

## Réunion annuelle du CU du CDPP, IRAP+distanciel, 09 septembre 2021

Présents (CU) : E. Astafyeva EA (IPGP), N. Aunai NAU (LPP), B. Grison BG (IAP, Rép. Tchèque), P. Henri PH (LPC2E), S. Hess SH (ONERA), K. Issautier KI (LESIA), O. Le Contel OLC (LPP), R. Maggiolo RMA (IASB, Belgique), R. Modolo RMO (LATMOS)

Abstente (CU) : C. Foullon CF (University of Exeter, UK)

Participants (CDPP): B. Renard BR (AKKA IRAP) V. Génot VG, F. Pitout FP, M. Bouchemit MB, N. André NAN, C. Jacquy CJ, A. Schulz AS, A. Rouillard AR, I. Plotnikov IP (IRAP), B. Cecconi BC (LESIA), N. Dufourg ND, D. Heulet DH, D. Boucon DB, D. Ly DL (CNES)

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

*En jaune, les points qui m'ont échappé ou qui me semblent douteux.*

*Ce compte-rendu ne retranscrit pas les nombreux transparents (200?) vus pendant la réunion mais insiste surtout sur les points qui ont soulevé des discussions.*

- Accueil – Introduction – Agenda

Première réunion du CU depuis 3 ans. Remerciements à Vincent pour ses années en tant que directeur du CDPP. NAN souligne qu'il a une moindre implication que VG (20%).

Plusieurs changements dans l'équipe : M. Gangloff et E. Budnik sont partis à la retraite. Cette dernière a été remplacée par B. Renard (AKKA, contrat de 4 ans, basé à l'IRAP)

CDD A. Schulz (IRAP) et au CNES J. Durand a été remplacée par D. Ly.

- AMDA : Nouvelles données

Ajout de nombreux jeux de données : PSP, SO (survol de Vénus notamment), Bepi Colombo, Cassini, JUNO, JUICE. Mariner 10, Phobos-2, AMPTE< FVA, ISEE, ATRTEMIS, Cluster, (Freja), Geotail, Interball Tail, Polar, JASON3, MMS, SWARM, Rosetta, Ulysses, Helios.

Finalisation : VEX MEX/IMA

Et encore à venir (RPW, SWA/HIS et SWA/EAS SO)

→ **Enrichissement notable de la base de données AMDA**

Une procédure de mise à jour automatique des jeux de données et contrôle de l'état de ces jeux a été implémentée. Cette procédure est maintenant intégrée dans le process d'ajout de nouvelles données.

Ce suivi de la qualité des données pourra encore être améliorée en renseignant la source (exclusivité AMDA ou non) des jeux de données (NAU). L'info existe mais doit être plus visible (NAN).

Common Workframe Language (CWF) est un outil existant pour tracer toutes les transformations (par exemple GSE-GSM) dans les métadonnées (BC). L'outil est peu utilisé mais n'a pas d'alternative, si ce n'est via SPASE (comme à l'ONERA) (SH,BC). Pour implémenter cela, il y a besoin de lister toutes les hypothèses (NAN).

→ **La recommandation du CU quant à l'importance d'avoir des données de qualité a été suivie (et dépassée). Le traçage des transformations des jeux de données serait une suite logique à cette démarche.**

BC propose de présenter un outil dans le futur en JSON pour décrire les spectres dynamiques (polygone ou nuage de points) (quel outil?)

- AMDA : Nouvelles fonctionnalités

Implémentation des catalogues pour le machine learning (ML) qui sont des tables avec une valeur. Les différents ajouts sont passés en revue (formats pour les exports, amélioration de l'édition des paramètres avancés, possibilités de téléchargements, retour du multi plot, optimisation des traitements, changement de repères, filtrage des points singuliers, inclusion de différents modèles).

Il manque des paramètres (Beta, pression dynamique, ...) dans l'onglet SPACE des paramètres avancés (vitesse d'Alfvén uniquement actuellement) (BG). Dans l'idéal un utilisateur devrait être capable d'ajouter lui-même un paramètre (NAU), ce qui revient à pouvoir définir non plus simplement un paramètre mais une fonction informatique.

→ **Le CU recommande l'ajout de paramètres dans l'onglet SPACE des paramètres avancés et d'étudier la possibilité pour un utilisateur de les développer soi-même.**

Discussion entre NAU et BR pour étudier la possibilité d'installer une instance d'AMDA au LPP. Ce serait utile, notamment pour avoir un retour sur la nouvelle procédure d'installation d'AMDA (en 3h).

