

Utilisation de Demetra+ à l'INSEE

Himpens Stéphanie
Méthodologue CVS – UMS-E



Le 21
décembre
2011

Séries désaisonnalisées

IPI : 91 séries (niveau **NACE3**)

Indices de chiffres d'affaires : 440 séries (niveau **NACE3**
dans l'industrie et **NACE5** dans les autres secteurs).

Toutes les séries sont **mensuelles**.



Date
21/12/2011

Organisation des désaisonnalisations

En conformité avec les recommandations des lignes directrices européennes, deux types de campagnes existent :

- ✿ Chaque année, recherche complète des paramètres de toutes les séries.
- ✿ Chaque mois, utilisation des paramètres choisis précédemment. Les points aberrants sont identifiés uniquement sur la période récente.

=> Ajustement de type **concurrent partiel**.



Date
21/12/2011

Organisation des désaisonnalisations

Désaisonnalisation de type indirecte :

- ✿ Les séries CVSCJO des agrégats sont estimés comme des moyennes pondérées des séries désaisonnalisés qui les constituent.



Date
21/12/2011

Méthodes et outils utilisés

Depuis janvier 2011, le logiciel Demetra+ est utilisé afin de désaisonnaliser les séries avec la méthode X-12-ARIMA.

Un ensemble de macros SAS a été développé afin :

- D'intégrer le logiciel à la chaîne de traitement lors des campagnes annuelles et mensuelles.
- De faciliter la recherche de modèles lors des campagnes annuelles.



Date
21/12/2011

Déroulement des campagnes annuelles

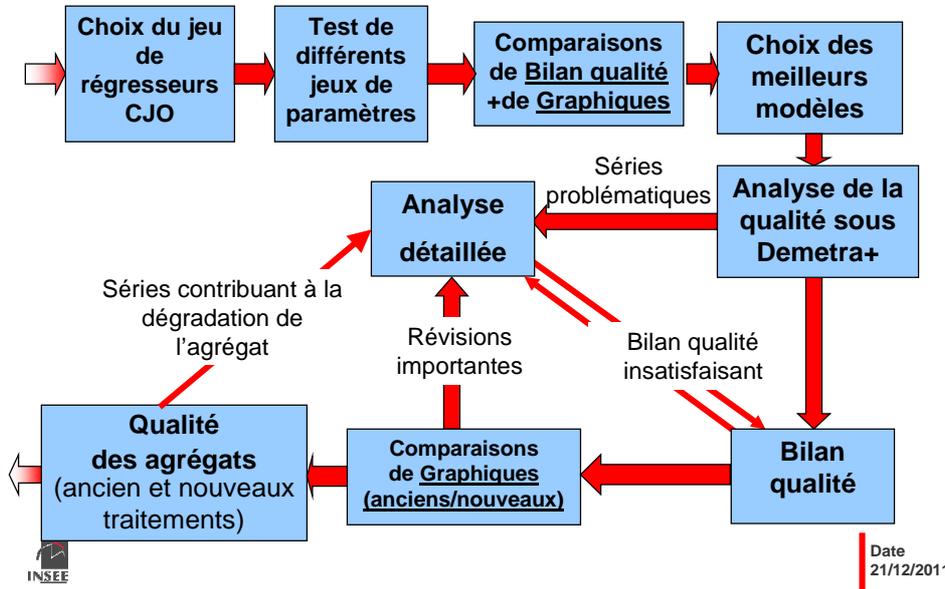
Les campagnes annuelles comportent 8 phases :

1. Spécification du jeu de régresseurs de correction des jours ouvrables à utiliser avec chaque série.
2. Plusieurs versions de workspaces sont élaborées (avec des modèles identifiés automatiquement, avec les anciens modèles etc.). Pour chaque version, un bilan de la qualité est édité.
3. Le meilleur modèle des différentes versions est sélectionné pour chaque séries à partir de la comparaison des qualités obtenues et des graphiques des composantes.
4. Certaines séries problématiques (modèles insatisfaisants, révisions fortes par rapport à la série publiée etc.) font l'objet d'une analyse approfondie sous Demetra+. Certaines spécifications sont modifiées interactivement.
5. Bilan de la qualité de l'indice avec les spécifications modifiées
6. Comparaisons des nouvelles séries avec les anciennes
7. Notation globale de la qualité (y compris des agrégats)
8. Une nouvelle recherche de spécifications est possible pour les séries présentant une détérioration de leur qualité de traitement



Date
21/12/2011

Déroulement des campagnes annuelles



Calcul et identification du jeu de régresseurs

Les différents jeux de régresseurs pour la correction des effets de calendriers sont construits selon une **formule de calcul propre au DSCT** prenant en compte le calendrier français.

