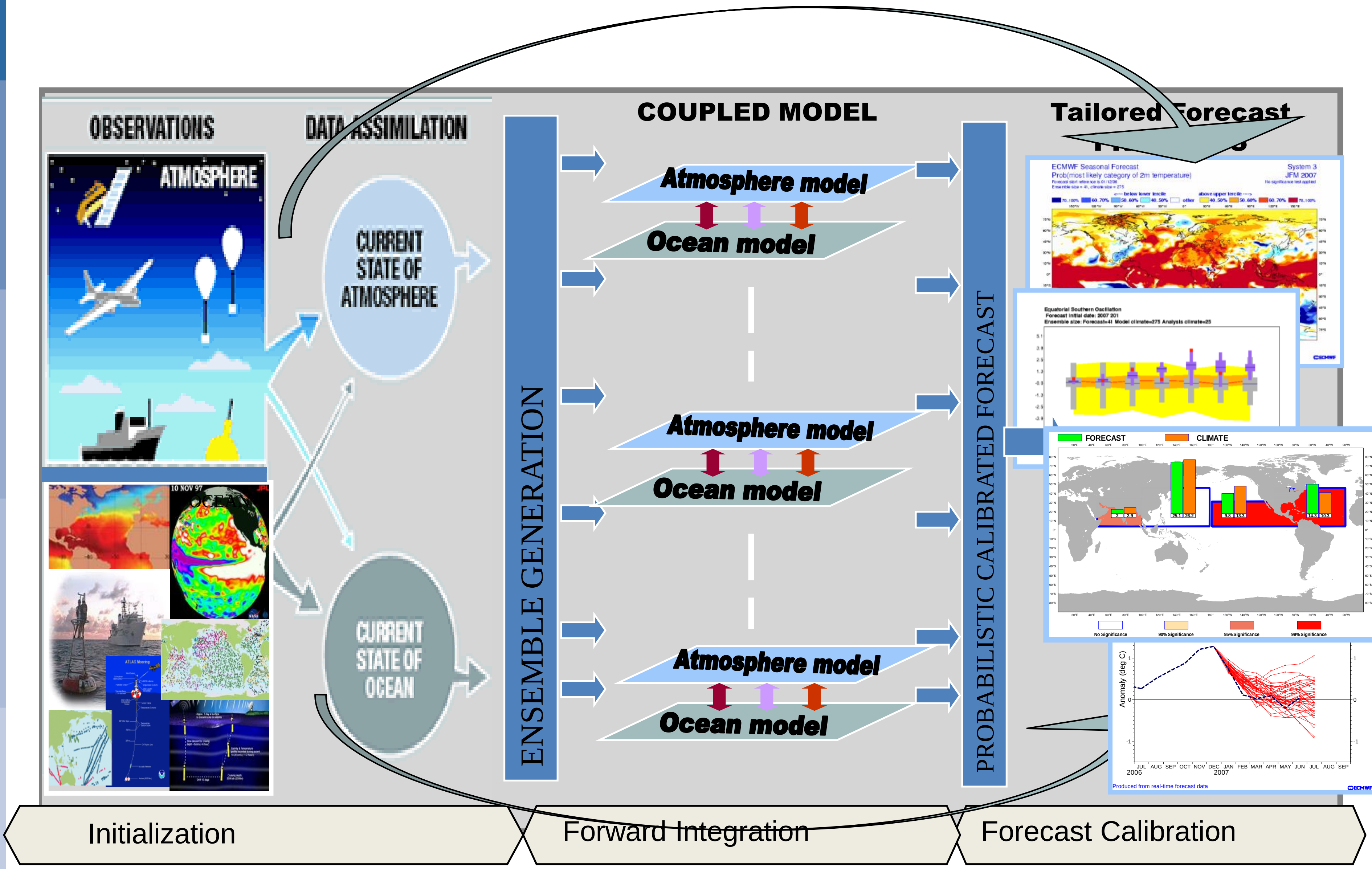
S2S models and Database

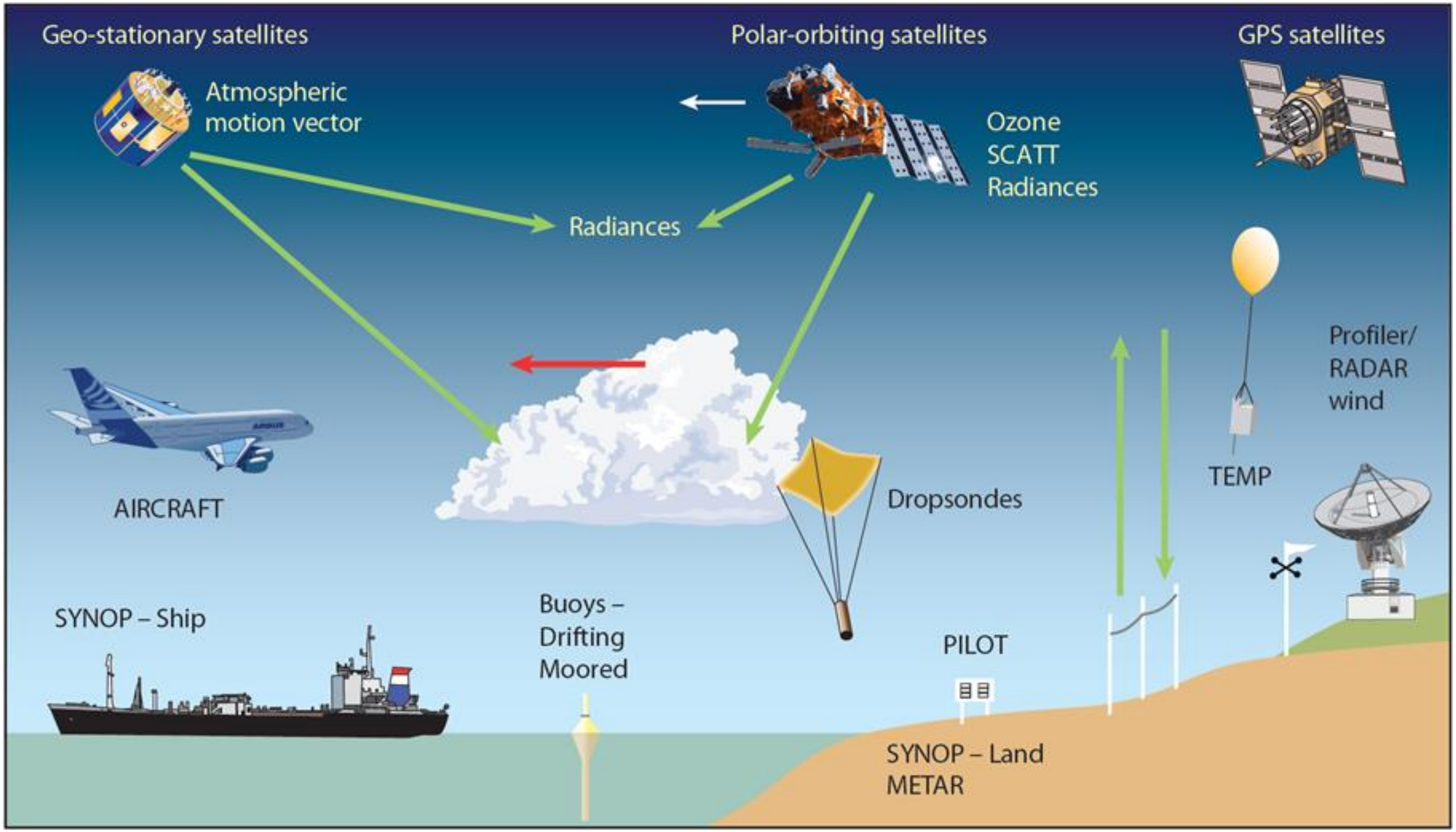
Frédéric Vitart

S2S Models

End-To-End forecasting System



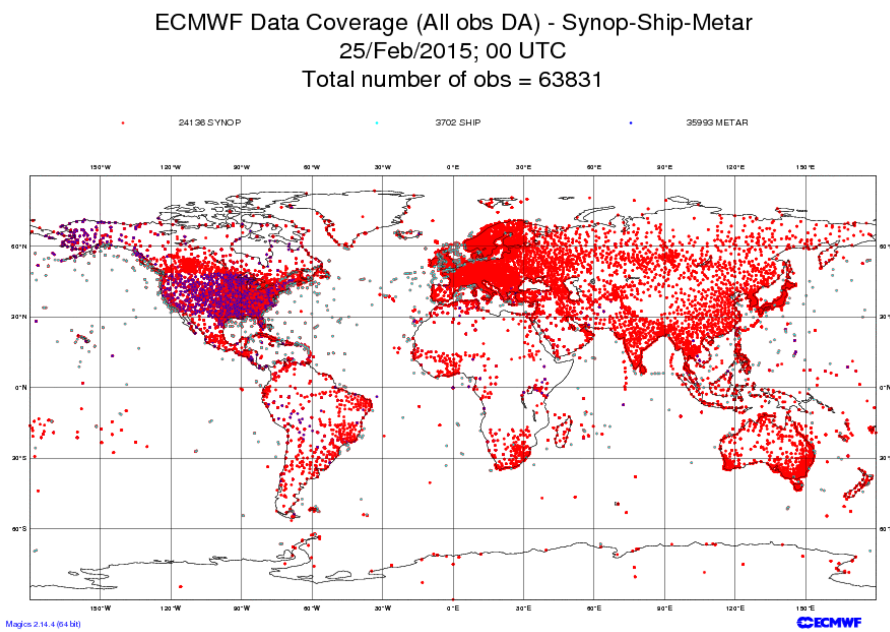
Informations to initialize the atmosphere

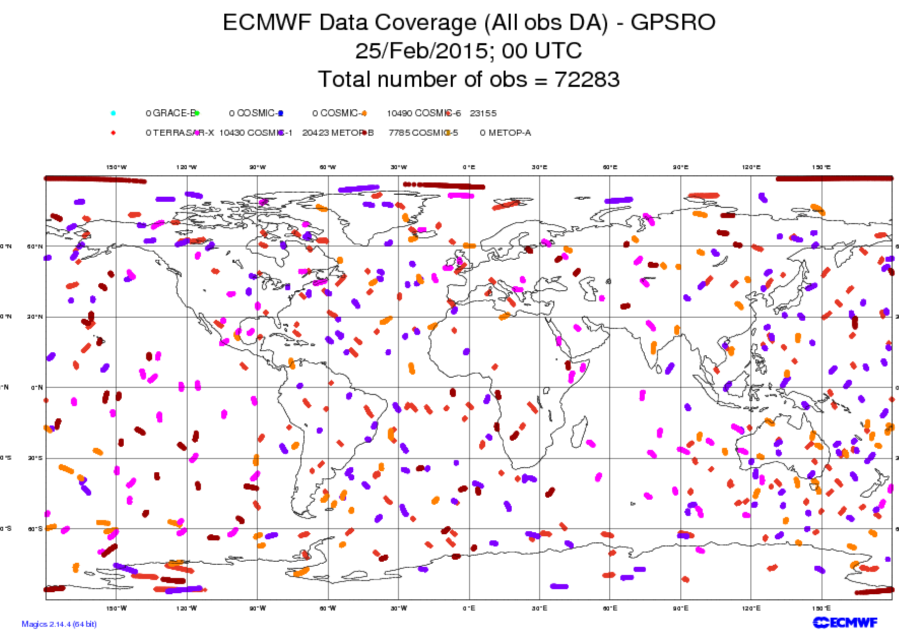


Observations coverage and accuracy

To make accurate forecasts it is important to know the current weather:

- ~ 155M obs (99% from satellites) are received daily;
- ~ 15M obs (96% from satellites) are used every 12 hours.





Information to initialize the ocean

- Ocean model Plus:

  SST

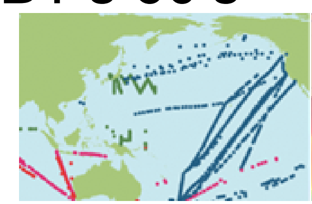
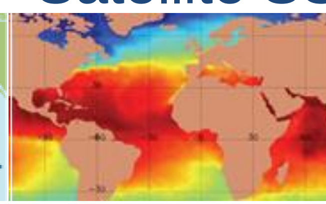
  Atmospheric fluxes from atmospheric reanalysis

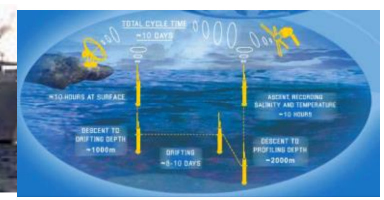
  Subsurface ocean information

Time evolution of the Ocean Observing System

1982 | 1993 | 2001

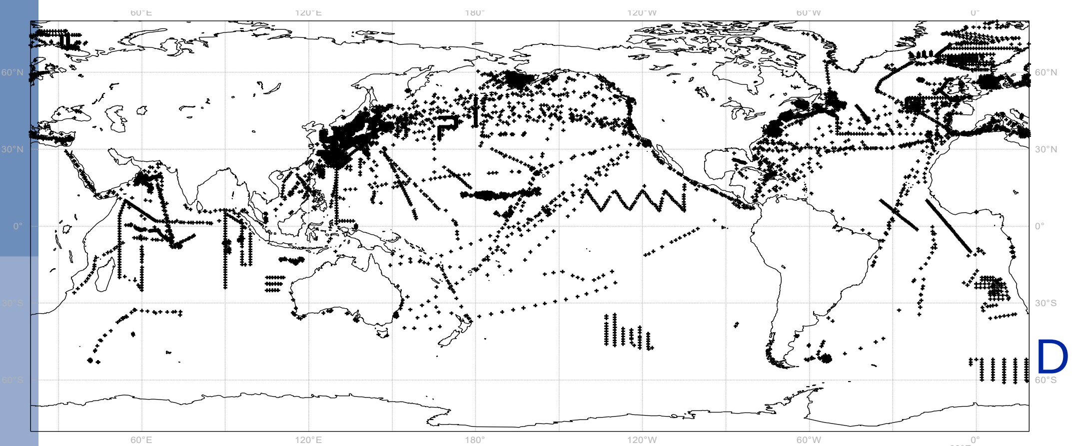
XBT's 60's | Satellite SST | Moorings/Altimeter ARGO



