also provide a selected list of ~30 publications resulting from this work available to the reader for more detail.

2. THE GLOBAL PREDICTIVE SEABED MODEL (GPSM)

2.1. Introduction to GPSM

The objective of seafloor prediction is to obtain optimal estimates of any given quantity where it has not been directly measured. Because the seafloor is vast and very sparsely sampled in most quantities, we require a prediction almost everywhere. Conventional interpolation techniques (e.g. splines in tension), that rely only on geospatial proximity, ignore almost everything we know about the seabed are ill-suited for our purpose.

Geospatial Machine Learning (GML) predictions are based on the proximity in multi-dimensional geologic predictor space, rather than just geospatial proximity. With GML we predict, or interpolate intelligently, taking advantage of areas that may be geographically very distant, but geologically very similar. GML can use all the information we have about the seabed to predict what we do not know. Not relying strictly on geographical proximity allows prediction in geographical areas that are denied, as well as simply difficult-to-access, such as the Arctic.

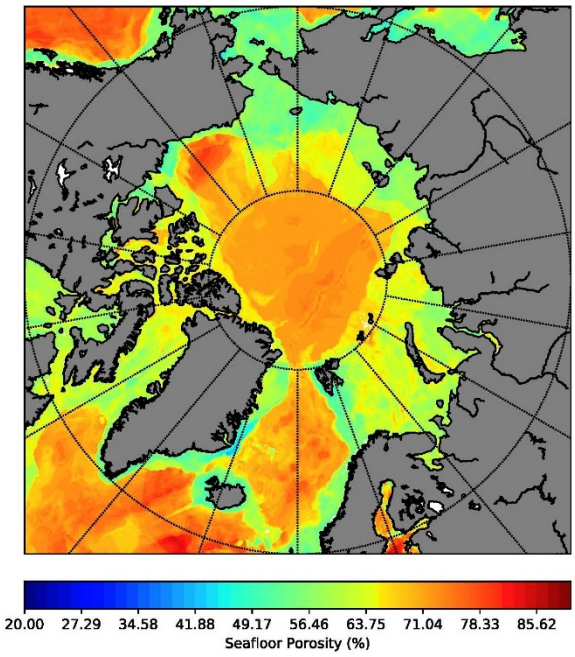


Figure 2-1. Seafloor porosity predicted by GPSM using geospatial machine learning. (Martin et al., 2015)

GML uses predict*ors* and a predict*and*. The quantity we wish to predict (the predictand) must also be observed at multiple locations (these observations are the sparse data). Increasing numbers of observations generally correspond to a denser sampling of predictor space, and more certain predictions.

The geologic predictors used in GML are quantities that are known both at the geospatial locations of the observations, as well as the locations of sites at which we want to predict. These can be quantities that have been measured (e.g. bathymetry), calculated (e.g. distance from shore), or predicted from other GML applications (e.g. porosity). Spatial statistics of these parameters may also be used as predictors. One must clearly be careful to avoid circularity; e.g. using predicted porosity as a predictor to predict a quantity, then subsequently using that same quantity to upgrade the prediction of porosity.

The relationship between the predictors and the predictand need not be known a priori – this is a strength of the GML approach, but also means that for any given application we do not know which predictors will be the most useful. Bad predictors (those poorly correlated with the predictand), can degrade the prediction. For this reason, we typically employ several strategies of predictor (also called feature) selection – determining the best predictors for any given seafloor quantity.

Quantitatively determining the best predictors, as well as quantifying the accuracy of the final prediction, is done through validation. To validate any given group of predictors, we withhold a random fraction of the observed data (usually 10%, for 10-fold validation), make predictions at those locations, and then compare the predicted with the observed values. This is repeated with a different random fraction, until each observation has, at some point, been withheld and predicted. The error between observed and predicted is a reliable, quantitative measure of predictive skill of the prediction method. Validation proves quantitatively and decisively the superiority of machine learning interpolation to conventional, geospatial-only interpolation.

The Naval Research Laboratory's Global Predictive Seafloor Model (GPSM) is a practical implementation of GML designed to provide near real-time estimates of Navy-relevant quantities from continuous seafloor property fields generated by machine learning algorithms (K-Nearest Neighbor, Random Forests, etc.) given often sparse measurements or direct observations compiled from widely available sources.

GPSM has been in use and continuously improved since providing the first global seafloor porosity field, published in 2015 (Martin et al., 2015; see Figure 2-1). More recent publications, some supported in whole or in part by this project include predictions of total organic carbon (Lee et al., 2019), seep probability (Phrampus et al., 2020), seafloor sediment accumulation rates (Restreppo et al., 2020), recent bathymetry change (Obelcz et al., 2020), sediment thickness vs. geologic age (Lee et al., 2020), density (Graw et al., 2021), and gas hydrate accumulation (Eymold et al., 2021). A more detailed discussion of how GPSM is used is presented in Section 2.2.

2.2. Specifics of GPSM Usage for Seabed Modeling

Past estimates of seafloor characteristics have been made from an assortment of data, including well logs (Hillman et al., 2017; Lee & Waite, 2008; Majumdar et al., 2017) and seismic surveys (Shipley et al., 1979; Wood & Ruppel, 2000) conducted by industry, military, and academic research groups. Despite their importance in determining the geo-acoustic and geo-mechanical properties for the seafloor, observations across much of the planet remain sparse, especially in sensitive areas such as the Arctic. Because extensive campaigns to accumulate higher resolution data of the seafloor remain prohibitive, probabilistic estimations of critical seafloor parameters such as porosity, sedimentation, and TOC can be made using geospatial machine learning (GML) techniques using GPSM, based on the sparse available data (Lee et al., 2019; Martin et al., 2015). By using these predictions as variable or uncertain input parameters for geophysical modeling, uncertainties from the input space can be propagated into numerical simulations to produce probabilistic maps of additional properties within the sediment column, such as acoustic velocity of the sediments, or the depth of the gas hydrate stability zone (GHSZ).

GPSM provides predictions for values of sedimentation rate (m/yr), total organic content (TOC) (%), heat flux (W/m^2), sediment porosity, seafloor depth (m), seafloor temperature (°C), among many other characteristics of interest. Current resolution for the model is up to 5 arc-minute spacing across the globe. By using multivariate data of seafloor physical properties, the model extrapolates from better-constrained areas to poorer-constrained areas based on similarity in geologic characteristics rather than just geographic distance. The data provided by the model include both the expected values and standard deviations on predictions for seafloor depth, seafloor temperature, TOC, heat flux, sedimentation rate, and porosity structure at each location in a desired location, such as the Blake Ridge study area (Figure 2-2), for example. Estimates for TOC and heat flux were made using the k-nearest neighbors (KNN) algorithm (Lee et al., 2019) based on datasets of TOC (Seiter et al., 2004) and heat flux (Goutorbe et al., 2011). Preliminary predictions for sedimentation

rate come from taking sedimentary thickness values from GlobSed (Straume et al., 2019) and dividing by their reported ages (Müller et al., 2019). Seafloor temperature estimates are provided through the World Ocean Atlas (Locarnini et al., 2018) and depths were taken from SRTM15+ down-sampled to 5 minutes (Tozer et al., 2019). Porosity profiles were determined by the Random Forest method using data from the Deep Sea Drilling Project, Ocean Drilling Program (ODP), Integrated Ocean Drilling Program (IODP), and Pangaea data sets (Martin & Wood, 2017; Martin et al., 2015).

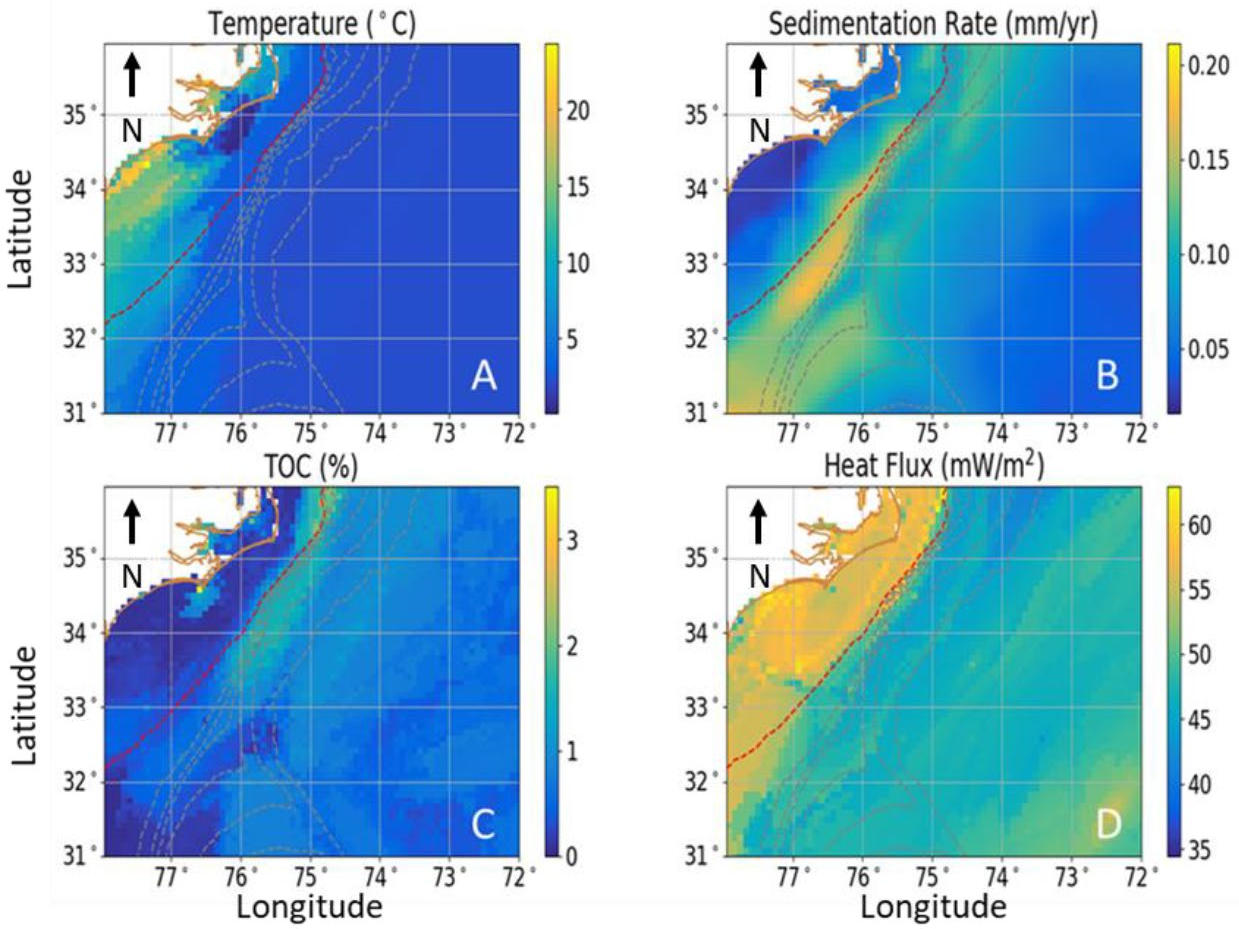


Figure 2-2. Map of average GPSM values for the Blake Ridge study area. Panel A shows the temperature in °C, Panel B shows the TOC means in %, Panel C shows sedimentation rate (ω) means in mm/yr, and Panel D shows the heat flux (Q) means in mW/m^2. All variables are mapped such that yellow colors indicate higher values and blue colors indicate lower ones. Onshore areas are shown in white and bathymetric contour intervals between 500–3,500 mbsl are plotted as dashed lines with the 500 mbsl line indicated in red. Based on Figure 2 of Eymold et al. (2021).

The GPSM results used in this report utilize publicly available data. For example, one such publicly available database is dbSEABED (http://instaar.colorado.edu/%7Ejenkinsc/dbseabed/), used to predict sediment type (terrigenous, biogenous, granitic, basaltic, calcareous, or siliceous) and grain size. dbSEABED catalogues millions of sediment samples collected throughout the world's oceans and over several decades. The following information and sediment parameters are included in the database: latitude, longitude, water depth, dominant bottom type of sediment (rock, gravel, sand, mud), and percentages of sediment types (gravel, sand, mud) (Figure 2-3).

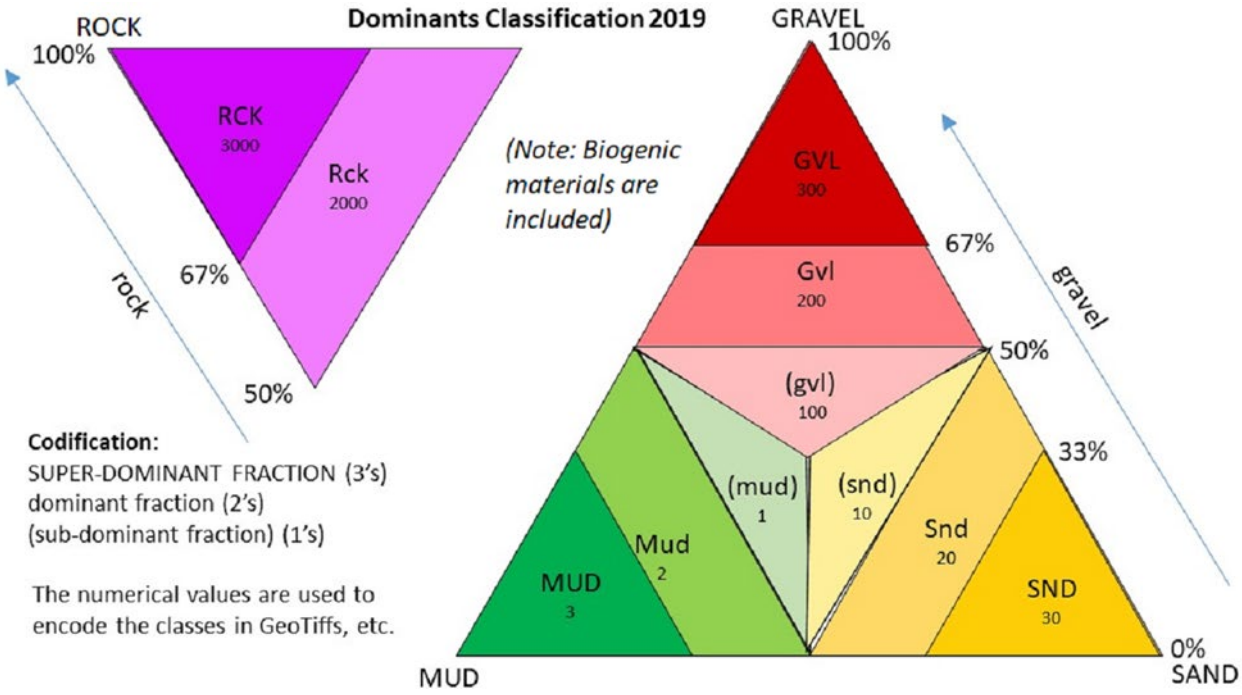


Figure 2-3. dbSEABED Dominants Classification 2019 color charts to be used with visual displays of the database.

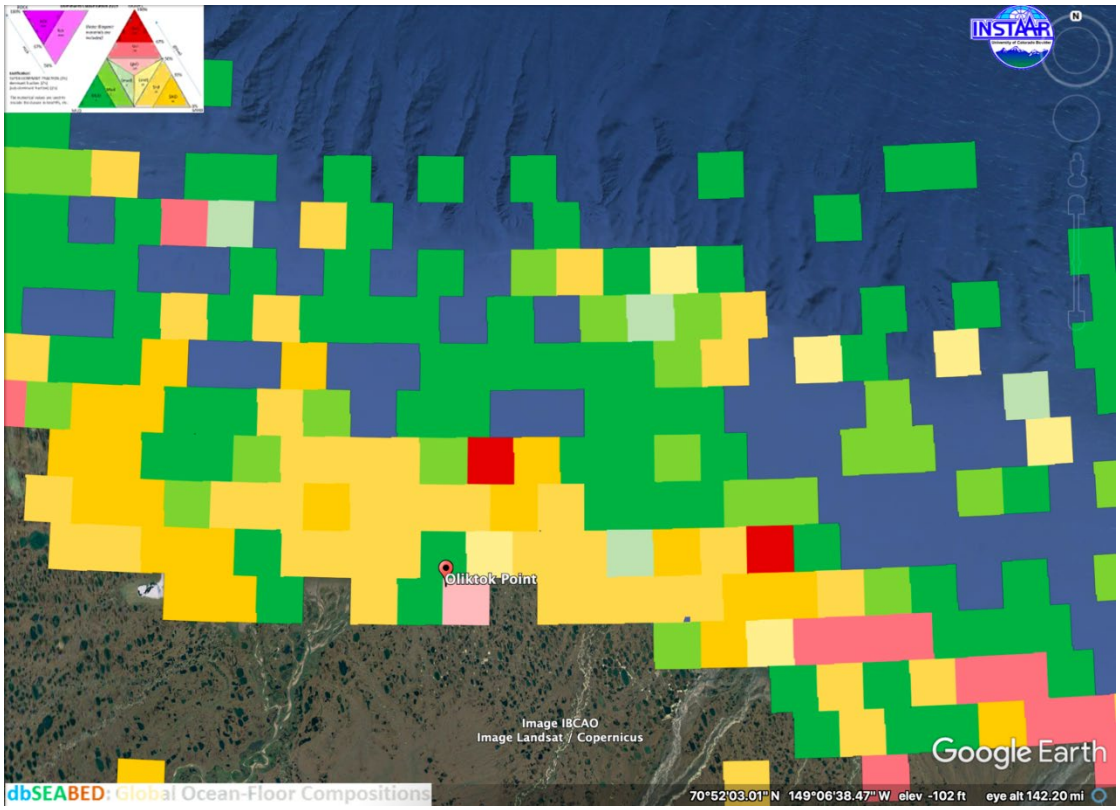


Figure 2-4. Map of available surficial sediment samples in dbSEABED in an example Alaskan North Slope region of interest.

For example, a search in dbSEABED of an example region of interest revealed the predominant surficial sediment type immediately north and offshore of Oliktok Point, Alaska, is sand and transitions to mud farther offshore, with infrequent pockets of gravel (Figure 2-4). Near the southeast boundary of the region, surficial sediment type is a mixture of gravel, sand, and mud.

3. DAKOTA

3.1. General Usage of Dakota Software

We use the statistical software package, Dakota, to initiate the PFLOTRAN simulations. Dakota was developed to optimize simulation codes and provide uncertainty quantification of parameter studies (Adams et al., 2019). Values for the mean, μ, and standard deviation, σ, of each seafloor parameter from GPSM are written as input to Dakota for each sample location. For our initial model demonstration at Blake Ridge (Eymold et al., 2021), five parameters were chosen for sampling: TOC, heat flux, sedimentation rate, methanogenesis rate, and a porosity scaling factor. Subsequently, we have added the capability to sample on the GPSM values for seafloor temperature and depth in the more recent version of the workflow. Dakota uses the Latin hypercube sampling (LHS) method to produce a distribution of variables for the simulations by dividing the range of each variable into N distributions of probability 1/N and then select N values for each variable in order to accurately represent the variability of all parameters (McKay et al., 1979). Based on the ranges of each parameter, the TOC, heat flux, sedimentation rate, temperature, and depth were sampled from a normal distribution, the scaling factor was sampled with a uniform distribution, and methanogenesis rate was sampled using a log-uniform distribution due to its spanning over three orders of magnitude. This method provides N values of each variable independent of the other variables to provide a set of distinct values of individual parameters for each simulation. To avoid non-physical values, upper and lower bounds for each variable can be set using flags in Dakota.

3.2. Specific Usage of the Dakota Software Options

Response functions can be defined in Dakota to determine the correlation coefficients between input variables and output values from all simulations run. We set two response functions for these simulations. First, we calculate the total mass of gas hydrate (H_n) formed in the complete sediment column at every location *n* following the formulation:

$$H_n = \sum_z S_h(z) \cdot \phi\,(z) \cdot V_{cell} \cdot \rho_h \qquad \textbf{(Eq. 3.1)},$$

where $S_h(z)$ is the gas hydrate saturation at depth z, $\phi\,(z)$ is the porosity at depth z, V_{cell} is the volume of the grid block (1 m^3), and ρ_h is the density of hydrate, set to a constant value of 920 kg/m^3. A similar calculation can be made for the mass of free gas in the sediment profile:

$$G_n = \sum_z S_g(z) \cdot \phi\,(z) \cdot V_{cell} \cdot X_g(z) \cdot \rho_g(z) \qquad \textbf{(Eq. 3.2)},$$

where $S_g(z)$ is the gas saturation at depth z, $\phi\,(z)$ is the porosity at depth z, V_{cell} is the volume of the grid block (1 m^3), $X_g(z)$ is the mole fraction of methane at each depth, and $\rho_g(z)$ is the pressure dependent density of methane at each depth.

Correlation coefficients were determined for the input values of sedimentation rate, TOC, heat flux, and methanogenesis rate and the output mass of hydrate formed during each simulation. The Pearson correlation coefficient was calculated by:

$$corr(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad \textbf{(Eq. 3.3)},$$

where x is the independent variable (e.g., sedimentation rate), y is the response function (e.g., mass of hydrate formed), i indicates the simulation number, n is the total number of simulations run, and $\bar{x}$ and $\bar{y}$ are the average independent variable and response function, respectively (Adams et al., 2019). The simple raw correlation values between the input parameters and H_n for Blake Ridge were plotted across the study area (Figure 3-1).

As a qualitative assessment, the variable with the strongest correlation with H_n can be determined at each location (Figure 3-2). Most of the study area has the highest correlation between TOC and formation mass (green), while parts of the northern portion exhibit the strongest relationship based on sedimentation rate (orange). Along many of the shallowest areas that formed hydrate, methanogenesis rate exhibits the strongest relationship (blue). The two formation maxima show a stronger relationship with heat flux than any of the other variables (purple).

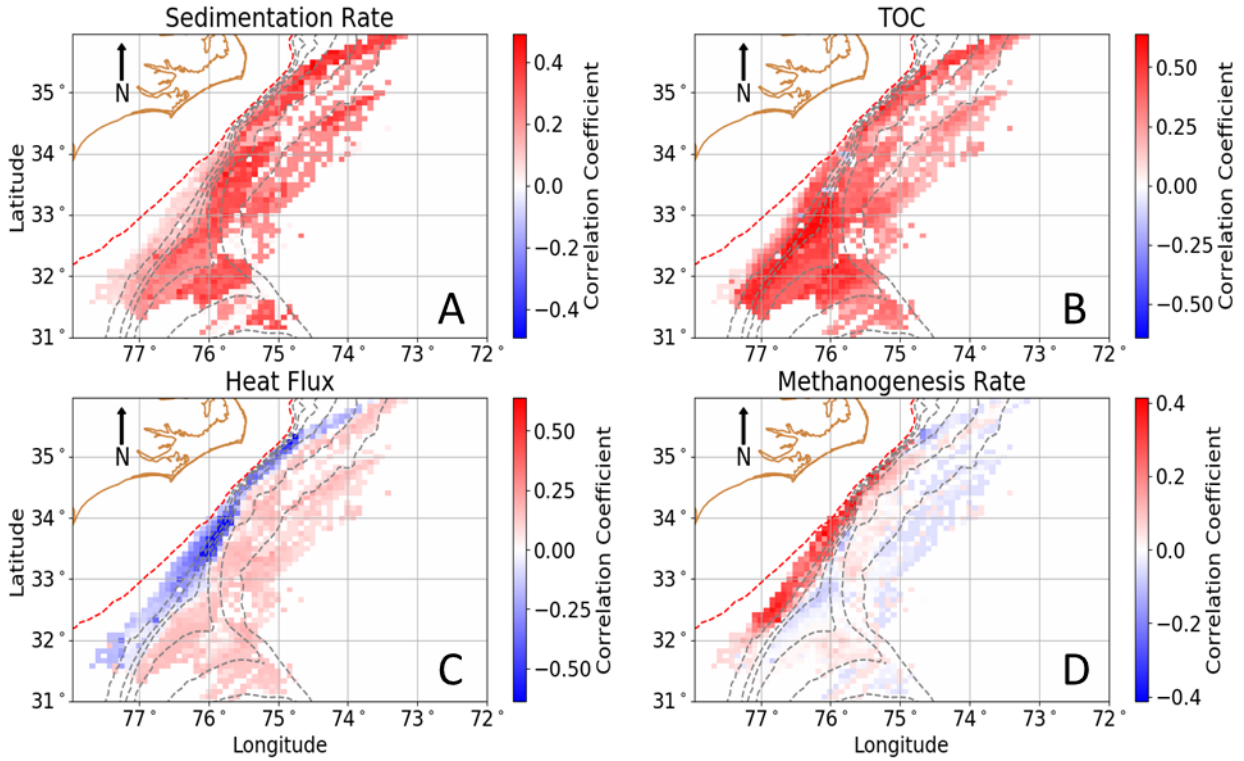


Figure 3-1. Correlation coefficients between hydrate formation mass and sedimentation rate (A), TOC (B), Heat Flux (C), and Methanogenesis Rate (D). All maps are shown such that red indicates positive correlations, blue indicates negative correlations, and white indicates a lack of correlation between variables. Adopted from Figure 7, Eymold et al. (2021).

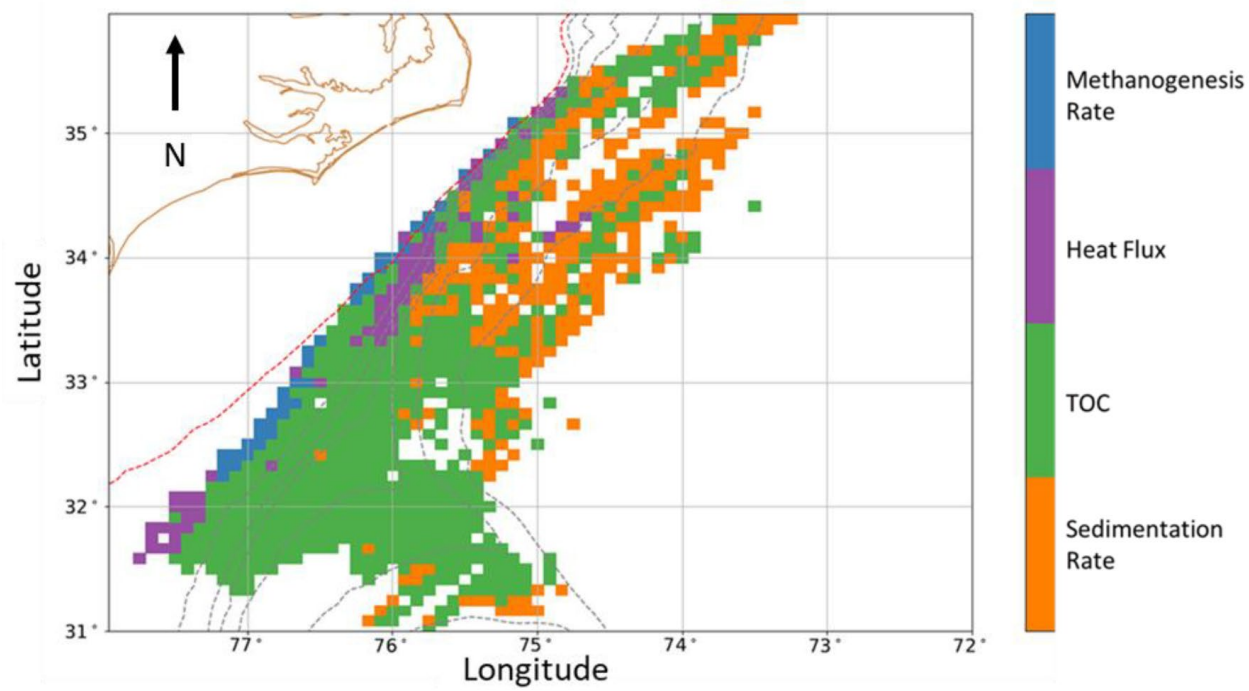


Figure 3-2. Map of variables with the strongest correlation coefficient at each location that formed gas hydrate in the Blake Ridge study area. Adopted from Figure 8, Eymold et al. (2021).

4. PFLOTRAN HYDRATE DEVELOPMENT

The open source, massively parallel multiphase flow and reactive transport finite volume simulator PFLOTRAN (Hammond et al., 2011; Lichtner and Hammond, 2012) was modified to simulate gas hydrate and ice formation in geologic media. This simulator was chosen both for its high-performance capabilities and its extensive in-house user and developer community at Sandia. Well-established workflows exist for probabilistic modeling of large-scale 3D geologic systems by coupling PFLOTRAN and Dakota (Mariner et al., 2017); leveraging this expertise and developing state-of-the-art simulation capability for simulating gas hydrate systems at large scales motivated the selection of PFLOTRAN for this development.

PFLOTRAN uses a "process model" coupling approach to multi-physics modeling, where individual process models solve independent systems of coupled partial differential equations. Process models are then coupled together sequentially either at the same timesteps (peer-to-peer) or by allowing one process model to sub-step another (parent-child). For example, when coupling single-phase flow to global-implicit reactive transport, PFLOTRAN would use both the RICHARDS flow mode process model and the GIRT (Global Implicit Reactive Transport) transport mode process model. The RICHARDS flow mode solves for the flow of liquid water in porous media through the solution of a single partial differential equation for mass conservation. The GIRT transport mode is sequentially coupled as a child to the flow mode, meaning GIRT can sub-step (take smaller time steps) the flow mode for a given time step and then sync back up at the end of the flow time step.

Newly implemented in PFLOTRAN is a flow mode called HYDRATE mode, which solves the coupled system of partial differential equations describing mass conservation of water and gas (methane) as well as conservation of energy in porous media. The system is solved using a fully

implicit method that is backward Euler in time, meaning the solution is unconditionally stable. The HYDRATE mode process model considers water in the liquid, steam, ice, and hydrate phases as well as gas in the dissolved, gaseous, and hydrate phases. Just like other flow modes, HYDRATE mode can be sequentially coupled to other process models such as GIRT.

4.1. Governing Equations for PFLOTRAN Hydrate

PFLOTRAN's HYDRATE mode solves a system of two mass conservation equations and one energy conservation equation to model two-component, equilibrium, multiphase flow in porous media considering phase changes and heat transfer. For each component, *i*, mass conservation takes the form:

$$\frac{d}{dt}\phi\left(s_l\rho_l x_l^i + s_g\rho_g x_i^g\right) + \nabla\cdot\left(\boldsymbol{q_l}\rho_l x_l^i + \boldsymbol{q_g}\rho_g x_g^i - \phi s_l D_l\rho_l\nabla x_l^i - \phi s_g D_g\rho_g\nabla x_g^i\right) = Q_i$$

(Eq. 4.1)

where ϕ is the porosity of the medium, s_l and s_g are the liquid and gas saturations (fraction of liquid or gas per unit volume of pore space) respectively, ρ_l and ρ_g are the liquid and gas densities, respectively, x_l^i and x_g^i are the mole fractions of component *i* in the liquid and gas phases, q_l and q_g are the advective (Darcy) fluxes of the liquid and gas phases, D_l and D_g are the effective diffusion coefficients in the liquid and gas phases, and Q_i represents a source or sink of component *i*. Hydrate and ice phases are considered immobile, and diffusion through solid phases is not considered in the current formulation.

Conservation of energy takes the form:

$$\sum_{\alpha=l,g,h,i}\left\{\frac{d}{dt}(\phi s_\alpha\rho_\alpha U_\alpha) + \nabla\cdot(q_\alpha\rho_\alpha H_\alpha)\right\} + \frac{d}{dt}\left((1-\phi)\rho_r C_p T\right) - \nabla\cdot(\kappa\nabla T) = Q \qquad \textbf{(Eq. 4.2)}$$

where contributions are considered from liquid (*l*), gas (*g*), hydrate (*h*), ice (*i*), and rock (*r*) phases; U_α is the internal energy and H_α is the enthalpy of each pore occupying phase; $C_p$is the heat capacity of the rock, and κ is the thermal conductivity of the rock.

The system of equation is subject to the following constraints:

$$\sum_{\alpha=l,g,h,i} s_\alpha = 1 \qquad \textbf{(Eq. 4.3)}$$

$$P_c(s_l) = P_g - P_l \qquad \textbf{(Eq. 4.4)}$$

$$P_g = K_H x_l^g \qquad \textbf{(Eq. 4.5)}$$

$$U_\alpha = H_\alpha - \frac{P_\alpha}{\rho_\alpha} \qquad \textbf{(Eq. 4.6)}$$

where P_c is the capillary pressure, which must equal the difference between nonwetting (gas) and wetting (liquid) phase pressures, and K_H is the Henry's law constant, which dictates the solubility of the gas component in the liquid phase when 2 phases are present. PFLOTRAN contains a large library of capillary pressure relationships as functions of liquid saturation; Van Genuchten capillary pressure functions are most used. For the Henry's Law constant, two options were added as part of this work. The default option is as follows (Carroll and Mather, 1997):

$$K_H = \exp\left(5.1345 + \frac{7837}{T} - \frac{1.509x10^6}{T^2} + \frac{2.06x10^7}{T^3}\right) \cdot 1000 \qquad \textbf{(Eq. 4.6)}$$

where *T* is temperature. The second option (Cramer, 1982) is as follows:

$$K_H = 1x10^5 \cdot (24582.4 + 6.71091x10^2 T + 6.87067T^2 - 1.773079x10^{-1}T^3 + 1.09652x10^{-1}T^4 - 3.19599x10^{-6}T^5 + 4.46472x10^{-9}T^6 - 2.40294x10^{-12}T^7) \qquad \textbf{(Eq. 4.7)}$$

Densities and internal energies of each phase are computed using equations of state. Multiple equation of state options exist for water and gases. The most commonly used equation of state for liquid water is the IAPWS IF97 equation of state (IAPWS, 2007). For methane, the Peng-Robinson equation of state is generally appropriate. In the current formulation, densities of hydrate and ice are constant and set to 920 kg/m^3. For hydrate, enthalpy is formulated as a function of temperature (Handa, 1998):

$$H_H = C_{pH}(T - T_F) + \frac{H_{H0}}{N_H + 1} \qquad \textbf{(Eq. 4.8)}$$

$$C_{pH} = 1.620(MW_{H_2O} N_H + MW_{CH_4}) \qquad \textbf{(Eq. 4.9)}$$

where C_{pH} is the heat capacity of hydrate, T_F is the freezing temperature of water, H_{H0} is the reference enthalpy of hydrate, N_H is the hydration number, set to 6 for structure I methane hydrate, MW_{H_2O} is the molecular weight of water, and MW_{CH_4} is the molecular weight of methane. For ice, internal energy is computed as follows (Fukusako and Yamamoto, 1993):

$$U_I = L_w + 185(T - 273.15) + 3.445(T^2 - 273.15^2), \quad T \geq 90K \qquad \textbf{(Eq. 4.10)}$$

$$U_I = L_w + 4.475(T^2 - 273.15^2), \quad T < 90K \qquad \textbf{(Eq. 4.11)}$$

where U_I is the internal energy of ice and L_w is the latent heat of fusion of water (set to -6017.1 J/mol).

For the mobile phases (liquid and gas), Darcy velocities of each phase are computed as follows:

$$\boldsymbol{q_\alpha} = -\frac{k k_\alpha^r}{\mu_\alpha} \nabla(P_\alpha - \rho_\alpha \boldsymbol{g} z), \quad (\alpha = l, g) \qquad \textbf{(Eq. 4.12)}$$

where k is the intrinsic permeability of the medium, k_α^r is the relative permeability of phase α (a function of phase saturation), μ_α is the viscosity of phase α, P_α is the pressure of phase α, $\boldsymbol{g}$ is the acceleration due to gravity, and z is the depth in the direction of gravity. The numerical method used here upwinds relative permeabilities and viscosities and harmonically averages permeabilities between neighboring grid cells.

For marine gas hydrate systems modeling, sedimentation is often an important component of the geologic system being modeled. Sedimentation has been incorporated into the mass and energy flux terms: in addition to the Darcy flux driven by pressure gradients, a constant sedimentation rate can be specified which moves pore constituents in the downward direction (which can be specified with the GRAVITY keyword) at the specified sedimentation rate.

The effective diffusion coefficients are formulated as follows:

$$D_\alpha = \tau D_\alpha^i \qquad \textbf{(Eq. 4.13)}$$

where the bulk diffusivity in phase α, D_α^i, is scaled by the tortuosity of the porous medium, τ, which ranges from 0 to 1.

Thermal conductivity of the composite medium, κ, is computed as a function of each pore constituent. Numerous options for parameterizing thermal conductivity exist across PFLOTRAN, but these only consider the presence of liquid and gas. Across all other flow modes, the default formulation of thermal conductivity is (Somerton et al., 1974):

$$\kappa = \kappa_{dry} + \sqrt{s_l}(\kappa_{sat} - \kappa_{dry}) \qquad \textbf{(Eq. 4.14)}$$

where κ_{dry} is the dry thermal conductivity of the rock and κ_{sat} is the saturated thermal conductivity of the rock. In hydrate mode, this thermal conductivity function is an option, but the default thermal conductivity is overridden to the following formulation:

$$\kappa = \kappa_{dry} + \phi(s_l\kappa_l + s_h\kappa_h + s_i\kappa_i + s_g\kappa_g) \qquad \textbf{(Eq. 4.15)}$$

where κ_l, κ_h, κ_i, and κ_g are the thermal conductivities of liquid, hydrate, ice, and gas phases, respectively.

Alternatively, a third option exists in the form (White et al., 2020):

$$\kappa = \kappa_{dry} + \phi(s_l\kappa_{sat} + s_h\kappa_h + s_i\kappa_i + s_g\kappa_g) \qquad \textbf{(Eq. 4.16)}$$

A common source of hydrate-generating methane in marine environments is microbial methanogenesis. Steady-state methanogenesis as a function of depth is included using the following formulation (Malinverno, 2010):

$$Q_m = k_\alpha \lambda \alpha e^{-\frac{\lambda}{\omega}(z - z_{SMT})} \qquad \textbf{(Eq. 4.17)}$$

where k_α is a conversion factor from organic matter to methane, λ is the microbial reaction rate, α is the organic matter content at the seafloor, ω is the sedimentation rate, and z_{SMT} is the depth to the base of the sulfate reduction zone.

Phase changes are handled using equilibrium flash calculations, and the solution method uses primary variable switching when the system enters new phase states. For example, when the system is fully liquid-saturated, the code solves for the three primary variables liquid pressure (P_l), mole fraction of gas (methane) in the aqueous phase (x^l_g), and temperature (T). If concentration of dissolved gas exceeds its solubility in the aqueous phase, a gas phase forms and therefore the phase state changes from liquid saturated to two-phase. At this point, mole fraction of gas in the aqueous phase is no longer unknown since it is determined by Henry's Law. Therefore, the second primary variable switches from mole fraction of gas in the aqueous phase to gas phase saturation (s_g).

Similarly, if a fully liquid-saturated system exists within the gas hydrate stability zone (GHSZ) and dissolved methane exceeds solubility, a gas hydrate phase forms and the second primary variable switches to hydrate saturation. Options are provided in the code for description of both the methane hydrate phase boundary and methane solubility in the gas hydrate stability zone. The three options for phase boundaries are as follows:

1) From Kamath (1984):

$$P_e[kPa] = e^{14.717-\frac{1.88679x10^3}{T}}, T < T_{freezing}$$

$$P_e[kPa] = e^{38.98-\frac{8.533x10^3}{T}}, T \geq T_{freezing}$$

(Eq. 4.18)

2) From Moridis (2003):

$$P_e[MPa] = e\text{^}(-43.8921173434628 + 0.776302133739303\,T - 7.27291427030502\text{x}10^{-3}T^2 + 3.85413985900724\text{x}10^{-5}T^3 - 1.03669656828834\text{x}10^{-7}T^4 + 1.09882180475307\text{x}10^{-10}T^5)\ , T < T_{freezing}$$

(Eq. 4.19)

$$P_e[MPa] = e\text{^}(-1.9413850446456x10^5 + 3.31018213397926x10^3\,T - 22.5540264493806\,T^2 + 0.0767559117787059\,T^3 - 1.30465829788791x10^{-4}T^4 + 8.86065316687571x10^{-8}\,T^5)\ , T \geq T_{freezing}$$

3) A simplified, invertible fit of Moridis (2003):

$$P_e[MPa] = e^{0.0334940999T-8.1938174346}, T < T_{freezing}$$ **(Eq. 4.20)**

$$P_e[MPa] = e^{0.1100383278\,T-29.1133440975}, T \geq T_{freezing}$$

where P_e is the three-phase equilibrium boundary pressure at a given temperature. While the second phase boundary (Moridis, 2003) is the most precise, we found that the third equation is the most useful because of its invertibility and smooth derivative. Invertibility is a useful feature because the option exists to adjust the methane solubility within the gas hydrate stability zone as follows (Davie et al., 2004):

$$K_H = K_{H0}e^{-\frac{T-T_3}{\alpha}}$$ **(Eq. 4.21)**

Where K_{H0} is the Henry's Law constant before adjustment, α is a fitting parameter (set to 14.4° C), and T_3 is the three-phase equilibrium temperature at a given pressure. The three-phase equilibrium temperature is derived from inverting the phase boundary equations; for phase boundary equation 2, an approximate inversion is used.

HYDRATE mode currently supports offsetting in the gas hydrate and ice phase boundary due to the presence of dissolved salt for a constant salinity or for varying salinity. Discussion of salt mass conservation implementation is covered in the next section.

Increasing NaCl concentrations shift the hydrate equilibrium phase boundary towards higher pressures and lower temperatures. This is accounted for in PFLOTRAN through the stability condition parameters outlined in Dickens and Quinby-Hunt (1997), and the resulting temperature depression equation for systems with inhibitors introduced by salt in Moridis (2014):

$$\Delta T_s = \Delta T_{D,r} \frac{\ln(1-x)}{\ln(1-x_r)} \qquad \textbf{(Eq. 4.22)}$$

where ΔT_s represents the subcooling for a given NaCl mole fraction, x, with a given reference temperature depression, $\Delta T_{D,r}$, at a reference mole fraction, x_r.

As hydrate is formed from methane and water, salt is excluded from the solid phase as hydrate is formed (Figure 4-1). As hydrate saturation increases, salinity increases, and the surrounding pore water becomes denser.

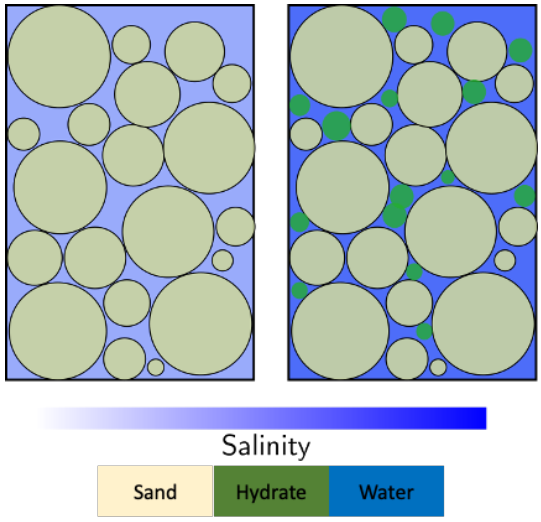


Figure 4-1. Hydrate formation increases salinity of the surrounding pore fluid.

In addition to being shifted due to salt presence, the gas hydrate phase boundary can be shifted due to the size of the pore in which hydrate or gas would precipitate. This behavior is captured through a modification the Gibbs-Thomson equation, which substitutes in a pseudo-capillary pressure of the hydrate phase for pore size through the Young-Laplace equation:

$$\Delta T_G = \frac{TP_c}{H_f \rho_H} \qquad \textbf{(Eq. 4.23)}$$

where ΔT_G is the shift in the three-phase equilibrium temperature due to the Gibbs-Thomson effect and H_f is the latent heat of fusion of hydrate, set to 6017.1 J/mol.

4.2. Additional Functionalities in PFLOTRAN Hydrate

In addition to modeling gas hydrate, HYDRATE mode also can model ice formation and dissolution. As a test problem of the capability to model gas hydrate and ice, a 1D sediment column was initialized with dissolved methane below solubility to form hydrate but within the gas hydrate stability zone. The heat flux at the top boundary was applied to reduce the temperature below the freezing point of ice. Since ice formation consumes only water, the dissolved concentration of methane increases locally as ice forms. Local increases in dissolved methane push methane concentrations past solubility to form hydrate, and therefore over time as the ice front propagates downward a corresponding hydrate saturation front ends up propagating with the ice front (Figure 4-2).

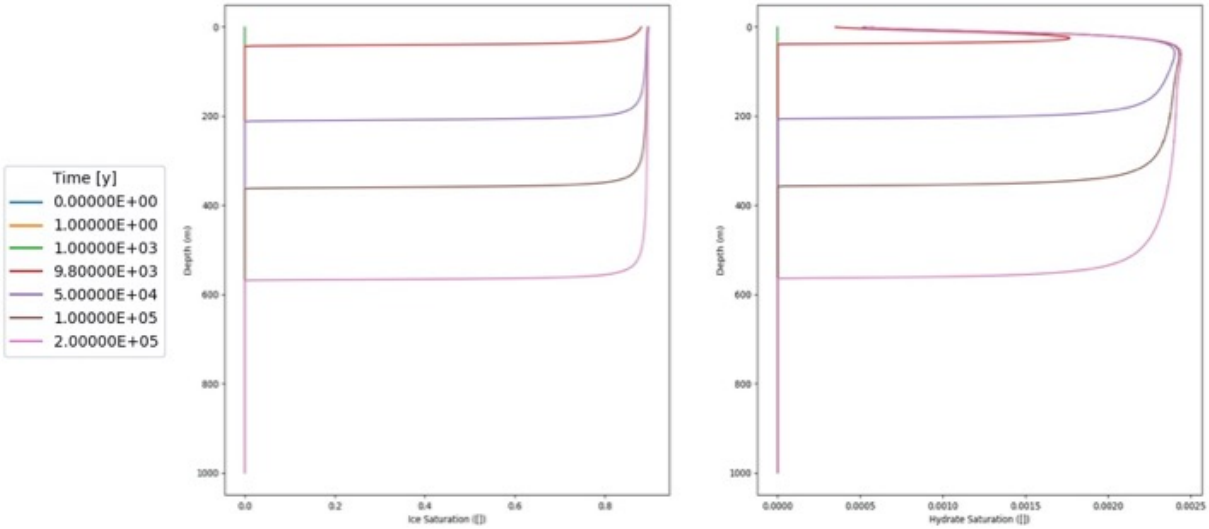


Figure 4-2. Ice saturation (left) and hydrate saturation (right) at several simulation times that demonstrate exclusion of methane as ice forms, allowing hydrate formation in the remaining pore space not occupied by ice.

PFLOTRAN's hydrate mode has been developed to consider a third component in addition to methane and water. At the time of publication this capability is not yet released in the main version of the code. The third component can either be used to model a separate dissolved species (such as a salt), or it can be used to model methane derived from a different source. When the third component is treated as methane from a different source, PFLOTRAN treats diffusive flux of methane by adding methane from both sources, and when forming a new phase (free gas or hydrate), the new phase is composed of methane from each source in the same ratio as these components exist in the aqueous phase.

