MIT EARTH RESOURCES LABORATORY
ANNUAL FOUNDING MEMBERS MEETING 2020

Earth Resources Laboratory

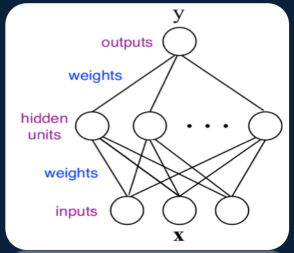
Bayesian Model Selection and the Application to Geothermal Development

Chen Gu
POSTDOCTORAL ASSOCIATE [EARTH, ATMOSPHERIC AND PLANETARY SCIENCES]

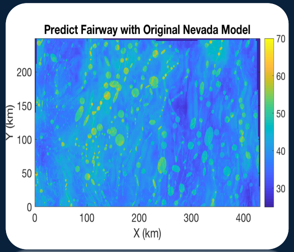
In collaboration with Youssef M. Marzouk, Stephon Brown, Michael Fehler, and M. Nafi Toksöz

Outline

- Research motivation – Geothermal opportunity estimate, uncertainty quantification and model selection



- Methodology – Bayesian inference and Bayesian model selection

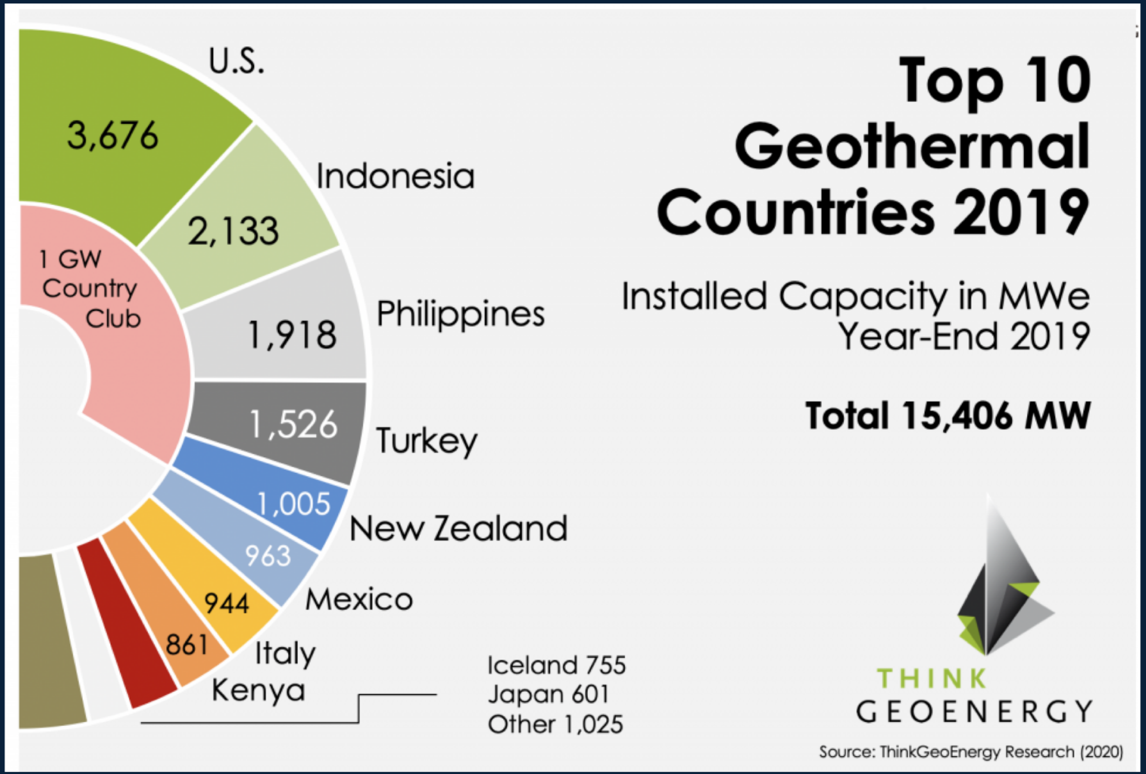


- Examples – Application to geothermal development

Great Basin

- The Great Basin region is a world-class geothermal province with ~720 MW electricity of current gross generation from ~24 power plants.
- And an estimated capability of ~**30,000** MW electricity!

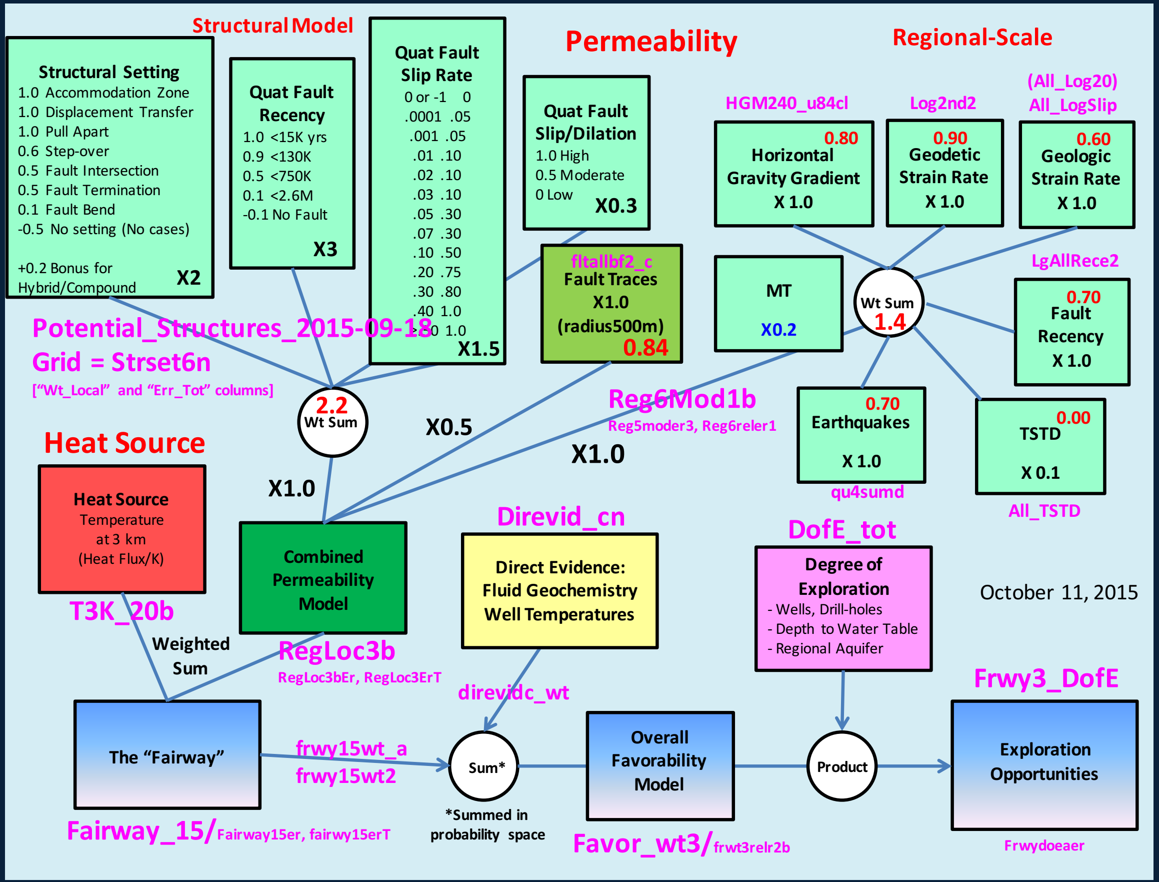
Figure: https://en.wikipedia.org/wiki/Great_Basin