Les jeux de régresseurs disponibles sont en nombre supérieur aux jeux prédéfinis.

Calcul et identification du jeu de régresseurs

Les jeux de régresseurs possibles dépendent du secteur de la série.

Secteur d'activité	Nombre de jeux de régresseurs CJO
Industrie	3
Commerce et services	4

Aucune identification du jeu de régresseurs CJO le plus approprié n'est disponible sous Demetra+ (seulement un test de significativité globale).



Date
21/12/2011

Calcul et identification du jeu de régresseurs

Le choix du jeu de régresseurs le plus approprié est effectué automatiquement via un programme SAS.

Il résulte d'une combinaison entre les résultats des tests de Fisher (obtention de la matrice de variance-covariance grâce à la version dos de X12) et les critères AICC (recalculés à partir de Demetra+).

Sortie du programme SAS :

	séries	Industrie	Jeu de régresseurs sélectionné
1	_4711B	Commerce de détail	Effet année bissextile
2	_4531Z	Commerce de détail	6 variables (jours ouvrables)+ Effet année bissextile
3	_4711D	Commerce de détail	6 variables (jours ouvrables)+ Effet année bissextile
4	_4774Z	Commerce de détail	2 variables (constraste semaine/weekend) + Effet année bissextile
5	_4791B	Commerce de détail	2 variables (constraste semaine/weekend) + Effet année bissextile
6	_4511Z	Commerce de détail	3 variables (Effets spécifiques des lundis, samedis et des autres jours de la semaine) + Effet année bissextile



Date
21/12/2011

Modifier des spécifications

Il est possible de lire et d'écrire des spécifications utilisables dans Demetra+ à partir de SAS.

Paramétrage du fichier de données à utiliser

Ordres du modèle ARIMA

Nom des séries	Type du fichier de données	Localisation du fichier de données	Méthode	Moyenne	Transformation	Ordres du modèle ARIMA					Intervalle de modélisation	Début de l'intervalle de modélisation	Fin de l'intervalle de modélisation
						phi	d	th	bphi	bd			
1_4511Z	XCLPRVDR	<<D:\user\cda_c.xls\X12			level	0	1	1	1	1	0		
2_4519Z	XCLPRVDR	<<D:\user\cda_c.xls\X12			level	0	1	1	0	1	1		
3_4520A	XCLPRVDR	<<D:\user\cda_c.xls\X12	true		log	0	1	1	0	1	1		
4_4520B	XCLPRVDR	<<D:\user\cda_c.xls\X12			log	0	1	1	1	1	0		
5_4531Z	XCLPRVDR	<<D:\user\cda_c.xls\X12			log	1	1	0	0	1	0		
6_4532Z	XCLPRVDR	<<D:\user\cda_c.xls\X12			log	0	1	1	0	1	1		
10_4711C2	XCLPRVDR	<<D:\user\cda_c.xls\X12			log	1	1	0	0	1	0 from	01/01/2002	
7_4540Z	XCLPRVDR	<<D:\user\cda_c.xls\X12			log	0	1	1	0	1	1		
8_4711A	XCLPRVDR	<<D:\user\cda_c.xls\X12			log	0	1	1	0	1	1		

- Un recodage des spécifications du fichier XML est effectué.
- Une identification automatique de certains paramètres est possible.



Date
21/12/2011

Lancer la désaisonnalisation depuis SAS

WSACruncher.exe

Cet outil permet de lancer les désaisonnalisations depuis SAS, logiciel dans lequel se trouve écrits la majorité des autres programmes.

Les résultats peuvent ensuite être ré-importés dans SAS dès les calculs terminés.

- ⇒ Cela permet d'éviter un changement d'environnement potentiellement source d'erreurs.
- ⇒ Demetra+ est uniquement ouvert afin de vérifier la qualité des traitements et les résultats.



