

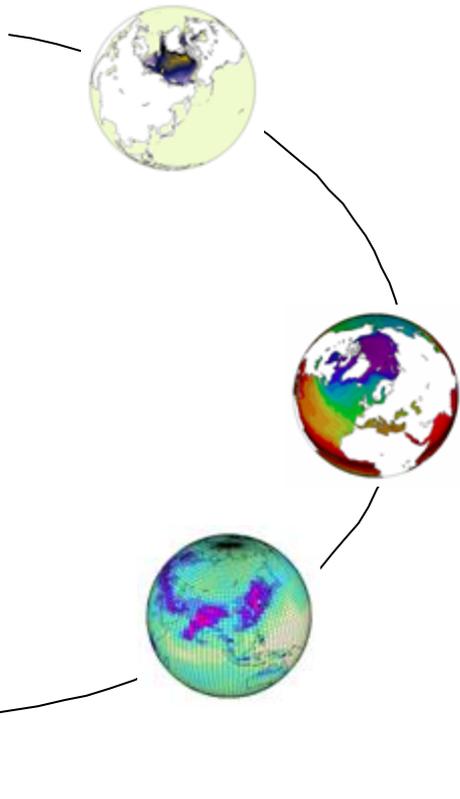
Reproductibilité dans les sciences du climat: Le cas de la modélisation du climat

Sylvie Joussaume

Emérite au CNRS

Institut Pierre Simon Laplace

Laboratoire des Sciences du Climat et de l'Environnement

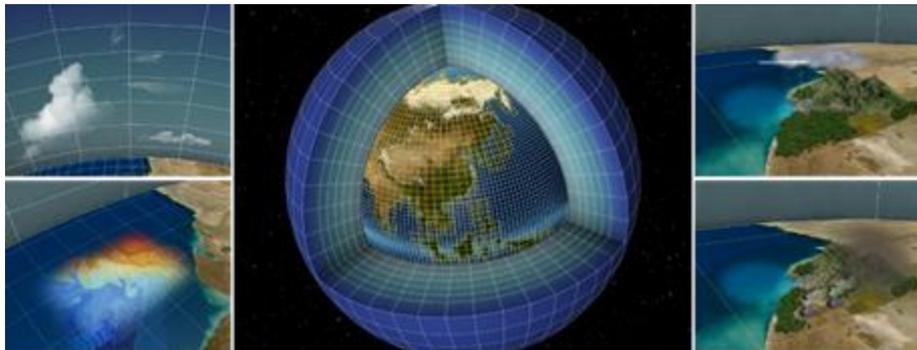


<https://climeri-france.fr/>



<https://is.enes.org/>

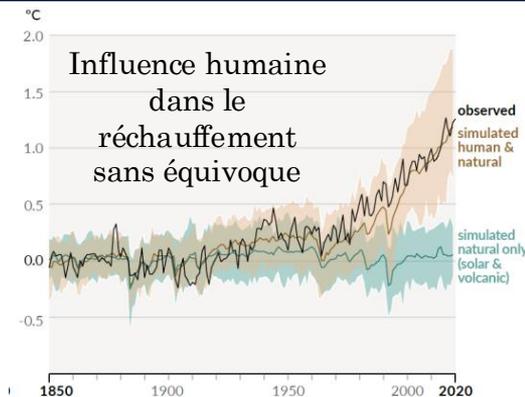
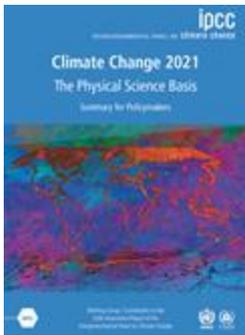
Modélisation du climat



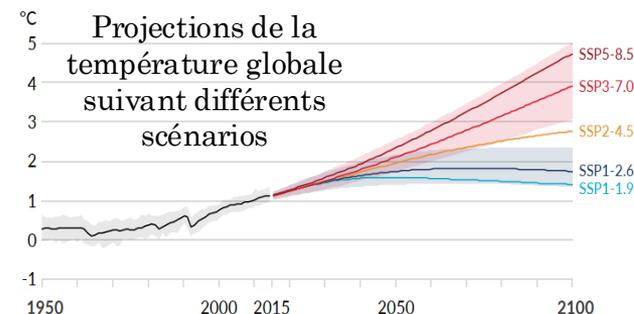
Comprendre et prévoir
la variabilité du climat
et ses changements