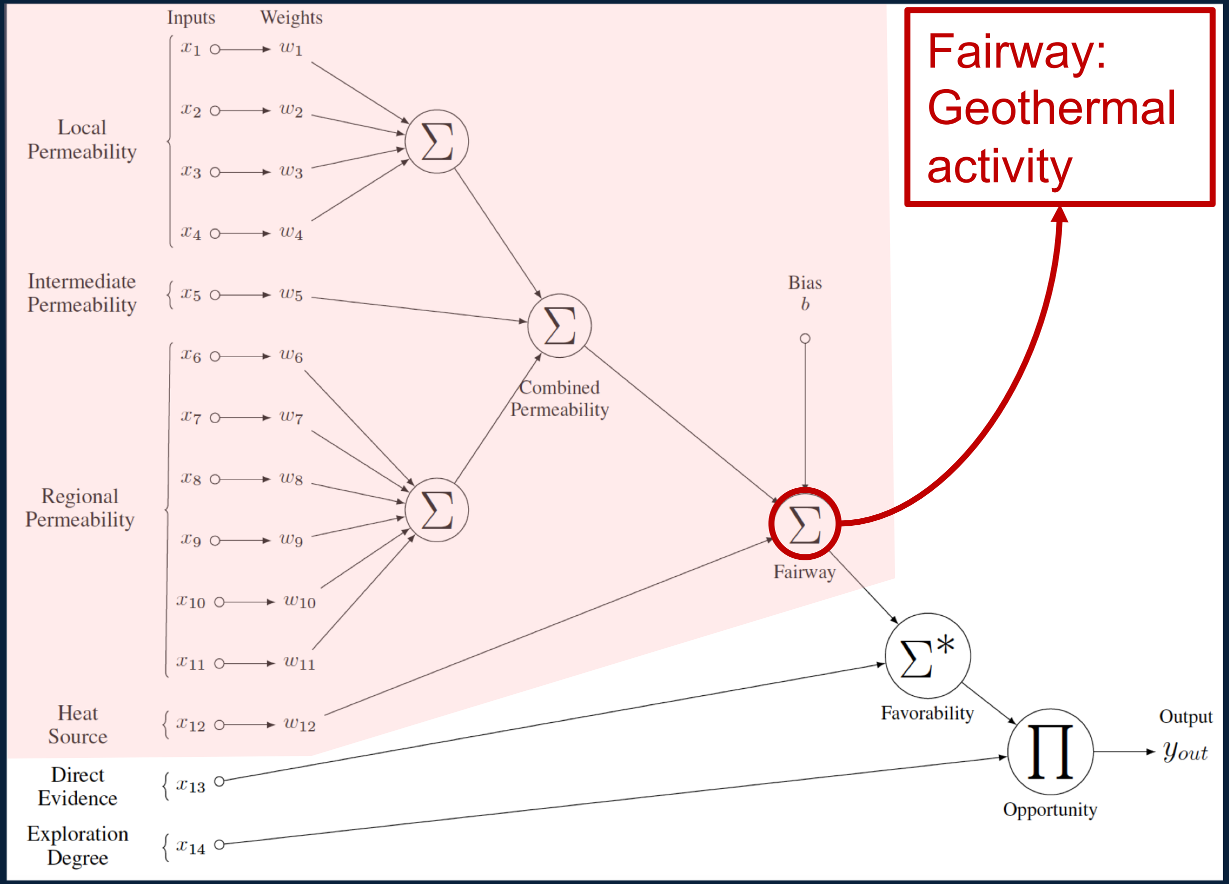
Geothermal opportunity pipeline

- Geological and geophysical measurements

Original Nevada model from geothermal experts



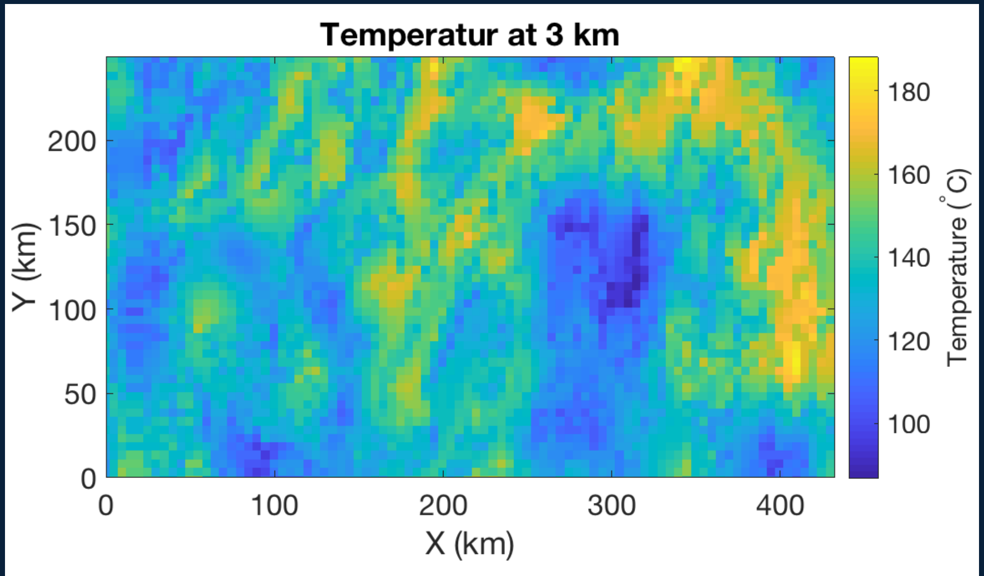
Feedforward network representation

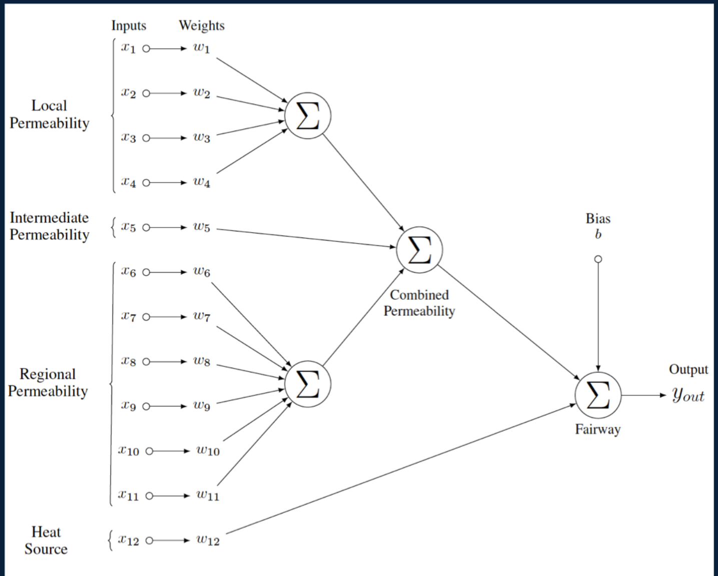


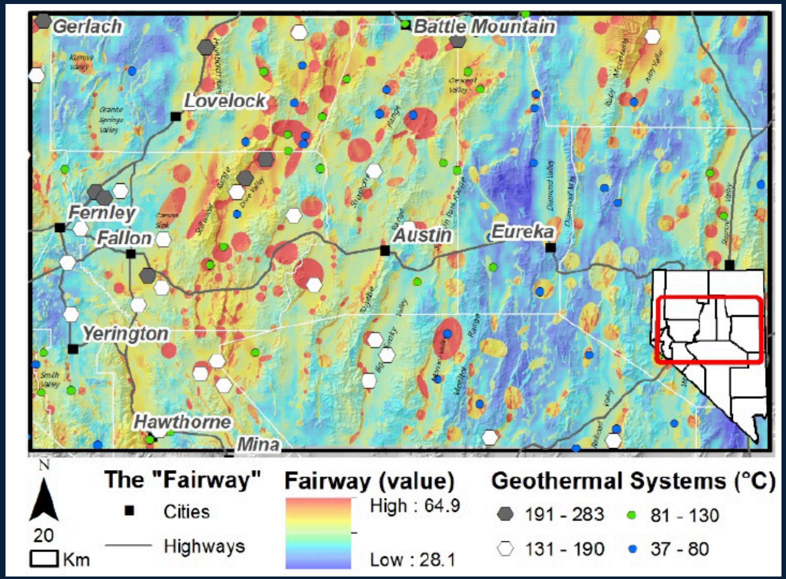
Geothermal play fairway analysis (PFA)

- Geothermal play fairway analysis (PFA) is a concept adapted from the petroleum industry to improve the exploration success rate (Doust, 2010).
- It involves integration of geologic, geophysical, and geochemical parameters indicative of geothermal activity (Faulds et al., 2017).