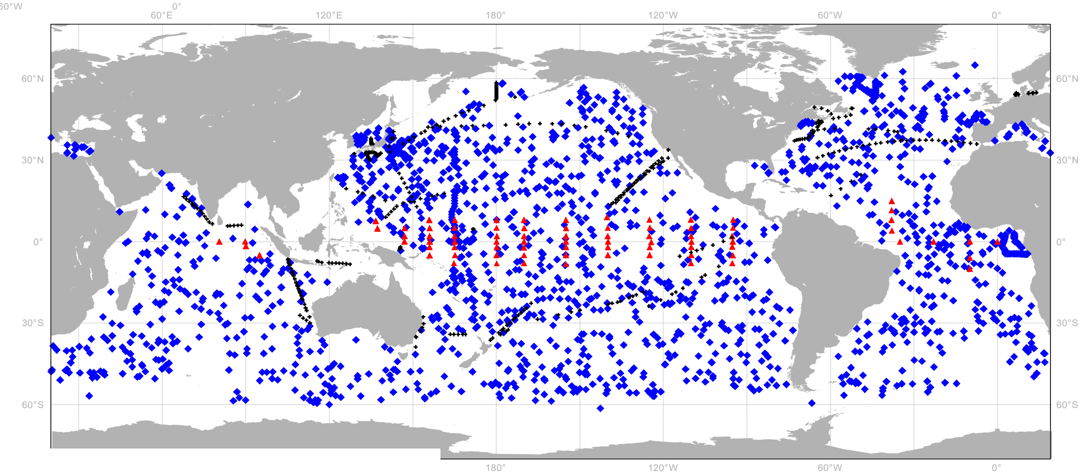

Ocean observing system

Data coverage for June 1982



Changing observing system is a challenge for consistent reanalysis

Data coverage for Nov 2005



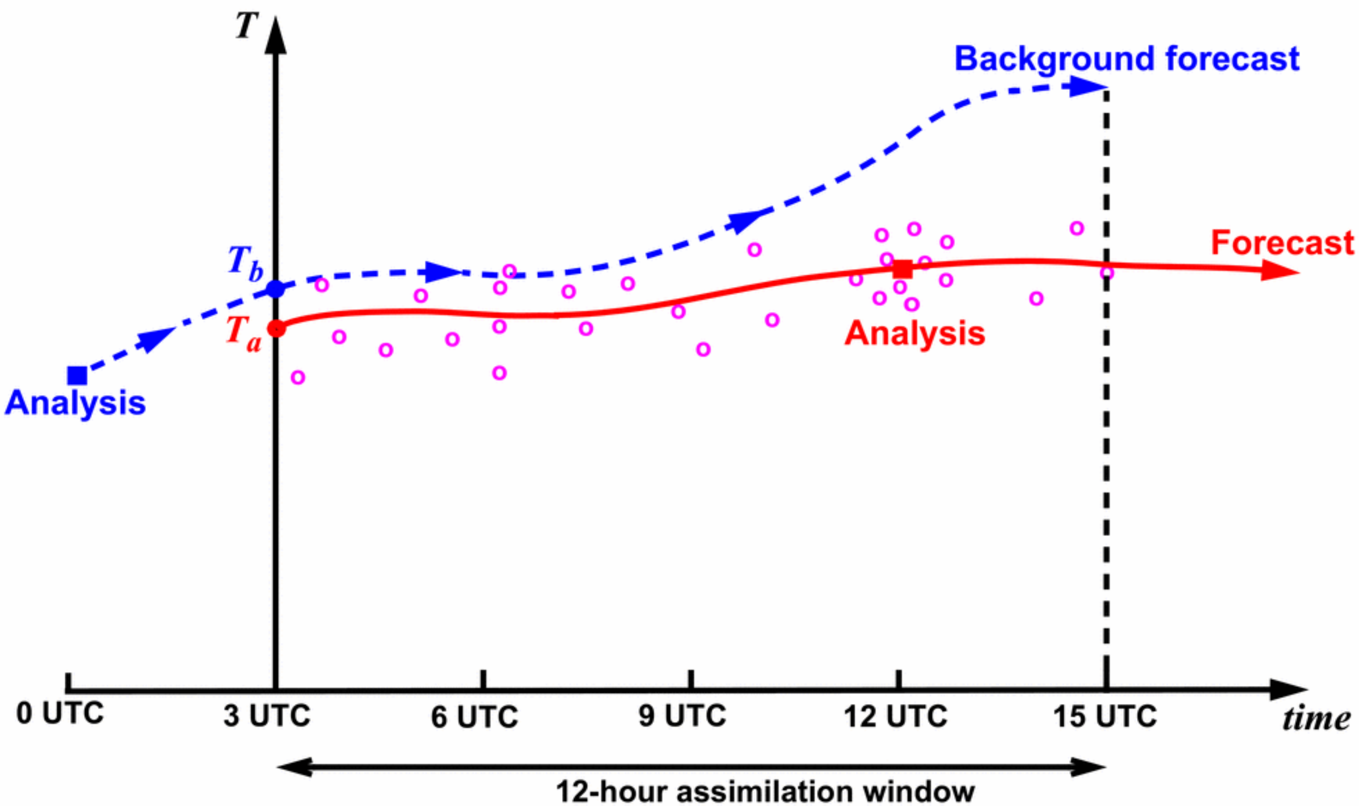
Today's Observations will be used in years to come

▲Moorings: SubsurfaceTemperature

◊ ARGO floats: Subsurface Temperature and Salinity

+ XBT : Subsurface Temperature

The ECMWF 4D-Var data-assimilation system



ECMWF extended-range forecasts

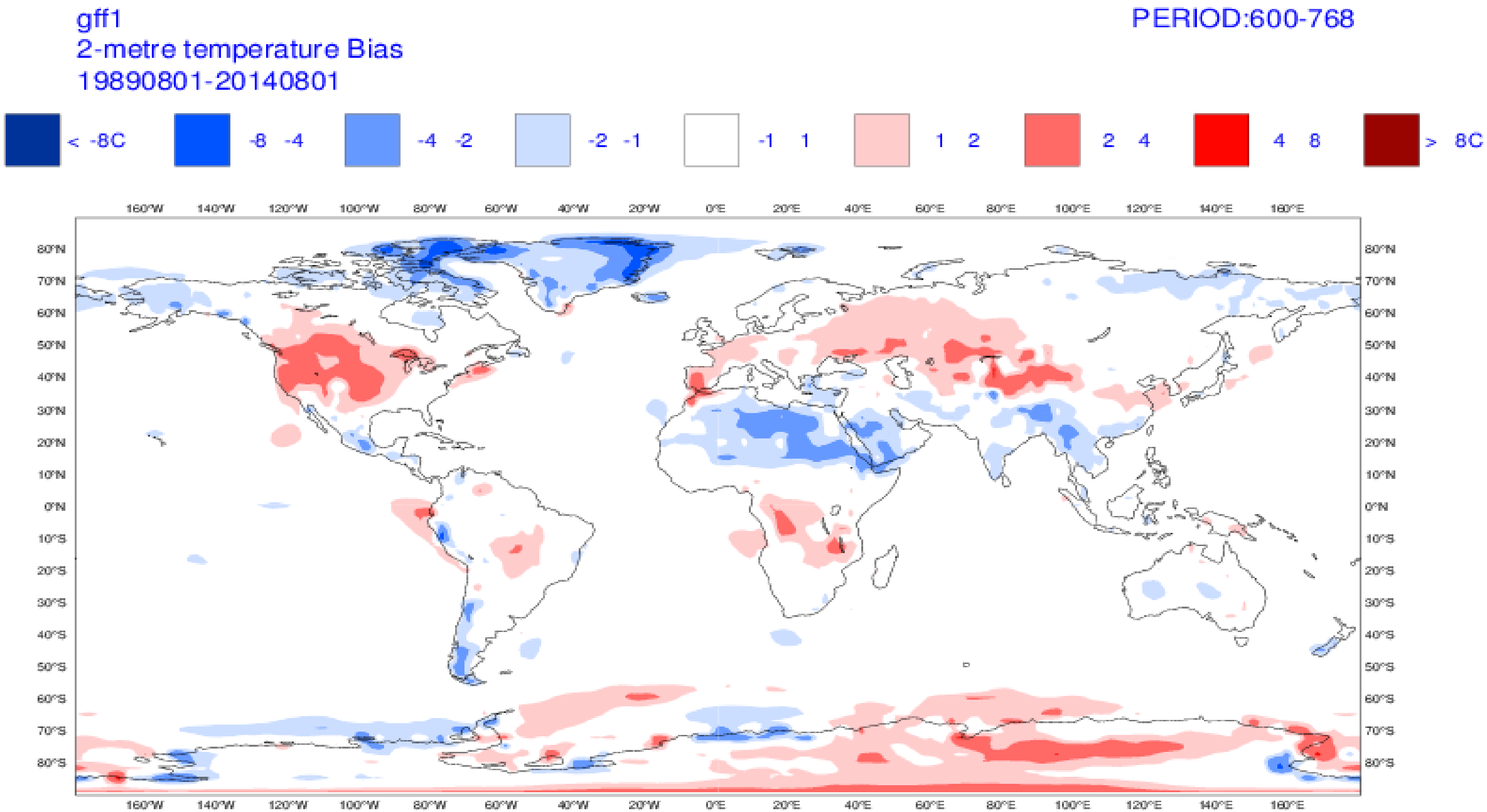
- A 51-member ensemble is integrated for 46 days twice a week (Mondays and Thursdays at 00Z)
- Atmospheric component: IFS with the latest operational cycle and with a 18 km resolution up to day 15 and 36 km after day 15.
- Ocean-atmosphere coupling from day 0 to NEMO (about 1/4 degree) every hour.

Initial conditions:

- Atmosphere: Operational 4-D var analysis + SVs+ EDA perturbations
- Ocean: 3D-Var analysis (NEMOVAR) + wind stress perturbations

Model Biases

Biases (eg 2mT as shown here) are often comparable in magnitude to the anomalies which we seek to predict



Re-forecast strategy

Re-forecasts are used for model calibration and also for skill assessment.

- A large reforecast database is needed for **calibration** to distinguish between random error and systematic errors and also to estimate flow dependent errors.

- A large reforecast database is also needed for verification and for flow dependant **skill assessment**, like assessing the concurrent impact of ENSO and specific phases of the MJO on the forecast skill scores. Signal to noise ration is also improved in long reforecast datasets (Shi et al, 2014)

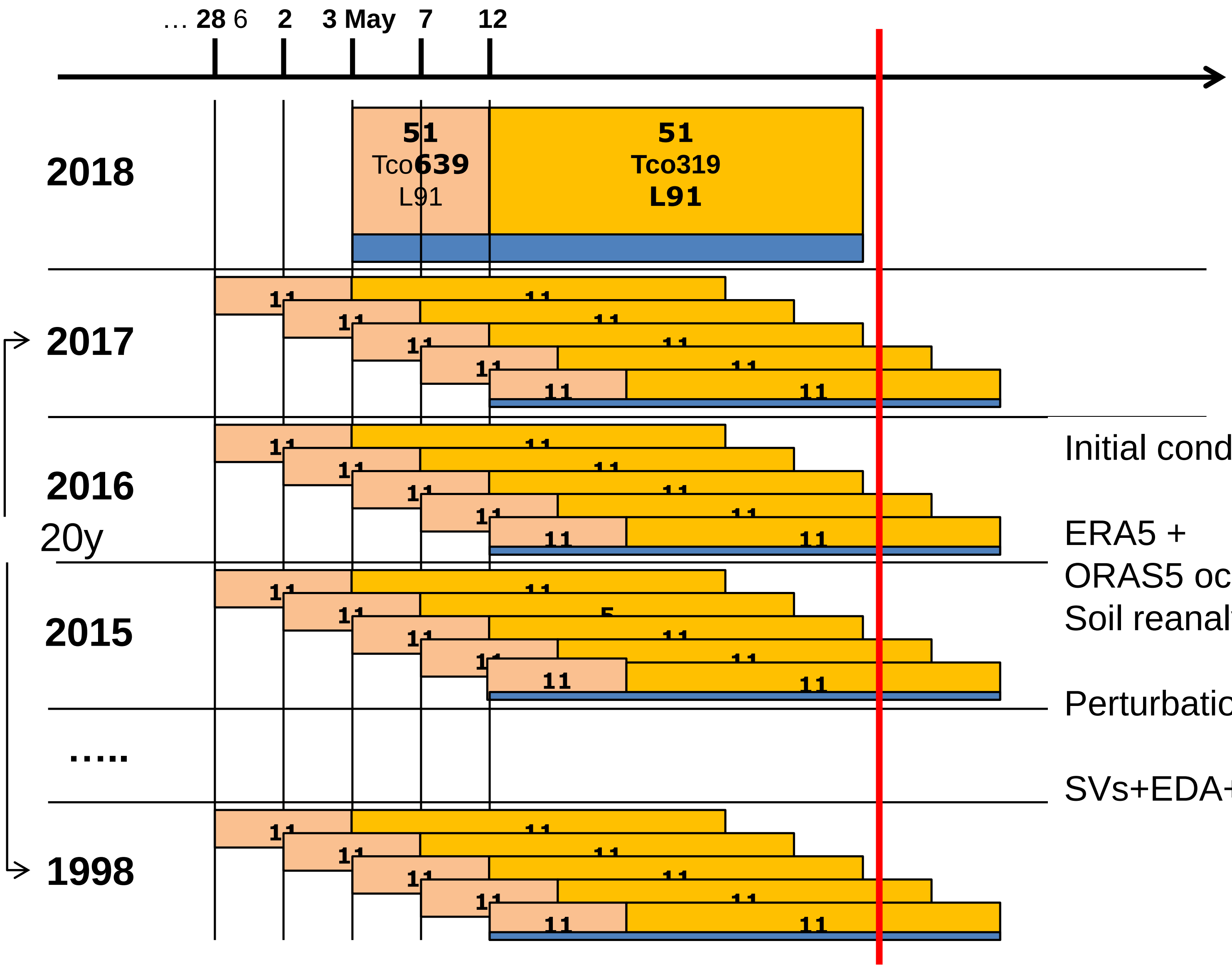
- Large ensemble size is also important for skill assessment , since some probabilistic skill scores are impacted by the ensemble size.

However

- Large re-forecast datasets with large ensemble size are often not affordable. Not clear what is more important: ensemble size, number of years?

