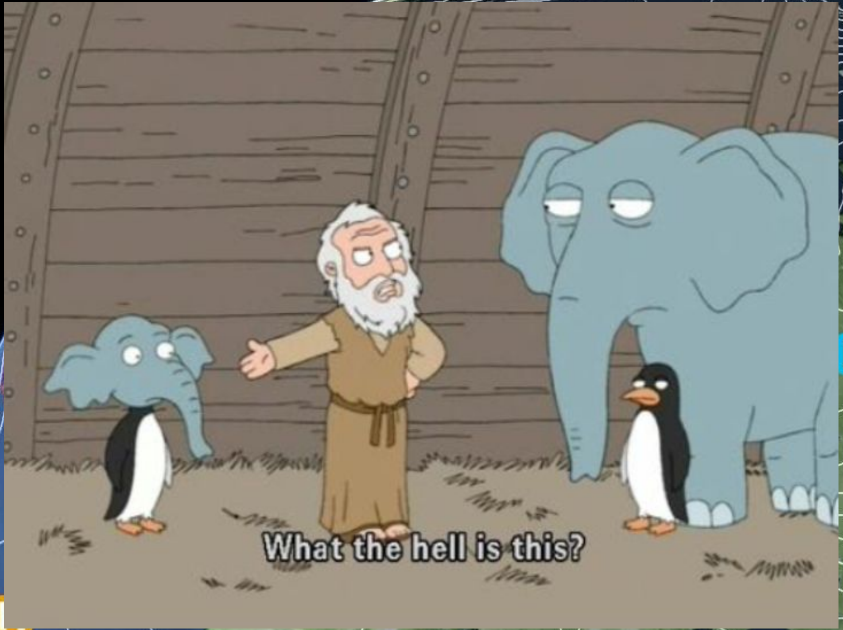
Met Office

A hybrid physical-ML model?

Ian Boutle, Mohit Dalvi, José Rodríguez, Martin Willett, Stephen Haddad, Tom Melvin, Matthew Chantry (ECMWF), Cyril Morcrette, Simon Vosper



© Crown Copyright 2025, Met Office

Met Office

Where we are currently

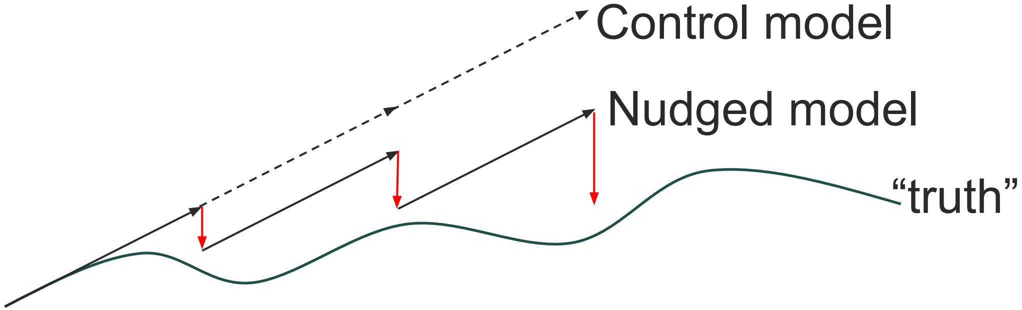
- AI/ML models are very fast
- They are becoming very good, at a limited set of diagnostics

- Physical models are also very good, and are fast enough for most purposes
- They provide a much wider and more comprehensive set of diagnostics, driving a whole range of customer requirements and downstream products

- Can we combine the two to give us the best of both worlds?

Met Office

Nudging



- Long used technique for coping with model error / drift in physical model simulations
- Add a small increment to temperature and wind fields which push them back towards some known “truth”
  - Physical model will respond in a consistent manner to this increment (as it does to increments from ‘real’ physical processes), and update all other variables accordingly
- Typically, “truth” would be, e.g. (re)analysis
- Can easily replace the (re)analysis with output from an ML model
  - If the ML model is better than the physical model at the nudging variables, this improvement will be imparted on the physical model
  - Because the ML model is so fast, this can be done in real-time to produce forecast products

