

**Recueil du Symposium de 2021 de Statistique Canada  
Adopter la science des données en statistique officielle pour répondre aux  
besoins émergents de la société**

**Le tableau de bord de la  
désaisonnalisation de Statistique Canada**

par François Verret

Date de diffusion : le 15 octobre 2021



Statistique  
Canada

Statistics  
Canada

Canada

## Le tableau de bord de la désaisonnalisation de Statistique Canada

François Verret<sup>1</sup>

### Résumé

La méthode X-12-ARIMA est utilisée pour réaliser la désaisonnalisation de séries chronologiques à Statistique Canada. Pour la plupart des programmes statistiques effectuant la désaisonnalisation, les experts des domaines spécialisés (EDS) sont responsables de la gestion du programme, ainsi que de la vérification, de l'analyse et de la diffusion des données, tandis que les méthodologistes du Centre de recherche et d'analyse en séries chronologiques (CRASC) sont chargés de l'élaboration et de la maintenance du processus de désaisonnalisation, de même que du soutien sur la désaisonnalisation aux EDS. Un rapport sommaire visuel appelé le tableau de bord de la désaisonnalisation a été développé à l'aide de R Shiny par le CRASC afin de développer les compétences en interprétation de données désaisonnalisées et de réduire les ressources nécessaires au soutien sur la désaisonnalisation. Il est présentement mis à la disposition des EDS afin de les aider à interpréter et à expliquer les séries désaisonnalisées. Le rapport sommaire inclut des graphiques des séries au fil du temps, en plus de résumer les différents effets saisonniers et de calendrier ainsi que leurs patrons. De plus, les diagnostics de désaisonnalisation clés sont exposés et l'effet net de l'ajustement saisonnier est décomposé en ses différentes composantes. Le présent article donne une représentation visuelle du processus de désaisonnalisation, tout en faisant la démonstration du tableau de bord et de ses fonctionnalités interactives.

Mots clés : série chronologique; X-12-ARIMA; rapport sommaire; R Shiny.

### 1. Introduction

Divers programmes statistiques de Statistique Canada collectent des données mensuelles ou trimestrielles pour former des séries chronologiques (une séquence d'observations recueillies à intervalles réguliers). Les séries chronologiques publiées couvrent un large éventail de sujets, notamment les prix à la consommation, les ventes (p. ex. du commerce de détail et de gros), l'emploi, le produit intérieur brut et le commerce international. En général, l'analyse de ces séries chronologiques infra-annuelles s'intéresse principalement aux variations d'une période à l'autre, et en particulier aux variations les plus récentes observées entre deux périodes. Le plus souvent, ces séries chronologiques montrent des tendances qui se répètent et qui peuvent être attribuées aux effets du calendrier et saisonniers (voir la courbe verte de la figure 1-1 qui présente les données des cinq dernières années d'une série chronologique mensuelle). Ces effets se répètent, sont prévisibles, et ne sont donc pas particulièrement intéressants sur le plan analytique. C'est pourquoi ils sont supprimés de la série et le résultat en est la série désaisonnalisée (courbe rouge de la figure 1-1). Les séries désaisonnalisées peuvent ensuite être décomposées en tendance-cycle (courbe noire de la figure 1-1) et en composantes irrégulières (différence entre les courbes rouge et noire).

---

<sup>1</sup>François Verret, Statistique Canada, 100, promenade du pré Tunney's, Ottawa (Ontario), Canada, K1Y 0T6, francois.verret@statcan.gc.ca

**Figure 1.-1**

**Cinq dernières années de données d'une série non désaisonnalisée (vert) et de la série désaisonnalisée correspondante (rouge) et de la tendance-cycle (noir)**