- Long re-forecasts suffer from inconsistent quality in the initial conditions (pre-sat. period).

The ENS re-forecasts to estimate the M-climate



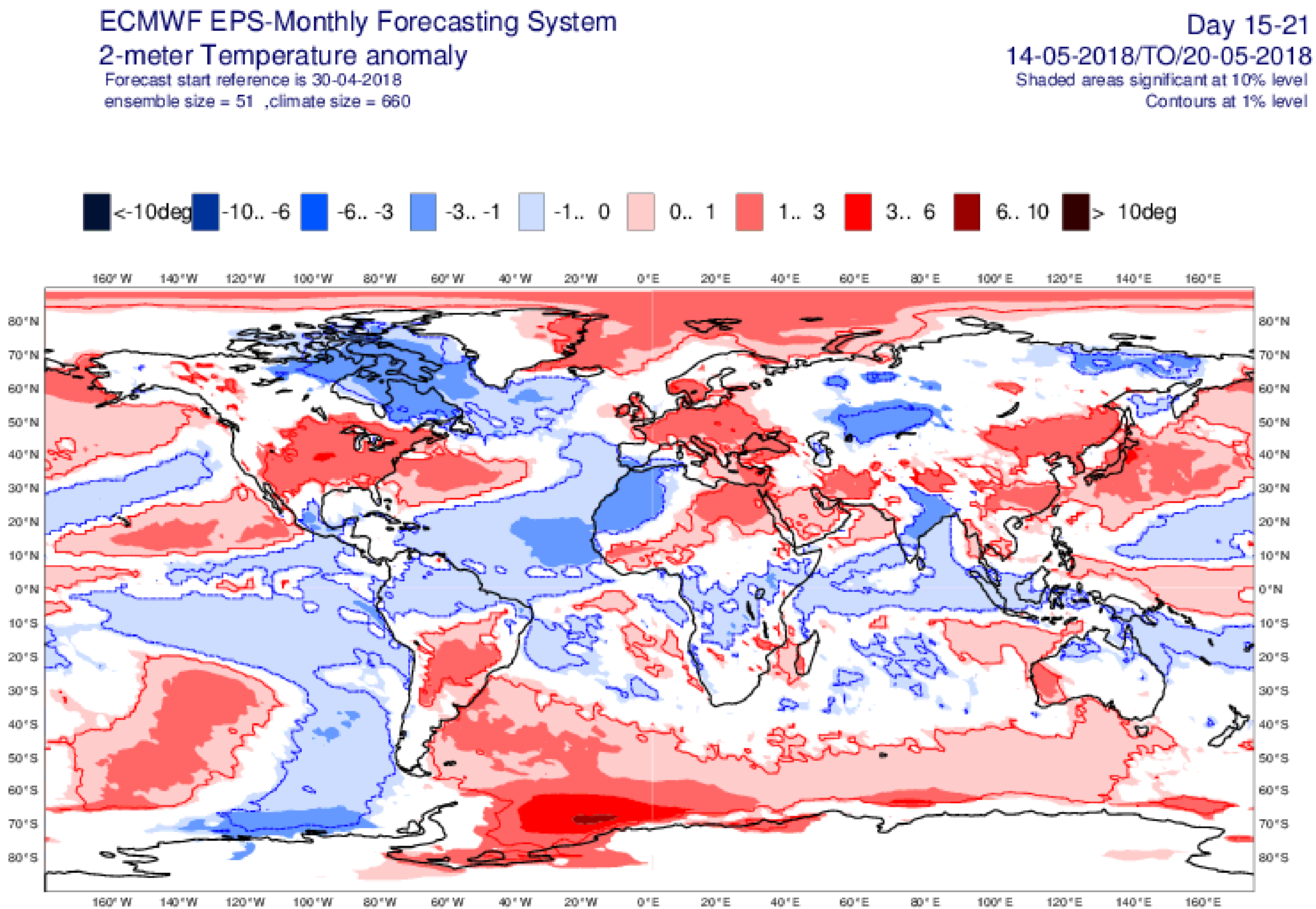
Initial conditions:

ERA5 +
ORAS5 ocean Ics+
Soil reanalysis

Perturbations:

SVs+EDA+SPPT+SKEB

Forecast Products



Anomalies (temperature, precipitation..)

S2S Database Models

Forecasts | **Re-forecasts**

Status on 5th January 2018	Time range	Resolution	Ens. Size	Frequency	Re-forecasts	Rfc length	Rfc frequency	Rfc size
BoM (ammc)	d 0-62	T47L17	3*11	2/week	fix	1981-2013	6/month	3*11
CMA (babj)	d 0-60	T106L40	4	daily	fix	1994-2014	daily	4
CNR-ISAC (isac)	d 0-32	0.75x0.56 L54	41	weekly	fix	1981-2010	every 5 days	5
CNRM (lfpw)	d 0-32	T255L91	51	weekly	fix	1993-2014	2/month	15
ECCC (cwao)	d 0-32	0.45x0.45 L40	21	weekly	on the fly	1995-2014	weekly	4
ECMWF (ecmf)	d 0-46	Tco639/319 L91	51	2/week	on the fly	past 20 years	2/week	11
HMCR (rums)	d 0-61	1.1x1.4 L28	20	weekly	on the fly	1985-2010	weekly	10
JMA (rjtd)	d 0-33	Tl479/Tl319L100	50	weekly	fix	1981-2010	3/month	5
KMA (rksl)	d 0-60	N216L85	4	daily	on the fly	1991-2010	4/month	3
NCEP (kwbc)	d 0-44	T126L64	16	daily	fix	1999-2010	day	4
UKMO (egrr)	d 0-60	N216L85	4	daily	on the fly	1993-2015	4/month	7

see s2sprediction.net for details and to access data

S2S Database Models

Forecasts | **Re-forecasts**

Status on 5th January 2018	Time range	Resolution	Ens. Size	Frequency	Re-forecasts	Rfc length	Rfc frequency	Rfc size
BoM (ammc)	d 0-62	T47L17	3*11	2/week	fix	1981-2013	6/month	3*11
CMA (babj)	d 0-60	T106L40	4	daily	fix	1994-2014	daily	4
CNR-ISAC (isac)	d 0-32	0.75x0.56 L54	41	weekly	fix	1981-2010	every 5 days	5
CNRM (lfpw)	d 0-32	T255L91	51	weekly	fix	1993-2014	2/month	15
ECCC (cwao)	d 0-32	0.45x0.45 L40	21	weekly	on the fly	1995-2014	weekly	4
ECMWF (ecmf)	d 0-46	Tco639/319 L91	51	2/week	on the fly	past 20 years	2/week	11
HMCR (rums)	d 0-61	1.1x1.4 L28	20	weekly	on the fly	1985-2010	weekly	10
JMA (rjtd)	d 0-33	Tl479/Tl319L100	50	weekly	fix	1981-2010	3/month	5
KMA (rksl)	d 0-60	N216L85	4	daily	on the fly	1991-2010	4/month	3
NCEP (kwbc)	d 0-44	T126L64	16	daily	fix	1999-2010	day	4
UKMO (egrr)	d 0-60	N216L85	4	daily	on the fly	1993-2015	4/month	7

All models have produce a real-time forecast every Thursday!

S2S Database Models

Forecasts | **Re-Forecasts**

Status on 5th January 2018	Time range	Resolution	Ens. Size	Frequency	Re-forecasts	Rfc length	Rfc frequency	Rfc size
BoM (ammc)	d 0-62	T47L17	3*11	2/week	fix	1981-2013	6/month	3*11
CMA (babj)	d 0-60	T106L40	4	daily	fix	1994-2014	daily	4
CNR-ISAC (isac)	d 0-32	0.75x0.56 L54	41	weekly	fix	1981-2010	every 5 days	5
CNRM (lfpw)	d 0-32	T255L91	51	weekly	fix	1993-2014	2/month	15
ECCC (cwao)	d 0-32	0.45x0.45 L40	21	weekly	on the fly	1995-2014	weekly	4
ECMWF (ecmf)	d 0-46	Tco639/319 L91	51	2/week	on the fly	past 20 years	2/week	11
HMCR (rums)	d 0-61	1.1x1.4 L28	20	weekly	on the fly	1985-2010	weekly	10
JMA (rjtd)	d 0-33	Tl479/Tl319L100	50	weekly	fix	1981-2010	3/month	5
KMA (rksl)	d 0-60	N216L85	4	daily	on the fly	1991-2010	4/month	3
NCEP (kwbc)	d 0-44	T126L64	16	daily	fix	1999-2010	day	4
UKMO (egrr)	d 0-60	N216L85	4	daily	on the fly	1993-2015	4/month	7

Common re-forecast period is 1999-2010!

Re-forecasts

Two strategies for re-forecasts in S2S database:

- **Fixed re-forecasts** (e.g. NCEP/BoM/JMA)

The model version used to produce the sub-seasonal forecasts is “frozen” for a number of years (e.g. CFS2). The re-forecasts have been produced once for all before the system became operational.

Advantage: More user friendly. The user can compute skill and calibration once for all.

- **“on the fly” re-forecasts** (e.g. ECMWF/UKMO/ECC..)

The model version changes frequently (at least once a year). Therefore re-forecasts have to be produced regularly since the model version of the re-forecasts has to be the same as the real-time forecasts.

Advantage: This methodology ensures the best model version has been used to produce the sub-seasonal forecasts.

Re-forecasts

Re-forecasts are assigned 2 dates:

- Model version date (date of production of the data)
- Hindcast date (start date of the re-forecast)

For fixed re-forecasts, Model version date is fixed (e.g. 20110301 for NCEP)

For on-the-fly re-forecasts, Model version date = YYYY-MM-DD

where YYYY= year of data production and MM-DD are same month and day as Hindcast date

For example, the ECMWF hindcast starting on 20000101 which was produced in 2019 is referred to as:

- Model version date = 20190101
- Hindcast date = 20000101

S2S Database

Model Description (s2s.ecmwf.int)

Models

Created by Richard Mladek, last modified by Craig MacLachlan on Jan 25, 2017

This table shows the centres that provide data to this project together with the **latest** configuration of their systems. Follow the link of each Data Provider for specific model description.

Status on 26th January 2017	Time range	Resolution	Ens. Size	Frequency	Re-forecasts	Rfc length	Rfc frequency	Rfc size	Volume of real-time forecast per cycle	Volume of reforecast per update
BoM (ammc)	d 0-62	T47L17	33	2/week	fix	1981-2013	6/month	33		6 TB
CMA (babj)	d 0-60	T106L40	4	daily	fix	1994-2014	daily	4		
CNR-ISAC (isac)	d 0-31	0.75x0.56 L54	41	weekly	fix	1981-2010	every 5 days	1		
CNRM (lfpw)	d 0-32	T255L91	51	weekly	fix	1993-2014	2/month	15		6.75 GB/start date
ECCC (cwao)	d 0-32	0.45x0.45 L40	21	weekly	on the fly	1995-2014	weekly	4		
ECMWF (ecmf)	d 0-46	Tco639/319 L91	51	2/week	on the fly	past 20 years	2/week	11		
HMCR (rums)	d 0-61	1.1x1.4 L28	20	weekly	on the fly	1985-2010	weekly	10		
JMA (rjtd)	d 0-33	T319L60	25	2/week	fix	1981-2010	3/month	5	3.8 GB	900 GB
KMA (rksl)	d 0-60	N216L85	4	daily	on the fly	1991-2010	4/month	3		
NCEP (kwbc)	d 0-44	T126L64	16	daily	fix	1999-2010	day	4		
UKMO (egrr)	d 0-60	N216L85	4	daily	on the fly	1993-2015	4/month	7		

Retrieval efficiency

- In order to retrieve data efficiently users should follow the instructions about S2S reforecasts retrieval efficiency.
- Find links here to the efficient retrieval scripts for different S2S datasets via ECMWF Web API

Model description

Implementation Date in S2S	Model version	Time range	Atmosphere Resolution	Ens. Size	Frequency	Re-forecasts	Rfc length	Time range Atmosphere Resolution	Rfc size	Ocean Resolution	Active Sea Ice
01/01/2015	CY40R1	d 0-32		51	2/week Mon/Thu	on the fly	past 20 years	once a week (Thu real-time)	5	1 degree	No
14/05/2015	CY41R1	d 0-46	32 km up to day 10 64 km after day 10	51	2/week Mon/Thu	on the fly	past 20 years	2/week (Mon/Thu real time)	11	1 degree	No
08/03/2016	CY41R2	d 0-46	16 km up to day 15 31 km after day 15	51	2/week Mon/Thu	on the fly	past 20 years	2/week (Mon/Thu real time)	11	1 degree	No
22/11/2016	CY43R1	d 0-46	16 km up to day 15 31 km after day 15	51	2/week Mon/Thu	on the fly	past 20 years	2/week (Mon/Thu real time)	11	1/4 degree	Yes
11/07/2017	CY43R3	d 0-46	16 km up to day 15 31 km after day 15	51	2/week Mon/Thu	on the fly	past 20 years	2/week (Mon/Thu real time)	11	1/4 degree	Yes
06/06/2018	CY45R1	d 0-46	16 km up to day 15 31 km after day 15	51	2/week Mon/Thu	on the fly	past 20 years	2/week (Mon/Thu real time)	11	1/4 degree	Yes
11/06/2019	CY46R1	d 0-46	16 km up to day 15 31 km after day 15	51	2/week Mon/Thu	on the fly	past 20 years	2/week (Mon/Thu real time)	11	1/4 degree	Yes

A few important links:

S2S project webpage: s2sprediction.net

S2S data portal + documentation at ECMWF: s2s.ecmwf.int

S2S data portal at CMA: s2s.cma.cn

Data Issues

Pages / Home / Resources

Issues with data

Created by Matthew Manoussakis, last modified by Manuel Fuentes on Aug 09, 2016

Meteo-France (lfpw)

Interpolation error in ALL data between 19 May and 16 June 2016

1. Brief description:

All S2S parameters were affected in that period. The corrected data was re-archived **on June 17 2016**.

The problem on the provider's side was caused by a bug in the last versions of EMOS and MARS (issued in March) that wa

2. Recommendation:
If lfpw data from the given period was downloaded in the period 19.5.2016-17.6.2016. it should be deleted and the correct v

BoM (ammc)

Wrong maximum values of surface air maximum temperature (Tmax) at some points for BoM data

1. Brief description:

Some **coastal grid points** may display **unrealistic** Tmax values. All of the spurious Tmax points are **flagged** as being **oc**

2. Recommendation:

To **eliminate** the spurious values, use Tmax over **land points only,** based on POAMA's land-sea mask .

