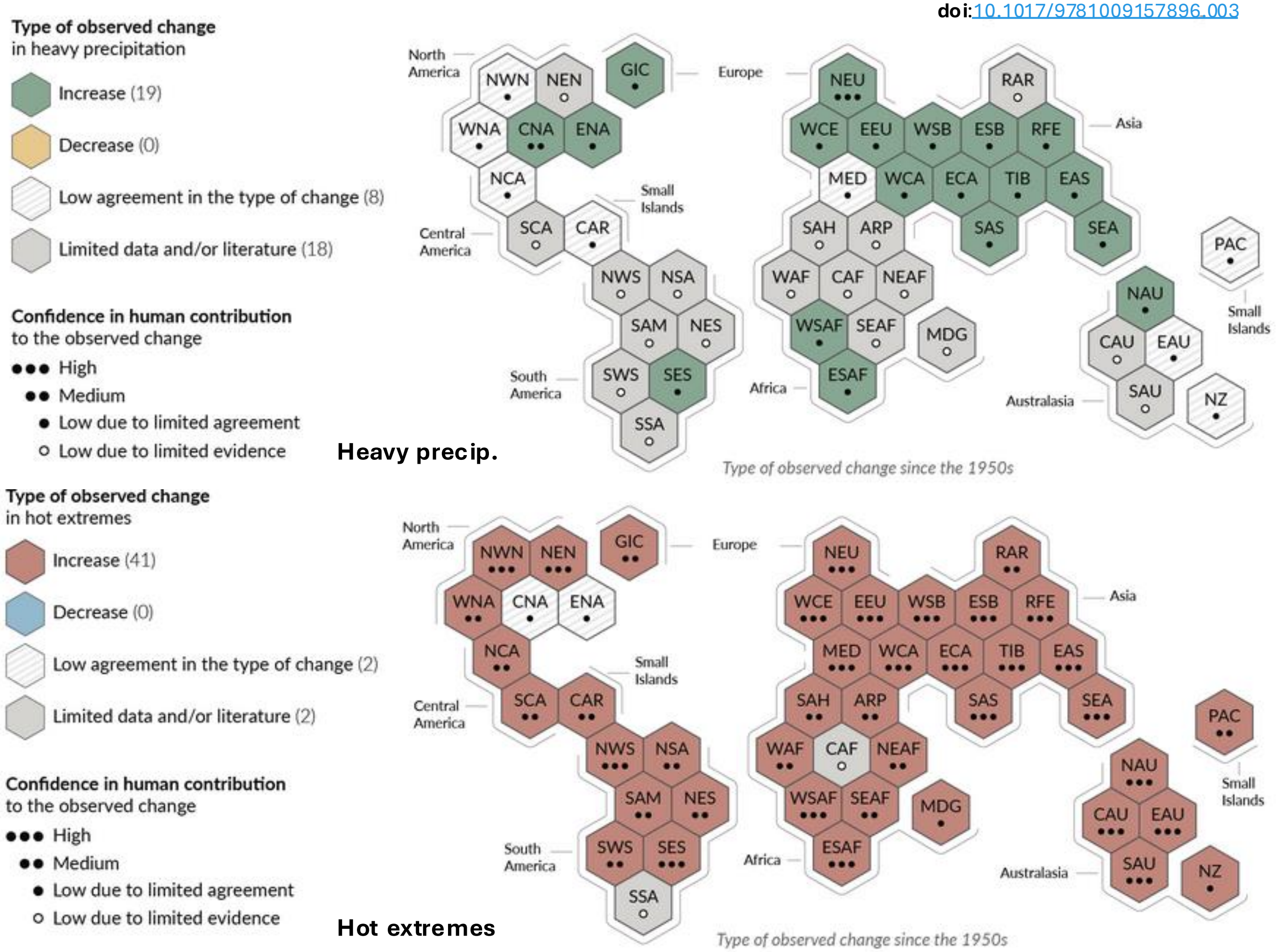
Session 04
Ensembles

doi:10.1017/9781009157896.003



Proyectos en ingeniería de datos e inteligencia artificial

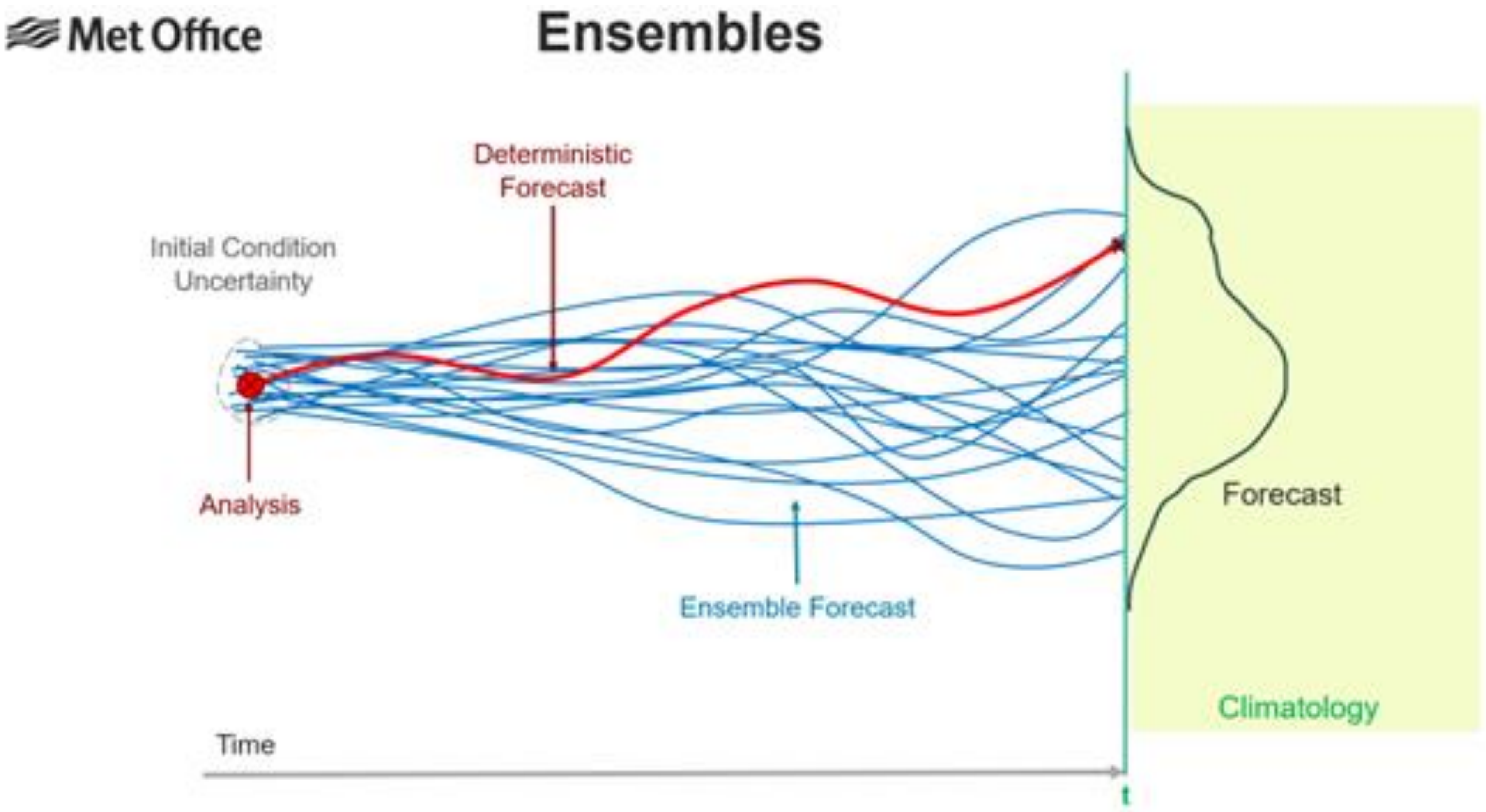
What you will learn

In this session, you will explore the concept of ensemble and learn how to leverage it using precipitation data from climate models.

The problem

A farmer needs to understand the range of possible weather conditions