Gas Hydrate Modeling A Statistician's Perspective: *Evaluating a Complex Model*

by
Jack Schuenemeyer
Southwest Statistical Consulting, LLC
Cortez, Colorado USA

2012 International Association of Mathematical Geoscientists Distinguished Lecturer

Indian Statistical Institute, Bangalore
November 2012

Acknowledgements

- US Bureau of Ocean and Energy Management (BOEM) for financial support
- Matt Frye, BOEM project leader, geologist
- Tim Collett, US Geological Survey, geologist
- Gordon Kaufman, MIT, math statistician
- Ray Faith, MIT, Computer scientist, retired
- John Grace, ESA, Engineer/economist
- Rick Colwell, Oregon State Univ., biologist

Purpose of Talk

- Introduce a stochastic model to estimate gas hydrates with example from the US Federal Offshore
- Discuss related statistical issues, including model evaluation

Specifically

- "Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification", NAP, 2012
  - Verification – Computational accuracy
  - Validation – Represents state of nature (truth)
  - Uncertainty Quantification
- Focus on specific "Questions of Interest" (QOIs)

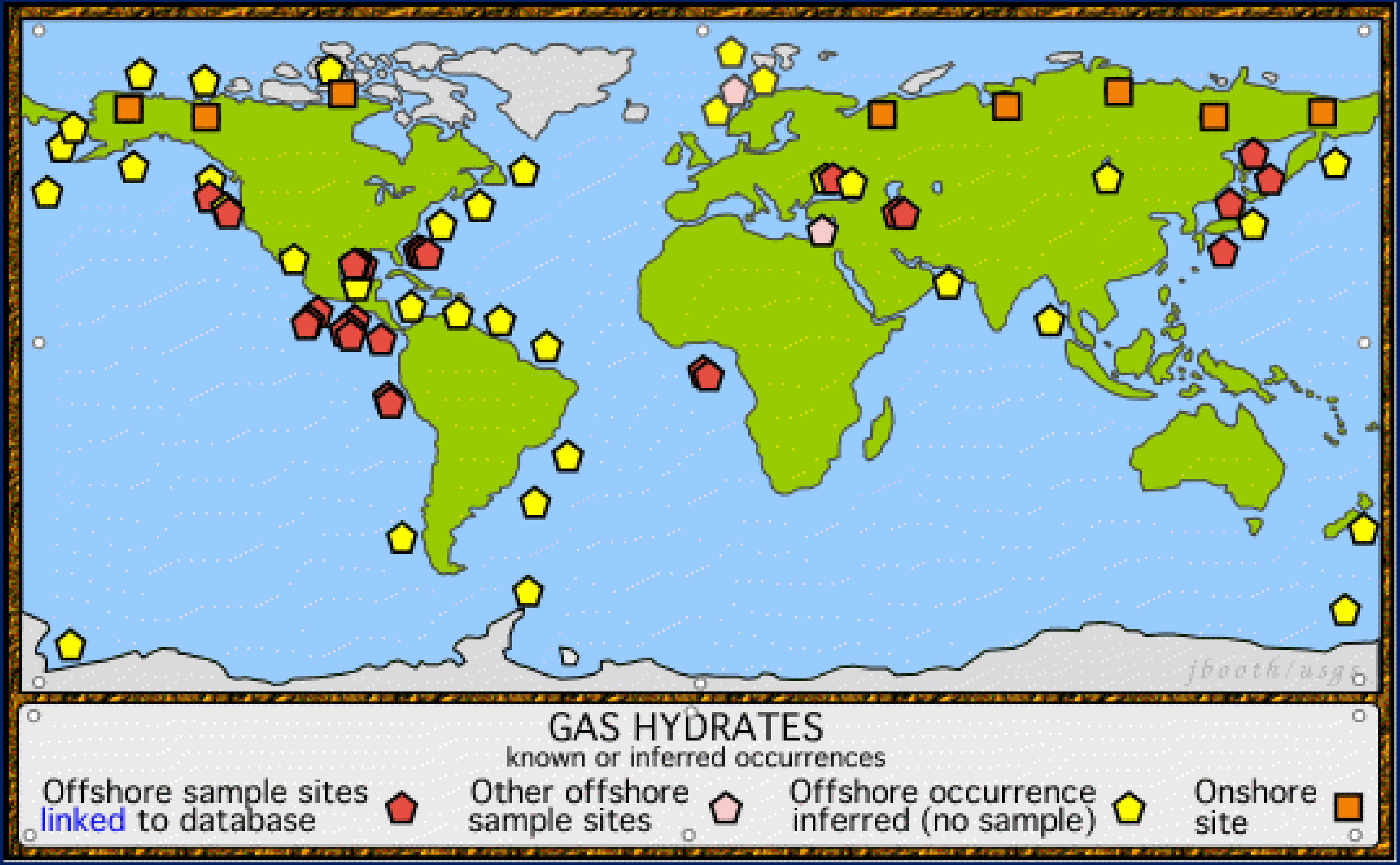
A Bit about Hydrates: Methane in Ice

Courtesy USGS

Crystalline solid consisting of gas molecules, usually methane, each surrounded by a cage of water molecules

Tim Collett, USGS

Gas Hydrate Locations



USGS

A Bit of History

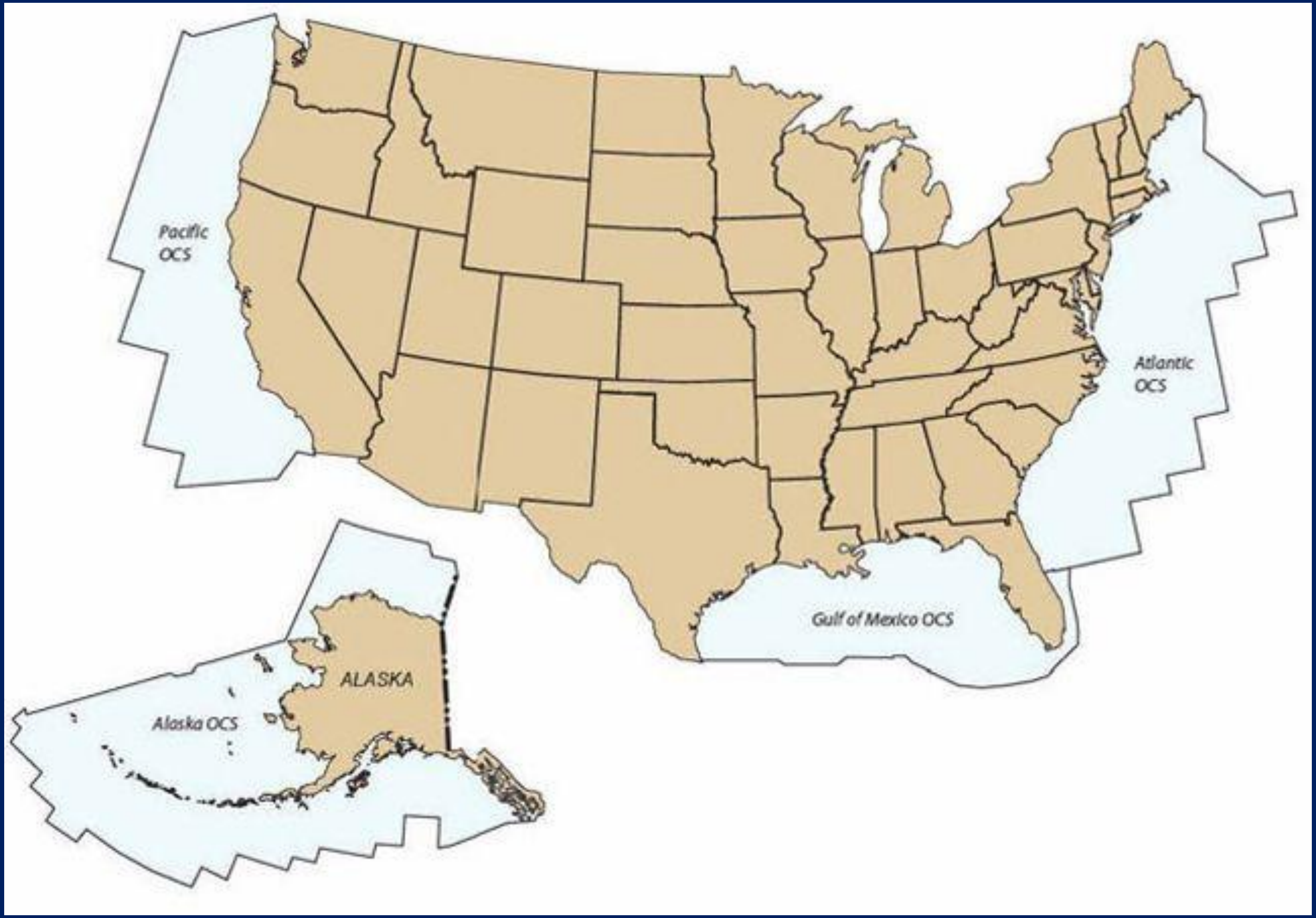
- 56 million years ago warming
- Hypothesis: Initial warming – volcanoes or Earth orbit fluctuations released methane
- Sudden, massive (to > 1,500 ppm) release of CO_2 into the atmosphere
- Planet warmed average of 5 deg C
- 150,000 years for oceans and forests to absorb excess

National Geographic, Oct. 2011, Robert Kunzig

Importance

- Potential supplement to natural gas
  - US natural gas consumption ~ 25 tcf/year
  - USGS 1995 assessment of in-place gas hydrate ~ 300,000 tcf; if 10% recoverable -> 1,000 years
- Natural gas 90% methane
- Natural gas heats 51% US homes
- 95% travels via pipe lines

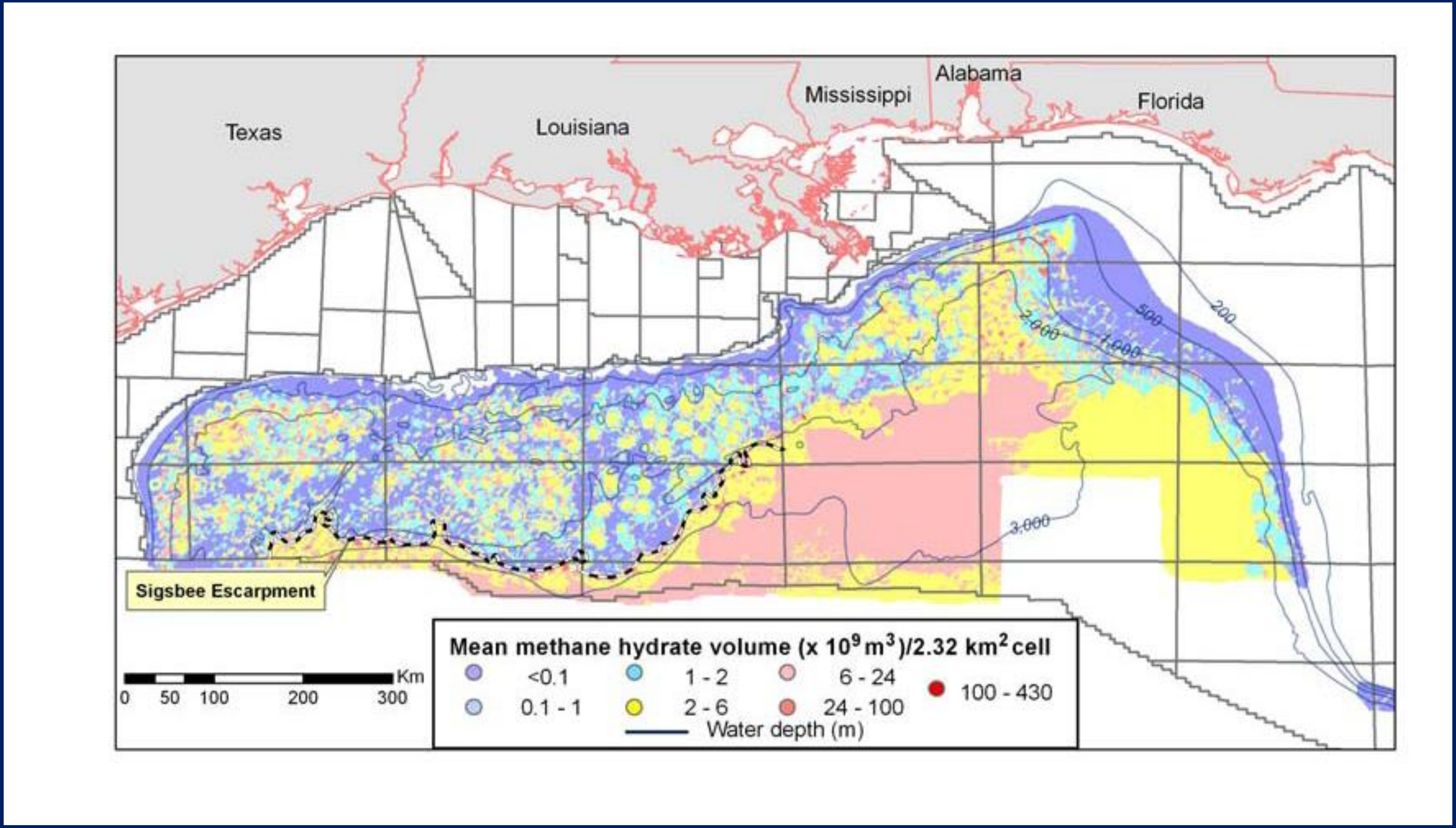
US Bureau of Ocean & Energy Management Assessment