Geo-X Input → **Nevada PFA Model** → **Fairway Output**

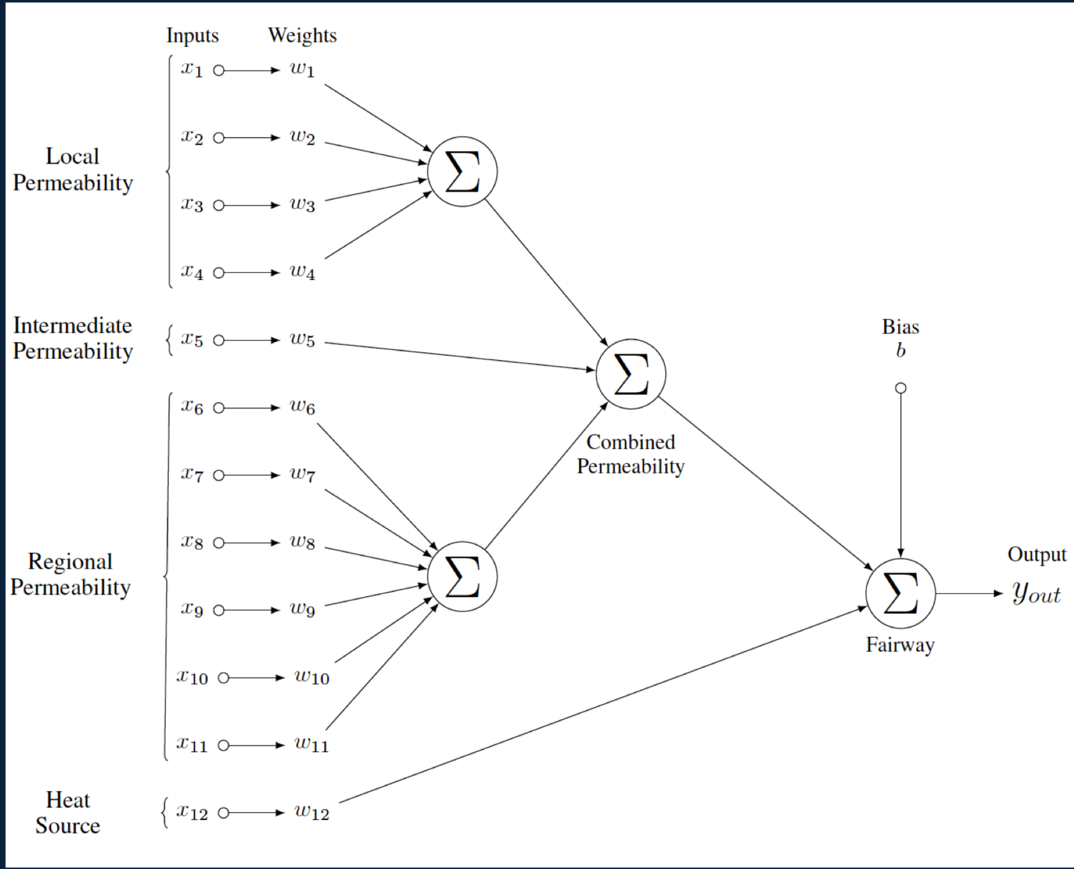






Uncertainty quantification for geothermal PFA

- Bayesian inference for the original PFA network ("Bayesian neural network")



Training data: $\mathcal{D} = \{(\boldsymbol{x}^n, y^n)\}_{n=1}^{N} = (\boldsymbol{X}, \boldsymbol{y})$

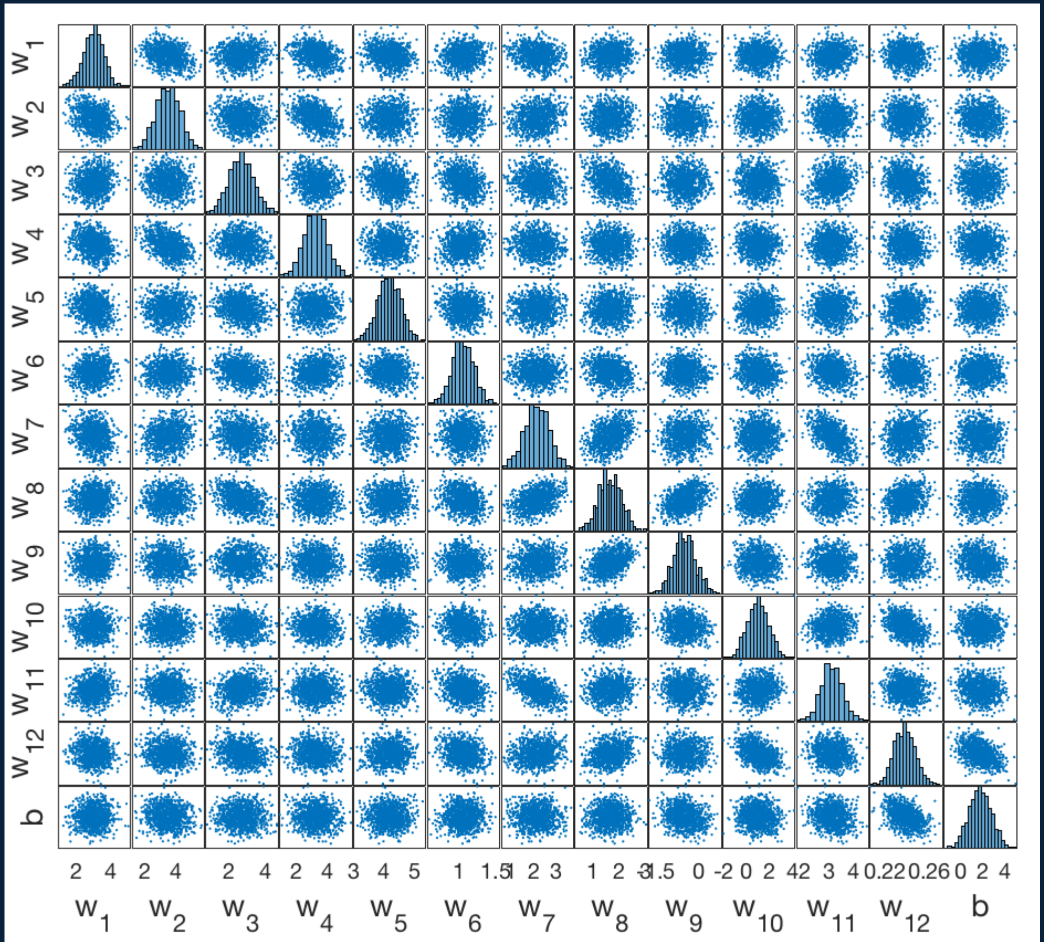
Forward model: $y = G\theta + \epsilon, \epsilon \sim \mathcal{N}(0, \Sigma_{obs})$