their crops may face in order to take appropriate protective measures.

The solution



Key concept: Ensemble

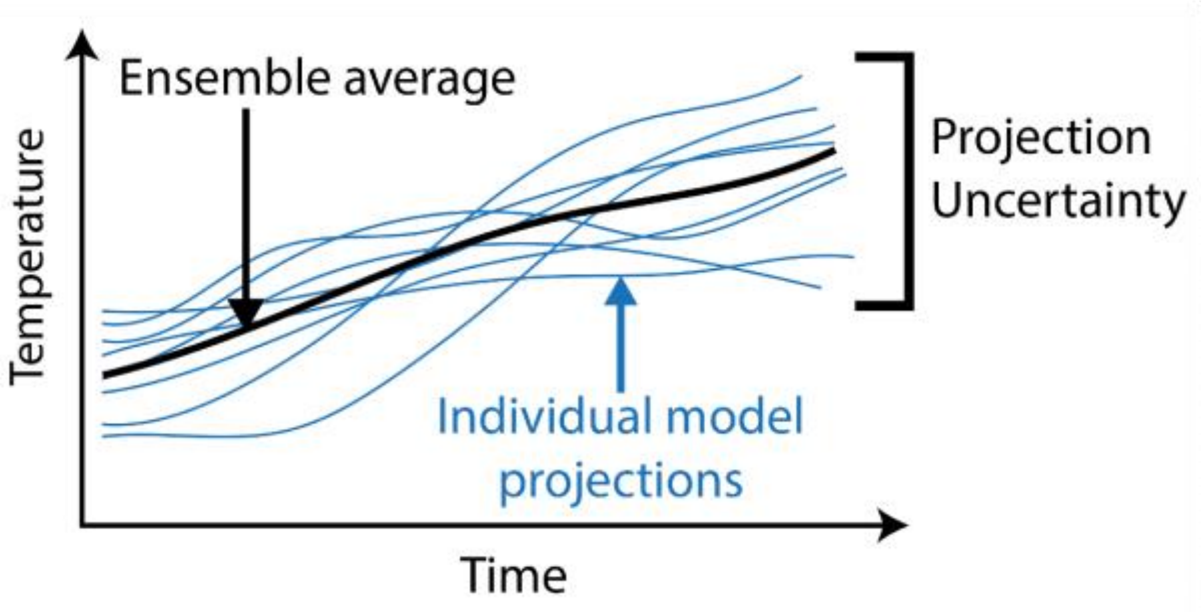
All the parts of a thing taken together, so that each part is considered only in relation to the whole

Ensemble [climatology]

A climate ensemble uses multiple climate models to simulate future climate, providing a range of possible outcomes and helping to estimate uncertainty.