Model

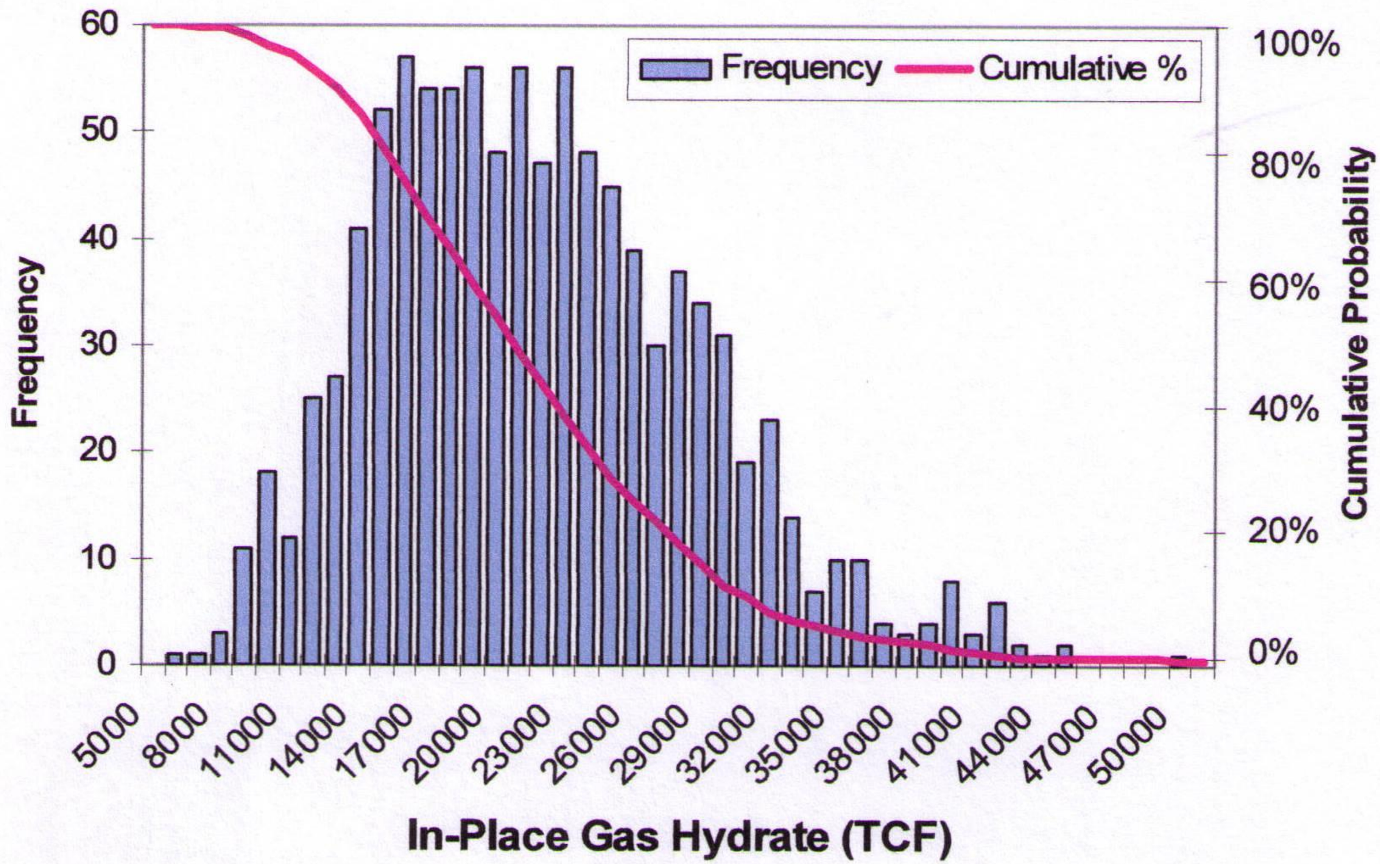
- Hierarchical model
- Cell based model (Cell area 2.3 km^2 to 5 km^2)
- Estimates in-place gas hydrates
- Biogenic process (thermogenic omitted)
- Stochastic as opposed to *scenario*
- US Federal offshore
- Below 300 meters water depth

Hydrate Volume By Cell, Gulf of Mexico

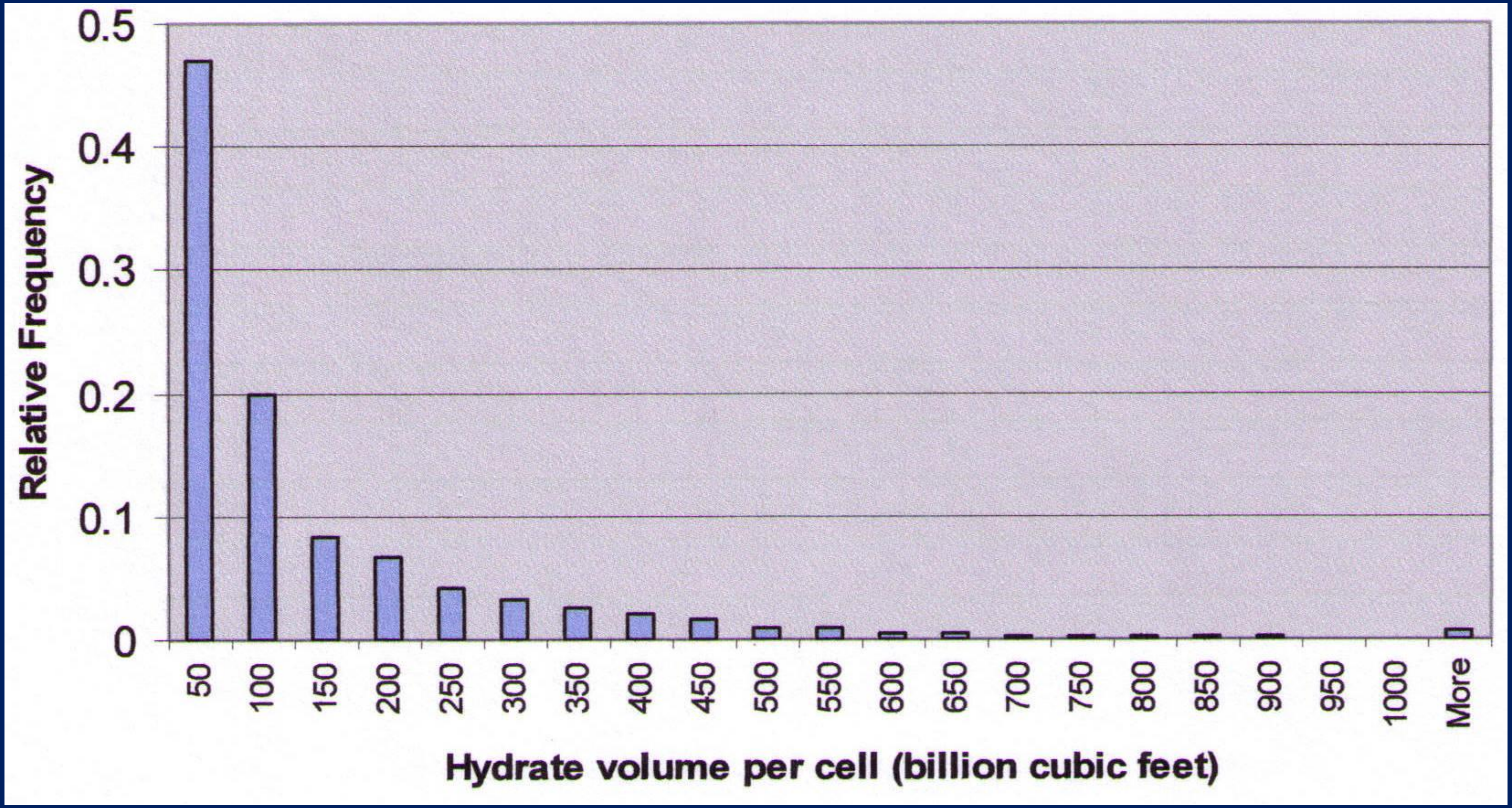


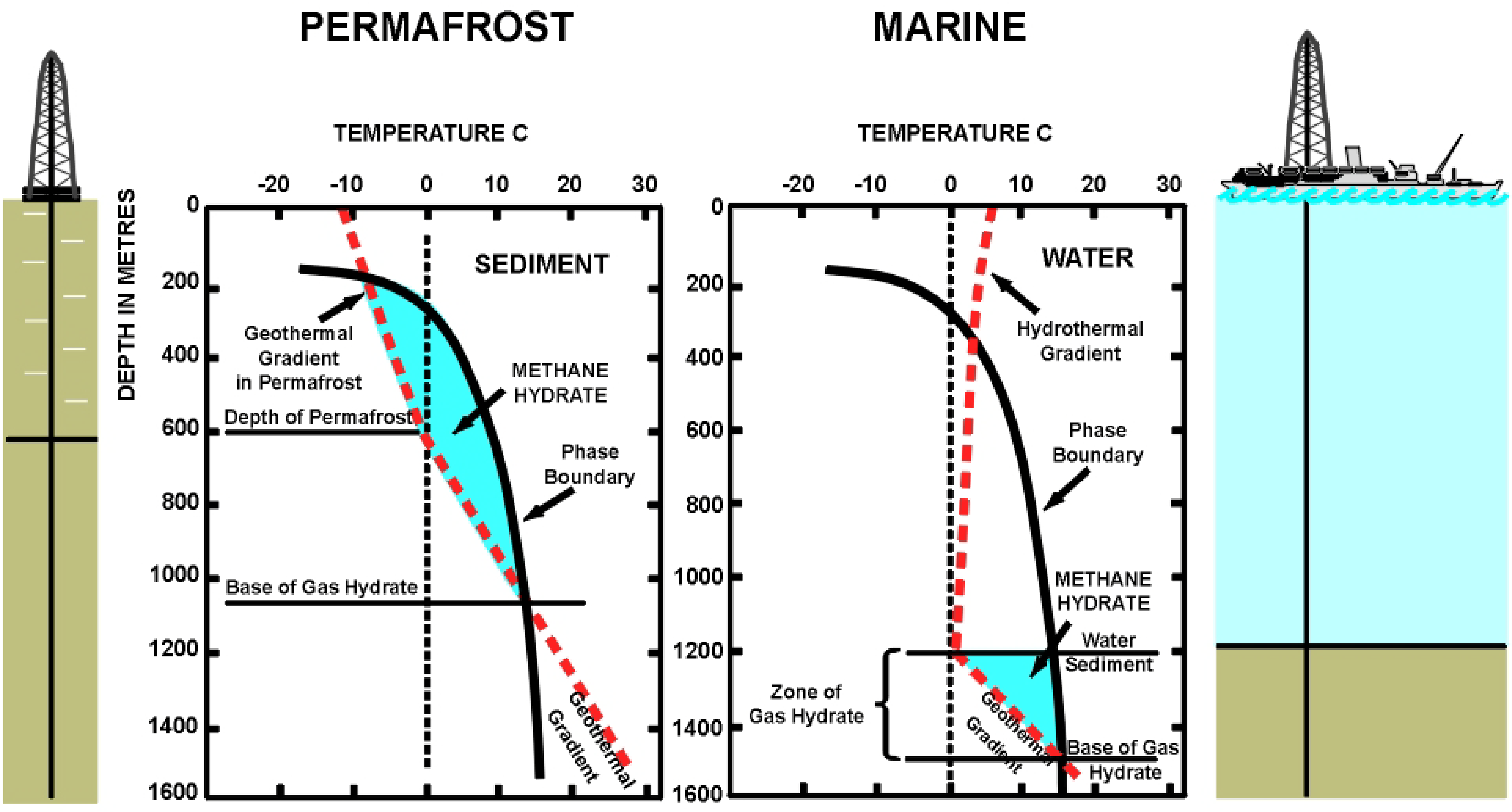
Gulf of Mexico (2008 results)

