Using stochastic models to improve predictions of internal tide-driven currents and temperature

Forum for Operational Oceanography

Matt Rayson, Lachlan Astfalck, Nicole Jones, Andrew Zulberti, Yutao Zheng, Aurelien Ponte, Michael Bertolacci, Edward Cripps

30th November 2023

Oceans Graduate School
University of Western Australia

ARC Research Hub for Transforming energy Infrastructure through Digital Engineering

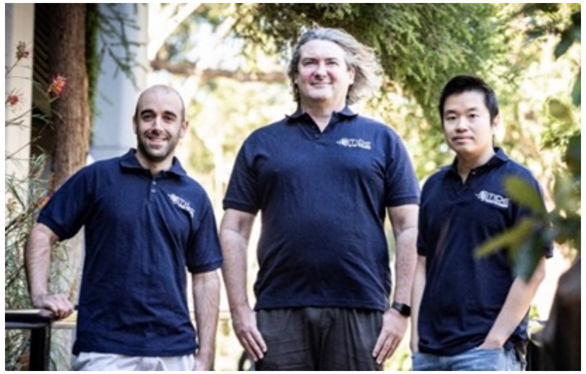
THE UNIVERSITY OF WESTERN AUSTRALIA

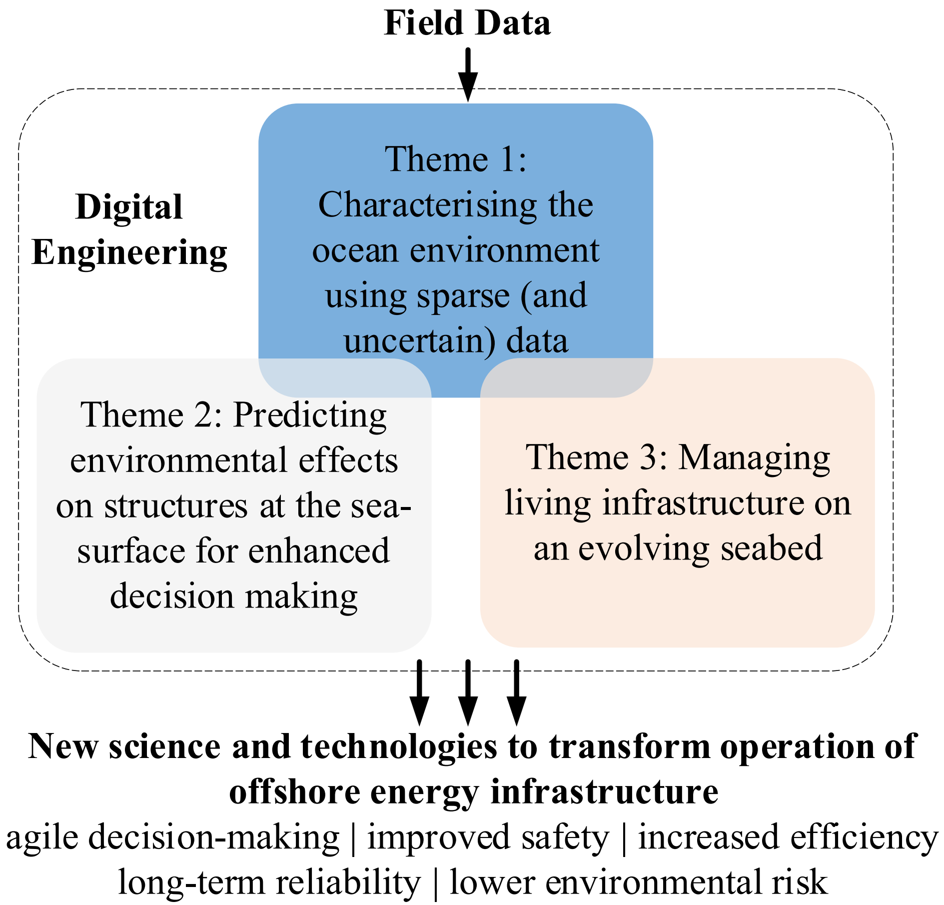
Transforming energy Infrastructure through Digital Engineering Research Hub

Development of new science and technology through ***digital engineering*** to optimise the management of offshore energy infrastructure – thereby making this activity cheaper and yet more reliable.

Digital Engineering is the creation, use and embedment of data in engineering.

Oceanography focus: Prediction of nonlinear internal waves and submesoscale eddies using "traditional" physical models and new "data science" methods





Our partners

Australian Government Australian Research Council

UNIVERSITY OF WOLLONGONG AUSTRALIA

INPEX

Woodside

BUREAU VERITAS 1828

Lloyd's Register

FUGRO

rps

wood.

