Slide 1

The Bureau
of Meteorology

Using Dynamic Ensembles and Machine Learning to Determine Tropical Cyclone Forecast Position Confidence Areas

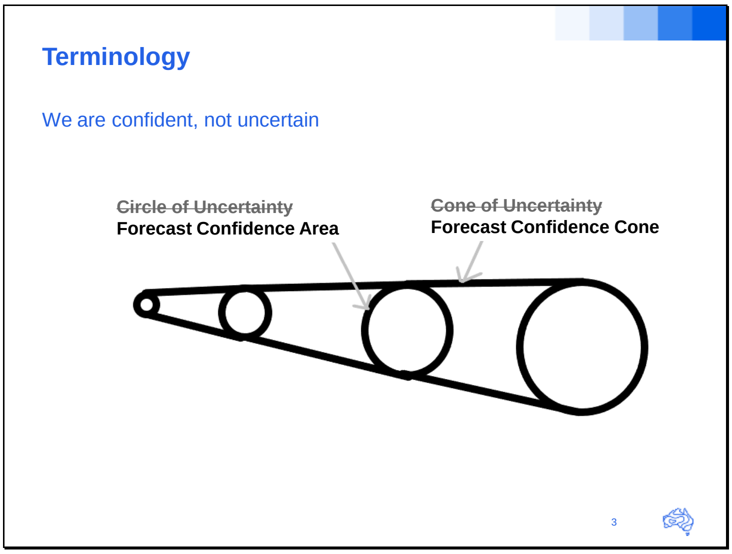
Rabi Rivett, Matt Boterhoven, Andrew Burton & Craig Earl-Spurr

Slide 2

Outline

- What is a confidence area? Some new terminology.
- What is a super ensemble?
- Deriving confidence areas using machine learning.
- Calibration.
- Comparison of old and new method.
- Uncertainty.
- Use in products.

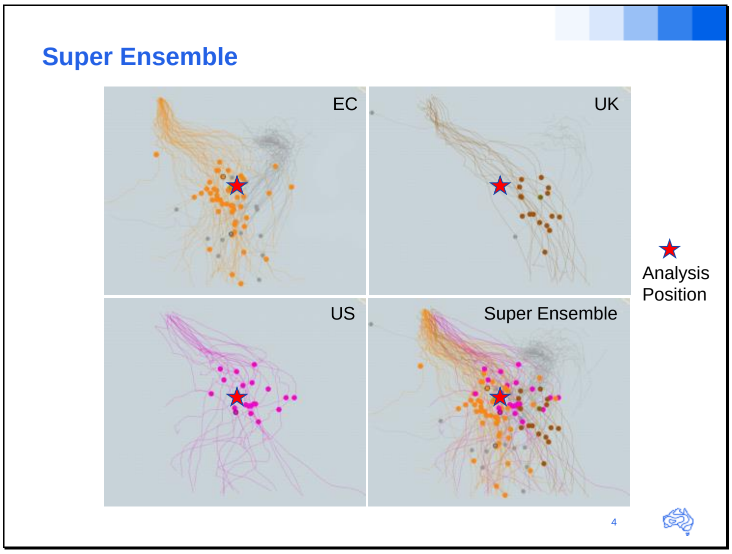
Slide 3



A standard way of defining forecast position uncertainty is with a distance, where the distance is based on a percentile of historical error. This is the circle of uncertainty and is the area in which the centre is expected to be a certain percentage of time. A swath of circle of uncertainties for a time range is called the "cone of uncertainty", and this is what is included in a lot of forecast track products.

But we want to use a positive term, therefore we are confident, not uncertain. So we are moving towards using the terms Forecast Confidence Area for a time and Forecast Confidence Cone for a time range.