Region	95%	mean	5%
Gulf of Mexico	11,112 TCF	21,444 TCF	34,423 TCF



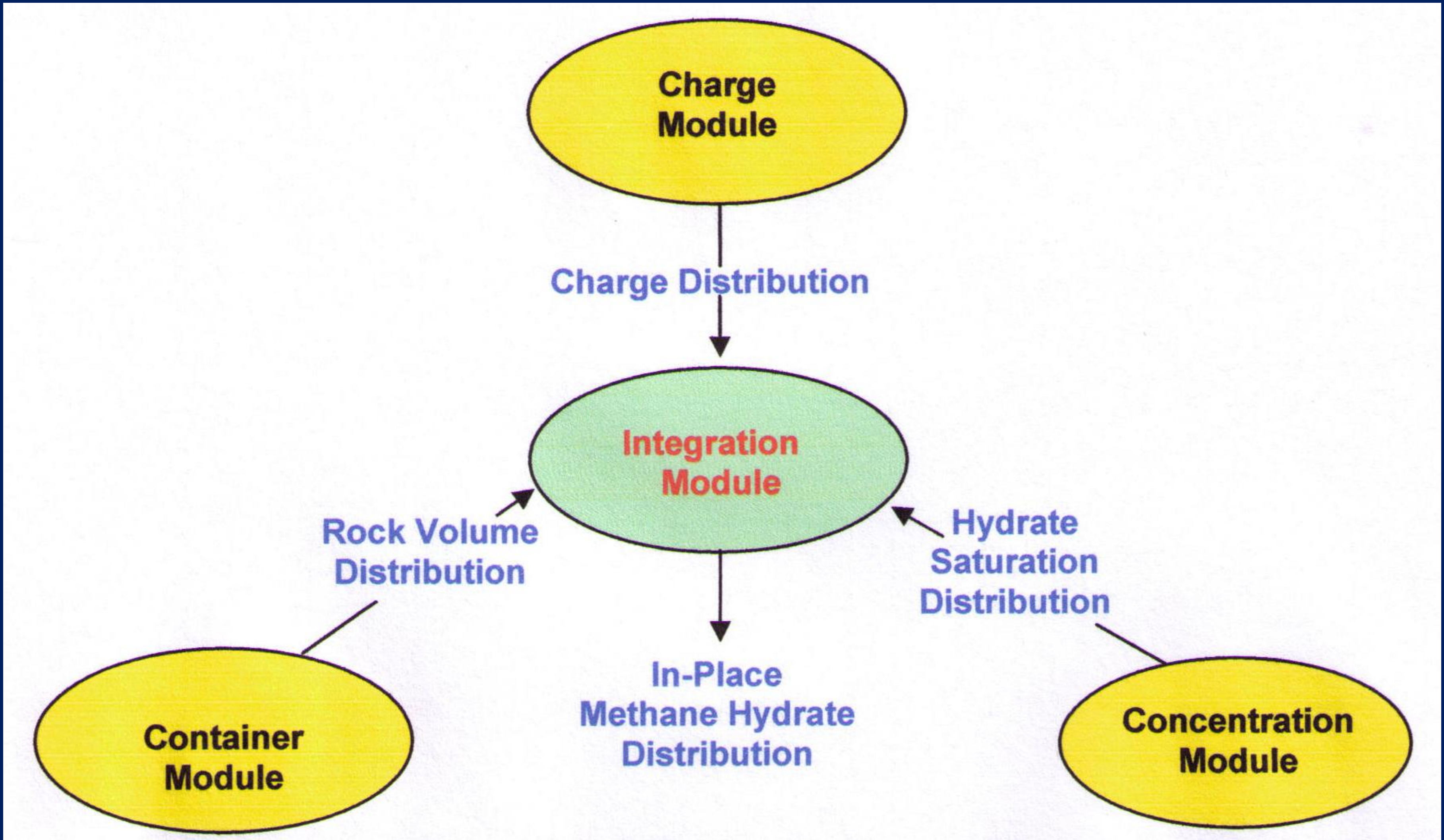
BOEM 2008, Gulf of Mexico





Tim Collett, USGS

The Gas Hydrate Assessment Model



Potential Volume

- Container Size: Hydrate Stability Zone (HSZ)
  - Water depth
  - Water bottom temp
  - Geothermal gradient
  - Undersaturated zone
- Saturation (How much room in container)
  - Sand & shale
  - Fraction sand
  - Sand & shale porosity – fn of HSZ thickness
  - Sand and shale concentration

Generation

- Source – Total Organic Carbon
- Depth
- Temperature
- Age
- Permeability
- Porosity
- Sediment density

Container

- Estimate Hydrate Stability Zone (HSZ)
- *HSZ* is a zero of:

$$f(HSZ\,|\,WD) = -[GTG \times \frac{HSZ}{1000}] - WBT + [\delta \times \ln(HSZ + WD) - \gamma] - \lambda$$

- where
  - *GTG* is geothermal gradient (degrees/km)
  - *WBT* is water bottom temperature; a function of water depth (*WD*)

Modified from Milkov and Sassen (2001)

Input Data

- By cell – location, depth, etc
- Model parameters – Excel spreadsheet
- Hard wired

By Cell Data
(131,745 cells, 2.32 km^2 each)

XID	LATITUDE	LONGITUDE	SANDP	WD	SALTISO	ANOM	BASIN_ID	CURVATURE	VERTICAL	BSR
cgom	27.60017	-88.17414	20	2471	3748	0	414	-27626038	1	0
cgom	27.50204	-88.11684	17	2476	4608	0	414	-1829776	1	0
cgom	27.48779	-88.10208	9	2479	4584	0	414	4977378	1	0
cgom	27.61441	-88.18892	20	2481	3445	0	414	17227454	1	0
cgom	27.61388	-88.17353	20	2493	3497	0	414	16362435	1	0
cgom	27.6582	-88.26411	13	2456	2250	1	420	168130768	0	2
cgom	27.64396	-88.2493	11	2462	2668	0	420	125241448	0	0
cgom	27.67296	-88.29432	18	2465	3356	0	420	61948868	0	0
cgom	27.67244	-88.27891	15	2470	2729	0	420	173680864	0	0
cgom	27.65767	-88.24871	15	2480	2056	1	420	164295200	0	2

Parameters – Excel Spreadsheet

Parameter	Type	Value1	Value2	Value3	Value4	Description
GeoThermal Gradient	Normal	25.546	4.634	0	50	Geothermal Gradient (mean,std,min,max)
BottomTempCoeffs1	Constant	21.19	-0.0251			WBT=wbtcoeff1*exp(wbtcoeff2*WD1)+wbtcoeff3+Error
BottomTempCoeffs2	Constant	1194	-0.78			
BottomTempCoeffs3	Constant	4.1				
BottomTempCutPts	Constant	400	1450			
BottomTempError	Normal	0	0.62	0.0556	0.184	Error term in above (Note: only standard deviation is used)

50 sub models

Sources of Data

- Hard data, i.e., water depth
- Published literature
- Analogs
- Expert judgment

Data is expensive!

Size of Model (Code)

- FORTRAN code – 3,000 lines of code
- Visual Basic – a few hundred lines
- R code – a few hundred lines

Statistical Issues

- Building a complex model
  - *How much complexity*?

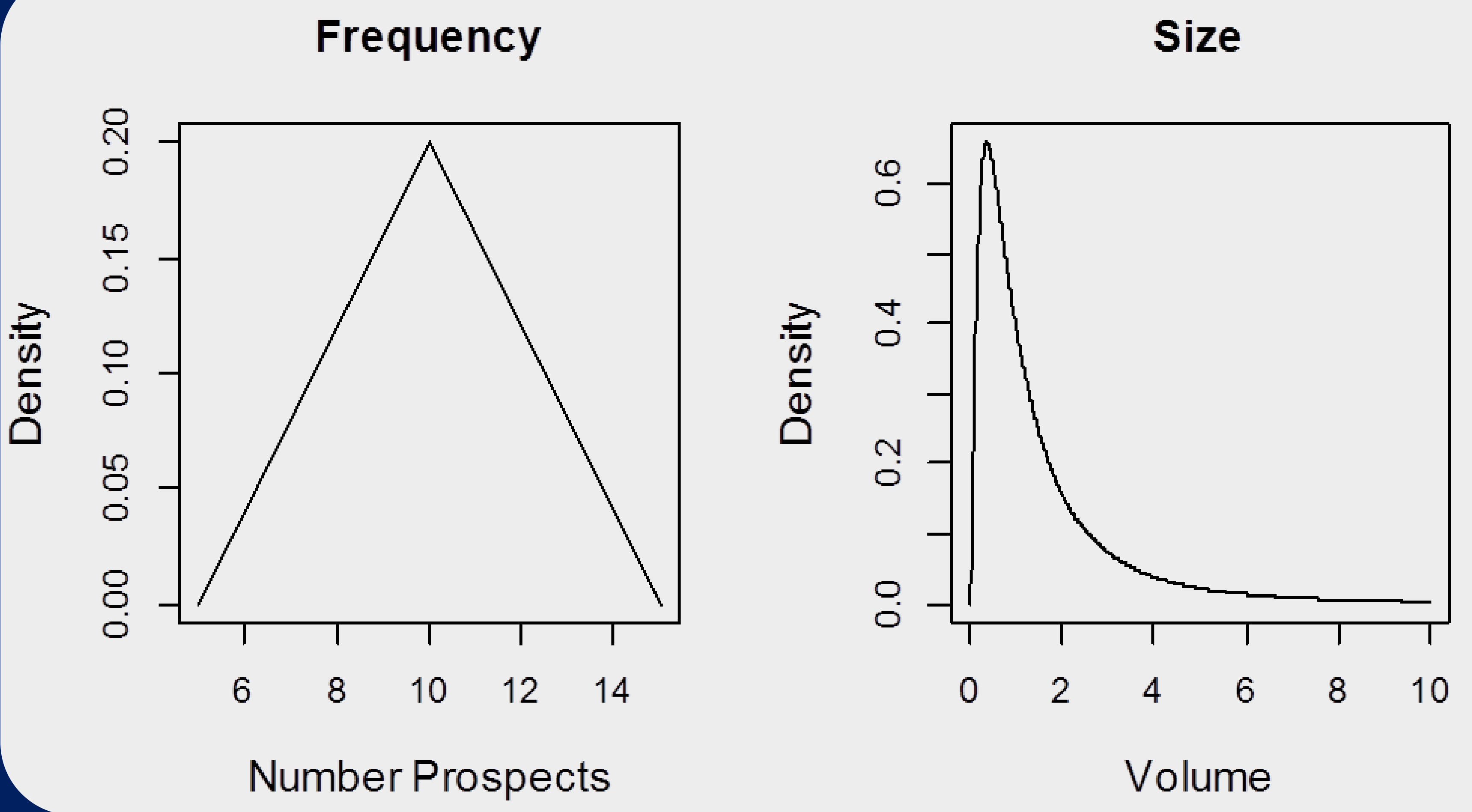
- Code/source verification
- Model validation
- Uncertainty quantification

