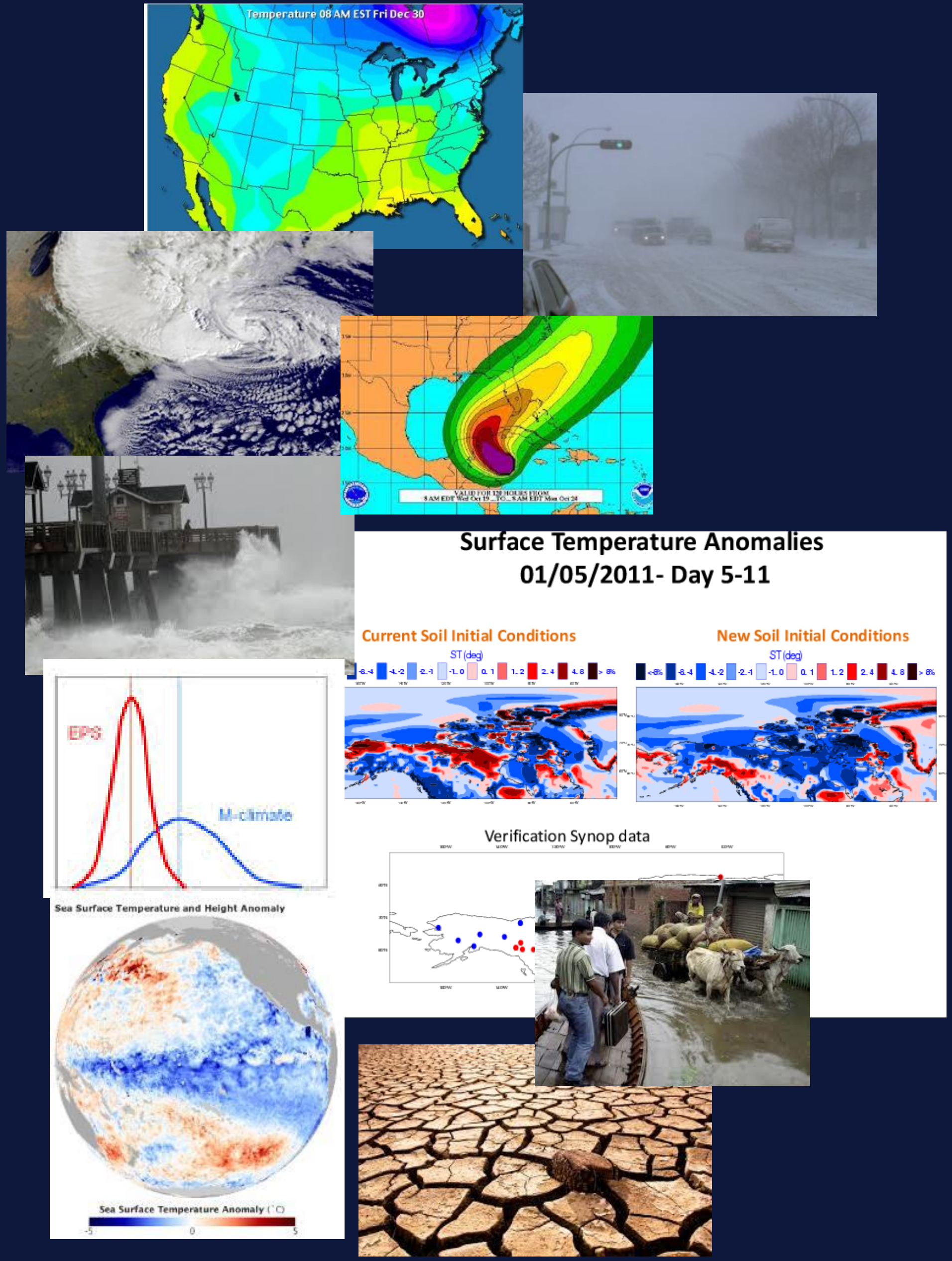
Probabilistic Forecasting System

Design Elements



- The ECMWF FC system
- The Initial Value Problem
- Ensembles: Error propagation and probabilistic forecasts
- Time scales and model components
- Calibration, skill assessment and reanalyses
- Measuring performance
- Examples of forecast products at different time ranges
- Wider context: Forecast strategies

M.A. Balmaseda

Designing the current and future systems

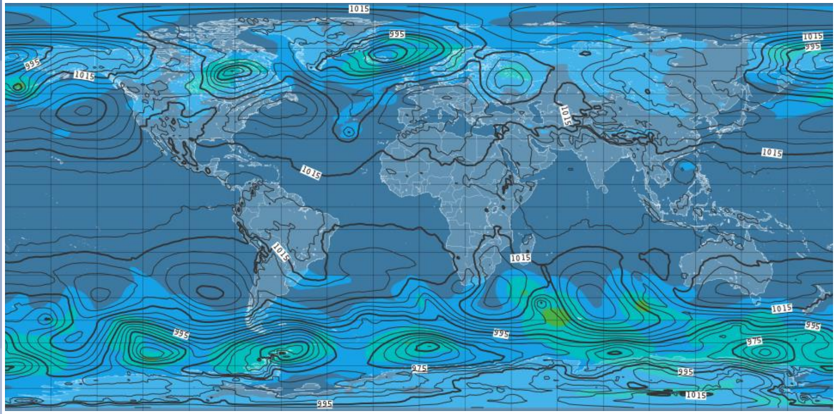
- This presentation has focused on design of medium-extended range –seasonal forecasting systems.

- The demand for service expansion is increasing
  - C3S: climate
  - CAMS: atmospheric composition
  - DestinE: extreme high resolution
  - GLOFAS and impact modelling

- Design of systems is even more important
  - Guiding principles: optimize quality, timeliness and maintenance
  - Seamless approach: to have a unique IFS with different plug-ins and configurations
- The scientific and technological landscape is rapidly changing.
  - What will be the role of ML in future forecasting system? Will the seamless approach still be valid?

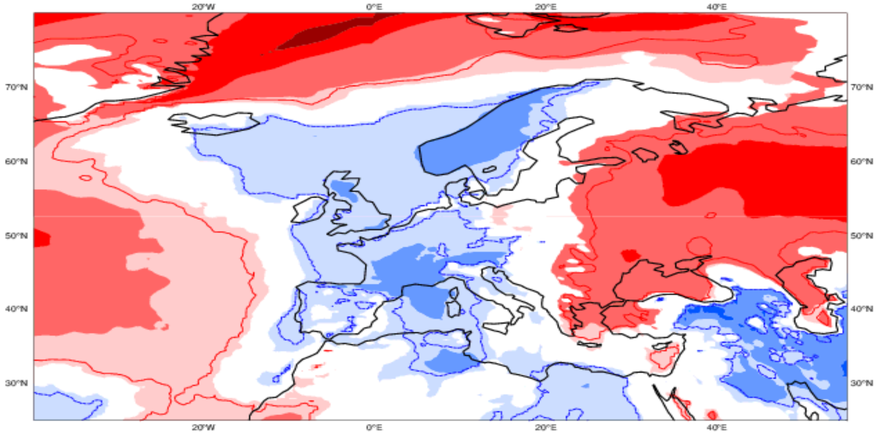
Deliverables: Global NWP from days to years

Medium range prediction



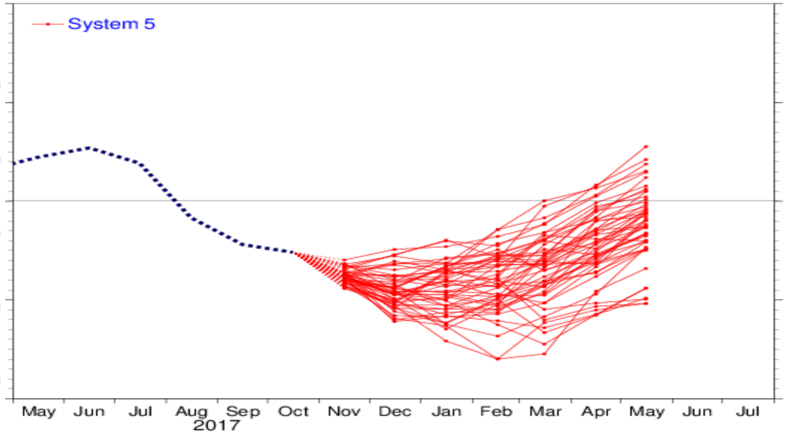
High resolution mean sea level pressure and ensemble spread
Forecast range: several days ahead

Subseasonal range –monthly-



Weekly anomaly – 2m temperature over Europe
Forecast range weeks 3-4

Long range –seasonal- prediction



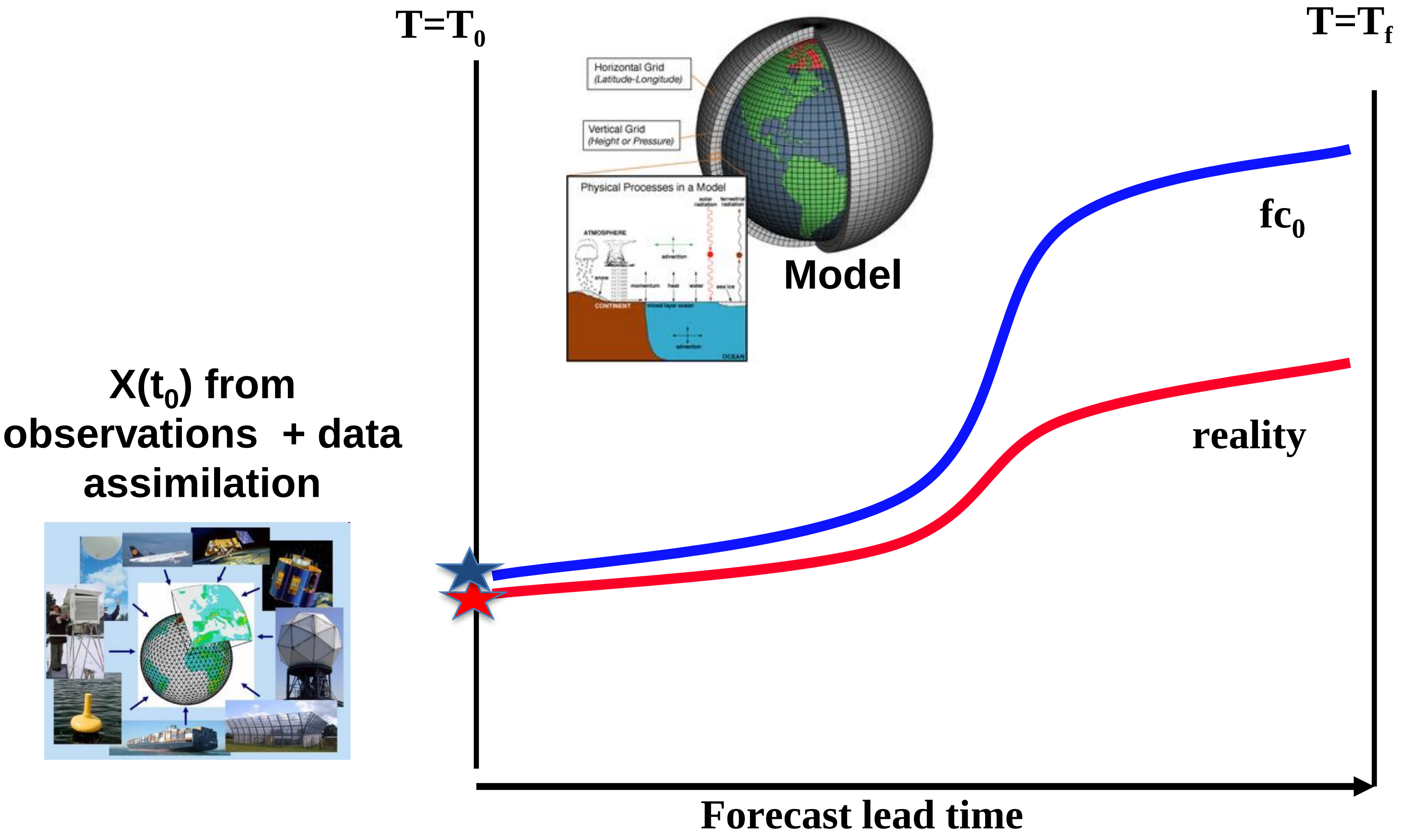
El Nino 3.4 SST anomaly plume – 1 November 2017

Up to 6-12 months

https://www.ecmwf.int/en/forecasts

1. Weather forecast as an initial value problem

$$X(t_f) = M[X(t_0)]$$



1. The initial value problem. Predictability drivers

- Wave propagation
- Advection of signals
- Persistence of signals : regimes, soil moisture, sea-ice
- Slow dynamical time scales: stratosphere, ocean, sea-ice

How to best represent these drivers?

Modelling

- Model resolution
- Physical parameterizations
- <u>Earth system components – complexity-</u>
  - Feedbacks among components
  - Multiple temporal scales

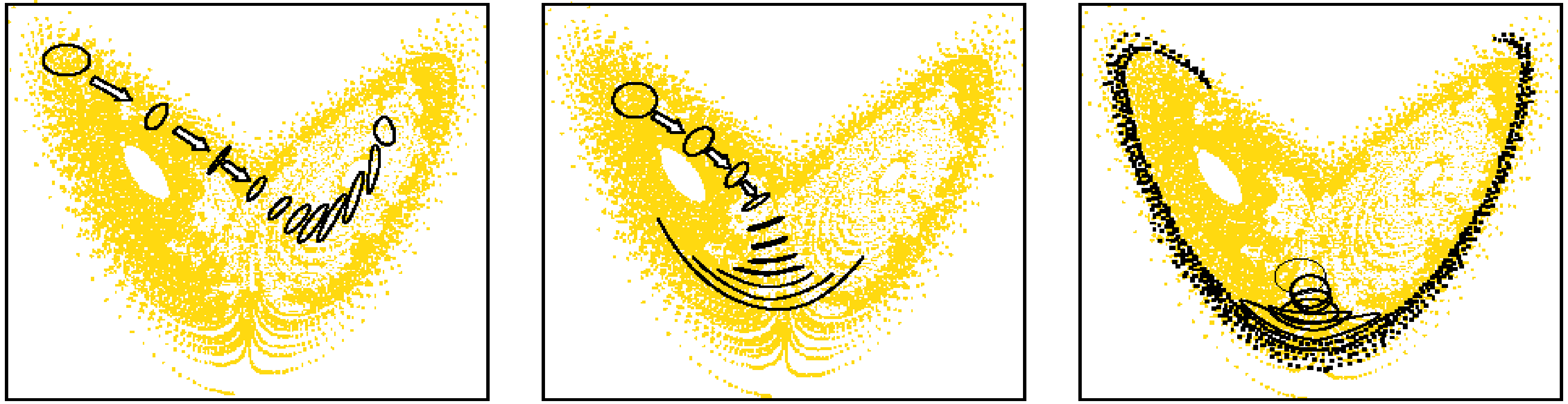
Initialization

- Observations
- Data Assimilation Methods (4D-var especially good for wave propagation)

