

Réunion Tech

- Production CMIP6
 - Publications CMIP6
 - Ressources et site Web
 - Discussion sur les workflow
 - CliMAF
 - C-ESM-EP
 - Eclis
 - dr2xml
-

14 novembre 2019

Production CMIP6

- Liste des champs fixes manquants à finaliser et à générer si besoin.
- Variable vt100 :
 - Version 1.00.31 de la data request disponible avec des modifications sur la variable vt100 mais a priori pas finalisées
- Liste des variables de glace non pondérées et éventuellement boguées en attente (cf diapo Errata Service)
- L'expérience amip-piForcing a été relancée avec la bonne date d'initialisation (1870 et non 1980)
- Le parent_variant_label est erroné pour certaines simulations pour lesquelles $r > 1$ (cf diapo Errata Service)
- Simulations HR en cours, post-traitements en cours ou terminés :
 - DECK terminé
 - Post-traitement en cours : AMIP et 1 scenario
- 4 simulations LS3MIP mal définies dans la DR (pas de nom différent pour les simulations quand les scénarios sont différents). Décision unilatérale de leur côté pour f1 → ssp126 / f2 → ssp585. Utilisation de f2 pour ssp585 et f21 pour ssp126 côté CNRM-Cerfacs.

Publications CMIP6 ESGF

- Publication CMIP6 réalisée à ce jour sur 24 nœuds. Détails sur l'ESGF et sur la page de monitoring : https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/
- Nouvelles versions de la pile logicielle ESGF (v4.0.3 et v4.0.4) :
 - * montée de niveau 4.0.3 (via Ansible) réalisée sur le nœud de production esg1 du CNRM le 01/10 (nouveau composant de publication)
 - * installation nouvelle release 4.0.4 sur le nœud de test esg2 du CNRM le 12/11 (publication CMIP6 ok, Cordex à tester encore). Après validation complète, cette nouvelle release sera installée sur le nœud de production esg1
- Reprise des réunions régulières bimensuelles du CDNOT (CMIP Datanode Operations Team) le 13/11 (dernière réunion tenue le 26/06).
- L'outil de statistiques fourni par la pile logicielle ESG ne fonctionne pas. Côté CNRM, les statistiques sont faites via des log maison. Les log ESG sont gardées pour pouvoir refaire les statistiques a posteriori.
- Meeting annuel ESGF F2F (Face2Face) organisé pour la 1ère fois en Europe : 23 au 25/03/20 sur la Météopole (couplé à l'AG IS-ENES3 prévue du 25 au 27/03). Co-organisation Météo-France pour ces 2 événements.
Informations disponibles pour les trois meeting sous : <https://is.enes.org/events/workshops>
Communication interne CNRM à faire sur ces événements.
- Quelques problèmes d'accès et de récupération des données via le site ESGF IPSL liés à des coupures électriques sur le site. Possibilité d'utiliser d'autres Index Node avec conseil d'avoir plusieurs openid sur des nœuds différents pour pallier ce type de problème (deux login différents nécessaires).
- Faire un rappel par mail concernant l'occupation scratch et home. Une fois les simulations terminées et publiées, il faut effacer le répertoire assemble sur le scratch et le répertoire de listing sur le home.

Publication ESGF au CNRM

■ Bilan publication CNRM au 13/11 :

- cf /cnrm/est/USERS/franchisteguy/CMIP6/CNRM/Publication/SuiviPubli/SuiviPubliGlobal.ods
- **258** bons à publier émis (+ 3 extensions 2100-2300 membres r1 ScenarioMIP ESM2-1)
- **233** simulations publiées (~325 To) soit ~75 % du nombre total de simulations à publier sur l'exercice. 14 MIPs dorénavant couverts côté CNRM-CERFACS
- Depuis la dernière réunion Tech (24/09) :
 - * dépublication amip-piForcing et aqua-p4K (rerun aqua-p4K republié et amip-piForcing en cours de publication)
 - * dépublication ESM2-1 r1i1p1f2 ssp370 et ssp585 puis republication des 2 rerun correspondants
 - * publication membres supplémentaires DECK CM6-1 historical (10 membres faits sur 15)
 - * poursuite publication simulations LS3MIP et LUMIP
 - * publication variables '3hr' CM6-1_ssp585 (ps, uas, vas) supplémentaires
 - * début publication DECK HR : historical (reste 2 tables) et piControl en cours de publication

