



École Nationale Supérieure d'Informatique
pour l'Industrie et l'Entreprise



Application du Modèle de Rupture pour la loi normale asymétrique aux données Régularité mensuelle TER (SUJET 5)

Réalisé par:

N'CIB Meriem

BARNICHA Mohammed Yassine

GHARBI Mohamed Skander

HAMMAMI Mohtadi

KAMMOUN Yasmine

AZZABI Mouhcine

BENTAHAR Zakariae

Encadré par:

GHORBANZADEH Dariush

2^{ème} Année MA

Année Universitaire: 2023-2024

Table des matières

Introduction générale	1
1 Contextualisation	2
1.1 Présentation des données	2
1.2 Modèle de rupture	3
2 Méthode de detection du point de rupture	5
2.1 Estimation des paramètres pour un point de rupture donné	5
2.2 Détermination de \hat{k} , point de rupture optimal	7
2.3 Test de Kolmogorov-Smirnov	8
3 Application aux données	9
3.1 Bretagne	10
3.1.1 Trains programmés	10
3.1.1.1 Moment de rupture \hat{k}	11
3.1.1.2 Test de Kolmogorov-Smirnov	12
3.1.1.3 Évaluation des paramètres des lois	12
3.1.2 Trains annulés	14
3.1.2.1 Moment de rupture \hat{k}	15
3.1.2.2 Test de Kolmogorov-Smirnov	16
3.1.2.3 Évaluation des paramètres des lois	16
3.2 Pays-de-la-Loire	18
3.2.1 Trains programmés	18
3.2.1.1 Moment de rupture \hat{k}	19
3.2.1.2 Test de Kolmogorov-Smirnov	20
3.2.1.3 Évaluation des paramètres des lois	20

TABLE DES MATIÈRES

3.2.2	Trains annulés	22
3.2.2.1	Moment de rupture \hat{k}	23
3.2.2.2	Test de Kolmogorov-Smirnov	24
3.2.2.3	Évaluation des paramètres des lois	24
3.3	Centre	26
3.3.1	Trains programmés	26
3.3.1.1	Moment de rupture \hat{k}	27
3.3.1.2	Test de Kolmogorov-Smirnov	28
3.3.1.3	Évaluation des paramètres des lois	28
3.3.2	Trains annulés	30
3.3.2.1	Moment de rupture \hat{k}	31
3.3.2.2	Test de Kolmogorov-Smirnov	32
3.3.2.3	Évaluation des paramètres des lois	32
3.4	Aquitaine	34
3.4.1	Trains programmés	34
3.4.1.1	Moment de rupture \hat{k}	35
3.4.1.2	Test de Kolmogorov-Smirnov	36
3.4.1.3	Évaluation des paramètres des lois	36
3.4.2	Trains annulés	38
3.4.2.1	Moment de rupture \hat{k}	39
3.4.2.2	Test de Kolmogorov-Smirnov	40
3.4.2.3	Évaluation des paramètres des lois	40
3.5	Provence Alpes Côte d'Azur	42
3.5.1	Trains programmés	42
3.5.1.1	Moment de rupture \hat{k}	43
3.5.1.2	Test de Kolmogorov-Smirnov	44
3.5.1.3	Évaluation des paramètres des lois	44
3.5.2	Trains annulés	46
3.5.2.1	Moment de rupture \hat{k}	47
3.5.2.2	Test de Kolmogorov-Smirnov	47
3.5.2.3	Évaluation des paramètres des lois	48
3.6	Normandie	50

TABLE DES MATIÈRES

3.6.1	Trains programmés	50
3.6.1.1	Moment de rupture \hat{k}	51
3.6.1.2	Test de Kolmogorov-Smirnov	51
3.6.1.3	Évaluation des paramètres des lois	51
3.6.2	Trains annulés	53
3.6.2.1	Moment de rupture \hat{k}	54
3.6.2.2	Test de Kolmogorov-Smirnov	54
3.6.2.3	Évaluation des paramètres des lois	55
3.7	Hauts-de-France	57
3.7.1	Trains programmés	57
3.7.1.1	Moment de rupture \hat{k}	58
3.7.1.2	Test de Kolmogorov-Smirnov	58
3.7.1.3	Évaluation des paramètres des lois	58
3.7.2	Trains annulés	60
3.7.2.1	Moment de rupture \hat{k}	61
3.7.2.2	Test de Kolmogorov-Smirnov	62
3.7.2.3	Évaluation des paramètres des lois	62
3.8	Auvergne-Rhône-Alpes	64
3.8.1	Trains programmés	64
3.8.1.1	Moment de rupture \hat{k}	65
3.8.1.2	Test de Kolmogorov-Smirnov	65
3.8.1.3	Évaluation des paramètres des lois	65
3.8.2	Trains annulés	67
3.8.2.1	Moment de rupture \hat{k}	68
3.8.2.2	Test de Kolmogorov-Smirnov	68
3.8.2.3	Évaluation des paramètres des lois	69
Conclusion et ANNEXE		71
3.9	Conclusion	71
3.10	ANNEXE	72
3.10.1	Code permettant d'estimer les paramètres des lois pour un k donné	72
3.10.2	Code permettant de trouver le point de rupture optimal	74
3.10.3	Test de Kolmogorov-Smirnov	75

Introduction

Les séries temporelles sont des données qui sont collectées au fil du temps et qui sont organisées de manière chronologique. Elles peuvent être utilisées pour analyser et comprendre les tendances et les modèles dans les données au fil du temps, ainsi que pour faire des prévisions sur les événements futurs. Les séries temporelles peuvent être trouvées dans de nombreux domaines, tels que la finance, l'économie, la météorologie, la santé...

Les modèles de rupture, également connus sous le nom de modèles de changement de régime, sont des outils statistiques qui permettent de détecter et de modéliser les changements dans la structure d'une série temporelle. Ces modèles sont souvent utilisés pour analyser des données économiques, financières, météorologiques ou sociales, où il peut y avoir des périodes de stabilité suivies de périodes de changement soudain.

Les modèles de rupture permettent d'identifier les moments où un changement significatif s'est produit dans les données et de modéliser le comportement de la série temporelle avant et après ce changement. Autrement dit, pour une séquence donnée de variables aléatoires $(X_i)_{1 \leq i \leq n}$, nous essayons de trouver un point de rupture k où les éléments X_1, \dots, X_k ont une fonction de distribution identique f_1 et les éléments X_{k+1}, \dots, X_n sont distribuées selon une autre densité de probabilité f_2 .

Dans le rapport qui suit nous allons nous pencher sur le Modèle de Rupture pour la loi normale asymétrique (Skew normal distribution) et l'appliquer aux données mensuelles de TER.

Chapitre 1

Contextualisation

1.1 Présentation des données

Notre jeu de données contient les informations relatives aux TER en circulation dans différentes régions de France. Ces informations sont mensuelles, et nous disposons d'un nombre d'observations différent selon la région concernée (*cf Figure 1.1*).

Région	Observations disponibles	Période concernée
Bretagne	134	Entre 01/2013 et 02/2024
Provence Alpes Côte d'Azur	134	Entre 01/2013 et 02/2024
Pays-de-la-Loire	134	Entre 01/2013 et 02/2024
Aquitaine	92	Entre 01/2013 et 08/2020
Centre	92	Entre 01/2013 et 02/2024
Grand Est	86	Entre 01/2017 et 02/2024
Bourgogne-Franche-Comté	74	Entre 01/2018 et 02/2024
Normandie	74	Entre 01/2018 et 02/2024
Hauts-de-France	74	Entre 01/2018 et 02/2024
Auvergne-Rhône-Alpes	74	Entre 01/2018 et 02/2024
Occitanie	68	Entre 07/2018 et 02/2024
Midi Pyrénées	66	Entre 01/2013 et 06/2018
Poitou Charentes	60	Entre 01/2013 et 12/2017
Picardie	60	Entre 01/2013 et 12/2017
Limousin	60	Entre 01/2013 et 12/2017
Franche Comté	60	Entre 01/2013 et 12/2017
Bourgogne	60	Entre 01/2013 et 12/2017
Basse Normandie	60	Entre 01/2013 et 12/2017
Auvergne	60	Entre 01/2013 et 12/2017
Haute Normandie	60	Entre 01/2013 et 12/2017

Rhône Alpes	60	Entre 01/2013 et 12/2017
Languedoc Roussillon	60	Entre 01/2013 et 12/2017
Nord Pas de Calais	58	Entre 01/2013 et 12/2017
Alsace	48	Entre 01/2013 et 12/2016
Champagne Ardenne	48	Entre 01/2013 et 12/2016
Centre Val-de-Loire	42	Entre 09/2020 et 02/2024
Nouvelle Aquitaine	42	Entre 09/2020 et 02/2024
Lorraine	13	Entre 02/2015 et 12/2016

FIGURE 1.1 : Nombre d'observations en fonction de la région

Dans notre rapport, nous considérons seulement les régions pour lesquelles nous avons plus de 70 observations (les autres régions sont traitées dans le code joint au rapport). Pour chacune de ces régions, nous nous intéressons en particulier aux variables '**nombre de trains programmés**' et '**nombre de trains annulés**'. Ces variables constituent des séries temporelles auxquelles nous appliquerons notre modèle de rupture.

1.2 Modèle de rupture

Nous souhaitons appliquer notre modèle de rupture aux séries temporelles correspondant au nombre de trains programmés et au nombre de trains annulés pour les régions sélectionnées.

Nous utilisons un **modèle de rupture pour la loi normale asymétrique**. La densité de la loi normale asymétrique est de la forme :

$$f(x, \mu, \sigma, \theta) = \frac{2\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\theta\left(\frac{x-\mu}{\sigma}\right)\right)}{\sigma}$$

où $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ et $\Phi(x) = \int_{-\infty}^x \phi(t)dt$.

Nous considérons le modèle suivant :

$$\begin{cases} X_t \sim \mathcal{SN}(\mu_1, \sigma_1, \theta) & \text{si } t \leq k \\ X_t \sim \mathcal{SN}(\mu_2, \sigma_2, \theta) & \text{si } t > k \end{cases}$$

où $\mu_1, \sigma_1, k, \mu_2, \sigma_2, \theta$ sont inconnus que l'on souhaite estimer.

Concrètement, nous suivons la démarche suivante :

- Soit k , un moment de rupture supposé. Nous évaluons les paramètres des lois normales asymétriques des variables aléatoires X_1, \dots, X_k et X_{k+1}, \dots, X_n en nous basant sur les échantillons dont nous disposons.
- Nous souhaitons déterminer \hat{k} , le moment de rupture le plus probable. Pour cela, nous évaluons les paramètres des lois normales asymétriques pour chaque k candidat (c'est à dire appartenant à $\llbracket 10, n - 10 \rrbracket$). Nous sélectionnons alors k tel que les lois obtenues en considérant ce moment de rupture soient le plus plausible.
- Ayant déterminé \hat{k} , nous testons la similarité en loi des échantillons $(x_1, \dots, x_{\hat{k}})$ et $(x_{\hat{k}+1}, \dots, x_n)$.

