

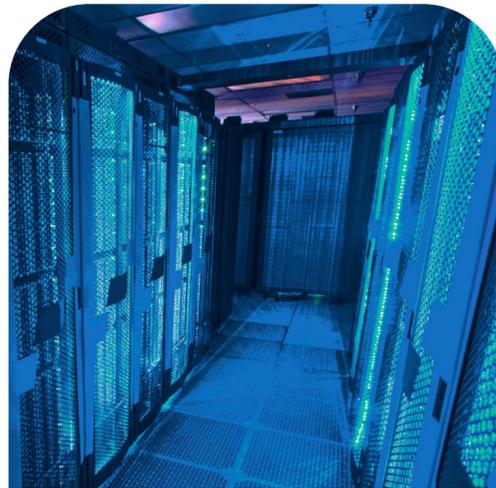


De la production à l'analyse de données issues des simulations climatiques

Levavasseur Guillaume

*Centre de Modélisation du Climat (CMC) de l'Institut Pierre-Simon Laplace (IPSL)
Centre de Calcul et de Donnée de l'IPSL (ESPRI)*

28 novembre 2023 - Formation migrations/climat



01

Une avalanche
de processeurs



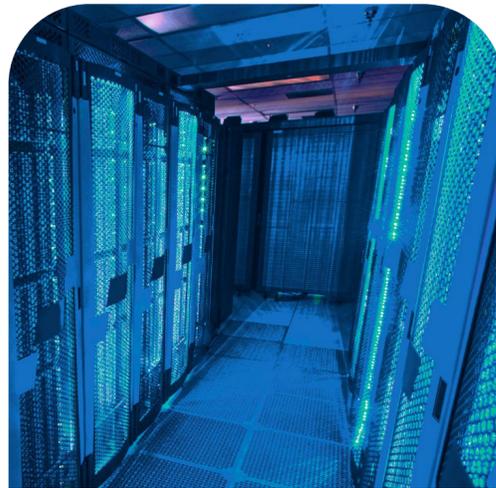
02

Une déferlante
de pétaoctets



03

Vers le 7^{ème}
rapport du GIEC



01

Une avalanche
de processeurs

“Difficile à voir. Toujours en mouvement est l’avenir.”

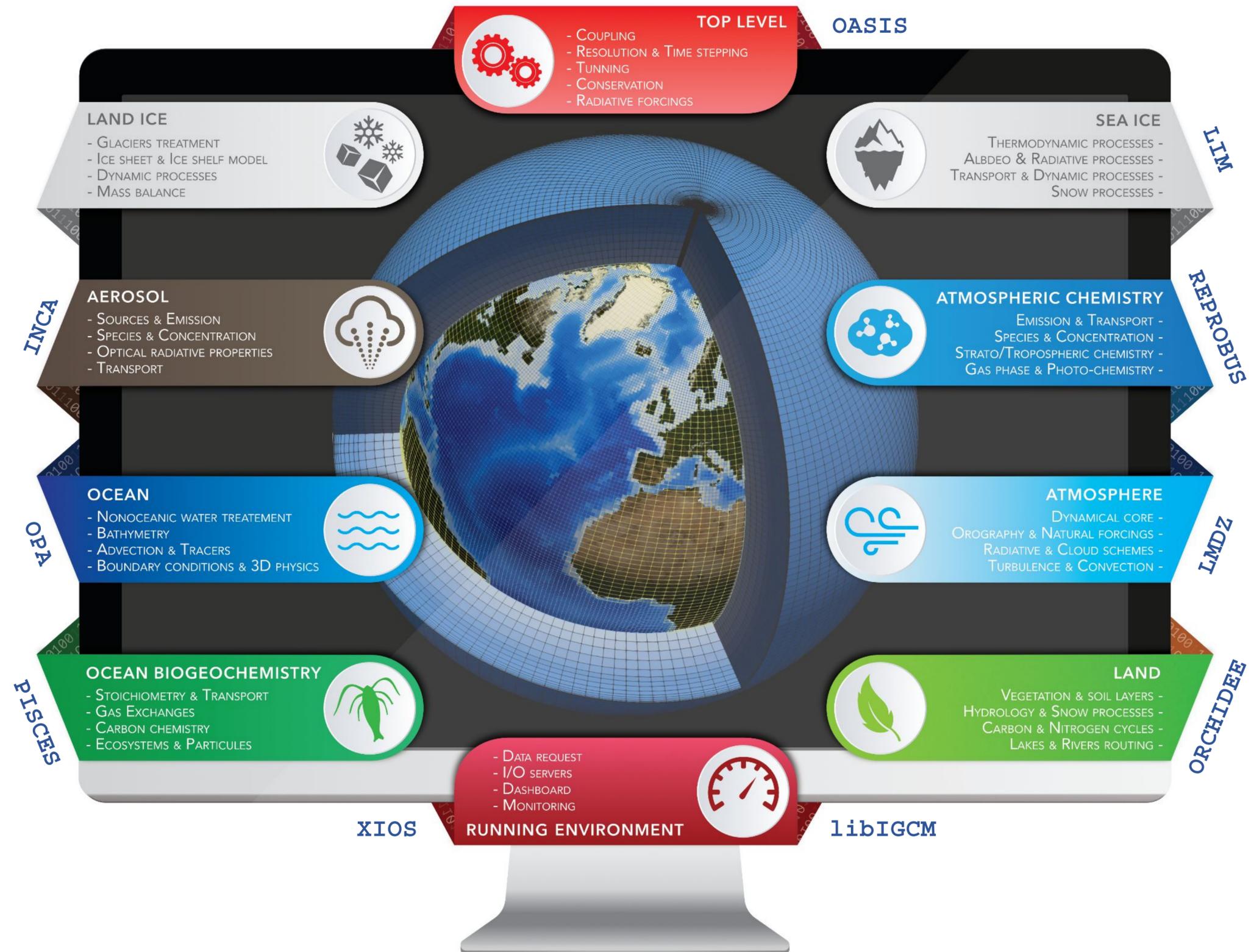
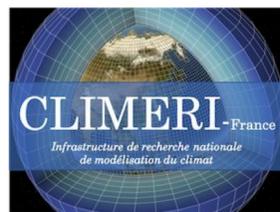
Yoda

Un modèle climatique couplé, kesako ?



Un “gros” algorithme informatique capable de simuler toutes (ou partie) des composantes du système terre.

- “couplé” = un code informatique pour chaque composante, qui “discutent” entre elles.
- Composantes/codes labellisés CNRS INSU
- Développements qui s’inscrivent dans des infrastructures de recherche nationale (CLIMERI-France) et européenne (ENES)



Ils font pas déjà la même chose à la météo ?



Bien que leurs algorithmes soient proches, ils diffèrent notamment par :

Leur horizon et résolution temporelle,

Leur échelle et résolution spatiale,

Leur paramétrisation,

Leurs usages.

Weather versus Climate

The difference between weather and climate is a matter of time

Weather

refers to short-term changes in the atmosphere.
It can change minute-to-minute, hour-to-hour and day-to-day

Climate

describes the average weather conditions in a specific area over a long period of time – 30 years or more

Satellites measure several aspects of Earth's weather as well as provide essential data over decades to monitor how our climate is changing

Day	Weather Icon
Mon	Sunny
Tue	Partly cloudy
Wed	Cloudy
Thu	Cloudy with rain
Fri	Cloudy with snow

