



Dossier Techniques de prévision et de
Conjoncture

**Analyse saisonnière et non
saisonnière de série
temporelle**

Ariinui TERIITEHAU

Darlène NIBOGORA

Naoufali MADI

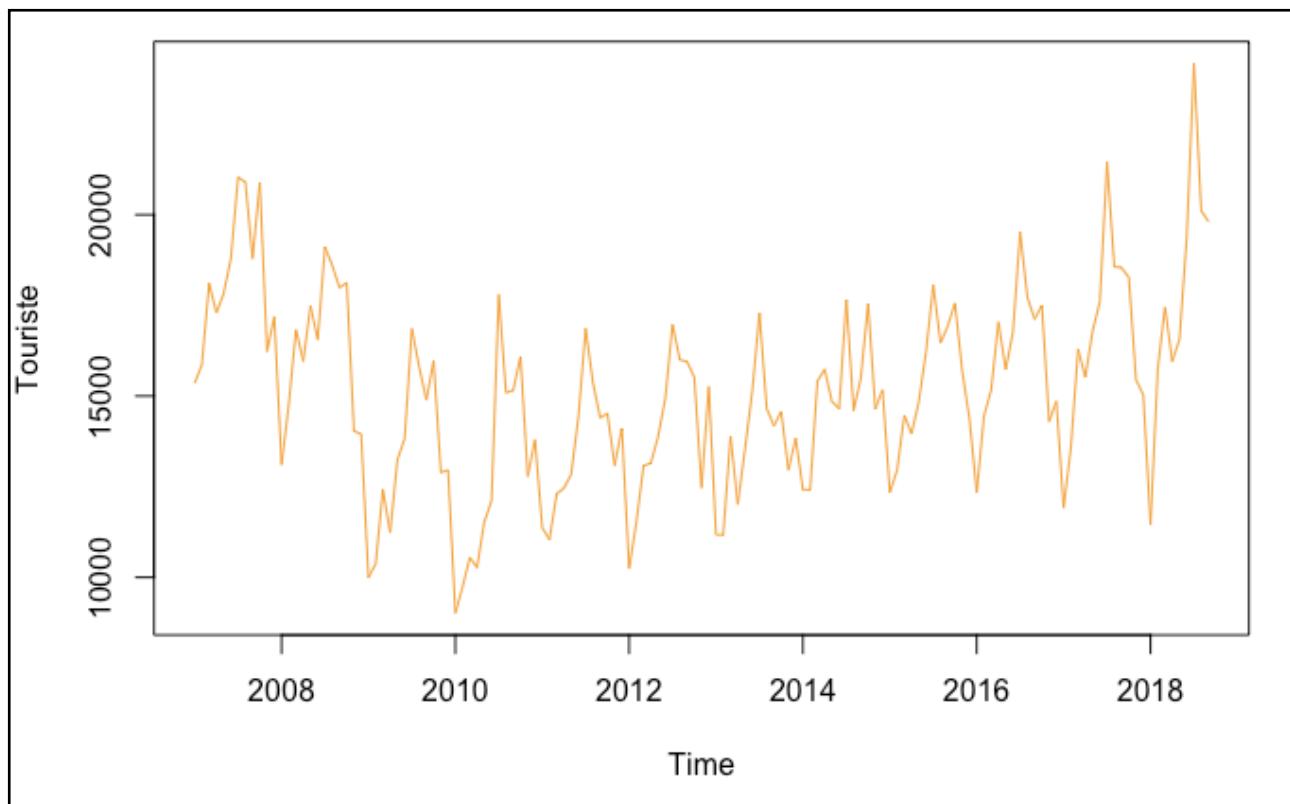
Master 1 EKAP

I.	Série saisonnière et analyse cyclique	3
A)	Analyse préliminaire	3
B)	Désaisonnalisation, CVS et tendance de la série	11
C)	Analyse cyclique	14
II.	Série saisonnière et prévision	22
A)	Comparaison du modèle SARIMA et X13	22
B)	Prévision	23
C)	Calcul des mesures d'erreur de prévision	25
D)	Test de Diebold-Mariano et multiple de Mariano et Preve (2012)	26
III.	Série non saisonnière et prévision	29
A)	Analyse préliminaire	29
B)	Estimation et prévisions des modèles	33
C)	Test des modèles de prévision avec la série brute	42
	Conclusion	45
	Bibliographique	46
	Annexe	48

I. Série saisonnière et analyse cyclique

A) Analyse préliminaire

Figure 1 : Série temporelle du nombre de touristes en Polynésie française



Source : données de l'ISPF - logiciel R

1) Description de la série

Les données proviennent de l'institut de la statistique de Polynésie Française (**ISPF**) entre janvier 2007 et septembre 2018. Nous avons 141 observations représentant les mensualités du nombre de touristes arrivé sur le territoire. **La figure 1** montre la série temporelle saisonnière et les variations du nombre de tourisme en Polynésie française.

Ils varient par des pics et des creux régulier chaque année. Les pics apparaissent durant les saisons chaudes (périodes estivales), ce qui expliquent une fréquentation plus élevées durant cette période et les périodes creuses se forment durant les saisons froides (automne-hiver), donc nous pouvons nous attendre à avoir une plus forte fréquentation des touristes provenant de l'hémisphère nord comme les américains du nord ou les européens.

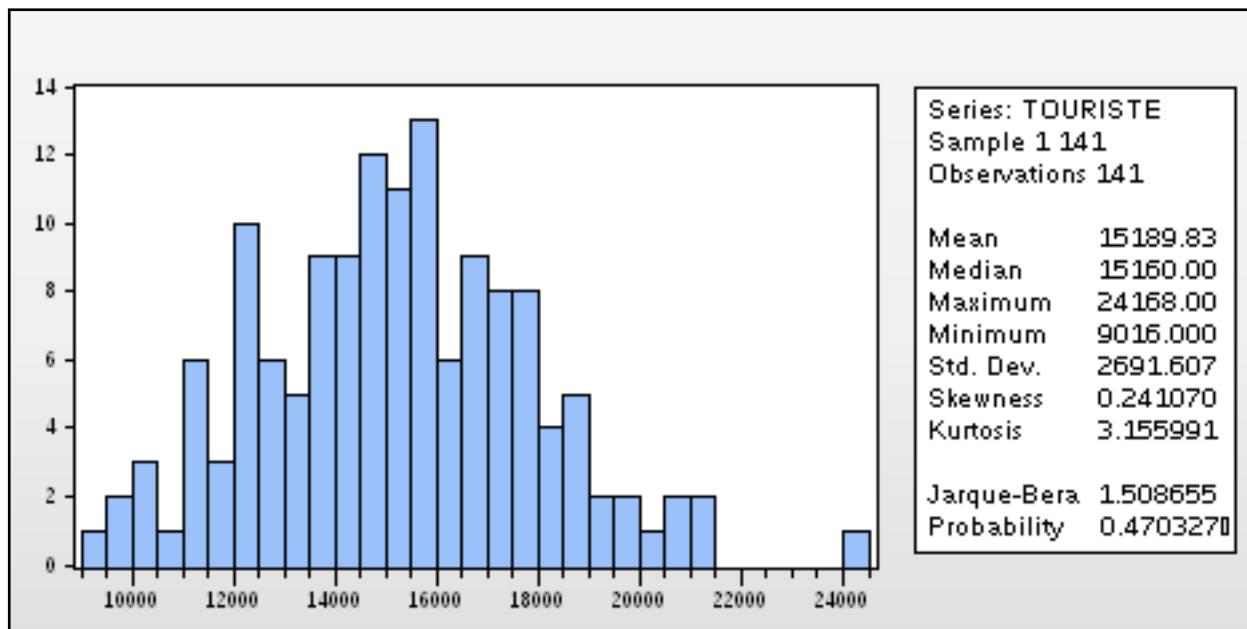
En effet, la fréquentation touristique suit les mêmes saisonnalités de cette même hémisphère malgré la stagnation de la température chaude durant toute l'année en Polynésie.

La série montre une tendance de 2007 à 2010, nous pouvons voir une décroissance sur le nombre de tourisme puis une croissance faible entre 2010 et 2018.

Dans la même période (2010-2018), l'écart entre le creux et le pic est de plus en plus importante lorsque nous nous rapprochons de la fin de la série. Cette hausse soudaine durant l'année 2018 s'explique par l'apparition de nouvelles compagnies aériennes low-cost comme « French Bee » qui encouragent la hausse du nombre de touristes à prix compétitif sur le marché.

2) Analyse statistique

Figure 2 : Distribution de la série et description statistique



Source : données de l'ISPF - logiciel Eviews

Pour avoir une meilleure compréhension de la série, nous nous intéressons à la description statistique dans la **figure 2**. Nous avons une moyenne de touristes de **15 189.83** et une médiane assez proche de **15 160**. La valeur maximum est égale à 24 168 touristes dont la valeur est atteinte durant l'été en juillet 2018 puis la valeur minimale est de 9 016 en janvier 2010. Nous trouvons un écart type de 2 691.60, étant très faible par rapport à la moyenne et médiane, nous pouvons dire que les valeurs sont peu dispersées autour de la moyenne donc nos observations sont homogènes.

De plus, le graphique à barre dans la figure 2, regroupe par intervalle toutes les observations mensuelles et ces fréquences. Ce graphique permet d'illustrer clairement les trois tests statistiques : kurtosis indiquant le coefficient d'aplatissement, skewness sur la asymétrie de distribution des valeur puis la normalité des valeurs selon le test Jarque-Bera. Nous avons des signes positifs pour le kurtosis et skewness ce qui signifient respectivement que la loi est plus pointue qu'une loi normale et que la distribution s'étale vers la droite. Le test Jarque-Bera a pour hypothèse nulle (H_0) si l'échantillon suit une loi normale dont le test suit la loi de Chi2 à deux degrés de liberté.

Ayant une p-value de 0.47 étant supérieur à 5% alors nous acceptons l'hypothèse nulle donc notre série suit une loi normale. Il est possible de confirmer par le test Shapiro-Wilk étant plus performant pour définir la normalité d'un échantillon, la p-value est égale à 0.6994 et est supérieur à 5% alors l'hypothèse de normalité est bien confirmé.

Nous notons que dans la série, elle présente une unique valeur atypique selon graphique de boîte à moustache (**figure 3, annexe**) et le test de Grubbs (p-value = 0.047). Cette valeur date pour le mois de juillet 2018. Cette augmentation du nombre de touristes en Polynésie s'explique par l'implantation de nouvelle compagnie aérienne low-cost vu précédemment.

3) Détection de la saisonnalité et schéma de décomposition

Tableau 1 : Tableau de Buys-Ballot de la série brute

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
oct	20891	18121	15972	16092	14519	15519	14576	17546	17561	17499	18271	
nov	16229	14040	12892	12784	13086	12470	12953	14646	15681	14295	15449	
déc	17190	13940	12958	13798	14106	15262	13832	15169	14364	14858	15030	
janv	15354	13102	9999	9016	11371	10238	11174	12422	12343	12340	11910	11457
févr	15859	14765	10372	9730	11038	11523	11177	12410	12949	14444	13564	15747
mars	18107	16829	12415	10547	12304	13075	13897	15410	14472	15168	16281	17452
avr	17290	15962	11230	10271	12458	13147	12011	15737	13956	17042	15523	15955
mai	17818	17485	13236	11525	12838	13879	13534	14853	14832	15738	16782	16559
juin	18783	16551	13824	12119	14424	14940	15120	14650	16223	16758	17599	19372
juil	21034	19111	16853	17790	16858	16979	17289	17656	18060	19517	21448	24168
août	20897	18601	15808	15087	15372	16002	14655	14603	16463	17719	18563	20110
sept	18789	17989	14888	15160	14402	15944	14175	15500	16927	17117	18539	19809
Total	218241	196496	160447	153919	162776	168978	164393	180602	183831	192495	198959	160629
\bar{X}	18186,7	16375	13371	12827	13565	14082	13699	15050	15319	16041	16580	17848
σ	1970,21	2009,5	2207,7	2777	1692,9	2035,9	1733,2	1603,3	1798,5	1936,4	2544,1	3568,8

Source : données ISPF - logiciel Excel

Dans cette partie, nous cherchons à détecter la saisonnalité de la série brute et son schéma de composition si elle est de nature « multiplicatif » ou « additif ».

La saisonnalité dans notre série va correspondre à un phénomène qui se répète à un intervalles de temps réguliers (périodique avec 12 mois). Cette **saisonalité**, nous la notons : S_t .

Dans le tableau 1 et la figure 1, nous remarquons la persistance des pics et creux pour chaque année. Ainsi, les creux se trouvant durant les mois de janvier (indiquer en rouge dans le tableau 1) et pics durant les mois de juillet (en bleu).

Dans ce phénomène de nombre touristique, il est évident de trouver une nature saisonnière comme cité dans la partie description de la série. Nous avions dit que la série représenté deux saisons : froide et chaude.

Vérification de la saisonnalité par jdmetra

De plus, nous devons confirmer le raisonnement précédent par des tests concrets. Ces tests sont automatiquement effectués en utilisant les modélisations de TRAMO-SEAT ou X13-ARIMA sur le logiciel Jdmetra dans **R** développé et publié par Eurostat. Différents tests sont réalisé sur la série et les résidus du modèle de TRAMO-SEATS.

Figure 4 : Test de détection de la saisonnalité avec TRAMO-SEATS

Combined test in the entire series	
Non parametric tests for stable seasonality	
Kruskall-Wallis test	P.value 0.000
Test for the presence of seasonality assuming stability	0.000 1
Evolutive seasonality test	0.058
Identifiable seasonality present	
Residual seasonality tests	
qs test on sa	P.value 1.000
qs test on i	1.000 2
f-test on sa (seasonal dummies)	1.000
f-test on i (seasonal dummies)	1.000 3
Residual seasonality (entire series)	1.000
Residual seasonality (last 3 years)	0.812 4
f-test on sa (td)	0.001
f-test on i (td)	0.030 5

Source: library ('RJDemetra') , logiciel R

s

Tout les tests ont pour hypothèse nulle H_0 : l'absence de saisonnalité, ont supposant que la série est stationnaire ou rendu stationnaire. Le test Kruskall-Wallis, test de Friedman et le test montrant la présence de stabilité de la saisonnalité s'interprètent de la même manière, c'est-à-dire si leur p-value est inférieure à 0.05 alors H_0 est refusé. Dans la **figure 4**, les même tests montrent des p-values (**1 et 5**) qui sont inférieurs à 0.05 donc nous refusons H_0 pour les trois tests.

Ensuite, nous nous intéressons aux tests QS (**2**), F-test sur les variables dichotomiques saisonnières (**3**) et la saisonnalité des résidus (**4**). Ici, les p-values sont exprimés différemment des deux tests précédents, ainsi si la p-value est supérieur à 0.05 alors on peut rejeter H_0 . Dans la même **figure 4**, nous remarquons que les tests ont une p-value supérieur à 0.05 (**2, 3 et 4**) donc nous pouvons aussi rejeter H_0 .

Nature de la série : multiplicative ou additive

Pour le choix de la nature de la série, nous allons nous appuyer sur plusieurs méthode pour déterminer si elle est additive ou multiplicative :

- *méthode du profil* : nous allons superposer plusieurs courbes de profil (par année) sur un même graphique. Si les courbes sont parallèles entre elles, le modèle est alors additifs.

- *méthode de la bande* : cette méthode consiste à tracer deux droites entre les minimas et les maximas de la série pour chaque saison. Si les deux droites sont parallèles alors la série présente un modèle additive. Dans le cas inverse, la série est un modèle multiplicative

- *méthode analytique* : cette méthode est la plus robuste pour déterminer la nature du modèle. Nous devons réaliser une régression linéaire simple de moindres carrées entre les moyennes (\bar{x}) et les écart-types (σ) pour chacune des périodes (avec les 12 années) tel que :

$$\sigma = a\bar{x} + b,$$

Si a est nul, alors nous avons à modèle additif et multiplicatif dans le cas contraire. Nous utiliserons la moyenne et l'écart-type que nous avons calculé dans le tableau de Buys-Ballot.

Rappelons qu'un modèle est *additif* si la décomposition de la série est : $Y_t = Z_T + S_T + \epsilon_t$

Et *multiplicatif* si : $Y_t = Z_T(1 + S_T)(1 + \epsilon_t)$

Avec $Z_t = T_t \times C_t$: la tendance de long-terme et mouvement cyclique sur cette tendance

ϵ_t : les erreurs qui sont des variables aléatoires centrées.

S_t : la saisonnalité dans la série

Par la première méthode, dans la **figure 5**, nous remarquons que l'ensemble des courbes semblent être parallèles entre elles, ce qui nous pousse à dire que nous avons un modèle additif.

Ensuite, en regardant la **figure 6**, nous avons tracé grossièrement deux droites par rapport en minima et maxima de la série. Nous pouvons dire que les droites semblent aussi être parallèles alors nous avons peut être un modèle additif.

Pour la dernière méthode analytique qui paraît la plus clair, nous trouvons le coefficient $a = 0.0958$, la série admet pour un modèle multiplicatif selon cette méthode. Ainsi dans la suite de l'étude, nous analyserons sur un modèle multiplicatif.

4) Détection des points atypiques par TRAMO et RegArima

Times Series Regression with ARIMA noise, Missing values and Outliers (TRAMO) et Signal Extraction in ARIMA Time Series (SEATS) sont des modèles développés par V. Gómez et A. Maravall pour la banque espagnol pour détecter les points atypiques et de proposer des paramètres $(p,d,q)x(P,D,Q)$ de SARIMA adéquat pour la série étudié [1]. En effet, les points atypiques rendent sensibles les modélisations ARIMA et sont nécessaire de les identifier et de les prendre en compte dans la suite de l'étude.

Il est important de rappeler les contrainte du modèle TRAMO et RegArima qui impose des contraintes :

- partie non saisonniers : $p, q \leq 3, d \leq 3$

- partie saisonnier : $P, Q \leq 1, D \leq 2$

Figure 5: Nombre de touristes selon le mois pour chaque année

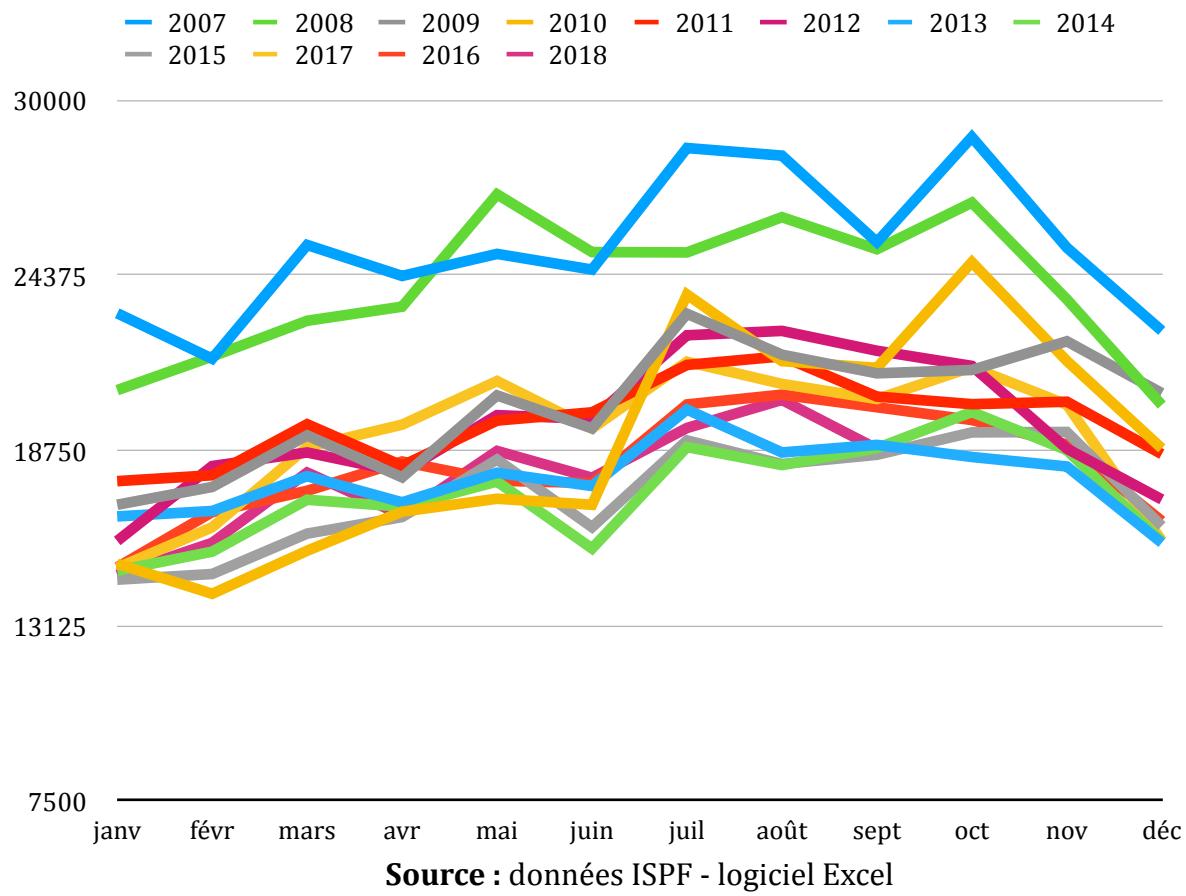
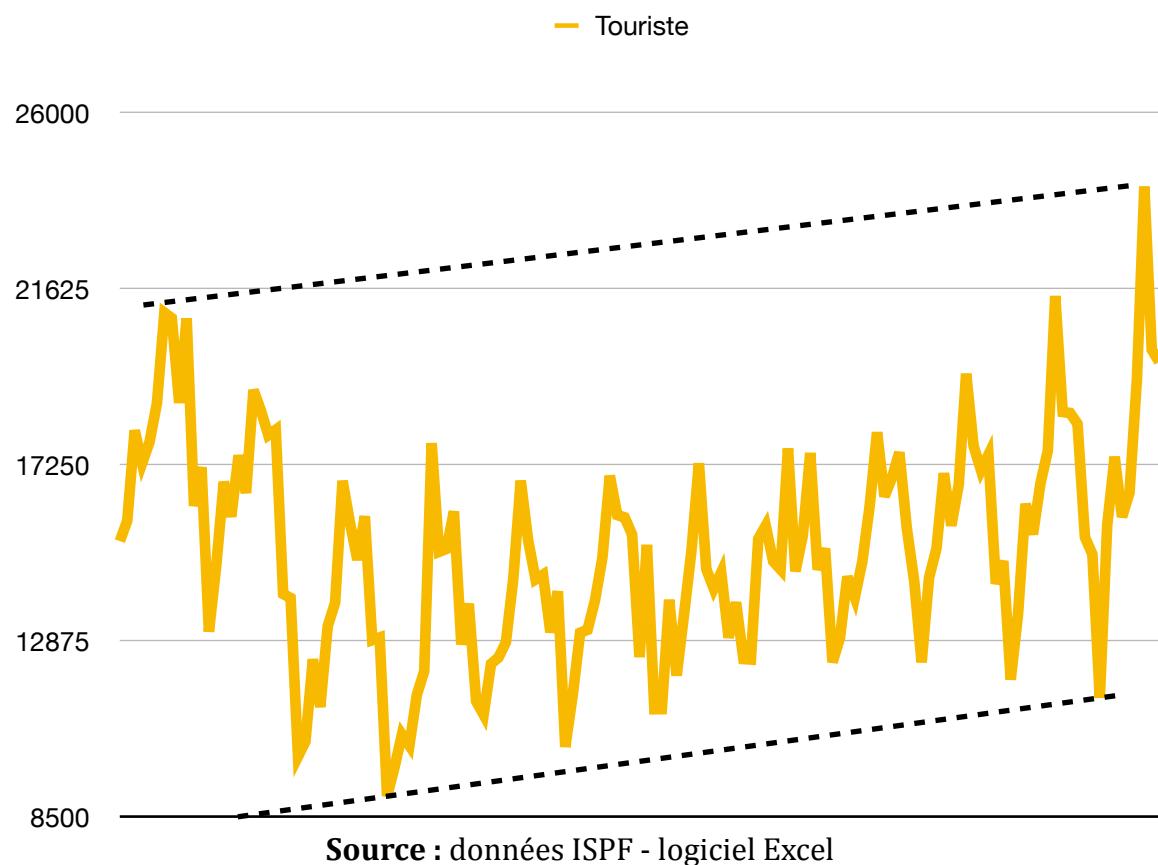


Figure 6: Nombre de touristes selon le mois pour chaque année



Pour détecter les points atypique dans notre étude avec TRAMO et RegArima, nous allons réutilisé la fonction jdemetra + dans le logiciel R. Nous obtenons les résultats pour TRAMO dans la **figure 7** et nous trouvons un point atypique datant le mois de juillet 2010. Nous constatons aussi que le type du point atypique est de LS (*level shift*), ce même point a un effet permanent sur le niveau de la série.