La mise en œuvre de chaque étape est détaillée dans le chapitre suivant.

Chapitre 2

Méthode de detection du point de rupture

La détection de point de rupture est une tâche cruciale dans de nombreux domaines tels que l'économie, la météorologie et la surveillance de processus industriels. Identifier avec précision ces points peut révéler des éléments importants sur les événements sous-jacents ou les changements de régime.

La méthode utilisée dans notre analyse repose sur la maximisation de la vraisemblance, une approche robuste qui consiste à ajuster les paramètres de modèles statistiques de manière à rendre la séquence de données observées aussi probable que possible sous ces modèles. Cette méthode est particulièrement efficace pour modéliser des séries avec des ruptures, car elle permet de comparer rigoureusement les hypothèses de continuité contre celles de changement à des instants spécifiques.

2.1 Estimation des paramètres pour un point de rupture donné

Soit $k \in [10, n - 10]$, un point de rupture supposé. On a alors :

$$\begin{cases} X_1, \dots, X_k \sim \mathcal{SN}(\mu_1, \sigma_1, \theta) \\ X_{k+1}, \dots, X_n \sim \mathcal{SN}(\mu_2, \sigma_2, \theta) \end{cases}$$

Sous hypothèse d'indépendance des observations, la vraisemblance s'écrit alors :

$$L(\underline{x}, \mu_1, \mu_2, \sigma_1, \sigma_2, \theta) = \prod_{i=1}^k f_1(x_i, \mu_1, \sigma_1, \theta) * \prod_{i=k+1}^n f_2(x_i, \mu_2, \sigma_2, \theta)$$

où

$$\begin{cases} f_1(x_i, \mu_1, \sigma_1, \theta) = \frac{2\phi\left(\frac{x_i - \mu_1}{\sigma_1}\right)\Phi\left(\theta\left(\frac{x_i - \mu_1}{\sigma_1}\right)\right)}{\sigma_1} \\ f_2(x_i, \mu_2, \sigma_2, \theta) = \frac{2\phi\left(\frac{x_i - \mu_2}{\sigma_2}\right)\Phi\left(\theta\left(\frac{x_i - \mu_2}{\sigma_2}\right)\right)}{\sigma_2} \end{cases}$$

soit :

$$L(\underline{x}, \mu_1, \mu_2, \sigma_1, \sigma_2, \theta) = \frac{2^n}{\sigma_1^k * \sigma_2^{(n-k)}} * \prod_{i=1}^k \phi\left(\frac{x_i - \mu_1}{\sigma_1}\right) * \Phi\left(\theta\left(\frac{x_i - \mu_1}{\sigma_1}\right)\right) * \prod_{i=k+1}^n \phi\left(\frac{x_i - \mu_2}{\sigma_2}\right) * \Phi\left(\theta\left(\frac{x_i - \mu_2}{\sigma_2}\right)\right)$$

Pour simplifier les calculs, nous passons à la log vraisemblance et nous obtenons finalement :

$$\begin{aligned} l(\underline{x}, \mu_1, \mu_2, \sigma_1, \sigma_2, \theta) = & \\ & \frac{n}{2} * \log(2) - \frac{n}{2} * \log(\pi) - k * \log(\sigma_1) - (n - k) * \log(\sigma_2) - \frac{1}{2 * \sigma_1^2} * \sum_{i=1}^k (x_i - \mu_1)^2 \\ & + \sum_{i=1}^k \log(\Phi(\theta(\frac{x_i - \mu_1}{\sigma_1}))) - \frac{1}{2 * \sigma_2^2} * \sum_{i=k+1}^n (x_i - \mu_2)^2 + \sum_{i=k+1}^n \log(\Phi(\theta(\frac{x_i - \mu_2}{\sigma_2}))) \end{aligned}$$

Ou encore

$$\begin{aligned} l(\underline{x}, \mu_1, \mu_2, \sigma_1, \sigma_2, \theta) = & \\ C^{te} - k * \log(\sigma_1) - (n - k) * \log(\sigma_2) - \frac{1}{2 * \sigma_1^2} * \sum_{i=1}^k (x_i - \mu_1)^2 + \sum_{i=1}^k & \log(\Phi(\theta(\frac{x_i - \mu_1}{\sigma_1}))) \\ - \frac{1}{2 * \sigma_2^2} * \sum_{i=k+1}^n (x_i - \mu_2)^2 + \sum_{i=k+1}^n & \log(\Phi(\theta(\frac{x_i - \mu_2}{\sigma_2}))) \end{aligned}$$

On peut alors estimer les paramètres des deux lois normales asymétriques par maximisation de la log vraisemblance :

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = \operatorname{argmax}_{\mu_1, \mu_2, \sigma_1, \sigma_2, \theta} l(\underline{x}, \mu_1, \mu_2, \sigma_1, \sigma_2, \theta)$$

sous conditions $\sigma_1, \sigma_2 > 0$

Si nous calculons les dérivées partielles de la log vraisemblance , nous nous rendons compte qu'il n'existe pas de solution explicite. Nous utiliserons donc la fonction minimize du module scipy.optimize sur l'opposé de la log vraisemblance afin de pouvoir estimer les paramètres des lois.

Le code de la fonction permettant d'estimer les paramètres des lois pour un point de rupture donné est fourni en **ANNEXE**

2.2 Détermination de \hat{k} , point de rupture optimal

À présent, nous voulons déterminer \hat{k} , le point de rupture optimal, parmi les points de rupture candidats $k \in [\![10, n - 10]\!]$

Pour cela, pour chaque k candidat, nous estimons les paramètres des lois normales asymétriques en utilisant la méthode décrite dans la section précédente : nous obtenons alors $\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k$ pour chaque k .

Nous pouvons alors calculer la log vraisemblance en fonction de k en utilisant ces paramètres estimés :

$$\Lambda(k) = l(\underline{x}^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k)$$

Une fois encore, nous cherchons à maximiser la vraisemblance et le point de rupture se trouve de la manière suivante :

$$\hat{k} = \operatorname{argmax}_k \Lambda(k)$$

Le code de la fonction permettant de trouver le point de rupture optimal est donné en **ANNEXE**

2.3 Test de Kolmogorov-Smirnov

Ayant trouvé \hat{k} , nous nous interrogeons sur la pertinence d'un tel point de rupture et donc sur la nécessité de séparer l'échantillon $(x_1, \dots, x_{\hat{k}}, x_{\hat{k}+1}, \dots, x_n)$ en deux sous échantillons $(x_1, \dots, x_{\hat{k}})$ et $(x_{\hat{k}+1}, \dots, x_n)$. Pour cela nous cherchons à tester la similarité en loi des échantillons $(x_1, \dots, x_{\hat{k}})$ et $(x_{\hat{k}+1}, \dots, x_n)$. Nous utiliserons le test de **Kolmogorov-Smirnov**.

Le test de Kolmogorov-Smirnov est un test d'hypothèse statistique permettant de tester la similarité en loi de deux échantillons. Autrement dit, il permet de tester si deux échantillons ont été générés ou non par la même loi.

$$\begin{cases} H_0 : \text{Les deux échantillons sont générés par la même loi} \\ H_1 : \text{Les deux échantillons sont générés par deux lois différentes} \end{cases}$$

Nous utiliserons la p-value du test avec un seuil de confiance $\alpha=0.05$

- **p-value** $<\alpha$: on rejette H_0 . Il y a des preuves suffisantes pour affirmer que les deux échantillons ont été générés par deux lois différentes.
- **p-value** $\geq \alpha$: on ne rejette pas H_0 . Il n'y a pas de preuves suffisantes pour affirmer que les deux échantillons ont été générés par deux lois différentes.

Bien sûr, le point de rupture est pertinent seulement dans le cas de H_1 , c'est à dire dans le cas où les deux échantillons ont été générés par deux lois différentes. **Autrement dit, le point de rupture est pertinent dans le cas où p-value** $<\alpha=0.05$

Pour effectuer nos test de Kolmogorov-Smirnov, nous utilisons la fonction `kstest` du module `scipy.stats`.

Chapitre 3

Application aux données

Nous allons à présent appliquer notre méthode de détection de point de rupture présentée au chapitre précédent à nos données.

Nous rappelons que dans notre rapport, nous traitons seulement les régions pour lesquelles nous disposons de plus de 70 observations (les autres régions sont traitées dans le code joint au rapport) et que notre modèle sera appliqué aux séries temporelles correspondant aux nombre de trains programmés et au nombre de trains annulés pour les régions sélectionnées (*cf Figure 3.1*).

Région	Observations disponibles	Période concernée
Bretagne	134	Entre 01/2013 et 02/2024
Provence Alpes Côte d'Azur	134	Entre 01/2013 et 02/2024
Pays-de-la-Loire	134	Entre 01/2013 et 02/2024
Aquitaine	92	Entre 01/2013 et 08/2020
Centre	92	Entre 01/2013 et 02/2024
Grand Est	86	Entre 01/2017 et 02/2024
Bourgogne-Franche-Comté	74	Entre 01/2018 et 02/2024
Normandie	74	Entre 01/2018 et 02/2024
Hauts-de-France	74	Entre 01/2018 et 02/2024
Auvergne-Rhône-Alpes	74	Entre 01/2018 et 02/2024

FIGURE 3.1 : Régions selectionnées

3.1 Bretagne

Nous examinons la performance du réseau ferroviaire dans la région de Bretagne, en nous concentrant sur les données opérationnelles recueillies sur une période s'étendant de 2013 à 2023. Les indicateurs clés comprennent le nombre de trains programmés, ceux ayant réellement circulé, ainsi que les incidents tels que les annulations et les retards à l'arrivée.

Dans cette partie, nous avons entrepris une étude approfondie des données sur le nombre de trains programmés en Bretagne, s'étendant de 2013 à 2023. L'analyse suivante présente une visualisation de ces données, révélant les tendances sous-jacentes et les potentiels points de rupture qui ont marqué la période observée.

3.1.1 Trains programmés

Dans cette partie, nous nous concentrons sur la visualisation du nombre de trains programmés. Cette visualisation nous aide non seulement à comprendre les tendances générales et les variations saisonnières, mais elle est également cruciale pour détecter les anomalies et les points de rupture qui pourraient indiquer des changements majeurs dans la planification des transports ou refléter des réponses à des événements externes.