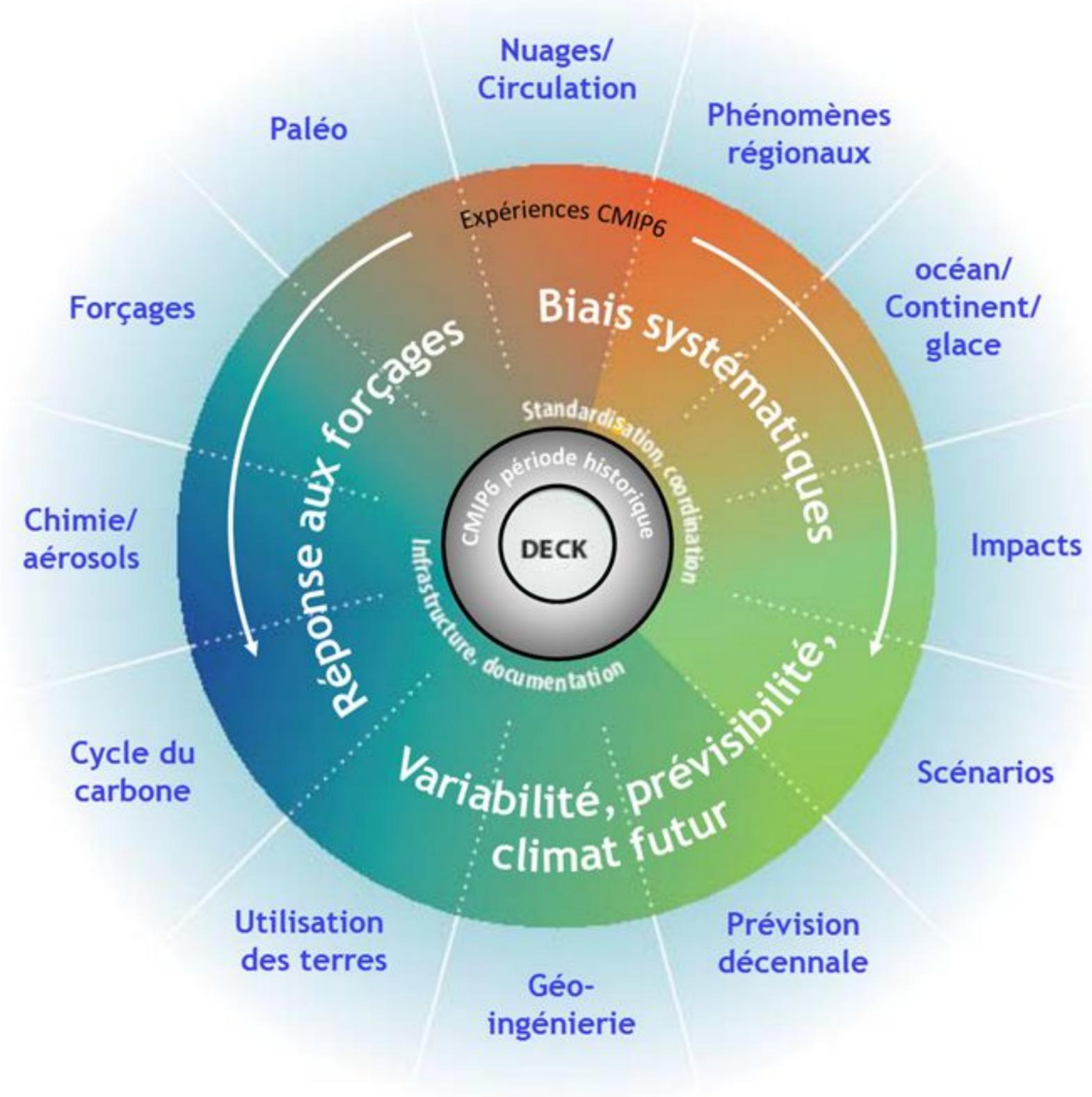
Line graph showing temperature trends from 1995 to 2040. The x-axis represents years (1995, 2000, 2005, 2010, 2015, 2020, 2025, 2030, 2035, 2040) and the y-axis represents temperature. The graph shows a clear upward trend, indicating global warming.

C'est quoi une "simulation" climatique ?

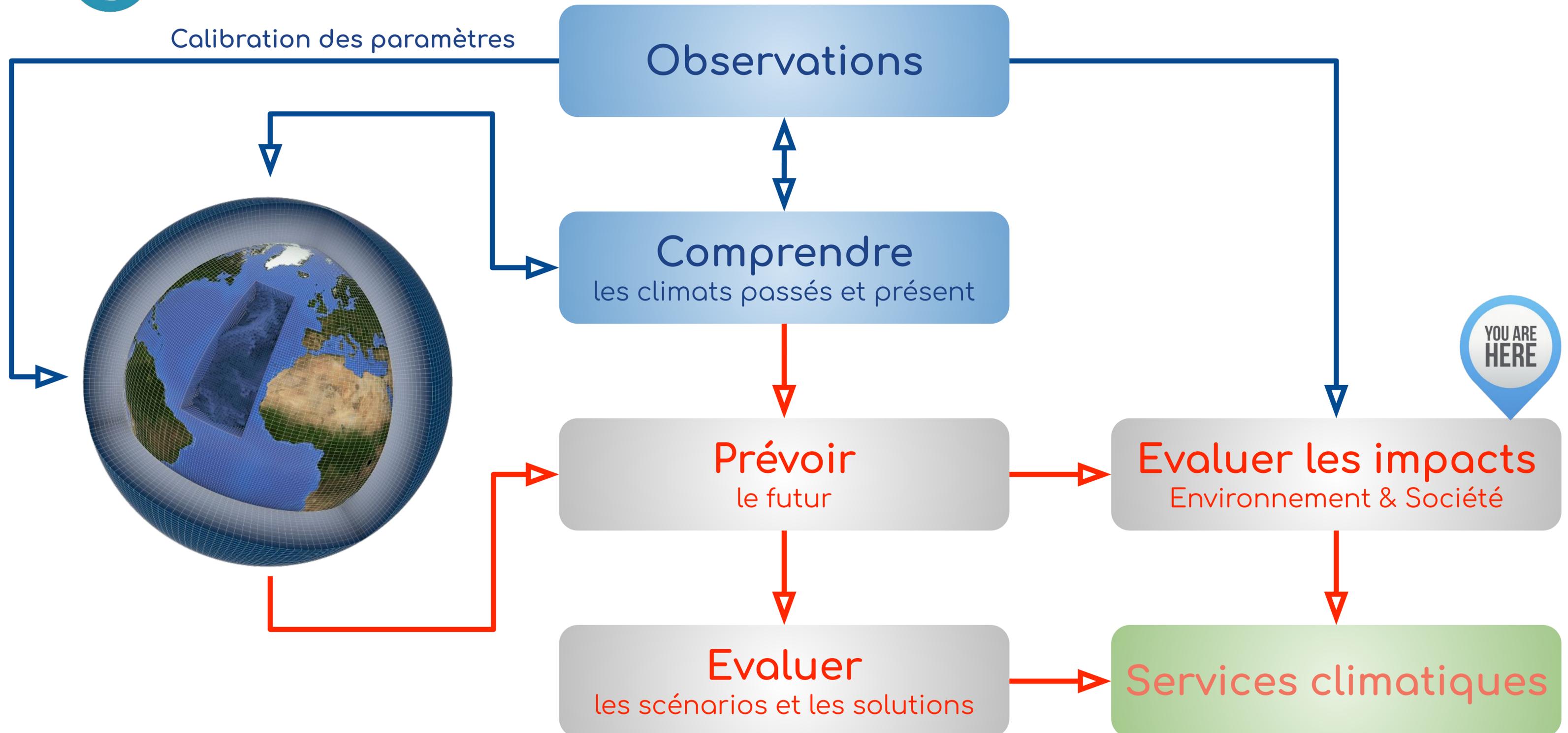


La "Data Request" est un outil coordonné par le WGCM

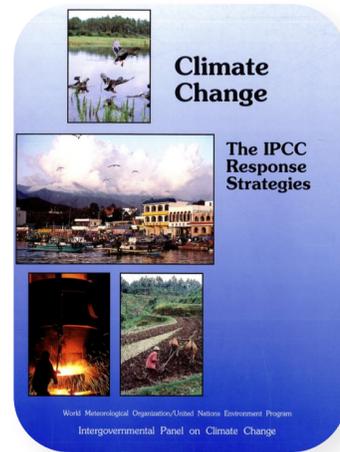
- **Recense et organise les protocoles de simulation** répartis sur différents groupes de diagnostics thématiques (ou "MIP": Model Intercomparison Project)
- **Listes les "forçages"** nécessaires à chaque protocole.
 - "historical" = Simulation du passé récent de 1850 à 2014, où l'on impose aux modèles les concentrations en gaz à effet de serre observées sur la période.
 - "piControl" = Simulation de contrôle pré-industrielle avec des conditions pré-industrielles qui n'évoluent pas. Ces conditions ont été choisies pour être représentatives de la période précédant le début de l'industrialisation à grande échelle, 1850 étant l'année de référence. La durée minimale est de 500 ans.
 - "ssp" = Simulation du futur proche jusqu'à 2100, où l'on impose des concentrations en gaz à effet de serre cohérentes avec différentes trajectoires socio-économiques (appelés scénarios RCP ou SSP pour Shared Socioeconomic Pathways).
- **Filtre les variables et leurs fréquences de sortie en fonction du contexte scientifique** des Grands Défis du WCRP.