Figure 7 : Résultat points atypiques avec le modèle TRAMO

```
RegARIMA
y = regression model + arima (0, 1, 1, 0, 1, 1)
Log-transformation: yes
Coefficients:
Estimate Std. Error
Theta(1) -0.4194 0.081
BTheta(1) -0.6110 0.083

Estimate Std. Error
LS (7-2010) 0.2061 0.053

Residual standard error: 0.06354 on 124 degrees of freedom
Log likelihood = 168.2, aic = 2128 aicc = 2128, bic(corrected for length) = -5.398
```

Source : Library Jdemetra, logiciel R

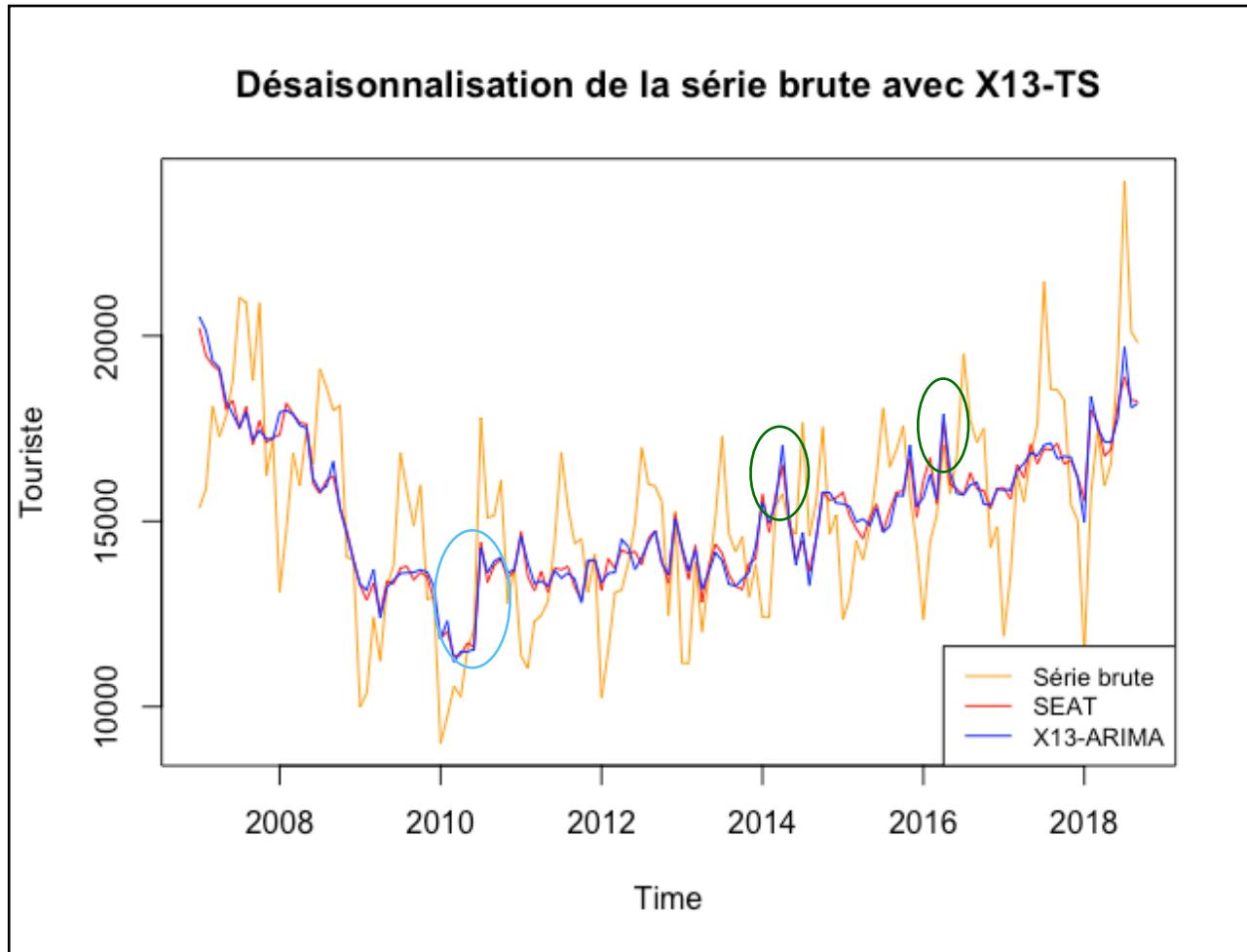
En effectuant avec RegARIMA (annexe, figure 8), nous obtenons le même point atypique que TRAMO et les coefficients et écart-types entre les deux modèles sont pratiquement le même.

Explication de l'unique valeur atypique :

- Après la crise financière de 2007 et 2008 provoquant une grande récession et touchant plusieurs secteur comme celui du tourisme, la fréquentation est à la hausse durant la période estivale après un fort recul lié à cette crise. Selon le bilan de **IOEM** (Institut d'Emission d'Outre-Mer), 2010 marque le retour de la clientèle américaine (Amérique du Nord) qui représente 30% de la totalité des tourisms et une augmentation de 3.4% par rapport à 2009, dont cette même clientèle est encouragé par un effet de taux de change euro/dollar avantageux. Ainsi, cette hausse soudaine est un phénomène récurrent après une crise économique.

B) Désaisonnalisation, CVS et tendance de la série

Figure 9 : Désaisonnalisation du nombre de touristes avec TRAMO-SEATS et X13



Source : traité avec library Jdemetra +, logiciel R

1) Correction des variations saisonnière (CVS)

Après avoir identifier les points atypique par TRAMO et RegArima, nous cherchons à désaisonnaliser notre série qui a pour objectif d'éliminer les composantes saisonnières afin de se focaliser sur la tendance ou le cycle de la série. Ainsi, cela nous permet d'avoir une meilleure vue sur l'analyse économique. La désaisonnalisation (correction des variations saisonnière ,CVS ou seasonal-adjusted, SA) est définie tel que :

$$SA_t = Z_t + \epsilon_t$$

Avec $Z_t = T_t \times C_t$: la tendance de long-terme et mouvement cyclique sur cette tendance

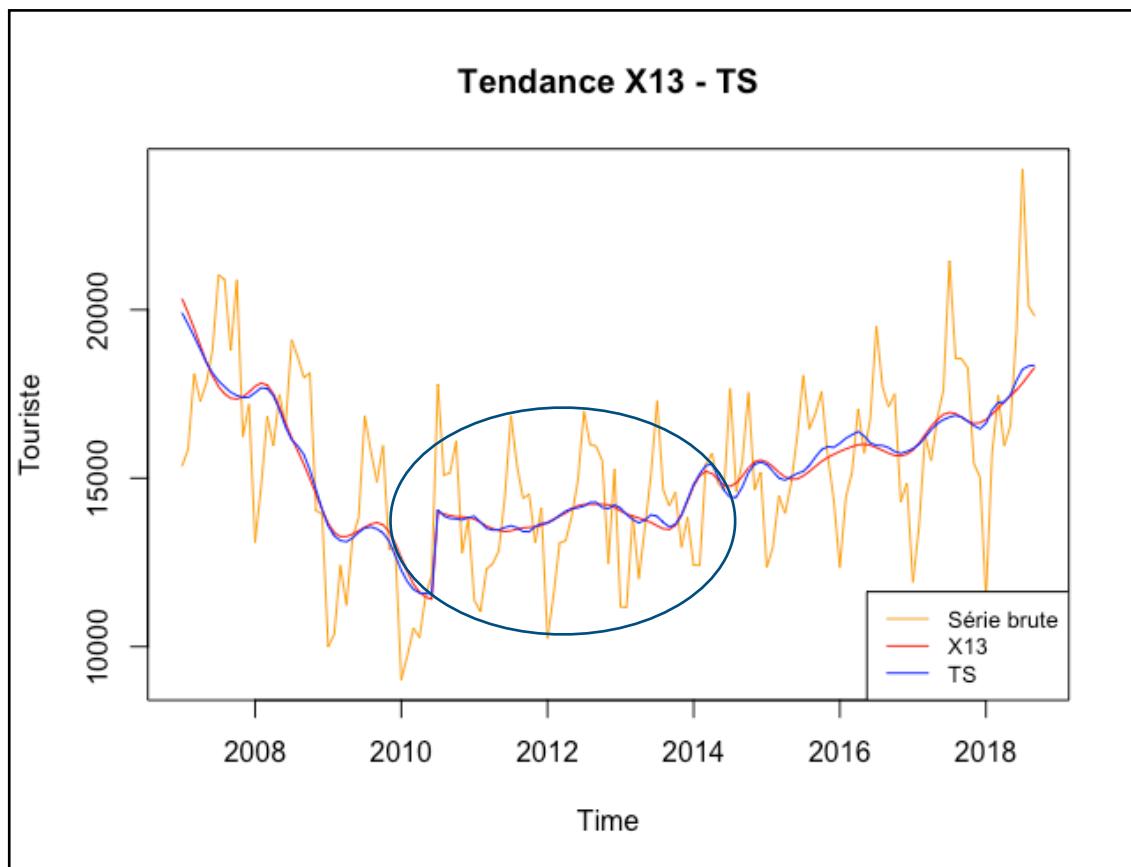
ϵ_t : les erreurs qui sont des variables aléatoires centrées.

Dans **la figure 9**, nous avons réalisé une désaisonnalisation avec X13 représentée en bleue et TRAMO-SEAT en rouge. Les deux nouvelles séries CVS montrent bien la suppression des parties saisonnières par la disparition des pics et des creux durant les mois de janvier et juillet. Les CVS montrent que les parties essentielles de la tendance de la série, nous voyons bien une décroissance entre 2007 et l'été 2010, puis une faible croissance entre l'été 2010 et 2018.

En observant les variations des deux séries CVS, nous remarquons qu'elles se confondent pratiquement sauf pour certains points. Par exemple, en regardant un extrait de la série entre 2008 et 2011 (**figure 10, annexe**), nous remarquons que la CVS X13 est plus lisse que celle de TRAMO-SEAT donc nous pouvons dire que X13 corrige mieux les parties saisonnières que TRAMO-SEAT. De plus, nous savons que X13 a la capacité de générer une CVS basé sur les modèles de SEATS, expliquant un meilleur lissage donc de meilleure stabilité d'ajustement. [2]

2) Tendance de la série

Figure 11 : série temporelle tendance de X13 et TRAMO SEAT



Source : Library Jdemetra, logiciel R

Dans cette partie, nous avons tracé la série tendancielle qui est Z_t de la série brute avec X13 et TRAMO-SEAT (**figure 11**). Nous remarquons que tous les pics et creux sont complètement corrigés et sont mieux lissés que les séries CVS. Dans le graphique (figure 11), nous avons illustré un cercle pour montrer que TS présente des petites variations entre 2011 et 2014, ce qui montre que le lissage de TS est moins bon que celui de X13. La valeur atypique au juillet 2010 est toujours lisible, nous verrons plus tard par quel procédé pour corriger le point.

3) Comparaison des deux modèles X13 et TRAMO-SEAT (TS)

Ensuite, nous intéressons aux résultats des modèles de X13 et TS dans la figure 12 (annexe) dont la série a subi une transformation logarithmique. Nous constatons que les paramètres $(p,d,q)x(P,D,Q)$ sont les mêmes pour les deux modèles :

$$SARIMA : (0,1,1)(0,1,1)_{s=12}$$

Ces paramètres favorisent les critères d'acceptation par rapport à la bonne qualité d'ajustement et de performance de prévision selon les critères AIC ou BIC que nous verrons plus tard dans cette étude.

En revanche, ils se différencient à la valeur des coefficients de la partie ***q et Q***, ce peut s'expliquer par une méthode de calcul différente dont X13 est lié à TS.

Nous nous intéressons au critère AIC (critère d'information d'Akaike). AIC est une mesure sur la qualité d'un modèle statistique, ce critère permet de nous proposer le meilleur modèle avec une mesure la plus faible. Dans notre cas, X13 est le modèle le plus fiable avec ce critère avec une valeur de 2124 contre 2128 pour TRAMO-SEAT.

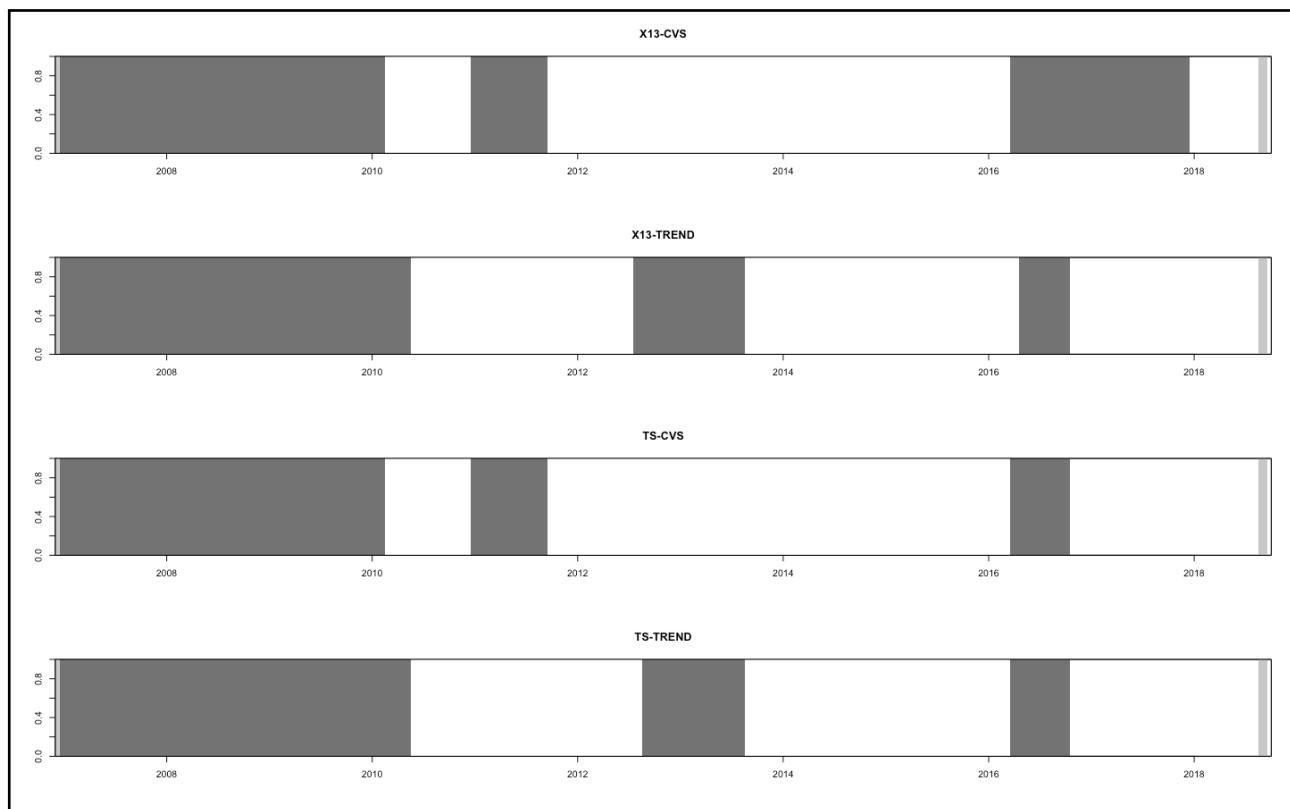
Selon les diagnostics, X13 trouve 7.36 cycles et 57.3 saisonnalités et TS trouve 7.57 cycles et 57.05. Nous pouvons dire que ces valeurs ne diffèrent pas vraiment sur le critère de cycle et de saisonnalité.

En outre, nous pouvons constater une droite verticale pour les CVS de X13 et TRAMO-SEAT pour le mois de juillet 2010 (cercle bleu) dans la **figure 9**. Cette même droite illustre bien le point atypique LS pour les deux modèles vu dans la partie précédente, en indiquant un changement de niveau dans la série (lisible dans la série tendancielle, **figure 11**).

En réalisant un autre modèle pour détecter les points atypiques de la série à partir de l'approche de Chen et Liu [3]. Dans la **figure 13 (annexe)**, ce modèle détecte deux nouveaux points aux dates suivantes : avril 2014 et avril 2016. Ces deux points atypiques sont des ‘additive outliers’, des points qui affectent une seule observation dans la série. Nous avons identifié ces points dans le graphique (figure 9, cercles verts), qui sont deux pics.

C) Analyse cyclique

Figure 15 : Datation du cycles des affaires des séries Trend et CVS de X13 et TS.



Source: library BCDating, logiciel R

1) Datation des points de retournement du cycle des affaires de la série CVS et Trend.

Dans cette partie, nous déterminerons les dates des cycles des affaires où les points de retournement se définissent lors d'une récession (croissance négative) ou d'une expansion (croissance positive).

Ainsi, la phase d'une expansion correspond donc entre le creux de la série et le pic qui marque sa fin d'expansion, à l'inverse pour la phase d'une récession. Notons aussi qu'un cycle

représente l'ensemble de deux phases consécutives d'un pic à un autre pic ou d'un creux à un autre creux.

La datation des points de retournement du cycle des affaires de la série suit l'algorithme de Bry-Boschan (BB), pour condition que la durée minimal d'une phase soit supérieure ou égale à 6 mois et la durée minimal d'un cycle soit supérieure ou égale à 15 mois [4]. De plus, l'algorithme supprime la composante irrégulière et les valeurs atypiques de la série CVS.

Nous avons identifiés graphiquement sans l'intervention de l'algorithme BB (*annexe, figure 14*) l'ensemble des pics et creux sur la série trend X13. Nous comptons au total 17 creux, pics et la présence de la valeur atypique.

Dans *la figure 15*, il y a quatre graphiques qui représentent les datations de points de retournement du cycle des affaires des séries CVS et TREND des modèles X13 et TRAMO-SEAT déterminer par l'algorithme BB. Les parties noirs sont les phases de récessions et les parties blanches sont les phases d'expansion.

Dans *la figure 16 (annexe)*, nous avons calculer trois caractéristiques pour identifier les phases d'un cycle de chaque série : la durée, l'amplitude et la sévérité. La durée d'une phase est le temps mesuré par le nombre de période qui séparent deux points entre un pic et un creux et l'amplitude représente la distance entre une pic et un creux. La sévérité qui mesure la perte du gain que subit l'économie la phase du cycle. La sévérité se calcul par le produit de la durée et l'amplitude en divisant par deux, qui peut être illustrer par un triangle rectangle (*annexe, figure 17*).

Nous avons 6 phases pour les quatre séries (*figure 15*), mais elles se différencient selon la duration des phases et la période. Par exemple, pour la série TREND de X13 et TS, la première phase est plus longue que les deux autres séries en se terminant pour le mois de juin 2010, alors que pour les séries CVS, elles se terminent pour le mois de mars 2010.

Les séries TREND de X13 et TS se ressemblent entre elles, tandis que les séries CVS se différencient sur la durée des deux dernières phases. La cinquième phase récession de X13-CVS dure 21 mois alors que pour TS-CVS ne dure que 7 mois.

En étudiant les pics et creux retenus par l'algorithme BB, il trouve les même pics pour X13-CVS et TS-CVS pour 01/2011 et 04/2016 et des creux pour 03/2010 et 10/2011. Mais le dernier creux se différencie dont X13-CVS date pour le 01/2018 et TS-CVS pour le 11/2016.

Les séries TREND trouvent aussi les mêmes creux mais pas les mêmes pics qui se différencient d'un mois. (*figure 16, annexe*) Par conséquent, la durée des phases de récession vont être différent entre les séries. Pour les quatre série, nous trouvons aussi trois cycles.

Puis, nous avons l'amplitude des phases qui sont pratiquement différentes entre les séries CVS et TREND. Par exemple, les séries CVS X13 et TS de la deuxième phase d'expansion ont respectivement une amplitude de 3398.4 et 3369.2 contre 2925.5 et 2727.4 pour les séries

TRENDS. Nous pouvons dire que pour les séries TRENDS, l'amplitude est plus faible car les séries sont plus lisses que X13. (*voir annexe, figure 17*)

Pour avoir une vue globale des deux caractéristiques et l'asymétrie des séries, nous avons calculer la moyenne de l'amplitude et durée des phases de récession et d'expansion. Dans le **tableau 2**, nous constatons que les durées moyennes d'expansion est fortement supérieur que celui des récessions, de même pour l'amplitude. Donc, nous pouvons dire qu'il y a aucune symétrie entre les phases de récession et les phases d'expansion des séries.

La moyenne des sévérités montre que le gain (phase expansion) est plus important que la perte (phase récession) pour chaque modèles et que les pertes sont amortis.

Tableau 2: Moyenne des amplitude, durée et sévérité

Expansion	CVS-X13	TREND-X13	CVS-TS	TREND-TS
Durée	32	29	32	29,0
Amplitude	4237,2	2678,5	4075,8	2775,2
Sévérité	67795,2	38838,25	65212,8	40240,4

Récession	CVS-X13	TREND-X13	CVS-TS	TREND-TS
Durée	15	9,5	8	9,5
Amplitude	2349,2	550,1	2081,9	684,2
Sévérité	17619	2612,975	8327,6	3249,95

2) Datation des points de retournement du cycle d'accélération.

Nous intéressons sur les points de retournement du **cycle d'accélération [5]**. Ce type de cycle distingue sur le taux de croissance de l'activité économique et définit par deux phases:

- phase d'accélération (housse du taux de croissance)
- phase de décélération (baisse du taux de croissance)

Pour effectuer une étude de datation des points de retournement du cycle d'accélération, nous devons réaliser une première différenciation de la série CVS sur X13 ou TS (*annexe, figure*

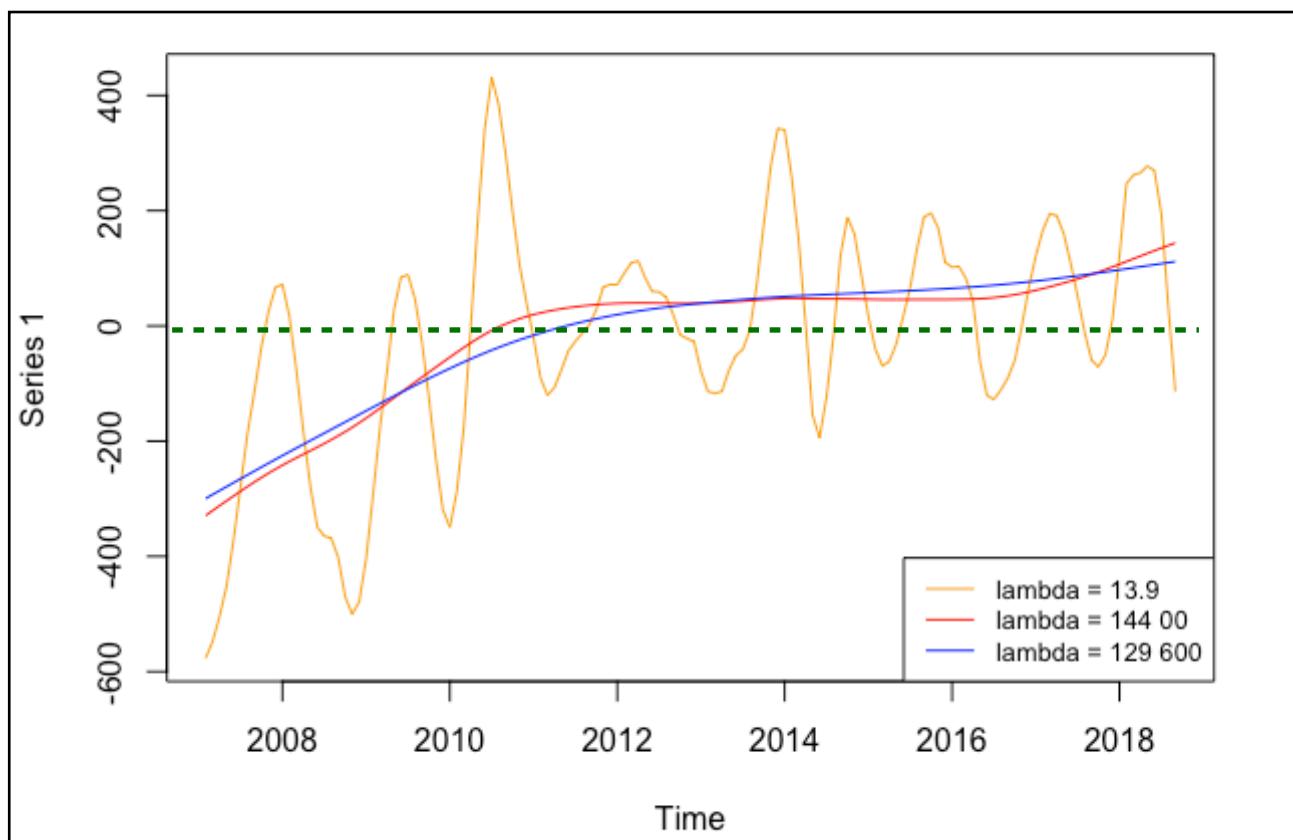
18). Puis, nous utilisons une méthode de filtre de Hodrick-Prescott (HP) pour décomposer la série en deux parties : tendance de long terme et cycle. Les contraintes de l'algorithme est modifié pour minimum de 9 mois de durée minimale d'une phase et 18 mois minimales pour un cycle.

De plus, nous devons déterminer le coefficient λ permettant de paramétrer les variation du taux de croissance. Plus λ est élevé et plus les variations seront supprimées dans le modèle donc nous aurons une série plus lisse.

Dans un premier temps, nous utiliserons un paramètre faible $\lambda = 13.9$ qui aura un comportement similaire à celui d'un **filtre low-pass** pour que le filtre ne conserve qu'une certaines fréquences inférieur ou égale à un seuil et supprime ce qui sont supérieur.

Dans un second temps, λ suivra les critères de HP avec une valeur de 14 400 et celui de Ravn et Uhlig pour $\lambda = 129\,600$.

Figure 19: filtre HP tendance de la série CVS X13 diff 1



Source: library mFilter , logiciel R

Dans la **figure 19**, nous remarquons que nombreux pics et creux dans la première différenciation CVS ont été corriger. Lorsque $\lambda = 13.9$, nous constatons quelque pics et creux

réguliers, alors qu'avec les deux autres tendances, nous avons deux séries lissées qui ne présentent pas de pics et creux.