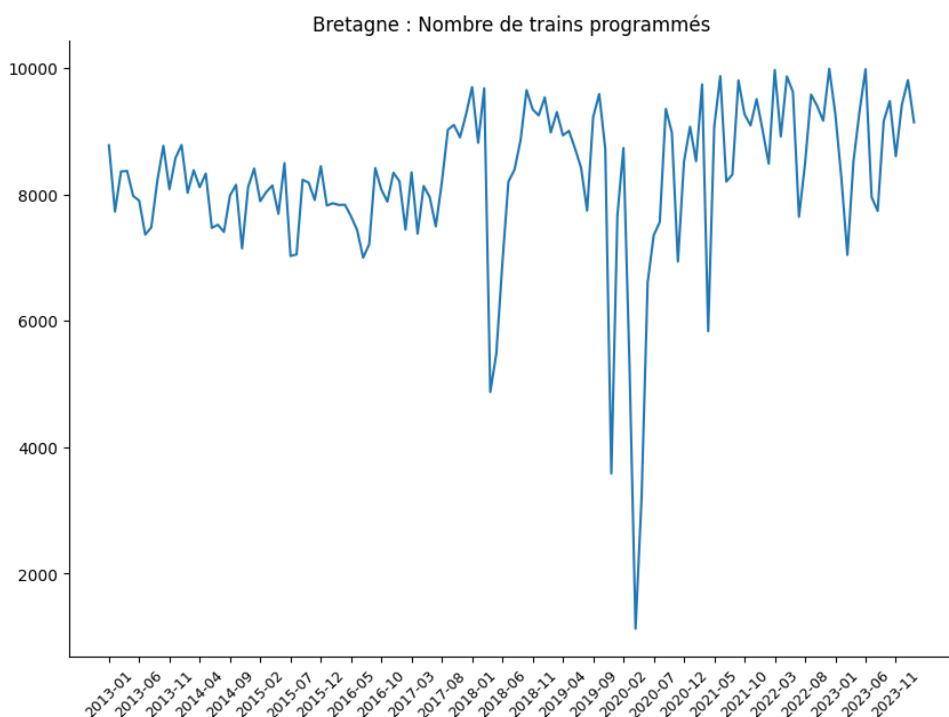


FIGURE 3.2 : Bretagne : Nombre de trains programmés de 2013 à 2024.

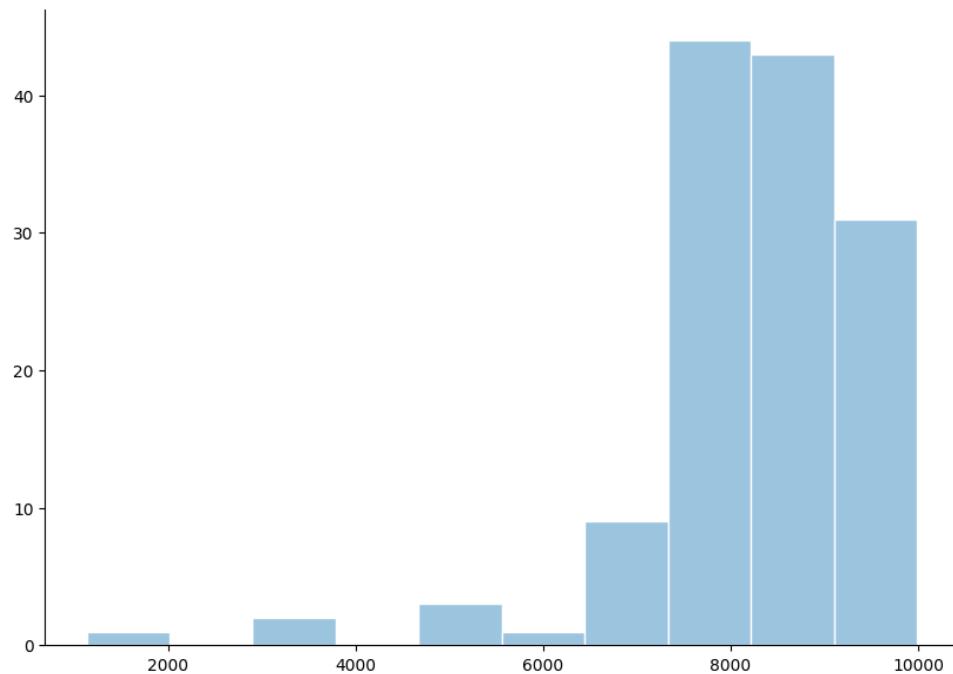


FIGURE 3.3 : Bretagne : Régions selectionnées.

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.1.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2017-08**.

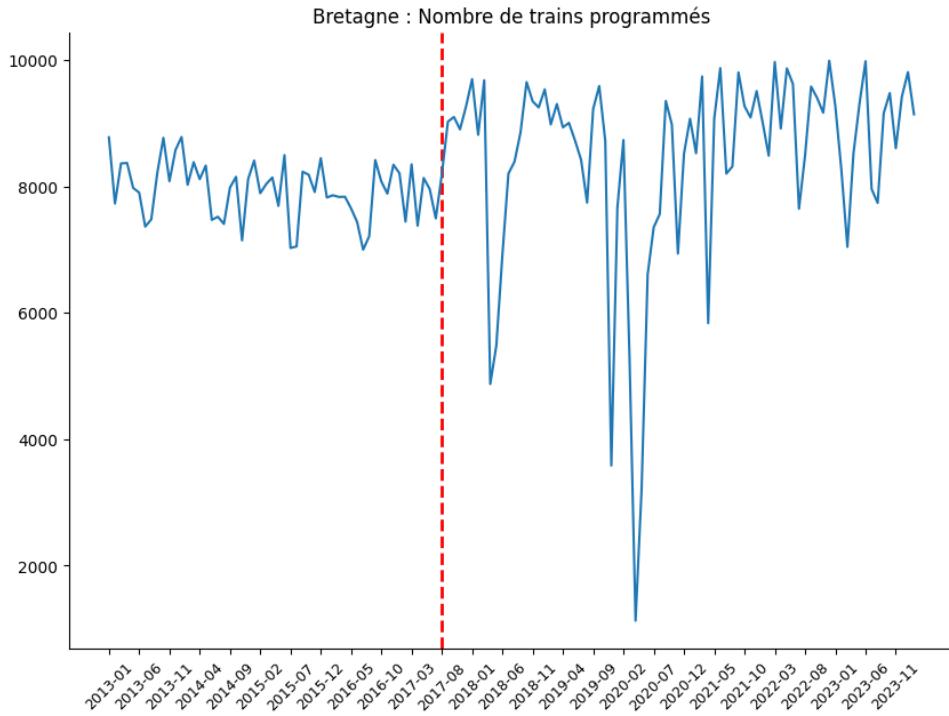


FIGURE 3.4 : Bretagne : La date de reprise.

3.1.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 4.314571e-11** : nous rejettions H_0 , ainsi le point de rupture est pertinent.

3.1.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (8627.61, 9957.25, 850.73, 2140.80, -6.96)$$

Nous constatons que la moyenne des trains programmés avant le point de rupture ($\hat{\mu}_1^k$) est inférieure à celle après le point de rupture ($\hat{\mu}_2^k$). Cela indique une augmentation du nombre moyen de trains programmés post-rupture. De plus, nous observons une augmentation significative de la variance après le point de rupture ($\hat{\sigma}_2^k$ par rapport à $\hat{\sigma}_1^k$), ce qui suggère une plus grande dispersion dans le nombre de trains programmés. Ces observations sont en accord avec la visualisation présentée à la Figure 3.4.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.5 et 3.6*)

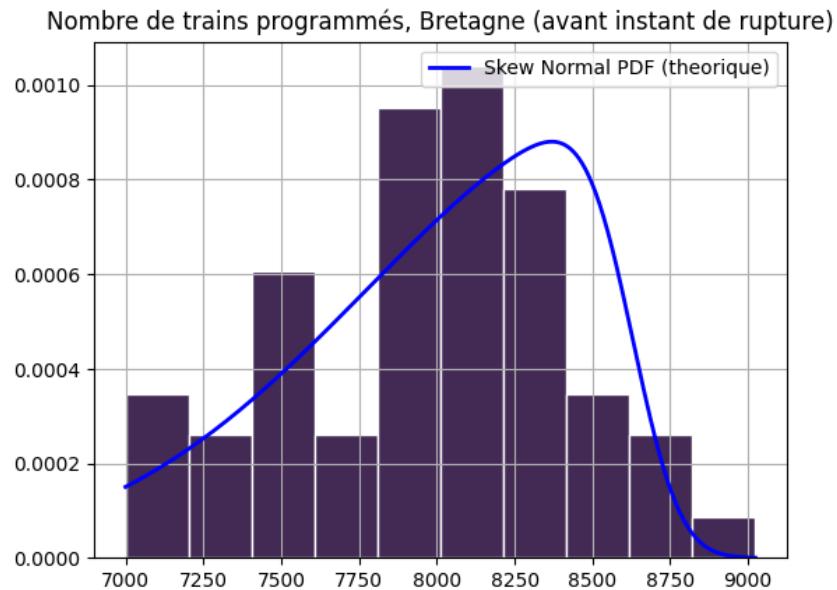


FIGURE 3.5 : Histogramme de l'échantillon avant point de rupture

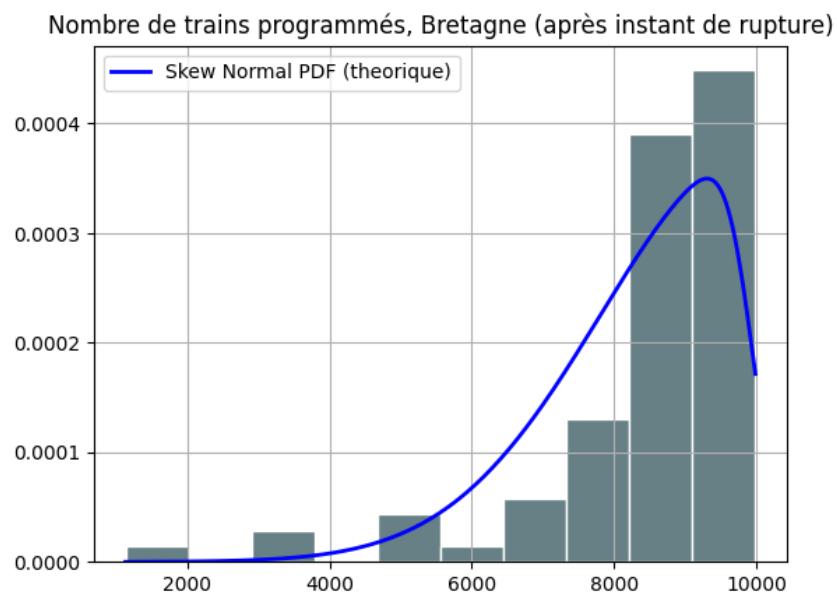


FIGURE 3.6 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ négatif, indiquant des lois normales asymétriques inclinées vers la droite, ce qui signifie une queue de distribution plus longue du côté gauche. Cette caractéristique semble concorder avec les histogrammes présentés, où l'on observe une concentration plus importante des valeurs à gauche de la moyenne.