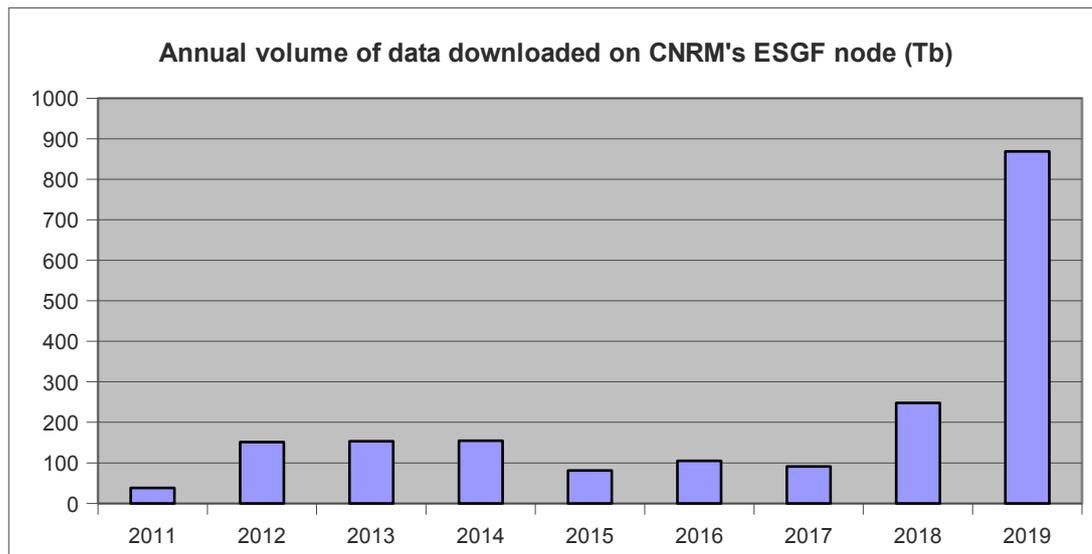
■ A venir côté CNRM (par ordre de priorité décroissant, **à valider**) :

- * simulations LS3MIP et LUMIP (7)
- * simulations DECK HR r1i1p1f2 (historical, piControl, 1pctCO2, abrupt-4xCO2)
- * fin publication membres supplémentaires DECK CM6-1 historical (reste 5 membres)
- * publication 5 membres supplémentaires DECK ESM2-1 historical
- * simulation HR amip-hist (GMMIP)
- * extension '2300' membres r1 ScenarioMIP ESM2-1
- * esm-hist r1i1p1f3 (modèle en émission)
- * complément DECK piControl LR (500 premières années seulement publiées) après émission bons à publier correspondants
- * (variables tables SIday et Simon après instruction définitive, cf [diapo Known Issues](#))
- * puis ordre chronologique d'arrivée des futurs bons à publier....

| Activity | |
|--------------------------|---------------------|
| <input type="checkbox"/> | AerChemMIP (14262) |
| <input type="checkbox"/> | C4MIP (2822) |
| <input type="checkbox"/> | CFMIP (3739) |
| <input type="checkbox"/> | CMIP (19845) |
| <input type="checkbox"/> | DAMIP (5161) |
| <input type="checkbox"/> | DCPP (20919) |
| <input type="checkbox"/> | GMMIP (1013) |
| <input type="checkbox"/> | GeoMIP (2038) |
| <input type="checkbox"/> | HighResMIP (209) |
| <input type="checkbox"/> | LS3MIP (382) |
| <input type="checkbox"/> | LUMIP (1548) |
| <input type="checkbox"/> | OMIP (189) |
| <input type="checkbox"/> | RFMIP (1858) |
| <input type="checkbox"/> | ScenarioMIP (20871) |