Met Office

Spectral nudging

- Based on Uhe & Thatcher (2015, GMD)
- Spectral low-pass filter adjusts large wavelengths, leaving small wavelengths unperturbed – choice of filter scale
- 2D latitude/longitude filter too expensive to implement – approximated as 2x 1D filters in each direction (latitude first)
- Nudging applied every 12 mins (3 timesteps at N1280)
- Nudging tapered or stopped near surface and model top
- Re-grid ML output onto model levels for nudging (could be better if ML predicted our model level values)

Met Office

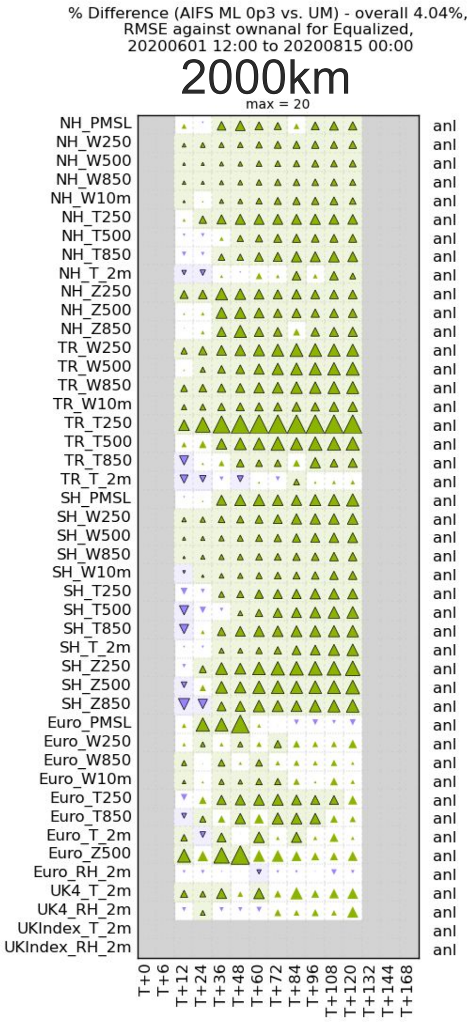
What we've done

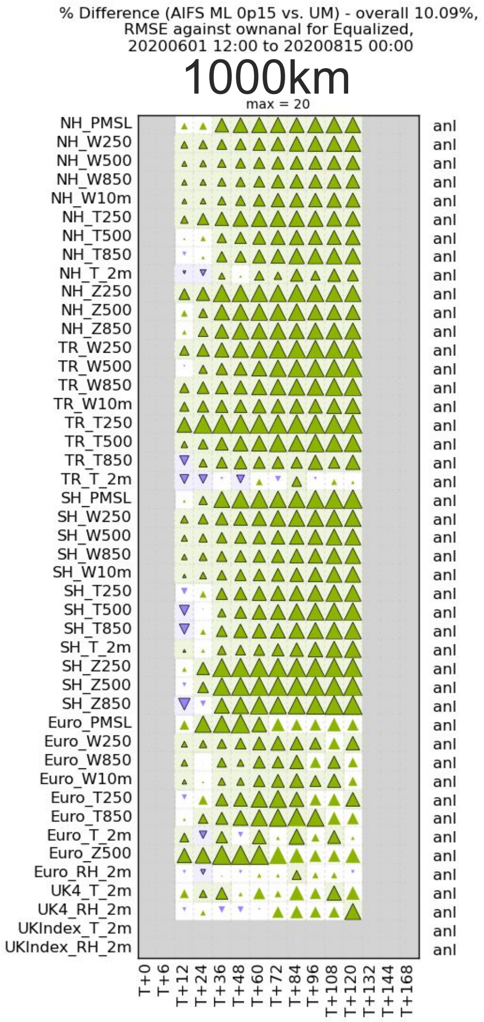
- As with everything ML, once you start doing something new, someone else comes along having already done it!
  - Environment Canada (Hussain et al) – nudge their model (GEM) with GraphCast
  - ECMWF – nudge their model (IFS) with their ML model (AIFS)
- Nudge our model (UM) with AIFS initialised by ERA5
  - Also looked at Pangu and NeuralGCM – AIFS is best
- Typical NWP 'first look' testing strategy
  - 24 non-overlapping 6 day case studies initialised from operational analysis
  - 12 winter and 12 summer cases (focus here only on the summer cases)
  - N1280 (~10km) operational resolution