Australian Government
Australian Institute of Marine Science

Australian Government Bureau of Meteorology

The Alan Turing Institute

HR Wallingford

Lancaster University

University of Oxford

University of Southampton

TCOMS

The University of Texas at Austin

Setting the scene

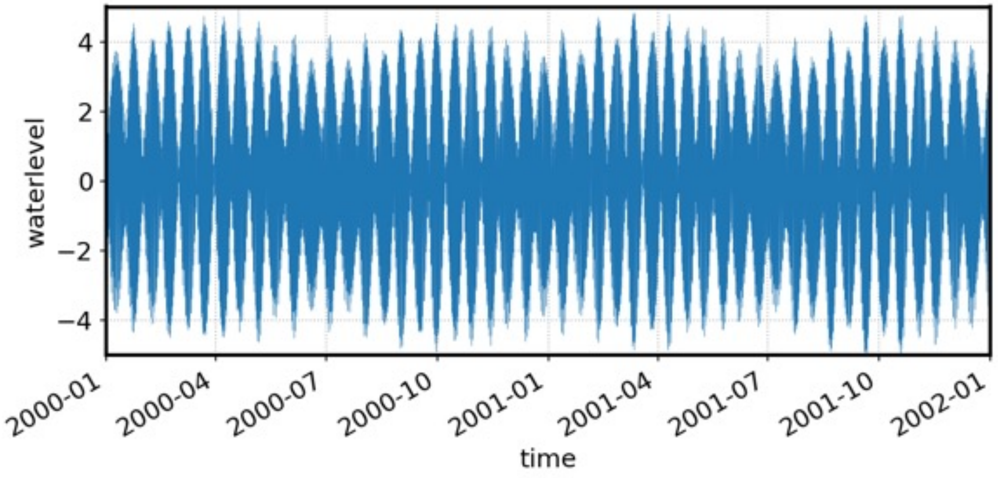
- **Tidal harmonic analysis** is useful for predicting water level in coastal regions
- It is also useful for predicting ocean currents and water temperature in some offshore locations where internal waves are active
- The utility of tidal predictions (i.e. their *useability* for decision-makers) is usually judged from some **bulk measure of residual properties** (rmse, skill, etc.)
- Threshold for a "good fit" is usually left to the decision-maker... resulting in ad-hoc decisions

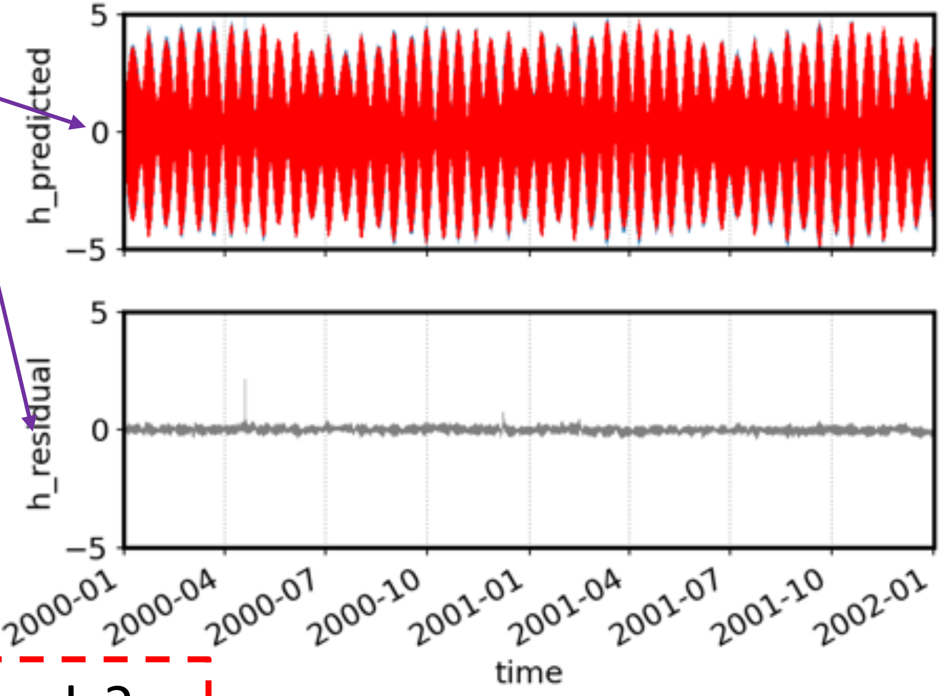
What should decision-makers (operators) do with the model residual?

