

Réunion annuelle du CU du CDPP, IRAP+distanciel, 12 décembre 2022

Présents (CU) : B. Grison BG (IAP, Rép. Tchèque), S. Hess SH (ONERA), K. Issautier KI (LESIA), O. Le Contel OLC (LPP), R. Maggiolo RMA (IASB, Belgique), R. Modolo RMO (LATMOS)

Excusés (CU) : E. Astafyeva EA (IPGP), N. Aunai NAU (LPP), P. Henri PH (LPC2E)

Participants (CDPP): N. André NA (IRAP, président), B. Renard BR (AKKA IRAP) V. Génot VG, F. Pitout FP, M. Bouchemit MB, C. Jacquy CJ, A. Rouillard AR, I. Plotnikov IP (IRAP), N. Dufourg ND, D. Boucon DB (CNES)

En gras, les points susceptibles de figurer dans la présentation au CD.

NA mentionne lors de l'introduction le format du **CU découpé en deux demi-journées**. En plus de la réunion du jour, le CU se réunira courant mars pour le retour des tests. **La place est donnée aujourd'hui à plus de discussion pour moins de longues présentations.**

OdJ:

09:30-10:00 Ingestion de nouvelles données observationnelles et préparation des tests des données Solar Orbiter

10:00-10:30 Activité archiver (exploitation et migration) au CNES

10:30-11:00 Ingestion de nouvelles données de simulations numériques

Pause (10-15 min)

11:15-11:45 Ingestion de catalogues et d'algorithmes de machine learning et préparation des tests du module dédié

11:45-12:15 Réflexion sur les catalogues et partage de codes python communautaires

12:15-12:45 Module paramètres dérivés, modèles analytiques, et préparation des tests dédiés

Pause déjeuner

14:00-14:30 Renouvellement du CU et identification des testeurs

14:30-15:00 Participation aux projets, liens avec les centres et bases de données, enjeux et perspectives

AMDA (BR)

- On peut citer les ajouts de données Parker Solar Probe (PSP) et notamment Solar Orbiter (SO) les données RPW L2. Ces dernières sont pour le moment **accessibles via une plateforme provisoire pour les tests du CU qui devront faire remonter les problèmes et commenter l'organisation/description des données.**

Quelques données dans AMDA sont encore sous embargo. Il manque encore quelques données SO.

- Un nouvel outil d'ingestion plus fiable et plus versatile a été développé en python. Il fait aussi remonter les problèmes dans les jeux de données (corrompus, redondants, ...)

BG demande si cet outil fonctionne sur une base régulière. BR : Pas encore mais c'est prévu à terme.

Le CU est satisfait de cet outil qui répond à un besoin exprimé lors d'une précédente réunion.

- Speasy est maintenant en version 1.0.4 (progrès considérable)

BG demande comment ajouter une courbe provenant d'un jeu de données personnelles à l'aide des fonctionnalités de tracé de Speasy. BR envisage d'ajouter une démo pour ce point précis.

Pour trouver un jeu de données, le plus simple est d'utiliser la complétion automatique.

OLC note que py spedas multi-mission et multi-instrument (sol et espace) est plus ancien est permet de faire beaucoup de choses, il pourrait être complémentaire de speasy.

BR : pyspedas peut-il accéder aux données AMDA ?

OLC : il permet d'accéder aux données CDAWeb au format CDF, une interface serait probablement nécessaire pour accéder aux données AMDA au format NetCDF.

NA mentionne que l'équipe japonaise de BepiColombo envisage d'utiliser aussi pyspedas. Y'a-t-il un choix à faire ou des recoupements à éviter entre py speasy et py spedas?

- Description des données SPASE : elles sont conformes et accessibles au HPDE de la NASA.

SH : l'ONERA utilise AMDA/HPDE qui marche bien dans presque tous les cas (IMPEX à finaliser)

Le CU note l'avancée du CDPP dans la description des données

- AMDA représente quelques 5.2 To de données bien comprimées (+2.9 To à venir). Cela semble peu (BG) : rôle de la compression, mais la migration à venir pour un accès plus rapide va entraîner un accroissement du volume des données (ainsi que les données SO).