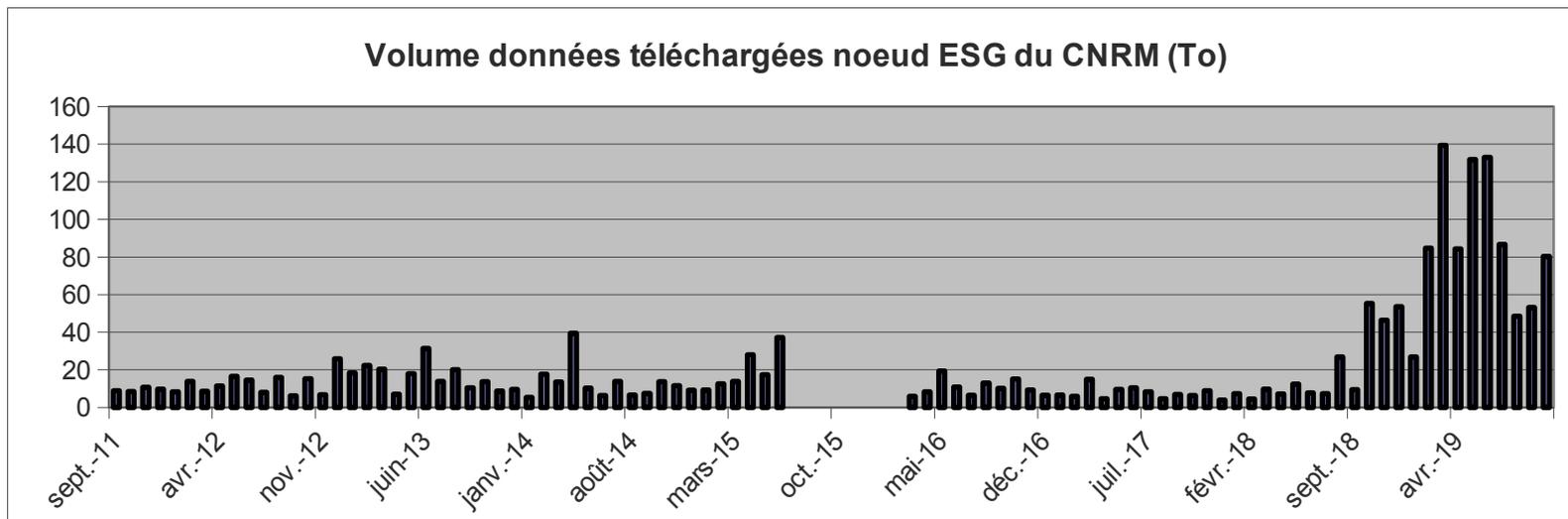
Nb dataset CNRM
sur l'ESGF (13/11)

Trafic noeud ESGF au CNRM (volume téléchargé)

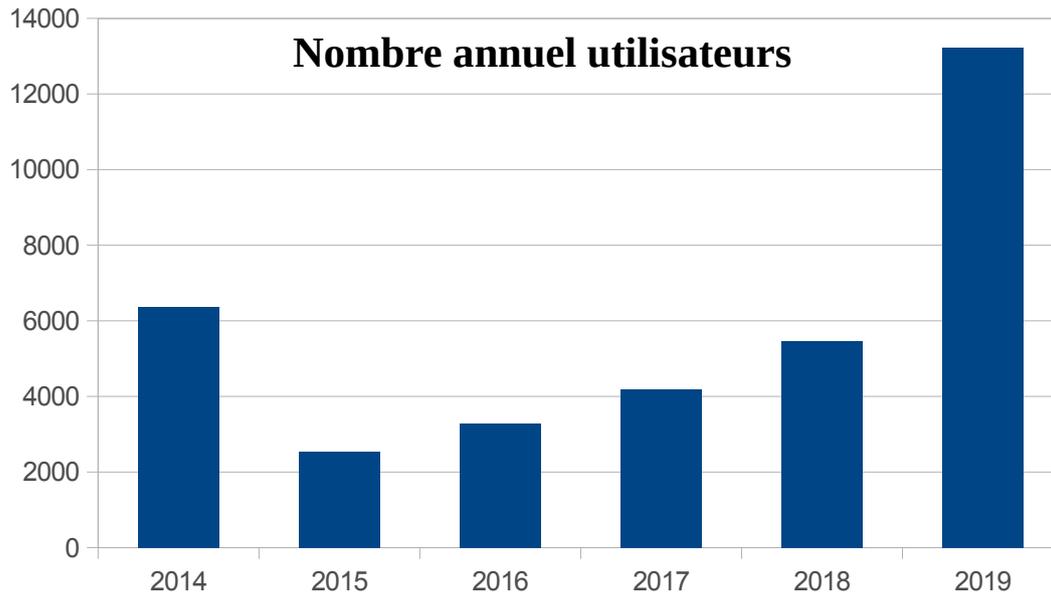


(01/11/19)