Model Complexity

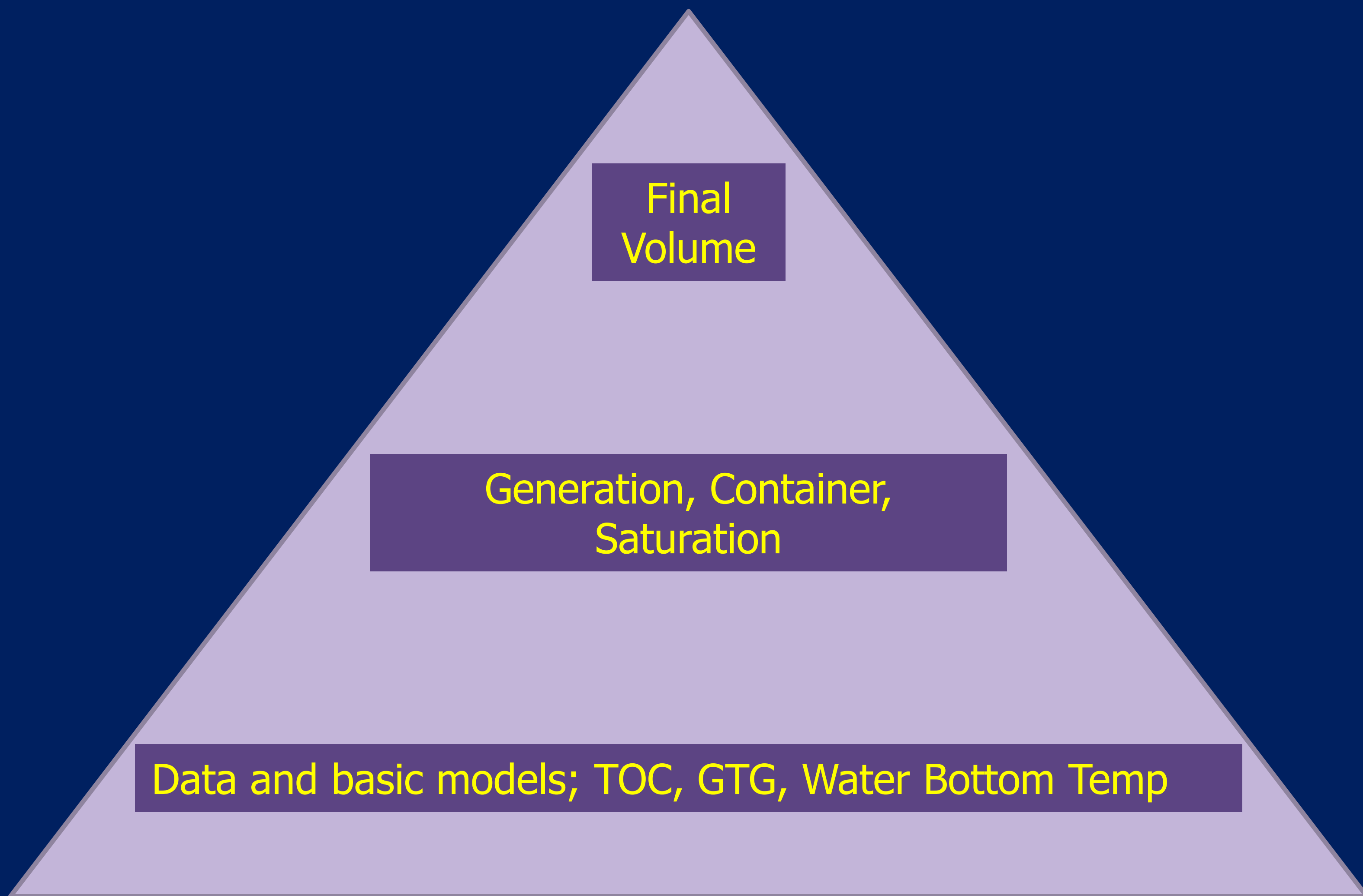
- Simple (see next slide)
- Moderately complex (Mass balance gas hydrate example)
- Excessively complex
  - Long run time
  - Understanding decreases
  - Knowledge base insufficient to make reasonable estimates of parameters

All models are wrong, some are useful – George Box

USGS 1995 Gas Hydrate Assessment Model



Validation Hierarchy



Validation

- Clear question of interest – Estimate in-place volume of gas hydrates
- Accuracy – Tough one!
  - Not manufacturing
  - Know results at some selected spots
- Applicability – Defined geographic area

Validation

- Model estimates versus observable
  - At the top level (hydrates in GOM) this is impossible, so
    - Pass a laugh test
    - Compare with other estimates (see means below)

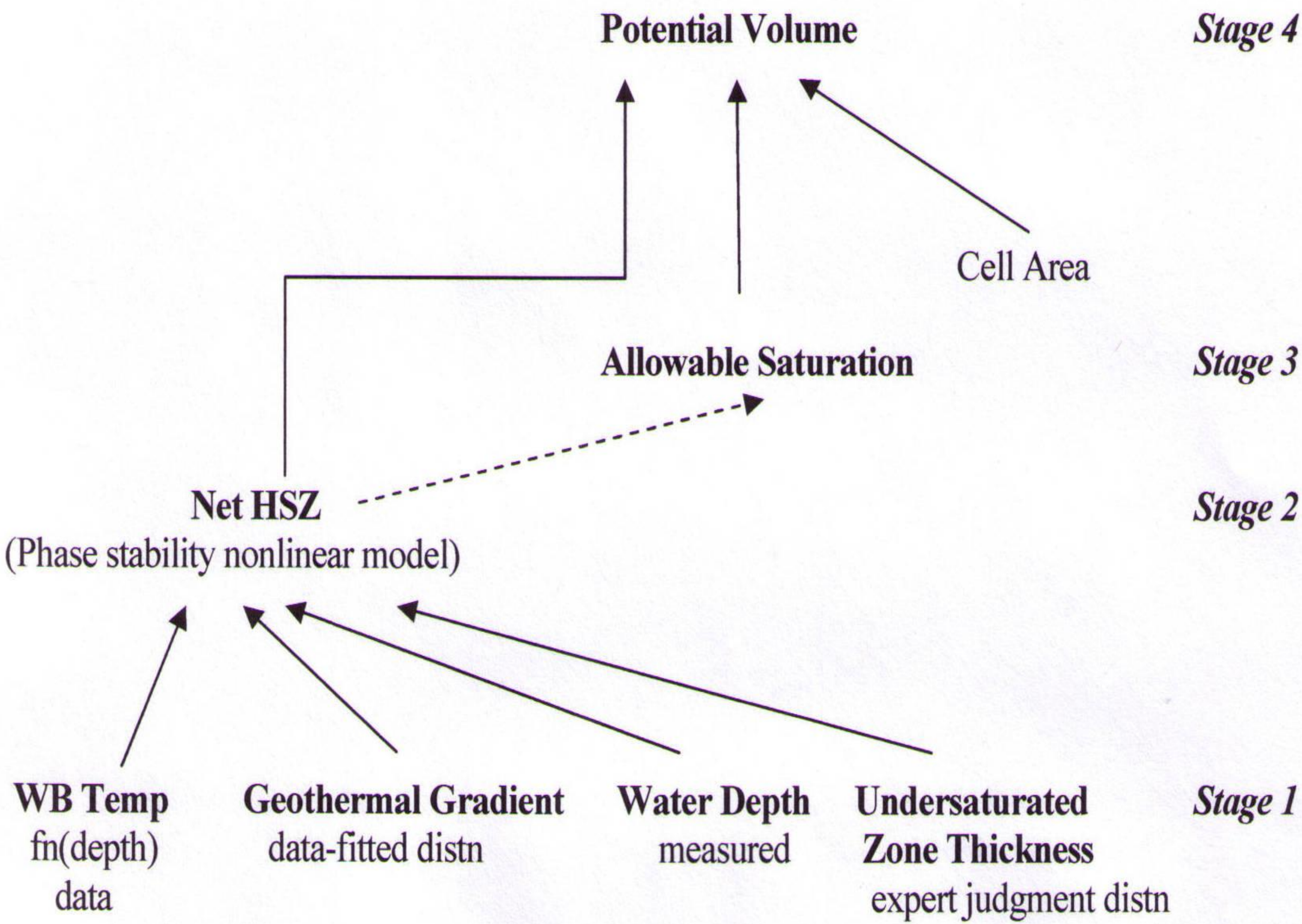
USGS,	1995 GOM	38,251 tcf
BOEM,	2008 GOM	21,444 tcf

Laugh Test - Example

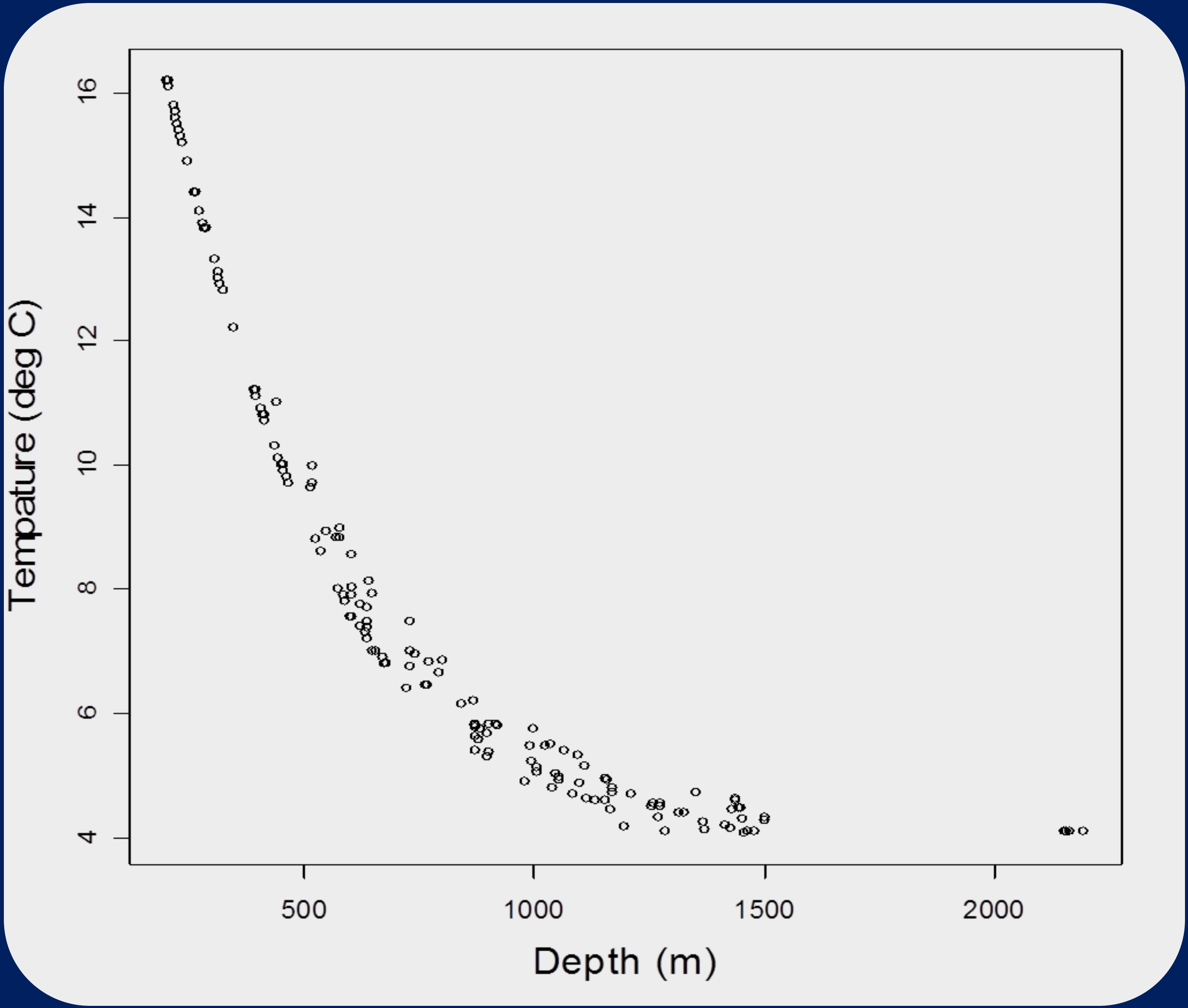
- Shale gas (technically recoverable in Poland)
  - Energy Information Adminstration reports a study estimating *187* trillion cubic feet (TCF) at the mean
  - US Geological Survey (USGS) reports a 05/95 range of *0 to 4* TCF!