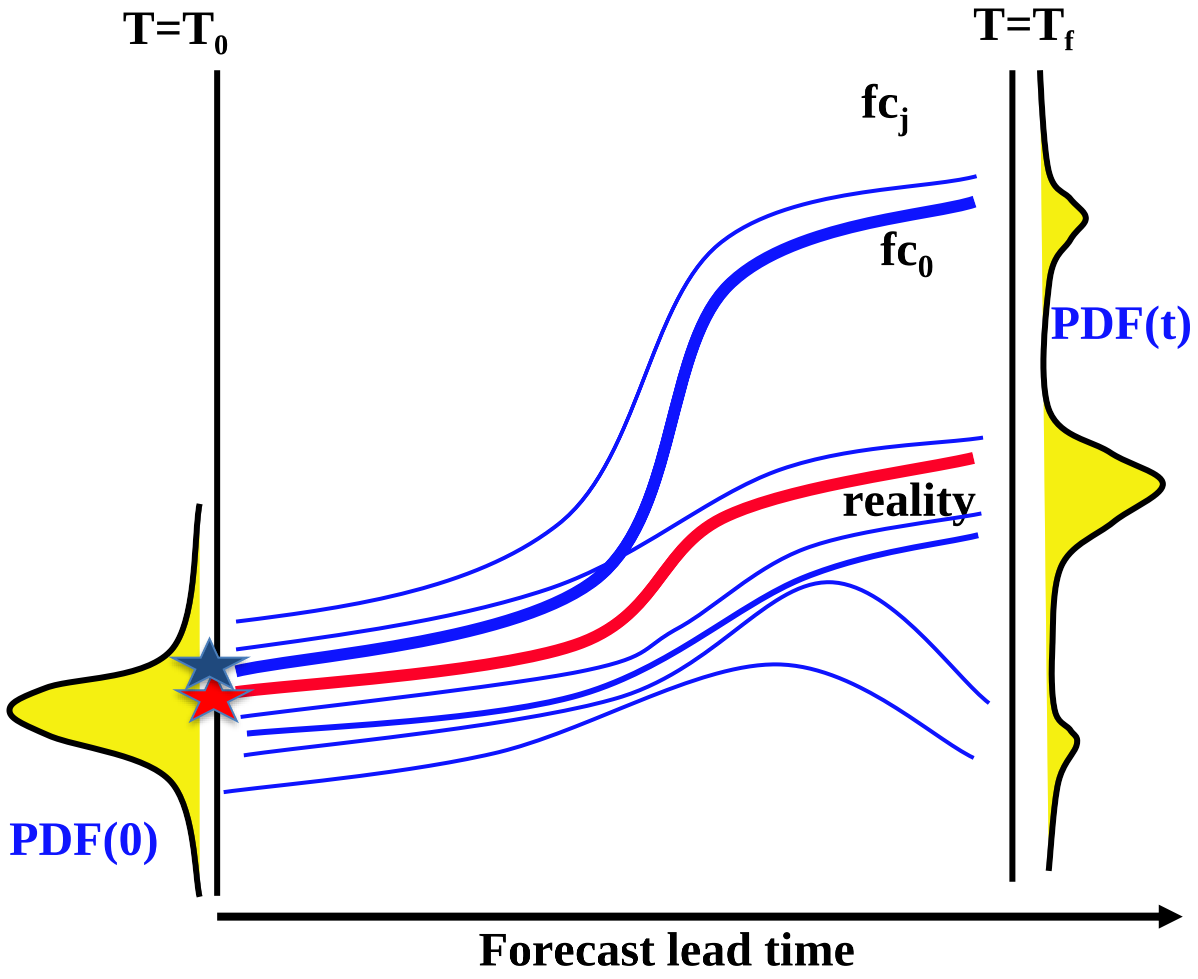
1. Initial Value Problem: Predictability limits

- Weather is intrinsically unpredictable in the deterministic sense: The atmosphere as a chaotic system

  <u>Uncertainty</u>: In a chaotic system, small uncertainty in the initial conditions leads to forecast uncertainty



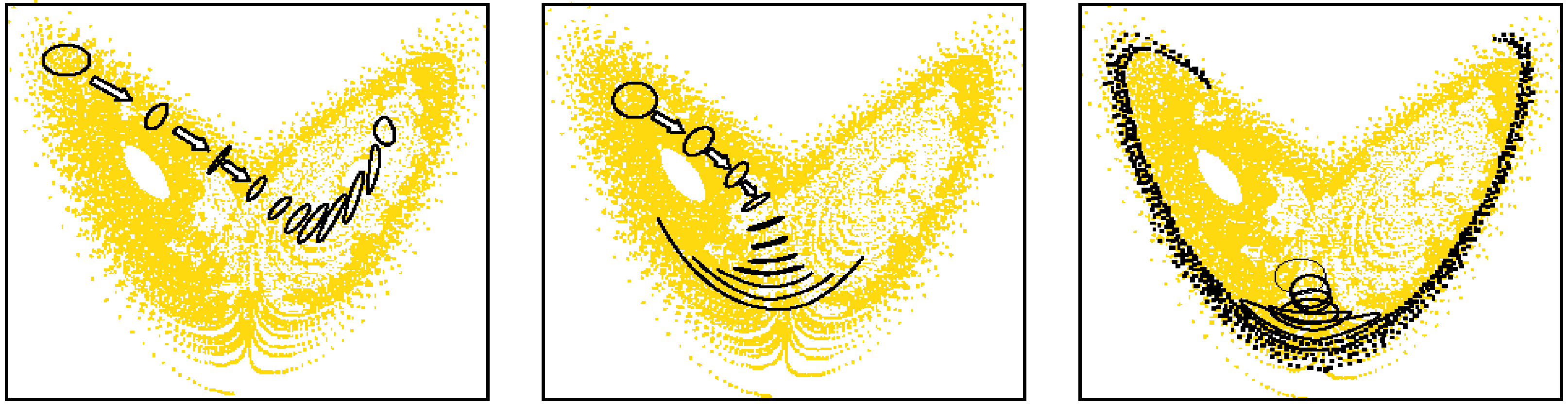
2. Ensemble Prediction: A pragmatic approach for propagation of uncertainty



1. Initial Value Problem: Predictability limits

- Weather is intrinsically unpredictable in the deterministic sense: The atmosphere as a chaotic system

  Uncertainty: In a chaotic system, small uncertainty in the initial conditions leads to forecast uncertainty



- Forecasting system deficiencies leads to forecast **error**

  – The initial conditions are not accurate enough, e.g. due to poor coverage and/or observation errors, or errors in the assimilation.

  – The model used to assimilate the data and to make the forecast describes only in an approximate way the true atmospheric phenomena (model error).

Distinguishing between forecast error and intrinsic predictability is a major challenge.

A few rules of the forecast game

- We should distinguish between
  - **Errors** - which we should aim at correcting: improving model and initialization
  - **Uncertainty** - we should aim at representing : improving the ensemble generation.
    - model uncertainty (currently stochastic physics: SPPT, SPP.)
    - initial conditions uncertainty (currently EDA + SV)
- **Irreducible errors** can also be accounted for in the forecasting system in order to provide reliable forecast products. (online empirical error treatment, tuning of model/IC errors , a-posteriori calibration).

  => TREATMENT OF MODEL ERROR

3. Predictability limits and drivers:

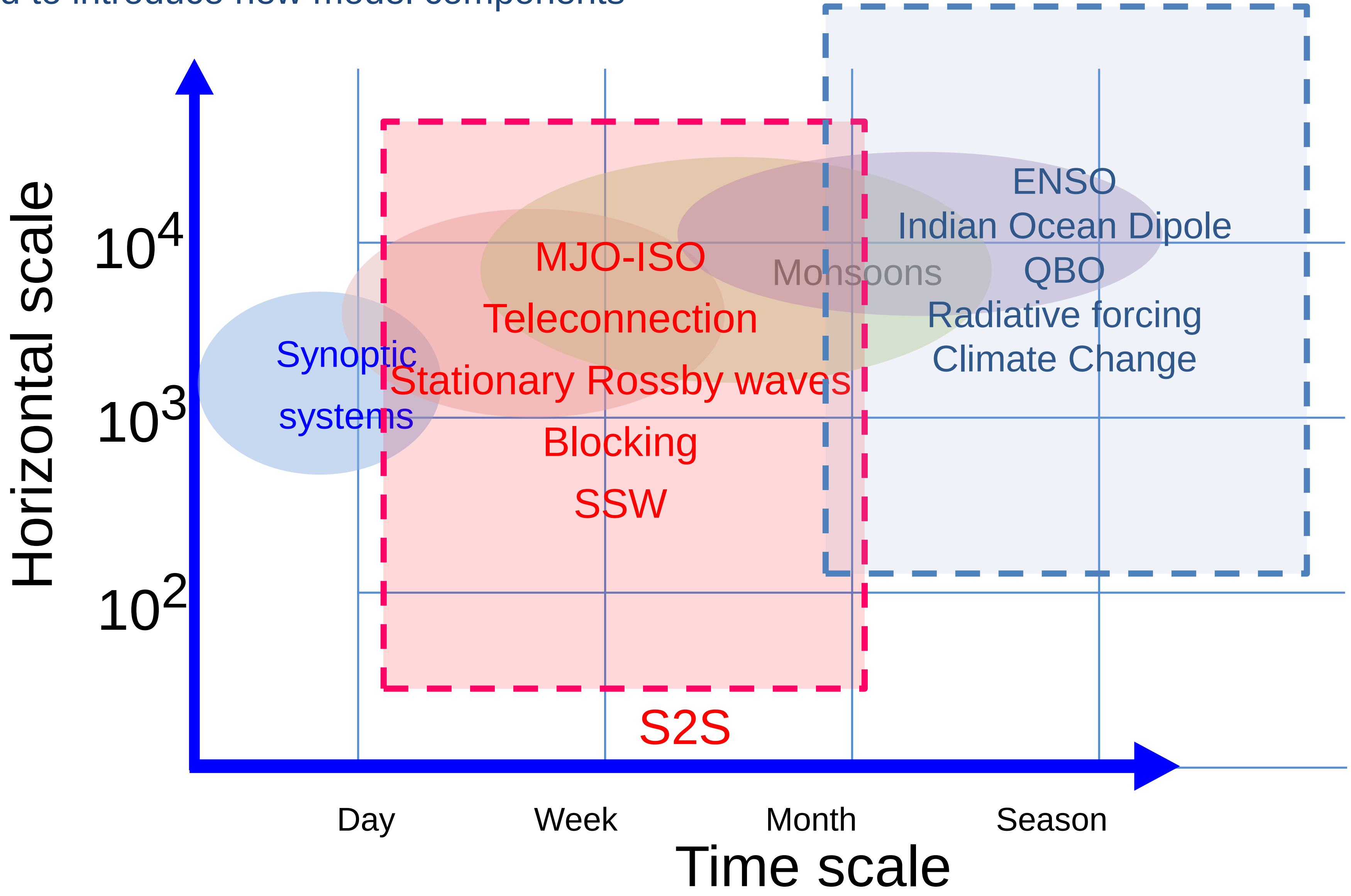
Probabilistic forecasts and multiple forecast ranges

- **Predictability of the fist kind: Fast time scales limit the predictability**, since error grows rapidly
    - Example: atmospheric convection; medium range: baroclinic instability.
    - Medium range problem, probabilistic by nature. Accurate/optimal representation of flow dependent uncertainty.

- **Predictability of the second kind or loaded dice paradigm:**
    - **Slower time scales** can act as a source of predictability: the atmospheric behaviour can be modulated by the state of slower neighbouring components, such as ocean, land, sea-ice, stratosphere
    - Including these slow components in the forecasting system allows extending the predictability horizon
        - Extended range: several weeks ahead
        - Seasonal forecast: several months ahead
    - By nature, these are PROBABILIST FORECASTS, but representation of the uncertainty in the atmospheric initial conditions is not so important in the ensemble generation.

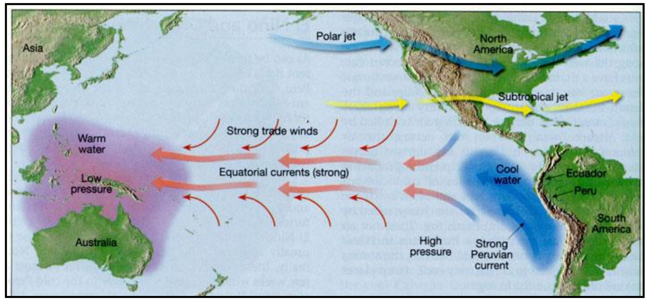
3. Slow time scales as predictability drivers:

Need to introduce new model components



ECMWF

Interannual Time scales: ENSO



ENSO: El Nino -Southern Oscillation

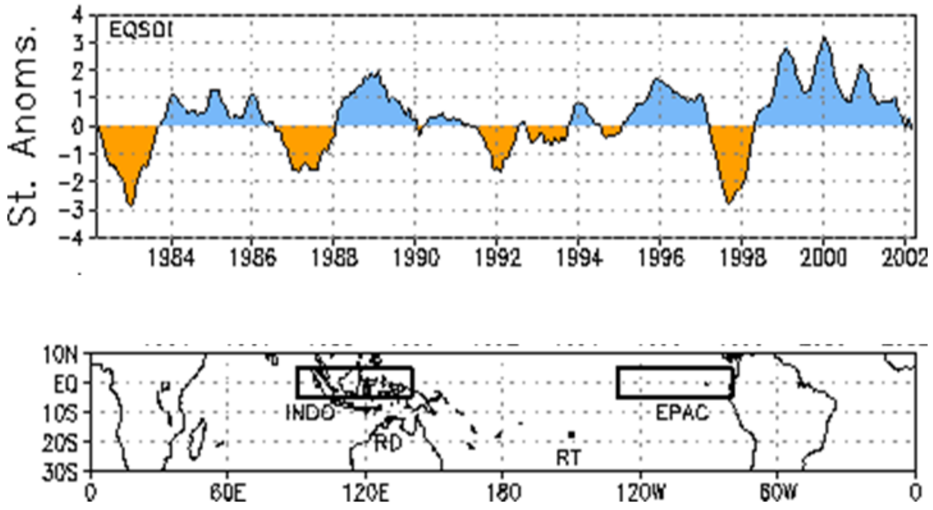
Largest mode of O-A interannual variability

Best known source of predictability at seasonal time scales

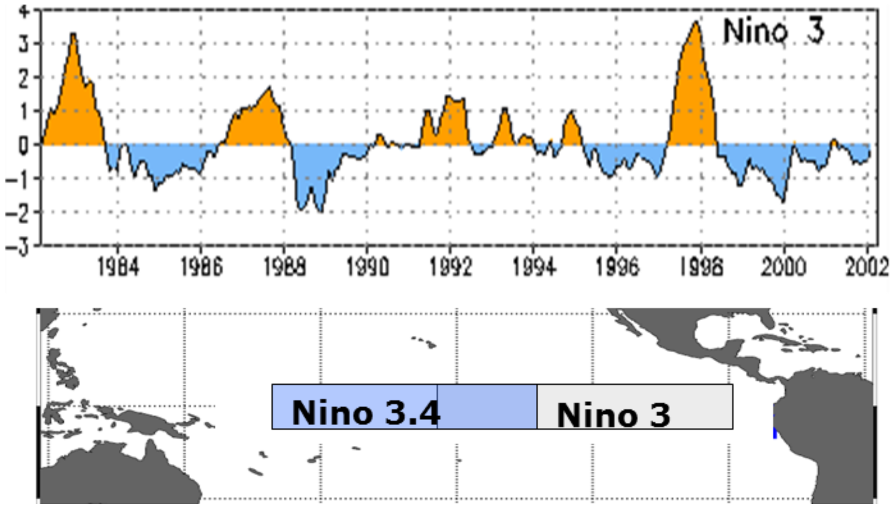
It affects global patterns of atmospheric circulation, with changes in rainfall, temperature, hurricanes, extreme events

SOI: Southern Oscillation Index (SLP Darwin – Tahiti)

Sea Level Pressure (SOI)

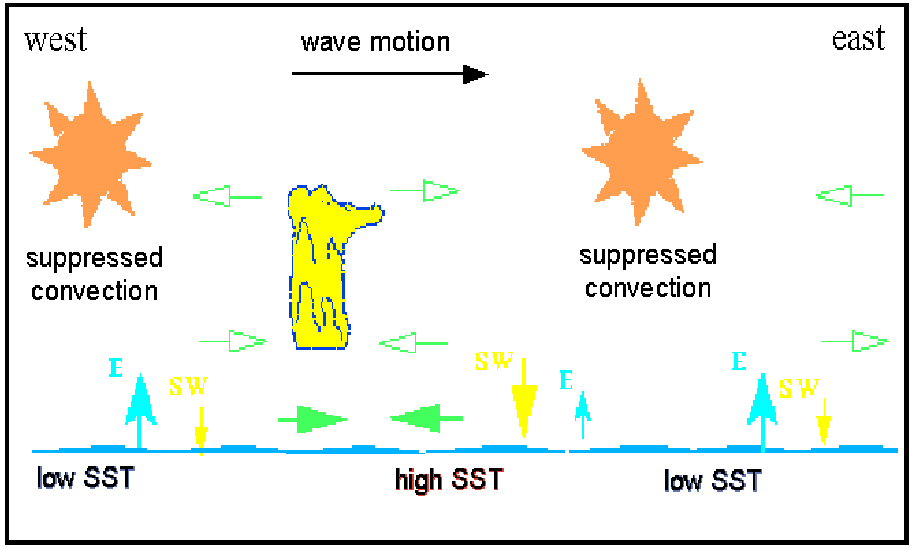


Sea Surface Temperature (Nino 3)