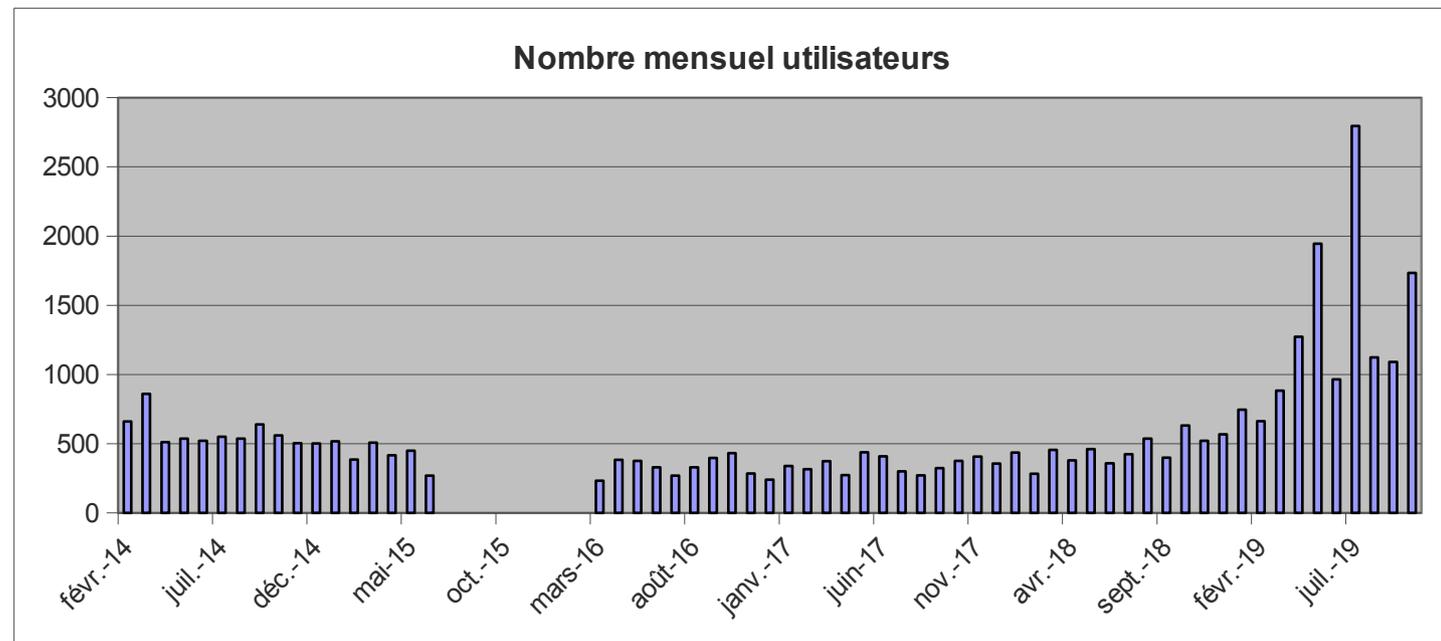
Volume 2019 provisoire téléchargé équivalent au volume 2011-2016



Trafic noeud ESGF au CNRM (nombre utilisateurs)



(01/11/19)