Met Office

Verification

- N1280 UM nudged with AIFS:
  - ~1000km nudging scale, NWP index +6% (obs), +10% (analysis)
  - ~2000km nudging scale, NWP index +2.5% (obs), +4% (analysis)
  - ECMWF/EnvCanada etc all using 2000km
- Slight degradation at early times can be mitigated by ramping up the nudging with time

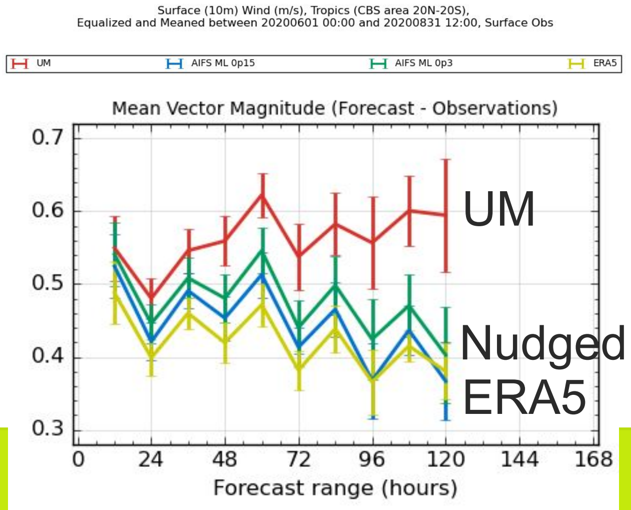


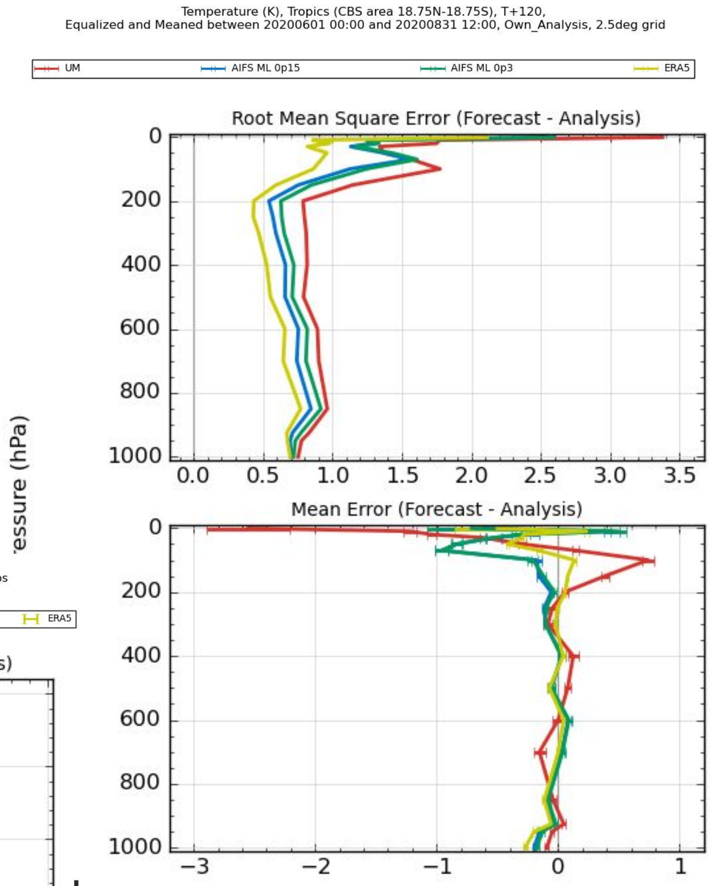


Met Office

Genuine improvement

- Improvements shown in RMS error do appear as genuine improvements for many fields
- Reduced mean errors also seen
- Nudging gets the mean state right, physics can worry about variability?