Example 1: Broome sea-level

THE UNIVERSITY OF WESTERN AUSTRALIA

$$y = \sum_j A_j \cos(\phi_j) + B_j \sin(\phi_j) + \varepsilon$$



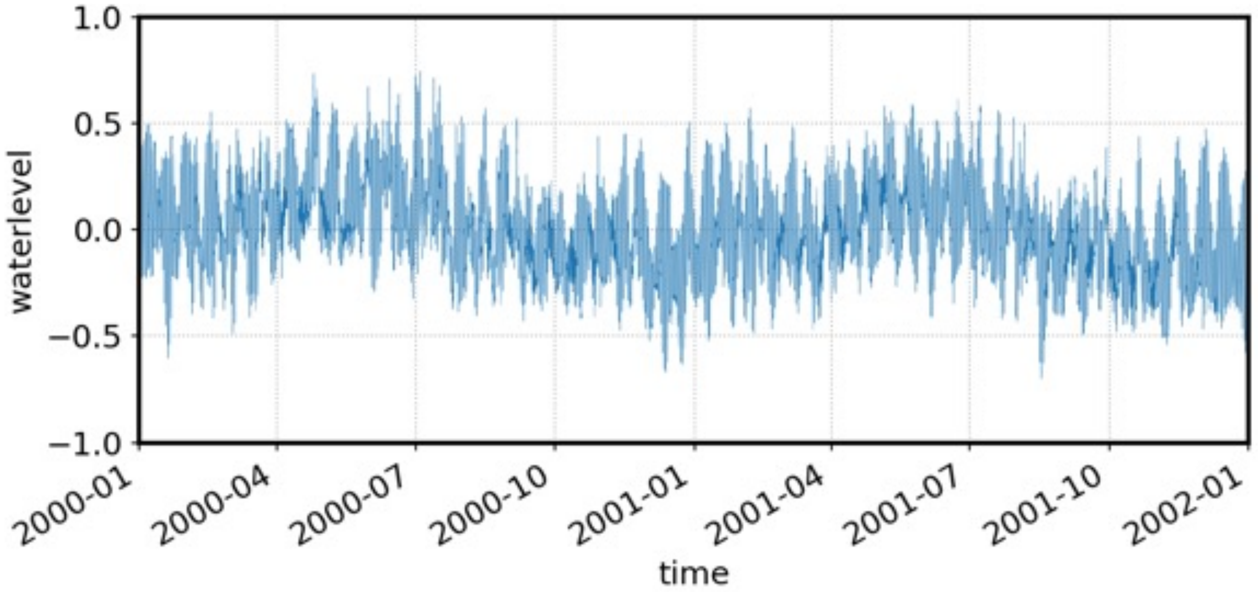


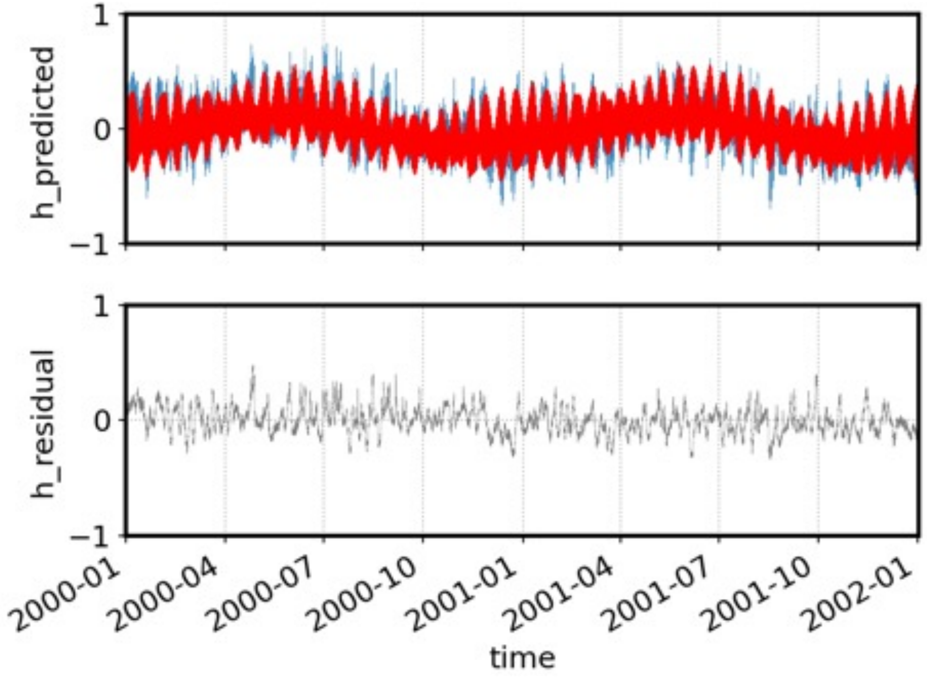
Question: Is the residual OK for my needs?

Data source: Australian Baseline Sea Level Monitoring Project (BoM)

Example 2: Hillarys sea-level

$$y = \sum_j A_j \cos(\phi_j) + B_j \sin(\phi_j) + \varepsilon$$



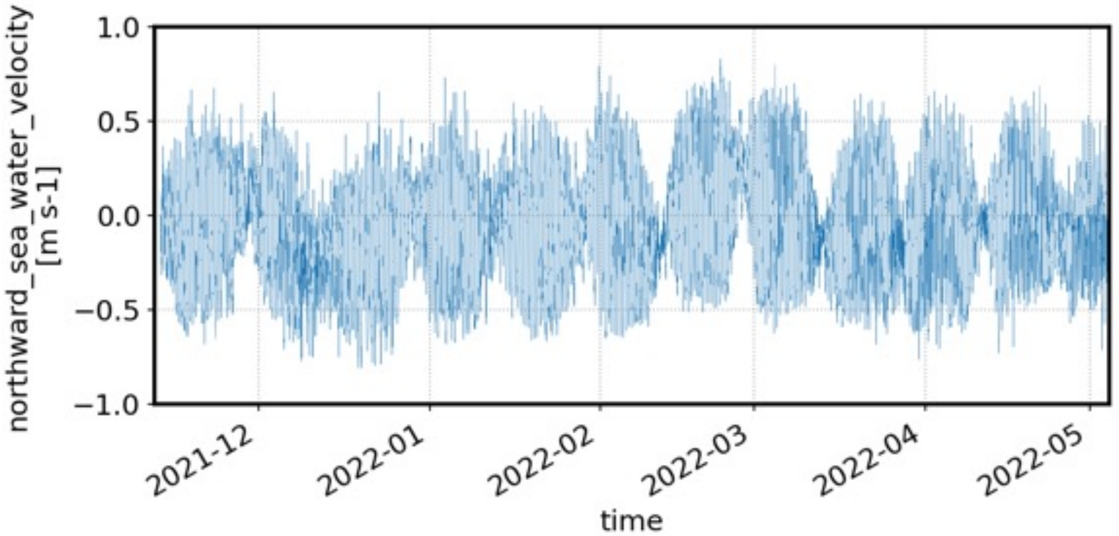


Question: Is the residual OK for my needs?