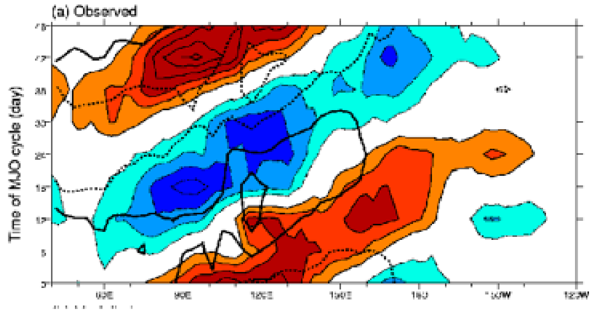


Madden Julian Oscillation: Coupled O-A tropical convection mode~ 30-60 days

Basis for Predictability at the subseasonal forecacast range (week 2 -4)



MJO: Coupled Mode



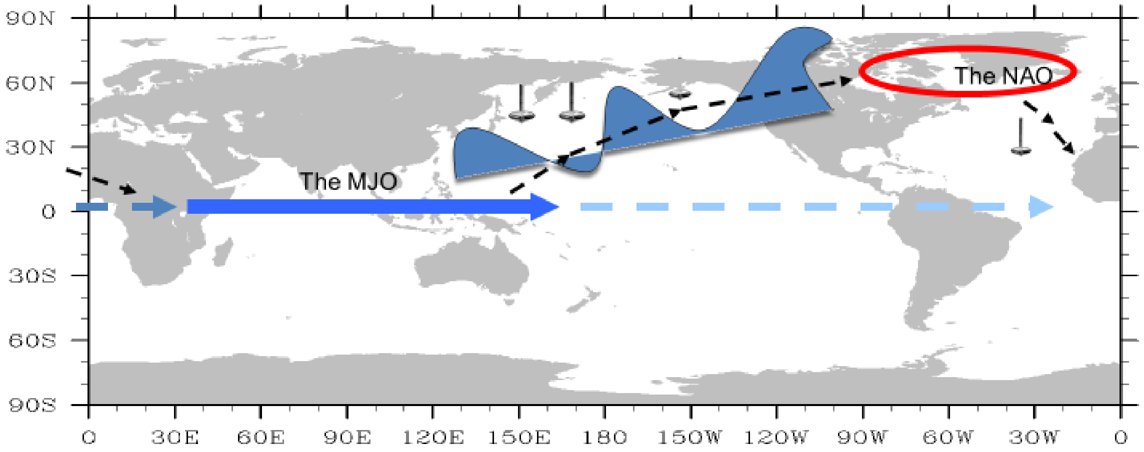
Composites of SST anomalies (contours) and OLR (colours) of MJO events. SST and convection are in quadrature.

The lead-lag relationship between SST and deep convection seems instrumental for setting the propagation speed of the MJO.

A two way coupling is required. Thin ocean layers are needed to represent this phase relationship.

Tropical-Mid Latitude interaction: a series of complex stepping stones

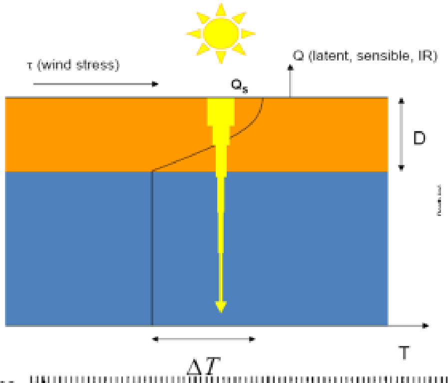
Wave propagation – Fronts and Storm Tracks –Stratospheric Bridge .



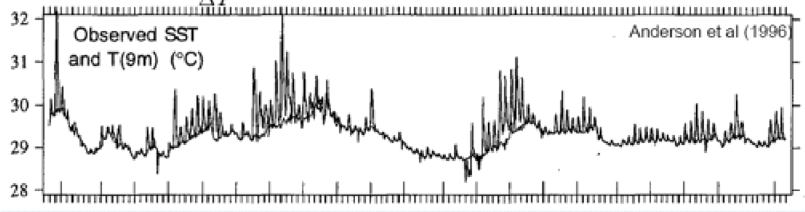
Adapted from Brunet, 2015

The ocean also affects fast processes:

Diurnal Warm Layers: amplification of diurnal cycle



- Stably stratified (warm) thin layers form during the day.
- They isolate the deeper ocean by reducing vertical mixing.
- The increase the value of peak temperature.
- They trigger convection events, which can rectify in MJO

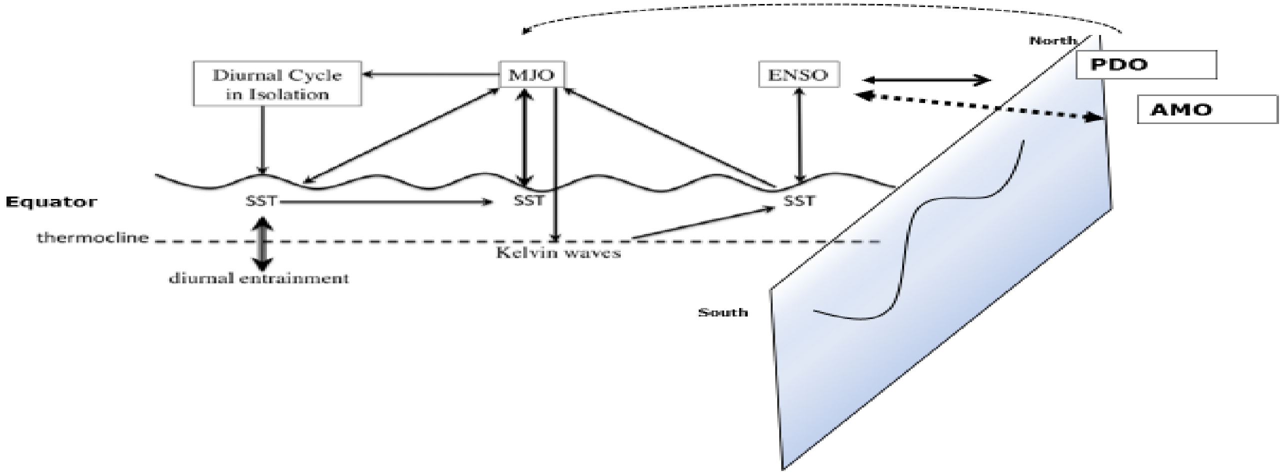


Air-Sea Interaction in Tropical Cyclones

Heat Flux exchange: ocean mixing and upwelling
Wind-Wave interaction
Ocean Initial conditions also matter

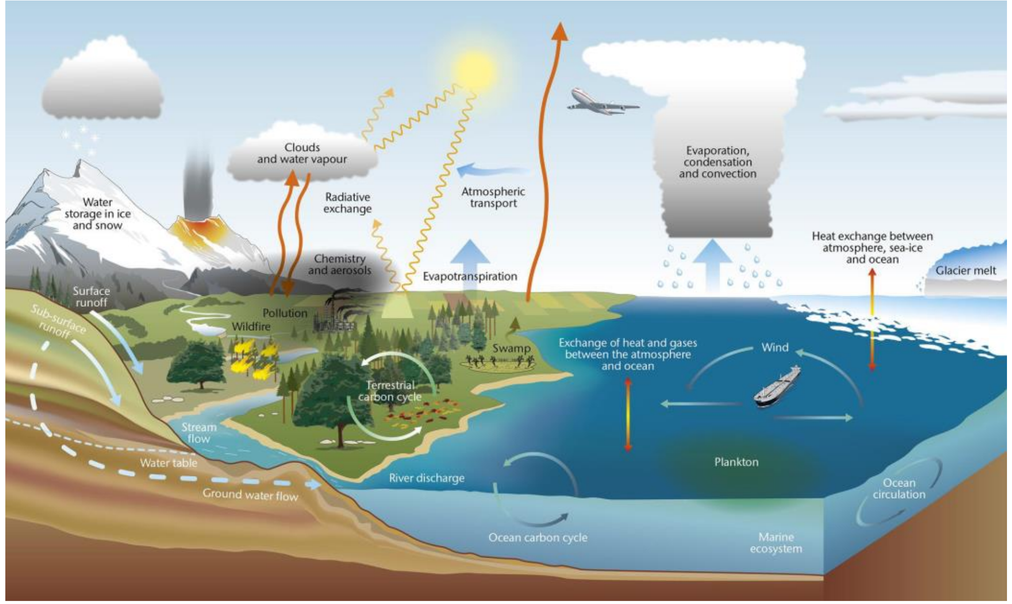
From Ginis 2008

Scale interaction key to variability:

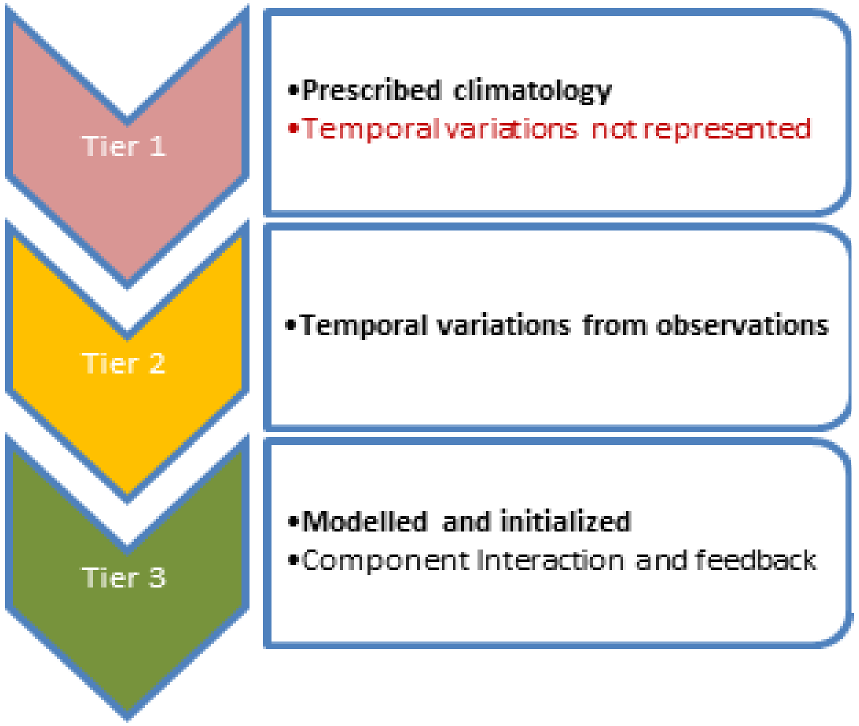


4. System Design: Earth System Complexity in a forecasting system

- Physical basis for inclusion
- Ability to model
- Ability to initialize
- Affordability



Hierarchical representation of an earth system component

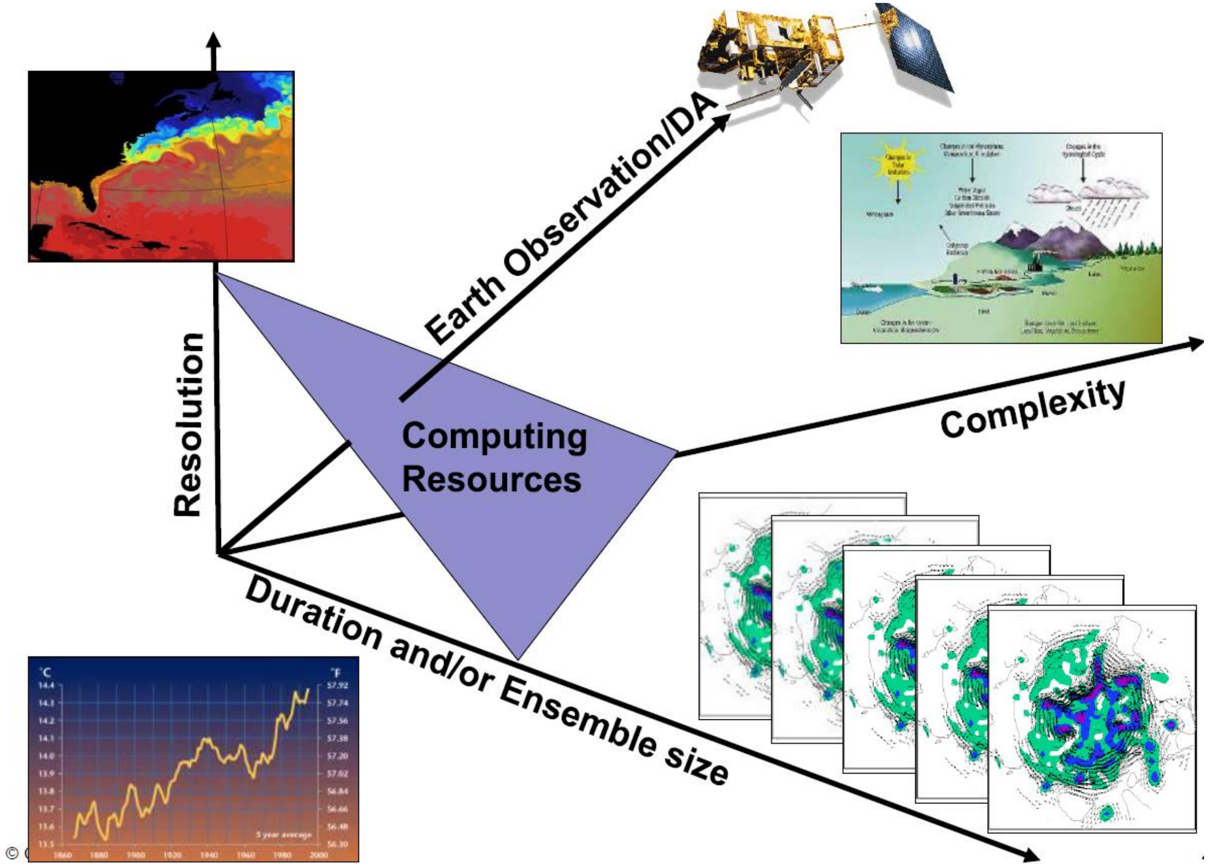


Vegetation, land use aerosols

C02 in reanalyses
Solar

Atmosphere, ocean, waves, sea-ice, soil moisture, snow,

5. System Design: balancing elements



Requirements

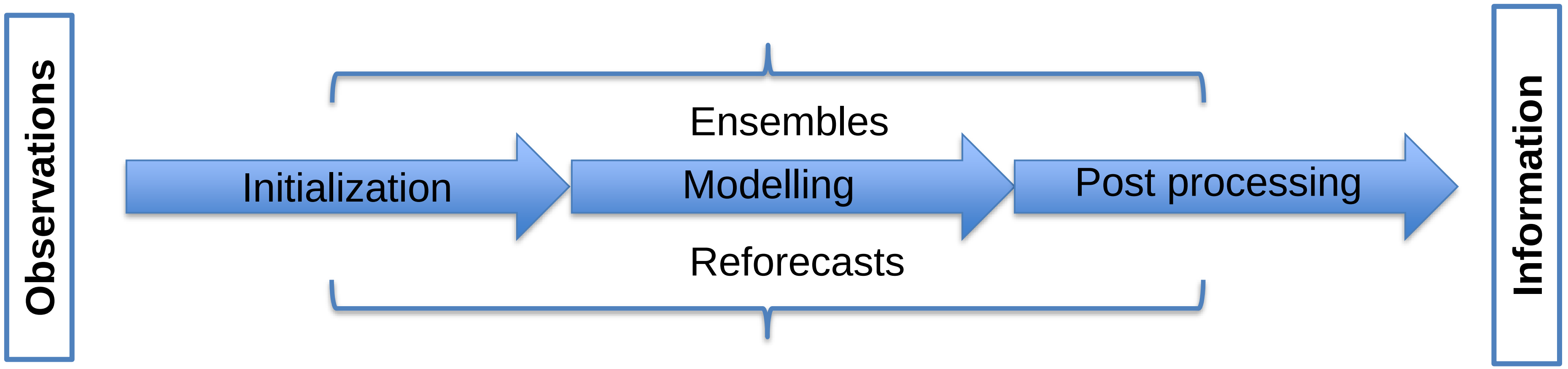
Resolution
Number of observations
Number of ensemble members
Model components
Forecast range and frequency
Reforecasts: calibration period, frequency and ensemble members

