Climate and Oceans Support
Program in the Pacific

ACCESS-S Workshop

MODULE: ACCESS-S Model Design Forecasts/Hindcasts

Climate and Oceans Support Program in the Pacific

Topics in this module

- Hindcasts, what are they and why do we need them?
- Forecasts – design
- Forecasts - Ensemble size and lagged ensembles
- How do we make a forecast
- Example forecasts

Expected learning outcomes

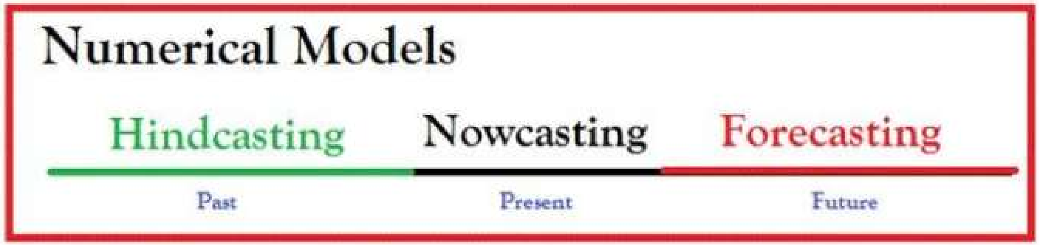
- Understanding of how forecasts and hindcasts are calculated in ACCESS-S
- Terminology associated with ensemble forecasts

These outcomes are important for understanding and interpreting ACCESS-S outputs and products, particularly model skill

What is a hindcast and why do we need it

A hindcast is a retrospective forecast or historical re-forecast

- Model skill (run the model lots of times to find how well it performs)
- Model climatology (e.g. calculate anomalies)



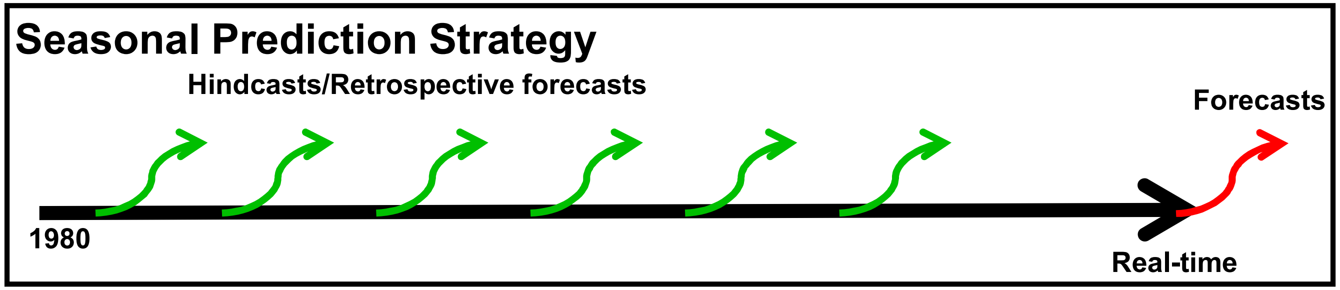
How is a hindcast generated

- Choose date in past and only use observations available then to estimate state

- Use model to project forward in time

- Repeat for many dates (e.g. multiple forecasts over a 30-year period). Important for climate so that we have a good sample size of infrequent climate features such as ENSO.

Why do we need hindcasts?



- Model bias can be estimated and removed using hindcasts
- Hindcast set:
  → provides estimate of how model climatology changes with lead time. Calibrate forecasts
  → needed for skill assessment (e.g. need enough samples of ENSO)
  → used for interfacing with downstream applications models e.g. crop, streamflow models
- NB. Need initial conditions for the hindcast set!

Climate and Oceans Support Program in the Pacific

ACCESS-S1 Hindcast design

Hindcast design

23 years: 1990-2012
4 start-dates every month: 1^{st}, 9^{th}, 17^{th}, 25^{th}
11-member burst ensemble
Lead times out to 6 months
Created about 500 TB of data!!

Larger ensembles than used by UKMO

Why only 23 years? Why not standard 1981-2010?