3.1.2 Trains annulés

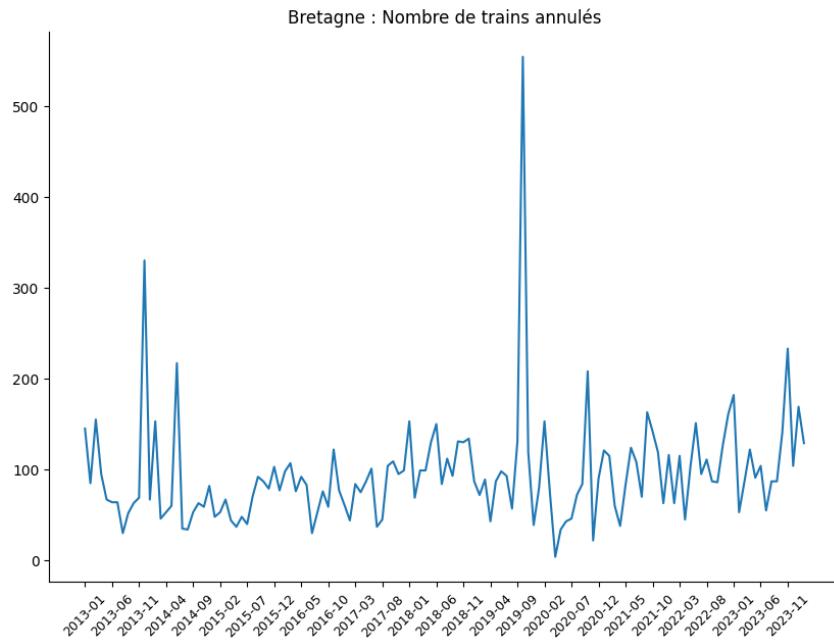


FIGURE 3.7 : Nombre de trains annulés entre 01/2013 et 02/2024

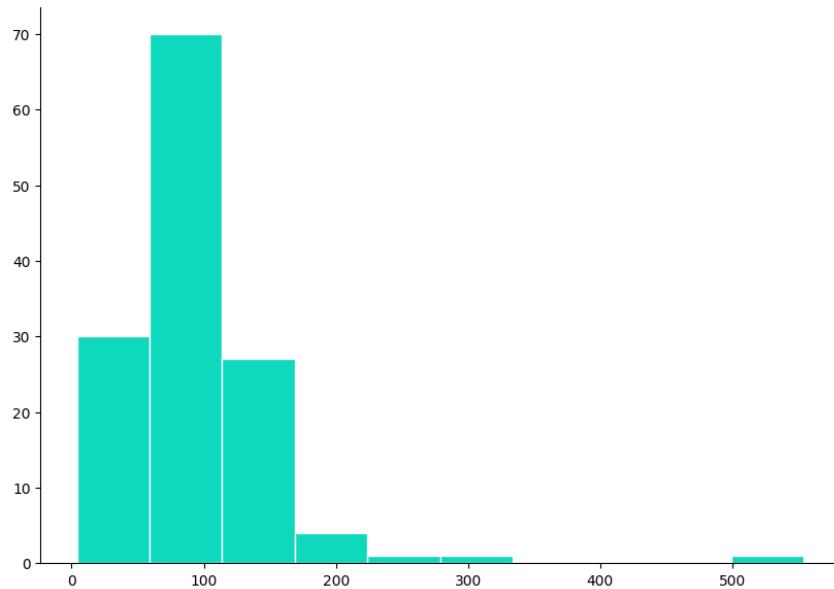


FIGURE 3.8 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.1.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2019-09**.

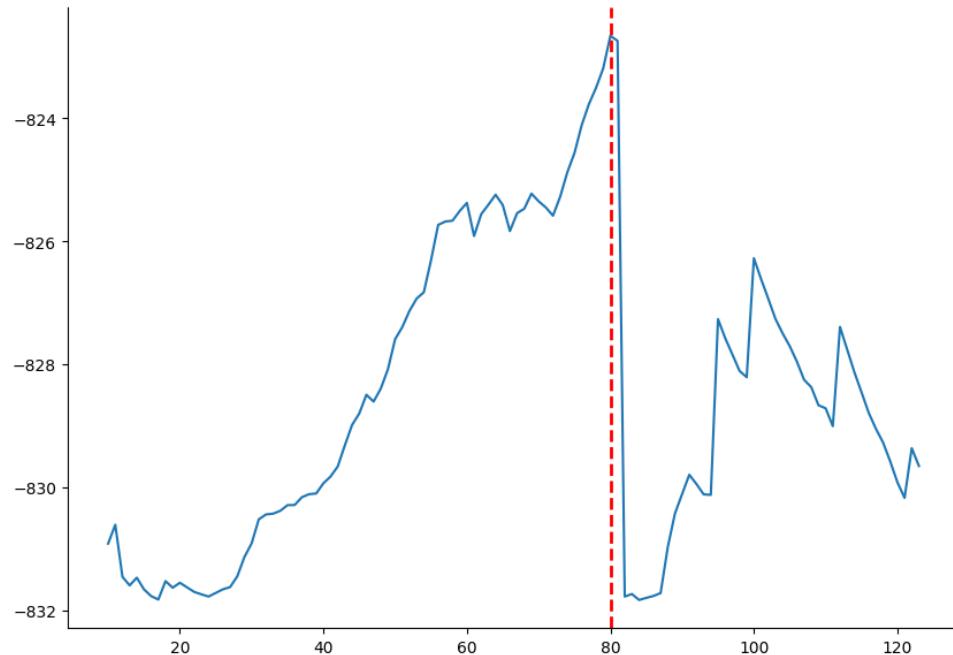


FIGURE 3.9 : Bretagne : Recherche du moment de rupture optimal.

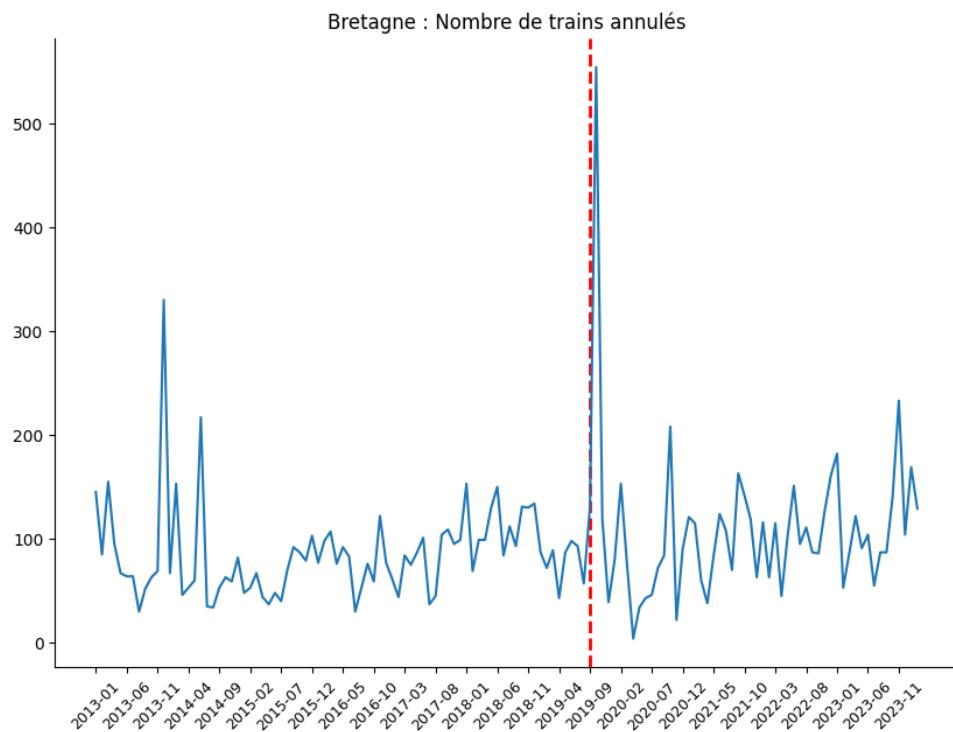


FIGURE 3.10 : Bretagne : La date de reupture.

3.1.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 0.01** : nous rejettons H_0 , ainsi le point de rupture est pertinent.

3.1.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (35.10, 26.35, 66.08, 113.60, 7.86)$$

Nous constatons que la moyenne du nombre de trains annulés avant le point de rupture ($\hat{\mu}_1^k$) est supérieure à celle après le point de rupture ($\hat{\mu}_2^k$), ce qui indique une réduction du nombre moyen de trains annulés post-rupture. De plus, l'augmentation significative de la variance après le point de rupture ($\hat{\sigma}_2^k$ comparée à $\hat{\sigma}_1^k$) suggère une plus grande dispersion dans le nombre de trains annulés. Ces observations sont cohérentes avec les tendances visualisées dans la Figure 3.10.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.5 et 3.6*)

