

# Désaisonnalisation de séries journalières avec la méthode X12

Dominique Ladiray

Institut National de la Statistique et des Études Économiques

Paris, 25/10/2016

# Ouline

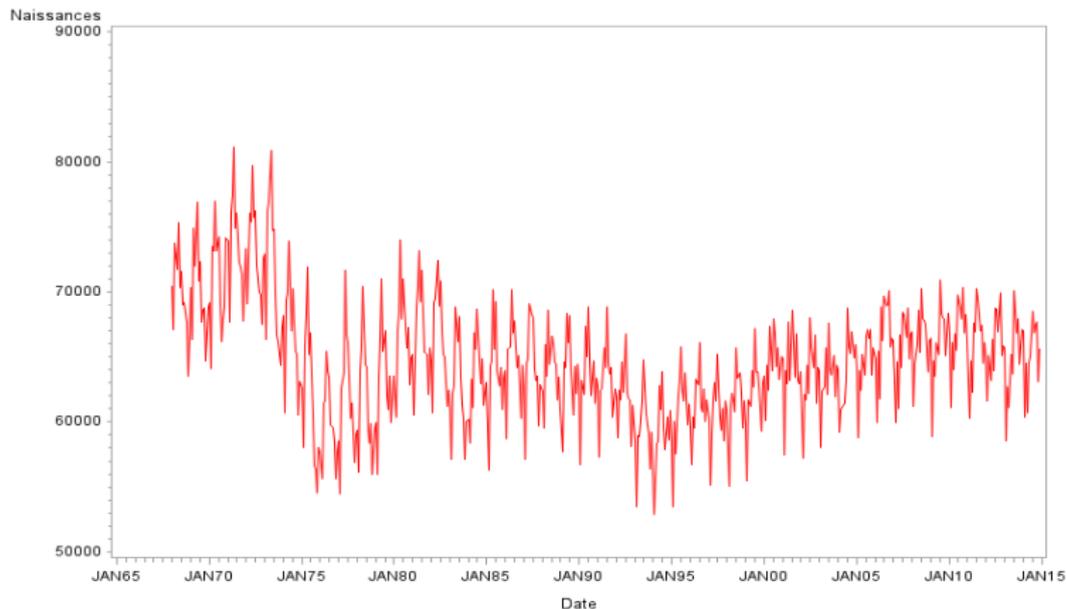
- 1 Introduction
- 2 Des périodicités multiples
- 3 Les pré-ajustements
- 4 L'algorithme de base de X-11
- 5 Un exemple
- 6 Conclusions

# Introduction

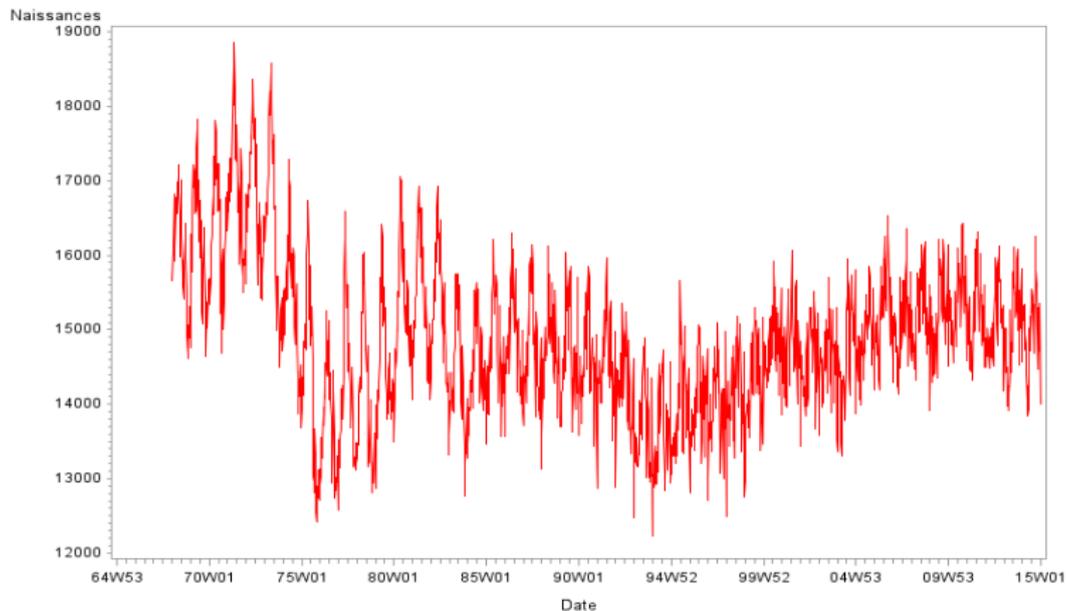
- Les données à hautes fréquences (hebdomadaires, journalières, horaires) sont utilisées depuis des décennies en Finance par exemple, mais essentiellement pour la prévision.
- Ces données sont de plus en plus disponibles et “frappent à la porte” de la statistique officielle, pour la production d'estimations rapides par exemple.
- En matière de désaisonnalisation, ces données posent plusieurs problèmes :
  - Elles présentent des périodicités non entières (52 ou 53 semaines, 365 ou 366 jours par exemples) ;
  - Elles sont très volatiles et présentent plus de points atypiques et de ruptures ;
  - Leur modélisation nécessite parfois de nombreux régresseurs (effets de calendrier, harmoniques, points atypiques).