La composante de tendance-cycle peut être considérée comme la moyenne du processus statistique, tandis que la composante irrégulière peut être considérée comme le résidu. Une série chronologique est le plus souvent décomposée soit de façon additive ( $Y = S + C + I$ , où  $Y$  est la série chronologique d'origine,  $S$  combine les effets de calendrier et saisonniers,  $C$  est la composante de tendance-cycle et  $I$  représente les composantes irrégulières) soit de façon multiplicative ( $Y = S \times C \times I$ ). À Statistique Canada, la désaisonnalisation (DS) est effectuée au moyen de la méthode X-12-ARIMA (Findley et coll., 1998) conformément aux lignes directrices concernant la qualité qui sont publiées (Statistique Canada, 2009). Pour chaque enquête ou programme statistique, chaque période de référence de DS est appliquée dans SAS au moyen d'un système de traitement de séries chronologiques (STSC; Ferland et Fortier, 2009). Le STSC utilise la procédure X12 SAS pour effectuer une DS. Il génère une « liste » X-12-ARIMA détaillée pour chaque série (c.-à-d. un fichier texte de plusieurs pages décrivant étape par étape et en détail le résultat de l'application de la méthode X-12-ARIMA à la série), ainsi qu'une multitude de fichiers SAS pouvant servir à diagnostiquer la DS.

Statistique Canada utilise la gestion matricielle pour optimiser l'efficacité organisationnelle (Statistique Canada, 2016). Ce type de gestion est notamment employé pour le processus de DS. Les experts des domaines spécialisés (EDS) sont responsables du programme ou du projet statistique. Ils coordonnent les activités des différents groupes concernés (collecte, technologie de l'information, méthodes statistiques, diffusion), exécutent le traitement des différentes étapes, et vérifient, analysent et diffusent les produits statistiques. Les méthodologistes du Centre de recherche et d'analyse en séries chronologiques (CRASC) sont responsables de l'élaboration et du maintien de l'environnement de traitement de la DS. Le CRASC met régulièrement à jour les options de DS, effectue des examens périodiques des diagnostics, et fournit un soutien continu aux fins d'analyse et d'interprétation des résultats. Pour faciliter le processus global de DS, le CRASC a créé un tableau de bord interactif. La section 2 de l'article décrit les objectifs et la méthode générale du tableau de bord. La section 3 présente le contenu et les principales fonctionnalités du tableau de bord, tout en expliquant le processus de DS pour une série mensuelle particulière. Une conclusion est donnée à la section 4, accompagnée d'une discussion sur les prochaines étapes du développement.

## **2. Tableau de bord de la désaisonnalisation – Objectifs et méthode**

Le premier objectif du tableau de bord est de renforcer les capacités d'interprétation des données désaisonnalisées. Il a été conçu pour :

- aider les EDS à comprendre les concepts et à éviter qu'ils traitent le processus de DS comme une boîte noire;
- servir d'outil de formation dans les cours sur la DS;
- servir d'outil d'information pour la haute direction.
- 

Le deuxième objectif est d'accroître l'indépendance des EDS et l'efficacité du processus, en réduisant leur dépendance à l'égard du CRASC. Le tableau de bord est un outil automatisé, conçu pour répondre aux questions les plus courantes. On attend également des gains d'efficacité, puisqu'il ne sera plus utile de recueillir et de résumer l'information pertinente provenant des listes de STSC et des sorties SAS.

Le tableau de bord consiste en une page automatisée par série. La page est conçue pour résumer le processus de DS de la manière la plus intuitive possible, en présentant des renseignements numériques à l'appui. Tout en étant intuitive, elle évite la simplification excessive et les approximations. Le tableau de bord vise à répondre aux questions les plus courantes pour une série donnée. Ces questions prennent habituellement la forme suivante.