IP note qu'AMDA permet la comparaison de données particules hautes énergies de différentes sources

BG demande s'il faudrait identifier les jeux de données plus rapidement que par le nom de l'instrument ? NA rappelle qu'il y a une info bulle, mais qu'on peut envisager un code couleur par famille (ondes, particules, ...)

Archive (ND, DB)

- Ajouts de nouveaux jeux de données (routine). On pourra noter l'arrivée des données Giotto/RPA

- Rappel : le SIPAD doit être complètement refondu (obsolescence matérielle et logicielle) vers REGARDS et STAF.

Ce processus est bien engagé. Pendant la migration vers REGARDS, l'ajout de nouvelles données est suspendu. La fin de la migration interviendra au plus tard fin T2 2023. Parmi les buts de cette migration, on peut citer la compatibilité avec SPASE et une plus grande interopérabilité (notamment avec AMDA).

La migration vers REGARDS sera suivie de l'évolution vers STAF.

BG demande si le seul impact concret est la suspension de l'ajout de nouvelles données ? Théoriquement oui (NG, DB) : le passage vers STAF devrait être transparent pour les utilisateurs, mais on ne sait jamais.

Une démo de REGARDS est prévue pour le prochain CU.

La migration a un impact très limité pour les utilisateurs.

- ND souligne que le CNES est cité en exemple pour les activités d'archivage. NA note que les observations de la Terre ont une plus grande force de frappe pour l'archivage
ND souligne que dans les sciences de l'univers leur travail est reconnu. La thématique plasma fonctionne bien et l'intérêt d'avoir un pôle qui a une pérennité plus grande que les missions est bien compris.

- ND est plus optimiste pour le calendrier REGARDS car il souhaite avoir un peu de temps entre les actions REGARDS et STAF (très chronophages) pour développer de nouvelles fonctionnalités, notamment l'accès à AMDA depuis l'archive et l'accès à l'archive depuis AMDA, et redémarrer l'ajout de nouvelles données.

NA soutient le développement des nouvelles fonctionnalités et BG rappelle que ces fonctionnalités sont une demande de longue date du CU.

Le CU remercie le CNES pour la prise en compte de l'intérêt des utilisateurs du CDPP en cette période chargée. Le CU souhaite que l'équipe ait les moyens de parvenir à réaliser ces développements entre la fin de la migration vers REGARDS et le début du déploiement de STAF.

Les registres SPASE sont multiples (CNES, CDPP, et HPDE de la NASA) et il faut les garder synchronisés, ce qui est complexe. SH note que cela entraîne la duplication de certaines données qui sont déjà sur HPDE. DC est au courant du problème mais sans cela les données NASA ne sont pas mises à jour à temps.

SH souligne que faire le lien avec le HPDE n'a rien de simple. BR la même expérience avec AMDA, c'est compliqué pour que ce soit homogène.

Le CU suggère que ce problème soit traité en commun (CNES, AMDA)

Simulations numériques (NA) :

Il y a déjà des résultats de simulations numériques dans AMDA (MHD, hybrides, IMPEX)

NA souhaite développer cette activité qui est importante pour les missions planétaires.

RMO souligne que cela pose de nombreuses questions : Où seront stockées ces données ? Comment sont-elles décrites ? Il faut une interopérabilité entre le CDPP et les producteurs de simulations. Combien de temps garder ces données ? (à la différence des observations, les simulations se périment)

SH souligne également les énormes besoins de stockage.

ND souhaite avoir cette réflexion, mais cela n'entraînera pas plus de budget.

BG demande une clarification pour savoir si cela entre dans l'activité d'archivage du CDPP.

Non, car ce n'est pas directement du spatial (NA). AR observe de plus en plus un brouillage entre données pures et simulations numériques.

SH a partagé par mail les conclusions d'un groupe de travail de la NASA/PDS autour de cette thématique.