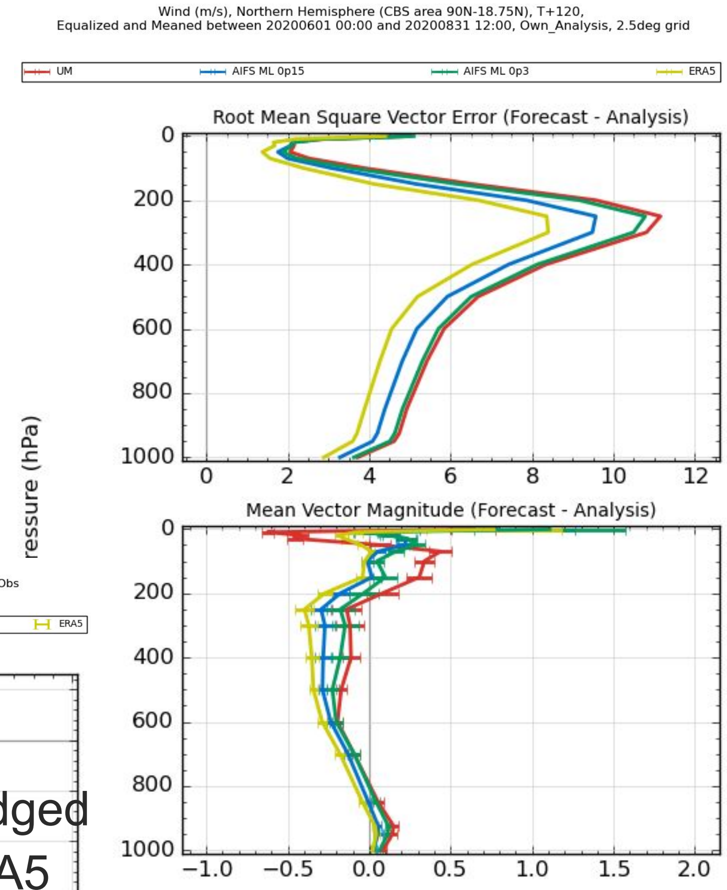


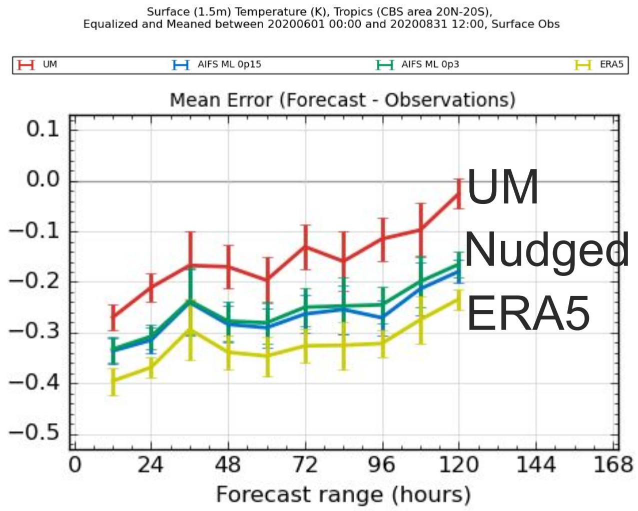


Met Office

Biases inherited from IFS

- Whilst RMS error generally looks good, some variables show increased mean-bias
- Largely this comes from the IFS/ERA5 biases
- Pulling our model in wrong direction in places it's better than ECMWF

