



## Passage de SAS à R : enjeux et outils ?

ANNA SMYK ET TANGUY BARTHÉLÉMY  
Division Recueil et Traitement de l'Information  
Département des Méthodes Statistiques

# Sommaire

---

## 1. Quelques nouvelles

## 2. Chaines de production de séries cvs-cjo

## 3. Outils R

## Version 2.2.4 de JDemetra+

---

Vous devriez utiliser la version 2.2.3 sortie en juillet 2020.

Une version 2.2.4 vient de sortir : aucune nouvelle fonctionnalité mais correction de bugs.

Vous trouverez sur [Github](#) une fiche d'installation des ressources.

# Communauté symphonie et Ressources

---

Sur [Symphonie](#) vous trouverez :

- Biblio
- Code R
- Slides de formation Insee
- Regresseurs CJO Insee

Sur [Github](#) vous trouverez :

- Biblio
- Code R spécifique Atelier
- Slides de présentation de l'atelier

L'espace commun AUS n'est plus mis à jour.

Tous les mois nous enverrons la liste des nouveautés et mises à jour à notre mailing list.

# Formations

---

Prochaines formations :

- module initiation 15, 16 juin
- module perfectionnement 29, 30 juin

Ce seront les dernières sous ce format !

A partir de l'automne : 2 nouveaux modules

- formation producteur de séries cvs-cjo : 3j (indispensable pour tous les nouveaux)
- formation séries temporelles avec R et JDemetra+ (2j+2j), débutant R possible

Contenu précis et dates à venir

# Sommaire

---

## 1. Quelques nouvelles

## 2. Chaînes de production de séries cvs-cjo

### 2.1 Généralités

### 2.2 Spécificités d'une chaîne de production de séries cvs-cjo

### 2.3 Configurations courantes au sein du SSP

### 2.4 Recommandations

## 3. Outils R

# Objectifs de l'atelier

---

- clarifier la structure et les points clés d'une chaîne de production de séries cvs
- vous présenter des outils R utiles pour la production (et l'expertise), afin que vous puissiez remplacer vos programmes SAS
- recueillir vos questions et besoins concernant les évolutions que vous comptez mettre en place prochainement

# Ce que l'on souhaite produire / exporter

---

- *In fine* : des séries cvs-cjo ou coefficients saisonniers
- Au cours du processus
  - des séries intermédiaires (série linéarisée,...)
  - des paramètres et diagnostics statistiques (boucle, intervention manuelle)
  - des scores d'anomalie, une priorisation
  - des indicateurs de comparaison des séries



# Spécificités d'une chaîne de production de séries cvs-cjo

---

En quoi diffère-t-elle de la plupart des processus ?

- Appel à un algorithme externe (très) sophistiqué

En général X13-Arima (X12-Arima si configuration non mise à jour)

Nécessite de choisir un logiciel permettant d'accéder à l'algorithme

- SAS
- Interface JDemetra+
- R

# Choix de l'algorithme et du logiciel y donnant accès

---

Choix impactant quant à

- la structure des données
- les versions accessibles
- les possibilités de rafraîchissement des données

JDemetra+ est logiciel recommandé par Eurostat à tous les pays de l'Eurosystème

**Objectif** : utiliser la dernière version de algorithme recommandé

# Situations courantes

---

En général :

- X-13 Arima
  - appel via interface (GUI) et Cruncher
  - programmes auxiliaires en SAS (ou déjà R..)

Problème :

- sortie de SAS, remplacer le lancement du Cruncher et les programmes auxiliaires par des programmes R

Plus rarement :

- X12 Arima
  - appel via SAS (Proc X12)
  - programmes auxiliaires en SAS

Problèmes :

- utiliser la dernière version de l'algorithme, avec le logiciel recommandé
- sortie de SAS, remplacer les programmes auxiliaires par des programmes R

# Campagnes Annuelles

---

Pour les campagnes **annuelles**, importance

- de l'expertise manuelle (interface)
- du sélective editing ([JDCruncher](#))
- de la possibilité d'une modification massive de spécifications ([rjdworkspace](#), [RJDemetra](#), manipulation des fichiers xml en R)
- indicateurs et graphiques liés aux révisions ([RJDemetra](#))

# Campagnes Infra-Annuelles

---

Pour les campagnes **infra-annuelles**, accent mis sur

- les politiques de rafraîchissement (Cruncher)
- la vitesse de production de l'output (Cruncher)
- la possibilité d'une modification massive de spécifications (outliers. . .)  
([rjdworkspace](#), [RJDemetra](#), manipulation des fichiers xml en R)
- indicateurs et graphiques liés aux révisions ([RJDemetra](#))

# Organisation recommandée

---

Avoir la possibilité d'utiliser l'interface (expertise) et cruncher (rafraîchissement avec contraintes et production output)

- implique l'utilisation d'une structure de **workspace**
- programmes auxiliaires sous R (avec en option la lecture directe des données avec [RJDemetra](#))

Pour qu'un workspace soit crunchable il lui faut des "metadata" complètes, ce ne sera pas le cas s'il est créé avec [RJDemetra](#).

# Structure de workspace : avantages et inconvénients

---

Avantages : possibilité d'utilisation de l'interface et du Cruncher

- questions clé : besoin de l'interface GUI ? Besoin du cruncher ? ...

Inconvénient : portabilité (chemin)

Chaîne de Production avec [RJDemetra](#) sans workspace ?

Cas “pathologique” : imbrication de boucles avec cvs et traitements annexes

# Sommaire

---

1. Quelques nouvelles

2. Chaines de production de séries cvs-cjo

**3. Outils R**

3.1 Export et expertise

3.2 Personnalisation des specifications



# Rafraîchissement des données et export séries

---

Rafraîchissement : Cruncher

- [rjwsacruncher](#)
- [JDCruncher](#)

Selective editing avec bilan qualité et expertise manuelle

# Impact numérique des changements de paramètres

---

Révisions :

- lecture de plusieurs workspaces avec [RJDemetra](#) pour comparer différentes versions d'une même série

# Sélection de regressseurs cjo

---

Programme de choix parmi les différents jeux de regressseurs proposés

# Customization de spécifications à grande échelle

---

Modification individualisée pour un grand nombre séries :

- attribution de regressseurs cjo
- outliers
- autres..

RJDemetra, [rjdworkspace](#)...

# Fusion de workspaces

---

Cas d'utilisation classique : fusionner workspace de référence (en vigueur) et workspace automatique selon un score

Et pouvoir cruncher le résultat

# Conclusion

---

## Identifier

- chaîne cible
- Programmes à remplacer / ajouter

Possibilité d'utiliser les templates fournis et/ou de nous demander une assistance pour la refonte (occasion de revoir des choix méthodologiques structurants : longueur des séries,...)

## Ressources :

Prochain atelier juin 2023, thèmes pressentis

- présentation de nouvelles chaînes, études de cas
- questions/réponses