Slide 4

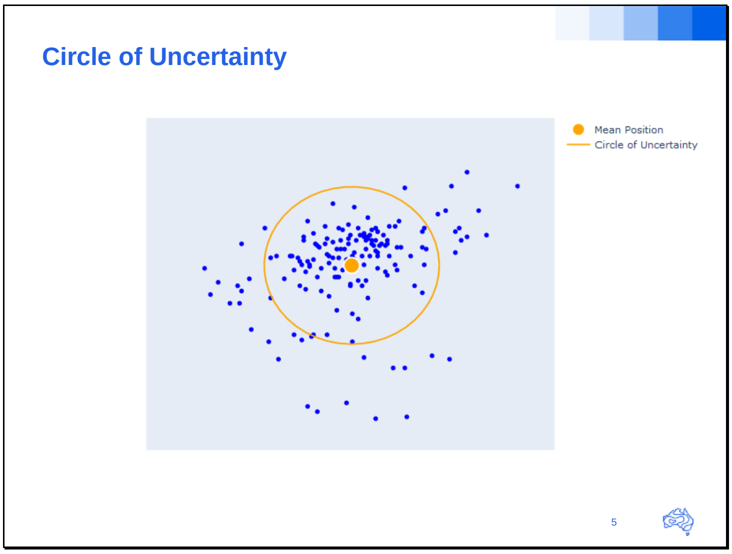


A super ensemble is an ensemble of ensembles, possibly lagged. This slide shows a super ensemble of EC, UK and US ensembles. The red star is the analysis position, and we see that the analysis position is near the edge or just outside the area of ensemble fixes.

In the Super Ensemble panel, we see that the analysis fix is in the middle. This is a cherry-picked example, but it does show that a consensus of ensembles gives a better forecast. One thing to note is that region cover by super ensemble fixes is larger than any of the individual ensembles.

In the following results and slides, the shown super ensemble is a combination of two runs of EC, UK, US and ACCESS-G ensembles.

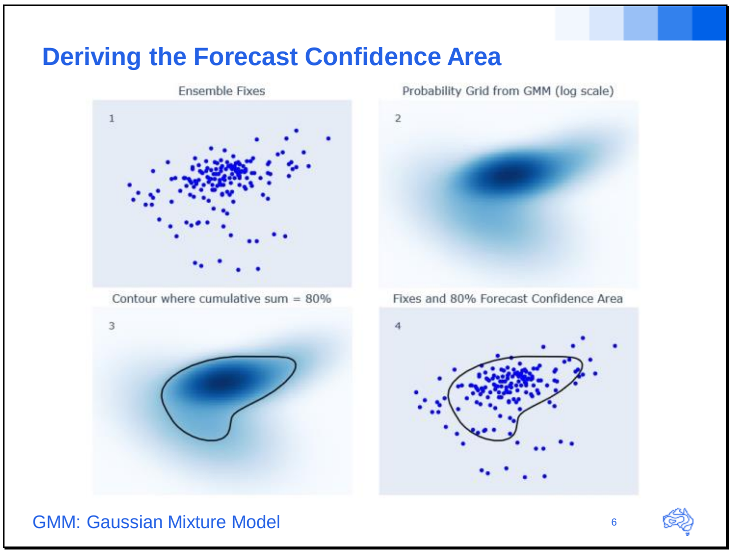
Slide 5



The standard method of deriving the forecast position and uncertainty is to derive the mean position from a consensus of models, and then use the historical error to define the uncertainty. This gives a circle of uncertainty.

Comparing the circle to the ensemble points, we see the circle method doesn't follow the dynamics of the situation. We can see the circle extends a bit too far north and southeast compared to the ensemble points.

Slide 6



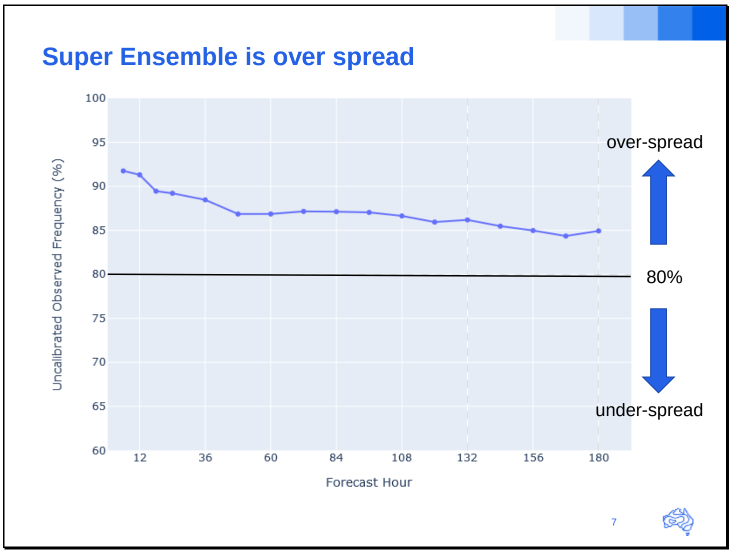
In panel 1 we have points from a super ensemble for a specific time.

In panel 2 we a probability grid, which is the probability of the centre being at that grid point. This is derived variational Bayesian estimation to derive a weighted Gaussian Mixture Model, which is the machine learning part of the method. A Gaussian Mixture Model is a weight sum of a number of different gaussian distributions with different positions, sizes and orientations.