As an example test problem, a 2D sediment column is initialized with gas hydrate that is all of shallow microbial origin (Figure 4-3). A dipping high permeability layer exists throughout the domain (outlined in Figure 4-4), and methane of deep thermogenic origin is injected along the dip as a boundary condition; the source is kept constant through time. Over time, methane with a deep thermogenic signature displaces methane with a microbial signature and forms hydrate. Since observational data can distinguish between methane with a microbial and thermogenic signature, this type of simulation would be useful for constraining methane migration mechanisms from source to reservoir.

The multi-component modifications to PFLOTRAN allow for simulation of density-driven flows in a system comprised of methane, salt, and water. Salinity impacts density, which is determined using the Batzle and Wang (1992) equation of state (Eqs. 1-2) for brines given pressure and temperature. ρ_w represents the water density at a given temperature, T (C), and pressure, P (MPa), while ρ_B represents brine density given salinity, S (ppm/10000000), water density, temperature, and pressure.

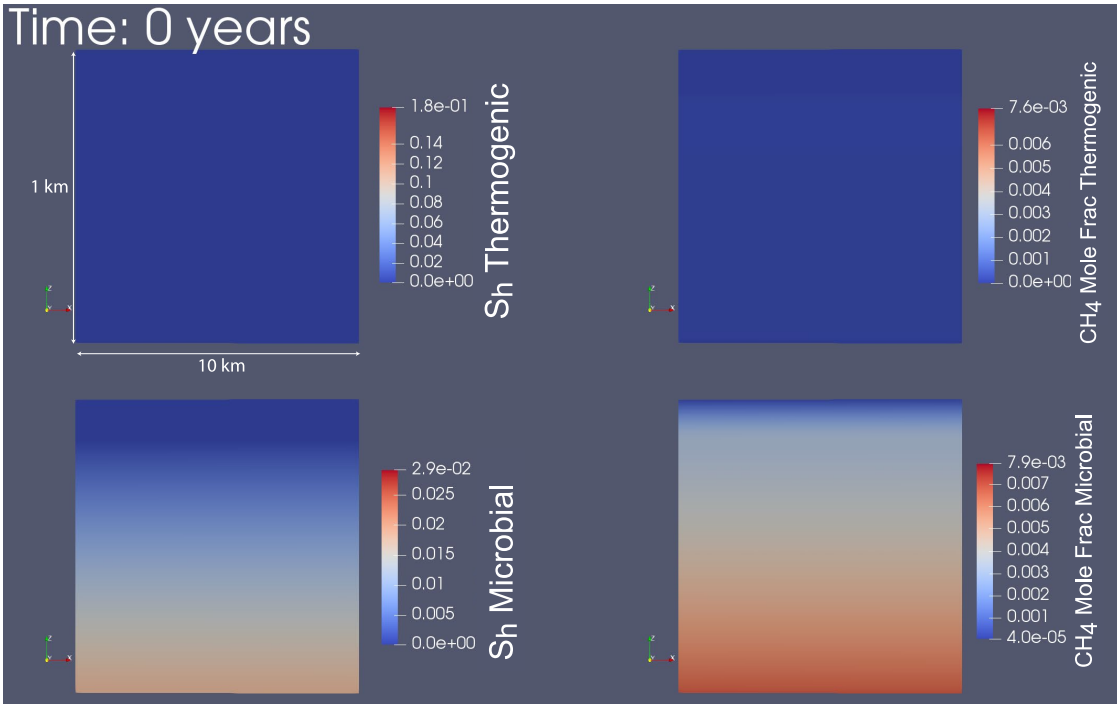


Figure 4-3. Initial distribution of thermogenic and microbial hydrate (left) and mole fraction of thermogenic and microbial methane (right) in a 1 km by 10 km 2D domain.

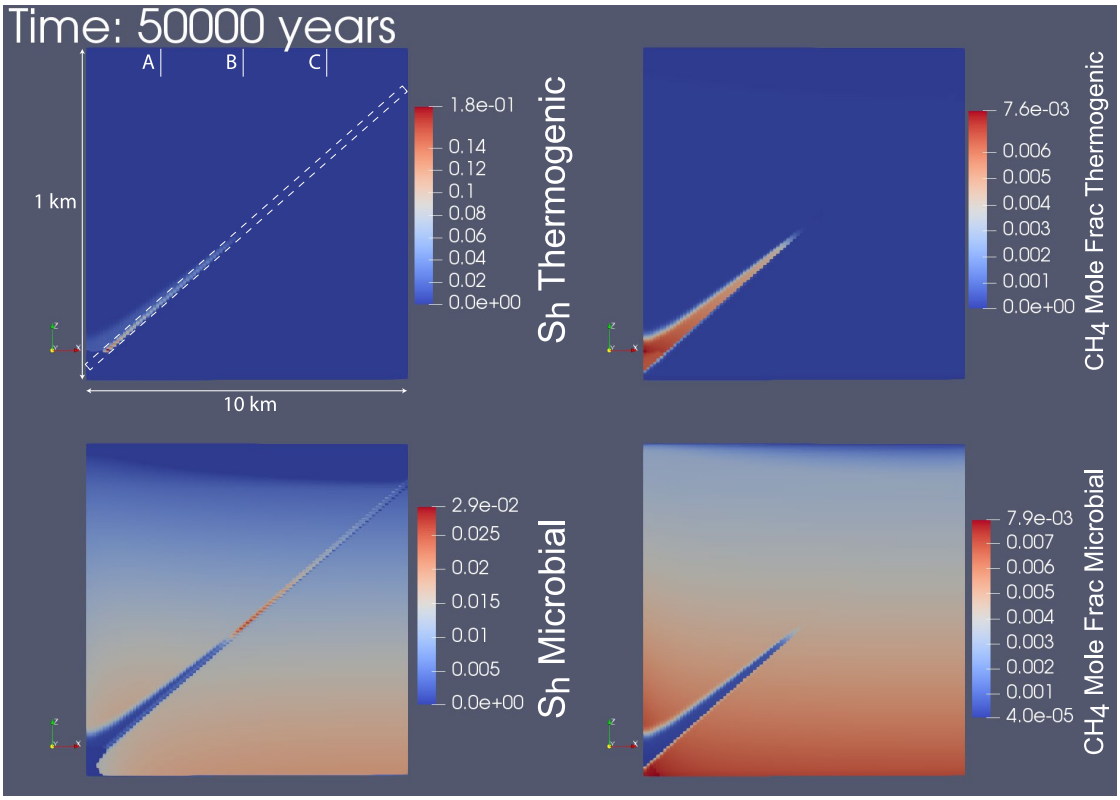


Figure 4-4. Distribution of thermogenic and microbial hydrate (left) and mole fraction of thermogenic and microbial methane (right) in a 1 km by 10 km 2D domain with a dipping high permeability layer where methane of deep thermogenic origin is injected along the dip.

$$\rho_w = 1 + 10^{-6}(-80T - 3.3T^2 + 0.00175T^3 + 489P - 2TP + 0.016T^2P - 1.3 \cdot 10^{-5}T^3P - 0.333P^2 - 0.002TP^2) \quad \textbf{(Eq. 4.24)}$$

$$\rho_B = \rho_w + S\left(0.668 + 0.44S + 10^{-6}\left(300P - 2400PS + T(80 + 3T - 3300S - 13P + 47PS)\right)\right) \quad \textbf{(Eq. 4.25)}$$

Density-driven flows in hydrate systems are initiated when hydrate formation is rapid enough to initiate a density contrast between fluids in a highly permeable layer. As hydrate forms, the salt is excluded in the pore space. This density contrast between the top of the rock layer and the background density results in convection initiated by the slight density change from hydrate formation (Figure 4-5).

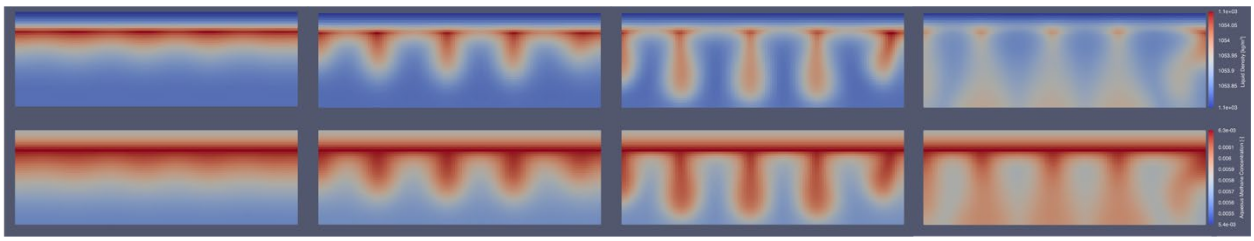
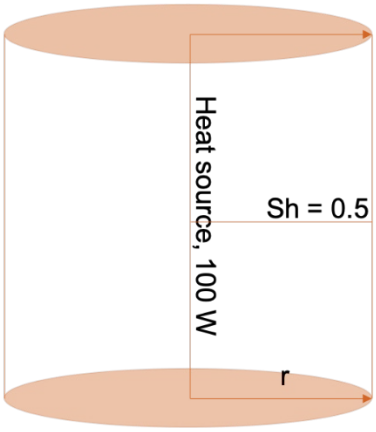


Figure 4-5. Liquid density (top) and aqueous methane concentration (bottom) for a system with convection initiated by hydrate formation.



Functionality was added to simulate CO_2 hydrate formation in a CO_2-water system. The CO_2 hydrate functionality utilizes the pre-existing Span and Wagner equation of state to compute the state properties of gaseous CO_2. A test problem was developed (Fig. 3) to compare the CO_2 hydrate functionality against the CH_4 hydrate functionality. The problem is a cylinder with an initial hydrate saturation of 0.50 and a heat source in the center set at 100 W. As the system heats, hydrate dissociates and partitions into the liquid and gas phases. A small hydrate saturation increase is seen at the dissociation front due to the endothermic nature of hydrate dissociation.

Figure 4-6. Hydrate dissociation test problem schematic.

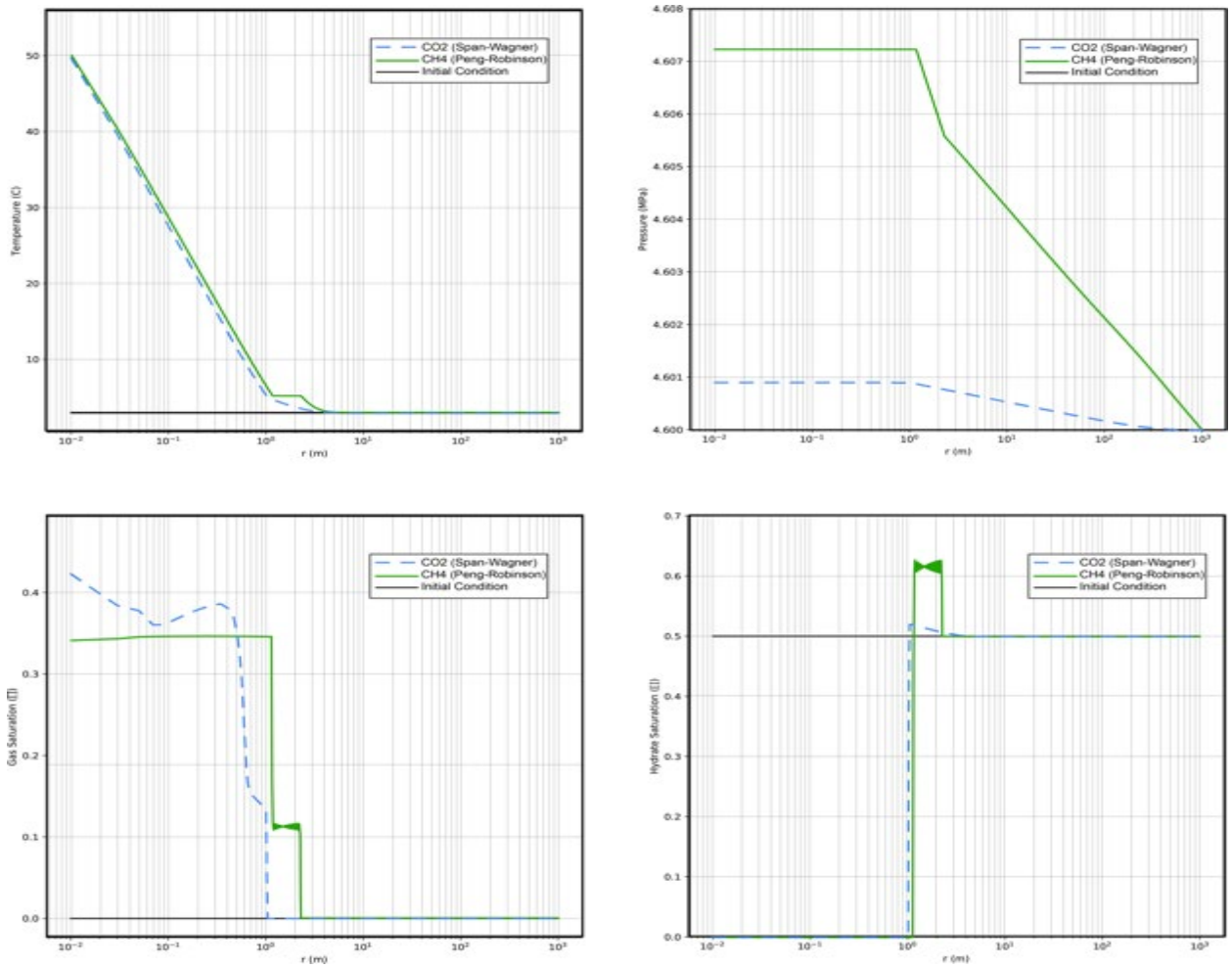


Figure 4-7. Temperature (top left), pressure (top right), gas saturation (lower left), and hydrate saturation (lower right) plots for a cylindrical dissociation test of CO2 and CH4 hydrate.

5. DESIGN OF THE SOFTWARE WORKFLOW

Creating probabilistic maps of geoacoustic and geomechanical sediment properties is achieved via a novel workflow that combines GPSM, Dakota, and PFLOTRAN software packages (Figure 5-1, indicated by the blue boxes). First, sparse global seabed data and high-resolution predictors are input to GPSM, which uses geospatial machine learning to produce high-resolution, 2D, gridded predictions of seabed characteristics with uncertainty. The uncertainty in gridded seabed characteristics are sampled on by Dakota and used to define the input parameters of several unique realizations of marine sediment thermodynamic simulations run by PFLOTRAN. The outcomes are 3D probabilistic maps of gas and hydrate distribution, which are used by GPSM – GeoPhysMod to calculate the corresponding 3D acoustic properties of the marine sediments. The software workflow can be run in pieces, or through the entire cycle, allowing flexibility in its use to produce the desired information.

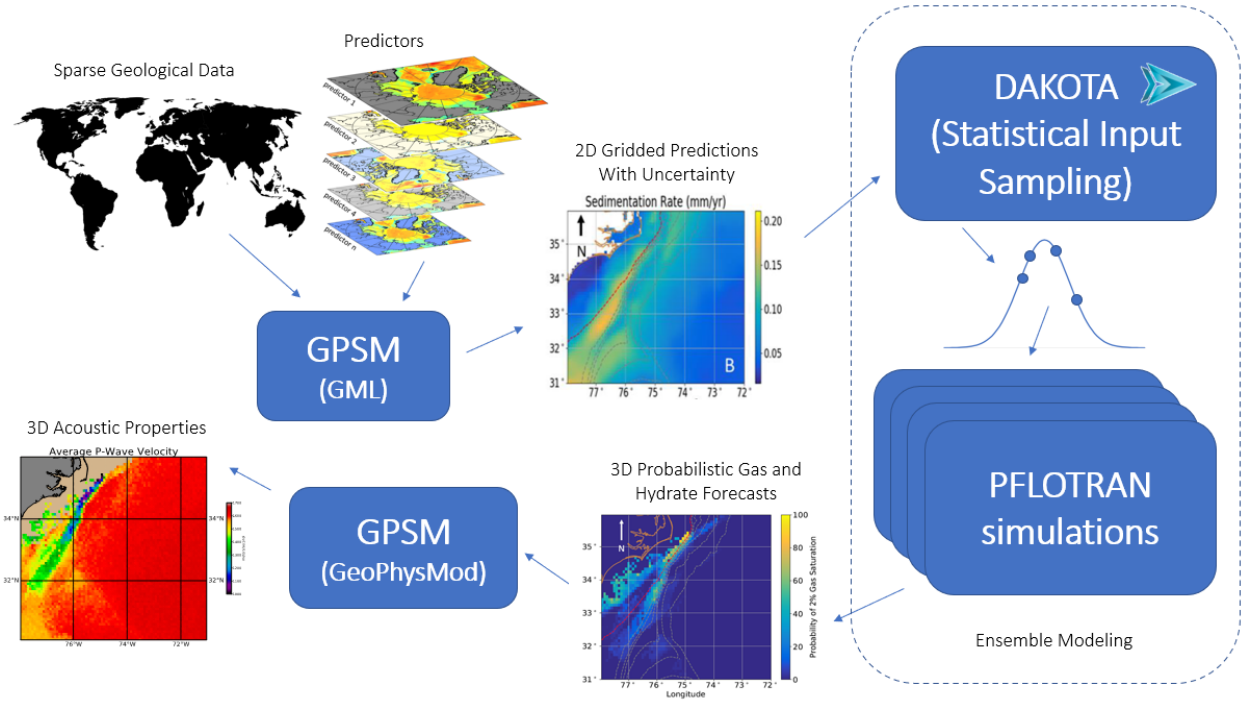


Figure 5-1. Schematic of the software workflow. Sparse seabed data and high-resolution predictors are input to GPSM, which uses geospatial machine learning to produce 2D gridded predictions of seabed characteristics with uncertainty. The uncertainty in gridded seabed characteristics are sampled on by Dakota and used to define the input parameters of several unique realizations of marine sediment thermodynamic simulations run by PFLOTRAN. The outcome are 3D probabilistic maps of gas and hydrate distribution, which are used by GPSM – GeoPhysMod to calculate the corresponding 3D acoustic properties of the marine sediments.

6. MODEL VALIDATION

6.1. International Code Comparison

Implementation of gas hydrate formation and dissociation capability in PFLOTRAN was validated by including simulation results in the 2nd International Gas Hydrate Code Comparison Study (IGHCC2; White et al., 2020). PFLOTRAN simulation results were included in this study for 2 test problems: a radial thermal stimulation problem and a radial depressurization problem. In each problem, a 1D radial domain is initialized with a constant gas hydrate saturation and then gas hydrate is driven to dissociate into water and free gas. In the thermal stimulation problem (Figure 6-1 through Figure 6-3), dissociation is driven by a head flux boundary condition (150 W) at r=0. As heat is applied, hydrate begins to dissociate at r=0 and then over time the hydrate dissociation front propagates away from the heat source. While the far boundary (at r=1500 m) remains unaffected by the heat source, a similarity solution exists. Codes were compared to each other in terms of the similitude variable, r^2/t, for a given set of outputs including hydrate saturation (Figure 6-1), temperature (Figure 6-2), and aqueous saturation (Figure 6-3). PFLOTRAN results (light blue dashed line) are difficult to make out in the comparison because they very nearly match the vast majority of the other codes for hydrate saturation, temperature, and liquid saturation.

In the depressurization problem (Figure 6-4 through Figure 6-7), the same 1D radial domain is subjected to an extraction well at r=0 with a constant extraction flow rate (100 g/s) applied throughout the duration of the simulation. Flow out of the domain reduces the pressure below the pressure for stability of gas hydrate, thus driving hydrate initially present to dissociate into water and gas. Just like the thermal stimulation problem, a similarity solution exists before the pressure wave propagates to the edge of the domain. Gas hydrate saturation (Figure 6-4), pressure (Figure 6-5), gas saturation (Figure 6-6), and temperature (Figure 6-7) are compared across simulators as functions of the similitude variable (r^2/t). The PFLOTRAN solution compares very well to the other simulators for this problem, but differences could be due to differences not covered by the problem description, including equations of state for liquid and gas or gas hydrate phase boundary assumptions.

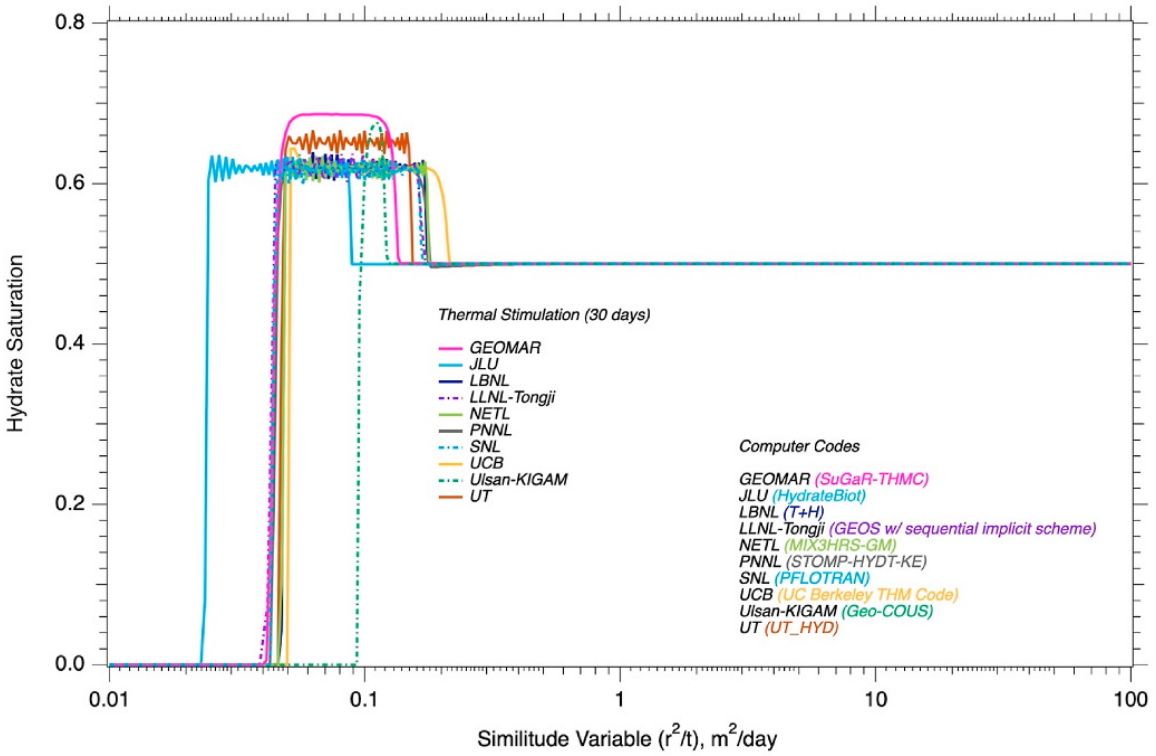


Figure 6-1. Comparison of several simulator results for hydrate saturation during and after thermal stimulation of the reservoir. Adopted from White et al., 2020.

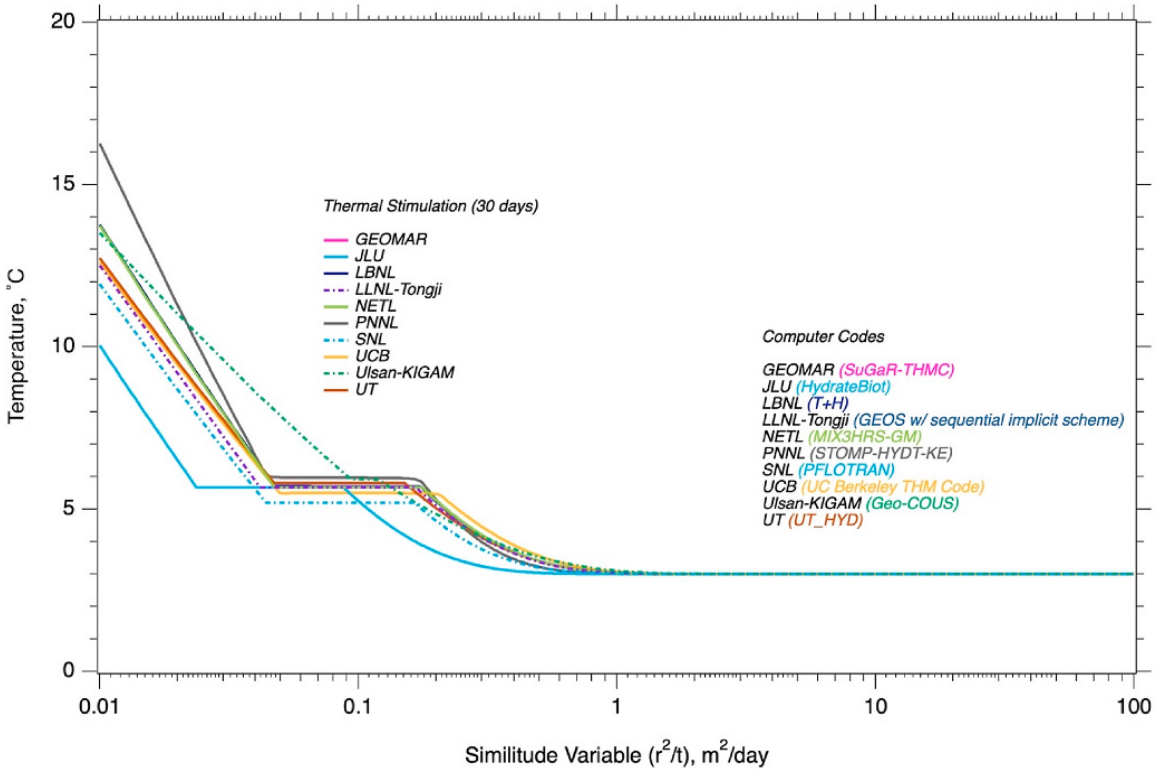


Figure 6-2. Comparison of several simulator results for temperature during and after thermal stimulation of the reservoir. Adopted from White et al., 2020.

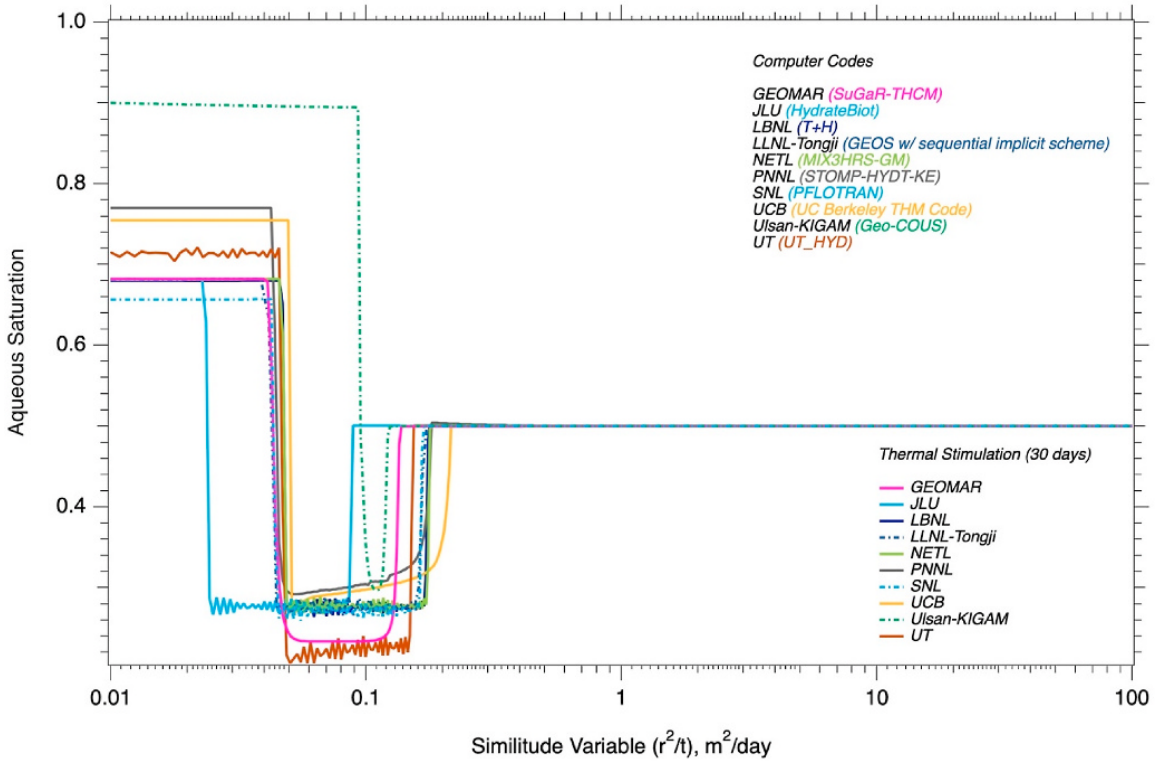


Figure 6-3. Comparison of several simulator results for liquid saturation during and after thermal stimulation of the reservoir. Adopted from White et al., 2020.

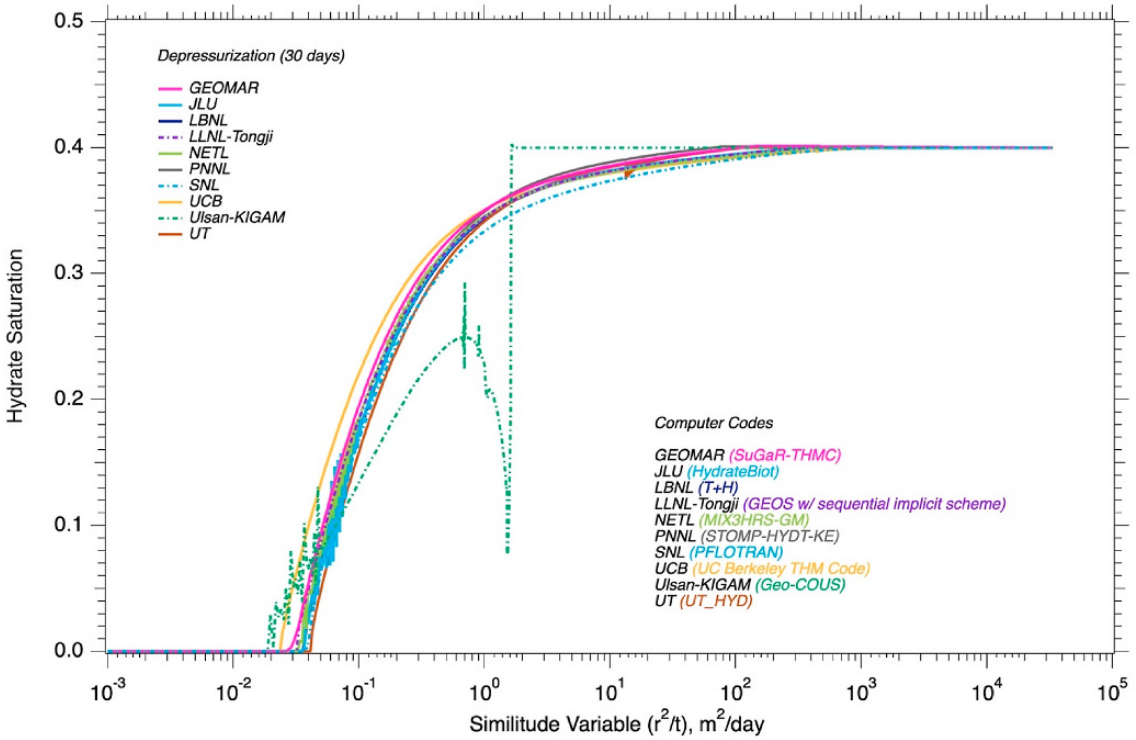


Figure 6-4. Comparison of several simulator results for hydrate saturation during and after depressurization of the reservoir. Adopted from White et al., 2020.

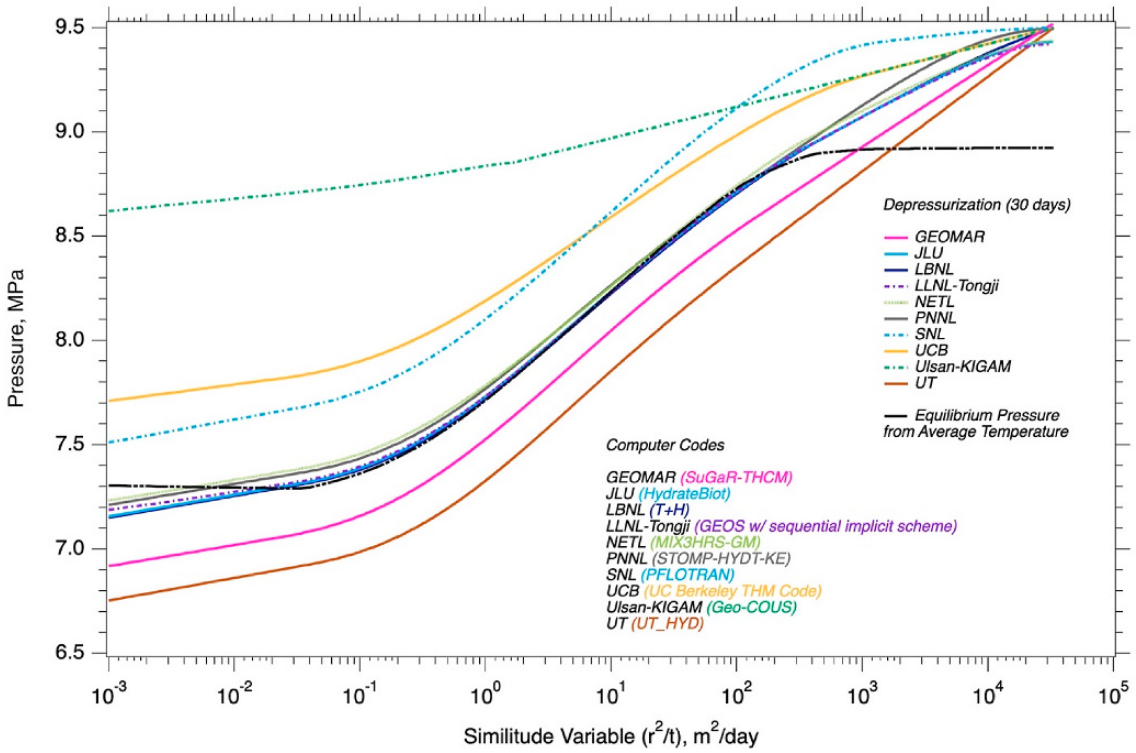


Figure 6-5. Comparison of several simulator results for pressure during and after depressurization of the reservoir. Adopted from White et al., 2020.

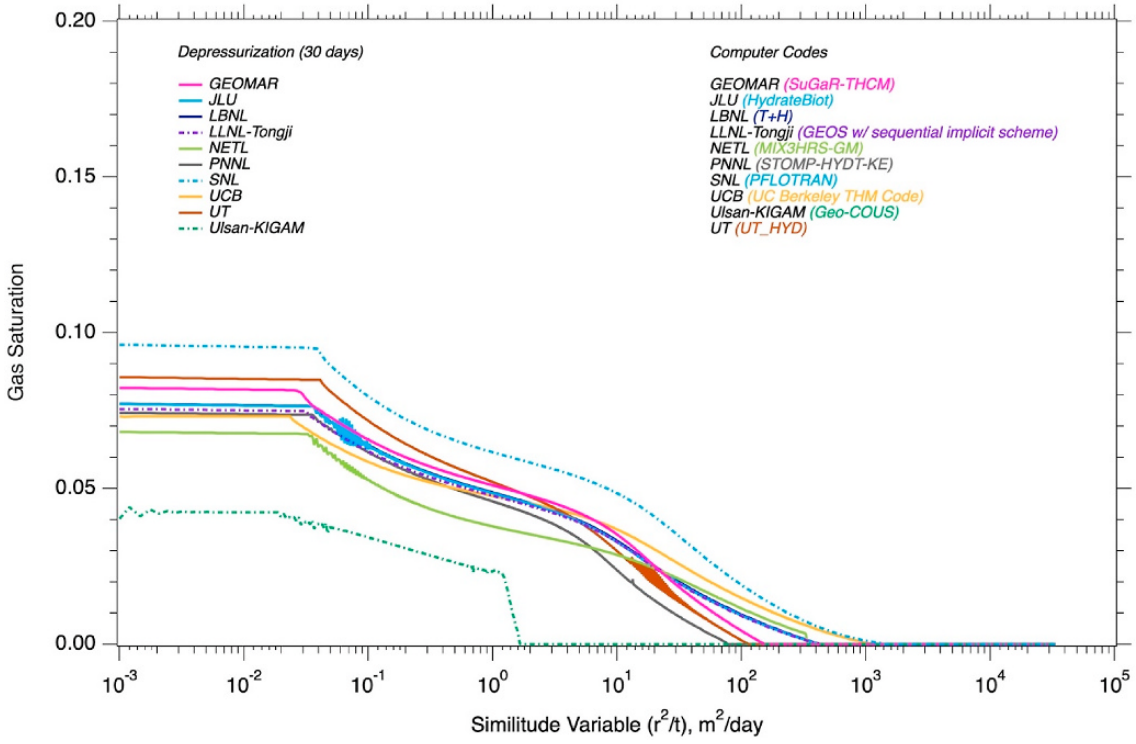


Figure 6-6. Comparison of several simulator results for gas saturation during and after depressurization of the reservoir. Adopted from White et al., 2020.

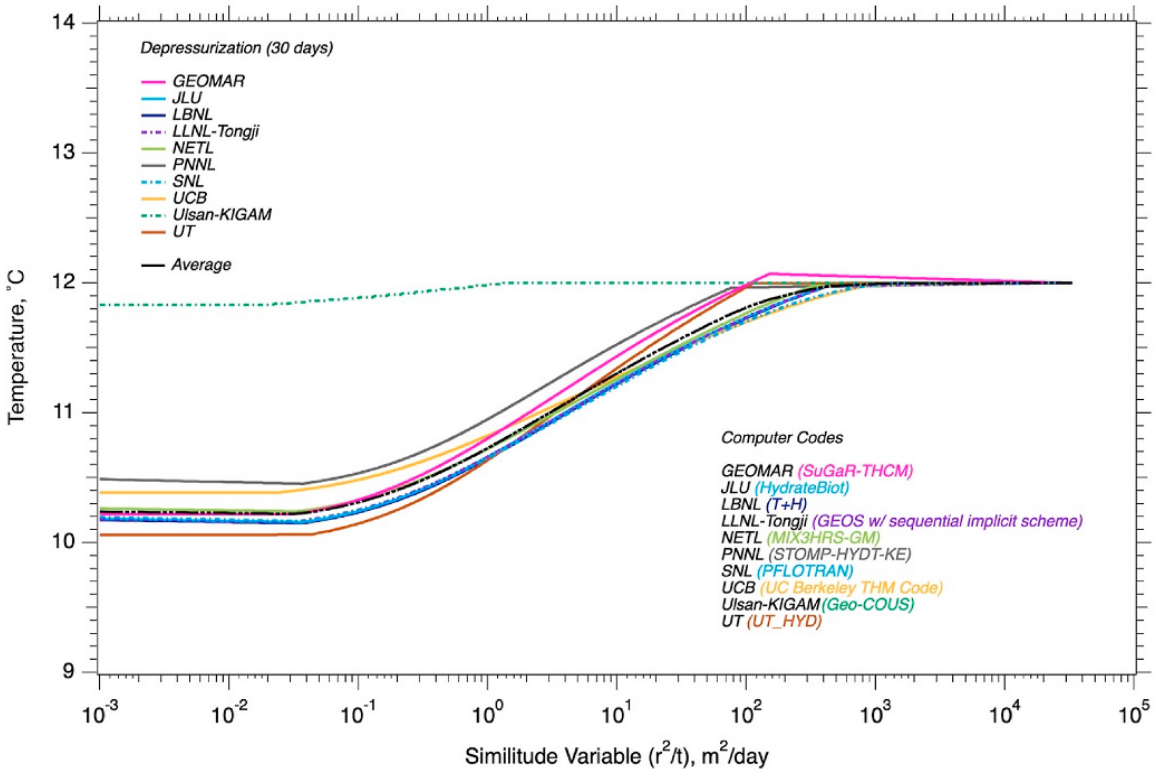


Figure 6-7. Comparison of several simulator results for temperature during and after depressurization of the reservoir. Adopted from White et al., 2020.

6.2. Blake Ridge

Validation of our results has been published previously (Eymold et al., 2021) and the comparison to Bhatnagar et al. (2007) from that work is reproduced here. ODP Site 997 in Blake Ridge (31°50.6' N, 75°28.1' W) is a well-characterized location of gas hydrate occurrence which has been studied extensively (Burwicz & Rüpke, 2019; Dickens et al., 1997; Egeberg & Dickens, 1999; Frederick & Buffett, 2013; Wood & Ruppel, 2000) and offers an opportunity to verify the simulations using PFLOTRAN's HYDRATE mode by comparing these results to previous work (Bhatnagar et al., 2007). Samples of hydrate have been observed and collected at ODP Site 997 that show an average methane hydrate content of 4-7% of porosity between depths of 186-451 mbsf (Egeberg & Dickens, 1999). Outputs from simulations using PFLOTRAN are compared to those simulated by Bhatnagar et al. (2007) in Figure 6-8 and are in good agreement with the observations from Egeberg & Dickens (1999). This evaluation will help foster new study designs and provide a template to verify the simulator against research from other groups to validate the seafloor model based on known values.

The water depth of the site is 2,781 mbsl, seafloor temperature is 3.4°C, and gas hydrate has been collected from depths of 180-240 and 380-450 mbsf (Paull et al., 1996). Previous 1-D simulations of hydrate formation at ODP Site 997 have been conducted, and we compare the results of PFLOTRAN to those from Bhatnagar et al. (2007).

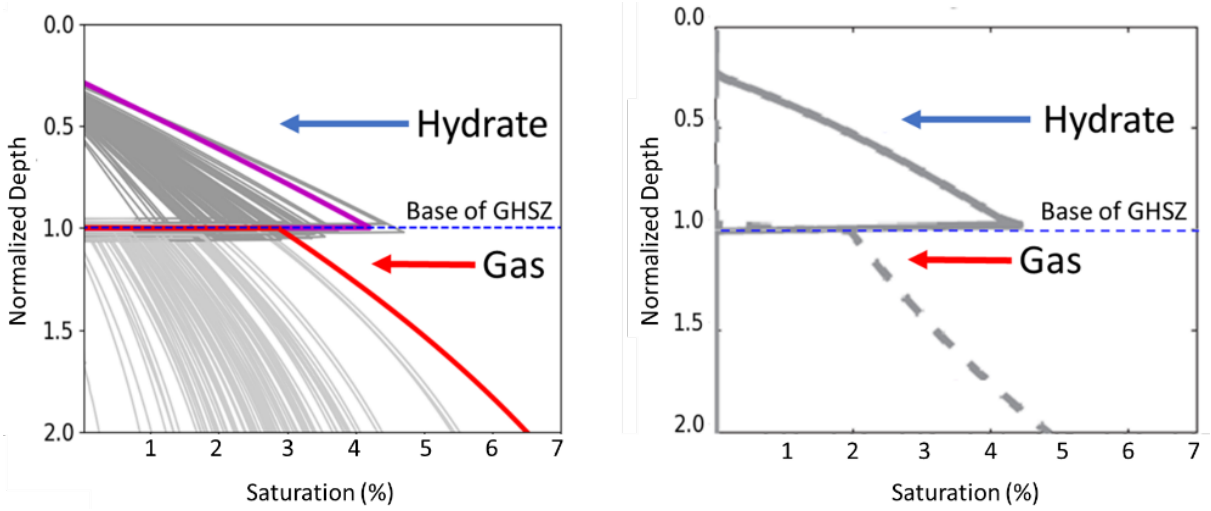


Figure 6-8. Hydrate saturation profiles from 100 simulations are shown in dark gray and the gas saturation profiles are plotted in light gray (left) on the same dimensionless depth scale as Bhatnagar et al. (2007), normalized to 458 mbsf (the base of the GHSZ in their simulations). The simulation that was run with PFLOTRAN using the exact Bhatnagar values displays comparable values of hydrate saturation (magenta) and gas saturation (red). For verification, the output profile shown in Figure 7 of Bhatnagar et al. (2007) is shown in the right panel. Adopted from Figure 11, Eymold et al. (2021).

The 1-D numerical model by Bhatnagar et al. (2007) simulated fluxes of methane leading to accumulation of gas hydrate in both the Cascadia Margin and Blake Ridge. The modeling framework used in that study was based on phase equilibrium and methane solubility, mass balance for methane, water, and organic matter, and porosity reduction due to sedimentation. Estimates for the extent of the GHSZ based on steady state results after approximately 12 Myr were provided from the simulations. The input parameters came from past studies (Davie & Buffett, 2001; Paull et al., 1996) and do not incorporate measurement or uncertainty errors in contrast to our simulations here.

Table 6-1. Model parameters for 1-D gas hydrate simulation from GPSM and those used in Bhatnagar et al. (2007). Simulations of Location 4903 used a fixed value of 1 × 10^{-14} s^{-1} for methanogenesis rate and 0.22 mm/yr for sedimentation rate but sampled on TOC and heat flux. Output profiles for hydrate saturation and gas saturation are plotted in Figure 5.2.1. (Table 3, Eymold et al. 2021)

Variable	GPSM Value	Bhatnagar Value	Units
Methanogenesis Rate	1×10^{-15} - 1×10^{-13}	1×10^{-14}	s^{-1}
Permeability	1×10^{-15}	1×10^{-15}	m^2
Seafloor Depth	2,734.20	2,781	m
Seafloor Temperature	2.34	3.4	°C
Sedimentation Rate	0.0297 ± 0.0063	0.22	mm/yr
Heat Flux	47.47 ± 1.34	62.5*	mW/m^2
TOC	0.828 ± 0.496	1.5	%

* Heat flux was enhanced in the PFLOTRAN simulations in order to adjust the temperature profile and match the GHSZ depth from Bhatnagar et al. (2007).

Most variables for ODP Site 997 used by Bhatnagar et al. (2007) were comparable to the GPSM values at the nearest location (Location 4903 at 31°52.5' N, 75°27.5'W, centered approximately 3.6 km to the northwest) (Table 6-1). The geothermal gradient used to match the Bhatnagar simulation was ~15 mW/m^2 higher than that predicted by GPSM, and the methanogenesis rate was within the range that we used to sample with Dakota for the Blake Ridge PFLOTRAN simulations. Seafloor depth was slightly shallower for GPSM and the seafloor temperature was 1°C colder. The most drastic differences between the parameter sets from GPSM and those used by Bhatnagar et al. (2007) were the lower sedimentation rate in GPSM (previously discussed) and the TOC, which was nearly double the value provided by GPSM.