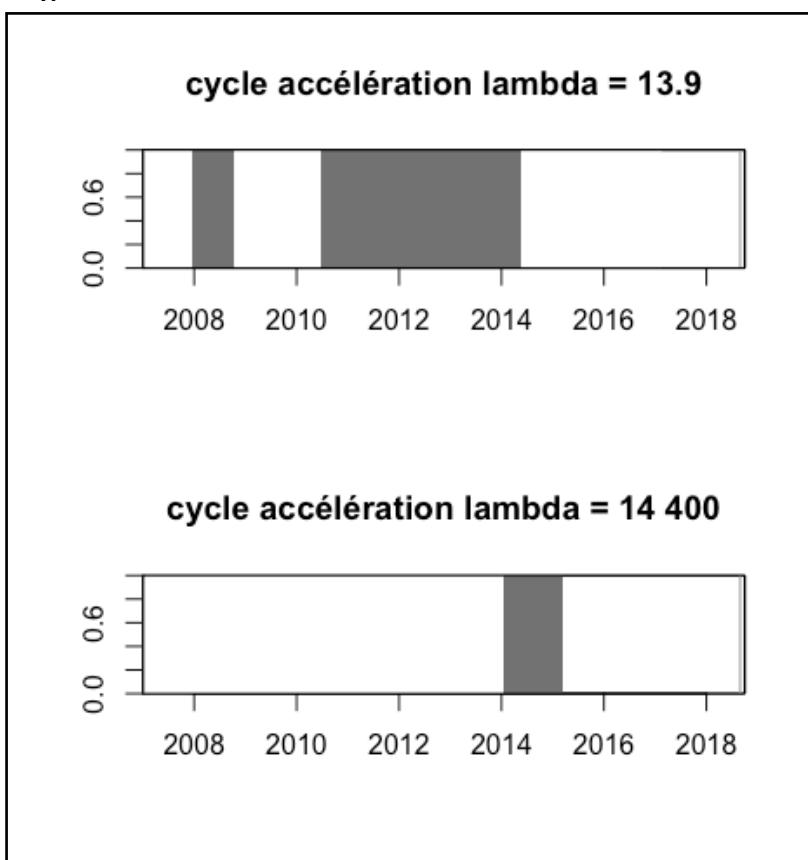
Lorsque $\lambda = 14\ 400$, nous avons un début de croissance entre 2007 et début 2011, puis une légère baisse entre 2011 et 2016 et enfin une faible croissance.

Pour $\lambda = 129\ 600$, la tendance augmente tout le long entre 2007 et 2018.

Ensuite, pour dater les points de retournements de la série, nous procéderons comme dans la partie précédente en utilisant l'algorithme de BB. Dans la **figure 20.A et 20.B (annexe**, pour la duration, l'amplitude et la sévérité), montrent les résultats de datation des points. Pour $\lambda = 129\ 600$, l'algorithme BB n'a pas pu calculer les dates de la série tendance car l'algorithme ne peut pas identifier les différentes phases selon les contraintes (durée minimal des phases et cycles) à cause du lissage de la série.

Ainsi, nous étudions que lorsque λ est égale à 13.9 et 14 400.

Figure 20.A: Différenciation de la série CVS X13



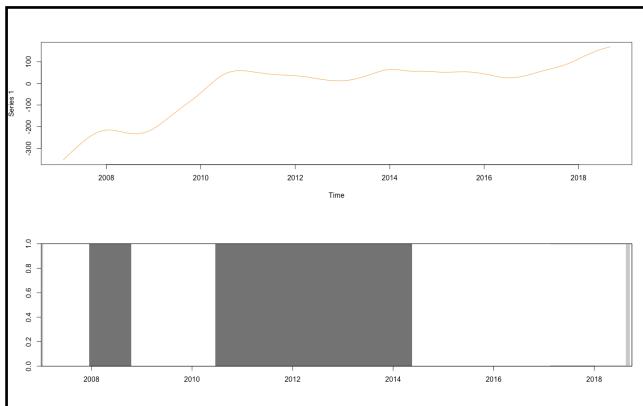
Source : library BBQ, logiciel R

Pour le premier graphique (**figure 20.A**), l'algorithme BB identifie 5 phases et nous avons une amplitude moyenne pour la phase récession de 564.6 et une durée moyenne de 26.5 mois. Pour la phase expansion, la moyenne d'amplitude est de 619.9 et une durée de 28.5 mois (**figure 20.B, annexe**). Il reconnaît une totalité de 3 pics et de 2 creux, donc il y a 3 cycles

Le deuxième graphique, nous avons des valeurs plus faibles pour l'amplitude avec une valeur nulle pour la phase expansion et 2 pour la phase récession. Ce sont des valeurs attendus car la série est plus lisse et donc la suppression des pics et creux. La duration n'est que 14 mois pour la phase récession. La sévérité moyenne pour la phase de récession est de 14.

De plus, lorsque $\lambda = 1\ 600$, nous trouvons les mêmes caractéristiques que le premier graphique quand $\lambda = 13.9$ en simulant l'algorithme BB. Le lambda est basé sur le ***cycl*e amérⁱcain trimestriel** (Hodrick et Preckott), ce cycle présente des variations de l'ordre de 5% par trimestre et la tendance de long terme ne devrait pas fluctuer de plus de 1/8eme de point sur la même période. Nous la représentons dans le graphique suivant :

Figure 20.C: Différenciation de la série CVS X13



Source : library BBQ, logiciel R

La série trend selon HP est plus lissé et corrige les pics et creux qui ne sont pas nécessaire.

Commentaire sur le cycl^ee des affaires :

Concrètement, le cycl^ee des affaires extrait le message principal pour faciliter l'interprétation des conjonctures sur l'évolution de fréquentation du tourisme en Polynésie.

Le ***graphique X13-TREND*** superpose la tendance de la série par X13 et la datation des cycles des affaires par le filtre HP pour avoir une idée plus clair de cette évolution.

L'indicateur de point de retournement détecte une conjoncture favorable en juin 2010, en septembre 2013 et en novembre 2016 et défavorable en août 2012 et en mai 2016. Ainsi, il y a trois phases d'expansion qui sont enregistrées d'amplitude égale en moyenne à 2678.5 d'une durée moyenne de 29 mois et trois phases de récession d'amplitude moyenne inférieur de 550.1 de 9.5 mois. Nous avons alors la phase d'expansion qui domine.

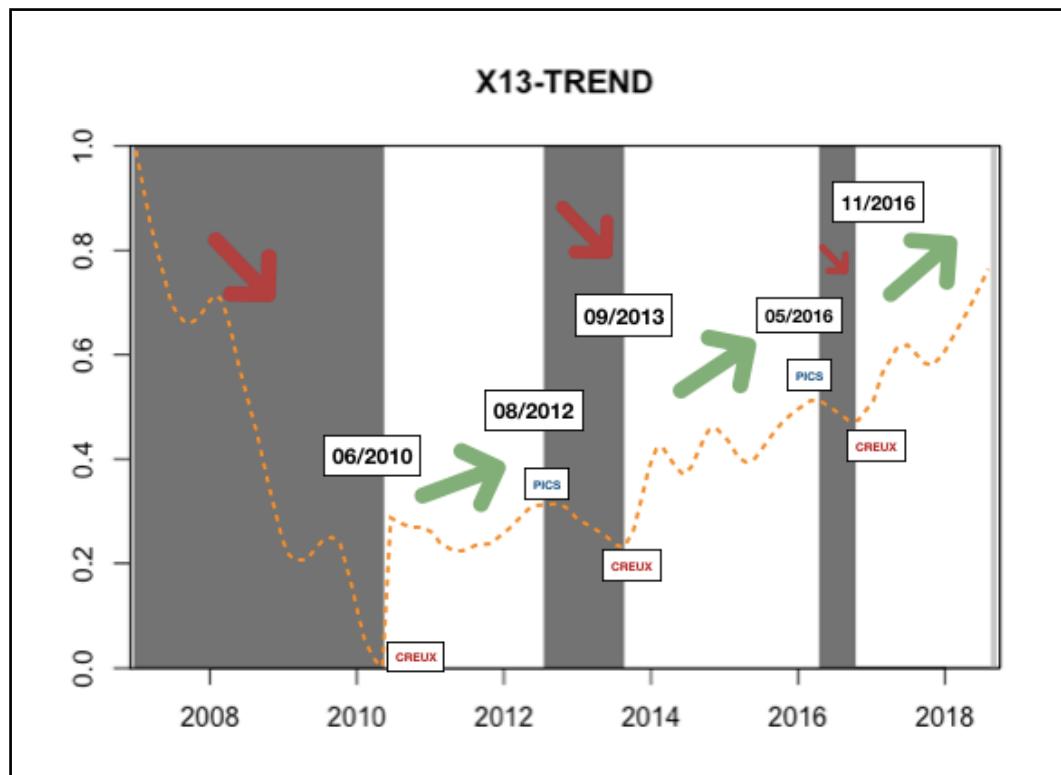
La grande première phase de récession entre 2007 et juin 2010 est la conséquence de la crise financière de 2007 qui trouve son origine dans l'effondrement du marché immobilier américain et la crise des crédits subprimes. Cette crise provoque une récession mondiale qui a conduit à un ralentissement dans les échanges internationaux, donc une diminution économique dans le secteur du tourisme.

La deuxième phase pour l'expansion est un phénomène récurrent après une grande récession économique. L'expansion du nombre de touristes s'explique essentiellement sur la hausse de fréquentation de touristes venant d'Amérique du Nord (Canada et Etats-Unis). Après la crise financière, le pouvoir d'achat des américains s'améliore et encourage à consommer des biens et des services.

La troisième phase montre une faible récession du nombre de tourisme malgré une hausse du tourisme mondial de 5% en 2013. Cette même phase montre une sévérité 4882.15, dont une perte de gain de 4882.15 touristes dans cette phase. Selon le bilan ISPF, il explique que cette baisse soudaine est à une baisse de la fréquentation des touristes européens, il perd 7.4% par rapport à 2012.

La longue phase d'expansion de 32 mois peut s'expliquer encore par l'accroissement des touristes américains et la valorisation du taux de change (voir série non saisonnière) francs pacifique/dollar. Il s'explique aussi par la hausse de tourisme sur de nouveau marché en Amérique du Sud soit plus de 16.5% par rapport à 2015. La courte phase de récession qui suit, il est démontré par la baisse de fréquentation des croisiéristes en Polynésie Française.

La dernière phase d'expansion montre une forte hausse du nombre de touristes. Cette hausse de fréquentation s'explique par l'émergence de prix attractifs des compagnies aériennes mais aussi la hausse de nouveau type de location comme « AirBnb » proposant des prix moins chers que les traditionnels hôtels. L'apparition de nouvelles compagnies low-cost comme « French Bee » encourage l'apparition de nouveau type de tourisme « sac-à-dos », des personnes voulant dépenser peu durant leur séjour.



II. Série saisonnière et prévision

Dans la partie prévision, nous parlerons des prédictions sur les 12 derniers mois de notre série brute avec plusieurs modèles. Nous procérons une estimation sur une nouvelle série entre janvier 2007 et septembre 2017 (**série training**) et des prévisions entre octobre 2017 et septembre 2018 afin de comparer avec la même série brute en calculant l'erreur estimée (**série test**).

A) Comparaison du modèle SARIMA et X13

Figure 21: Modèle X13, SARIMA, AUTO-SARIMA

Modèle X13 <pre> RegARIMA y = regression model + arima (0, 1, 1, 0, 1, 1) Log-transformation: no Coefficients: Estimate Std. Error Theta(1) -0.3691 0.091 BTheta(1) -0.6539 0.076 Estimate Std. Error Mean 30.93 21.67 Monday 260.62 131.09 Tuesday -355.19 133.15 Wednesday 349.31 134.25 Thursday -334.48 134.45 Friday 99.59 133.21 Saturday 303.33 134.77 Leap year 426.53 406.87 LS (7-2010) 2786.02 722.97 Residual standard error: 793.7 on 104 degrees of freedom Log likelihood = -942.5, aic = 1909 aicc = 1912, bic(corrected for length) = 13.8 </pre>	<pre> > print(mod_arima) Series: training ARIMA(0,1,1)(0,1,1)[12] Box Cox transformation: lambda= 0 Coefficients: SARIMA (0,1,1)(0,1,1) ma1 sma1 -0.3395 -0.7120 s.e. 0.0886 0.1032 sigma^2 estimated as 0.004313: log likelihood=148.14 AIC=-290.28 AICc=-290.06 BIC=-282.02 </pre>	<pre> > print(mod_autoarima) Series: training ARIMA(0,1,1)(0,1,1)[12] Coefficients: Auto SARIMA ma1 sma1 -0.4010 -0.6745 s.e. 0.0908 0.1049 sigma^2 estimated as 881312: log likelihood=-961.28 AIC=1928.57 AICc=1928.78 BIC=1936.83 </pre>
---	--	---

Source: Dossier technique de prévisions et conjoncture

Nous avons réalisé 2 modèles supplémentaire SARIMA avec les paramètres $(0,1,1)(0,1,1)$ et un modèle auto régressif.

Le modèle auto régressif renvoie automatiquement le meilleur modèle SARIMA selon la valeur des critères BIC ou AIC, ainsi les même critères doit être minimiser parmi plusieurs modèles pour pouvoir sélectionner un modèle [7]. De plus, le modèle auto-régressif utilise une méthode de lissage exponentiel de la série. Ce modèle trouve le même paramètres (p,d,q) $(P,D,Q)_{12}$ que X13 et SARIMA.

Dans la **figure 21**, nous voyons que la valeur des coefficients sont différentes parmi les trois modèles, mais varient faiblement. Donc, nous nous pencherons essentiellement sur les deux critères (BIC ou AIC) que les modèles renvoient. AIC est plus importante pour AUTO-SARIMA et X13, mais BIC est plus faible. Nous avons AIC et BIC qui sont plus faible pour SARIMA, donc selon ces critères, il serait peut-être plus judicieux de choisir ce modèle afin de procéder aux prévisions.

B) Prévision

Nous réalisons la prévision de la **série training** avec les modèles suivants : X13, SARIMA, SARIMA AUTO (auto régressif), ETS, NNETAR (Neural Network Time Series Forecasts), TBATS, BATS, STLM, Holt-Winters, NAIIVE.

Les modèles se différencient selon leur méthode de lissage ou d'autres caractéristiques et NAIIVE va servir de référence pour étudier la qualité des prévision. Nous portons une attention que la série training est un de **type multiplicatif** pour la réalisation des prévisions.

Pour prévoir les séries des modèles, nous procédons à une méthode algorithmique avec une prévisions récursive d'un pas d'un mois sur une années. Dans **la figure 22**, il y a les différentes étapes de l'algorithme pour calculer de la prévision. Forecast représente la fonction dans R pour prédire. Nous partons avec une série initiale training et nous incorporons en fur-à-mesure la nouvelle valeur pas-à-pas pour les 12 mois.

Figure 22: algorithme de prévisions récursive

Entrées : série y_{mod} et réel w

Initialisation : affecter à y_{mod} la **série training**

Pour i allant de 1 à 12 faire

$$\left| \begin{array}{l} y_{mod} = y_{mod}[08/2017 + i] \\ w = forecast(modele(y_{mod})) \\ y_{mod}[09/2017 + i] = w \end{array} \right.$$

Fin

Sorties : afficher la valeur de y_{mod}

Source : Dossier technique de prévisions et conjoncture

Dans la **figure 23**, nous avons représenté graphiquement toutes les séries de prévisions estimées avec la série brute en rouge pour avoir une idée de l'erreur de prédiction et le **tableau 3 (annexe)** montre toutes les valeurs prédictives des 12 mois.

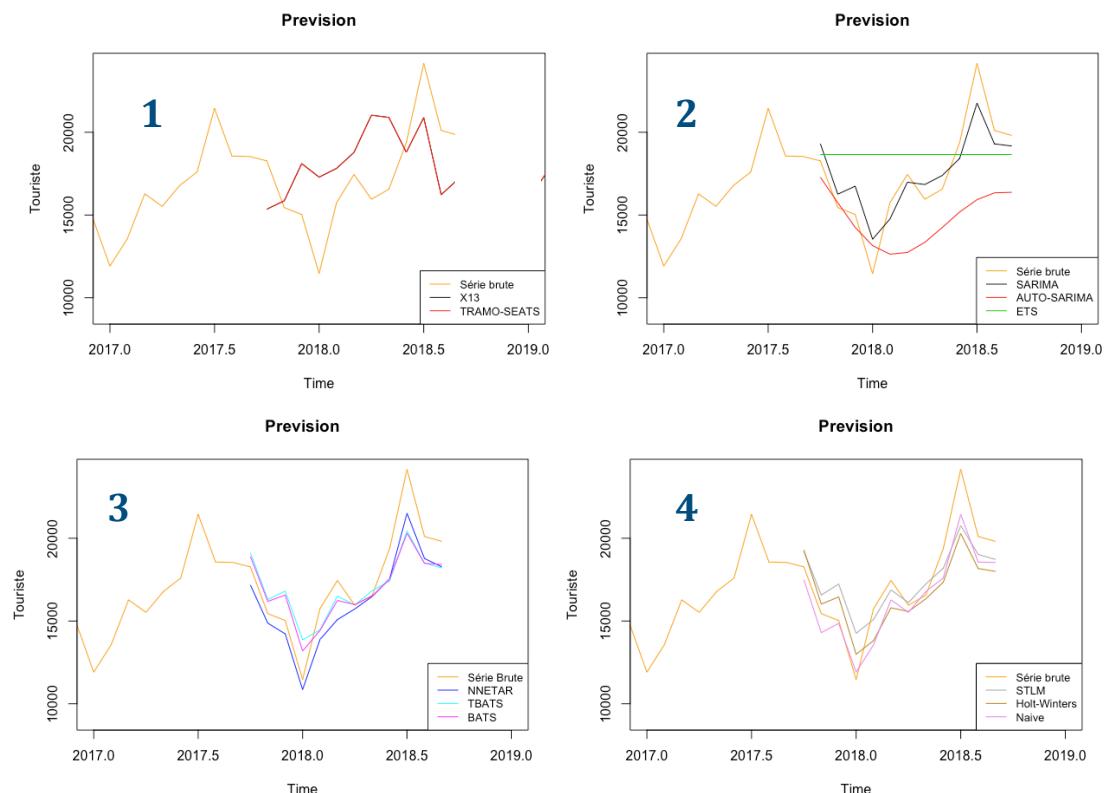
Pour X13 (ou TS), nous trouvons un grand écart entre la série test et la prévision X13 (en rouge, graphique 1) entre novembre 2017 et juin 2018, la prédiction avec X13 est peut être mauvaise.

L'auto SARIMA (en rouge, graphique 2) montre une série plus lisse par rapport aux autres modèles et un écart élevé plutôt vers la partie inférieure de la série brute et nous avons la série ETS qui prédit une droite horizontale (en vert, graphique 2)

Les modèles SARIMA, NNETAR, TBATS, BATS, STLM et HW semblent être bien meilleur. Nous constatons graphiquement que l'erreur de prédiction pourrait être faible car nous voyons très peu d'écart entre la série brute et les mêmes modèles.

Pour identifier le meilleur modèle, nous analyserons l'erreur de prévision dans la partie suivante.

Figure 23: graphique des prévision avec série brute et modèle

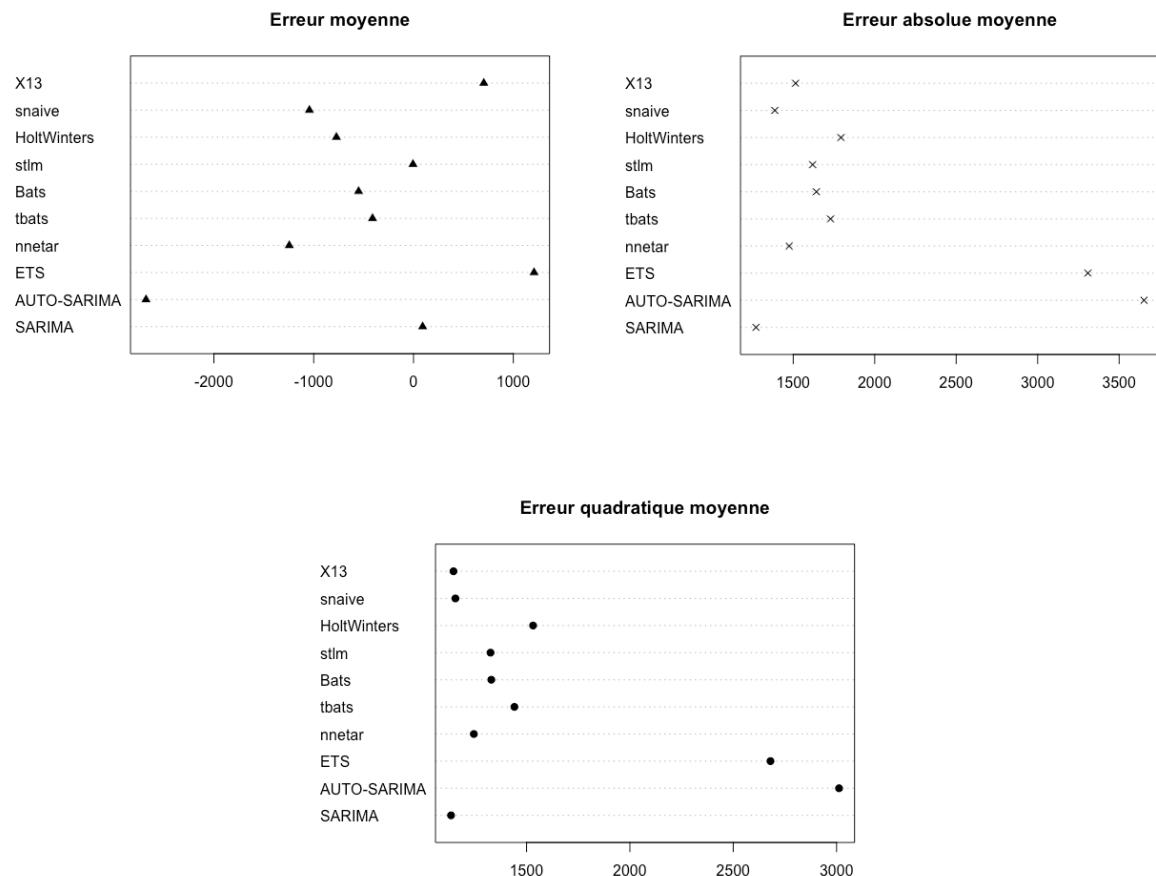


Source: Dossier technique de prévision et conjoncture

Remarque : Lors de la prévision des modèles TS et X13, nous remarquons que les modèles prédissent les mêmes valeurs et se confondent dans le graphique (*figure 23*). Dans la suite de l'étude, nous décidons de garder que la prévision de X13

C) Calcul des mesures d'erreur de prévision

Figure 24: graphique des mesures d'erreur de prévision



Les mesures d'erreur de prévision sont un moyen efficace pour examiner les erreurs de prédition .Nous nous baserons sur 3 indices : l'erreur moyenne (ME), l'erreur quadratique moyenne (MSE) et l'erreur absolue moyenne (MAE).

Dans le **tableau 4 (annexe)** montre les mesures d'erreur de prévision des 10 modèles et nous illustrons ces mesures dans la **figure 24** pour comparer plus facilement les modèles.

D'après le critère de l'erreur moyenne, il est judicieux de choisir quand la valeur est proche de 0. Ainsi, en regardant le graphique (*figure 24*), nous avons deux modèle qui est proche de 0 qui sont SARIMA avec 91,34 et STLM avec une valeur très faible de -5.25.

Selon le critère de MSE et MAE, il faudrait sélectionner des valeurs minimales (**figure 24**) parmi tous les modèles. Pour MAE, ARIMA et SNAIVE présentent des valeurs faibles. Pour le critère MSE, nous avons SARIMA, SNAIVE et X13 qui ont comme valeurs faibles.

Par ailleurs, la prévision naïve est considéré la meilleur pour les prévisions de séries, il fournit une référence utile pour d'autres méthodes de prévisions. Il est intéressant de comparer la série NAIVE avec les autres modèles. La ME du modèle est de -1043,17 (**tableau 4, annexe**), en la mettant en valeur absolue, la ME passe devant les modèles SARIMA, TBATS, BATS, STLM et HW. Notons que graphiquement, ces modèles reproduisent presqu'à l'identique la série test. Selon la MAE, nous avons les modèles X13 et SARIMA qui ont une valeur inférieure à la NAIVE. Pour finir, avec RMSE, nous trouvons uniquement le modèle SARIMA dont le RMSE est inférieur à celui de NAIVE.

Pour conclure, nous remarquons que le modèle SARIMA ressortent le mieux selon ME, MSE et MAE, mais aussi la comparaison avec le modèle de référence NAIVE dont SARIMA est bien meilleur.

D) **Test de Diebold-Mariano et multiple de Mariano et Preve (2012)**

Pour vérifier la qualité des prévisions, nous réaliserons deux test : le test de Diebold-Mariano (DM) et le test de Mariano et Preve (PM).

Le test de Diebold-Mariano [8] permet de savoir si les erreurs sont statistiquement significatives entre deux modèles par l'indicateur de RMSE. L'hypothèse nulle (H_0) impliquera que les deux modèles ont la même précision de prévisions sinon ils ont des niveaux de précision différents.

Par conséquent, nous réalisons le test avec un seuil de 5%, lorsque p-value est inférieur à 5%, nous rejetons H_0 . Si p-value est supérieur à 5%, nous acceptons H_0 . Lors des tests, nous utiliserons l'hypothèse alternative des deux cotés dont la précision des deux modèles seraient complètement différents et une prévision de 1 pas dans R.

Tableau 5: p-value du test de Diebold-Mariano

SARIMA	AUTO-S	ETS	NNETAR	TBATS	BATS	STLM	Holt-w	X13	NAIVE
SARIMA	0.04	0.04	0.44	0.07	0.21	0.09	0.09	0.28	0.63
AUTO-S		0.70	0.04	0.04	0.03	0.04	0.03	0.06	0.03
		ETS	0.07	0.06	0.06	0.04	0.09	0.03	0.06
			NNETAR	0.37	0.54	0.66	0.18	0.92	0.54
				TBATS	0.30	0.39	0.60	0.49	0.20
					BATS	0.90	0.04	0.74	0.32
						STLM	0.46	0.62	0.47
							Holt-W	0.49	0.07
								X13	0.76
									NAIVE

Le **tableau 5** montre les p-values du test de DM. En prenant la prévision de référence NAIVE, nous remarquons que H0 est rejeté entre NAIVE et AUTO SARIMA et elle est accepté avec tout les autres modèles. Nous remarquons pour les p-value qui sont proches de 5%, par exemple ETS et HW (p-value de 0.06 et 0.07), leurs RMSE sont très élevés et supérieurs à celui de NAIVE avec des valeurs respectives 3308.45 et 1793.07. Ainsi, il serait plus intéressant de s'intéresser au seuil de 10%.

Le modèle SARIMA retenu selon les trois critères de mesures d'erreur de prévision montre que H0 est accepté à l'exception du modèle AUTO SARIMA, ETS, TBATS, STLM et HW avec un seuil de 10%.