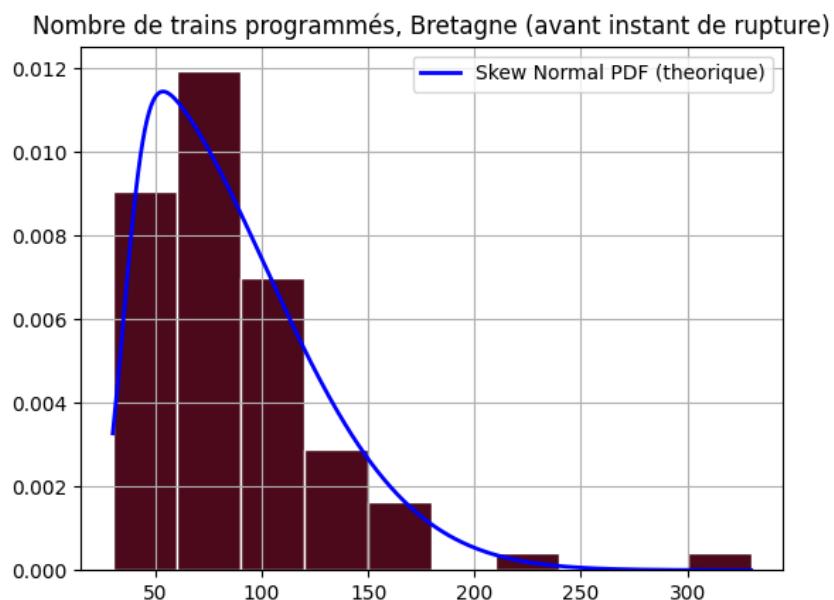


FIGURE 3.11 : Histogramme de l'échantillon avant point de rupture

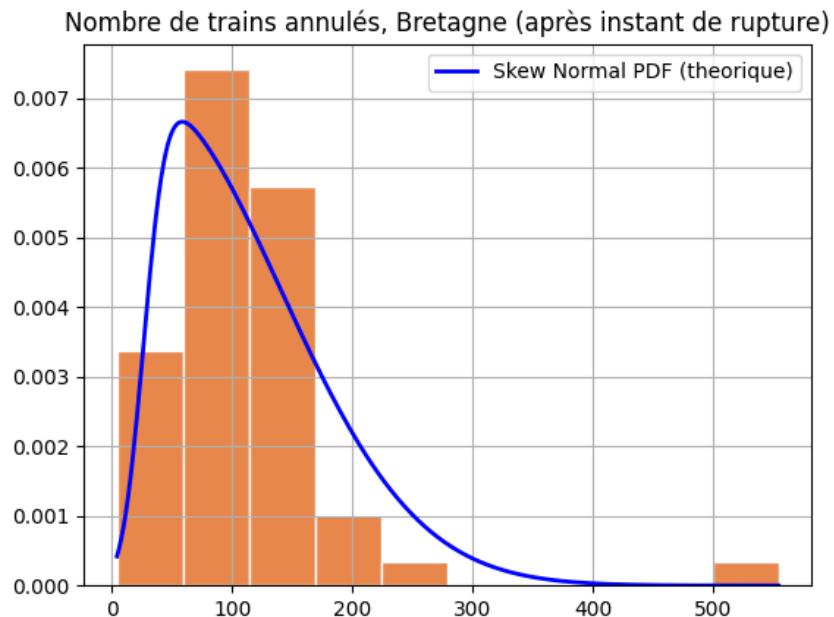


FIGURE 3.12 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ positif, ce qui suggère des lois normales asymétriques avec une queue de distribution plus étendue vers la droite. Cette caractéristique est cohérente avec les histogrammes présentés pour l'échantillon avant et après le point de rupture. En effet, les histogrammes indiquent une asymétrie où la queue droite de la distribution s'étend au-delà de la moyenne plus que la queue gauche, conformément à une asymétrie positive.

3.2 Pays-de-la-Loire

3.2.1 Trains programmés

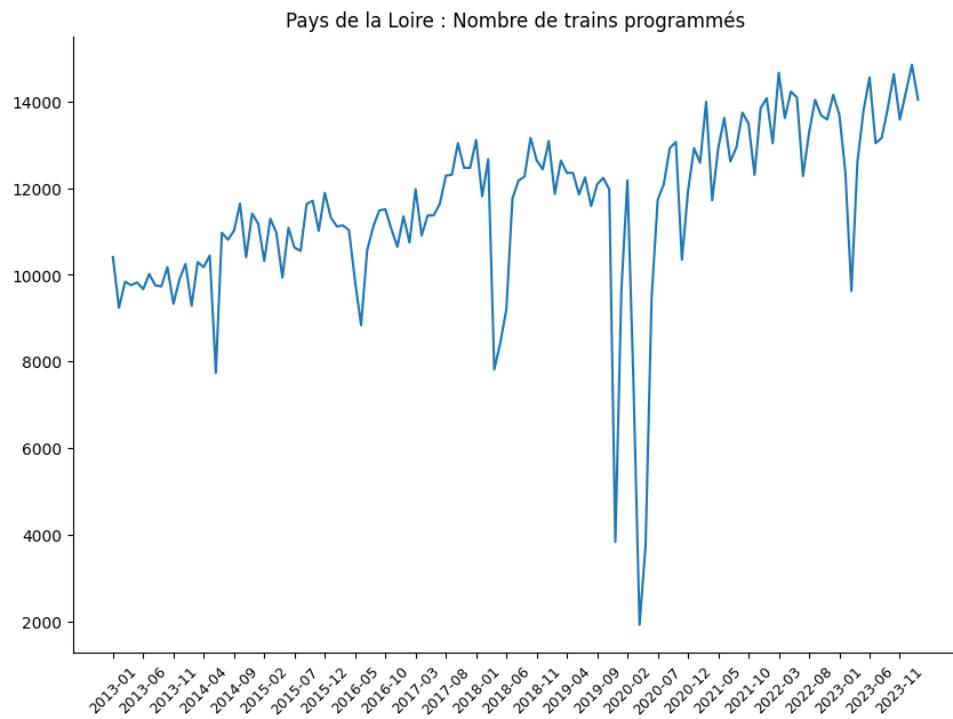


FIGURE 3.13 : Nombre de trains programmés de 2013 à 2024.

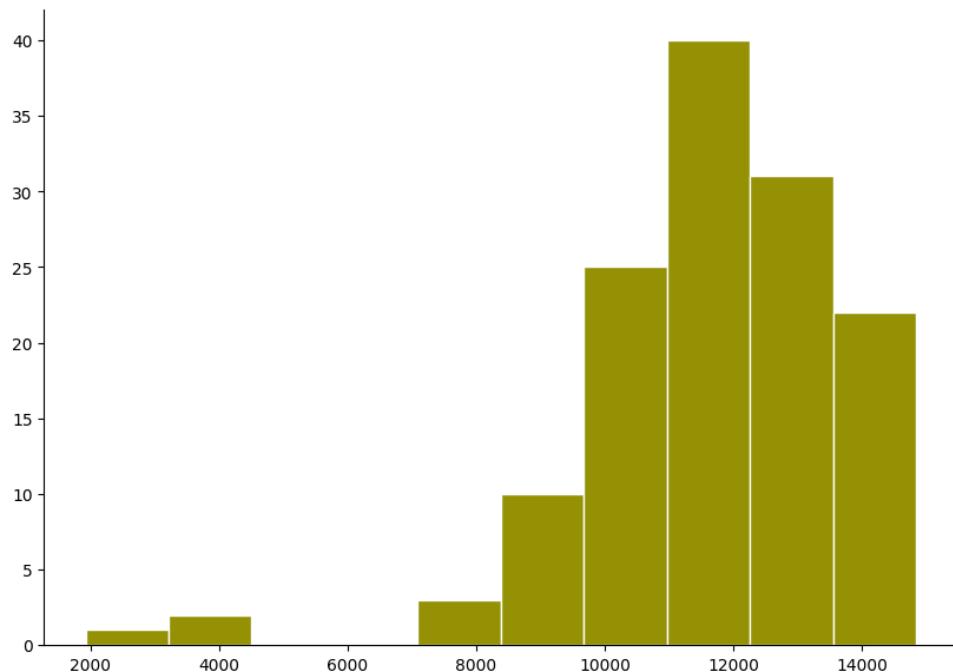


FIGURE 3.14 : Régions sélectionnées.

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.2.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2017-05**.

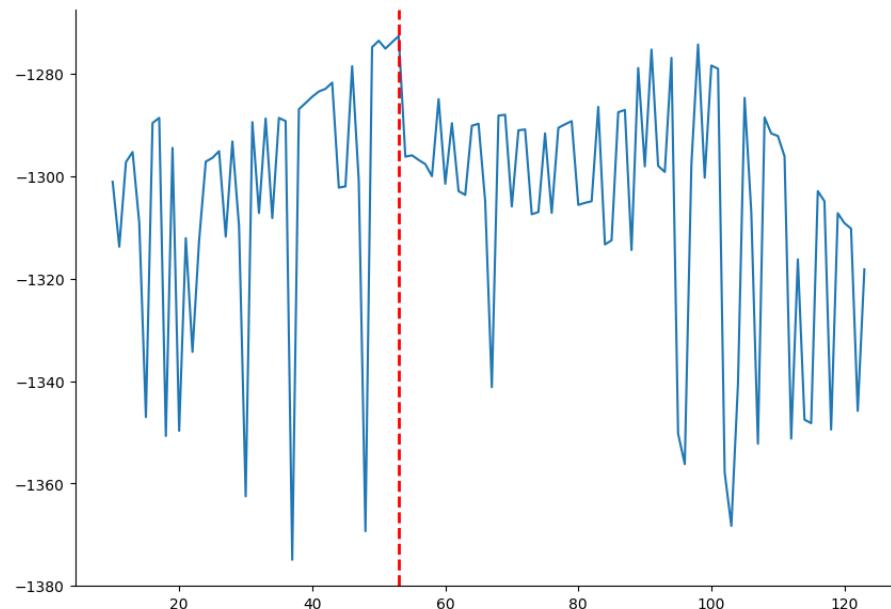


FIGURE 3.15 : Recherche du moment de rupture optimal.

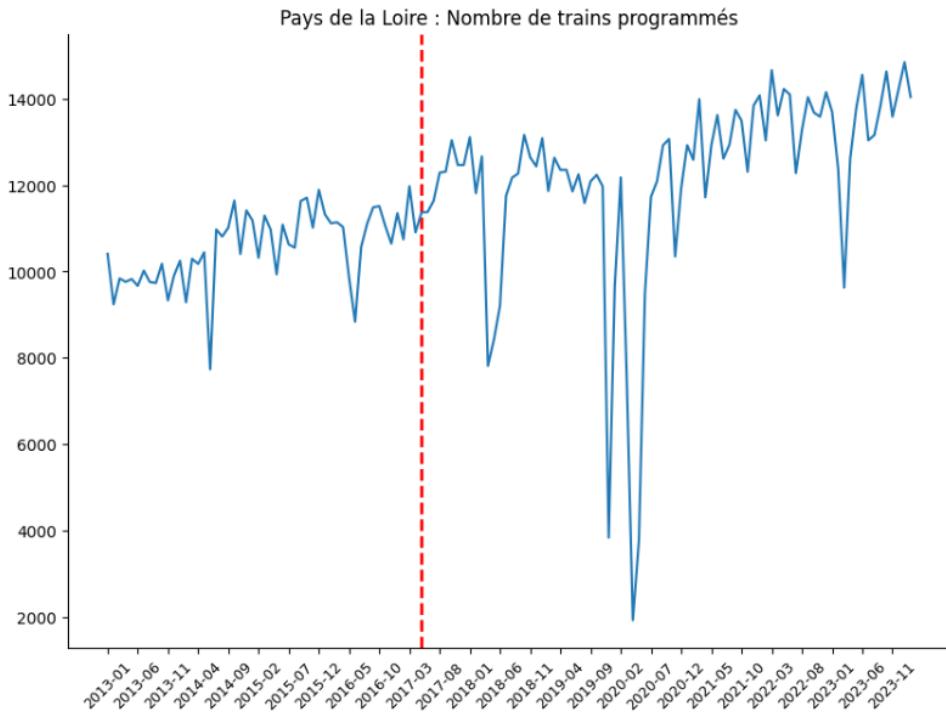


FIGURE 3.16 : La date de rupture.