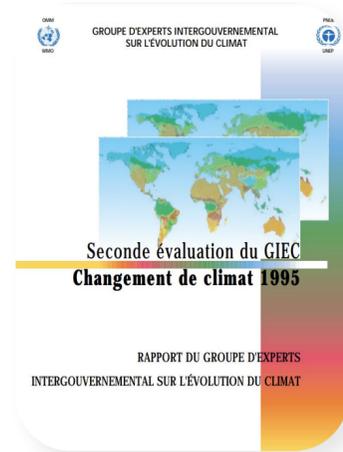
Mais si c'est pas de la météo, ça sert à quoi ?



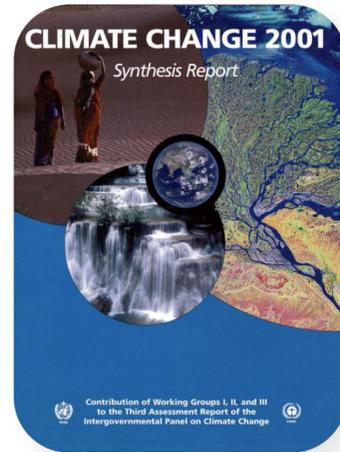
Plus de 30 ans de modélisation du climat



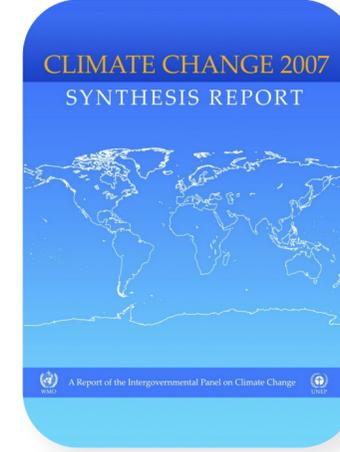
1^{er} rapport du GIEC



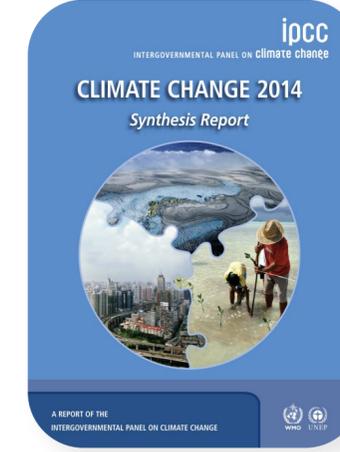
2^{ème} rapport du GIEC



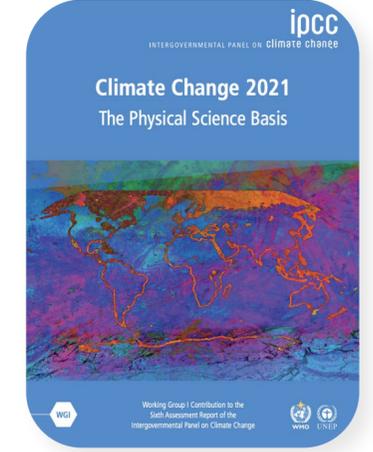
3^{ème} rapport du GIEC



4^{ème} rapport du GIEC



5^{ème} rapport du GIEC



6^{ème} rapport du GIEC

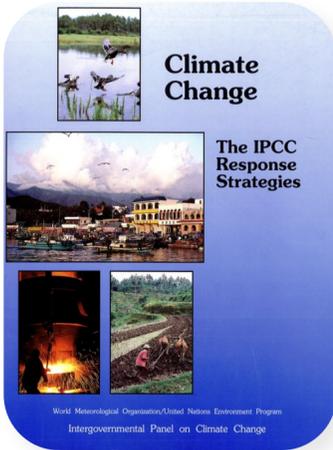


Coupled
Model
Intercomparison
Project
(CMIP)

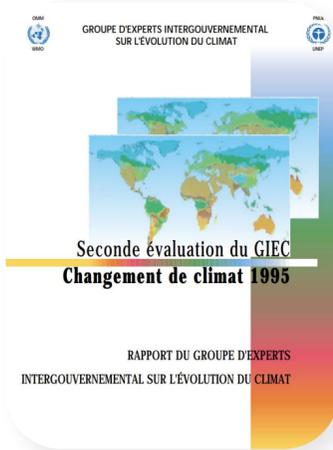


COordinated
Regional climate
Downscaling
EXperiment
(CORDEX)

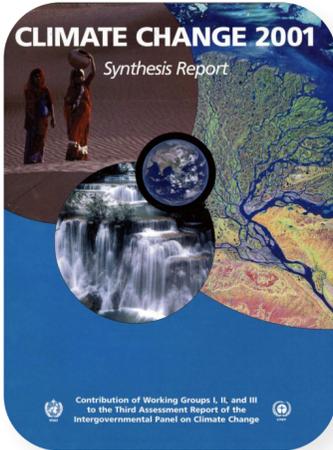
Une complexité et un nombre croissant de processus