Parameters: $\theta \in \Theta$, weights of the network

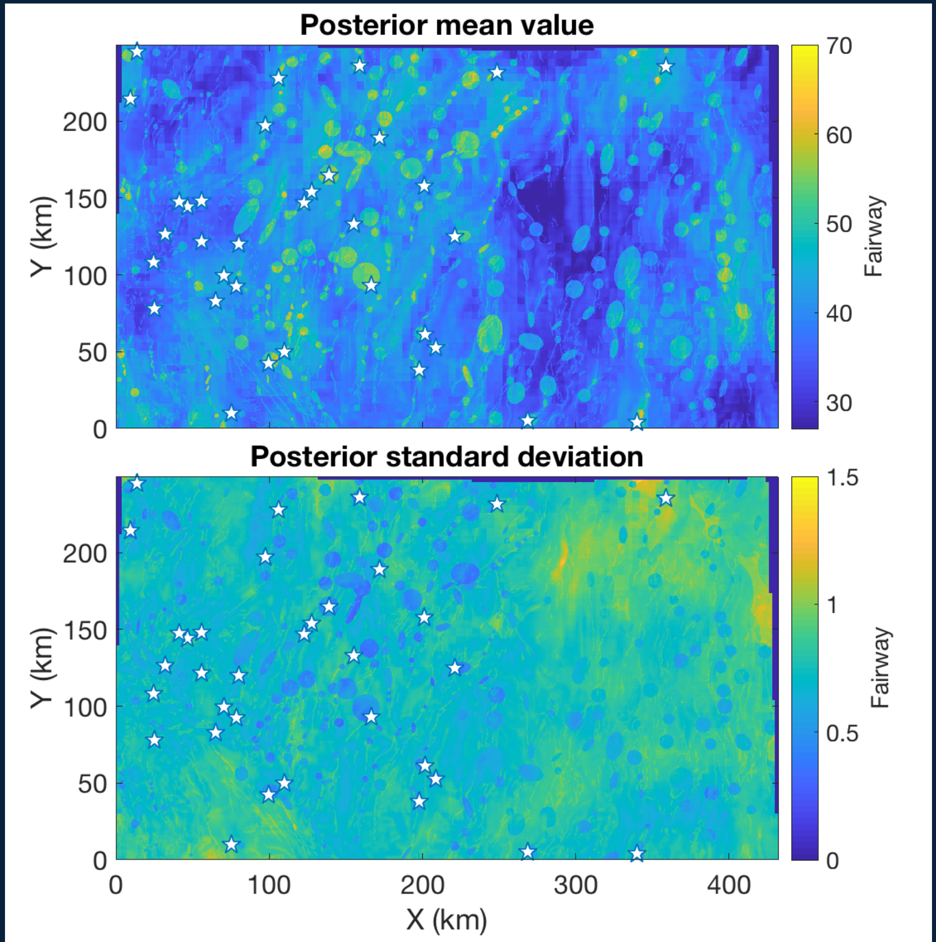
Prior	$p(\theta) = \exp\left(-\frac{1}{2}\theta^T \Sigma_{pr}^{-1}\theta\right)$
Posterior	$p(\theta\|\mathcal{D}) \propto p(\boldsymbol{y}\|X,\theta)p(\theta)$ $= \exp\left(-\frac{1}{2}(y - G\theta)^{\mathrm{T}}\Sigma_{obs}^{-1}(y - G\theta)\right)\exp\left(-\frac{1}{2}\theta^T\Sigma_{pr}^{-1}\theta\right)$
Prediction	$p(y'\|\mathcal{D}, x') = \int p(y'\|x', \theta)p(\theta\|\mathcal{D})d\theta$