Errata Service, known issues

- variables **seaice** mal pondérées : feu vert David pour publication variables P1 (siu, siv, sithick, siconc, sisnthick et sivol) avec message de warning pour siu et siv à insérer sur site web CMIP6
https://docs.google.com/spreadsheets/d/1Mg0qgOYIGOtTs9yr7JNw_OLnBTCOXbyakqA_OJkSPHQ/edit#gid=0
Hiérarchisation de l'ordre de publication des données pour ces simulations à faire : DECK, scenarii...
- attribut **parent_variant_label** erroné pour certaines simulations ('variant_label' au lieu de 'parent_variant_label'). A faire encore : liste des simulations concernées, texte pour site web (rubrique 'known issues')
- Instruction à terminer suite à un retour utilisateur sur l'absence des champs lat_bnds / lon_bnds dans nos fichiers (attributs a priori présents pour grille native seulement, à confirmer encore). La variable aire des mailles existe dans le fichier et devrait suffire s'il ne s'agit que de pondérer par cette aire.
- Rappel : tableau (usage interne) recensement erreurs sous /cnrm/est/USERS/franchisteguy/CMIP6/CNRM/Production/Issues (InfoIssuesInterne.odt)

ESDOC

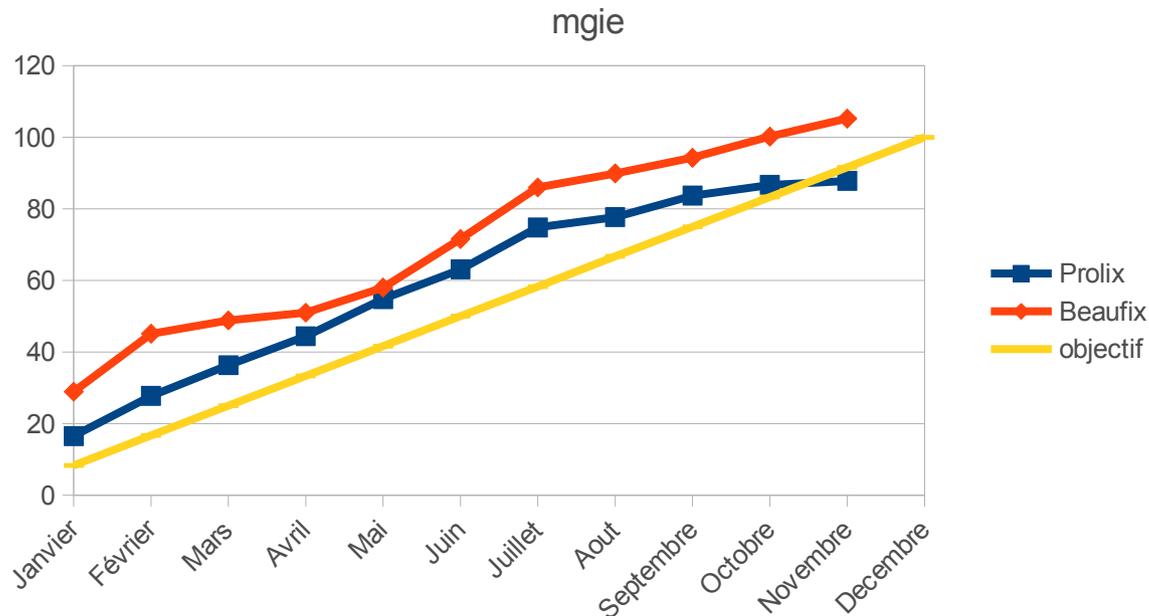
- ESDOC alimenté à ce jour par 10 centres, cf <https://explore.es-doc.org/cmip6/models>
- Coté CNRM, pour ESM2-1 :

| | landice | aerosol | atmoschem | land | coupling | atmos | ocean | ocnbgchem | seaice | toplevel | citation |
|-------------|---------|---------|-----------|------|-----------|-----------|-------------|-----------|--------|----------|----------|
| Fichier xls | - | Ok | Ok | Ok | Ok | Ok | Ok a priori | En cours | | | |
| EsDoc | - | Ok | Ok | Ok | A publier | A publier | | | | | |

- Toujours en attente de réponse de l'équipe esdoc sur certaines issues (cf <https://github.com/ES-DOC-INSTITUTIONAL/cnrm-cerfacs/issues>)
- A suivre :
 - 1) récolte puis vérification des fichiers restants pour esm2-1 (ocean, ocnbgchem, seaice, toplevel et citation), dépôt sur GitHub ESDOC et envoi à l'équipe ESDOC de vos remarques éventuelles sur les fichiers à renseigner
 - 2) extension ensuite à cm6-1 puis cm6-1-hr