Date
21/12/2011

Recherche de modèles

Éléments utilisés afin de vérifier la qualité d'un ensemble de séries :

✿ **Sous SAS :**

Les bilans qualité.

Comparaisons des bilans qualité des différentes versions.

Comparaisons graphiques.

✿ **Sous Demetra+ :**

Les diagnostics synthétiques et détaillés de Demetra+.

Comparaisons graphiques ponctuelles.



Date
21/12/2011

Le Bilan qualité

Des notes (valeurs numériques) sont calculés à l'aide de macros SAS.

La qualité globale du modèle ARIMA est évaluée par une moyenne pondérée des autres notes.

Dans les colonnes ci-dessous, la note est égale à 100 si le test est accepté à 5% pour la série, à 50 si le test est rejeté à 5% mais accepté à 1% et à 0 si il est rejeté dans les deux cas.

	series names	Note de qualité globale des outliers	...	Note de qualité globale du modèle ARIMA	Note associé au test de Box-Pierce	Note associée au test de Ljung-Box	Note sur la statistique de Ljung-Box sur le carré des résidus	Note sur la statistique de Box-Pierce sur le carré des résidus	Note sur la normalité	Note sur le test du Kurtosis	Note sur le test du Skweness
1	.4511Z	100		67	100	100	100	100	50	100	0
2	.4519Z	97		100	100	100	100	100	100	100	100
3	.4520A			100	100	100	100	100	100	100	100
4	.4520B	100		100	100	100	100	100	100	100	100



Date
21/12/2011

Le Bilan qualité

Les différents groupes de diagnostics sont :

- La qualité de la modélisation ARIMA.
- La qualité des points aberrants (nombre et concentration).
- La qualité de décomposition (effets résiduels et orthogonalité des composantes).
- La qualité des révisions de type "History" (+stabilité des outliers détectés).



Date
21/12/2011

Comparaison de bilans qualités et de résultats

Les fichiers sont élaborés de façon à mettre en avant les séries présentant une **amélioration** ou une **dégradation** de leur qualité.

Comp_Spécif / **Comp_Qualité** / Comp_Révisions

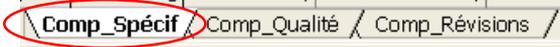
Série	Pond	Tendance	Différences Qualité Nouveau/Ancien					Diff Note Qualité Décomp	Diff Qualité Dem+
			Diff Note ARIMA	Diff Note Outliers	Diff Note CJO res	Diff Note CVS res			
4511Z	103 915 133	0	-11.1	0.0	0.0	0.0	0.0	0	
4519Z	9 705 758	0	0.0	0.0	0.0	0.0	0.0	0	
4520A	16 907 061	0	0.0	0.0	0.0	0.0	0.0	0	
4520B	2 175 770	0	0.0	0.0	0.0	0.0	0.0	1	
4531Z	16 759 511	0	-22.2	0.0	0.0	0.0	0.0	0	
4532Z	12 325 051	0	-11.1	5.6	0.0	0.0	0.0	0	
4540Z	4 837 440	0	22.2	5.6	100.0	0.0	0.0	0	
4711A	1 933 163	0	0.0	0.0	0.0	0.0	0.0	0	
4711B	5 472 522	0	0.0	8.3	100.0	0.0	0.0	0	
4711C	5 069 115	0	22.2	0.0	0.0	0.0	0.0	-1	



Date
21/12/2011

Comparaison de bilans qualités et de résultats

Il est ensuite possible de connaître les différences dans les deux jeux de spécifications grâce à un autre onglet.



Différences entre les deux jeux de spécifications (0 = éléments différents, 1 = éléments identiques)

Détails des deux jeux de spécifications

Série	Séries suivies en tendance	Résumé				Modèle		Transformation	
		Modèles ARIMA égaux	Type transformation égaux (Add/Mult)	Jeux de régresseurs cjo égaux	Nombre outliers égaux	Nouveau	Ancien	Nouveau	Ancien
4511Z	0	0	1	0	1	(0 1 1) (1 1 0)	(0 1 1) (0 1 1)	None	None
4519Z	0	1	1	0	1	(0 1 1) (0 1 1)	(0 1 1) (0 1 1)	None	None
4520A	0	1	1	0	0	(0 1 1) (0 1 1)	(0 1 1) (0 1 1)	Logarithm	Logarithm
4520B	0	1	1	0	1	(0 1 1) (0 1 0)	(0 1 1) (0 1 0)	None	None
4531Z	0	0	0	1	0	(1 1 0) (0 1 1)	(1 1 0) (0 1 0)	Logarithm	None

