# Using Probabilistic Machine Learning to Estimate Ocean Mixed Layer Depth

**Recovery from sparse in-situ observations informed from satellite data.**

***Dallas Foster***
*Oregon State University*
*Advisors: David John Gagne, Daniel Whitt*

NCAR

**July 29th 2020**

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

# MOTIVATION

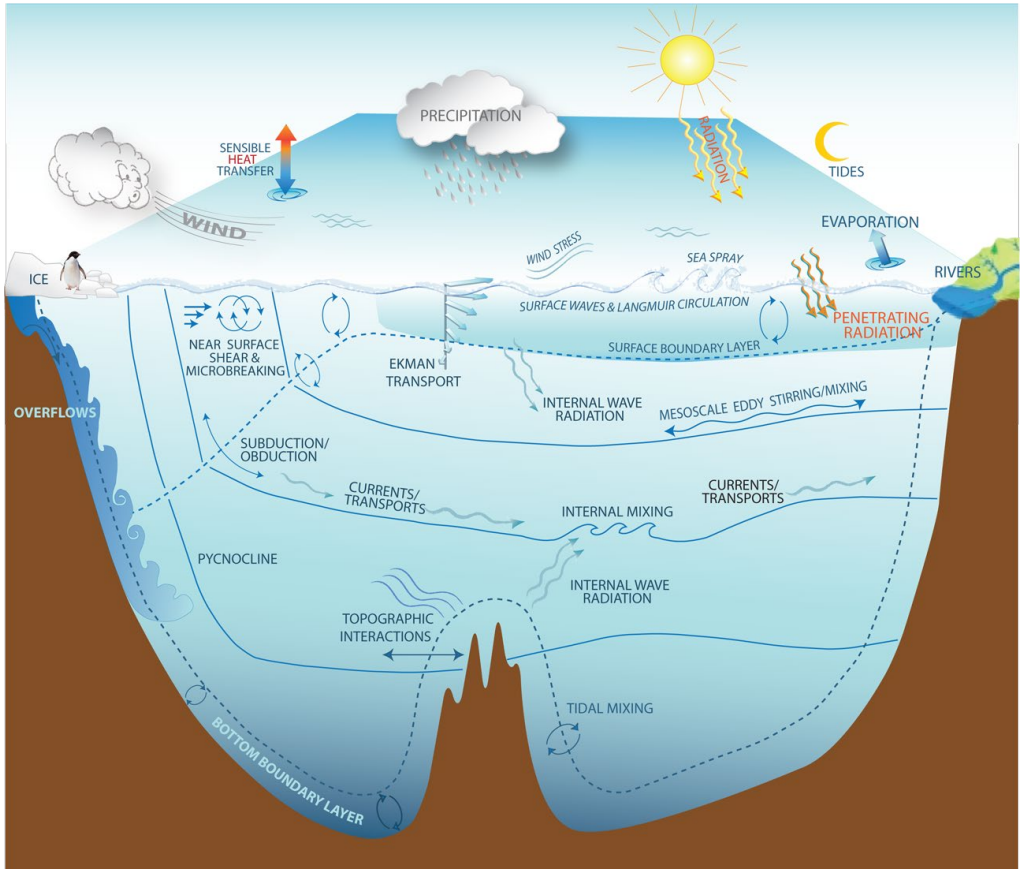
**Ocean and Earth system** processes are **highly sensitive to** ocean surface **mixed layer depth (MLD)**

- water mass formation and circulation
- air-sea exchange
- Biogeochemistry

**Observational data** is increasingly available, but still **relatively sparse**

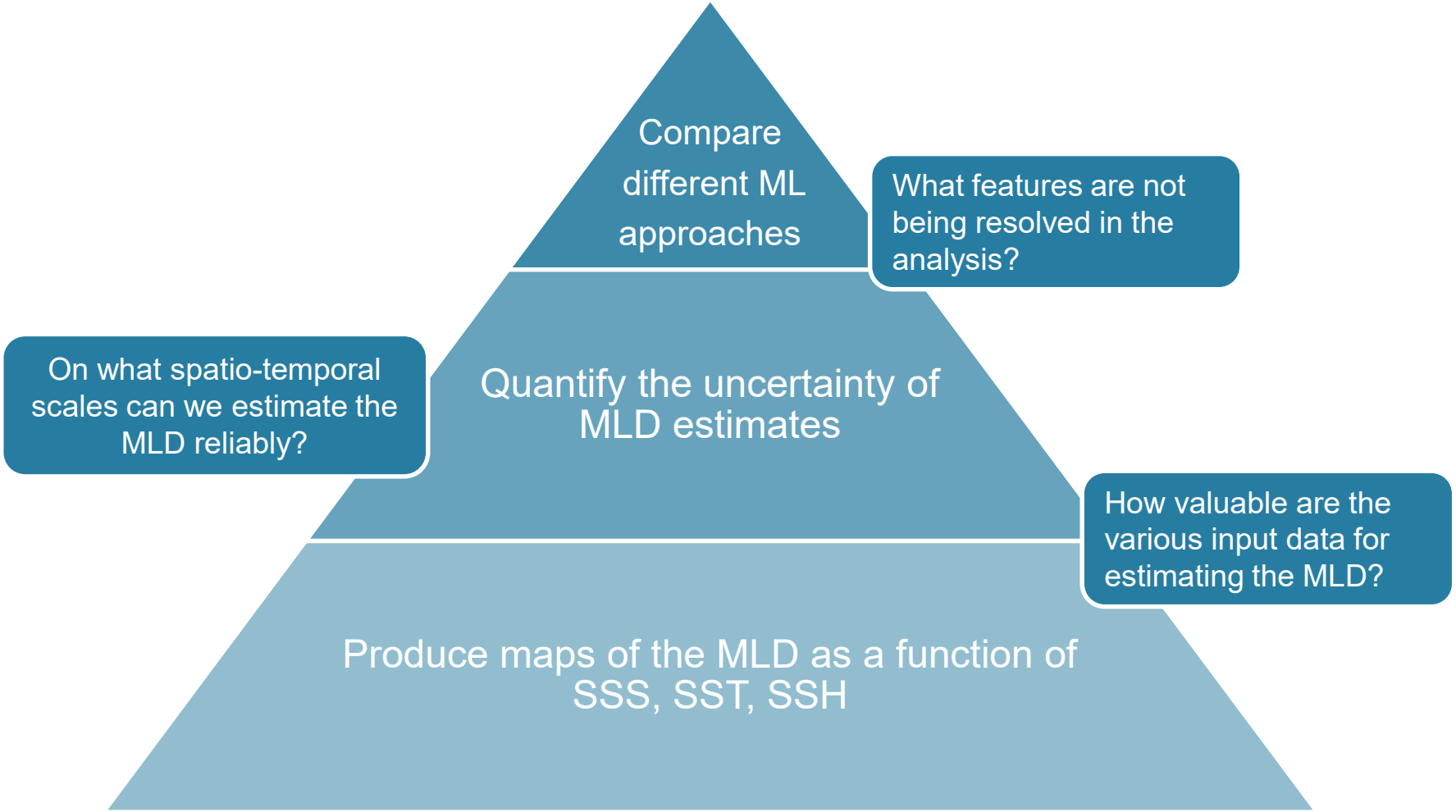
**Existing methods** perform optimal interpolation, but **do not inform with satellite sea surface data**.