Validation – Bottom Up

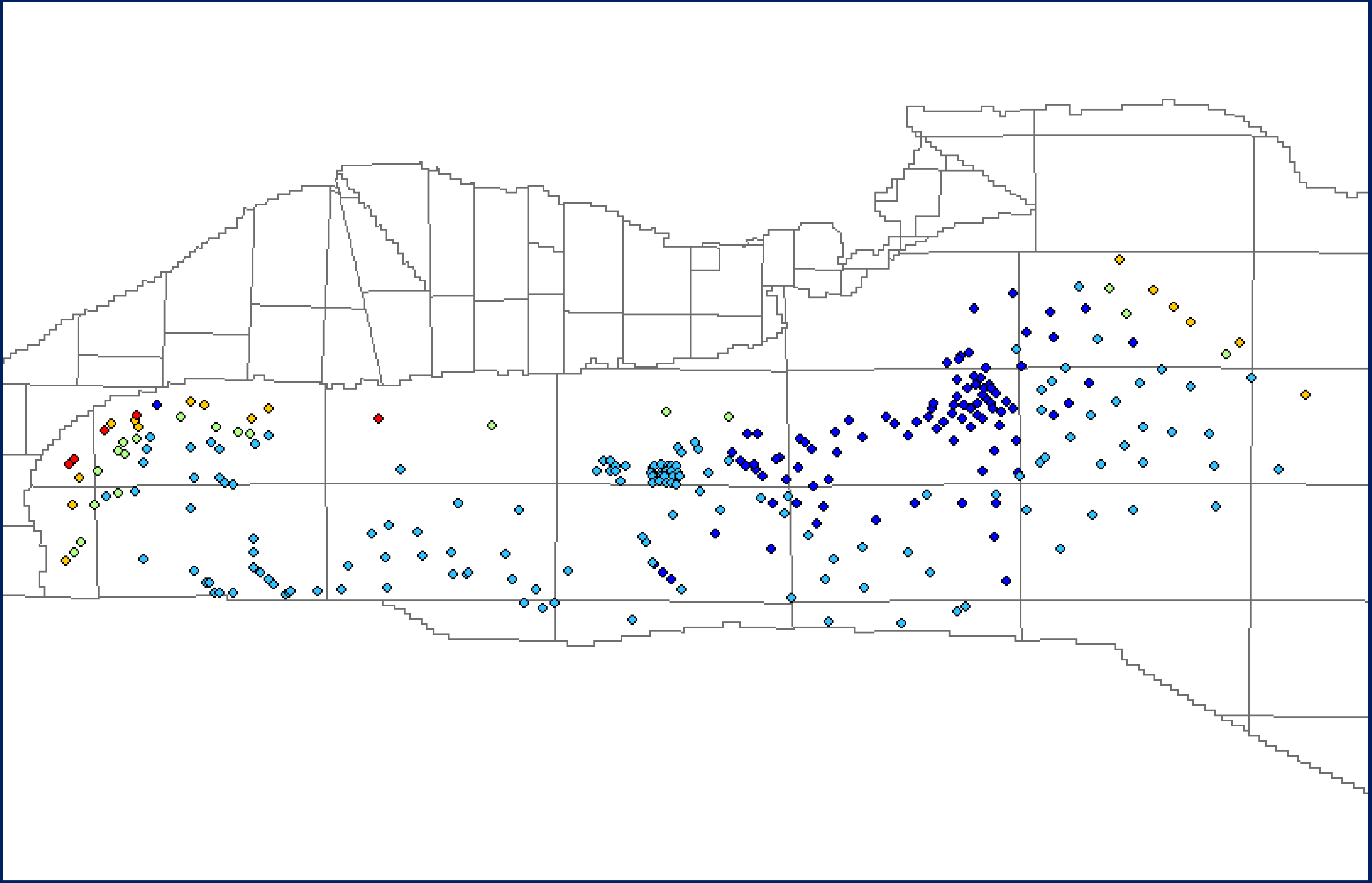
- Lowest level variables/models (data or model constructed directly from data):
  - Water bottom (seafloor) temperature
  - Geothermal gradient (rate of increasing temperature with respect to increasing depth in the Earth's interior)
  - Sediment thickness
  - Total organic carbon
  - Percent sand


Potential Volume
Stage 4
Cell Area
Allowable Saturation
Stage 3
Net HSZ
(Phase stability nonlinear model)
Stage 2
WB Temp
fn(depth)
data
Geothermal Gradient
data-fitted distn
Water Depth
measured
Undersaturated
Zone Thickness
expert judgment distn
Stage 1

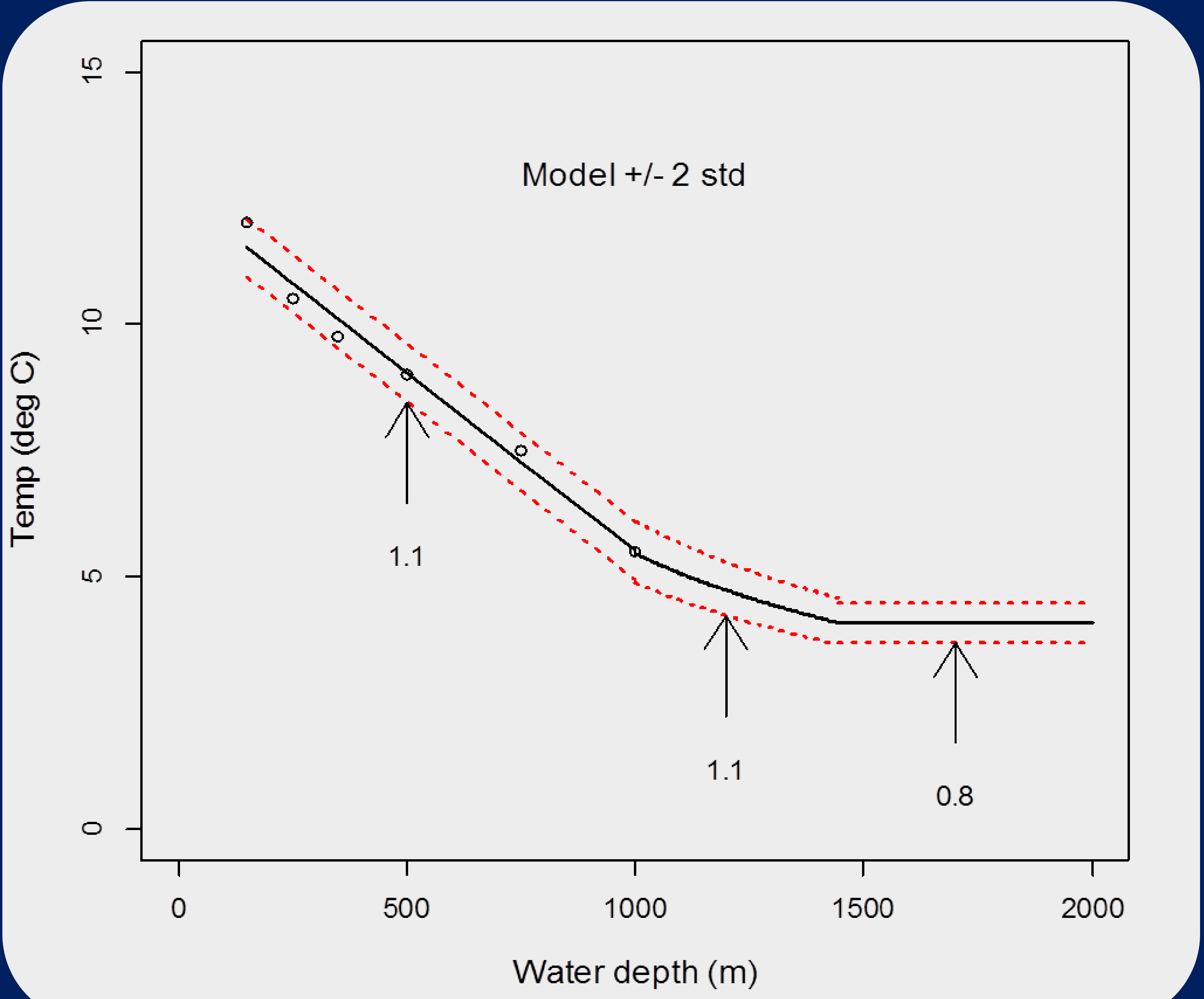
Water Bottom Temperature (WBT)



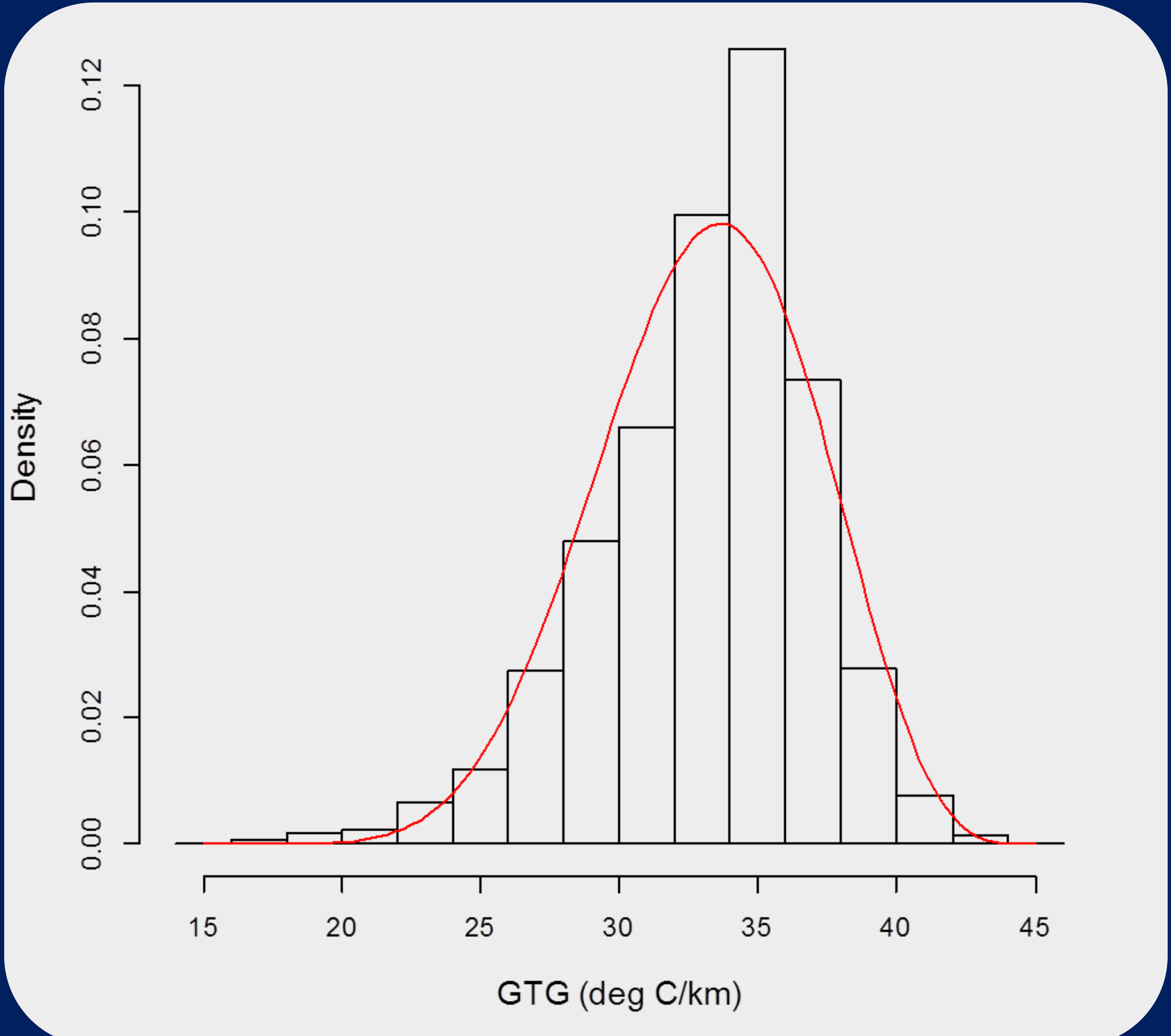
WBT Locations (GOM)



