

Réunion Tech

- Production CMIP6
 - Publications CMIP6
 - Ressources et site Web
 - CliMAF
 - C-ESM-EP
 - Eclis
 - dr2xml
-

23 janvier 2020

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
- ~~Problèmes avec certaines variable 6hrLev de l'historique (ua, va...) pour l'axe des temps~~
- Le parent_variant_label est erroné pour certaines simulations pour lesquelles $r > 1$ ([liste des simulations concernées à faire encore](#))
- Fichier de recensement des publications par MIP utilisant les sorties des modèles du CNRM-Cerfacs pour CMIP6 :
<https://docs.google.com/document/d/1HZyJWuGMQxVonX5gQltD9wP2TRjHHkE2/edit#>
- Fichier de recensement des variables non produites qui manquent ([si besoin, des reruns pourront être envisagés pour ces variables, pour l'instant, fait à titre d'information](#)) :
<https://lite.framacalc.org/9dxk-ocyj7tesh5>

Publications CMIP6 ESGF

- Publication CMIP6 réalisée à ce jour sur 28 nœuds. Détails sur l'ESGF et sur la page de monitoring : https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/
- Nouvelle version de la pile logicielle ESGF (v4.0.4) : installée sur le nœud de test esg2 du CNRM (publication Cordex à tester encore avant décision de déploiement sur nœud de production esg1)
- CDNOT (CMIP Datanode Operations Team) : papier sur travaux CDNOT pour CMIP6 en cours de rédaction, évolution conditions accès CMIP5 (accès libre, idem CMIP6, mis en œuvre au CNRM le 09/12), réflexion sur évolution ErrataService pour permettre aux utilisateurs de remonter les issues via l'outil (utilisateurs → producteurs au lieu de producteurs → utilisateurs)
- Premières fonctionnalités 'compute' livrées par ESGF CWT (Compute Working Team) : cf <https://esgf.llnl.gov/esgf-compute-announcement.html>
- Meeting annuel ESGF F2F (Face2Face) 23 au 25/03 puis 1ère AG IS-ENES3 25 au 27/03 au CIC, informations AG IS-ENES3 (DL inscription 21/02) sous <https://is.enes.org/events/workshops/is-enes3-first-general-assembly>

Publication ESGF au CNRM

■ Bilan publication CNRM au 22/01 :

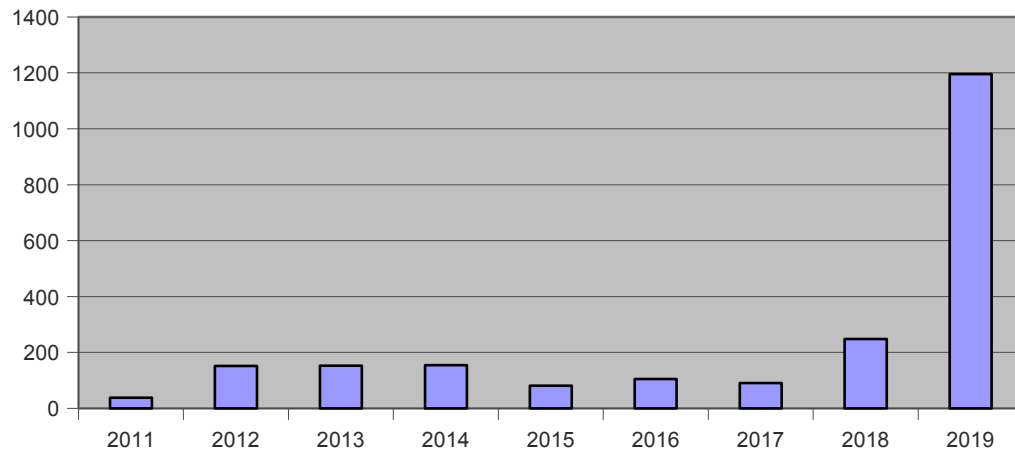
- détails sous /cnrm/est/USERS/franchisteguy/CMIP6/CNRM/Publication/SuiviPubli/SuiviPubliGlobal.ods
- **276** bons à publier émis au CNRM (+ 3 extensions 2100-2300 membres r1 ScenarioMIP ESM2-1)
- **258** simulations publiées soit ~88 % du nombre total de simulations à publier sur l'exercice (**290** prévues à ce jour). 14 MIPs couverts côté CNRM-CERFACS
- Depuis la dernière réunion Tech (13/11) :
 - * publi rerun amip-piForcing
 - * publi DECK HR : historical, piControl, amip-hist, amip, 1pctCO2, abrupt-4xCO2
 - * publi ScenarioMIP HR ssp245, ssp585 (sauf 4 variables table 6hrLev hus, ta, ua, va)
 - * simulations LS3MIP(4) et LUMIP(3)
 - * publication membres supplémentaires DECK CM6-1 historical (28 membres d'ores et déjà publiés)
 - * dépublication puis publication rerun CM6-1 amip-a4SST-4xCO2 (4xCO2 par rapport au preindus au lieu des valeurs historiques run amip)