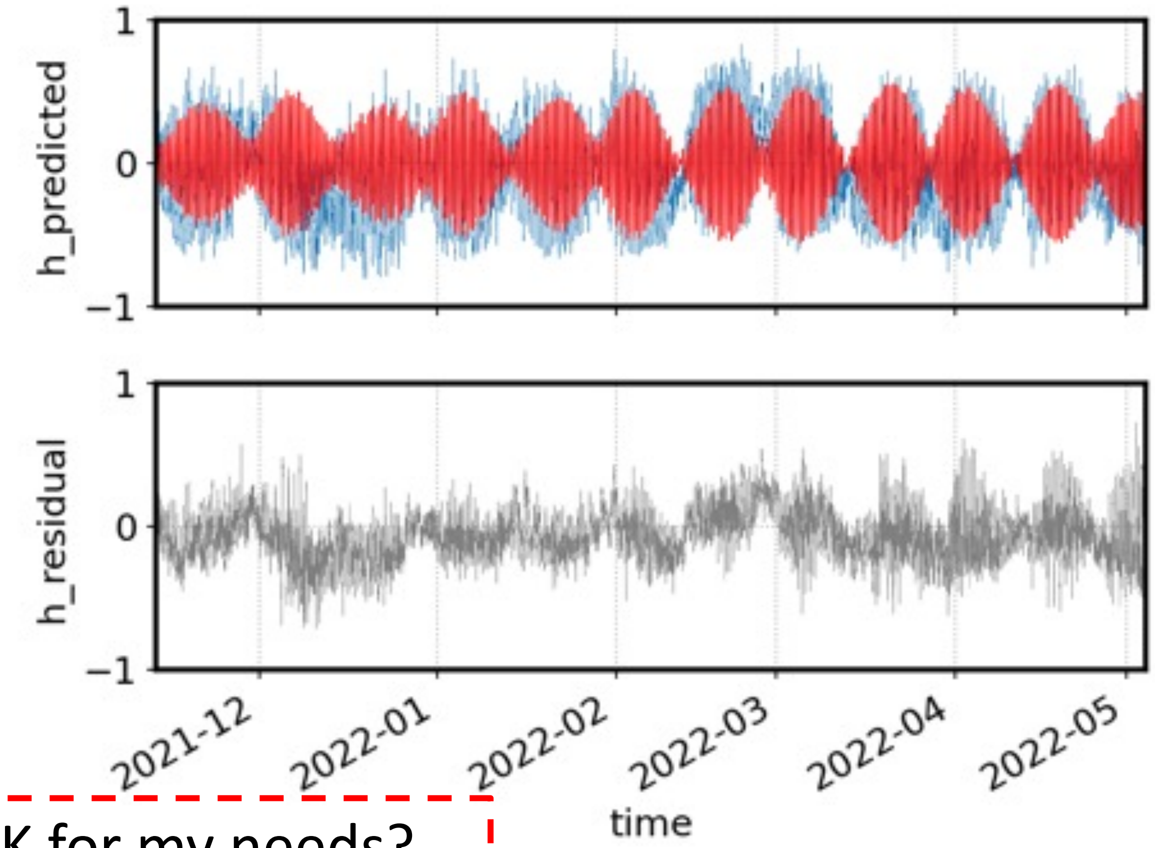
Data source: Australian Baseline Sea Level Monitoring Project (BoM)

Example 3: NWS ocean currents

$$y = \sum_j A_j \cos(\phi_j) + B_j \sin(\phi_j) + \varepsilon$$



Data source: IMOS National Mooring Network (NWSBRW)



Question: Is the residual OK for my needs?

Question 2: If no, can I retain the residual for decision-making?

A solution to keeping the residual

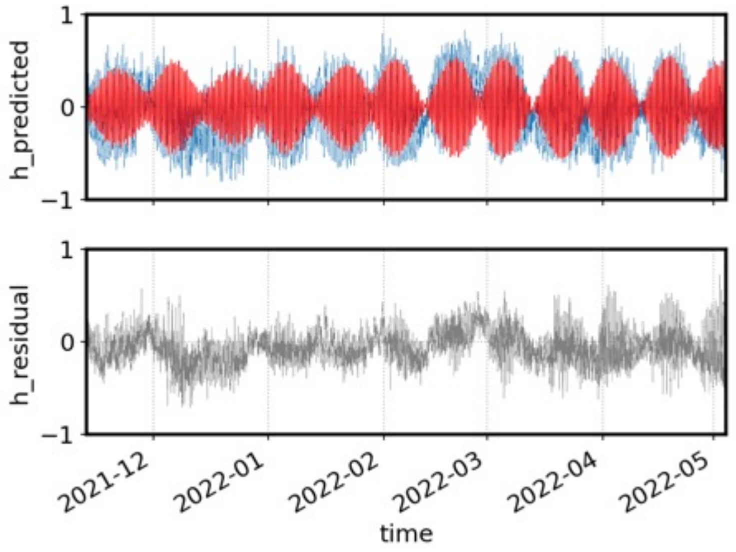
Model the tides deterministically (i.e. harmonics) and the residual as a *stochastic process* by retaining the same *statistical properties*

