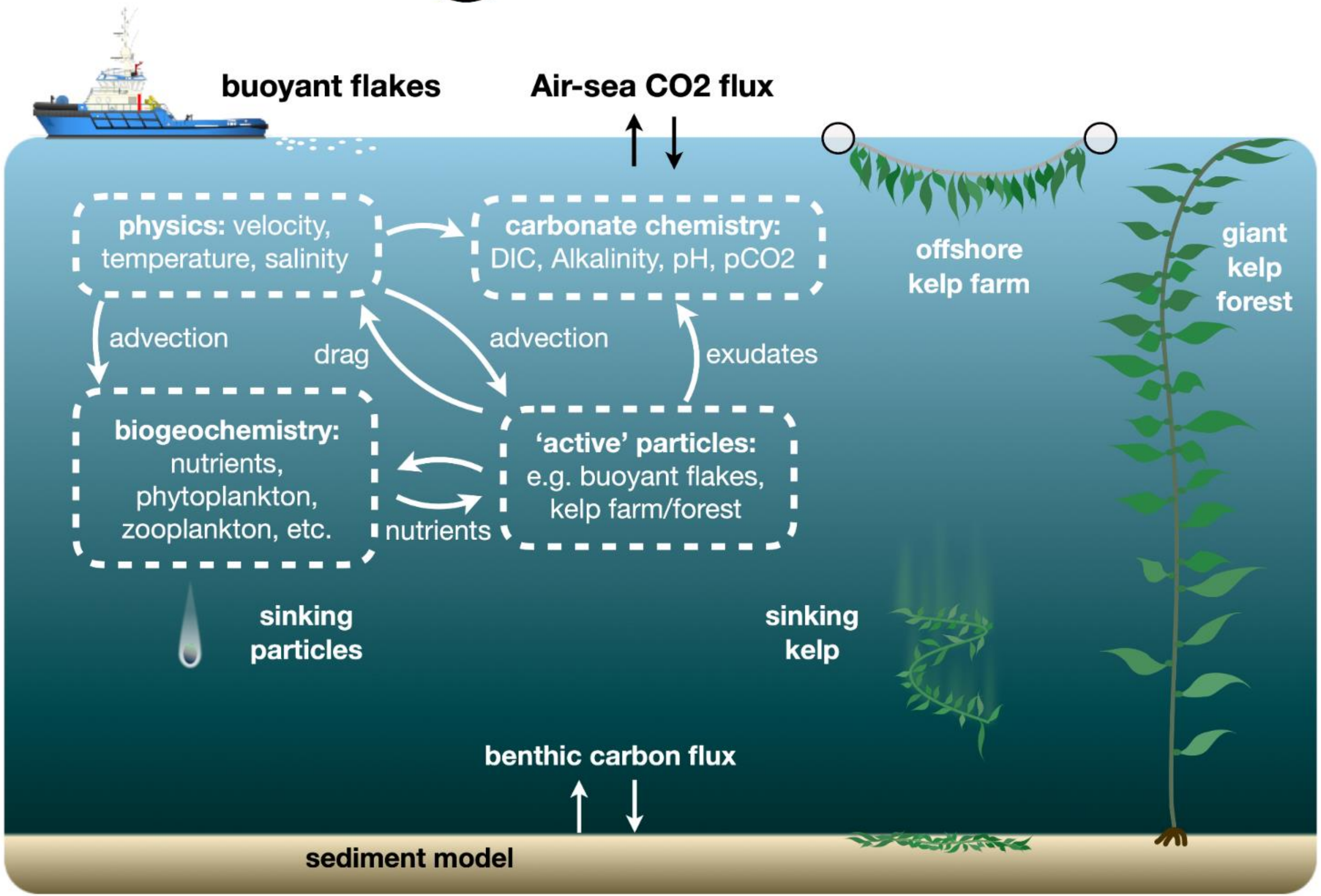
Parameter optimisation of ocean biogeochemical models

Chloe Y. Huang
Ocean Dynamics Group (DAMTP)

OceanBioME



julia

https://github.com/OceanBioME

Model setup & Aims

- Worked on two types of models -
  - Biogeochemical (BGC) models (eg. LOBSTER, PISCES, NPZD)
  - Carbon Chemistry model
- Hence, the aim was to build a tool that can optimise parameters given raw data (or a ‘truth run’) presented as the truth for both types of models
  - For BGC models, this meant optimising parameters that were informed by the ‘real world’
  - For the Carbon Chemistry model, this was optimising the model itself

Ensemble Kalman Processes (EKP)

- A class of derivative-free Bayesian optimization techniques based on Ensemble Kalman Filters (EnKF)
- For θ, y, where $\theta \in R^{N\theta}$ are the parameters and $y \in R^{Ny}$ is the noisy observation, we seek to optimise problems of the form