A venir côté CNRM (par ordre de priorité décroissant) :

- * fin publication membres supplémentaires DECK CM6-1 historical (r29 et r30) et membres supplémentaires DECK ESM2-1 historical (r6 à r10)
- * publication ScenarioMIP HR ssp126 et ssp370
- * publication 4 membres supplémentaires CM6-1 ssp245 (r7 à r10)
- * extension '2300' membres r1 ScenarioMIP ESM2-1
- * esm-hist r1i1p1f3 (modèle en émission)
- * complément DECK piControl LR (500 premières années seulement publiées) après émission bon à publier correspondant
- * variables P1 tables SIday et SImon ESM2-1 (fait pour CM6-1)
- * puis ordre chronologique d'arrivée des futurs bons à publier....

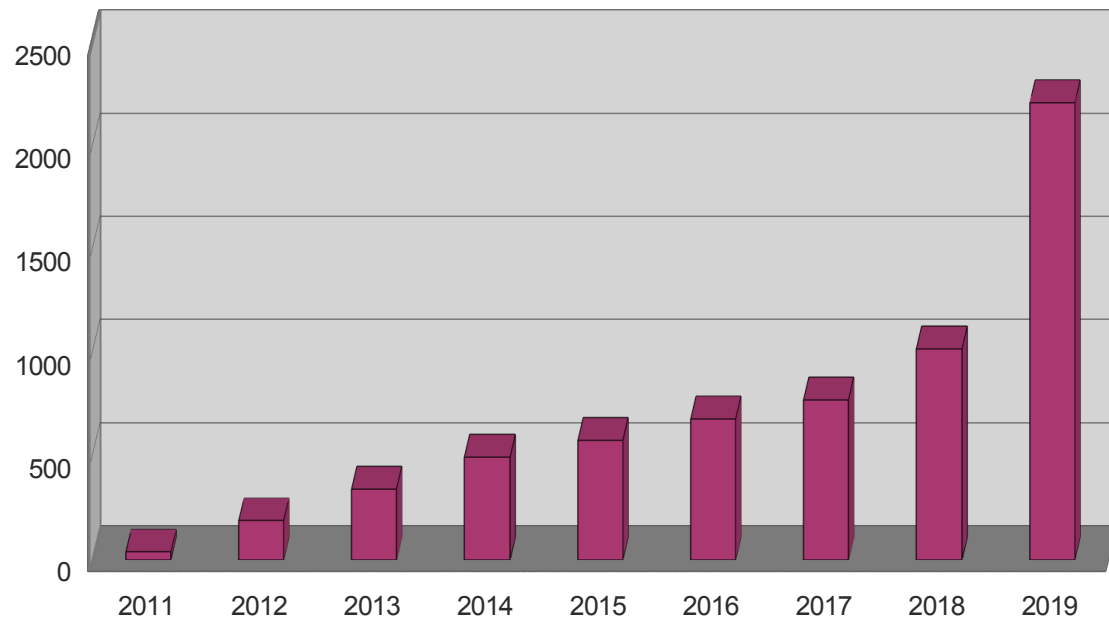
Trafic noeud ESGF au CNRM (volume téléchargé)