Uncertainty quantification for geothermal PFA

Posterior distribution of parameters

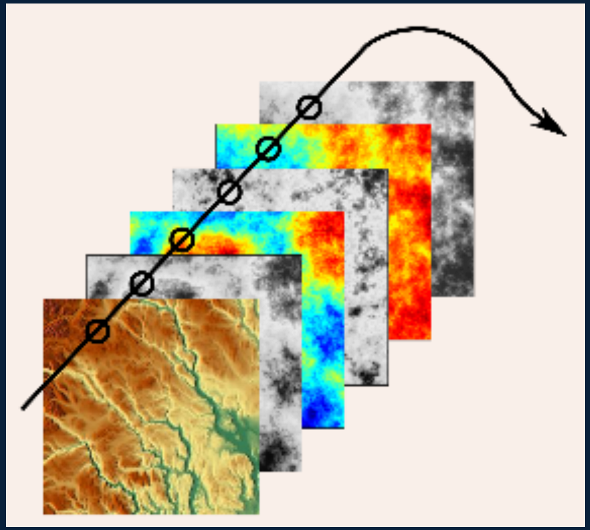


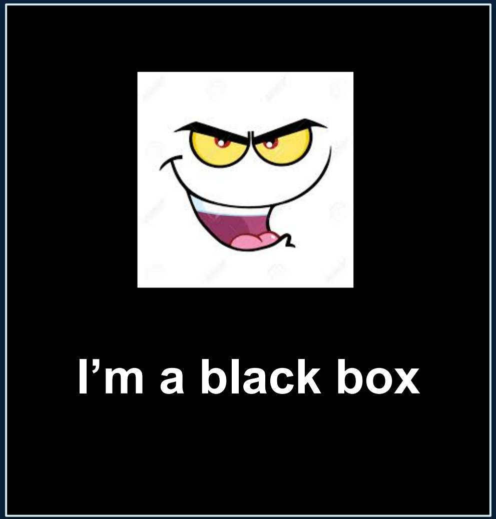
Posterior predictive fairway map

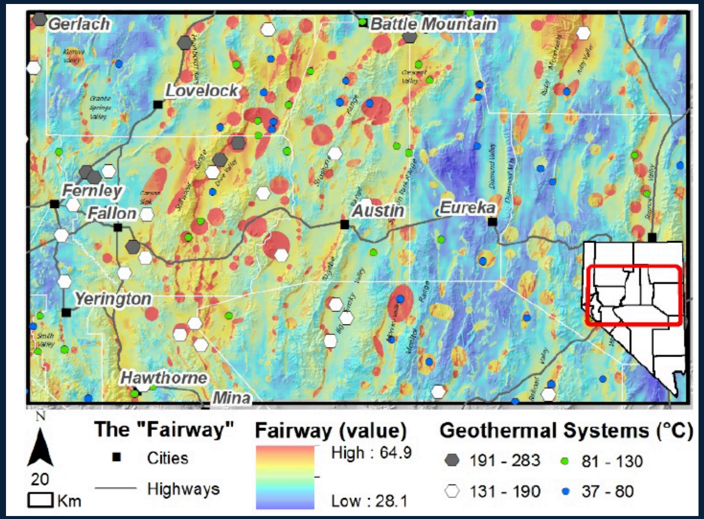


Multiple models from Steve…

Geo-X Input X → Unknown PFA Model *m* → Fairway Output Y



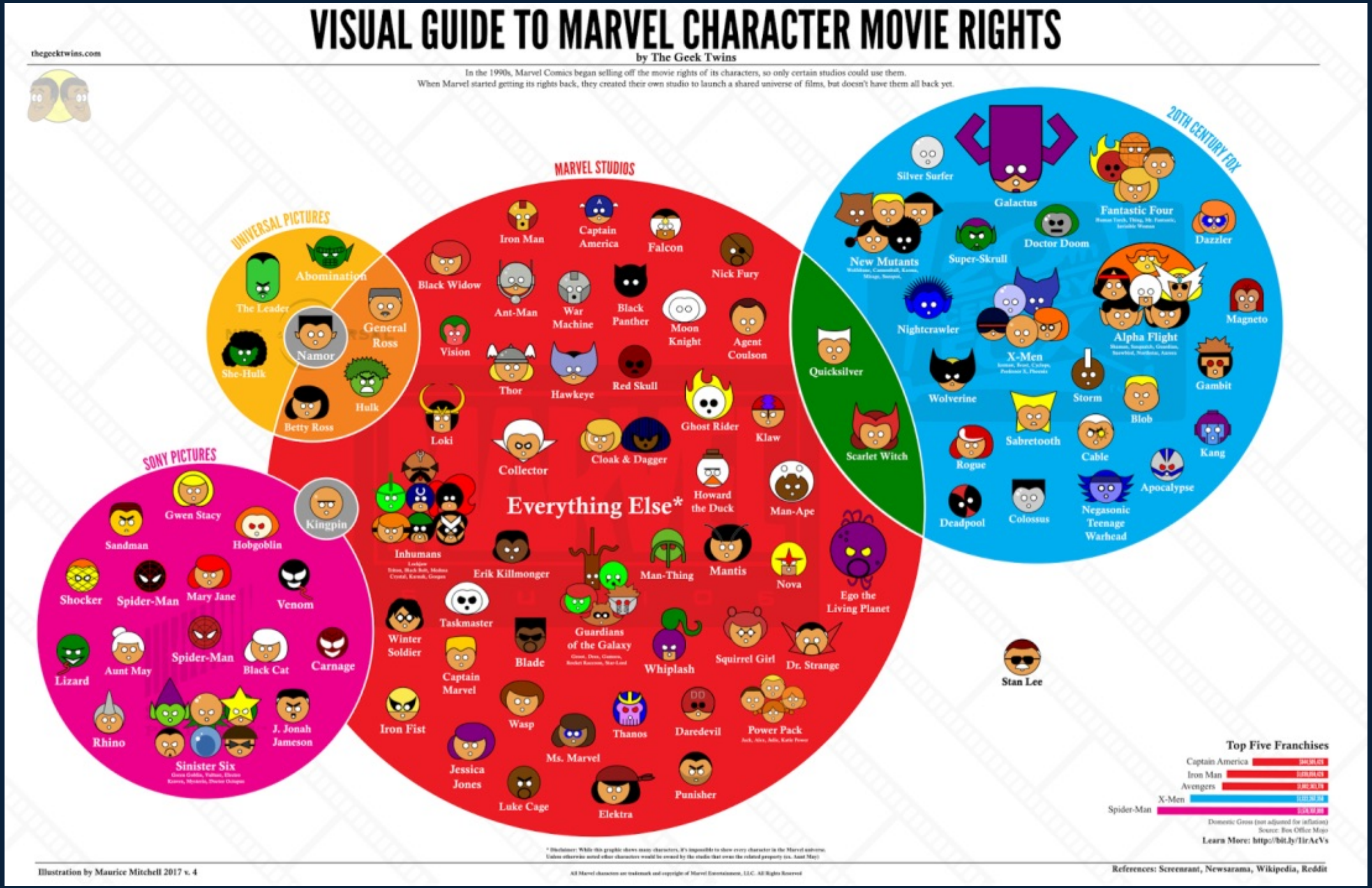




UQ in a multi-model universe

Q: Who's your favorite superhero in Marvel Universe?

A: It will be easier to tell if we can compare them in a same scale.



A big picture of model selection universe

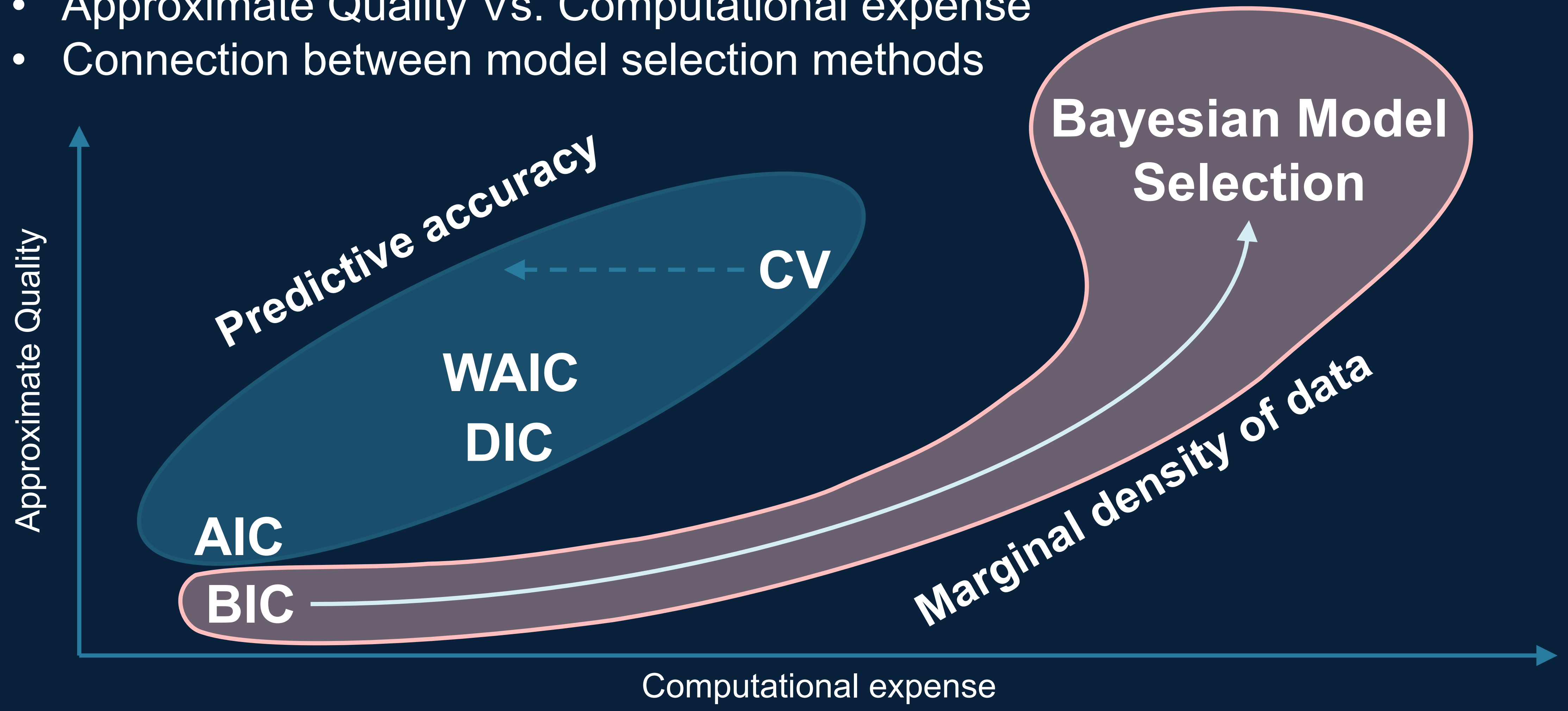
- Choose the best model from a set of candidate models $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$

$$\mathcal{M}_j = \{p(y; \theta_j) : \theta_j \in \Theta_j\}$$

Common method	Goodness of fit +	Penalty/ Bias correction
Akaike Information Criterion (AIC) **(DIC, WAIC)**	$\frac{1}{n}\sum_{i=1}^{n} \log p(y_i \mid \hat{\theta}_j)$	$-2d_j$
Leave-one-out Cross-Validation **(K-fold Cross-validation)**	$\sum_{i=1}^{n} \log p_{(-i)}(y_i \mid \hat{\theta}_j)$	$\sum_{i=1}^{n} \log p(y_i \mid \hat{\theta}_j) - \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} \log p_{(-k)}(y_i \mid \hat{\theta}_j)$
Bayesian Information Criterion (BIC) **(Bayesian Model Selection)**	$\sum_{i=1}^{n} \log p(y_i \mid \hat{\theta}_j)$	$-\frac{d_j}{2}\log n$

A big picture of model selection universe

- Approximate Quality Vs. Computational expense
- Connection between model selection methods



Why Bayesian model selection?

Reason 1:
Easy to understand

Reason 2:
Consistent

Reason 3:
Automatic Ockam's razors

Reason 4:
Conceptually same for multi-model selection problems

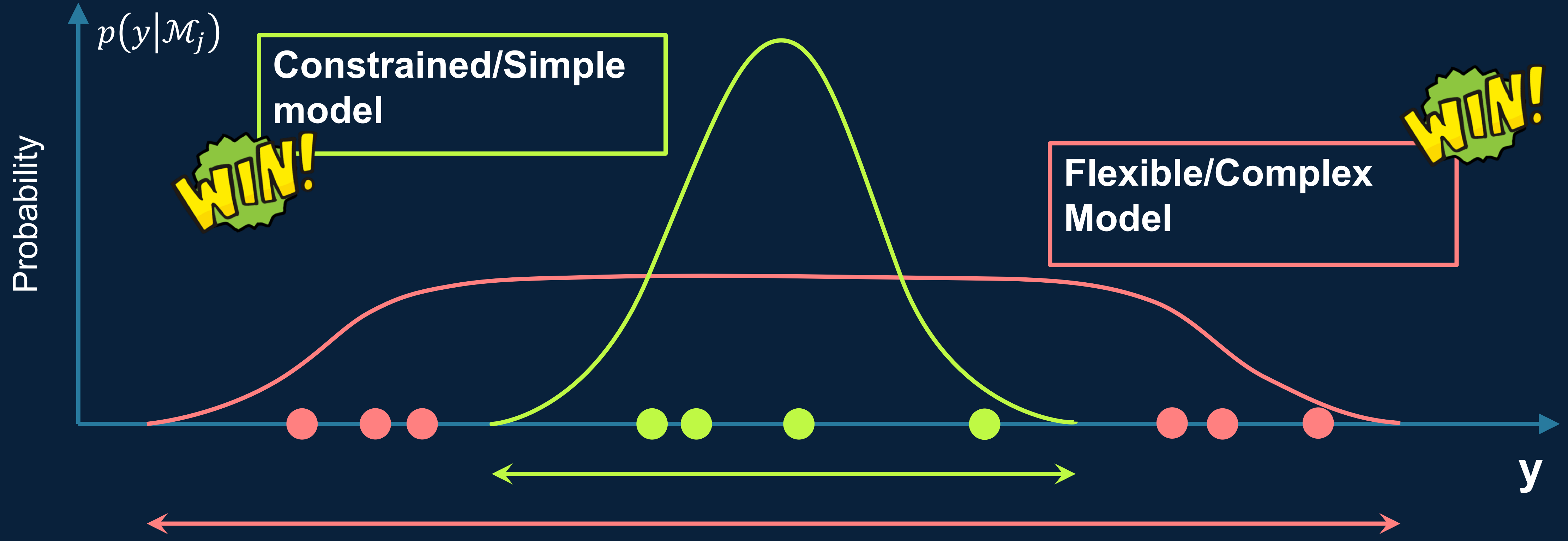
Reason 5:
Do not require nested models, std. distr., or regular asymptotics