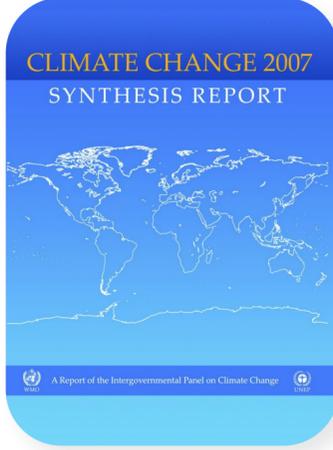
1^{er} rapport du GIEC



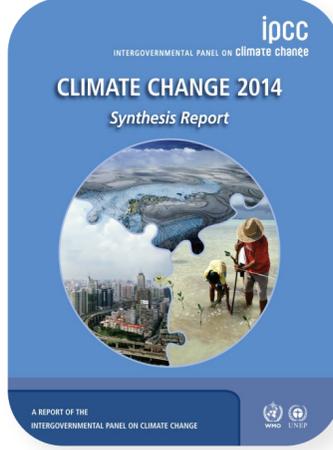
2^{ème} rapport du GIEC



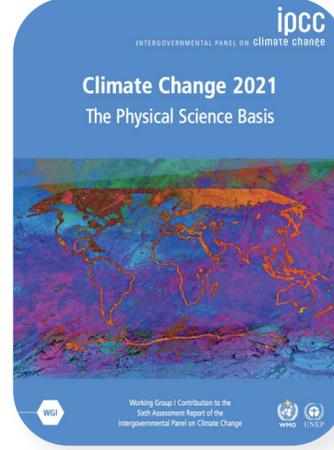
3^{ème} rapport du GIEC



4^{ème} rapport du GIEC



5^{ème} rapport du GIEC



6^{ème} rapport du GIEC



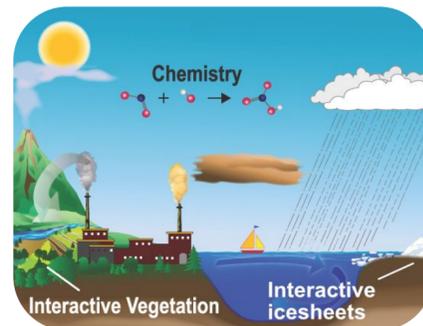
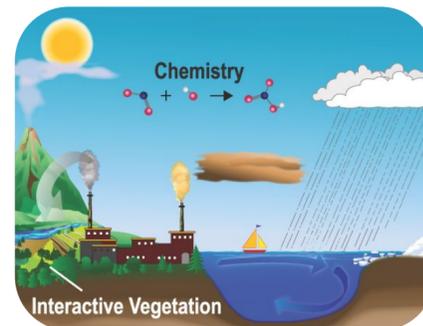
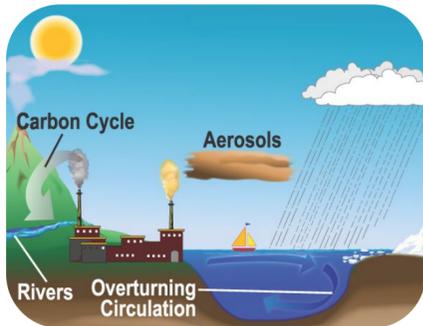
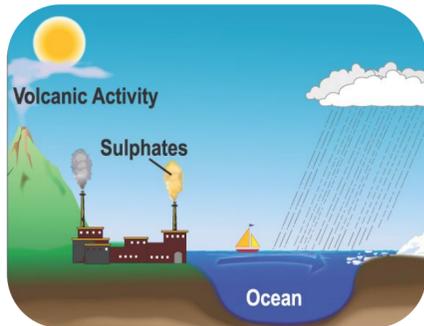
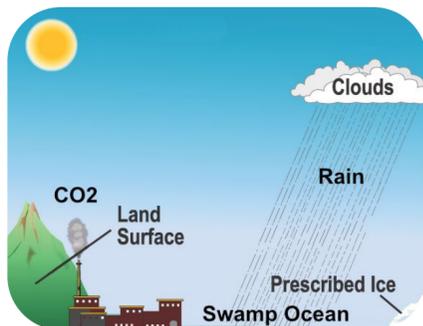
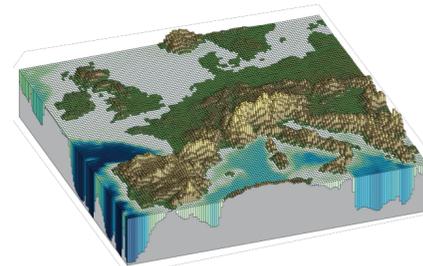
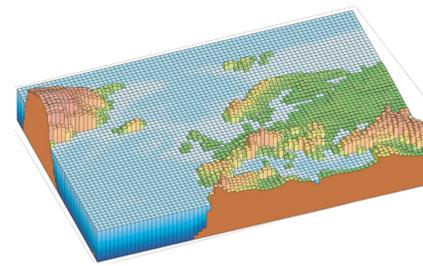
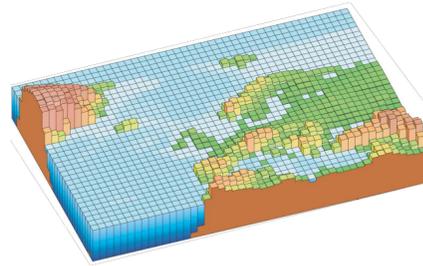
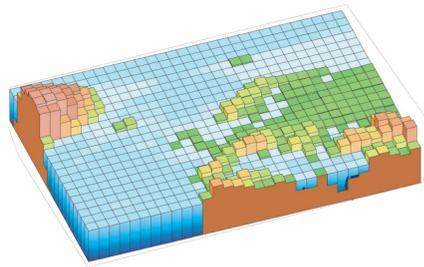
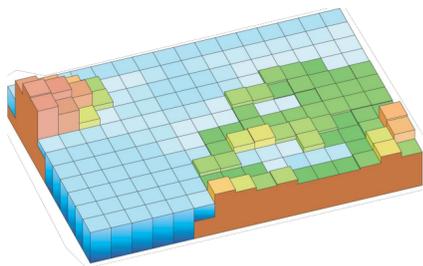
~450Km
~15 niveaux

~450Km
~20 niveaux

~200Km
~40 niveaux

~200Km
~50 niveaux

~50Km
~70 niveaux

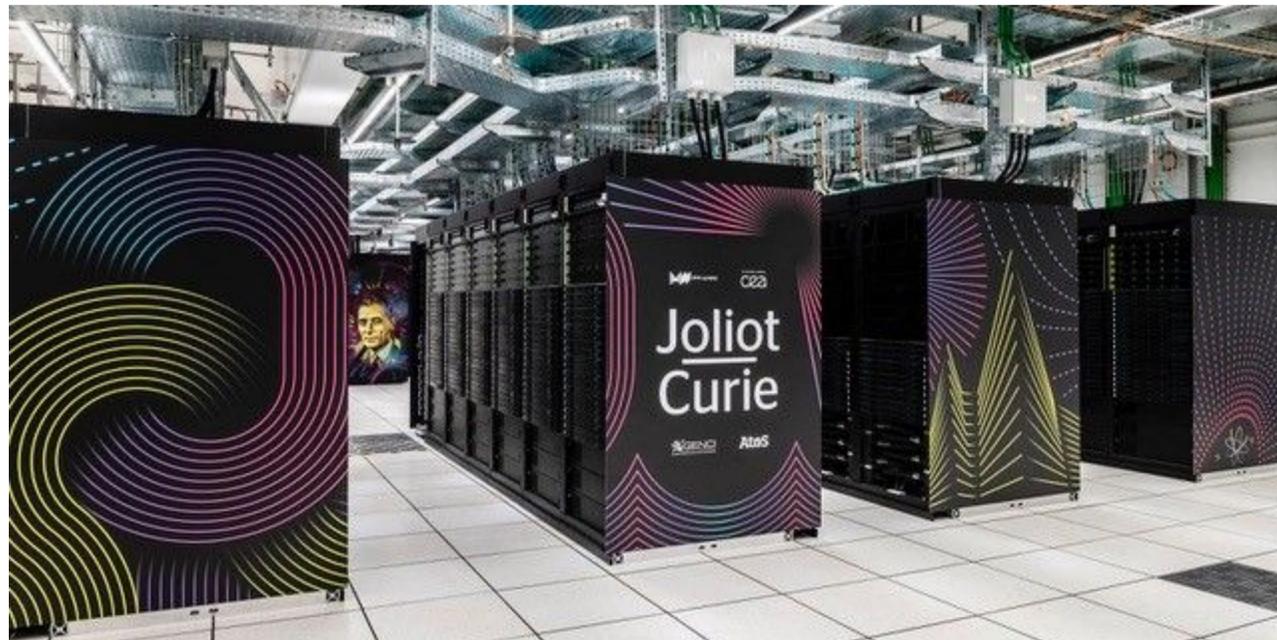


Une compréhension du système climatique très gourmande en calcul !



Production IPSL CMIP6 réalisée sur le Très Grand Centre de Calcul (TGCC) du CEA

- ~3 800 simulations réalisées
- ~50 000 années simulées
- ~100M heures.CPU (~115 ans de calcul à temps plein sur un CPU → ~1,5 ans sur supercalculateur)





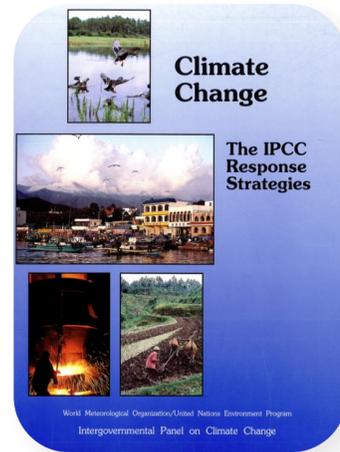
02

Une déferlante
de pétaoctet

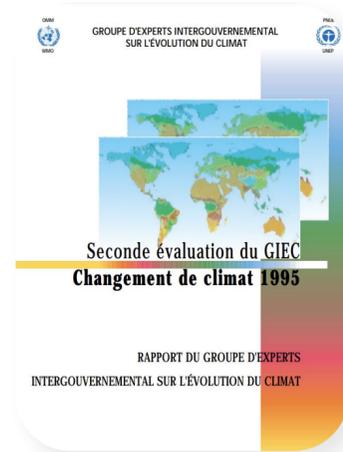
“Tu n’arrives pas à raisonner en quatre dimensions.”

Dr. Emmett Brown

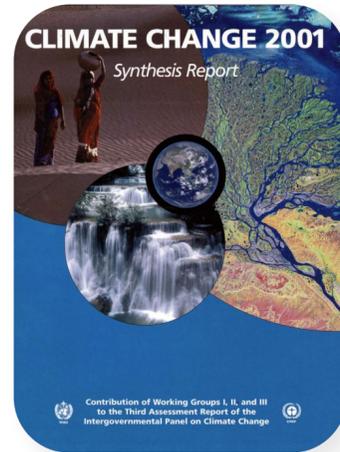
Le deuxième effet "KissCool"...