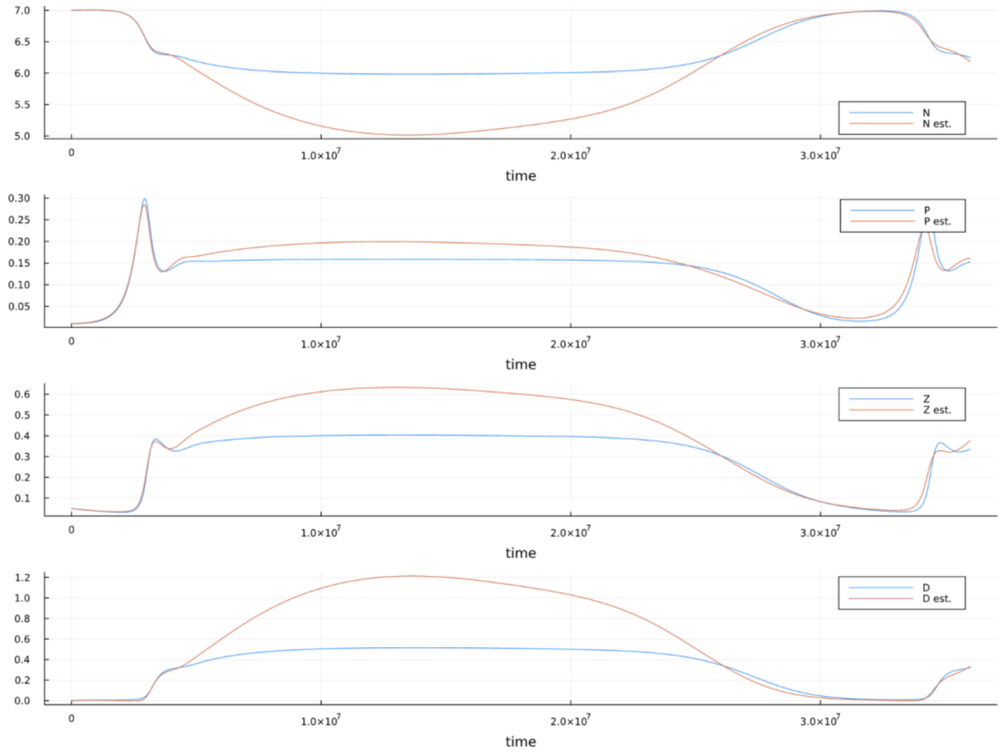
$$y = \mathcal{G}(\theta) + \eta.$$

where $\mathcal{G}$ is the forward map, and η is noise (which is ideally Gaussian)

- i.e seek to find the true value of θ

Application to the models

- EnKF approaches are often used in climate modelling as its strength lies in application to non-linear and non-Gaussian problems
- The models and data I am working with are very nonlinear with unknown, but bounded, error distribution.



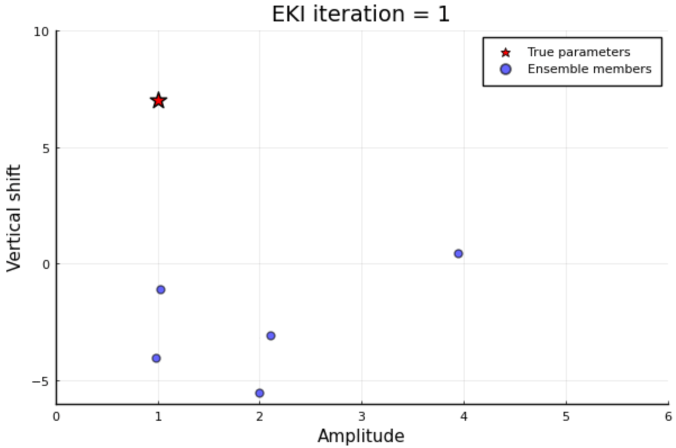
Process

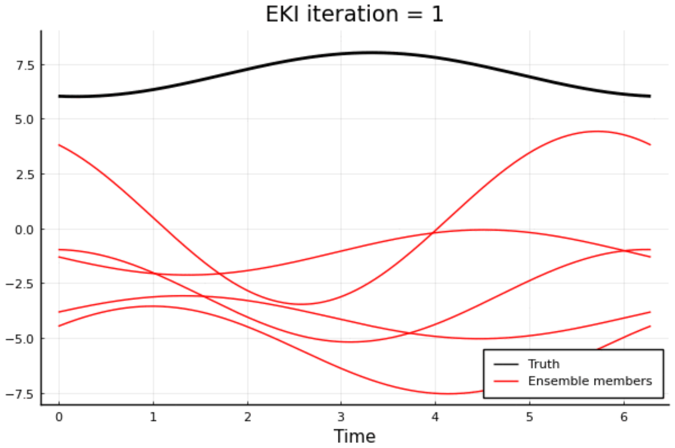
- Models produced results either as a timeseries (several vectors length 1000+) or as individual values of pH/pCO_2 corresponding to specific parameters (also totalling to vectors length ~10,000)
- Hence, it is not feasible to optimise over the raw data as output for the forward map
- Thus, for computational speed and otherwise, need to take statistics of the output (eg. mean, variance) to optimise over, which then I set as the forward map ($\mathcal{G}$)

Process cont.