Want to quantify the **sub-seasonal** relationship between **Sea Surface Salinity (SSS)**, **Temperature (SST)**, **Sea Level Height Anomaly (SSH)** and **MLD**

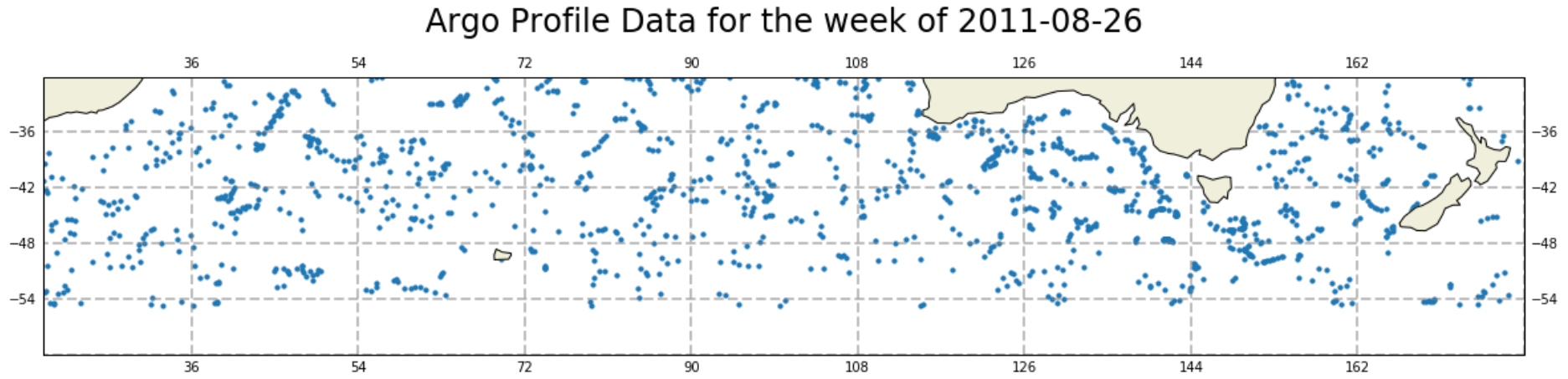



https://www.gfdl.noaa.gov/oceanprocschem/

# GOALS AND QUESTIONS

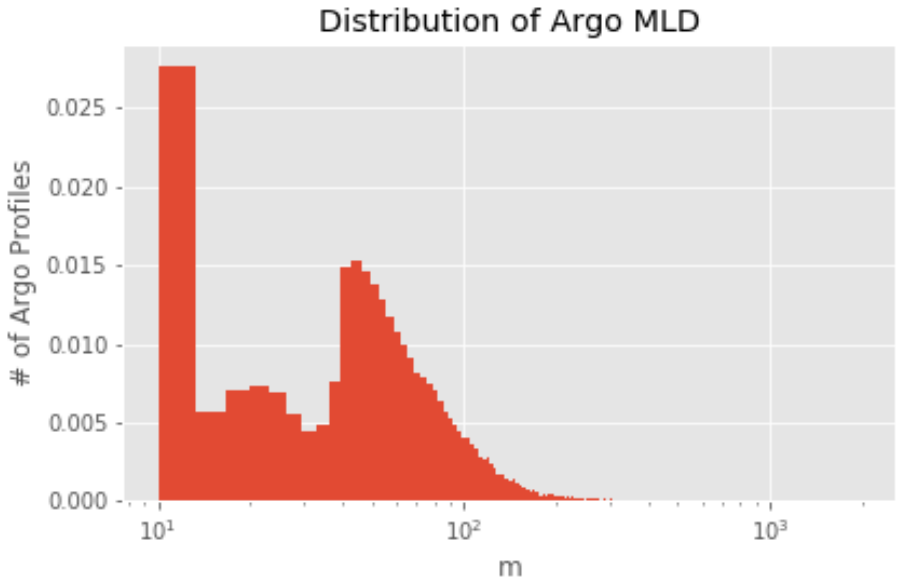


# DATA




Argo float MLD data is:

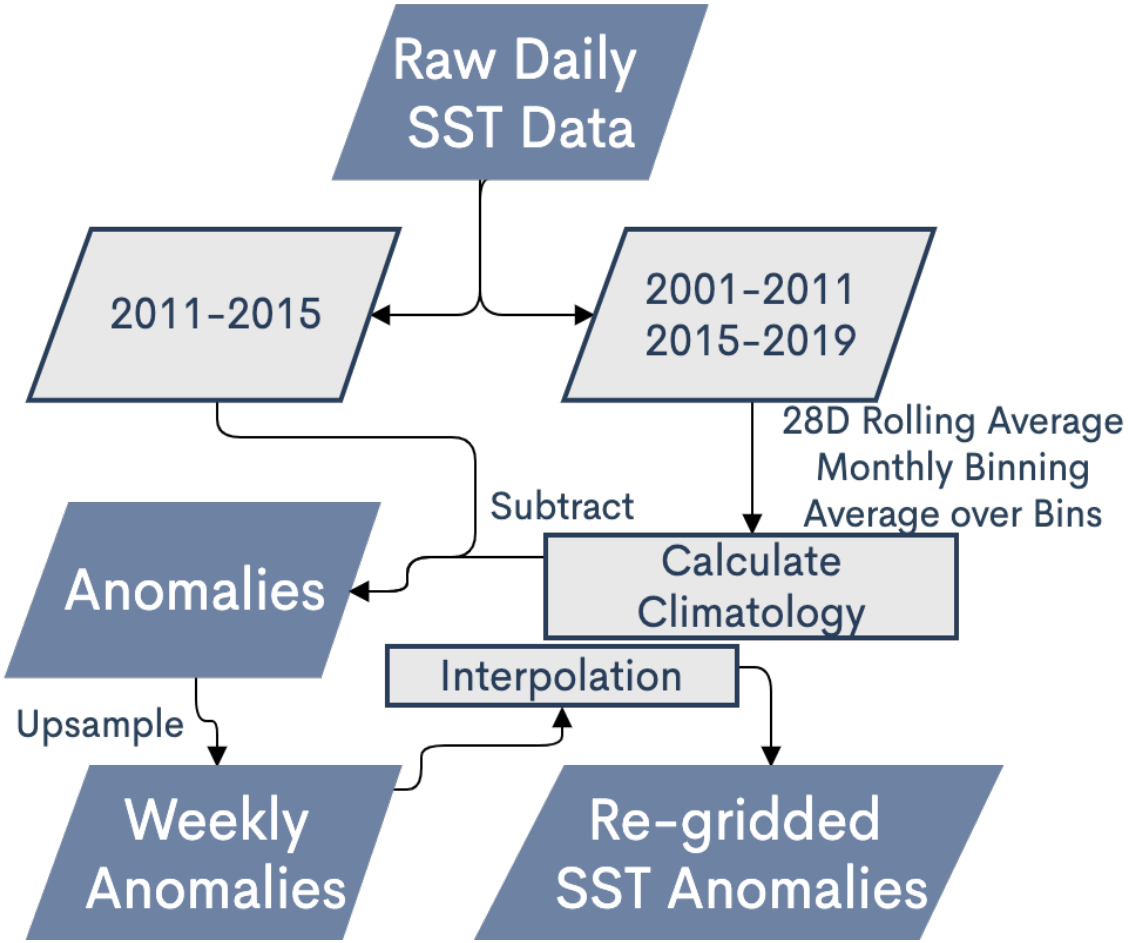
- **Sparse**
- **Non-normal**



# DATA PREPROCESSING

**Preprocessing Steps:**

1. **Divide Data**
   - Validate divisions
2. **Calculate Climatology**
   - Apply rolling average
   - Bin data into months
   - Average over bins
3. **Calculate Anomalies**
   - Bin data into months
   - Subtract binned climatology
   - Remove diurnal cycle
4. **Resample**
5. **Interpolate**



# TERMINOLOGY

| Terminology | Interpretation |
|---|---|
| **Machine Learning** | A framework of building and **fitting nonlinear models to data.** |
| **Uncertainty Quantification** | Techniques to determine how likely certain outcomes are if some aspects of the system are not exactly known. |
| $S, T, H, x, y$ | Variables **SST, SSH, SHA** and the 2-dimensional spatial coordinates. |
| $d_{obs}, d, \sigma, \Sigma$ | MLD **observations** (sparse grid), **estimates**, **uncertainties** (full grid). |
| $\theta, p(\theta)$ | Model **parameters** and probability distribution (**prior distribution**). |
| $p(d_{obs}\|d, \theta)$<br>$p(d\|d_{obs}, \theta)$ | Conditional **marginal likelihood** and **posterior distributions**. Represent the relative likelihood of the observations (estimates) given estimates (observations) and parameters. |