Constrains

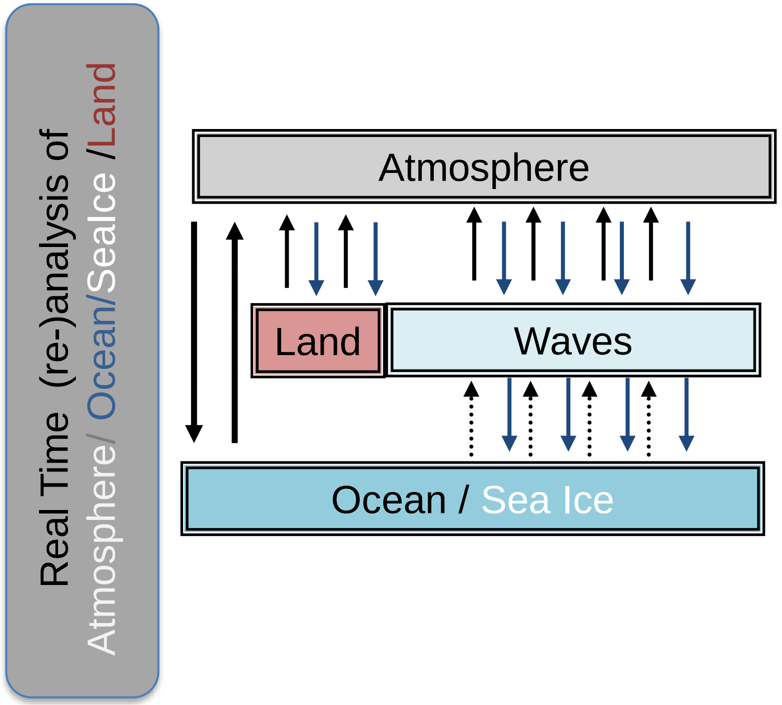
Computer resources
Data handling platforms and algorithms
Code maintenance and development
Expertise

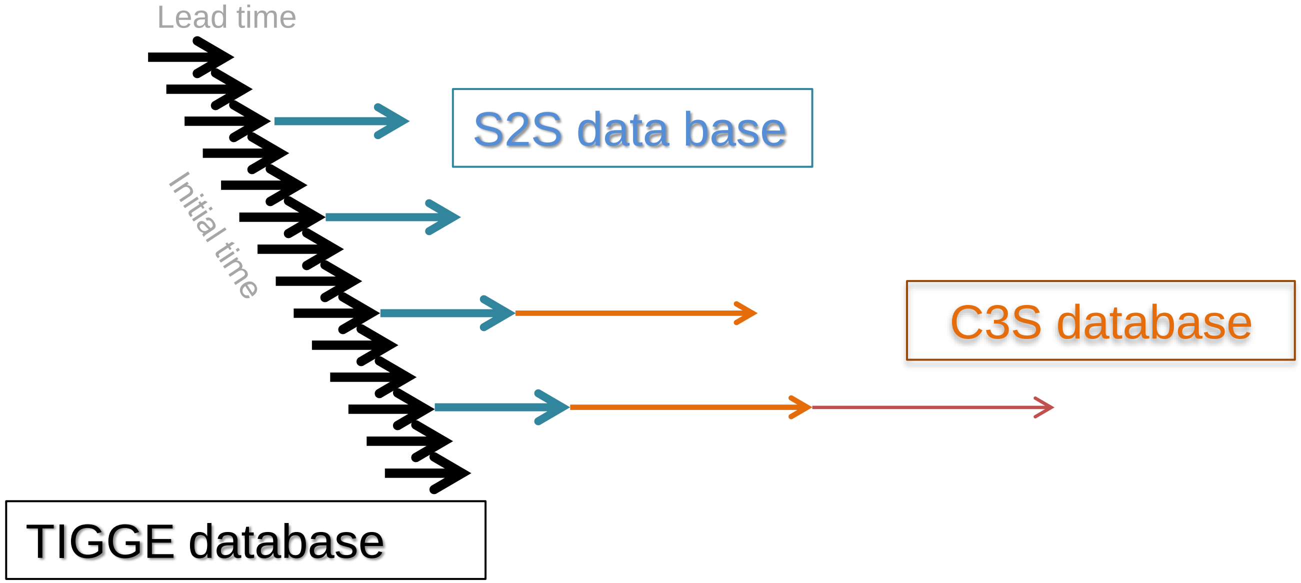
Courtesy of Andy Brown

6. System Design: End to End Forecasting System
From observations to societal information



6. System Design: Seamless Probabilistic Prediction



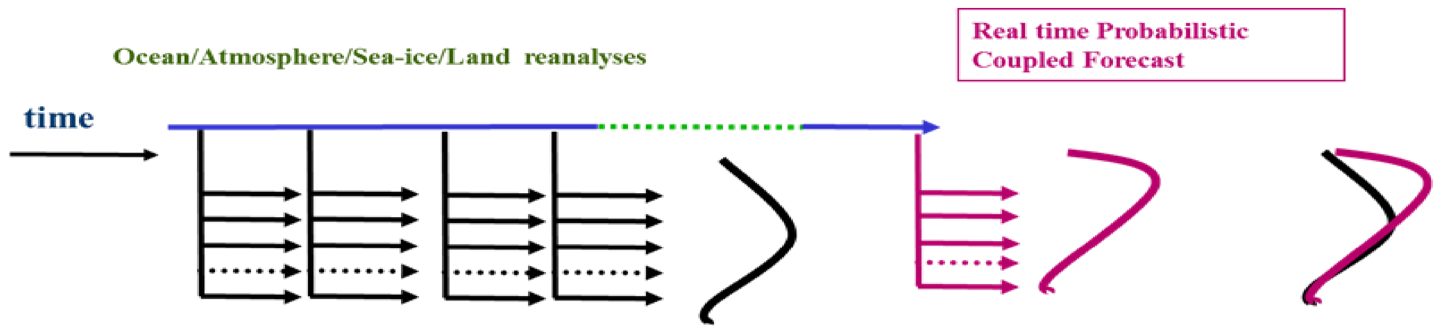


- Same model and initial conditions for different lead times.
- Resolution dependent on forecast range
- Main advantage: simplicity and cost
- Each prediction is an ensemble of N members (N~50)
  - Subseasonal N=100 since 2024
  - Next seasonal system will also have N=100

System	Lead Time	Prod Frequency
Medium Range	15 days	twice daily
Monthly	46 days	daily
Seasonal	7 months	Twice monthly
Annual	12 months	quarterly
ENSO outlooks	24 months	twice a year

7. System Design: Calibration and Skill Assessment
Reforecasts as integral part of a forecasting system

- Calibration of forecast output for useful products.
- Skill estimation
- Prediction of extreme events
- To guide fc system development by identification of critical fc errors



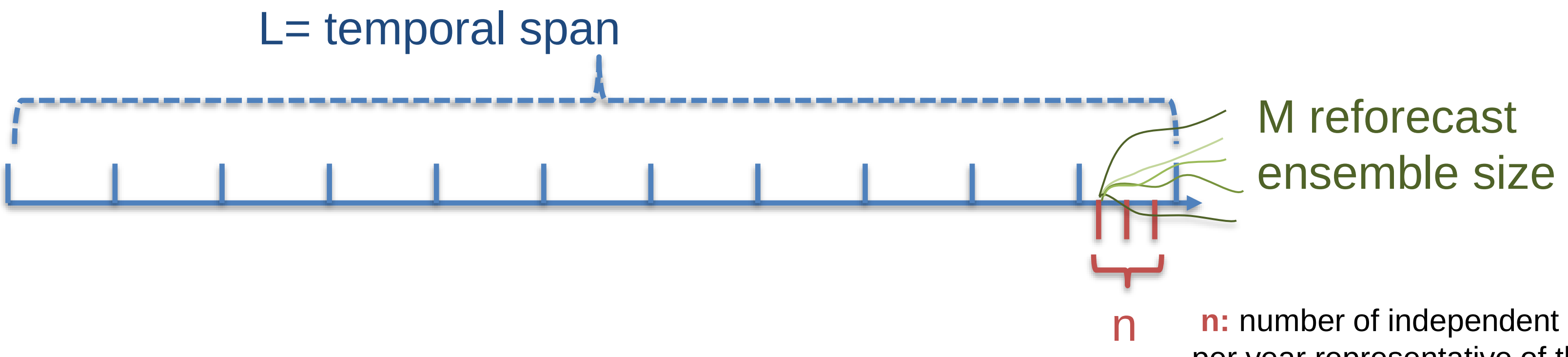
Reforecast errors should be consistent with the errors of the real time forecast.

Reforecast design needs to consider **temporal sampling and initialization**

7.Reforecast requirements: Temporal Sampling

- The calibration needs and verification periods depend on the forecast lead time and products
  - **Medium range:**
    - ECMWF products are not calibrated a posteriori (except for EFI-Extreme Forecast Index)
    - Skill can be estimated from a number of cases over a couple of seasons.
  - **Subseasonal range:**
    - Forecast PDF needs a-posteriori calibration (around 20 years)
    - Strong conditional skill, several cases spanning different seasons and interannual variability
  - **Seasonal range:**
    - Forecast PDF needs a-posteriori calibration (30 years or more)
    - Skill and error depends on season. The calibration data set should cover several ENSO episodes, QBO phases…

7. Reforecast requirements: Sampling



n: number of independent dates per year representative of the real time forecast situation

- **Temporal span (L) and frequency (n):**
  - **i) L,n depend on the forecast range**
  - **ii) L requires existence of initialization and verifying dataset**
  - L, n Need to sample enough independent cases and different regimes (e.g. seasonal cycle)
  - Reforecast span L may be different for skill assessment (Ls) and reference climatology (Lc)

- **Reforecast Ensemble size (M) : different for calibration (M_c) and skill assessment (M_s)**
  - Calibration: reforecast climatology size N_{clim} ~ real time forecast ens size N_{fc} (N_{clim}=L x n x M_c)
  - Skill assessment: : M_s sufficiently large to score probabilistic forecast. Balance between L,n,Mc

Examples

On Ls,Lc:

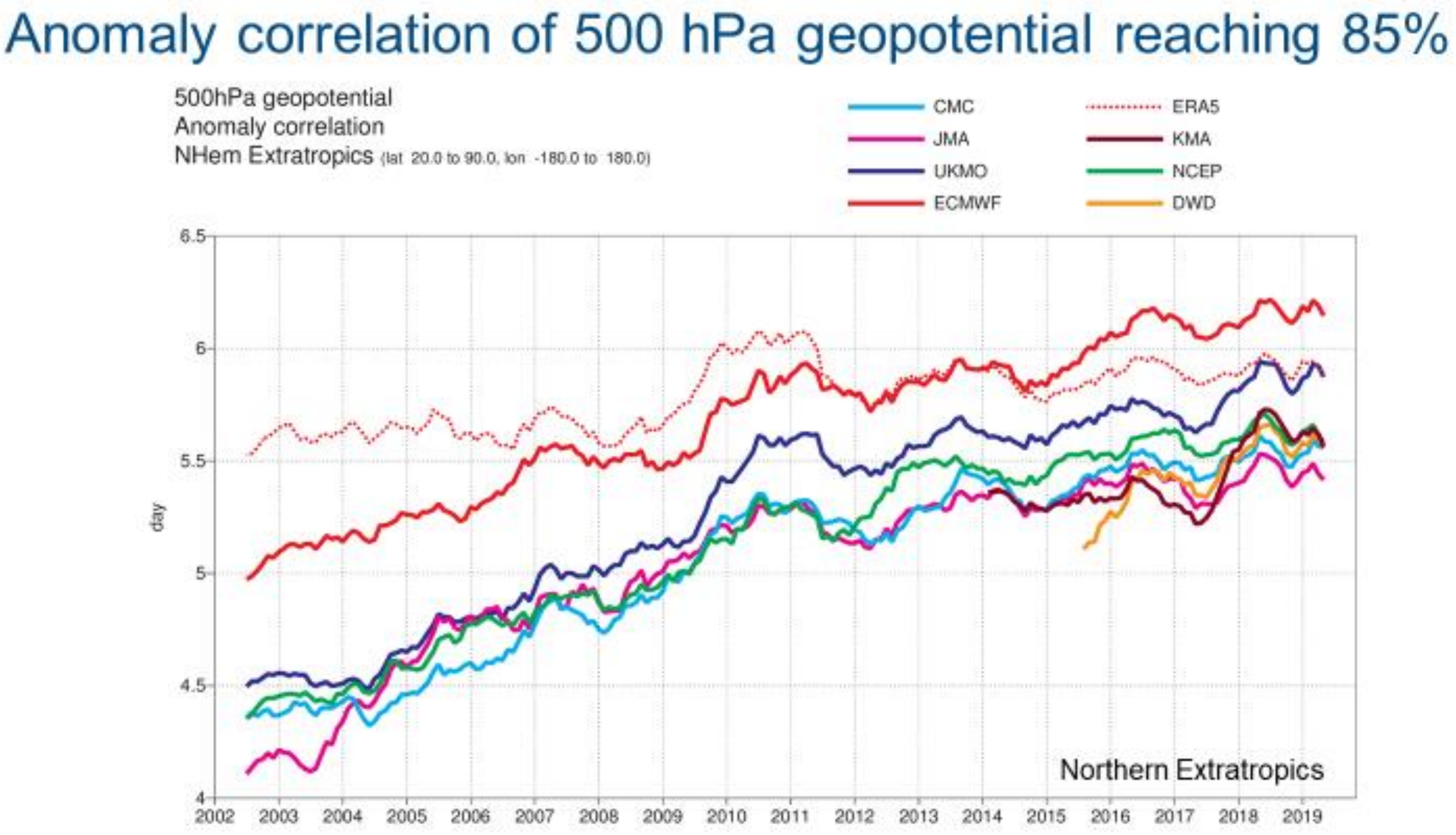
SEAS5: Ls= 37 yrs, Lc=24 yrs
Sub-S Ls=Lc=20 yrs
ecPoint : Lc=1yr

On n:

SEAS5	n=1 per month
Subseaonal Range	n=2 per week
Medium Range	n>> 1 (if Lc=1yr)

8. Measuring skill and Estimating the limits of predictability.

Example 1: measuring skill and benchmarking



8. Measuring skill and Estimating the limits of predictability

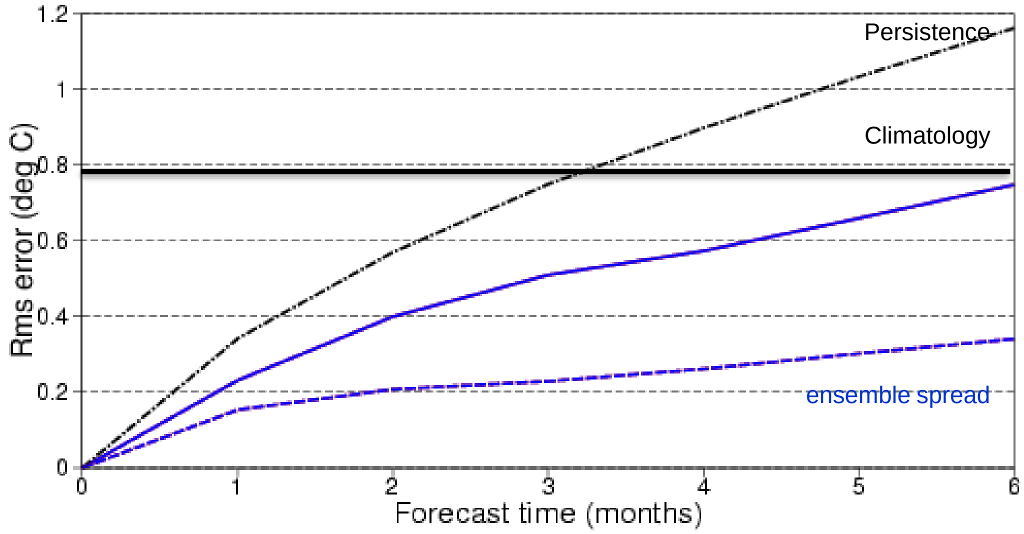
How skilful is a forecast?