Volume annuel données téléchargées sur le noeud ESGF du CNRM (To)



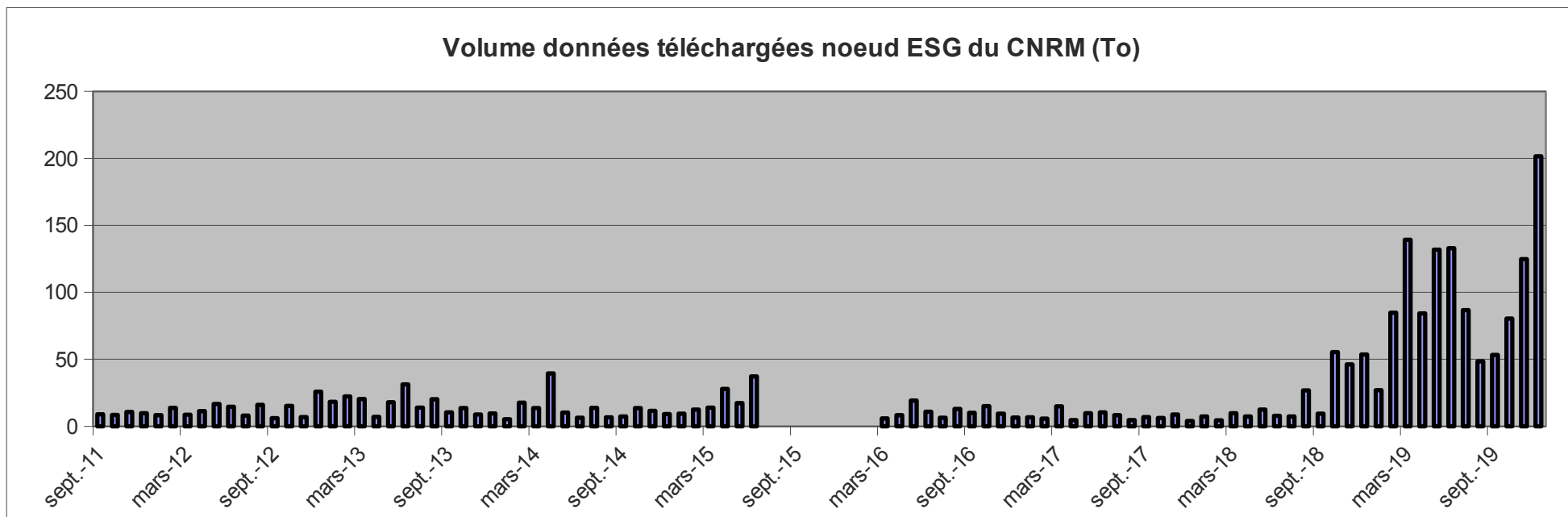
Volume téléchargé 2019 (1.2 Po) supérieur au volume cumulé 2011-2018 (1 Po)

Volume cumulé annuel données téléchargées sur le noeud du CNRM (To)