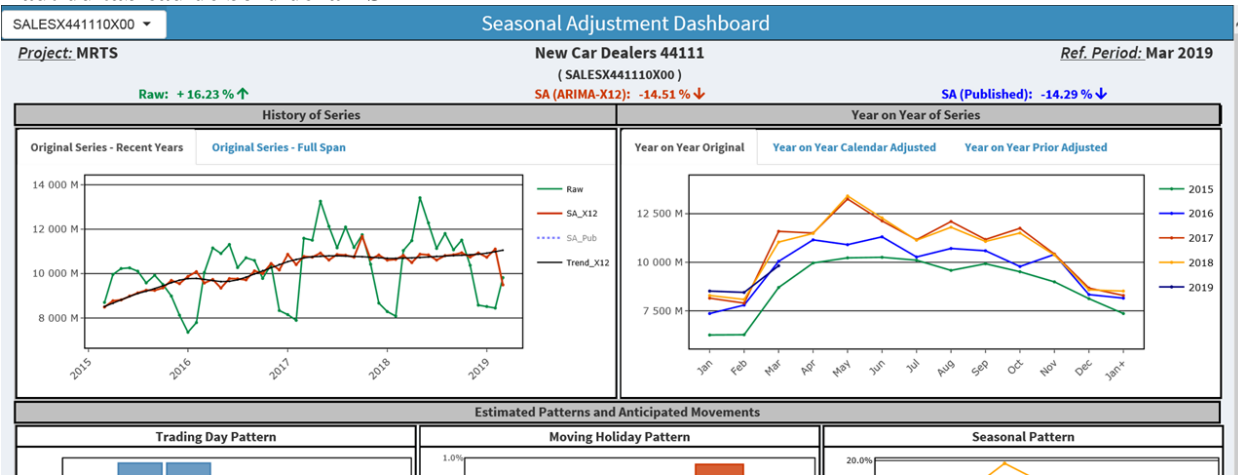
- Le processus de désaisonnalisation fonctionne-t-il comme il le devrait?
- La tendance saisonnière varie-t-elle?
- Pourquoi l'effet de la désaisonnalisation est-il différent cette année par rapport à l'an dernier?
- Constatons-nous un effet de *<insérer l'événement ici>* dans notre estimation?

La plateforme de programmation utilisée pour produire le tableau de bord de la DS a beaucoup évolué au fil des ans. La première version du tableau de bord était une feuille de calcul Excel où les données d'entrée provenant du STSC étaient copiées et collées manuellement. Des macros Visual Basic ont été ajoutées plus tard pour faciliter le reformatage des données d'entrée. Une version du tableau de bord a ensuite été programmée dans Power BI, mais des problèmes de personnalisation, de verrouillage du contenu et d'interactivité se sont posés. L'évaluation de SAS-JMP a montré que son utilisation pourrait entraîner des problèmes similaires. Le produit final de tableau de bord a été programmé au moyen de la plateforme R Shiny en raison de ses nombreux avantages. Tout d'abord, les entrées sont tirées directement des sorties SAS du STSC. Ensuite, la sortie du tableau de bord R Shiny est une page html, ce qui signifie que l'utilisateur n'a pas besoin de connaissances logicielles particulières pour utiliser le tableau de bord. Enfin, la plateforme R Shiny propose de nombreuses fonctions interactives intéressantes, comme le pointage de curseur, le zoom, les menus déroulants et l'utilisation de totalisations pour superposer des graphiques semblables.

### **3. Tableau de bord de la désaisonnalisation – Contenu**

La démonstration du contenu et des fonctionnalités du tableau de bord sera réalisée dans cette section à partir de la série sur les ventes des concessionnaires d'automobiles neuves de l'Enquête mensuelle sur le commerce de détail (EMCD), pour le mois de référence de mars 2019 (notons que la valeur de la série en mars 2019 ne correspond pas à la valeur diffusée afin que certaines fonctionnalités du tableau de bord puissent être montrées). La figure 3-1 présente le haut de la page du tableau de bord pour cette série. Il est possible de choisir la page de la série de l'EMCD à afficher dans le menu supérieur gauche. Ce menu donne des renseignements sur le programme statistique, la série et la période de référence d'intérêt. Un résumé des variations de février 2019 à mars 2019 dans la série brute/non désaisonnalisée est également donné dans l'en-tête (texte vert); il est accompagné des variations de la série désaisonnalisée résultant de l'application de la méthode X-12-ARIMA (texte rouge), ainsi que des variations de la série désaisonnalisée qui seront diffusées (texte bleu). Il pourrait y avoir des différences entre les deux dernières variations quand un processus de réconciliation est appliqué aux diverses séries désaisonnalisées du programme statistique (Quenneville et Fortier, 2011). La méthode de réconciliation est appliquée de façon à ce que les séries constituantes s'additionnent en une série agrégée quand elles sont désaisonnalisées indépendamment.