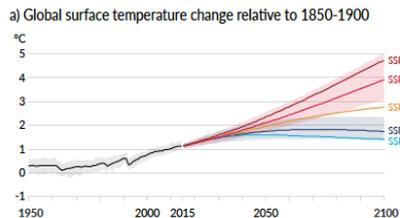
Rapports du GIEC : Simulations internationales de référence
Coordonnées par le Programme Mondial de Recherche sur le Climat (WCRP)

6^{ème} rapport
2021



a) Global surface temperature change relative to 1850-1900





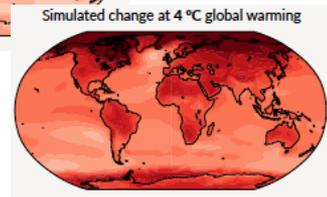
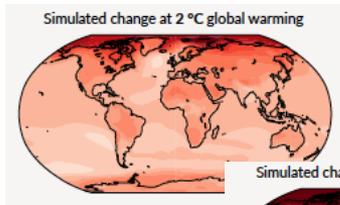
Projections

IPCC AR6
(2021)

CMIP6
48 groupes
120 modèles
312 expériences

Scénarios futurs

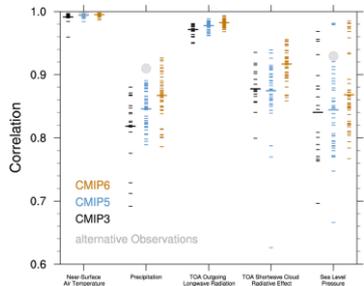
Prévisions décennales



Evaluation

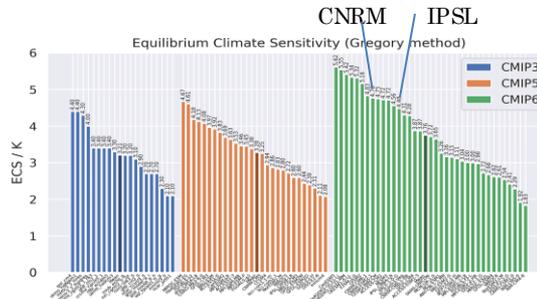
Mécanismes

Pattern correlation with observations



Simulations historiques

Climats passés



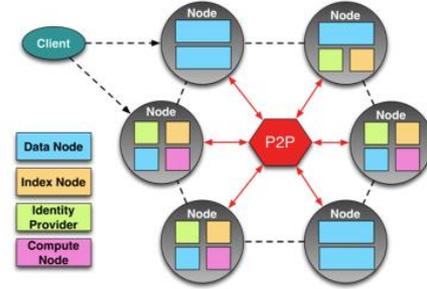
Rôle des nuages

Volcans

Usage des terres

Chimie, aérosols,
cycle du carbone

Base de données internationale Earth System Grid Federation



15 000 utilisateurs

CMIP6 :

25 Po de données, 14 M « datasets »
nœuds EU: moyenne 400 To/mois

Données FAIR

Netcdf-CF

Vocabulaire commun

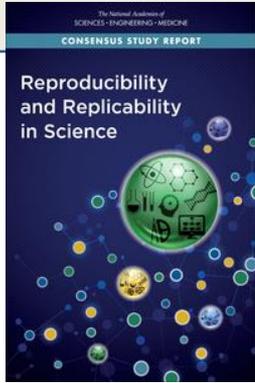
Documentation: modèles & expériences

Nombreux services avals:
« services climatiques »



PI: IPSL

Concepts: Reproductibilité & Réplicabilité



NAS, 2019

CONCLUSION 3-1:

For this report, **reproducibility** is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.

This definition is synonymous with “computational reproducibility,” and the terms are used interchangeably in this report.

Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

« Despite efforts to coalesce around the use of these terms, lack of consensus persists across disciplines. »

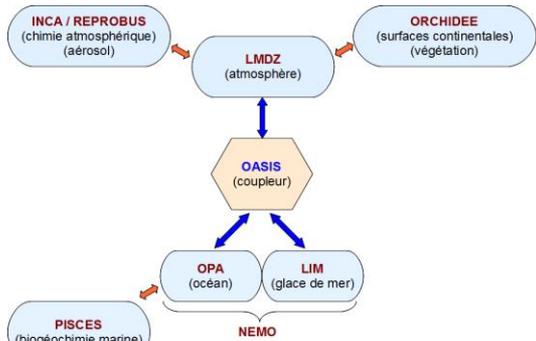
Reproductibilité et Réplicabilité : cas des simulations climatiques internationales

Production des données

Conditions
initiales
Atmosphère
Océans

Conditions
de
l'expérience
« Forçage »

Codes



Simulation



GENCI
Le calcul intensif au service de la connaissance

Exploitation des données

Données

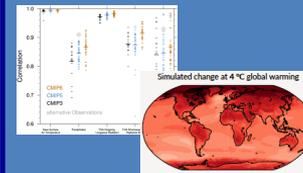
Fichiers produits
par le modèle

Température, vents,
pression,
précipitations

Localisation
temps

Traitement des Données

Analyses
Statistiques
Cartes ...



Chaque modèle
& au niveau international

Reproductibilité : Au niveau des données (1/2)

Données

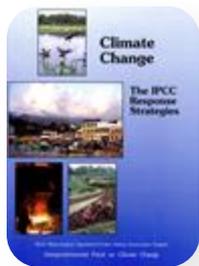
1- Données accessibles

> Même standard de données et métadonnées, même vocabulaire

2- Données ouvertes :

NB: A évolué au fil des projets: « confiance sur la bonne utilisation des données »

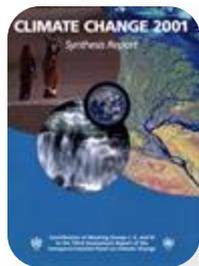
Simulations climatiques internationales et rapports du GIEC



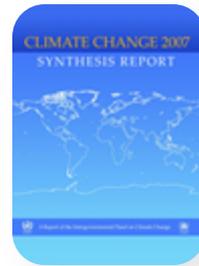
1^{er} rapport du GIEC



2^{ème} rapport du GIEC



3^{ème} rapport du GIEC



4^{ème} rapport du GIEC



5^{ème} rapport du GIEC



6^{ème} rapport du GIEC

AMIPI 1990 AMIP2

CMIP1 1995

CMIP2 2001

CMIP3 2007

CMIP5 2014

CMIP6 2021

24 modèles
x 12 expériences

63 modèles
x 101 expériences

128 modèles
x 312 expériences

0.5 To

35 To

5 Po

25 Po

Données
accessibles
Sur projets

Données ouvertes
pour usage non
commercial

Données ouvertes
peu fermées pour
non commercial

Données ouvertes
Pour tous usages

Reproductibilité : Au niveau des données (2/2)

Données

1- Données accessibles

> Même standard de données et métadonnées, même vocabulaire

2- Données ouvertes :

NB: A évolué au fil des projets: « confiance sur la bonne utilisation des données »

3- Versions des données : qualité, pouvoir reproduire les diagnostics

Logiciel commun pour gérer les versions de données (depuis CMIP5)

Service de correction des données « errata » (depuis CMIP6)

Service de doi (par modèle/par expérience)(2600) et PID (fichiers, 40 M) (depuis CMIP6)

> reconnaissance des producteurs de données

Traçabilité des données

« Authenticité » des données

Reproductibilité : Au niveau du traitement des données

Traitement des données (ex rapports du GIEC)

Among objectives of IPCC:

*Archive and **provide transparency, traceability, and stability of data and scenarios** used by the IPCC in its reports, available at the DDC or elsewhere.*

4- Archivage long-terme des données : utilisées dans les rapports (« closing date »)
Archive spécifique « IPCC Distribution Data Center » (DKRZ, DE) depuis 1997

5- Logiciels d'analyse des données, de production de diagnostics
Garantir la reproductibilité des analyses du GIEC (mêmes données, mêmes méthodes)

Europe: Logiciel de diagnostics communs ESMValTool et de « recettes » mises à disposition