By setting the sedimentation rate to 0.22 mm/yr and the methanogenesis rate to 1×10^{-14} s^{-1} to match the values used by Bhatnagar et al. (2007), the PFLOTRAN outputs for both hydrate and gas saturation closely approximated the previous study (Figure 6-8). The maximum hydrate saturation values varied depending on TOC for each simulation, but the base of the GHSZ averaged 464.1 mbsf and was in good agreement with the past prediction of 458 mbsf. The bottom of the hydrate saturation profile predicted by GPSM ranged from 436-486 mbsf (left panel of Figure 6-8), depending on the heat flux used in each simulation. While the gas saturation profiles were comparable in location and scale, the discrepancy in shape and exact value may be attributed to the different equations of state used in each study.

The structure of the hydrate saturation profiles depends on the methanogenesis rate used; by following the higher methanogenesis rate of $\lambda = 1 \times 10^{-13}$ s^{-1} (Malinverno, 2010), the GHSZ extended deeper and maximum hydrate saturation was lower. In these simulations where the Bhatnagar value of $\lambda = 1 \times 10^{-14}$ s^{-1} was used instead, the saturation profile was sharper and had a higher maximum value, suggesting that hydrate formation strongly depends on this rate. This reaction rate is temperature dependent and will be highest just beneath the z_{SMT} in the profile (Wehrmann et al., 2011; You et al., 2019). Future simulations should vary the z_{SMT} based on geography and λ based on depth (temperature) if information is available.

Despite differences based on the GPSM values in a nearby location, the Dakota-PFLOTRAN workflow described here produces comparable results to those predicted at ODP Site 997 by Bhatnagar et al. (2007). As this workflow is applied to other geographic locations in the future, these verification tests will be increasingly necessary to refine the model parameters.

7. MODEL DEMONSTRATIONS

7.1. Blake Ridge Gas Hydrate Province

The results from simulations at Blake Ridge study area were published previously (Eymold et al., 2021) and are reproduced here. The Blake Ridge study area (31°N-36°N and 78°W-72°W) domain covers 84 × 72 (N = 6,048) geographic locations equally spaced at 5 arc-minutes in both latitude and longitude, with parameters generated via GPSM. Of these 6,048 locations, 5,293 were located offshore and were simulated 100 times each using 48 processors on a local Linux cluster, with a 1,000 m 1-D sediment profile at 1 m depth discretization.

Gas hydrate formation occurs at depths greater than 500 mbsl in the Blake Ridge study area (Figure 7-3). Two peak areas of formation occur, with one located in the area north of 35°N 75°W between 500-1,500 m depth and a second located east of 34°N 76°W between 1,000-1,500 m depth. The area north of 35°N 75°W has shown widespread methane seepage at sites down to 600 m depth east of

Cape Hatteras, possibly related to hydrate dissociation (Skarke et al., 2014). The peak east of 34°N 76°W is located near the Carolina Rise, where BSRs are extensive and gas hydrate is prone to dissociation due to warming Gulf waters (Phrampus & Hornbach, 2012; Ruppel et al., 1995). The greatest amount of hydrate (3,746.9 kg) occurred at 35°17.52'N 74°52.5'W and a seafloor depth of 1,324.5 mbsl.

Essentially no hydrate occurred in the southwest portion of the study area between 500-1,500 m depth after the simulations reach steady state, suggesting that regions without occurrence may indicate unfavorable conditions for hydrate formation. This region is characterized by lower TOC values than in the study area above 33°N along with elevated values of sedimentation rate. Since rapid burial prevents microbial breakdown of organic material at the seafloor and is associated with increased levels of preservation at depth (Burwicz & Rüpke, 2019), faster sedimentation rates could lead to more methane generation in the sediment column below the SMT. However, since the source of methanogenesis in our model is exclusively the organic matter brought to depth by burial, exceptionally rapid sedimentation rates may drive the TOC through the GHSZ faster than it can be metabolized into methane. This could explain simulations that resulted in no hydrate (or free gas) formation despite high values of TOC and appropriate pressure and temperature conditions.

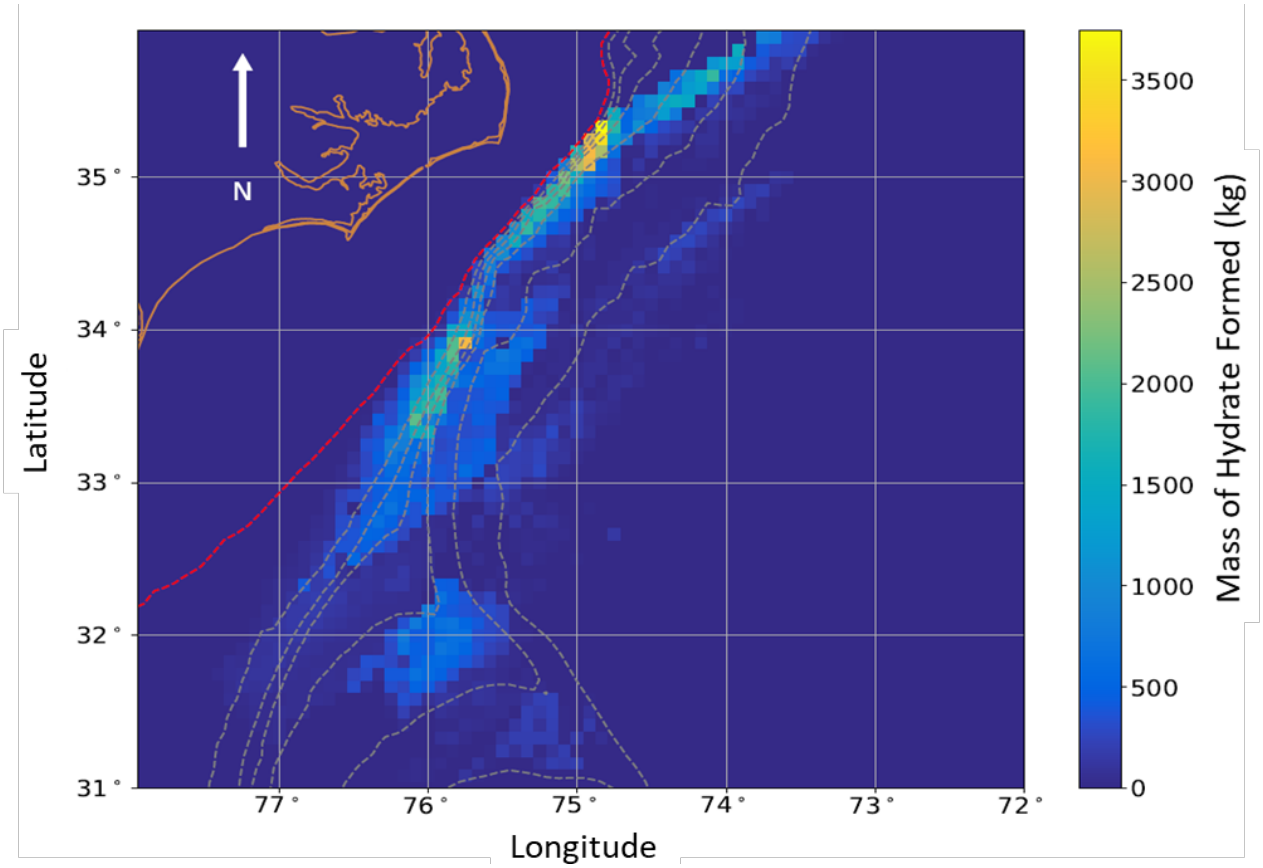


Figure 7-1. Map of average hydrate formed in sediment columns in Blake Ridge, determined from the integration of hydrate formation in the entire sediment column for all simulations (Equation 3.1). The mass of hydrate formed is shown in kg and was calculated based on the ρ_h = 920 kg/m^3 and the hydrate saturation at each depth in the profile. Adopted from Figure 5, Eymold et al. (2021).

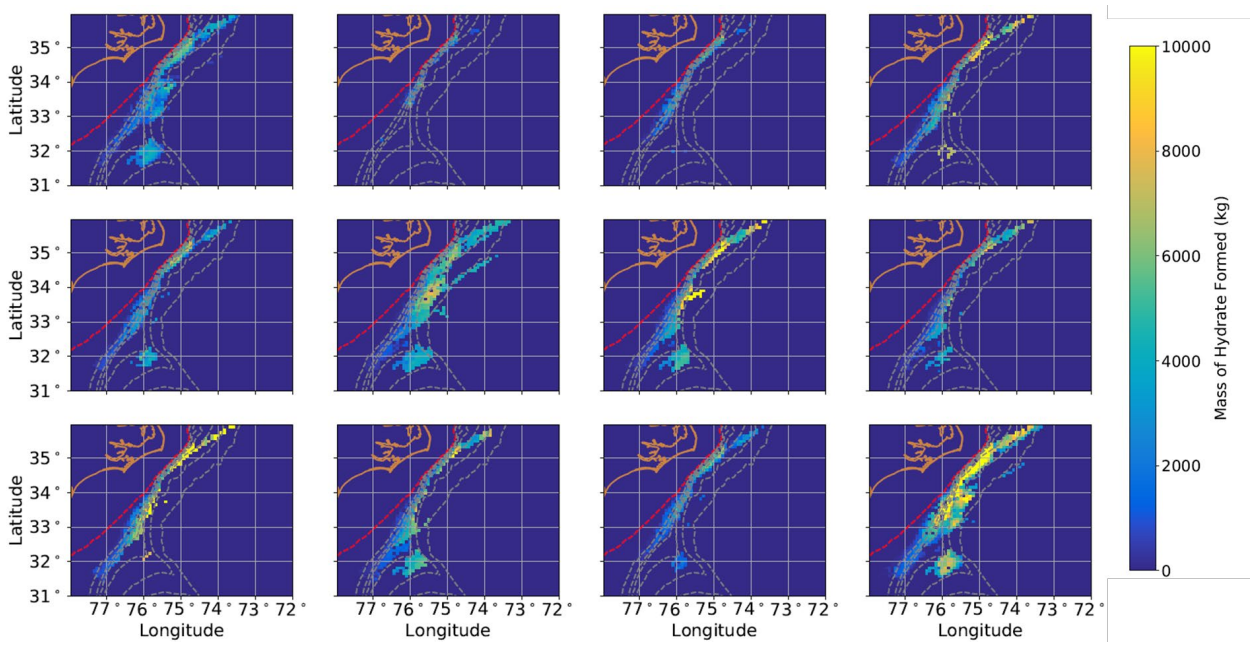


Figure 7-2. Hydrate formation from 12 realizations of Blake Ridge simulations. Areas that formed gas hydrate are plotted from light blue to yellow as mass of hydrate formed increases. Adopted from Figure 6, Eymold et al. (2021).

Each of the 100 simulations differs based on the variables used at each location. The formation results in each of the 100 simulations are stored as a single mapped output for all locations in the study area. By going through these individual runs, hydrate formation can be similar between the simulations but is not identical between runs due to variation of the parameters used for each location (Figure 7-2). Hence, these realizations and the formation rates can be used to produce statistical forecast maps of the study area going forward.

Calculations were also made to determine the percentage of simulations where free gas was formed in the study area (Eqn. 3.2). The average mass of free gas formed in each simulation (Figure 7-3) and the number of simulations where at least 1% gas saturation and 2% gas saturation occurred in the sediment column (Figure 7-4) can be used to evaluate the likelihood that gas is present in locations near Blake Ridge. While no hydrate formed in locations shallower than 500 mbsl, much of the nearshore region routinely formed free gas during the 10 Myr simulations. Near both hydrate formation peaks, free gas almost always reached at least the 1% saturation threshold.

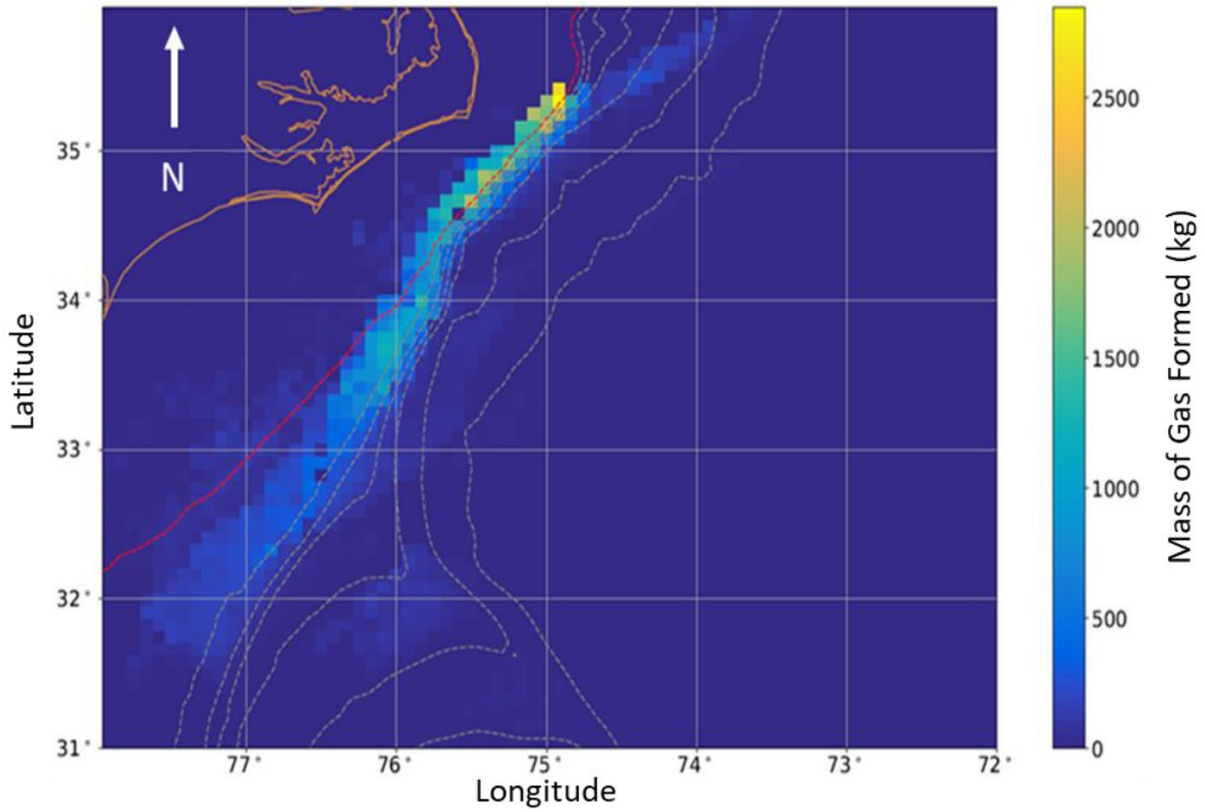


Figure 7-3. Map of average total mass of free gas formed in sediment columns in Blake Ridge. The mass of gas formed is shown in kg and was calculated based on the specific mole fraction (X_g), density (e), and saturation (S_g) of methane at each depth in the profile. Adopted from Figure 9, Eymold et al. (2021).

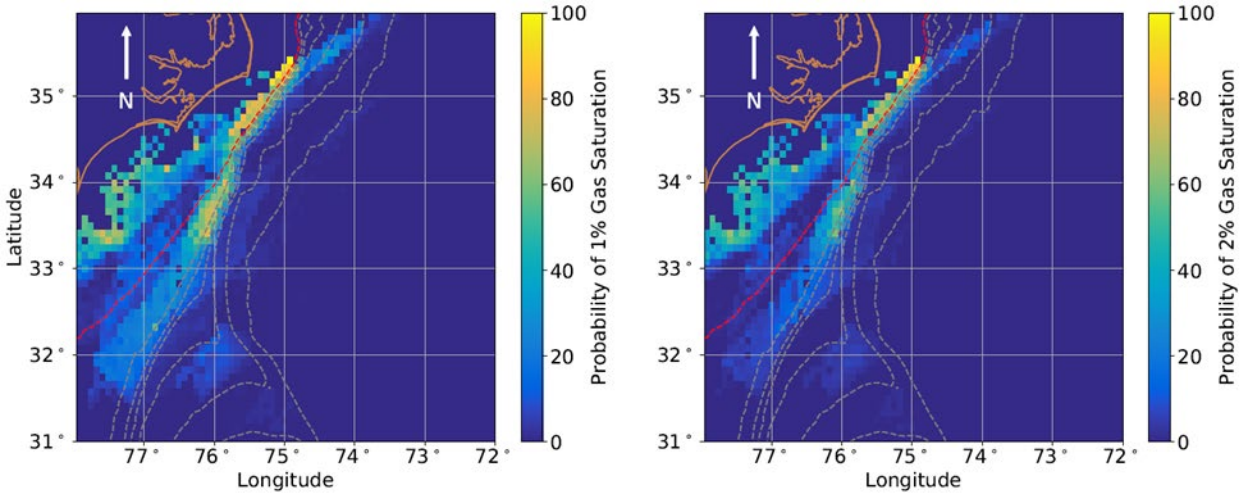


Figure 7-4. Map of percentage of simulations where gas saturation reached at least 1% (left) and at least 2% (right) in the sediment column. Adopted from Figure 10, Eymold et al. (2021).

7.2. North Atlantic Gas Seep Provinces

Ocean Drilling Program (ODP) boreholes give an insight into locations where methane gas and methane hydrate occur along the U.S. Atlantic margin. Along the upper continental slope, offshore of New Jersey, ODP boreholes report subsurface microbial methane, indicating methanogenesis in the area (Shipboard Scientific Party, 1994a,b,c,d, 1998a). South of this region, offshore of North Carolina and South Carolina, seismic profiles from the ODP show a strong BSR at some locations, inferring the presence of hydrate in the area (Dickens et al., 1997; Holbrook et al., 1996; Shipboard Scientific Party, 1996a,b). In global surveys summarizing free gas distribution in marine sediments, Fleischer et al. (2001) and Judd (2003) both note occurrences of free gas along the U.S. Atlantic margin. Evidence of gas seeps has also been documented by Hill et al. (2004) and Brothers et al. (2014) who note instances of pockmarks and blowouts caused by gas escaping from the seafloor.

More recently, investigations by Skarke et al. (2014) have found multiple instances of methane gas leakage from the seafloor along the U.S. Atlantic margin at a higher concentration than previously thought. Skarke et al. (2014) identified 570 gas plumes on the northern U.S. Atlantic margin, with 440 of these seeps (77%) lying between the shelf break and 600 m water depth. The location of these plumes would lie just shallow of the methane hydrate stability zone. The seeps in this area were further explored by Prouty et al. (2016) who suggest that seepage may have begun as early as 15 kya in the Baltimore Canyon slope field, and between 1-3 kya at the deeper Norfolk seep field.

The global model of seafloor TOC was extended by Eymold et al. (2021) to model the formation of gas hydrates at Blake Ridge. Combining global models from GPSM for seafloor TOC, heat flux, and sedimentation rate, Eymold et al. (2021) modeled the burial of organic carbon on the seafloor, and the resulting generation of methane gas and hydrate. They found two main areas of gas hydrate formation along the U.S. Atlantic margin which corresponded to locations where methane seepages had previously been located. For a majority of the locations modeled, Eymold et al. (2021) found that TOC was the most strongly correlated variable to forming gas hydrate (other variables were sedimentation rate, methanogenesis rate, and heat flux).

We used the GPSM-Dakota-PFLOTRAN workflow to predict seafloor TOC and gas and hydrate abundance at several sites along the continential shelf break. For GPSM, we used k-nearest neighbor regression with k = 5 following Lee et al. (2019). For the PFLOTRAN simulations, A period of 120,000 years was chosen, representing the time period of a glacial cycle. Dakota was used to provide a distribution of results by sampling the PFLOTRAN input parameters TOC, sedimentation rate, and heat flux (Table 7-1). Specifically, Dakota used Latin hypercube sampling to provide a distribution of input variables to model with PFLOTRAN. Table 6.2.1 summarizes the input variables used in the PFLOTRAN/Dakota model, many following the work of Eymold et al. (2021).

Table 7-1. Input parameters for PFLOTRAN simulation.

Parameter	Status	Value	Units
Sedimentation Rate	From Restreppo et al. (2020)	Variable	m/s
TOC	Modeled from GPSM (max 5%)	Variable	%

Parameter	Status	Value	Units
Heat Flux	0.048±0.002 (Fuchs et al., 2021)	Variable	W/m^2
Porosity	Calculated	Variable	-
Methanogenesis Rate	Fixed	$5x10^{-14}$	1/s
Geothermal Gradient	Variable, base on porosity and heat flux	Variable	°C/m
Pressure Change Limit	Max change per time step	100000	Pa
Temperature Change Limit	Max change per time step	1	°C
Gravity	Fixed	-9.8	m/s^2
Labile Portion of TOC	75% of TOC	Variable	%
Conversion Factor of Methane	Fixed	2241	kg/m^3
Diffusion Coefficient of Methane	Fixed	$1x10^{-9}$	m^2/s
Gas Viscosity	Methane	$1.1x10^{-5}$	Pa.s
Tortuosity	Fixed	1.4	-
Rock Density	Fixed	2700	kg/m^3
Thermal Conductivity (dry)	Fixed	1	W/m/°C
Thermal Conductivity (wet)	Fixed	1	W/m/°C
Heat Capacity	Fixed	830	J/kg/°C
Permeability	Fixed	$1x10^{-15}$	m^2
SMT Depth	Fixed	15	mbsf
Van Genuchten Pressure	Fixed	$5.8x10^{-4}$	1/Pa
Van Genuchten Pore Size Factor	Fixed	0.189	-
Liquid Residual Saturation	Fixed	0.1	-

Parameter	Status	Value	Units
Max Capillary Pressure	Fixed	$1x10^{8}$	Pa
Gas Residual Saturation	Fixed	0.15	-
Initial Liquid Pressure	Initial Condition	Variable	Pa
Initial Mole Fraction	Initial Condition	0.001	-
Initial Temperature	Initial Condition	Variable	°C
Initial Hydrostatic Liquid Pressure	Initial Condition	Variable	Pa
Initial Dirichlet Hydrate Saturation	Initial Condition	$1x10^{-8}$	-
Initial Dirichlet Temperature	Initial Condition	Variable	°C
Neumann Liquid Flux	Bottom Boundary Condition	0	m/yr
Neumann Gas Flux	Bottom Boundary Condition	0	m/yr
Neumann Energy Flux	Bottom Boundary Condition	Q	W/m^2
Hydrostatic Liquid Pressure	Top Boundary Condition	Variable	Pa
Dirichlet Mole Fraction	Top Boundary Condition	0.001	-
Dirichlet Temperature	Top Boundary Condition	Variable	°C

Dakota sampled TOC, sedimentation rate, and heat flux using a normal distribution. As discussed earlier, TOC predictions and standard deviations were modeled using GPSM predictor grids from Phrampus et al. (2020) and observed data points from Seiter et al. (2004). To avoid values of zero, a lower bound of TOC was set to 0.01%. An upper bound of TOC was set to 5%, like the GPSM model, to avoid unrealistically high values. Sedimentation rates and standard deviations were determined from Restreppo et al. (2020) who modeled global oceanic sediment accumulation rates using GPSM at a 5 x 5 arc-minute resolution. Bounds on sedimentation rates were set to a minimum of $1x10^{-14}$ m/s ($3.16x10^{-5}$ cm/yr) and to a maximum of 1 m/s. For the studied area, heat flux was sampled from Global Heat Flow Compilation Group (2013), and was set to 48 ± 2 mW/m^2. A methanogenesis rate of $\lambda = 5x10^{-14}$ s^{-1} was chosen based on estimates of $\lambda = 1x10^{-14}$ s^{-1} from Bhatnagar et al. (2007) and $\lambda = 1x10^{-13}$ s^{-1} from Malinverno (2010). This value for methanogenesis rate also lies between the constraints of $1x10^{-15} \leq \lambda \leq 1x10^{-13}$ s^{-1} set by Eymold et al. (2021). Porosity was calculated at depth using the relationship for marine sediments presented by Kominz et al. (2011):

$$\phi = 0.775e^{-\frac{z}{1251}} \quad \textbf{(Eq. 7.1)}$$

where φ is porosity and z is depth below seafloor in m. Daigle et al. (2020) discuss the validity of this porosity model.

In determining which points to model hydrate formation, the study area was narrowed to 36.6°N–38°N and 74.5°W–73.1°W. This area shows some of the higher TOC estimates in the modeled region. Table 6.2.2 summarizes the locations where PFLOTRAN and Dakota were used to model hydrate and gas generation. Four points were chosen along the dip of the continental shelf. Since the goal of the geomechanical model was to model hydrate dissociation at the feather edge of the hydrate stability zone, four additional points were chosen from this area where seafloor depth was between 485 and 510 m (Figure 7-5). In addition, one location was chosen just shallower than the feather edge to show gas generation with no hydrate.

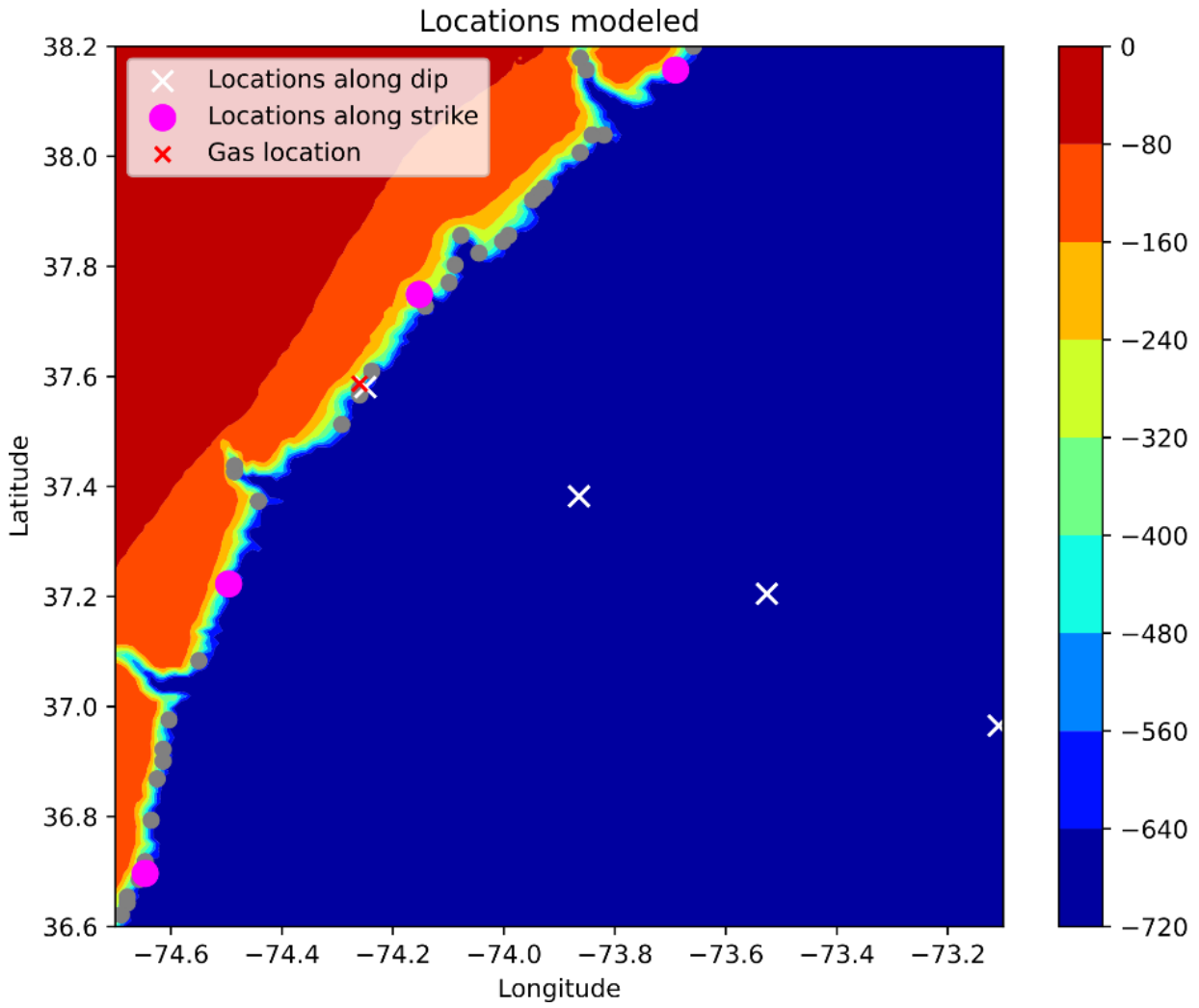


Figure 7-5. Locations chosen to model with PFLOTRAN and Dakota. Colorbar is water depth in m. Gray dots represent locations along the continental slope where seafloor depth is between 485 and 510 m. From these, four locations along the strike of the continental shelf were modeled (magenta dots). Five locations were chosen along the shelf dip. Four of these locations produced hydrate (white x) and one location shallower than 485m produced only gas (red x). To emphasize

the shelf edge where the feather edge of the hydrate stability zone is expected, a maximum depth of 720m was graphed. Refer to Table 6.2.2 for depths at each location.

The Global Multi-Resolution Topography (GMRT) Synthesis was used to determine seafloor depths (Ryan et al., 2009). The GMRT Synthesis provides high resolution bathymetry data for almost 10% of the global ocean. Individual locations can be queried through the GMRT PointServer Web Service to retrieve accurate water depths at specific latitude and longitude locations.

Table 7-2. Locations chosen for PFLOTRAN simulation.

Location #	Lat, Lon	Water depth (m)	Note
Along shelf strike			
1	36.6966, -74.6463	-508	Feather edge
2	37.2228, -74.4960	-510	Feather edge
3	37.7490, -74.1523	-497	Feather edge
4	38.1570, -73.6906	-487	Feather edge
Along shelf dip			
5	37.5865, -74.2600	-434	
6	37.5805, -74.2494	-651	Feather edge
7	37.3822, -73.8646	-1901	
8	37.2049, -73.5260	-2595	
9	36.9652, -73.1084	-3061	

Seafloor temperatures were calculated at a given depth through a polynomial regression. Data from Boyer et al. (2018) provided 47 temperature profiles within the 36.6°N–38°N and 74.5°W–73.1°W study area. Since the area of focus was the GHSZ feather edge, near the continental slope, a visual inspection of profile location was used to choose temperature profiles near the continental slope (Figure 7-6). This left 10 temperature profiles, and a 6th order polynomial regression was used to create a model of temperature with depth (Equation 7.2):

$$T(z) = 5.450 \times 10^{-16} z^6 - 1.641 \times 10^{-12} z^5 + 1.763 \times 10^{-9} z^4 - 7.746 \times 10^{-7} z^3 + 1.075 \times 10^{-4} z^2 - 8.606 \times 10^{-3} z + 12.504 \quad \textbf{(Eq. 7.2)}$$

where here z is water depth.

The weight percent of total organic carbon (TOC) at the seafloor predicted with GPSM is shown in Figure 7-7a. The standard deviation of these predictions is in Figure 7-7b. The highest TOC values were predicted along the line from (35.4°N, 75.0°W) to (39.0°N, 72.0°W). Comparing this region in the prediction map to the standard deviation map, known seafloor TOC values are sparse in this

area and TOC standard deviations are higher than the rest of the predicted grid. The locations chosen along the shelf strike can be found in this area. In addition, the mean inexperience is shown in Figure 7-7c. The mean inexperience is calculated as the average distance in parameter space from the predicted location to its 5 nearest neighbors. Compared to the standard deviation map, a relatively low mean experience can be seen along the line from (35.4°N, 75.0°W) to (39.0°N, 72.0°W). Thus, although there is a high variance of predictions, there are other locations globally that are parametrically close to locations along this shelf strike line.

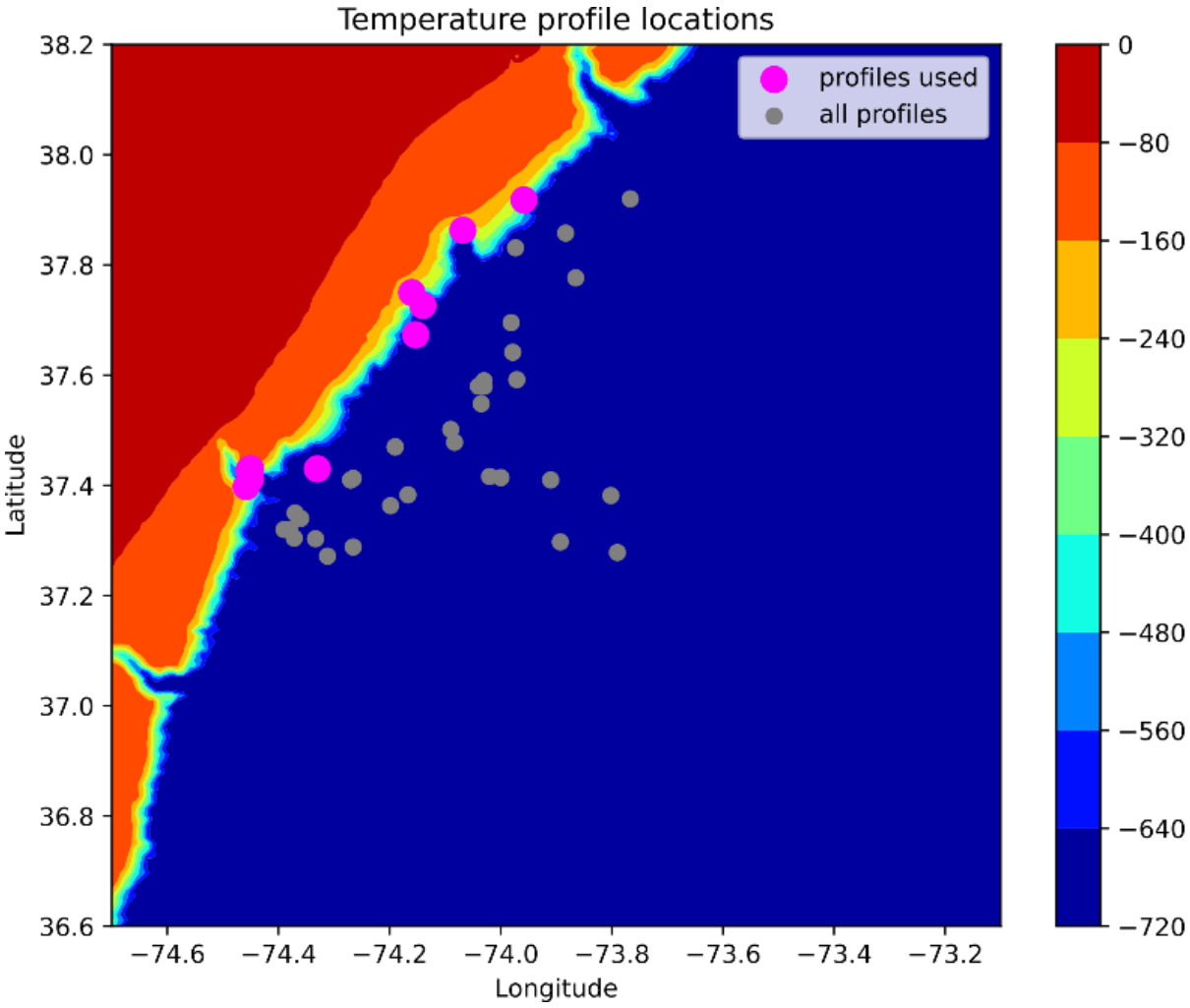


Figure 7-6. Temperature profiles from Boyer et al. (2018) within the study area. Those used to construct Eq. 7.2 are shown with magenta circles.

Average sedimentation rates were just over 0.007 cm/yr with standard deviations around 0.1 cm/yr (Restreppo et al., 2020). Location 9 had a much lower predicted sedimentation rate than the other eight locations, corresponding to its further distance from the shelf edge and seacoast, as well as its overall depth. Average TOC predictions were around 1-2% dry weight. Location 1 had a relatively low average TOC prediction of 0.095% dry weight as the location is on the southern tip of the increased TOC zone between (35.4°N, 75.0°W) and (39.0°N, 72.0°W). Heat flux for all simulations were between 42.315 and 52.168 mW/m^2.

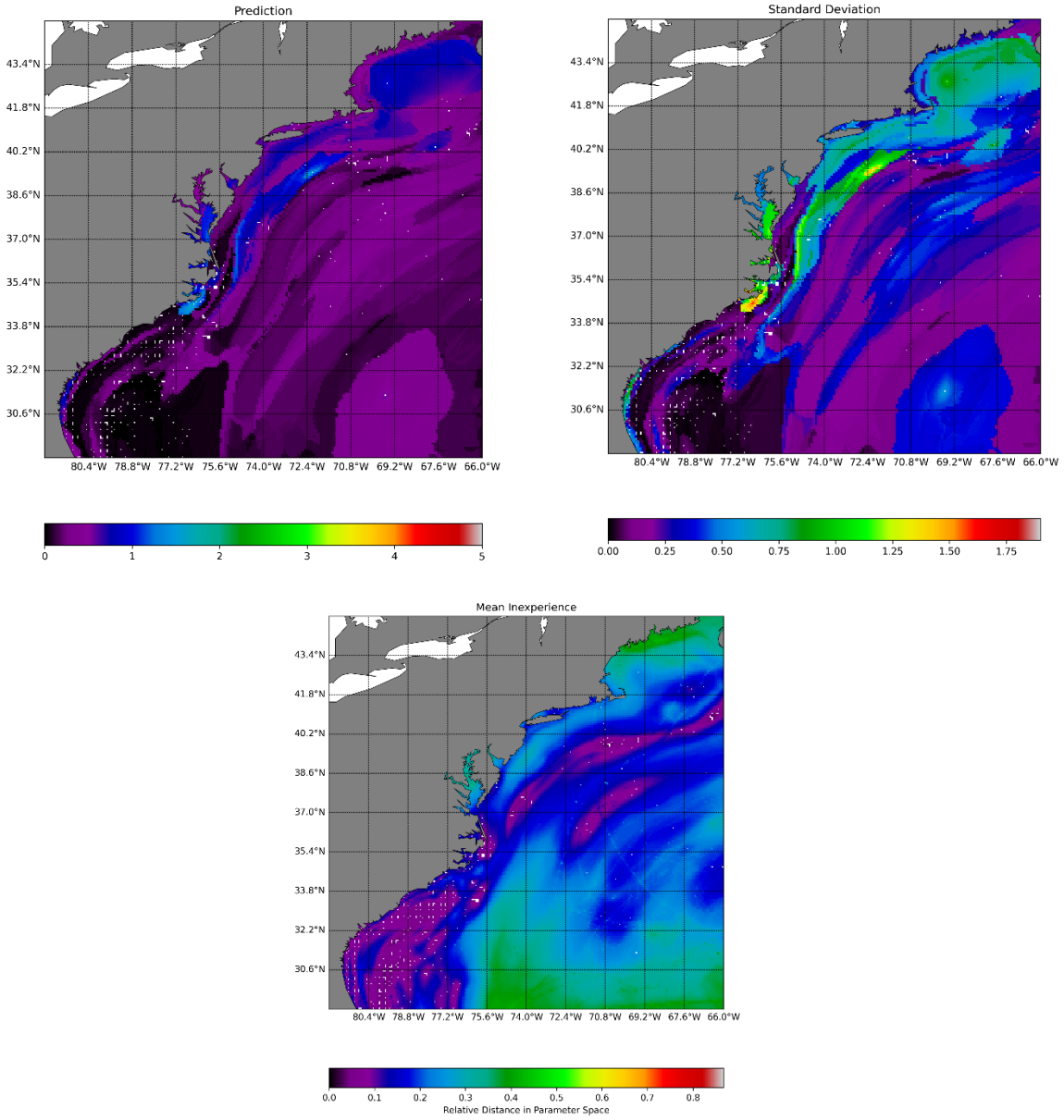


Figure 7-7. (a) Mean predicted TOC. (b) Standard deviation. (c) Mean inexperience.

At each location, 50 simulations were run using PFLOTRAN with a sampling of inputs done by Dakota. For an individual location, the resulting output profiles of gas, hydrate, and temperature could be plotted against depth, and a base of hydrate stability can be calculated. Figure 7-8 shows the hydrate saturation profiles with depth for the nine study locations and Figure 7-9 shows the gas saturation profiles with depth for the nine study locations.

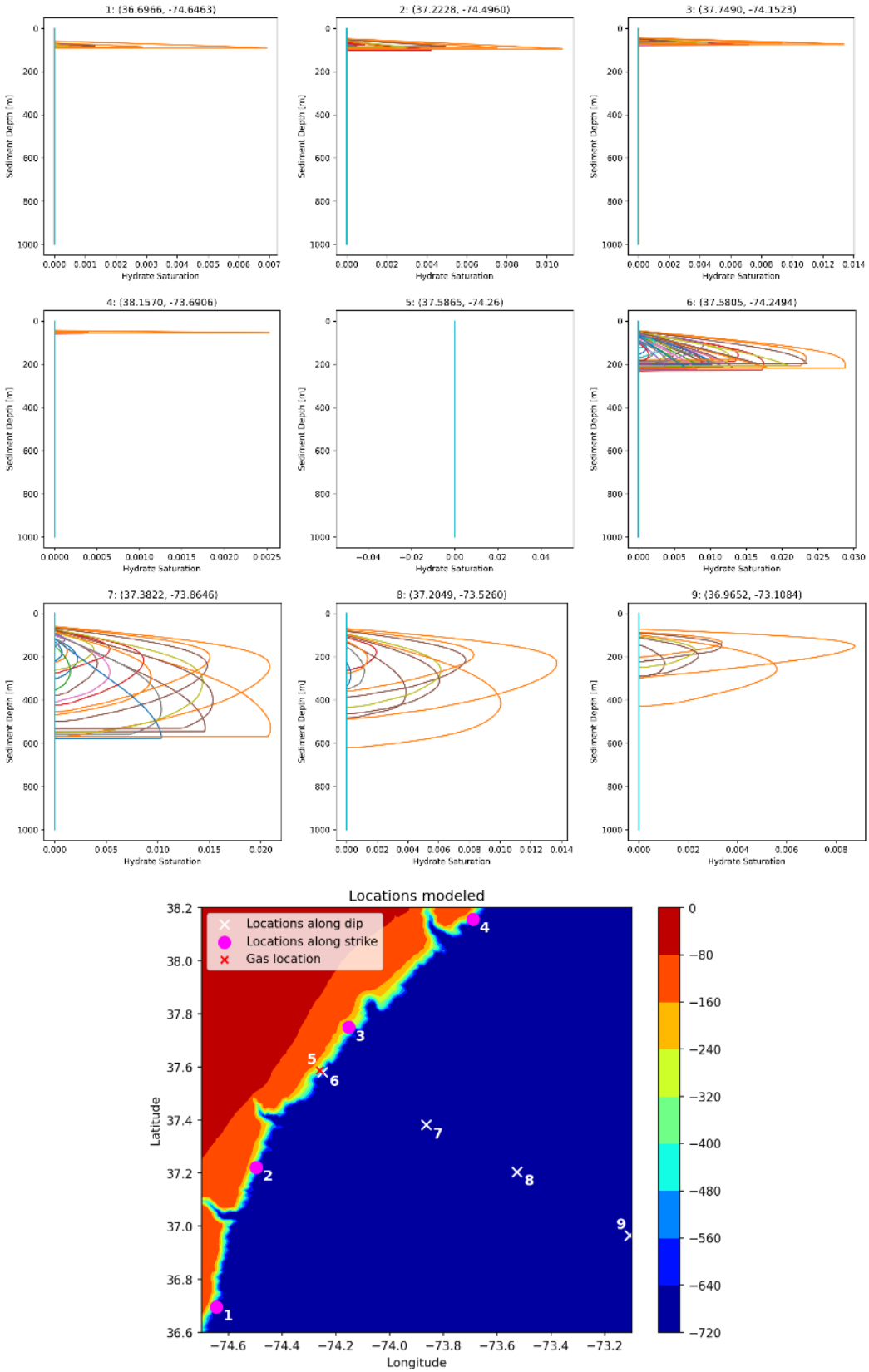


Figure 7-8. Top: predicted hydrate saturations at each location. Bottom: location map for reference.

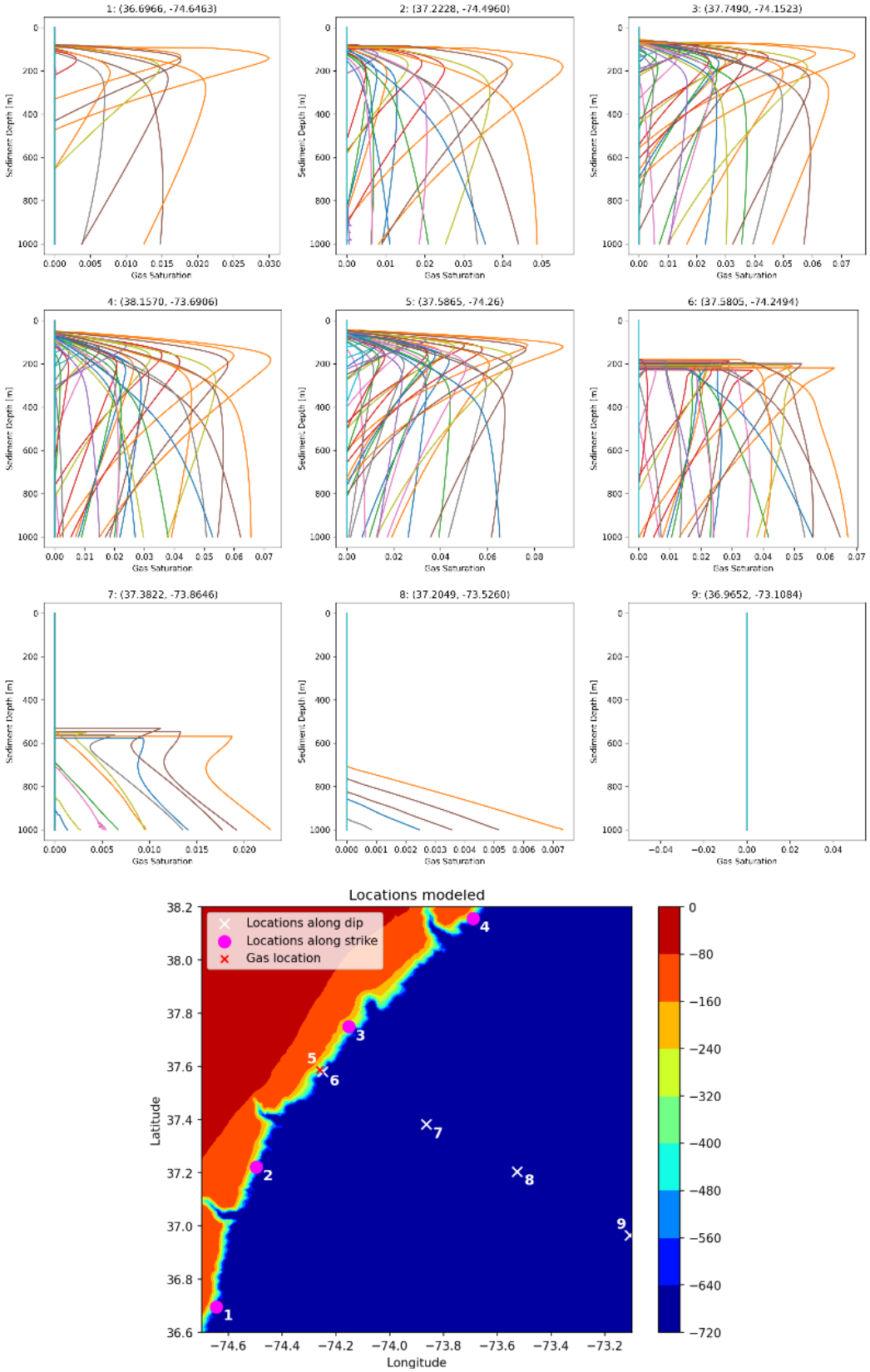


Figure 7-9. Top: predicted gas saturations at each location. Bottom: location map for reference.