Concernant l'élimination des points singuliers (spike filtering) dont la démonstration a été faite sur un plot, PH se fait préciser si le traitement concerne uniquement la visu ou si cela se propage dans les calculs (et le téléchargement). CDPP: à vérifier mais cela semble bien se propager.

Lors de la discussion sur les modèles numériques, NAU souhaite savoir s'ils sont implémentés en local, et accessibles par les utilisateurs. Ce dernier point ne semble pas être simple à mettre en œuvre. BGR fait préciser un terme (position = distance) et les infos utilisées en entrée du modèle de magnétopause. Importance pour l'utilisateur d'avoir les hypothèses des modèles.

- AMDA : Speasy

AS présente cette interface python qui permet d'accéder aux données disponibles sur AMDA sans avoir à passer par l'interface web. Cela facilite le développement des tâches de ML et l'ensemble est disponible sur Github. Le CU souligne que le tutoriel est très bien fait : pour un non spécialiste il est très facile de faire ses premiers plots rapidement. NAN souligner que Speasy a développé la collaboration entre différents labos.

La discussion sur le cache de données montre que cela peut être un frein pour l'identification de l'usage des données. NAU souligne qu'il peut y avoir un cache au niveau local et à celui du labo (plusieurs niveaux de cache).

→ **Nouveau moyen d'accéder à la bdd AMDA via un outil (Github) très (re)connu.**

- Machine Learning

Le CDPP a développé une plateforme de code de gestion de code python des différentes simulations (LITF Lost In The Forest) disponible sur Github. Dans le cadre de l'identification de sous-orage isolés, OLE signale à CJ l'existence d'une liste d'évènements sur la base de THEMIS.

NAU rappelle que pour tout ce qui est catégorisation d'évènements, si deux experts ne sont d'accord, ce n'est pas le ML qui permettra de les accorder : il y a une part de subjectivité et c'est donc important d'y intégrer un maximum de physique. Par exemple commencer par identifier les régions avant de passer aux phénomènes.

- AMDA : Interopérabilité

Mise à jour des descriptifs des données AMDA (recommandation précédente du CU). Utilisation de descripteurs SPASE dans un dépôt local à AMDA, à déposer sur le site officiel.

HAPI permet de tracer les données AMDA dans AUTO PLOT.

SHE souligne que l'ONERA est interfacé au SPASE d'AMDA pour divers produits. Il a noté quelques métadonnées manquantes qui sont à renseigner.

→ **Les descriptifs des données AMDA ont été mis à jour (recommandation du CU suivie). Continuer à remplir les données manquantes et s'assurer des conséquences vis à vis des utilisateurs d'un changement d'adresse du répertoire.**

**Mise en place d'un DOI** pour quelques jeux de données AMDA (modèle de propagation 1D-MHD. KI demande si cela peut être mis en place comme support pour les chercheurs qui ont besoin d'un DOI pour leurs données lors de publications. Le CDPP pense que c'est un peu lourd comme projet.

- Publications et statistiques

**L'analyse des publications fait suite à une demande du CU.** Les différentes thématiques du CDPP sont bien présentes. Le nombre de publications référées baisse légèrement avec le temps (méthode de comptage se base sur la bonne volonté des auteurs).

Le module le plus utilisé de l'IHM (interface homme machine) reste le plot. ws\_print permet de compte les utilisations via speasy

La pluridisciplinarité se retrouve dans les jeux de données utilisés. En 2019 : Cluster, rosetta, ACE et MAVEN. En 2021 : Cluster, SO, VEX, modèle 1D-MHDtao, ACE

- Évolution

Le CDPP consulte le CU pour savoir s'il faut maintenir l'effort sur les plots. **Le CU encourage à maintenir l'effort sur le module de plot car c'est le module le plus utilisé.**

À propose des fonctions d'analyse sophistiquée, NAU souhaite faire partager les codes par la communauté et KI avertit que différents utilisateurs peuvent en avoir une interprétation différente

BC note que le débat sur les ondes (toujours absentes d'AMDA) pourrait être réglé avec MASER qui devrait être labellisé (et qui utilise autoplot). La difficulté de son intégration viendra de l'échantillonnage irrégulier d'où une grande quantité de données à rééchantillonner)

- SCIQLOP

Le CDPP soutien ce projet porté par le LPP qui illustre la collaboration entre les différents labos.

**Le CU soutient cette coopération inter-laboratoires**

- ARCHIVE

Après l'exceptionnelle année 2018 (fin de la mission DEMETER), l'utilisation continue d'augmenter. Le rythme est maintenant de 190 commandes par mois. On peut noter un regain d'intérêt pour les données INTERBALL et une utilisation internationale (Chine, USA puis France).