# Exemple

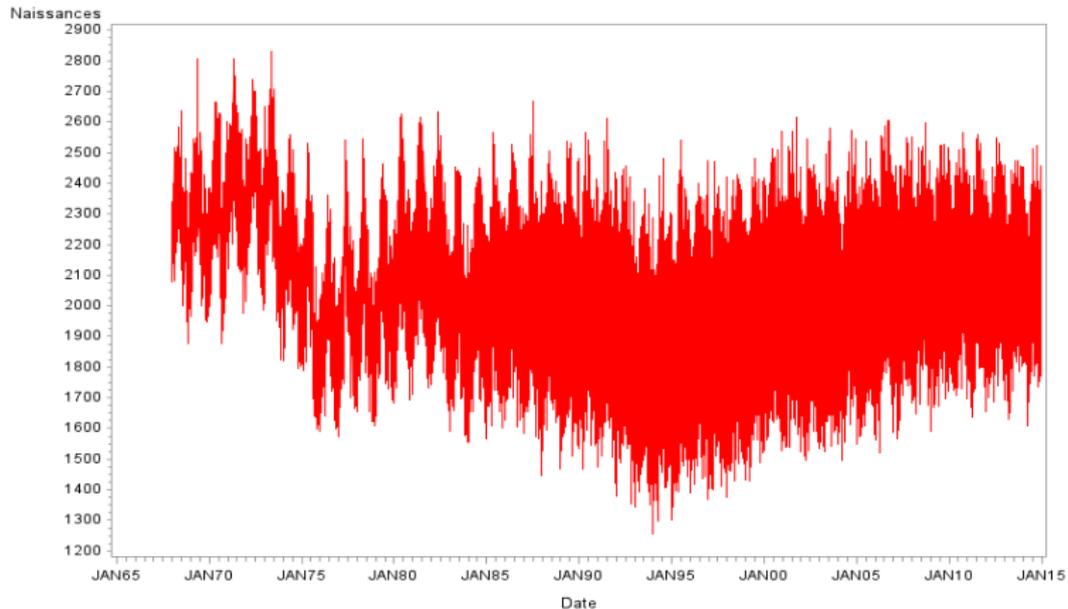
Série des naissances en France depuis 1968, version mensuelle.