The BHZ is very evident in Figure 7-8in sites 1-4 and site 6. There is a spread of BHZ depths at each location due to the different temperature profiles of each simulation. The depth at which the temperature profile and hydrate phase envelope intersect, causing hydrate to no longer be stable, varies for each run (Figure 7-9). Sites 1-4 and site 6 lie along the strike of the continental shelf and near the feather edge of the hydrate stability zone. The BHZ can be seen graphically for these sites where the hydrate saturation drops from its maximum value to zero.

At sites 7-9, some hydrate saturation profiles show this sudden drop in saturation, but many of the hydrate profiles decrease gradually making it difficult to predict the depth of the BHZ on hydrate saturation alone. This is discussed by Xu and Ruppel (1999) who note the difference between the hydrate stability zone and the actual zone of hydrate occurrence which relies on "the availability of methane in excess of its solubility in seawater." In profile 5, no hydrate forms This location is upslope of the feather edge and outside of the hydrate stability zone.

Table 7-3. Summary of hydrate profiles modeled with PFLOTRAN.

Site #	Total runs	Success runs	Hydrate runs	BHZ avg. (mbsf)	BHZ min. (mbsf)	BHZ max. (mbsf)	Hydrate sat. at BHZ	Temp at BHZ (°C)
Along strike								
1	50	49	5	86.1	80.5	91.5	0.0024	15.20
2	50	46	12	87.8	77.5	100.5	0.0029	8.44
3	50	44	16	67.1	55.5	77.5	0.0026	13.85
4	50	48	3	53.5	49.5	57.5	0.0011	7.58
Along dip								
5	50	47	0	nan	nan	nan	nan	nan
6	50	47	34	196.2	123.5	228.5	0.0080	17.72
7	50	50	20	414.9	145.5	576.5	0.0036	17.40
8	50	50	11	404.3	261.5	619.5	0.0001	22.96
9	50	50	7	282.8	200.5	427.5	0.0001	12.30

Table 7-3 summarizes the hydrate profiles shown in Figure 7-8. The number of successful runs at each location is noted, as well as the number of runs where hydrate formed. Only simulations in which hydrate formed were used to calculate the depth of the BHZ and the hydrate saturation at the BHZ. In addition, the average temperature at the BHZ was calculated by averaging the temperature profiles at the depth of the calculated BHZ. As mentioned, in sites 7-9, some hydrate saturation profiles did not show the sudden drop indicating the BHZ. Thus, the average BHZ depths for these

three sites are likely inaccurate and much shallower than the actual depth of the BHZ. A better estimation of the BHZ depth for sites 7 and 8 is the maximum predicted depth of the BHZ. As expected the BHZ calculated at sites 7 and 8 is deeper than the BHZ for sites found along the shelf strike.

The gas saturations are summarized in Table 7-4. Gas generation occurs within the modeled 1000 m sediment column in 8 of the simulations. The only simulation in which no gas was generated during the 50 runs was at the deepest location, site 9. When calculating the maximum gas saturation and the depth of the maximum gas saturation in Table 7-4, only maximum saturations shallower than 1000 m were considered. This is why site 7 does not have a maximum gas saturation value or depth. The maximum gas saturation for site 8 occurs in only one run, and can likely be ignored as well. Comparing the remaining six locations, site 5 has the largest average maximum gas saturation. As discussed, this site is up-slope of the feather edge and outside of the hydrate stability zone.

Table 7-4. Summary of gas profiles modeled with PFLOTRAN.

Site #	Total runs	Success runs	Gas runs	Max gas saturation	Depth of max gas saturation (mbsf)
Along strike					
1	50	49	11	0.0136	249.95
2	50	46	23	0.0188	339.68
3	50	44	33	0.0300	248.95
4	50	48	37	0.0262	292.56
Along dip					
5	50	47	37	0.0383	205.39
6	50	47	32	0.0266	407.98
7	50	50	11	nan	nan
8	50	50	5	0.0073	998.50
9	50	50	0	nan	nan

The predictive map of seafloor total organic carbon (TOC) from GPSM was comparable to the work done by Lee et al. (2019) and Eymold et al. (2021). The global map of seafloor TOC produced by Lee et al. (2019) also predicted an area of increased TOC off the east coast of the United States in the same region ((35.4°N, 75.0°W) to (39.0°N, 72.0°W)) predicted in this model. The map of seafloor TOC predicted by Eymold et al. (2021) covers about a quarter of the area of my prediction. In this prediction, trends of TOC are similar to the trends predicted here as Eymold et al. (2021)

also predicts an increase in TOC at the southern end of the (35.4°N, 75.0°W) to (39.0°N, 72.0°W) region. It is unsurprising that the areas with increased TOC predictions are consistent between my work and the work done by Lee et al. (2019) and Eymold et al. (2021). Although the data sets used between the three predictions differed slightly, the general methodology was consistent across the models.

To ensure the accuracy of the seafloor TOC model, it is additionally important to compare the results to field data in the area. Investigations by Skarke et al. (2014) have found multiple instances of methane gas leakage from the seafloor along the U.S. Atlantic margin in the form of gas plumes (Figure 6.2.6). The sites of these methane seeps correspond with areas predicted by GPSM to have higher values of seafloor TOC, specifically in the region between (35.4°N, 75.0°W) and (39.0°N, 72.0°W). This region of highly concentrated methane seeps identified by Skarke et al. (2014) also aligns with the locations of the sites (specifically along the shelf break) where hydrate and gas formation were modeled with PFLOTRAN/Dakota.

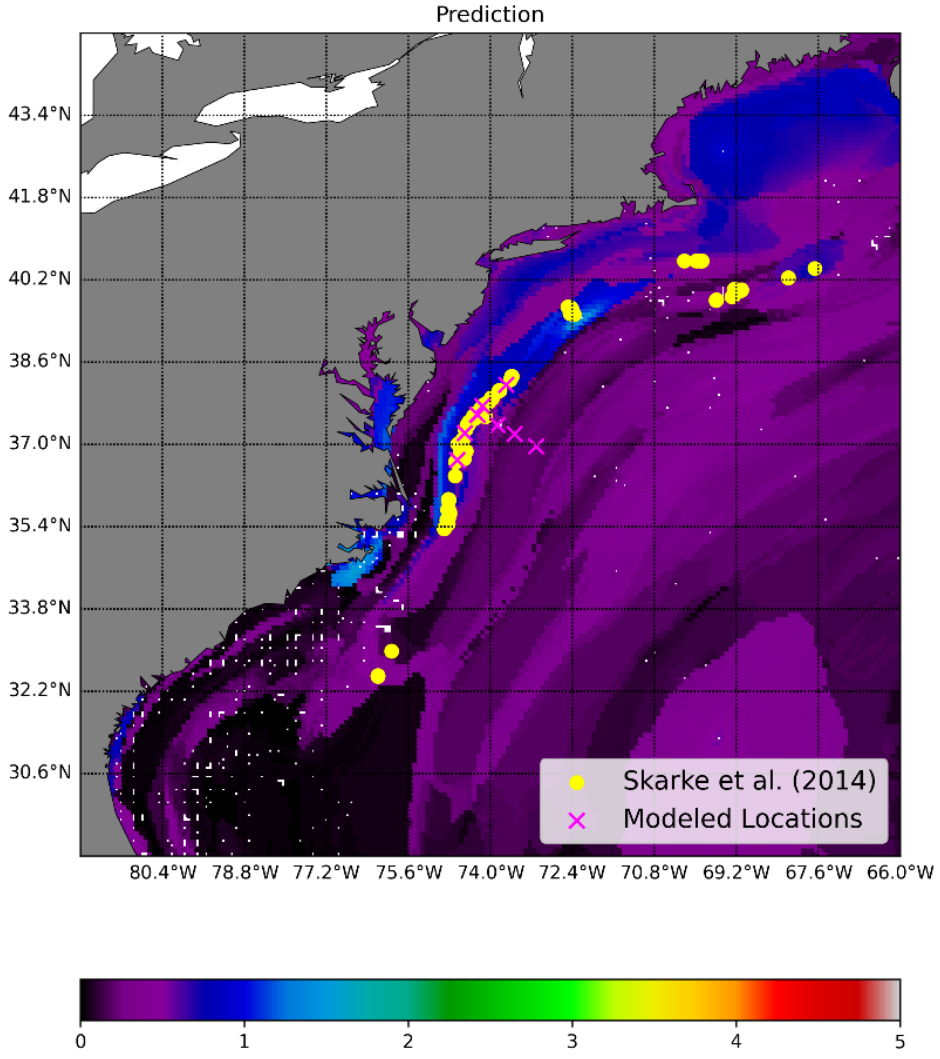


Figure 7-10. Output maps from GPSM over the area 29°N–45°N and 82°W–66°W. Locations with known TOC values are marked by small white dots. Methane seeps (large yellow dots) identified by Skarke et al. (2014) are also plotted and support the increased TOC prediction between (35.4°N, 75.0°W) and (39.0°N, 72.0°W). Modeled locations are marked with a magenta x.

7.3. Gulf of Mexico – Shallow Gas

We focused on predicting the probability of encountering shallow gas and hydrates in the Gulf of Mexico using the GPSM-Dakota-PFLOTRAN workflow. In shallow marine sediments, gas accumulations and hydrates are significant geohazards to subsea infrastructure, drilling, and production. Therefore, these predictions are crucial to ensure offshore drilling safety and submarine infrastructure security. The objective is to generate a prediction with uncertainty estimates to allow geohazard assessment before any shallow hazard surveys are performed.

We used k-nearest neighbor regression for this task following Lee et al. (2019) and Eymold et al. (2021). The optimal value of k was selected at k = 5 based on its highest R^2 value and similar runtime to other values (Table 7-5). TOC predictions and standard deviations in the Gulf of Mexico are shown in Figure 7-11.

Table 7-5. Selection of k based on the coefficient of determination (R^2) and runtime.

K value	R^2	Runtime
		[min]
4	0.6131	113.34
5	0.6165	93.76
6	0.6134	87.16
7	0.6144	108.25
8	0.6121	93.03
10	0.6074	101.47
12	0.6046	90.35

Dakota sampled TOC, heat flux, and sedimentation rate from a normal distribution. The scaling factor was sampled with a uniform distribution, and methanogenesis rate was sampled using a log-uniform distribution (Eymold et al. 2021). Furthermore, boundaries were set for each variable to avoid non-physical values such as negative sedimentation rates. This method provides N values of each variable independent of the other variables to provide a set of distinct values for each simulation (Eymold et al. 2021). These inputs were then used in PFLOTRAN. Input parameters are shown in Table 7-6.

To narrow the focus of the simulation runs and accurate estimate parameters, we chose geographic locations with economic and geologic significance. Cross-referencing oil production data from the Enverus database, we selected locations in offshore lease blocks with the highest oil production (Enverus 2021). These blocks included Alaminos Canyon, Green Canyon, and Garden Banks, which can be seen below on Figure 7-12. These blocks were also chosen because they had both significant production history and direct measurements of heat flux and sedimentation rates from other sources.

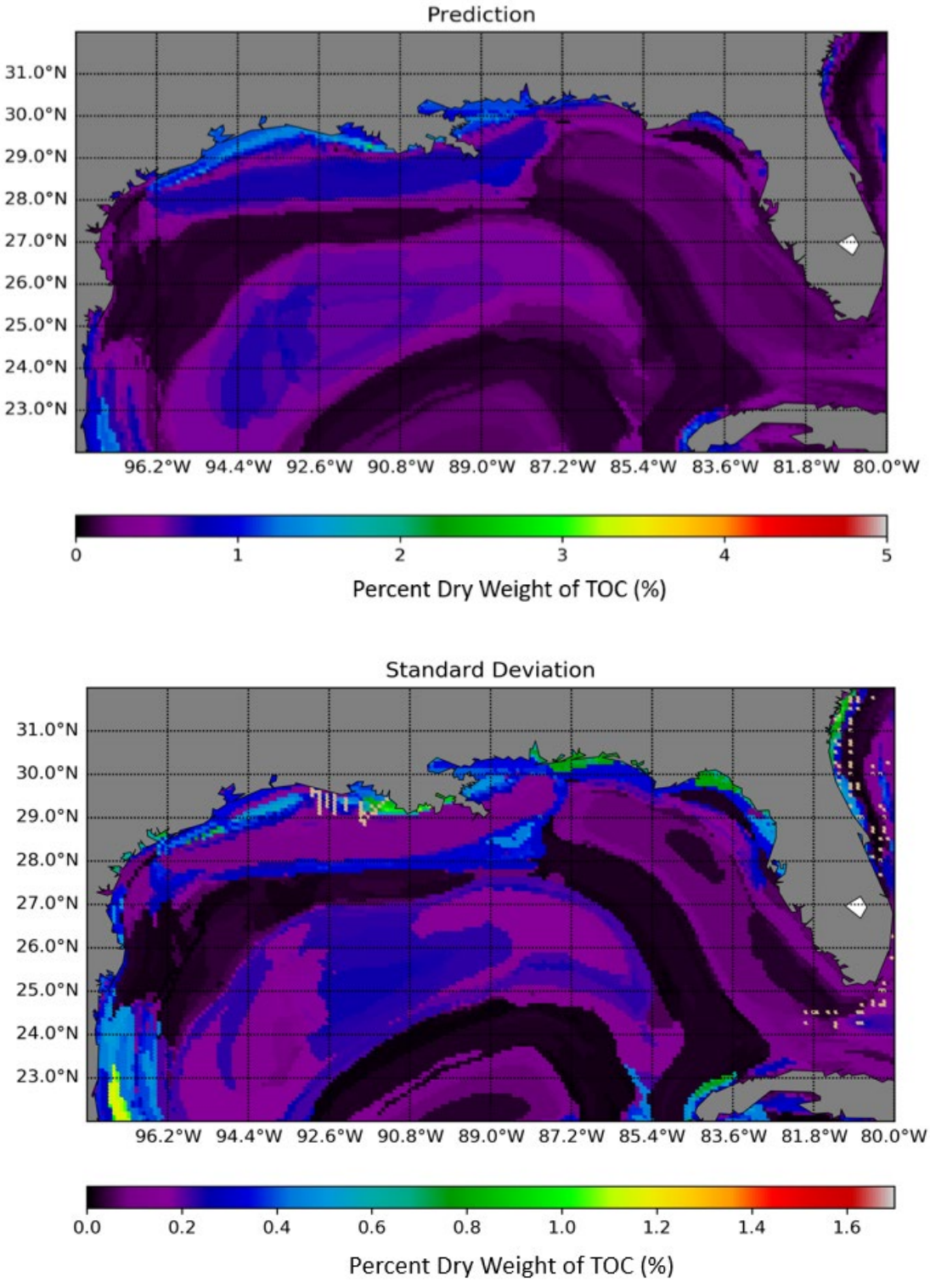


Figure 7-11. Top: TOC prediction. Bottom: standard deviation of prediction.

Table 7-6. Input values for PFLOTRAN simulations.

Parameter	Status	Value	Units
Sedimentation Rate	Predictions from Restreppo et al. (2020)	Variable	m/s
TOC	Predictions from GPSM	Variable	%
Heat Flux	IHFC Global Heat Flow Database	Variable	W/m^2
Porosity Scalar	Kominz et al. (2011) Model: $0.663*e^{(-depth/1333)}$	Variable	-
Methanogenesis Rate	Log uniform sampled between 1×10^{-15} to 1×10^{-13}	Variable	s^{-1}
Geothermal Gradient	Variable, based on porosity/heat flux	Variable	°C/m
Pressure Change Limit	Max change per time step	1×10^5	Pa
Temperature Change Limit	Max change per time step	1 x 105	°C
Gravity	Fixed	-1.125	m/s^2
Labile Portion of TOC	Set to 75% TOC	Variable	%
Conversion Factor of Methane	Fixed	2,241	kg/m^3
Diffusion Coefficient of Methane	Fixed	1×10^{-9}	m^2/s
Gas Viscosity	Methane	1.1×10^{-5}	Pa-s
Tortuosity	Fixed	1.4	-
Rock Density	Fixed	2,700	kg/m^3
Thermal Conductivity (dry)	Fixed	1	W/m/°C
Thermal Conductivity (wet)	Fixed	1	W/m/°C
Heat Capacity	Fixed	830	J/kg/°C
Permeability	Fixed	1×10^{-15}	m^2
SMT Depth	Fixed	15	mbsf

Parameter	Status	Value	Units
Van Genuchten Pressure	Fixed	5.8×10^{-4}	m^{-1}
Van Genuchten Pore Size Factor	Fixed	0.189	-
Liquid Residual Saturation	Fixed	0.1	-
Max Capillary Pressure	Fixed	1×10^{8}	Pa
Gas Residual Saturation	Fixed	0.15	-
Initial Liquid Pressure	Initial Condition	Variable	Pa
Initial Mole Fraction	Initial Condition	1×10^{-3}	-
Initial Temperature	Initial Condition	Variable	°C
Initial Hydrostatic Liquid Pressure	Initial Condition	Variable	Pa
Initial Dirichlet Hydrate Saturation	Initial Condition	1×10^{-8}	-
Initial Dirichlet Temperature	Initial Condition	Variable	°C
Neumann Liquid Flux	Bottom Boundary Condition	0	m/yr
Neumann Gas Flux	Bottom Boundary Condition	0	m/yr
Neumann Energy Flux	Bottom Boundary Condition	Q	W/m^2
Hydrostatic Liquid Pressure	Top Boundary Condition	Variable	Pa
Dirichlet Mole Fraction	Top Boundary Condition	1×10^{-3}	-
Dirichlet Temperature	Top Boundary Condition	Variable	°C

Furthermore, to provide contrast to these deep-water locations, several shallow locations near the continental shelf were selected in blocks, Galveston Area, Ship Shoal, High Island Area, Viosca Knoll, and East Cameron Area. These locations were also supported by direct measurements of heat flux and sedimentation rates from other sources. In addition, locations were selected from the International Ocean Discovery Program (IODP) database because these were well supported by accurate and direct measurements of necessary parameters (Expedition 308 Scientists, 2006).

To accurately determine water depth at these locations, we used the Global Multi-Resolution Topography synthesis (GMRT) (Ryan et al. 2009). This global database was of the highest resolution

and has the most up to date bathymetry for the entire globe. Therefore, it gave us an accurate value of water depth at specific latitude and longitude locations. Similarly, to determine seafloor temperature, we used the GIS database from the Marine Conservation Institute. This database provided us with an accurate value of seafloor temperature at specific latitude and longitude locations, with a cell size of 30 arc seconds (Boyer et al, 2005). To determine sedimentation rate and its standard deviation, we used predictions from Restreppo et al. (2020), who applied a machine learning algorithm to predict oceanic sediment accumulation rates for all coasts and oceans. The results of this prediction model are summarized in Figure 7-13.

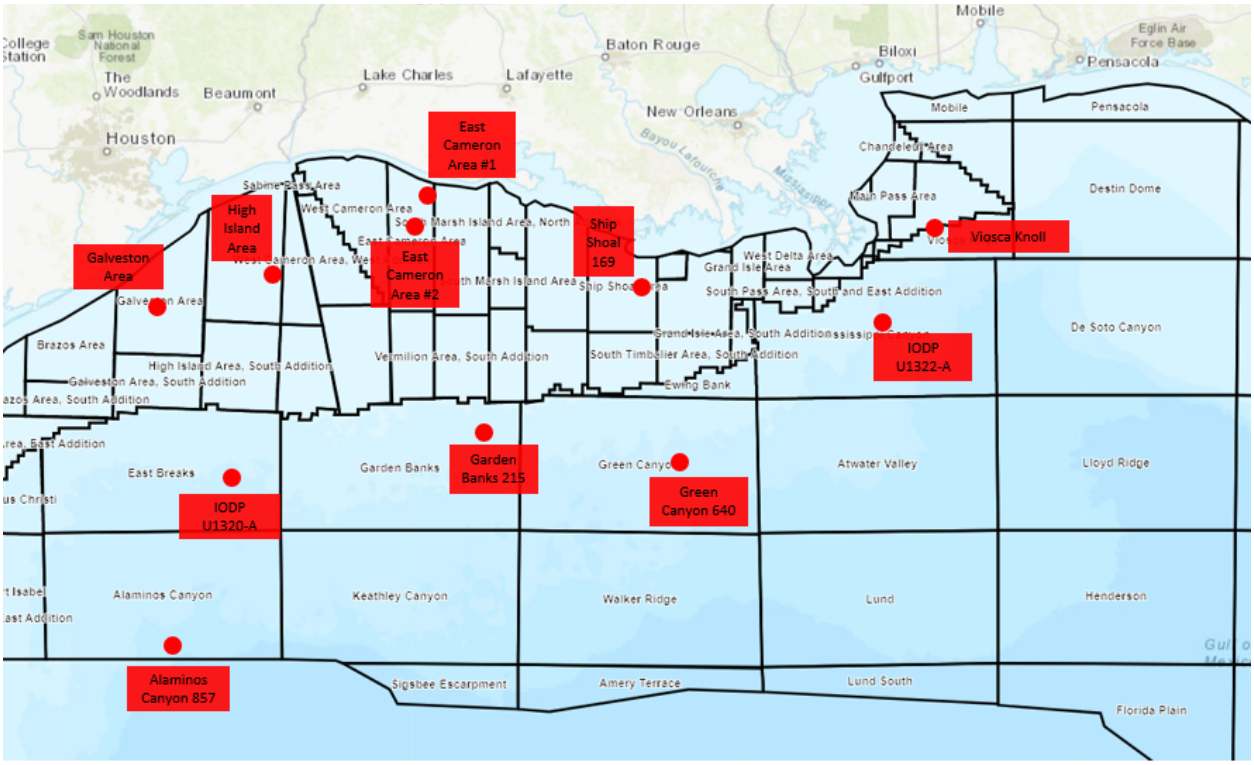


Figure 7-12. Simulation locations with lease blocks.

To determine heat flux and its standard deviation, we referenced the International Heat Flow Commission's Global Heat Flow Database (Global Heat Flow Compilation Group, 2013). This database has approximately 279 direct measurements of heat flow in the Gulf of Mexico region. This provides us with an accurate observation of heat flux at specific latitude and longitude, which is significant because heat flux varies greatly as you move from the continental shelf to the slope in the Gulf of Mexico. Furthermore, to supplement this data in areas that were not densely sampled, we referenced a US Department of the Interior Minerals Management Service report (U.S. Resource Evaluation Division, 2008). This report included raw data on the geothermal gradient in the Gulf of Mexico, as seen below in Figure 7-14, originally collected in units of °C/km. Based on this histogram, we applied the mean geothermal gradient of 32.44 mW/m^2 and the standard deviation of 17.35 mW/m^2 to locations where no direct measurements were present in the global heat flow database.

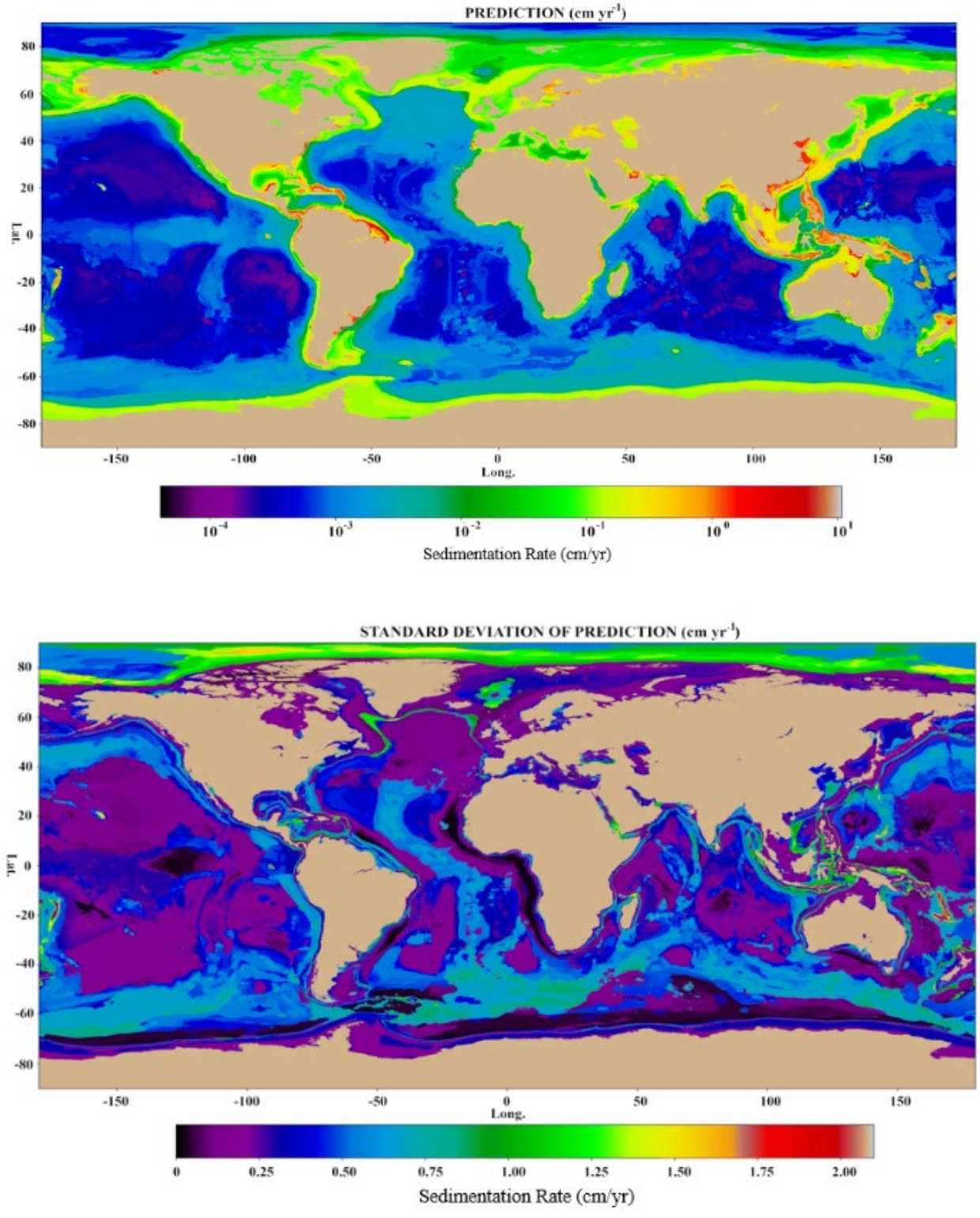


Figure 7-13. Top: sedimentation rate prediction. Bottom: standard deviation. Values from Restreppo et al. (2020).

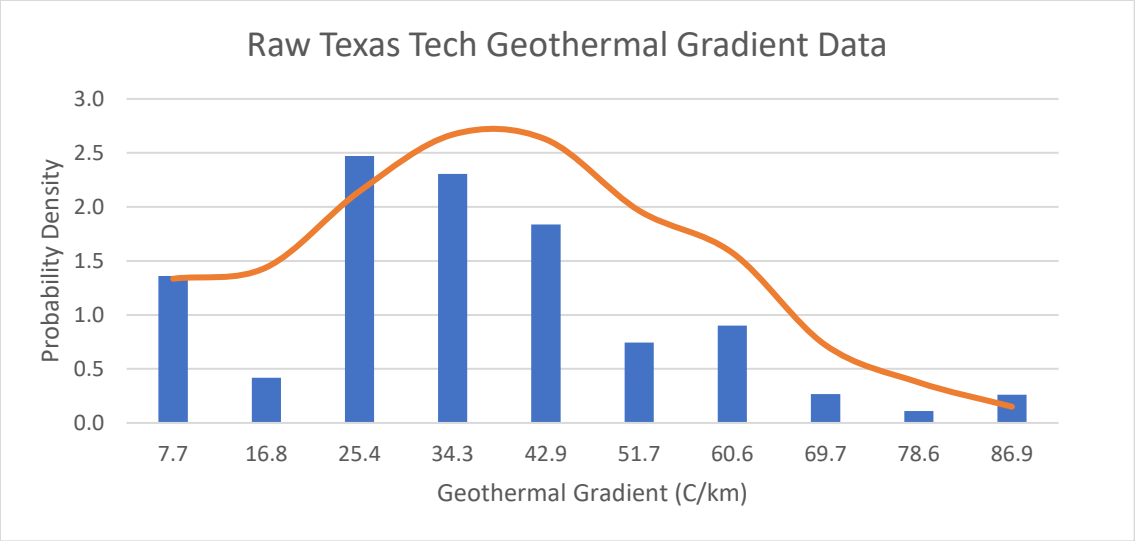


Figure 7-14. Raw Texas Tech geothermal gradient distribution (U.S. Resource Evaluation Division, 2008).

Based on the availability and accuracy of parameters, the selected locations for simulation are summarized in Table 7-7 and Table 7-8. These tables include exact parameter values taken from the sources previously listed.

Table 7-7. Prediction locations.

Block		Alaminos Canyon 857	IODP 2004-2013 308-U1320-A	Green Canyon 640	Garden Banks 215	IODP 2004-2013 308-U1322-A
Profile		9299	12327	12587	13216	14766
Latitude	[°]	26.0939572	27.2664	27.3168347	27.5994467	28.1542281
Longitude	[°]	-94.904053	-94.4032	-90.753767	-92.298389	-89.103724
Water Depth	[m]	-2460	-1430	-1307	-770	-1157
Seafloor Temperature	[°C]	4	5	4	7	4.5
Sedimentation Rate	[cm/yr]	0.10308577	0.25487322	0.21047477	0.25127935	0.23858934
Sedimentation Rate Standard Deviation	[cm/yr]	0.35268494	0.33345547	0.37109157	0.33762705	0.35216802
Heat Flux	[mW/m^2]	47	32.43511	32.43511	34	32.43511
Heat Flux Standard Deviation	[mW/m^2]	2	17.35227	17.35227	2	17.35227

Table 7-8. Prediction locations, continued.

Block		Galveston Area	Ship Shoal 169	High Island Area	Viosca Knoll	East Cameron Area (#2)	East Cameron Area (#1)
Profile		15773	15821	16216	16928	17091	17736
Latitude	[°]	28.593	28.6187564	28.7138	29.0049	29.0932	29.3488
Longitude	[°]	-94.9804	-90.995106	-94.0069	-88.5556	-92.8321	-92.6947
Water Depth	[m]	-31	-19	-27	-468	-22	-16
Seafloor Temperature	[°C]	23	24	23	10	23	23
Sedimentation Rate	[cm/yr]	0.48083304	0.23699815	0.31938822	1.93456078	0.301677951	0.458751461
Sedimentation Rate Standard Deviation	[cm/yr]	0.4152395	0.20393202	0.43229884	0.36461788	0.41631946	0.40278223
Heat Flux	[mW/m^2]	52	41	46	44	52	48
Heat Flux Standard Deviation	[mW/m^2]	2	2	2	2	2	2

By referencing multiple sources for accurate estimations of input parameters and assuming microbial methanogenesis is the source of shallow hydrocarbons, we can propagate the seafloor properties downward using Dakota and PFLOTRAN. This generates probabilistic models of gas and hydrate presence. In these simulations, we assumed that the parameters remained constant for the duration of the simulation. Although this is not geologically reasonable, it allows us to identify where gas saturation should occur.

The results from the Dakota and PFLOTRAN runs are represented graphically in Figure 7-15–Figure 7-19. On the left side of each figure are histograms of results data. The first histogram represents the number of simulation runs that resulted in gas being formed, as well as how much gas was formed. The x-axis illustrates the volume of gas, in cubic meters, integrated over the simulation domain, which is a rectangular prism with the base of a 5 x 5 arc minute box on the seafloor that extends vertically downwards 1,000 m. The second histogram represents the reaction rate constant for the methanogenesis function, which was given as log normal distribution in Dakota. The third histogram represents the distribution of TOC values, which should follow a normal distribution and depend on the mean and standard deviation predicted by GPSM. Although these distributions will follow a normal distribution, this trend may not appear in some histograms because the sample size is relatively low when using only ten simulations runs on PFLOTRAN. Furthermore, the plots on the right side are known as horsetail plots. These plots represent the distinct gas or hydrate saturation profile you expect to see based on the variation of parameters provided by Dakota. Each plot line represents one set of parameters input in one simulation run. Overall, these histograms and horsetail plots are illustrating the probability of seeing gas and hydrates within the parameter space.

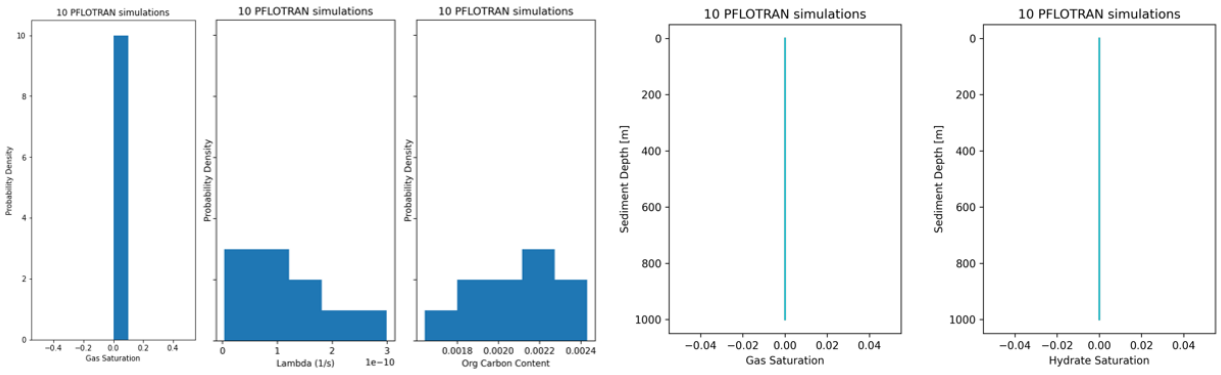


Figure 7-15. Results for Alaminos Canyon.

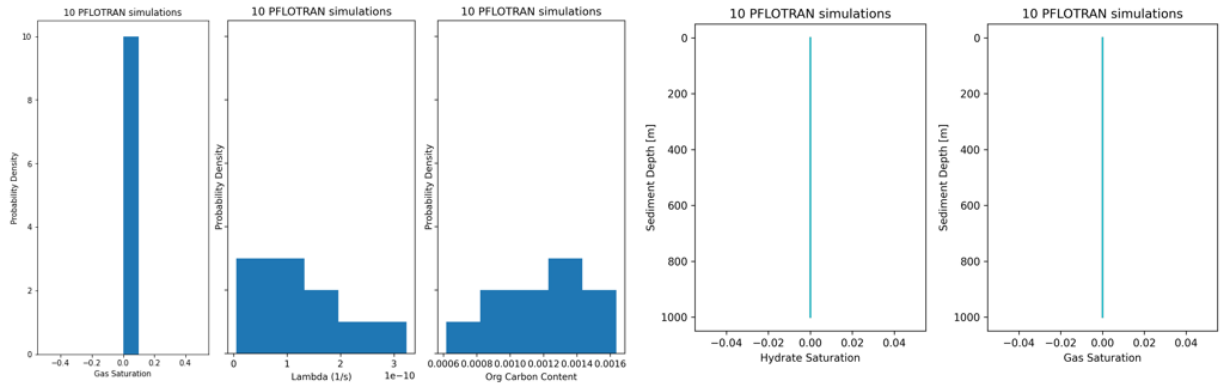


Figure 7-16. Results for IODP 308-U1320-A.

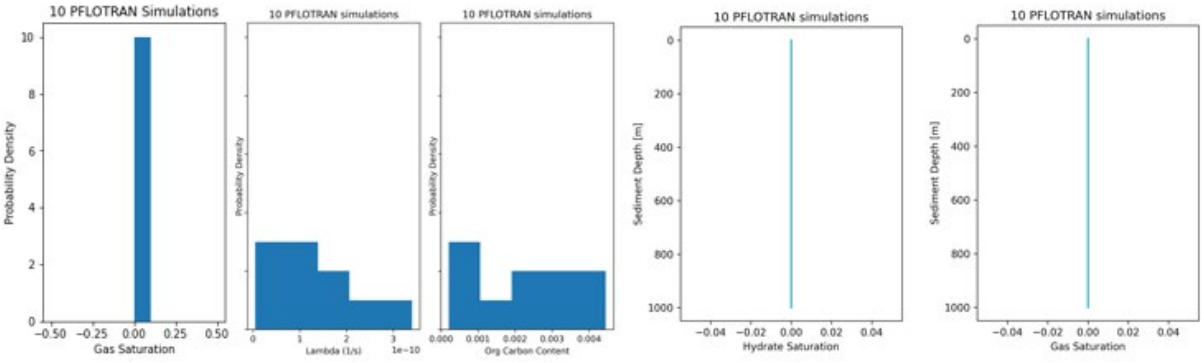


Figure 7-17. Results for Green Canyon.

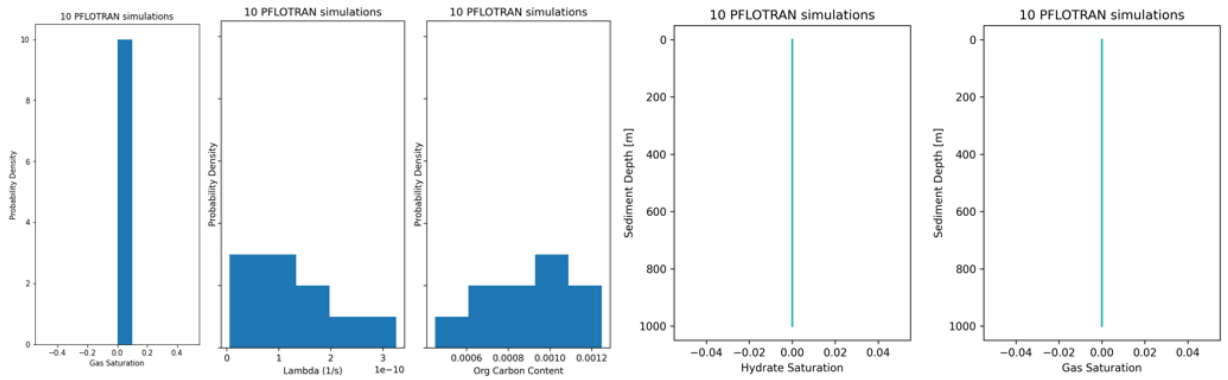


Figure 7-18. Results for Garden Banks.

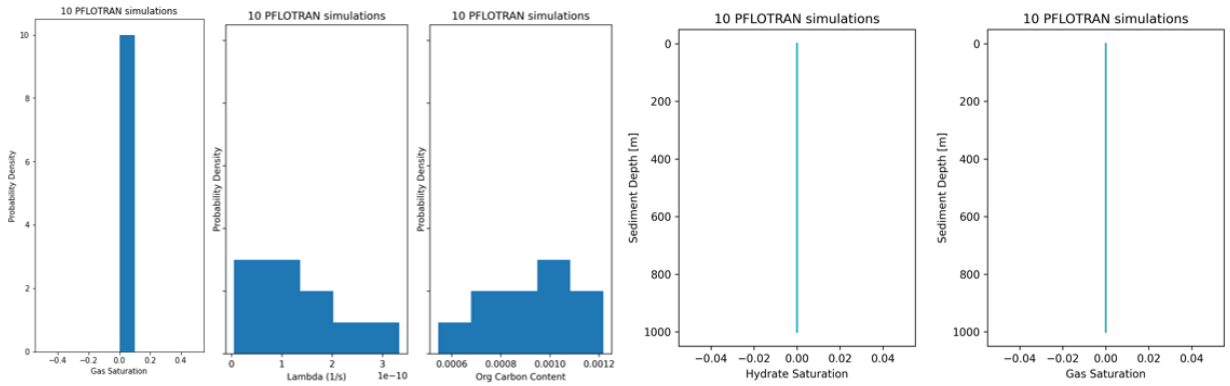


Figure 7-19. Results for IODP 308-U1322-A.

In the deep-water region of the Gulf of Mexico, the simulations run at locations within Alaminos Canyon, Green Canyon, Garden Banks, IODP 308-U1320-A, and IODP 308-U1322-A represent probabilistic models for gas accumulation and hydrate presence in these deep-water marine environments. The gas saturation histogram shows that it is highly probably that no gas is formed because the runs' volumes of gas present were frequently zero. In addition, the distributions of the TOC values also were normally distributed around low values, less than 0.25% dry weight. In addition, the horsetail plots show that all the saturation profiles for both gas and hydrates remain at approximately zero throughout the 1,000-meter zone.

In the continental shelf region of the Gulf of Mexico, the simulations run at locations within the Galveston Area, Ship Shoal, High Island Area, Viosca Knoll, and East Cameron represent probabilistic models for gas accumulation and hydrate presence in these shallow marine environments. The results for each of these locations are illustrated in Figure 7-20 - Figure 7-25.

In Figure 7-20, the gas saturation histogram shows that it is approximately equally likely for no gas to be present, or between 20-25 m^3 to be present. In addition, the TOC distribution shows that this Galveston Area location was more likely to have higher values of TOC because the distribution ranged from 0.4-0.8% dry weight. The horsetail plots show that this location will likely not have hydrate present, but is likely to have gas saturation present. Most the profiles slowly increase with depth, showing multiple profiles ending around 0.025 at 1,000 m.

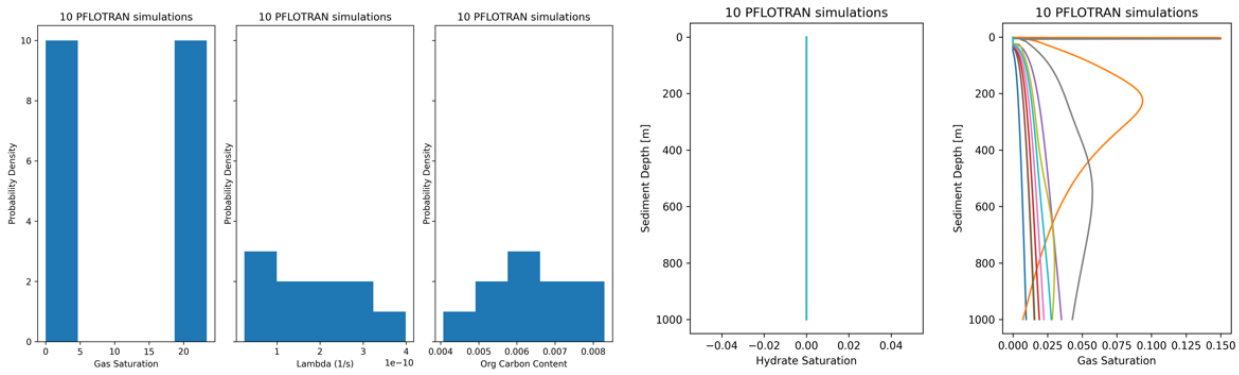


Figure 7-20. Results for Galveston Area.

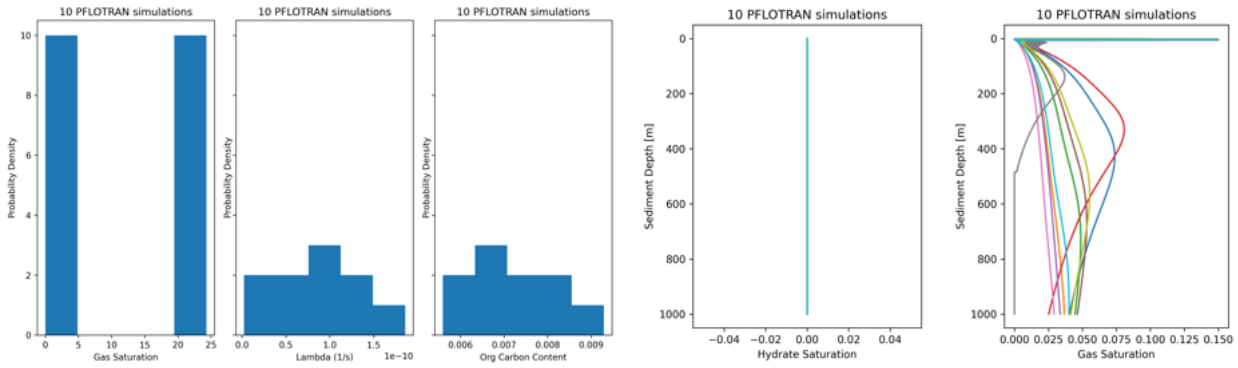


Figure 7-21. Results for Ship Shoal.

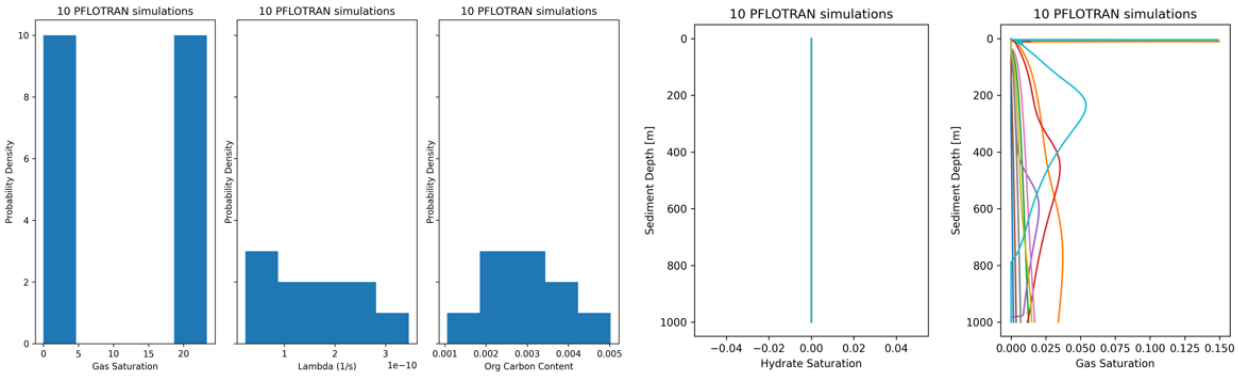


Figure 7-22. Results for High Island Area.