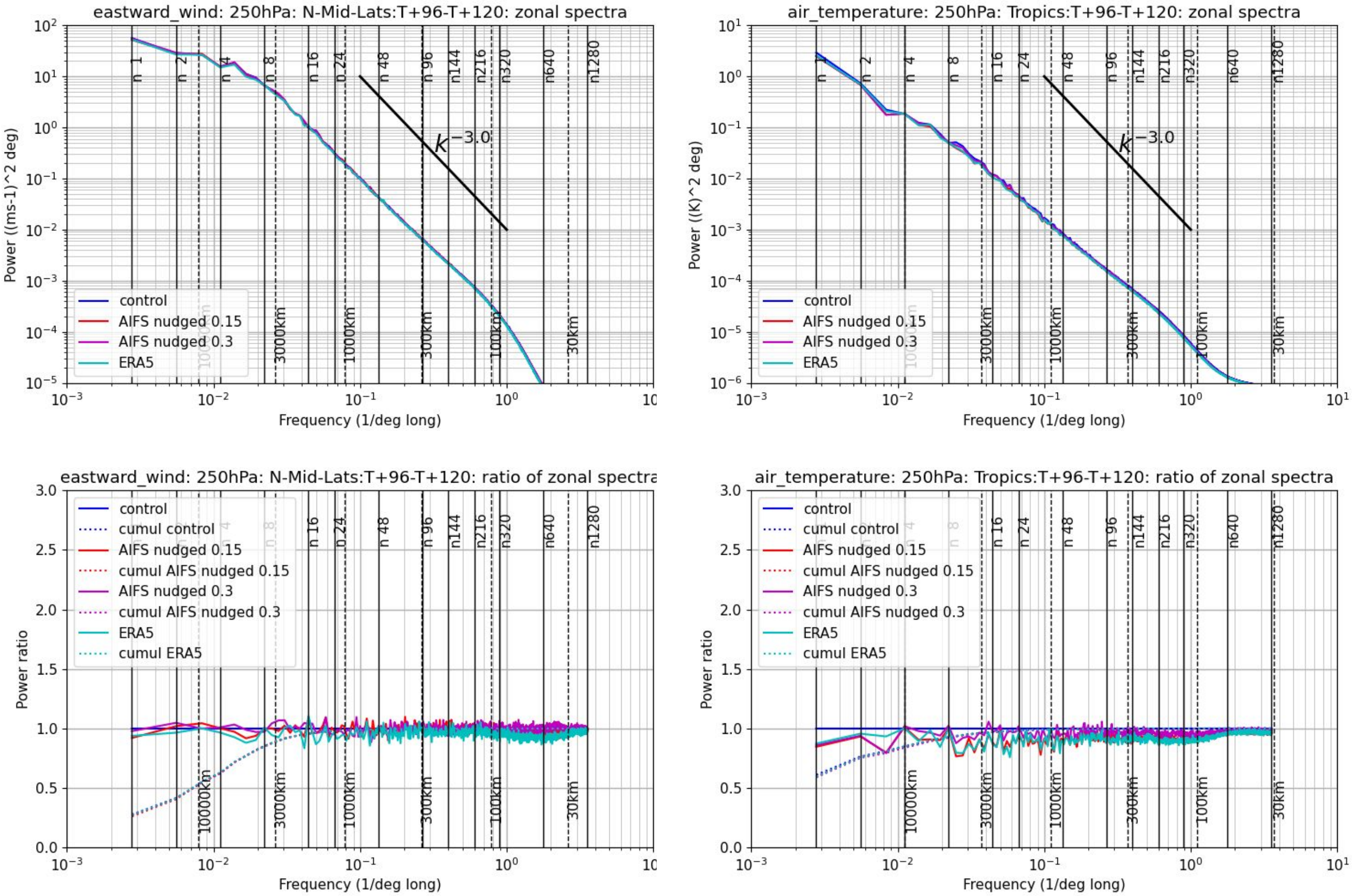




Met Office

Spectra

- Wind spectra look very close
- Some very slight loss of power in temperature spectra – is this from ERA at large scale?
- Small scale must be a response to large scale?