In panel 3 we have the contour where the sum of all grid cells in the contour is 80%. This is where the centre is expected to be 80% of the time and this is what we call the 80% Forecast Confidence Area.

In panel 4, we can see the ensemble fixes and the confidence area.

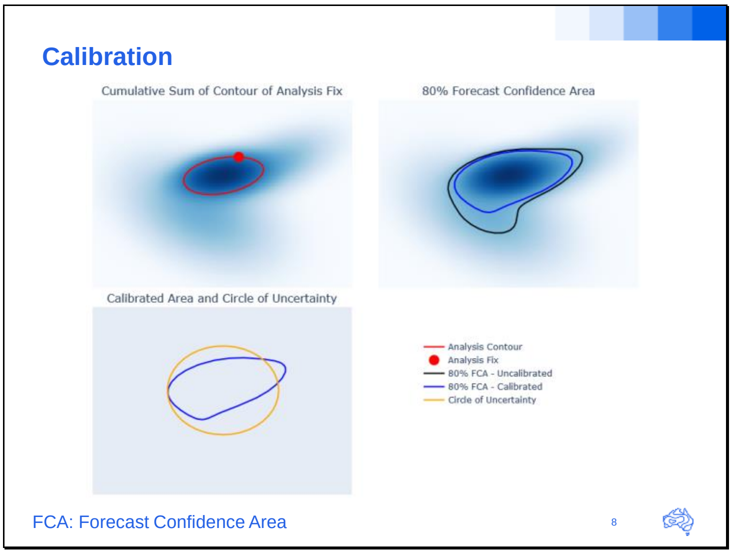
Slide 7



If we look at a couple years of past super ensemble forecasts and analysis positions, and calculate the frequency at which the analysis position is within the 80% forecast confidence area, we find the observed frequency is higher than 80%.

This indicates that the super ensemble is over-spread, which means we need to calibrate.

Slide 8



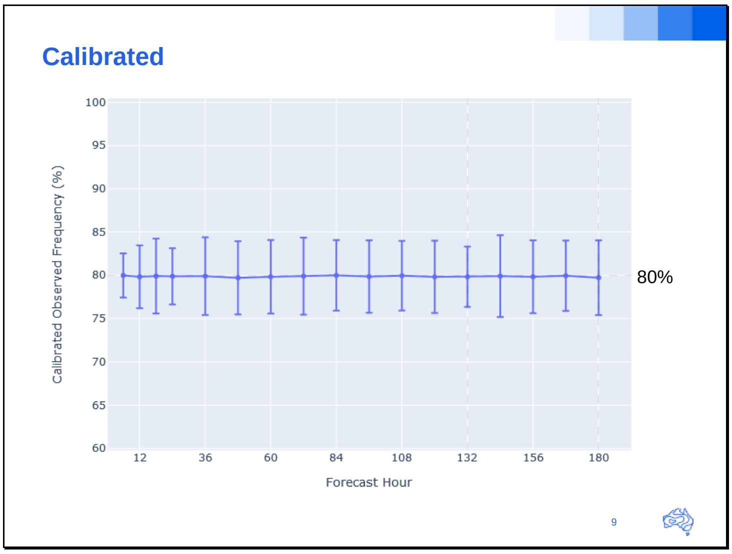
Since the super ensemble is over-spread, we need to calibrate.

In the top left panel, the red dot is the analysis fix and the red ellipse is the contour of the analysis fix. We find the sum of all grid cells within the contour and call this the analysis cumulative probability. To calibrate, we find the analysis cumulative sum values for a couple years of past forecasts and find the 80th percentile of these, which we call calibrating cumulative probability. Depending on the forecast hour, this is a value like 72%.

For a forecast probability grid, if we find the contour which sums up to that 72% value, instead of 80%, then we have a calibrated confidence area.

The top right panel shows the black uncalibrated area and the smaller blue calibrated area. The bottom left panel shows the calibrated area and the circle of uncertainty, showing an overall smaller shape.

