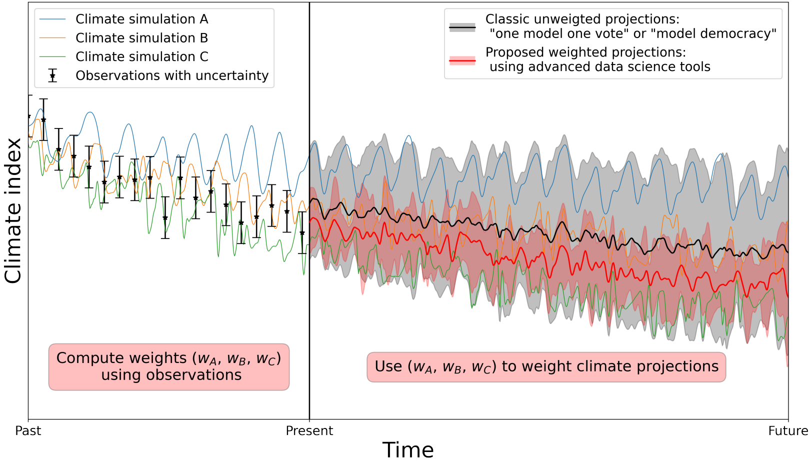
Narrowing uncertainties in climate projections using data science tools

P. Tandeo[1], P. Ailliot[1,2], C. Bello[1], B. Chapron[3], T. Chau[4], R. Knutti[5], P. Le Bras[6,1], P. Naveau[4], V. Monbet[7], J. Ruiz[8], F. Sévellec[6], A.-M. Treguier[6]

IMT Atlantique, France[1]
Univ. Brest, France[2]
IFREMER, France[3]
LSCE-IPSL, France[4]
ETH Zurich, Switzerland[5]
IUEM, France[6]
Univ. Rennes I, France[7]
Univ. Buenos Aires, Argentina[8]

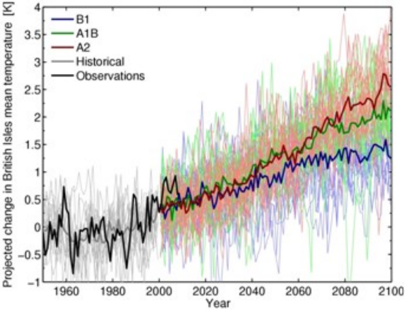
April 14, 2021
IMT-Atlantique & RIKEN Online Joint Seminar Series

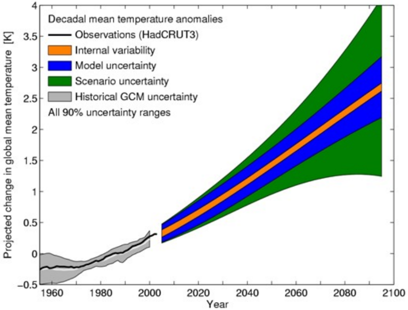
Context and main goal



- In the IPCC → ensemble of **unweighted projections** ("one model one vote" or "model democracy", [Knutti, 2010])
- Idea → **learn weights** from historical observations and simulations, then **propagate weights** to climate projections

Climate uncertainties revealed by CMIP

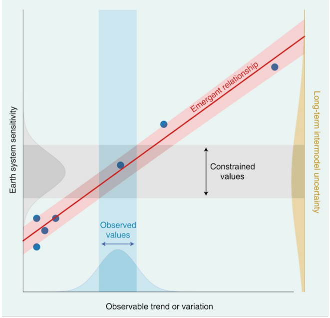




Source: [Hawkins and Sutton, 2011]

- Climate projections are sensitive to internal, model and scenario uncertainties
- **Potential to narrow uncertainties**, especially in **regional climate predictions** [Hawkins and Sutton, 2009]

New at IPCC: emergent constraints



Source: [Eyring et al., 2019]

