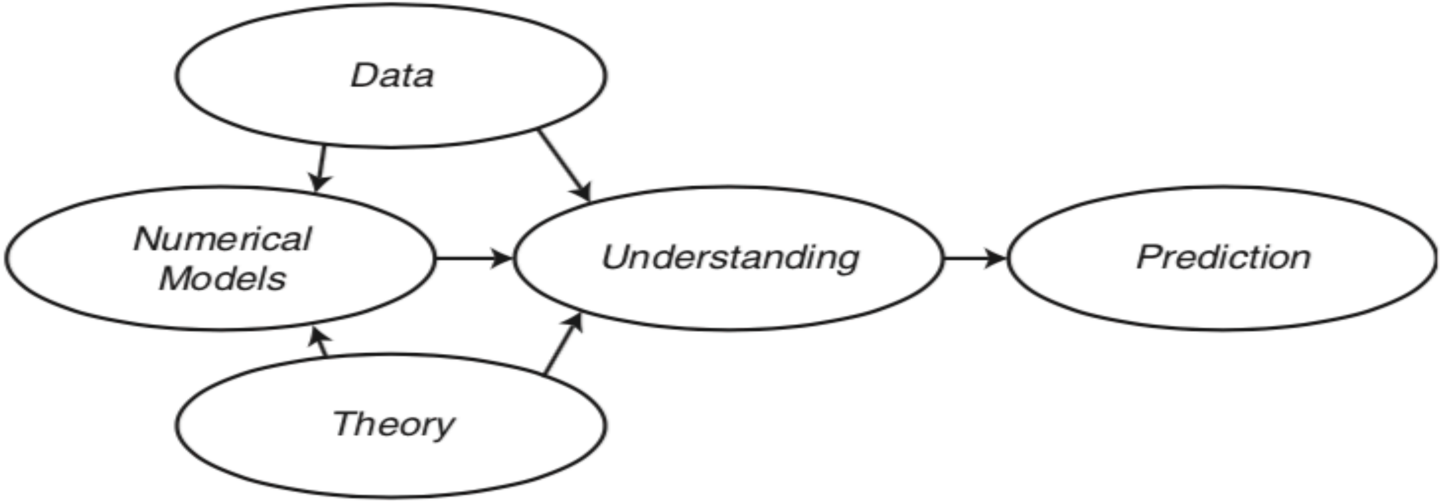
Integrating physics and statistics: a Bayesian approach to predictive uncertainty of solitons

Edward Cripps
School of Mathematics and Statistics,
CI Data Analytics, ARC Research Hub Offshore Floating Facilities
UWA

Data Science: What is it?

Our Definition: Using probability and data to reconcile real physical systems with mathematical approximations



The soliton problem

Soliton: a non-linear internal oceanic wave generated by tidal forcing over topography.

Interest for ocean engineering and oceanography because:

- Important driver of extreme currents.
- Induce large stress on offshore infrastructure.
- Influence dynamic position systems during operations.
- Drive sediment suspension

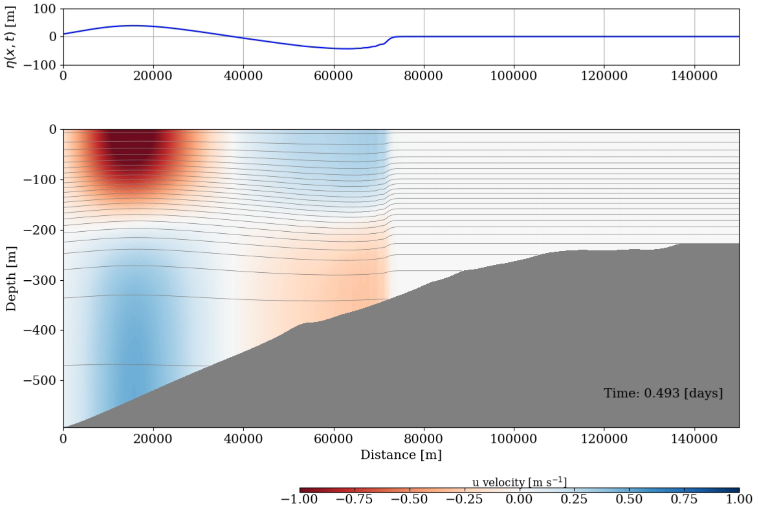
Motivating Industrial Question:

What will be a *plausible range* of maximum amplitudes induced by a given soliton.