**Figure 3-1**  
**Haut du tableau de bord de la DS**

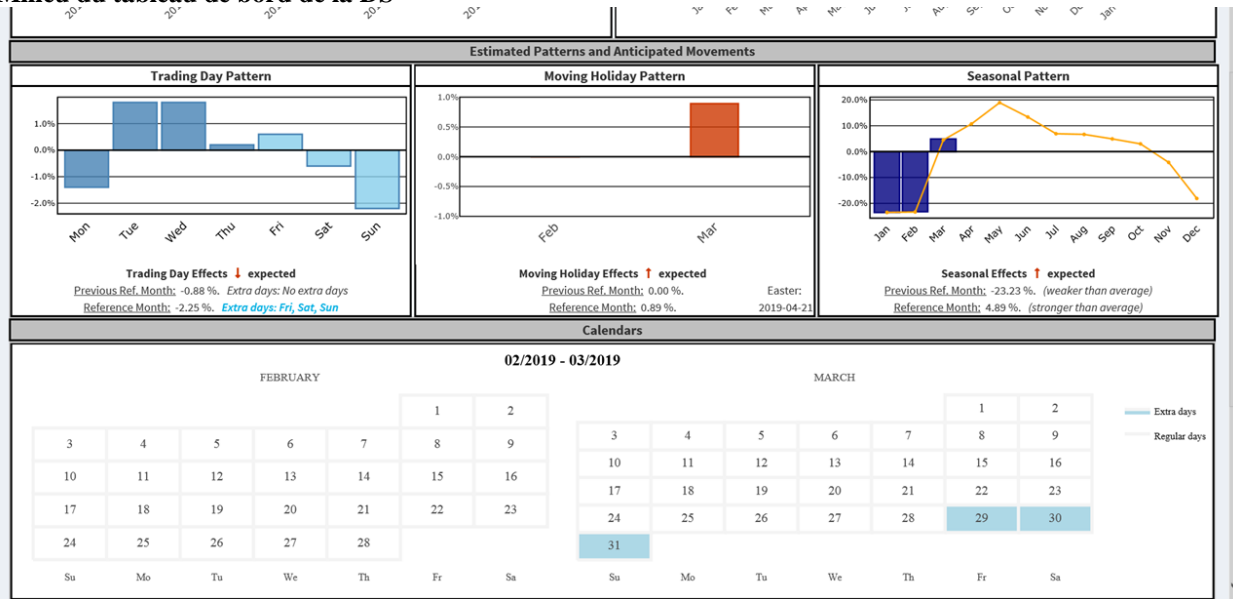


La première ligne de graphiques du tableau de bord présente le panneau « Historique de la série » à gauche et « La série d'une année à l'autre » à droite. Le graphique de gauche est celui présenté à la figure 1-1. Par défaut, la totalisation « Série originale – Années récentes » est sélectionnée, mais on peut choisir d'afficher l'étendue complète de la série en sélectionnant la totalisation « Série originale – Toute la période ». Le graphique d'une année à l'autre présenté à droite correspond à la courbe verte du graphique de gauche, tracé sur lui-même année par année. Des graphiques d'une année à l'autre sont produits pour mettre en évidence les tendances infra-annuelles. Par défaut, la totalisation « Originale d'une année à l'autre » est sélectionnée. Les deux autres totalisations sont présentées à la figure 3-3 et seront abordées dans les paragraphes suivants.