er 2020

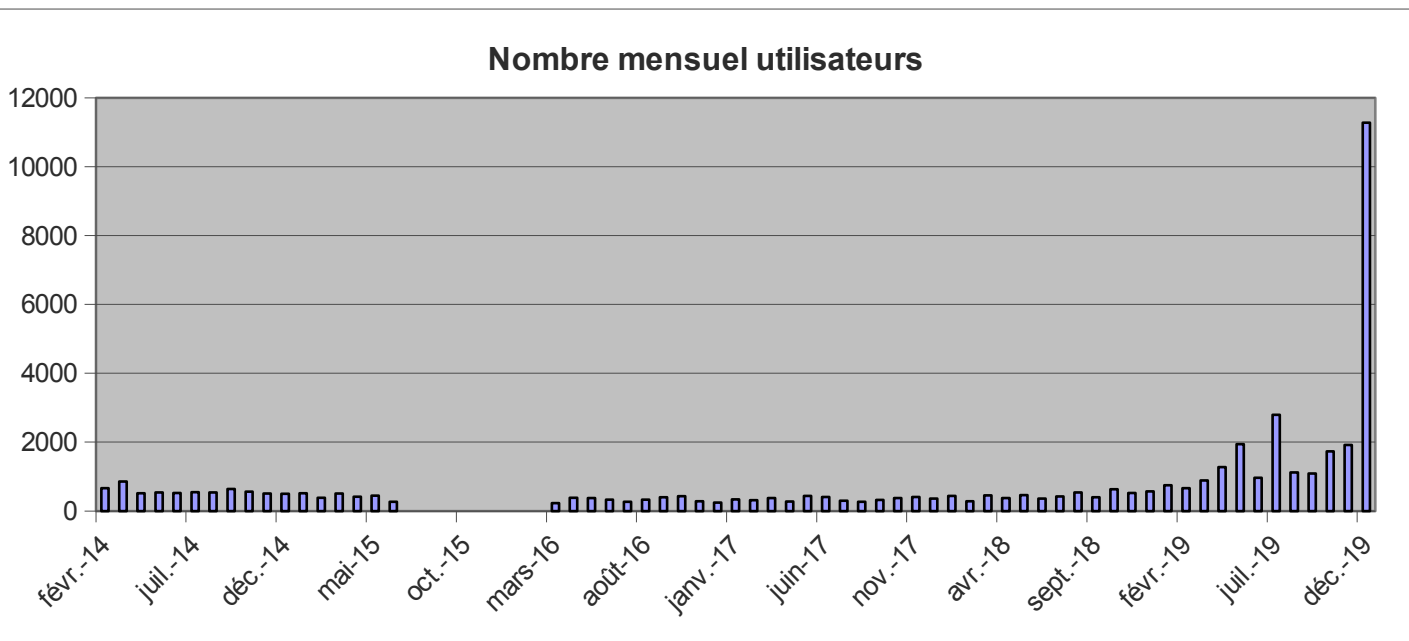
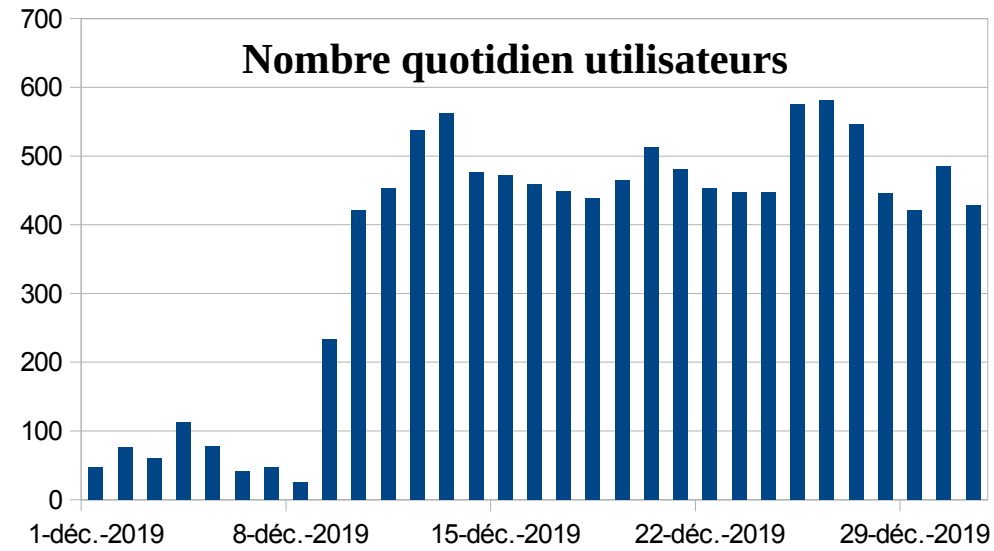
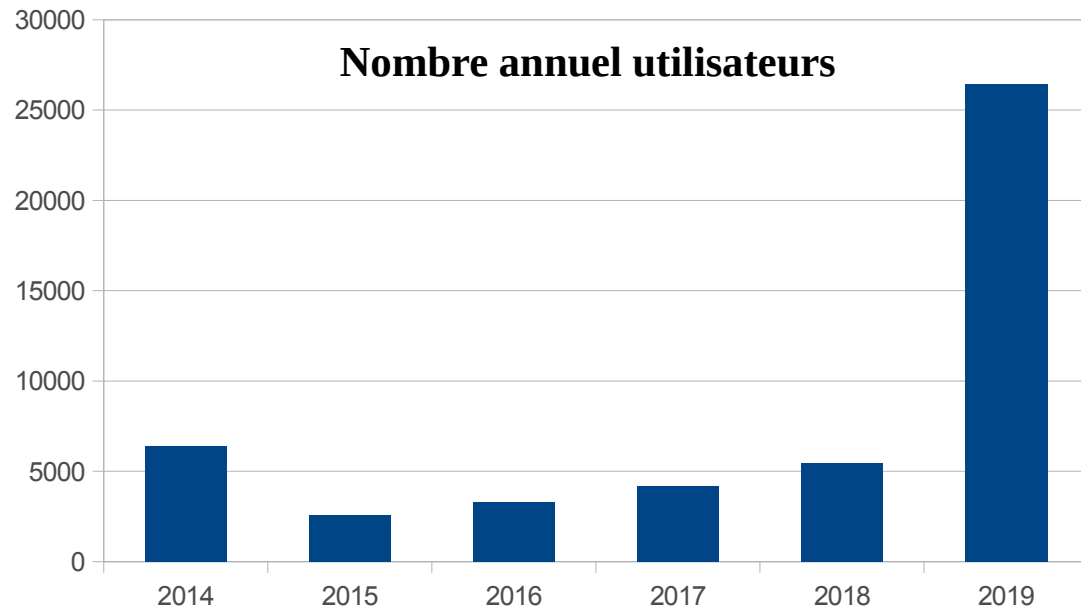
Trafic noeud ESGF au CNRM (volume téléchargé)