Because reliant on UKMO ocean initial conditions and they are not available prior to 1990

Climate and Oceans Support Program in the Pacific

Forecasts

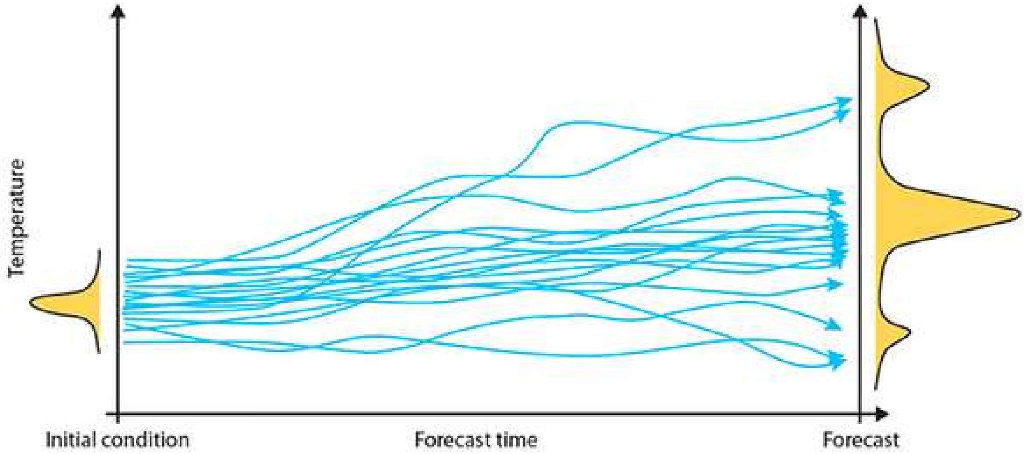
Real-time forecasts

- Use most recent observations to estimate current state and use model to integrate forward in time
- Daily data is aggregated to weekly – seasonal timescales
- *Examples of forecasts include the seasonal climate outlook or ENSO*

ACCESS-S1 Forecast design

Larger ensembles than used by UKMO

Forecast design



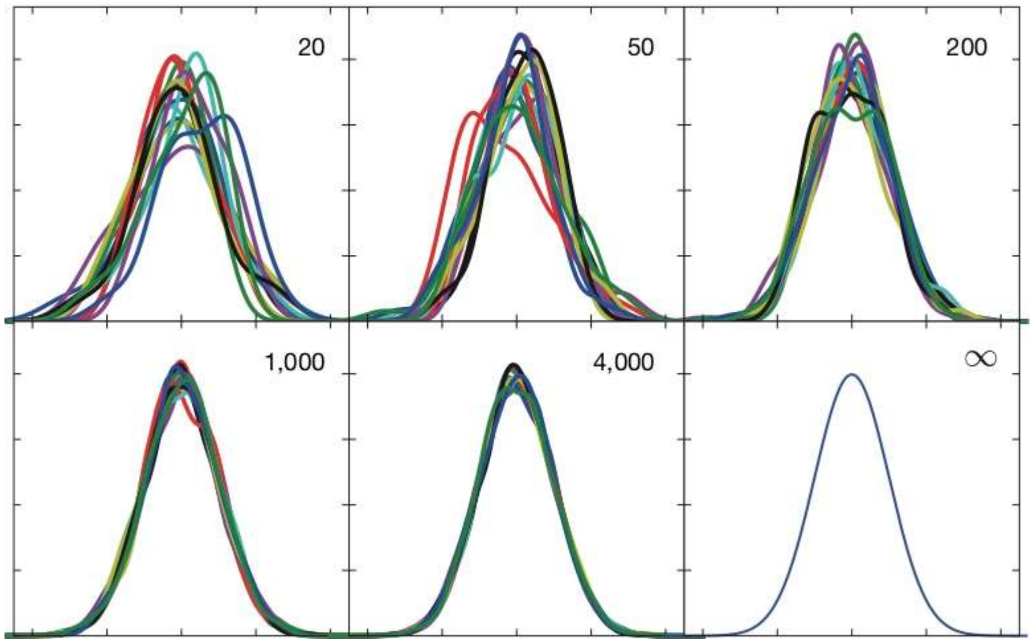
Every day the following are created:

- **11**-members run out for 6 months (210 days)
- **22**-members run out for 6 weeks (*To provide better skill in the subseasonal)*

- Primarily three factors that limit the skill of seasonal forecasts:
  - coupled model error
  - error in the estimate of the initial state
  - the unpredictable nature of atmospheric synoptic variability

- By running many ensembles we sample the effect of these uncertainties in a seasonal forecast system
- Perturb the initial conditions in agreement with the known statistics of error sources.
- The spread of the ensemble should then provide some measure of the level of uncertainty attached to the forecast.