- *Stochastic process:* random draw of one possible outcome. This can be done many times e.g. Monte Carlo analysis.
- *Similar statistical properties*: signal variance partitioned into similar frequencies i.e. similar power spectral density

A solution is a machine learning technique called **Gaussian Process Regression**

Advantages of the method:

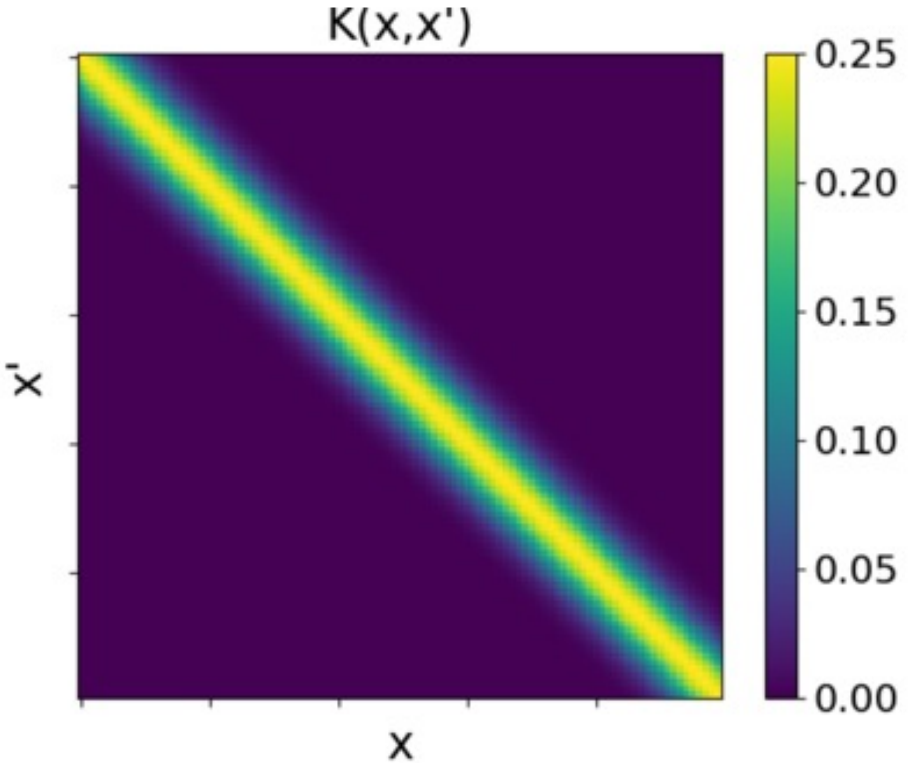
- Makes stochastic forecasts → **uncertainty quantification**
- It can be fast to train (compared to deep learning)
- It is much cheaper to run than traditional ocean models
- It is **interpretable** (parameters may be set by an expert without data)
- Predictions look realistic (not too smooth, not too rough)