La deuxième ligne de graphiques (voir la figure 3-2, qui présente le milieu de la page du tableau de bord) affiche chacun des effets de calendrier et saisonniers qu'il faut supprimer de la série originale pour obtenir la série désaisonnalisée. Pour que les effets de calendrier soient mieux compris, le calendrier de la période de référence (mars 2019) et celui de la période de référence précédente (février 2019) sont présentés à la ligne suivante. Le premier effet de calendrier présenté est l'effet des jours ouvrables. Il s'agit de l'effet produit par la composition du mois pour ce qui est de chacun des sept jours de la semaine. La première moitié de X-12-ARIMA est l'ajustement d'un modèle de régression ARIMA (reg-ARIMA) à la série. Le graphique à barres verticales indique que le modèle a estimé que les ventes sont moins nombreuses les lundis et dimanches et plus nombreuses les mardis et mercredis. Février 2019 comporte un nombre égal de chacun des sept jours de la semaine, ce qui pourrait conduire à un effet des jours ouvrables nul. Toutefois, l'effet des jours ouvrables tient également compte de la durée variable du mois de février en raison des années bissextiles et est légèrement négatif (-0,88 %), car février 2019 a 28 jours, ce qui est inférieur à la durée moyenne de 28 ¼ d'un mois de février. Mars 2019 a un vendredi, un samedi et un dimanche de plus, comme indiqué en bleu clair sur le deuxième calendrier et sur le graphique à barres verticales, ce qui entraîne une diminution prévue de 2,25 % des ventes. Globalement, nous nous attendons donc à une diminution dans les séries non désaisonnalisées de février à mars 2019 en raison de l'effet des jours ouvrables.

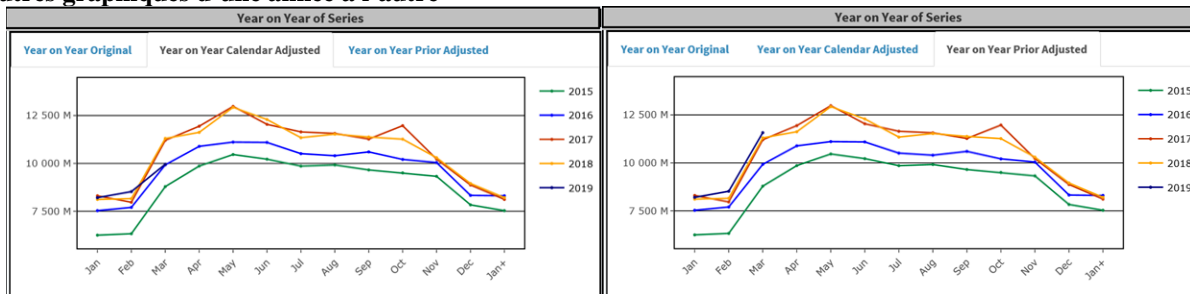
Le deuxième effet de calendrier est l'effet des jours fériés mobiles. Les concessionnaires d'automobiles neuves sont habituellement fermés à Pâques. Comme Pâques peut être en mars ou en avril, on s'attend à ce que le mois de Pâques ait moins de ventes que d'habitude et l'autre mois plus. En 2019, Pâques était le 21 avril. Nous nous attendons donc à ce que les ventes soient plus élevées que d'habitude en mars, de près d'un pour cent de plus selon l'estimation du modèle reg-ARIMA. Globalement, nous nous attendons donc à ce que les séries non désaisonnalisées augmentent de 0,89 % en raison de l'effet des jours fériés mobiles. Nous pouvons voir le résultat de l'élimination de ces deux effets en comparant le graphique ajusté en fonction du calendrier d'une année à l'autre (deuxième tableau du panneau supérieur droit, voir le premier graphique de la figure 3-3) au graphique Originale d'une année à l'autre (figure 3-1, panneau supérieur droit). Le graphique ajusté en fonction du calendrier montre une tendance plus régulière parce que le « bruit » attribué aux effets des jours ouvrables et des jours fériés mobiles a été retiré de la série.