Ensemble

The models within an ensemble, though comparable, are different and therefore produce a range of projections.



**We will see how to create an ensemble in the following sessions…*

Ensemble

Assumption. The ensemble average is expected to perform better than individual model runs.

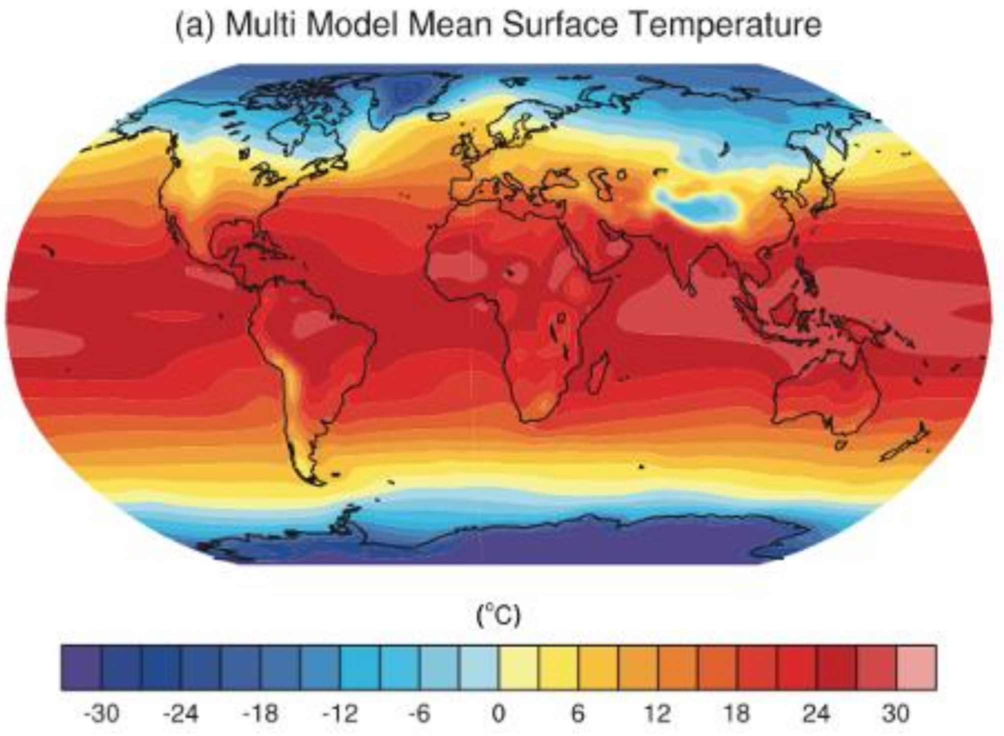
10.1175/2008MWR2381.1

Why is important to measure the level of uncertainty?

- Single deterministic forecasts can be misleading, as they do not convey the range of possible scenarios.

- Single deterministic forecasts can be misleading, as they do not convey the range of possible scenarios.
- It helps decision-makers—like farmers—prepare for a **range of plausible outcomes**, not just the most likely one.

Types: multi-model ensemble



Types: multi-model ensemble

How do we combine them into one single projection of the future?

- One model, one vote
  - All models are not equally good

Knutti, R. The end of model democracy?. Climatic Change 102, 395–404 (2010). https://doi.org/10.1007/s10584-010-9800-2

Climatic Change (2010) 102:395–404
DOI 10.1007/s10584-010-9800-2

EDITORIAL

The end of model democracy?

An editorial comment

Reto Knutti

Received: 24 November 2009 / Accepted: 4 January 2010 / Published online: 13 January 2010