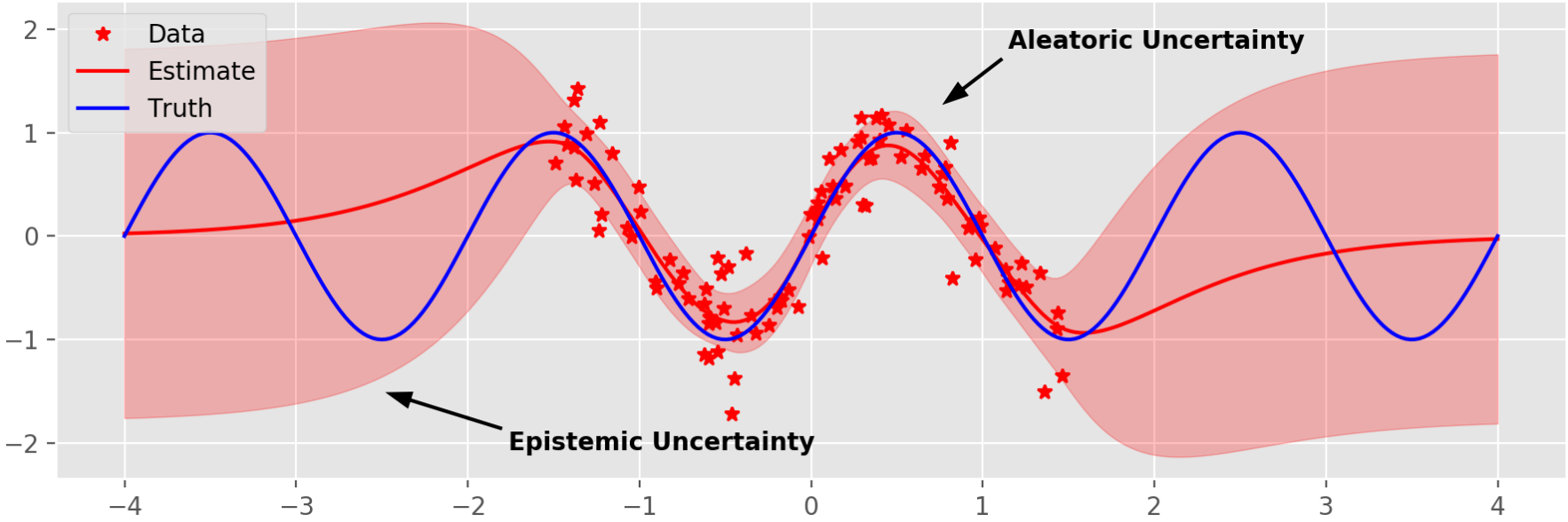
# UNCERTAINTY QUANTIFICATION

## Aleatoric Uncertainty

- Inherent noise in data
- Irreducible

## Epistemic Uncertainty

- Lack of knowledge, data
- Deficiency of model




Model must account for **aleatoric** and **epistemic** uncertainty

- **Monte Carlo** sampling of model
- **Bayesian** interpretation of model weights
- Specification of **noise model**

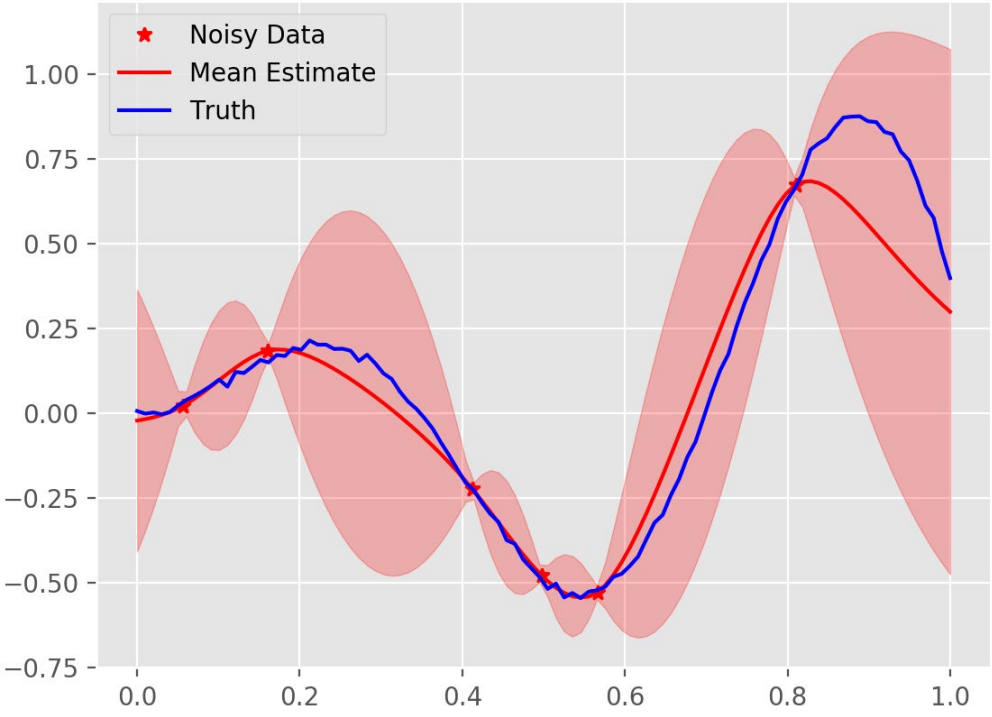
# GAUSSIAN PROCESS

A **Gaussian Process (GP)** $y$, observed at points $x$ is a sample from a **multivariate normal distribution**,

$$y(x) \sim N(0, K(x, x'))$$

$K$ is a **covariance function** that specifies the **spatial relationships** between points