# Version hebdomadaire



# Version journalière



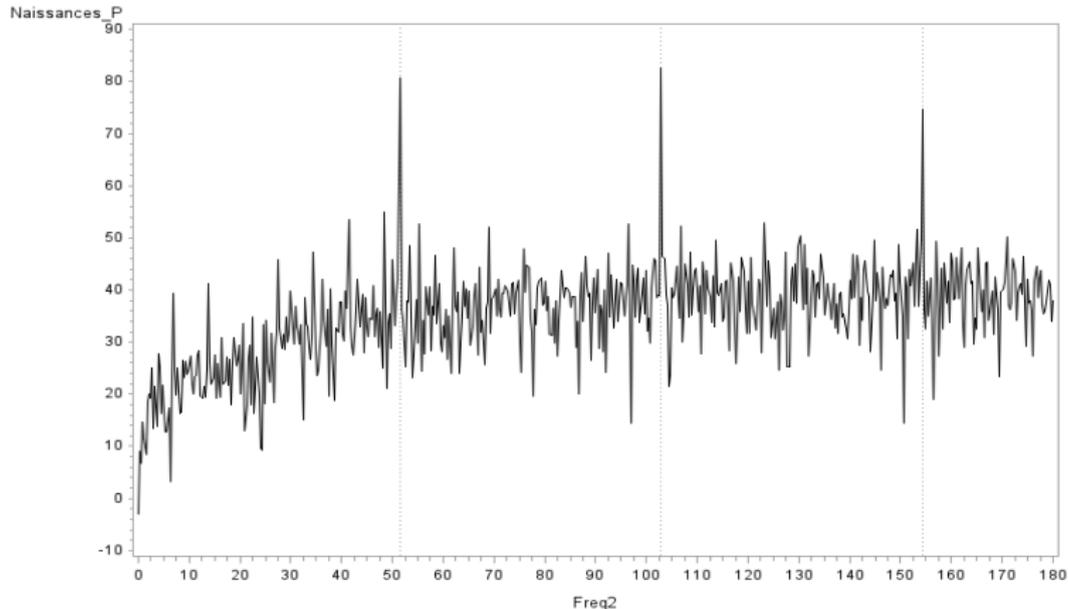
## Des périodicités multiples

Données	Période (nombre d'observations par cycle)						
	Minute	Heure	Jour	Semaine	Mois	Trimestre	Année
Annuelles							1
Trimestrielles							4
Mensuelles						3	12
Hebdomadaires					4.348125	13.044375	52.1775
Journalières				7	30.436875	91.310625	365.2425
Horaires			24	168	730.485	2191.455	8765.82
Demi-Horaires			48	336	1460.97	4382.91	17531.64
Minutes		60	1440	10080	43829.1	131487.3	525949.2
Secondes	60	3600	86400	604800	2629746	7889238	31556952

Une série journalière aura donc une périodicité hebdomadaire (7 jours) et une périodicité annuelle (en moyenne 365.2425 jours).

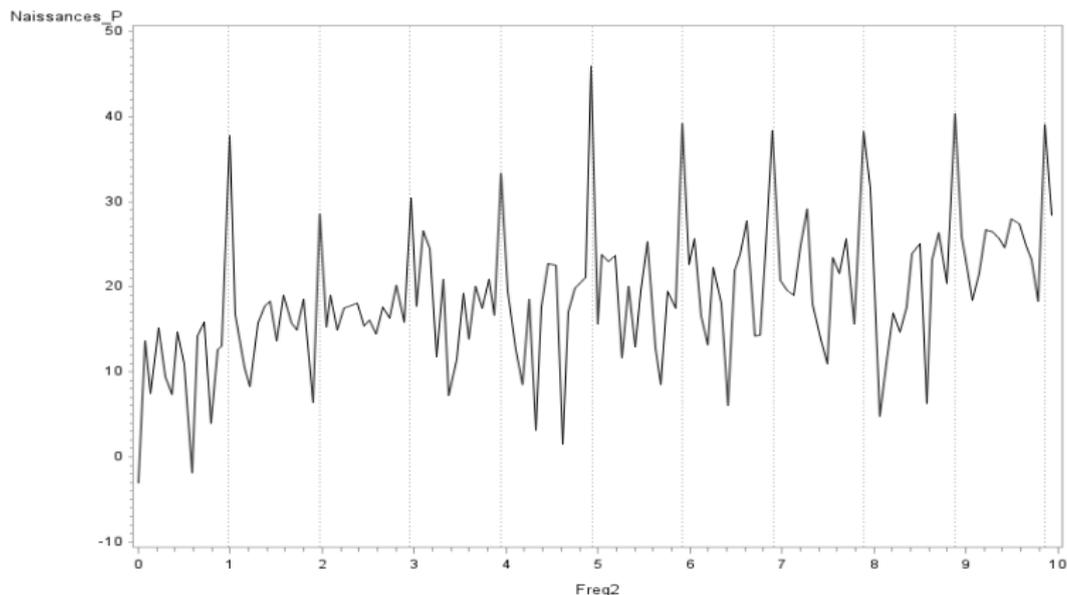
## Périodogramme de la série Naissances

Pics au multiples de  $2\pi/7$ !!!!!! La périodicité annuelle est cachée.