Ressources calcul et stockage, site web

- (Rappel renouvellement calculateurs : arrêt Beaufix/Prolix prévu le 20/09/2020)
- Consommation calcul mgie (au 13/11) : quota Beaufix 2019 dépassé (105.2%) et il reste 1.7 Mh sur Prolix (NB : environ 24 Mh déjà consommées en queue Libre sur Beaufix sur mgie en 2019). 10 Mh relâchées par Mercator : fait-on remonter un besoin d'heures supplémentaires pour 2019 (réponse CTI/DSI avant le 15/11) ? [Mail envoyé \(14/11\) pour demander 2Mh supplémentaires.](#)
- Demande calcul 2020 à suivre : bilan des simulations CMIP6 restant à réaliser à consolider avant l'élaboration de cette demande. [Mail à faire pour que chacun fasse le bilan des simulations qui restent à réaliser.](#)



Workflow

Compte-rendu diffusé suite au Workshop « Building Reproducible Workflow for Earth Sciences » (14-16 octobre)

- Présentation de plusieurs workflow utilisés en recherche ou en opérationnel, PNT ou climat.
- Côté climat :
 - MetOffice : post-traitements CMIP6
 - CESM2 : workflow de réalisation des simulations (elles sont ensuite post-traitées)
- Quelques outils pour construire ou suivre l'exécution des workflow :
 - ecFlow : remplaçant de SMS utilisé pour suivre les simulations type Olive
<https://confluence.ecmwf.int/display/ECFLOW>
 - Cylc / Rose : équivalent de SMS/Olive (mais avec gestion de source)
<https://cylc.github.io/cylc/>
 - CIME : workflow python utilisé pour CESM2 et interfacé avec Cylc pour le suivi (autres interfaçages possibles) :
<http://esmci.github.io/cime>
- Majorité de systèmes codés en Python (portabilité)
- Compromis entre portabilité – reproductibilité – performance

CliMAF

- Formation :
 - Poursuite des boccas CliMAF
- Documentation :
 - La documentation et les exemples nécessitent une mise à jour (ex : utilisation de crm avec des pattern)
- Contributions à CliMAF :

CliMAF : Dernière version disponible (V1.2.12)

- Contenu :
 - Modifications pour le Cerfacs et l'IPSL
 - Modifications relatives à la C-ESM-EP
 - Corrections de bogues (dont lecture des champs fixes)
 - Ajout de fonctions permettant de ne pas refaire de sélection avec cdo (devrait résoudre en partie les problèmes de lenteurs) :
ccdo_fast, space_average_fast, time_average_fast, llbox_fast, annual_cycle_fast, clim_average_fast
 - PEP8
- Compatible avec la C-ESM-EP (v1 et v2)
- Disponible depuis le 11 septembre.

- Bogues connus :
 - Problème dans certains cas avec sellonlat : lié à cdo
 - Lenteurs dans les cas où les fichiers sont gros : utilisation des fonctions *_fast à généraliser
 - Problème dès que les calendriers utilisés ne sont pas standards : lié à la fonction check des dataset et aux comparaison de dates
 - Problème de lecture d'un champ fixe dans un projet (voir si cela fonctionne avec fds plutôt qu'avec la définition d'un projet)
 - Débogage nécessaire de la fonction check des dataset
 - [Problème avec cdo quand le nombre de fichiers à traiter est grand \(limitation cdo\)](#)