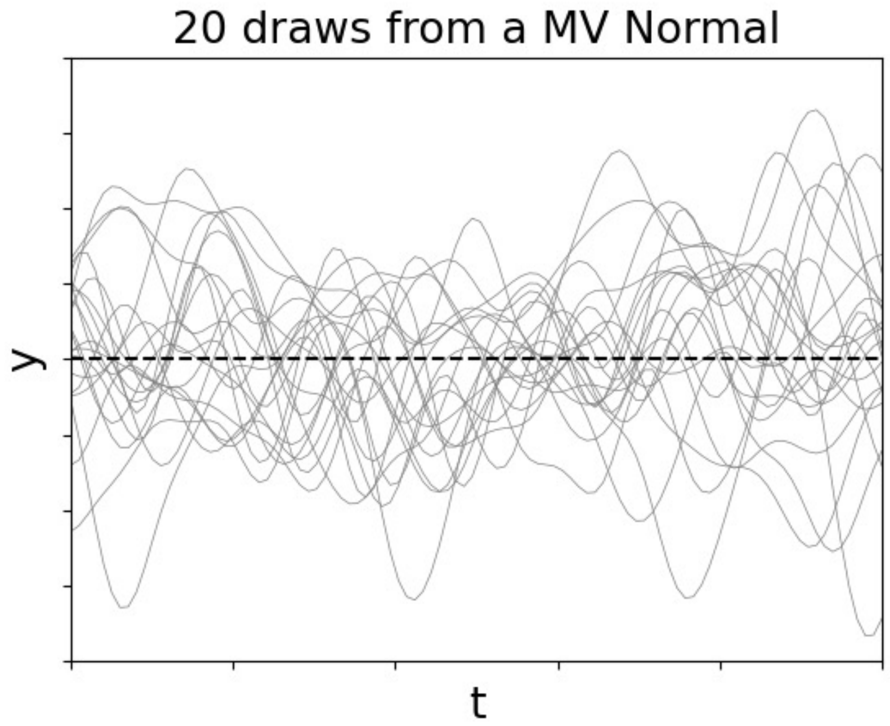


Multivariate Normal Distribution

$$y \sim MVN(\boldsymbol{\mu}, \Sigma)$$

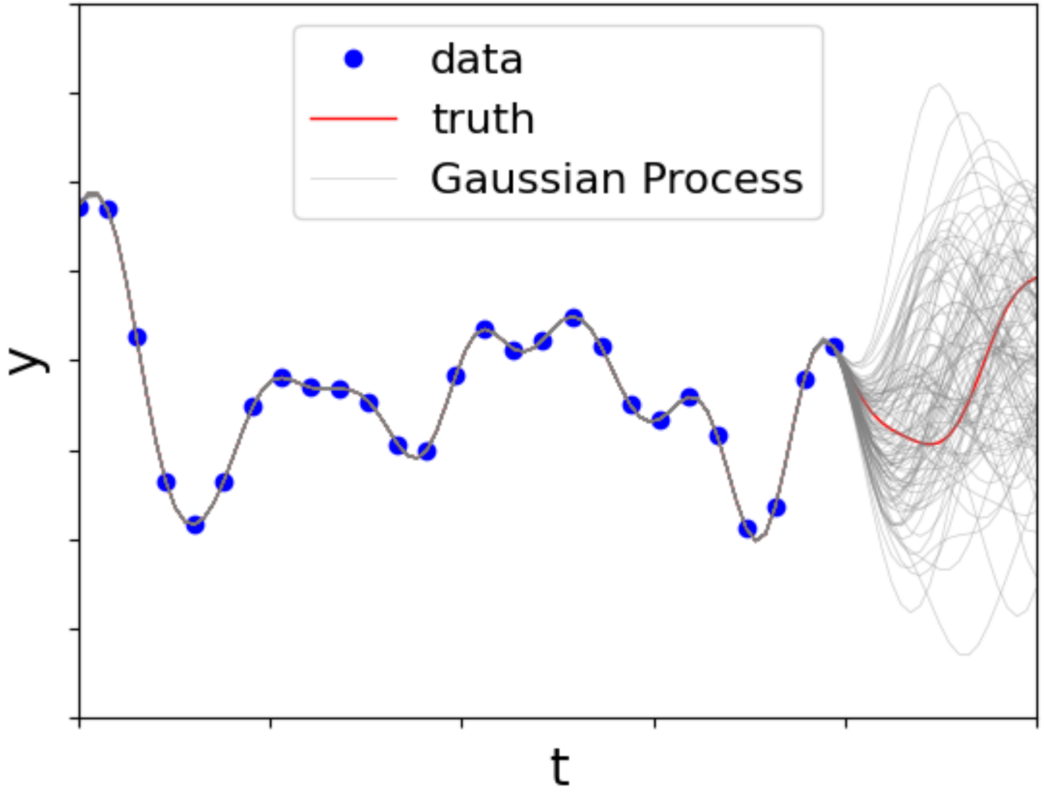
- $\boldsymbol{\mu}$ = mean function (e.g. tidal harmonics)
- Σ = covariance kernel, $\Sigma \approx \mathrm{K}(\mathrm{x}, \mathrm{x}'; \theta)$ (the statistical relationship between time points)





Gaussian Process Regression

- *Loose definition: Multivariate normal distribution conditioned on some data*

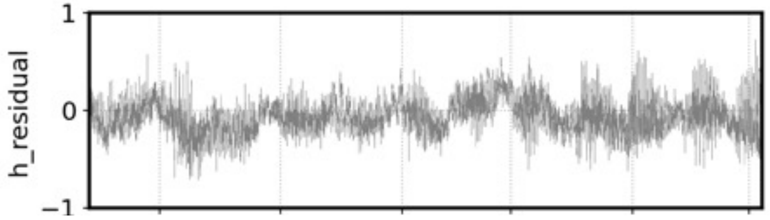


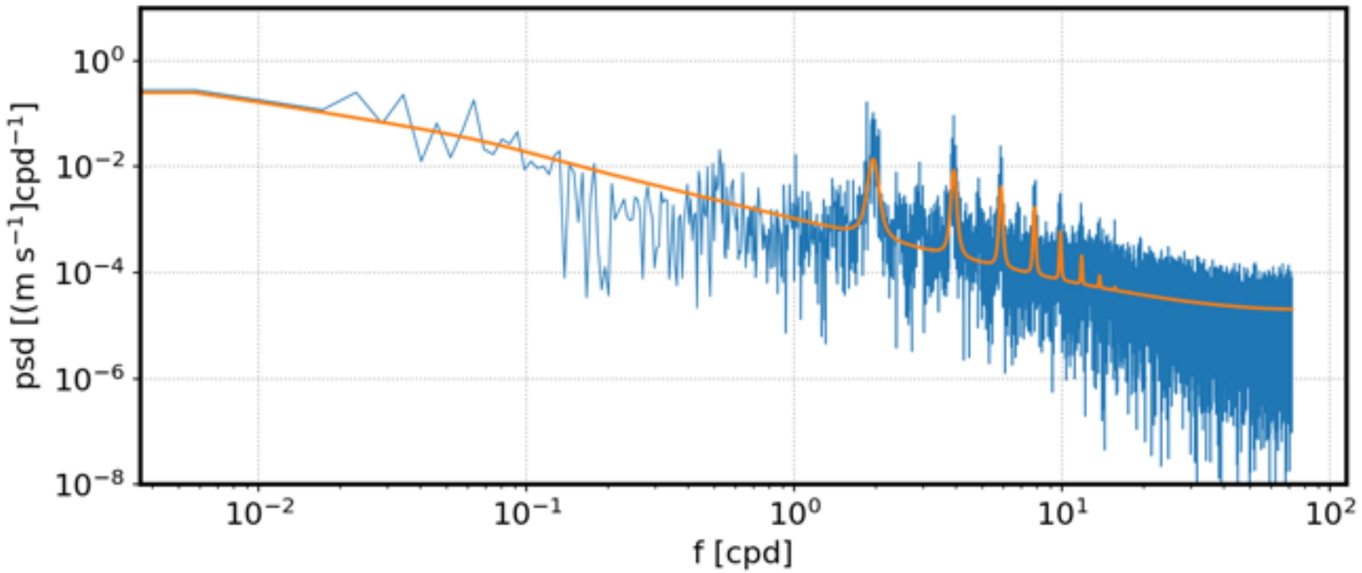
The key to the method producing useful forecasts is in the **choice of covariance kernel**, $\Sigma \approx \mathrm{K}(\mathrm{x}, \mathrm{x}'; \theta)$