User perspective: The answer to the question depends on the application.

System Design perspective: better than a “cheaper” bench-mark

- Climatology
- Persistence
- Other empirical model
- Other GCMS -> multi-model comparison

RMS error of Nino3 SST anomalies



How hard should we try?
e.g. have we reached the predictability limit?

- **Climatology** is considered the lowest limit of predictability
  - If model skill worse than climatology there is room for improvement

- **Persistence** can be indicative of potential predictability.

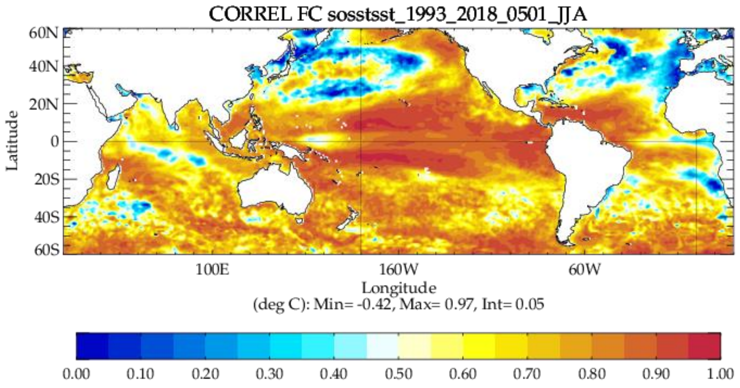
- **Ensemble Spread**: In a perfect model , ensemble spread is considered upper level of predictability
  - But sometimes model is overdispersive: ensemble spread larger than RMS error?

8. Estimating predictability limits

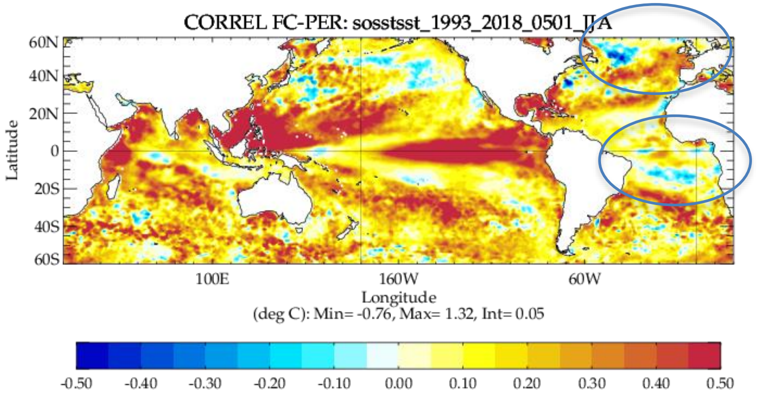
Example 2: Seasonal predictions of SST, Initialized in May, verified in JJA

A) SKILL (ACC) Benchmarked against persistence

Model



Model - Persistence

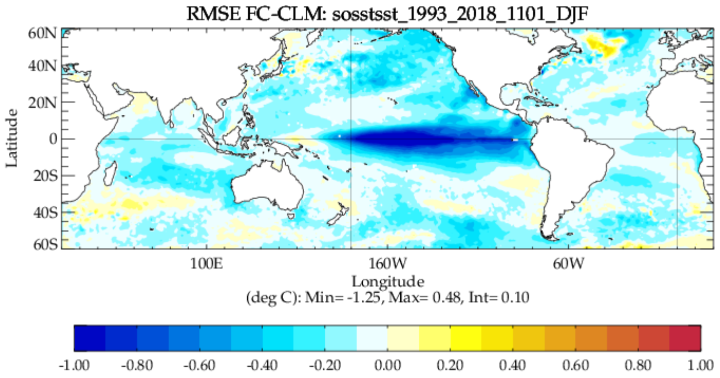


Model better

Model worse

B) RMSE: Benchmarked against Climatology

Model - CLIMATOLOGY

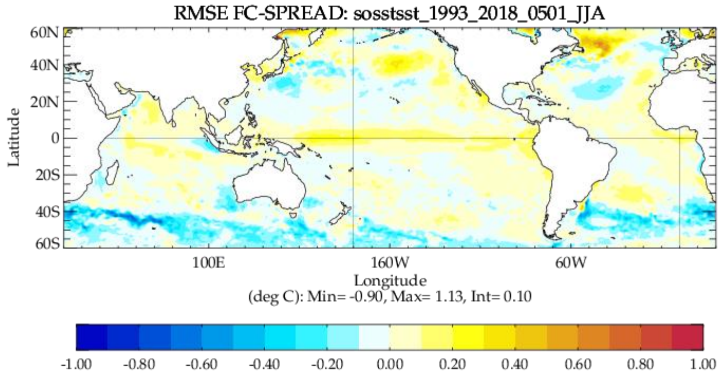


Model better

Model worse

C) Reliability: Benchmarked against Ensemble Spread

Reliability: RMSE Model - Ensemble Spread



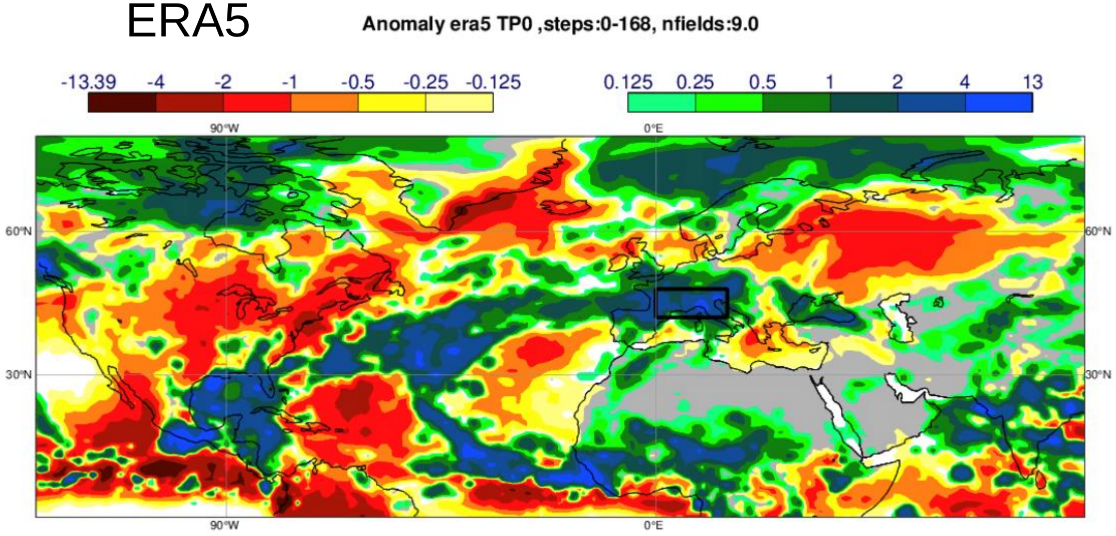
Overdispersive

Underdispersive

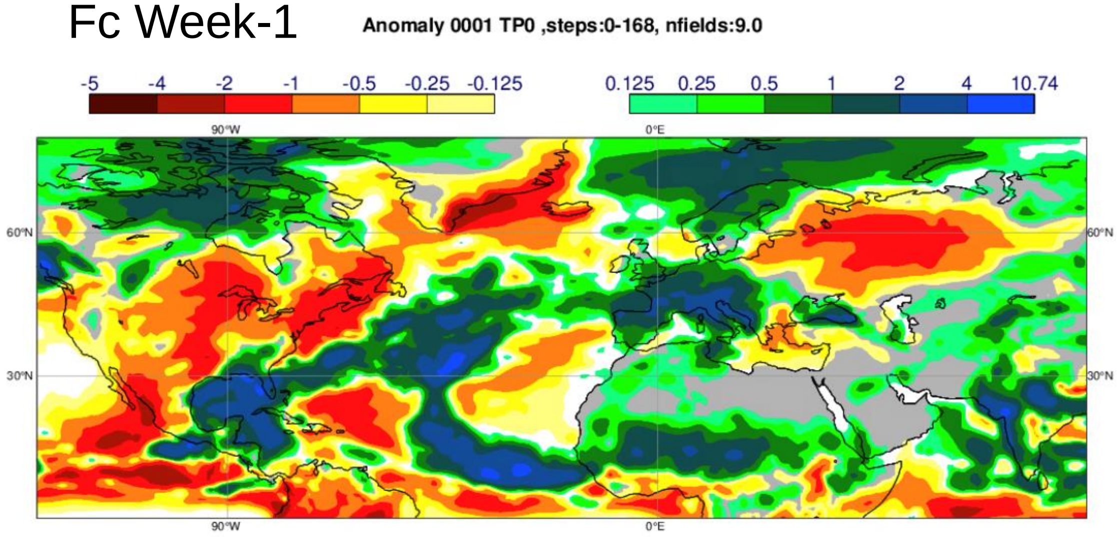
8. Separating uncertainty from model error

Precipitation anomaly 26 August – 27 October (mm/day)

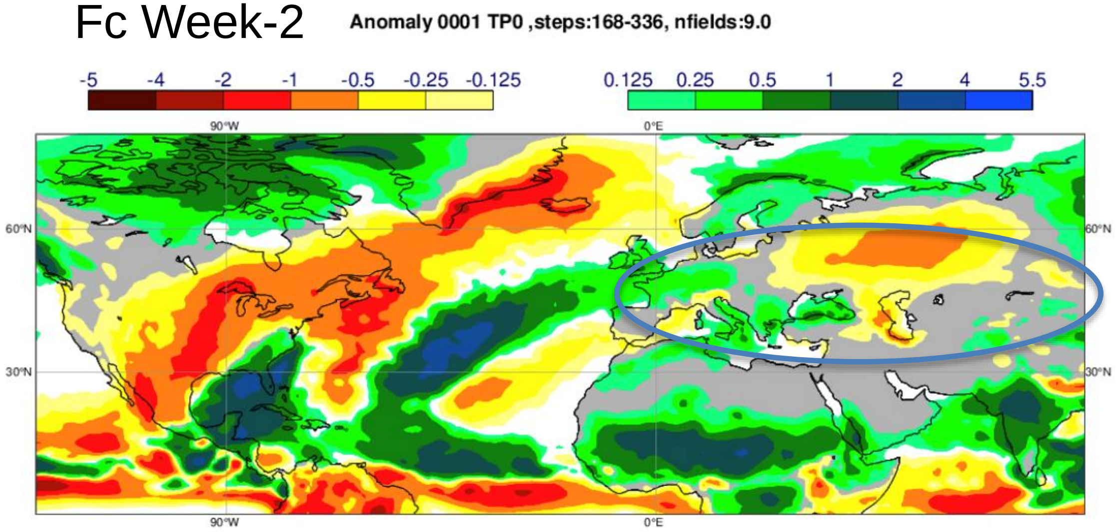
ERA5



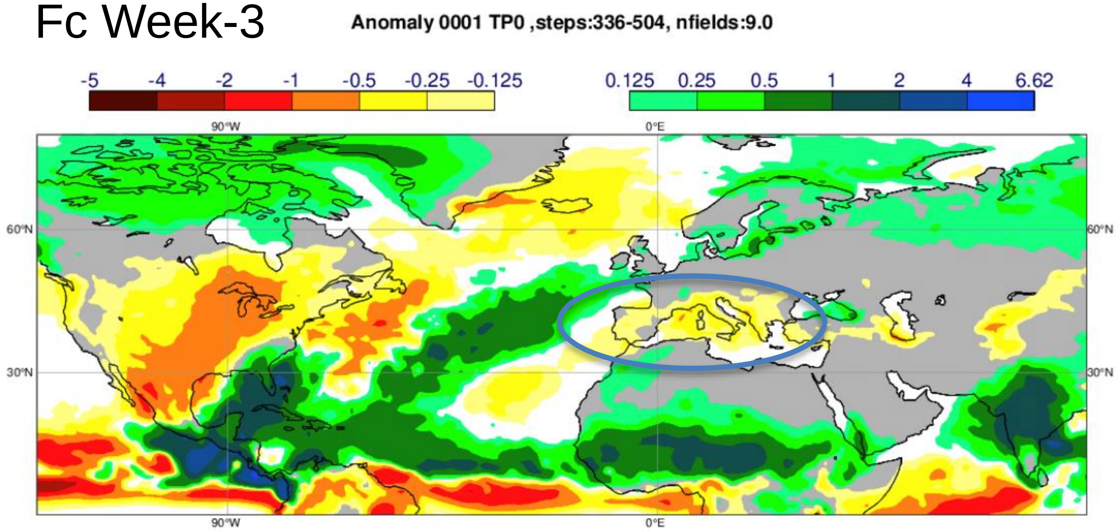
Fc Week-1



Fc Week-2



Fc Week-3



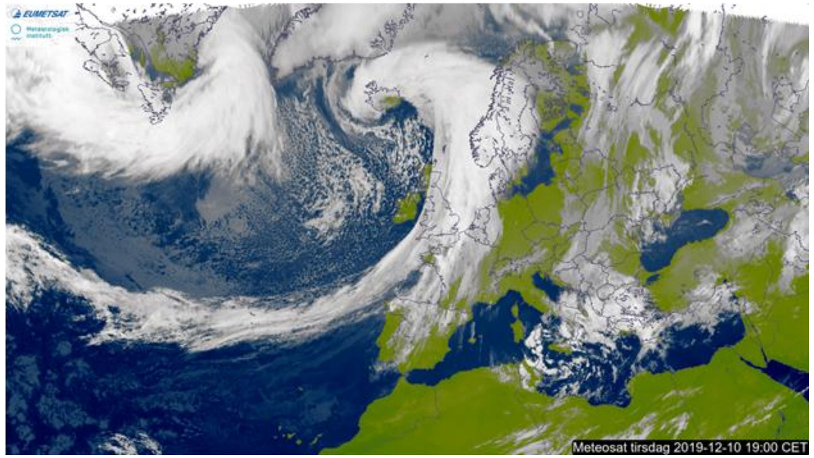
Weaker signal consistent with large uncertainty loss of deterministic predictability

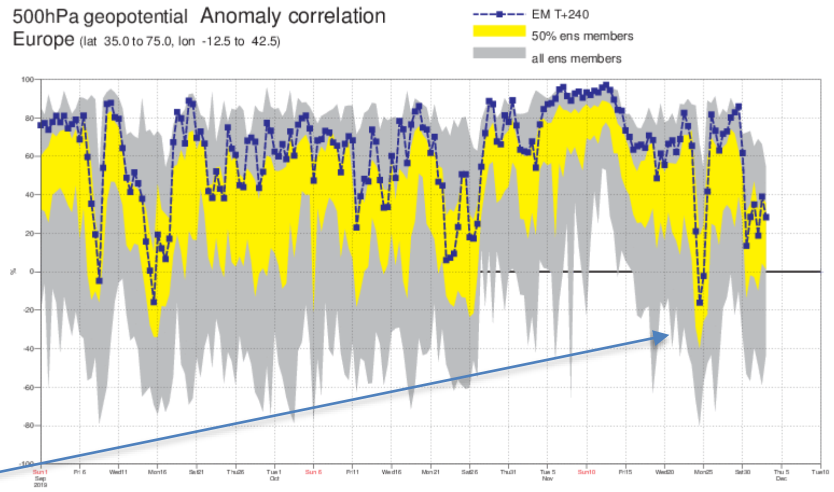
Wrong sign of signal, suggesting model error