Les modèles ETS et AUTO-ARIMA sont les seuls prévisions qui ont des H0 rejetées avec tout les autres modèles sauf entre elles (pour seuil de 10%). Dans la figure 23, nous avions vu que les deux séries présentaient une droite horizontale et une série plus lisse que les autres, ce sont des résultats attendus avec le rejet de H0.

De plus, nous remarquons que les modèles HW, NNETAR, TBATS, STLM et X13 ont une acceptation de H0 entre elles. Cela peut s'expliquer que les 5 modèles utilisent une méthode lissage exponentiel particulière à court terme lors de la prévision.

Test de Mariano et Preve

Ce test multiple de Mariano et Preve [9] est un test lié à DM . Il permet de calculer la précision prédictive de plusieurs modèles. Nous avons H0 est l'ensembles des prévisions ont la même précision. L'hypothèse alternative est que l'égalité de la précision prédictive (Equal predictive accuracy (EPA)) n'est pas validé.

Ce test permet de classer l'ensemble des modèles selon la qualité de prédiction par rapport au test de Wald corrigé que nous nommons Sc et la moyenne des erreurs aux carrés (mean loss).

Dans la **figure 25**, nous avons réaliser le test multiple de MP et nous avons p-value qui est égale à 1 alors nous acceptons H0. Le test a rejeté la prévision AUTO-SARIMA et ne peut calculer Sc de X13.

Ensuite, le test nous renseigne le classement du meilleur modèle et nous trouvons bien au premier rang la prévision de SARIMA, suivi de Holt-Winters et ETS.

Pour conclure, les deux tests que nous venons de réaliser, ils nous ont permis d'identifier la meilleur prévision parmi l'ensemble des modèle dont supposé dans la partie précédente selon les mesures d'erreur de prévision. SARIMA présente une meilleur prédiction dans le futur à court terme avec la méthode de prédiction de pas-à-pas.

Figure 25: Test multiple de Mariano-Preve

Models with outstanding predictive ability:		
	Rank	Sc Mean loss
FCarima	1	-2.4033 1617082
FCets	3	2.0444 10945858
FCnnetar	6	-0.9682 2175683
FCtbats	8	1.1343 2989459
FCbats	5	0.1243 2695885
FCstlm	4	-0.7860 2621689
FCHW	2	2.0569 3215095
FCna	7	-0.3259 1923634
FCX13		2293227

p-value: 1

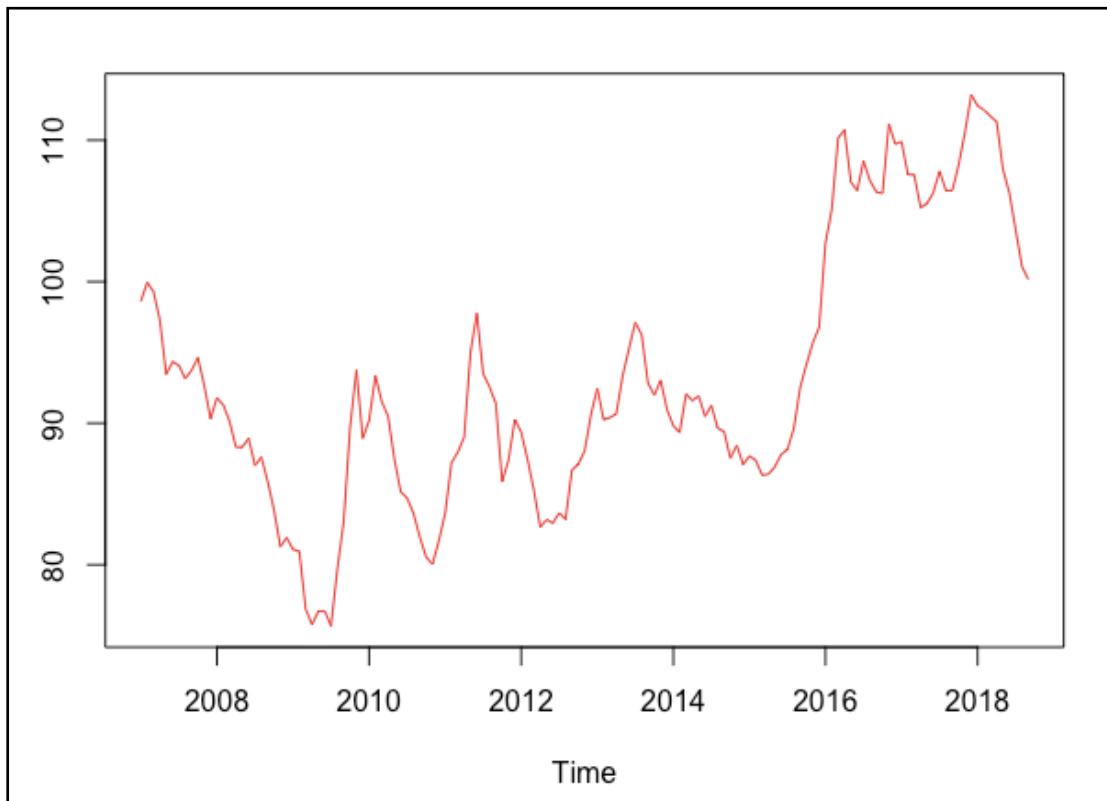
Number of eliminated models: 1

Source : Dossier Technique de Prévision et

III. Série non saisonnière et prévision

A) Analyse préliminaire

Figure 26: Série temporelle du taux de change \$/XPF



Source: Site FXTOP, auteur : Laurent PELÉ, copyright

1) Description de la série

Après avoir traité la série saisonnière du nombre de touristes arrivés dans le mois, nous nous focalisons maintenant sur une série non saisonnière.

Nous avons choisi la série non saisonnière, le taux de change du dollar vers le franc pacifique (**XPF**) en calculant la moyenne dans le mois (en fin de clôture). Le taux de change (**tc**) est une variable cohérente dans une économie ouverte où les fluctuations du **tc** ont des impacts sur tous les secteurs économiques et dans le cas de cette étude celui du tourisme.

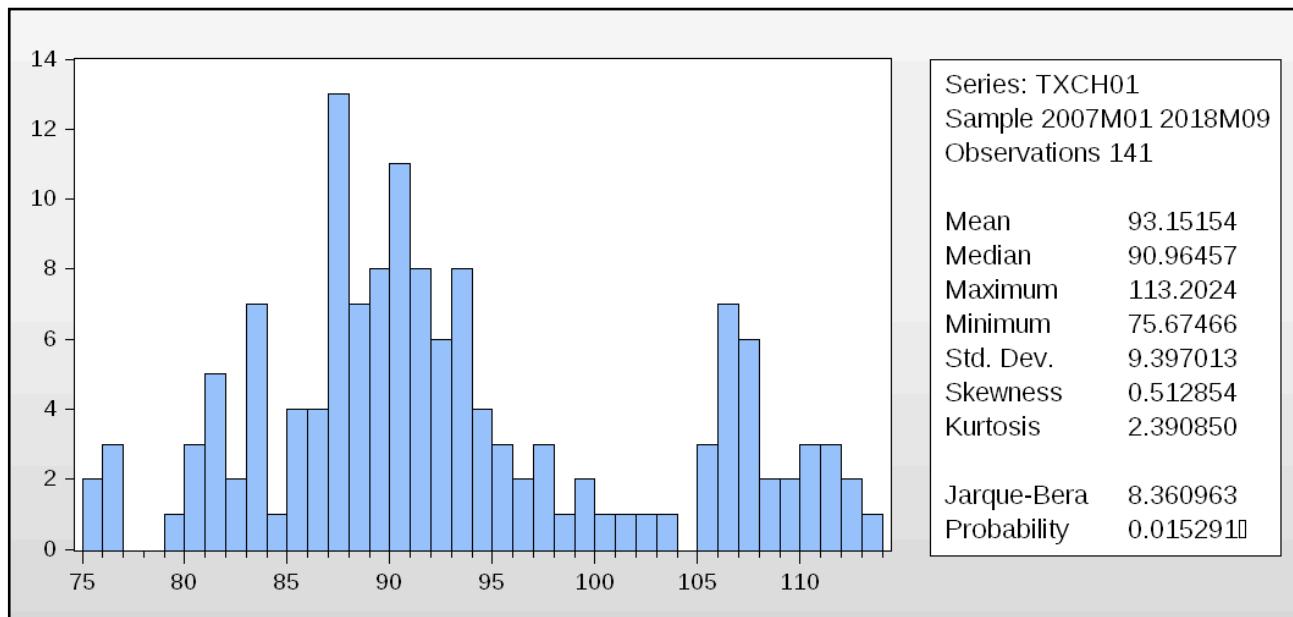
La monnaie XPF est utilisée dans trois territoires d'Outre-Mer (Nouvelle Calédonie, Wallis, Futuna et la Polynésie française) et suit la même fluctuation du taux de change dollar/euro [10].

En outre, nous utilisons le dollar dans notre étude, car c'est la première devise la plus utilisée dans le monde pour les transactions et que la proportion du touriste américain est la plus importante en Polynésie française.

Les données de la série ont été extraites du site FXTOP et nous les traiterons dans la même période que la série saisonnière (01/2007-09/2018 avec 141 observations). Nous les présentons graphiquement dans la **figure 26**.

Visuellement, le tc décroît entre le début de la série et au troisième trimestre de 2009, le mois de juin trouve le tc le plus faible de la série (min = 75.67), puis il augmente brusquement au moins de novembre 2009 pour former un pic. Tc forme d'autres pics en mois de juin 2011, juillet 2013 et plusieurs entre avril 2014 et décembre 2017. Après novembre 2009, le tc présente une tendance croissante.

Figure 27 : Distribution de la série et description statistique



Source : données de FXTOP - logiciel Eviews

En outre, la **figure 27** illustre la distribution de la série et la description statistique. La moyenne du tc est de 93.15 et une médiane assez proche de 90.96. L'écart-type est de 9.39 qui est très faible par rapport à la moyenne et la médiane, donc la dispersion de l'échantillon se trouve autour de la moyenne. Nous pouvons dire que les observations sont homogènes.

Puis, le skewness et le kurtosis sont de signe positifs et indiquent que la distribution de l'échantillon est plus pointue qu'une loi normal et qu'elle s'étale vers la droite. Selon le test Jacque-Bera, l'échantillon ne suit pas une loi normale ($p\text{-value} < 0.05$). Graphiquement, les barres présentent des fréquences élevées pour le tc entre 106 et 107, alors que la moyenne de l'échantillon est de 93.15 ce qui nous pousse à dire que la série ne suit donc pas une loi normale.

Notons aussi que la série ne présente qu'une unique valeur atypique pour le mois de décembre 2017, cette observation est aussi la valeur maximum de la série (**figure 28, annexe**).

2) Détection des points atypiques par TRAMO

Pour détecter les valeurs atypiques, nous utilisons le logiciel Jdemetra sous R par TRAMO. La fonction trouve 4 valeurs atypiques de types LS (level shift) :

- Décembre 2009
- Mai 2011
- Octobre 2011
- Mai 2016

Il est important d'expliquer la présence des différentes valeurs atypiques. Depuis la grande Récession de la crise financière de 2007 à 2009, pour stimuler l'économie, la FED [11] devait recourir à des taux de 0% pour sortir de la crise par trois phases d' « assouplissement quantitatif » par des rachat des actifs douteux et stimuler l'activité. La FED procède à des politique monétaire et annonce le quantitative easing 1 (QE1) pour des achats non stérile de la dette comme la dette publique et s'engage de maintenir sur une longue période de taux de très bas. Fin de 2009, l'économie US sort de la récession et à des répercussion sur le tc dollar/euro, ce qui explique le premier point atypique de décembre 2009 de la haute de tc de dollar/XPF.

Pour le deuxième point atypique, la FED lance le QE2 en achetant 600 milliards \$ de dette publique et de continuer à réinvestir afin d'ancrer les anticipations sur le FED Funds puis de les maintenir en baisser les taux jusqu'au maturité les plus longues (environ 2 ans).

Le troisième point atypique montre un creux, la FED lance le 21 septembre 2011 l'opération twist de réallocation de son bilan pour réduire le taux d'intérêt de long terme, en échangeant des obligations publiques à courte terme contre de long terme. Cette politique monétaire était appliquée pour réduire le chômage, éviter une phase de récession et minimiser les craintes d'une explosion inflationniste.

Le dernier point atypique est élevé en mai 2016, ce phénomène s'explique par l'anticipation de l'euro et l'expansion de l'économie mondiale. La Banque Centrale Européen (BCE) accroît l'activité et l'inflation dans la zone euro, la FED procède à des politiques monétaires en amplifiant la réactivité du tc et sur d'autres actifs. [12]

Les points typiques dépendent des politiques monétaires américaines (FED) et européennes (Banque Centrale).

3) Vérification de la stationnarité I(0) et I(1)

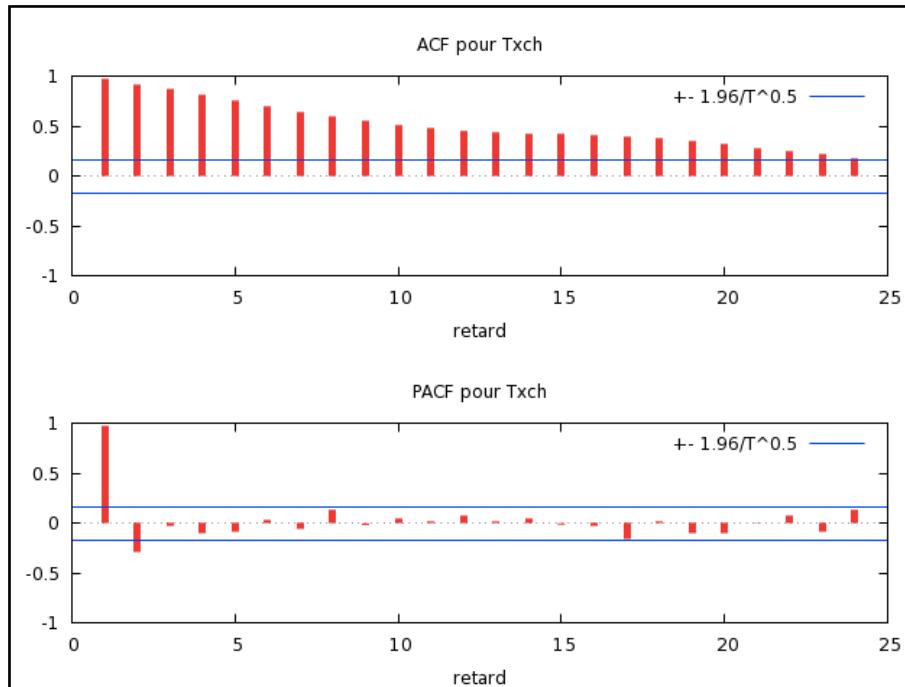
Pour étudier la stationnarité de la série non saisonnier du tc, nous allons tester les corrélogrammes de la série sans et avec différenciation . Les corrélogrammes sont des graphes de la fonction d'autocorrélation (FAC) et de la fonction d'autocorrélation partiel (FACP).

Brièvement, La FAC est composé de l'ensemble de l'autocorrélation tel que :

$$\rho_k = \text{corr}(z_t, z_{t+k}), \text{ avec } k \text{ qui représente le nombre de décalage et } z_t \text{ la série.}$$

La FACP mesure la corrélation partielle entre z_t et z_{t-h} influençant les autres variables décalées de h périodes ($z_{t-1}, \dots, z_{t+h-1}$) ayant été retirée.

Figure 30 : corrélogramme de la série sans différenciation



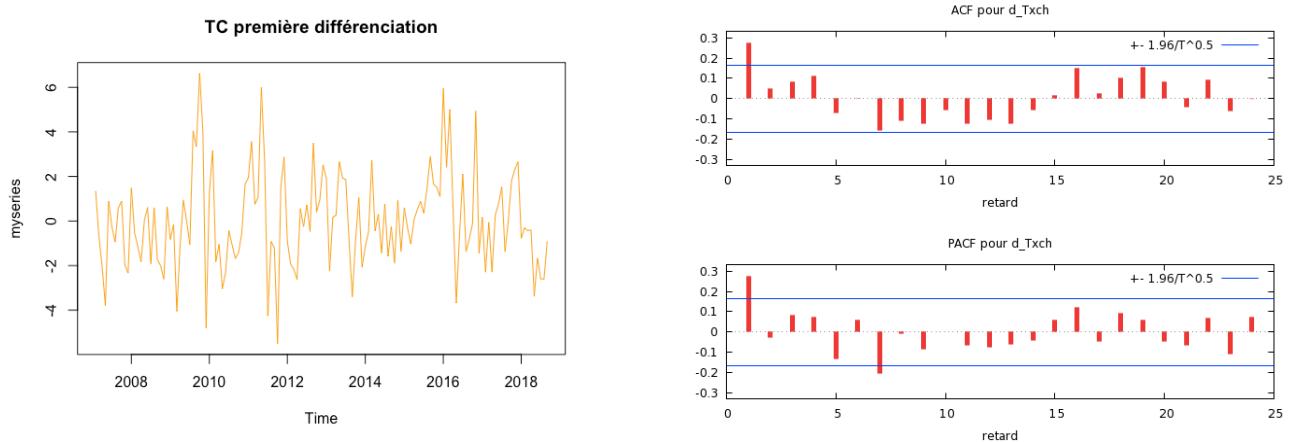
Source : logiciel Gretl

La FAC est présent dans la **figure 30** avec un décalage de 24, il y a une décroissance lente des retards de la série non saisonnière sans différenciation I(0). La droite horizontale en bleu sur le graphique indique le seuil critique au-delà duquel l'autocorrélation est considérée significative. Tous les retards (indiqués en rouge) sont fortement significatifs car ils sont au-dessus de l'intervalle (calculé en fonction de la loi de Student). Les tests sont réalisés pour les 24 retards dans la **figure 30.A (annexe)**. De même, pour FACP, nous avons les deux premiers retards qui sont fortement significatifs. À partir de ces résultats, nous pouvons conclure que la série n'est pas stationnaire alors il est primordial de procéder à une première différenciation de la série I(1).

En procédant à I(1), la **figure 31** montre la série différenciée en premier degrés et son corrélogramme. Graphiquement par rapport à la nouvelle série, nous pouvons voir que l'échantillon évolue autour d'une moyenne mobile, ce qui nous mène à penser que la série est stationnaire. Le graphique du corrélogramme, la FAC et FACP indiquent que seul le premier retard est significatif et montre une rupture par rapport à FAC et FACP de I(0). Ces résultats

nous indiquent que la série est bien stationnaire à degrés 1. Donc nous pouvons continuer l'étude avec I(1).

Figure 31 : TC avec I(1) et son corrélogramme



Source : Logiciel R et Gretl

B) Estimation et prévisions des modèles

4) Estimation des modèles linéaires

Nous allons estimer les modèles AR(1), AR(1,1,0), AR(1,1,1), AUTO ARIMA et la méthode LED Holt-Winters sur la série test de la même période que pour la série saisonnière (janvier 2007 à septembre 2017). Nous allons commenter les paramètres et de leurs résidus.

En premier lieu, la modélisation de AR(1) est une estimation par la méthode moindre au carré ordinaire (MCO), c'est une MCO avec une spécification de la variance-covariance et prend en compte que la partie « auto-régressif » p . Dans cette modélisation, nous avons choisi une spécification ordinaire.

Puis, le modèle LED Holt-Winters [13] s'agit d'un lissage exponentielle double. Cette méthode LED permet de traiter les séries présentant un tendance linéaire mais sans saisonnalité. Nous avons les paramètre de lissage qui sont définis par alpha tel que : $\alpha = \beta^2$ et $\gamma = \frac{2\beta}{1 - \beta^2}$.

En comparant tous les modèles (**figure 32, annexe**), nous trouvons tous les coefficients de AR(p) positifs à l'exception du modèle AUTO ARIMA. Les modèles AR(1), AR(1,1,0), AR(1,1,1) et AUTO ARIMA vérifient les conditions de stationnarité et d'invisibilité car la valeur absolue du coefficient de AR est inférieure à 1. Notons que le modèle AUTO ARIMA trouve les

paramètres (1,0,2). Nous avons aussi AR(1) qui présente une qualité de modélisation faible avec un R2 de 6%.

Le modèle LED HW a pour paramètre $\alpha = 0.49$ et $\beta = 0.12$ et les constantes de lissage a et b sont de valeurs négatives : a = -0.15 et b = -0.15. Nous constatons que β est proche de 1 alors la prévision sera plus souple et fortement influencé par les observations les plus récentes.

Nous nous focalisons sur le logarithme de vraisemblance (log likelihood) pour comparer les modèles. ARIMA(1,1,0) a le log likelihood le plus faible parmi les quatre modèles (sauf pour HW). Selon le critère AIC, le modèle AUTO ARIMA a la valeur minimum parmi les modèles ARIMA (1,1,0) et ARIMA (1,1,1), nous pouvons supposer que AUTO ARIMA pourrait proposer de meilleure estimation lors de la prévision.

En outre, nous constatons que tous les résidus suivent un loi normal selon les test de Jarque-Bera pour un seuil de 5%. Nous avions aussi vérifier que les résidus vérifient bien l'hypothèse de non-corrélation associé au bruit blanc (**figure 33, annexe**).

Graphiquement (**figure 34, annexe**), les résidus des modèles AR(1), AR(1,1,0), AR(1,1,1) et AUTO ARIMA se confondent pratiquement avec la série I(1) et nous remarquons un écart entre les pics de la série brute et les résidus. Nous pouvons dire que les modèles ont bien estimée la série.

Cependant, entre les résidus du modèle HW et la série brute I(1), nous voyons une différence tout le long de la période. Ceci peut s'expliquer par le paramètre beta qui est proche d'une valeur nul.

5) Prévision avec variables explicatives

Dans cette partie, nous allons procéder des prévisions en exploitant des variables explicatives. Les variables explicatives ont été choisis après une étude économétrique du tourisme vu dans la littérature [14 et 15]. Nous proposons comme variables explicatives, le nombre de arrivée de touristes en Polynésie française vu dans la partie saisonnière et nous avons présentés dans l'annexe les fluctuations de certaines variables (**figure 35.A**).

Nombre de touriste

Le nombre de touriste américain et le nombre de touriste au total, nous supposons que le tc et cette variable ont un lien cohérent, ie que le pouvoir d'achat d'un tourisme peut varier en fonction des fluctuations du tc entre le dollar et XPF. Plus le XFP sera avantageux pour le touriste américain alors plus le touriste aura la capacité de voyager et dépenser en Polynésie française.

Variable chambre offert et vendu

Nous avons choisis aussi le nombre de chambre vendu dans les hôtels de luxe (plus de 2 étoiles)et non luxe (2 étoiles et moins). Puis, nous calculons le ratio de l'offre et chambre vendu des deux types de chambre vendu sur le marché. C'est un bon indicateur pour expliquer les fluctuations du tc par les biens vendus ou offerts en Polynésie Française pouvant expliquer le pouvoir d'achat d'un tourisme.

Variable Google Trend

La variable Google Trend permet de suivre les tendances et les comportement des consommateurs par rapport à leur recherche de requête sur le site. Les données se révèlent être pertinentes pour prévoir l'arrivée des touristes sur le territoire. Les mots-clés recherchés sont = « Hotel » + « tahiti ». Google réalise une mise en échelle des valeurs qui varient entre 0 et 100, avec 100 représente le record de trafic mensuel pour la requête de la recherche des mots-clés. Nous avons sélectionner que les Etats-Unis comme secteur de recherche.

Différenciation

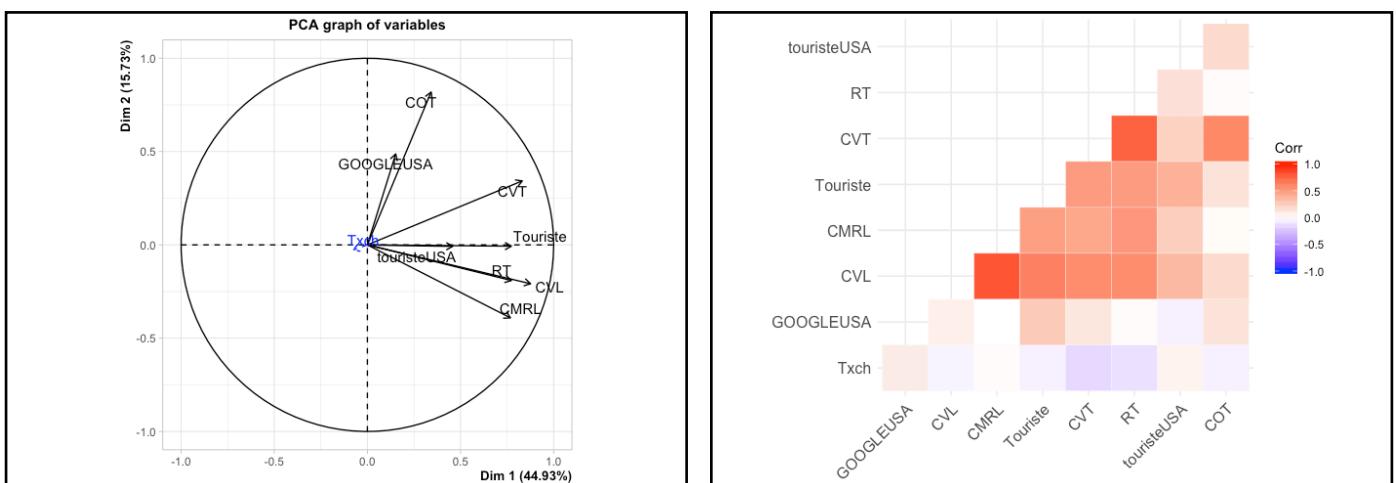
Tout d'abord, nous devons procéder à la différenciation de même degrés que la série tc I(1) pour les séries explicatives. Pour vérifier la stationnarité des séries, nous utilisons le test Dickey-Fuller augmenté (AFD) [16] sous le logiciel Gretl. H0 est la présence d'une racine unitaire dans la série et que la série n'est pas stationnaire. Nous rejetons H0 au seuil de 5% lorsque la p-value est inférieur à la valeur critique. En vérifiant pour toutes les séries, nous

trouvons qu'elles sont toutes stationnaires de degrés 1 (**figure 35.B, annexe**), donc nous pouvons passer à la partie suivante par le choix des variables.