3.2.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value = 1.42e-20** : nous rejetons H_0 .

3.2.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (11737.96, 14503.91, 1459.96, 3231.32, -7.89)$$

Nous observons que la moyenne du nombre de trains programmés dans les Pays de la Loire avant le point de rupture ($\hat{\mu}_1^k$) est moins élevée que celle après le point de rupture ($\hat{\mu}_2^k$), révélant une augmentation notable du nombre moyen de trains programmés après la rupture. Parallèlement, la variance après le point de rupture ($\hat{\sigma}_2^k$) dépasse significativement celle d'avant la rupture ($\hat{\sigma}_1^k$), ce qui indique une plus grande hétérogénéité dans la programmation des trains suite à cette date.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.17 et 3.18*)

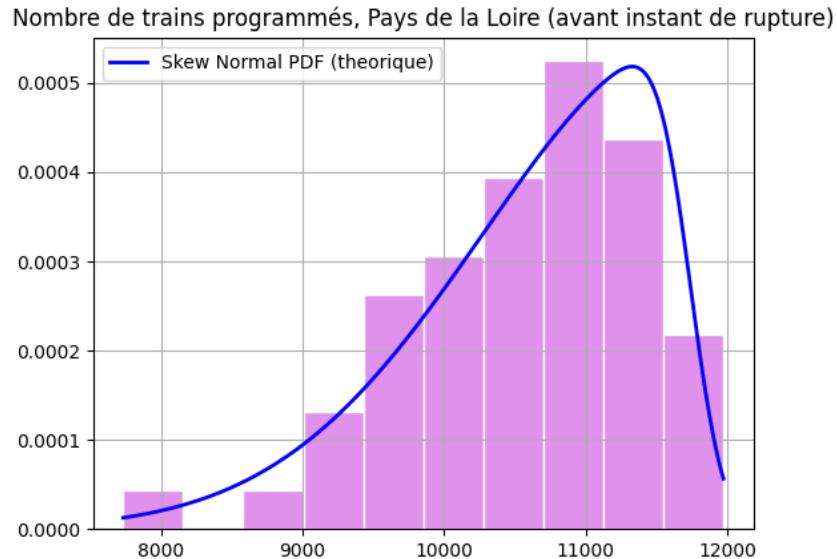


FIGURE 3.17 : Histogramme de l'échantillon avant point de rupture

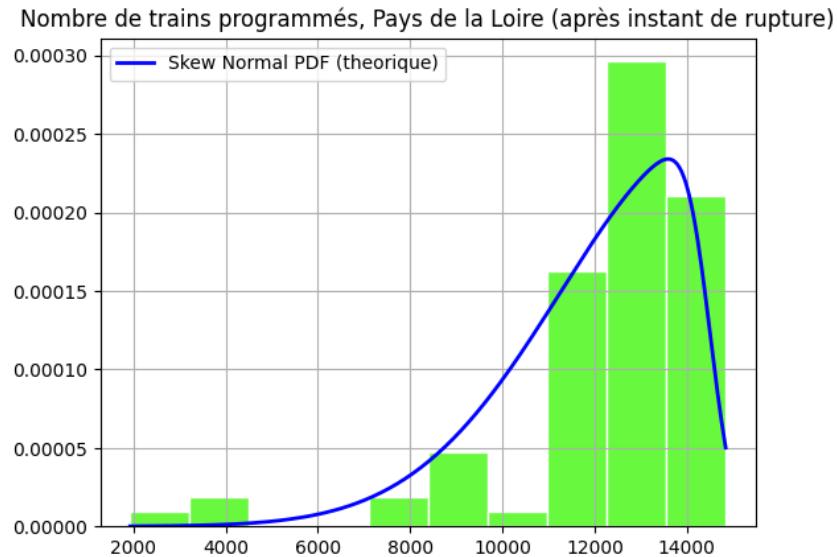


FIGURE 3.18 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ négatif, suggérant des distributions normales asymétriques inclinées vers la droite, impliquant ainsi une queue de distribution plus longue du côté gauche. Cette asymétrie est validée par les histogrammes présentés, où l'on constate une concentration accrue des valeurs à droite de la moyenne, spécialement dans les données après le point de rupture. Cette tendance indique que dans les Pays de la Loire, les trains programmés tendent à présenter une distribution avec une incidence plus marquée de valeurs extrêmement élevées suite au point de rupture, comme le montrent

les Figures 3.17 et 3.18.

3.2.2 Trains annulés

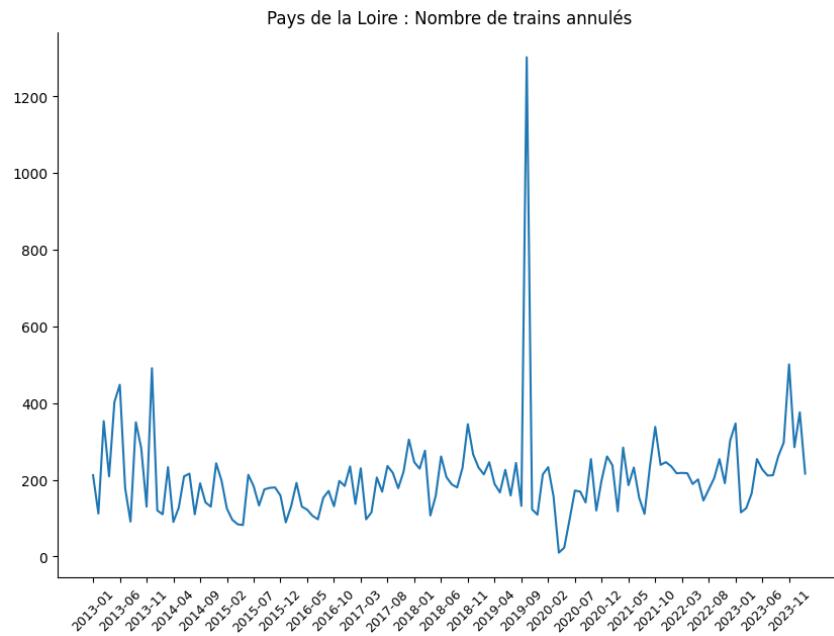


FIGURE 3.19 : Nombre de trains annulés entre 01/2013 et 02/2024

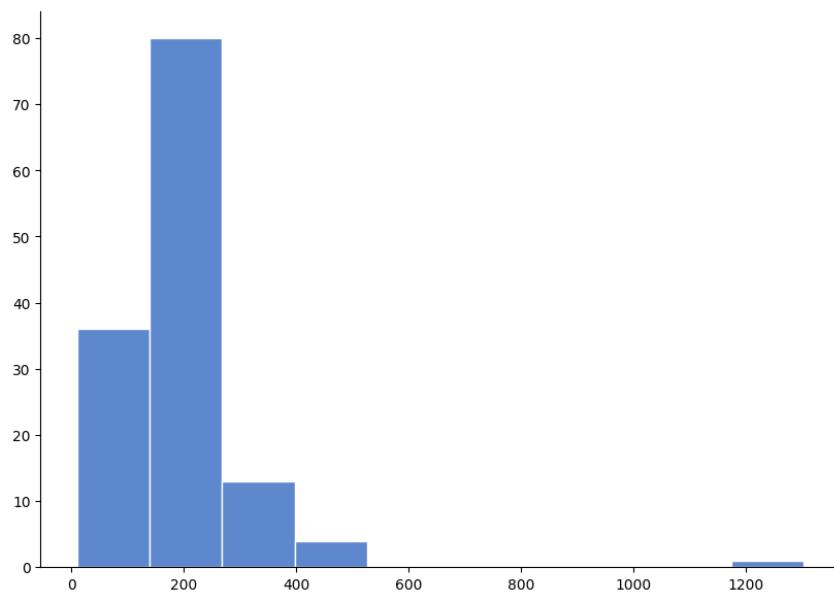


FIGURE 3.20 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.2.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2019-09**.

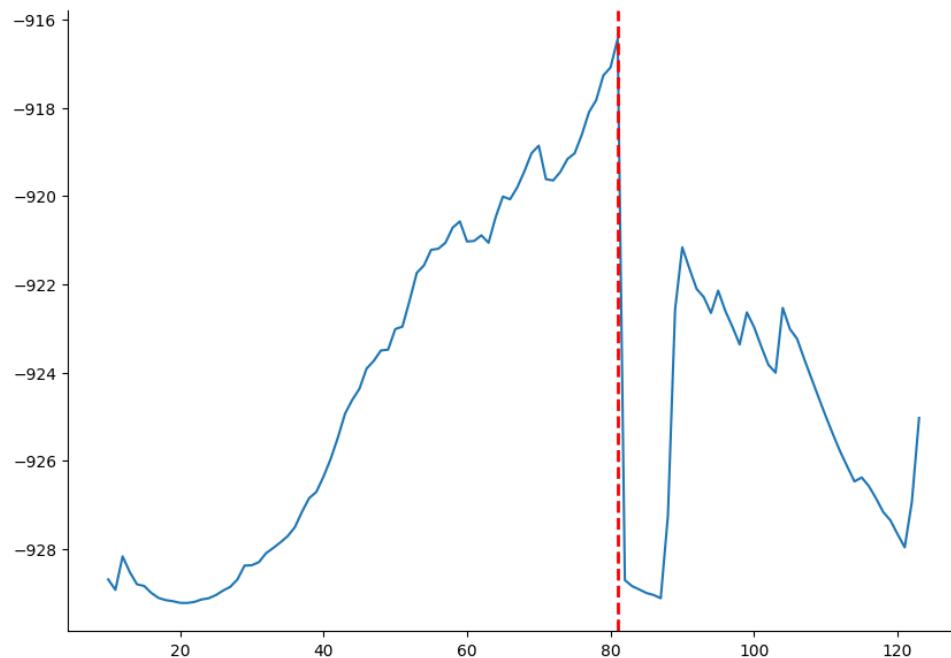


FIGURE 3.21 : Recherche du moment de rupture optimal.