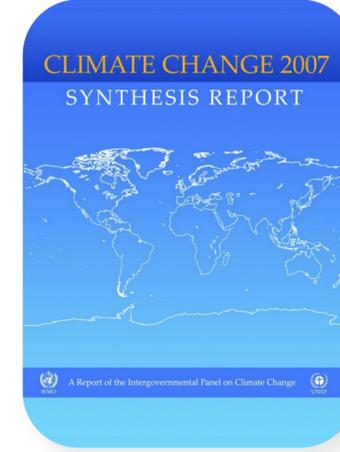
1^{er} rapport du GIEC



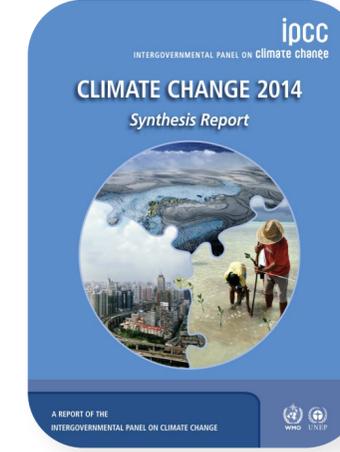
2^{ème} rapport du GIEC



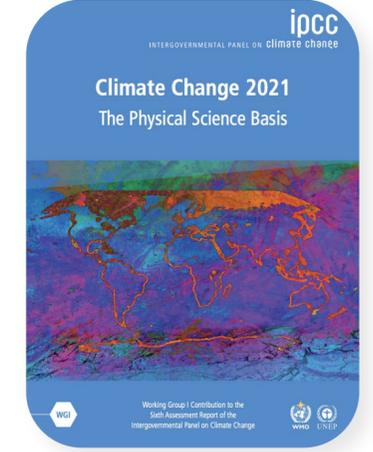
3^{ème} rapport du GIEC



4^{ème} rapport du GIEC



5^{ème} rapport du GIEC



6^{ème} rapport du GIEC



24 modèles
x 12 expériences
= ~40 Téraoctets

63 modèles
x 101 expériences
= ~1,5 Pétaoctets

128 modèles
x 312 expériences
= ~14,5 Pétaoctets



Les **ensembles de simulations** échantillonnent les états initiaux, les forçages et la physique (e.g., la simulation du climat de 1850 à 2014 de CMIP6 doit être réalisée au minimum 10 fois par modèle).

Les **ensembles multi-modèles** assurent que tous les modèles aient réalisés certaines simulations obligatoires (e.g., scénarios futurs).

La modélisation du climat dans le monde : un esprit d'équipe



+ 50 groupes de modélisation à travers le monde ont participé à CMIP6





Votre point de départ :

pcmdi.llnl.gov/mips

- Documentation modèles et expériences
- Sorties de modèles
- ToU et login
- Accès
- Organisation CMIP & ESGF

Program for Climate Model Diagnosis & Intercomparison

ENHANCED BY Google



[Home](#) [About](#) [Research](#) [CMIP6](#) [MIPs](#) [Publications](#) [Software](#) [CMIP Data \(ESGF Portal\)](#)

Model Intercomparison

AMIP - Atmospheric Model Intercomparison Project

AMIP is a standard experimental protocol for global atmospheric general circulation models (AGCMs).

CMIP - Coupled Model Intercomparison Project

The Coupled Model Intercomparison Project (CMIP) studies output from coupled ocean-atmosphere general circulation models that also include interactive sea ice.

CMIP2

The Coupled Model Intercomparison Project Phase II

CMIP3

The Coupled Model Intercomparison Project Phase III

CMIP5

The Coupled Model Intercomparison Project Phase V

CMIP6

The Coupled Model Intercomparison Project Phase VI

[Related Intercomparison Projects](#)

obs4MIPs

obs4MIPs

obs4MIPs (Observations for Model Intercomparisons Project) is an activity to make observational products more accessible for climate model intercomparisons

input4MIPs

input4MIPs

input4MIPs (input datasets for Model Intercomparison Projects) is an activity to make available via ESGF the boundary condition and forcing datasets needed for CMIP6. Various datasets are needed for the pre-industrial control (piControl), AMIP, and historical, simulations, and additional datasets are needed for many of the [CMIP6-endorsed model intercomparison projects \(MIPs\) experiments](#). Earlier versions of many of these datasets were used in the [5th Coupled Model Intercomparison Project \(CMIP5\)](#)



Lawrence Livermore National Laboratory
7000 East Avenue • Livermore, CA 94550

Operated by Lawrence Livermore National Security, LLC, for the
Department of Energy's National Nuclear Security Administration.



... et plongeons dans le vif du sujet !



Noeud ESGF français : <https://esgf-node.ipsl.upmc.fr>

Home Technical Support

You are at the [ESGF-NODE.IPSL.UPMC.FR](https://esgf-node.ipsl.upmc.fr) node

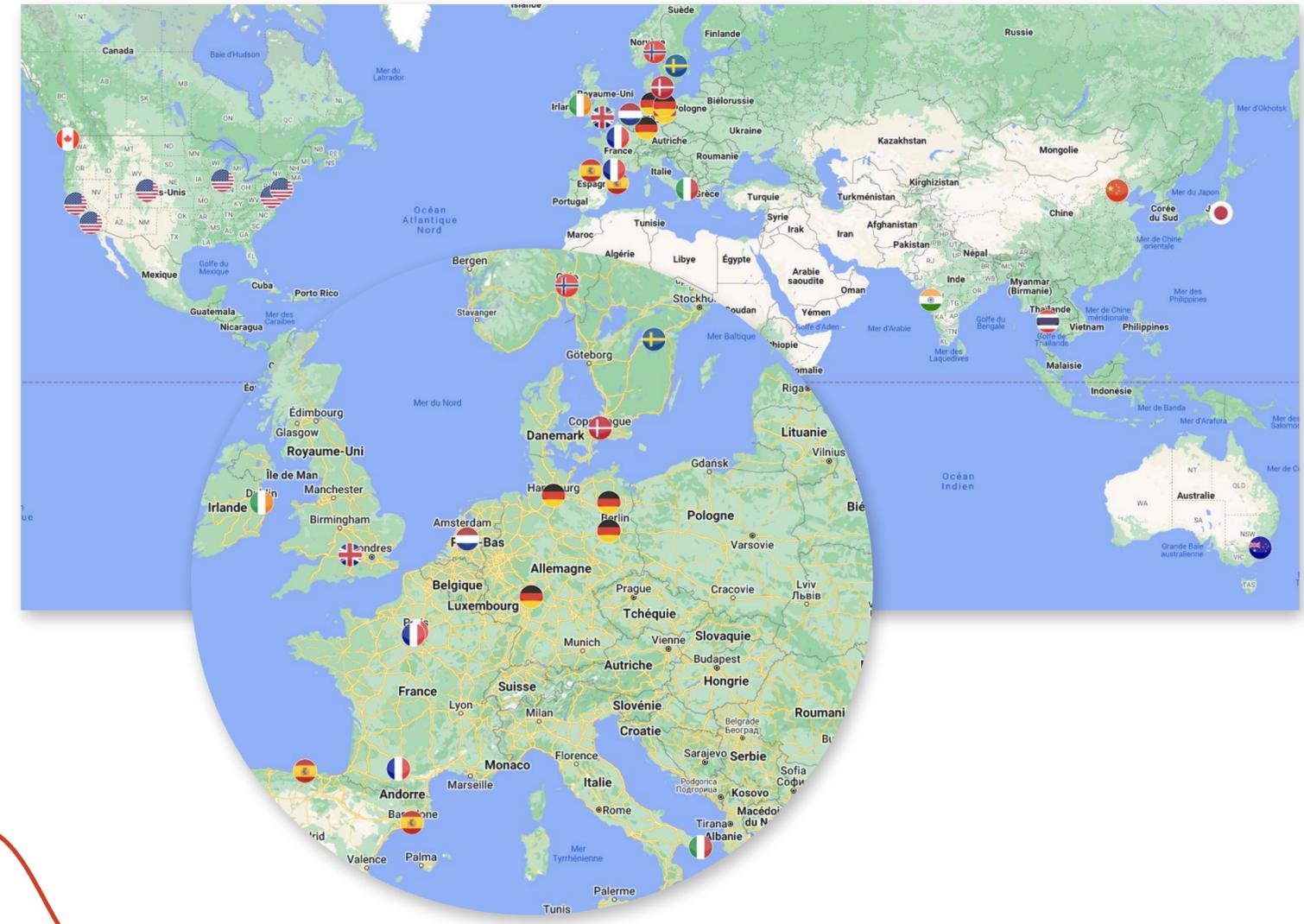
WARNING: Not all models include a variant "r11p1f1", and across models, identical values of variant_label do not imply identical variants. To learn which forcing datasets were used in each variant, please check modeling group publications and documentation provided through ES-DOC.