## Zoom sur le périodogramme de la série Naissances

Zoom sur basses fréquences : pics aux fréquences  $2k\pi/365$ .



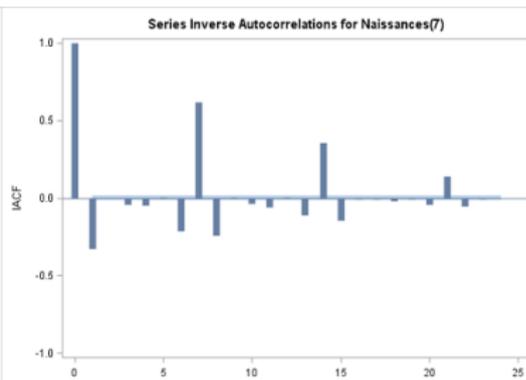
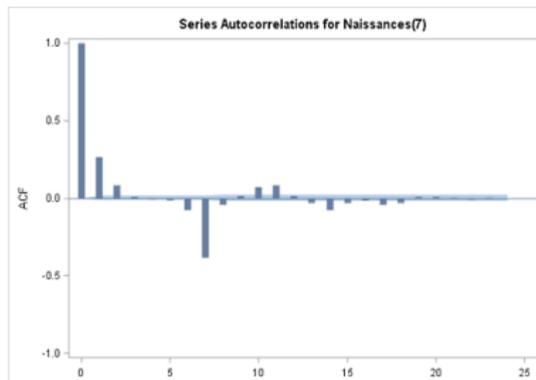
## Rappel sur la méthodologie usuelle

X-13ARIMA-SEATS et TRAMO-SEATS désaisonnalisent une série (mensuelle ou trimestrielle) en deux étapes :

- 1 “Nettoyage” (linéarisation) de la série via une modélisation Reg-ARIMA :
  - Détection du type de modèle additif/multiplicatif ;
  - Détection et correction des points atypiques et des ruptures ;
  - Détection et correction des effets de calendrier ;
  - Préviation de la série.
- 2 Décomposition de la série par filtres linéaires (moyennes mobiles ou filtres de Wiener-Kolmogorov).

## Le modèle des résidus

ARIMA ou ARFIMA ou ??? ACF et IACF de  $(1 - B^7)X_t$ .