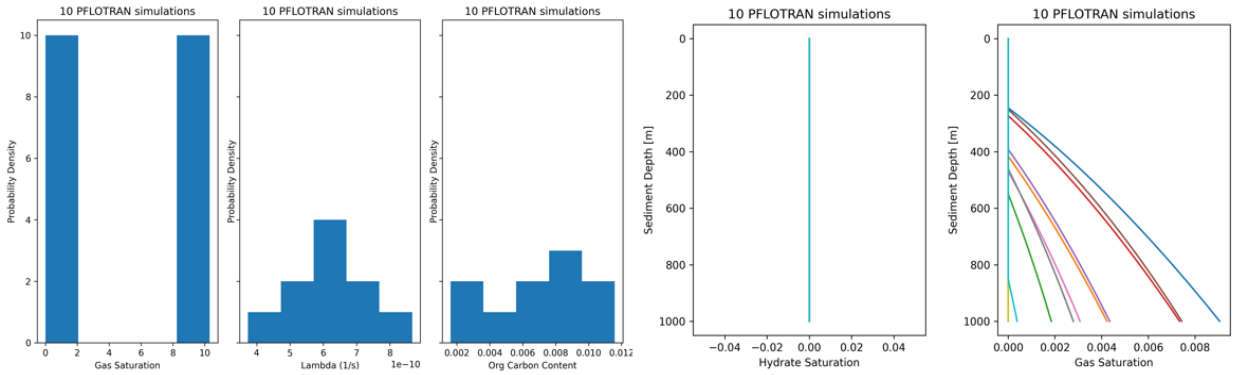


Figure 7-23. Results for Viosca Knoll.

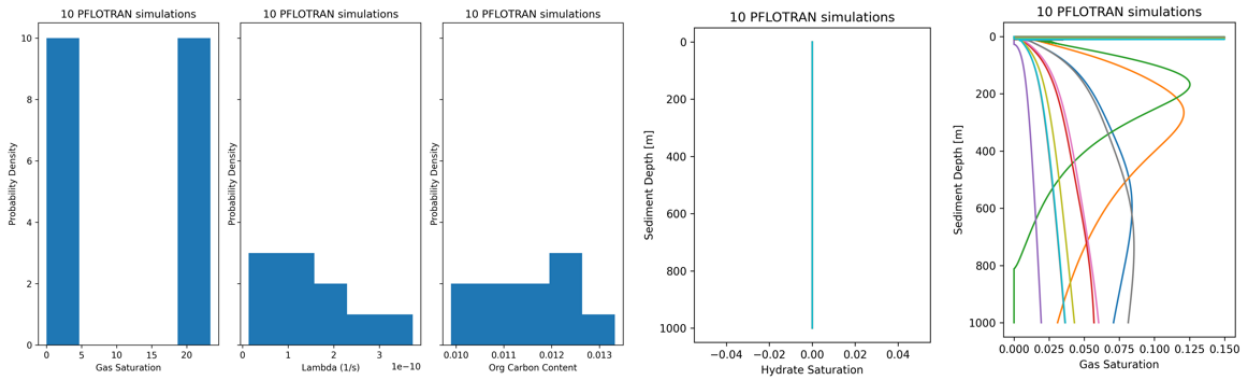


Figure 7-24. Results for East Cameron (profile 17091).

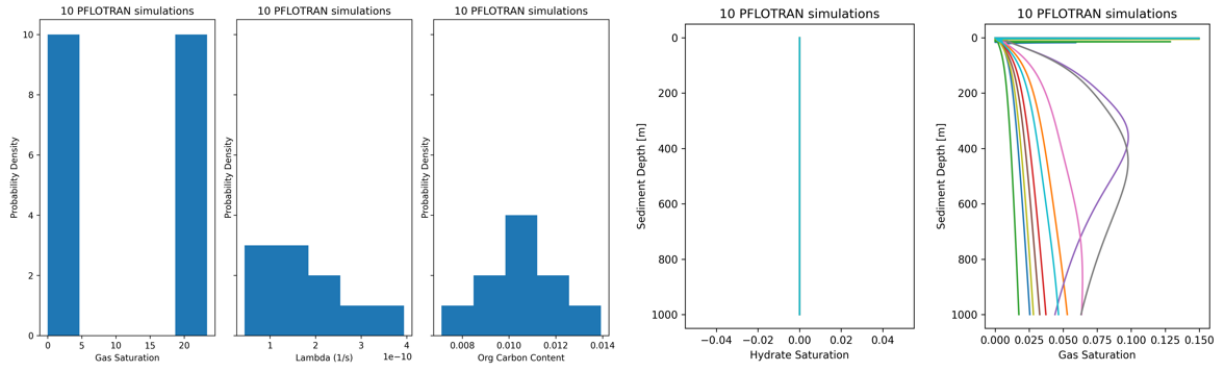


Figure 7-25. Results for East Cameron (profile 17736).

Figure 7-21 shows the results for the Ship Shoal location. Similar to Figure 7-21, the gas saturation histogram shows that it is approximately equally likely for no gas saturation to be present, or between 20-25 m^3 to be present. In addition, the TOC distribution has high values ranging from 0.6-0.9% dry weight. The horsetail plots show that this location will likely not have hydrate present, but is likely to have gas saturation present. Most the profiles have the highest saturation of gas present between 200-600 m from the seafloor.

In Figure 7-22, the gas saturation histogram for High Island area shows that it is approximately equally likely for no gas saturation to be present, or between 20-25 m^3 to be present. However, the TOC values are not as high as previous locations, ranging from 0.1-0.5% dry weight. The upper bounds of this distribution is still higher than the deep water locations which were unlikely to have gas present. The horsetail plots show that this location will likely not have hydrate present, but is likely to have gas saturation present. However, the gas saturation is most likely to be relatively low around 0.015, with the exception of a few profiles that show gas present between 200-400 m.

In Figure 7-23, the gas saturation histogram shows that it is approximately equally likely for no gas saturation to be present, or between 8-10 m^3 to be present. In addition, the TOC distribution has values ranging from 0.2-1.2% dry weight. This distribution is much wider than previous locations, and is likely to have high TOC values. In addition, the highest frequency of the distribution falls on the higher end of the scale of TOC values. The horsetail plots show that this location will likely not have hydrate present, but is likely to have gas saturation present. These profiles increase with depth. However, the maximum gas saturation for each profile ranges between 0-0.008, which is lower than other locations selected in the Gulf of Mexico.

In Figure 7-24, the gas saturation histogram shows that it is approximately equally likely for no gas saturation to be present, or between 20-25 m^3 to be present. In addition, the TOC distribution has very high values ranging from 1.0-1.3% dry weight. The horsetail plots show that this location will likely not have hydrate present, but is likely to have gas saturation present. These profiles show gas is likely to be present between 200-400 m below the seafloor in some simulation runs. In addition, these saturation profiles also range up to a maximum of 0.125 gas saturation.

In Figure 7-25, the gas saturation histogram shows that it is approximately equally likely for no gas saturation to be present, or between 20-25 m^3 to be present. In addition, the TOC distribution has relatively high values ranging from 0.8-1.4% dry weight. The horsetail plots show that this location will likely not have hydrate present, but is likely to have gas saturation present. These profiles show gas saturation increasing with depth, and in some runs, experiencing a spike in gas saturation 200-400 m below the seafloor in some simulation runs. In addition, these saturation profiles also range up to a maximum of 0.100 gas saturation.

After graphing the results from these simulation runs, we observe that there is no hydrate present at any of these locations. After evaluating the exact values that make up the above distributions of gas saturation and hydrate saturation, the minimum and maximum depths that gas occurs is represented in Table 7-9 below.

Table 7-9. Average maximum and minimum depth of gas and hydrate present and range of TOC values.

Block	Profile	Average Minimum Depth of Gas Present (m)	Average Maximum Depth of Gas Present (m)	Maximum Gas Saturation	Average Gas Saturation Over Zone of Occurrence	Range of TOC (Fraction Dry Weight)
Alaminos Canyon 857	9299	-	-	-	-	0.0018 - 0.0024

Block	Profile	Average Minimum Depth of Gas Present (m)	Average Maximum Depth of Gas Present (m)	Maximum Gas Saturation	Average Gas Saturation Over Zone of Occurrence	Range of TOC (Fraction Dry Weight)
IODP 2004-2013 308-U1320-A	12327	-	-	-	-	0.0005 - 0.0020
Green Canyon 640	12587	-	-	-	-	0.000 - 0.004
Garden Banks 215	13216	-	-	-	-	0.0006 - 0.0012
IODP 2004-2013 308-U1322-A	14766	-	-	-	-	0.0006 - 0.0012
Galveston Area	15773	0.5	972.9	0.15	0.0213	0.004 - 0.008
Ship Shoal 169	15821	51.95	997.1	0.15	0.0355	0.006 - 0.009
High Island Area	16216	23.55	926.3	0.15	0.0122	0.001 - 0.005
Viosca Knoll	16928	100.35	509.95	0.00905	0.00283	0.002 - 0.0012
East Cameron Area (#2)	17091	19.25	993.3	0.15	0.0442	0.010 - 0.013
East Cameron Area (#1)	17736	0.5	994.9	0.15	0.0364	0.008 - 0.014

Considering the results above, we can draw conclusions about the relationship between geologic location and gas and hydrate presence, as well as the minimum amount of TOC needed to indicate gas or hydrate presence. Overall, locations on the continental shelf along the Gulf of Mexico coast are more likely to have presence of gas accumulation compared to those located in deep-water. However, neither the locations along the shelf nor in deep-water were likely to have hydrate. Because we selected lease blocks with already existing high oil production, it may be likely that these blocks do not have hydrate present because major oil and gas operators are more likely to drill in locations without obvious shallow hazards. Areas with high drilling activity in the Gulf of Mexico, such as Alaminos Canyon, Green Canyon, and Garden Banks, are less likely to have shallow gas hazards. Furthermore, any seafloor seeps identified in areas of high drilling activity likely originate from deep reservoirs, not from shallow gas accumulation because high drilling activity disturbs the deep reservoir temperature and pressure conditions, which leads to seeps.

Moreover, our predictions illustrate that gas is more likely to be present along the shelf where the total organic content values are high. Gas is less likely to be present in deep water where the total organic content values are low. From these results, it can be asserted that approximately 0.25% dry weight is a minimum amount of TOC that indicates the presence of a shallow gas hazard. Galveston, Ship Shoal, Viosca Knoll, and East Cameron range well above 0.25% dry weight, and correlate to a long average depth interval where gas saturation is likely to be present. Out of the locations along the shelf, the High Island area has a wider TOC range which drops below 0.2% dry weight. However, the range also well exceeds this minimum and spans to 0.5% dry weight. By comparison, all the locations along the slope and deep-water have TOC ranges around 0.1% dry weight, with Green Canyon as the expectation with a wider range that does go below 0.2% but also spans up to 0.4% dry weight. Furthermore, these regions continental shelf where sedimentation rate is greater than 0.1 cm/yr are more likely to have shallow gas hazards. This trend can be visualized in Figure 7-26 below. Overall, these results illustrate that gas is more likely to be present along the shelf where the total organic content values are high.

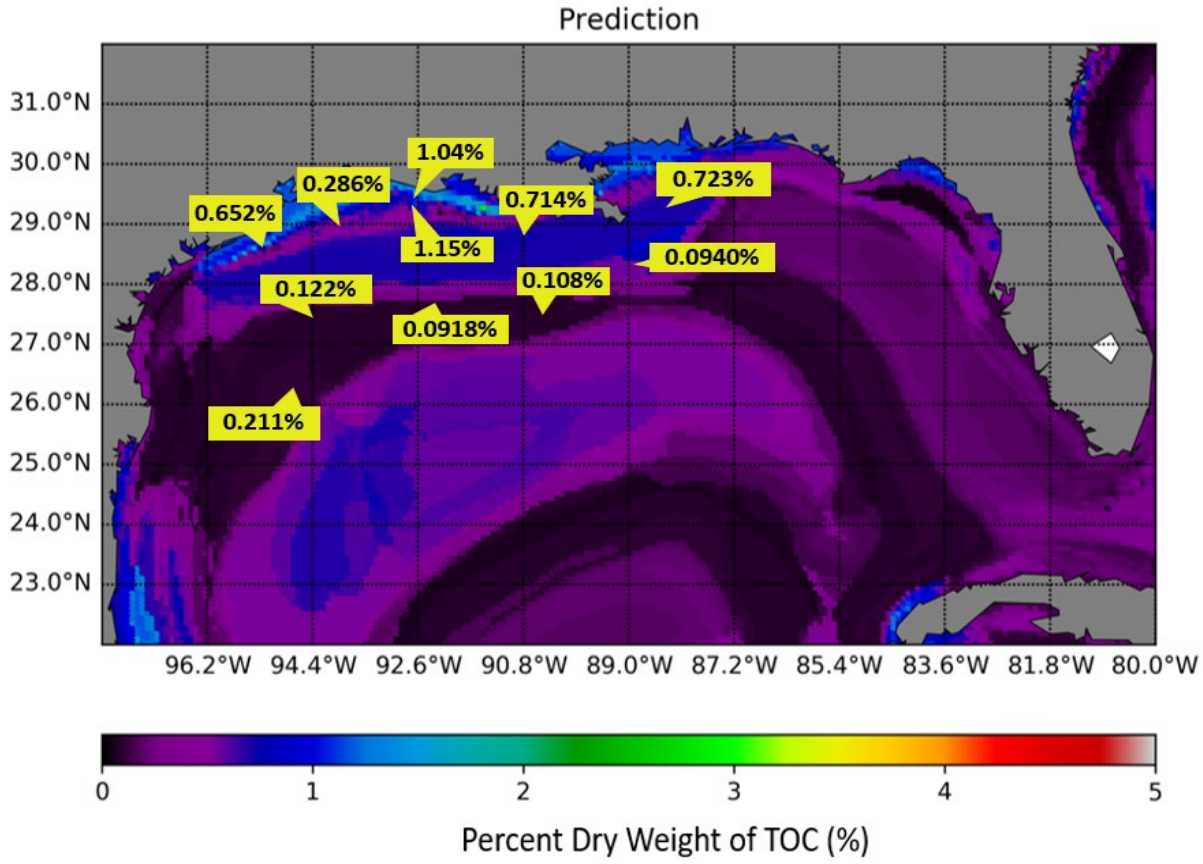


Figure 7-26. TOC values at selected simulation locations.

Based on the PFLOTRAN simulations, we observe that a threshold seafloor TOC concentration of 0.25 wt% is necessary for shallow gas to occur (Figure 7-27). For TOC values lower than this, the shallow hazard risk is low.

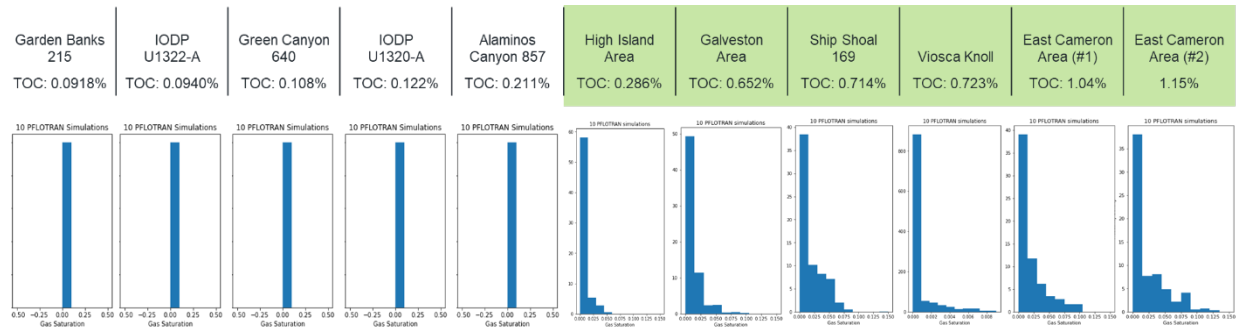


Figure 7-27. Predicted gas concentrations arranged in order of increasing seafloor TOC from left to right. Locations highlighted in green are predicted to have shallow gas.

Our predictions illustrate that gas is more likely to be present along the shelf where TOC values are high, compared to low TOC values in deep water. We can support the assertion that 0.25% dry weight is a minimum amount of TOC that indicates the presence of a shallow gas hazard. Furthermore, 10% gas saturation is often considered the mobility threshold. Comparing the likely gas saturation profiles with the TOC distributions, we can conclude that 0.25% dry weight TOC may be needed for 10% gas saturation. Furthermore, regions along the shoreline and continental shelf where sedimentation rate is greater than 0.1 cm/yr are more likely to have shallow gas hazards. Areas with high drilling activity in the Gulf of Mexico, such as Alaminos Canyon, Green Canyon, and Garden Banks, are less likely to have shallow gas hazards. Any seafloor seeps identified in areas of high drilling activity likely originate from deep reservoirs, not from shallow gas accumulation.

This work provides granular predictions of shallow geohazards on a basin scale, and offers a holistic approach to identifying shallow hazard using big data and machine learning techniques. Leveraging geospatial machine learning models improves the predictions of subsurface hydrocarbons, despite scarce sampling of total organic content in the seafloor. Furthermore, these predictions also include associated uncertainties, which improves engineering analysis.

7.4. Greenland-Norwegian Sea

The seafloor between Greenland, Iceland, Scandinavia, and Svalbard contains numerous sites where free gas and gas hydrate have been detected in sediments. Off the east coast of Greenland, evidence of hydrate occurrence comes in the form of gassy sediments and BSR signals (Minshull et al., 2020; Knies and Mann, 2002). Additional BSRs have been observed during analysis of the Storegga Slide off the northwest coast of Norway (Brown et al., 2006). The presence of gas and hydrate off the coast of Svalbard is also widely documented by sonar (Veloso-Alarcón et al., 2019; Westbrook et al., 2009), hydroacoustic analysis of the Hornsund Fracture Zone (Mau et al., 2017), and sediment core collection (Treude et al., 2020). Therefore, this region represents another study area to demonstrate the Dakota-PFLOTRAN workflow.

The GPSM values for a study area between 30°W and 30°E and 65°N and 80°N were used to construct a study area for the region between Scandinavia and Greenland, termed the Greenland-Norwegian Sea study area. Sedimentation rate is generally highest in shallow locations, with exceptionally high rates found north of Iceland (Figure 7-28). The TOC values tend to be higher nearshore but TOC exceeds 2% in the region between Svalbard and Norway in the eastern part of

the map (Figure 7-29). High heat fluxes are associated with many locations near Greenland and Iceland, as well as off the west coast of Svalbard (Figure 7-30). Similarly, warmer seafloor temperatures are associated with nearshore locations throughout the study area (Figure 7-31). Finally, regions with depths <1,000 mbsl extend off the coasts for hundreds of kilometers and the plate boundary can be seen along the middle of the study area (Figure 7-32).

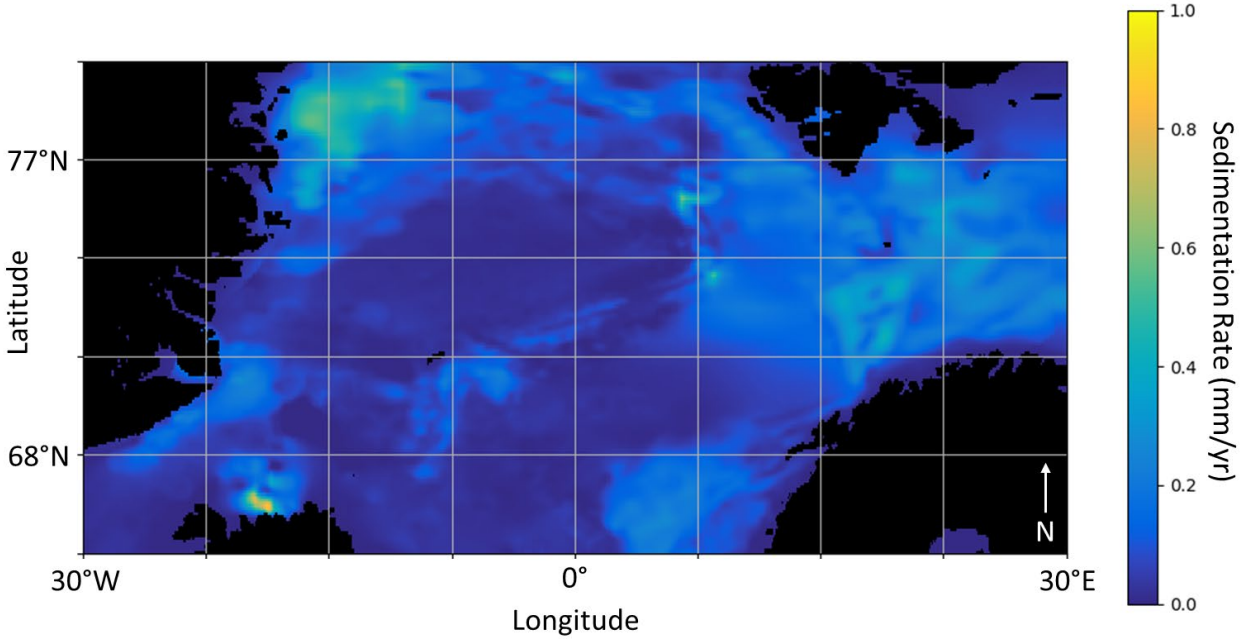


Figure 7-28. Annual sedimentation rates for Greenland-Norwegian Sea study area based on Straume et al., (2019). Higher sedimentation occurred near northeast Greenland and north of Iceland.

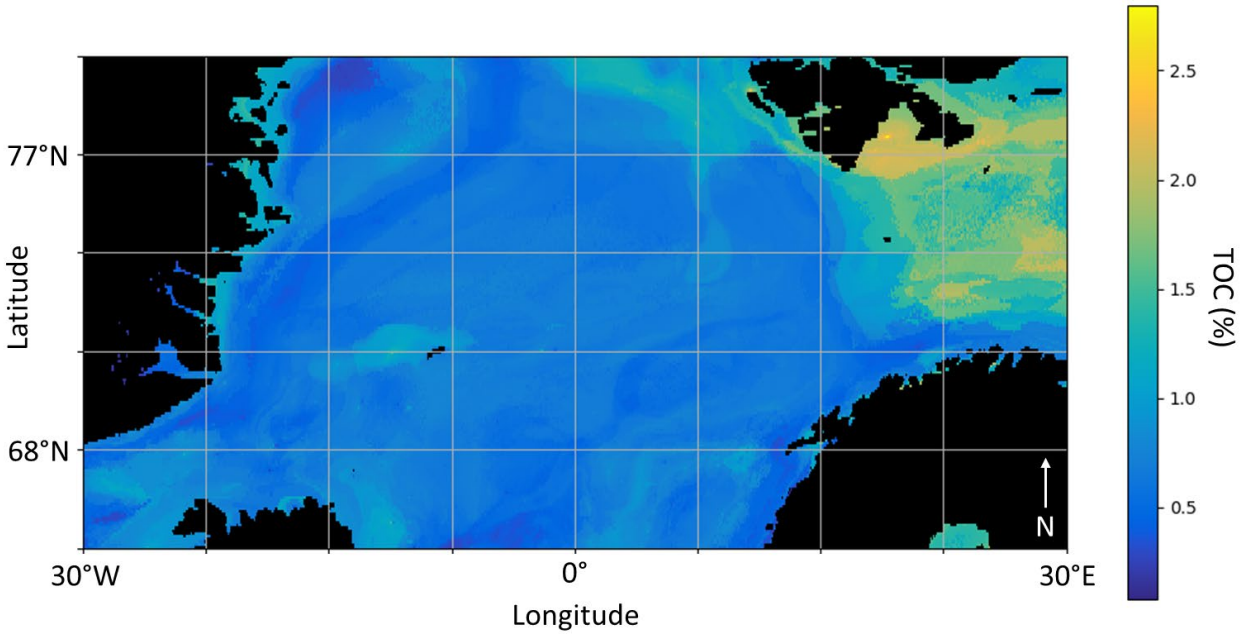


Figure 7-29. Total Organic Carbon (TOC) for the Greenland-Norwegian Sea study area. TOC values above 2% are common for the section between Svalbard and Norway.

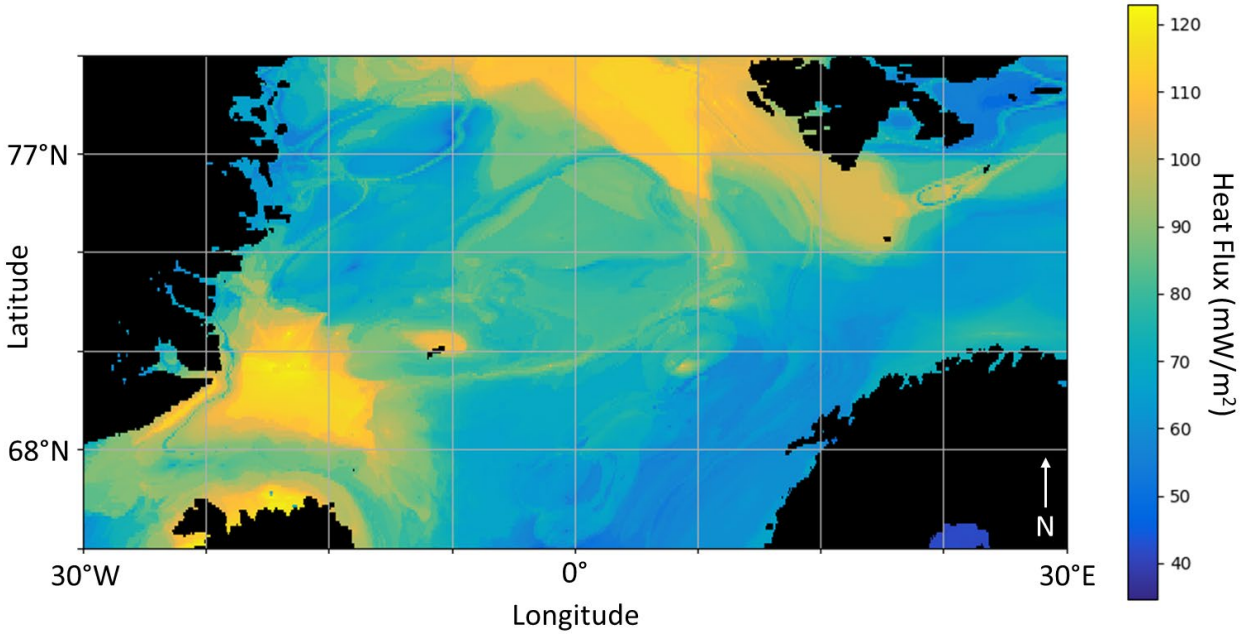


Figure 7-30. Heat Flux for Greenland-Norwegian Sea study area. Elevated values of heat flux occur along mid ocean ridge and near Iceland hotspot.

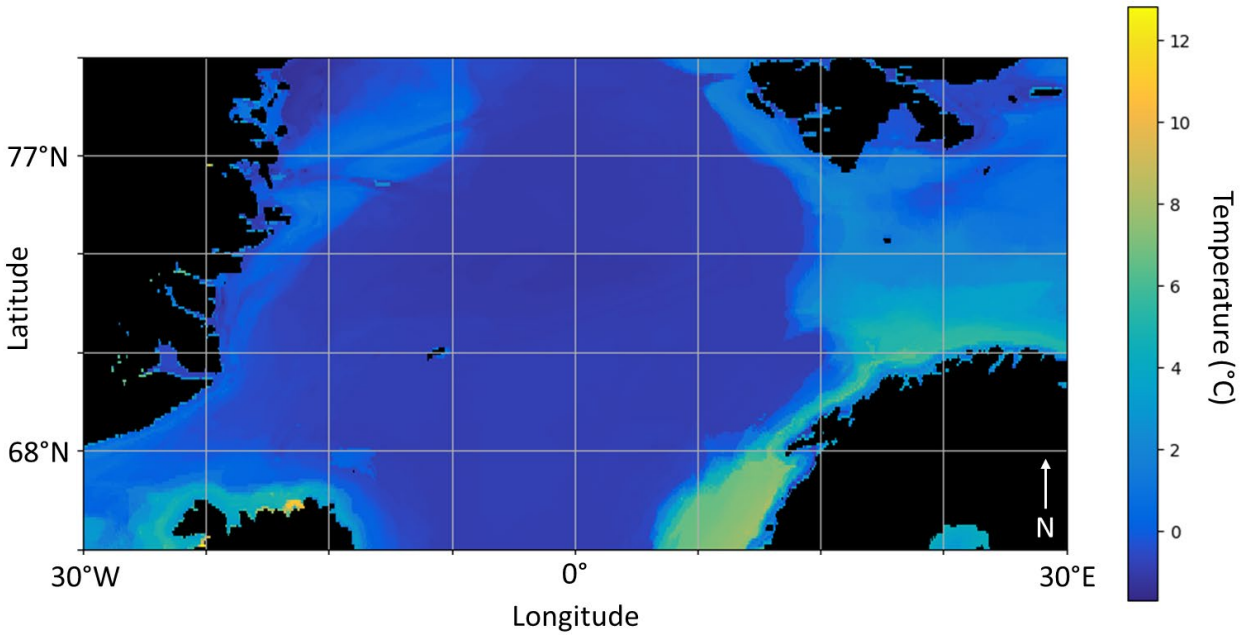


Figure 7-31. Seafloor temperature for Greenland-Norwegian Sea study area. Higher temperatures coincide with nearshore areas along coastlines.

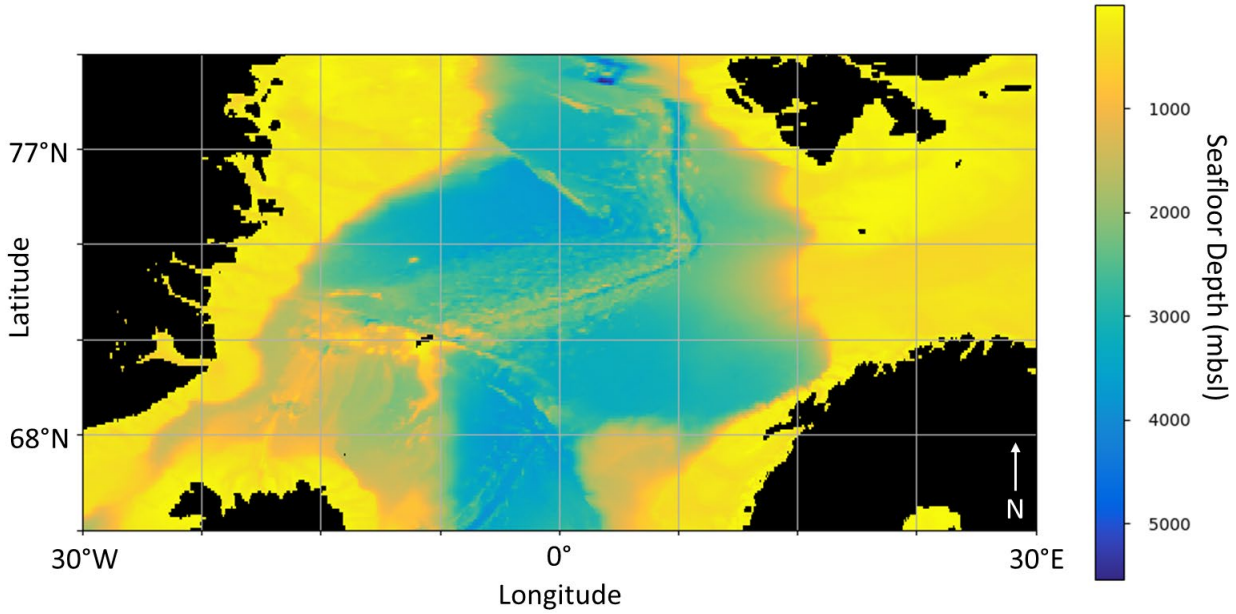


Figure 7-32. Seafloor depth for Greenland-Norwegian Sea study area. The mid ocean ridge is clearly visible, running as the thin line from Svalbard down to Iceland.

The Greenland-Norwegian Sea study area encompasses 129,600 locations, 96,857 of which are offshore and can be simulated. Dakota was used to sample on the sedimentation rate, TOC, heat flux, temperature, and depth associated with every location and the distributions were used to initiate PFLOTRAN simulations. We ran five simulations on each of these locations for 500 kyr to determine the formation of free gas and gas hydrate. The masses of gas hydrate and free gas formed in the sediment column during the simulations were calculation following Equation 3.1 and 3.2, and the average was taken between all five simulations at each location.

The average gas formation shows three regions of significant free gas occurrence north of Iceland, along the northeast coast of Greenland, and the highest occurrences between Svalbard and Norway (Figure 7-33). Nearly all locations between the Norwegian and Barents Seas formed free gas during the simulation time when seafloor depth was <1,000 mbsl (maximum of 9,381.0 kg). Effectively no gas forms in regions approaching the mid ocean ridge as the available TOC and sedimentation rates decrease. Generally, the boundaries of the gas forming areas envelope the regions which will be suitable for gas accumulation in the Greenland-Norwegian Sea study area.

The formation of gas hydrate in the Greenland-Norwegian Sea study area was much more restricted compared to free gas occurrence (Figure 7-34). Elevated amounts of formation occurred off the northeast coast of Greenland between 17°W - 10°W (maximum of 2,554.6 kg), with lower masses forming between the Barents and Norwegian Seas from 15°E-30°E. Intriguingly, high formation was predicted for the west coast of Svalbard (7°E-9°E, north of 77°N) and may be associated with gas migration along the Hornsund Fracture Zone (Mau et al., 2017). This region was also included in the areas of high gas formation, suggesting this system is crucial to seafloor methane considerations.

The Greenland-Norwegian Sea study area was selected to demonstrate the Dakota-PFLOTRAN workflow and evaluate its performance in an Arctic region. Hydrates have been either directly observed or inferred in nearshore locations along Arctic coasts (Hester and Brewer, 2009). Because all regions in the Arctic are warming (Meredith et al., 2019), increased temperatures threaten the stability of hydrate and free gas contained in sediments. Methane dissociation has the potential to

decrease ocean pH (Biastoch et al., 2011) and poses a risk to climate change if the gas is able to reach the atmosphere (Ruppel and Kessler, 2017). Predictions based on high resolution GPSM data can provide improved predictions for the location, volume, and risk of hydrate and gas occurrence that more accurately represent the previously less-constrained Arctic areas.

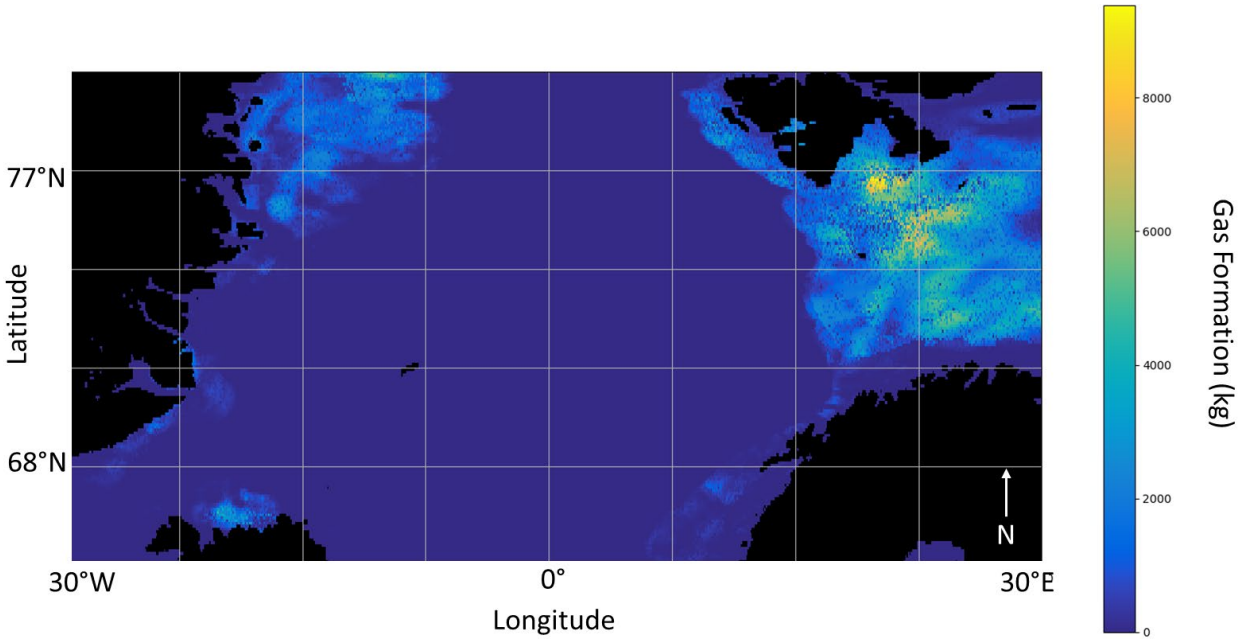


Figure 7-33. Average gas formation for each sediment column. Gas formation was more abundant and more frequent in the Greenland-Norwegian Sea study area compared to hydrate formation.

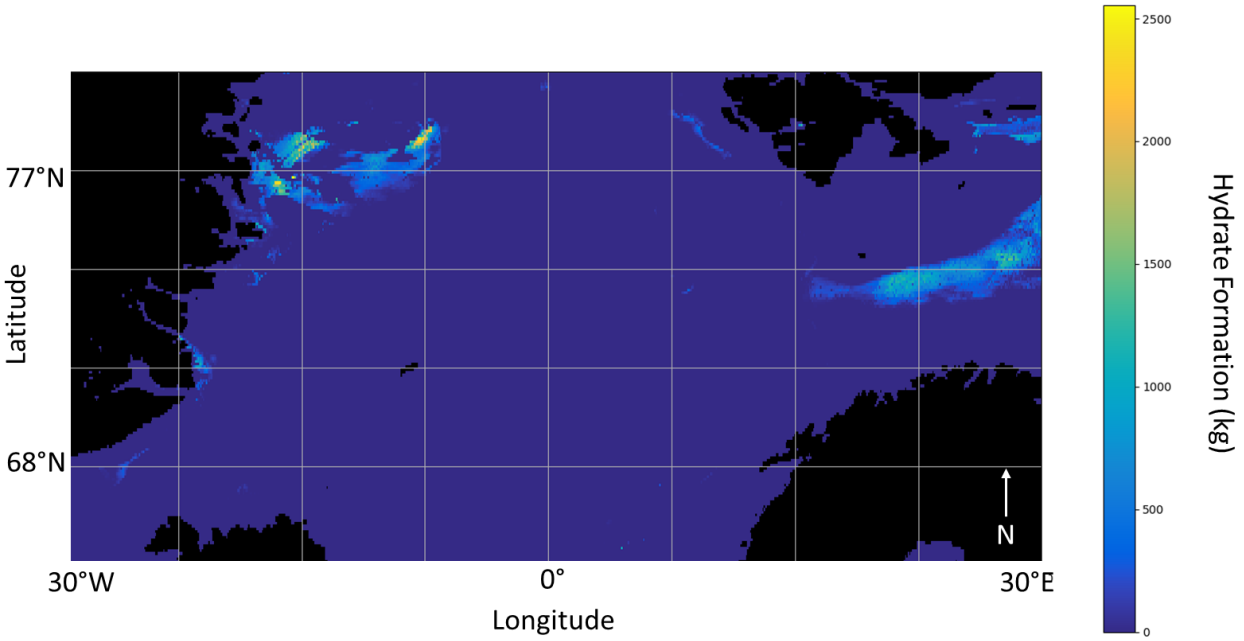


Figure 7-34. Average hydrate formation for each sediment column. Locations off eastern shore of Greenland and between Svalbard and Norway were the only regions to form substantial amounts.

7.5. Machine Learning Driven Regression Model

In this work, we tested three machine learning (ML) models including artificial neural networks (ANNs) and convolutional neural networks (CNNs) to predict free gas and gas hydrate formation. PFLOTRAN simulations as described in Section 7 and Sections 8.1-8.2 were used to generate training data of free gas and gas hydrate formation as a function of sedimentation rate (mm/yr), TOC (%), Q (mW/m^2), depth (mbsl), and temperature (°C). A total of 24000 datasets were generated. All physical input and output data as shown in Figure 7-35 were normalized from 0 to 1 for ML training. Three different groups including no hydrate, only free gas, and both free gas and hydrate formation are separately and randomly split into training/validation/testing sets with a ratio of about 7:1.8:1.2.

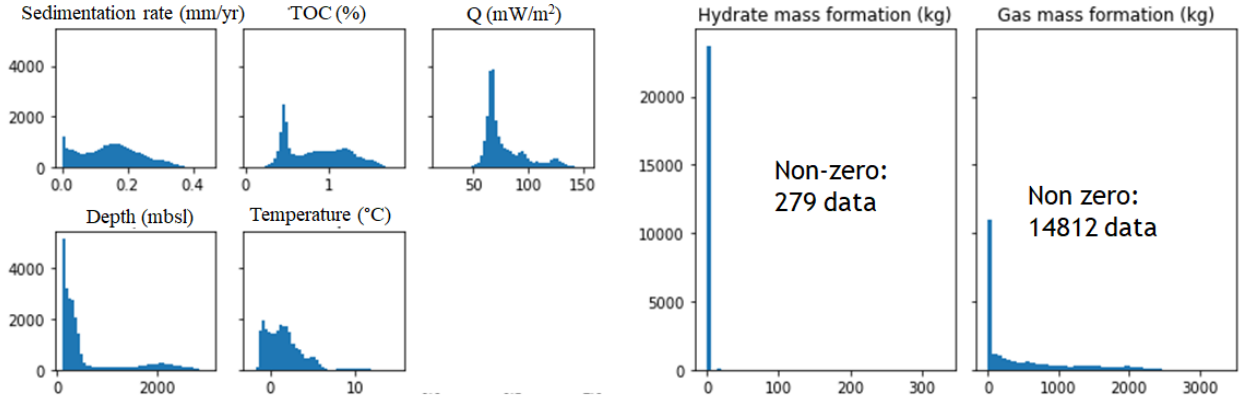


Figure 7-35. (Left) Distribution of five varying input parameter values for PFLOTRAN simulation including sedimentation rate (mm/yr), TOC (%), Q (mW/m^2), depth (mbsl), and temperature (°C). (Right) Distribution of formed hydrate and free gas mass with a number of non-zero cases. The horizontal and vertical axes are corresponding to the range of parameter values and frequency of parameter values, respectively. A total of 24000 simulation data are generated.

In this work we used two different CNN architectures to construct ML-driven models for predicting hydrate formation. CNNs have been widely used for image classification and segmentation and also applied for regression tasks of various scientific and engineering problems (e.g., Anwar et al., 2018, Ling et al., 2016). As shown in Figure 7-36(a&b) two CNN architectures with different complexities are chosen to develop robust ML-driven models for prediction of hydrate & gas formation. The first CNN architecture includes four CNN-based models with each having different numbers of CNN layers, filters and kernel sizes. Maxpooling is used in only one CNN branch. This particular structure shown in Figure 7-36(a) is to extract various relationships of five input parameters at different scales. All four CNN models are concatenated, followed by a series of dense layers with different sizes to construct a regressor with an output dimension of 2. The second CNN architecture (Figure 7-36b) consists of three CNN1D layers with 32, 64, 128 filters and a kernel size of 2, followed by 4 dense layers with a size of 256, 96, 16, and 2 per each layer. This second CNN is a vanilla type standard CNN structure for feature extraction and regression. Hence, we can compare the performance of two CNN-based models. A number of filters in CNN1D layer, a number of epochs from 5,000 to 20,000, learning rate, activation function, dropout, and a number of dense layers and size are manually evaluated to find out optimal hyperparameters.

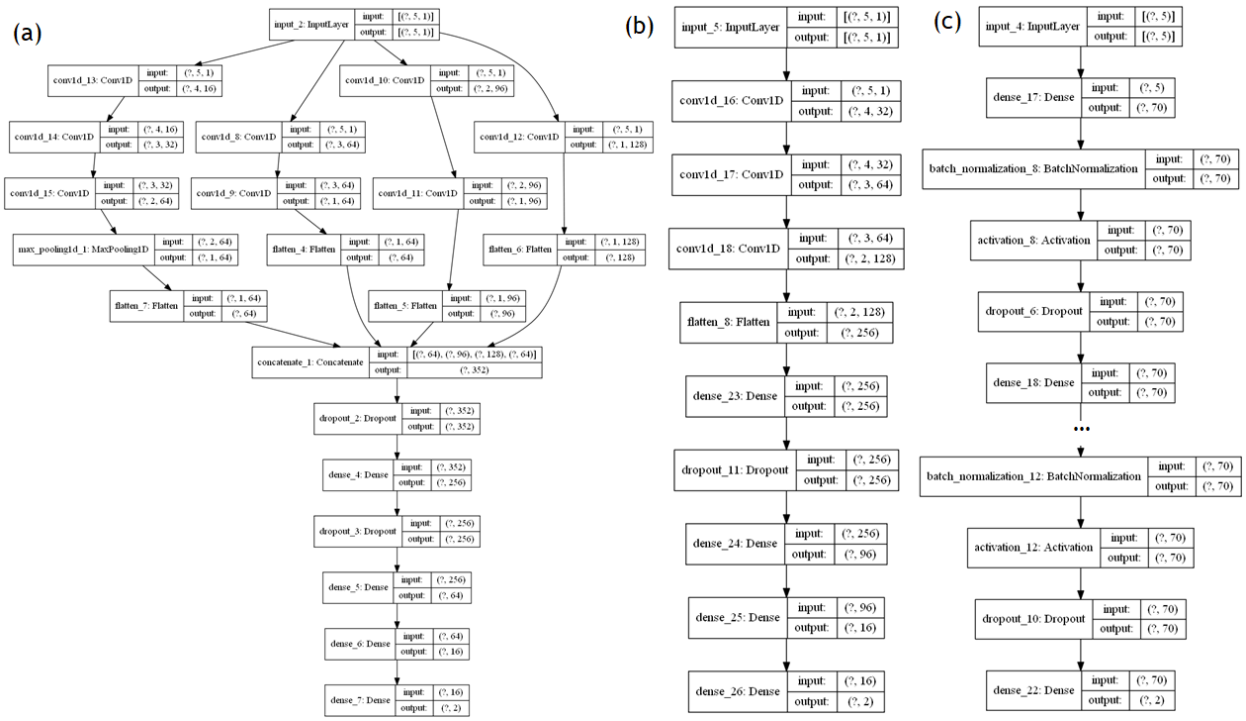


Figure 7-36. (a) Stacked four CNN architectures with different levels of depth and filter sizes for multiscale pattern extractions. Conv1D was used. (b) Single CNN architecture with three Conv1D layers followed by a series of dense layers. (c) ANN architecture with a repeat of hidden layers using the same number of nodes per each layer.

ANNs consist of input, hidden layer(s) with nodes, and output with a number of hidden layers and nodes as primary hyperparameters. Due to its simple structure and as universal approximator ANNs have been widely used as a regression model (e.g., Abiodun et al., 2019). Advances in ANNs-based applications have been investigated for hyperparameter optimization, hybrid neural networks models, and integrated approaches with other neural network or deep learning methods among others. Here, we use a relatively simple ANN architecture consisting of a number of hidden layers (5-8) with the same number of nodes per each layer (40-80) and batch normalization. We did a limited sensitivity analysis to optimize the hyperparameters.