ACCESS-S1 Forecast Ensemble Size



Users V Developers

- *Users – the more ensembles the better*
  - often interested in the tails of the distribution (forecast skill increases with ensemble numbers)
- *Developers – its expensive to have lots of ensembles*
  - easier to work with smaller amounts of data

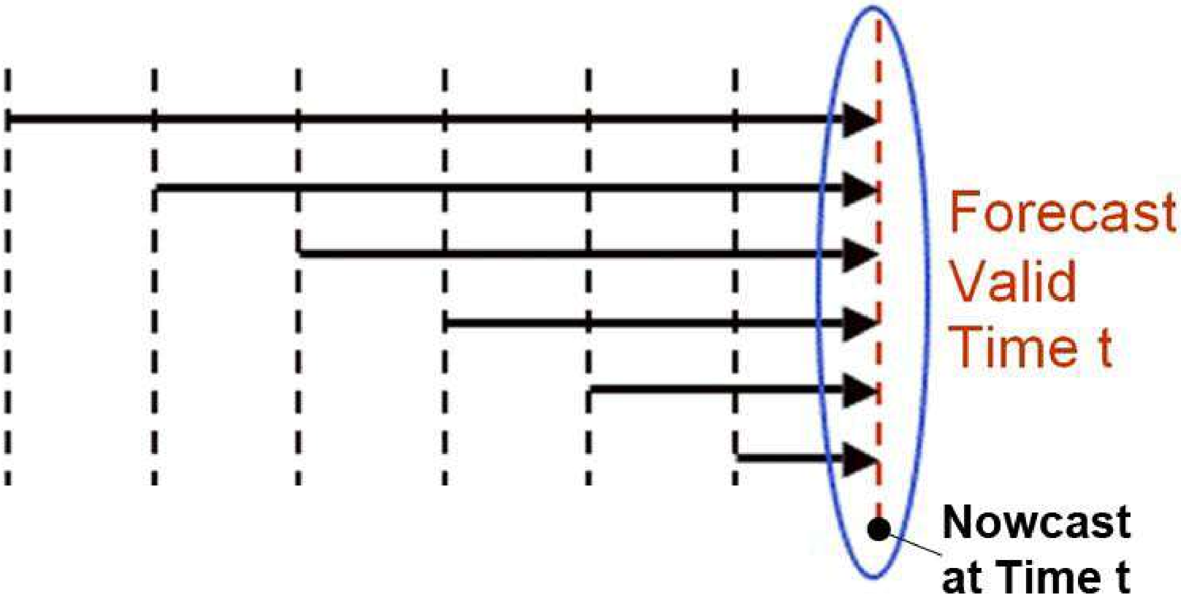
ACCESS-S1

- 99 ensemble members per forecast
- Provides a good estimate of uncertainty
- Provides optimum skill and reliability

https://www.ecmwf.int/en/newsletter/157/news/how-many-ensemble-members-are-desirable

ACCESS-S1 Lagged Ensemble

In order to inflate the ensemble size, real-time forecasts are produced by combining all the forecasts produced over a window of several days to produce a "lagged ensemble" in which ensemble members have different lead times.



Every day the following are created:

- **11**-members run out for 6 months (210 days)
- **22**-members run out for 6 weeks

99-member ensemble for forecast products:

- Seasonal ensemble: create latest 99 ensembles from the last 9 days of ensembles 1-11 (15 day window)
- Multi-week ensemble: create latest 99 ensembles from the last 3 days of ensembles 1-33 (5 day window)

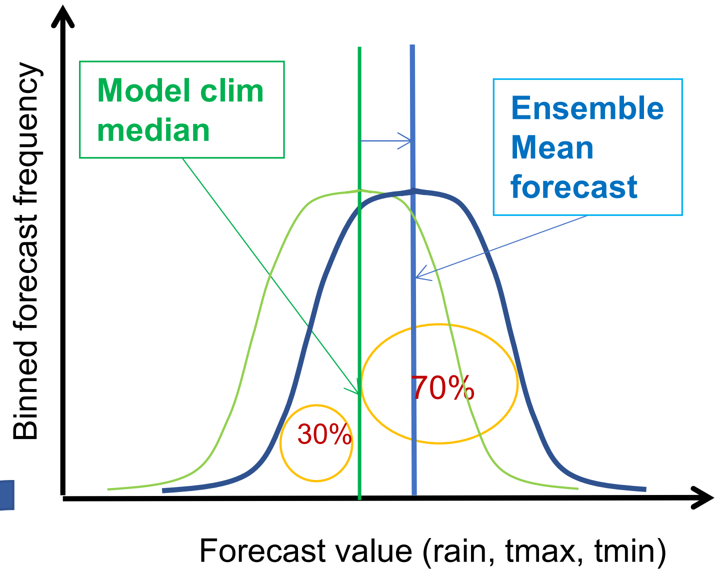
Making a probability above median and tercile forecast