The soliton problem

Model the amplitude, $A(x,t)$, using the variable-coefficient *KdV equation:*

$$\frac{\partial A}{\partial t} + c(x)\frac{\partial A}{\partial x} + \alpha(x)A\frac{\partial A}{\partial x} + \beta(x)\frac{\partial^3 A}{\partial x^3} + \frac{c(x)}{2Q(x)}\frac{\partial Q}{\partial x}A = 0$$



Initial Conditions (unknown):

- Density Stratification
- Initial wave amplitude

Important to acknowledge:

- Inputs are uncertain
- KdV is a simplification/approximation of reality
- Even if exact, the solution is still approximate because it is intractable.

The approximation problem

ARC Research Hub for Offshore Floating Facilities
THE UNIVERSITY OF WESTERN AUSTRALIA
WESTERN SYDNEY UNIVERSITY

Uncertainty and probability:

- Any form of approximation induces uncertainty
- Probability theory provides a mathematical description of uncertainty.
- Basic to all decision making under uncertainty.

"...the theory of probabilities is basically just common sense reduced to calculus; it makes one appreciate with exactness that which accurate minds feel with a sort of instinct, often without being able to account for it." Pierre Simon Laplace (1749–1827)

Bayesian Statistics

A coherent probabilistic fusion of data and scientific knowledge

new data: $p(y \mid \theta)$

prior knowledge: $p(\theta)$

$$p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta)$$

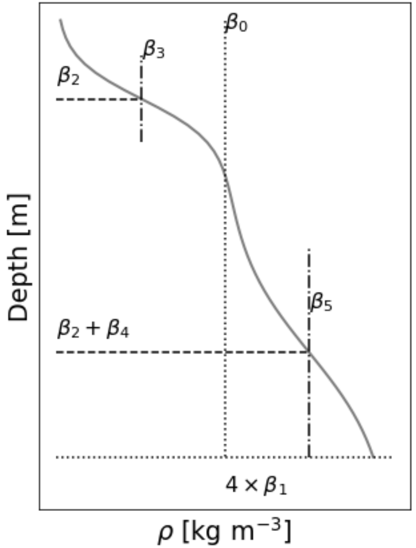
posterior $\propto$ likelihood • prior

We build a model for the temporal evolution of density stratification, given data observed on the North West Shelf.

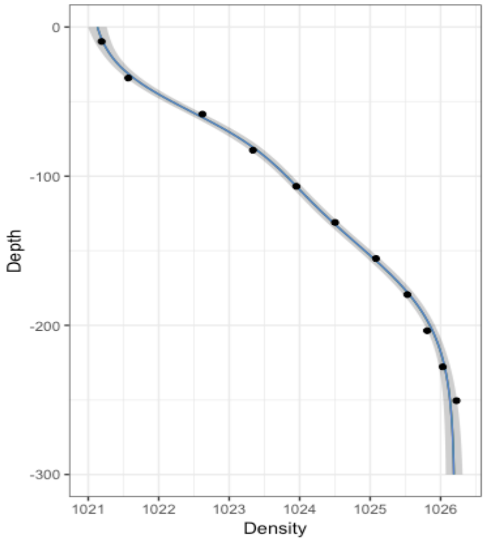
KdV Inputs: density stratification

ARC Research Hub for Offshore Floating Facilities
THE UNIVERSITY OF WESTERN AUSTRALIA
WESTERN SYDNEY UNIVERSITY