Pour sélectionner des variables explicatives afin d'avoir des modèles de prévision performante, nous allons justifier nos choix des variables en utilisant deux méthodes : Analyse en Composantes Principales (ACP) et la régression linéaire multiple par Step-Wise avec les séries différenciées de degrés 1.

Analyse en Composantes Principales (ACP)

Figure 36 : ACP et matrice de corrélation



Source: Library Factominer, logiciel R

La figure 36 montre la corrélation entre les variables. Les cases de couleur dégradé rouge montre un corrélation positive et les cases de couleurs violet montre une corrélation négative. Nous avons le tc qui a une relation faiblement positive avec Google USA et touristeUSA et une relation moyennement négative avec CVT (chambre vendu non luxe), COT (chambre offert non luxe) et RT (ratio offre/demande de chambre non luxe).

D'après le cercle de l'ACP, nous remarquons que le tc est plus proche du premier axe et très proche du centre de gravitation. Si le tc était éloigné du centre alors il est bien expliquer par l'axe. En examinant les variables qui contribuent le plus la première axe, nous avons le nombre de tourisme et tourisme américain. Pour le deuxième axe, nous avons google USA et COT qui contribuent positivement. Ainsi, nous devons proposer un modèle avec l'un des deux variables de chaque axe pour éviter les corrélations entre les variables.

Stepwise

Ensuite, nous procédons à la régression linéaire classique par step-wise dans la direction backward, cette méthode nous permettra de retenir le meilleur modèle MCO par le critère AIC. Nous avons la régression linaire suivante :

$$Y_{tc} = \beta_0 + \beta_1 X_{USA} + \beta_2 X_{RT} + \beta_3 X_{CVT} + \beta_4 X_{touriste} + \beta_5 X_{CMRL} + \beta_6 X_{CVL} + \beta_7 X_{GOOGLE}$$

Dans la **figure 37(annexe)**, le step-wise retient la variable CVT pour le dernier modèle avec un AIC le plus faible (AIC = 214.64). Le deuxième modèle qui retient notre attention avec un AIC de 215.06, a deux variables, Google USA et CVT. Nous retenons aussi le quatrième modèle avec les variables Tourisme USA, CMRL, Google USA et CVT.

Choix des variables

Par conséquent, afin de choisir les variables qui définiront notre modèle au mieux, nous combinons par parcimonie selon les deux méthodes de choix de variables. Par rapport à stepwise, nous ne gardons pas la variable CMRL car la corrélation avec le tc est nulle (**figure 36**).

Modèle 1 : tc = CVT + GOOGLE USA

Modèle 2 : tc = CVT

Modèle 3: tc = Touriste USA + CVT

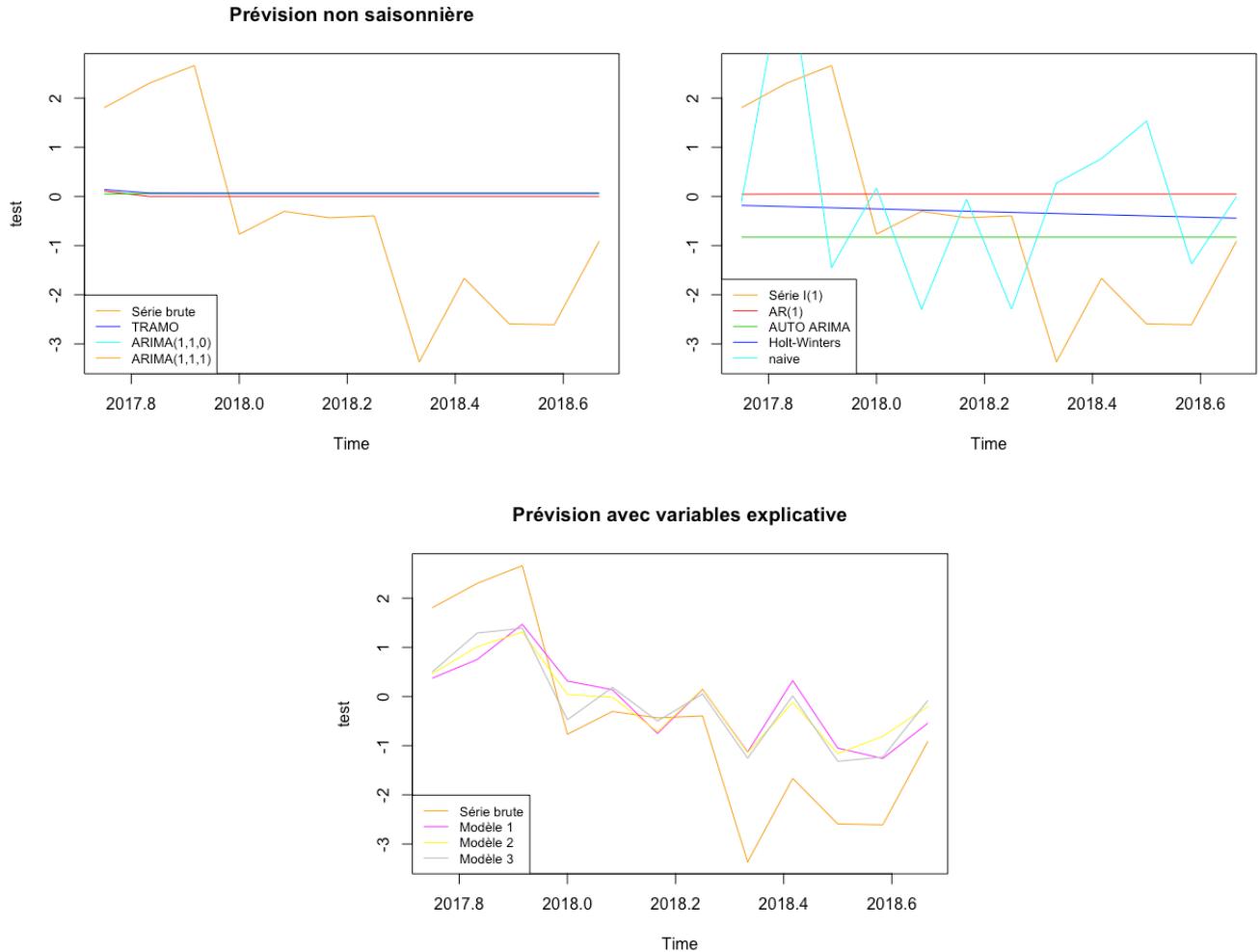
Commentaire sur les 3 modèles :

Pour les trois modèles, nous avons réalisé la méthode MC0 avec une variance-covariance ordinaire comme le modèle AR(1) (figure 37). En comparant les modèles selon le R2, le modèle 1 est bien meilleur parmi les trois modèles et celui de AR(1), avec un R2 de 8%. Le modèle 2 présente un log de vraisemblance plus faible que les deux autres. Puis, les trois modèles ont le coefficient constant et ar(1) positive et les coefficients des variables explicatives sont faibles mais aussi négative pour CVT..

Nous avons les résidus des modèles qui suivent une loi normal selon le test de Jarque-Bera. Nous avons présentés graphiquement les résidus des 3 modèles avec la tc différencié (**figure 38, annexe**) et nous constatons que les résidus se superpose de la tc différencié avec un décalage vers la gauche.

6) Prévision des modèles

Figure 39 : Prévision des modèles avec série tc I(1)



Source : Dossier Techniques de prévision et conjoncture

Nous réalisons la prévision des modèles vus précédemment sur une année avec un pas de 1 mois en utilisant le même algorithme dans la série saisonnière. Nous décidons d'ajouter le modèle X13 dans l'étude pour se donner une idée de la prévision.

La **figure 39** montre la prédiction des modèles entre octobre 2017 et septembre 2018, nous avons représenté la **série test** en orange et les valeurs des prévision dans le **tableau 6 (annexe)**. Les prévisions TRAMO, ARIMA(1,1,0), ARIMA(1,1,1) et AR(1) forment une droite horizontale sur l'axe nulle. AUTO ARIMA est aussi une droite horizontale se trouvant au-dessous de l'axe et stagne à -0.82. Holt-Winters prédit une décroissance faiblement et négativement.

Dans **la figure 40**, nous constatons visuellement que les écarts entre les prévisions et la série test sont négatives pour la majorité des prévisions entre octobre 2017 et décembre 2017. Elles sont faiblement positives entre janvier 2018 et avril 2018 et fortement de mai 2018 à septembre 2018.

Les modèles de prévision avec variable explicative fluctuent en fonction des variations de la série test et sont proche entre eux. Entre octobre 2017 et décembre 2017, les 3 modèles se trouvent au-dessus de la série test puis au-dessous jusqu'au septembre 2018(à l'exception du mois mai 2018). Nous constatons que l'écart entre les trois modèles et la **série test** est faible. (**figure 40**)

Pour finir, nous prédissons le modèle naïve qui va servir de référence dans la mesure des erreurs et nous remarquons Naive fluctue aussi de la même manière que la **série test**. Nous pouvons dire que la naïve, les modèles 1, 2 et 3 pourraient être bien meilleur pour prédire la **série test**.

7) Calcul des mesures d'erreur de prévision

Le tableau 7 présente le calcul des mesures d'erreur de prévision des modèles.

Tableau 7 : Mesures d'erreur de prévision

Modele	ME	MAE	MSE
X13	0,53	1,64	1,93
ARIMA(1,1,0)	0,59	1,68	1,96
ARIMA(1,1,1)	0,59	1,68	1,95
AUTO	-0,31	1,47	1,89
Holt-W	0,21	1,44	1,81
AR(1)	0,57	1,68	1,95
CVT+GOOGLE	0,42	1,17	1,32
CVT	0,42	1,14	1,28
USA+CVT	0,4	1,01	1,17
Naive	0,53	2,18	2,49

Tout d'abord, les modèles 1, 2 et 3 ont des valeurs MAE et MSE les plus faibles (**tableau 7**). Nous trouvons HW qui a une ME la faible et proche 0. Le modèle AR(1), ARIMA(1,1,0) et ARIMA(1,1,1) présentent la MAE et MSE le plus élevé dans le tableau.

En comparant avec le modèle référence Naive, tout les modèles ont des valeurs inférieurs selon les deux critères MAE et MSE. Nous avons ARIMA(1,1,1), ARIMA(1,1,0) et AR(1) qui se rapprochent du modèle Naive car MAE et MSE sont supérieurs des autres modèles. Nous pouvons supposer que les modèles ayant la MAE, MSE et la ME proche de celui de la Naive sont considérés comme de bonne prévision. Nous allons la démontrer dans la partie suivante.

8) Test de Diebold-Mariano (DM) et test multiple de Mariano et Preve (MP)

Tableau 8 : Test de Diebold-Mariano

X13	ARIMA(1,1,0)	ARIMA(1,1,1)	AUTO_ARIMA	Hotl-W	Modele 1	Modele 2	Modele 3	AR(1)	NAIVE
X13	0.09	0.10	0.88	0.23	0.02	0.01	0.01	0.12	0.07
	ARIMA(1,1,0)	0.43	0.80	0.20	0.02	0	0	0.15	0.09
		ARIMA(1,1,1)	0.80	0.23	0.02	0	0.01	0.91	0.09
			AUTO_ARIMA	0.64	0.13	0.10	0.07	0.81	0.08
				Hotl-W	0.06	0.03	0.02	0.20	0.03
					Modele 1	0.58	0.05	0.02	0.01
						Modele 2	0.07	0.00	0.01
							Modele 3	0.00	0.01
								AR(1)	0.08
									NAIVE

Test de Diebold-Mariano

Tableau 8 indique la p-value du test de DM entre tout les modèles. Rappelons que H0 indique que les deux modèles de prévisions ont la même précision et si la p-value est inférieur à 5% donc nous refusons H0.

En premier lieu, nous nous intéressons à la dernière colonne du tableau 8, le modèle NAIVE a un niveau de précision différente avec les modèles 1, 2, 3 et HW et un même niveau avec le reste des modèles. Nous avons une p-value plus forte avec le modèle ARIMA(1,1,0) et ARIMA(1,1,1). A ce stade, nous pouvons dire que le modèle ARIMA(1,1,0) sera de meilleur prévision selon les trois critères (ME, MSE et MAE) et le test DM en prenant pour référence NAIVE.

Ensuite, les trois modèles avec des variables explicatives ont un niveau de précisions différentes avec tout les modèles sauf AUTO-ARIMA. Nous avions vu que les trois modèles étaient très proche mutuellement et suivent les variations de la **série test**, ainsi ils ont un même niveau de précision.

Puis, nous constatons que la p-value de AUTO-ARIMA est élevé avec X13, ARIMA(1,1,0), ARIMA(1,1,1), AR(1) et HW. Dans la figure 39, les prévisions étaient toutes des droites horizontales se trouvant avoisinante de l'axe des abscisse, ce qui explique le même niveau de précision.

Test de Mariano-Preve

Pour finir, nous utilisons le test MP afin de classer la précision du modèle. Le test MP (**figure 41**) renseigne que ARIMA(1,1,0) est la meilleure prévision suivie de la NAIVE. AR(1) est le seul modèle qui a été supprimé du test, ainsi ayant une p-value égalant à 1, nous acceptons H0 que l'ensemble des prévisions ont la précision.

Nous avons les trois modèles de prévision avec variables explicatives et HW qui sont classés en dernier. Le modèle 1 est considéré comme le meilleur parmi les trois, nous avions vu que le même modèle 1 avait un R2 le plus élevé à 8%.

Les modèles ARIMA (1,1,0), (1,1,1) et AUTO ARIMA sont considérés comme de bonne prévision par rapport aux tests effectués. Ces modèles utilisent les valeurs passées de la série pour prédire les valeurs des futures observations dans le court-terme, c'est pour cela qu'ils prédisent des lignes horizontales.

Figure 41 :Test MP

```
#####
Models with outstanding predictive ability:

      Rank      Sc  Mean loss
FCnots      5 -1.7244   3.7285
FCnoar      1  1.1563   3.8384
FCnoarima    3  0.2358   3.8094
FCnoauto    4  0.4952   3.5804
FCnoHW       7 -2.5195   3.2923
FCnona      2  2.0163   6.2175
forecast1     6  0.5881   1.7444
forecast2     8  2.0233   1.6281
forecast3          1.3643

p-value: 1

Number of eliminated models:  1
#####
```

Source :Dossier Techniques de prévision et conjoncture.

C) Test des modèles de prévision avec la série brute

Nous avons testé précédemment les séries avec leur différenciation. Maintenant, nous testerons de nouvelles séries de prévision avec leur transformation afin d'étudier l'écart avec la série brute.

La différenciation de la série se calcule tel que :

$$Z_t = y_t - y_{t-1}, z_t \text{ est la série différencié et } y_t \text{ la série brute}$$

Et nous cherchons la nouvelle série de prévision tel que : $y_t = Z_t + y_{t-1}$

Après avoir calculer les nouvelles séries de prédition, la **figure 42** montre les nouvelles prévisions avec la série brute sans différenciation.

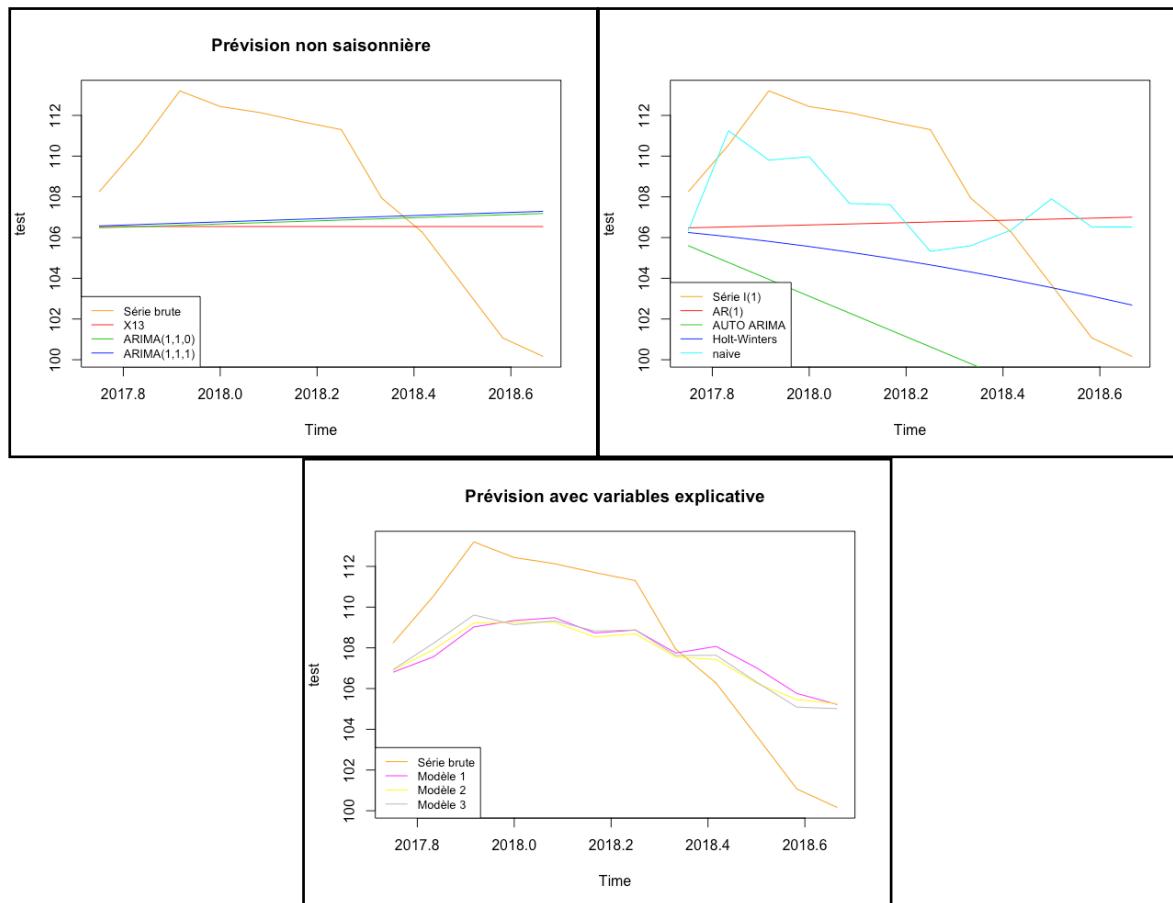
Nous avons les modèles X13 (TS) qui prédit une droite horizontale sur les 12 mois. ARIMA(1,1,0), ARIMA(1,1,1) et AR(1) croit faiblement. AUTO ARIMA décroît fortement et prédit au-dessous de la série brute. Le modèle HW prédit aussi une décroissance mais moins faible que celui d'AUTO ARIMA.

Les 3 modèles avec variables explicatives se confondent pratiquement entre elles et suivent la tendance de la série brute.

Par rapport au modèle NAIVE, nous constatons que le modèle reproduit la série brute des 12 mois précédent (*figure 43, annexe*) entre 09/2016 et 10/2017 pour prédire les 12 mois de *la série test*.

L'écart des erreurs des modèles suivent la même tendance (*figure 44, annexe*), L'ensemble des erreurs sont négatifs entre 10/17 et 05/18 puis ils deviennent positifs entre 06/18-08/18. A l'exception de la série AUTO-ARIMA dont l'écart des erreur reste négatif tout le long de l'année avec un écart très élevé.

Figure 42 : Prévision des modèles avec série brute



Source : Dossier Techniques de prévision et conjoncture

Maintenant, nous nous intéressons aux mesures d'erreur de prévision (**tableau 9, annexe**). Toutes les prévisions ont une erreur moyenne (ME) négative. Le ME ayant une valeur proche de 0 sont les 3 modèles de prévisions avec variable explicative. Nous pouvons dire aussi que les même modèles ont la MAE et MSE la plus faible.

Le modèle AUTO a la ME le plus éloigné de 0 donc une MAE la plus élevé (7.17) et MSE avec 7.66. Il serai possible de ne pas retenir ce modèle à cause des grosses valeurs de MAE et MSE.

En comparant les modèles avec le modèle de référence NAIIVE, ME est proche de 0 avec -0.65 puis la MAE de NAIIVE est 3.45 et MSE est de 3.95. Selon MAE et MSE, ARIMA(1,1,0), ARIMA(1,1,1), X13 et AR(1) sont proche et supérieur de la NAIIVE, nous supposons que ces modèles sont de bonnes qualités.

En réalisant le test MP (**figure 45**), AR(1), modèle 2 et AUTO ARIMA ne sont pas retenus dans le test et p-value est supérieur de 0.05 donc nous acceptons H0.

Le modèle X13 (y_modnnts) est considéré la meilleur prédition avec une erreur moyenne au carré de 21.60. Au deuxième position, nous avons ARIMA(1,1,1) (y_modnnoarima) puis ARIMA(1,1,0) (y_modnoar). Nous trouvons le même classement en tête par rapport au test avec la série différencié sauf pour AUTO ARIMA.

Le modèle 1 est toujours considéré comme le meilleur modèle parmi les trois modèles avec variable explicative.

Figure 42 : Prévision des modèles avec série brute

Models with outstanding predictive ability:		
	Rank	Sc Mean loss
Y_modnnts	1	-0.4475 21.6099
Y_modnoar	3	0.6082 22.2990
Y_modnnoarima	2	-0.2414 22.0362
Y_modHW	4	0.8085 24.2750
Y_modnnona	6	1.5062 15.6241
Y_mod1	5	0.9886 10.1407
Y_mod3		8.4525
p-value:	1	
Number of eliminated models:	3	

Source : Dossier Techniques de Prévision et Conjoncture

Conclusion

Pour conclure, ce dossier a permis de nous exercer sur les différentes techniques de prévisions et conjoncture. Ainsi, nous avons désaisonnalisé une série en deux parties : CVS et Trend, puis d'analyser les cycles de cette série par les datations de point de retournement. Puis, nous avons réaliser une étude sur les prédictions des séries saisonnière et non saisonnières.

Nous avons comme résultat que les modèles ARIMA (1,1,1) et ARIMA(1,1,0) sont considérés de bonne qualité selon les test MP et DM pour la prévision lorsque nous avons étudiés avec les séries et différencié puis transformé. En se rappelant du critères AIC, il serait préférable de choisir le modèle ARIMA(1,1,0) car le AIC est inférieur à celui de ARIMA(1,1,1).

La modélisation ARIMA utilise de façon très simple les données récentes ou plus anciennes afin de réaliser les prévision adéquates concernant le comportement futur. Ainsi, la série doit être stationnaire et ne doit présenter aucune tendance pour pouvoir prédire.

Cependant, il faut porter une attention particulière sur les modèles de prévision avec variable explicative qui ne sont pas considéré dans les meilleurs selon le test MP. Le modèle étant le meilleur avec les variables de nombre de chambre vendu hors luxe et la variable Google TREND, a pu prédire avec une erreur moyenne de -0,41 soit une erreur moyenne de 4%.

Dans l'ensemble de ce dossier, nous avions vu que des simples paramètres de ARIMA(1,1,0) et SARIMA(0,1,1)(0,1,1) pouvaient prédire au mieux le future. La partie saisonnière prenait en compte que la partie q et Q, moyenne-mobile et non saisonnière la partie auto-régressif p avec une intégration de degrés 1.