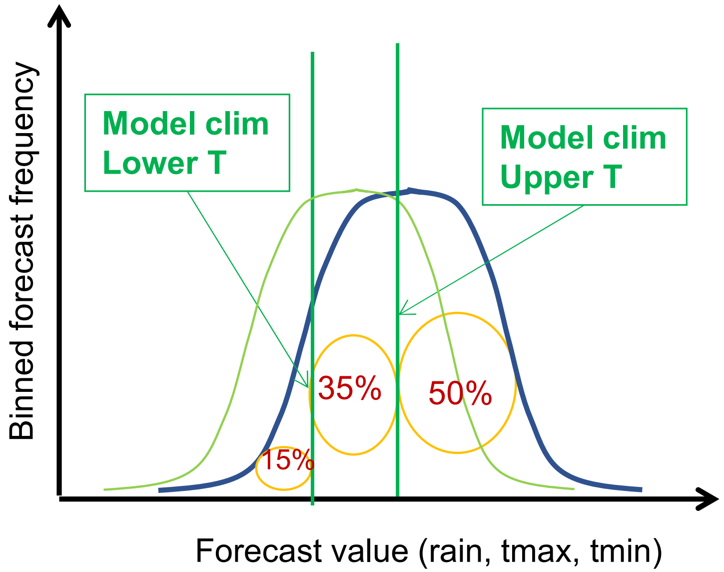
Climate and Oceans Support Program in the Pacific

Model climatology computed from the hindcast – No real world observations used

When an ensemble forecast is produced the area under its distribution is compared to the area under the distribution from the hindcasts.

In this case the forecast is significantly to the right of the model median and 70% of the forecast distribution is to the right of the hindcast ensemble median.





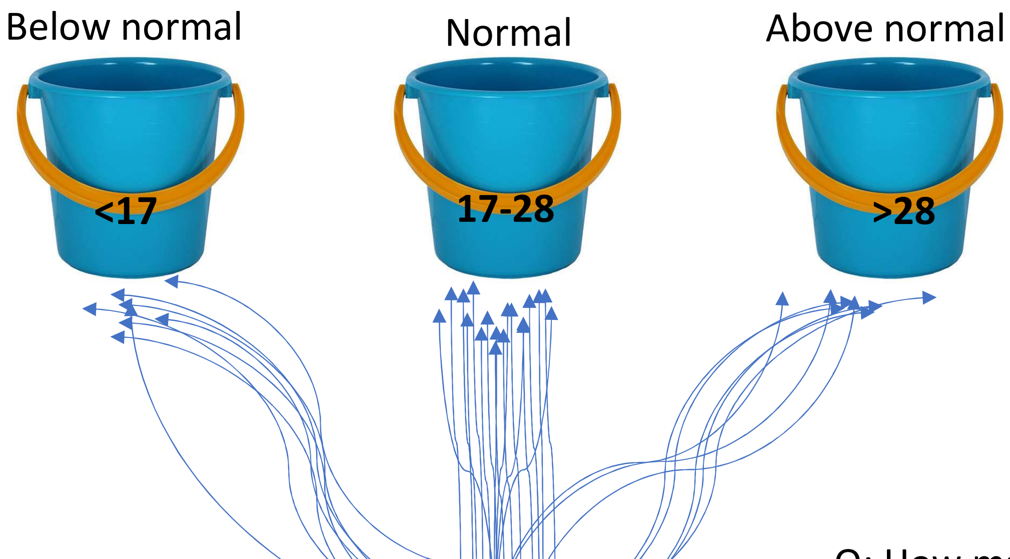
Forecast: 70% probability above median

Forecast: T1 15% T2 35% T3 50%

Tercile example

1, 11, 15, 17, 19, 20, 24, 28, 34, 42, 47

Hindcast



Forecast 99 ensembles

Q: How many ensembles are in each bucket?

A: Most ensemble members forecast between 17 and 28 so are in the middle bucket – or the normal tercile

Example: June 2021 Monthly Forecast from 17 May 2021

- 17 May 2021 is a Monday, and is one of the days from which forecasts are stored
- 11 forecasts generated on the 17th, but we need 99 ensembles
- Recall that Hindcasts (23 years from 1990-2012) were generated on the 1st, 9th, 17th and 25th of each month

Example: June 2021 Monthly Forecast from 17 May 2021

Hindcasts for June available from runs on 17 May

Initialised 17 May 1990: 11 predictions for June 1990

Initialised 17 May 1991: 11 predictions for June 1991

Initialised 17 May 1992: 11 predictions for June 1992

…

Initialised 17 May 2012: 11 predictions for June 2012

A total of 23 years x 11 predictions for June (e.g. rain, temp) at each grid point = 253 grids for each variable

Example: June 2021 Monthly Forecast from 17 May 2021

June Climate statistics are generated for each variable from 253 grids

Mean, median, terciles etc.