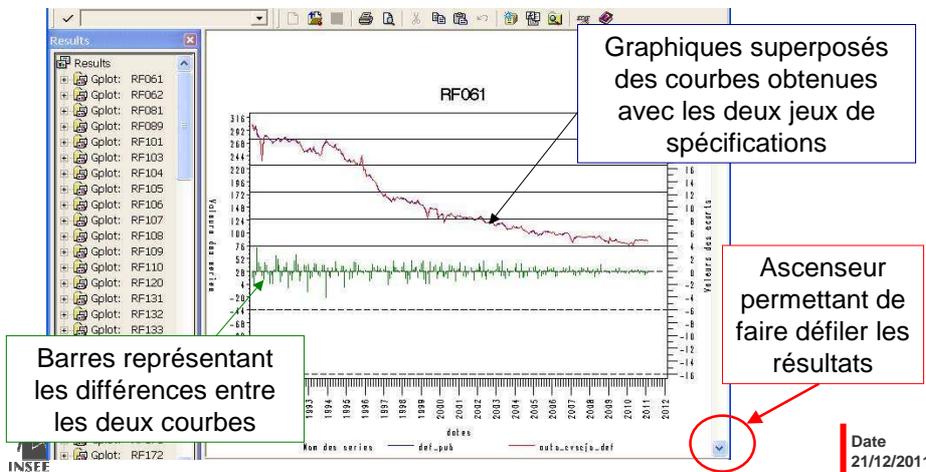
Séries suivies en tendance



Date
21/12/2011

Comparaison de bilans qualités et de résultats

Les graphiques superposant les résultats obtenus avec les différentes versions de spécifications sont tracés sous SAS. Il est possible de les faire défiler rapidement.



Date
21/12/2011

La qualité sous Demetra+

Le fait de disposer d'une évaluation synthétique de la qualité ne dispense pas de vérifier les diagnostics dans Demetra+.

Une vérification systématique des diagnostics du nœud « Main Result » de Demetra+ est effectuée pour chacune des séries

L'ensemble des diagnostics détaillés est consulté en cas de qualité insatisfaisante

The screenshot shows the Demetra+ software interface. On the left, a tree view under 'Spec:' lists various components: Main results (highlighted with a red circle and a red arrow pointing down), Charts, Table, S-I ratio, Pre-processing (RegArima), Pre-adjustment series, Arima, Regressors, Residuals, Statistics, Distribution, Decomposition (X11), and Diagnostics. On the right, the 'Source : User' section displays the 'Pre-processing' status: 'Estimation span: [1-]', 'Series has been log', 'Trading days effect', 'No easter effect', and 'No outliers found'. Below this, the 'Diagnostics' section shows a 'summary' of 'Good' in green text, and 'basic checks' with 'definition: Good (I)' and 'annual totals: Unc' in orange text. At the bottom right, the date '21/12/2011' is visible.

La qualité sous Demetra+

Pour les séries problématiques :

- Modifications interactives possibles afin de tester un jeu de spécifications différent.
- Possibilité de comparer les qualités de différents traitements grâce aux fonctionnalités d'import ou de modification interactive de spécifications.
- Les composantes obtenues peuvent être superposées grâce à l'outil Chart.

La qualité des agrégats

Des moyennes pondérées des notes servent à évaluer la qualité des agrégats (désaisonnalisation indirecte). Les résultats sont résumés dans un fichier Excel :

	Pondérations	Note globale	Qualité du modèle reg-ARIMA				
		Note sur 20 points (modèle ARIMA et qualité de décomposition)	Tests du bruit blanc		Tests d'hétéroscédasticité		Examen des points aberrants
			Ljung-Box	Box-Pierce	Ljung-Box ² (sur le carré des résidus)	Box-Pierce ² (sur le carré des résidus)	
G	553 181 753	19	4	4	4	4	4
45 Commerce et réparation	166 625 724	18	4	4	4	4	4
45.11Z Commerce de voitures et de véhicules automobiles légers	103 915 133	18	4	4	4	4	4
45.19Z Commerce d'autres véhicules automobiles	9 705 758	20	4	4	4	4	4
45.20A Entretien et réparation de véhicules automobiles légers	16 907 061	20	4	4	4	4	4



Date
21/12/2011

2 types de spécifications

- ✿ Un jeu de spécifications **initiales** (par exemple identification automatique du modèle ARIMA requise ou non, test pour la transformation...)
- ✿ Un jeu de spécifications **finale**s (quels sont les éléments identifiés par la méthode : le modèle ARIMA, la transformation, les points aberrants, etc.)