Nouveaux services: encore peu développés

Jupyter Notebooks,

« Web Processing Services » : traiter les données là où elles sont stockées

> environnements virtuels – ex projet PIA3 Gaia Data

R & R : Au niveau de la production des données (1/2)

Production des données

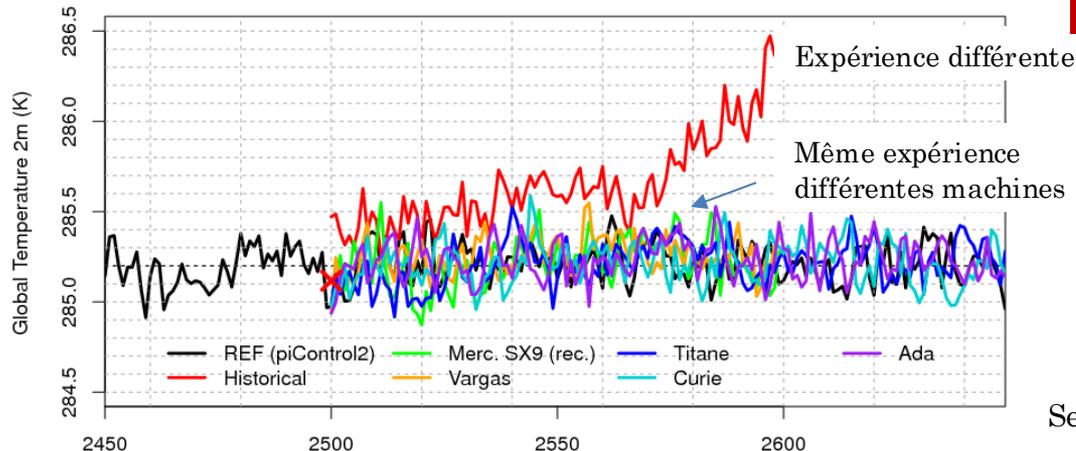
Au niveau de chaque modèle:

On s'assure que les résultats sont « **reproductibles** » :

Mêmes conditions initiales, même code, même expérience, même ordinateur > OK au bit près (vérifié très régulièrement !)

ATTENTION: ne marche pas dès qu'on change de ordinateur ou même d'environnement dans un ordinateur donné (ex OS)

Reproductibilité au niveau du modèle
Impossible (bit) par un groupe extérieur

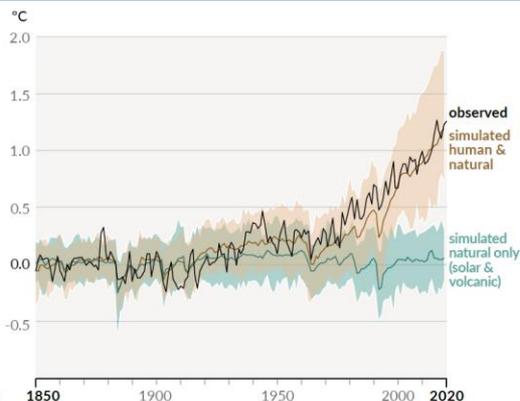


Système chaotique:
pas la même trajectoire
MAIS même climat

Production des données

Au niveau international:

- 1- **Mêmes conditions d'expérience:** travail communautaire international
conditions de forçage des simulations : gros travail, interdisciplinaire
> Intercomparaison des simulations : **étudier la « robustesse » des résultats**
- 2- **Documentation** des conditions d'expériences et de la version des modèles (ES-DOC)
Surtout pour faciliter la compréhension des données (processus pris en compte)
Documentation des modèles: **Souvent difficile à obtenir des groupes**



Robustesse et compréhension des résultats

Conclusions

Traitement des
données

Données

Production des
données

- Un effort important fait pour les simulations internationales partagées (enjeu GIEC)
Mais trop lourd pour être fait pour toutes les simulations publiées
- R&R: surtout au niveau des données
standards des données et métadonnées, traçabilité
- Se développe de plus en plus pour le traitement des données
- Au niveau de la production des données: en partie
Limitations intrinsèques : calculateurs et lourdeur des expériences
Reproductibilité bit à bit ou statistique au niveau d'un même modèle
Réplicabilité / Robustesse entre modèles
- Motivé d'abord par la science et le besoin de connaissance
(pas juste pour R&R)
- Un concept à stabiliser