Précisions apportées par VG : il faut que les simulations soient décrites par les API ce qui n'est pas simple et il faut un fichier descriptif d'aide à l'ingestion de simulations. Pour le stockage, il faudrait laisser la responsabilité aux simulateurs (les SI des labos ne sont pas forcément équipés pour répondre à ce besoin)

OLC exprime le besoin côté utilisateur d'avoir un module qui permettrait de tracer les paramètres physiques d'une simulation (ex : n, V, T, B) le long de l'orbite d'un ou plusieurs

satellites à partir des fichiers d'orbites archivés au CDPP. Cela constituerait par exemple une aide à la définition de mission ou de projet de recherche sur les données.

BR : 3D view permet déjà de tracer les données obtenues par des satellites virtuels se déplaçant dans une simulation

AR note que si on veut qu'un utilisateur puisse rebondir sur des simulations ou avoir des run à la demande c'est extrêmement coûteux en RH.

PH souhaite faire passer un message sur cette thématique. Un atelier joint de l'ASN et de l'ASOV il y a un mois qui pointait les besoins de mise à disposition de résultats de modèles et l'exemple du CDPP a été bien mis en évidence, comme ce que l'INSU chercherait à voir se développer (dans d'autres domaines de l'astro). La communauté PNST est en avance sur cet aspect : il serait bon de le faire valoir à l'INSU et de conserver voire développer un peu plus cet aspect.

NA conclut en rappelant qu'il faut être présent sur cette thématique demander au CU de faire des retours si il y a des discussions sur le sujet.

AMDA doit intégrer ces bases de données de simulations comme des « remote databases ».

Catalogue et machine learning (NA, BR) :

Dans AMDA, on peut trouver :

- modules dédiés (ML)
- objets partagés (catalogues)

Les modules sont isolés les uns des autres dans la plateforme orchestra (transparent pour l'utilisateur). Et chaque module est lié à une mission. BG demande si on peut faire remonter les erreurs ? Le ré-entraînement n'est pas possible pour le moment (BR) mais à terme le module doit pouvoir faire intervenir l'utilisateur (NA).

BG voit surtout l'intérêt de créer un jeu de données dans le cas des régions identifiées par ML pour une mission donnée (le temps de calcul à la volée n'est pas négligeable). Mais **ce serait intéressant pour les utilisateurs d'utiliser ces modules sur d'autres missions.**

NA compte sur le CNES pour gagner du temps de calcul. AR rappelle la réception en février 2023 d'un nouveau serveur à l'IRAP pour les calculs intensifs.

Catalogues (thème sollicité par le CU), NA, BR :

On peut facilement tracer les catalogues qui ont un format simple ce qui permet d'en ajouter dans AMDA. On ne voit pas rapidement tous les catalogues disponibles dans AMDA (il y a eu une page web dédiée). BG trouve qu'il n'y a pas beaucoup de publicité faite autour de tout ce qui est ajouté dans AMDA (ex catalogue de traversées magnétopause).

Action VG : rendre plus visibles les catalogues.

KI demande si on est certain d'avoir des catalogues à jour ? c'est une limite inhérente à chaque catalogue (VG) sauf si on a un lien avec les auteurs de catalogues (NA).

Partage de code Python

Après un rapide tour de table, on note que les membre présents du CU ne développent pas beaucoup en python.

OLC note qu'il y a une page qui présente les outils Aidapy <https://gitlab.com/aidaspace/aidapy> ou pypedas <https://github.com/spedas/pypedas> (même si tout n'est pas bien commenté). OLC propose d'avoir une page web dédiée aux développements de codes sur le site du CDPP pour centraliser les informations et où les développeurs laisseraient un lien vers leur git. Se pose alors le problème d'être proche de la communauté (NA). CJ propose d'avoir un forum mais ND prévient qu'un forum nécessite un travail d'administration.

BG note que NAU était particulièrement demandeur mais est absent aujourd'hui.

Paramètres dérivés (demande du CU)

BG et KI souhaitent avoir plus de fonctions dans l'onglet « space » (qui sont donc des paramètres multi-missions)

RMO propose d'avoir une fonction de calcul de MVA et note qu'il n'est pas possible de partager un paramètre. BR mentionne que ce n'est pas encore au point. RMO propose de créer un groupe de travail si besoin.

Le CU sera appelé à tester la nouvelle fenêtre et commenter sa présentation.