Bibliographique

- [1] https://publications.banque-france.fr/sites/default/files/medias/documents/working-paper_147_2006.pdf
- [2] <https://www.census.gov/srd/www/x13as/>
- [3] <https://www.istat.it/it/files/2014/06/Joint-Estimation-of-Model-Parameters-and-Outlier-Effects-in-Time-Series.pdf>
- [4] <https://www.cairn.info/revue-d-economie-politique-2012-3-page-365.htm>
- [5] <https://www.cairn.info/revue-d-economie-politique-2012-3-page-365.htm>
- [6] <https://www0.gsb.columbia.edu/faculty/rhodrick/prescott-hodrick1997.pdf>
- [7] <https://cran.r-project.org/web/packages/forecast/vignettes/JSS2008.pdf>
- [8] <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.4490&rep=rep1&type=pdf>
- [9] <https://www.sciencedirect.com/science/article/pii/S0304407612000152>
- [10] <https://www.wgtn.ac.nz/law/research/publications/about-nzacl/publications/special-issues/hors-serie-volume-iv,-2004/Maurice.pdf>
- [11] <https://www.cairn.info/revue-de-l-ofce-2009-4-page-275.htm>
- [12] <https://www.cairn.info/revue-de-l-ofce-2016-3-page-223.htm>
- [13] <https://www.sciencedirect.com/science/article/pii/S0169207003001134?via%3Dhub>
- [14] https://static.googleusercontent.com/media/www.google.com/fr//googleblogs/pdfs/google_predicting_the_present.pdf
- [15] https://mpra.ub.uni-muenchen.de/90205/1/MPRA_paper_90203.pdf
- [16] Bourbonnais R., Terraza M., « Analyse des séries temporelles, applications à l'économie et à la gestion », Dunod, 2ème édition, 2008, pp. 157-166

Bilan

L'IEOM : https://www.ieom.fr/IMG/pdf/ra2010_polynesie.pdf

Données

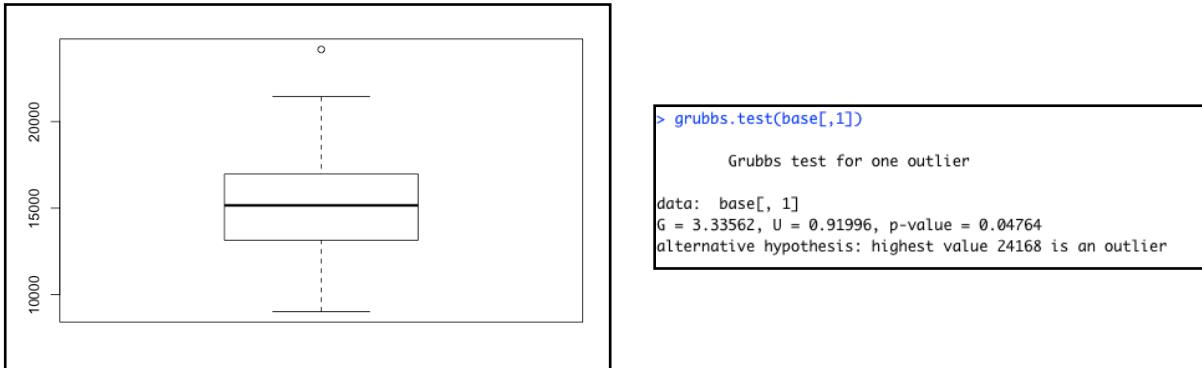
ISPF : <http://www.ispf.pf/Publications.aspx?Collection=Tourisme>

FXTOP : <https://fxtop.com/fr/historique-taux-change.php?A=1&C1=USD&C2=XPF&MA=1&DD1=01&MM1=01&YYYY1=2007&B=1&P=&I=1&DD2=24&MM2=04&YYYY2=2020&btnOK=Chercher>

FXTOP droit : <https://fxtop.com/fr/legal.php#copyright>

Annexe

Figure 3 : Boîte à moustache de la série brute et test de Grubbs



Source : données de l'ISPF - logiciel R

Figure 8 : Résultat points atypiques avec le modèle RegARIMA

```
RegARIMA
y = regression model + arima (0, 1, 1, 0, 1, 1)
Log-transformation: yes
Coefficients:
Estimate Std. Error
Theta(1) -0.4029 0.084
BTheta(1) -0.5784 0.091

Estimate Std. Error
Mean 0.002096 0.002
Monday 0.009761 0.009
Tuesday -0.016402 0.009
Wednesday 0.017016 0.009
Thursday -0.018592 0.009
Friday 0.005391 0.009
Saturday 0.018507 0.009
LS (7-2010) 0.209849 0.051

Residual standard error: 0.05941 on 117 degrees of freedom
Log likelihood = 177.2, aic = 2124 aicc = 2126, bic(corrected for length) = -5.267
```

Source : Library Jdemetra, logiciel R

Figure 10 : CVS SEAT - TRAMO SEAT entre 2008 - 2011

Désaisonnalisation de la série brute avec X13-TS



Source : Library Jdemetra, logiciel R

Figure 12 : Résumé X13 et TRAMO-SEAT

```
RegARIMA
y = regression model + arima (0, 1, 1, 0, 1, 1)
Log-transformation: yes
Coefficients:
             Estimate Std. Error
Theta(1)    -0.4029    0.084
BTheta(1)   -0.5784    0.091

             Estimate Std. Error
Mean        0.002096   0.002
Monday      0.009761   0.009
Tuesday     -0.016402   0.009
Wednesday   0.017016   0.009
Thursday    -0.018592   0.009
Friday      0.005391   0.009
Saturday    0.018507   0.009
LS (7-2010) 0.209849   0.051

Residual standard error: 0.05941 on 117 degrees of freedom
Log likelihood = 177.2, aic = 2124 aicc = 2126, bic(corrected for length) = -5.267
```

Decomposition

Monitoring and Quality Assessment Statistics:

M stats

M(1) 0.222

M(2) 0.165

M(3) 0.634

M(4) 0.286

M(5) 0.664

M(6) 0.119

M(7) 0.272

M(8) 0.444

M(9) 0.284

M(10) 0.412

M(11) 0.337

Q 0.342

Q-M2 0.363

Final filters:

Seasonal filter: 3x5

Trend filter: 13 terms Henderson moving average

Diagnostics

Relative contribution of the components to the stationary portion of the variance in the original series, after the removal of the long term trend

Trend computed by Hodrick-Prescott filter (cycle length = 8.0 years)

Component

Cycle 7.306

Seasonal 57.367

Irregular 3.543

TD & Hol. 0.585

Others 30.513

Total 99.314

Combined test in the entire series

Non parametric tests for stable seasonality

	P.value
Kruskall-Wallis test	0.000
Test for the presence of seasonality assuming stability	0.000
Evolutative seasonality test	0.007

Identifiable seasonality present

Residual seasonality tests

P.value

qs test on sa 1.000

qs test on i 1.000

f-test on sa (seasonal dummies) 0.933

f-test on i (seasonal dummies) 0.949

Residual seasonality (entire series) 0.984

Residual seasonality (last 3 years) 0.665

f-test on sa (td) 0.963

f-test on i (td) 1.000

RegARIMA

y = regression model + arima (0, 1, 1, 0, 1, 1)

Log-transformation: yes

Coefficients:

	Estimate	Std. Error
Theta(1)	-0.4194	0.081
BTheta(1)	-0.6110	0.083

	Estimate	Std. Error
LS (7-2010)	0.2061	0.053

Residual standard error: 0.06354 on 124 degrees of freedom

Log likelihood = 168.2, aic = 2128 aicc = 2128, bic(corrected for length) = -5.398

Decomposition**Model**

D : 1 - B - B¹² + B¹³

MA : 1 - 0.419383 B - 0.610983 B¹² + 0.256236 B¹³

SA

D : 1 - 2.000000 B + B²

MA : 1 - 1.387279 B + 0.410307 B²

Innovation variance: 0.6681277

Trend

D : 1 - 2.000000 B + B²

MA : 1 + 0.040171 B - 0.959829 B²

Innovation variance: 0.0548858

Seasonal

D : 1 + B + B² + B³ + B⁴ + B⁵ + B⁶ + B⁷ + B⁸ + B⁹ + B¹⁰ + B¹¹

MA : 1 + 1.390292 B + 1.447151 B² + 1.365349 B³ + 1.165815 B⁴ + 0.919899 B⁵ + 0.656910 B⁶ + 0.398886 B⁷ + 0.182194 B⁸ - 0.019651 B⁹ - 0.151679 B¹⁰ - 0.434299 B¹¹

Innovation variance: 0.04121986

Irregular

Innovation variance: 0.3268184

Diagnostics

Relative contribution of the components to the stationary portion of the variance in the original series, after the removal of the long term trend

Trend computed by Hodrick-Prescott filter (cycle length = 8.0 years)

Component

Cycle 7.576

Seasonal 57.059

Irregular 2.472

TD & Hol. 0.000

Others 30.246

Total 97.353

Combined test in the entire series

Non parametric tests for stable seasonality

	P.value
Kruskall-Wallis test	0.000
Test for the presence of seasonality assuming stability	0.000
Evolutive seasonality test	0.058

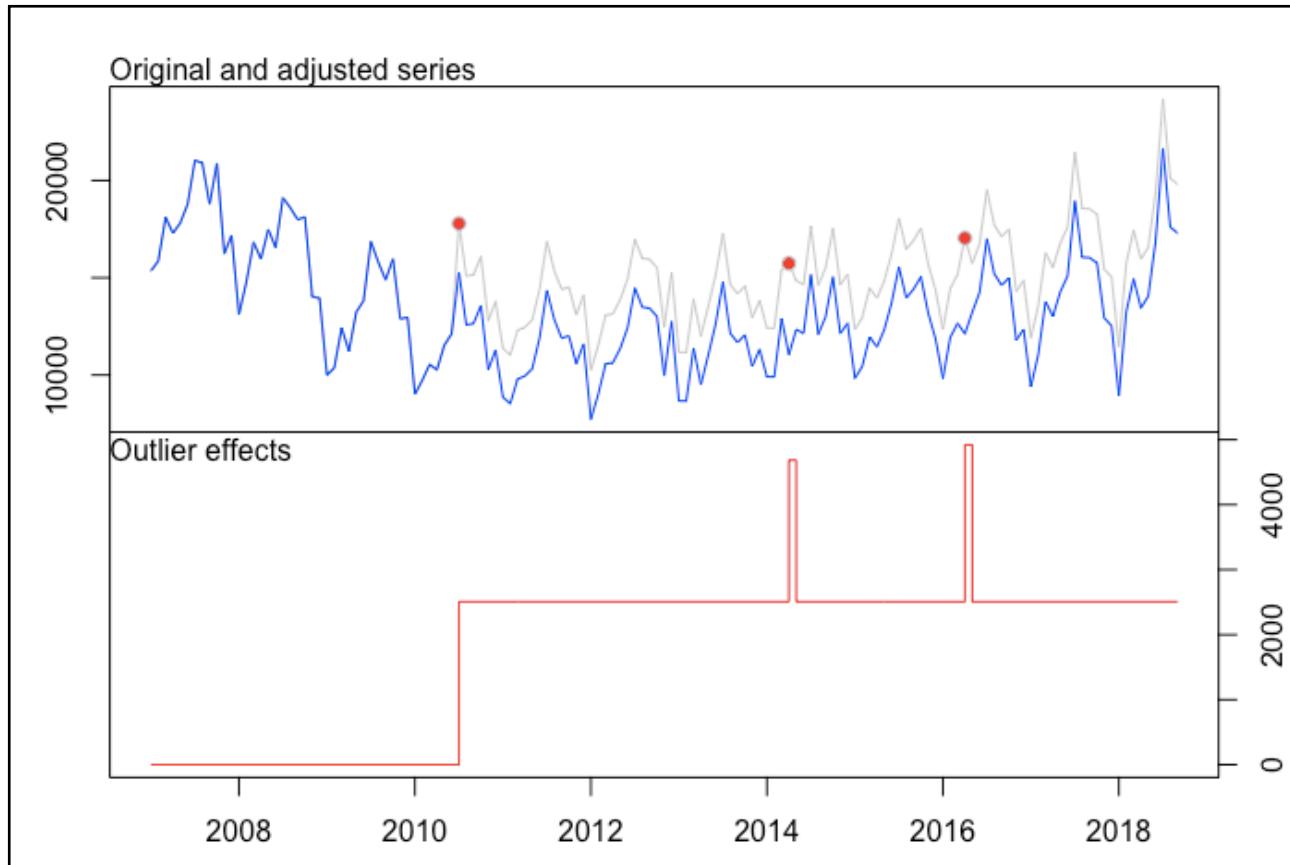
Identifiable seasonality present

Residual seasonality tests

	P.value
qs test on sa	1.000
qs test on i	1.000
f-test on sa (seasonal dummies)	1.000
f-test on i (seasonal dummies)	1.000
Residual seasonality (entire series)	1.000
Residual seasonality (last 3 years)	0.812
f-test on sa (td)	0.001
f-test on i (td)	0.030

Additional output variables

Figure 13 : Détection de outliers selon Chen et Liu



```
Series: myseries
Regression with ARIMA(0,1,1)(0,1,1)[12] errors
```

Coefficients:

	ma1	sma1	LS43	A088	A0112
	-0.4814	-0.4296	2506.9747	2177.660	2414.0588
s.e.	0.0972	0.1100	661.4544	659.938	648.4261

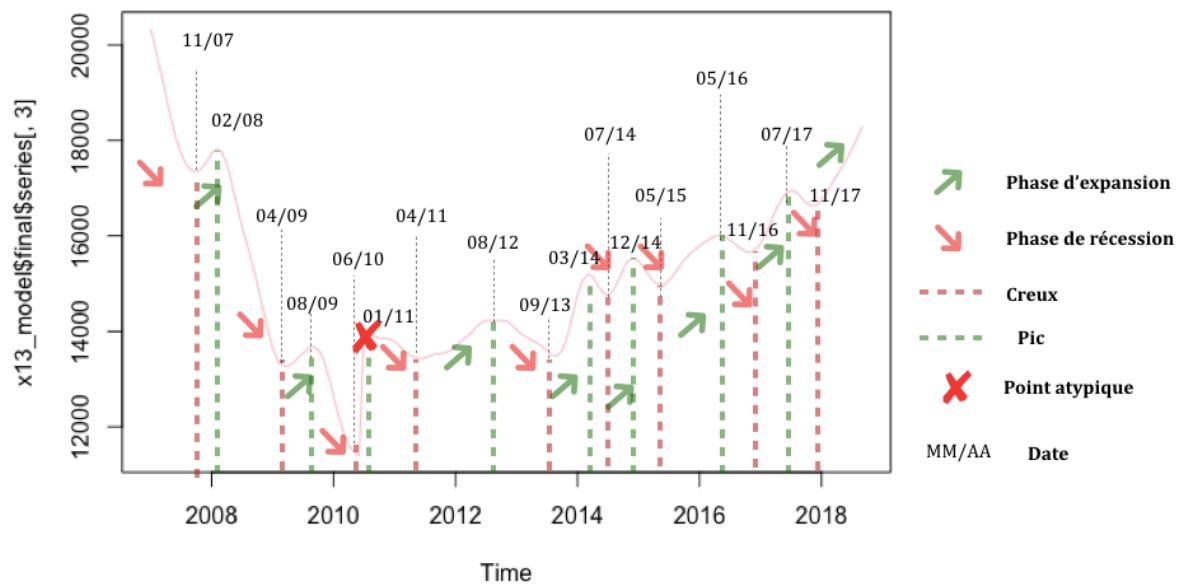
sigma^2 estimated as 804648: log likelihood=-1050.71
AIC=2113.42 AICc=2114.11 BIC=2130.53

Outliers:

	type	ind	time	coefhat	tstat
1	LS	43	2010:07	2507	3.790
2	A0	88	2014:04	2178	3.300
3	A0	112	2016:04	2414	3.723

Source : Library tso, logiciel R

Figure 14 : Tendance de la série avec modèle X13 et identification des phases



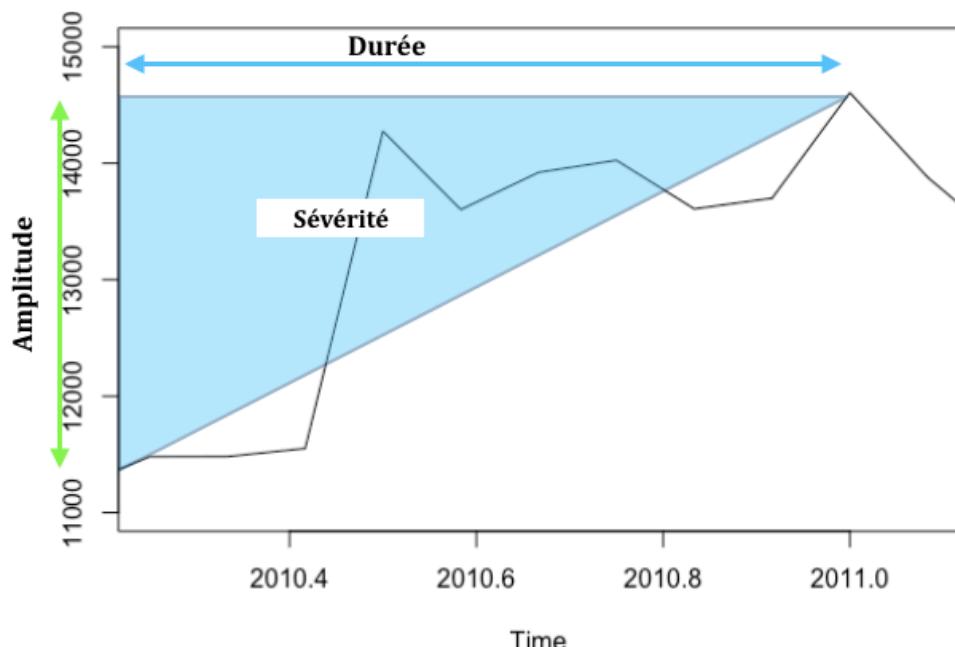
Source : Dossier technique de prévision et conjoncture

Figure 16: Calcul des durée et amplitude des séries

X13 - CVS	<table border="1"> <thead> <tr> <th>Phase</th> <th>]Start ;End]</th> <th>Duration</th> <th>LevStart</th> <th>LevEnd</th> <th>Amplitude</th> </tr> </thead> <tbody> <tr> <td>1 Recessions</td> <td><NA> 2010M3</td> <td>NA</td> <td>NA</td> <td>11207</td> <td>NA</td> </tr> <tr> <td>2 Expansion</td> <td>2010M3 2011M1</td> <td>10</td> <td>11207</td> <td>14606</td> <td>3398.4</td> </tr> <tr> <td>3 Recessions</td> <td>2011M1 2011M10</td> <td>9</td> <td>14606</td> <td>12804</td> <td>1801.4</td> </tr> <tr> <td>4 Expansion</td> <td>2011M10 2016M4</td> <td>54</td> <td>12804</td> <td>17880</td> <td>5075.9</td> </tr> <tr> <td>5 Recessions</td> <td>2016M4 2018M1</td> <td>21</td> <td>17880</td> <td>14983</td> <td>2897.0</td> </tr> <tr> <td>6 Expansion</td> <td>2018M1 <NA></td> <td>NA</td> <td>14983</td> <td>NA</td> <td>NA</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Amplitude</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>Exp=]T;P]</td> <td>4237.2</td> <td>32</td> </tr> <tr> <td>Rec=]P;T]</td> <td>2349.2</td> <td>15</td> </tr> </tbody> </table>	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude	1 Recessions	<NA> 2010M3	NA	NA	11207	NA	2 Expansion	2010M3 2011M1	10	11207	14606	3398.4	3 Recessions	2011M1 2011M10	9	14606	12804	1801.4	4 Expansion	2011M10 2016M4	54	12804	17880	5075.9	5 Recessions	2016M4 2018M1	21	17880	14983	2897.0	6 Expansion	2018M1 <NA>	NA	14983	NA	NA		Amplitude	Duration	Exp=]T;P]	4237.2	32	Rec=]P;T]	2349.2	15	<table border="1"> <thead> <tr> <th>Peaks</th> <th>Troughs</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>1 <NA></td> <td>2010M3</td> <td><NA></td> </tr> <tr> <td>2 2011M1</td> <td>2011M10</td> <td>9</td> </tr> <tr> <td>3 2016M4</td> <td>2018M1</td> <td>21</td> </tr> </tbody> </table>	Peaks	Troughs	Duration	1 <NA>	2010M3	<NA>	2 2011M1	2011M10	9	3 2016M4	2018M1	21
	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude																																																											
1 Recessions	<NA> 2010M3	NA	NA	11207	NA																																																												
2 Expansion	2010M3 2011M1	10	11207	14606	3398.4																																																												
3 Recessions	2011M1 2011M10	9	14606	12804	1801.4																																																												
4 Expansion	2011M10 2016M4	54	12804	17880	5075.9																																																												
5 Recessions	2016M4 2018M1	21	17880	14983	2897.0																																																												
6 Expansion	2018M1 <NA>	NA	14983	NA	NA																																																												
	Amplitude	Duration																																																															
Exp=]T;P]	4237.2	32																																																															
Rec=]P;T]	2349.2	15																																																															
Peaks	Troughs	Duration																																																															
1 <NA>	2010M3	<NA>																																																															
2 2011M1	2011M10	9																																																															
3 2016M4	2018M1	21																																																															
X13 - TREND	<table border="1"> <thead> <tr> <th>Phase</th> <th>]Start ;End]</th> <th>Duration</th> <th>LevStart</th> <th>LevEnd</th> <th>Amplitude</th> </tr> </thead> <tbody> <tr> <td>1 Recessions</td> <td><NA> 2010M6</td> <td>NA</td> <td>NA</td> <td>11404</td> <td>NA</td> </tr> <tr> <td>2 Expansion</td> <td>2010M6 2012M8</td> <td>26</td> <td>11404</td> <td>14230</td> <td>2825.5</td> </tr> <tr> <td>3 Recessions</td> <td>2012M8 2013M9</td> <td>13</td> <td>14230</td> <td>13478</td> <td>751.1</td> </tr> <tr> <td>4 Expansion</td> <td>2013M9 2016M5</td> <td>32</td> <td>13478</td> <td>16010</td> <td>2531.5</td> </tr> <tr> <td>5 Recessions</td> <td>2016M5 2016M11</td> <td>6</td> <td>16010</td> <td>15661</td> <td>349.1</td> </tr> <tr> <td>6 Expansion</td> <td>2016M11 <NA></td> <td>NA</td> <td>15661</td> <td>NA</td> <td>NA</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Amplitude</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>Exp=]T;P]</td> <td>2678.5</td> <td>29.0</td> </tr> <tr> <td>Rec=]P;T]</td> <td>550.1</td> <td>9.5</td> </tr> </tbody> </table>	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude	1 Recessions	<NA> 2010M6	NA	NA	11404	NA	2 Expansion	2010M6 2012M8	26	11404	14230	2825.5	3 Recessions	2012M8 2013M9	13	14230	13478	751.1	4 Expansion	2013M9 2016M5	32	13478	16010	2531.5	5 Recessions	2016M5 2016M11	6	16010	15661	349.1	6 Expansion	2016M11 <NA>	NA	15661	NA	NA		Amplitude	Duration	Exp=]T;P]	2678.5	29.0	Rec=]P;T]	550.1	9.5	<table border="1"> <thead> <tr> <th>Peaks</th> <th>Troughs</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>1 <NA></td> <td>2010M6</td> <td><NA></td> </tr> <tr> <td>2 2012M8</td> <td>2013M9</td> <td>13</td> </tr> <tr> <td>3 2016M5</td> <td>2016M11</td> <td>6</td> </tr> </tbody> </table>	Peaks	Troughs	Duration	1 <NA>	2010M6	<NA>	2 2012M8	2013M9	13	3 2016M5	2016M11	6
	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude																																																											
1 Recessions	<NA> 2010M6	NA	NA	11404	NA																																																												
2 Expansion	2010M6 2012M8	26	11404	14230	2825.5																																																												
3 Recessions	2012M8 2013M9	13	14230	13478	751.1																																																												
4 Expansion	2013M9 2016M5	32	13478	16010	2531.5																																																												
5 Recessions	2016M5 2016M11	6	16010	15661	349.1																																																												
6 Expansion	2016M11 <NA>	NA	15661	NA	NA																																																												
	Amplitude	Duration																																																															
Exp=]T;P]	2678.5	29.0																																																															
Rec=]P;T]	550.1	9.5																																																															
Peaks	Troughs	Duration																																																															
1 <NA>	2010M6	<NA>																																																															
2 2012M8	2013M9	13																																																															
3 2016M5	2016M11	6																																																															
TS - CVS	<table border="1"> <thead> <tr> <th>Phase</th> <th>]Start ;End]</th> <th>Duration</th> <th>LevStart</th> <th>LevEnd</th> <th>Amplitude</th> </tr> </thead> <tbody> <tr> <td>1 Recessions</td> <td><NA> 2010M3</td> <td>NA</td> <td>NA</td> <td>11350</td> <td>NA</td> </tr> <tr> <td>2 Expansion</td> <td>2010M3 2011M1</td> <td>10</td> <td>11350</td> <td>14719</td> <td>3369.2</td> </tr> <tr> <td>3 Recessions</td> <td>2011M1 2011M10</td> <td>9</td> <td>14719</td> <td>12869</td> <td>1849.7</td> </tr> <tr> <td>4 Expansion</td> <td>2011M10 2016M4</td> <td>54</td> <td>12869</td> <td>17652</td> <td>4782.4</td> </tr> <tr> <td>5 Recessions</td> <td>2016M4 2016M11</td> <td>7</td> <td>17652</td> <td>15338</td> <td>2314.2</td> </tr> <tr> <td>6 Expansion</td> <td>2016M11 <NA></td> <td>NA</td> <td>15338</td> <td>NA</td> <td>NA</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Amplitude</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>Exp=]T;P]</td> <td>4075.8</td> <td>32</td> </tr> <tr> <td>Rec=]P;T]</td> <td>2081.9</td> <td>8</td> </tr> </tbody> </table>	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude	1 Recessions	<NA> 2010M3	NA	NA	11350	NA	2 Expansion	2010M3 2011M1	10	11350	14719	3369.2	3 Recessions	2011M1 2011M10	9	14719	12869	1849.7	4 Expansion	2011M10 2016M4	54	12869	17652	4782.4	5 Recessions	2016M4 2016M11	7	17652	15338	2314.2	6 Expansion	2016M11 <NA>	NA	15338	NA	NA		Amplitude	Duration	Exp=]T;P]	4075.8	32	Rec=]P;T]	2081.9	8	<table border="1"> <thead> <tr> <th>Peaks</th> <th>Troughs</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>1 <NA></td> <td>2010M3</td> <td><NA></td> </tr> <tr> <td>2 2011M1</td> <td>2011M10</td> <td>9</td> </tr> <tr> <td>3 2016M4</td> <td>2016M11</td> <td>7</td> </tr> </tbody> </table>	Peaks	Troughs	Duration	1 <NA>	2010M3	<NA>	2 2011M1	2011M10	9	3 2016M4	2016M11	7
	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude																																																											
1 Recessions	<NA> 2010M3	NA	NA	11350	NA																																																												
2 Expansion	2010M3 2011M1	10	11350	14719	3369.2																																																												
3 Recessions	2011M1 2011M10	9	14719	12869	1849.7																																																												
4 Expansion	2011M10 2016M4	54	12869	17652	4782.4																																																												
5 Recessions	2016M4 2016M11	7	17652	15338	2314.2																																																												
6 Expansion	2016M11 <NA>	NA	15338	NA	NA																																																												
	Amplitude	Duration																																																															
Exp=]T;P]	4075.8	32																																																															
Rec=]P;T]	2081.9	8																																																															
Peaks	Troughs	Duration																																																															
1 <NA>	2010M3	<NA>																																																															
2 2011M1	2011M10	9																																																															
3 2016M4	2016M11	7																																																															
TS - TREND	<table border="1"> <thead> <tr> <th>Phase</th> <th>]Start ;End]</th> <th>Duration</th> <th>LevStart</th> <th>LevEnd</th> <th>Amplitude</th> </tr> </thead> <tbody> <tr> <td>1 Recessions</td> <td><NA> 2010M6</td> <td>NA</td> <td>NA</td> <td>11556</td> <td>NA</td> </tr> <tr> <td>2 Expansion</td> <td>2010M6 2012M9</td> <td>27</td> <td>11556</td> <td>14283</td> <td>2727.4</td> </tr> <tr> <td>3 Recessions</td> <td>2012M9 2013M9</td> <td>12</td> <td>14283</td> <td>13558</td> <td>724.9</td> </tr> <tr> <td>4 Expansion</td> <td>2013M9 2016M4</td> <td>31</td> <td>13558</td> <td>16381</td> <td>2823.0</td> </tr> <tr> <td>5 Recessions</td> <td>2016M4 2016M11</td> <td>7</td> <td>16381</td> <td>15738</td> <td>643.5</td> </tr> <tr> <td>6 Expansion</td> <td>2016M11 <NA></td> <td>NA</td> <td>15738</td> <td>NA</td> <td>NA</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Amplitude</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>Exp=]T;P]</td> <td>2775.2</td> <td>29.0</td> </tr> <tr> <td>Rec=]P;T]</td> <td>684.2</td> <td>9.5</td> </tr> </tbody> </table>	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude	1 Recessions	<NA> 2010M6	NA	NA	11556	NA	2 Expansion	2010M6 2012M9	27	11556	14283	2727.4	3 Recessions	2012M9 2013M9	12	14283	13558	724.9	4 Expansion	2013M9 2016M4	31	13558	16381	2823.0	5 Recessions	2016M4 2016M11	7	16381	15738	643.5	6 Expansion	2016M11 <NA>	NA	15738	NA	NA		Amplitude	Duration	Exp=]T;P]	2775.2	29.0	Rec=]P;T]	684.2	9.5	<table border="1"> <thead> <tr> <th>Peaks</th> <th>Troughs</th> <th>Duration</th> </tr> </thead> <tbody> <tr> <td>1 <NA></td> <td>2010M6</td> <td><NA></td> </tr> <tr> <td>2 2012M9</td> <td>2013M9</td> <td>12</td> </tr> <tr> <td>3 2016M4</td> <td>2016M11</td> <td>7</td> </tr> </tbody> </table>	Peaks	Troughs	Duration	1 <NA>	2010M6	<NA>	2 2012M9	2013M9	12	3 2016M4	2016M11	7
	Phase]Start ;End]	Duration	LevStart	LevEnd	Amplitude																																																											
1 Recessions	<NA> 2010M6	NA	NA	11556	NA																																																												
2 Expansion	2010M6 2012M9	27	11556	14283	2727.4																																																												
3 Recessions	2012M9 2013M9	12	14283	13558	724.9																																																												
4 Expansion	2013M9 2016M4	31	13558	16381	2823.0																																																												
5 Recessions	2016M4 2016M11	7	16381	15738	643.5																																																												
6 Expansion	2016M11 <NA>	NA	15738	NA	NA																																																												
	Amplitude	Duration																																																															
Exp=]T;P]	2775.2	29.0																																																															
Rec=]P;T]	684.2	9.5																																																															
Peaks	Troughs	Duration																																																															
1 <NA>	2010M6	<NA>																																																															
2 2012M9	2013M9	12																																																															
3 2016M4	2016M11	7																																																															