Figure 7-37shows the prediction of hydrate and gas masses using the best ML-based models with three different architectures described above for testing datasets (~3,000 data). Overall, the prediction of hydrate formation has been predicted very well for all three models. Particularly, two CNN based models predict extremely well although there are only 279 cases out of 24,000 with hydrate formation. As described, all ML models are trained together to predict hydrate and gas formation. It is noted that there are many multiple CNN-based solutions founded to make a similar prediction as shown in Figure 7-37, while ANN-based models are relatively limited to achieve similar performance. This simple comparison shows that CNN-based models is superior to ANN-based approach to extract information better for this problem. In contrast, the prediction of gas formation has more complex behaviors. First, for the most of data both CNN-based approaches have a good prediction compared to the PFLOTRAN results (i.e., truth) as seen in most of data points near the 1-to-1 line. However, scattered patterns of low range of gas mass formation is observed in all three models, indicating that under certain conditions the training data is not enough to train the ML-based models well. It should be noted that training results are much better than those in Figure 7-37

where testing data unseen during training is displayed. Although it is further needed to investigate the conditions where the current ML-based models are struggled to make a good prediction of gas formation quantitatively, data screening for training may help to improve ML training in the future.

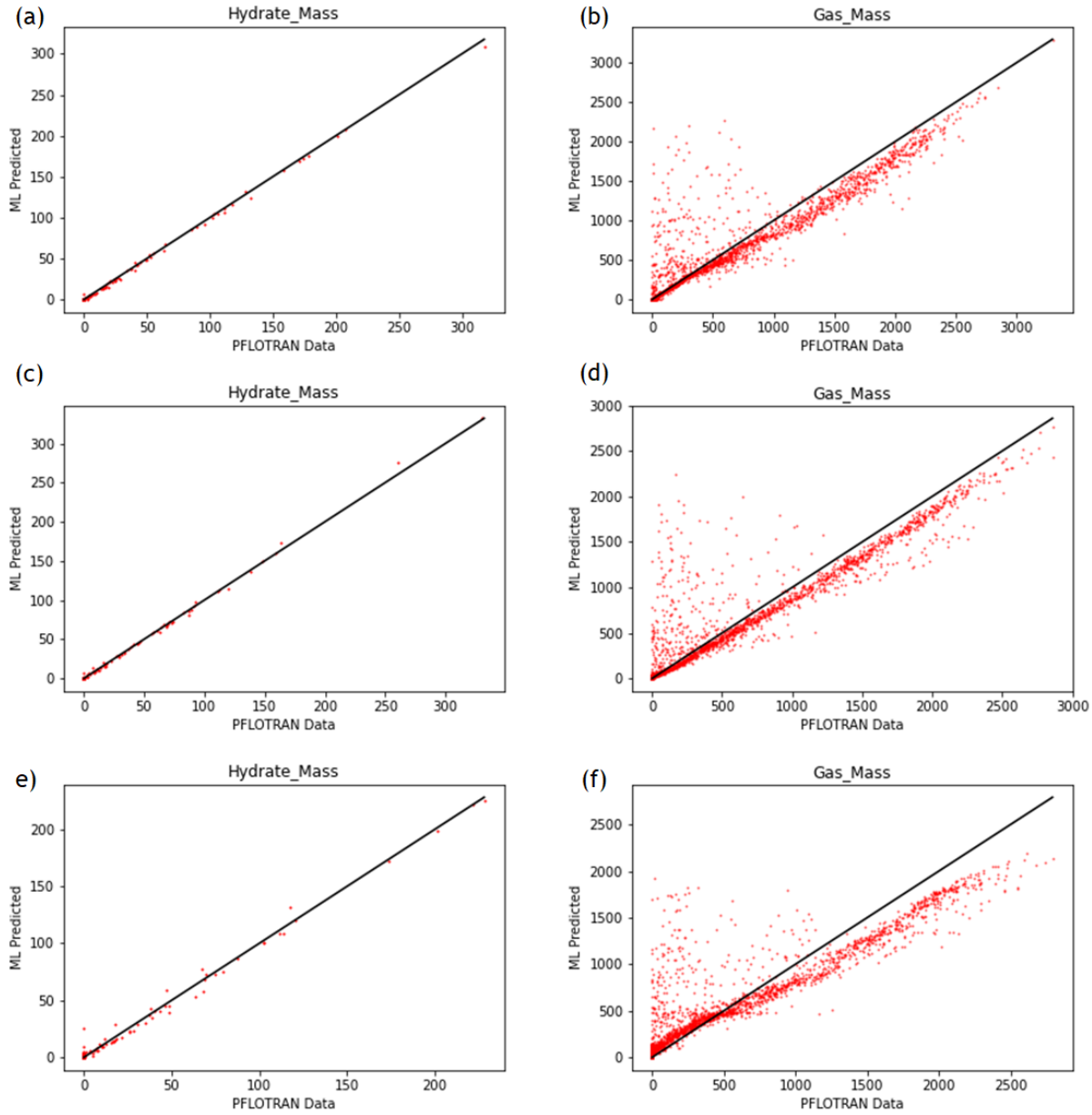


Figure 7-37. Parity plots of ML prediction vs. PFLOTRAN data (i.e., truth) for hydrate and gas formation. (a&b) stacked CNN architecture, (c&d) single CNN architecture, and (e&f) ANN architecture. All ML model architectures are shown in Figure 7-36.

Although prediction of formed gas mass looks scattered, a fraction of these scattering data is about 4% of all testing data. As in the hydrate formation, CNN-based models have higher equally likely sets of hyperparameters compared to ANN-based models, indicating that if we adopt for automatic hyperparameter optimization, it is likely to find a number of models for good prediction of both hydrate and gas formation with CNN-based models. It should be noted that ML-based regression

models developed here can be used as a surrogate model for fast prediction (e.g., a few seconds for 100 model prediction), so we can perform highly computationally demanding jobs quickly as well as initial screening for high fidelity modeling studies.

7.6. Alaskan North Slope

Submarine permafrost in the Arctic has been estimated to trap 60 billion tons of methane and contain 560 billion tons of organic carbon within seafloor sediments and soil (Sayedi et al. 2020). Much of the North Slope exists within the gas hydrate stability zone and the presence of gas hydrate has been determined from core samples and logging data (Collett et al., 2011). Gas hydrate associated with permafrost in this region has been hypothesized to form by the conversion of trapped free gas following glaciation (Collett, 2002; Boswell et al., 2011). While research shows that dissociation of permafrost-trapped methane is unlikely to reach the atmosphere through the water column, the rates of release or microbial oxidation could change in response to global warming of the Arctic (Sparrow et al., 2018). A better understanding of the effects of potential methane gas release on long term climate concerns requires probabilistic predictions of subsurface methane gas formation. We applied the Dakota-PFLOTRAN workflow to a case study on Alaska's North Slope to develop new capabilities and evaluate the performance in predicting shallow gas formation in the Arctic.

To further demonstrate the model capabilities in the Arctic, we developed a flexible version of the Dakota-PFLOTRAN workflow that runs on Sandia's Common Engineering Environment (CEE) Linux system. This updated workflow allows for flexible sampling setups and incorporates changes in relative sea level (RSL) over the simulation time. We used the GPSM values for 1,565 locations between 157°W and 142°W and 70°N and 72°N to cover the shallow coast off of the North Slope of Alaska. The sedimentation rate was consistently low along the coast in the study area, with rates reaching 0.16 mm/yr in the southeastern portion (Figure 7-35). TOC values were scattered but exceeded 2% in the eastern half of the coastline as well as further north offshore (Figure 7-36). The study area exhibited modest heat flux values which were highest nearer to shore but ranged between 45-75 mW/m^2 (Figure 7-37). Due to the Arctic setting, seafloor temperatures were almost always just above 0 °C even in shallow depths, but some coastal areas such as near the village of Utqiagvik (71.28°N, 156.79°W) exhibited temperatures exceeding 20 °C (Figure 7-38). Finally, the seafloor depth followed the trace of the coastline, with deeper portions of the study area in the northwest (Figure 7-39). Any or all of these variables could be toggled on or off for Dakota sampling in the new workflow.

The RSL time history was introduced to account for the changes in sea level due to glaciation over the past 120 kyr (Figure 7-40). This time history is used to apply pressure conditions that vary through time due to changing water depth as well as adjusting the temperature when the location would be exposed to atmospheric as opposed to oceanic conditions based on site specific depth. Due to glaciation, much of the time history is at lower sea level relative to today's conditions.

To initiate this flexible workflow, we maintained a constant porosity of 70% for the sediment column and simulated for one full glaciation cycle of 120 kyr. All locations with depths ≤ 200 mbsl were selected and 25 simulations were run at each site. While the workflow allows for Dakota to sample on all five GPSM parameters, several limitations were observed for this study area. Due to the shallow depths associated with the coastline, frequent numerical issues occurred when very shallow depths were selected from the Dakota distributions coupled with high TOC values. As a

temporary solution, we chose to retain the GPSM averages for seafloor depth at each location rather than sampling on this parameter. Similarly, the cold temperatures associated with the Arctic resulted in ice formation during simulations where the Dakota-provided seafloor temperature was negative. Therefore, we emplaced a minimum temperature of 0.1 °C for all North Slope simulations.

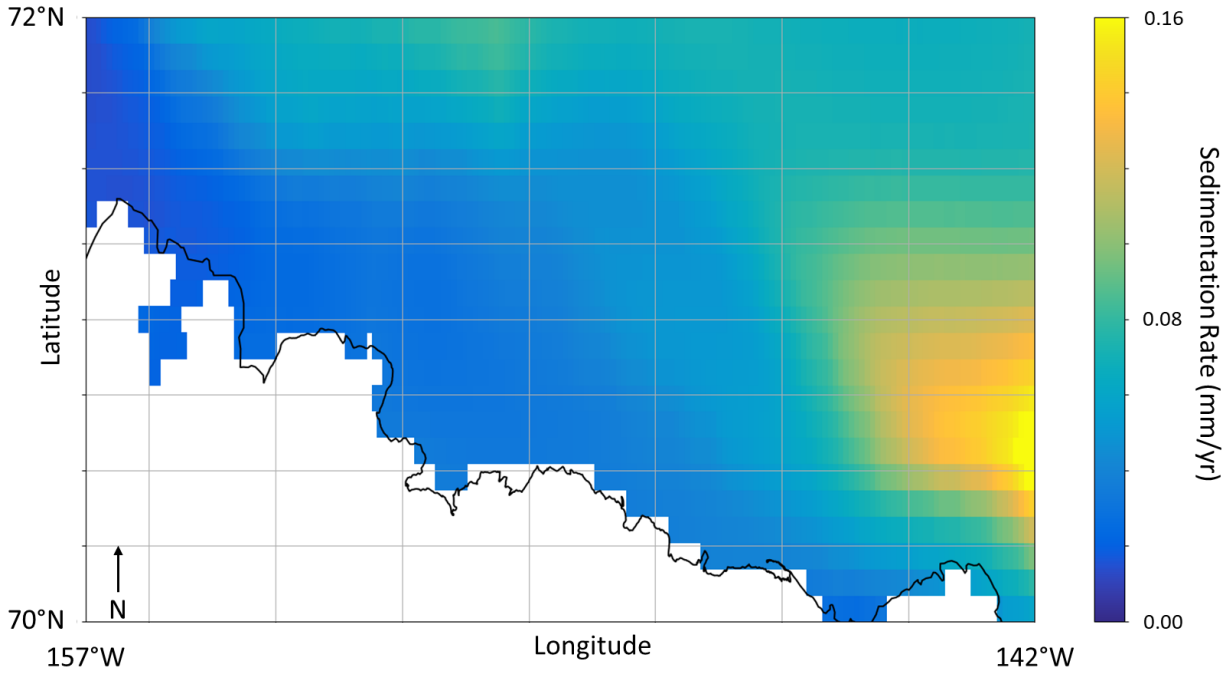


Figure 7-38. Sedimentation rate for North Slope study area.

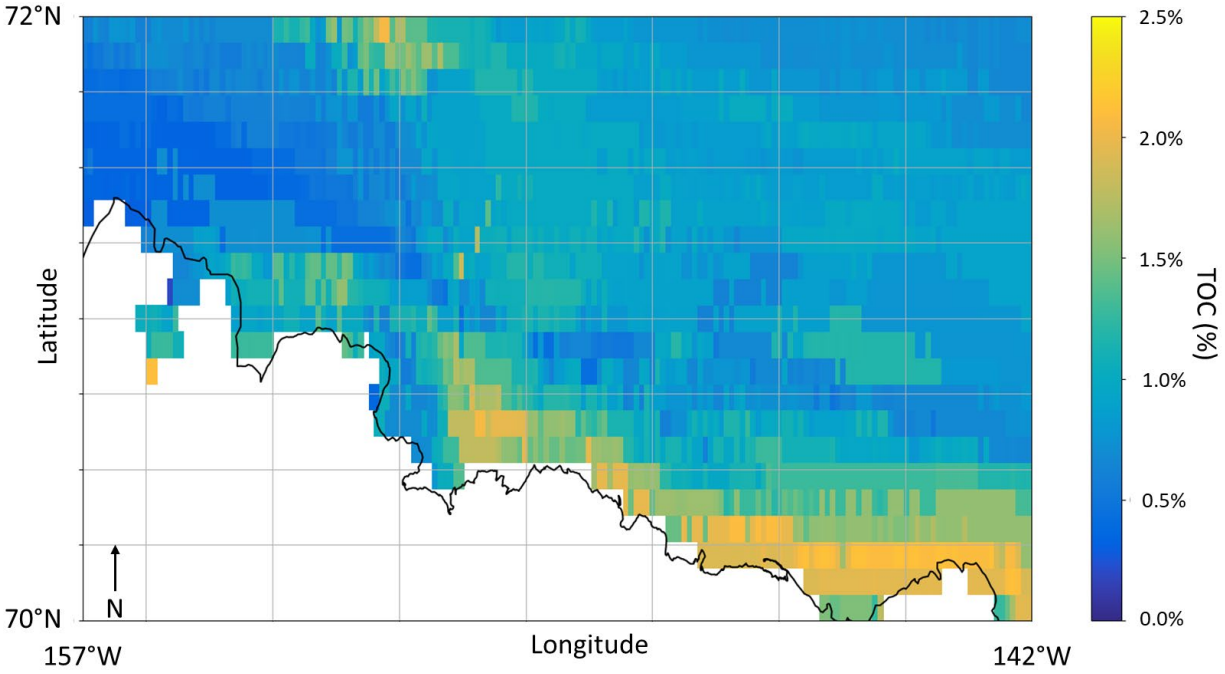


Figure 7-39. TOC values for North Slope study area.

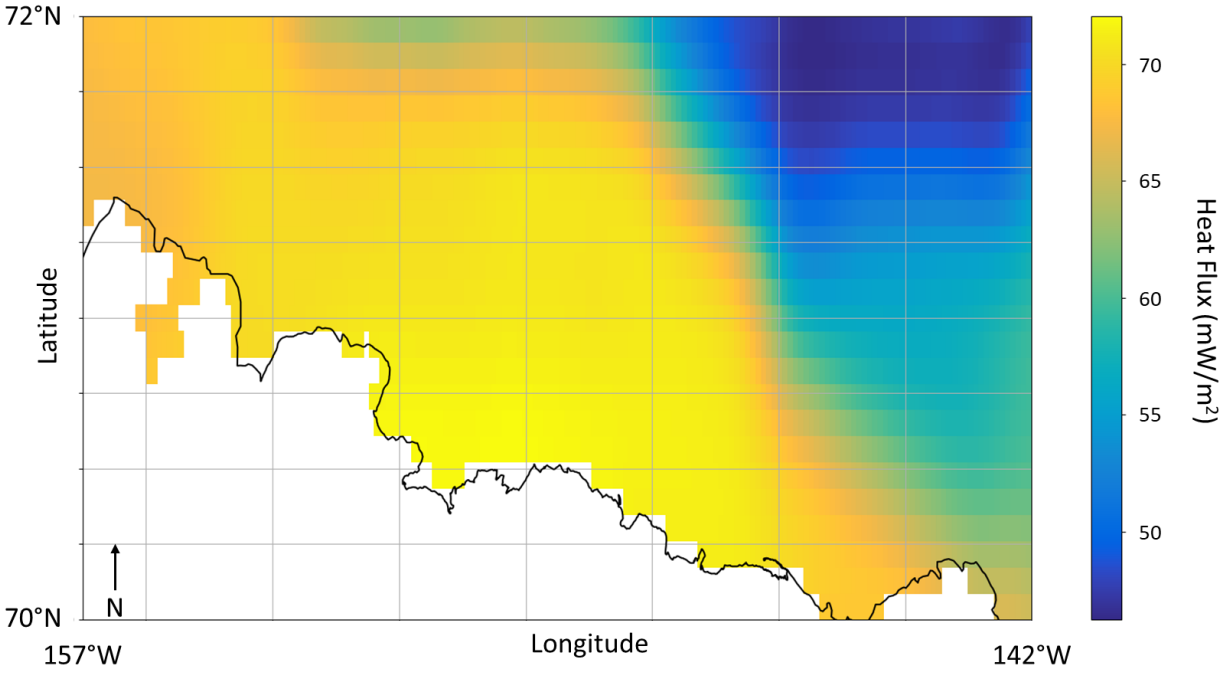


Figure 7-40. Heat flux values for North Slope study area.

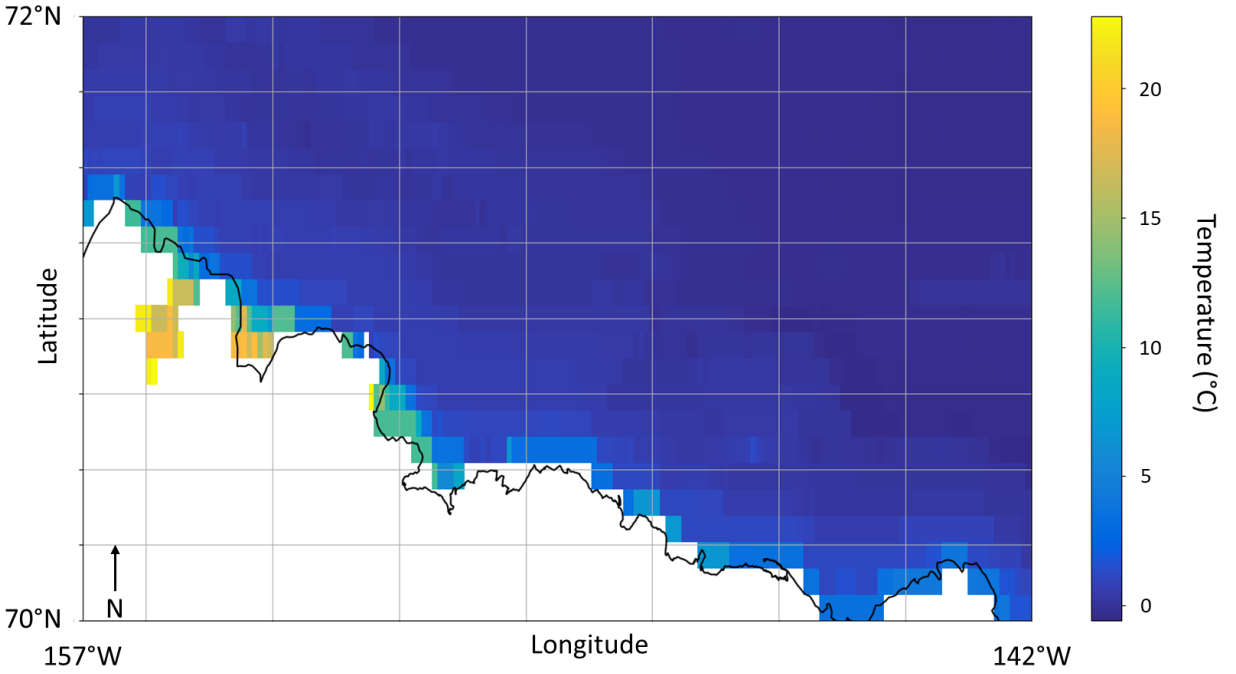


Figure 7-41. Seafloor temperature for North Slope study area.

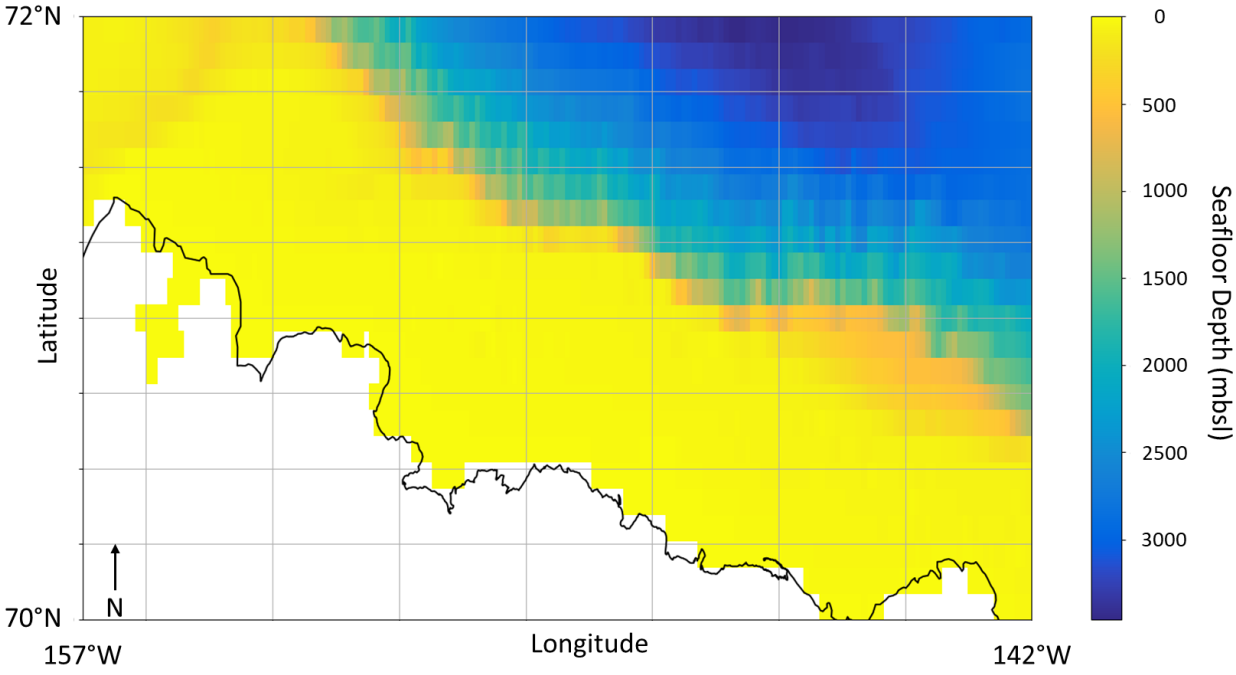


Figure 7-42. Seafloor depth for North Slope study area.

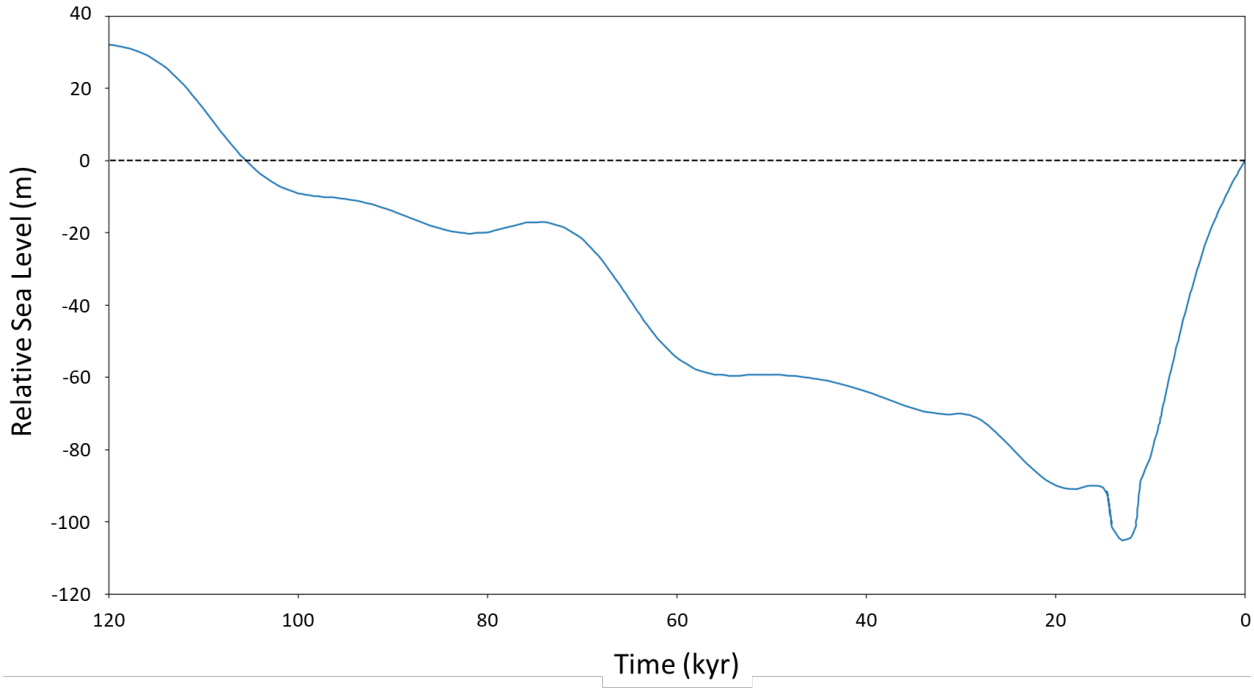


Figure 7-43. Relative sea level due to changes in glaciation. Modern day sea level is shown as a dashed black line.

The results from the simulations (Figure 7-41) suggest that free gas is ubiquitous in the sediments along the North Slope, though in low quantities (maximum of 132.7 kg). The low mass of gas likely results from the short simulation times which should be increased in the future. Locations along the 200 m contour exhibited slightly higher formation masses than locations nearshore. We hypothesize that this is due to more elevated sedimentation rates away from the coastline and further evaluation should extend the simulations to all depths in the study area.

By including the RSL time history into PFLOTRAN, we can apply more realistic boundary conditions for the Arctic but more adjustments are necessary. The 120 kyr simulation times are too short to allow for simulations to reach steady state and the duration should be extended. Repeating the RSL time history will incorporate the effects of multiple glacial cycles over these longer simulation times. The temperature restriction emplaced is unrealistic and next simulations should include salinity to reduce the freezing temperature of water and prevent ice formation in the sediment columns. Finally, the constant porosity was applied to allow greater flexibility of the workflow to run on CEE but future simulations will incorporate the site-specific porosity profiles available in GPSM. Overall, these efforts demonstrate that the workflow has been updated to include more customization for specific studies and will be able to run on external Linux systems once the workflow is made publicly available.

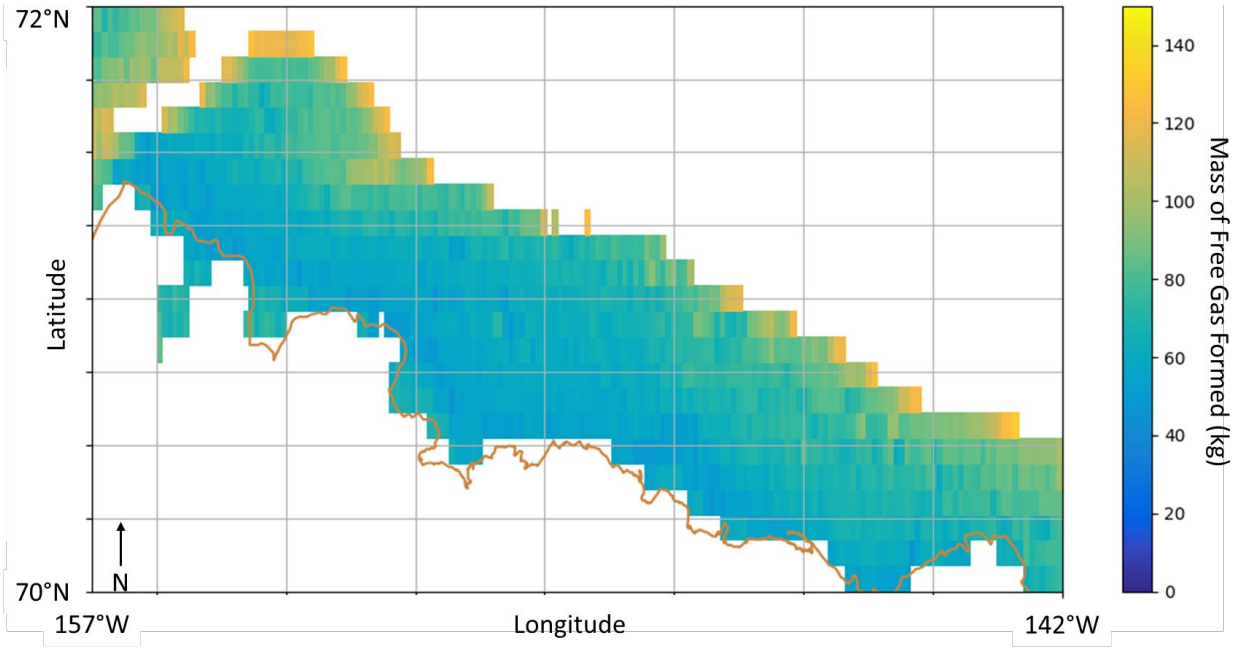


Figure 7-44. Average gas formation for the North Slope of Alaska. All simulations yielded gas at every location in the study area.

7.7. Global Forecasting of Free Gas and Gas Hydrate Occurrence

To showcase the generalizability of the GPSM-Dakota-PFLOTRAN framework, we applied the framework at a global scale. This study presents results from a global application of this data-driven workflow for integrated probabilistic gas hydrate systems modeling, which includes parameter distribution prediction (GPSM), parameter sampling (Dakota), mechanistic simulation (PFLOTRAN), statistical analysis (Dakota and Python scripting), and 3D visualization (ParaView). Coarse mapping of gas hydrate and free gas at this scale can yield insight into which locations to sub-sample for more rigorous uncertainty quantification; this kind of study can also be used to identify which regions of the globe could be most sensitive to global climate change with respect to their gas or hydrate deposits.

In this study, 500 simulation realizations were run for each of 35,136 distinct locations across the globe (excluding land, 1-degree). For each location, data was extracted from GPSM-generated maps of expected value and standard deviation of heat flux (Figure 7-42) and seafloor total organic carbon (Figure 7-43). For each simulation realization at each location, boundary conditions, initial conditions, and sediment material properties were set according to Dakota-derived sample parameter sets which were generated based off expected value and standard deviations prescribed for each parameter at a given location. Five parameters were sampled for this study: seafloor depth, seafloor temperature, heat flux, sedimentation rate, and seafloor organic carbon content. These parameters informed the boundary and initial conditions of simulations run to steady state which consider non-isothermal effects, multiphase flow, in-situ microbial methane generation,

sedimentation, and methane diffusion/advection. The end result of this workflow is a set of 500 sediment column simulations at each location with predictions for each simulation of steady-state free gas and gas hydrate quantities (Figure 7-44).

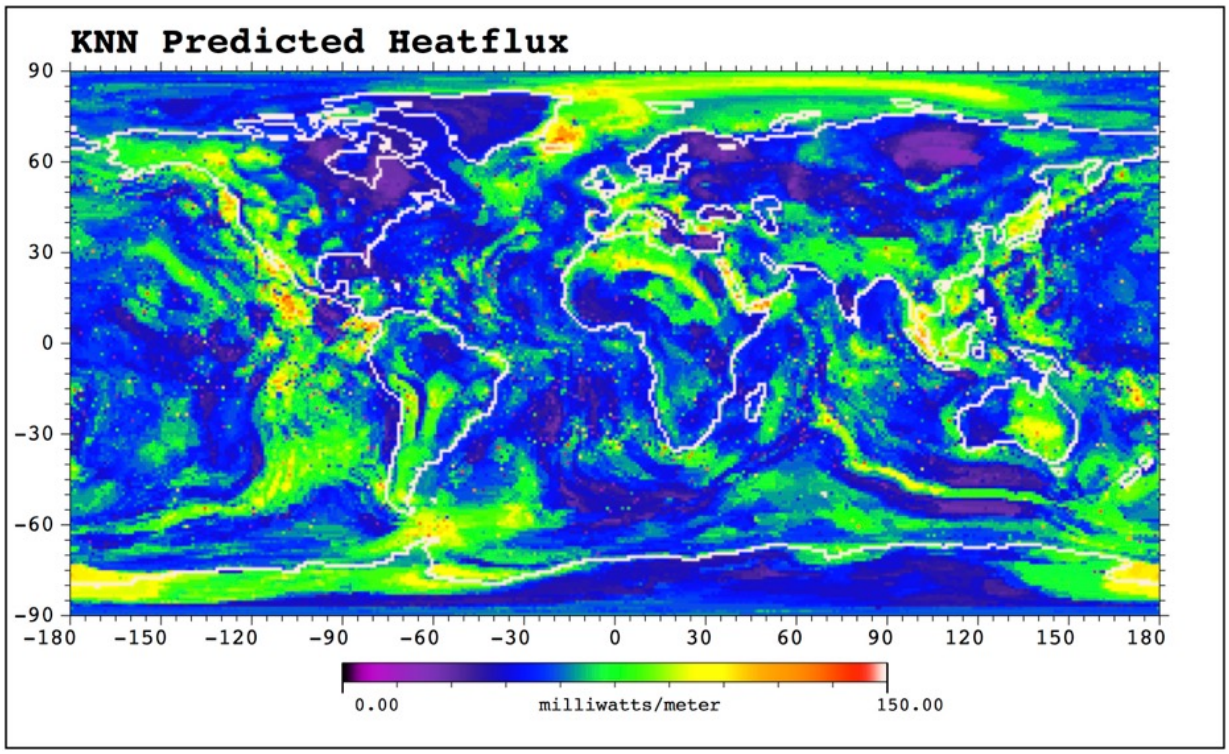


Figure 7-45. Predicted heat flux using K-nearest neighbors algorithm produced by GPSM.

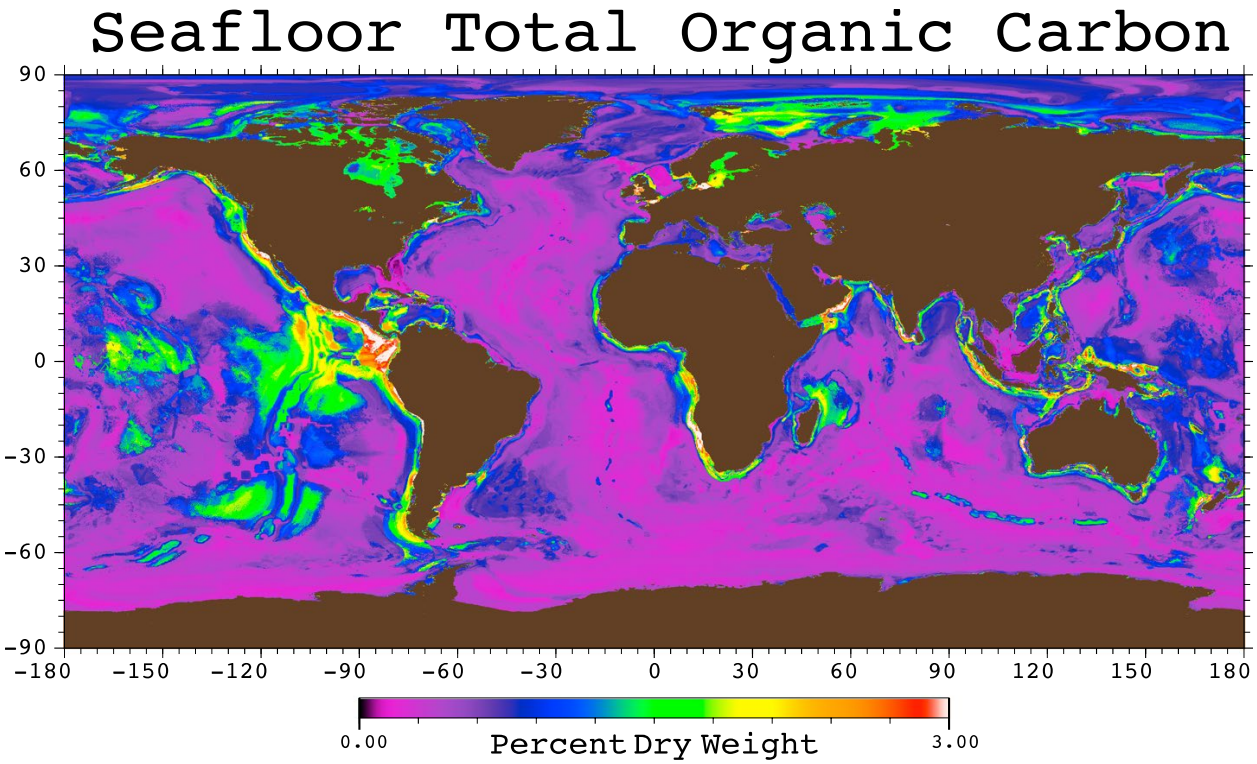


Figure 7-46. GPSM prediction of seafloor total organic carbon (Lee et al., 2019).

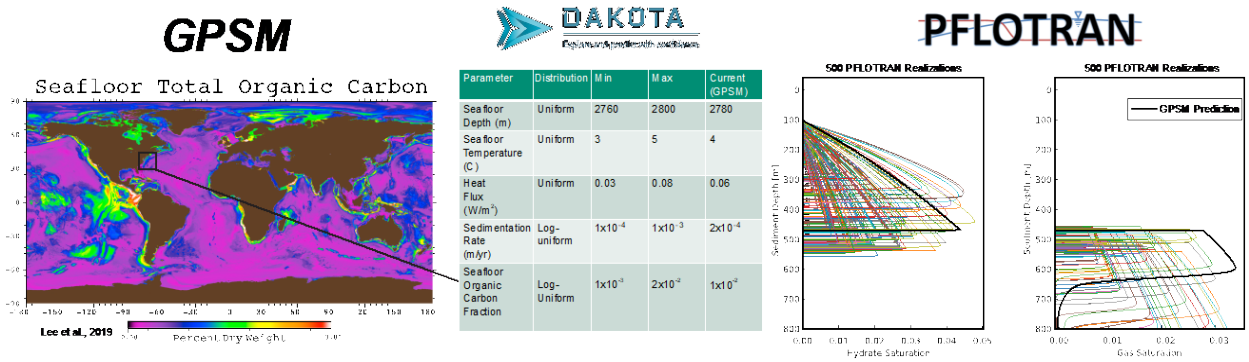


Figure 7-47. Example of the GPSM-Dakota-PFLOTRAN workflow

Once these results are produced by PFLOTRAN, the data can be post-processed by Dakota and visualized with Python scripting to tease out potential relationships between input parameters and outputs of interest. As an example, Figure 7-44 illustrates a snapshot of 500 simulations at one location taken near Blake Ridge offshore North Carolina. For this single location, Dakota post-processing generated a tabular output file containing sampled parameter values for each individual realization along with output variables of interest at a given point in the domain. In this example, the domain point chosen was 500 mbsf which is approximately the current base of the gas hydrate stability zone at Blake Ridge. Figure 7-45 plots a series of scatterplots illustrating the relationships between sampled input parameters (along the bottom axis) and specified output variables (along the left axis). At this location, the strongest positive correlation appears to be between heat flux and sediment temperature. Free gas and hydrate saturations both appear to be positively correlated to

seafloor organic carbon fraction and negatively correlated to sedimentation rate. In addition to scatterplots, statistical measurements of correlation coefficients between input parameters and output variables were also generated by Dakota and plotted in Figure 7-46. These correlation coefficients indicate that organic carbon fraction has a significant impact on dissolved methane concentrations, gas saturations, and hydrate saturations. Correlation coefficients also identify a positive relationship between seafloor temperature and heat flux on sediment temperature, and relationships between other input/output combinations appear mixed.

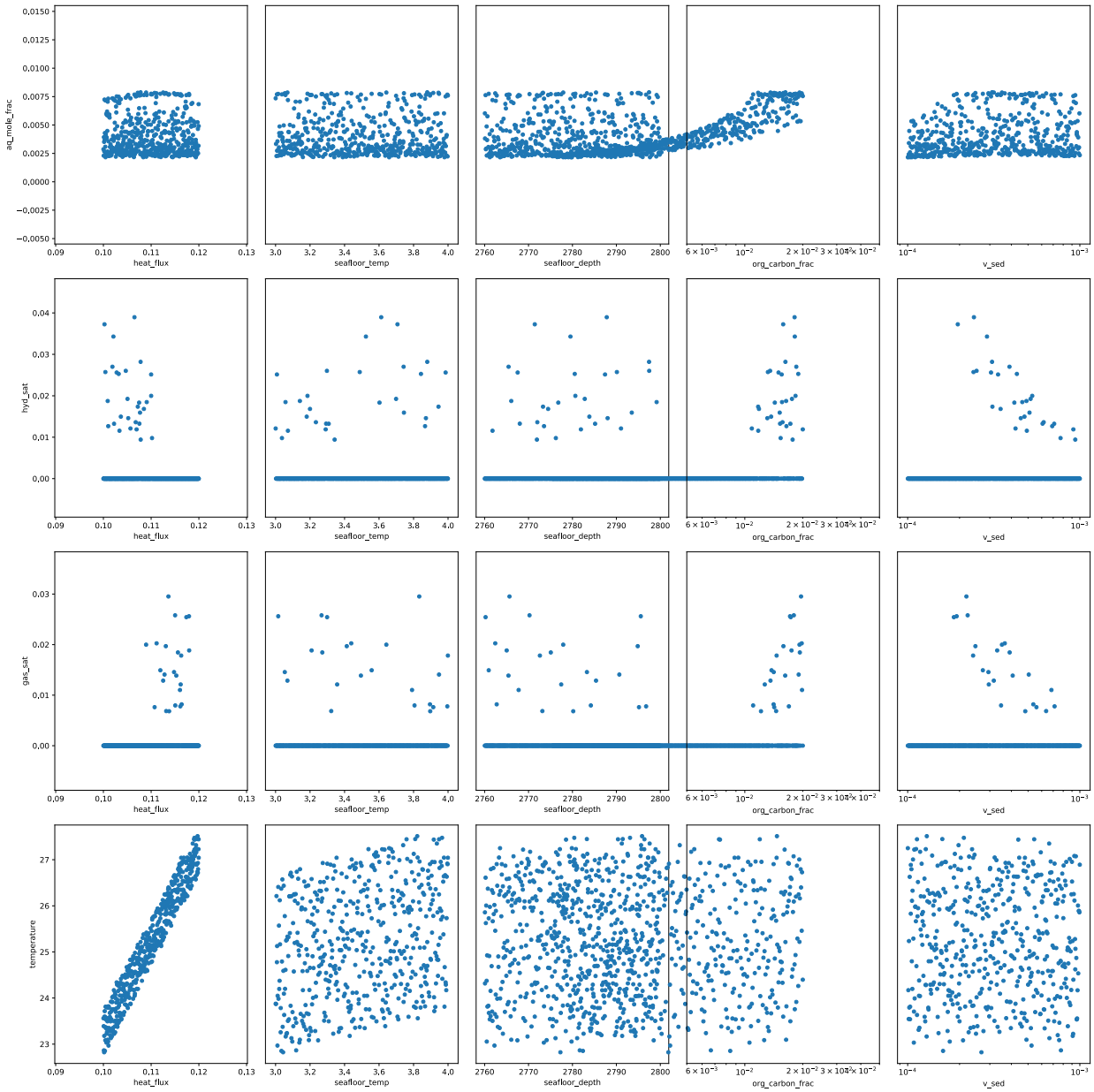


Figure 7-48. Scatterplots of input variables (along the bottom axis) vs outputs (left axis) at 500 mbsf at a location near Blake Ridge.

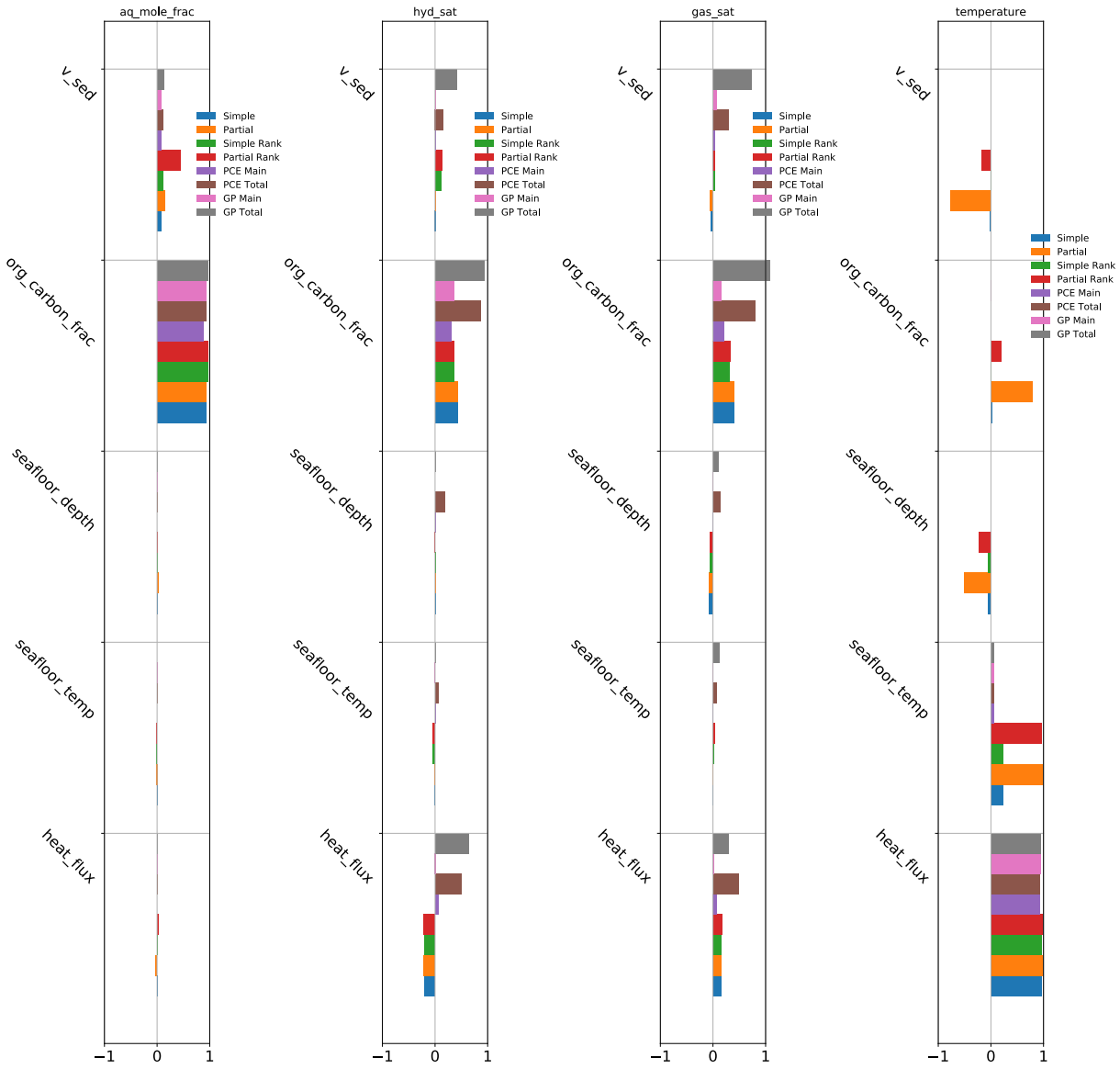


Figure 7-49. Dakota-derived correlation coefficients between input parameters and output variables.