** POAMA stands for Predictive Ocean Atmosphere Model for Australia.*

List of parameters provided

https://software.ecmwf.int/wiki/display/S2S/Provided+parameters

	ECMWF	BoM	CNRM	NCEP	CMA	JMA	HMCR	ISAC_CNR	KMA	ECCC	UKMO
Geopotential height							no 10hPa				
Temperature							no 10hPa				
U velocity							no 10hPa				
V velocity							no 10hPa				
Specific Humidity							no 10hPa				
Vertical Velocity											
Potential Vorticity											
10m U											
10m V											
CAPE											
Skin temperature											
Snow depth											
Snow density											
Snow fall				No re-fcst							
Snow albedo											
Soil moisture top 20cm											
Soil moisture top 100 cm											
Soil temperature top 20 cm											
Soil Temperature top 100cm											
Surface air max temp											
Surface air min Temp											
Surface air Temp											
Surface air dewpoint temp											
SST											

STEP 1: REGISTER

Register from ECMWF data portal

- First name/last name/email address
- Accept term and conditions of use of the database

Two ways to get S2S data

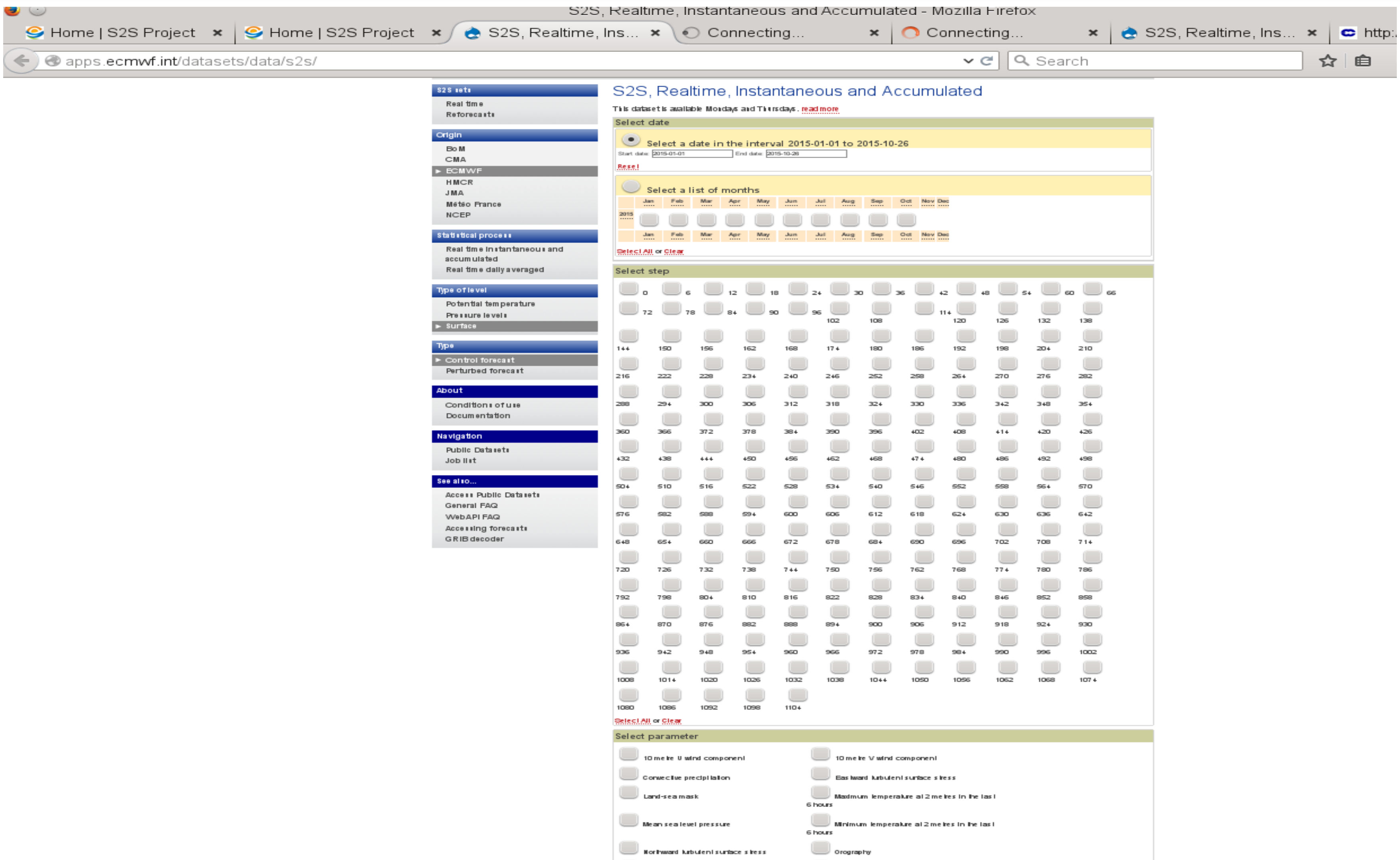
- **Web INTERFACE**: **http://apps.ecmwf.int/datasets/data/s2s/**

This is a "discovery" tool. Recommended for first time users. It gives a good idea of the content of the database, its structure and most importantly what is available. Easy to use. Good for small retrievals.

- **WEBAPI:** **https://software.ecmwf.int/wiki/display/WEBAPI/WebAPI+FAQ**

This is a more advanced tool for data retrieval. Users install a "webapi key" on their computer. This allows them to run scripts to perform intensive S2S data retrievals. Recommended for advanced users with intensive data retrievals. Retrievals can be optimized.

WEB INTERFACE: http://apps.ecmwf.int/datasets/data/s2s



Getting DATA from scripts: WEBAPI

WEBAPI is an application programming interface (API) for a web server. It allows to download S2S data directly from you computer using PYTHON scripts.

- You may start with the examples available on S2S sample scripts or by creating your request using the S2S Datasets Web Interface .
- Please note the following:
- We strongly advice you to start with a simple request. (1-2 parameters 1 time step 1-2 steps etc)
- The PYTHON request will be a dictionary with "keys" and "values" that represent your selection. (eg "step":"00", "time": 00")
- The request is strongly connected to the availability of the data

Example of WEBAPI SCRIPT

```
#!/usr/bin/env python
from ecmwfapi import ECMWFDataServer
server = ECMWFDataServer()
server.retrieve({
    "class": "s2",
    "dataset": "s2s",
    "date": "2015-01-01/2015-01-05/2015-01-08/2015-01-12/2015-01-15/2015-01-19/2015-01-22/2015-01-26/2015-01-29",
    "expver": "prod",
    "levtype": "sfc",
    "origin": "ecmf",
    "param": "165",
    "step": "0/24/48/72/96/120/144/168/192/216/240/264/288/312/336/360/384/408/432/456/480/504/528/552/576/600/624/648/672
/696/720/744/768",
    "stream": "enfo",
    "target": "CHANGEME",
    "time": "00",
    "type": "cf",
})
```

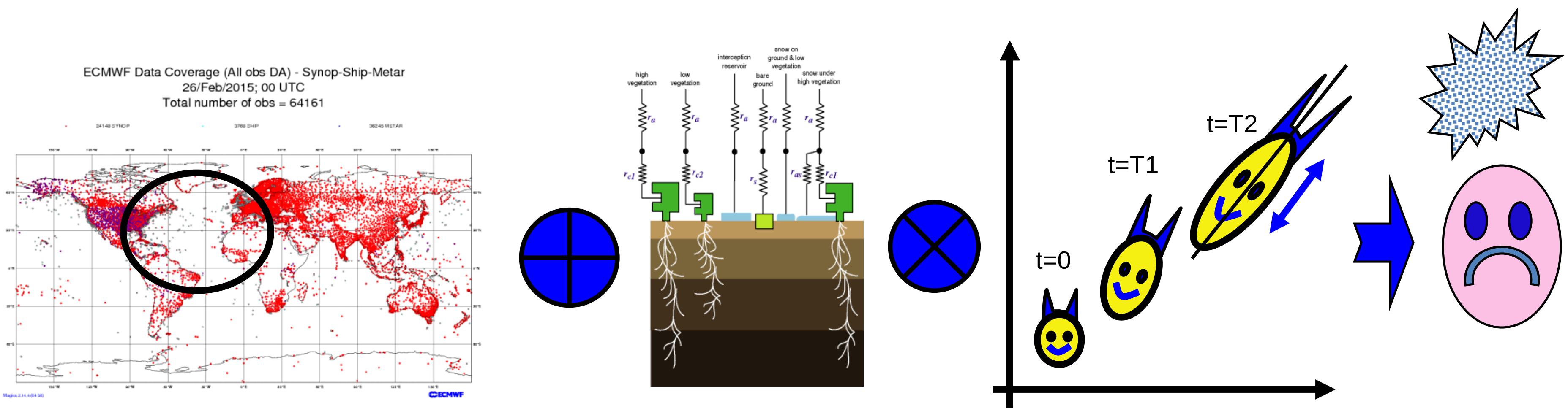
You can also add other commands: “grid”: “1.5/1.5”,
"area": "15/-180/-15/180",

Why ensemble prediction?

Why do forecasts fail?

Forecasts can fail because:

- The initial conditions are not accurate enough, e.g. due to poor coverage and/or observation errors, or errors in the assimilation (initial uncertainties).
- The model used to assimilate the data and to make the forecast describes only in an approximate way the true atmospheric phenomena (model uncertainties).



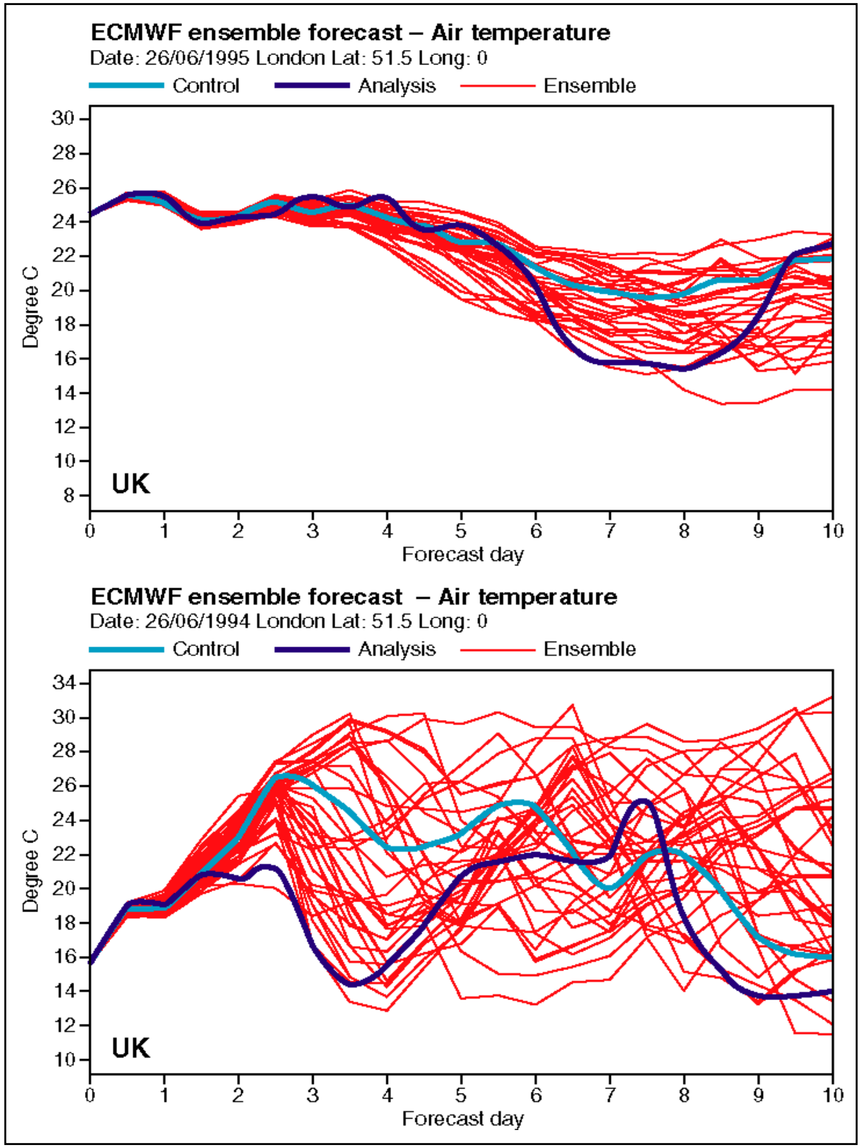
Chaos and weather prediction

The atmosphere is a chaotic system

- Small errors can grow to have major impact
- We can never perfectly measure the current state of the whole atmosphere

Ensemble Forecasts

- Parallel set of forecasts from slightly different initial conditions and model formulation
- Assess uncertainty of today's forecast



What is the aim of ensemble forecasting?