Enter Text: Search Reset Display results per page [\[More Search Options \]](#)

Search Constraints: IPSL Show All Replicas Show All Versions Search Local Node Only (Including All Replicas)

Total Number of Results: 1349970
-1- 2 3 4 5 6 Next >>
Please login to add search results to your Data Cart
Expert Users: you may display the search URL and return results as XML or return results as JSON

1. **CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.intvaw.gr**
Data Node: vesg.ipsl.upmc.fr
Version: 20190110
Total Number of Files (for all variables): 1
Full Dataset Services: [\[Show Metadata \]](#) [\[List Files \]](#) [\[THREDDS Catalog \]](#) [\[WGET Script \]](#) [\[Show Citation \]](#) [\[PID \]](#) [\[Further Info \]](#)
2. **CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.treeFracNdlDcd.gr**
Data Node: vesg.ipsl.upmc.fr
Version: 20190110
Total Number of Files (for all variables): 1
Full Dataset Services: [\[Show Metadata \]](#) [\[List Files \]](#) [\[THREDDS Catalog \]](#) [\[WGET Script \]](#) [\[Show Citation \]](#) [\[PID \]](#) [\[Further Info \]](#)
3. **CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.treeFracNdlEvg.gr**
Data Node: vesg.ipsl.upmc.fr
Version: 20190110
Total Number of Files (for all variables): 1
Full Dataset Services: [\[Show Metadata \]](#) [\[List Files \]](#) [\[THREDDS Catalog \]](#) [\[WGET Script \]](#) [\[Show Citation \]](#) [\[PID \]](#) [\[Further Info \]](#)
4. **CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppC-amv-ExTrop-pos.r13i1p1f1.Emon.treeFracBdlDcd.gr**
Data Node: vesg.ipsl.upmc.fr
Version: 20190110
Total Number of Files (for all variables): 1
Full Dataset Services: [\[Show Metadata \]](#) [\[List Files \]](#) [\[THREDDS Catalog \]](#) [\[WGET Script \]](#) [\[Show Citation \]](#) [\[PID \]](#) [\[Further Info \]](#)



Chaque item résultat est un "jeu de donnée".

Les clés du vocabulaire contrôlé utilisées pour l'interface de recherche sont aussi appelées "facets".

L'**Earth System Grid Federation (ESGF)** est un **réseau de serveurs décentralisé** et fédéré avec une collaboration internationale. C'est un système de nœuds distribués qui **partagent dynamiquement leurs catalogues (ou index) locaux**. C'est la base de données **complète** des données de référence en modélisation du climat.



climate4impact.eu

is-enes  Exploring climate model data
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Home Data Discovery Help Feedback Register Login

PROJECT: CMIP 6 NODES & SUBSETTING SCOPED VIEW EXTENDED VIEW

VARIABLE

FREQUENCY

EXPERIMENT

MODEL

MEMBER

Temperature

- ta - Air temperature (166375)
- tas - Temperature (153950)
- tasmax - Max. Temperature (101151)
- tasmin - Min. Temperature (100839)

Precipitation

- pr - Precipitation (150622)
- prsn - Snow (82460)
- prc - Convective precipitation (54553)

Humidity

- huss - Specific humidity (90063)
- hurs - Rel. Humidity (58694)
- rhsm - Min, Rel. Humidity (-)
- rhs - Rel. Humidity (-)
- hus - Spec. Humidity (113287)
- hur - Rel. Humidity (72564)

Wind

- sfcWind - Wind (111198)
- uas - Eastward wind (103651)
- vas - Northward wind (102942)
- sfcWindmax - Max Wind (57235)

Radiation

- rsds - SW Radiation Dn (120295)
- clt - Cloud (109820)
- rlds - LW Radiation Dn (100112)
- rsus - SW Radiation Up (76017)
- rlus - LW Radiation Up (74732)
- rsdsdiff - Diff. Radiation (12611)

Pressure

- psl - Sea level pressure (156559)
- ps - Pressure (103356)
- pfull - Pressure (11849)

Evaporation

- evspsbl - Act. Evap. (68406)
- evpsblpot - Pot. Evap. (-)
- evspsblsol - Sol Evap. (29777)
- evspsblveg - Canopy Evap. (28396)

OPEN YOUR NOTEBOOK SCIENTIFIC PRESETS SUBSETTING DOWNLOAD

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084



Contrôle des standards et des métadonnées

Vérification, validation et modification des (méta) données selon la **convention "Climate and Forecast" (CF):**

- ✓ **Définit le format de fichier et ses règles de structuration** (metadonnée + donnée)

Contrôle des nomenclatures, des arborescence et versionnement

Respect des **spécifications des projets** (e.g., CMIP, CORDEX), appelées "Data Reference Syntaxes" (DRS) qui encadrent notamment :

- ✓ **le nommage des fichiers,**
- ✓ **leur organisation/arborescence,**
- ✓ **le versionnement des jeux de données.**

Contrôle des vocabulaires

CMIP/CORDEX imposent l'utilisation de **termes prédéfinis et autorisés** qui ont été **présélectionnés par la communauté scientifique**. Seuls les termes approuvés, souvent définis comme des paires clé/valeur, peuvent être utilisés pour formater et publier les données.

Contraintes

Les utilisateurs doivent connaître le vocabulaire pour naviguer.

Pattern

`<variable>_<table_id>_<source_id>_<experiment_id>_<variant_label>_<grid_label>[_<period_start>--<period_end>].nc`

Exemple

`tas_Amon_IPSL-CM6A-LR_historical_r1i1p1f1_gr_185001-201412.nc`



Network Common Data Form (netCDF)

Est un format de fichier qui permet de stocker des **données scientifiques multidimensionnelles** (sous forme de tableau de données) et de façon "**auto-documentée**" c'est-à-dire que chaque fichier dispose d'une en-tête qui décrit la disposition des données dans le reste du fichier.

(exemple ci-contre)

Conséquence directe

Données et métadonnées sont dans le même fichier !

Donnée = valeur(s) numérique(s) d'une quantité/information physique/scientifique.

Métadonnée = tout le reste ! (l'entête et le nom du fichier, l'arborescence, la documentation des modèles, l'errata, citation, etc.)

Vocabulaire contrôlé = sous-ensemble de métadonnées dédié à la gestion et à l'accès aux données (organisation des fichiers, standardisation, interfaces utilisateur, etc.)

```
netcdf tas_Amon_HadGEM3-GC31-LL_historical_rlilplf3_gn_185001-194912 {
dimensions:
    time = UNLIMITED ; // (1200 currently)
    bnds = 2 ;
    lat = 144 ;
    lon = 192 ;
variables:
    double time(time) ;
        time:units = "days since 1850-01-01" ;
        time:calendar = "360_day" ;
        time:standard_name = "time" ;
[...]
    double lat(lat) ;
        lat:units = "degrees_north" ;
        lat:standard_name = "latitude" ;
[...]
    double lon(lon) ;
        lon:units = "degrees_east" ;
        lon:standard_name = "longitude" ;
[...]
    float tas(time, lat, lon) ;
        tas:standard_name = "air_temperature" ;
        tas:units = "K" ;
[...]
// global attributes:
    :experiment_id = "historical" ;
    :frequency = "mon" ;
    :further_info_url = "https://furtherinfo.es-doc.org/CMIP6.MOHC.HadGEM3-GC31-LL.historical.none.rlilplf3" ;
    :grid = "Native N96 grid; 192 x 144 longitude/latitude" ;
    :institution = "Met Office Hadley Centre, Fitzroy Road, Exeter, Devon, EX1 3PB, UK" ;
    :institution_id = "MOHC" ;
    :nominal_resolution = "250 km" ;
    :source_id = "HadGEM3-GC31-LL" ;
    :realm = "atmos" ;
    :tracking_id = "hdl:21.14100/b6959414-d5ed-4cd9-a627-59238e52132d" ;
    :variable_id = "tas" ;
    :variant_label = "rlilplf3" ;
[...]
data:

tas =
    248.884, 248.8906, 248.8589, 248.8403, 248.8538, 248.8433, 248.8284,
    248.8313, 248.8093, 248.811, 248.7976, 248.7747, 248.77, 248.7769,
    248.7715, 248.7466, 248.7534, 248.759, 248.7498, 248.7473, 248.7329,
    248.729, 248.7227, 248.7249, 248.7302, 248.7314, 248.7395, 248.7363,
    248.7188, 248.7231, 248.7283, 248.7334, 248.7273, 248.7517, 248.7434,
    248.7356, 248.7314, 248.7427, 248.7412, 248.7339, 248.7146, 248.7373, [...]
```