**Figure 3-2**  
Milieu du tableau de bord de la DS



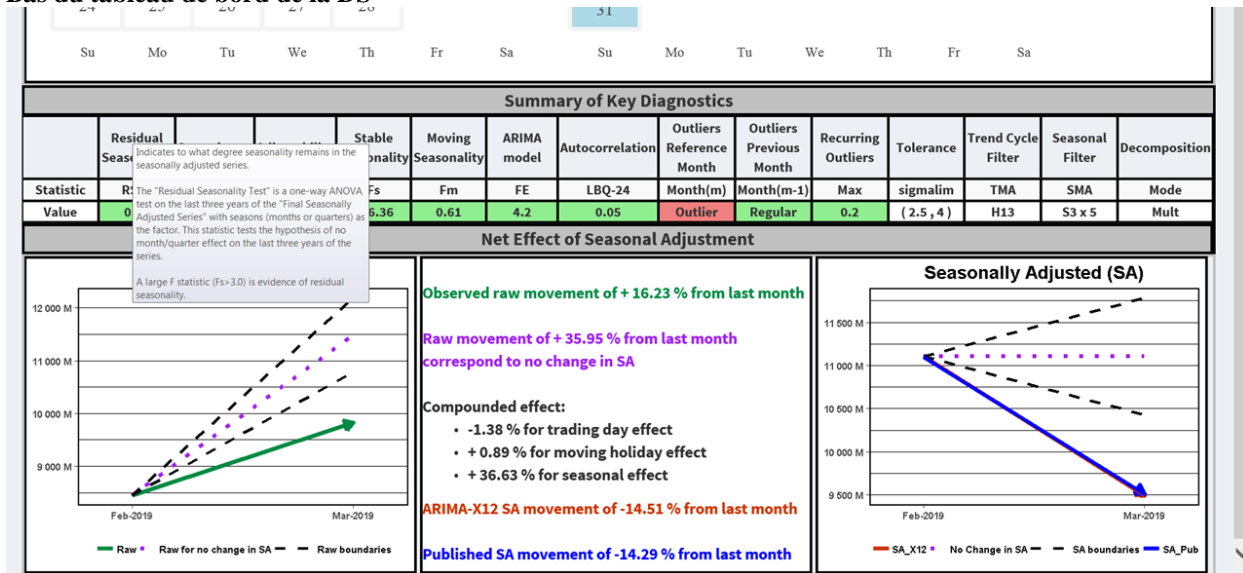
Le modèle reg-ARIMA est également utilisé par X-12-ARIMA pour identifier les valeurs aberrantes de la composante irrégulière. Cela correspond habituellement à des événements particuliers, comme une grève. La valeur du mois d'intérêt correspond en fait à une valeur aberrante très négative. On le constate clairement si l'on regarde le graphique d'Historique de la série (panneau de gauche de la figure 3-1) ou le graphique de la série Originale d'une année à l'autre (premier graphique de la figure 3-3). Dans le premier, le point de données désaisonnalisé de mars 2019 est bien inférieur au point de données de la tendance-cycle. Dans le second, les variations de février à mars 2019 sont nettement plus faibles que les années précédentes. En corrigeant cette valeur aberrante, on obtient le graphique ajusté a priori d'une année à l'autre, qui est le deuxième graphique présenté à la figure 3-3. Cela rend cette tendance encore plus régulière, ce qui permet de voir plus clairement la tendance saisonnière se dessiner.

**Figure 3-3**  
Autres graphiques d'une année à l'autre



La seconde moitié de X-12-ARIMA consiste à appliquer la méthode X11 qui, compte tenu des composantes estimées par le modèle reg-ARIMA, isole les composantes saisonnières, la tendance-cycle et les composantes irrégulières. La composante saisonnière est présentée dans le dernier graphique de la deuxième ligne de graphiques du tableau de bord (graphique en haut à droite de la figure 3-2). La composante saisonnière peut changer avec le temps. Les barres violettes indiquent la tendance estimée pour les mois observés de l'année en cours et la courbe jaune indique la tendance de l'année précédente à des fins de comparaison. Dans cet exemple, nous voyons clairement que nous nous attendons à une très forte hausse de février à mars 2019 dans les séries non désaisonnalisées en raison de la composante saisonnière. Pour ce qui est de la tendance-cycle et des composantes irrégulières, aucun panneau n'y est spécifiquement consacré. Toutefois, comme nous l'avons dit, l'estimation de la tendance-cycle est présentée en noir sur le graphique de l'Historique des séries et celle des composantes irrégulières peut être déduite de la différence entre les courbes désaisonnalisées et les courbes de tendance-cycle.