Hence, the general process of optimising the model required the following:

1. Determine a prior mean and variance
2. Determine appropriate statistics to process the data
3. Generate truth data
4. Optimise parameters





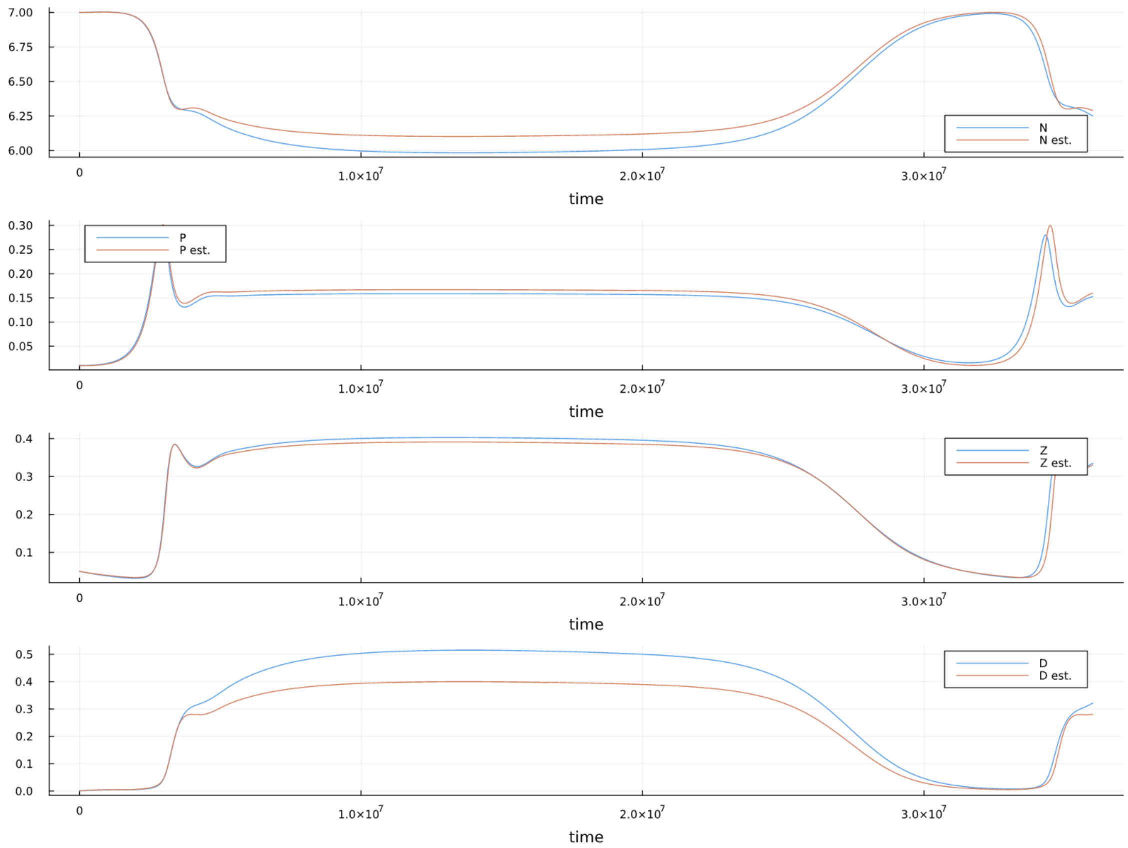
Source: https://clima.github.io/EnsembleKalmanProcesses.jl/dev/

Process - biogeochemical models

- Optimised over a ‘truth run’ with specified parameters and artificially added Gaussian noise
- Chose statistics of the timeseries such as max - min, rms, value at specific time, and time at which maximum was achieved
- Chose priors randomly to be MVN distributed with mean as the true parameters and standard deviation to be some proportion of the true parameters (here std = 0.3 * true)
- Error is given by $(\bar{g} - y)\Gamma^{-1}(\bar{g} - y)$