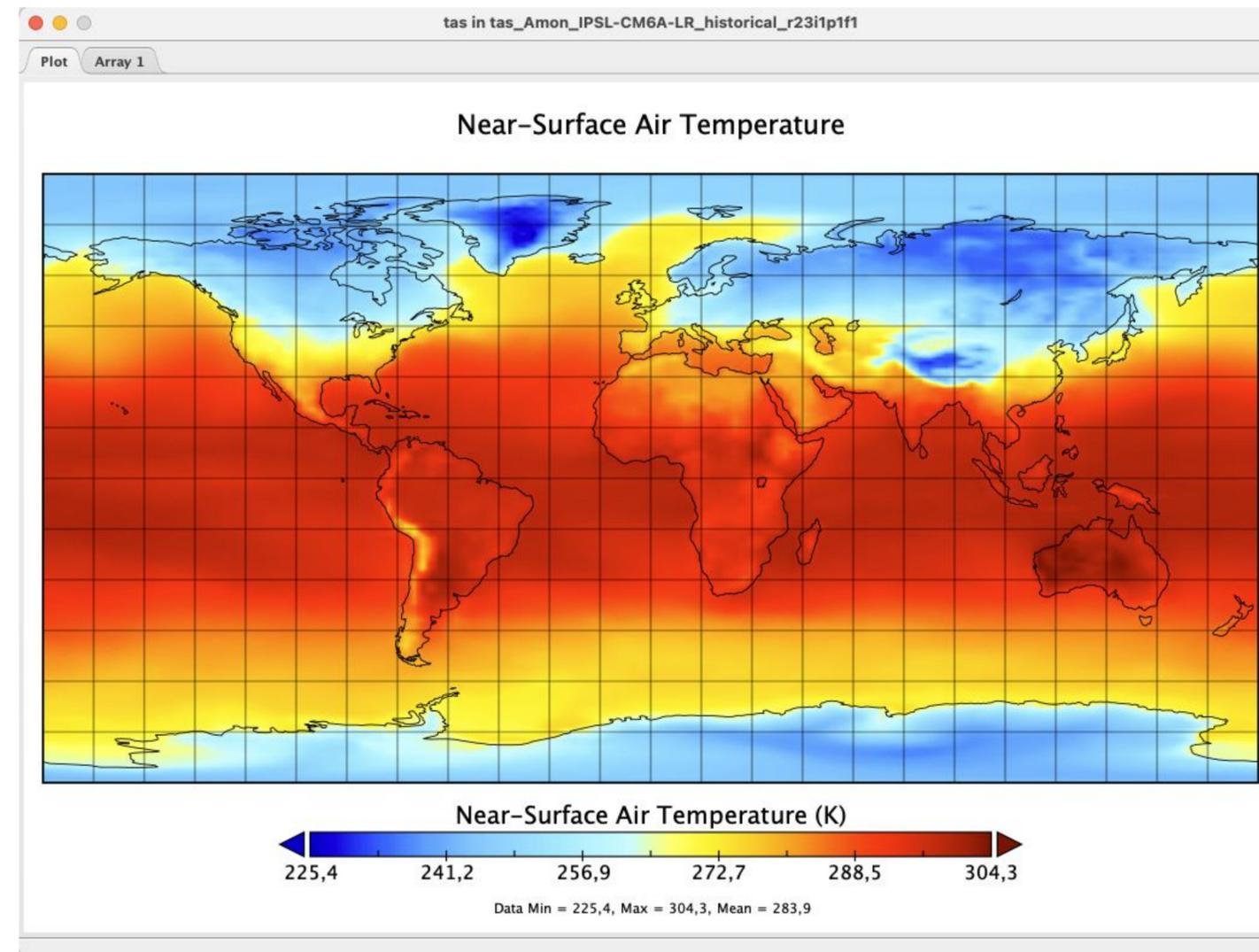
Les données issues des simulations climatiques ayant par construction une architecture multidimensionnelle, il est très compliqué de les représenter dans les formats de fichier de la bureautique classique (e.g. Excel, CSV, etc.).

netCDF s'accompagne donc d'un ensemble de bibliothèque logicielles qui permettent d'ouvrir et d'exploiter ces fichiers dans différents langages (Python, R, Unix, etc.) ou via des outils dédiés..

Pour une première approche, préférez l'outil graphique Panoply :

giss.nasa.gov/tools/panoply/

Le portail Climate4Impact propose aussi des outils de visualisation en ligne.



Pour les plus avertis (courageux ?) d'entre vous



Un sous-ensemble de données CMIP6 a également été dupliqué sur les clouds privés Amazon (AWS) et Google. Le format de stockage utilise un autre paradigme que celui des “fichiers” que nous connaissons tous. L'avantage est que ce sous-ensemble CMIP6 est exploitable à distance sans avoir à télécharger la donnée, directement depuis un interpréteur Python (e.g., Notebook) :

<https://pangeo-data.github.io/pangeo-cmip6-cloud>

```
import intake

# for Google Cloud:
col = intake.open_esm_datastore("https://storage.googleapis.com/cmip6/pangeo-cmip6.json")
# for AWS S3:
#col = intake.open_esm_datastore("https://cmip6-pds.s3.amazonaws.com/pangeo-cmip6.json")

col
```

```
# form query dictionary
query = dict(experiment_id=['historical', 'ssp245', 'ssp585'],
             table_id='Amon',
             variable_id=['tas'],
             member_id = 'r1i1p1f1',
             grid_label='gr')

# subset catalog and get some metrics grouped by 'source_id'
col_subset = col.search(require_all_on=['source_id'], **query)
col_subset.df.groupby('source_id')[['experiment_id', 'variable_id', 'table_id']].nunique()
```

- ✓ Pas besoin de ressources locales (CPU ou To)
- ✓ Agnostique de la localisation
- ✓ Agnostique de l'arborescence
- ✓ Un système de requêtes natif
- ✓ Sauvegardez vos propres catalogues.
- ✓ Agrégations natives (temps, membres).





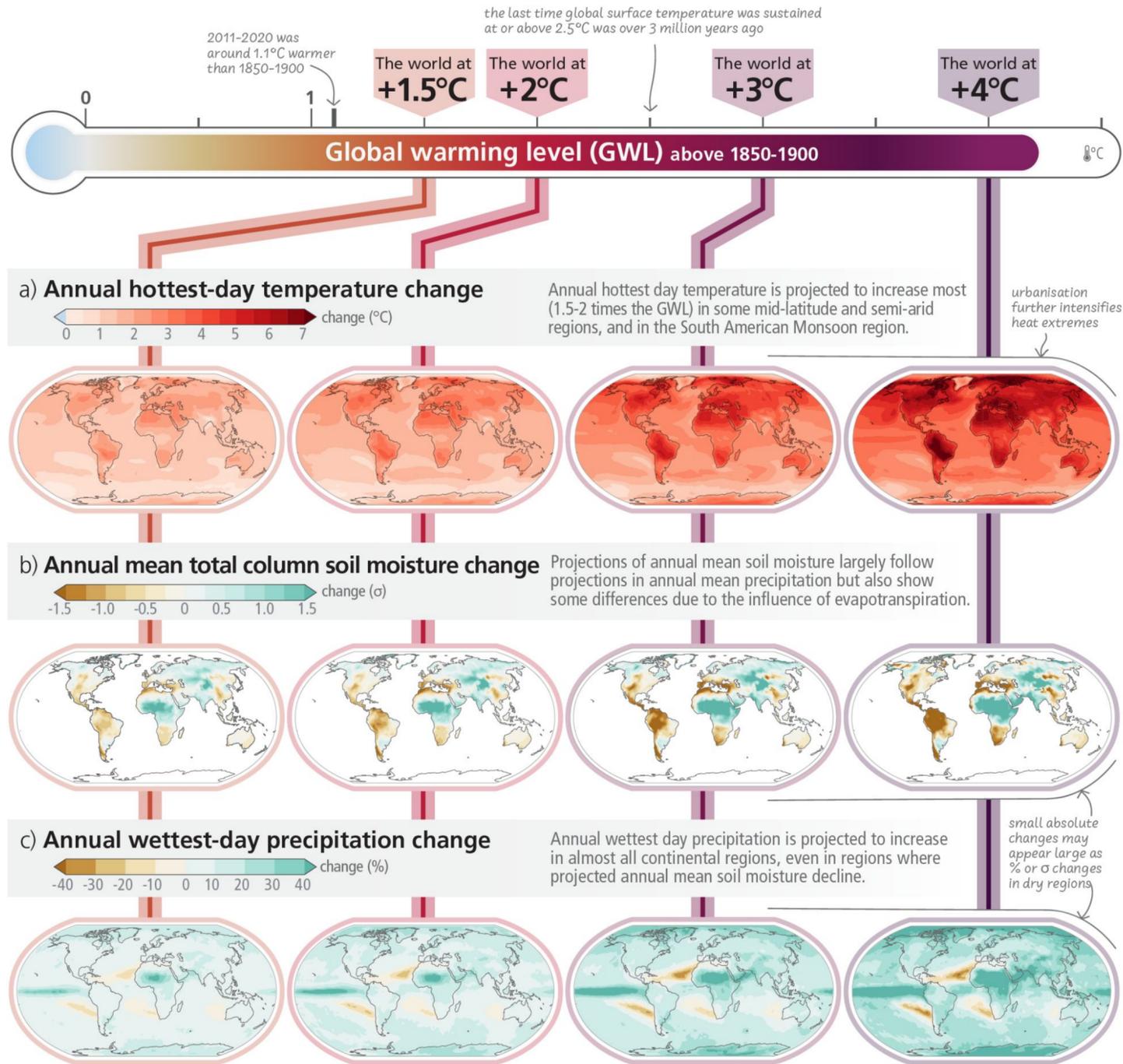
03

Vers le 7^{ème}
rapport du GIEC

“Un grand pouvoir implique de grandes responsabilités.”