Allows us to **predict the mean and variance** of $y$ at new points $x_*$



# MODEL FRAMEWORK

**Modeling Steps**:

1. **Generate dense MLD field**

$$d = f(S, T, H) + \epsilon$$

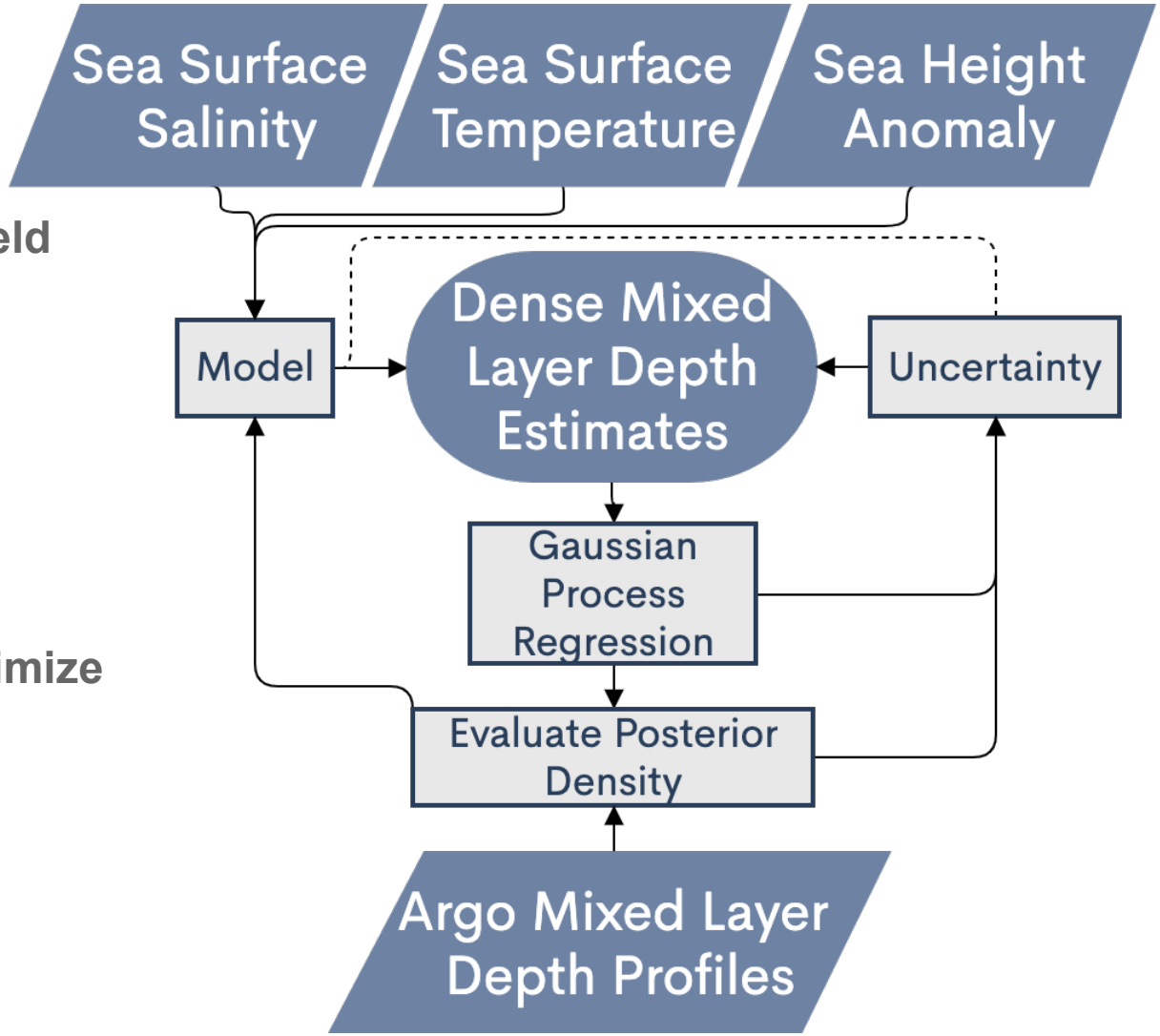
2. **GPR to sparse field**

$$d_{obs} = L \cdot d + \sigma,$$
$$\sigma \sim N(0, \Sigma)$$

3. **Compute loss and maximize weights according to**

$$p(d|d_{obs}, \theta)$$

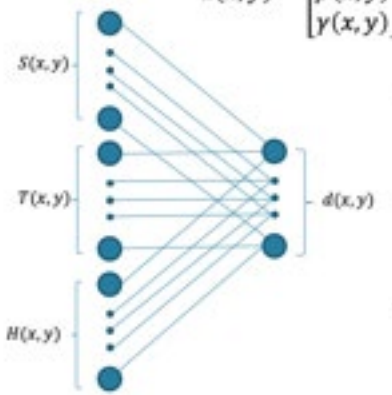
4. **Repeat**



# TRADITIONAL MODELS

## LINEAR MODEL

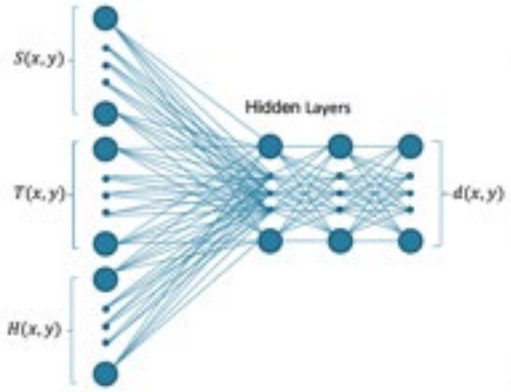
$$d(x,y) = \begin{bmatrix} \alpha(x,y) \\ \beta(x,y) \\ \gamma(x,y) \end{bmatrix} \cdot \begin{bmatrix} S(x,y) \\ T(x,y) \\ H(x,y) \end{bmatrix} + c(x,y) + \epsilon(x,y)$$




- Little to **no spatial correlation** between grid points
- Relatively **few parameters** required
- Trade off between **performance** and **overfitting**



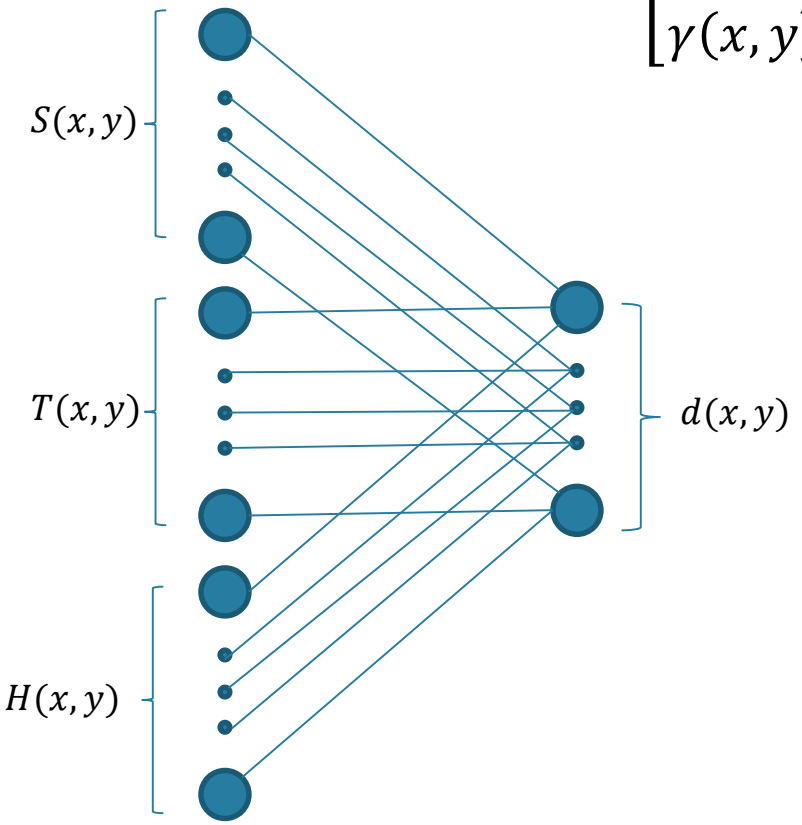
## FEED FORWARD ARTIFICIAL NEURAL NETWORK




- Universal function approximator
- Comprised of a **series of simple nonlinear functions**
  $h_i = f(Ah_{i-1} + b)$
- Weights are updated through **backpropagation**
- **Surplus of parameters**

