Physical model

*Physical model* is defined by a set of differential equations. Now, they are usually described by computer code models which are deterministic with unknown parameters. For example, the physical model we are interested in is Terrestrial Ecosystem Model (TEM).
Terrestrial Ecosystem Model (TEM)

TEM  A complex process that describes the dynamics of carbon, nitrogen, soil, water and other vegetation related variables.
Calibration

Definition: the process of fitting the model to the observed data by adjusting the parameters. It is a reverse process to regression, where instead of a future dependent variable being predicted from known explanatory variables, a known observation of the dependent variables is used to predict a corresponding explanatory variable.

Methods:

- Traditional Method which is ad hoc fitting
  - Data assimilation: the process by which observations are incorporated into a computer model of a real system
- Bayesian calibration which provides parameter estimates and uncertainty of parameters
Calibration

The advantages of Bayesian calibration over traditional method

- Allows for all sources of uncertainty
  - inaccurate inputs
  - insufficient knowledge of the parameters in the model
  - inherent randomness of the system

- Attempt to correct for any inadequacy of the model which is revealed by a discrepancy between the observed data and the model predictions from even the best-fitting parameter values
Statistical Emulation

Emulator is the statistical relationship

- Resembles the behaviour of the real system.
- An approximation of complex computer model.
- Simpler/easier surrogate of the computer model.
- Simpler and easier to quantify uncertainties
Statistical Emulation

- Suppose we can afford to obtain output $y(s_i)$ at input $x \in S = \{s_1, \ldots, s_n\}$, $n$ is training sample size.
- Next, we establish some statistical relationship between output function $y(x)$ and the obtained $y(s_i)$.
- Quantify the uncertainty of $y(x)$ from the statistical relationship of $y(x)$ and $y(s_i)$. 
Methods: Explanation in flow-chart

- Treat the complex TEM as a black-box
- Get training samples from the real model
- Build statistical emulator based on the training samples
Model

The Bayesian calibration framework in [1]:

\[ \mathbf{z}(\mathbf{x}) = \mathbf{y}(\mathbf{x}, \theta) + \delta(\mathbf{x}) + \epsilon(\mathbf{x}) \]  

- \( \theta \): the true real process calibration parameters, to be calibrated
- \( \mathbf{x} \in \mathcal{X} \): the variables inputs, such as temperature, precipitation.
- \( \mathbf{z}(\mathbf{x}) \): the real observations
- \( \mathbf{y}(\mathbf{x}, \theta) \): the emulator of the computer model
- \( \delta(\mathbf{x}) \): discrepancy between \( \mathbf{z}(\mathbf{x}) \) and \( \mathbf{y}(\mathbf{x}, \theta) \) besides measurement error \( \epsilon(\mathbf{x}) \)
Notations

- \( n \) : number of observations (\( z \))
- \( N \) : number of computer output (\( y \))
- \( t_1, t_2, ..., t_N \) : set of calibration parameters, given by sampling from the prior
- \( x_1^*, x_2^*, ..., x_N^* \) : set of variable inputs at which we get the computer output \( y(x^*, t) \).
- \( x_1, x_2, ..., x_n \) : set of variable inputs at which we observe the output \( z \).
- \( y(x_i^*, t_k) \) : set of computer outputs given \( x_1^*, x_2^*, ..., x_N^* \) and \( t_1, t_2, ..., t_N \)
- \( z(x_j) \) : the observations at \( x_1, x_2, ..., x_n \)
Gaussian process emulator $y(x, \theta)$

Gaussian process is a stochastic process whose realizations consist of random values associated with every point in a range of times (or of space) such that each such random variable has a normal distribution. Moreover, every finite collection of those random variables has a multivariate normal distribution.
Gaussian process emulator $y(x, \theta)$

When computer model is computationally expensive, $y(x, \theta)$ is known in practical but only at small number of input sites. Then, we treat it as unknown function in the Bayesian framework. Here, Gaussian process emulator is used to model $y(x, \theta)$

$$y = \eta(x, \theta) \sim N(m_1(x, \theta), c_1((x, \theta), (x', \theta)))$$

Measurement error $\epsilon(x)$ is observational error which is iid $N(0, \sigma^2)$, where $\sigma^2$ is estimated by experts.
Discrepancy model: \( \delta(x) \)

\( \delta(x) \) is normal distribution:

\[
\delta \sim N(m_2(x), c_2(x, x'))
\]

The measurement error \( \epsilon \sim N(0, \lambda_0) \), \( \lambda_0 \) is elicited from expert.

- \( m_1(x) = h_1(x)^T \beta_1 \)
- \( m_2(x) = h_2(x)^T \beta, \beta = (\beta_1^T, \beta_2^T)^T \)
- \( c_1(., .) \) and \( c_2(., .) \) have hyperparameters \( \phi_1 \) and \( \phi_2 \)
- prior information about \( \theta \) is independent of the others.
Bayesian Calibration

To Be continued
References