Record volume maximal quotidien téléchargé battu le 21/12/19 : 15.3 To (ancien record 14.7 To le 19/12)

Record mensuel battu en décembre 2019 : 201.6 To (ancien record 139.3 To en mars 2019)

Trafic noeud ESGF au CNRM (nombre utilisateurs)



Effet date limite soumission articles AR6

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. Fait pour CM6-1, à faire encore pour ESM2-1
- 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)
- Champs ***4co2 ESM-2-1** erronés (calculés 1xCO2) à dépublier et à déclarer sur l'ErrataService (456 datasets concernés). Info sur site web CMIP6
- Rappel : tableau (usage interne) recensement erreurs sous /cnrm/est/USERS/franchisteguy/CMIP6/CNRM/Production/Issues (InfoIssuesInterne.odt)
- Nombre de demandes reçues sur contact.cmip@meteo.fr en 2019 : 26
- Envoi à Laure liste des simulations CNRM dépubliées puis republiées (fait le 23/01), informations disponibles sur l'Errata-Service et sur le site web CNRM CMIP6.

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 (à commencer) puis cm6-1-hr

Ressources calcul et stockage, site web

- (Rappel renouvellement calculateurs : arrêt Beaufix/Prolix prévu le 19/10/2020 pour le moment, voir mail A. Beuraud 10/01)
- Bilan consommation mgie 2019 : +2 Mh au-dessus du quota Beaufix 2019 mgie (- 1.6 Mh quota Prolix), 28.7 Mh consommées en queue Libre sur Beaufix
- Demande calcul 2020 (élaborée pour nœuds 128 Go en novembre dernier) :
 - * Beaufix : réalisation simulations CMIP6 restantes (5 Mh) + post-traitement nœud 256
 - * Prolix : 500 000 h
 - * Belenos : reproductibilité 200 ans LR et rerun éventuels (1 Mh) + post-traitements nœuds 512 Go
 - * Taranis : 200 000 h
 - arbitrage lors du CPCC du 31/01

Portage Calcul2020

- Les prochaines réunions aborderont le portage.