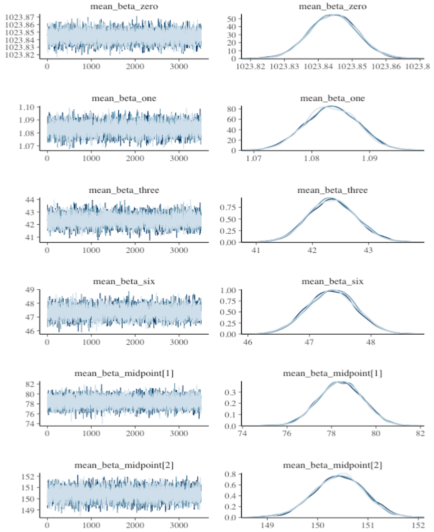
$$\rho(z) = \beta_o + \beta_1(\tanh(\frac{\beta_2 - z}{\beta_3}) + \beta_1 \tanh(\frac{\beta_4 - z}{\beta_5}))$$



2. Fit w/ uncertainty



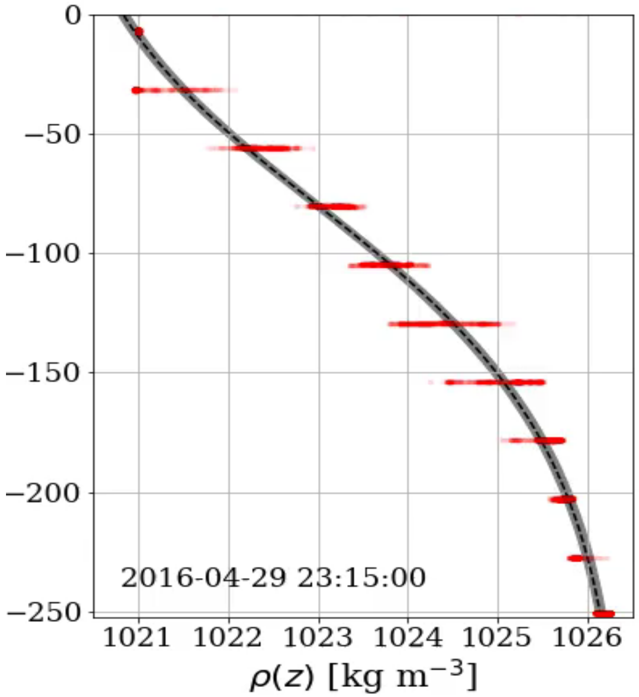
3. Parametric inference

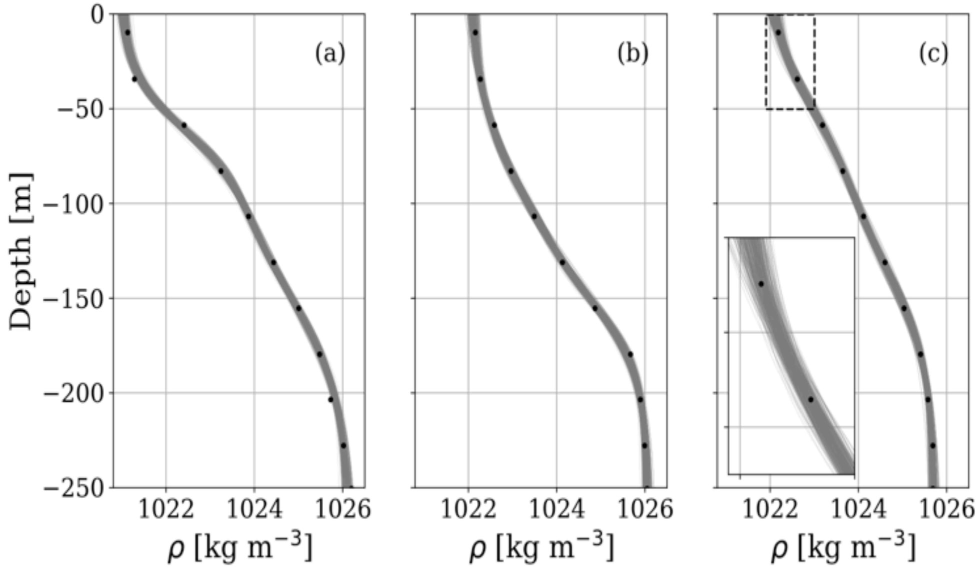


Manderson, A., Rayson, M., Cripps, E., et al. (2019)
``Uncertainty quantification of density and stratification estimates with implications for predicting ocean dynamics''.
Journal of Atmospheric and Oceanic Technology, 36(7), pp 1313--1330