NDU sollicite l'avis du CU sur le devenir des indices géomagnétiques stockés. Une question de KI nous apprend que les utilisateurs ne les téléchargent pas. Il est souligné que ces données ne sont pas prioritaires, d'autant plus que la collection d'indices disponible est limitée (BGR). Cependant un utilisateur pourra vouloir ajouter à sa commande des indices (NAN).

**Le CU encourage le CDPP à identifier les indices utiles et explorer sur la possibilité de les récupérer via AMDA (et non plus les stocker dans l'archive).**

Côté industriel, le contrat VALDO a été renouvelé.

- ARCHIVE / Exploitation

L'exploitation des l'archive concerne 169 To de données avec un nombre de jeux de données en augmentation

L'exploitation récurrente concerne des nouveaux jeux de données pour STEREO et PSP. Pour Solar Orbiter, cela attendra la fin de la migration des archives.

Concernant les données GEOS, BGR souhaite vérifier que les données S300 ULF et S331 Magnetometer ont des systèmes de coordonnées compatibles.

**Action : BGR vérifie la compatibilité des différents systèmes de coordonnées des jeux de données GEOS.**

Devant les difficultés les autres données Cassini et Rosetta sont en voie d'être abandonnées. Il semble que les données soient archivées ailleurs (outils de l'ESA et de la NASA).

Le passé montre que si les actions d'archivage en sont pas enclenchées en amont ou au moins pendant la mission, l'archivage devient beaucoup plus compliqué et donc plus coûteux en temps.

- ARCHIVE / Migration

Cet énorme chantier qui concerne à la fois la transformation des bases de données Oracle vers REGARDS, des serveurs redhat vers VM SIS et de l'infrastructure STAD vers le DATALAKE a débuté en 2019 et devrait être achevé au plus tôt mi-2022. Aucune nouvelle mission n'est prise en charge avant que ce soit terminé.

REGARDS sera compatible avec le format SPASE très utilisé au CDPP.

**Le CU s'assure que les données seront accessibles pendant la migration.** BC souhaite savoir si DATALAKE aura une interface contrôlée ou ouverte. DH précise que c'est un outil interne au CNES qui devrait être intégré avec DATATERA.

La catalogue SPASE du CDPP est produit et maintenu par le CNES est déposé sur trois sites ce qui nécessite une bonne organisation pour éviter les problèmes

SHE souhaite savoir si les informations ajoutées sont bien remontées au consortium SPASE. Oui, mais pas encore de retour (DB). BC souligne que c'est un besoin récurrent donc il faut bien identifier le correspondant.

**Le CU encourage le CDPP à vérifier que les nouvelles informations sont bien remontées au consortium SPASE**

- 3D-view

Il y a eu ces dernières années beaucoup d'évolutions dans le cadre de financement CNES pour la Tierce Maintenance Appliquative (TMA) et de l'ANT Tempête.

Un des succès les plus visibles est l'utilisation de 3d-view pour un clip de l'ESA (JUICE). LA promotion est assurée via des écoles, en plus de publications.

Le code a été déposé sur GitHub (visibilité de l'outil) et enregistré sur ASCL. Partage également du code de « Magnetopause shear angle and reconnection locations », aka Trattner et al. (2007) implémenté dans 3Dview.

Dans les discussions pour améliorer la visibilité des outils et des modèles à ajouter, le CU évoque la possibilité de faire la possibilité de la publicité pour les outils du CDPP sur la page d'accueil d'AMDA qui est l'outil le plus utilisé. BGR note que dans l'idéal on devrait pouvoir obtenir une scène 3d-view quand on utilise AMDA. VGE souhaite que 3d-view s'ouvre avec une scène par défaut. Le CU souligne que se lancer dans 3D-view est ardu, voire décourageant (juste une simple barre de menus au lancement). Et que les modèles d'un outil puissent se retrouver dans les autres.

**Le CU suggère d'évaluer la possibilité qu'un plot à AMDA puisse permettre de générer une scène 3D-view ou mieux de lancer l'outil.**

La discussion dérive sur **TREPS dont la limite de taille a été augmentée suivant une résolution du CU.**

- CST

**Les dernières améliorations contiennent de nombreuses suggestions du CU lors des derniers tests comme la recherche d'instruments dans le texte ou sur les cartes.** Les vues ont été améliorées et sont plus confortables. De nombreux développements sont encore en cours.

EAS suggère d'ajouter des cartes globales de TEC. Bonne idée pour les ovales auroraux.