Data from all simulations were then aggregated at a global scale to get a picture of the global distribution of gas hydrate and free gas. The gas hydrate map (Figure 7-47) predicts the highest gas hydrate saturations along continental margins and in the Arctic. This is remarkably consistent in terms of location of hydrate deposits with a USGS map of gas hydrate inferred and recovered around the globe (Figure 7-48). While consistency between these two maps does not validate the accuracy of the specific quantities predicted by the simulations, it does demonstrate that a data-driven approach to gas hydrate systems quantification can produce physically realistic results. A similar map of shallow free gas was produced through this workflow; this map predicts more sparse free gas accumulations but for the most part these free gas accumulations are collocated with gas hydrate (Figure 7-49).

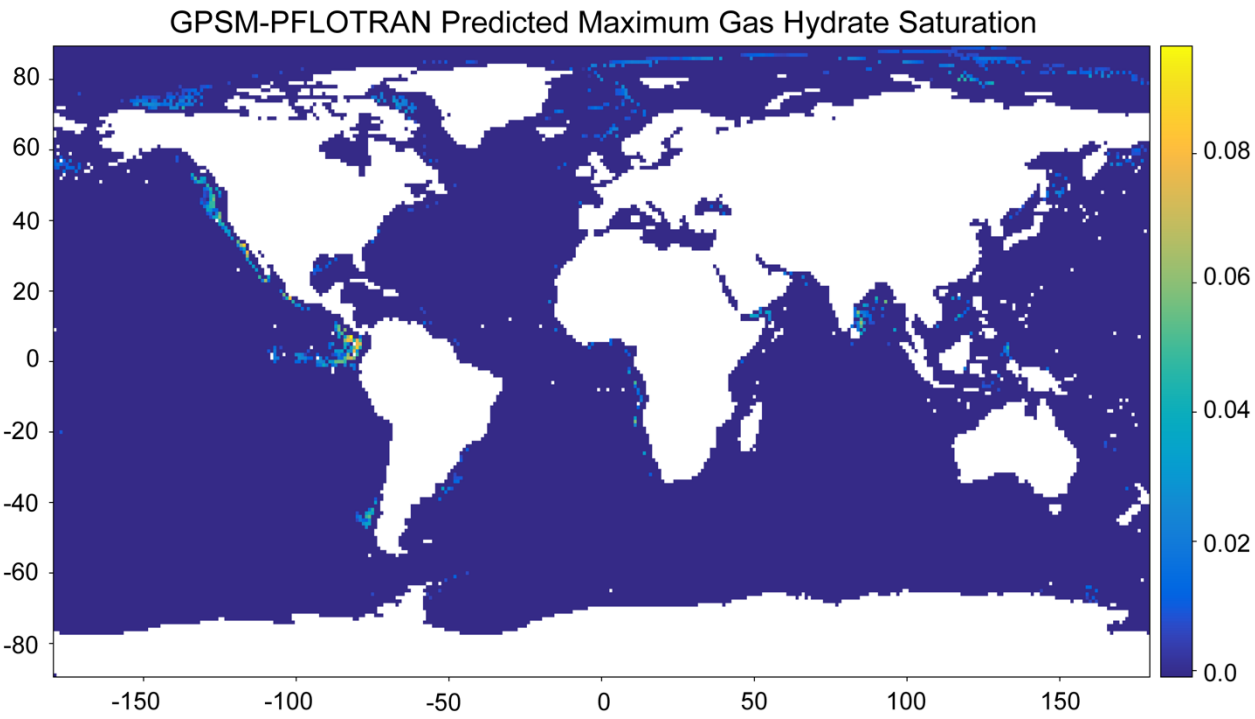


Figure 7-50. Maximum gas hydrate saturation predicted from the workflow.

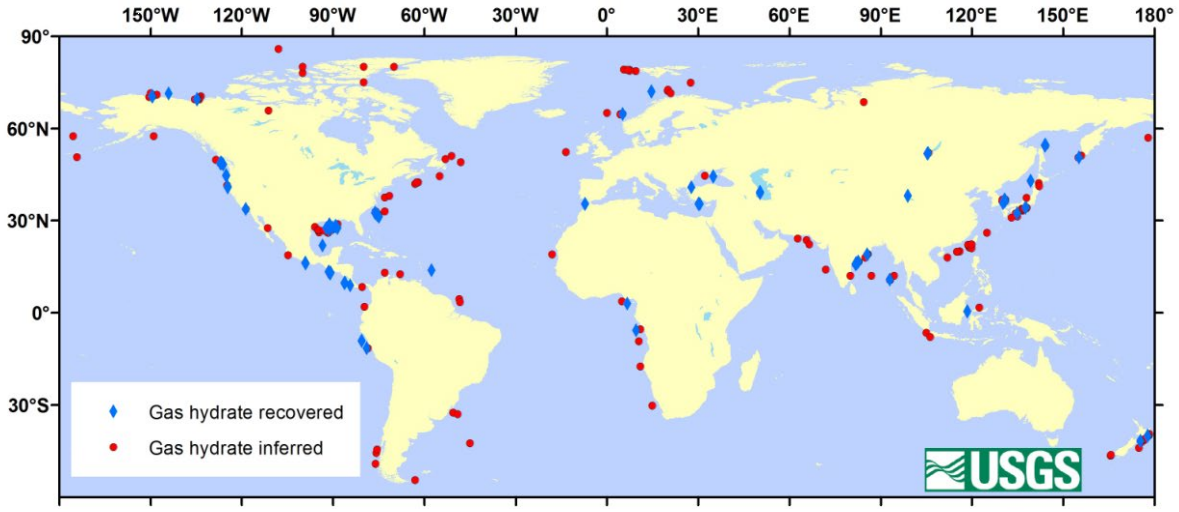


Figure 7-51. Global distribution of inferred and recovered gas hydrate in sediments (adopted from USGS.gov).

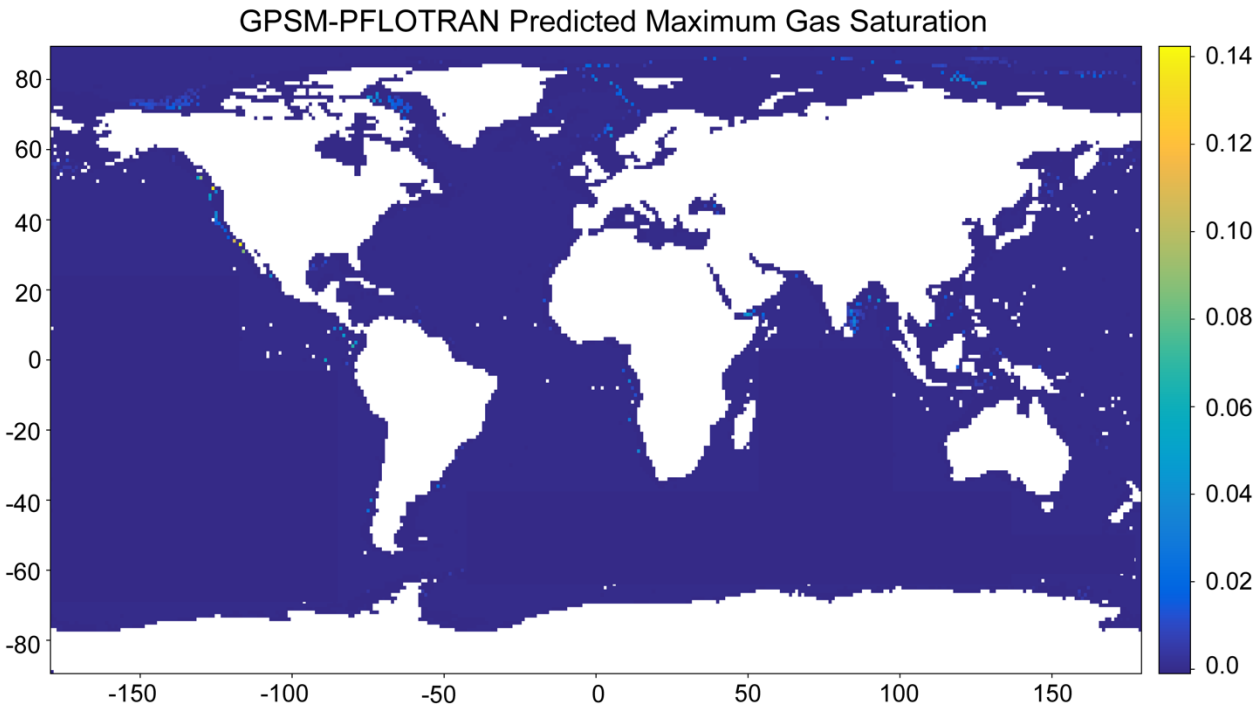


Figure 7-52. Maximum free gas saturation predicted from the workflow

One significant caveat with this work is that the physical conceptual model (steady sedimentation to steady-state, diffusion-dominated environment, standard marine mud lithology/physical properties) was applied in the same way to every location on the map. This is obviously physically unrealistic but it served as a first order approximation and as a launching point for advancing this framework at a global scale. To produce more realistic results, in the future flexibility needs to be added to the framework to heterogeneously apply different conceptual models and different sediment lithologies depending on the region of the globe being simulated.

7.8. Seafloor Acoustic Calculations

From Section 7.1, seafloor heat flux, porosity, total organic carbon, and gas and hydrate saturations with depth were input into the US Naval Research Laboratory's GPSM (GeoPhysMod) program. This program ingests geological and geophysical values at the seafloor and with depth in the subsurface to estimate various resulting geophysical and geoacoustic profiles (e.g. density, P-wave velocities). Here, we demonstrate the effect hydrate and free methane gas has on sediment density and P-wave velocity.

To model porosity vs depth, we used the Martin & Wood (2017) mechanical compaction model where the primary mechanism for porosity reduction is vertical effective stress resulting in grain rearrangement and rotation. This compaction model requires estimates of the total amount of sediment grains that are clay grains, called fraction clay. However, we do not have a predicted estimate of fraction clay for the seafloor. Therefore, we solve for the fraction clay by the rearrangement of Equation 7.3 found within Martin and Wood (2017) using our GPSM GML predicted seafloor porosity value. Using this fraction of clay, we solve for the residual porosity (Equation 7.4) assuming clay fractions are constant with depth.

$$\phi_0 = f_{cly}\phi_{0cly} + (1 - f_{cly}) * \phi_{0snd} \quad \textbf{(Eq. 7.3)}$$

$$\phi_r = f_{cly}\phi_{rcly} + (1 - f_{cly}) * \phi_{rsnd} \quad \textbf{(Eq. 7.4)}$$

In Equation 7.3 and 7.4, $f_{cly}$is the fraction of sediment grains that are clay for a parcel of sediment. For a sediment column containing 100% clay, we assume depositional (ϕ_{0cly}) and residual (ϕ_{rcly}) porosities to be 0.85 and 0.05 respectively. Conversely, for a sediment column containing 100% sand, we assume depositional (ϕ_{0snd}) and residual (ϕ_{rsnd}) porosities to be 0.39 and 0.22 respectively. These porosities of sand and clay are used to calculate the final depositional (ϕ_0) and residual (ϕ_r) porosity for a given grid cell or location.

For each depth (1 meter discretization) and corresponding porosity we calculate the density of the sediment where pore space is occupied by some fraction of hydrate and/or gas. Hydrate is considered to be a pore fill in this model, and therefore porosity will decrease slightly in depth voxels where hydrate saturation increases. The density of the sediment is calculated using Equation 7.5. The density of the pore fill is determined by density of the fluid and includes gas using conventions discussed in Batzle and Wang (1992) and Mavko et al. (1998). Density of the grain material includes hydrate. Grains, excluding hydrate, are calculated from the weighted average of clay and sand components where clay and sand densities are 2.6 and 2.625 g/cm^3 respectively. Fraction clay percentages used here are determined in the same routine as porosity fraction clay values (Equation 7.3; f_{cly}), whereby the remaining components are assumed hydrate or sand.

$$\rho_{sed} = (\rho_{pore} * \phi_z) + \rho_{grain} * (1 - \phi_z) \quad \textbf{(Eq. 7.5)}$$

Finally, to calculate the p-wave velocity (Equation 7.6), the effective moduli (bulk;k and shear;μ) of the sediment are needed. The bulk moduli of the sediment were calculated using the Gassmann equation (Gassmann, 1951). The shear moduli of the sediment were calculated using Helgerud (1999). Both the bulk and shear modulus of the sediment were then implemented into Equation 7.6 to calculate the p-wave velocity.

$$V_p = \sqrt{\frac{k + \frac{4}{3}\mu}{\rho_{sed}}} \quad \textbf{(Eq. 7.6)}$$

From the Blake Ridge region (Section 7.1) 100 ensembles of various porosity, gas, and hydrate saturations with depth were given as input to GeoPhysMod. We ran these 100 ensembles for each location. Two locations from an ensemble were pulled to highlight p-wave velocity and density. These profiles represent locations where 1) there is no hydrate or gas formation (green in Figure Figure 7-50) and 2) where there are reasonably high values for gas and hydrate saturation (blue in Figure 7-50). The blue and green lines on Figure 6.7.1 correspond to latitude and longitude locations 34°37.5'N 75°27.48'W and 32° 57.48' N 75° 2.5' respectively. Heat flux and TOC for the no gas and hydrate formation were 53.631 W/m^3 and 1.4249 percent dry weight (pdw). Heat flux and TOC for

the scenario where gas and hydrate formation occurred was 71.82 W/m^3 and 1.8754 pdw. The maximum hydrate and gas saturations occur at 303 and 586 mbsf, respectively.

Generally, hydrate does not significantly affect the density of the sediment. Hydrate only slightly increases the velocity of the sediment. However, the inclusion of gas can decrease the density of the sediment and the P-wave velocity of the sediment (Figure 7-50). Our 1-D results of the P-wave and density of the sediment with depth are consistent with these findings.

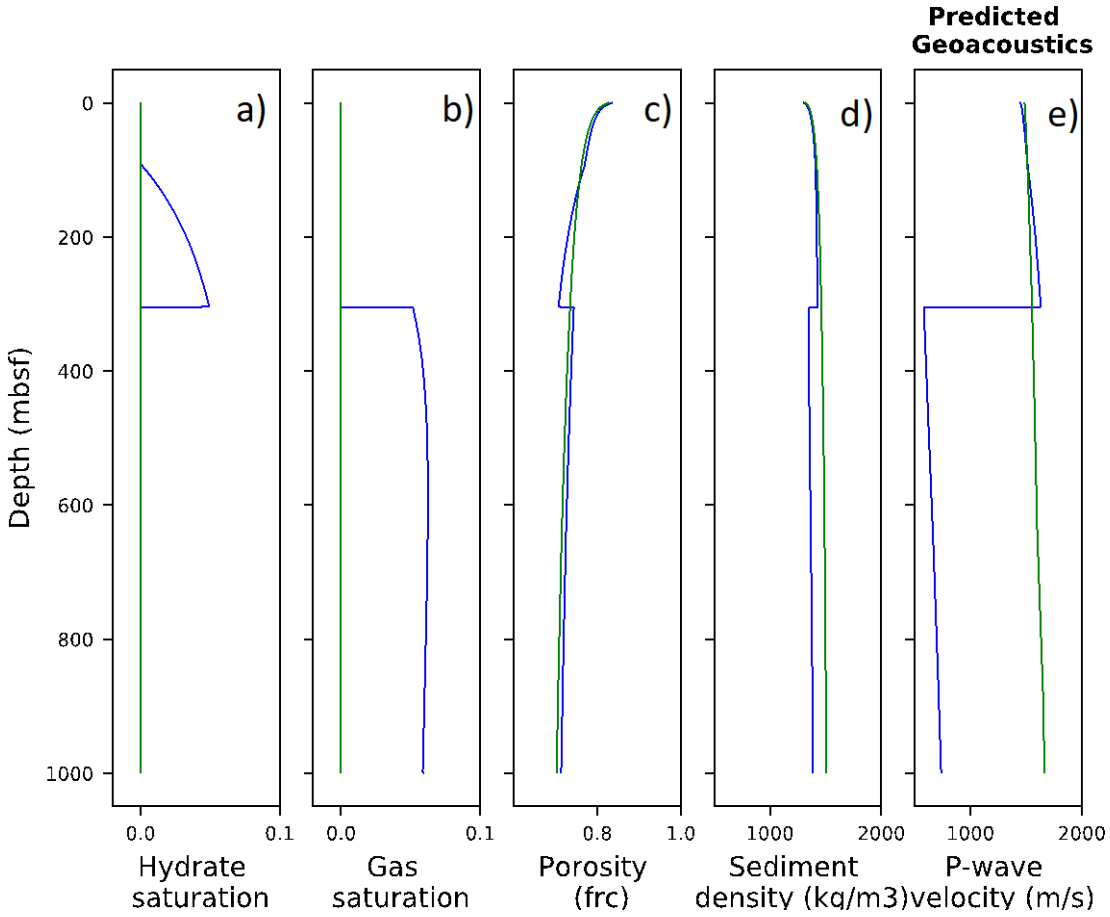


Figure 7-53. Predicted profiles for two locations (34°37.5′N 75°27.48′W and 32° 57.48' N 75° 2.5'). Profiles show depth versus (a) hydrate saturation, (b) gas saturation, (c) porosity, (d) sediment density, and (e) sediment p-wave velocity. The blue lines are for 34°37.5′N 75°27.48′W (1311 mbsl – with gas and hydrate generation), while the green lines are for 32° 57.48' N 75° 2.5' W (4113 mbsl – no gas generation).

In addition, a 2D map of the depth-averaged p-wave velocity from the Blake Ridge region (Section 7.1) was created to show how p-wave velocity of the sediments varies in space (Figure 7-51).

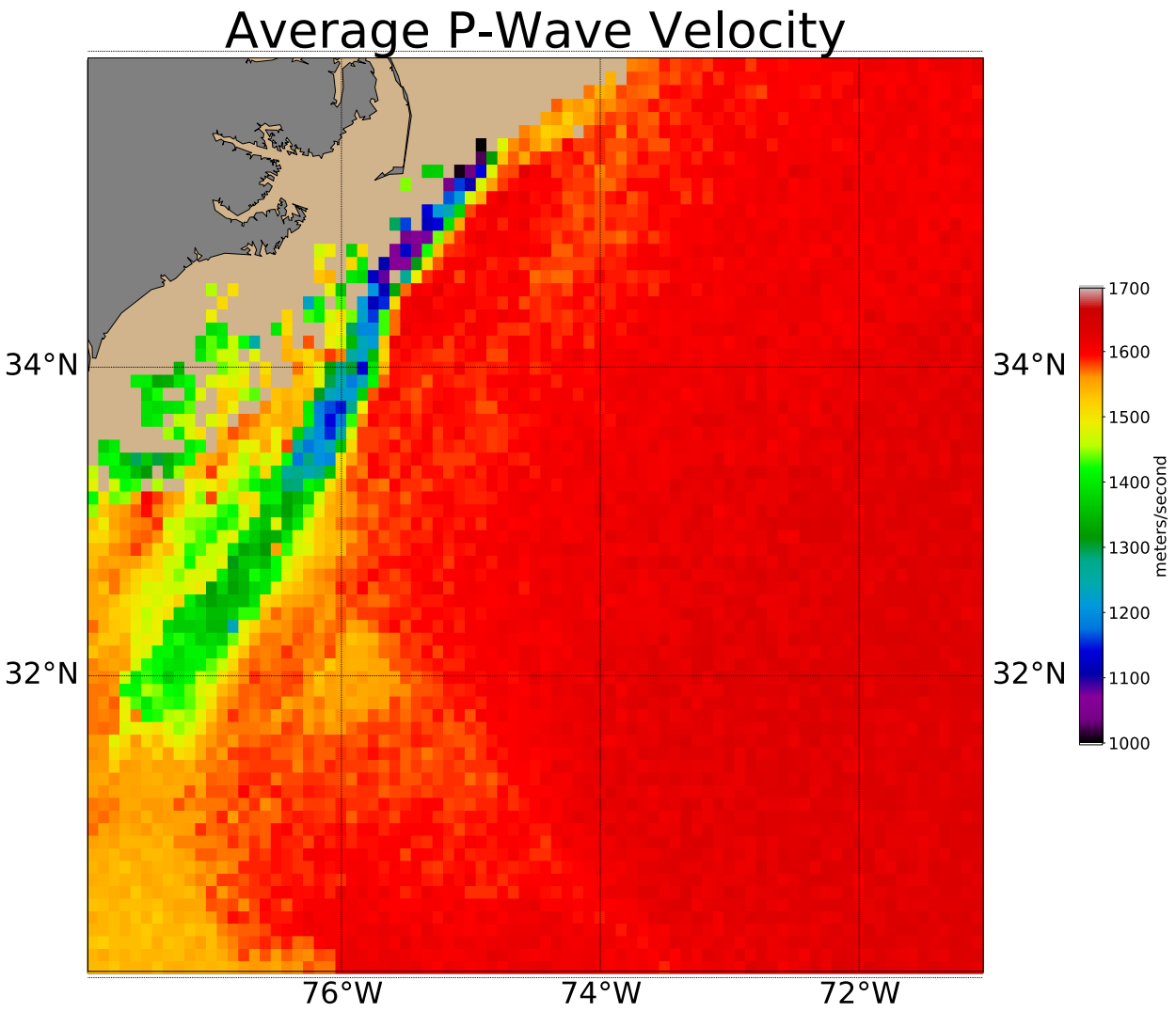


Figure 7-54. Variation of the depth-averaged p-wave velocity calculated at the Blake Ridge gas hydrate province, using GPSM GeoPhysMod, and based on the simulations presented in Section 7.1.

8. K-MEANS CLUSTERING

8.1. Motivation and Conceptual Reasoning

As a novel implementation of the workflow, we have developed the capability to drive PFLOTRAN simulations based on k-means clustering of the GPSM data set. K-means clustering assigns members of a data set into k distinct groups based on similarities in parameter space (Arthur and Vassilvitskii, 2007). To avoid biasing the clustering algorithm based on the range of magnitudes among component variables, the GPSM data for the Greenland study area were normalized by subtracting the mean for the entire study area for each parameter then dividing by the standard deviation. The algorithm was initiated by randomly selecting k sets of normalized values of sedimentation rate, TOC, heat flux, seafloor temperature, and seafloor depth at k locations from the study area which serve as centroids. We used the L_2 norm to calculate the squared Euclidean distance between these centroids and every location's parameter set and assign each location to the centroid for which this distance is minimal to define a cluster. Once all nodes were assigned a value, the centroids for each cluster were re-calculated by taking the average of all members and the clustering algorithm was run again until the total distance for all members did not decrease. The full algorithm was run for 100 iterations to achieve a global minimum for the data.

To benchmark this capability, we clustered the data from the Greenland study area (Figure 8-1) into 48 clusters (k = 48) and run simulations on each of the 48 processors available on a local Linux cluster. The visualization of the clusters can be mapped by color coding the individual regions representing each cluster based on free gas formation and gas hydrate formation. From these centroid outputs, we rescale the normalized data and determine the μ and σ for each parameter that are then used in our Dakota-PFLOTRAN workflow (Table 8-1). Similar to the simulations described in Section 7.4, we sampled 500 times on the sets of μ and σ for each cluster and ran PFLOTRAN for 500 kyr to evaluate the formation rate of free gas and gas hydrate. These formation rates were then used to determine the probability of occurrence for each cluster and mapped back to their corresponding geographic locations.

While the individual simulations of 50 samples described in Section 7.4 take weeks to complete on our Linux cluster, the k-means driven simulations take less than 8 hours to run 500 samples for all centroids. The process therefore enables dimension reduction of the model and elucidates the strongest patterns underlying the data. By verifying the results from the k-means driven simulations against individual simulation outputs, we can make preliminary assessments of the probability for free gas and gas hydrate occurrence in a large study area very efficiently.

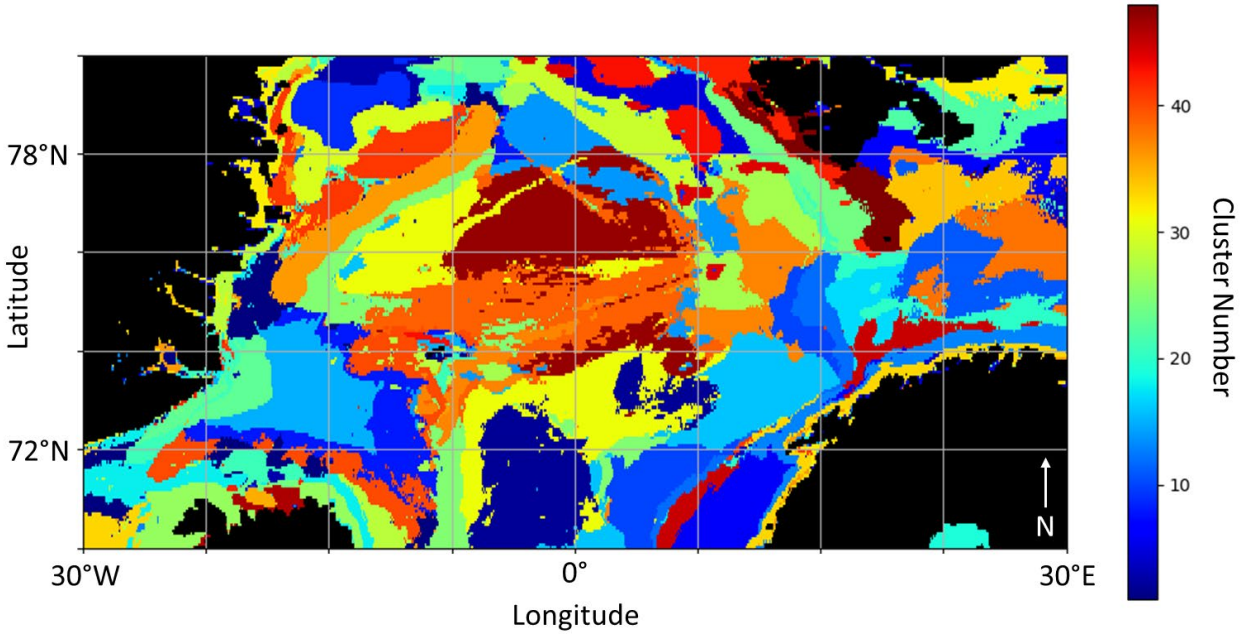


Figure 8-1. Clustering outputs for Greenland-Norwegian study area (48 clusters). Associated GPSM μ and σ values are provided in Table 8-1. Land is shown in black.

Table 8-1. Greenland k-means centroid values of averages and uncertainties. The number of locations assigned to each cluster are indicated in the Members column.

Cluster	Members	Sed Rate μ (mm/yr)	Sed Rate σ (mm/yr)	TOC μ (%)	TOC σ (%)	Heat Flux μ (mW/m^2)	Heat Flux σ (mW/m^2)	Depth μ (mbsl)	Depth σ (mbsl)	Temp μ (°C)	Temp σ (°C)
1	2220	0.0467	0.0260	0.554	0.092	87.91	6.26	-448.8	273.1	-0.10	0.50
2	4328	0.0195	0.0103	0.531	0.072	65.18	3.55	-3373.0	254.1	-0.88	0.02
3	765	0.3285	0.0544	0.499	0.218	90.99	5.59	-168.1	97.3	-0.44	0.46
4	816	0.0400	0.0420	1.257	0.164	65.94	5.25	-230.5	141.4	1.75	0.88
5	1276	0.1476	0.0352	1.113	0.127	110.99	3.55	-2093.1	557.4	-0.84	0.23
6	1729	0.0974	0.0377	0.630	0.142	60.35	1.24	-298.3	109.2	7.04	0.46
7	1587	0.1227	0.0355	1.893	0.177	76.01	7.86	-138.9	80.0	0.41	0.52
8	2870	0.0395	0.0265	0.658	0.099	93.28	5.04	-1594.2	311.1	-0.76	0.22
9	1381	0.4675	0.0438	0.491	0.136	77.80	9.34	-247.4	219.3	-0.35	0.59
10	2359	0.1628	0.0379	0.738	0.091	61.00	3.89	-1499.5	289.2	-0.64	0.36
11	2149	0.2668	0.0370	1.607	0.157	65.71	3.73	-344.6	106.4	2.54	0.71
12	1233	0.1834	0.0422	0.608	0.079	63.70	5.17	-1028.8	435.2	1.77	0.71
13	1504	0.0534	0.0331	0.582	0.158	68.46	7.57	-326.8	331.4	5.02	0.69
14	2885	0.0847	0.0217	0.659	0.082	87.96	4.21	-2854.5	300.7	-0.97	0.05
15	3272	0.0381	0.0305	0.652	0.119	111.88	4.15	-1406.5	314.0	-0.71	0.17
16	3220	0.0648	0.0283	0.661	0.077	59.96	3.04	-2674.2	273.0	-0.68	0.48
17	973	0.3184	0.0401	0.869	0.184	63.54	2.91	-400.2	89.2	3.15	0.87
18	1758	0.0333	0.0220	0.835	0.107	89.33	4.75	-330.0	175.7	0.79	0.68
19	360	0.0035	0.0091	1.388	0.103	42.97	4.41	-34.2	46.6	3.86	0.40
20	1229	0.1875	0.0422	1.114	0.136	71.20	5.72	-319.6	87.1	3.29	0.81
21	1699	0.1598	0.0368	0.694	0.124	88.56	5.95	-478.7	346.2	0.07	0.58
22	1507	0.0544	0.0266	1.766	0.171	59.53	5.35	-132.2	89.3	-0.08	0.56
23	1724	0.1696	0.0434	0.519	0.098	108.25	6.42	-653.8	461.4	-0.23	0.42
24	1541	0.2129	0.0390	0.898	0.168	102.08	4.75	-418.2	304.8	1.57	0.74
25	3131	0.0422	0.0194	0.598	0.094	71.82	4.26	-1813.0	355.6	-0.89	0.09
26	1390	0.0108	0.0203	0.840	0.093	97.84	8.95	-150.2	124.3	3.80	0.79
27	2253	0.1765	0.0293	0.675	0.079	87.45	5.45	-2140.0	343.8	-0.85	0.25
28	1300	0.0341	0.0239	1.006	0.125	78.57	5.40	-316.9	219.3	-0.38	0.46
29	1971	0.0876	0.0368	0.742	0.101	109.86	3.37	-2610.7	339.3	-0.92	0.07
30	1432	0.2866	0.0494	0.677	0.125	71.01	5.83	-216.9	69.7	0.25	0.68
31	6468	0.0217	0.0113	0.661	0.052	70.35	2.91	-3044.8	258.2	-0.93	0.06
32	1584	0.0321	0.0220	1.201	0.128	59.96	5.54	-164.5	97.1	-0.39	0.46
33	1698	0.0196	0.0210	0.618	0.155	68.11	6.09	-396.8	424.7	2.30	0.81
34	1802	0.2455	0.0502	1.541	0.147	83.10	8.09	-114.7	64.8	0.61	0.59
35	110	0.6097	0.1231	0.646	0.033	112.76	5.87	-258.0	97.4	4.53	1.03
36	2524	0.0909	0.0358	0.514	0.093	69.84	4.89	-375.5	256.6	-0.53	0.41
37	3341	0.1250	0.0268	0.669	0.059	76.98	4.45	-2167.4	367.6	-0.89	0.12
38	2092	0.1879	0.0327	1.774	0.186	65.69	3.16	-323.7	112.2	1.81	0.77
39	5291	0.0288	0.0174	0.696	0.061	79.86	2.98	-2538.3	240.2	-0.96	0.07
40	2481	0.0371	0.0259	0.845	0.113	86.79	4.26	-1216.4	311.6	-0.54	0.37
41	2222	0.1334	0.0397	0.790	0.111	64.87	4.48	-262.2	135.6	0.55	0.52
42	977	0.0756	0.0349	1.074	0.121	102.85	4.70	-281.7	238.5	1.33	0.74
43	1112	0.0507	0.0286	1.170	0.113	111.96	3.48	-2756.5	730.1	-0.90	0.02
44	510	0.2858	0.0687	0.754	0.140	102.93	8.38	-2366.0	378.5	-0.82	0.24
45	1303	0.2026	0.0392	0.587	0.125	63.51	6.49	-381.5	195.7	5.01	0.73
46	280	0.1683	0.0954	0.735	0.165	114.96	4.06	-188.4	121.7	6.12	2.23
47	5722	0.0228	0.0146	0.652	0.050	81.96	2.69	-3311.4	257.2	-0.99	0.06
48	1478	0.1191	0.0482	1.449	0.148	100.67	4.33	-137.7	89.0	1.40	0.65

8.2. Verification of Clustering Procedure at Norwegian Sea

We verify this clustering process by calculating the probability of gas and hydrate formation for each cluster in two ways: 1) the number of runs from the k-means driven simulations that resulted in gas/hydrate formation divided by the total number of runs (500) and 2) the number of runs from the individual simulations that resulted in gas/hydrate formation divided by the total number of runs at all member locations ($50 \times N$, where N is the number of member locations for a given cluster). By comparing these two formation rates, we demonstrate that this clustering-based approach performs exceptionally well in capturing the expected likelihood of gas formation and exhibits more complicated behavior with regards to predicting hydrate occurrence.

Table 8-2. Probabilities of free gas and gas hydrate associated with individual and k-means simulations.

Cluster	Gas Formation		Hydrate Formation		Cluster	Gas Formation		Hydrate Formation	
	Individual	Kmeans	Individual	Kmeans		Individual	Kmeans	Individual	Kmeans
1	17.12%	6.52%	0.00%	0.00%	**25**	0.00%	0.00%	0.00%	0.00%
2	0.00%	0.00%	0.00%	0.00%	**26**	19.71%	34.98%	0.00%	0.00%
3	99.22%	87.68%	0.00%	0.00%	**27**	0.00%	0.00%	0.00%	0.13%
4	54.53%	52.78%	0.43%	0.43%	**28**	36.62%	16.36%	0.00%	5.38%
5	0.55%	0.80%	0.00%	0.00%	**29**	0.00%	0.00%	0.00%	0.00%
6	74.49%	51.19%	0.00%	0.00%	**30**	99.44%	98.41%	0.23%	30.17%
7	98.93%	99.49%	1.01%	1.01%	**31**	0.00%	0.00%	0.00%	0.00%
8	0.00%	0.00%	0.00%	0.00%	**32**	66.60%	46.61%	0.41%	10.35%
9	98.91%	67.89%	0.00%	0.00%	**33**	18.79%	4.21%	0.00%	0.00%
10	0.04%	0.00%	0.00%	0.00%	**34**	99.28%	100.00%	0.00%	0.00%
11	99.86%	100.00%	12.64%	12.64%	**35**	100.00%	100.00%	0.00%	0.00%
12	12.00%	4.21%	0.00%	0.00%	**36**	37.16%	14.14%	0.00%	16.72%
13	47.87%	15.37%	0.00%	0.00%	**37**	0.00%	0.00%	0.00%	0.00%
14	0.00%	0.00%	0.00%	0.00%	**38**	99.62%	100.00%	25.06%	57.60%
15	0.18%	0.00%	0.00%	0.00%	**39**	0.00%	0.00%	0.00%	0.00%
16	0.00%	0.00%	0.00%	0.00%	**40**	0.00%	0.20%	0.00%	0.00%
17	96.51%	89.98%	0.85%	0.85%	**41**	88.93%	67.85%	0.22%	53.20%
18	24.74%	15.99%	0.00%	0.00%	**42**	78.30%	58.50%	0.00%	0.20%
19	8.06%	24.14%	0.00%	0.00%	**43**	0.00%	0.20%	0.00%	0.00%
20	98.70%	98.61%	0.46%	0.46%	**44**	0.00%	0.00%	0.00%	6.86%
21	66.39%	32.50%	0.00%	0.00%	**45**	79.66%	44.57%	0.00%	0.08%
22	99.14%	87.61%	3.48%	3.48%	**46**	87.86%	85.53%	0.00%	0.00%
23	49.77%	17.01%	0.00%	0.00%	**47**	0.00%	0.00%	0.00%	0.00%
24	87.99%	67.72%	0.00%	0.00%	**48**	97.83%	97.64%	0.00%	0.07%

The probabilities for free gas and gas hydrate formation in each cluster are shown in Table 8-2. The outputs from the k-means driven simulations show that the study area includes vast regions which are expected to form free gas in the sediment column, with 34 clusters including at least 1 simulation with formation (Table 8-2). Disagreement (i.e., one set of simulations predicts formation whereas the other predicts no formation) exists only in clusters 10, 15, 40, and 43, but note that the maximum gas formation rate for any of these clusters is 0.20%. The agreement between formation

rate predictions indicate a Pearson correlation coefficient of r=0.9563 (p<0.01). By mapping these gas formation rates onto their source location, comparisons can be made between the individual simulation outputs (Figure 8-2) and the k-means simulation outputs (Figure 8-3). The two maps are nearly identical and indicate exceptionally high formation rates in nearshore locations along eastern Greenland, northern Iceland, and between Svalbard and Norway.

While the outputs for hydrate formation are similar, more discrepancy exists between the predictions based on k-means driven simulations and individual simulations and the correlation is not significant (r=0.1971, p=0.1794). Under the individual simulations (Figure 8-4), 20 clusters yielded at least one simulation whereas hydrate formed while 17 clusters yielded hydrate formation based on the k-means driven simulations (Figure 8-5). The maximum formation rate for individual simulations was 42.76% (Cluster 11) and for k-means simulations was 57.60% (Cluster 38). Disagreement exists in 17 of the 48 clusters, suggesting that the greater complexities associated with gas hydrate systems are not as easily captured by the k-means process compared to the free gas system.

These probability maps can be used for quick and efficient prediction of free gas and gas hydrate occurrence without the need for rigorous simulation. The strong agreement between the individual and k-means driven forecasts for free gas formation indicates that this technique is highly effective, but the discrepancy between forecasts for gas hydrate formation based on k-means simulations demands higher scrutiny. We hypothesize that formation of gas hydrate is much more sensitive to location-dependent parameters (especially to heat flux and depth at each site) and by generalizing the parameter space, the nuances of the system are often missed.

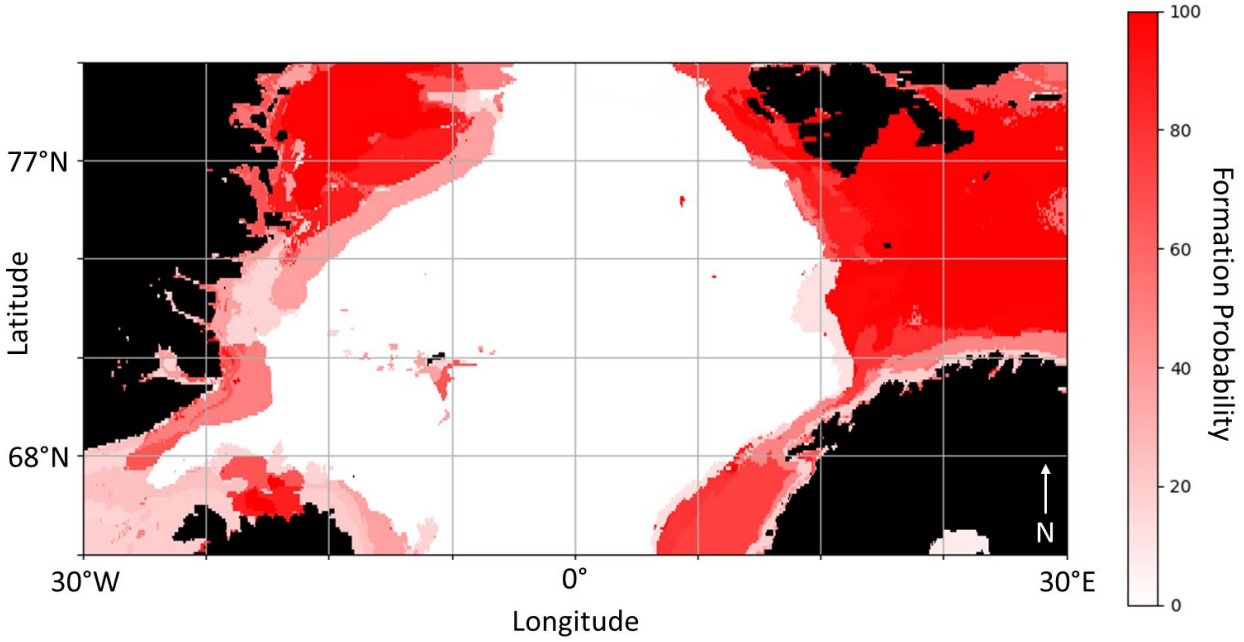


Figure 8-2. Free gas formation probability from individual simulations.

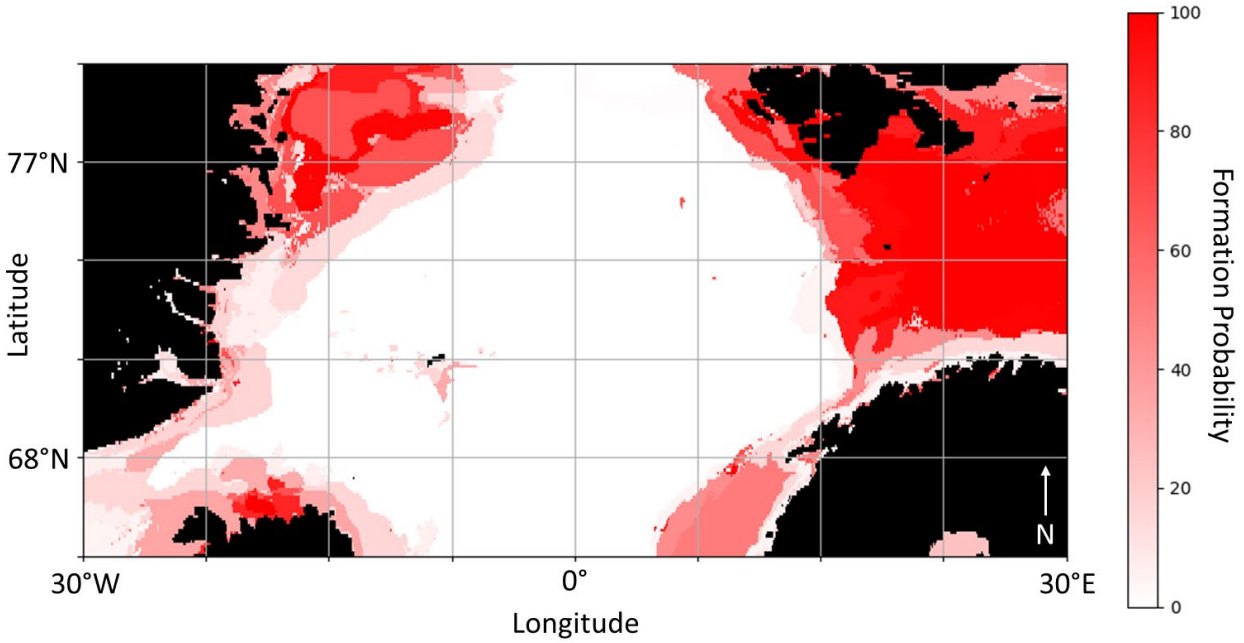


Figure 8-3. Free gas formation probability from k-means simulations.

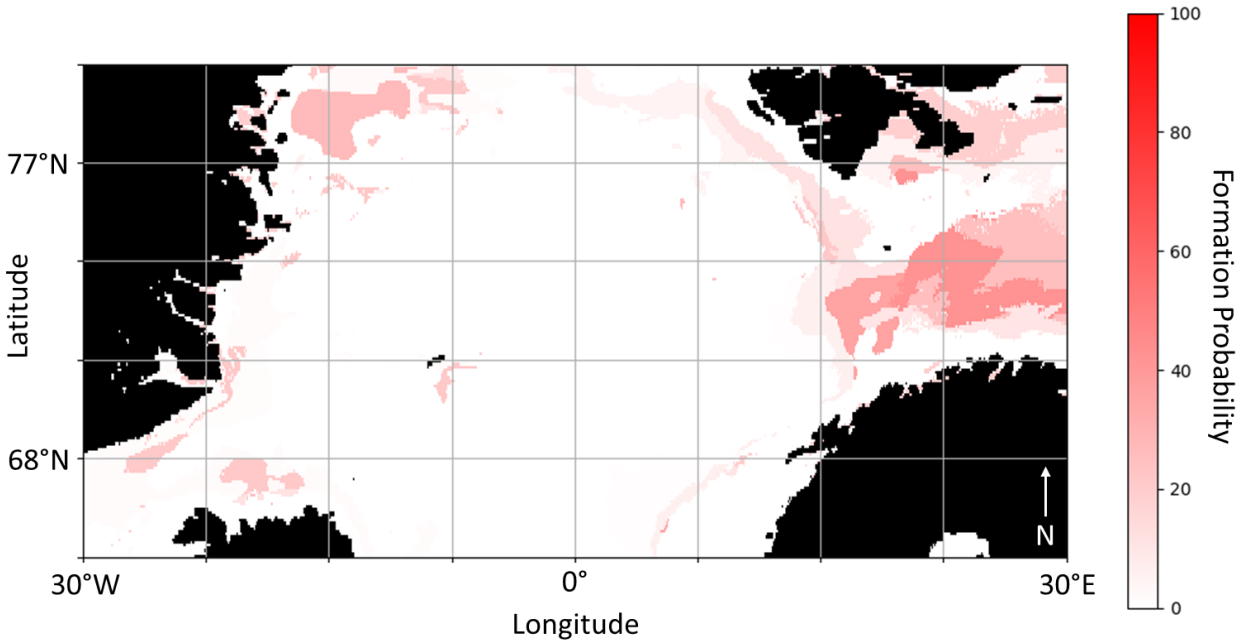


Figure 8-4. Gas hydrate formation probability from individual simulations.