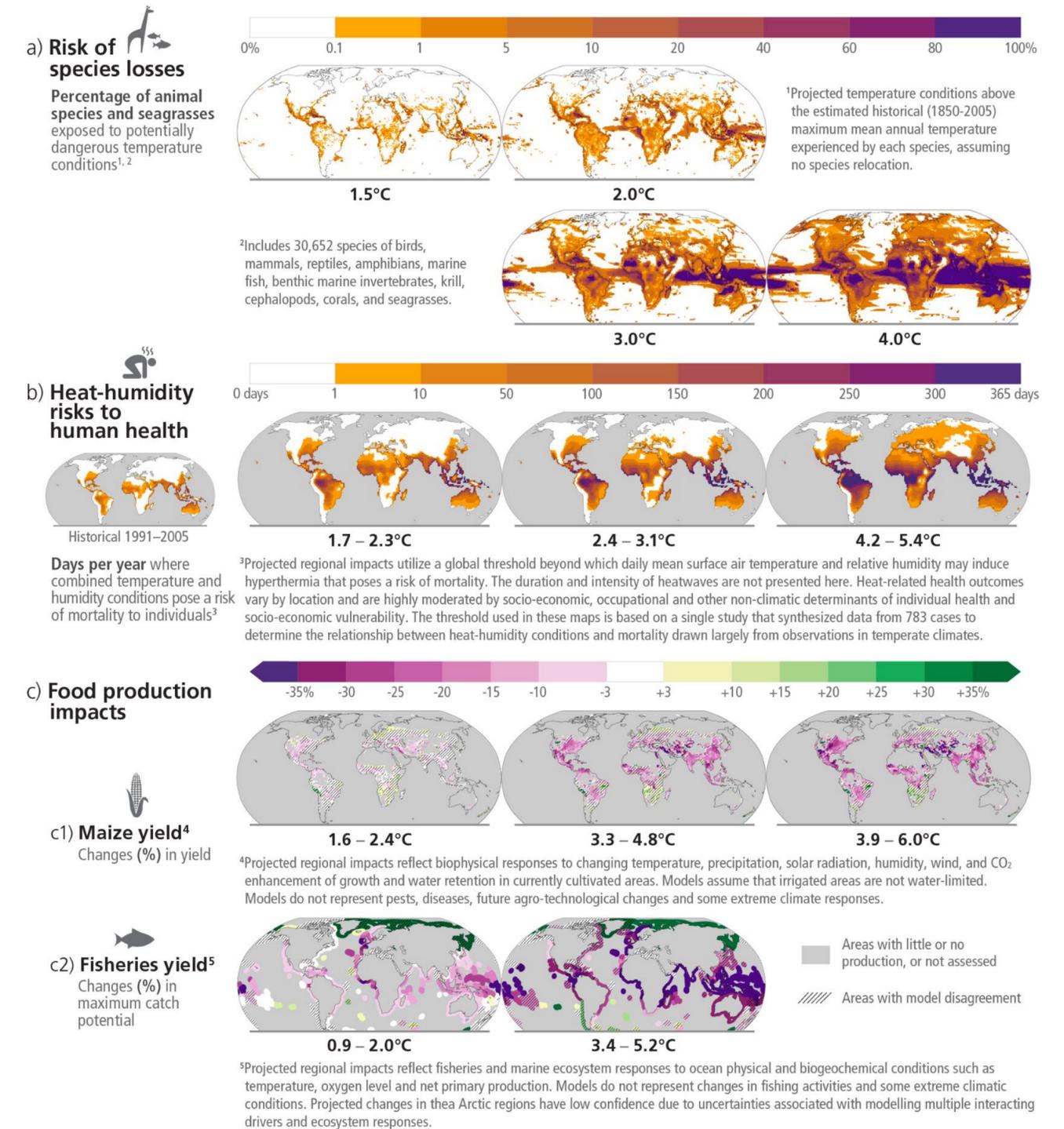
Peter Parker

Le climat ne déplace pas les montagnes... mais les humains assurément !

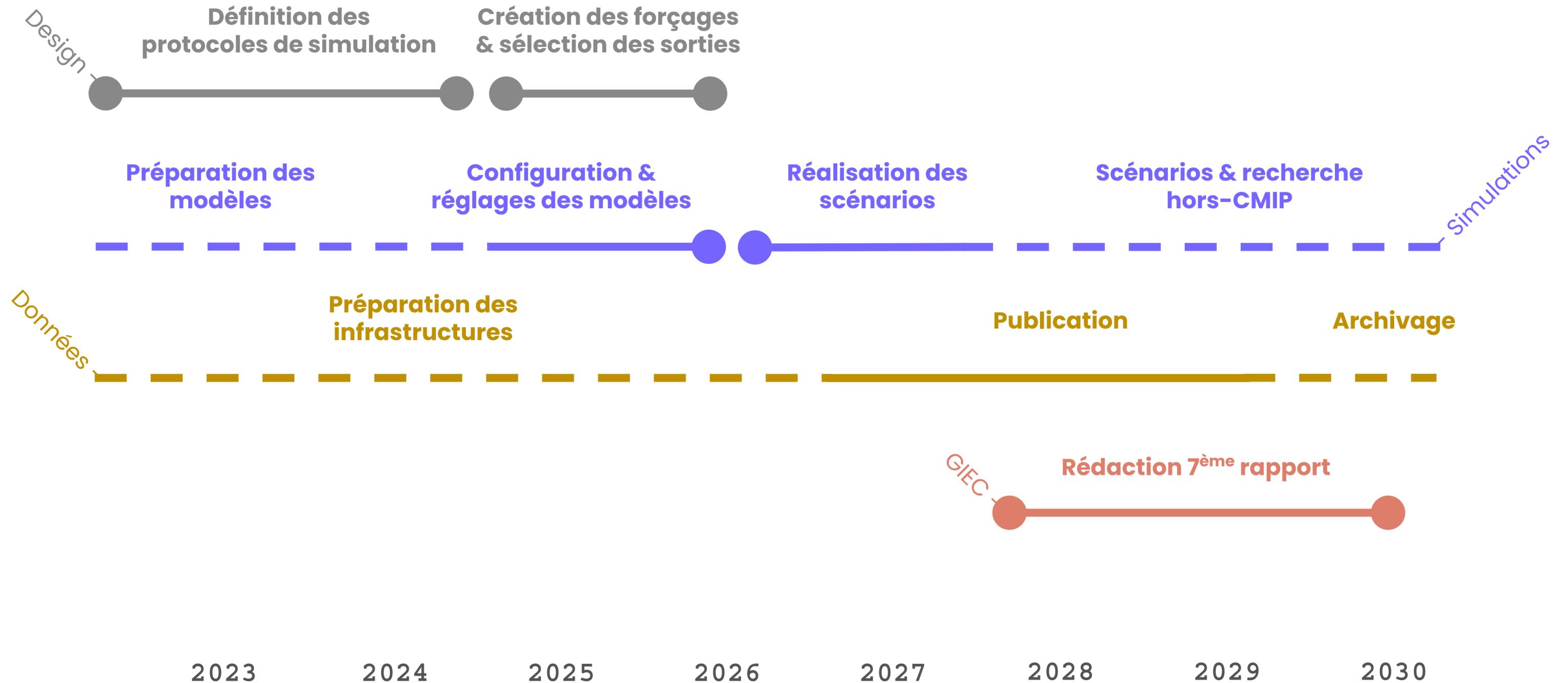


Future climate change is projected to increase the severity of impacts across natural and human systems and will increase regional differences

Examples of impacts without additional adaptation



Source: IPCC AR6





MERCI – THANK YOU

Institut Pierre-Simon Laplace (IPSL)