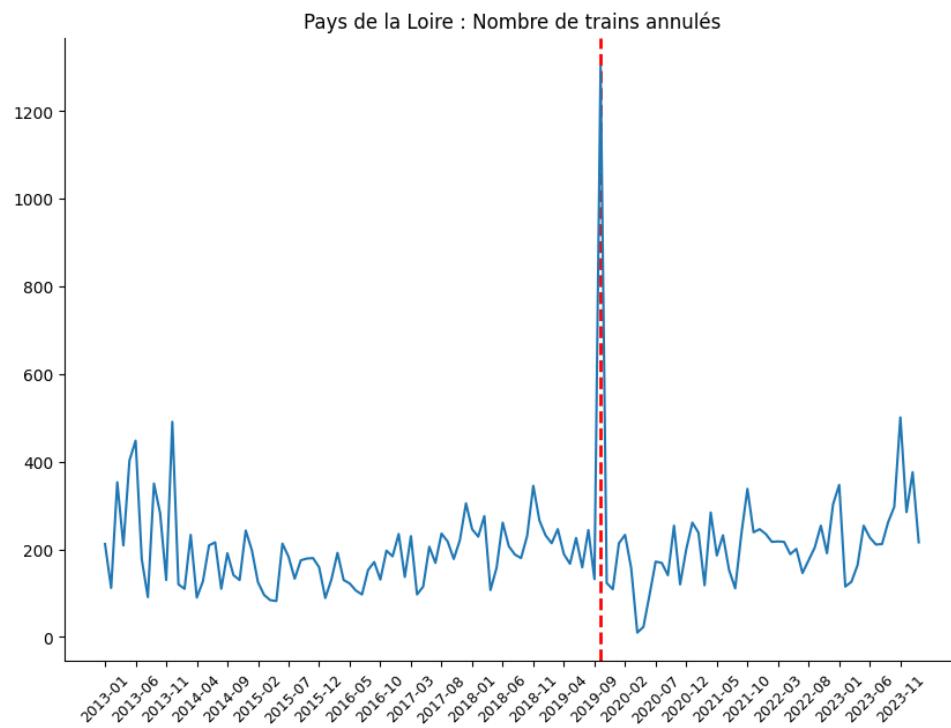


FIGURE 3.22 : La date de reupture.

3.2.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 0.17** : nous ne pouvons pas rejeter H_0 . Le point de rupture ne semble pas pertinent.

3.2.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (92.92, 68.23, 125.48, 239.38, 6.28)$$

Nous constatons que la moyenne du nombre de trains annulés dans les Pays de la Loire avant le point de rupture ($\hat{\mu}_1^k$) est plus élevée que celle après le point de rupture ($\hat{\mu}_2^k$), révélant ainsi une diminution du nombre moyen de trains annulés suite à la rupture. En outre, l'augmentation de la variance post-rupture ($\hat{\sigma}_2^k$ par rapport à $\hat{\sigma}_1^k$) indique une variabilité plus importante dans le nombre de trains annulés après le point de rupture. Le paramètre de skewness positif ($\hat{\theta}^k$) suggère une distribution asymétrique avec une queue plus prononcée vers la droite. Cette distribution est visuellement corroborée par la Figure 3.22, qui montre un changement significatif dans la distribution des annulations de trains à la date indiquée par le trait rouge en pointillé.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.23 et 3.24*)

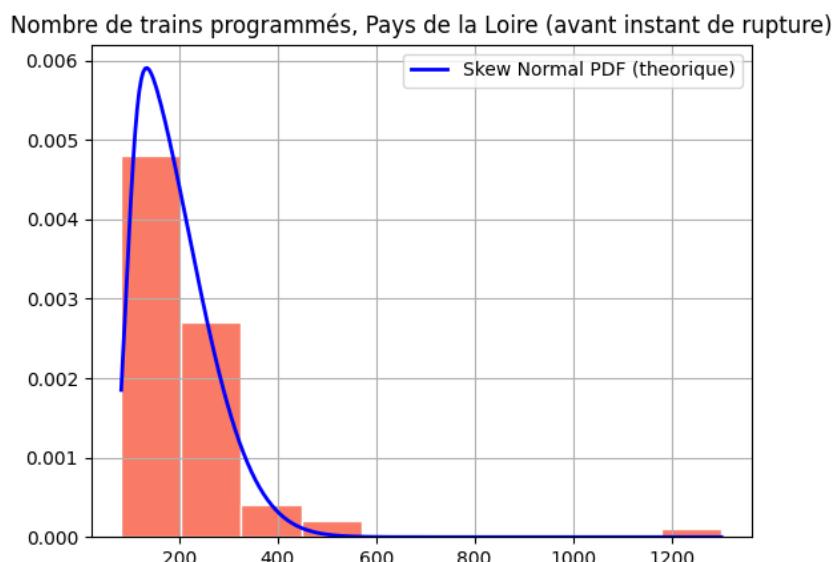


FIGURE 3.23 : Histogramme de l'échantillon avant point de rupture

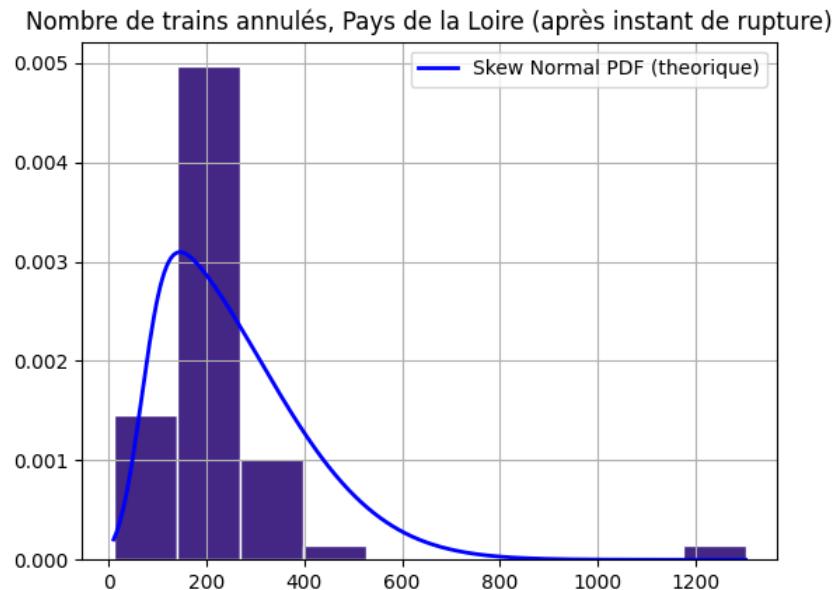


FIGURE 3.24 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ positif, indiquant des distributions normales asymétriques avec une queue plus longue s'étendant vers la droite. Cette caractéristique se vérifie dans les histogrammes pour l'échantillon avant et après le point de rupture, où l'on constate une concentration des valeurs dépassant la moyenne sur le côté droit de la distribution.

3.3 Centre

3.3.1 Trains programmés

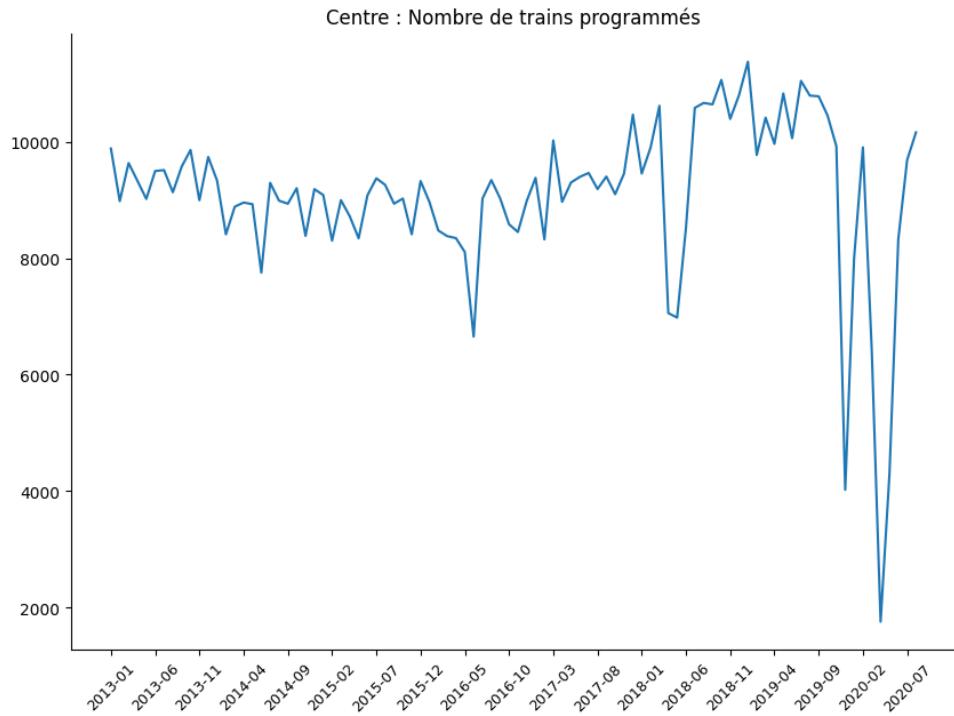


FIGURE 3.25 : Nombre de trains programmés de 2013 à 2024.

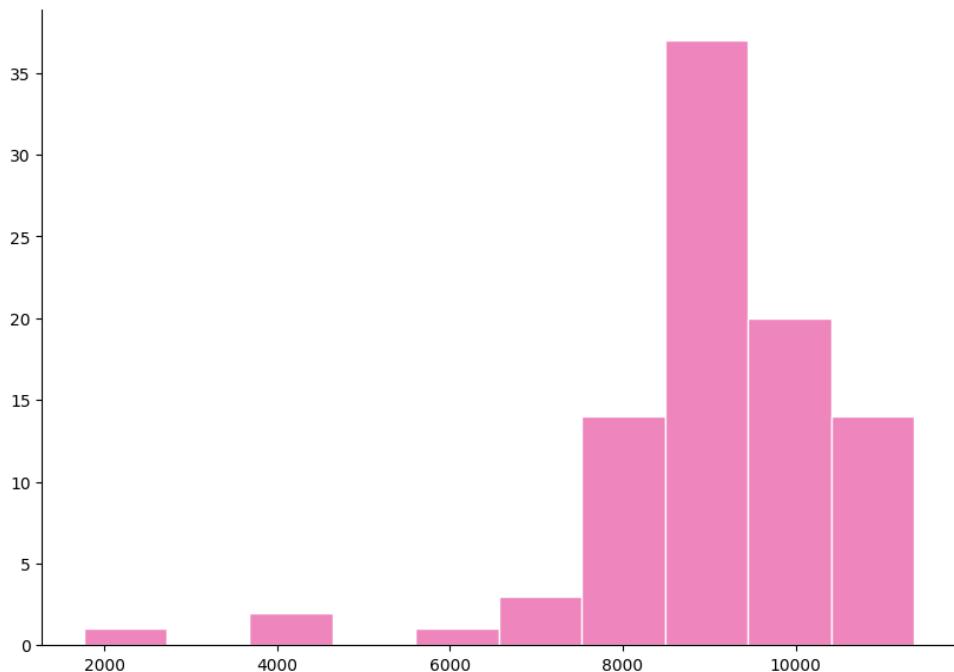


FIGURE 3.26 : Régions sélectionnées.