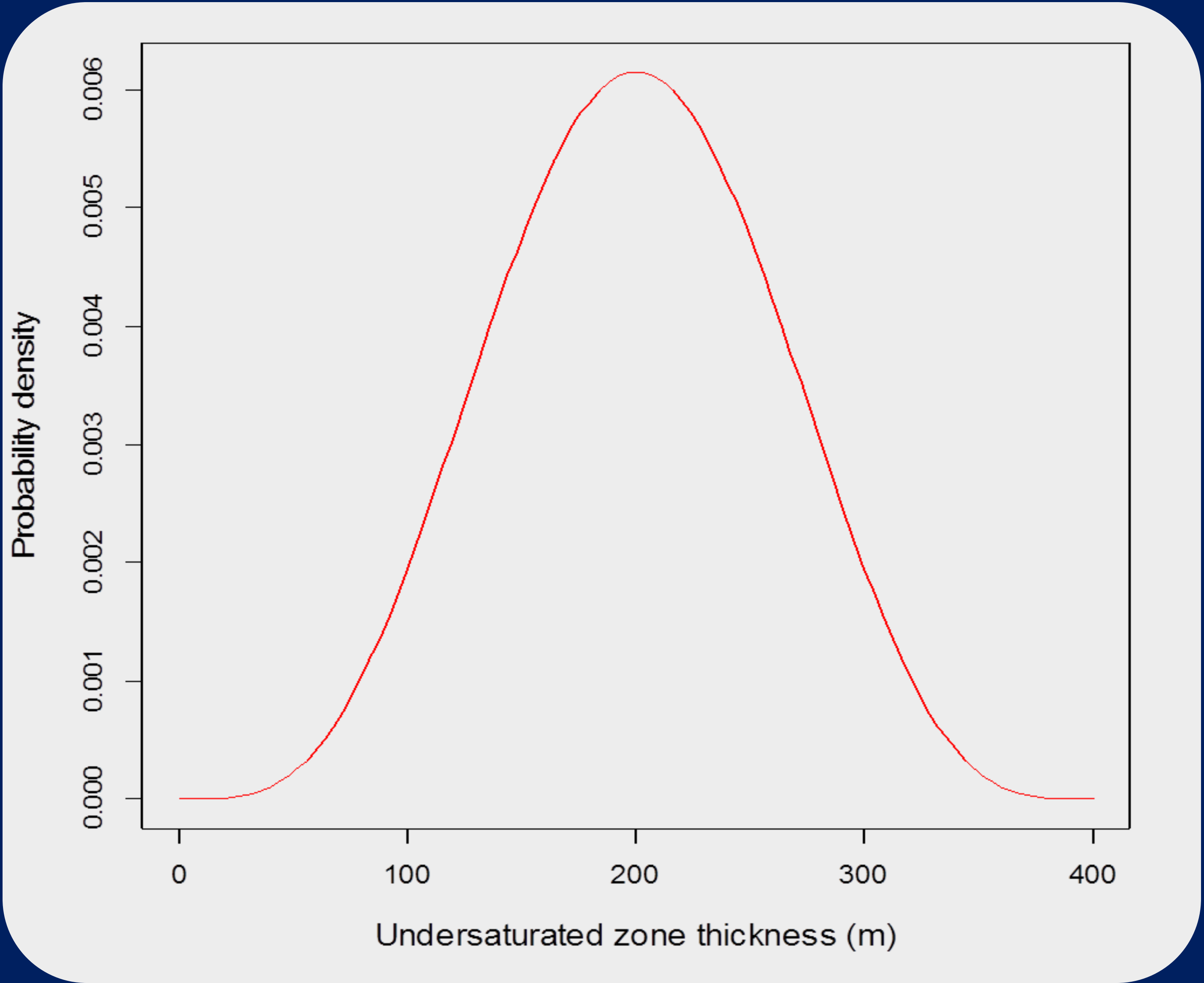
Water Bottom Temperature (*WBT*)



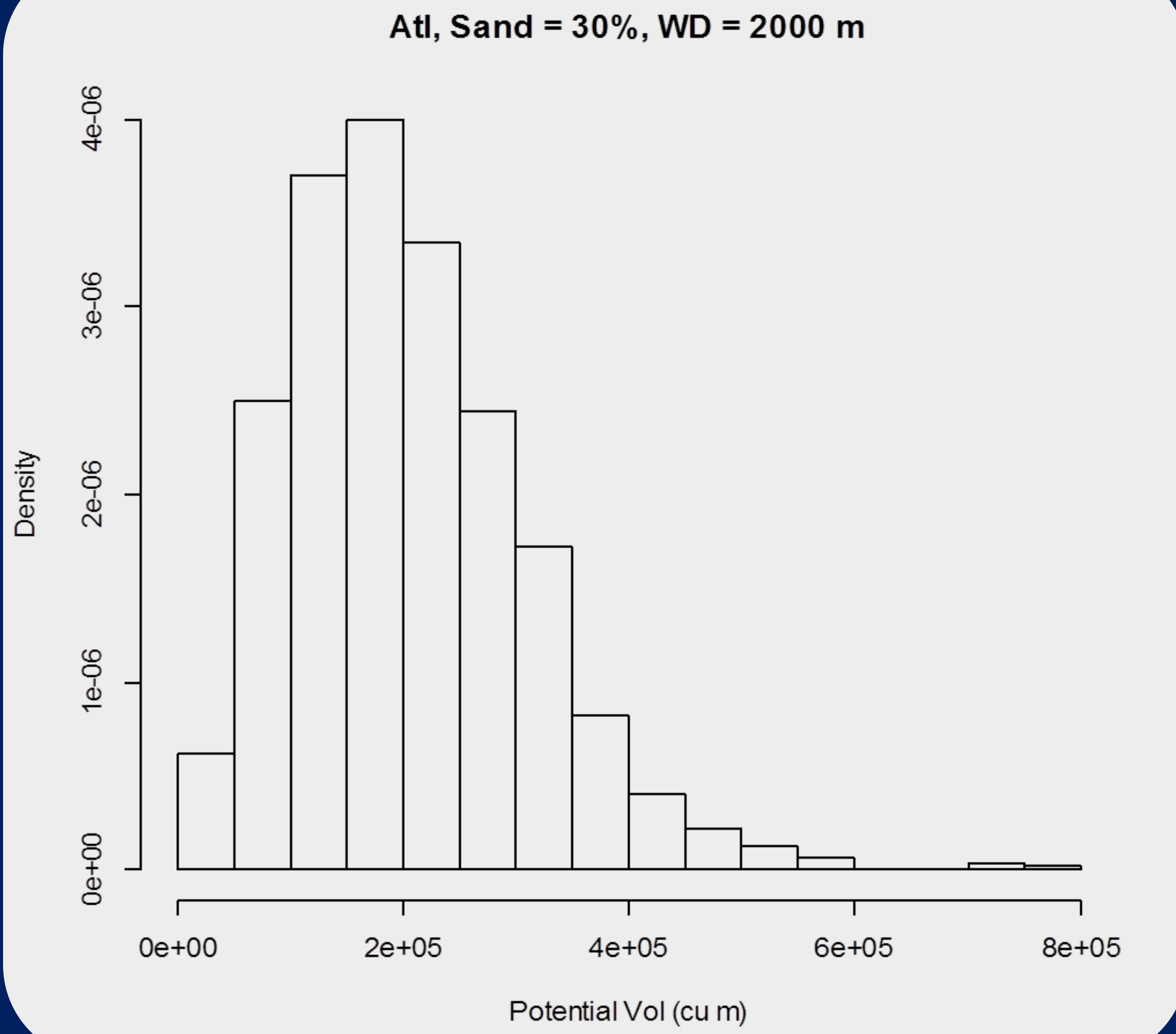
Geothermal Gradient



Undersaturated Zone Thickness

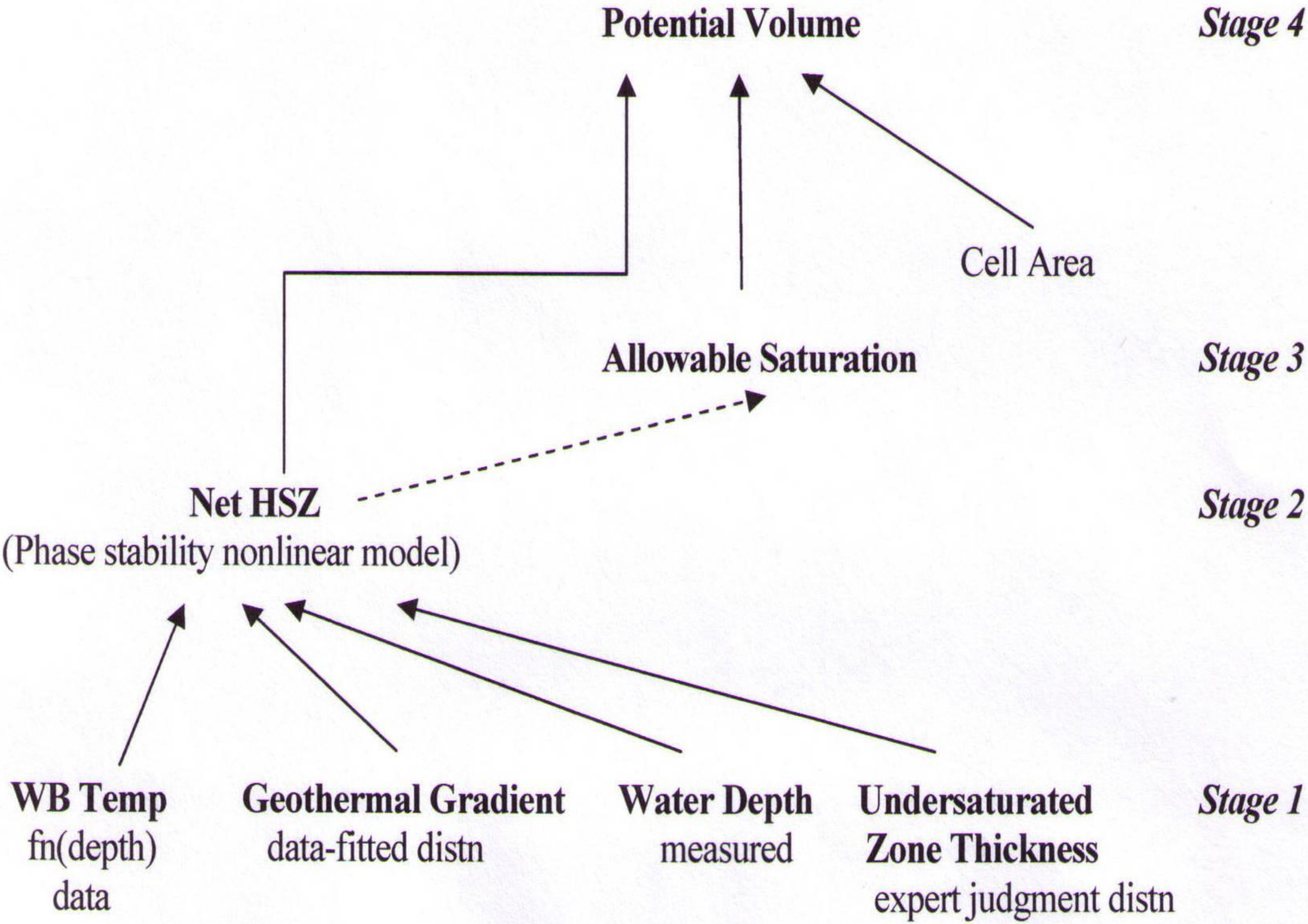


Potential Volume



Uncertainty Quantification on Potential Volume

- Range 0 m^3 to 800,000 m^3
- Consists of 12 sub models


Potential Volume
Stage 4
Cell Area
Allowable Saturation
Stage 3
Net HSZ
(Phase stability nonlinear model)
Stage 2
WB Temp
fn(depth)
data
Geothermal Gradient
data-fitted distn
Water Depth
measured
Undersaturated
Zone Thickness
expert judgment distn
Stage 1