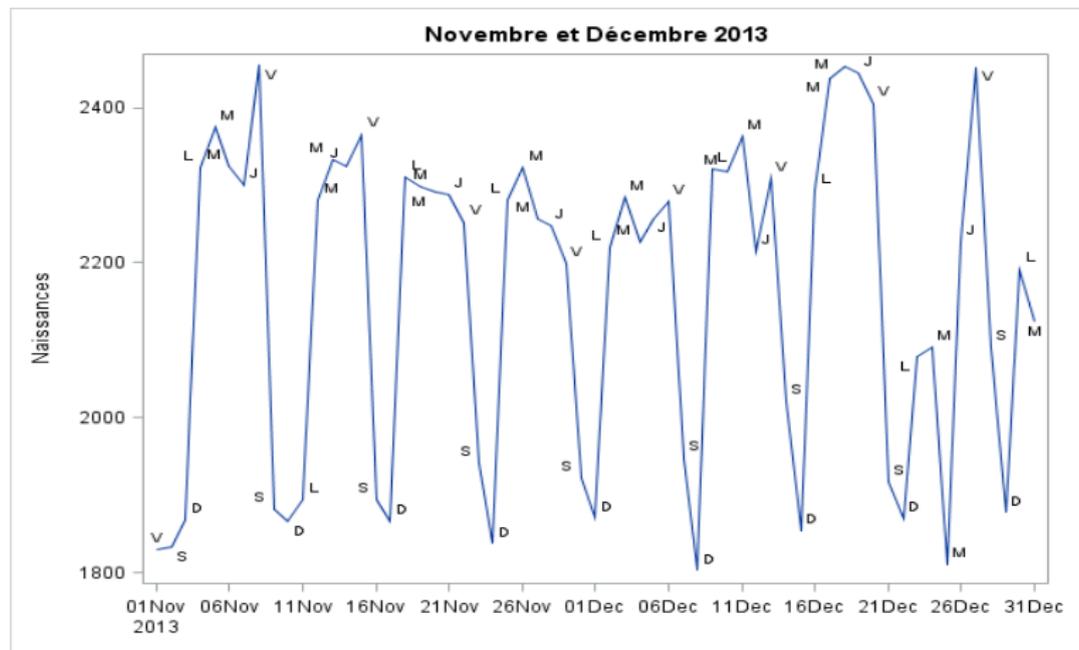


## Les effets de calendrier

- En principe, des données journalières ne présentent pas d'effets de calendrier “purs”.
- Mais les jours fériés peuvent avoir un impact non négligeable.
- Exemples du 11 novembre et de Noël sur notre série des naissances.

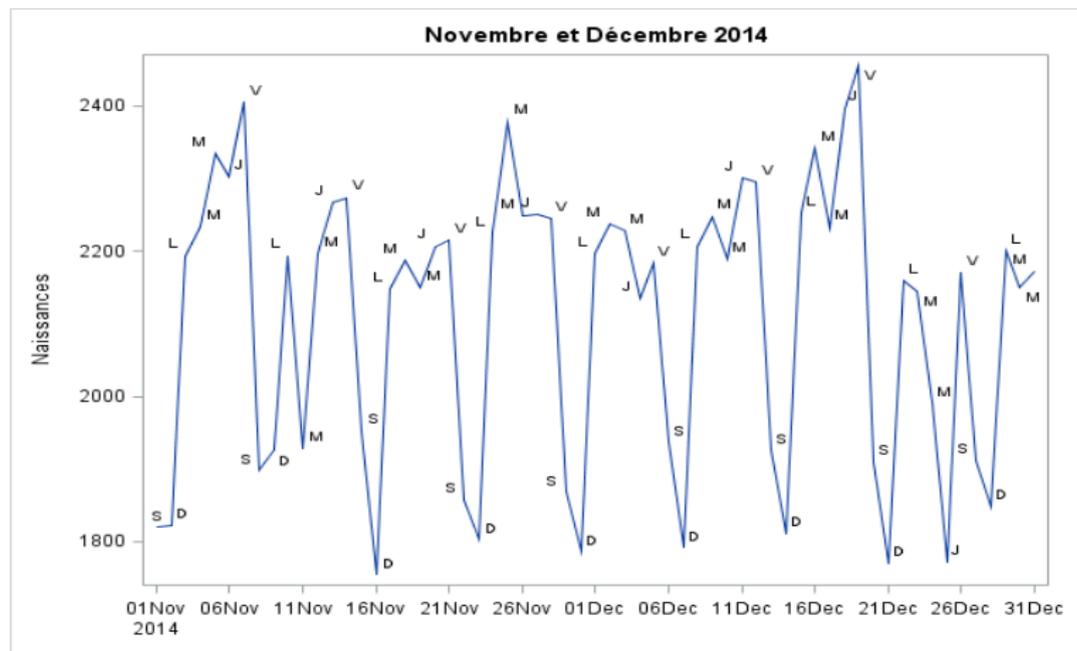
## Zoom sur Novembre-Décembre 2013

Périodicité hebdomadaire très claire. Et fort effet des jours fériés.



## Zoom sur Novembre-Décembre 2014

Confirmé en 2014.



## Le modèle pour pré-ajustements

- Potentiellement un grand nombre de régresseurs dans le modèle :
  - 11 par an pour les jours fériés, régresseurs pour modèles à impact, points atypiques ....
  - C'est le point délicat. Ne rien oublier (cf infra).
- La modélisation ARIMA peut s'avérer non pertinente.
- Aller vers des modèles plus simples pour détecter et corriger les non-linéarités.
  - modéliser la saisonnalité (indicatrices, sinus-cosinus) et supposer des résidus "simples";
  - utiliser des modèles de régression avec résidus de Holt-Winters ou de lissage exponentiels avec saisonnalité.