1 The trillion dollar garden party—an analogy

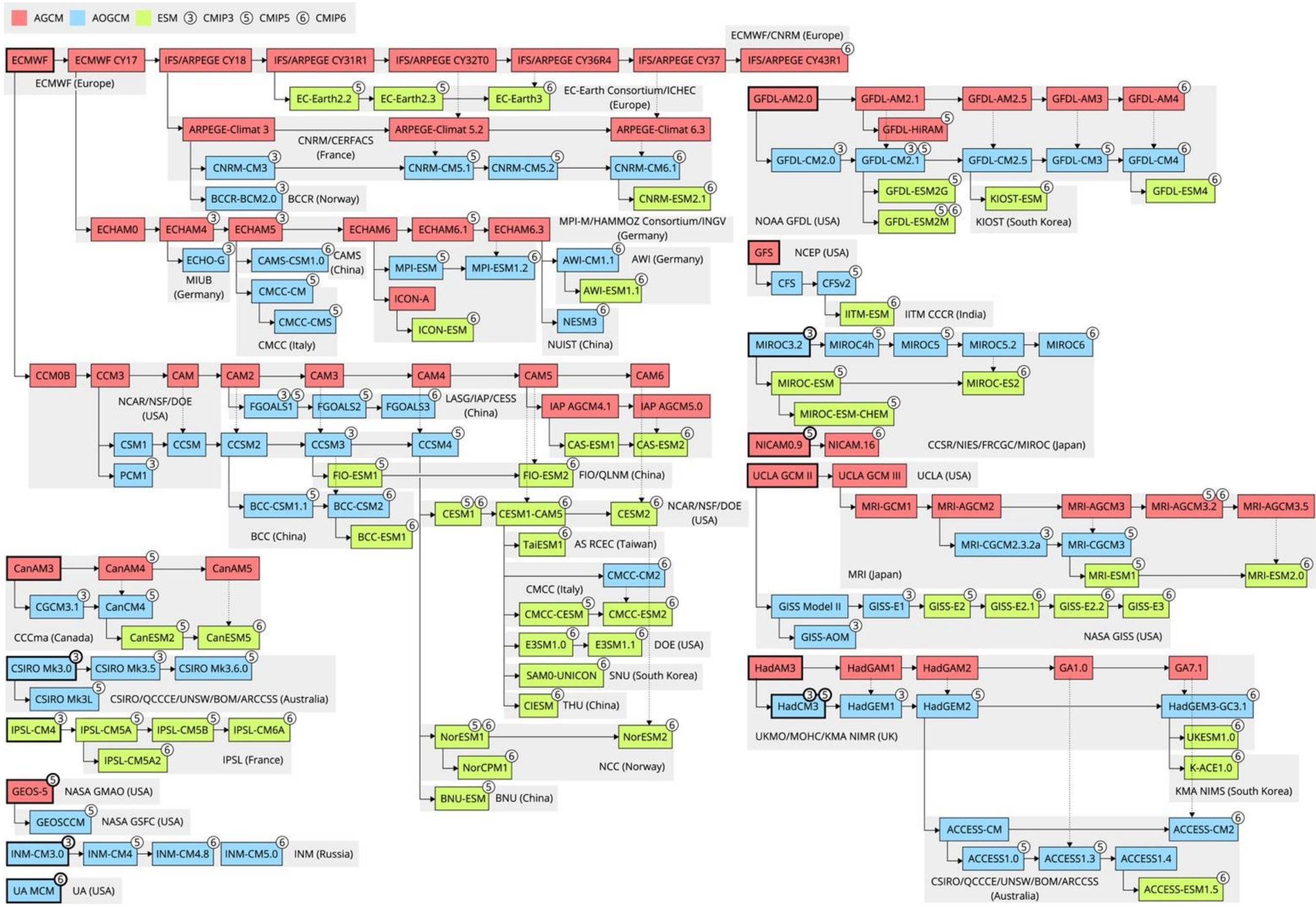
Imagine you are hosting a garden party tomorrow and you are trying to decide whether or not to put up a tent against the rain. You read the weather forecast in the newspaper and you ask the farmer next door, and you look at the sky (knowing that persistence is often not a bad weather forecast). So you get three predictions, but how would you aggregate them? Would you average them with equal weight? You might trust the forecast model more (or less) than the farmer, not because you understand how either of them generates their prediction, but because of your past experience in similar situations. But why seek advice from more than one source in the first place? We intuitively assume that the combined information from multiple sources improves our understanding and therefore our ability to decide. Now having read one newspaper forecast already, would a second and a third one increase your confidence? That seems unlikely, because you know that all newspaper forecasts are based on one of only a few numerical weather prediction models. Now once you have decided on a set of forecasts, and irrespective of whether they agree or not, you will have to synthesize the different pieces of information and decide about the tent for the party. The optimal decision probably involves more than just the most likely prediction. If the damage without the tent is likely to be large, and if putting up the tent is easy, then you might go for the tent in a case of large prediction uncertainty even if the most likely outcome is no rain.

Although it may seem far-fetched at first, the problem of climate projection is in fact similar in many respects to the garden party situation discussed above. So far, projections from multiple climate models were often aggregated into simple averages, standard deviations and ranges. One example is the recent Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC),

Types: multi-model ensemble

How do we combine them into one single projection of the future?