Reason 6:
Can account for model uncertainty

Bayesian model selection – Basic concepts

- The marginal probability of data y given model $\mathcal{M}_j$

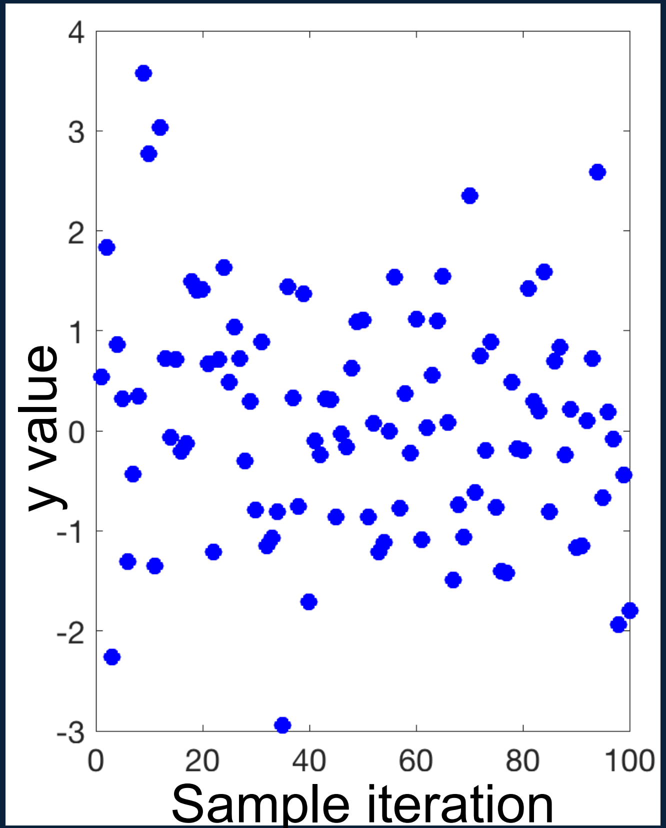
$$p(\mathcal{M}_j|y) \propto \boxed{p(y|\mathcal{M}_j)} = \int p(y|\theta_j, \mathcal{M}_j)p(\theta_j|\mathcal{M}_j)d\theta_j$$



Bayesian model selection – Basic concepts

- Bayes Factor (BF)

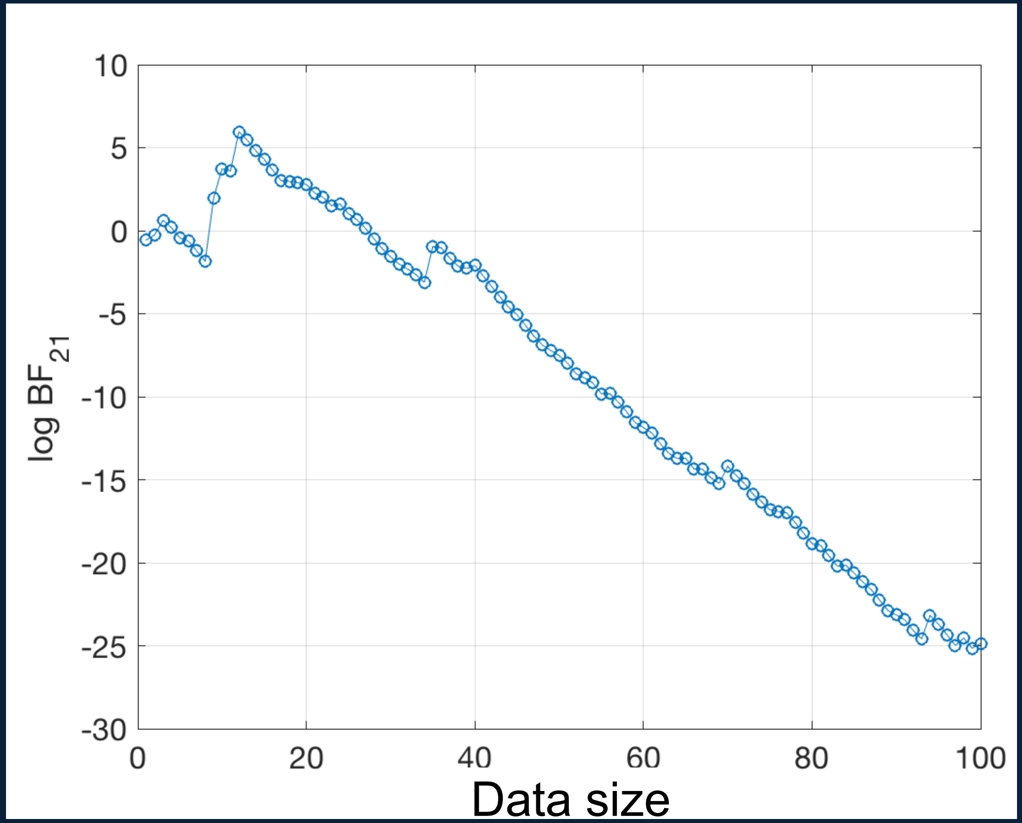
$$BF_{ji} = \frac{p(y|\mathcal{M}_j)}{p(y|\mathcal{M}_i)} = \frac{\int p(y|\theta_j,\mathcal{M}_j)p(\theta_j|\mathcal{M}_j)d\theta_j}{\int p(y|\theta_i,\mathcal{M}_i)p(\theta_i|\mathcal{M}_i)d\theta_i}$$