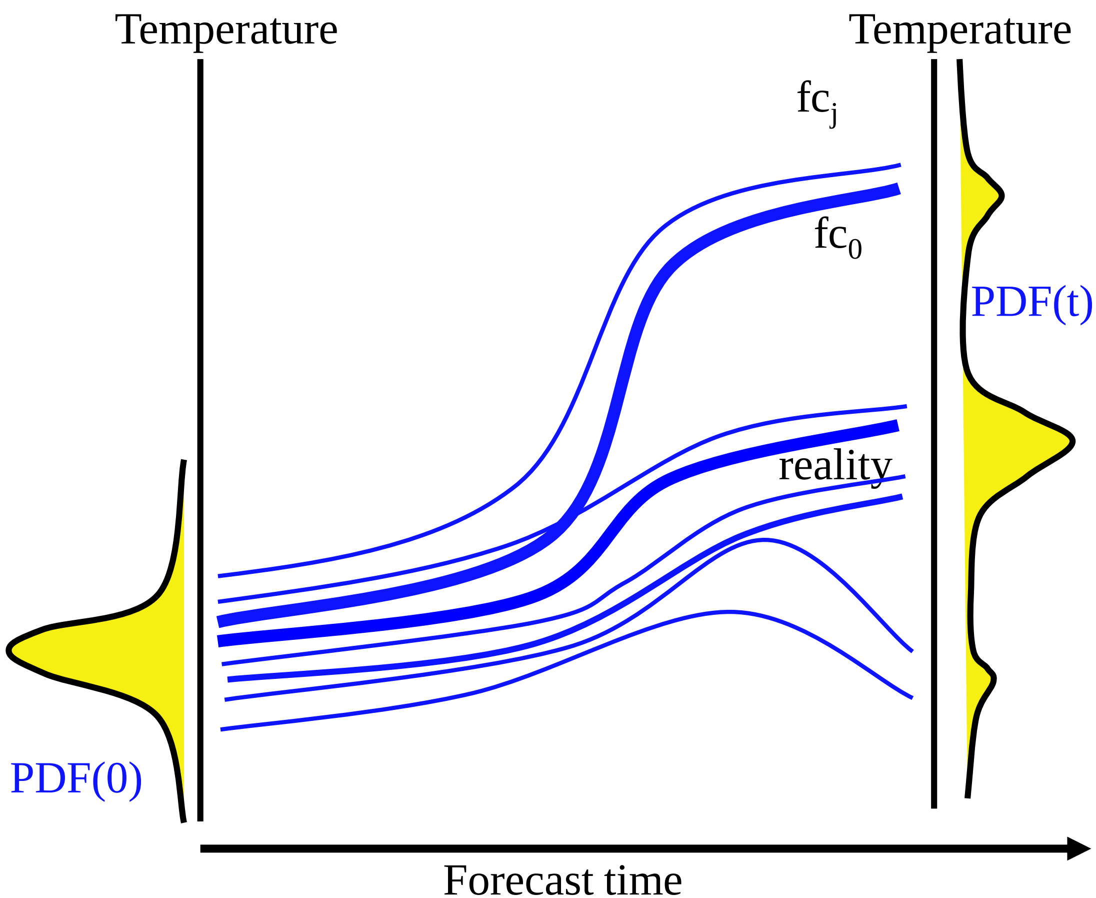
We have seen that single forecasts can fail due to a combination of **initial** and **model uncertainties**, and that the NWP problem is made extremely complex by the **chaotic nature of the atmosphere**.

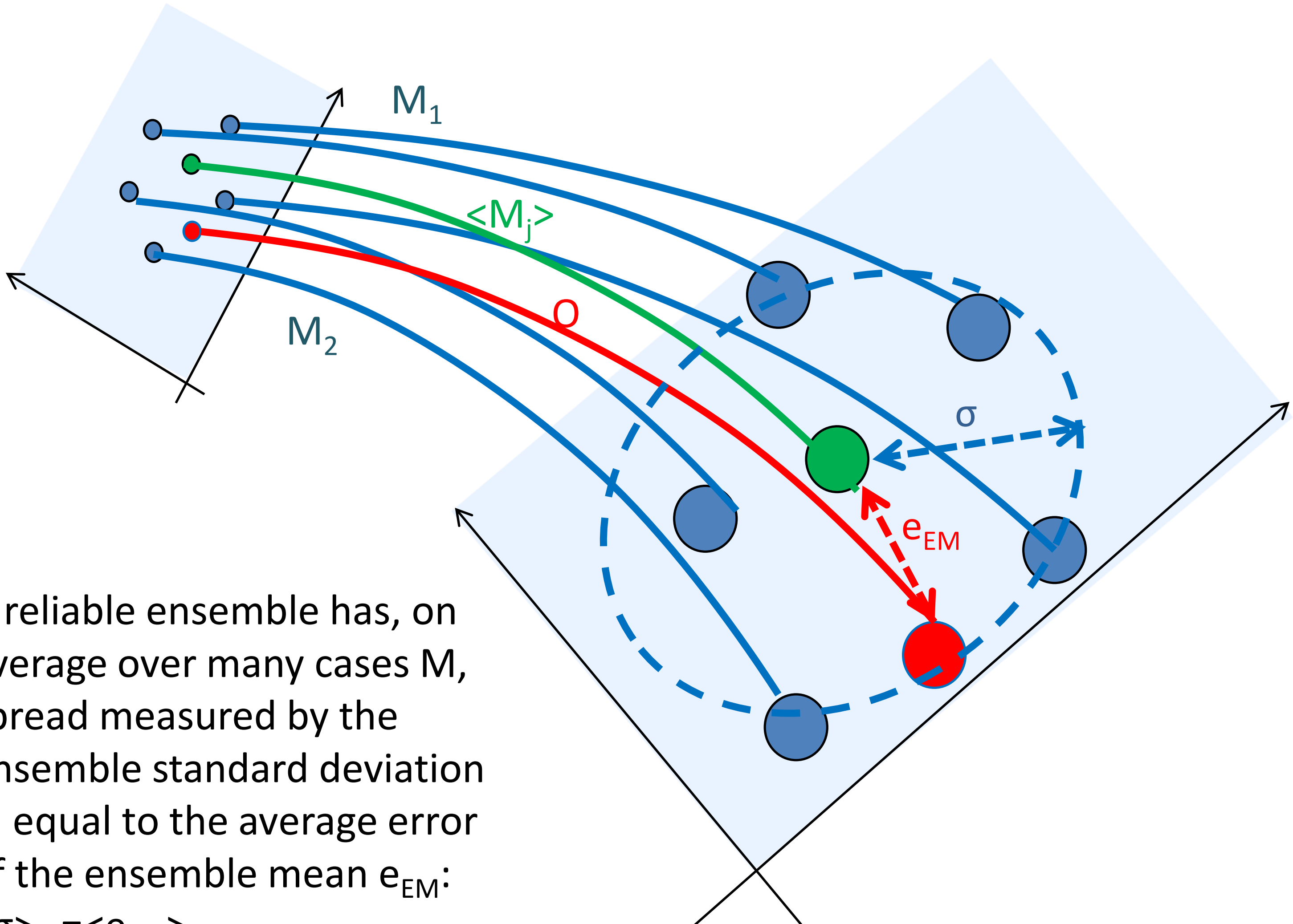
- Does it make sense to issue single forecasts?
- Can something better be done?
- More generally, what is the aim of weather and climate forecasting?
- Should it be to predict only the most likely scenario, or should it aim to predict also its uncertainty (give a 'confidence band'), for example expressed in terms of weather scenarii or probabilities that different weather conditions can occur?

Ensemble prediction

Ensemble prediction aims to estimate the probability density function of forecast states, taking into account all possible sources of forecast error:

- ❖ Observation errors and imperfect boundary conditions
- ❖ Data assimilation assumptions
- ❖ Model errors



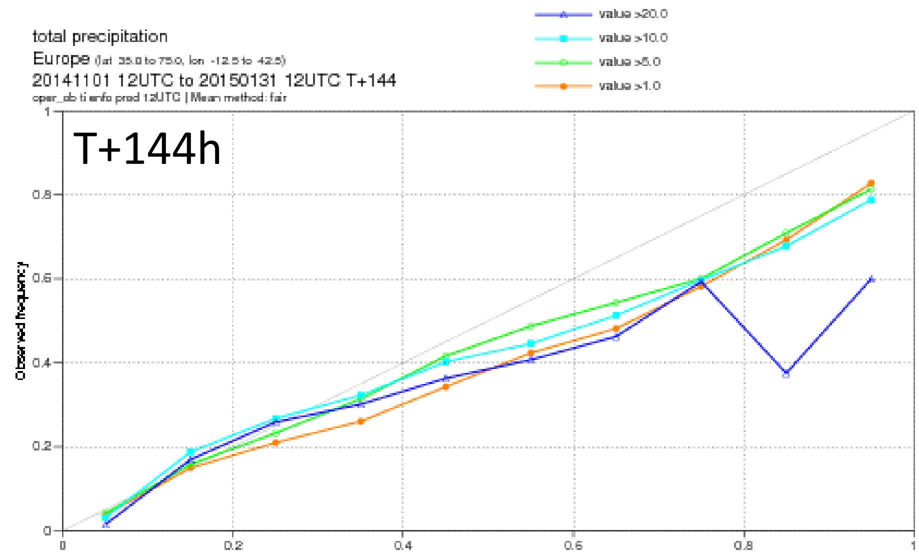


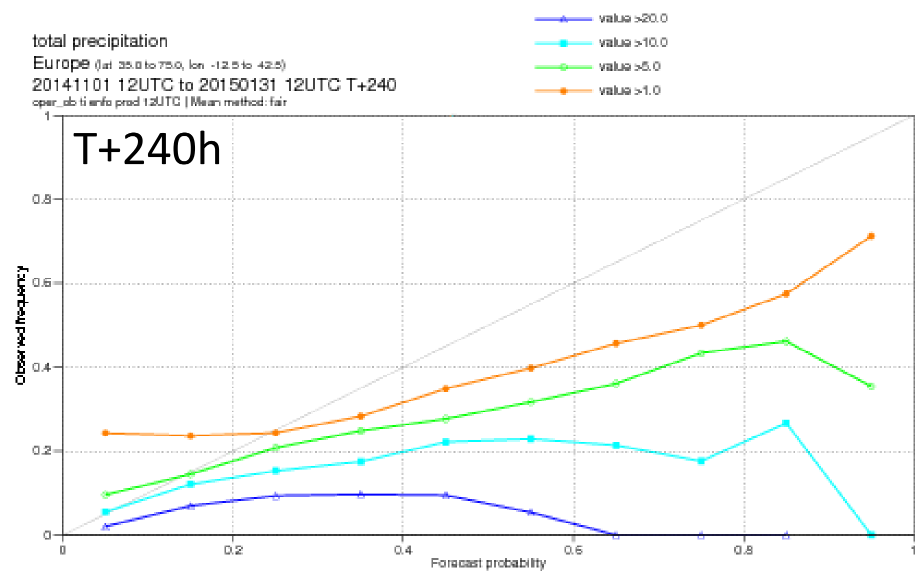
A reliable ensemble has, on average over many cases M, spread measured by the ensemble standard deviation σ, equal to the average error of the ensemble mean e_{EM}:
$\langle\sigma\rangle_M = \langle e_{EM}\rangle_M$

In a reliable ensemble, <fc-prob>~<obs-prob>

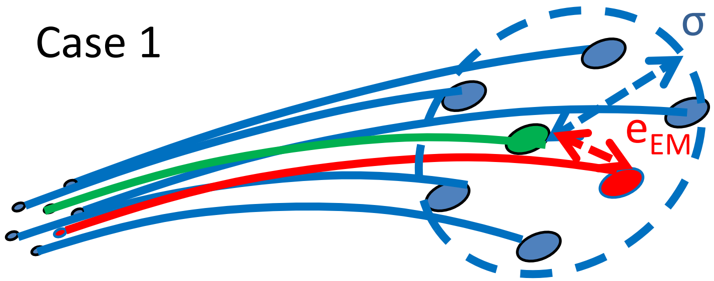
One way to check the ensemble reliability is to assess whether the average forecast and observed probabilities of a certain event are similar.

These plots compare the two probabilities at t+144h and t+240h for the event ‘24h precipitation in excess of 1/5/10/20 mm’ over Europe for ND14J15 (verified against observations).