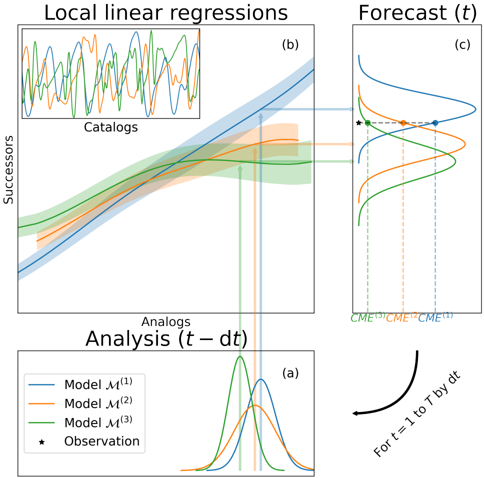
Pros:

- Easy to implement (projection using linear regression)
- Easy to understand (synthetic graphical representation)
- Do not weight climate simulations (not directly)

Cons:

- Causality not obvious (especially for large horizons)
- Low number of samples to fit the regression
- Questionable linear relationship and homoscedasticity

Proposed approach: use advanced data science methods



Three main steps:

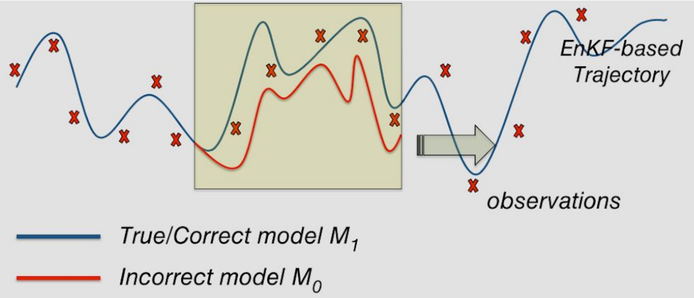
- (a) **Data assimilation** (ensemble Kalman filter)
- (b) **Data-driven forecasting** (local linear regression)
- (c) **Distance obs-forecasts** (contextual model evidence)

Ingredient 1: (a) EnKF + (c) contextual model evidence

Contextual model evidence in data assimilation (CME):

$$\mathcal{L}\left(\mathrm{y}(t)|\mathcal{M}_{(i)}\right) \propto \exp\left(-\mathrm{d}_{(i)}(t)^\top \Sigma_{(i)}(t)^{-1}\mathrm{d}(t)\right) \qquad (1)$$

with the **innovation** defined by its mean and covariance:
$\mathrm{d}_{(i)}(t) = \mathrm{y}(t) - \mathrm{Hx}^f_{(i)}(t)$ and $\Sigma_{(i)}(t) = \mathrm{HP}^f_{(i)}(t)\mathrm{H}^\top + \mathrm{R}$.



Source: [Carrassi et al., 2017]

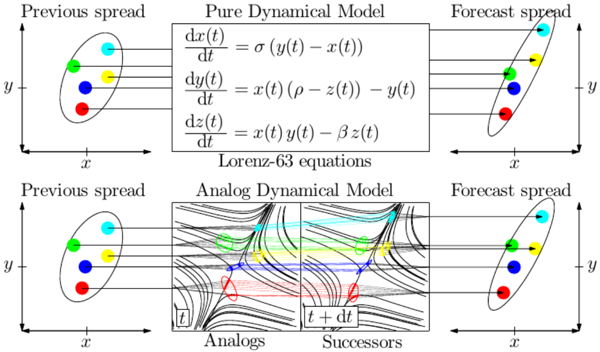
Ingredient 2: (a) EnKF + (b) analog forecasting

Analog forecasting within data assimilation (AnDA):

$$\mathsf{x}(t) = \mathcal{A}\left(\mathsf{x}(t - \mathrm{d}t), \boldsymbol{\eta}(t)\right) \quad (2)$$

$$\mathsf{y}(t) = \mathcal{H}\left(\mathsf{x}(t)\right) + \boldsymbol{\epsilon}(t) \quad (3)$$