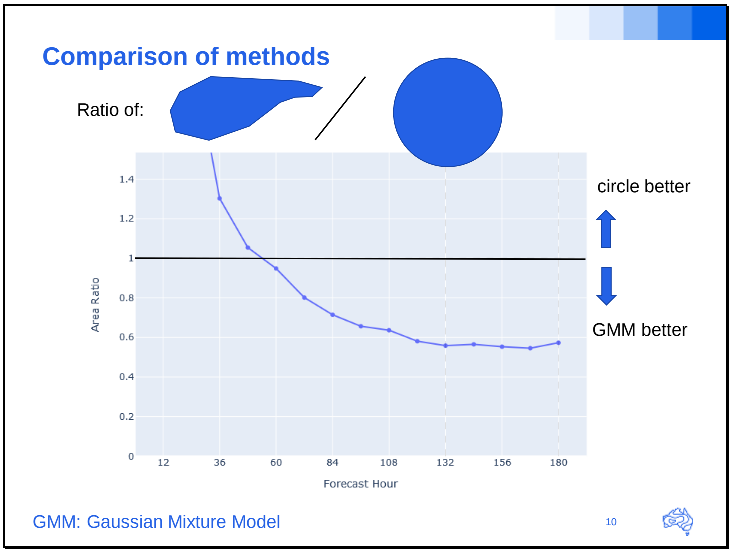
Slide 9



So what does the observed frequency look like now?

Using k-fold cross validation, we find that the mean observed frequency is very close to the target 80%, with uncertainty of about 5%.

Slide 10



What about comparing old to new?

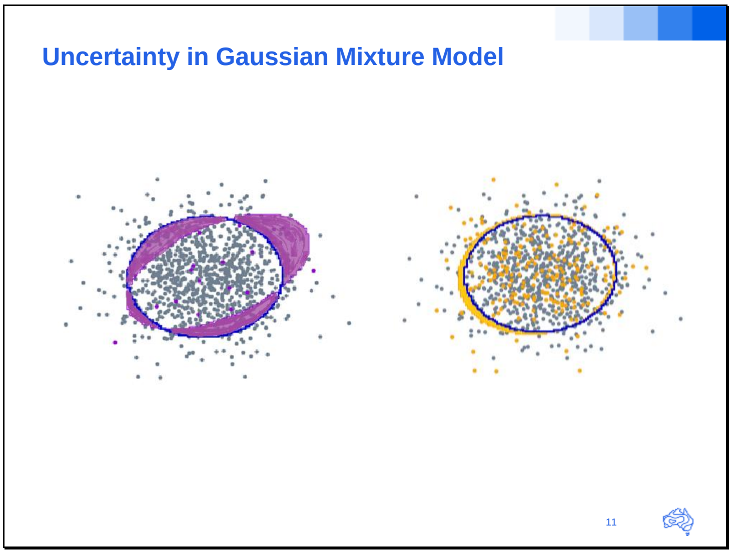
The metric we are using is the average spatial area of the confidence area, assuming both methods are calibrated. The smaller the area that needs to be covered to achieve the same observed frequency, the better the method.

This chart shows the ratio of the spatial area of the gaussian mixture model derived confidence area and the circle of uncertainty. The gaussian mixture model confidence areas are derived for a super ensemble. The circle of uncertainty is the 80th percentile of error of the shifted super ensemble mean. It is shifted using an analysis position at forecast hour 12.

If the ratio is greater than 1, then the circle method is better, and if it is less than 1, the GMM method is better. We expect the shifted ensemble mean to outperform guidance which is 12 to 18 hours old, so the ratio is large in the short term.

We see that before forecast hour 54, the circle of uncertainty is better, while after than the gaussian mixture model confidence area is better. Out operational forecasts use a blend of these two methods.

Slide 11

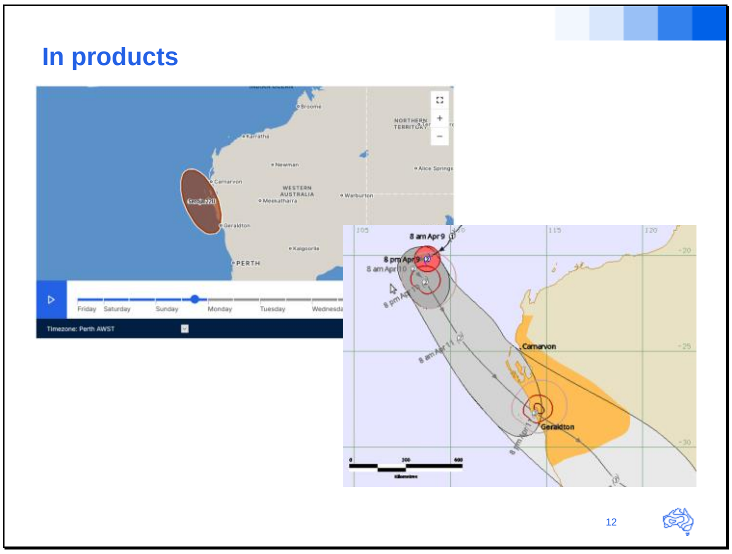


On these charts, the grey dots are a simulated 1000-member ensemble, and the blue oval is the corresponding Confidence Area. This can be considered truth.