- A noter :
 - Le volume des \$HOME sera limité à 50Gb/utilisateur sur les prochains calculateurs
 - Chaque utilisateur aura la responsabilité du transfert de ses propres données entre anciennes et nouvelles machines Bull
 - Les répertoires sous /scratch seront nettoyés tous les deux à trois mois
 - A priori, il n'y aura pas de file system pérenne
 - Modifications à prévoir pour EcliS en particulier (répertoires de relance)

- A faire :
 - Lister les besoins en termes d'espace pérenne en particulier
 - Lister les fichiers se trouvant sous /scratch/CMIP6/V2 ou dans d'autres dossiers, en rapport avec CMIP6 qui n'ont pas été archivés et doivent l'être
 - Reproductibilité CM6-1 LR (200 ans) envisagée

- Formation :
 - Poursuite des baux CLiMAF

- Contributions à CLiMAF :

- Version disponible :
 - V1.2.12 (11/09/2019) :
 - Compatible avec la C-ESM-EP
 - Corrections de bogues
 - Ajout d'opérateurs cdo sans sélection cachée : `ccdo_fast`, `space_average_fast`, `time_average_fast`, `llbox_fast`, `annual_cycle_fast`, `clim_average_fast`

- 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 comparaisons 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
 - Sélection multiple du même dataset

- Prochaines versions :
 - V1.2.13 (février 2020) :
 - ▶ Mise à jour de la documentation
 - ▶ Mise à jour des exemples
 - ▶ Correction de bogues
 - ▶ Modification de l'appel à cdo
 - ▶ Amélioration des figures
 - ▶ Possibilité d'appel des opérateurs sans fichiers d'entrée
 - V2.0 :
 - ▶ Compatibilité python 3
 - ▶ Écriture de tests unitaires

CliMAF : Et pour la suite...

- Homogénéisation de la récupération des données
- Extension des projets
- Remplacement de NCL
- Interrogation du service d'errata
- Récupération de données sur des sites distants
- Remplacement de CDO (?)

- Interfaçage / comparaison avec ESMValTool : des discussions ont commencé entre Jérôme et l'équipe de développement de ESMValTool (meeting à l'EGU). Voir pour prévoir un moment de discussion à l'occasion de l'AG IS-ENES.
- Installer ESMValTool pour prise en main.

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

- Versions :
 - V6.44 (janvier-février 2020) :
 - Ajout de la configuration pour Epona
 - Réécriture de la partie post-traitement
 - Ajout d'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 variable PROJECT pour indiquer le projet auquel appartient la simulation
 - V7.0 (en développement) :
 - Permet de faire des run annuels
 - Travail sur le nudging
 - Adaptation pour les run aladin et ESM
- Pause dans les gros développements pour le moment. A suivre dans les prochaines versions :
 - Corrections de bogues (si besoin)
 - Adaptations mineures (si besoin)
 - Mise à jour de la documentation
 - Mise à jour des params de référence
 - Interfaçage de xiofilechecker (vérification cohérence xml / fichiers produits)
 - Interfaçage d'un débogueur
 - Suppression de LDR_EXPID
 - [Modifier le répertoire de relance \(contenu, positionnement, stratégie d'archivage\) pour s'adapter aux changements d'architecture](#)

Eclis : Et pour la suite...

- 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.