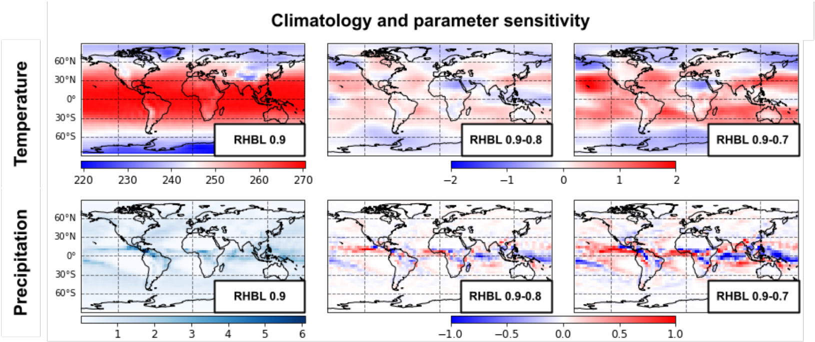
with $\mathcal{A}$ the **analog forecasting** operator [Lguensat et al., 2017].



Source: [Tandeo et al., 2015]

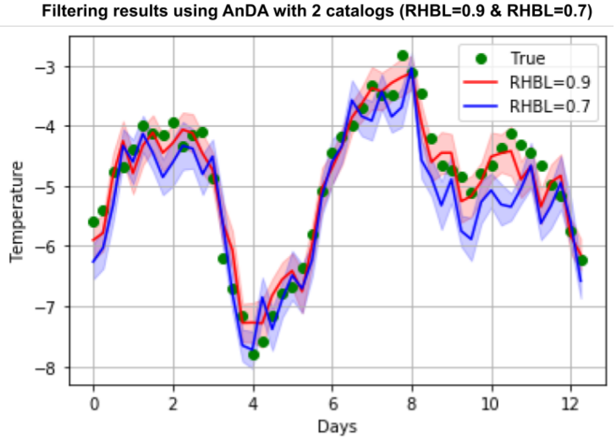
Ingredients 1 + 2: (a) EnKF + (b) AF + (c) CME

- Tested on a **simplified GCM** (SPEEDY, [Molteni, 2003]):
  - 7 vertical levels, 96×48 horizontal grid
  - simple physics (convection, clouds, radiation, boundary layer)
- Relative Humidity threshold in the Boundary Layer:
  - RHBL $= 0.9 \rightarrow$ the "true" model
  - RHBL $= 0.8 \rightarrow$ slightly imperfect model
  - RHBL $= 0.7 \rightarrow$ imperfect model

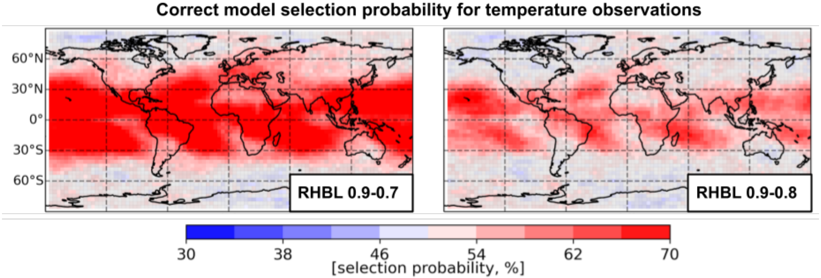


Ingredients 1 + 2: (a) EnKF + (b) AF + (c) CME

- Analog data assimilation details:
  - EnKF with 40 members with adaptive inflation [Miyoshi, 2011]
  - 30-years catalogs for 3 parameterizations (RHBL 0.9, 0.8, 0.7)
  - 3D local domains (3 vertical levels, 3×3 horizontal grid)
  - 3 years of noisy observations from RHBL 0.9 (std $= 0.7K$)