- Note that choice of forward map is up to the user in the developed tool

Results - NPZD model



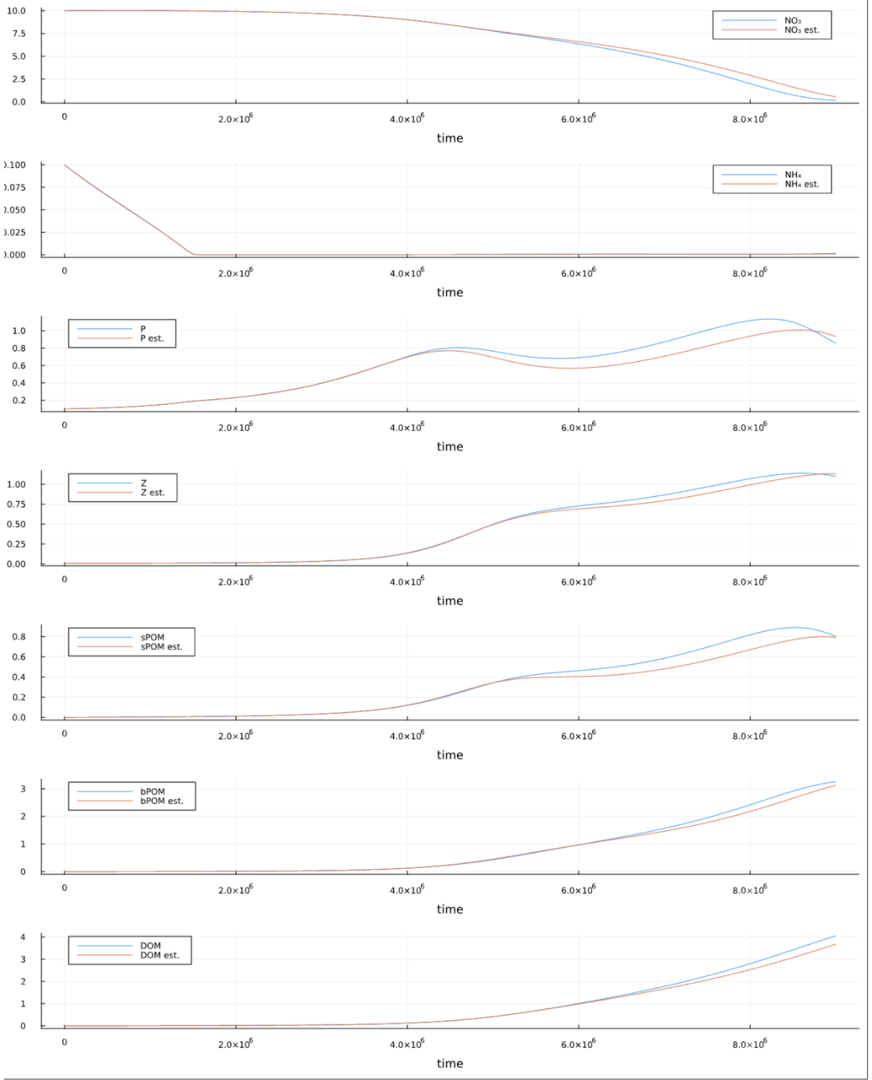

```
true params
pairs(::NamedTuple) with 11 entries:
  :initial_photosynthetic_slope => 2.26042e-6
  :base_maximum_growth          => 8.08912e-6
  :nutrient_half_saturation     => 2.3868
  :base_respiration_rate        => 7.63889e-7
  :phyto_base_mortality_rate    => 1.16898e-7
  :maximum_grazing_rate         => 2.49097e-5
  :grazing_half_saturation      => 0.5573
  :assimulation_efficiency      => 0.9116
  :base_excretion_rate          => 1.18056e-7
  :zoo_base_mortality_rate      => 3.9294e-6
  :remineralization_rate        => 1.40394e-6
------

final params
pairs(::NamedTuple) with 11 entries:
  :initial_photosynthetic_slope => 3.32702e-6
  :base_maximum_growth          => 9.32256e-6
  :nutrient_half_saturation     => 5.63589
  :base_respiration_rate        => 5.57802e-7
  :phyto_base_mortality_rate    => 4.38366e-7
  :maximum_grazing_rate         => 2.85554e-5
  :grazing_half_saturation      => 0.63103
  :assimulation_efficiency      => 1.04519
  :base_excretion_rate          => 1.6805e-7
  :zoo_base_mortality_rate      => 4.19398e-6
  :remineralization_rate        => 1.41295e-6
------
```