Date
21/12/2011

Les options de rafraichissement

- Elles lisent le chemin du fichier des séries initiales afin d'actualiser les données du SProcessing.
- Elles prennent en considération les choix des spécifications initiales.

=> Les paramètres des spécifications initiales interagissent avec l'option de rafraichissement sélectionnée



Date
21/12/2011

Options de Rafraichissement

Prenons l'exemple suivant :

- Supposons que l'on décide de lancer un traitement avec identification automatique de tous les paramètres sauf les points aberrants (l'option correspondante est alors désactivée).
- Puis, que nous exécutons un rafraichissement de type « outlier ».

On pourrait croire qu'une identification des points aberrants sur toute la série va être pratiquée puisque l'option de rafraichissement est de type « outliers »...



Date
21/12/2011

Options de Rafraichissement

Il n'est est rien ! L'option est toujours désactivée.



=> Aucune identification automatique des points aberrants n'avait été requise afin de rechercher le modèle initial !



Date
21/12/2011

Options de Rafraichissement

Conséquences :

- ✿ En cas de rafraichissement global d'un multi-processing, le traitement effectivement appliqué à chaque série peut être différent en fonction des paramètres de leurs spécifications initiales.
- ✿ Il n'existe pas de façon simple de vérifier les éléments fixés et identifiés pour chaque série.



Date
21/12/2011

Options de Rafraichissement

Ce que nous voulons (campagnes mensuelles) :

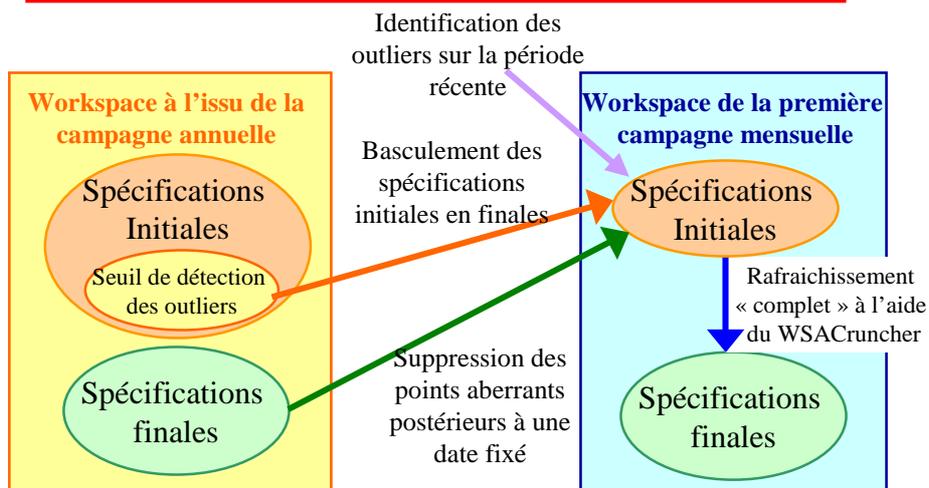
- ✿ Que les modèles soient fixés (modèles ARIMA, points aberrants avant une date fixée, régresseurs pour la correction des effets de calendriers, transformation).
- ✿ Que les points aberrants soient identifiés uniquement à partir d'une date fixée (demetra+ ne permet pas ce type de rafraichissement).