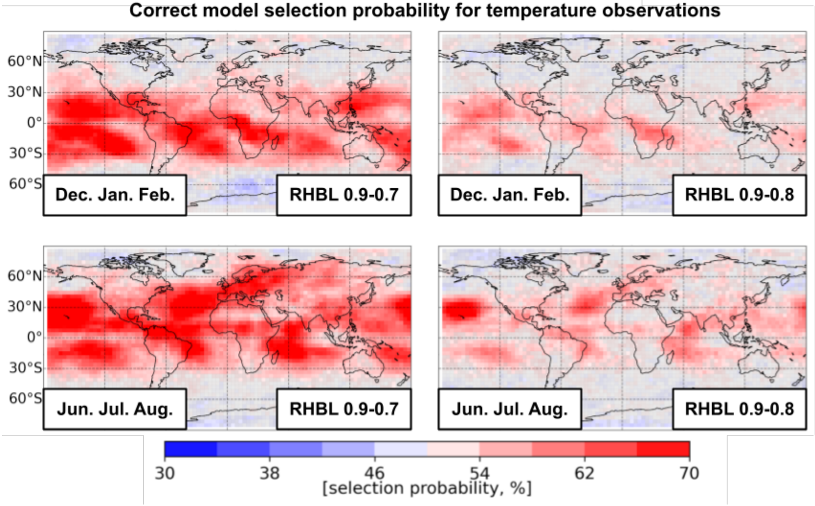


Ingredients 1 + 2: (a) EnKF + (b) AF + (c) CME



- Results about model identification (in space):
  - tropical-subtropical regions affected by model imperfections
  - degree of imperfection is captured (RHBL 0.7 $<$ 0.8)

Ingredients 1 + 2: (a) EnKF + (b) AF + (c) CME



- Results about model identification (in space and time):
  - sensibility to the RHBL parameter is evolving in time
  - detection of model imperfection more important in summers (i.e., when there is more convection observed)

Ingredients 1 + 2: (a) EnKF + (b) AF + (c) CME

Conclusions:

- Combination of **advanced data-science methods**
- Able to compare **short-term model dynamics**
- Ruiz et al., will be submitted soon to the *Journal of Climate*

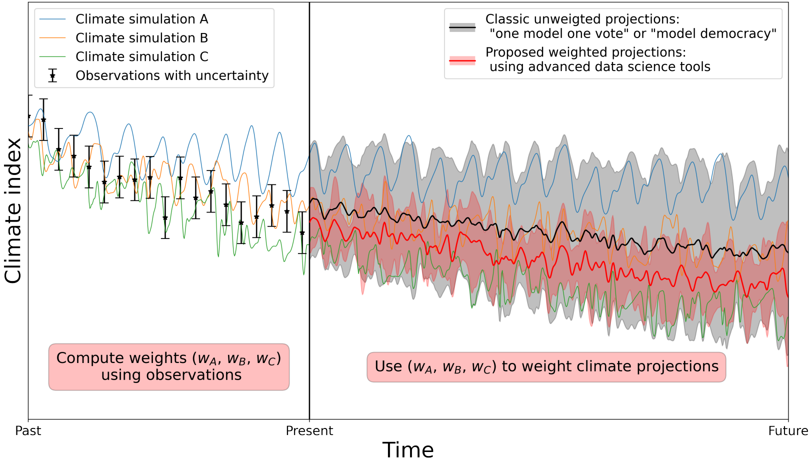
Pros:

- **Local approach** (sub-domain, given period, partial variables)
- **Low-cost procedure** (no need to run climate models)
- Capture **spatiotemporal differences** in model identifications

Cons:

- Need historical numerical simulations
- Need tuning (analogs, inflation, domain, observations)
- May seem complicated (but not so much!)