Results - LOBSTER model

```
true params
pairs(::NamedTuple) with 19 entries:
  :phytoplankton_preference             => 0.5
  :maximum_grazing_rate                 => 9.26e-6
  :grazing_half_saturation              => 1.0
  :light_half_saturation                => 33.0
  :nitrate_ammonia_inhibition           => 3.0
  :nitrate_half_saturation              => 0.7
  :ammonia_half_saturation              => 0.001
  :maximum_phytoplankton_growthrate     => 1.21e-5
  :zooplankton_assimilation_fraction    => 0.7
  :zooplankton_mortality                => 2.31e-6
  :zooplankton_excretion_rate           => 5.8e-7
  :phytoplankton_mortality              => 5.8e-7
  :small_detritus_remineralisation_rate => 5.88e-7
  :large_detritus_remineralisation_rate => 5.88e-7
  :phytoplankton_exudation_fraction     => 0.05
  :nitrification_rate                   => 5.8e-7
  :ammonia_fraction_of_exudate          => 0.75
  :ammonia_fraction_of_excriment        => 0.5
  :phytoplankton_redfield               => 6.56
------

final params
pairs(::NamedTuple) with 19 entries:
  :phytoplankton_preference             => 0.506737
  :maximum_grazing_rate                 => 9.81307e-6
  :grazing_half_saturation              => 0.87784
  :light_half_saturation                => 34.8271
  :nitrate_ammonia_inhibition           => 1.99728
  :nitrate_half_saturation              => 0.67116
  :ammonia_half_saturation              => 0.000835212
  :maximum_phytoplankton_growthrate     => 1.27035e-5
  :zooplankton_assimilation_fraction    => 0.613767
  :zooplankton_mortality                => 2.11147e-6
  :zooplankton_excretion_rate           => 6.12352e-7
  :phytoplankton_mortality              => 4.80005e-7
  :small_detritus_remineralisation_rate => 5.6358e-7
  :large_detritus_remineralisation_rate => 5.83504e-7
  :phytoplankton_exudation_fraction     => 0.0641984
  :nitrification_rate                   => 6.05876e-7
  :ammonia_fraction_of_exudate          => 0.792522
  :ammonia_fraction_of_excriment        => 0.532765
  :phytoplankton_redfield               => 6.63159
------
```