# Algorithme de base de X-11 (1)

Série brute mensuelle :  $X_t = TC_t + S_t + I_t$

1. **Estimation de la tendance-cycle par une moyenne mobile  $2 \times 12$  :**

$$C_t^{(1)} = M_{2 \times 12}(X_t)$$

2. **Estimation de la composante saisonnier-irrégulier :**

$$(S_t + I_t)^{(1)} = X_t - TC_t^{(1)}$$

3. **Estimation de la composante saisonnière par une moyenne  $3 \times 3$  sur chaque mois :**

$$S_t^{(1)} = M_{3 \times 3} \left[ (S_t + I_t)^{(1)} \right]$$

et normalisation

$$\tilde{S}_t^{(1)} = S_t^{(1)} - M_{2 \times 12} \left( S_t^{(1)} \right)$$

4. **Estimation de la série corrigée des variations saisonnières :**

$$A_t^{(1)} = (C_t + I_t)^{(1)} = X_t - \tilde{S}_t^{(1)}$$

## Algorithme de base de X-11 (2)

5. **Estimation de la tendance-cycle par une MM de Henderson sur 13 termes :**

$$TC_t^{(2)} = H_{13} \left( A_t^{(1)} \right)$$

6. **Estimation de la composante saisonnier-irrégulier :**

$$(S_t + I_t)^{(2)} = X_t - TC_t^{(2)}$$

7. **Estimation de la composante saisonnière par une MM  $3 \times 5$  sur chaque mois :**

$$S_t^{(2)} = M_{3 \times 5} \left[ (S_t + I_t)^{(2)} \right]$$

et normalisation

$$\tilde{S}_t^{(2)} = S_t^{(2)} - M_{2 \times 12} \left( S_t^{(2)} \right)$$

8. **Estimation de la série corrigée des variations saisonnières :**

$$A_t^{(2)} = (C_t + I_t)^{(2)} = X_t - \tilde{S}_t^{(2)}$$

## Un algorithme multi-périodicités

Il est assez facile de généraliser l'algorithme.

1. **Estimation de la tendance-cycle par une MM  $7 \times 365$  :**

$$C_t^{(1)} = M_{7 \times 365}(X_t)$$

2. **Estimation de la composante saisonnier-irrégulier globale :**

$$(S_t + I_t)^{(1)} = X_t - TC_t^{(1)}$$

- 3a. **Estimation de la composante saisonnier-irrégulier "7" (MM 365)**

$$SI_t^{(1),7} = M_{365} \left[ (S_t + I_t)^{(1)} \right]$$

- 3b. **Estimation de la composante saisonnière "7" par une MM  $3 \times 3$**

$$\text{sur chaque période (jour) : } S_t^{(1),7} = M_{3 \times 3} \left[ SI_t^{(1),7} \right]$$

$$\text{et normalisation : } \tilde{S}_t^{(1),7} = S_t^{(1),7} - M_7 \left( S_t^{(1),7} \right)$$

- 3c. **Idem pour la composante saisonnière "365"**

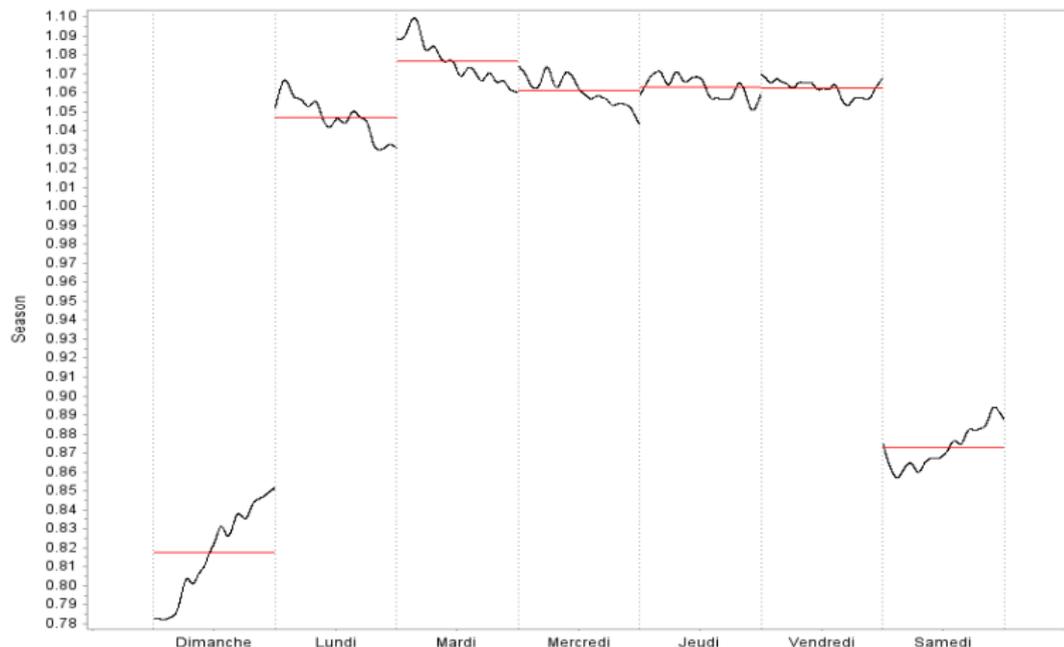
4. **Estimation de la série corrigée des variations saisonnières :**