More Concerns

- Bias
  - Representative data!
- Uncertainty
  - Input data
  - Model components
  - Propagation of error
  - Consistence with knowledge
- Dependency

More Concerns

- Dependency/correlation
  - Input – model components – aggregation
- Spatial correlation
  - Data/coverage
- Use of analogs
- Expert judgment

Still More Concerns

- Hard data
  - Occasionally data rich – satellite
  - Usually data poor – drilling expensive
  - Historical data sometimes unknown quality
  - Often spatially clustered
- "Soft" data – expert opinion
  - Eliciting information
  - Analogs
  - Integrating hard and soft data

Finally the Conclusions

- Mass balance reasonable approach
- “Easily” upgradable
- Incorporates geology and biology
- Probabilistic
- Preliminary results seem reasonable
- Output serve as input to technically recoverable estimate
- Transparent
- Reasonable run time

Thank you

- Questions – comments – suggestions

- Jack’s contact info: jackswsc@q.com
- Web site: www.swstatconsult.com

- IAMG: www.iamg.org

Dependency Concerns

- Many past oil, gas and other resource assessments have assumed:
  - Pairwise independence between assessment units (plays, cells, basins, etc.)
  - Total (fractile) dependence

Middle Ground on Dependency

- Develop *a statistical model* using geologic data to estimate correlations between neighboring cells, i.e., spatial extent of total organic carbon
- Use *expert judgment* based upon geology and analogy to specify associations
- Assume that cells are totally dependent within basins and independent between basins

Implications

- Perception of resource base is different depending on level of assumed or inferred association

- Risk that a government or company is willing to assume differs

Consider One Variable - Total Organic Carbon (TOC)

- Suppose a TOC = 3 wt % is selected from a random draw, i^{th} trial, $i = 1, 1000$
- Assumption
  - Independence – only applies to one cell
  - Basin dependence – applies to all cells in basin
  - Total (fractile) dependence – applies to all cells

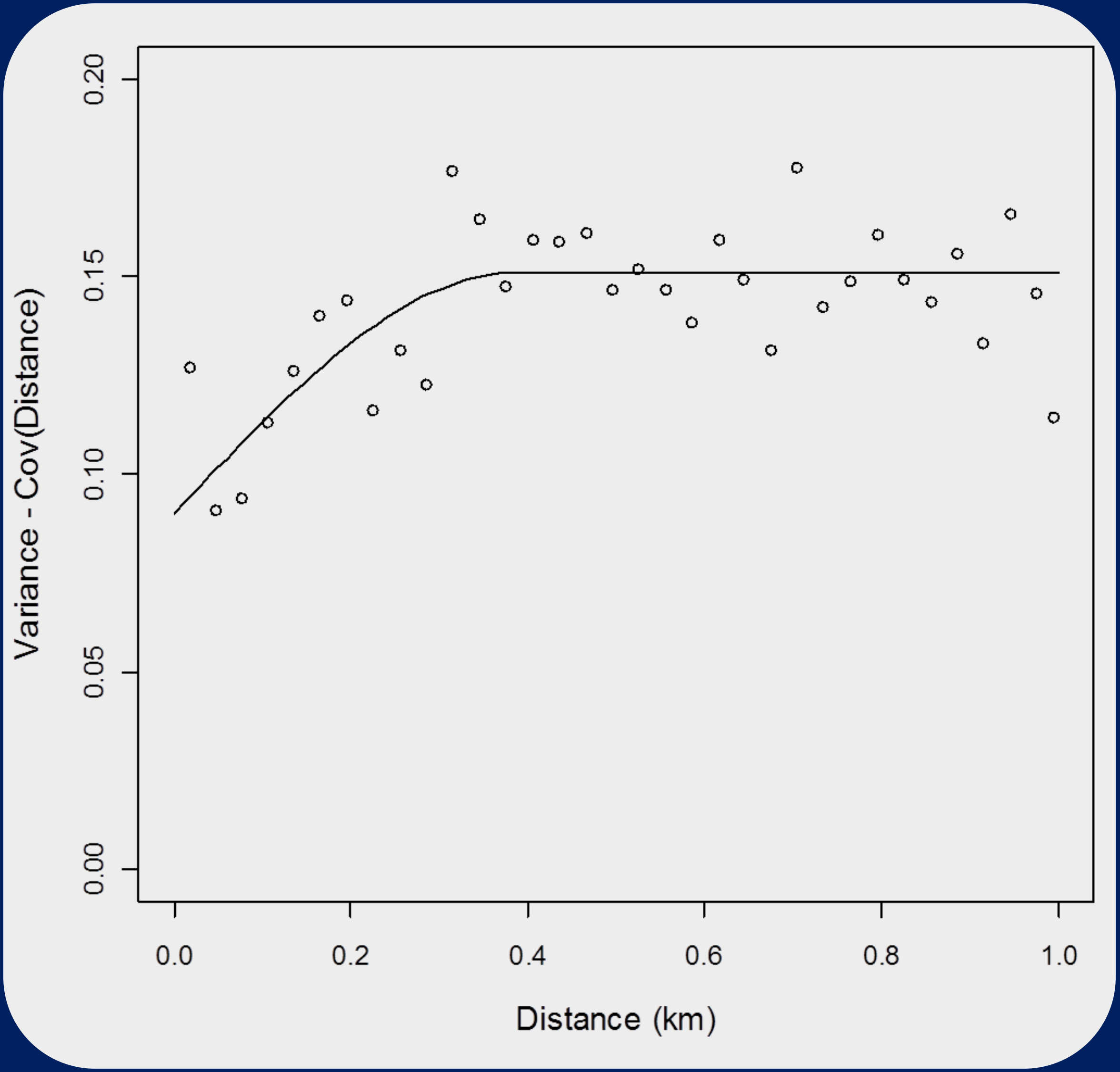
Dependency Issues

- Cell size (large, moderate, small)
  - Too small – sparse data
  - Too large – mask variability and dependence; impose artificial level of homogeneity
- Independence between nearby cells!
  - Uncertainty estimates of resource much too small

The BIG Question???

- How far does dependence extend?
- Denote the same variable measured a distance h from each other as Z(s) and Z(h+s); s is a location.
- Issue: Anisotropy (correlation varies by direction)
- Typical covariance model is shown in the following slide

Covariance (Semivariogram) Model



Correlation

- Does significant correlation exist
  - For 2 km
  - For 20 km
  - For 200 km
  - For 2,000 km
- Is the correlation the same in all directions? Probably not.

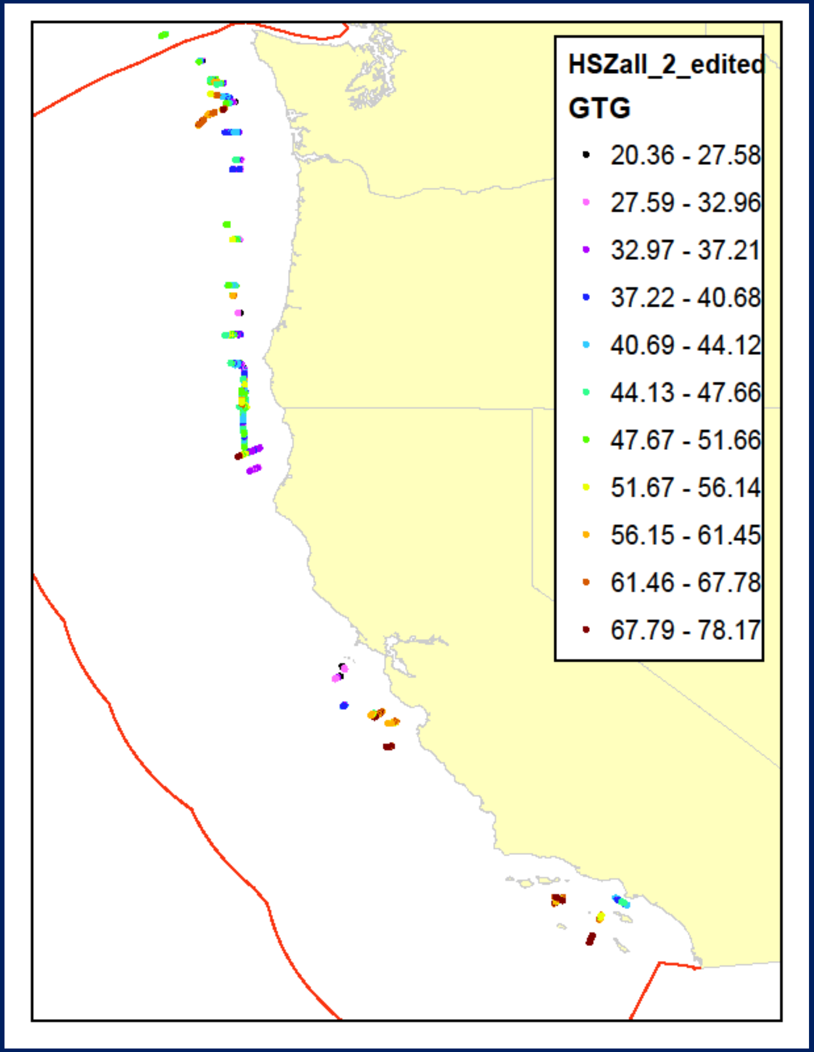
Other Considerations

- Influence of Bottom Surface Reflectors (BSRs)
- Global spatial trends
- Geologic discontinuities
- Use of expert judgment to define associations between cells