CliMAF : Développements

- V2.0 (mise à disposition visée : au plus tard mi-janvier) :
 - Compatibilité Python 3
 - Renommage de la fonction slice
 - Réécriture de mcd0.sh pour avoir plus de flexibilité
 - Mise à jour de la doc et des exemples
 - Rédaction de tests unitaires
 - Vérifier le fonctionnement de la création d'objets sans utilisation du cache
- Et pour plus tard (en particulier réunion du 13/11/19) :
 - Arrêter d'utiliser glob.glob (alternative nécessaire au moins pour les données CMIP6 et CORDEX côté IPSL et ailleurs)
 - Implémenter une méthode de suppression en fonction de la durée de création
 - Simplifier et déboguer les projets *_extent (récupération d'un historique sur une période, d'un scénario pour la suite)
 - Homogénéiser la sélection des données (ds, cdataset, cens, eds, ...)
 - Remplacer les scripts NCL par des scripts python :
 - ▶ Documentation PyNGL (disponible python2 et python3) pour faciliter le passage de NCL à python (un script écrit avec PyNGL est plus proche dans sa construction de NCL mais plus long en lignes de code que l'équivalent via cartopy ou matplotlib, mais il y a plus de fonctionnalités) :
 - Documentation : <https://www.dkrz.de/up/services/analysis/visualization/sw/pyngl-pynio>
 - Guide de transition : https://www.ncl.ucar.edu/Document/Manuals/NCL_to_Python/Transition_Guide_NCL_PyNGL.pdf
 - ▶ Exemple de script utilisant cartopy déjà disponible sur lustre (tracé de trois cartes superposées : couleurs, contours, hachures) :
`/cnrm/est/USERS/rigoudyg/dev/scripts_plot/plotmap.py`
 - Organiser un Jeudi du Climat sur l'utilisation de python (installation package...) ou sur le passage NCL – python ?
 - Interrogation du service d'errata
 - Récupération de données sur des sites distants
 - Regarder s'il est utile/envisageable d'appeler cdo depuis le module python dédié : non installé sur les serveurs
 - Discussion sur le remplacement de cdo (à long terme, soucis de performances sur des volumes de données importants)

Atlas et C-ESM-EP

- C-ESM-EP :
 - Version V2.0 de la C-ESM-EP disponible par défaut sur le dépôt GitHub
 - Atlas réalisés par Laure Coquart sur les simulations CMIP6 disponibles sous : <http://cerfacs.fr/giec6/>
 - Version disponible sur le lustre du CNRM sous `/cnrm/est/USERS/rigoudyg/C-ESM-EP`.
Utilisable après copie chez soi.
Les jobs sont lancés sur la partition P8HOST sur aneto (valable pour tous les serveurs).
 - Vérifier qu'il est possible de mettre l'option adéquate dans les plot pour le régional :
 - ▶ Ne donne pas des résultats satisfaisants à première vue (données Aladin projetée sur la grille des observations et non l'inverse)
 - ▶ `file:///cnrm/est/USERS/rigoudyg/NO_SAVE//C-ESM-EP/standard_comparison_aladin_rigoudyg/Atmosphere_Surface/atlas_Atmosphere_Surface_standard_comparison_aladin.html`
- Atlas / Générateur automatique de diagnostics :
 - Développements toujours en cours
 - Discussion en cours avec Jérôme, cahier des charges à faire avant réunion sur le sujet

Eclis : Versions

- V6.43 (depuis le 9 août) :
 - Augmentation du realization_index
 - Correction du plugin atm_lbc_nudging pour pouvoir commencer des run Aladin en milieu de mois
 - Corrections de bogues
 - Correction de cm_files
 - Changement de la version de NCTIME par défaut : 4.6.4 (correction du bogues liées aux dates supérieures à l'an 4000)
 - Ajout de variables disponibles via le fichier param pour gérer les caractéristiques du job assemble_and_QC : ELAPSAQC, ELAPSFRONTAQC, MEMAQC, MEMFRONTAQC, QOS AQ C, QUEUEAQ C
- V7.0
 - Poursuite des développements
 - Développements des branches V6.* et V7.0_beta* remontés régulièrement
 - Début du travail sur le nudging avec récupération des différentes versions, préparation pour les run aladin
 - Durée des run unitaires : limitation dans UPDCLIM et non dans ECLIS

Eclis : Prochaine version (V6.44)