$$A_t^{(1)} = (C_t + I_t)^{(1)} = X_t - \tilde{S}_t^{(1),7} - \tilde{S}_t^{(1),365}$$

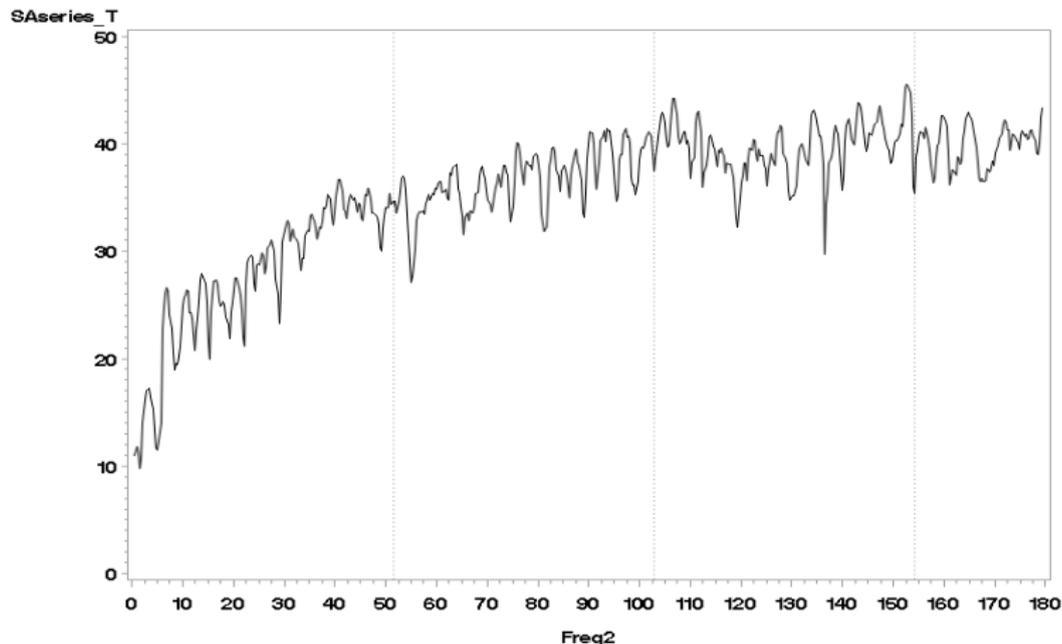
## Le modèle Reg-ARIMA

- On utilise ici une version ancienne de Tramo pouvant traiter une seule saisonnalité mais jusqu'à 10000 observations ;
- Le modèle des résidus imposé ici est un  $(0, 1, 1)(0, 1, 1)_7$  ;
- Jean Palate a présenté bien mieux hier : prise en compte des deux saisonnalités hebdomadaire et annuelle avec un modèle "transformé" permettant in fine de mimer un ARFIMA saisonnier.
- Cela permettra de régler les problèmes présentés ci-après.