Remember, this is for **runs from 17 May**.

The **same process** applies for ensembles from **1, 9, and 25 May** in terms of a June "climatology" at lag of one month or a partial month.

Climatology is prediction-based, not observation-based.

Climate and Oceans Support Program in the Pacific

Example: June 2021 Monthly Forecast from 17 May 2021

Now consider the June forecast generated on 17 May 2021

An ensemble of 99 members -

11 members from **17 May** which reference the 17 May climatology
11 members from **16 May** which reference the **9 May climatology***
11 members from **15 May** which reference the 9 May climatology
11 members from **14 May** which reference the 9 May climatology
11 members from **13 May** which reference the 9 May climatology
11 members from **12 May** which reference the 9 May climatology
11 members from **11 May** which reference the 9 May climatology
11 members from **10 May** which reference the 9 May climatology
11 members from **9 May** which reference the 9 May climatology
* It is the ***closest prior climatology*** that is used for each date

Example: June 2021 Monthly Forecast from 17 May 2021

So over 99 ensemble members we get:

- 99 June monthly anomaly values (for each variable) which are averaged to produce the predicted anomaly
- The counts in each tercile range where the fraction of 99 is converted to a percentage (%)
- The counts above/below median where the fraction of 99 is converted to a percentage (%)

Example: Feb-Apr 2021 Seasonal Forecast from 11 Jan 2021

- 11 January 2021 is a Monday, so it is one of the days from which forecasts are stored
- Recall that Hindcasts (23 years from 1990-2012) were generated on the 1st, 9th, 17th and 25th of each month
- Let us first consider the hindcasts relevant to this forecast

Example: Feb-Apr 2021 Seasonal Forecast from 11 Jan 2021

Hindcasts for Feb-Apr available from runs on 9 January

Initialised 9 Jan 1990: 11 predictions for FMA 1990

Initialised 9 Jan 1991: 11 predictions for FMA 1991

Initialised 9 Jan 1992: 11 predictions for FMA 1992

…

Initialised 9 Jan 2012: 11 predictions for FMA 2012

A total of 23 years x 11 predictions for FMA (e.g. rain, temp) at each grid point = 253 grids for each variable

Example: Feb-Apr 2021 Seasonal Forecast from 11 Jan 2021

Feb-Apr Climate statistics are generated for each variable from 253 grids

Mean, median, terciles etc.

Remember this is for runs from **9 January**.

The **same process applies** for ensembles from **1, 17, and 25 Jan** in terms of a Feb-Apr "climatology" at lag of one month or a partial month.

Climate and Oceans Support Program in the Pacific

Example: Feb-Apr 2021 Seasonal Forecast from 11 Jan 2021

Now consider the Feb-Apr forecast generated on 11 Jan 2021

An ensemble of 99 members -

11 members from **11 Jan** which reference the 9 Jan climatology
11 members from **10 Jan** which reference the 9 Jan climatology
11 members from **9 Jan** which reference the 9 Jan climatology
11 members from **8 Jan** which reference the **1 Jan climatology***
11 members from **7 Jan** which reference the 1 Jan climatology
11 members from **6 Jan** which reference the 1 Jan climatology
11 members from **5 Jan** which reference the 1 Jan climatology
11 members from **4 Jan** which reference the 1 Jan climatology
11 members from **3 Jan** which reference the 1 Jan climatology
* Once again it is the ***closest prior climatology*** to the date

Example: Feb-Apr 2021 Seasonal Forecast from 11 Jan 2021

So over 99 ensemble members we get:

- 99 Feb-Apr seasonal anomaly values (for each variable) which are averaged to produce the predicted anomaly
- The counts in each tercile range where the fraction of 99 is converted to a percentage (%)
- The counts above/below median where the fraction of 99 is converted to a percentage (%)

Climate and Oceans Support Program in the Pacific

Forecast and hindcast summary

- Hindcasts are historical model runs using data available at the time
- There are 23 years of hindcasts available
- Hindcasts are used to compute a climatology
- Hindcasts are used to calculate model skill
- Each forecast contains 99 ensemble members (multi-week has 33 runs X 3 days, seasonal has 11 runs x 9 days)
- Forecasts are compared to hindcasts to calculate statistical products (terciles, change of above median) or deterministic products (sea surface temperature anomaly)
- Due to IT limitations the way an ACCESS-S forecast is put together is very complicated