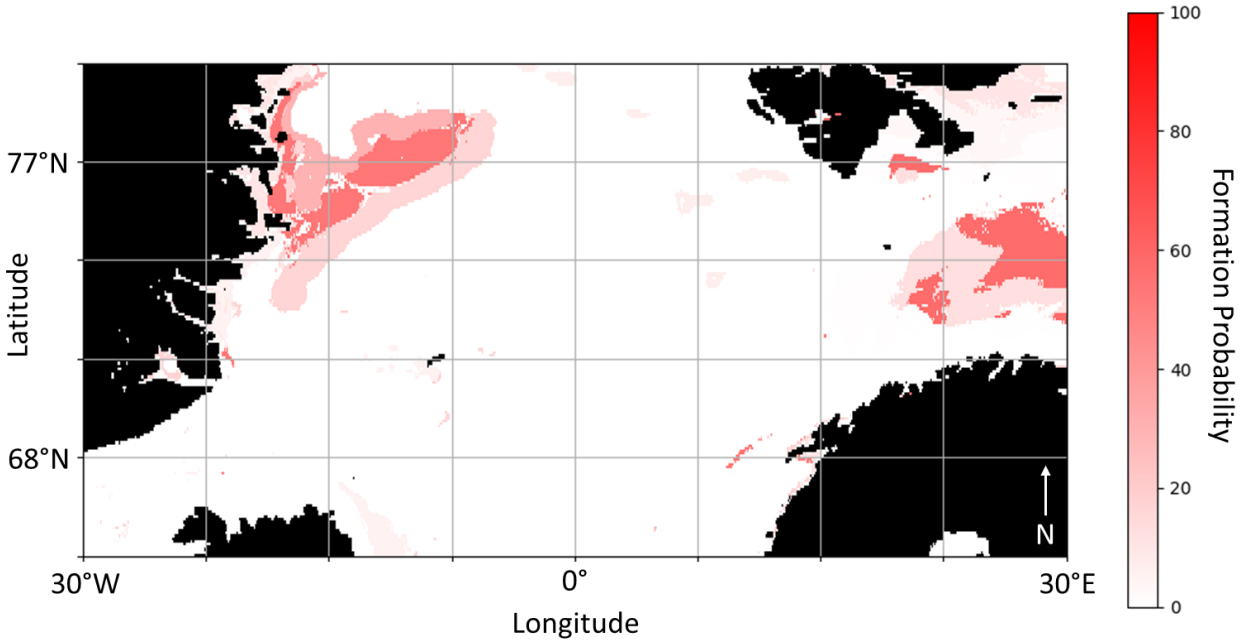


Figure 8-5. Gas hydrate formation probability from k-means simulations.

8.3. Global Clustering Extension and Forecasting

Because the presence of free gas in the sediment column will significantly drop acoustic velocity of these materials even if saturation reaches only 1-5% (Ellis & Singer, 2007; Murphy et al., 1993), corresponding locations with high probability of free gas formation are likely to affect seafloor acoustics and warrant further research. The k-means simulations can quickly indicate geographic areas which would benefit from a more intense simulation strategy while avoiding areas that are unlikely to form gas. In this way, we envision utilizing the strategy on the full GPSM dataset to cluster the entire globe into 48 regions (Figure 8-6) based on their common parameters (Table 8-3).

Predictions made from these clusters can be used in multiple ways. First, we can construct full global predictions based on free gas occurrence. While we have chosen to define occurrence as the presence of free gas anywhere in the sediment columns, we can also be more restrictive where we use the technique to identify areas exhibiting a minimum mass of gas formation or locations where gas formation reaches a target depth. Additionally, we can prioritize new study areas that would benefit from full simulations. Even in the Greenland study area described in Sections 7.4 and 8.2, we found that free gas and hydrate formation was much more prevalent in the northeastern section of the map and plan to fully evaluate this region further without needlessly extending the individual simulations to locations unlikely to form either phase.

Future work will include conducting the global k-means simulations and verifying the results against individual simulations within a subset of clusters. The tandem of k-means driven and individual simulations based on GSPM data allows for probabilistic forecasts of free gas and gas hydrate formation in diverse study areas. Ongoing research on the workflow demonstrates that k-means driven simulation represents an efficient computational technique to predict the occurrence of free gas and may be applicable to other simulation strategies beyond free gas and gas hydrate systems.

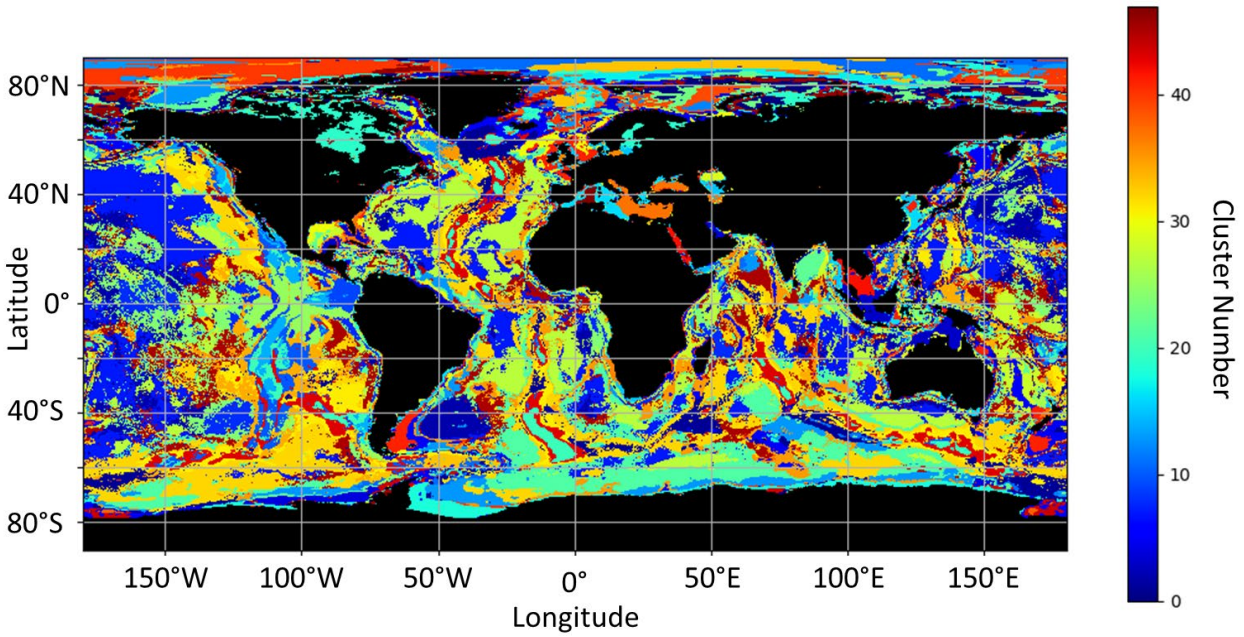


Figure 8-6. Clustering outputs for entire GPSM data set (48 clusters). Associated GPSM μ and σ values are provided in Table 8-3. Land is shown in black.

Table 8-3. Global k-means centroid values of averages and uncertainties.

Cluster	Members	Sed Rate μ (mm/yr)	Sed Rate σ (mm/yr)	TOC μ (%)	TOC σ (%)	Heat Flux μ (mW/m^2)	Heat Flux σ (mW/m^2)	Depth μ (mbsl)	Depth σ (mbsl)	Temp μ (°C)	Temp σ (°C)
1	93470	0.0708	0.0159	1.227	0.271	66.45	10.42	-511.2	674.4	-0.13	1.21
2	194820	0.0151	0.0110	0.393	0.133	78.67	6.40	-2890.2	512.9	1.37	1.06
3	278157	0.0065	0.0071	0.720	0.111	53.05	5.24	-5433.2	503.3	1.25	0.52
4	49548	0.0222	0.0179	0.406	0.264	65.37	13.41	-66.3	111.0	25.20	2.55
5	50191	0.1404	0.0215	0.510	0.246	65.47	11.72	-1206.0	1029.8	0.59	1.70
6	134194	0.0127	0.0098	0.497	0.186	58.73	9.29	-781.3	617.3	2.67	1.17
7	72492	0.0229	0.0148	0.640	0.284	100.36	12.26	-756.9	624.7	1.41	1.97
8	570054	0.0033	0.0028	0.382	0.080	58.47	4.96	-5370.0	355.0	1.37	0.51
9	125908	0.0091	0.0104	1.585	0.197	57.03	9.77	-4376.9	692.8	1.59	0.66
10	19593	0.0274	0.0185	2.749	0.333	135.03	20.15	-2665.8	821.7	2.36	0.81
11	81909	0.0072	0.0087	1.004	0.186	114.48	9.87	-3367.2	525.8	1.50	0.63
12	147986	0.0437	0.0112	0.726	0.155	83.56	6.97	-3911.4	509.0	-0.32	0.84
13	9491	0.3428	0.0704	0.949	0.503	84.48	19.95	-424.2	724.8	0.57	1.36
14	208396	0.0348	0.0095	0.409	0.144	58.84	6.94	-3713.3	532.7	0.45	0.96
15	46201	0.0125	0.0166	1.484	0.301	150.98	14.79	-2957.7	759.8	2.19	1.58
16	49157	0.0243	0.0178	1.595	0.353	67.43	13.88	-424.9	641.3	7.75	2.50
17	47084	0.0346	0.0240	0.701	0.288	78.22	15.52	-418.1	695.4	14.60	2.35
18	51081	0.0885	0.0212	0.930	0.269	119.88	17.50	-3141.7	820.2	-0.38	1.06
19	112161	0.0296	0.0150	0.432	0.157	65.35	8.93	-551.7	473.4	-0.65	0.91
20	90938	0.0111	0.0114	1.589	0.262	44.58	10.91	-207.8	363.2	0.59	1.54
21	80121	0.0126	0.0130	0.527	0.225	142.29	10.04	-2913.9	758.2	1.57	1.32
22	417363	0.0098	0.0066	0.313	0.066	55.46	4.50	-4442.6	401.0	0.45	0.73
23	79338	0.0782	0.0159	0.725	0.260	59.21	11.66	-3079.3	689.3	0.88	1.43
24	46182	0.1315	0.0207	1.353	0.335	69.42	11.83	-428.2	643.4	-0.03	1.52
25	208629	0.0059	0.0066	1.069	0.130	52.25	7.24	-4642.0	549.4	1.48	0.48
26	78013	0.0119	0.0127	1.753	0.226	101.45	13.52	-3560.3	640.3	1.70	0.69
27	56004	0.0294	0.0210	2.212	0.286	62.99	12.53	-2363.8	1274.7	2.77	1.46
28	395172	0.0054	0.0050	0.388	0.089	47.14	4.24	-4912.0	429.1	1.59	0.67
29	131839	0.0204	0.0132	0.893	0.188	53.92	8.36	-2695.9	601.1	2.45	1.06
30	17065	0.0191	0.0132	1.479	0.493	63.06	9.72	-69.7	168.5	21.04	3.50
31	52191	0.0751	0.0195	0.575	0.260	57.65	11.14	-659.6	672.8	4.62	2.28
32	192491	0.0070	0.0063	0.393	0.124	89.52	6.48	-3896.0	469.6	1.38	0.66
33	525427	0.0066	0.0057	0.371	0.094	68.92	4.64	-4409.5	418.5	0.99	0.73
34	92053	0.0313	0.0115	0.636	0.121	106.48	8.64	-3782.3	533.5	-0.34	0.72
35	126379	0.0086	0.0093	1.004	0.184	83.44	8.25	-3766.7	564.5	1.66	0.66
36	229696	0.0087	0.0065	0.388	0.118	55.53	6.01	-3063.9	544.8	1.48	0.94
37	21214	0.0202	0.0177	3.150	0.413	63.78	12.81	-1291.4	1350.5	4.60	3.36
38	33029	0.0309	0.0173	0.640	0.287	39.55	8.34	-1333.0	1178.5	13.16	2.29
39	20036	0.1745	0.0320	0.821	0.392	120.87	18.56	-2087.7	1303.0	-0.04	1.49
40	31354	0.2146	0.0272	1.268	0.448	65.54	11.15	-430.1	503.9	0.42	1.53
41	153439	0.0177	0.0098	0.736	0.136	60.49	6.57	-2295.8	569.6	-0.22	0.52
42	72940	0.0193	0.0145	0.486	0.252	67.23	11.19	-377.9	514.8	7.62	1.67
43	14892	0.0380	0.0317	0.639	0.267	121.09	18.11	-206.4	349.5	24.19	3.14
44	137115	0.0095	0.0089	0.355	0.113	114.02	8.02	-3391.9	529.5	1.52	0.79
45	35660	0.0845	0.0206	0.541	0.230	93.80	14.07	-678.8	615.0	-0.37	1.33
46	335181	0.0073	0.0067	0.644	0.107	59.08	6.23	-4301.8	402.0	1.40	0.62
47	127879	0.0204	0.0124	1.070	0.194	63.55	10.59	-401.2	479.5	0.56	1.45
48	5303	0.2014	0.0547	0.653	0.237	84.60	21.07	-1906.1	1131.6	13.79	3.19

9. SUMMARY OF OUTCOMES AND FUTURE WORK

We have developed a novel software workflow that produces probabilistic maps of geoacoustic and geomechanical sediment properties from an integration of spatially continuous forecasts of seabed characteristics produced with geospatial machine learning algorithms, and sediment physical numerical modeling. We've demonstrated the use of this workflow at several sites, such as the Blake Ridge, North Atlantic margin, the Greenland Norwegian Sea, the Alaskan North Slope (Beaufort Sea), the Gulf of Mexico, as well as globally at lower resolution. The probabilistic maps produced using this method provide the most reliable estimates of free methane gas and gas hydrate distribution over large regions of the seafloor. The ability to make probabilistic predictions is important because understanding the uncertainty of any forecast is usually just as important as the forecast itself. The accuracy of the forecasts can be improved over time through the collection of more observations in regions where there is high uncertainty. Therefore, the probabilistic maps produced using this workflow provide value along two axes: 1) in the predictions themselves, and 2) in their ability to highlight regions where data collection campaigns should be focused to provide the greatest benefit.

Over the course of this work, we have shared the development process and scientific outcomes, resulting in more than 30 publications and presentations (selected works listed in Appendix A): 10 scientific, peer-review journal publications, 16 international and domestic conferences, and 3 invited presentations.

Although this report has focused on the objective of producing probabilistic maps to predict seabed acoustic properties, the workflow can also provide insight in other scientific disciplines. For example, the suite of maps can be used to improve shallow tomographic models of the Arctic basin by constraining shallow structure, thereby improving the capability to use ocean bottom seismometers for nuclear treaty compliance monitoring. The maps of gas hydrate distribution can be used to provide the most reliable resource quantification for natural gas along U.S. continental margins, or globally. Maps of free methane gas can be used to estimate the potential source of greenhouse gas emission from the seabed, into the water column, and potentially into the atmosphere where they can contribute to climate warming. Finally, the entire software workflow is not limited to seabed and marine sediment modeling. The GPSM predictions and PFLOTRAN simulations can easily be extended to terrestrial locations, where the application space grows considerably. For example, such a workflow can be used to produce probabilistic simulations for the performance of geologic nuclear waste repositories, where input uncertainty is often high, and decisions are based on the statistics of certain outcomes.

REFERENCES

[1] Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Umar, A.M., Linus, O.U., Arshad, H., Kazaure, A.A., Gana, U. and Kiru, M.U., 2019. Comprehensive review of artificial neural network applications to pattern recognition. IEEE Access, 7, pp.158820-158846.

[2] Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hooper, R.W., Hough, P.D., Hu, K.T., Jakeman, J.D., Khalil, M., Maupin, K.A., Monschke, J.A., Ridgway, E.M., Rushdi, A.A., Seidl, D.T., Stephens, J.A., Swiler, L.P., and Winokur, J.G., "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.12 User's Manual," Sandia Technical Report SAND2020-12495, November 2020.

[3] Allen, T. W., and C. T. Whitman (2017), Arctic Imperatives: Reinforcing U.S. Strategy on America's Fourth Coast, Independent Task Force Report No. 75, Council on Foreign Relations, New York. www.cfr.org

[4] Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M. and Khan, M.K., 2018. Medical image analysis using convolutional neural networks: a review. Journal of medical systems, 42(11), pp.1-13.

[5] Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (pp. 1027–1035). Society for Industrial and Applied Mathematics.

[6] Batzle, M., & Wang, Z. (1992). Seismic properties of pore fluids. Geophysics, 57(11), 1396-1408.

[7] Bhatnagar, G., Chapman, W. G., Dickens, G. R., Dugan, B., and Hirasaki, G. J. (2007). Generalization of gas hydrate distribution and saturation in marine sediments by scaling of thermodynamic and transport processes. American Journal of Science, 307(6):861–900.

[8] Biastoch, A., Treude, T., Rüpke, L.H., Riebesell, U., Roth, C., Burwicz, E.B., Park, W., Latif, M., Böning, C.W., Madec, G. and Wallmann, K. (2011). Rising Arctic Ocean temperatures cause gas hydrate destabilization and ocean acidification. Geophysical Research Letters, 38(8).

[9] Boswell, R., Rose, K., Collett, T. S., Lee, M., Winters, W., Lewis, K. A., & Agena, W. (2011). Geologic controls on gas hydrate occurrence in the Mount Elbert prospect, Alaska North Slope. Marine and Petroleum Geology, 28(2), 589-607.

[10] Boyer, T. P., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Locarnini, R. A., Mishonov, A. V., Paver, C. R., Reagan, J. R., Seidov, D., Smolyar, I. V., Weathers, K. W., and Zweng, M. M. (2018). WORLD OCEAN DATABASE 2018. page 207.

[11] Boyer TP, Levitus S, Garcia HE, Locamini RA, Stephens C, et al. (2005) Objective analyses of annual, seasonal, and monthly temperature and salinity for the World Ocean on a 0.25° grid. International Journal of Climatology 25: 931–945.

[12] Brothers, D. S., Ruppel, C., Kluesner, J. W., Brink, U. S. t., Chaytor, J. D., Hill, J. C., Andrews, B. D., and Flores, C. (2014). Seabed fluid expulsion along the upper slope and outer shelf of the U.S. Atlantic continental margin. Geophysical Research Letters, 41(1):96–101. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013GL058048.

[13] Brown, H. E., Holbrook, W. S., Hornbach, M. J., & Nealon, J. (2006). Slide structure and role of gas hydrate at the northern boundary of the Storegga Slide, offshore Norway. Marine geology, 229(3-4), 179-186.

[14]Burwicz, E., Reichel, T., Wallmann, K., Rottke, W., Haeckel, M., and Hensen, C. (2017), 3-D basin-scale reconstruction of natural gas hydrate system of the Green Canyon, Gulf of Mexico, Geochem. Geophys. Geosyst., 18, 1959– 1985, doi:10.1002/2017GC006876.

[15]Burwicz, E., & Rüpke, L. (2019). Thermal State of the Blake Ridge Gas Hydrate Stability Zone (GHSZ)—Insights on Gas Hydrate Dynamics from a New Multi-Phase Numerical Model. Energies, 12(17), 3403.

[16]Carroll, J. J., and Mather, A. E. (1997). A Model for the Solubility of Light Hydrocarbons in Water and Aqueous Solutions of Alkanolamines, Chem. Engng. Sci., 52, 545-552.

[17]Claypool G.E., Kaplan I.R. (1974) The Origin and Distribution of Methane in Marine Sediments. In: Kaplan I.R. (eds) Natural Gases in Marine Sediments. Marine Science, vol 3. Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-2757-8_8

[18]Collett, T. S. (2002). Energy resource potential of natural gas hydrates. AAPG bulletin, 86(11), 1971-1992.

[19]Collett, T.S., Lee, M.W., Agena, W.F., Miller, J.J., Lewis, K.A., Zyrianova, M.V., Boswell, R. and Inks, T.L. (2011). Permafrost-associated natural gas hydrate occurrences on the Alaska North Slope. Marine and Petroleum Geology, 28(2), 279-294.

[20]Cramer, S. D. (1984). Solubility of methane in brines from 0 to 300. degree. C. Industrial & Engineering Chemistry Process Design and Development, 23(3), 533-538.

[21]Daigle, H., Cook, A., Fang, Y., Bihani, A., Song, W., and Flemings, P. B. (2020). Gas-Driven Tensile Fracturing in Shallow Marine Sediments. Journal of Geophysical Research: Solid Earth, 125(12).

[22]Davie, M. K., Zatsepina, O. Y., & Buffett, B. A. (2004). Methane solubility in marine hydrate environments. Marine geology, 203(1-2), 177-184.

[23]Dickens, G. R., Paull, C. K., and Wallace, P. (1997). Direct measurement of in situ methane quantities in a large gas-hydrate reservoir. Nature, 385(6615):426–428.

[24]Dickens, G. R., & Quinby-Hunt, M. S. (1997). Methane hydrate stability in pore water: a simple theoretical approach for geophysical applications. Journal of Geophysical Research: Solid Earth, 102(B1), 773-783.

[25]Eckstein, M. (2018), Navy to Release Arctic Strategy This Summer, Will Include Blue Water Arctic Operations, U.S. Naval Institute, USNI News, April 19, 2018, https://news.usni.org/2018/04/19/navy-to-release-arctic-strategy-this-summer-will-include-blue-water-arctic-operations, accessed April 25, 2018.

[26]Egeberg, P. K., & Dickens, G. R. (1999). Thermodynamic and pore water halogen constraints on gas hydrate distribution at ODP Site 997 (Blake Ridge). Chemical Geology, 153(1-4), 53-79.

[27]Ellis, D. V., & Singer, J. M. (2007). Acoustic Waves in Porous Rocks and Boreholes. In Well Logging for Earth Scientists (pp. 499-530): Springer.

[28]Enverus (2021). "Workspace: Production." https://app.drillinginfo.com/production/#/default. Accessed 3-1-2021.

[29]Expedition 308 Scientists (2006): Expedition 308 Scientists (2006). Expedition 308 summary. Proceedings of the Integrated Ocean Drilling Program, 308, 1-70. https://doi.org/10.2204/iodp.proc.308.101.2006

[30]Eymold, W. K., Frederick, J. M., Nole, M., Phrampus, B. J., and Wood, W. T. (2021). Prediction of Gas Hydrate Formation at Blake Ridge Using Machine Learning and Probabilistic Reservoir Simulation. Geochemistry, Geophysics, Geosystems, 22(4).

[31]Fleischer, P., Orsi, T. H., Richardson, M. D., and Anderson, A. L. (2001). Distribution of free 25 gas in marine sediments: a global overview. Geo-Marine Letters, 21(2):103–122.

[32]Frederick, J. M., & Buffett, B. A. (2013). Use of cosmogenic 129I to constrain numerical models of fluid flow in marine sediments: Application to the Blake Ridge Hydrate Province. Geochemistry, Geophysics, Geosystems, 14(5), 1343-1357.

[33]Fuchs, S., Norden, B., and Commission, I. H. F. (2021). The Global Heat Flow Database: Release 2021. GFZ Data Services.

[34]Fukusako, S., and Yamada, M. (1993). Recent Advances in Research on Water-Freezing and Ice-Melting Problems, Experimental Thermal and Fluid Science, 6, 90-105.

[35]Gassmann, F. (1951) Elastic Waves through a Packing of Spheres. Geophysics, 16, 673-685. http://dx.doi.org/10.1190/1.1437718

[36]Graw, J. H., W. T. Wood, B. J. Phrampus (2021), Predicting global marine sediment density using the random forest regressor machine learning algorithm, Journal of Geophysical Research: Solid Earth, 126 (1), e2020JB020135.

[37]Global Heat Flow Compilation Group (2013). Component parts of the World Heat Flow Data Collection. PANGAEA, https://doi.org/10.1594/PANGAEA.810104.

[38]Hammond, G. E., Lichtner, P. C., and Mills, R. T. (2014), Evaluating the performance of parallel subsurface simulators: An illustrative example with PFLOTRAN, Water Resour. Res., 50, 208– 228, doi: 10.1002/2012WR013483.

[39]Handa, Y. P. (1986). Compositions, Enthalpies of Dissociation, and Heat Capacities in the Range 85 to 270K for Clathrate Hydrates of Methane, Ethane, and Propane, and Enthalpy of Dissociation of Isobutane Hydrate, as Determined by a Heat- Flow Calorimeter. J. of Chem. Thermo., 18, 915-921.

[40]Hedges, John I., and Richard G. Keil (1995). Sedimentary organic matter preservation: an assessment and speculative synthesis, Marine Chemistry, Volume 49, Issues 2–3, 1995, pp 81-115, ISSN 0304-4203, https://doi.org/10.1016/0304-4203(95)00008-F.

[41]Helgerud, M. B., Dvorkin, A., Nur, A., Sakai, A., & Collett, T. (1999). Elastic-wave velocity in marine sediments with gas hydrates: Effective medium modeling. Geophysical Research Letters, 26:13, 2021-2024. https://doi.org/10.1029/1999GL900421

[42]Hester, K. C., & Brewer, P. G. (2009). Clathrate hydrates in nature. Annual review of marine science, 1, 303-327.

[43]Hill, J. C., Driscoll, N. W., Weissel, J. K., and Goff, J. A. (2004). Large-scale elongated gas blowouts along the U.S. Atlantic margin. Journal of Geophysical Research: Solid Earth, 109(B9).

[44]Holbrook, W. S., Hoskins, H., Wood, W. T., Stephen, R. A., Lizarralde, D., and Leg 164 Science Party (1996). Methane Hydrate and Free Gas on the Blake Ridge from Vertical Seismic Profiling. Science, 273(5283):1840–1843.

[45]International Association for the Properties of Water and Steam (IAPWS), Revised Release on the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam, Lucerne, Switzerland, 2007.

[46]Judd, A. G. (2003). The global importance and context of methane escape from the seabed. Geo-Marine Letters, 23(3):147–154.

[47]Judd, A.G. (2004), Natural seabed gas seeps as sources of atmospheric methane. Env Geol 46, 988–996. https://doi.org/10.1007/s00254-004-1083-3

[48]Kamath,V.A. (1984). Study of heat transfer characteristics during dissociation of gas hydrates in porous media. Ph.D. dissertation, Univ. of Pittsburgh, Pittsburgh, PA.

[49]Knies, J., & Mann, U. (2002). Depositional environment and source rock potential of Miocene strata from the central Fram Strait: introduction of a new computing tool for simulating organic facies variations. Marine and Petroleum Geology, 19(7), 811-828.

[50]Kominz, M. A., Patterson, K., and Odette, D. (2011). Lithology Dependence of Porosity In Slope and Deep Marine Sediments. Journal of Sedimentary Research, 81(10):730–742.

[51]Lee, T. R., Wood, W. T., and Phrampus, B. J. (2019). A Machine Learning (kNN) Approach to Predicting Global Seafloor Total Organic Carbon. Global Biogeochemical Cycles, 33(1):37–46.

[52]Lee, T. R., B. J. Phrampus, J. Obelcz, W. T. Wood, A. Skarke (2020), Global marine isochore estimates using machine learning, Geophysical Research Letters, 47 (18), e2020GL088726.

[53]Ling, J.; Kurzawski, A.; Templeton, J. 2016. Reynolds aver-aged turbulence modelling using deep neural networks with embedded invariance. Journal of Fluid Mechanics 807: 155-166.

[54]Malinverno, A. (2010). Marine gas hydrates in thin sand layers that soak up microbial methane. Earth and Planetary Science Letters, 292(3):399–408.

[55]Martin, K. M., Wood, W. T., & Becker, J. J. (2015). A global prediction of seafloor sediment porosity using machine learning. Geophysical Research Letters, 42(24).

[56]Martin, K., & Wood, W. (2017). A new model of marine sediment compression. Earth and Planetary Science Letters, 477, 21–26. https://doi.org/10.1016/j.epsl.2017.08.008

[57]Mau, S., Römer, M., Torres, M.E., Bussmann, I., Pape, T., Damm, E., Geprägs, P., Wintersteller, P., Hsu, C.W., Loher, M. and Bohrmann, G. (2017). Widespread methane seepage along the continental margin off Svalbard-from Bjørnøya to Kongsfjorden. Scientific reports, 7(1), 1-13.

[58]Mavko, G., Mukerji, T., & Dvorkin, J. (1998). The rock physics handbook. Cambridge University Press. Cambridge, UK https://doi.org/10.1017/CBO9780511626753

[59]McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2), 239-245.

[60]Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M.M.C. and Ottersen, G. (2019). Polar Regions. Chapter 3, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate.

[61]Minshull, T.A., Marín-Moreno, H., Betlem, P., Bialas, J., Buenz, S., Burwicz, E., Cameselle, A.L., Cifci, G., Giustiniani, M., Hillman, J.I. and Hölz, S. (2020). Hydrate occurrence in Europe: A review of available evidence. Marine and Petroleum Geology, 111, 735-764.

[62]Moridis, G. J. (2003). Numerical studies of gas production from methane hydrates. Spe Journal, 8(04), 359-370.

[63]Moridis, G. J. (2014). User's manual for the hydrate v1. 5 option of TOUGH+ v1. 5: A code for the simulation of system behavior in hydrate-bearing geologic media (No. LBNL-6869E). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

[64]Murphy, W., Reischer, A., & Hsu, K. (1993). Modulus decomposition of compressional and shear velocities in sand bodies. Geophysics, 58(2), 227-239.

[65]Obelcz, J., W. T. Wood, B. J. Phrampus, T. R. Lee (2020), Machine learning augmented time-lapse bathymetric surveys: A case study from the Mississippi river delta front, Geophysical Research Letters, 47 (10), e2020GL087857.

[66]Phrampus, B. J., & Hornbach, M. J. (2012). Recent changes to the Gulf Stream causing widespread gas hydrate destabilization. Nature, 490(7421), 527-530.

[67]Phrampus, B. J., Lee, T. R., and Wood, W. T. (2020). Predictor Grids for "A Global Probabilistic Prediction of Cold Seeps and Associated Seafloor Fluid Expulsion Anomalies (SEAFLEAs)". Type: dataset.

[68]Phrampus, B. J., T. R. Lee, W. T. Wood (2020), A global probabilistic prediction of cold seeps and associated SEAfloor FLuid Expulsion Anomalies (SEAFLEAs), Geochemistry, Geophysics, Geosystems, 21 (1), e2019GC008747.

[69]Prouty, N. G., Sahy, D., Ruppel, C. D., Roark, E. B., Condon, D., Brooke, S., Ross, S. W., and Demopoulos, A. W. J. (2016). Insights into methane dynamics from analysis of authigenic carbonates and chemosynthetic mussels at newly-discovered Atlantic Margin seeps. Earth and Planetary Science Letters, 449:332–344.

[70]Restreppo, G. A., Wood, W. T., and Phrampus, B. J. (2020). Oceanic sediment accumulation rates predicted via machine learning algorithm: towards sediment characterization on a global scale. Geo-Marine Letters, 40(5):755–763.

[71]Restreppo, G. A., W. T. Wood, J. H. Graw, B. J. Phrampus (2021), A machine-learning derived model of seafloor sediment accumulation, Marine Geology, 106577.

[72]Ruppel, C. D., & Kessler, J. D. (2017). The interaction of climate change and methane hydrates. Reviews of Geophysics, 55(1), 126-168.

[73]Ruppel, C., Von Herzen, R., & Bonneville, A. (1995). Heat flux through an old (~ 175 Ma) passive margin: Offshore southeastern United States. Journal of Geophysical Research: Solid Earth, 100(B10), 20037-20057.

[74]Ryan, W. B. F., Carbotte, S. M., Coplan, J. O., O'Hara, S., Melkonian, A., Arko, R., Weissel, R. A., Ferrini, V., Goodwillie, A., Nitsche, F., Bonczkowski, J., and Zemsky, R. (2009). Global Multi-Resolution Topography synthesis. Geochemistry, Geophysics, Geosystems, 10(3).

[75]Ryan et al. (2009): https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2008GC002332

[76]Sayedi, S.S., Abbott, B.W., Thornton, B.F., Frederick, J.M., Vonk, J.E., Overduin, P., Schädel, C., Schuur, E.A., Bourbonnais, A., Demidov, N. and Gavrilov, A. (2020). Subsea permafrost carbon stocks and climate change sensitivity estimated by expert assessment. Environmental Research Letters, 15(12), 124075.

[77]Scalfi, Laura, Benoit Coasne, and Benjamin Rotenberg. (2021), On the Gibbs–Thomson equation for the crystallization of confined fluids. The Journal of Chemical Physics 154, no. 11 (2021): 114711.

[78]Seiter, K., Hensen, C., Schröter, J., and Zabel, M. (2004). Organic carbon content in surface sediments—defining regional provinces. Deep Sea Research Part I: Oceanographic Research Papers, 51(12):2001–2026.

[79]Shipboard Scientific Party (1994a). Site 902. Proceedings of the Ocean Drilling Program: Initial Reports, 150:63–127.

[80]Shipboard Scientific Party (1994b). Site 903. Proceedings of the Ocean Drilling Program: Initial Reports, 150:129–205.

[81]Shipboard Scientific Party (1994c). Site 904. Proceedings of the Ocean Drilling Program: Initial Reports, 150:129–205.

[82]Shipboard Scientific Party (1994d). Site 906. Proceedings of the Ocean Drilling Program: Initial Reports, 150:309–357.

[83]Shipboard Scientific Party (1996a). Site 995. Proceedings of the Ocean Drilling Program: Initial Reports, 164:175–240.

[84]Shipboard Scientific Party (1996b). Site 997. Proceedings of the Ocean Drilling Program: Initial Reports, 164:277–334.

[85]Shipboard Scientific Party (1998a). Site 1073. Proceedings of the Ocean Drilling Program: Initial Reports, 174A:153–191.

[86]Skarke, A., Ruppel, C., Kodis, M., Brothers, D., and Lobecker, E. (2014). Widespread methane leakage from the sea floor on the northern US Atlantic margin. Nature Geoscience, 7(9):657–661.

[87]Somerton, W. H., Keese, J. A., & Chu, S. L. (1974). Thermal behavior of unconsolidated oil sands. Society of Petroleum Engineers Journal, 14(05), 513-521.

[88]Sparrow, K.J., Kessler, J.D., Southon, J.R., Garcia-Tigreros, F., Schreiner, K.M., Ruppel, C.D., Miller, J.B., Lehman, S.J. and Xu, X. (2018). Limited contribution of ancient methane to surface waters of the US Beaufort Sea shelf. Science advances, 4(1), eaao4842.

[89]Span, R., & Wagner, W. (1996). A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa. Journal of physical and chemical reference data, 25(6), 1509-1596.

[90]Straume, E., Gaina, C., Medvedev, S., Hochmuth, K., Gohl, K., Whittaker, J. M., Abdul Fattah, R., Doornenbal, J. C., & Hopper, J. R. (2019). GlobSed: Updated total sediment thickness in the world's oceans. Geochemistry, Geophysics, Geosystems, 20(4), 1756-1772.

[91]Treude, T., Krause, S., Steinle, L., Burwicz, E., Hamdan, L.J., Niemann, H., Feseker, T., Liebetrau, V., Krastel, S. and Berndt, C. (2020). Biogeochemical consequences of nonvertical methane transport in sediment offshore northwestern Svalbard. Journal of Geophysical Research: Biogeosciences, 125(3), e2019JG005371.

[92]U.S., Resource Evaluation Division, and Matthew Frye. Preliminary Evaluation of In-Place Gas Hydrate Resources: Gulf of Mexico Outer Continental Shelf, 2008, pp. 1–136.

[93]Veloso-Alarcón, M.E., Jansson, P., De Batist, M., Minshull, T.A., Westbrook, G.K., Pälike, H., Bünz, S., Wright, I. and Greinert, J. (2019). Variability of acoustically evidenced methane bubble emissions offshore western Svalbard. Geophysical Research Letters, 46(15), 9072-9081.

[94]Wehrmann, L. M., Risgaard-Petersen, N., Schrum, H. N., Walsh, E. A., Huh, Y., Ikehara, M., D'Hondt, S., Ferdelman, T. G., Ravelo, A. C., & Takahashi, K. (2011). Coupled organic and inorganic carbon cycling in the deep subseafloor sediment of the northeastern Bering Sea Slope (IODP Exp. 323). Chemical Geology, 284(3-4), 251-261.

[95]Westbrook, G.K., Thatcher, K.E., Rohling, E.J., Piotrowski, A.M., Pälike, H., Osborne, A.H., Nisbet, E.G., Minshull, T.A., Lanoisellé, M., James, R.H. and Hühnerbach, V. (2009). Escape

of methane gas from the seabed along the West Spitsbergen continental margin. Geophysical Research Letters, 36(15).

[96]White, M. D., Kneafsey, T. J., Seol, Y., Waite, W. F., Uchida, S., Lin, J. S., ... & Zyrianova, M. (2020). An international code comparison study on coupled thermal, hydrologic and geomechanical processes of natural gas hydrate-bearing sediments. Marine and Petroleum Geology, 104566.

[97]Wood, W. T., & Ruppel, C. (2000). 26. SEISMIC AND THERMAL INVESTIGATIONS OF THE BLAKE RIDGE GAS HYDRATE AREA: A SYNTHESIS1. Proceedings of the Ocean Drilling Program. Scientific results. Ocean Drilling Program, 253-264.

[98]Xu, W. and Ruppel, C. (1999). Predicting the occurrence, distribution, and evolution of methane gas hydrate in porous marine sediments. Journal of Geophysical Research: Solid Earth, 104(B3):5081–5095.

[99]You, K., Flemings, P. B., Malinverno, A., Collett, T., & Darnell, K. (2019). Mechanisms of Methane Hydrate Formation in Geological Systems. Reviews of Geophysics.

APPENDIX A.

Several journal publications (10), conference abstracts (16), and invited presentations (3) have been published/given as a result of this project's outreach efforts. The selected citations are listed here.

[1] Daigle, H., M. Broten, B. J. Phrampus, W. T. Wood, M. Nole, J. M. Frederick, and A. S. Lapari (**2020**), Predicting Gas and Hydrate Presence on the U.S. Atlantic Margin Using Geospatial Machine Learning, Abstract OS029-0001, presented at the 2020 Fall Meeting, AGU, (virtual), 1-17 December.

[2] Eymold, W. K., Frederick, J. M., Nole, M., Phrampus, B. J., and Wood, W. T. (**2021**). Prediction of Gas Hydrate Formation at Blake Ridge Using Machine Learning and Probabilistic Reservoir Simulation. Geochemistry, Geophysics, Geosystems, 22(4).

[3] Eymold, W.K. (**2020**). Probabilistic Forecasts of the Occurrence of Methane Hydrate and Free Gas Near Blake Ridge. Oral presentation at the 14th Annual 2020 Postdoctoral Technical Showcase, Albuquerque, NM.

[4] Eymold, W. K., J. M. Frederick, M. Nole, B. J. Phrampus, and W. Wood (**2020**), Probabilistic Predictions of Gas Hydrate Formation near Blake Ridge using Dakota and PFLOTRAN, Abstract OS029-0002, presented at the 2020 Fall Meeting, AGU, (virtual), 1-17 December.

[5] Frederick, J. M. (**2021**), The Fate of Submerged Permafrost and its Sequestered Carbon: Marine Sources of Greenhouse Gases. Invited seminar speaker, Department of Energy Fugitive Emissions and Clean Air Working Groups, 25 May.

[6] Frederick, J. M. (**2021**), Mapping seafloor acoustic properties in the Arctic and beyond: The role of evolving thermodynamics and free gas distribution, U.S. Navy Undersea Warfighting Development Center, 29 June.

[7] Frederick, J. M., W. K. Eymold, M. Nole, B. J. Phrampus, W. T. Wood, and H. Daigle (**2020**), Probabilistic Predictions of Offshore Gas Hydrate and Submarine Permafrost Distribution Along the Alaskan North Slope, Abstract OS034-08, presented at the 2020 Fall Meeting, AGU, (virtual), 1-17 December.

[8] Frederick, J. M., W. K. Eymold, M. Nole, W. T. Wood, B. J. Phrampus, and H. Daigle (**2020**), A Probabilistic Modeling Framework for Forecasting Gas Hydrate Using Geospatial Machine Learning, Presented at the Gordon Research Conference on Natural Gas Hydrate Systems, Galveston, Texas, 24-28 February.

[9] Frederick, J. M., W. K. Eymold, W. T. Wood, M. Nole, B. J. Phrampus, H. Daigle, H. Yoon, and B. Young (**2019**), A Probabilistic Modeling Framework Utilizing Ocean Drilling Data to Forecast Current and Future Thermodynamic Conditions on and Below the Seafloor, Abstract PP11D-1421, presented at the 2019 Fall Meeting, AGU, San Francisco CA, 9-13 December.

[10] Frederick, J. M., W. K. Eymold, W. T. Wood, M. Nole, B. J. Phrampus, H. Daigle, H. Yoon, B. Young, and K. Sale (**2019**), Forecasting Marine Sediment Properties On and Near the Arctic Shelf with Geospatial Machine Learning, presented at the 2019 Symposium on Arctic Maritime Domain Awareness, National Geospatial-Intelligence Agency, Springfield VA, 21-22 November.

[11]Graw, J. H., Wood, W. T., & Phrampus, B. J. (**2021**). Predicting global marine sediment density using the random forest regressor machine learning algorithm. Journal of Geophysical Research: Solid Earth, 126, e2020JB020135. https://doi.org/10.1029/2020JB020135

[12]Graw, Jordan H., Warren T. Wood, Benjamin J. Phrampus, Taylor R. Lee, and Jeffrey Obelcz, (**2020**). Predicting Global Marine Sediment Density Using Machine Learning, presented at the Seismological Society of America Annual meeting, Albuquerque, NM, Apr.27-30.

[13]Lee, T. R., Phrampus, B. J., Obelcz, J., Wood, W. T., & Skarke, A. (**2020**). Global marine isochore estimates using machine learning. Geophysical Research Letters, 47, e2020GL088726. https://doi.org/10.1029/2020GL088726

[14]Lee, Taylor, Benjamin Phrampus, Warren Wood, Adam Skarke, (**2020**). A new estimate of carbon sequestered within the global methane hydrate stability zone using machine-learning predicted inputs, presented at the American Geophysical Union Fall Meeting, San Francisco, CA 1-17 December.

[15]Lee, Taylor Lee, Warren Wood, Adam Skarke, Benjamin Phrampus, Jeff Obelcz, (**2019**). Global machine learning predictions of sediment unit thickness for present to middle Miocene using Deep Sea Drilling Project data, presented at the American Geophysical Union Fall Meeting, San Francisco, CA 9-13 December.

[16]Nole, M., W. K. Eymold, B. J. Phrampus, W. T. Wood, Frederick, J. M., and T. Lee (**2020**), A Hybrid Machine Learning and Reservoir Simulation Approach to Estimating Global Gas Hydrate Occurrence Using Open-Source Software, Presented at the Gordon Research Conference on Natural Gas Hydrate Systems, Galveston, Texas, 24-28 February.

[17]Nole, M., H. Daigle, J. M. Frederick (**2019**), An Open-Source Framework For Probabilistic Gas Hydrate Systems Modeling, Abstract OS41B-1665, presented at the 2019 Fall Meeting, AGU, San Francisco CA, 9-13 December.

[18]Obelcz, Jeffrey, Warren Wood, Benjamin Phrampus, Derek Sawyer, Taylor Lee, (**2020**). Global analysis of relationship between machine learning predicted underwater landslide scar fraction and sediment shear strength profiles, presented at the American Geophysical Union Fall Meeting, San Francisco, CA 1-17 December.

[19]Phrampus, B. J., Lee, T. R., & Wood, W.T. (**2020**). A global probabilistic prediction of cold seeps and associated SEAfloor FLuid Expulsion Anomalies (SEAFLEAs). Geochemistry, Geophysics, Geosystems, 21, e2019GC008747. https://doi.org/10.1029/2019GC008747

[20]Phrampus, Benjamin J. and Warren T. Wood, (**2020**). Stochastic and Machine Learning Predictions of Bathymetry, presented at the American Geophysical Union Fall Meeting, San Francisco, CA 1-17 December.

[21]Phrampus, B. J., T. Lee, and W. T. Wood (**2019**), Compiling a census for SEAFLEAs, Eos, 100, https://doi.org/10.1029/2019EO119605

[22]Phrampus, Benjamin J., Warren T. Wood, (**2019**). Utilizing Machine and Deep Learning to Predict Bathymetry, presented at the American Geophysical Union Meeting, San Francisco, California, 9-13 December.

[23]Restreppo, Giancarlo A., Warren T. Wood, Jordan H. Graw, Benjamin J. Phrampus. (**2021**). A machine-learning derived model of seafloor sediment accumulation, Marine Geology, Volume 440, 106577, https://doi.org/10.1016/j.margeo.2021.106577

[24]Restreppo, G.A., Wood, W.T. & Phrampus, B.J. (**2020**). Oceanic sediment accumulation rates predicted via machine learning algorithm: towards sediment characterization on a global scale. Geo-Mar Lett, https://doi.org/10.1007/s00367-020-00669-1

[25]Restreppo, Giancarlo A., Warren T. Wood, Jordan H. Graw, Benjamin J. Phrampus, (**2020**). Comparing a Machine-Learning Based Model of Benthic, Marine Mass Accumulation with Fluvial Sediment Flux into the Global Oceans, presented at the American Geophysical Union Fall Meeting, San Francisco, CA 1-17 December.

[26]Sayedi, S. S., B. Abbott, B. Thornton, J. M. Frederick, J. Vonk, P. Overduin, C. Schadel, E. Schuur, A. Bourbonnais, A. Gavrilov, S. He, G. Hugelius, M. Jakobsson, M. Jones, D. Joung, G. Kraev, R. Macdonald, A. D. McGuire, C. Mu, M. O'Regan, K. Schreiner, C. Stranne, E. Pizhankova, A. Vasiliev, S. Westermann, J. P. Zarnetske, T. Zhang, M. Ghandehari, S. Baeumler, B. Brown, R. Frei, and A. Maslakov (**2020**), Subsea permafrost carbon stocks and climate change sensitivity estimated by expert assessment, Environmental Research Letters, Vol. 15, No. 12, doi:10.1088/1748-9326/abcc29.

[27]Wei, L., Cook, A., Daigle, H., Malinverno, A., Nole, M., & You, K. (**2019**). Factors controlling short-range methane migration of gas hydrate accumulations in thin coarse-grained layers. Geochemistry, Geophysics, Geosystems, 20(8), 3985-4000.

[28]White, M. D., Kneafsey, T. J., Seol, Y., Waite, W. F., Uchida, S., Lin, J. S., ... & Zyrianova, M. (**2020**). An international code comparison study on coupled thermal, hydrologic and geomechanical processes of natural gas hydrate-bearing sediments. Marine and Petroleum Geology, 104566.

[29]Wood, Warren, Jeffrey Obelcz, Justin Tran, John Yu, Vishnu Karthik, and Benjamin Phrampus. (**2020**), Updates on global sediment thickness using geospatial machine learning, presented at the American Geophysical Union Fall Meeting, San Francisco, CA 1-17 December.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	01977	sanddocs@sandia.gov

Email—External (encrypt for OUO)

Name	Company Name
Dr. Warren T. Wood	U.S. Naval Research Laboratory
Dr. Hugh Daigle	University of Texas at Austin

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.