$\mathcal{M}_1 : y \sim \mathcal{N}(0,1)$

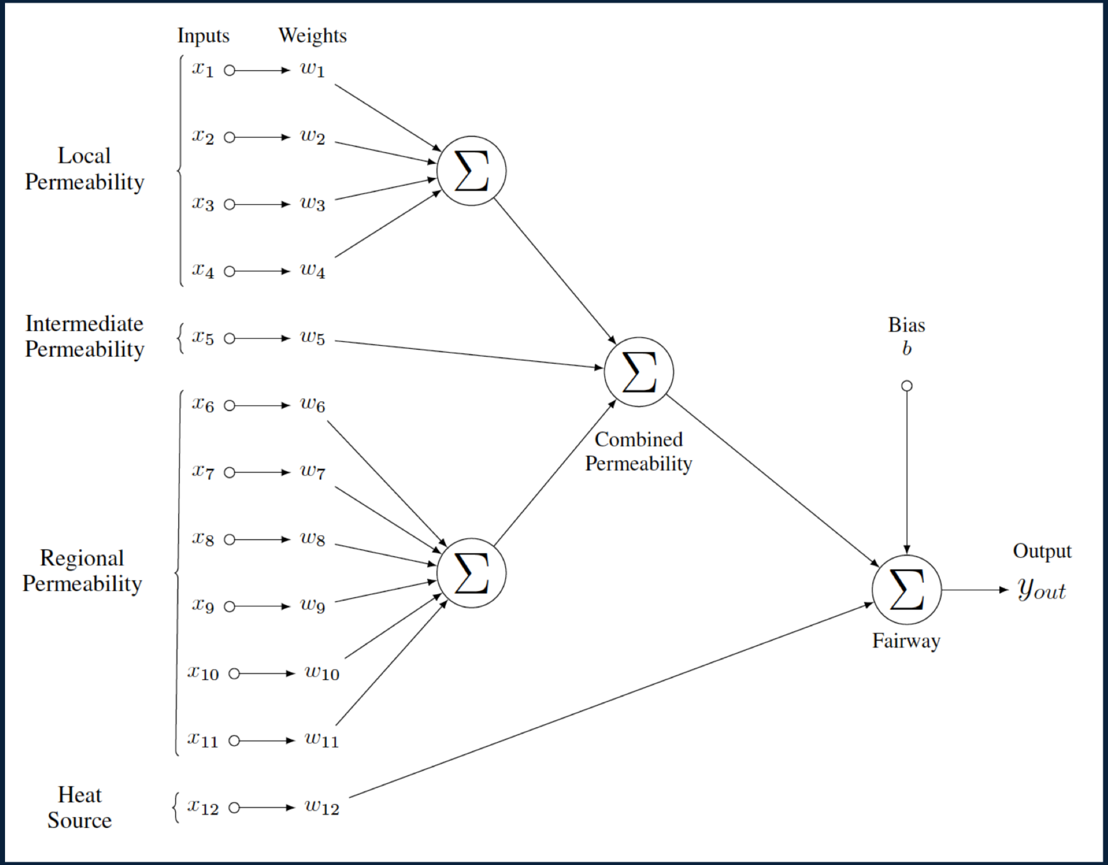
Vs.

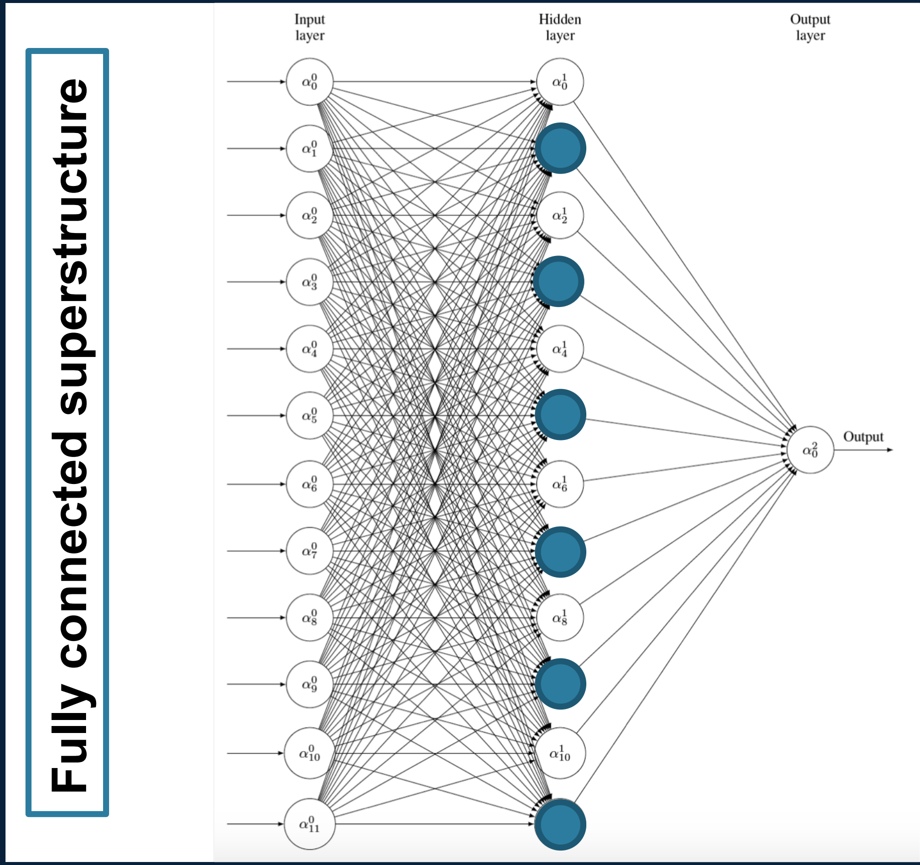
$\mathcal{M}_2 : y \sim \mathcal{N}(\mu,1)$



Application to Nevada geothermal development

$\mathcal{M}_1$: the original PFA network Vs. $\mathcal{M}_2$: a random sampled subnet





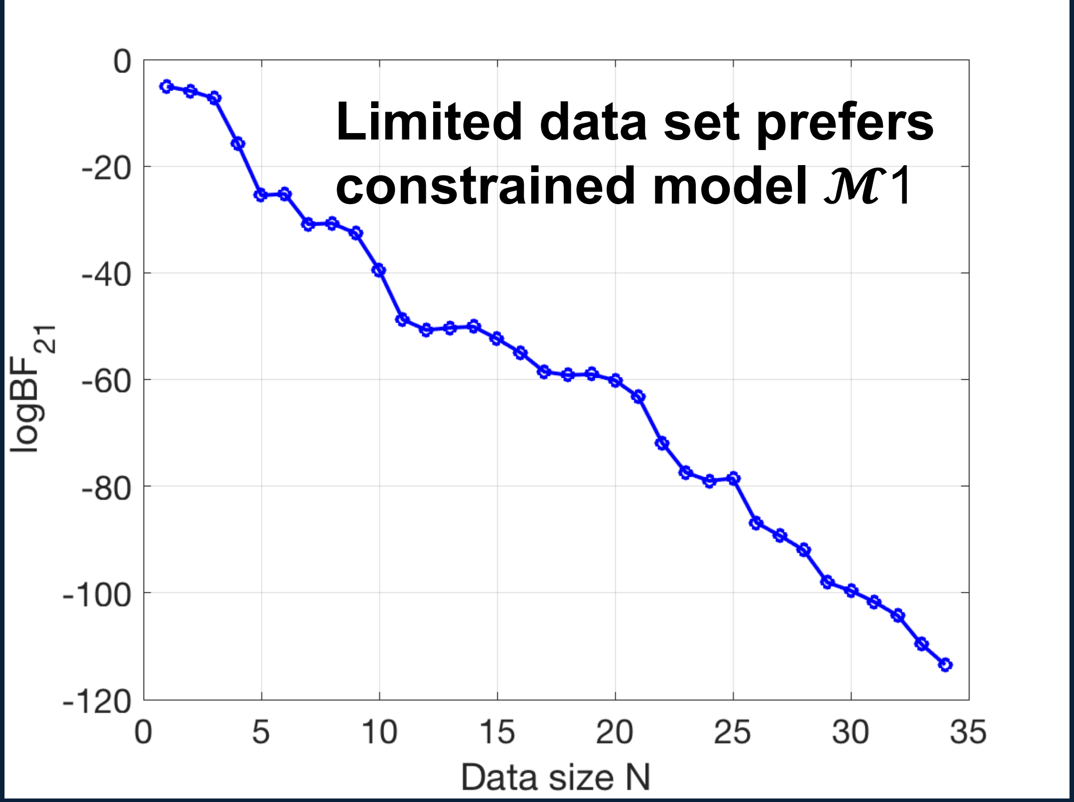
Application to Nevada geothermal development

$$BF_{21} = \frac{p(y|\mathcal{M}_2)}{p(y|\mathcal{M}_1)} = \frac{\int p(y|\theta_2, \mathcal{M}_2)p(\theta_2|\mathcal{M}_2)d\theta_2}{\int p(y|\theta_1, \mathcal{M}_1)p(\theta_1|\mathcal{M}_1)d\theta_1}$$

$\mathcal{M}_1$: the original PFA network (13 parameters)

Vs.

$\mathcal{M}_2$: a random sampled subnet (241 parameters)



Conclusion

- The uncertainty quantification (UQ) of the geothermal play fairway analysis (PFA) is important to increase the exploration success rate and reduce the risk.
- Bayesian model selection calculates marginal probability of data of different models and therefore can compare models on the same scale; it selects the best model within the candidate models, and quantify model uncertainties.
- This work applies Bayesian model selection method to compare multiple geothermal prediction models and quantify the uncertainties.

