Adventures in Emulation

Peter Challenor, Yiannis Andrianakis [1]
Robin Tokmakian [2]

[1]National Oceanography Centre
[2]Naval Postgraduate School

Outline

Emulators

Stommel Model

Extremes

Emulation as Parameterisation

Emulators

- There has been a lot of discussion at the workshop about emulators but just to recap
- We have a simulator
$$y = f(x)$$
- Treat $f(.)$ as an unknown random function and use Bayesian methods to estimate it
- Note it is possible to redo all this work in a frequentist way see book by Santner et al

Gaussian Processes

- We model the simulator with a Gaussian Process
- This has mean
$$y = f(x) = h(x)^T \beta$$
- Variance σ^2
- And correlation function $c(x, x')$ Usually, but not invariably, we take
$$c(x, x') = exp(-(x - x')^T C(x - x'))$$
$$C^{-1} = diag(\delta)$$
- so we have a set of GP parameters (β, σ, δ) which we need to estimate

The Prior

$$\mu(x) = h(x)^T \beta$$

h(.) is a known vector of regressor (or basis) functions
e.g. $h(x)^T = (1, x, x^2)$
β is a vector of unknown parameters

$$\rho(x_1, x_2) = \sigma^2 c(||x_1, x_2||)$$

$$c(x_1, x_2) = exp\left(-\frac{||x_1, x_2||^2}{\delta}\right)$$

$$\pi(\beta, \sigma^2) \propto \sigma^{-2}$$

The Posterior

$\eta(x) \sim t_{n-q}$

$$E(\eta(x)) = h(x)^T\beta' + t(x)^T A^{-1}(y - H\beta')$$

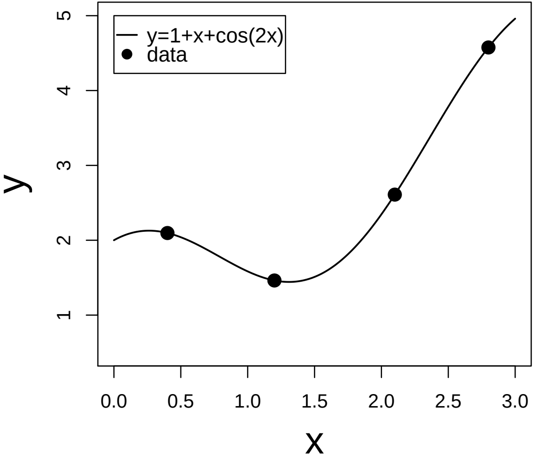
$$\beta' = \left(H^T A^{-1} H\right)^{-1} H^T A^{-1} y$$

H is the matrix $\{h(x_1), \cdots, h(x_n)\}^T$
These are the regression terms at x

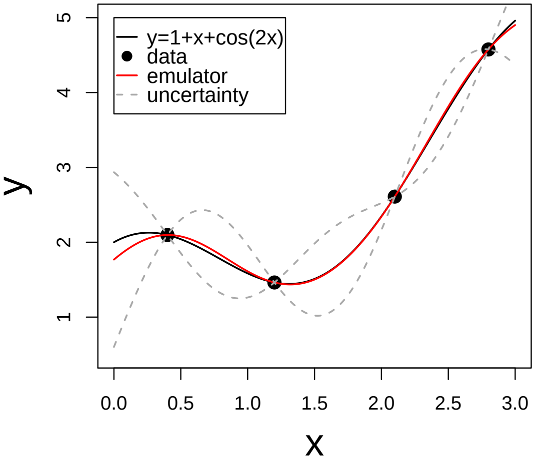
$$t(x) = \{c(x, x_1), \cdots, c(x, x_n)\}$$

This is the correlation of x with the data, x_i
A is the matrix $\{c(x_i, x_j)\}$
This is the correlation matrix of the data with itself
And there are similar, but more complex, expressions for the variance
We estimate δ by maximising the marginal likelihood

Example



Example



y=1+x+cos(2x)
data
emulator
uncertainty
y
x
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4
5

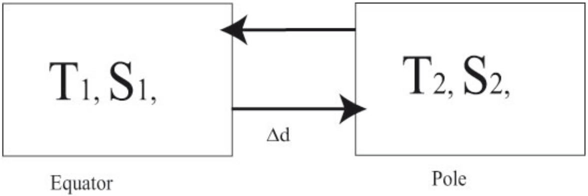
Some Adventures

- Highly non-linear models (the Stommel model)
- Extremes
- Emulation as Parameterisation

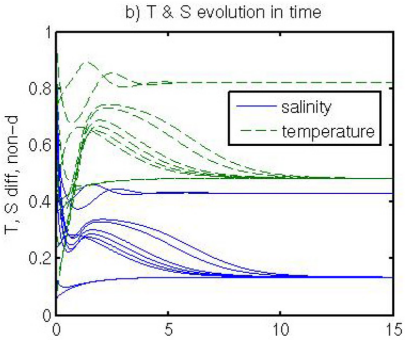
Outline

The Stommel Model

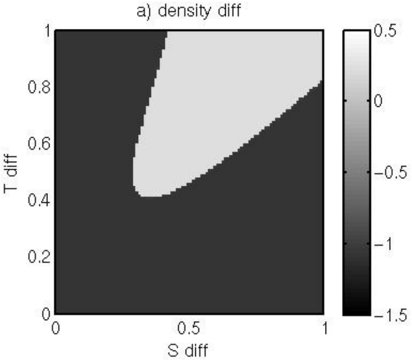
- The Stommel model for the Atlantic overturning circulation has only two states
- Our emulators assume ‘smoothness’
- Can we emulate such a highly non-linear model?