KdV Inputs: density stratification

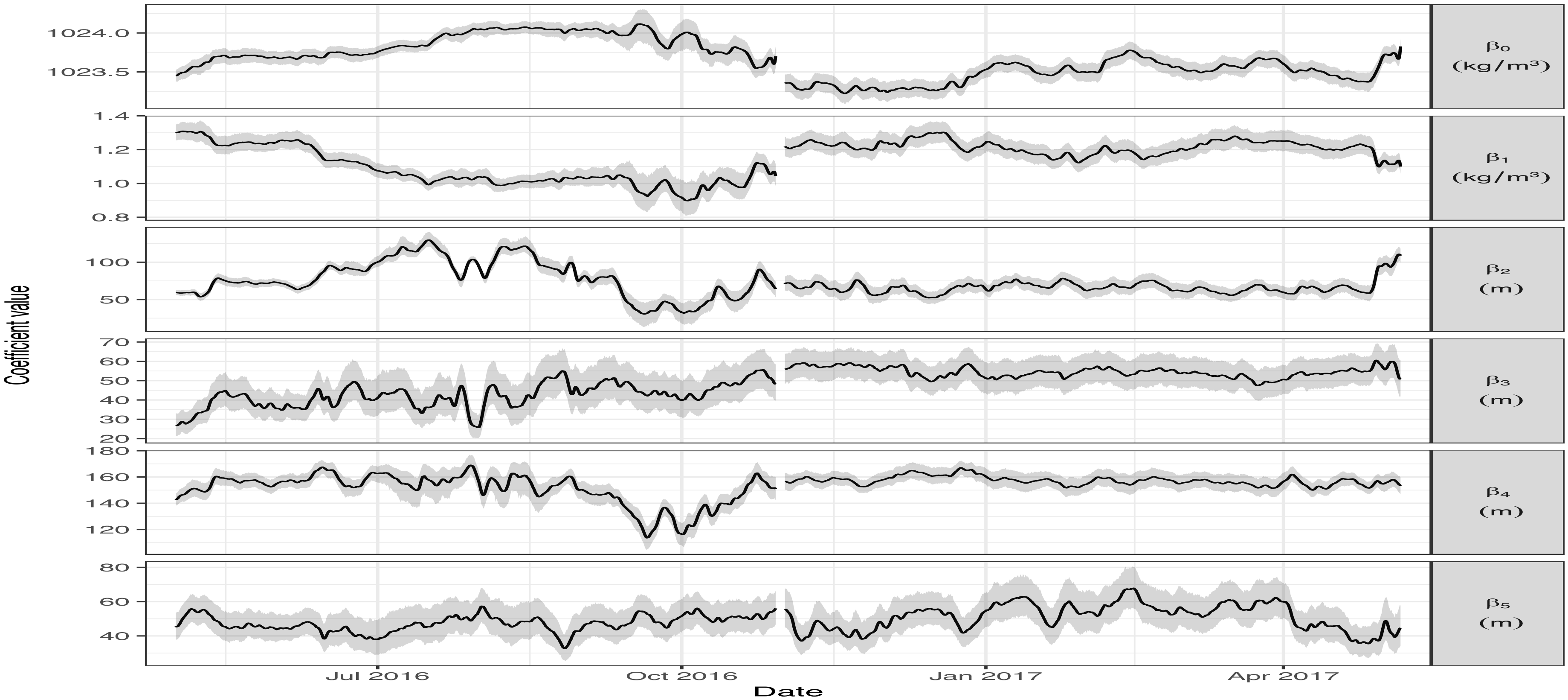
Temporal evolution: embed parametric in hierarchical model