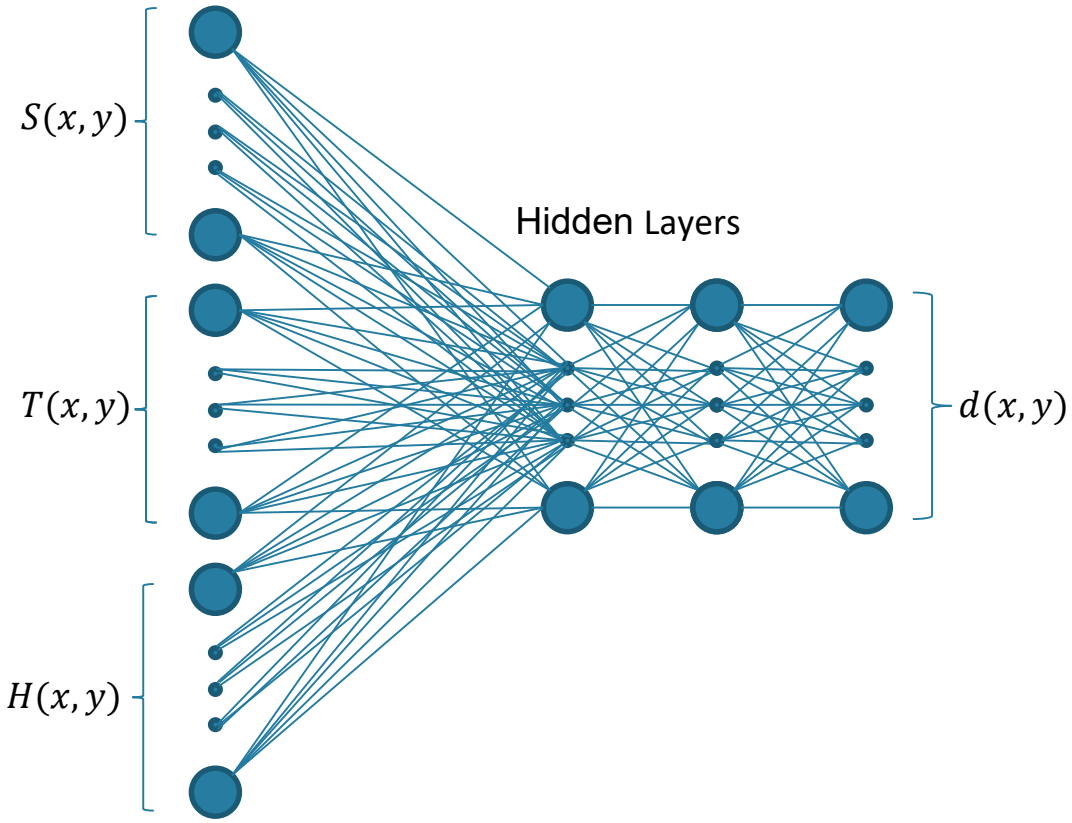
# LINEAR MODEL

$$d(x,y) = \begin{bmatrix} \alpha(x,y) \\ \beta(x,y) \\ \gamma(x,y) \end{bmatrix} \cdot \begin{bmatrix} S(x,y) \\ T(x,y) \\ H(x,y) \end{bmatrix} + c(x,y) + \epsilon(x,y)$$




- Little to **no spatial correlation** between grid points
- Relatively **few parameters** required
- Trade off between **performance** and **overfitting**

# FEED FORWARD ARTIFICIAL NEURAL NETWORK



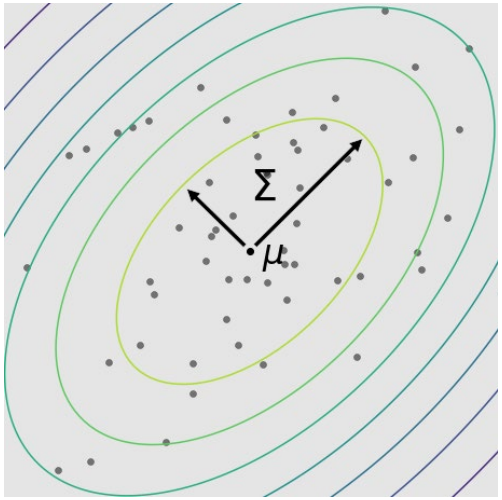

- Universal function approximator
- Comprised of a **series of simple nonlinear functions**

  $h_i = f(Ah_{i-1} + b)$
- **Surplus of parameters**

# PROBABILISTIC METHODS

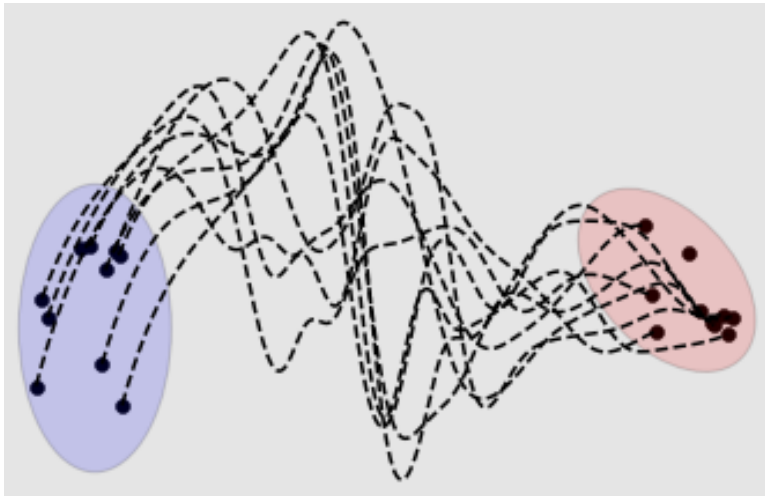
## Parameterization Methods

- Have to make a decision about the output's distribution
- Simple to implement, when possible
- Examples:
  - Least Squares Regression
  - Variational Neural Networks



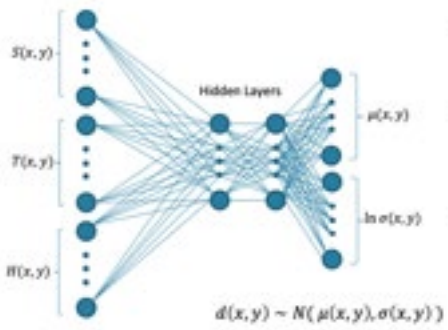

## Sampling Methods

- Initial distribution must be supplied
- Model must be run many times
- Examples:
  - Dropout
  - Variational AutoEncoders
  - Bayesian Neural Networks



# PROBABILISTIC METHODS

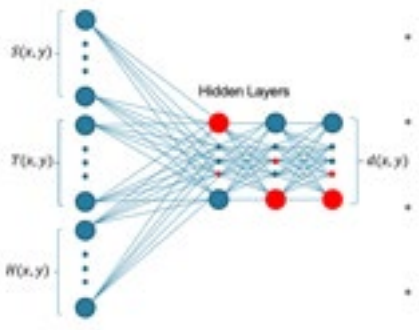
## VARIATIONAL NEURAL NETWORK




- Requires even **more parameters**
- Requires **parameterization of noise model**
- Better **captures aleatoric uncertainty.**



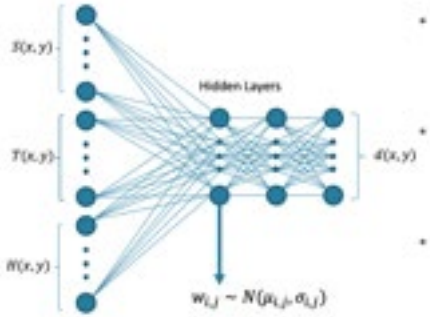
## DROPOUT




- **Randomly** set some weights to zero
- Creates an **ensemble of models**
- Computationally **inexpensive**
- **Requires sampling** to generate statistics



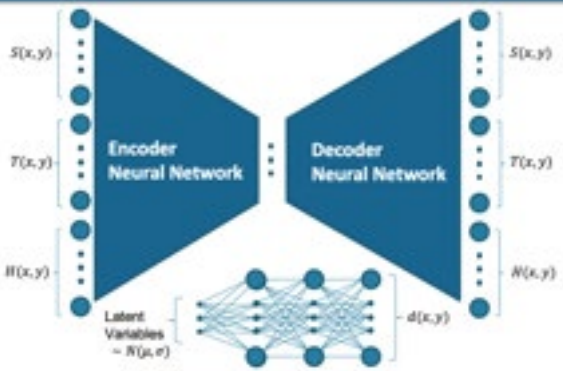
## BAYESIAN NEURAL NETWORK (FLIPOUT)




- Parameterizes a **prior** noise model **for each weight**
- Requires **double parameters** and Monte Carlo **sampling**
- Can help capture **epistemic uncertainty**