**Figure 3.-4**  
**Bas du tableau de bord de la DS**



Sur l'avant-dernière ligne du tableau de bord (voir la figure 3-4), nous trouvons un récapitulatif des diagnostics de la DS et des options utilisées. La couleur verte indique les bons diagnostics de DS, la couleur jaune les diagnostics acceptables et la couleur rouge les diagnostics nécessitant une attention particulière. Si l'on pointe le curseur sur un diagnostic donné, son explication s'affiche. À la figure 3-4, le curseur de la souris est pointé sur le diagnostic de saisonnalité résiduelle. Dans cet exemple, nous voyons aussi en rouge que la valeur de la composante irrégulière pour le mois d'intérêt est une valeur aberrante. Ce n'était pas une valeur aberrante le mois précédent, comme l'indique la couleur verte. En plus du processus de détection des valeurs aberrantes du modèle reg-ARIMA (première moitié de la méthode X-12-ARIMA), la méthode X11 comprend un processus de détection des valeurs extrêmes (deuxième moitié de la méthode X-12-ARIMA). Si elles se produisent, ces valeurs extrêmes sont également signalées pour le mois d'intérêt et le mois précédent.

La dernière ligne du tableau de bord présente un résumé de l'effet net sur la désaisonnalisation. À gauche, nous avons en vert les variations observées de février à mars 2019 dans la série originale; les variations qui produiraient une variation de DS plate basée sur les effets estimés de calendrier et saisonniers sont indiquées en violet, avec des intervalles de confiance (approximatifs) autour en noir. Le panneau du milieu donne des renseignements numériques. Nous nous attendons à une augmentation de 35,95 % causée par l'effet saisonnier. Cependant, nous avons observé une augmentation de seulement 16,23 %, ce qui entraîne une variation négative dans la série désaisonnalisée (-14,51 %). Cela est représenté sur le dernier panneau en rouge (la flèche bleue correspond à la variation obtenue par la méthode itérative du quotient), avec la variation DS plate en violet et l'intervalle de confiance correspondant en noir. Le fait d'avoir une variation DS plate dans la direction opposée de la variation originale, comme c'est le cas ici, est une situation typique dans laquelle le CRASC doit faire appel à des EDS. Le tableau de bord explique très clairement les raisons du phénomène et pour quelle raison ce n'est pas un problème.

#### 4. Conclusion et prochaines étapes

Après avoir été mis à la disposition des experts des domaines spécialisés, le tableau de bord de la DS a clairement et rapidement atteint ses objectifs de renforcement des capacités et d'augmentation de l'efficacité. Il s'est révélé un outil de production inestimable, surtout dans le traitement des données relatives à la pandémie. Pendant la production de statistiques, il a suscité des réactions très positives de la part des EDS, des participants à la formation sur les séries chronologiques et de la haute direction. Il a d'abord été produit pour quatre enquêtes économiques mensuelles essentielles : l'Enquête mensuelle sur le commerce de détail, l'Enquête mensuelle sur les industries manufacturières, l'Enquête mensuelle sur le commerce de gros et l'Enquête mensuelle sur les services de restauration et débits de boissons. Ces quatre enquêtes ont de nombreuses similitudes. En particulier, les données sont déclarées pour tout le