# La composante saisonnière hebdomadaire

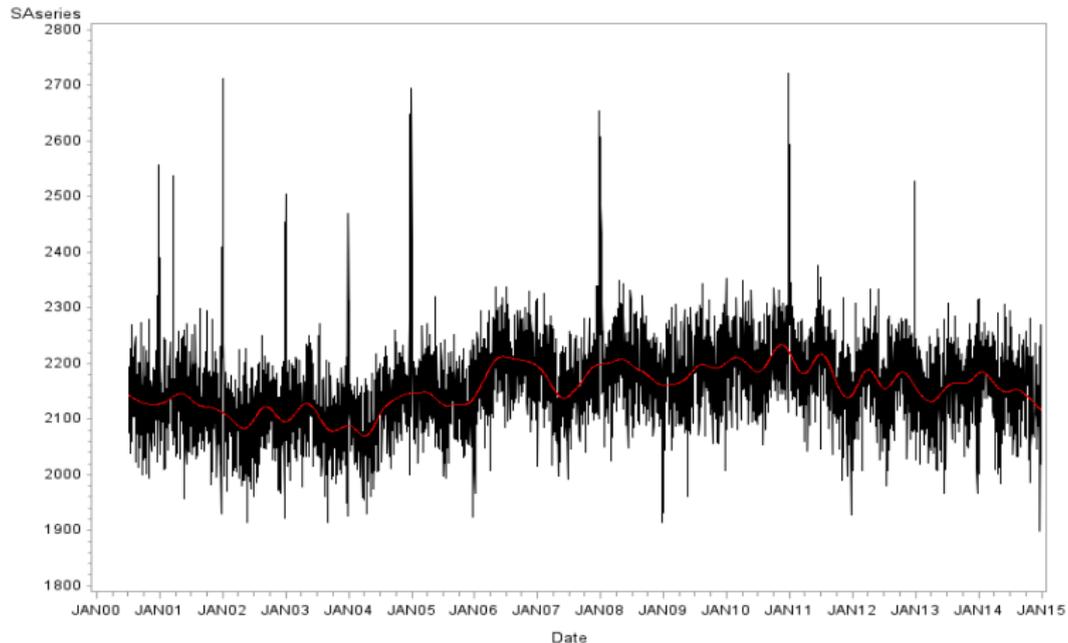


## Le spectre de la série désaisonnalisée



# La série désaisonnalisée

Quelques soucis ....



## Les fins d'année

Effets de calendrier mal pris en compte en fin d'année ....

<u>Date</u>	<u>Séries</u>
Friday, December 22, 2000	2556.9
Tuesday, March 20, 2001	2538.6
Monday, December 24, 2001	2409.1
Monday, December 31, 2001	2711.9
Wednesday, January 2, 2002	2407.4
Tuesday, December 24, 2002	2454.1
Tuesday, December 31, 2002	2504.3
Wednesday, December 24, 2003	2436.6
Wednesday, December 31, 2003	2470.0
Thursday, December 23, 2004	2648.2
Friday, December 31, 2004	2694.8
Saturday, January 1, 2005	2418.3
Monday, December 24, 2007	2655.1
Monday, December 31, 2007	2607.3
Tuesday, January 1, 2008	2434.9
Thursday, January 3, 2008	2482.6
Friday, January 4, 2008	2433.6
Thursday, December 23, 2010	2721.6
Saturday, December 25, 2010	2477.0
Tuesday, December 28, 2010	2402.2
Friday, December 31, 2010	2593.2
Saturday, January 1, 2011	2495.4
Sunday, January 2, 2011	2424.1
Monday, December 24, 2012	2528.2

## Remarques

- Noël est ici considéré comme un point atypique, éliminé puis réintroduit dans la série CVS ;
- Les effets de Pâques n'ont pas été introduits dans le modèle de régression ;
- Le modèle à utiliser pour les prévisions n'est pas évident (ARIMA ?) ;
- L'extension de Tramo présentée par Jean Palate devrait régler ces problèmes.
- .... et en tout cas fournit un premier prototype complet ;

## Conclusions

- Il est possible d'adapter l'algorithme de X11 aux séries journalières ;
- La philosophie itérative de X11 est d'ailleurs très proche de celle de STL ;
- Les effets des fêtes mobiles (Pâques) doivent être modélisés ;
- Le modèle à utiliser pour les prévisions n'est pas évident (ARIMA ?) ;
- Encore beaucoup de choses à régler :
  - L'ordre des moyennes mobiles ;
  - le choix automatique des moyennes (ratios RIC et RIS).