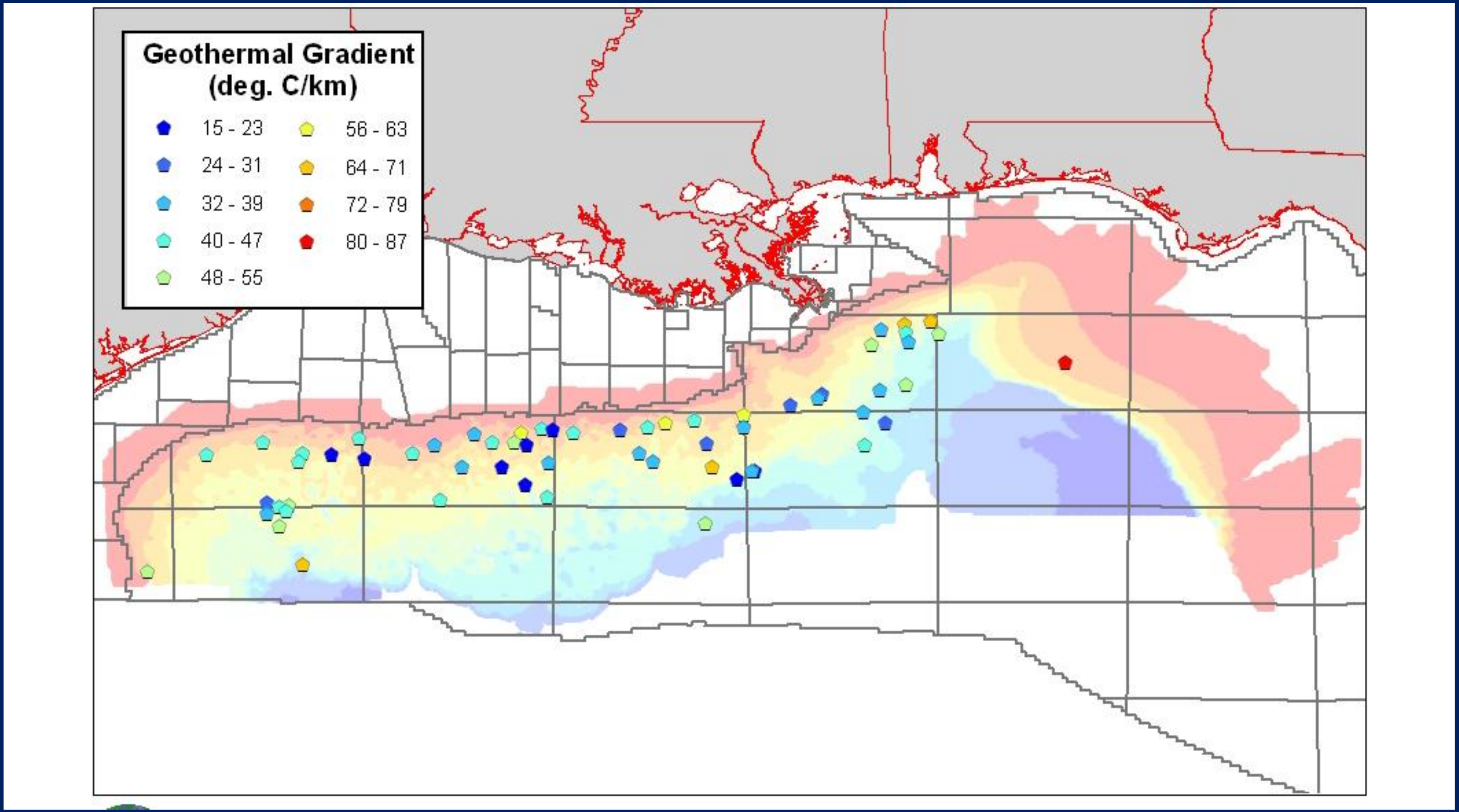
Representative Data

- A statistician's wish
  - To be able to design an experiment
  - To collect representative data to address
- In real life time, cost, accessibility intervene

Geothermal Gradient (GTG) Pacific Well Sites

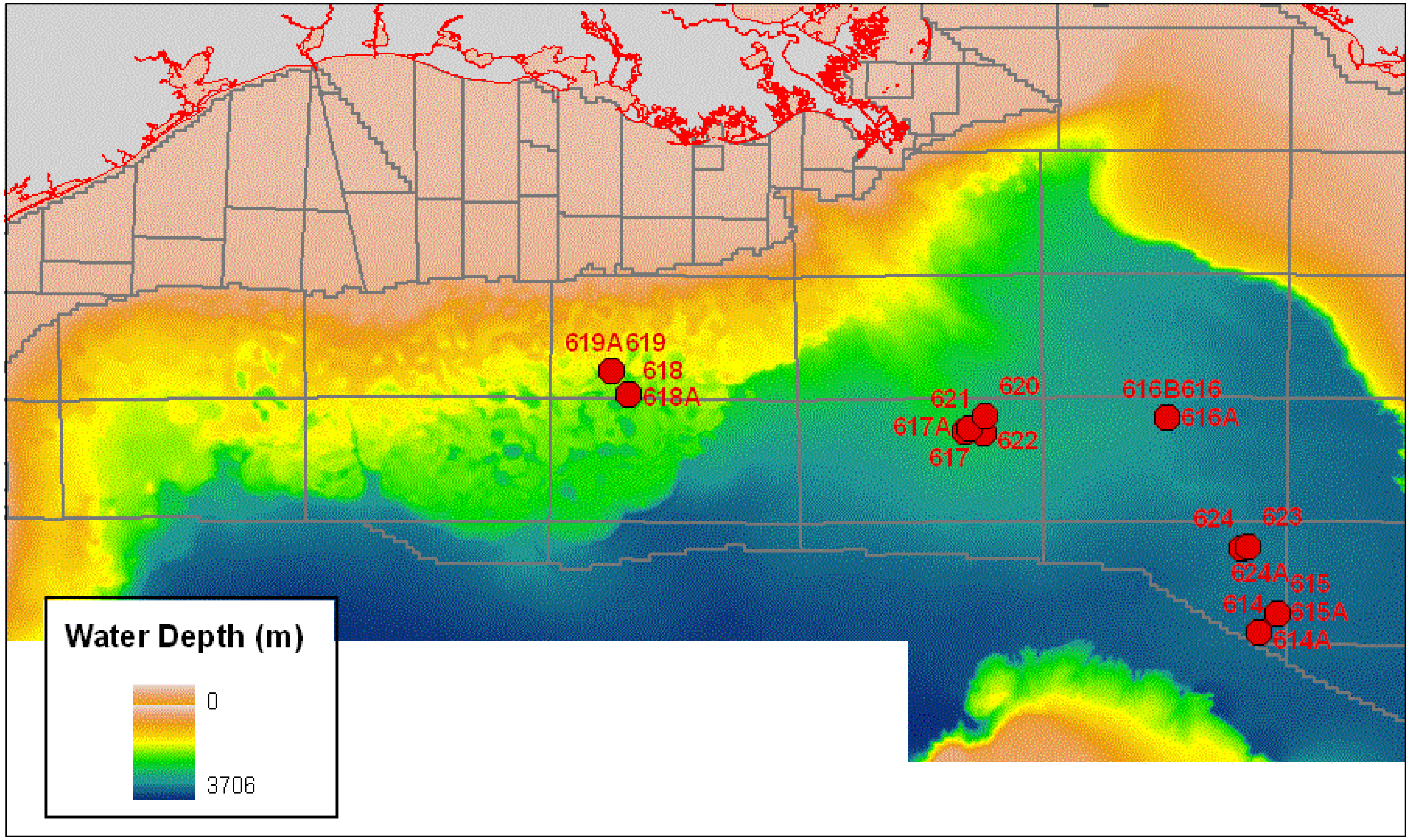


Gulf of Mexico



No obvious trend

Total Organic Carbon Sites



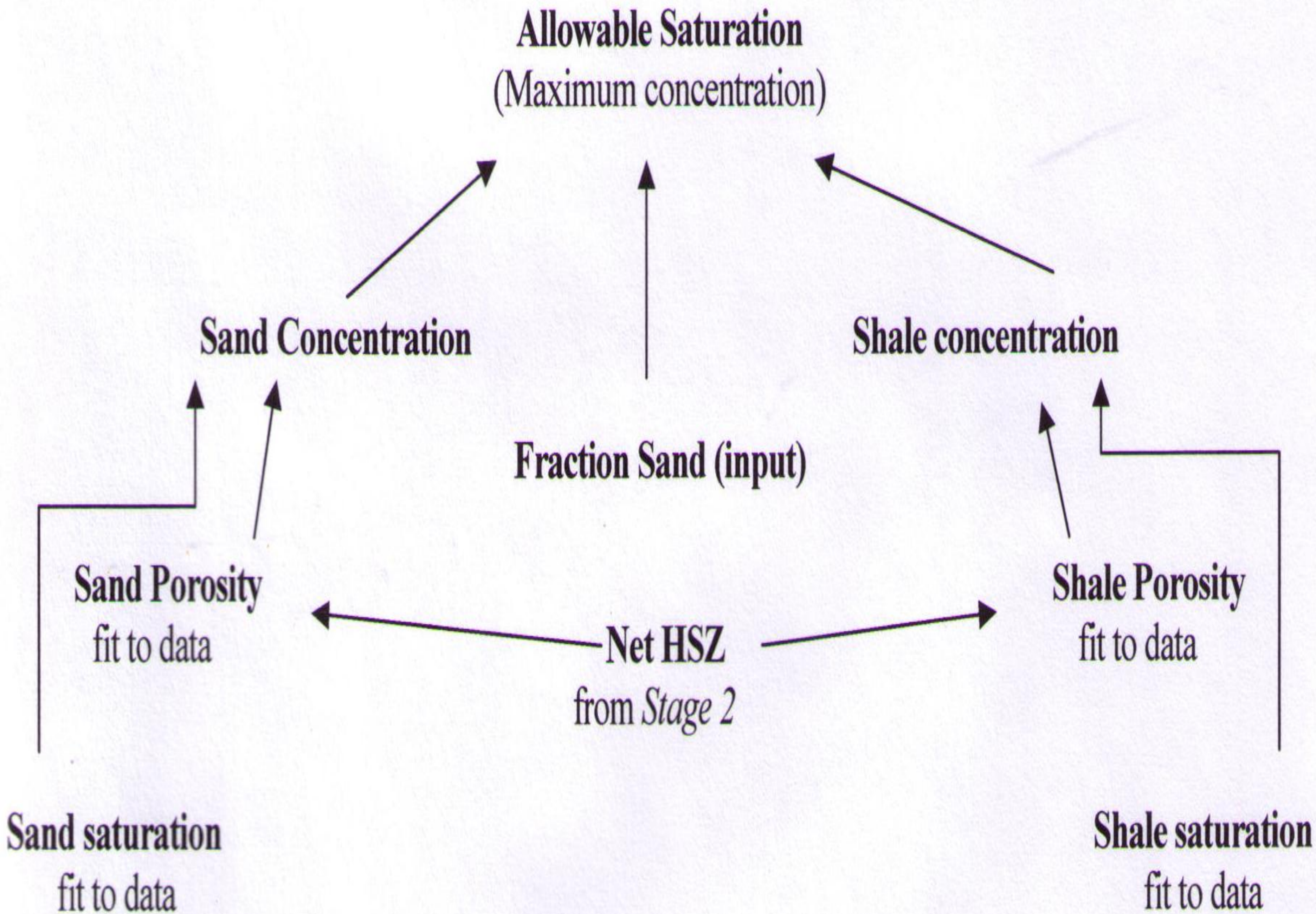
Partial Solutions

- Documentation
  - However …
- Evaluation
  - Results seem reasonable – not all scientific results seem reasonable at first
  - Consistent with measurements where hard data exists
  - Make available to public

Thank you

- Questions – comments – suggestions

- Jack’s contact info: jackswsc@q.com
- Web site: www.swstatconsult.com


Allowable Saturation
(Maximum concentration)
Sand Concentration
Shale concentration
Fraction Sand (input)
Sand Porosity
fit to data
Shale Porosity
fit to data
Net HSZ
from Stage 2
Sand saturation
fit to data
Shale saturation
fit to data

International Association for Mathematical Geosciences Student Chapters

- Student Chapters
  - You can plan events
  - Interact with other students
  - Receive financial help attending conferences
  - Meet practicing scientists
- Earth, climate, and environmental sciences are exciting disciplines, which combined with mathematics and statistics, can help solve important problems that will benefit us and future generations

International Association for Mathematical Geosciences

- Sandia National Labs – Albuquerque, NM
- Beijing Univ. of Geosciences – Beijing
- Dept of Earth Sciences, Univ. of Ottawa
- Dallas Geophysical Society, Dallas, TX
- TNO (energy company), Netherlands
- Univ. of Twente, Netherlands
- Univ. of Freiberg, Germany
- University of Georgia, Athens, GA
- India – 5 talks