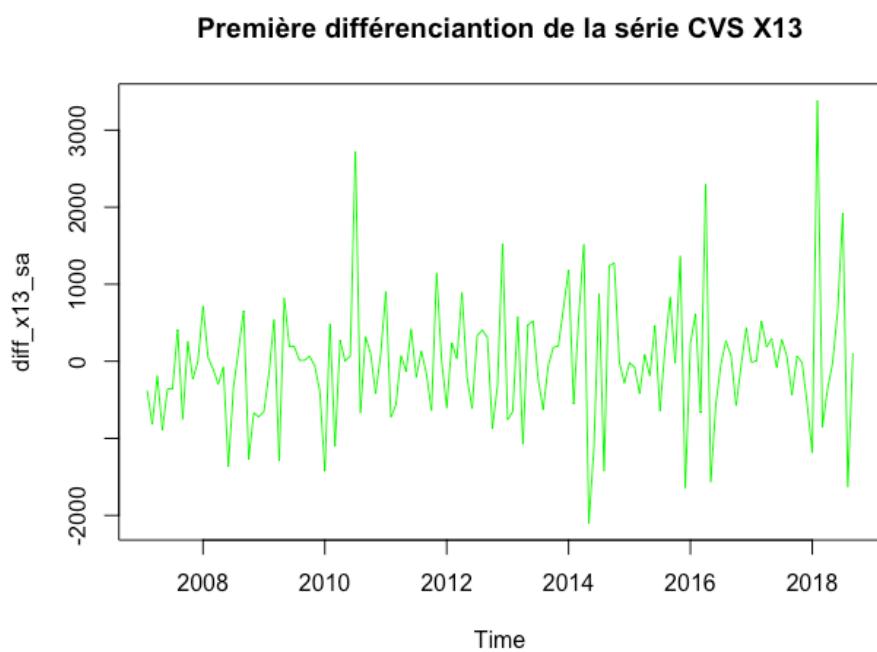
Source : Dossier technique de prévision et conjoncture

Figure 17: Illustration des trois critères : durée, amplitude et



Remarque : Dans cette figure, nous avons représenté les trois critères sur la deuxième phase d'expansion du modèle X13 de la série CVS pour avoir une idée plus clair de la sévérité. En effet, la sévérité représente l'aire d'un triangle rectangle, comme nous sommes dans une phase d'expansion alors la sévérité dans ce cas montre un gain de l'économie du nombre de tourisme.

Figure 18: Différenciation de la série CVS X13



Source: Dossier technique de prévision et conjoncture

Figure 20.B: Différenciation de la série CVS X13

```
> show(datHP)
  Peaks Troughs Duration
1 2008M1 2008M11      10
2 2010M7 2014M6       47
3 2017M3    <NA>     <NA>
> summary(datHP)
  Phase ]Start ;End] Duration LevStart LevEnd Amplitude
1 Expansion   <NA> 2008M1      NA      NA    83      NA
2 Recessions 2008M1 2008M11     10      83   -490    573.3
3 Expansion  2008M11 2010M7     20    -490    368    858.0
4 Recessions 2010M7 2014M6     47      368   -188    555.8
5 Expansion  2014M6 2017M3     33    -188    194    381.8
6 Recessions 2017M3    <NA>      NA    194      NA      NA

  Amplitude Duration
Exp=]T;P]      619.9    26.5
Rec=]P;T]      564.6    28.5
```

```
> show(datHP)
  Peaks Troughs Duration
1 2014M2 2015M4      14
> summary(datHP)
  Phase ]Start ;End] Duration LevStart LevEnd Amplitude
1 Expansion   <NA> 2014M2      NA      NA    48      NA
2 Recessions 2014M2 2015M4     14      48     46      2
3 Expansion  2015M4    <NA>      NA      46      NA      NA

  Amplitude Duration
Exp=]T;P]        NaN      NaN
Rec=]P;T]        2       14
```

Source: Dossier technique de prévision et

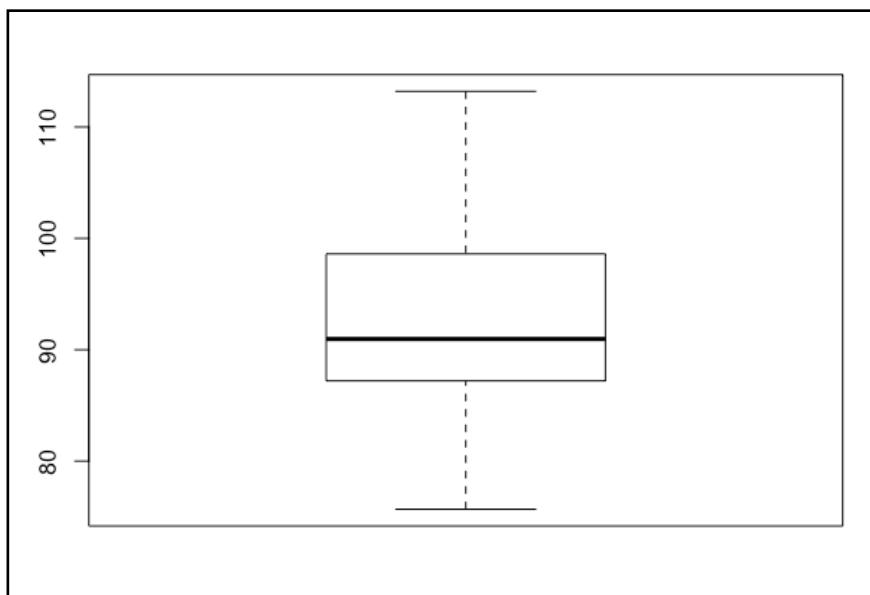
Tableau 3: Prévision des modèles

	ARIMA	AUTO	ETS	NNETAR	TBATS	BATS	STLM	HW	NAIVE	X13
oct-17	19307,62	17279,83	18656,02	17181,22	19134	18903,59	19223,92	19297,78	17499	18613,06
nov-17	16259,62	15744,62	18656,02	14871,79	16302,14	16183,53	16567,69	16019,79	14295	16601,33
déc-17	16743,57	14254,88	18656,02	14225,28	16805,27	16573,92	17235,64	16455,96	14858	17111,14
janv-18	13534,87	13151,43	18656,02	10856,38	13865,71	13186,21	14257,71	12988,65	11910	14757,26
févr-18	14765,38	12634,21	18656,02	13882,18	14464,67	14432,45	15095,63	13818,43	13564	15475,24
mars-18	16984,5	12738,77	18656,02	15086,75	16511,53	16243,46	16879,16	15800,1	16281	17871,32
avr-18	16842,18	13348,07	18656,02	15729,9	15962,13	15991,71	16113,22	15576,18	15523	17637,99
mai-18	17388,44	14245,88	18656,02	16479,4	16834,59	16513,09	17245,85	16346,15	16782	17930,22
juin-18	18416,78	15179,91	18656,02	17557,33	17420,05	17541,06	18188,18	17347,31	17599	19740,58
juil-18	21763,14	15926,8	18656,02	21516,7	20442,66	20305,69	20765,97	20289,34	21448	21786,43
août-18	19297,91	16339,82	18656,02	18780,61	18500,76	18485,79	19021,96	18171,97	18563	20246,62
sept-18	19171,1	16371,66	18656,02	18272,65	18207,54	18417,17	18721,04	17996,43	18539	20058,33

Tableau 4: Mesures d'erreur de prévision

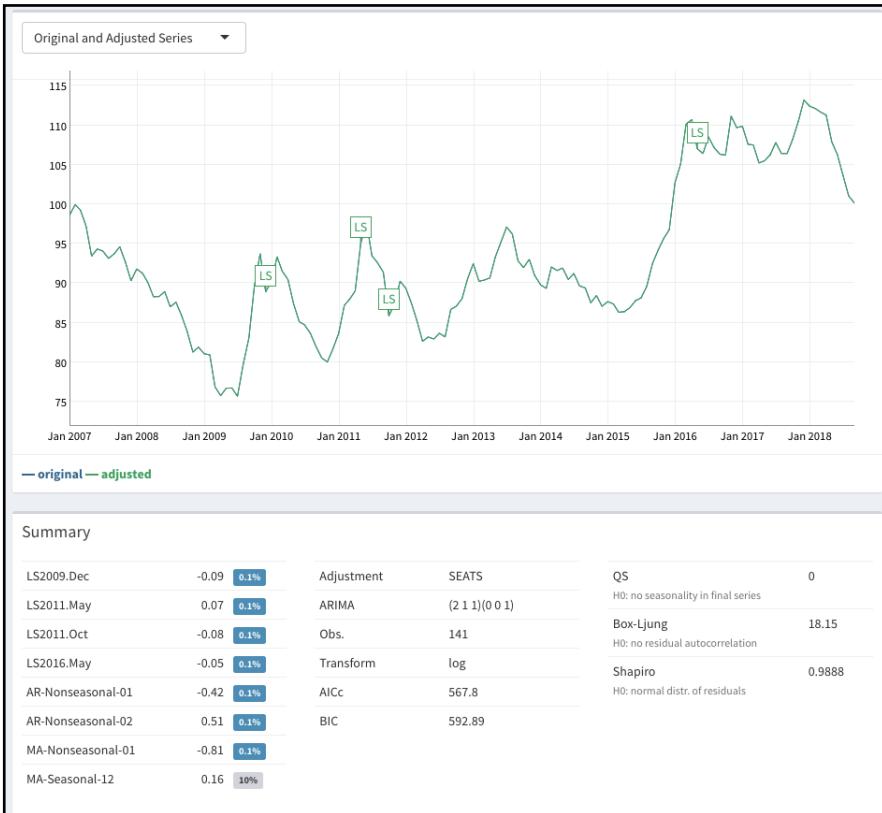
Modele	ME	MAE	RMSE
SARIMA	91,34	1134,54	1271,65
AUTO-SARIMA	-2680,26	3011,94	3652,77
ETS	1207,77	2680,26	3308,45
nnetar	-1244,9	1244,9	1475,02
tbats	-410,66	1441,14	1729,01
Bats	-550,11	1329,6	1641,91
stlm	-5,25	1325,76	1619,16
HoltWinters	-772,58	1531,77	1793,07
snaive	-1043,17	1155,83	1386,95
X13	704,21	1146,43	1514,34

Figure 28 : Distribution de la série et description statistique



Remarque : En réalisant une boîte à moustache avec le logiciel R, graphique ne montre aucune valeur atypique. Nous tentons de tester la série pour trouver des valeurs atypiques par le test de Grubbs. Ainsi le test identifie une valeur atypique pour le mois de décembre 2017.

Figure 29 : détection des outliers par TRAMO



Source : logiciel Jdemetra dans R

Figure 30.A : Calcul de FAC et FACP de I(0) et I(1) avec Q_stat

Fonction d'auto-corrélation pour Txch
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error $1/T^{0.5}$

RETARD	ACF	PACF	Q	[p. crit.]
1	0.9707	***	0.9707	***
2	0.9260	***	-0.2816	***
3	0.8774	***	-0.0231	
4	0.8238	***	-0.1019	
5	0.7651	***	-0.0834	
6	0.7078	***	0.0319	
7	0.6503	***	-0.0545	
8	0.6012	***	0.1390	*
9	0.5573	***	-0.0143	
10	0.5190	***	0.0417	
11	0.4859	***	0.0172	
12	0.4617	***	0.0731	
13	0.4431	***	0.0204	
14	0.4311	***	0.0493	
15	0.4217	***	-0.0123	
16	0.4130	***	-0.0242	
17	0.3971	***	-0.1510	*
18	0.3795	***	0.0238	
19	0.3562	***	-0.0981	
20	0.3243	***	-0.1000	
21	0.2877	***	0.0037	
22	0.2540	***	0.0733	
23	0.2160	**	-0.0849	
24	0.1834	**	0.1292	
			1154.0643	[0.000]

Fonction d'auto-corrélation pour d_Txch
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error $1/T^{0.5}$

RETARD	ACF	PACF	Q	[p. crit.]
1	0.2735	***	0.2735	***
2	0.0517		-0.0250	
3	0.0841		0.0827	
4	0.1107		0.0728	
5	-0.0703		-0.1317	
6	0.0016		0.0573	
7	-0.1568	*	-0.2051	**
8	-0.1094		-0.0074	
9	-0.1229		-0.0846	
10	-0.0563		0.0001	
11	-0.1231		-0.0655	
12	-0.1016		-0.0740	
13	-0.1230		-0.0618	
14	-0.0541		-0.0417	
15	0.0169		0.0613	
16	0.1483	*	0.1227	
17	0.0272		-0.0482	
18	0.1019		0.0912	
19	0.1570	*	0.0599	
20	0.0815		-0.0450	
21	-0.0396		-0.0669	
22	0.0905		0.0685	
23	-0.0586		-0.1067	
24	-0.0017		0.0719	

Source : Logiciel GRETL

Figure 32 : Résumé des modélisations

```
Date: Sun Apr 26 20:18:24 2020  
Dependent var.: training  
Method: Ordinary Least Squares (OLS)  
Variance-Covariance: Ordinary  
No. of observations (mean eq.): 127  
Sample: 2007(3) to 2017(9)
```

Modèle AR(1)

Mean equation:

	coef	std.error	t-stat	p-value
mconst	0.035843	0.188320	0.1903	0.849358
ar1	0.245106	0.086585	2.8308	0.005412

Diagnostics and fit:

	Chi-sq	df	p-value
Ljung-Box AR(2)	0.34756	2	0.84048
Ljung-Box ARCH(1)	0.53719	1	0.46360
Jarque-Bera	2.90407	2	0.23409

	SE of regression	R-squared	Log-lik.(n=127)
	2.12140	0.06025	-274.71892

```
> mod_ar=arima(training,order=c(1,0,0))  
> mod_ar
```

Call:

Modèle ARIMA(1,1,0)

```
arima(x = training, order = c(1, 0, 0))
```

Coefficients:

	ar1	intercept
s.e.	0.2439	0.0642
	0.0854	0.2448

sigma^2 estimated as 4.407: log likelihood = -276.58, aic = 559.17

```
> mod_arima=arima(training,order=c(1,0,1))  
> mod_arima
```

Call:

Modèle ARIMA(1,1,1)

```
arima(x = training, order = c(1, 0, 1))
```

Coefficients:

	ar1	ma1	intercept
s.e.	0.0473	0.2089	0.0649
	0.3237	0.3157	0.2348

sigma^2 estimated as 4.397: log likelihood = -276.43, aic = 560.87

```
> mod_auto
Series: training
ARIMA(1,0,2) with zero mean
```

Modèle AUTO ARIMA

Coefficients:

	ar1	ma1	ma2
-	-0.9891	1.3652	0.4043
s.e.	0.0157	0.0959	0.0931

σ^2 estimated as 4.198: log likelihood=-272.37

AIC=552.74 AICc=553.07 BIC=564.15

Holt-Winters exponential smoothing with trend and without seasonal component.

Call:

```
HoltWinters(x = training, gamma = F)
```

Smoothing parameters:

Modèle Holt-Winters

alpha: 0.4942074

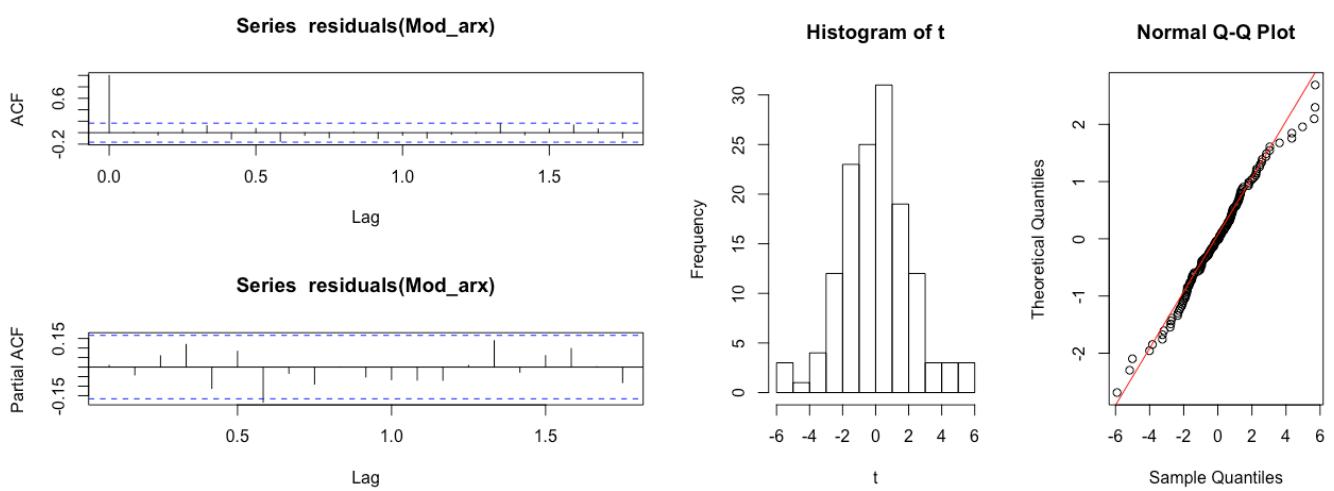
beta : 0.128832

gamma: FALSE

Coefficients:

	[,1]
a	-0.1589283
b	-0.0236468

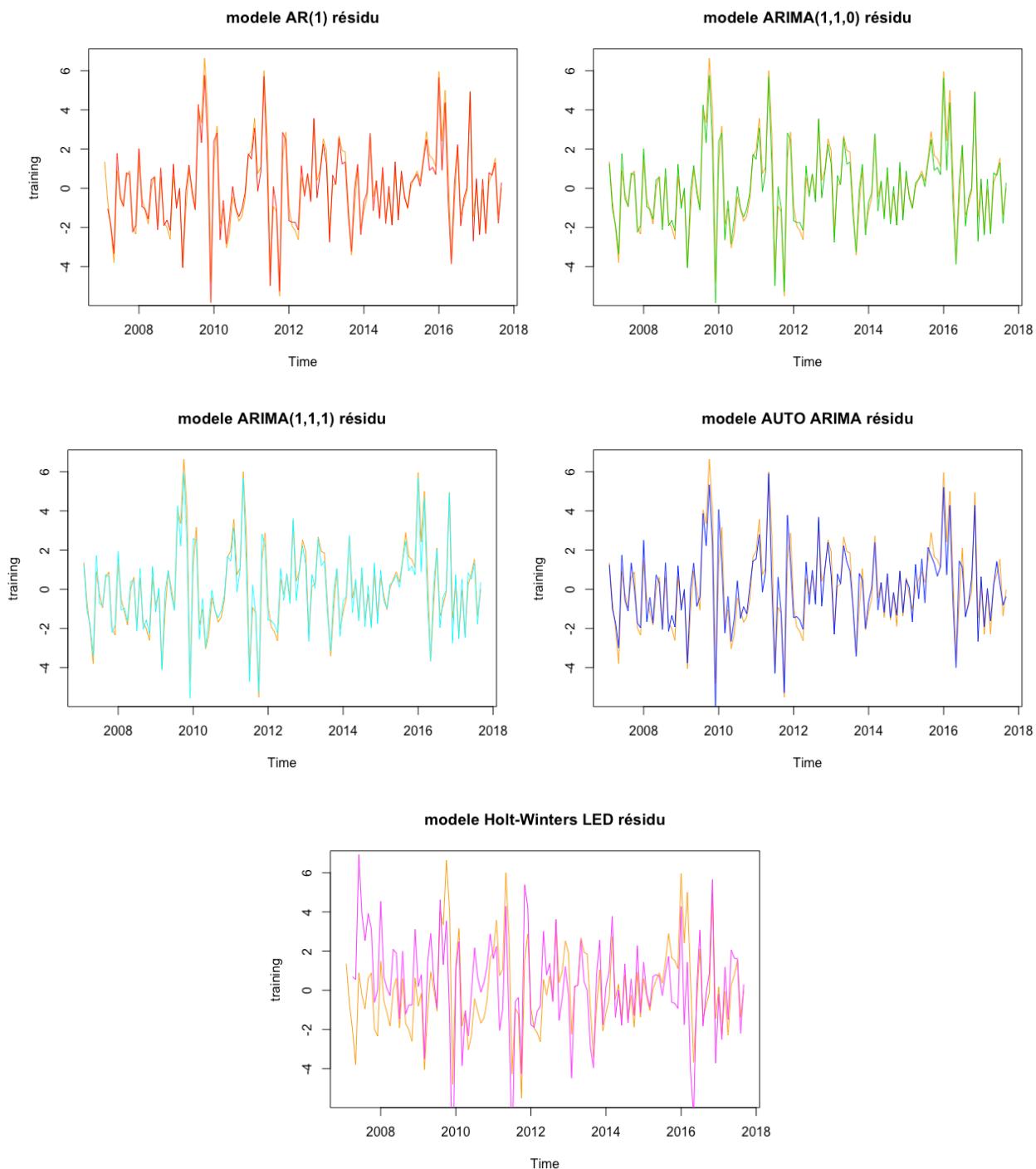
Figure 33 : Etude des résidus



Source : Logiciel sous R

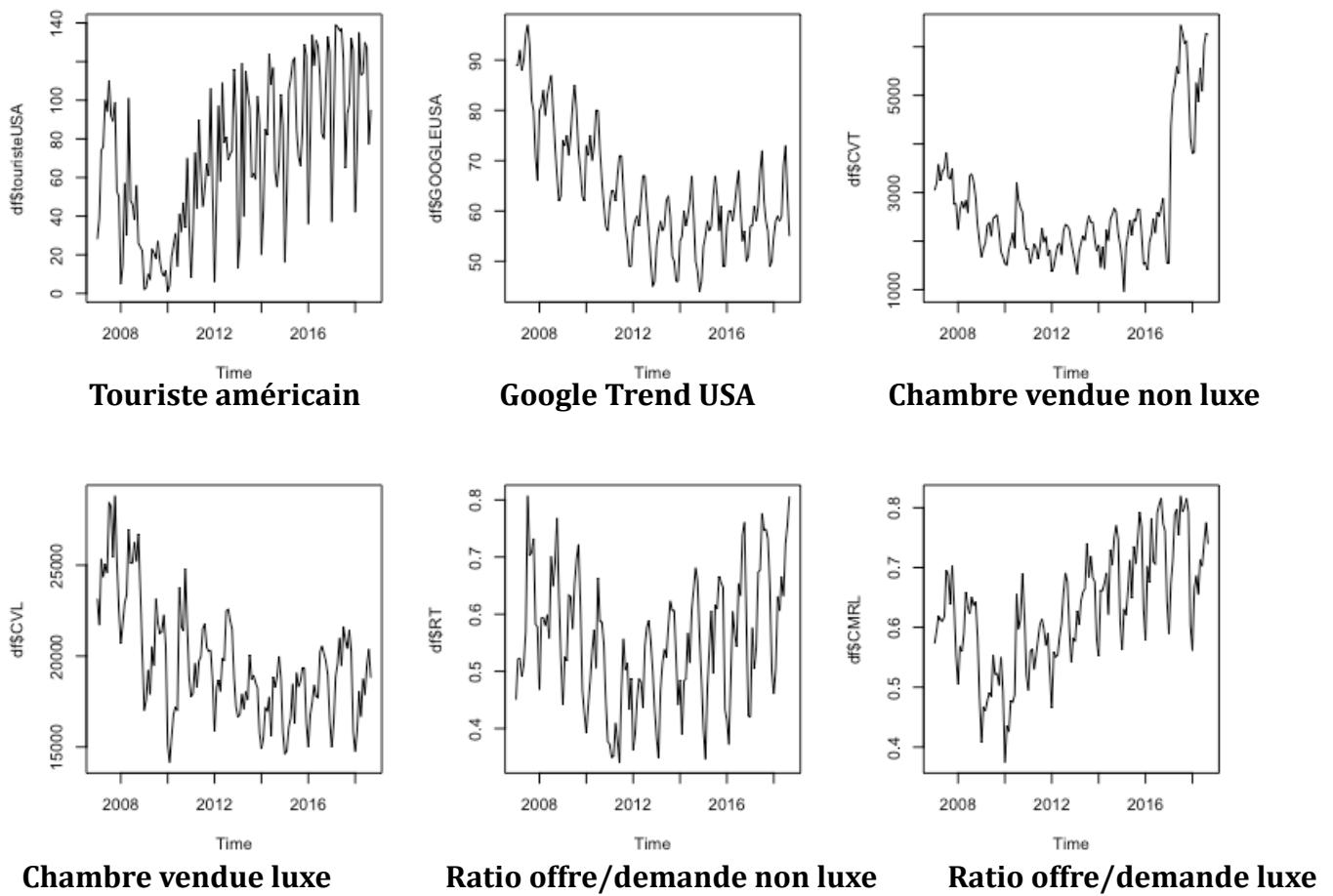
Remarque : Tout les résidus des modèles suivent une loi normale, nous avions vérifier la FAC, FACP, l'histogramme et le Q-Q plot par rapport à la droite de Henry. En regardant la FAC et FACP montrent que le premier retard est significative alors l'ensemble des résidus et indépendant et identiquement distribué (iid). Nous constatons aucune corrélation entre les résidus. Pour confirmer la normalité, nous avons aussi utilisé le test de Jarque-Bera qui annonce que les résidus suivent bien loi normale.