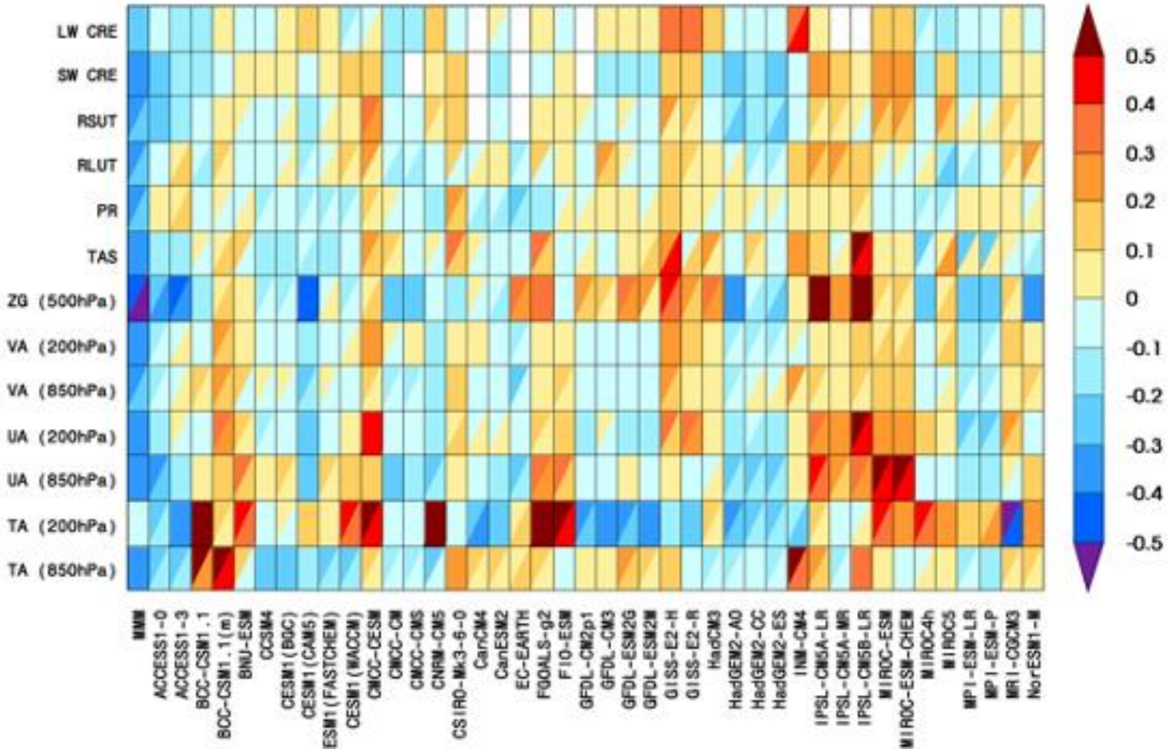
- One model, one vote
  - All models are not equally good
- Weighting models
  - Different models can be very similar



Types: multi-model ensemble

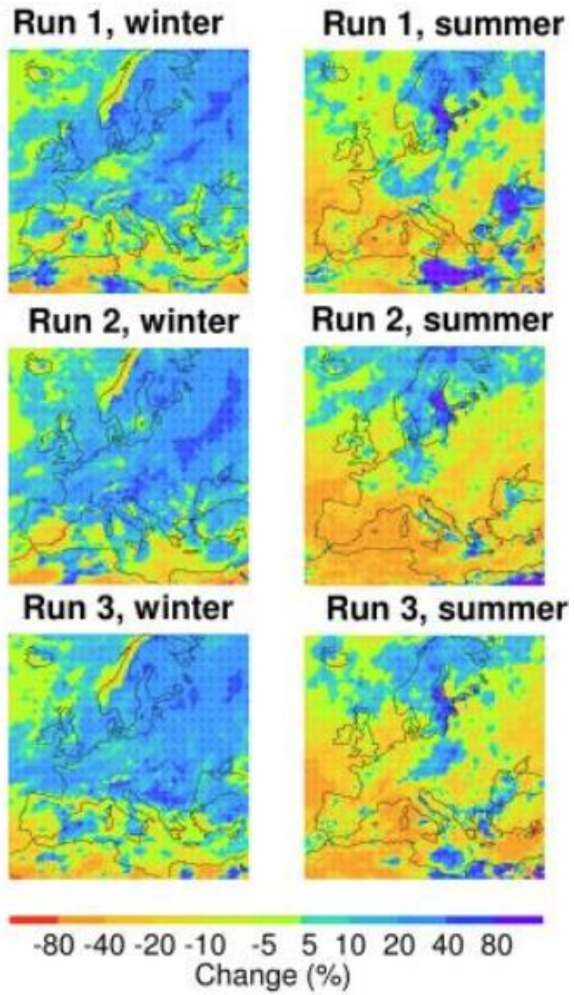
How do we combine them into one single projection of the future?

- One model, one vote
  - All models are not equally good
- Weighting models
  - Different models can be very similar
- Ranking models by performance
  - For each variable



Types: initial conditions ensemble

- 1 model run more than once give slightly different responses [climate models]
- For weather forecasts, it can lead to a significant change in the output.