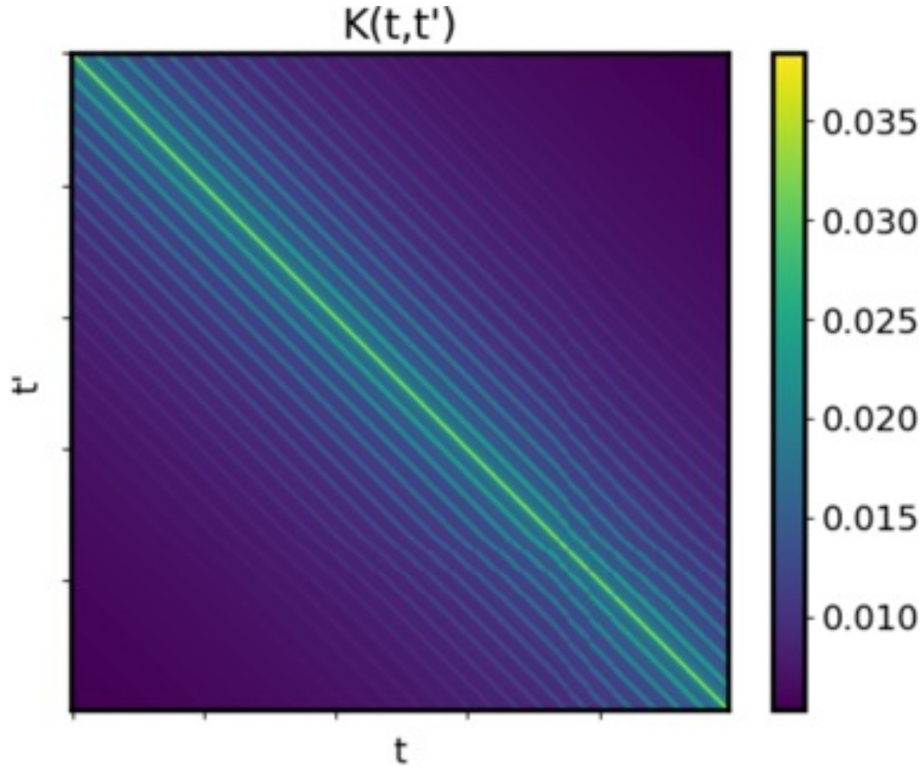
→ We do this based on the power spectrum of the tidal harmonic residual

Power Spectrum ←→ MV Normal

- *Similar statistical properties*: signal variance partitioned into similar frequencies i.e. similar power spectral density as the model residual
  - → "**realistic looking** predictions"