Courtesy of Linus Magnuson

9. Different kinds of predictability. Example of a medium range forecast bust

Wind Storm 2019-12-10

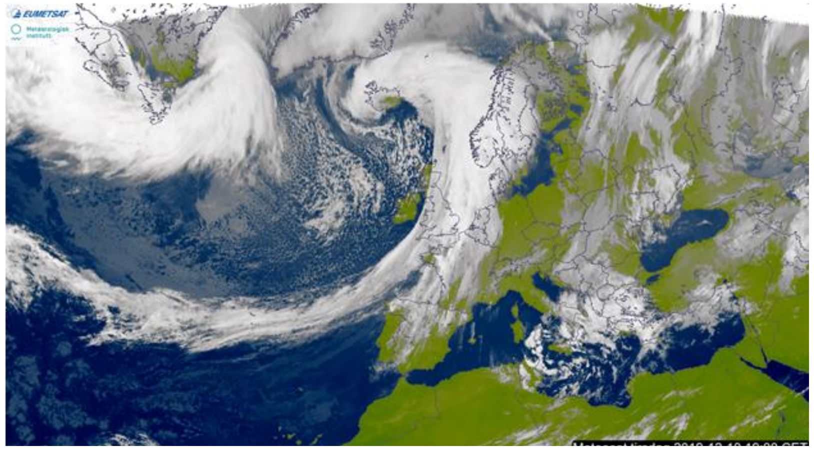


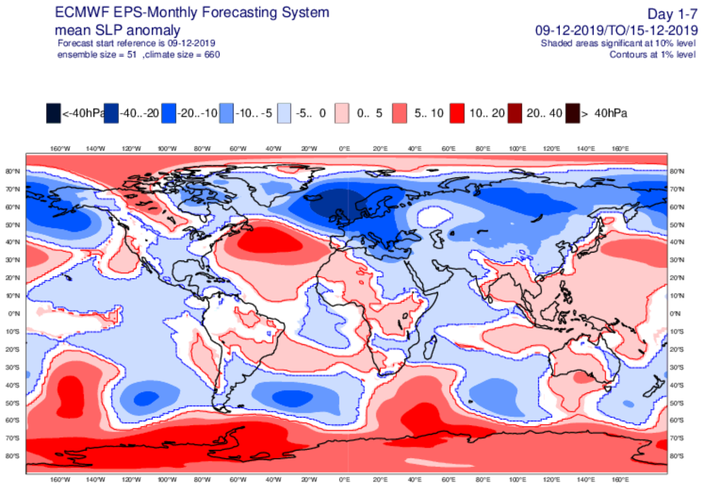


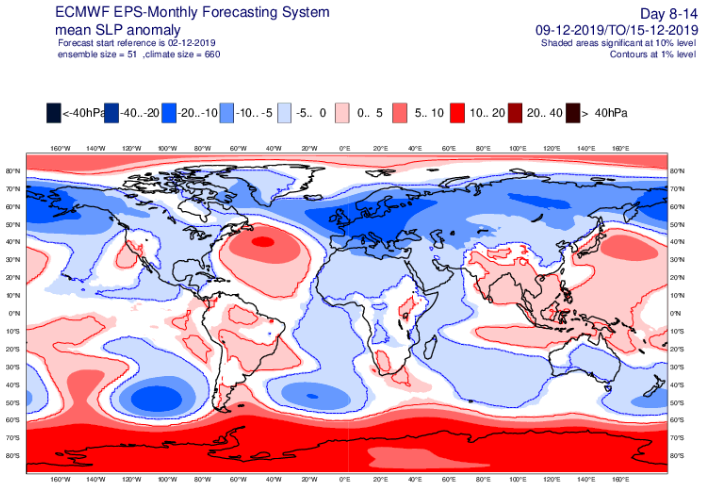
The skill for predicting this specific storm at the medium range was unusually low, as indicated by the anomalous low value of the anomalous correlation skill – forecast bust.

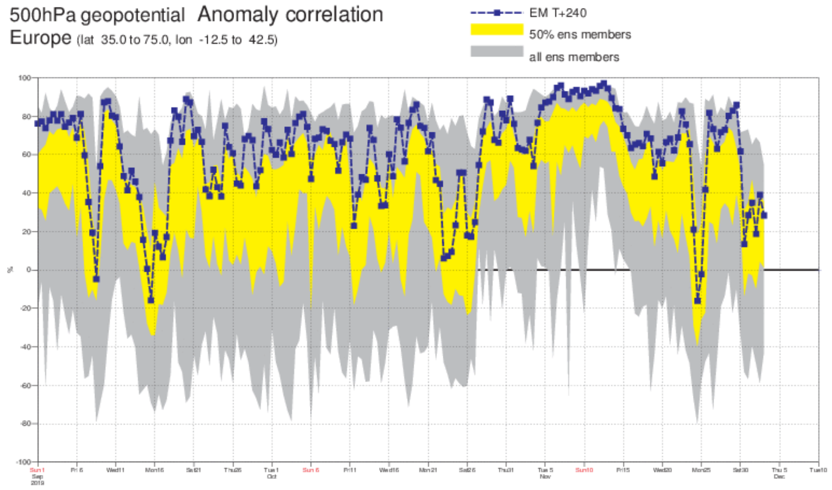
9. Different kind of predictability.

Wind Storm 2019-12-10

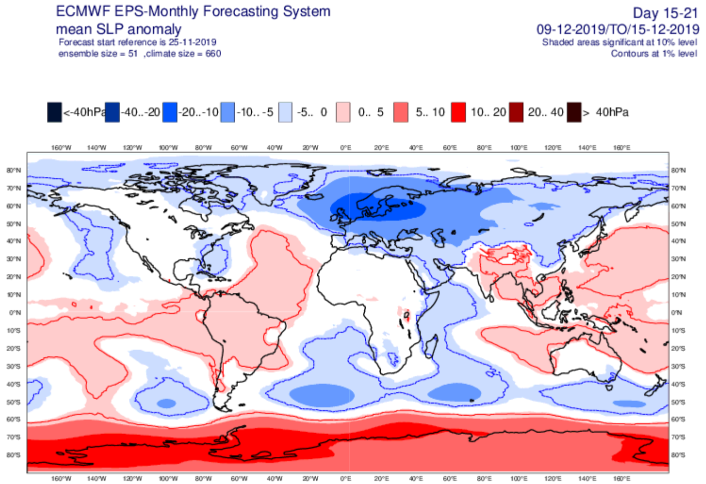








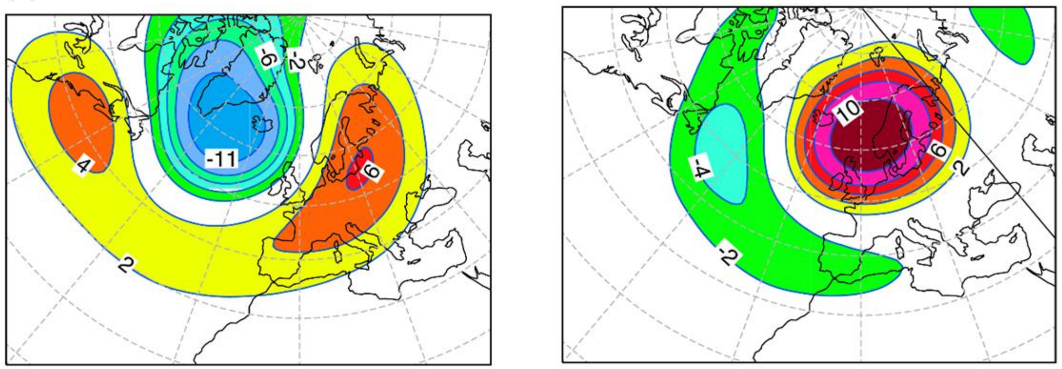
The subseasonal forecast was consistently predicting strong probability of zonal flow at week 3



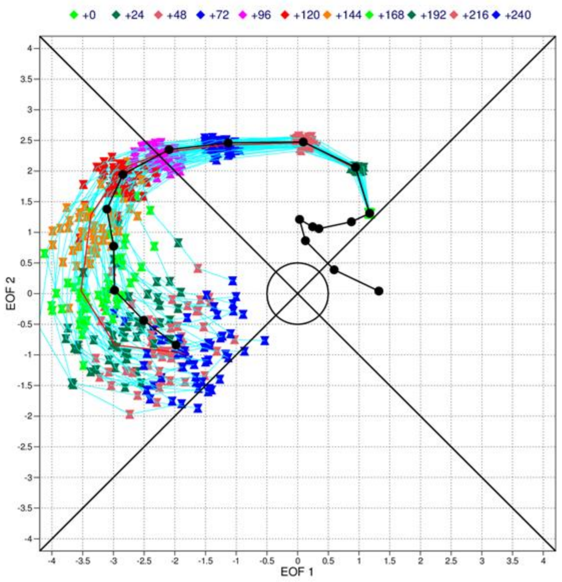
9. Different kinds of predictability.

A small detour: Interpreting extended range forecasts in reduced space

EOFs of North Atlantic Sector



EOF1 /NAO   EOF2/BLO

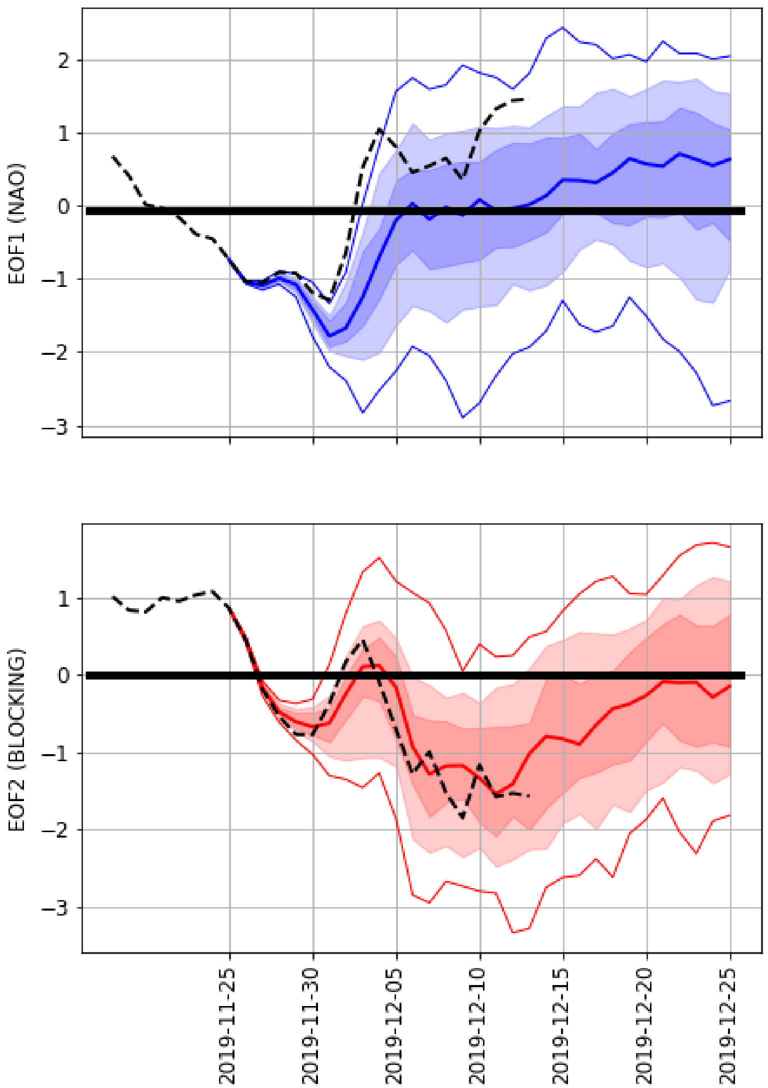


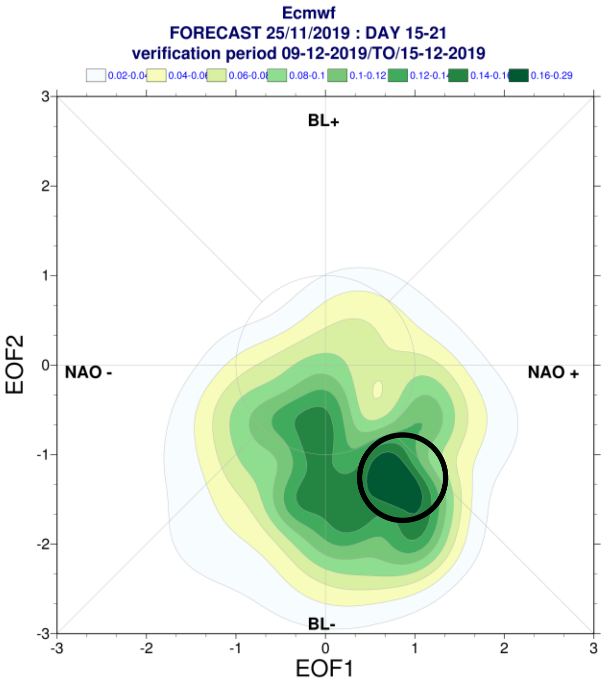
The reduction can also be done in terms of regimes

Ferranti et al 2018

9. Different kind of predictability.

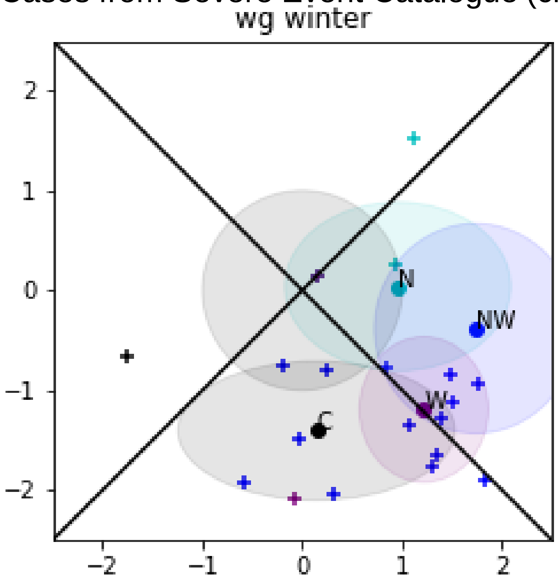
Projections on EOFs in forecast from 25 November