mois (ou sont converties par calendarisation pour correspondre à des mois entiers) et les modèles reg-ARIMA utilisés pour leurs séries de flux comprennent généralement une variable explicative des jours ouvrables et une variable explicative des jours fériés mobiles. Le tableau de bord a ensuite été étendu à l'Enquête sur la population active (EPA) et au programme sur les Statistiques de l'assurance-emploi (SAE). Les données de ces deux programmes statistiques sont déclarées pour une semaine de référence du mois. Par conséquent, un effet de semaine de référence est présent dans le modèle reg-ARIMA au lieu d'un effet des jours ouvrables (Statistique Canada, 2020). Les tableaux de bord de la DS de l'EPA et du SAE comprennent alors un panneau des effets de la semaine de référence au lieu d'un panneau des effets des jours ouvrables. Il est également prévu d'étendre le tableau de bord à l'Enquête sur l'emploi, la rémunération et les heures de travail et au programme de l'Indice des prix à la consommation. Cette étape sera suivie d'un déploiement progressif à d'autres programmes statistiques (y compris des programmes trimestriels).

Le développement continu de contenus est aussi prévu. Chaque nouveau programme statistique couvert par le tableau de bord de la DS s'accompagne de l'intégration de caractéristiques propres à l'enquête visée (p. ex. effets de calendrier supplémentaires). De plus, les sorties graphiques peuvent toujours être raffinées, et R Shiny propose de nombreuses options pour ce faire. On ajoutera une totalisation supplémentaire au panneau des graphiques d'une année à l'autre; elle montrera l'effet de la correction de la série pour tenir compte des valeurs extrêmes (voir la valeur d'octobre 2017 dans l'exemple de cet article). L'ajout d'une deuxième page pour chaque série est aussi envisageable pour les utilisateurs avancés. Plus important encore, l'ajout d'une page pour les séries qui sont des agrégats des séries présentées pourrait se révéler très utile. Il n'est pas facile de résumer les divers diagnostics de DS pour un agrégat de séries, mais l'utilisation des sorties de la spécification composite X-12-ARIMA (U.S. Census Bureau, 2017) serait un bon point de départ.

## Remerciements

De nombreux membres actuels et précédents du CRASC ont contribué au tableau de bord de la désaisonnalisation. L'auteur tient à remercier Steve Matthews, Michel Ferland, Nada Habli, Joanne Leung, Yu Wang, Lorcan Mischler, Joana Bérubé et Marouane Seffal pour leur dévouement et leurs nombreuses contributions à l'outil.

## Bibliographie

- Ferland, M. et S. Fortier (2009), « Recent Developments in Statistics Canada's Time Series Processing System : Transition to SAS® PROC X12 », *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, p. 1921-1932.
- Findley, D., B. Monsell, W. Bell, M. Otto, et B. Chen (1998), « New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program », *Journal of Business & Economic Statistics*, 16(2), p. 127-152.
- Quenneville, B. et S. Fortier (2011), « Restoring Accounting Constraints in Time Series – Methods and Software for a Statistical Agency », dans W. R. Bell et coll. (éd.) *Economic Time Series: Modeling and Seasonality*, Chapman and Hall/CRC, p. 231-254.
- Statistique Canada (2009), « Désaisonnalisation et estimation de la tendance-cycle », *Lignes directrices concernant la qualité de Statistique Canada*, 5<sup>e</sup> édition, no 12-539-X au catalogue, disponible à l'adresse <https://www150.statcan.gc.ca/n1/pub/12-539-x/2009001/seasonal-saisonnal-fra.htm>
- Statistics Canada (2016), « Chapitre 2.1 : Structure organisationnelle et gestion matricielle », *Compendium de pratiques de gestion pour les organismes statistiques du Programme international d'organismes statistiques de Statistique*, no 11-634-X au catalogue, disponible à l'adresse <https://www150.statcan.gc.ca/n1/pub/11-634-x/2016001/section2/chap1-fra.htm>
- Statistique Canada (2020), « Section 6 : Traitement des données », *Guide de l'Enquête sur la population active*, no 71-543-G au catalogue, disponible à l'adresse <https://www150.statcan.gc.ca/n1/pub/71-543-g/71-543-g2020001-fra.htm>



U.S. Census Bureau (2017), X-13ARIMA-SEATS Reference Manual, Version 1.1, Washington, DC, disponible à l'adresse <https://www.census.gov/ts/x13as/docX13AS.pdf>