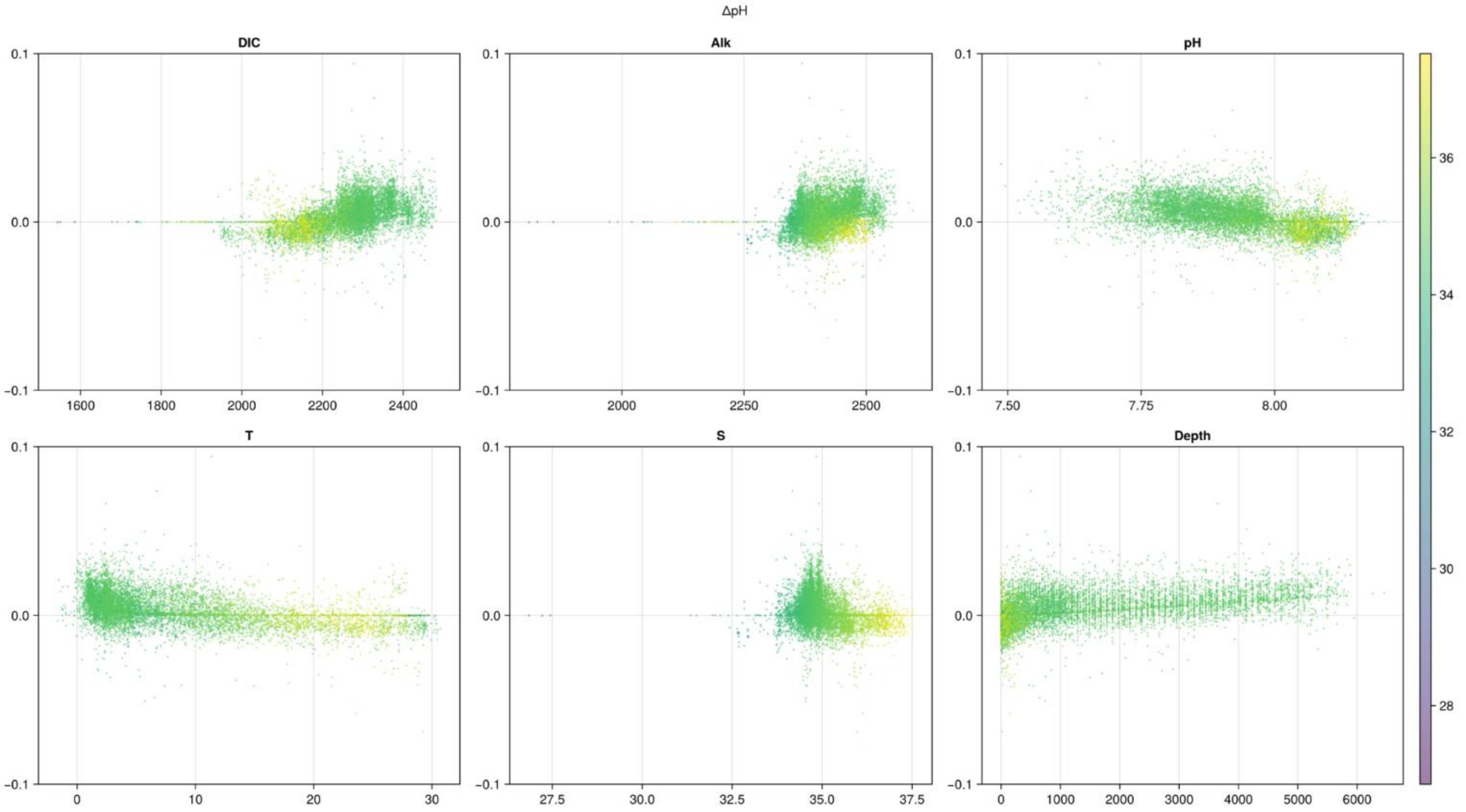
Process - CarbonChemistry model

- Optimised the model itself over GLODAPv2 (Global Ocean Data Analysis Project) data
- The Carbon Chemistry model contains series of several equations which describe equilibrium constants

eg. Solubility constant K0:
$\ln(k_0/k^\circ) = -60.2409 + 9345.17 / T + 23.3585 (\log(T) - \log(100)) + 0.0 T^2 + (0.023517 + -0.00023656 T + 4.7036e\text{-}7 T^2)S$

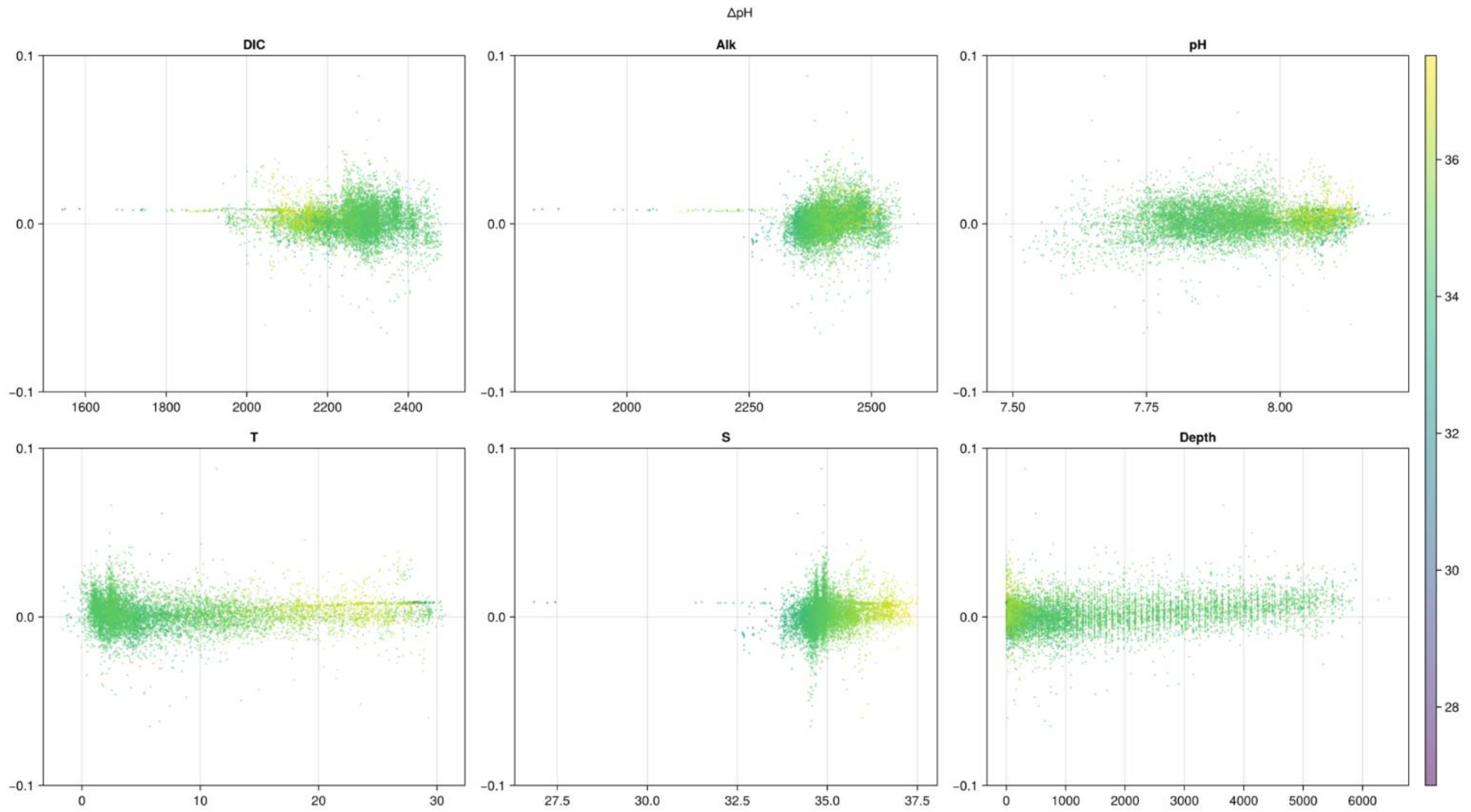
- Optimised over the coefficients of these equations to minimise error between predicted pH and pCO_2 of the model and GLODAP measurements