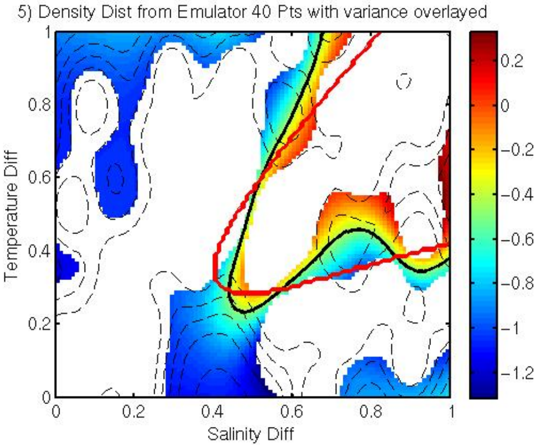
Emulating the Stommel Model



Emulating the Stommel Model



Emulating the Stommel Model



Outline

Emulators

Stommel Model

Extremes

Emulation as Parameterisation

Extremes

- Often we are interested in future climate extremes not mean values

The Statistics of Extremes

- In a similar way to the central limit theorem extremes have their own limits

$$\lim_{N\to\infty} F^N\left(\frac{x-a_N}{b_N}\right) = G(x)$$

- Distributions that satisfy these limits exactly are known as *max stable* distributions
- There is only one possible form: the *Generalised Extreme Value* distribution

$$P(X<x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

- (Sometimes the limit is given as three different forms:- Fréchet, Gumbel, and Weibull)

- These are obtained by performing extreme value analysis on model output via downscaling or a ‘weather generator’
- These estimates don’t have the simulator uncertainty included
- Can we emulate extremes directly across an ensemble of climate models?
- It is unrealistic to use GP’s for this as we know that extremes have very non-Gaussian distributions

Max stable processes

- In a similar way to Gaussian processes we can define max stable processes whose marginals are the extreme value distributions
- These are much more complex than Gaussian processes
- Brown-Resnick processes may be suitable for building extremal emulators

$$\eta(t) = \bigvee_{i=1}^{\infty} \left\{ U_i + W_i(t) - \frac{\sigma^2(t)}{2} \right\}$$

An Alternative

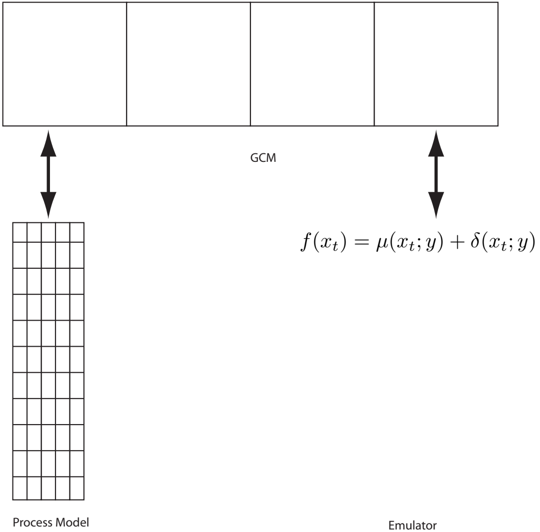
- The Generalised Extreme Value (GEV) distribution has three parameters
- Extremes do not have Gaussian distributions
- But the parameters of the GEV can be assumed to do so
- Therefore we can jointly emulate the three parameters
- Since these three jointly define the extremes we have, in effect, a GP emulator for extremes
- The max stable process defined this way is less adaptable than the Brown-Resnick process

Outline

GCM’s and Process Models

- GCM’s do not model well sub-grid scale processes
- Often we have good process models
- These are too expensive to embed in the GCM
- Replace the process model with an emulator

Emulation as parameterisation



We have a GCM represented by

$$x_{t+1} = g(x_t; \theta)$$

And a process model

$$x_t' = f(x_t; \phi, y_t)$$

where y_t are nuisance parameters We would like to calculate

$$x_{t+1} = g(f(x_t; \phi, y_t); \theta)$$

But it is too expensive

Conventionally we parameterise $f(.)$ with a simple (but possibly non-linear) parameterisation

$$f^{\dagger}(x_t) \approx f(x_t)$$

so we have

$$x_{t+1} = g(f^{\dagger}(x_t); \theta)$$

We propose to replace $f^{\dagger}$ with an emulator of f

$$f^{*} = \mu + \delta$$

where μ is a mean function and δ is a zero mean Gaussian process

Nuisance Parameters

- The process model $f(.)$ is not only driven by the GCM state parameters but also by other *nuisance* parameters
- We need values for these to drive the emulator
- (unless we find them non-active during the emulation process)

Nuisance Parameters

Encapsulate our prior knowledge about y in a prior
(since y may be affected by the value of x we make our prior conditional on the value of x)
so we have $p(y|x)$
There are two ways we can use this information

1. Deterministically - replace y by $E(y|x)$
2. Stochastically - replace y by a sample from $p(y|x)$

Emulation as Parameterisation Summary

- Emulators can be used to parameterise sub-grid scale processes
- They (approximately) preserve all the physics
- Need some examples

Conclusions

- Emulators are powerful tools for analysing climate models
- There are a lot of new exciting applications
- Look forward to many more adventures