Earth Resources Laboratory

Thank you!

Email: guchch@mit.edu

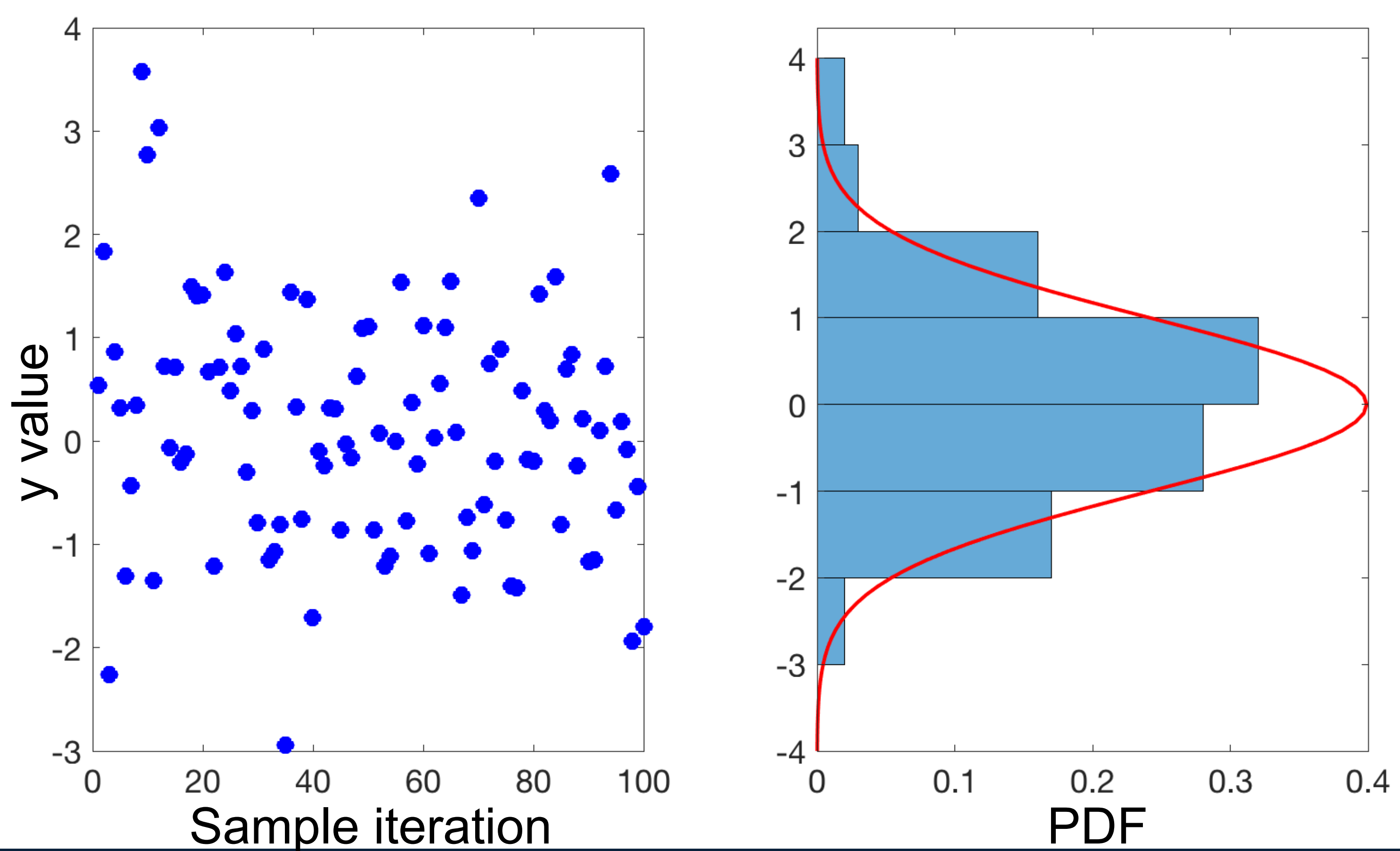
Conclusion

- The uncertainty quantification (UQ) of the geothermal play fairway analysis (PFA) is important to increase the exploration success rate and reduce the risk.

- Bayesian model selection calculates marginal probability of data of different models and therefore can compare models on the same scale; it selects the best model within the candidate models, and quantify model uncertainties.

- This work applies Bayesian model selection method to compare multiple geothermal prediction models and quantify the uncertainties.

Bayesian model selection – Normal mean



$\mathcal{M}_1: y \sim \mathcal{N}(0,1)$

Vs.

$\mathcal{M}_2: y \sim \mathcal{N}(\mu, 1)$

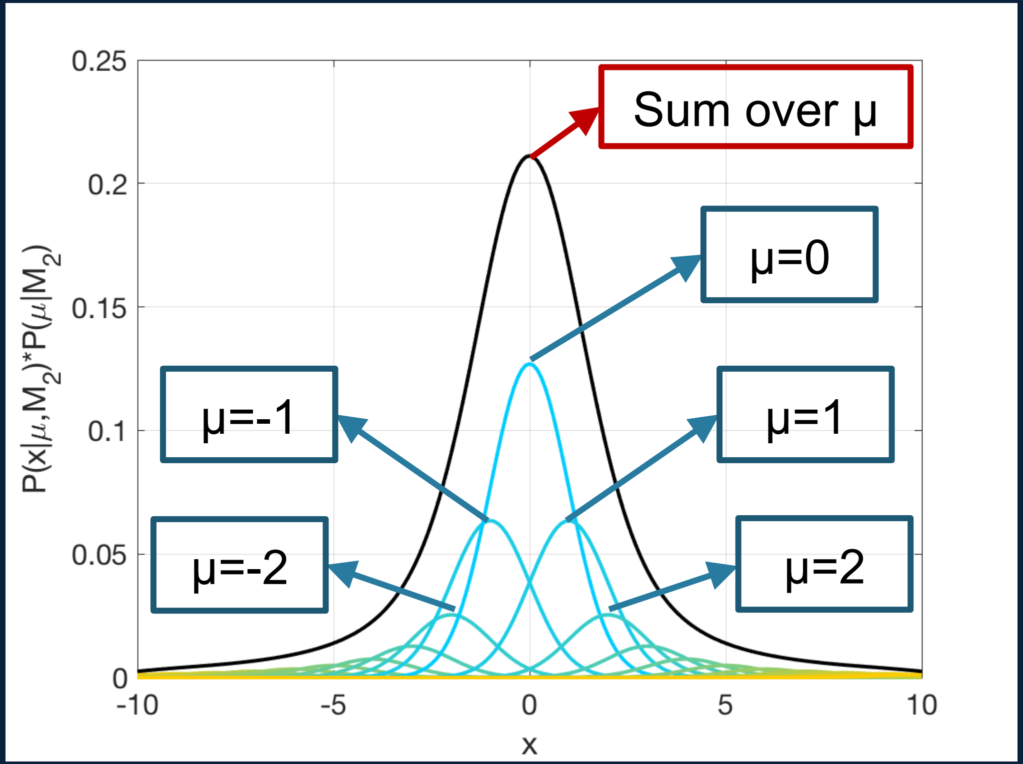
Bayesian model selection – Normal mean

$$BF_{21} = \frac{p(y|\mathcal{M}_2)}{p(y|\mathcal{M}_1)} = \frac{\int p(y|\mu, \mathcal{M}_2)p(\mu|\mathcal{M}_2)d\mu}{p(y|\mathcal{M}_1)}$$

$\mathcal{M}_1: y \sim \mathcal{N}(0,1)$

Vs.

$\mathcal{M}_2: y \sim \mathcal{N}(\mu, 1)$



Bayesian model selection – Normal mean

$$BF_{21} = \frac{p(y|\mathcal{M}_2)}{p(y|\mathcal{M}_1)} = \frac{\int p(y|\mu,\mathcal{M}_2)p(\mu|\mathcal{M}_2)d\mu}{p(y|\mathcal{M}_1)}$$

$\mathcal{M}_1 : y \sim \mathcal{N}(0,1)$

Vs.

$\mathcal{M}_2 : y \sim \mathcal{N}(\mu,1)$