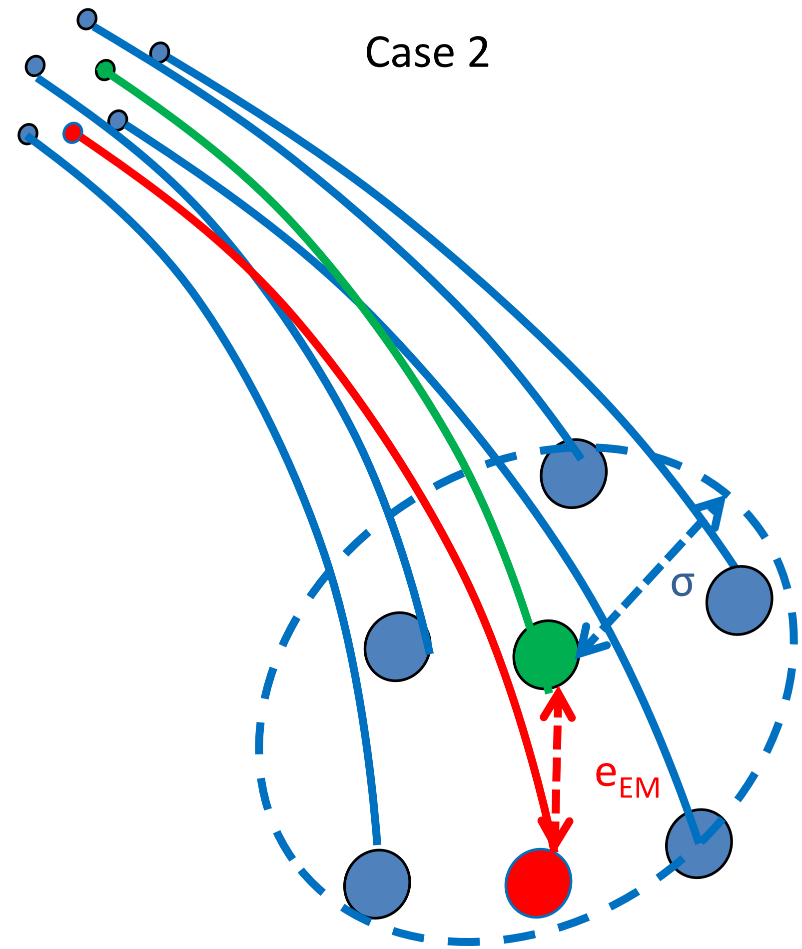




Case 1



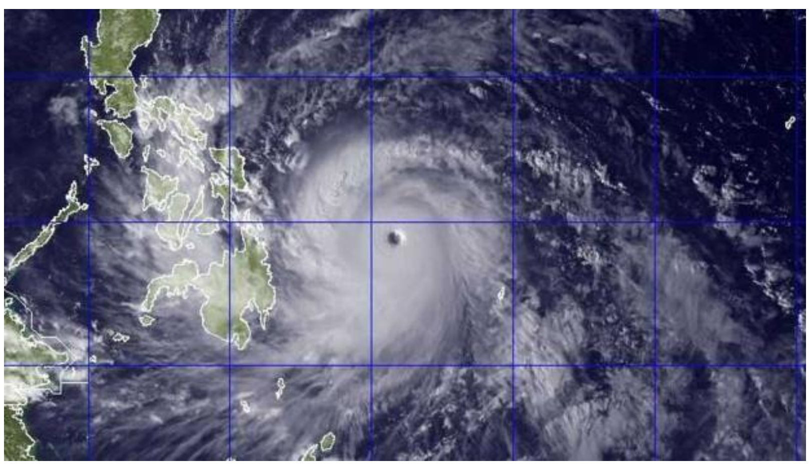
Case 2

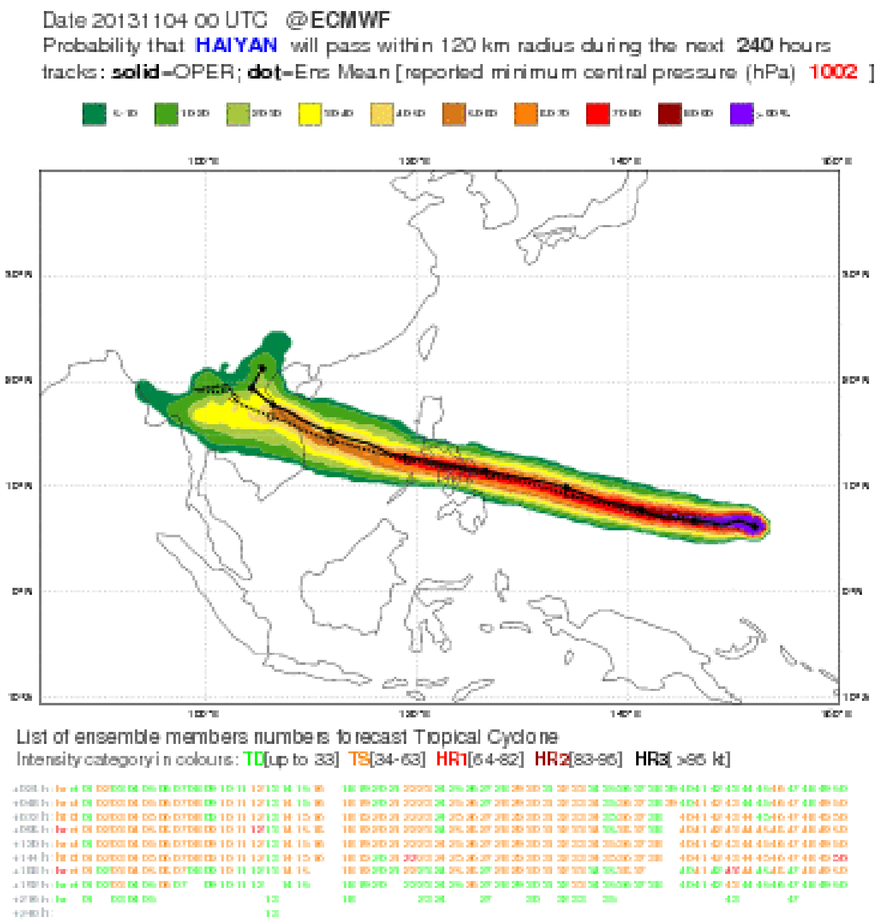


In a reliable ensemble, small ensemble standard deviation indicates a more predictable case, i.e. a small error of the ensemble mean e_{EM}.

Track dispersion & predictability: Haiyan (Nov 2013)

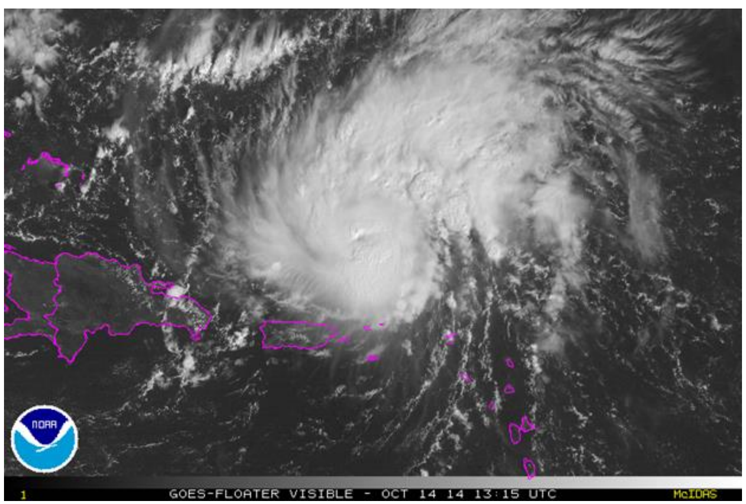
Haiyan (Nov 2013) - Dispersion of ENS tracks in the 10d forecast issued on 2014.10.13@12 was very small for the whole 10 day range, indicating high confidence on direction of travel.

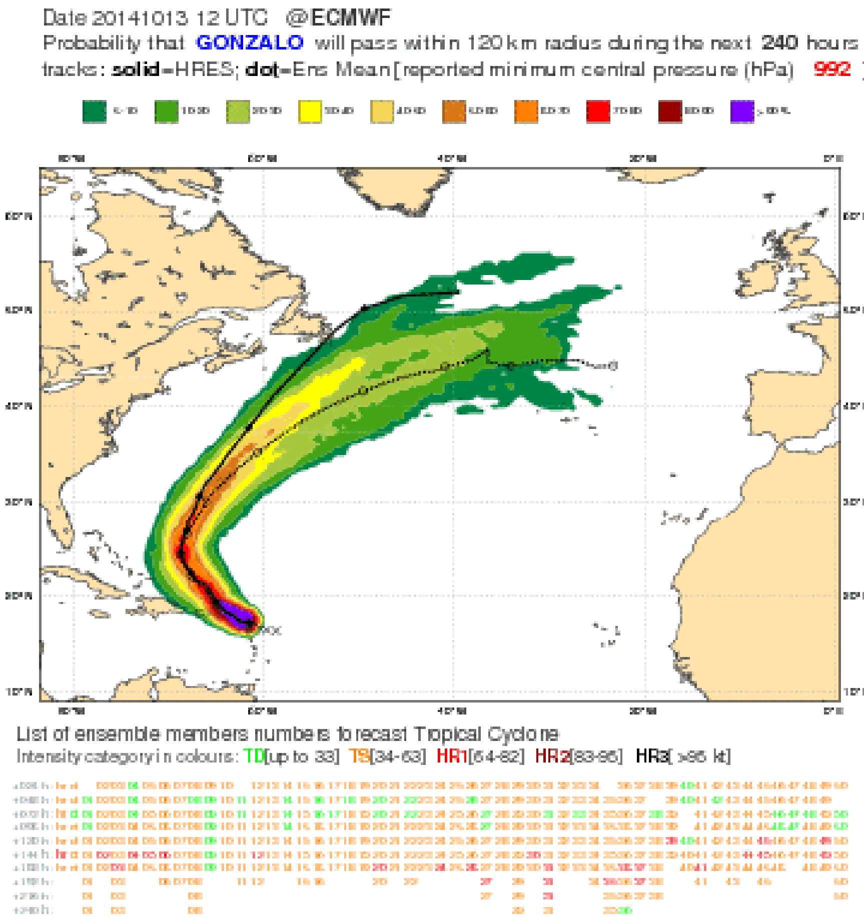




Track dispersion & predictability: Gonzalo (Oct 2014)

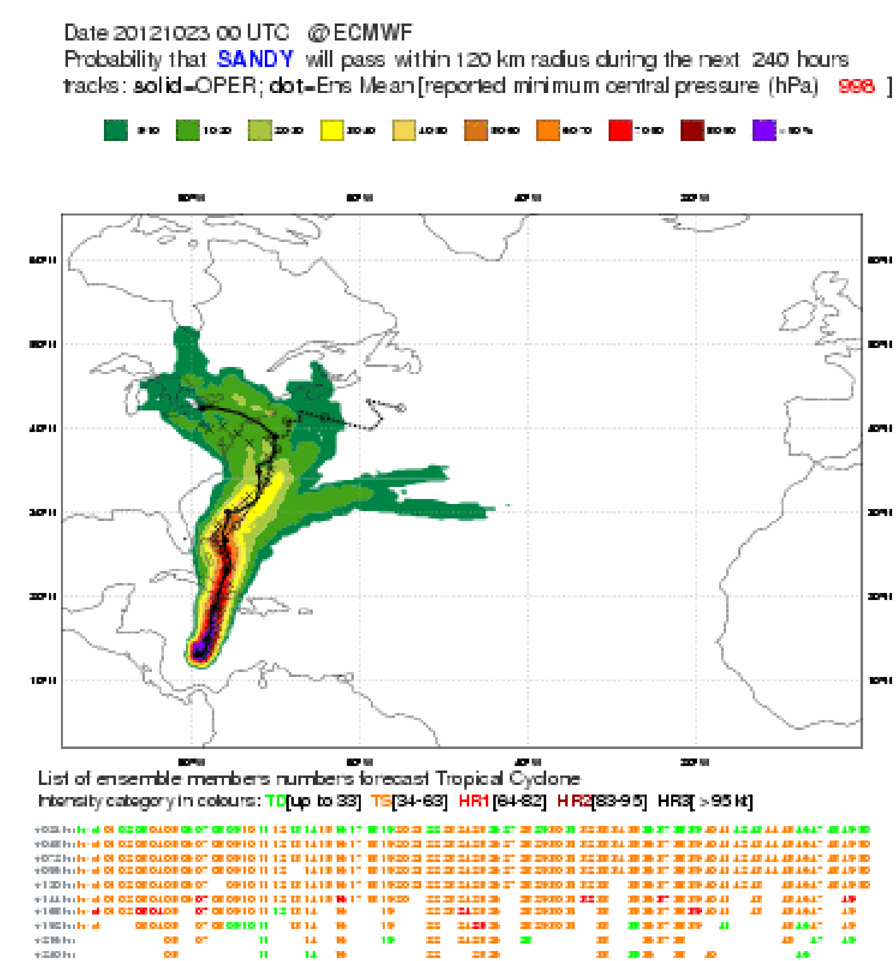
Gonzalo (Oct 2014) - Dispersion of ENS tracks in the 10d forecast issued on 2014.10.13@12 was relatively small for the whole 10 day range, indicating more confidence on   direction of travel.





Example: Track dispersion & predictability: Sandy (Oct 2012)

Sandy (Oct 2012) - Dispersion of ENS tracks in the 10d forecast issued on 2012.10.23@00 was relatively large after forecast day 5, indicating high uncertainty on direction and landfall location.