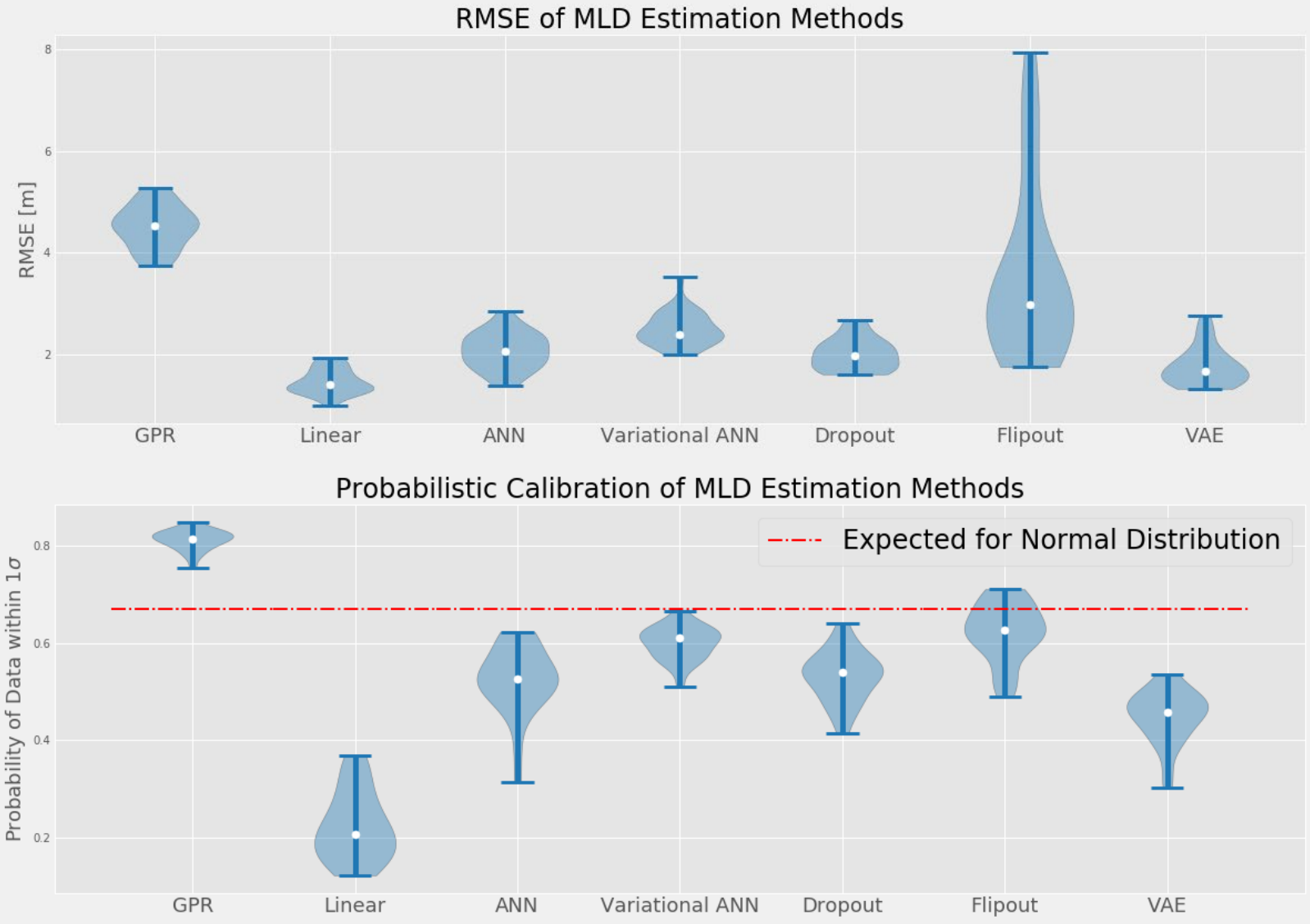


## VARIATIONAL AUTO ENCODERS

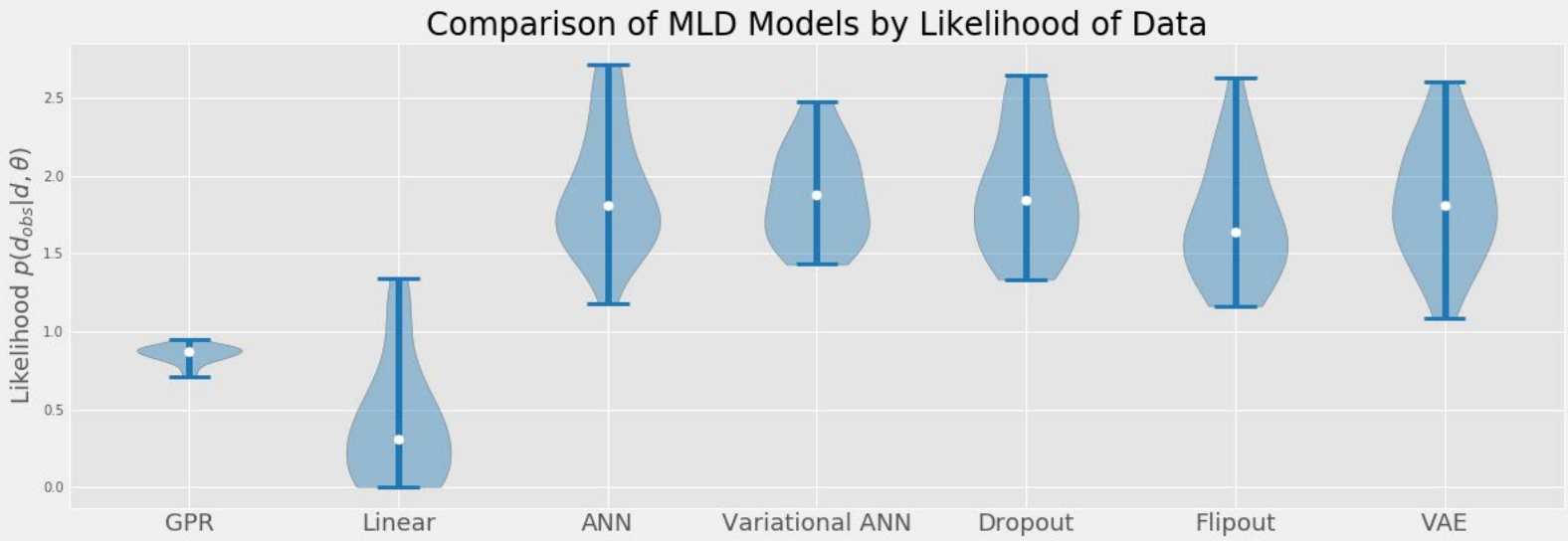


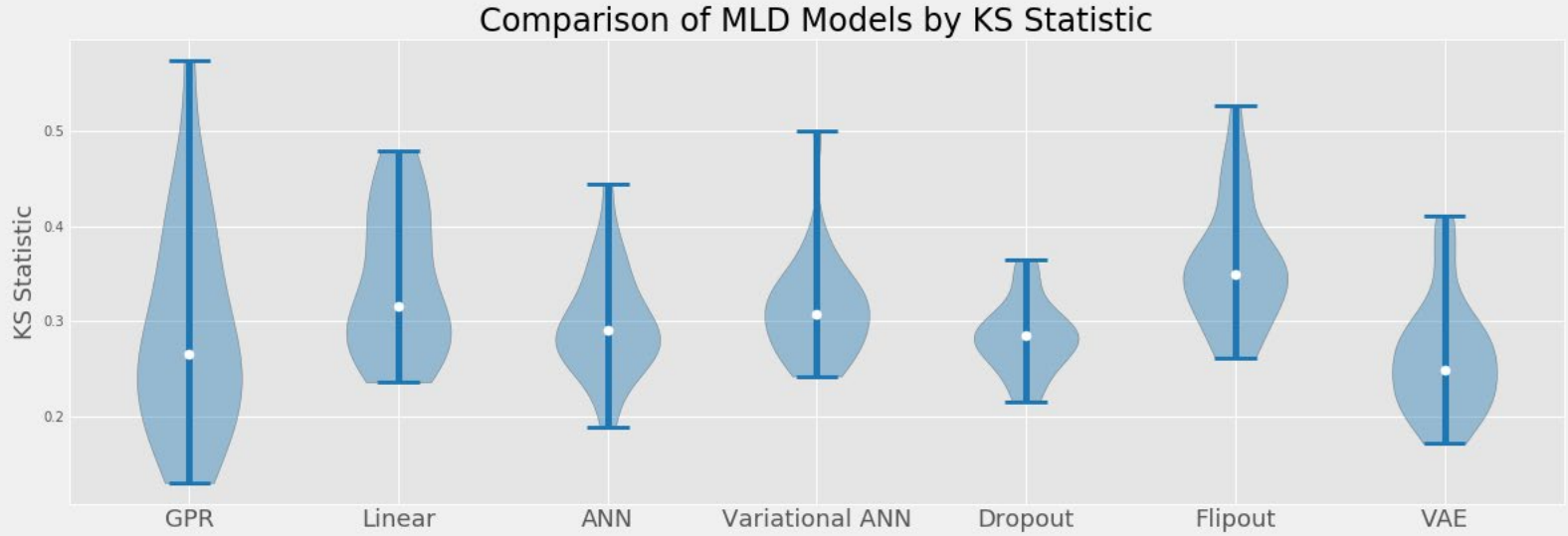

- Learns a **dimension reduction** of input data
- Gaussian **noise parameterization**