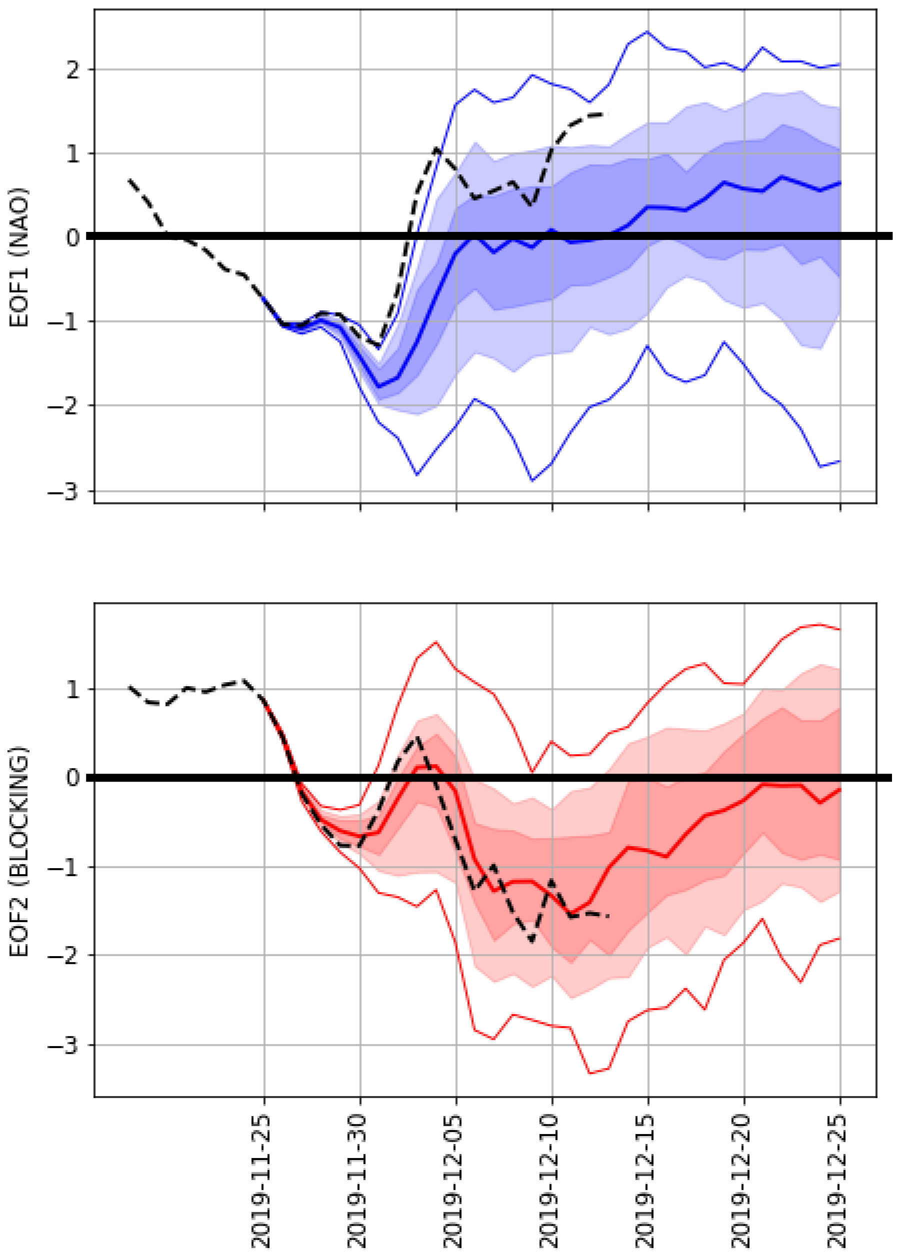
Climatology of extreme

Extreme short forecasts (circles)
Cases from Severe Event Catalogue (crosses)

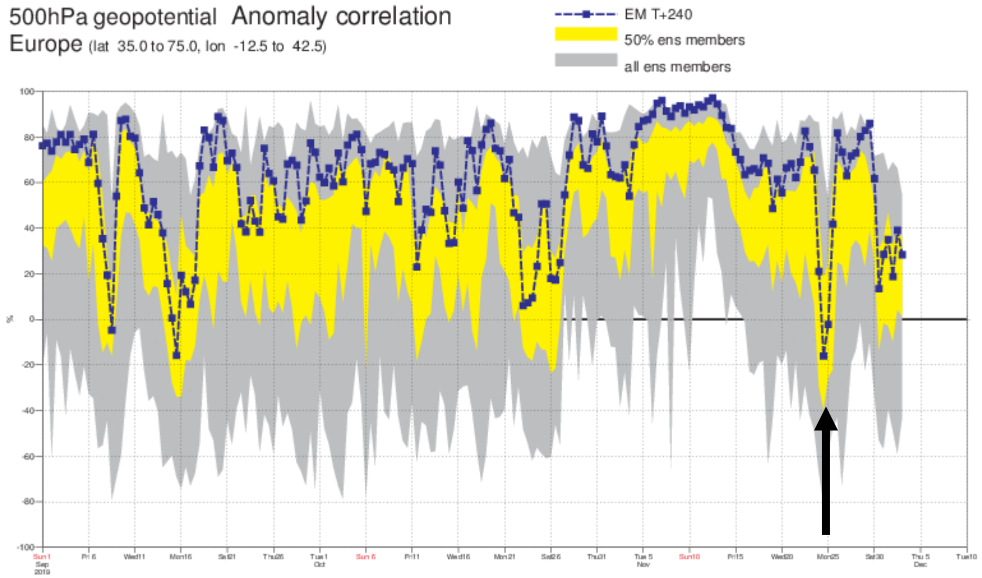


Courtesy of Linus Magnusson

Projections on EOFs in forecast from 25 November 2019



ACC for 10 day forecasts



The predictability of regimes was high in this occasion, indicating higher than average risk of stormy weather.

However, the skill for prediction of the specific storm was unusually low, as indicated by the anomalous low value of the anomalous correlation skill – forecast bust.

Courtesy of Linus Magnusson

FORECAST STRATEGIES: THE WIDER CONTEXT

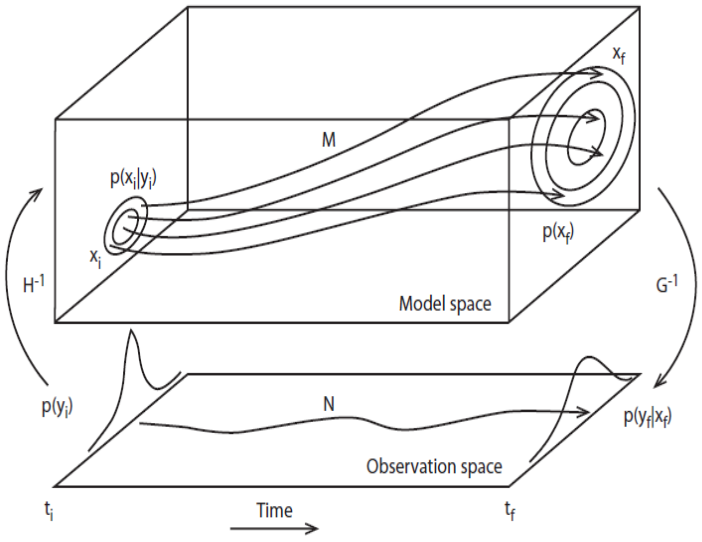
Treatment of system errors (until recently)

2) Propagating information, uncertainty and error into the future: *Forecast model*

- ✓ Stochastic parameterizations for sub-grid processes.
- X Other missing processes and earth system components not represented
- X Model bias is not targeted

1) Initialization
Data Assimilation

- ✓ Initial uncertainty considered.
- ✓ Model uncertainty starts being considered.
- ✓ Observation uncertainty considered
- ✓ Observation bias considered
- X Model bias was often ignored in atmosphere.



Model: $x = \bar{x} + \acute{x} + \varepsilon_x$

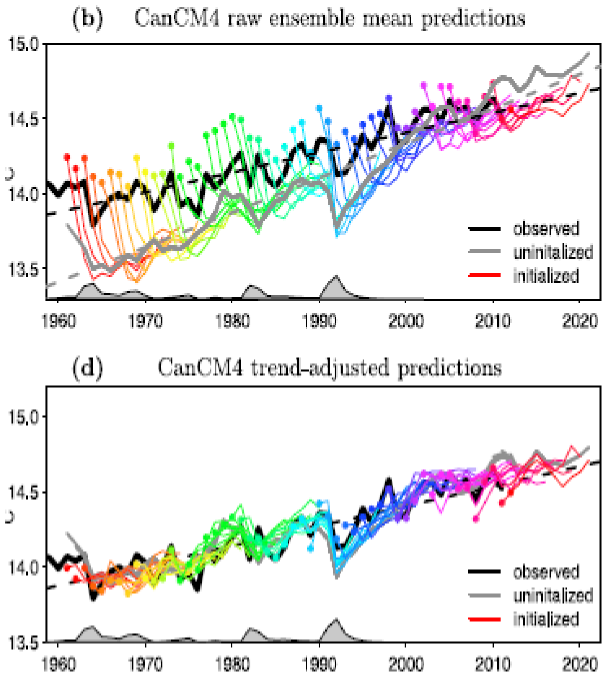
Observations: $y = \bar{y} + \acute{y} + \varepsilon_y$

3) Calibration
Forecast Assimilation

- Model Bias accounted for: removed a posterirori. Stockdale et al 1997
- Model uncertainty considered (ensemble)
- Observation error neglected *
- Residuals can be non stationary, non gaussian.
  Limitation to forecast skill calibration is more difficult

ECMWF

Calibration is complex if errors are non stationary



From Kharin eta la 2012

$$\tilde{x} = \bar{y} + \mathbf{K}(x - \bar{x}) + \mathbf{F}\varepsilon_x + \mathbf{T}(t) + \mathrm{G}(y_0)$$

Bias correction ($\bar{\boldsymbol{x}} \neq \bar{\boldsymbol{y}}$)

K: linear transformation of anomalies

F: Adjustment of ensemble spread

T: detrending

G: other flow dependent corrections

Error in mean state errors degrades variability and forecast skill, making forecast errors non stationary and calibration difficult. Too many parameters

Can model error be treated more explicitly during the forecast process?

Stephenson et al 2005
Kharin et al 2012
Fukar et al 2014

Mean state error influencing model fidelity and skill

Correcting model biases leads to better representation of variability (or model fidelity) :

(several papers: D'Andrea and Vautard 2000, Balmaseda et al 2010, Scaife 2011, ….)

Correcting bias in tropical SST improves seasonal forecast skill of ENSO, tropical cyclones…

Magnusson et al 2012, Vecchi et al 2014:

Correcting biases in atmosphere improves seasonal atmospheric predictability:

Kharin and Scinocca 2012

Correcting North Atlantic SST bias improves subseasonal skill over North Atlantic and Europe

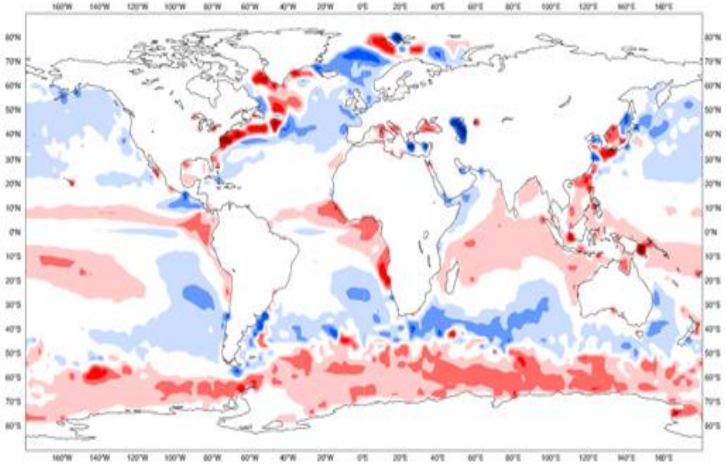
Roberts et al 2021, Vitart and Balmaseda 2018

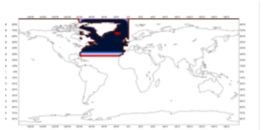
Non linear interactions: North Atlantic SST mean errors impact subseasonal forecast skill

SST Biases Week 4 (day 26-32)

DJF (162 start dates)

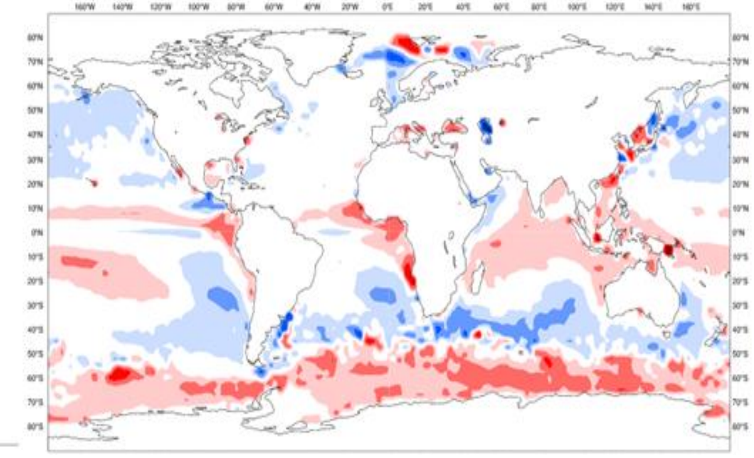
CONTROL





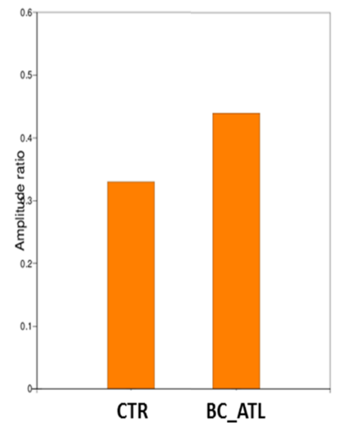
SST corrected o
dark area

SST bias corrected (BC) in NATL



3rd pentad after MJO Phase 7

NAO- Teleconnections



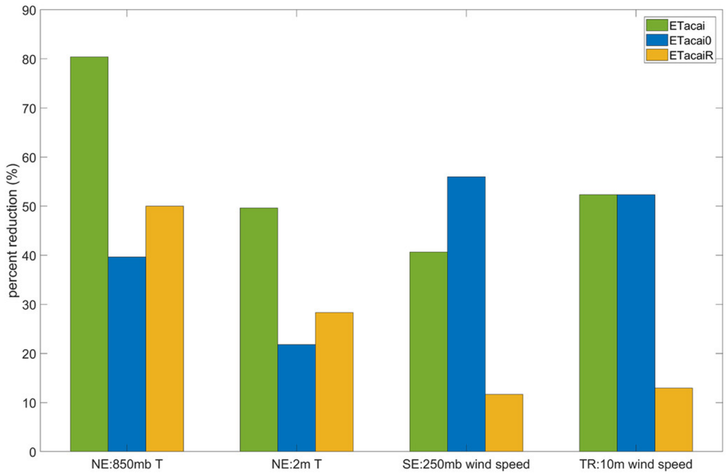
Correcting bias on SST over North Atlantic impacts the skill over Europe by improving MJO/ NAO –ve teleconnections

From Vitart and Balmaseda 2018

Recent efforts on treatment of model error

- At ECMWF
  - Assimilation Phase
    - Treatment of model biases usually considered in ECMWF ocean reanalysis (Balmaseda et al 2007)
    - the stratospheric model error started being treated around 2020, using Weak Constrain 4D-Var (Laloyaux et al 2020).
  - Forecast Phase: none yet, although there are ongoing efforts for empirical representation of model error

- In other operational centers:
  - Met Office (Piccolo et al 2020) and NRL (Crawford et al 2020) use past assimilation increments to represent model error, both random and systematic components



NRL system

Percent reduction in day-10 forecast bias

- Total
- Bias only
- Random component only

From Crawford et al 2020

Treatment of system errors

2) Propagating information, uncertainty and error into the future: *Forecast model*

Before 2024

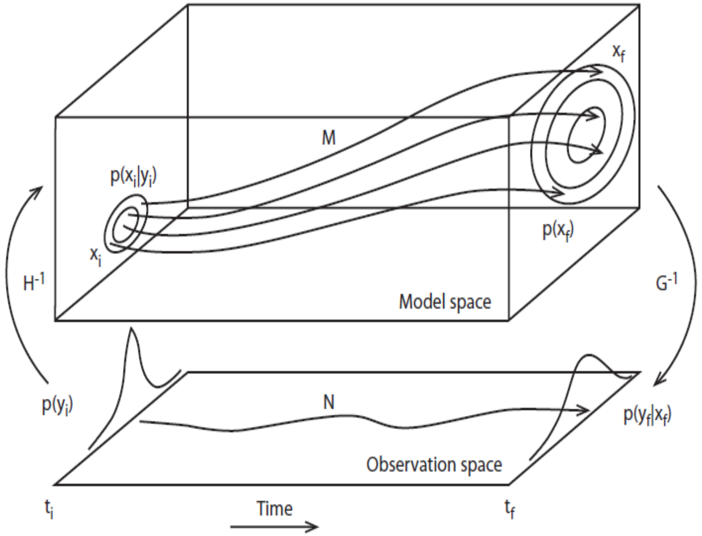
After 2024

✓ Stochastic parameterizations for sub-grid processes.