KdV Inputs: density stratification

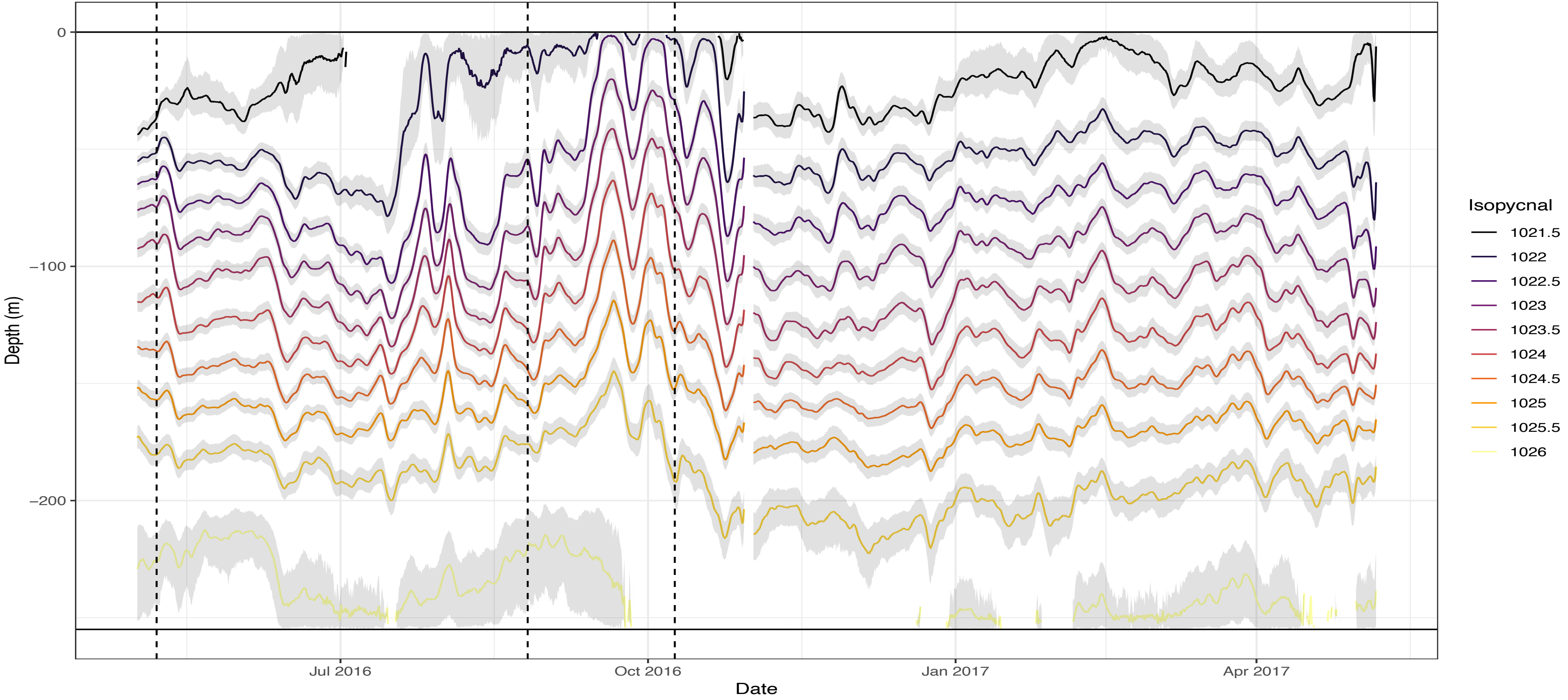
Temporal evolution of predictive distribution of density profile characterists



KdV Inputs: density stratification

ARC Research Hub for Offshore Floating Facilities
THE UNIVERSITY OF WESTERN AUSTRALIA
WESTERN SYDNEY UNIVERSITY

Temporal evolution of the predictive distribution of isopycnals



Uncertainty Analysis of KdV:

ARC Research Hub for Offshore Floating Facilities
THE UNIVERSITY OF WESTERN AUSTRALIA
WESTERN SYDNEY UNIVERSITY

Physical model: The Korteweg-de Vries equation for continuously stratified fluids:

$$\frac{\partial A}{\partial t} + c(x)\frac{\partial A}{\partial x} + \alpha(x)A\frac{\partial A}{\partial x} + \beta(x)\frac{\partial^3 A}{\partial x^3} + \frac{c}{2Q}\frac{\partial Q}{\partial x}A = 0$$