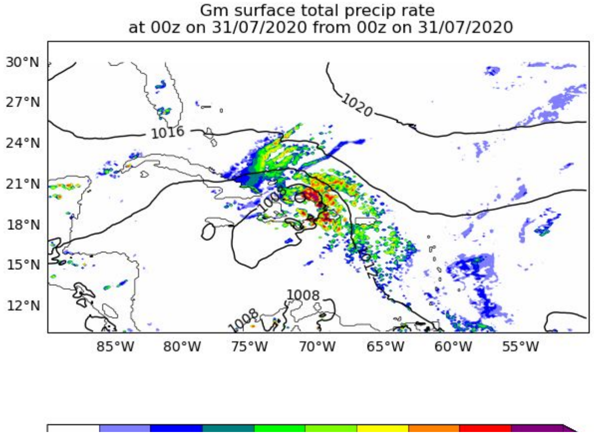


Met Office

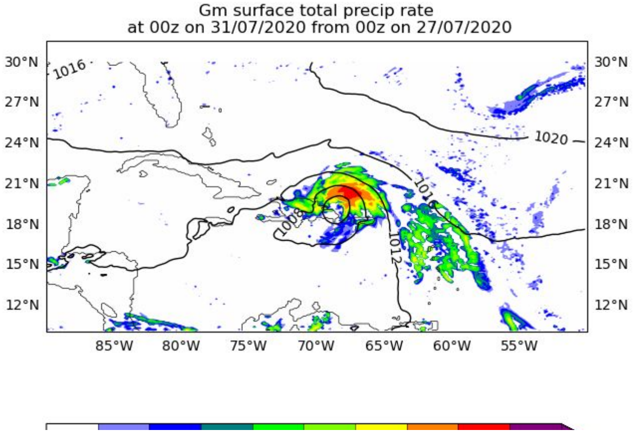
Forecast charts

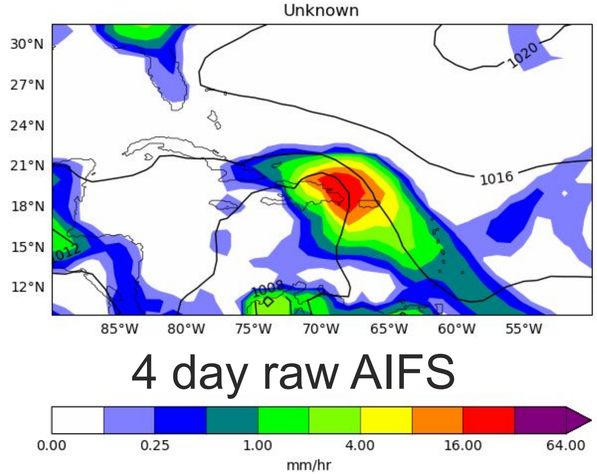
- All look better than raw AIFS
- Degraded TC intensity at 1000km

Analysis

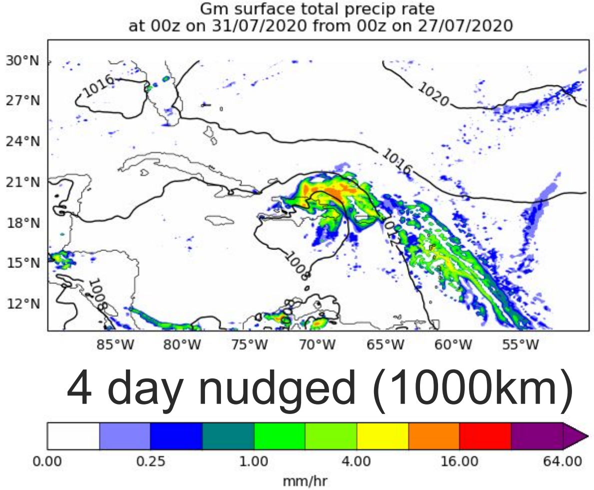


4 day raw UM

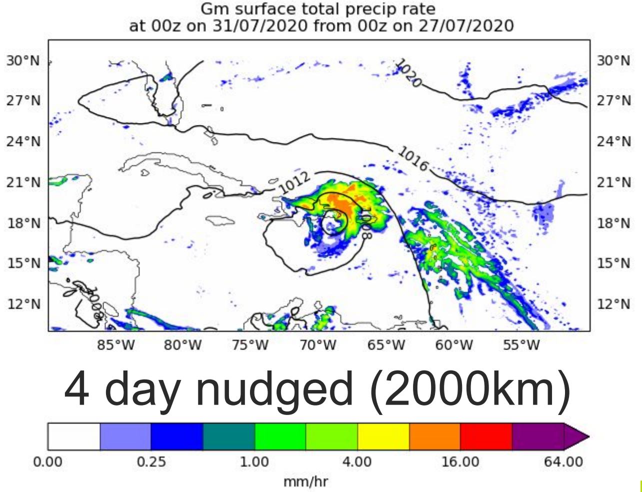




4 day raw AIFS



4 day nudged (1000km)



4 day nudged (2000km)

Discussion

Pros:

- Obvious statistical improvement to model
- Everyone else is doing it – risk being left behind if we don't
- (Genuine improvement in long-range skill)
- (Driving UKV from this will improve skill there)

Cons:

- Ties our model too closely to ECMWF
- Not great in places where UM is better than IFS
- Smoothing still apparent in some features

The ML model:

- Ideally would use Fastnet – can train on vertical levels of our model
- Still going to inherit biases from ECMWF – can fine-tuning with our own analyses improve this?
- Would have to initialise from our analysis in real-time

Met Office

Summary

- Very easy method of improving model skill and capitalising on enhanced large-scale skill shown by ML models
- ML is basically acting like a sophisticated bias-correction
- Needs lots more analysis of other objective & subjective measures of skill

- With an ensemble ML model, could easily apply technique to physical ensemble system
  - Would you actually need to run all members of the physical ensemble, or could you choose cleverly to sample the distribution?