The purple dots are a small subset of 1000 member ensemble, and the purple area is the difference between confidence area derived from the purple dots and the grey dots. Likewise for the yellow dots and area, but this time with a lot more members. If we do this experiment enough times, we can determine uncertainty in the forecast confidence area as a function of the number of members in the super ensemble. And like these two examples show, the more members the less the uncertainty.

This is a reason why our standard super ensemble combination includes as many members as possible, including lagged members, that do not decrease the skill.

Slide 12

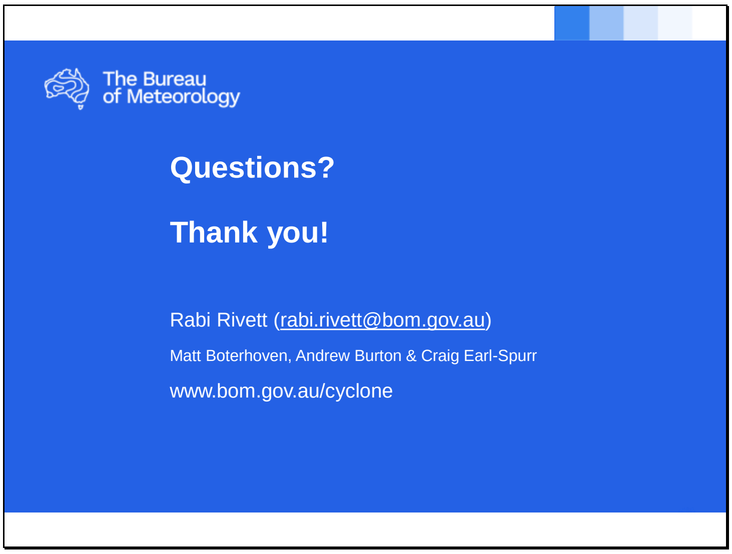


Top left is our 7 day tropical cyclone forecast, which is an outlook product. It consists of Forecast Confidence Areas and the probability of the system being at "tropical cyclone strength" every 12 hours for the next 7 days. "Tropical cyclone strength" means gale force winds more than half-way around the centre.

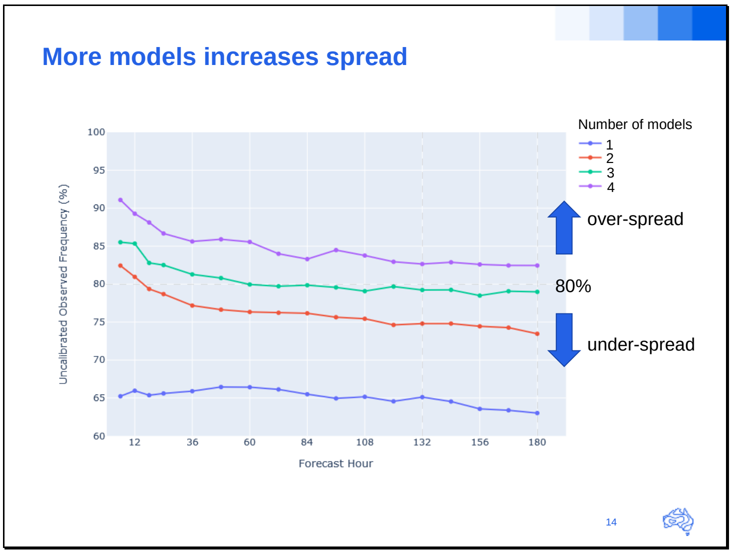
Bottom right is our Forecast Track Map which is issued when warnings are needed, so about 48 hours from causing gales on the coast or 24 hours of being a tropical cyclone, and covers the next 5 days. The Forecast Confidence Cones, otherwise known as "cone of uncertainty" is derived from 6 hourly Forecast Confidence Areas.

The two products use the same Forecast Confidence Areas and remain consistent.

Slide 13


The Bureau
of Meteorology
Questions?
Thank you!
Rabi Rivett (rabi.rivett@bom.gov.au)
Matt Boterhoven, Andrew Burton & Craig Earl-Spurr
www.bom.gov.au/cyclone

Slide 14



More models increases spread
Number of models
1
2
3
4
over-spread
80%
under-spread
Uncalibrated Observed Frequency (%)
100
95
90
85
80
75
70
65
60
12
36
60
84
108
132
156
180
Forecast Hour
14