# RESULTS


RMSE of MLD Estimation Methods
RMSE [m]
8
6
4
2
GPR
Linear
ANN
Variational ANN
Dropout
Flipout
VAE
Probabilistic Calibration of MLD Estimation Methods
Probability of Data within 1σ
0.8
0.6
0.4
0.2
Expected for Normal Distribution
GPR
Linear
ANN
Variational ANN
Dropout
Flipout
VAE

# RESULTS


Comparison of MLD Models by Likelihood of Data
Likelihood $p(d_{obs}|d,\theta)$
2.5
2.0
1.5
1.0
0.5
0.0
GPR
Linear
ANN
Variational ANN
Dropout
Flipout
VAE


Comparison of MLD Models by KS Statistic
KS Statistic
0.5
0.4
0.3
0.2
GPR
Linear
ANN
Variational ANN
Dropout
Flipout
VAE

# ISSUES AND FUTURE WORK

**Model Development**

- Machine Learning models need further training
- Evaluate different parameterization of VNN, Flipout, VAE models
- Train and evaluate global models

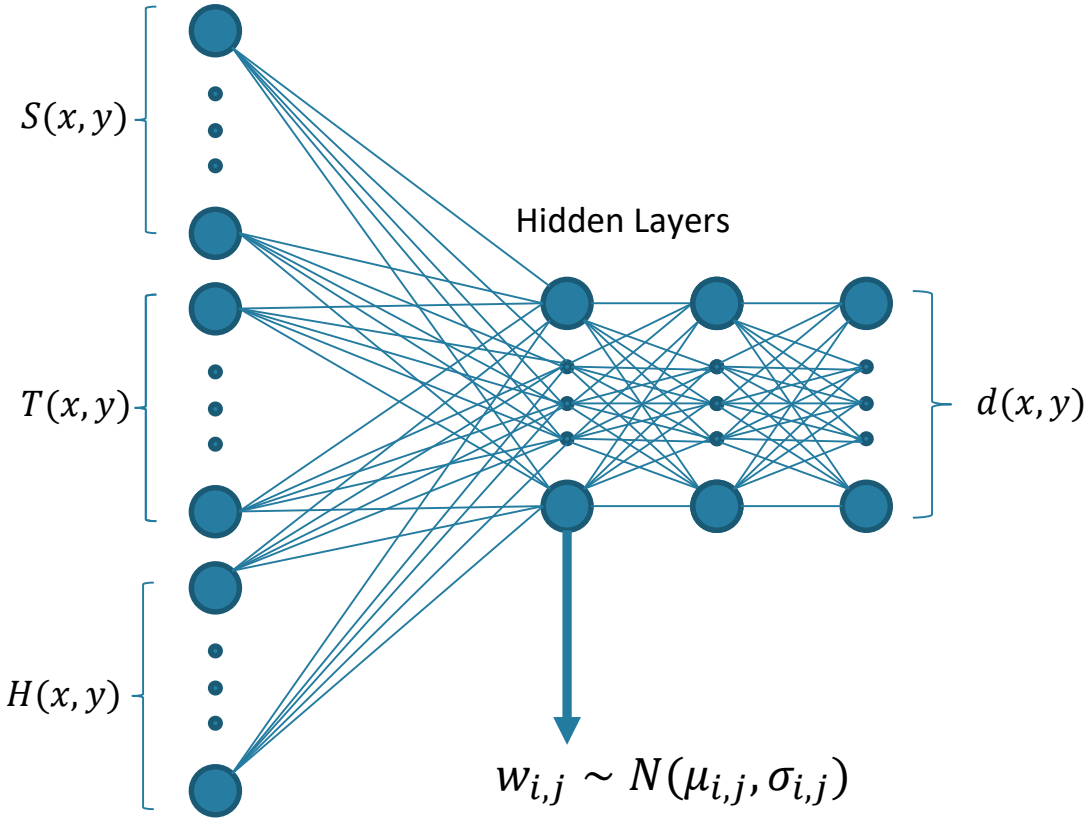
**Analysis**

- Further estimate spatial resolution and accuracy of models
- Investigate temporal relationships, predictability
- Build framework for optimal assimilation of model and data

# CONCLUSION

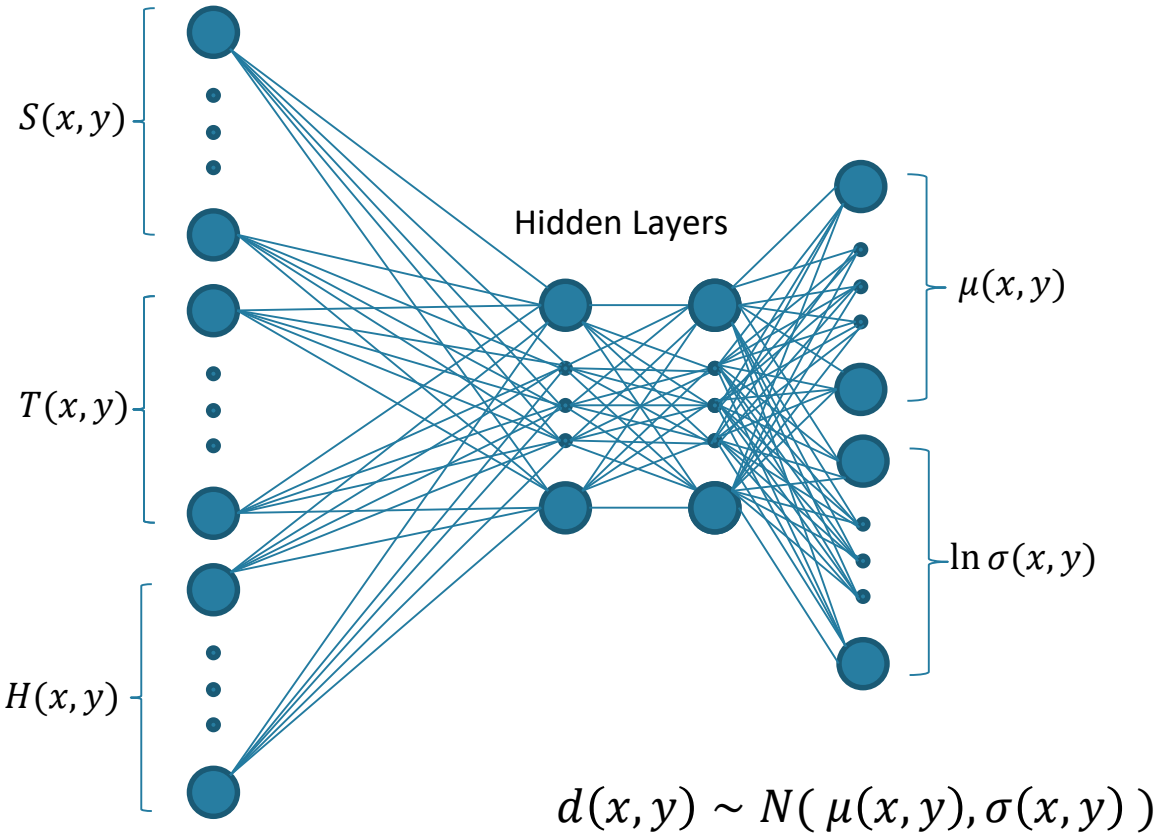
- Useful information can be extracted from surface data to estimate ocean mixed layer depths anomalies (MLD).
- Machine learning models are a promising approach to constructing models for estimating MLD.
- Simple noise parameterizations might be all that is necessary to get decent probabilistic estimates.
  - More analysis is needed!