Uncertain Inputs:

- $\alpha(x)$ and $\beta(x)$ require density profile, $\rho_t(z)$, z = depth vector
- Initial condition $A(x,0)$ requires initial wave amplitude, a_0

Statistical models for input uncertainty: $p(\rho_t(z))$ and $p(a_{0t})$

e.g.,

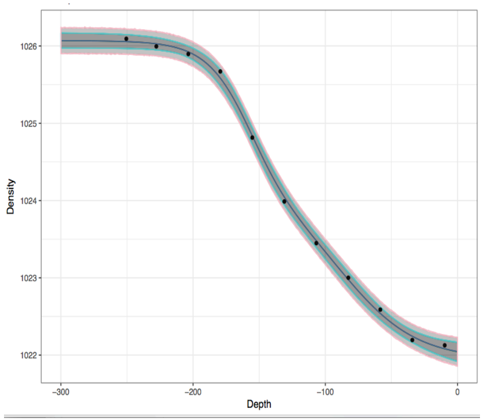
- $y_t = \rho_t + \epsilon_t,$

Combining the physics, statistics and computing for industrial impact

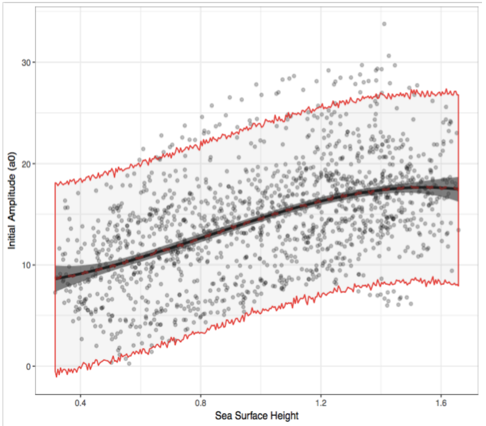
$p(A_{mt}|Data_t, KdV)$: the predictive distribution of A_{mt}, integrating out inputs

KdV uncertainty quantification

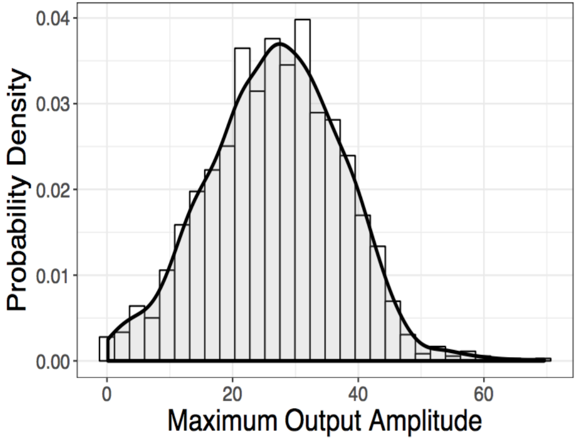
Predictive distribution of maximum amplitude at a given time.



+



=



Computationally demanding: Parallel, distributed and cloud computing

Software development: Necessary for industrial uptake

The people

To achieve the previous slides needs input from a wide range of expertise:
Oceanographers, statisticians, computer engineers, engineers.

Manderson

Rayson

Barlow

Girolami

Gosling

Hodkiewicz

Ivey

Jones

Current Extensions and the bigger picture

ARC Research Hub for Offshore Floating Facilities
THE UNIVERSITY OF WESTERN AUSTRALIA
WESTERN SYDNEY UNIVERSITY

Bayesian Analysis of Computer Code Output

Denote by $y = f(x : \theta)$ as the simulated output of a computer model that relies on inputs x and tuning parameters, θ.

- Uncertainty Analysis: propagate $p(x)$ through $f(x : \theta)$
- Inversion: given observations, z, identify optimal values of θ via its posterior distribution: $p(\theta | x, y, z)$
- Numeric Solver Uncertainty Different solvers/grids etc, can yield different approximate evaluations of the model. Probability can turn the problem into one of statistical inference.
- Sensitivity Analysis: determines which inputs are most influential on simulator output.