Progress - Carbon Chemistry



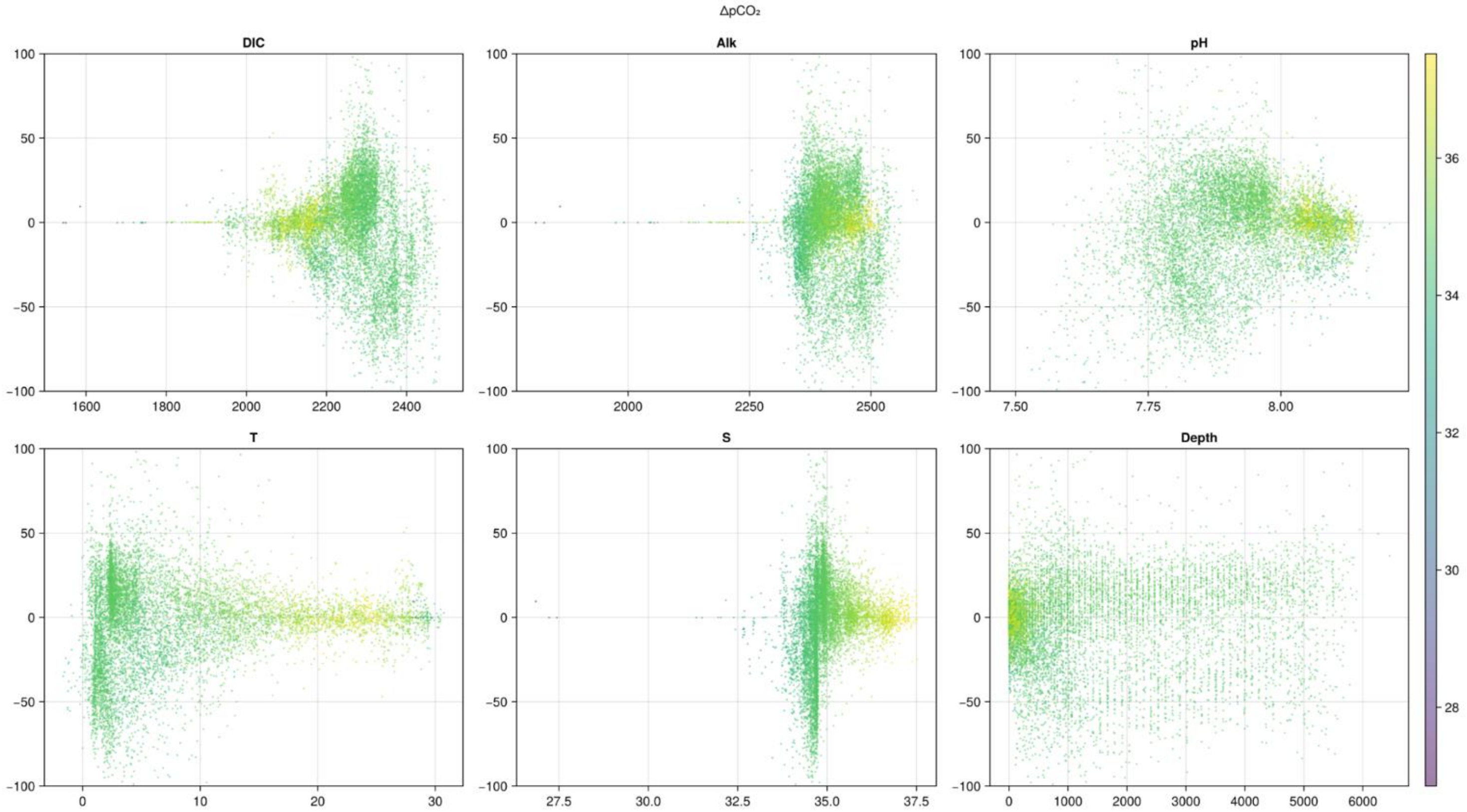
pH error of original model

Progress - Carbon Chemistry



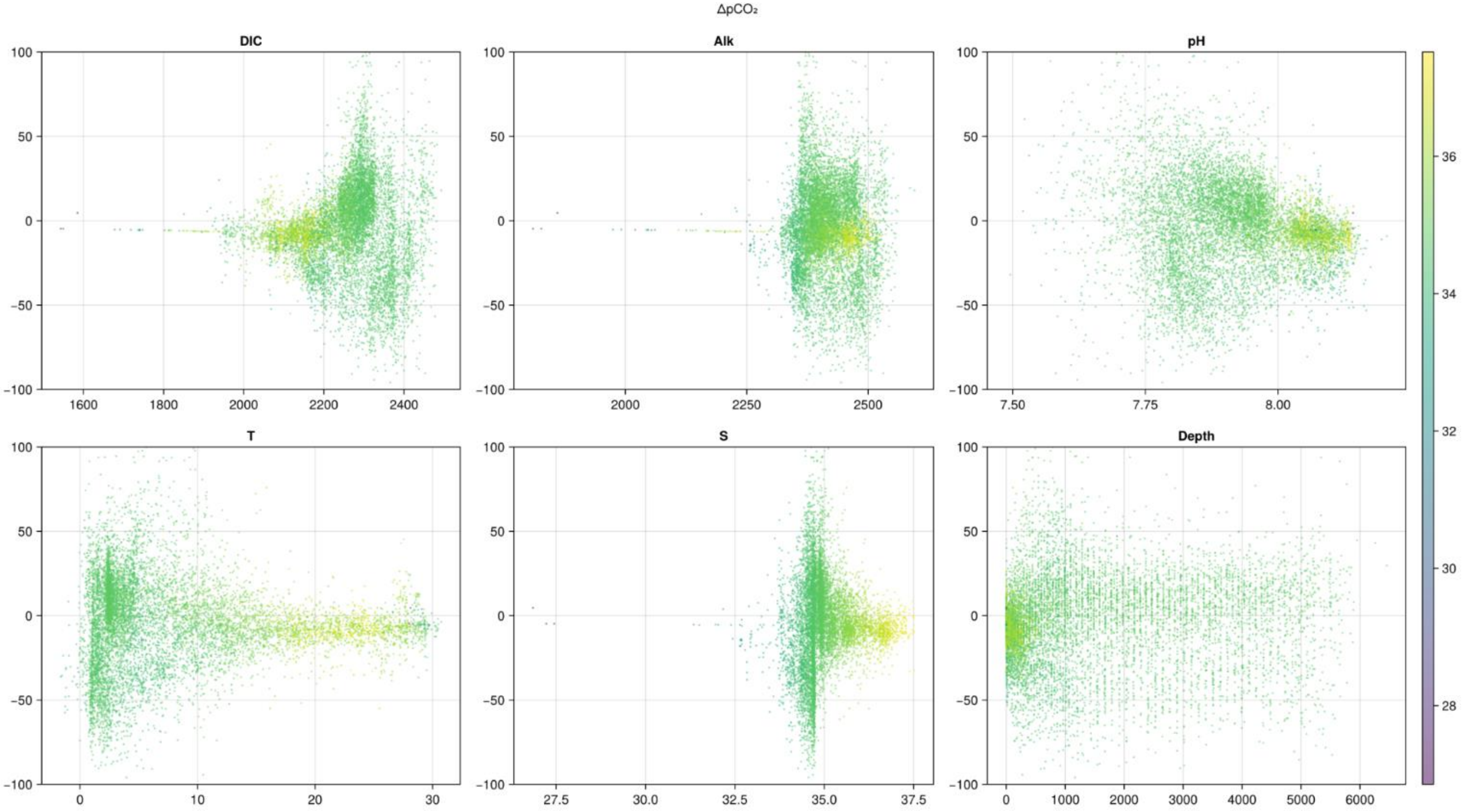
pH error of new model

Progress - Carbon Chemistry



pCO_2 error of original model

Progress - Carbon Chemistry



pCO_2 error of new model

Reflection - CarbonChemistry

- Overall, results of optimisation were subpar
- Realised that EKP was not necessarily the best approach for this type of problem - the model was much too unstable for convergence and often returned a root error
- the model itself required multidimensional data inputs (in this case taken from the GLODAP measurements) and meant that the choice of statistic to take from raw pH/pCO_2 data was not obvious
- Lacked knowledge in statistics/the ocean and time to make significant progress on any alternative approaches

Reflection

- Learnt in great detail on how to program in Julia and how computational models are created (eg. wrote a timestepper)
- Thought deeply about optimising code and complexity
- Applied statistical methods and computational methods to real world data

In the future:

- Would like to try some new approaches to optimising the CarbonChemistry model
- Get more precise results on a more complicated model like combined PISCES + physics

References

- EnsembleKalmanProcesses.jl, Available at: https://clima.github.io/EnsembleKalmanProcesses.jl/dev/

- Huang, D.Z., Schneider, T. and Stuart, A.M. (2022) ‘Iterated Kalman methodology for inverse problems’, Journal of Computational Physics, 463, p. 111262. doi:10.1016/j.jcp.2022.111262.

- Strong-Wright, J. et al. (2023) ‘OceanBioME.jl: A flexible environment for modelling the coupled interactions between ocean biogeochemistry and physics’, Journal of Open Source Software, 8(90), p. 5669. doi:10.21105/joss.05669.