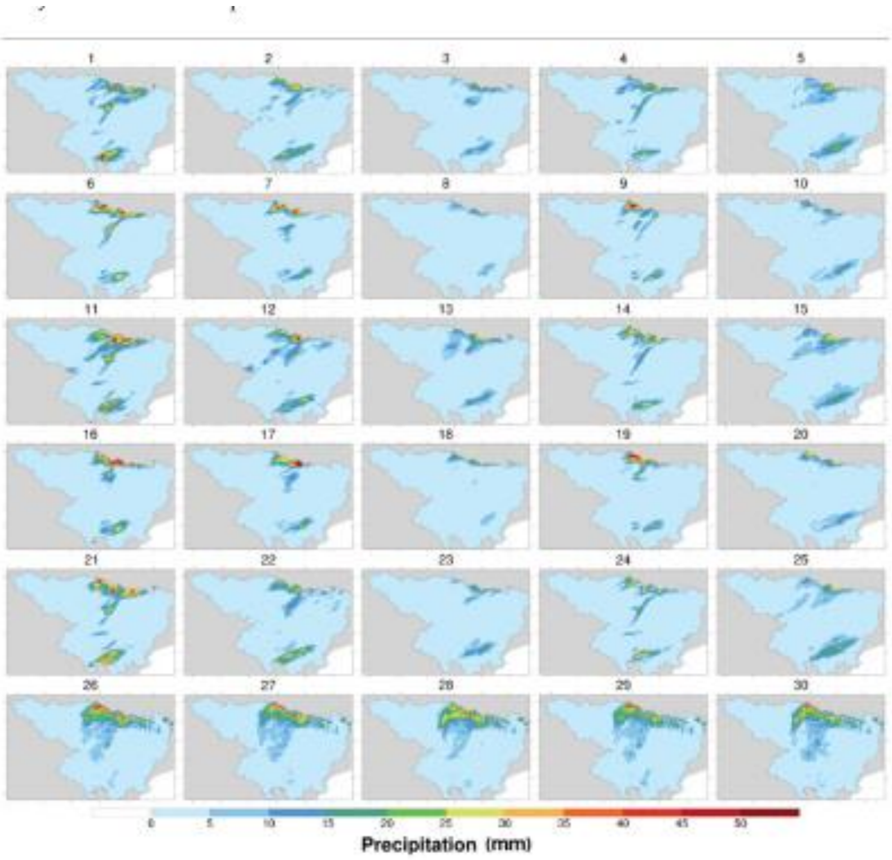
Types: multi-physics ensemble

ELSEVIER
Atmospheric Research
Volume 190, 1 July 2017, Pages 55-67

Performance of multi-physics ensembles in convective precipitation events over northeastern Spain

E. García-Ortega [a], J. Lorenzana [b], A. Merino [a], S. Fernández-González [c], L. López [a], J.L. Sánchez [a]

Member	mp	lw-sw	cu	Member	mp	lw-sw	cu	Member	mp	lw-sw	cu
1	PLS	RRT	KFS	11	PLS	RRT	BMJ	21	PLS	RRT	GDE
2	SM6	RRT	KFS	12	SM6	RRT	BMJ	22	SM6	RRT	GDE
3	GCS	RRT	KFS	13	GCS	RRT	BMJ	23	GCS	RRT	GDE
4	NTS	RRT	KFS	14	NTS	RRT	BMJ	24	NTS	RRT	GDE
5	MDS	RRT	KFS	15	MDS	RRT	BMJ	25	MDS	RRT	GDE
6	PLS	NGD	KFS	16	PLS	NGD	BMJ	26	PLS	NGD	GDE
7	SM6	NGD	KFS	17	SM6	NGD	BMJ	27	SM6	NGD	GDE
8	GCS	NGD	KFS	18	GCS	NGD	BMJ	28	GCS	NGD	GDE
9	NTS	NGD	KFS	19	NTS	NGD	BMJ	29	NTS	NGD	GDE
10	MDS	NGD	KFS	20	MDS	NGD	BMJ	30	MDS	NGD	GDE



More on Ensemble Predictions

The Skill of Ensemble Prediction Systems

FRÉDÉRIC ATGER

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received 6 April 1998, in final form 29 September 1998)

ABSTRACT

The performance of ensemble prediction systems (EPSs) is investigated by examining the probability distribution of 500-hPa geopotential height over Europe. The probability score (or half Brier score) is used to evaluate the quality of probabilistic forecasts of a single binary event. The skill of an EPS is assessed by comparing its performance, in terms of the probability score, to the performance of a reference probabilistic forecast. The reference forecast is based on the control forecast of the system under consideration, using model error statistics to estimate a probability distribution. A decomposition of the skill score is applied in order to distinguish between the two main aspects of the forecast performance: reliability and resolution. The contribution of the ensemble mean and the ensemble spread to the performance of an EPS is evaluated by comparing the skill score to the skill score of a probabilistic forecast based on the EPS mean, using model error statistics to estimate a probability distribution.

The performance of the European Centre for Medium-Range Weather Forecasts (ECMWF) EPS is reviewed. The system is skillful (with respect to the reference forecast) from +96 h onward. There is some skill from +48 h in terms of reliability. The performance comes mainly from the contribution of the ensemble mean. The contribution of the ensemble spread is slightly negative, but becomes positive after a calibration of the EPS standard deviation. The calibration improves predominantly the reliability contribution to the skill score. The calibrated EPS is skillful from +72 h onward.

The impact of ensemble size on the performance of an EPS is also investigated. The skill score of the ECMWF EPS decreases steadily with reducing numbers of ensemble members and the resolution is particularly affected. The impact is mainly due to the ensemble spread contributing negatively to the skill. The ensemble mean contribution to the skill decreases marginally when reducing the ensemble size up to 11 members.