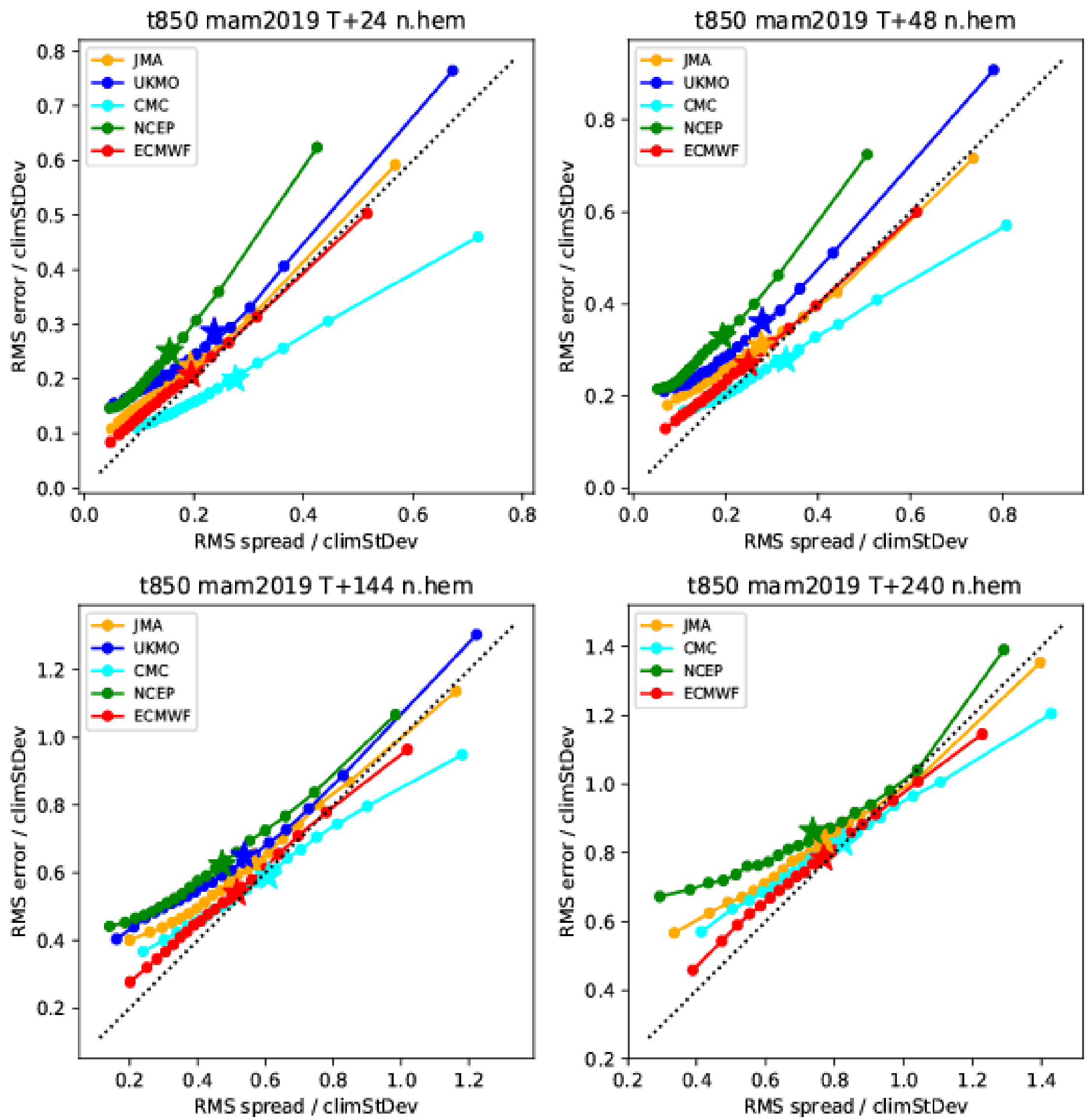


ENS spread as an index of predictability

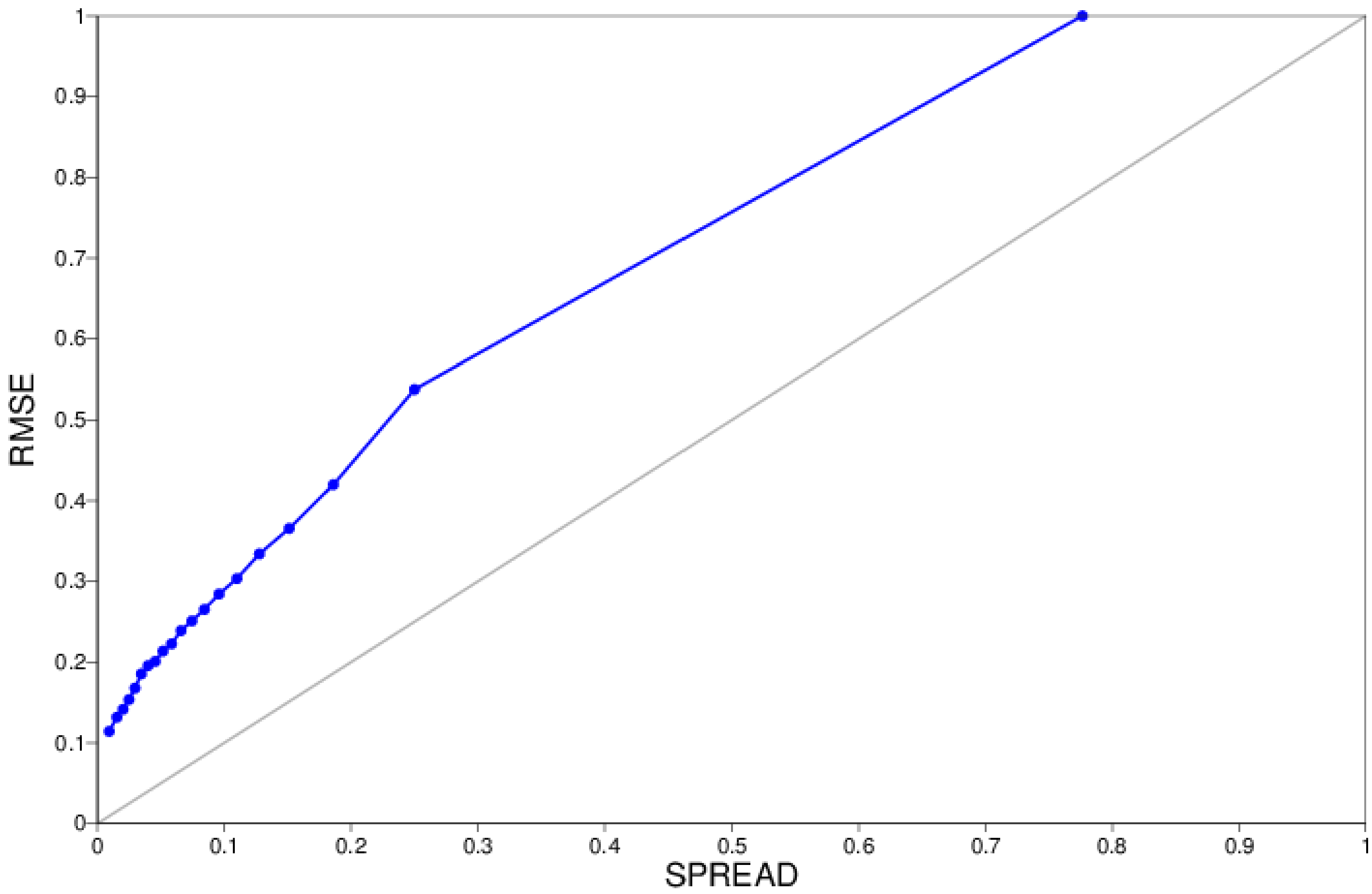
Small ensemble spread should identify predictable conditions

Generally true for short and medium range ensemble forecasts

Is it true for S2S time range?



SST Spread/skill relationship for SST forecasts at week 4



How do we produce ensembles?

World
Meteorological
Organization

What should an ensemble prediction simulate?

Two schools of thought:

- Monte Carlo approach: sample all sources of forecast error, perturb any input variable and any model parameter that is not perfectly known. Take into consideration as many sources as possible of forecast error.

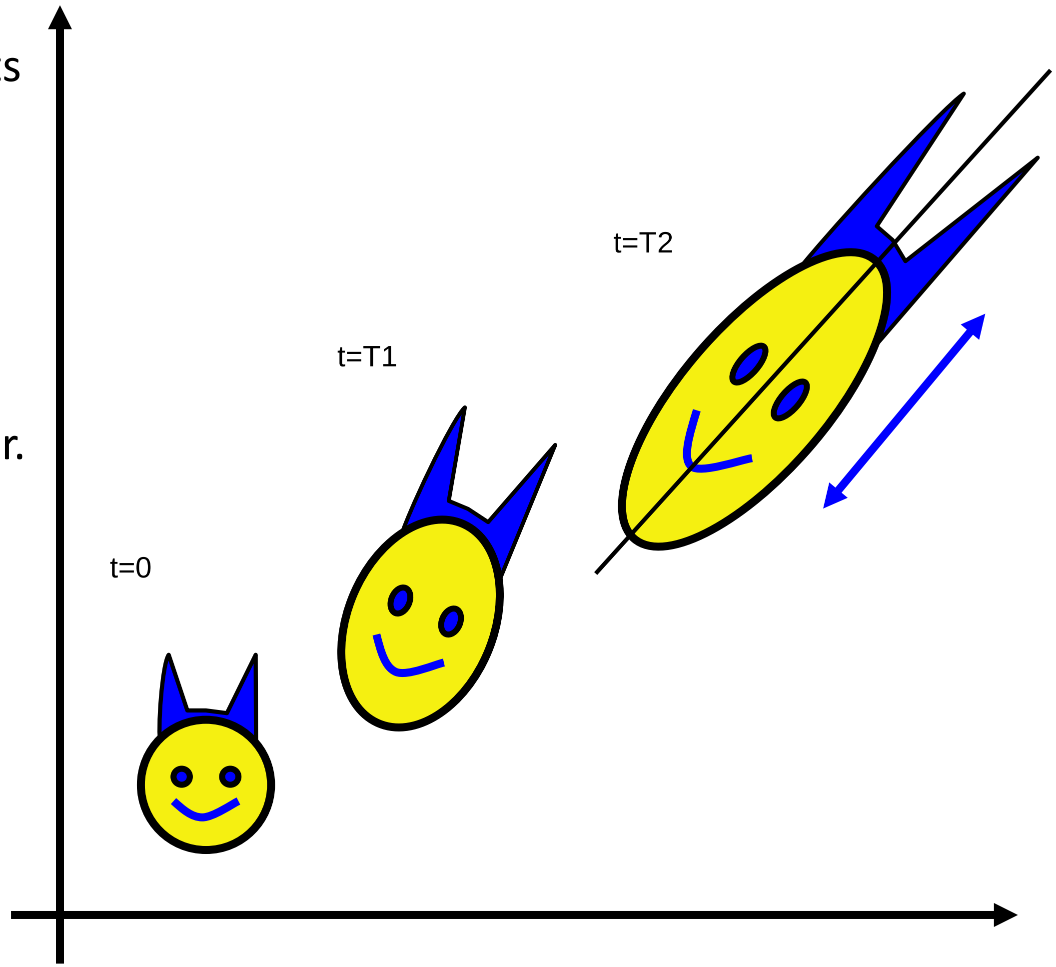
- Reduced sampling: sample leading sources of forecast error, prioritize. Rank sources, prioritize, optimize sampling: growing components will dominate forecast error growth.

There is a strong constraint: limited resources (man and computer power)!

How should initial uncertainties be defined?

The initial perturbations' components pointing along the directions of maximum growth amplify most.

If we knew the directions of maximum growth we could estimate the potential maximum forecast error.



Current methods for perturbing initial conditions

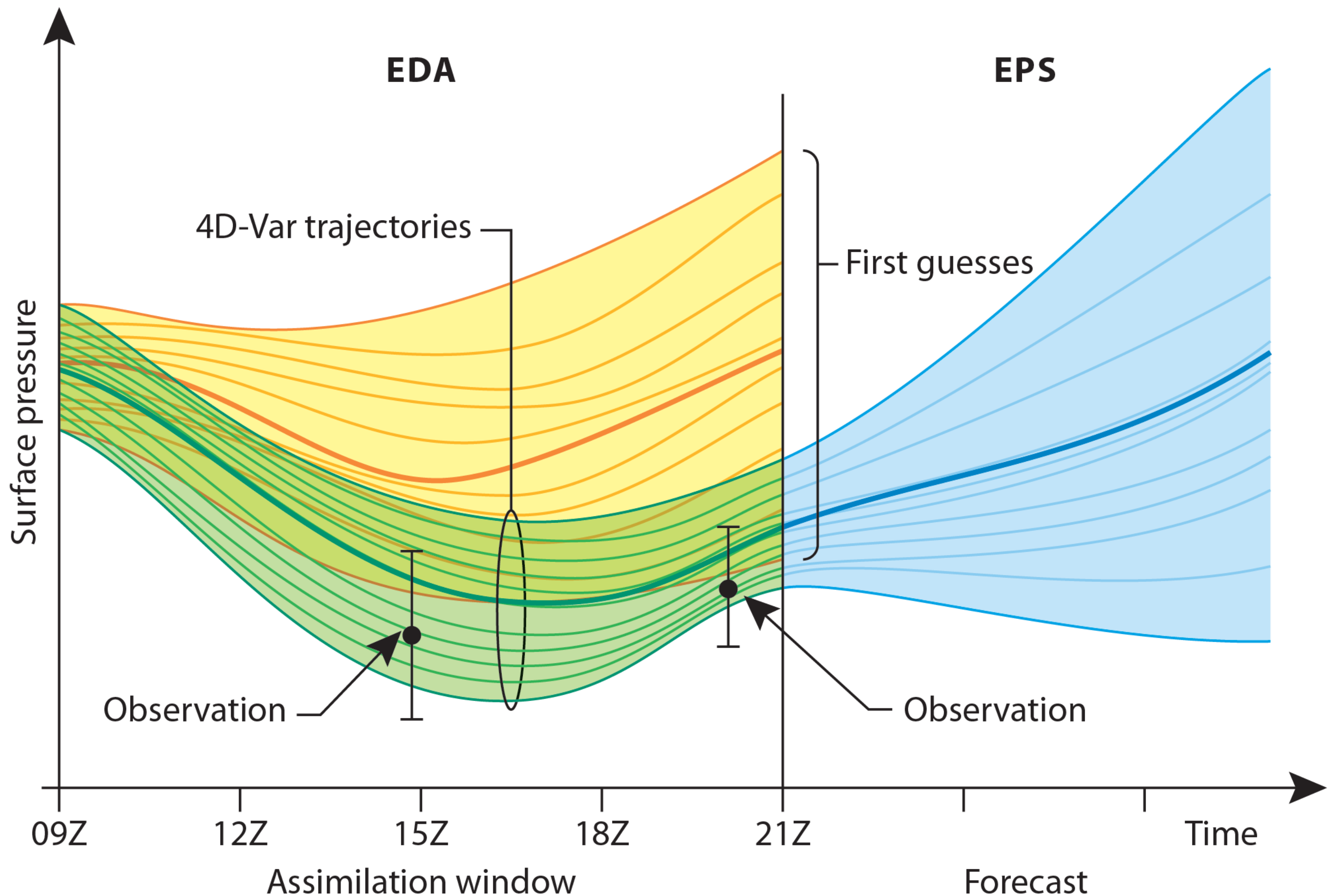
1. Lag ensemble: each ensemble member starts from a different analysis produced at a different time (e.g. 6 hour apart).

2. Singular vectors (e.g ECMWF) or breeding vectors (NCEP

3. Ensemble of data assimilations (slightly different analyses). E.g. 51 member EDA at ECMWF

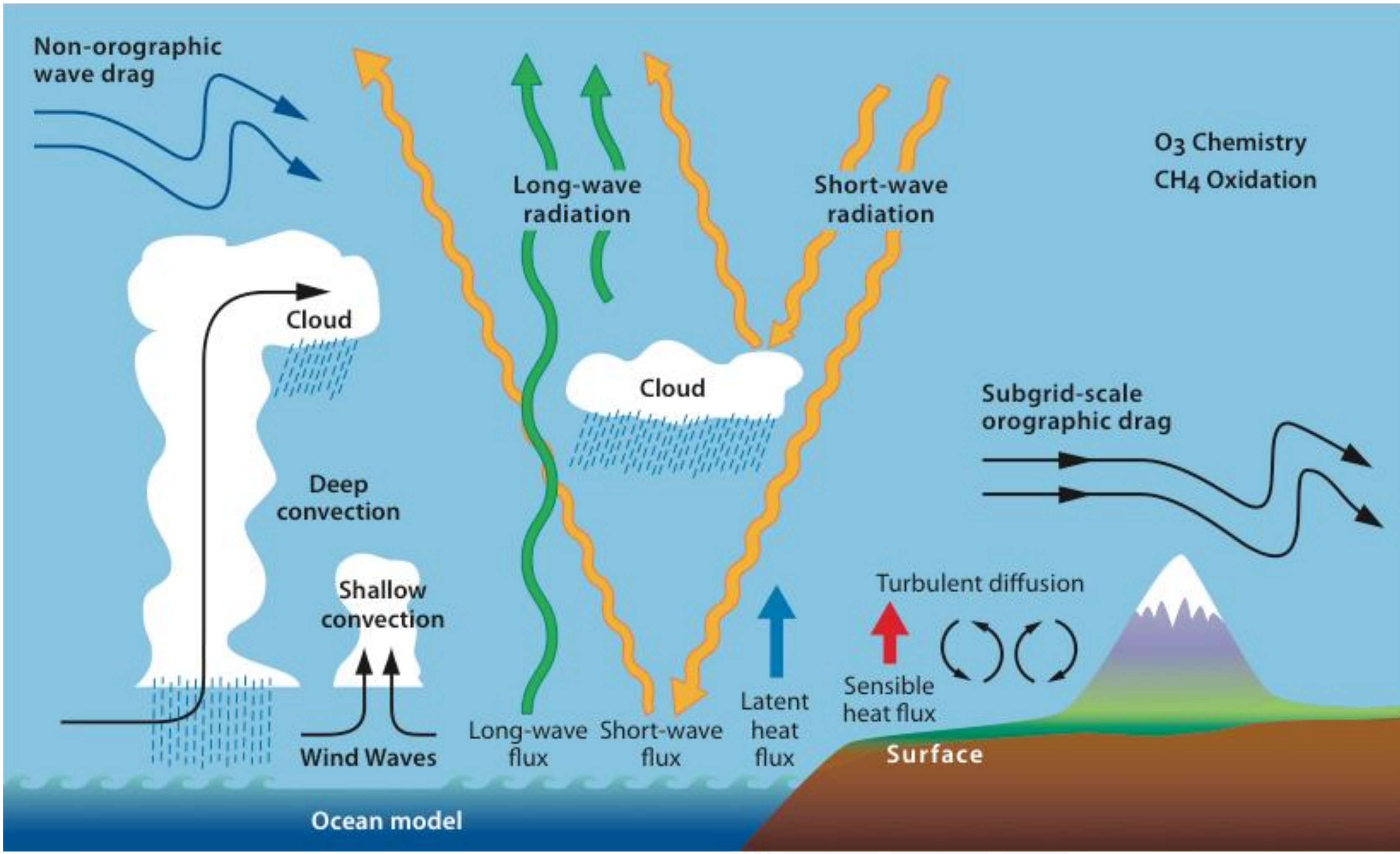
4. Combination of 1, 2 or 3. For example at ECMWF, combination of 1 and 3.