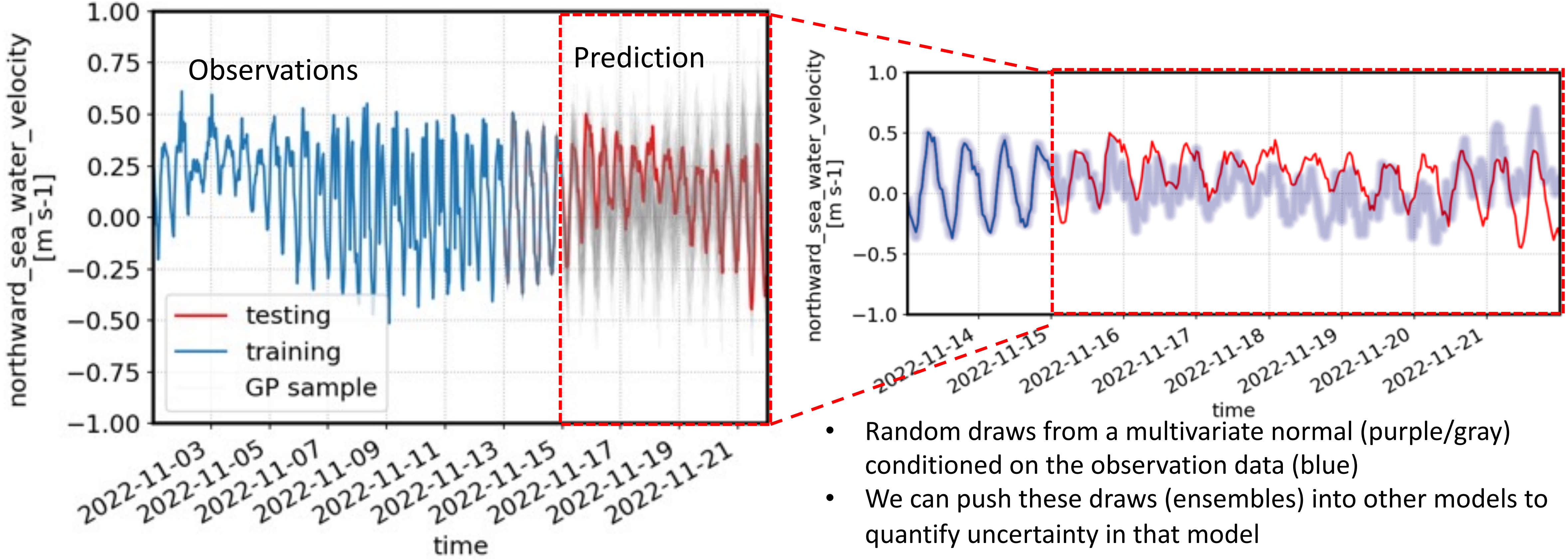






Stochastic predictions

Data source: IMOS National Mooring Network (NWSBRW)



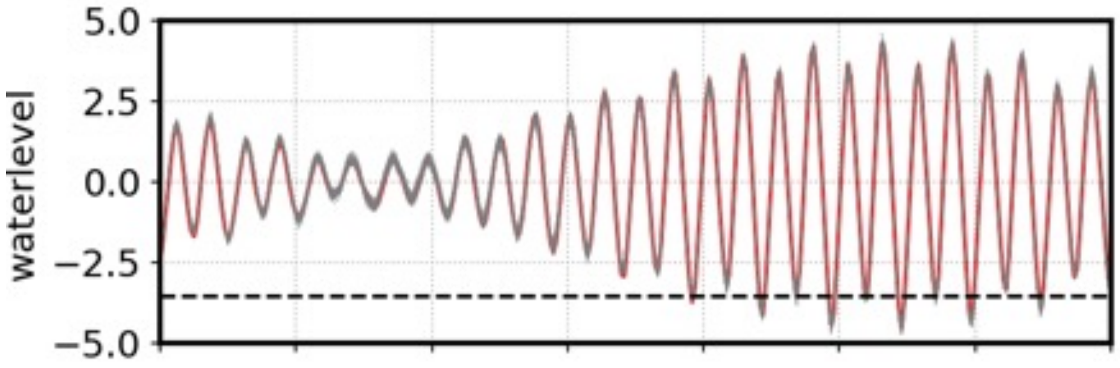
- Random draws from a multivariate normal (purple/gray) conditioned on the observation data (blue)
- We can push these draws (ensembles) into other models to quantify uncertainty in that model

Application Example

Uncertainty quantification of tidal forecast

- 14-d forecast at Broome tide gauge
- Gray lines are the stochastic predictions, red is truth

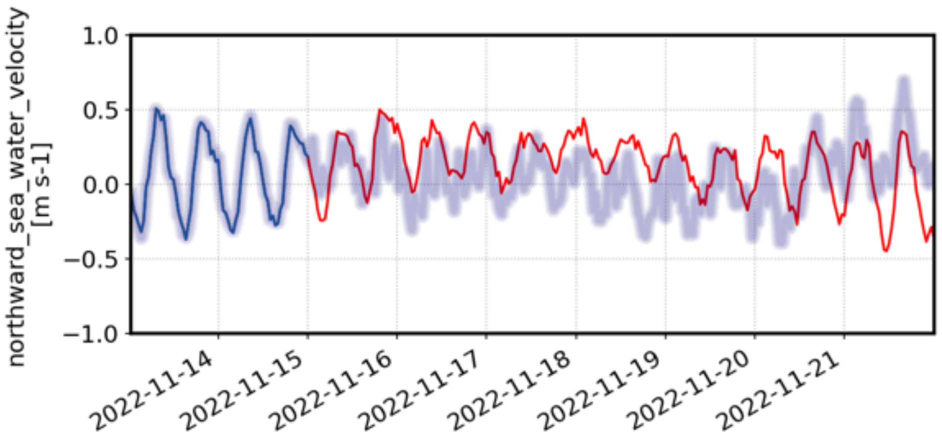
Broome Sea Level



<u>Example question:</u> *what is the probability of water level dropping below value -3.5 m in the next 14 days?*

Conclusions

- Machine learning method to make stochastic predictions with similar statistical (spectral) properties as observations of e.g. **sea level, currents, water temperature**.
  - → **Realistic looking** predictions
- **Uncertainty quantification** is a key by-product
- Extends the *useability* of tidal harmonic predictions to regions with mixture of tides and other ocean processes e.g. eddies, internal waves
- This method *has* been used operationally by offshore operators for ocean current prediction
- **Pushing model uncertainty into decision-making frameworks remains a hurdle**
  - requires forward propagation of many outcomes into response/decision models using Monte Carlo methods
  - This is time consuming and very specific to each use case
  - Ensemble forecasts suffer similar issues



matt.rayson@uwa.edu.au