The performance of the U.S. National Centers for Environmental Prediction (NCEP) EPS is also reviewed. The NCEP EPS has a lower skill score (vs a reference forecast based on its control forecast) than the ECMWF EPS especially in terms of reliability. This is mainly due to the smaller spread of the NCEP EPS contributing negatively to the skill. On the other hand, the NCEP and ECMWF ensemble means contribute similarly to the skill. As a consequence, the performance of the two systems in terms of resolution is comparable.

The performance of a poor man's EPS, consisting of the forecasts of different NWP centers, is discussed. The poor man's EPS is more skillful than either the ECMWF EPS or the NCEP EPS up to +144 h, despite a negative contribution of the spread to the skill score. The higher skill of the poor man's EPS is mainly due to a better resolution.

1. Introduction

Until the advent of ensemble prediction systems (EPSs) the evaluation of the quality of numerical weather forecasts was essentially based on a space–time comparison between forecast and verifying values, with only one forecast value and one verification value occurring at the same time and the same place. Since December 1992, both the U.S. National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF) have produced operational forecasts based on ensemble prediction (Tracton and Kalnay 1993; Palmer et al. 1993). An EPS is a prediction system designed to provide an ensemble of N forecasts of the meteorological state, considered as N independent realizations of a predicted probability distribution. The quality evaluation of an EPS should thus be based on the verification of a probability distribution. This implies that the forecast error cannot be estimated from a simple comparison between a forecast value and a verifying value. While the forecast is a distribution of values, the basic verification is still a single value. Two different approaches can be followed to solve this dilemma:

- The quality of a single probability distribution forecast (one time, one location) is estimated from the conditional probability that the actual verification occurs, given the probability distribution (Wilson 1995). In this Bayesian approach, the performance depends on two independent aspects: (i) how close the distribution

Corresponding author address: Dr. Frédéric Atger, Météo-France (SCEM/PREVI), 42, av. G. Coriolis, 31057 Toulouse Cedex, France.
E-mail: frederic.atger@meteo.fr

10.1175/1520-0493(1999)127<1941:TSOEPS>2.0.CO;2

The Benefits of Multianalysis and Poor Man's Ensembles

NEILL E. BOWLER, ALBERTO ARRIBAS, AND KENNETH R. MYLNE

Met Office, Exeter, United Kingdom

(Manuscript received 18 September 2007, in final form 18 February 2008)

ABSTRACT

A new approach to probabilistic forecasting is proposed, based on the generation of an ensemble of equally likely analyses of the current state of the atmosphere. The rationale behind this approach is to mimic a poor man's ensemble, which combines the deterministic forecasts from national meteorological services around the world. The multianalysis ensemble aims to generate a series of forecasts that are both as skillful as each other *and* the control forecast. This produces an ensemble mean forecast that is superior not only to the ensemble members, but to the control forecast in the short range even for slowly varying parameters, such as 500-hPa height. This is something that it is not possible with traditional ensemble methods, which perturb a central analysis.

The results herein show that the multianalysis ensemble is more skillful than the Met Office's high-resolution forecast by 4.5% over the first 3 days (on average as measured for RMSE). Similar results are found for different verification scores and various regions of the globe. In contrast, the ensemble mean for the ensemble currently run by the Met Office performs 1.5% worse than the high-resolution forecast (similar results are found for the ECMWF ensemble). It is argued that the multianalysis approach is therefore superior to current ensemble methods. The multianalysis results were achieved with a two-member ensemble: the forecast from a high-resolution model plus a low-resolution perturbed model. It may be possible to achieve greater improvements with a larger ensemble.

1. Introduction

Ensemble forecasting has its roots in attempts to understand the limits of deterministic prediction of the atmospheric state (Lewis 2005). By running a number of forecasts from a set of initial conditions, which are consistent with our knowledge of the current state of the atmosphere, we hope to gain an insight into the uncertainty in the forecast. Generally, this has been performed by creating a set of perturbations to add to a given best guess (or analysis) of the current state of the atmosphere (Toth and Kalnay 1993; Buizza and Palmer 1995).

An additional benefit of ensemble forecasting is that the ensemble mean forecast typically outperforms a forecast based on a single run of a numerical model. The latter forecasts are often described as "deterministic" forecasts. Because each ensemble member has a different realization of certain less-predictable small-scale features, the ensemble mean forecast will not contain such features, because these have been averaged out. This averaging is a curse as well as a blessing, because it means that the ensemble mean forecast will become increasingly smooth as the forecast progresses and the uncertainty increases. Thus, one needs to be very careful how the ensemble mean forecast is used (Smith 2003). This means that the ensemble mean is of little use on its own, and it is often supplemented by the probability of various events occurring derived from the whole ensemble. Nonetheless, any improvement to the ensemble mean forecast has a large effect on the quality of the ensemble forecast (Buizza et al. 2005).