Next step: application to climate simulations



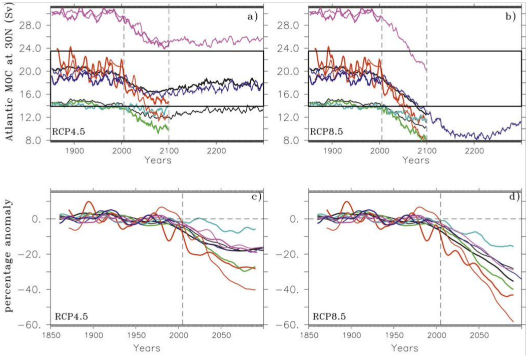
- Data → compare current observations to CMIP simulations
- Method → combine data-science methods (DA, AF, CME)
- Goal 1 → create weighted projections of climate metrics
- Goal 2 → reduce the uncertainty of climate projections

Next step: application to climate simulations

Specificity of climate simulations [Knutti et al., 2019]:

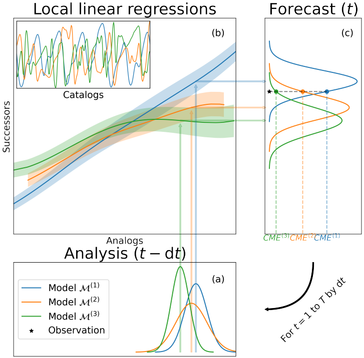
- **Interdependence** → many CMIP models share ideas, parts of code, or whole components (e.g., the sea ice model)
- **Performance** → some CMIP models are "good" at representing a specific climate index, other models are not
- Simulations are sometimes **biased** and need **standardization**

AMOC variations in CMIP5 for 2 RCP scenarios



Source: [Cheng et al., 2013]

Next step: application to climate simulations



Caveats and improvement of the methodology:

- (a) $\rightarrow$ deal with **model interdependence** (e.g., work with clusters of models), deal with **non-parametric** distributions
- (b) $\rightarrow$ find differences in the **short-term dynamics** of climate metrics (especially in the extremes), find **relevant dt**
- (c) $\rightarrow$ define more **flexible metrics** (e.g., optimal transport), find **relevant observations** (long time series, knowing noise)

Thank you for your attention! Any questions?

MAFALDA:

Multi-climate-model Analog Forecasting for Attributing Likelihoods using Data Assimilation

French ANR program

JCJC "young researcher"

Under evaluation (2nd round)

Carrassi, A., Bocquet, M., Hannart, A., and Ghil, M. (2017).
Estimating model evidence using data assimilation.
Quarterly Journal of the Royal Meteorological Society, 143(703):866–880.

Cheng, W., Chiang, J. C., and Zhang, D. (2013).
Atlantic meridional overturning circulation (AMOC) in CMIP5 Models: RCP and historical simulations.
Journal of Climate, 26(18):7187–7197.

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C., Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C., Simpson, I. R., Stouffer, R. J., and Williamson, M. S. (2019).
Taking climate model evaluation to the next level.
Nature Climate Change, 9(2):102–110.

Hawkins, E. and Sutton, R. (2009).
The potential to narrow uncertainty in regional climate predictions.
Bulletin of the American Meteorological Society, 90(8):1095–1108.

Hawkins, E. and Sutton, R. (2011).
The potential to narrow uncertainty in projections of regional precipitation change.
Climate dynamics, 37(1):407–418.

Knutti, R. (2010).

The end of model democracy?
Climatic Change, 102:395–404.

Knutti, R., Baumberger, C., and Hadorn, G. H. (2019).
Uncertainty Quantification Using Multiple Models—Prospects and Challenges.
In Computer Simulation Validation, pages 835–855. Springer.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., and Fablet, R. (2017).
The Analog Data Assimilation.
Monthly Weather Review, 145(10):4093–4107.

Miyoshi, T. (2011).
The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter.
Monthly Weather Review, 139(5):1519–1535.

Molteni, F. (2003).
Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments.
Climate Dynamics, 20(2-3):175–191.

Tandeo, P., Ailliot, P., Ruiz, J. J., Hannart, A., Chapron, B., Easton, R., and Fablet, R. (2015).
Combining analog method and ensemble data assimilation: application to the Lorenz-63 chaotic system.
In Machine Learning and Data Mining Approaches to Climate Science, pages 3–12.