=> SAS 



Date
21/12/2011

Sécurisation des résultats



Année N

Mois M

Date
21/12/2011

Déroulement des campagnes mensuelles

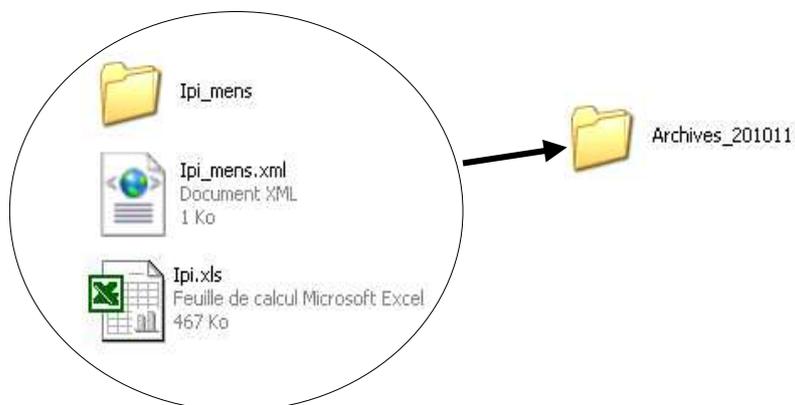
1. Modification du fichier de données lié au workspace « principal ».
2. Lancement de la désaisonnalisation dans Demetra+ à partir de SAS grâce au WSACruncher.
3. Contrôle de la qualité des traitements dans Demetra+.
4. Changement interactif des spécifications qui le nécessitent.
5. Export des résultats.
6. Archivage du workspace, du fichier de données lié et des spécifications.



Date
21/12/2011

Pourquoi sécuriser l'archivage ?

Les éléments à sauvegarder :



Date
21/12/2011

Pourquoi sécuriser l'archivage ?

Sont contenus dans le fichier XML du SAProcessing :

- Les Spécifications
- Le chemin du fichier de données initiales

Chaque modification du fichier de données source peut être la cause d'une modification des données du workspace archivé.

Lien vers le fichier de données initial



Date
21/12/2011

Comment garantir la conservation des données archivées ?

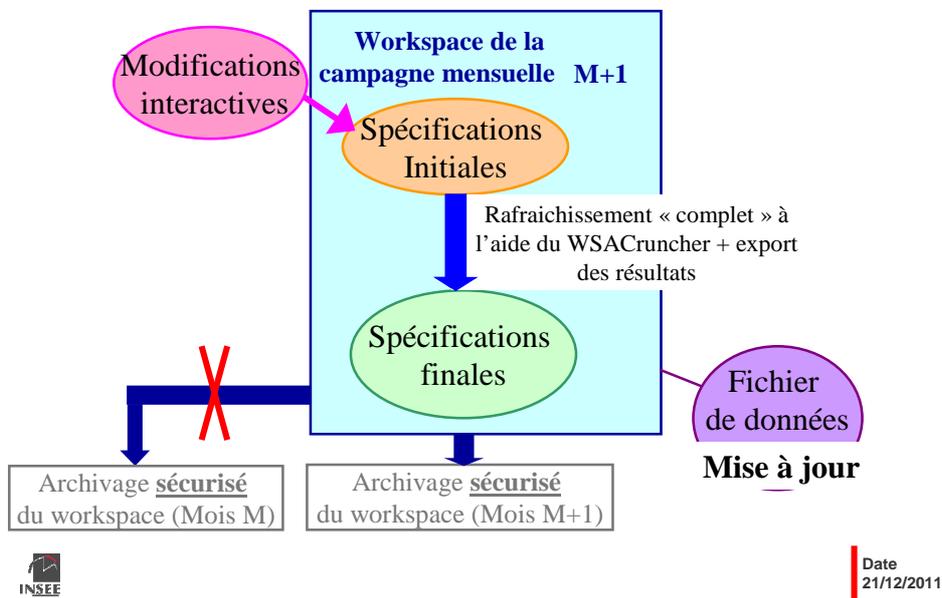
Afin de sécuriser un workspace :

- Le workspace et le fichier de données correspondant doivent être copiés dans un endroit approprié
- Le chemin du fichier contenant dans le fichier XML du SAProcessing doit être modifié afin qu'il pointe vers le fichier copié dans l'archive.



Date
21/12/2011

Déroulement des campagnes mensuelles



Merci de votre attention !

Contact
M.Himpens Stéphanie
Tél. 01 41 17 64 54
Courriel : stephanie.himpens@insee.fr

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr



Informations statistiques :
www.insee.fr / Contacter l'Insee
09 72 72 4000
(coût d'un appel local)
du lundi au vendredi de 9h00 à 17h00