- Développements finalisés :
 - Ajout d'un flag pour indiquer s'il s'agit d'une simulation de production et pour quel projet
- Développements en cours :
 - Finalisation de job_postpros_new et assemble_and_QC_new (utilisation de mtool)
 - Test changement de version de mtool
 - Mettre un warning dans relan si jamais il y a déjà eu un post-traitement pour éviter d'avoir des données plusieurs fois sur Hendrix
 - Ajout de la possibilité de post-traiter seulement les données correspondant à une période spécifiée par l'utilisateur
- Récupération des développements en cours sur la V7.0 concernant le post-traitement :
 - Réécriture d'une partie des scripts utilisés par le post-traitement en python pour plus de flexibilité
 - Ajout de la possibilité de faire une partie des post-traitement avant la concaténation des fichiers (exigence HR utilisable pour d'autres simulations)
 - Ajout de possibilités pour la concaténation par taille
- Mise à disposition visée : décembre 2019
- Objectif : première version portée sur le nouveau calculateur avec la V7.0, les versions plus anciennes ne seront plus disponibles

Eclis : Et pour la suite...

- Développements prévus :
 - Suppression de LDR_EXPID
 - Interfaçage de xiofilechecker (vérification de l'adéquation des fichiers xml et des fichiers produits)
 - Interfaçage d'un débogueur (voir la documentation d'Aurore sur le sujet)
- Développements à discuter à l'occasion :
 - Vérifier quelles variables devraient être modifiées par Eclis dans les settings de dr2xml (en lien avec les développements dr2xml)
 - Intégration d'un équivalent de verifrelan dans Eclis ? Voir si les fonctionnalités de slurm le permette comme on ne peut pas faire de cron sur le calculateur. A discuter en fonction.
 - Les post-traitement peuvent se faire au fil de l'eau, voir s'il faut le généraliser au vu de la charge de lustre sur les calculateurs.
- Documentation :
 - Mettre à jour la documentation sur la façon de faire des runs avec Eclis et dr2xml
 - Communiquer sur l'utilisation de LDR_EXPID et expid_in_filename (problèmes avec différents outils lors de l'utilisation par la suite : get, CliMAF, C-ESM-EP...)
 - Mettre à jour les param de référence sous */scratch/CMIP6/V2*

dr2xml : Versions

- V1_CMIP6_dev :
 - Version parallèle qui embarque toutes les nouveautés des branches récentes sans les modifications pouvant impacter le contenu ou les métadonnées des fichiers
- V2.0 (3 juin 2019) :
 - Compatibilité python 3
 - Axe de pression configurable
 - Fichier contenant les pattern des variables dev et perso
 - Possibilité de réduire les méta-données
- V2.1 :
 - Disponible dès qu'elle sera deboguée
 - Contenu : cf diapo suivante

dr2xml : Version en développement (V2.1)

- Développements finalisés :
 - Adaptation pour Cordex : V2.0_xml_cordex
 - Il reste à voir l'ajout des objets de type char (de la même façon que pour NEMO)
 - Gestion des caractères spéciaux :
 - Il faut encoder les fichiers contenant des caractères spéciaux en utf-8 (et même tous les fichiers texte :-)) :

```
iconv -f encodage_initial -t utf-8 fichier_init -o fichier_out
```
 - Gestion interne de la lecture et de l'écriture des fichiers xml avec conservation de l'ordre.
 - Début de réalisation d'une doc sphinx
 - Correction de l'absence de l'attribut branch_method pour les simulations n'ayant pas de parent_id
- Reste à faire :
 - Encore quelques problèmes de non reproductibilité à corriger
- Mise en place d'une base de tests permettant de vérifier la reproductibilité entre deux versions du code et de Python
- Mise à disposition visée : fin novembre 2019

dr2xml : Et pour la suite...

- Généralisation de dr2xml à d'autres projets (en fonction des demandes) :
 - Actuellement, présence de tests pour vérifier si on se trouve dans le cas Cordex ou dans un CMIP6).
 - Utilisation d'un fichier de settings propre au projet
 - Nécessite de savoir quelles variables des settings dr2xml sont propres à CMIP6
 - Généraliser la structure pour qu'elle soit plus adaptée aux différents projets

- Utilisation d'un fichier settings propre aux variables dev ?
 - Ajout des axes personnalisés
 - Gestion des grilles natives pour les variables dev

- Poursuite de la rédaction de la doc

- Simplification de l'installation (pip, ...)

Suite...

- Prochaine réunion : 16 janvier à 14h en salle 151