- Continuer à suivre le développement de simu-lanceurs pour voir si d'autres outils pourraient être à même de remplacer Eclis. En cas de refonte d'Olive côté GMAP, suivre pour voir ce qui sera développé (même démarche à mener éventuellement avec l'IPSL et l'outil Cylc).

dr2xml

- Versions disponibles :
 - V2.1 (27/11/2019) :
 - ▶ Adaptations pour Cordex
 - ▶ Début de documentation du code
 - ▶ Utilisation d'accents possible dans les fichiers de settings (encodage utf-8 requis)
 - ▶ Ajout de l'attribut `branch_method` (y compris quand il n'y a pas d'expérience parent)
 - V1_CMIP6_dev :
 - ▶ Version proche de celle utilisée pour CMIP6 (pour le rendu des fichiers de sorties)
 - ▶ Embarque des nouveautés (variables dev...)
 - Base de tests :
 - ▶ Possibilité d'inclure des configurations particulière pour les tests de non régression.
- Pause dans les gros développements pour le moment. A suivre dans les prochaines versions (2.1.*) :
 - Poursuite de la documentation
 - Écriture des tests unitaires
 - Corrections de bogues (si besoin)
 - Adaptations mineures (si besoin)

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 : [mardi 10 mars à 14h30](#)

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, lbox_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

CliMAF : Prochaine version (V1.2.13)

- Modifications disponibles :
 - Réécriture de mcdo.sh en python (mais toujours appel depuis le shell)
 - Correction de bogues pour la récupération d'une sous-période d'une variable dérivée disponible dans le cache
 - Modifications sur les figures (Stéphane)
 - Les opérateurs peuvent maintenant avoir un input vide (Stéphane)
- Modifications en attente :
 - Ajout du script de tracé en python
 - Correction du bogue liée à cprotect (Jérôme)
 - Modifications supplémentaires liée à l'AR6 (Stéphane)
 - Accélération du cache (Stéphane)
- Documentation :
 - Reprise des issues relatives à la documentation
- Disponibilité : février 2020

CliMAF : Développements

- V2.0 :
 - Compatibilité Python 3
 - Renommage de la fonction slice
 - 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
 - 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 (voir pyngle) :
 - ▶ Exemple de script déjà disponible sur lustre (tracé de trois cartes superposées : couleurs, contours, hachures) :
`/cnrm/est/USERS/rigoudyg/dev/scripts_plot/plotmap.py`
 - 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

Eclis : Versions

- V6.44 (janvier-février 2020) :
 - Ajout des fichiers de configuration pour Epona
 - Ajout d'un flag pour indiquer s'il s'agit d'une simulation de production et pour quel projet (variable PROJECT)
 - Réécriture du script d'assemblage et de contrôle qualité en Python avec l'utilisation de la dernière version de mtool (les scripts de post-traitement peuvent toujours être écrits dans d'autres langages). Amélioration de la gestion des périodes et ajout de la possibilité de faire des post-traitements avant la concaténation.
 - Modification de la version de mtool utilisée par défaut.
 - Ajout de la possibilité de post-traiter seulement les données correspondant à une période spécifiée par l'utilisateur
 - 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
- 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

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 (sauf exception).
 - **Ajout de l'attribut `branch_method` dans les fichiers netCDF dans tous les cas (pas seulement s'il y a une expérience parent). S'il n'y a pas d'expérience parent, la valeur prise est "no parent".**
- V2.1 (27 novembre 2019) :
 - Correction pour la gestion des caractères spéciaux : nécessité d'encoder les fichiers concernés (et même tous les fichiers :-)) en utf-8 :

```
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. Ce changement occasionne un léger surcoût (plus faible en Python 3 qu'en Python 2).
 - Ajout de l'attribut `branch_method` dans les fichiers netCDF dans tous les cas (pas seulement s'il y a une expérience parent). S'il n'y a pas d'expérience parent, la valeur prise est "no parent".
 - Adaptation pour la réalisation de simulations Cordex.
 - Ajout d'une documentation Sphinx qui doit encore être largement étoffée.
- Tests :
 - Mise en place d'une base de tests permettant de vérifier la reproductibilité entre deux versions du code et de Python.
 - Pour des configurations particulières, n'hésitez pas à m'indiquer les fichiers utiles pour tester les changements.