Renouvellement du CU

KI propose son remplacement par **Léa Griton** qui a un très bon profil pour le CU (simulations, analyse de données, code).

Le CU est d'accord pour ce remplacement qui sera proposé au CD.

RMO souhaite aussi passer la main mais il n'y pas de nouveaux recrutements au LATMOS.

OLC n'a pas non plus de remplaçant identifié. Le CU suggère de demander directement à Lina Hadid si elle est intéressée.

BG rappelle que Claire Foullon n'a pas été remplacée donc on peut proposer plus de candidats au CD. NA suggère Beatrice Sanchez-Cano pour l'aspect côté 3DView et Propagation Tool. KI note que Léa Griton connaît ces outils du CDPP. BG rappelle que l'an dernier le CU avait marqué sa préférence pour garder un CU francophone mais Beatrice Sanchez-Cano peut participer aux tests ou apporter son expertise ponctuellement.

Miho Janvier est également citée dans la discussion, au moins pour les tests, car elle a une expertise en particules énergétiques.

Action NA : se charge de contacter Miho Janvier.

Discussion ouverte

- Organisation des tests (et identification de testeurs)

Tous les membres du CU sont invités à tester le module des paramètres dérivés.

Le module de ML est facile à prendre en main. Le CDPP pourra solliciter Philippe Garnier. SH est intéressé par ces tests.

Python (via jupyter notebook) peu d'experts au sein du CU (Nicolas Aunai sera sans doute disponible pour les tests) : besoin de renforts (Lina Hadid, Léa Griton ?). RMA propose de faire participer sa doctorante aux tests python. RMO est également partant pour ces tests. OLC propose également de faire participer S. Alqeeq (post-doctorant au LPP) et M. Baraka (doctorant) qui travaillent en notebook python principalement sur les données MMS et qui ont participé au workshop Heliopy organisé par A. Masson en mai à l'ESAC (Madrid). Il propose de contacter Thomas Chust (Lead CoI RPW/LFR) qui travaillent également sur notebook python.

AMDA/Solar Orbiter : particules énergétiques RPW. BG se propose d'y participer pour équilibrer les équipes de testeurs.

Action BG : voir avec les membres du CU pour répartir les tests.

Et encore :

- KI s'enquière de la date du prochain CD. NA le prévoit pour fin juin.
 - NA note qu'il y a eu un soutien en RH, avec deux renforts mais que pour le technique MB est la seule à faire partie du labo.
 - Le CDPP souhaite à l'avenir resserrer les liens avec le CSA et la JAXA (Bepi)
 - Il y a maintenant une possibilité de recrutement au CNAP pour du multipoint / bdd.
 - NA confirme qu'il devra prendre du recul avec l'arrivée de Bepi. Il y a donc un aspect direction à préparer. KI demande quelle est l'importance de l'élargissement de la convention. NA souligne l'importance d'avoir une direction motivée et ND qu'il n'y a pas de soucis avec une direction ailleurs qu'à l'IRAP (on a beaucoup progressé en distanciel) car cela va être compliqué de trouver quelqu'un en local.
 - KI souligne l'existence de chaires d'excellence de la section 17 (après 3 ans, intégration au corps des DR) d'ici 2024-25 et l'importance pour la communauté de faire remonter un profil.
 - BG demande si il y a une préférence pour un CU distanciel, présentiel, ou en alternance. Il y a un consensus pour pouvoir se rencontrer régulièrement. RMO est favorable à au moins une alternance et un prochain CU hybride.
 - ND détaille les actions pour préparation de nouveaux contrat pour la maintenance corrective et évolutive (définition des specs début 2023). Le contrat plan du CNES se termine en 2023, et il a besoin d'identifier les besoins pour la suite. Le support archive en représente une part importante. A priori le même niveau de ressource sera demandé.
 - RMO souhaite quelques détails sur le retour des tests. NA souhaite un retour oral lors du prochain demi-CU avec quelques transparents.
- NA explicitera la forme des tests en début d'année 2023 et proposera un sondage pour le prochain CU focalisé sur le retour des tests (fin mars?).