**À L'AIDE : Et là j'ai perdu le son...**

- PropTool

Des catalogues d'évènements in situ ainsi que les flux solaires X et gamma ont été intégrés. L'intégration de catalogues et les différentes améliorations sont en cours. Il y a de nombreuses perspectives de développement pour cet outil.

- Helio Propoa

Modèle léger qui appelle AMDA pour la visualisation (et réciproquement les données du modèle 1d – mhd sont disponibles dans AMDA). Participe au projet européen H2020 SafeSpace.

- SSA

Le CDPP est un sous-traitant du RAL. Dans le cadre de SWESNET, **la question se pose d'être opérationnel 24/24.** Quels sont les besoins à prévoir ?

AR détaille l'intégration des nouveaux outils ainsi que la modélisation des ondes de choc dans la couronne.

Suite à une demande de précision de BG l'accent est porté sur les SEP au lancement du CME. Les SEP pendant le trajet du CME seront traités dans le futur.

Serpentine est un nouveau projet européen lancé en 2021 environ 300k€ (étude des SEP). Ressources CDPP : AMDA, PropTool – 3Dview, 1 post doc qui est déjà arrivé et une thèse.

**Le CU note que la participation du CDPP à différents projets européens apporte financement et visibilité.**

- Maintenance des outils CDPP :

Les contrats arrivent à échéance fin 2022. Besoin de préparer le renouvellement.

AMDA, 3d-view, TREPS, PropTool ont une activité de maintenance nominale.

Le Space Weather Tool n'a plus eu de développement depuis 2019, car l'outil n'a pas rencontré son public.

AR explique que l'outil est difficile à prendre en main et que la simplification faite n'est pas suffisante. Cependant la modélisation scientifique continue à être faite.

Deux voies sont possibles : prendre quelques paramètres et les intégrer au PropTool (version minima) ou l'intégrer à des simulations 3d du vent solaire car aucun outil dans le portail SSA est capable de suivre un CME magnétisé (ambitieux). La main d'oeuvre disponible pousse à la première solution.

**Le CU serait satisfait d'une fusion avec le PropTool si on peut garder le plus de fonctionnalités possibles du SWT.**

- Divers CNES

Le budget est maintenu et les RH CNES semblent bien dimensionnés.

Renouvellement de la convention : prendre exemple sur celle de MEDOC pour gagner en efficacité. Objectif signature en 2022.

Faut-il intégrer le LPC2E et le LATMOS en labo partenaires ? **Les membres concernés du CU sont appelés à voir avec leur direction.**

PHE demande des précisions sur nature de la convention et le statut du CDPP pour le CNRS, car il trouve que la section 17 ne le connaît pas très bien. NAN : l'annexe décrit les apports de chaque labo et c'est un service de l'INSU.

- Communication du CDPP

Le CDPP a préparé un questionnaire sur la visibilité des outils du CDPP.

**Le CU est d'accord pour donner son avis avant diffusion**

Le CDPP envisage de préparer une newsletter.

**Le CU souligne l'importance de faire court et incisif.** RMA propose de communiquer via les réseaux sociaux (oiseau bleu par exemple).

KI souligne que l'école d'été SO était très bien.

- Renouvellement du CU :

Quelques membres « historiques » sont prêts à passer la main (OLE, KI). OLE rappelle l'importance d'avoir un contact par labo et va demander si M. Berthommier souhaite rejoindre le CU. KI va voir s'il y a des candidats à l'IAS

PH demande s'il faut internationaliser le CU. **Personne n'est contre mais tous sont attachés à avoir une journée de présentation en français. Des profils plus divers (étudiant, étranger) pourraient être intégrés lors des tests.**

NAN souhaite avoir un regard extérieur le plus large possible et suggère de prendre des non-permanents. RMO souligne que les non-permanents sont souvent sur des contrats courts donc ce n'est pas l'idéal.

**Le CU ne soutient pas l'intégration de membres non-permanents.**

NAU souhaite revenir sur l'importance de mettre les outils du CDPP sur github pour la visibilité et aussi pour avoir plus de retour (facile de signaler un problème)

BG rappelle que CF souhaite quitter le CU depuis un moment mais cela n'a pas pu être officialisé de par l'absence de CU.

Personne ne s'oppose à son départ. **Le CU propose donc au CD d'accepter sa démission.**

Candidat de remplacement : personne n'est identifiée dans l'immédiat (activité test réduite et tous les labos sont représentés) mais un profil solaire/ vent solaire/ propagation serait dans les besoins actuels.

**Le CU ne demande pas d'appointer un nouveau membre dans l'immédiat.**