Ensemble assimilation and prediction



Model error: where does it come from?

Processes represented in the model:

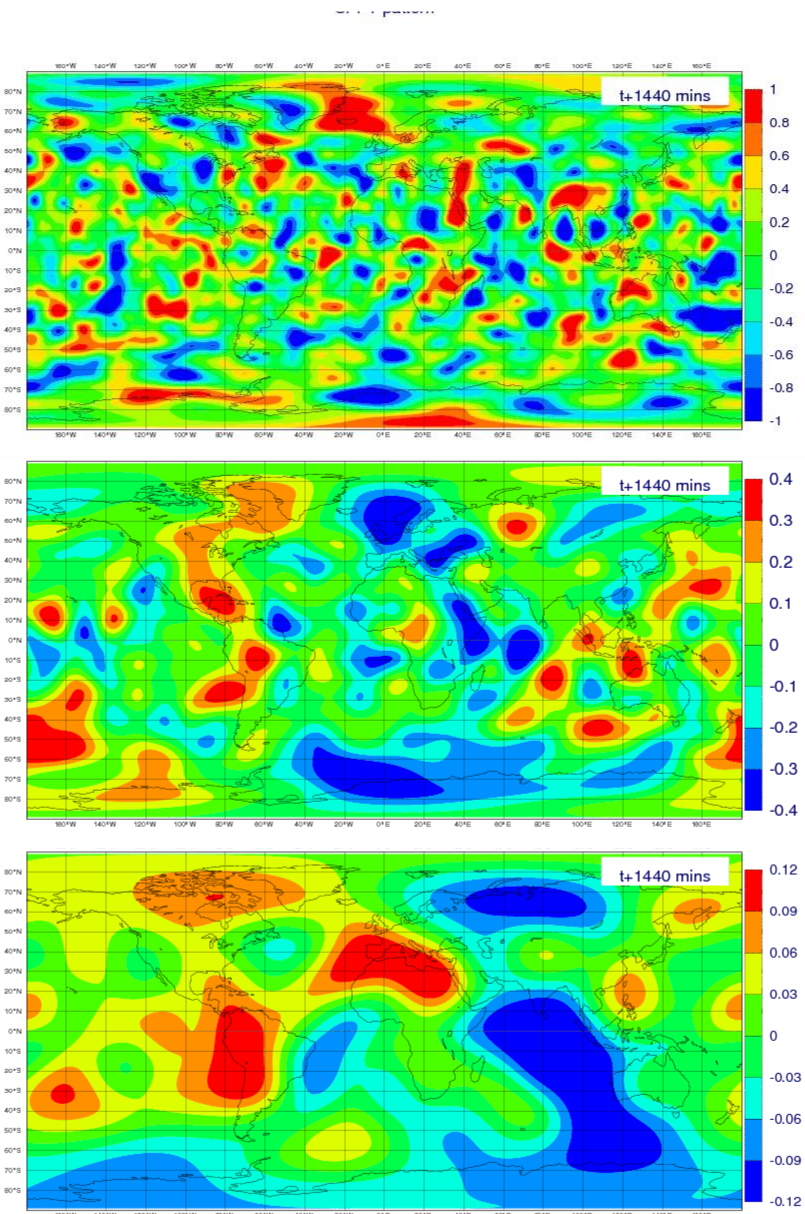


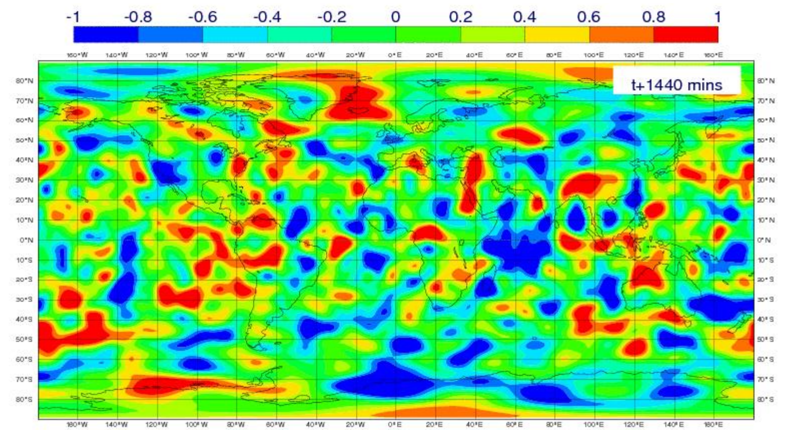
Model error: how to simulate it?

How can we represent model errors?

- Multi-model ensembles
- Multi-physics ensembles
- Perturbed parameter ensembles
- “Stochastic parameterizations”

SPPT pattern

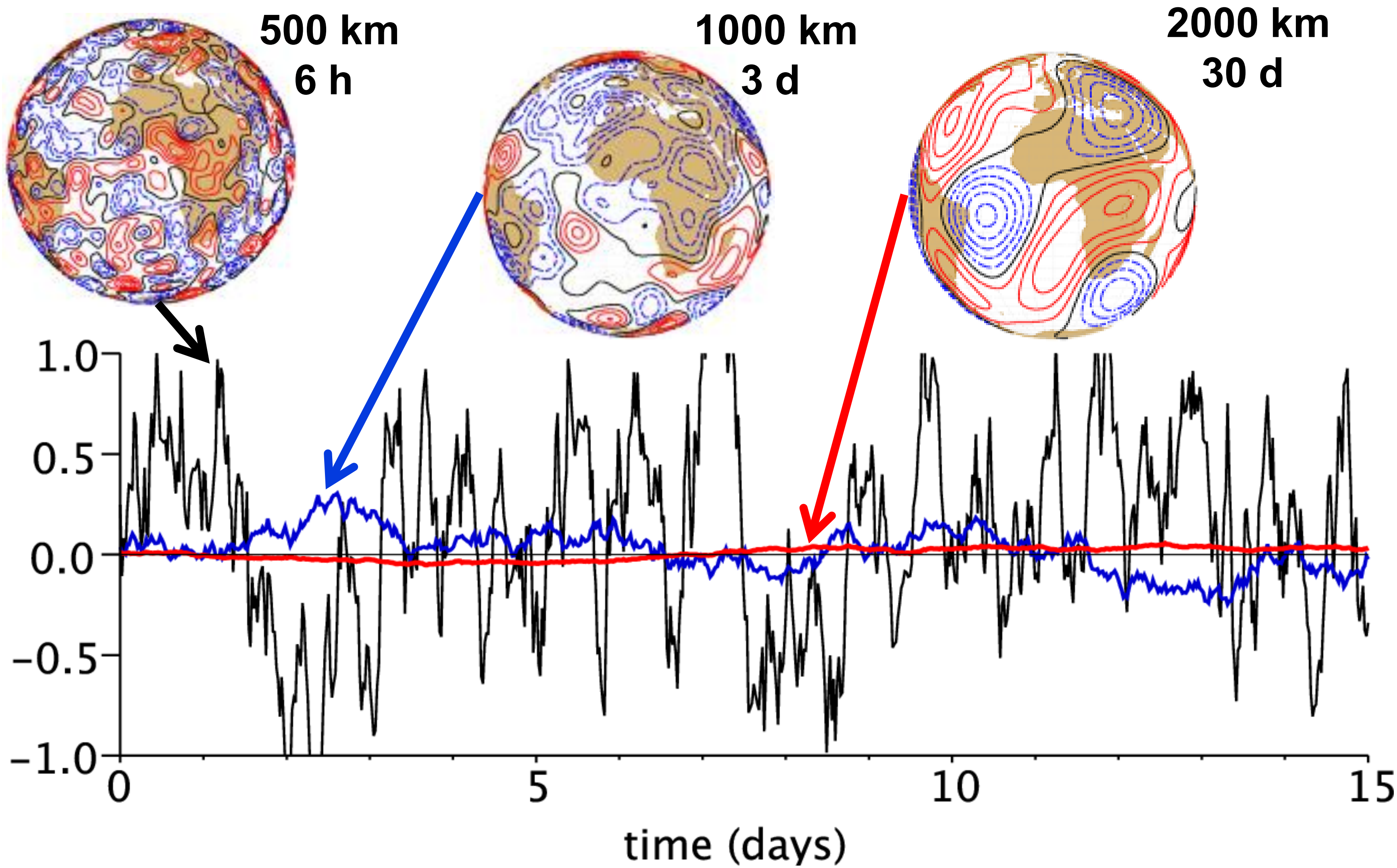




3 correlation scales:

i)	6 hours,	500 km,	$\sigma = 0.52$
ii)	3 days,	1 000 km,	$\sigma = 0.18$
iii)	30 days,	2 000 km,	$\sigma = 0.06$

SPPT pattern



Ensembles in the S2S database

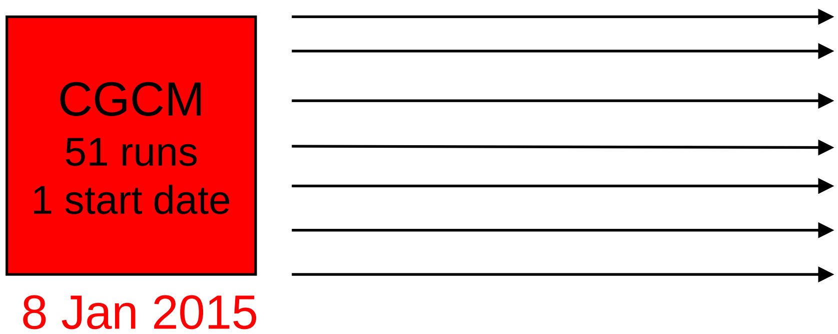
World
Meteorological
Organization

S2S Models

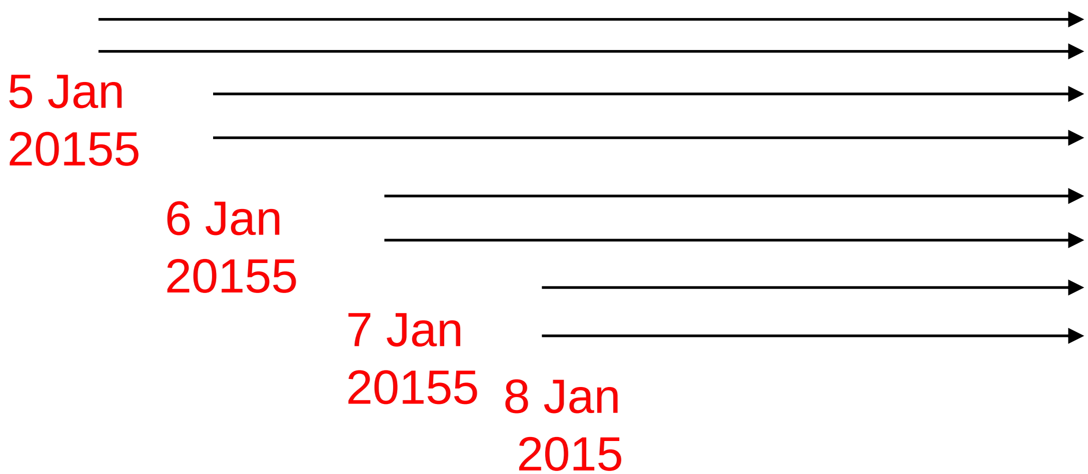
	Time-range	Resol.	Ens. Size	Freq.	Hcsts	Hcst length	Hcst Freq	Hcst Size
ECMWF	D 0-32	T639/319L91	51	2/week	On the fly	Past 20y	2/weekly	11
UKMO	D 0-60	N216L85	4	daily	On the fly	1989-2003	4/month	3
NCEP	D 0-44	N126L64	4	4/daily	Fix	1999-2010	4/daily	1
EC	D 0-35	0.6x0.6L40	21	weekly	On the fly	Past 15y	weekly	4
CAWCR	D 0-60	T47L17	33	weekly	Fix	1981-2013	6/month	33
JMA	D 0-34	T159L60	50	weekly	Fix	1979-2009	3/month	5
KMA	D 0-60	N216L85	4	daily	On the fly	1996-2009	4/month	3
CMA	D 0-45	T106L40	4	daily	Fix	1992-now	daily	4
Met.Fr	D 0-60	T127L31	51	monthly	Fix	1981-2005	monthly	11
CNR	D 0-32	0.75x0.56 L54	40	weekly	Fix	1981-2010	6/month	1
HMCR	D 0-63	1.1x1.4 L28	20	weekly	Fix	1981-2010	weekly	10

Burst ensemble vs lag approach

Burst approach: 1 start date, large ensemble size



Lag ensemble approach: multiple start date, small ensemble size



Burst ensemble vs lag approach

Burst approach:

Advantage: Uses Freshest initial conditions
More control on the ensemble generation

Disadvantage: Too costly to run daily
“flip-flop” forecasts

Lag approach:

Advantage: Forecasts can be updates every day
Smooth evolution of the forecasts

Disadvantage: less skilful because it uses “old” initial conditions

Conclusions

- Ensemble generation includes perturbations in the initial conditions + perturbations in the model physics.

- Various strategies for ensemble generation: burst vs lag ensemble. Not clear which one is optimal for S2S.