Figure 34 : Résidus des modèles



Source: Dossier Techniques de prévision et conjoncture

Figure 35 : Série temporelle des variables explicatives



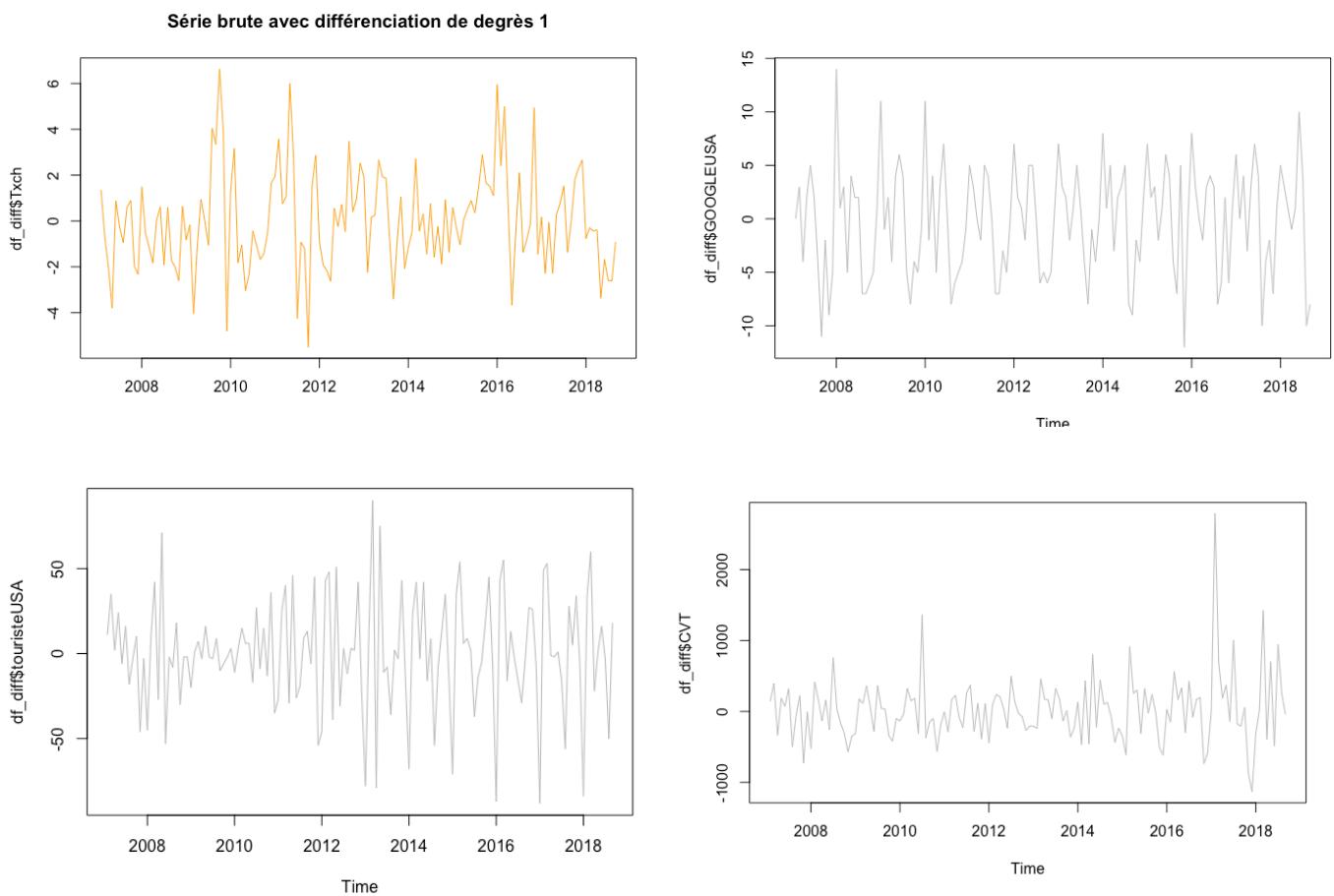
Source : ISPF et google Trend

Remarque : Nous remarquons que le nombre de touristes américains suit les mêmes tendances que tc . En effet, après la grande récession entre 2007-2009, le nombre de touristes diminue. Pour la variable Google Trend USA, le taux de requête diminue entre 2007 et 2014 et stagne entre 2014 et 2018.

Ensuite nous avons la tendance du nombre de chambre vendue non luxue qui diminue faiblement entre 2007 et 2012, puis stagne jusqu'au 2017 et augmente soudainement. Ceci peut s'expliquer par l'arrivée de la nouvelle compagnie qu'on avait vu pour la partie saisonnière.

Cependant, la tendance du nombre de chambre de luxe vendue diminue de 2007 à 2015 et stagne.

Figure 35.B : Série stationnaire des variables explicatives



Source : Dossier Techniques de prévision et conjoncture

Après la vérification de la stationnarité des variables, nous avons présenté les fluctuations des série différées de degré 1 dans la **figure 35.B**.

Figure 37 : Step-wise

```
Step: AIC=216.72
Txch ~ touristeUSA + CMRL + GOOGLEUSA + CVT

      Df Sum of Sq    RSS    AIC
- touristeUSA  1     3.9145 616.82 215.61
- CMRL         1     4.7522 617.65 215.80
- GOOGLEUSA    1     8.7305 621.63 216.70
<none>                  612.90 216.72
- CVT          1    26.6183 639.52 220.67

Step: AIC=215.61
Txch ~ CMRL + GOOGLEUSA + CVT

      Df Sum of Sq    RSS    AIC
- CMRL         1     6.4335 623.25 215.06
- GOOGLEUSA    1     7.8802 624.70 215.39
<none>                  616.82 215.61
- CVT          1    24.2153 641.03 219.00

Step: AIC=215.06
Txch ~ GOOGLEUSA + CVT

      Df Sum of Sq    RSS    AIC
- GOOGLEUSA    1     7.0422 630.29 214.64
<none>                  623.25 215.06
- CVT          1    17.8908 641.14 217.03

Step: AIC=214.64
Txch ~ CVT

      Df Sum of Sq    RSS    AIC
<none>                  630.29 214.64
- CVT          1    15.588 645.88 216.06
```

Source : logiciel R

Figure 38 : modélisation avec variables explicatives

```
Date: Mon Apr 27 13:51:55 2020
Dependent var.: y
Method: Ordinary Least Squares (OLS)
Variance-Covariance: Ordinary
No. of observations (mean eq.): 128
Sample: 2 to 129

Mean equation:

            coef  std.error  t-stat  p-value
mconst      0.07451382  0.18737460  0.3977  0.691555
ar1         0.23813311  0.08659003  2.7501  0.006849
df_diff.CVT -0.00047095  0.00043688 -1.0780  0.283132
df_diff.GOOGLEUSA  0.05272218  0.03754937  1.4041  0.162796

Diagnostics and fit:

            Chi-sq df p-value
Ljung-Box AR(2)  0.24628  2 0.88414
Ljung-Box ARCH(1) 0.38985  1 0.53238

SE of regression   2.11215
R-squared          0.08057
Log-lik.(n=128)   -275.33055
```

Modele 1 : CVT + GOOGLE USA

```
Date: Mon Apr 27 13:53:46 2020
Dependent var.: y
Method: Ordinary Least Squares (OLS)
Variance-Covariance: Ordinary
No. of observations (mean eq.): 128
Sample: 2 to 129

Mean equation:

            coef  std.error  t-stat  p-value
mconst  0.05917546  0.18778132  0.3151  0.753189
ar1     0.23753039  0.08692478  2.7326  0.007195
mxreg1 -0.00039331  0.00043505 -0.9041  0.367705

Diagnostics and fit:

            Chi-sq df p-value
Ljung-Box AR(2)  0.46773  2 0.79147
Ljung-Box ARCH(1) 0.76840  1 0.38071

SE of regression   2.12034
R-squared          0.06595
Log-lik.(n=128)   -276.32600
```

Modele 2 : CVT

Date: Mon Apr 27 13:54:44 2020
 Dependent var.: y
 Method: Ordinary Least Squares (OLS)
 Variance-Covariance: Ordinary
 No. of observations (mean eq.): 128
 Sample: 2 to 129

Mean equation:

	coef	std.error	t-stat	p-value
mconst	0.05834772	0.18795708	0.3104	0.756754
ar1	0.24533117	0.08745832	2.8051	0.005841
df_diff.touristeUSA	0.00515756	0.00587954	0.8772	0.382072
df_diff.CVT	-0.00048157	0.00044692	-1.0775	0.283334

Diagnostics and fit:

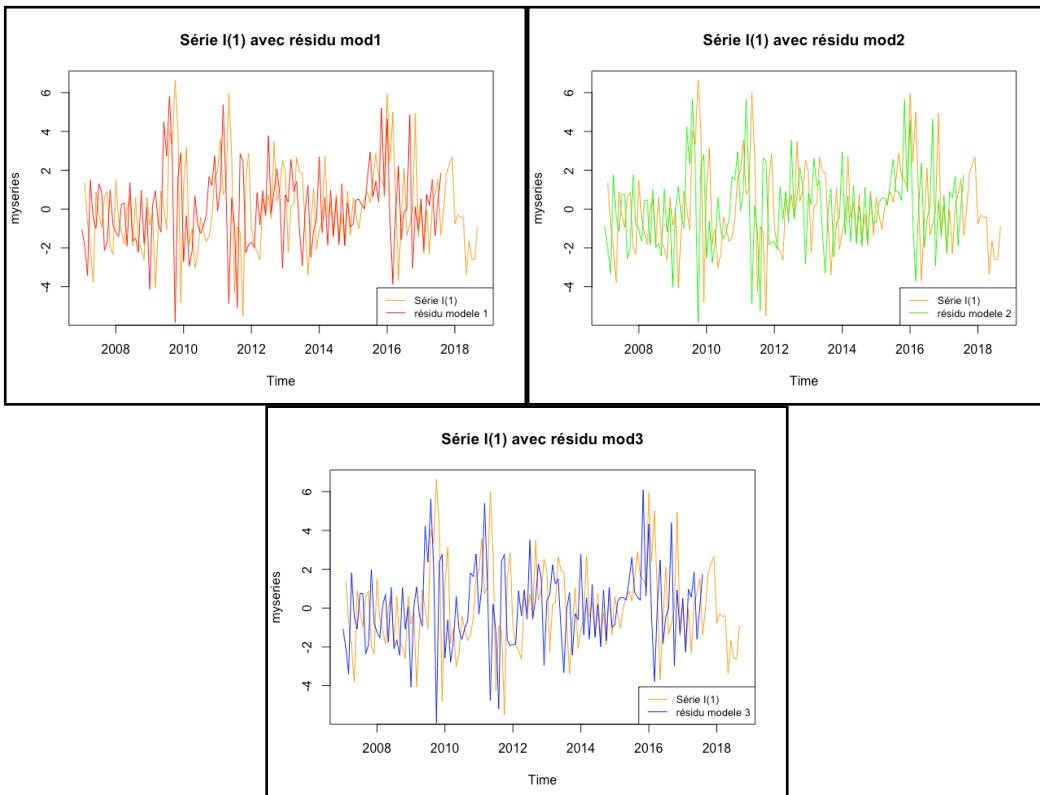
	Chi-sq	df	p-value
Ljung-Box AR(2)	0.48877	2	0.78319
Ljung-Box ARCH(1)	0.77909	1	0.37742

SE of regression 2.12230
 R-squared 0.07171
 Log-lik.(n=128) -275.94414

Modèle 3 : TOURISTE USA + CVT

Source : logiciel R, library ARCH

Figure 38 : modélisation avec variables explicatives



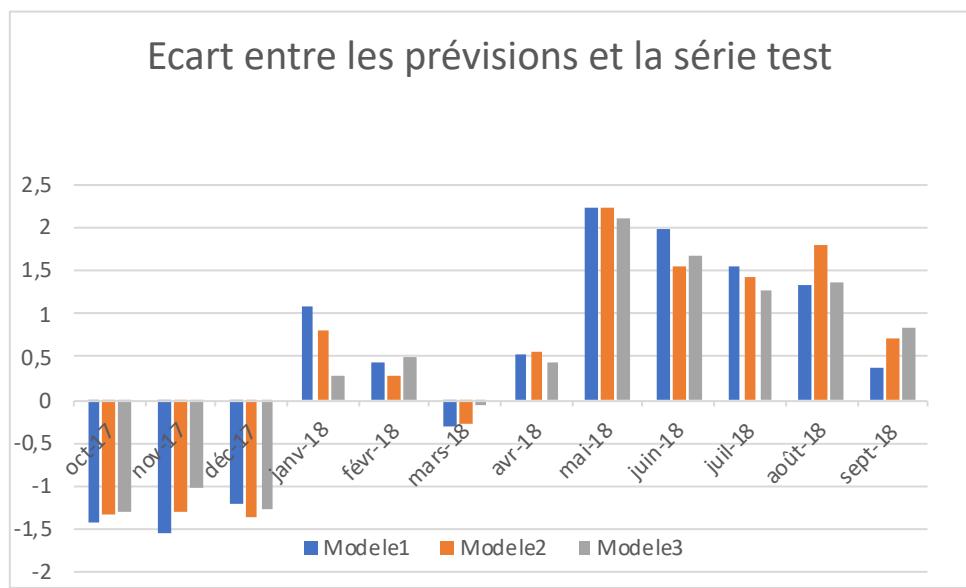
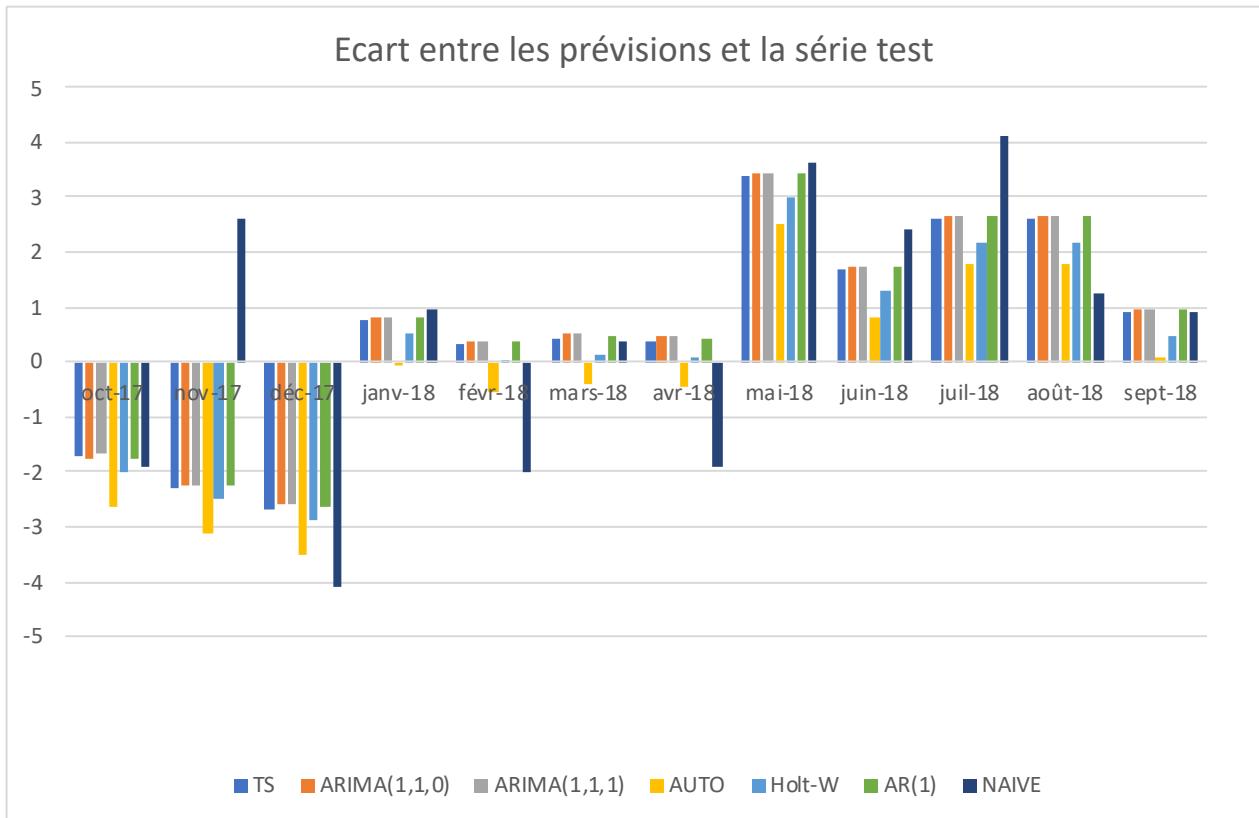
Source : Dossier Techniques de prévision et conjoncture

Tableau 6 : Prévision des modèles

Date/ Résidus	TS	ARIMA(1,1,0)	ARIMA(1,1,1)	AUTO	Holt-W	AR(1)	NAIVE
Octobre 17	0,10862	0,04686	0,13863	-0,82846	-0,18258	0,04421	-0,11773
Novembre 17	0,00000	0,06001	0,06838	-0,82846	-0,20622	0,04738	4,93517
Décembre 17	0,00000	0,06321	0,06491	-0,82846	-0,22987	0,04758	-1,44703
Janvier 18	0,00000	0,06399	0,06471	-0,82846	-0,25352	0,04759	0,16814
Février 18	0,00000	0,06418	0,06470	-0,82846	-0,27716	0,04759	-2,29836
Mars 18	0,00000	0,06423	0,06471	-0,82846	-0,30081	0,04759	-0,05782
Avril 18	0,00000	0,06424	0,06471	-0,82846	-0,32446	0,04759	-2,29127
Mai 18	0,00017	0,06424	0,06471	-0,82846	-0,34810	0,04759	0,27320
Juin 18	0,00016	0,06424	0,06471	-0,82846	-0,37175	0,04759	0,76871
Juillet 18	0,00011	0,06424	0,06471	-0,82846	-0,39540	0,04759	1,53195
Août 18	0,00008	0,06424	0,06471	-0,82846	-0,41904	0,04759	-1,37449
Sept 18	0,00005	0,06424	0,06471	-0,82846	-0,44269	0,04759	-0,00702

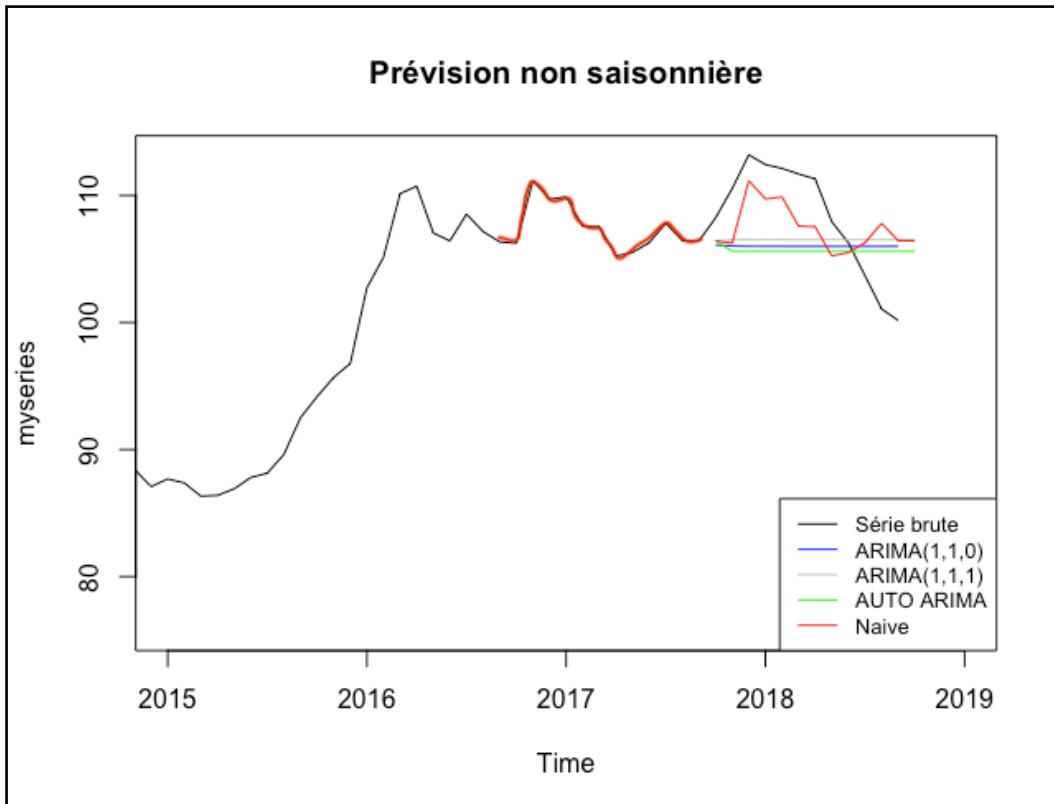
Date/Résidus	Modele1	Modele2	Modele3	Série test
Octobre 17	0,37168	0,46520	0,49913	1,80540
Novembre 17	0,75405	1,01177	1,29195	2,30197
Décembre 17	1,46782	1,31059	1,38898	2,66231
Janvier 18	0,31672	0,04306	-0,47312	-0,76730
Février 18	0,13698	-0,01502	0,18558	-0,30346
Mars 18	-0,75208	-0,71527	-0,50046	-0,43521
Avril 18	0,14656	0,15578	0,05169	-0,39541
Mai 18	-1,12855	-1,14057	-1,25418	-3,36489
Juin 18	0,32495	-0,12054	0,01399	-1,66467
Juillet 18	-1,05092	-1,16139	-1,31743	-2,59138
Août 18	-1,25956	-0,80734	-1,22954	-2,60985
Sept 18	-0,54170	-0,20185	-0,07743	-0,90961

Figure 40 : Ecart entre les prévisions et la série test



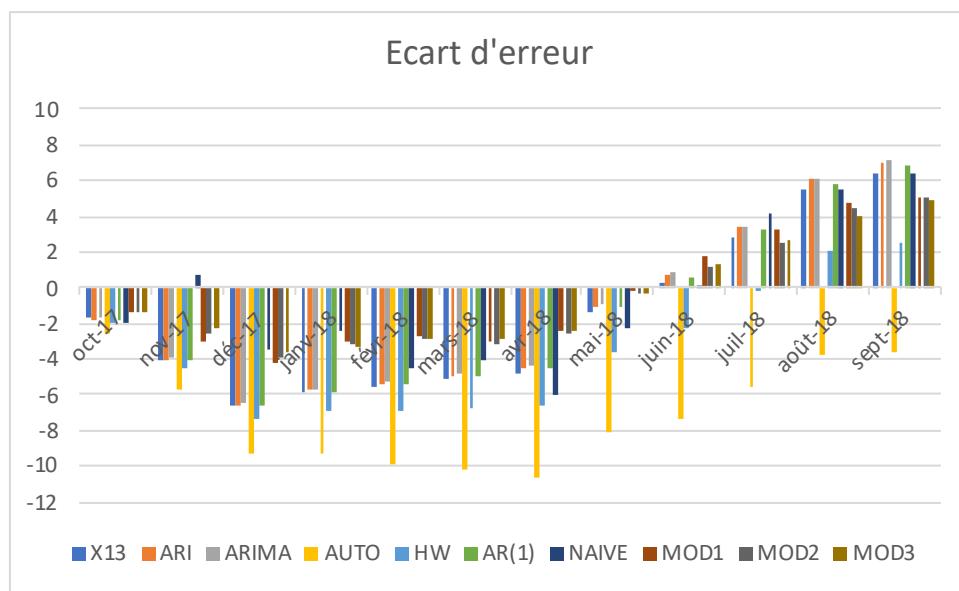
Source : Logiciel excel

Figure 43 : Prévision avec série brute et Naive



Source : Dossier techniques de Prévision et Conjoncture

Figure 44 : Ecart des erreurs



Source : Dossier techniques de Prévision et Conjoncture

Tableau 9 : Mesures d'erreur de prévision

Modele	ME	MAE	MSE
X13	-1,68047	4,177875	4,648647
ARIMA(1,1,0)	-1,39402	4,253137	4,72218
ARIMA(1,1,1)	-1,29117	4,220737	4,694278
AUTO	-7,17422	7,174225	7,667399
Holt-W	-3,53955	4,301897	4,926964
AR(1)	-1,48347	4,238952	4,705389
CVT+GOOGLE	-0,41943	2,900668	3,184448
CVT	-0,57861	2,783254	3,071217
USA+CVT	-0,50419	2,649835	2,907313
Naive	-0,65102	3,45609	3,952732