X Other missing processes and earth system components not represented

X Model bias explicitly modelled (e.g. ML)

X Model bias is not targeted



1) Initialization
Data Assimilation

- ✓ Initial uncertainty considered.
- ✓ Model uncertainty starts being considered.
- ✓ Observation uncertainty considered
- ✓ Observation bias considered

(before 2020)

(after 2020)

- X Model bias often ignored in atmosphere.
- ✓ **Model errors considered during assimilation**

Model: $x = \bar{x} + \acute{x} + \varepsilon_x$

Observations: $y = \bar{y} + \acute{y} + \varepsilon_y$

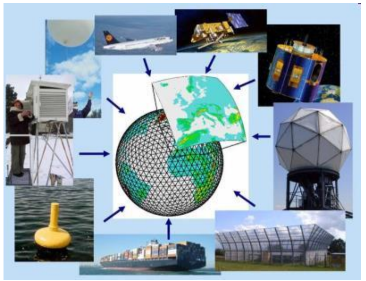
3) Calibration
Forecast Assimilation

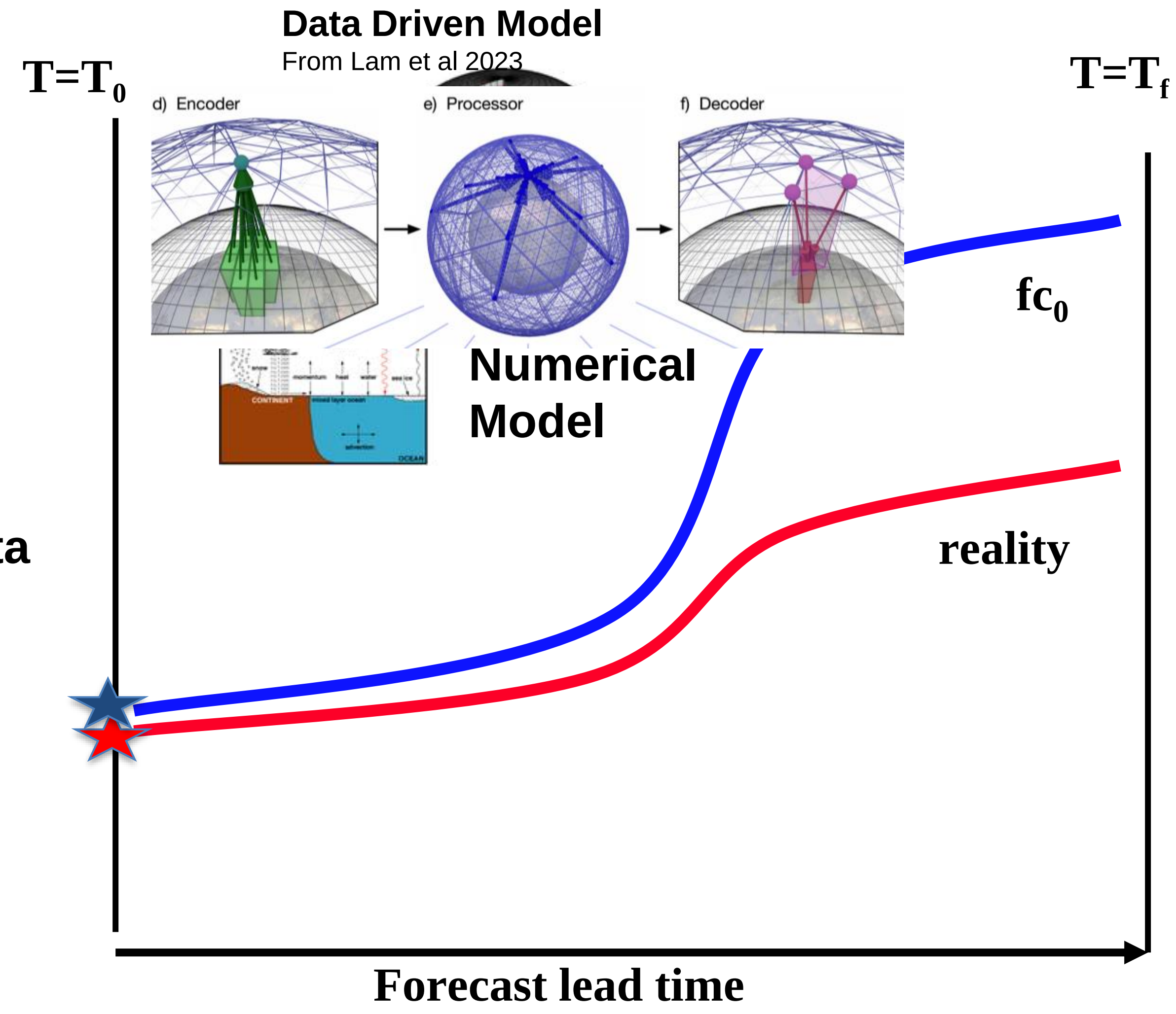
- Model Bias accounted for: removed a posterirori. Stockdale et al 1997
- Model uncertainty considered (ensemble)
- Observation error neglected *
- Residuals can be non stationary, non gaussian.
  Limitation to forecast skill calibration is more difficult

ECMWF

Data Driven Models: A revolution in weather prediction

$X(t_0)$ from observations + data assimilation



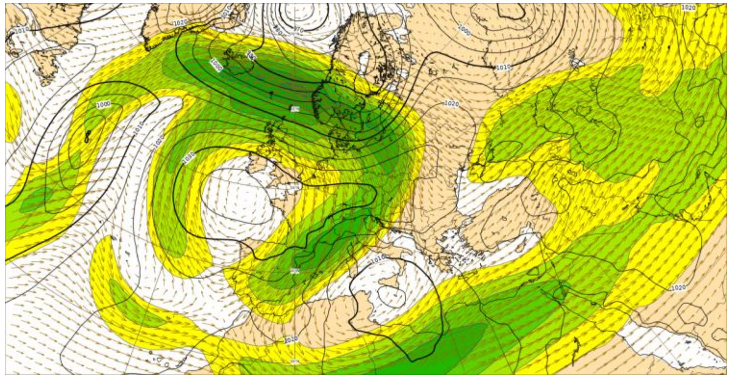


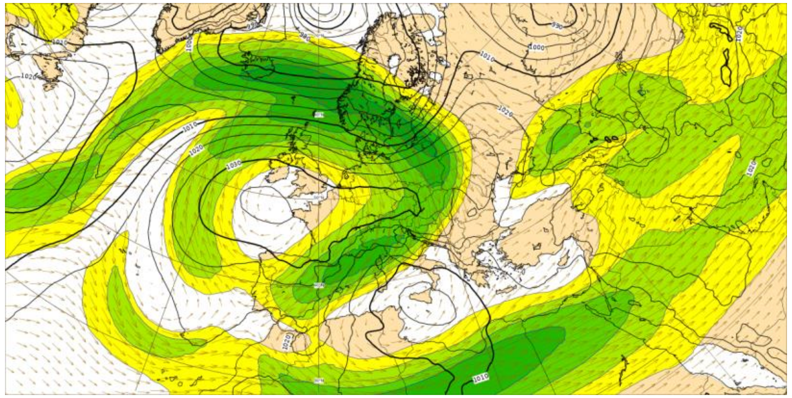
- Substantial skill gains in the medium range.
- Much cheaper

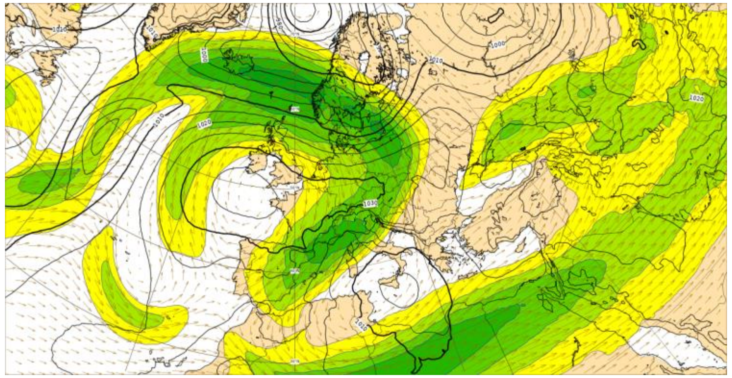
- Need reanalyses (most competitive models are based in ERA5), which have been produced using traditional numerical models

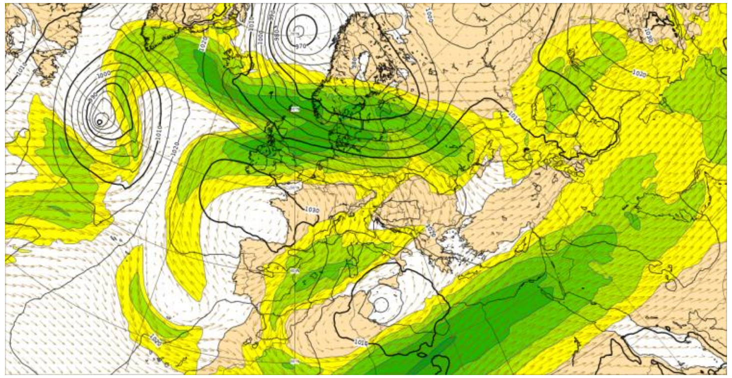
- Challenges ahead for ML: higher resolution, more user-relevant variables, and longer lead time

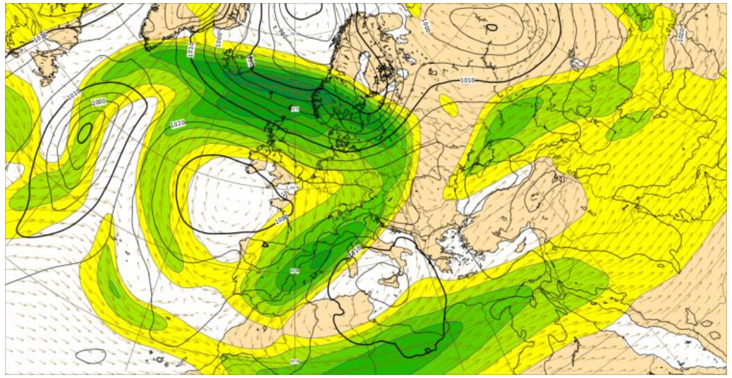
- Can we use ML for data assimilation?

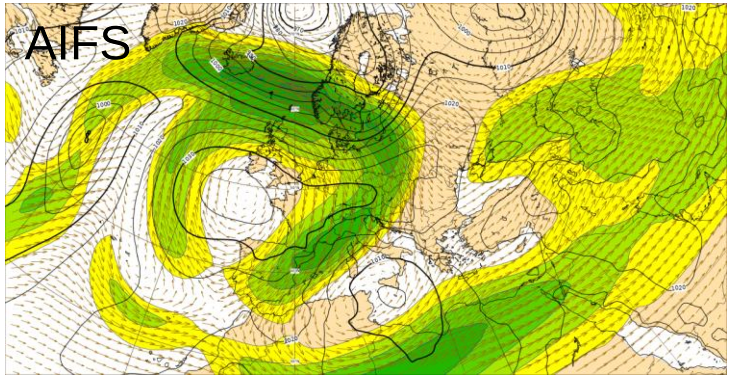


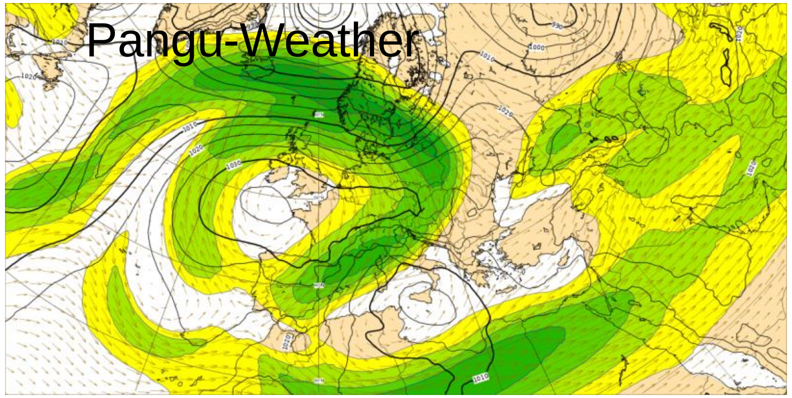


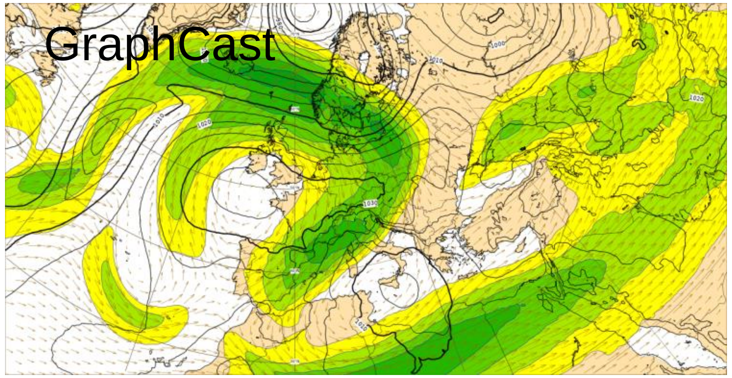


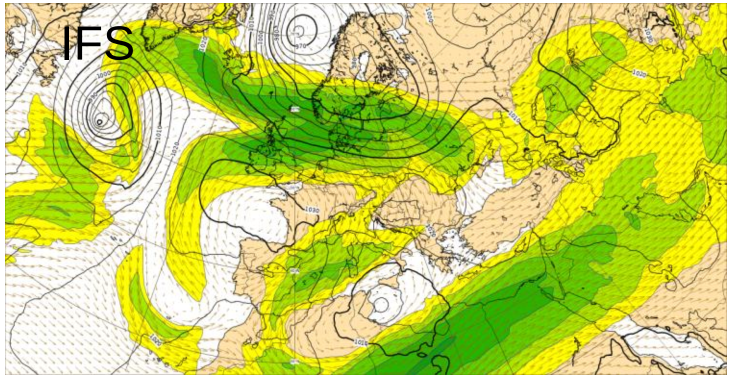


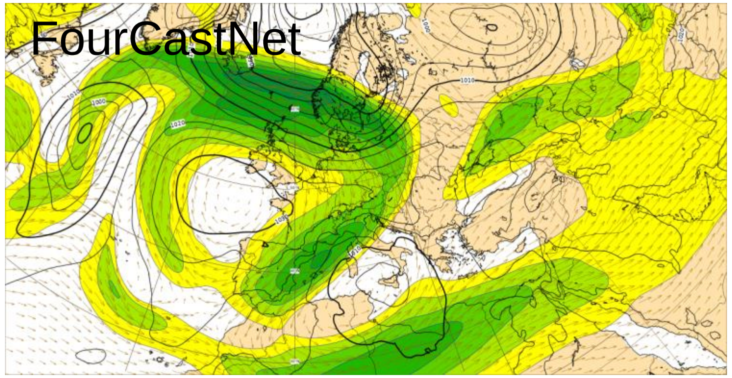



AIFS


Pangu-Weather


GraphCast


IFS


FourCastNet

Summary

- Weather and climate prediction as initial value problem of probabilistic nature. Ensemble prediction

- Predictability drivers and limiters. **Importance to distinguish error from uncertainty**

- Slow time scales extend the forecast horizon: Criteria to include slow earth system components in a forecasting system: physical basis, ability to model and to **initialize**. The **initialization distinguish a prediction from a climate scenario projection**

- Different sort of predictability: A specific event may be difficult to predict in the deterministic sense in the short range, but its probability of occurrence may be well predicted at the extended range.

- Reforecast needed for calibration, skill assessment, detection of extremes. They are an integral part of forecasting system

- Importance to balance different elements: complexity, ensemble members, forecast length, resolution, reforecasts.

- Some examples of forecast products in reduced space

- A brief outlook of alternative forecasting strategies. Future: Coupling Dynamical with ML models?
- Data driven models: a revolution for NWP

Question for break out groups

How will the forecasting systems look like in 10 years time?

How will you design them?

Possible questions to consider:

- What will be the role of ML and traditional GCMs?
- Will we need a seamless approach?
- Will we need reforecasts? If so, what will be the reforecast period?
- What will be expected enhancements in terms of skill, products, forecast horizon?

Thanks for your attention!

Any questions?