The size of an ensemble is typically much smaller than the number of degrees of freedom in a numerical model [the number of grid points of an operational numerical model is currently $O(10^8)$]. This means that the focus in ensemble forecasting has been to choose perturbations to the deterministic analysis, which grow very rapidly. The two schemes initially used for medium-range forecasting are error breeding (Toth and Kalnay 1993) and singular vectors (Buizza and Palmer

Corresponding author address: Neill Bowler, Met Office, FitzRoy Road, Exeter, EX1 3PB, United Kingdom.
E-mail: neill.bowler@metoffice.gov.uk

DOI: 10.1175/2008MWR2381.1

10.1175/2008MWR2381.1

Ensembles beyond climate science

- Finance
  - Anomalous trade prices
  - Credit card fraud
- Remote sensing and geosciences
  - Thermal anomalies [forest fire detection]
  - Gravitational anomalies [subsurface water detection]

Assignment

The Poor Man’s Ensemble

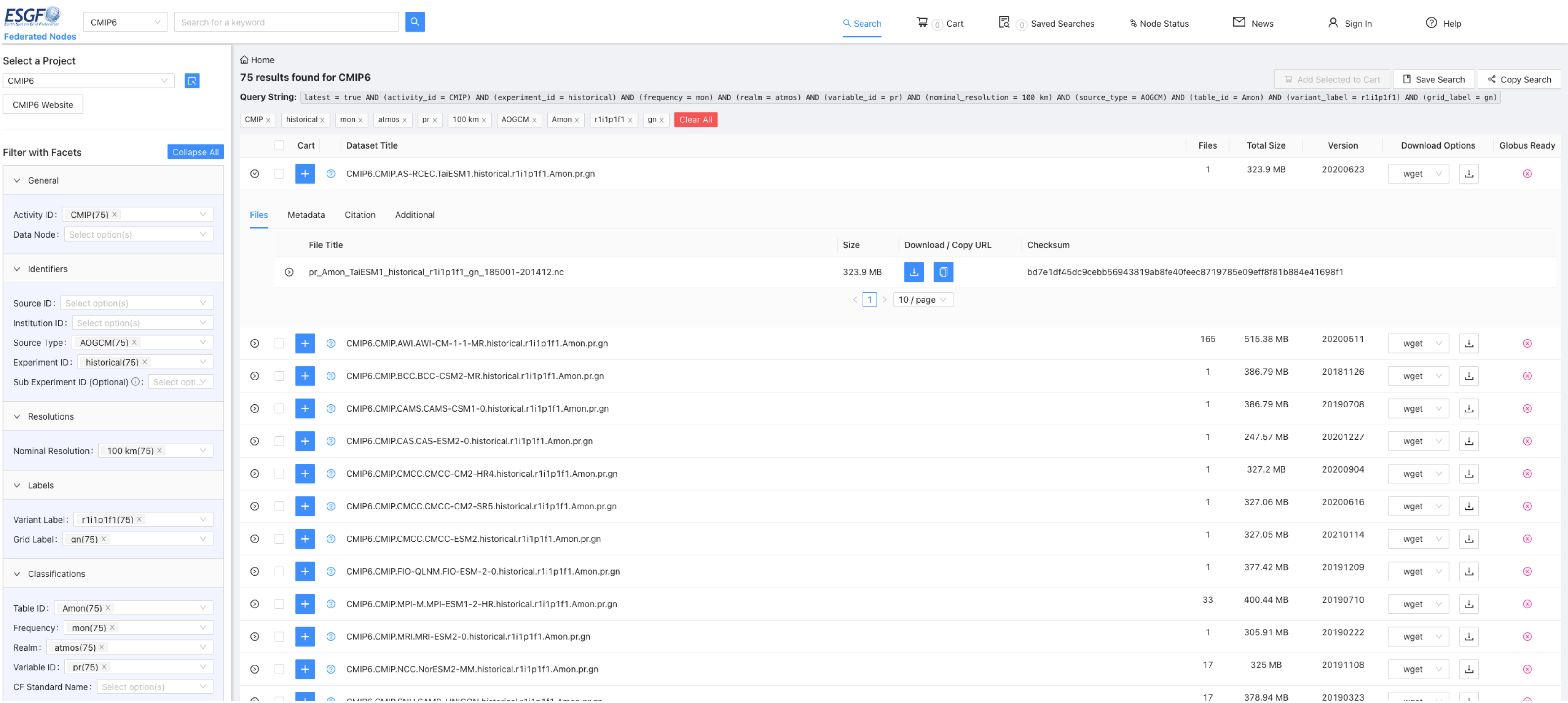
Creating a climate model ensemble: MME

Should be the same:

- Scenario/SSP
- Spatial & temporal resolution

Creating a climate model ensemble: MME

1. Get *pr* data from 4 CMIP6 models [same spatial resolution]
2. Select the period [e.g. 1850-2014]
3. Select seasonal data [DJF, MAM, JJA, SON]
4. Compute seasonal mean
5. Create the MME
6. Compute the Standard Deviation



https://esgf-metagrid.cloud.dkrz.de/search

File list

pr_Amon_CMCC-CM2-HR4_historical_r1i1p1f1_gn_185001-201412.nc

pr_Amon_TaiESM1_historical_r1i1p1f1_gn_185001-201412.nc

pr_Amon_CESM2-WACCM_historical_r1i1p1f1_gn_185001-201412.nc

pr_Amon_FIO-ESM-2-0_historical_r1i1p1f1_gn_185001-201412.nc