# BAYESIAN NEURAL NETWORK (FLIPOUT)




- Parameterizes a **prior** noise model **for each weight**
- Requires **double parameters** and Monte Carlo **sampling**
- Can help capture **epistemic uncertainty**

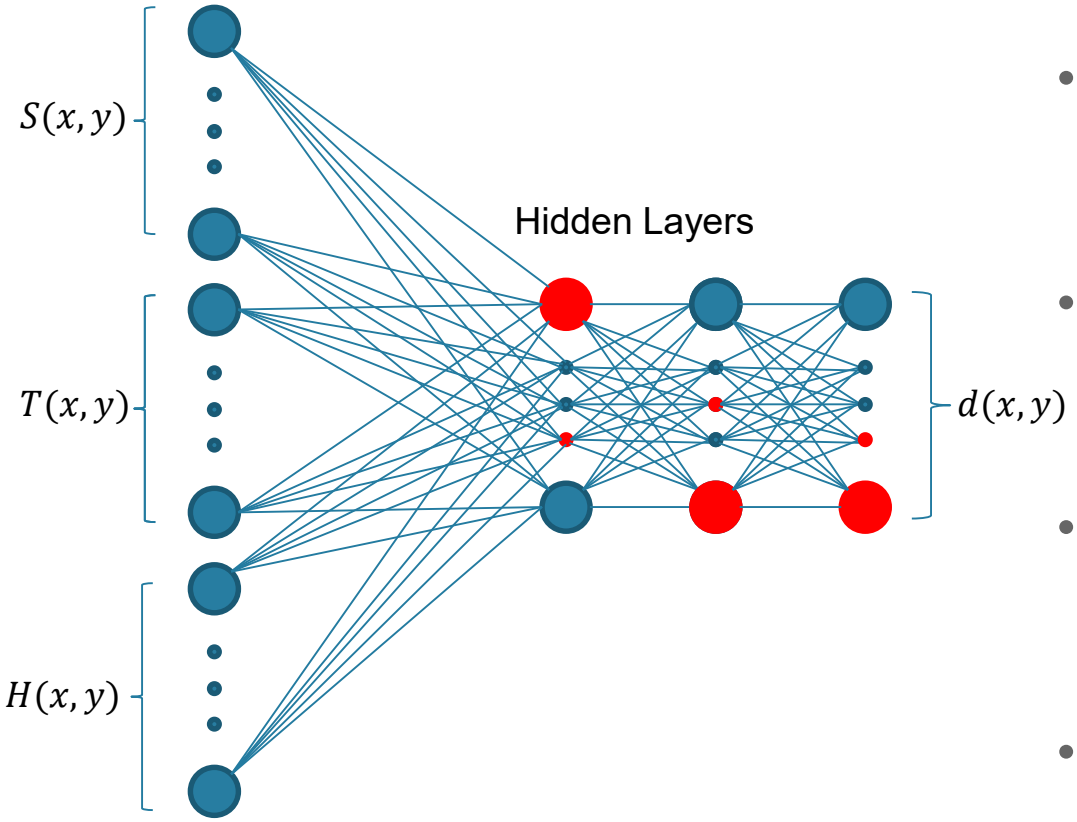
# VARIATIONAL NEURAL NETWORK




$$d(x,y) \sim N(\,\mu(x,y), \sigma(x,y)\,)$$

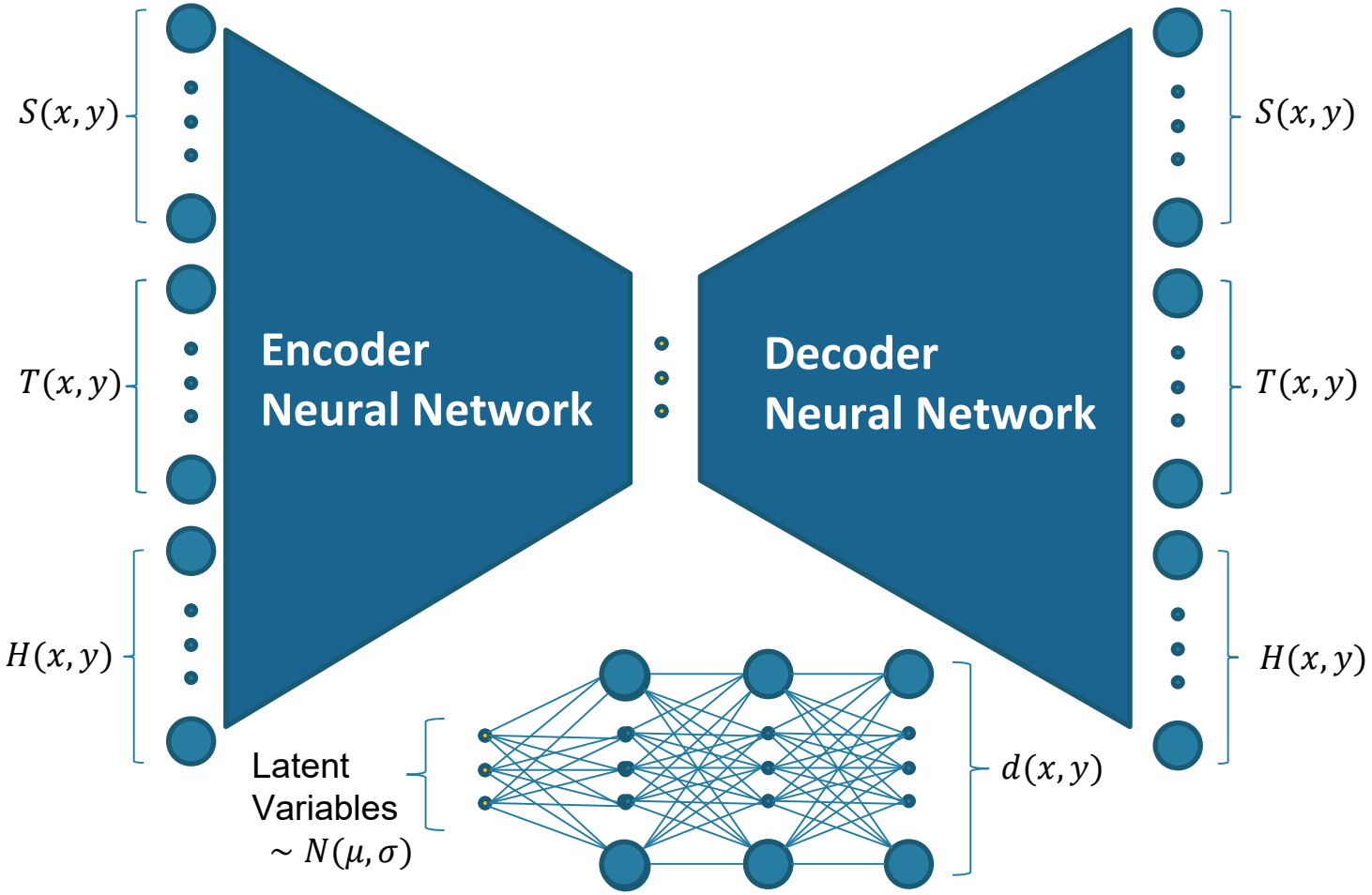
- Requires even **more parameters**
- Requires **parameterization of noise model**
- Better **captures aleatoric uncertainty**.

# DROPOUT




- **Randomly** set some weights to zero
- Creates an **ensemble of models**
- Computationally **inexpensive**
- **Requires sampling** to generate statistics

# VARIATIONAL AUTO ENCODERS




- Learns a **dimension reduction** of input data
- Gaussian **noise parameterization**