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.3.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2017-10.**

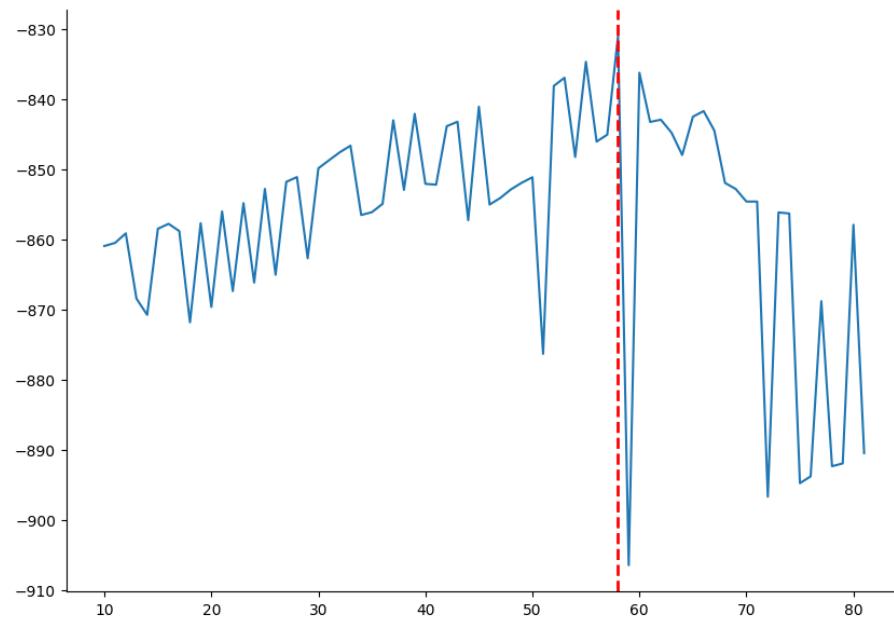


FIGURE 3.27 : Recherche du moment de rupture optimal.

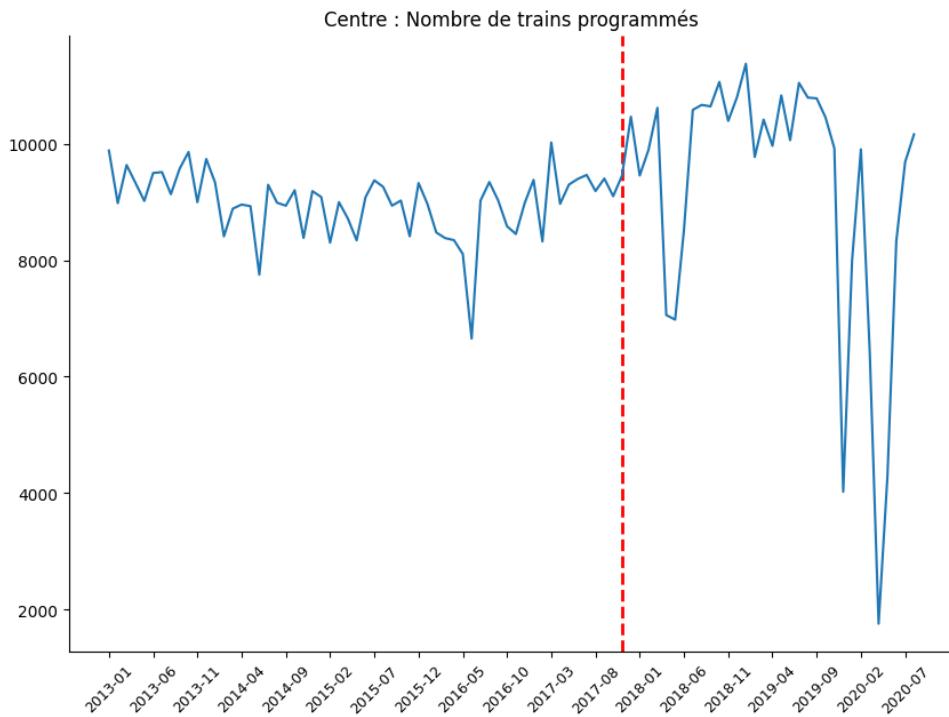


FIGURE 3.28 : La date de reupture.

3.3.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 4.651e-08** : nous rejetons H_0 .

3.3.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (9.68897541e+03 \ 1.13077069e+04 \ 9.30181789e+02 \ 2.86048081e+03 - 4.93819171e+00)$$

Nous observons que la moyenne du nombre de trains programmés dans la région Centre avant le point de rupture ($\hat{\mu}_1^k$) est inférieure à celle après le point de rupture ($\hat{\mu}_2^k$), suggérant une augmentation significative du nombre moyen de trains programmés post-rupture. De même, la variance post-rupture ($\hat{\sigma}_2^k$) est nettement plus élevée que la variance pré-rupture ($\hat{\sigma}_1^k$), indiquant une diversité accrue dans le nombre de trains programmés après cette date. Le skewness négatif ($\hat{\theta}^k$) reflète une asymétrie de la distribution, avec une tendance à des valeurs extrêmement basses plus fréquentes. Ces observations sont cohérentes avec le changement marqué illustré dans la Figure 3.28, où la ligne rouge en pointillé indique le moment de la rupture.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.29 et 3.30*)

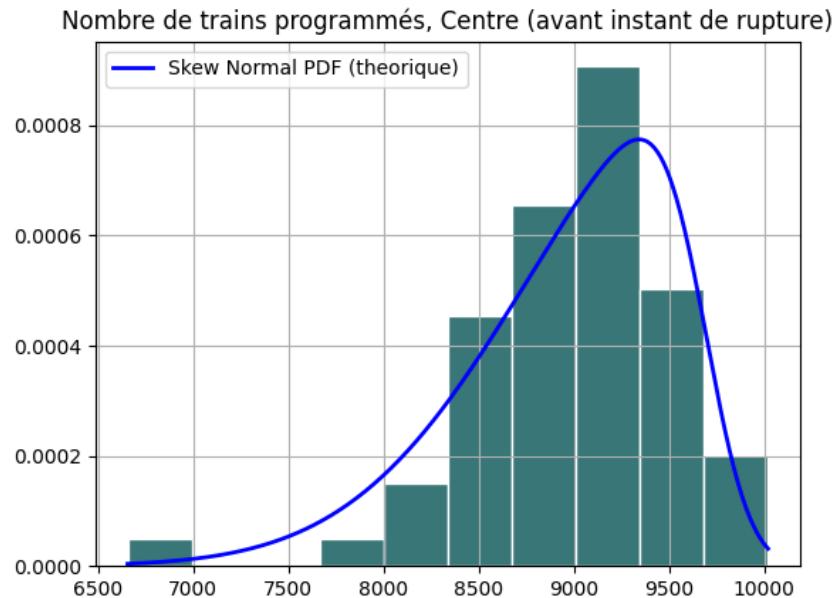


FIGURE 3.29 : Histogramme de l'échantillon avant point de rupture

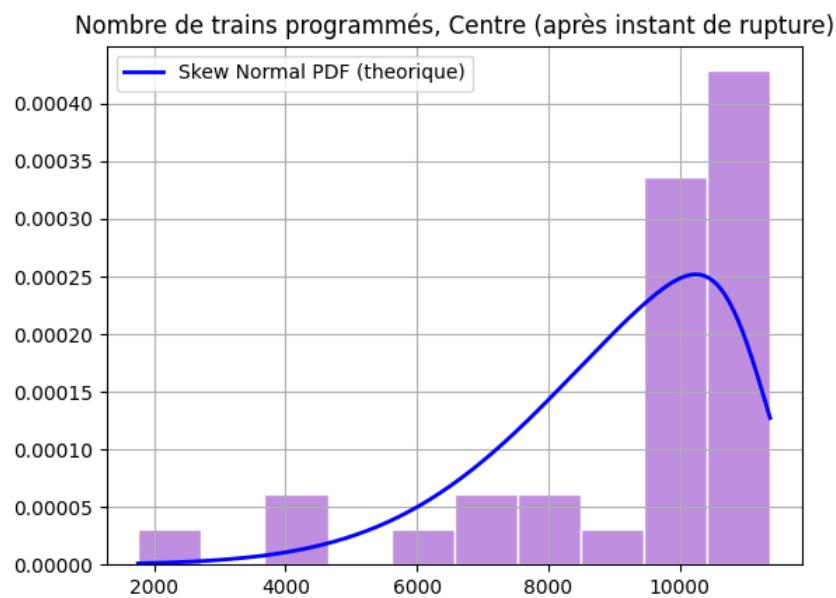


FIGURE 3.30 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ négatif, ce qui indique des distributions normales asymétriques inclinées vers la gauche, résultant en une queue de distribution allongée du côté droit. Cette asymétrie se vérifie dans les histogrammes, où nous observons une augmentation des valeurs supérieures à la moyenne, en particulier après le point de rupture.

Cette observation suggère que dans la région Centre, la distribution des trains programmés post-rupture est marquée par une fréquence accrue de valeurs élevées, comme l'illustrent les Figures 3.29 et 3.30.

3.3.2 Trains annulés

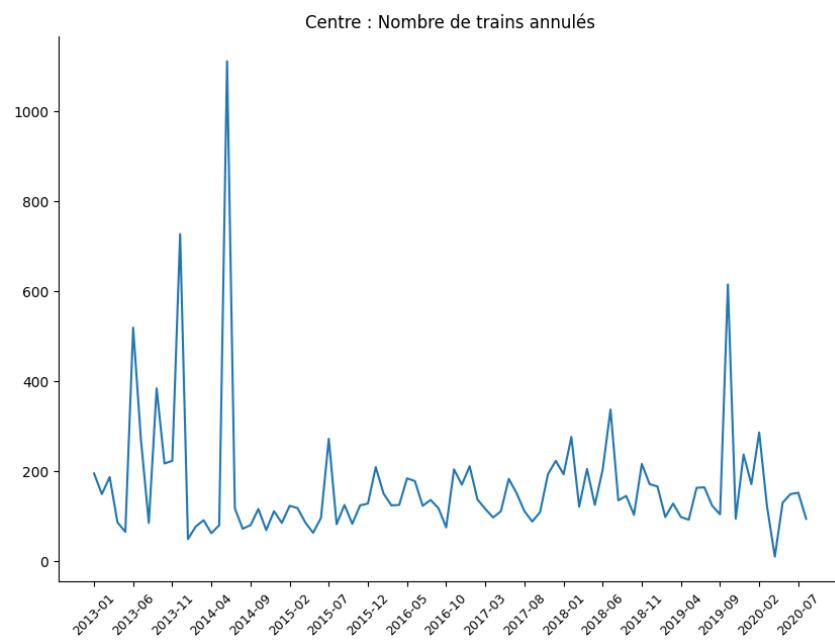


FIGURE 3.31 : Nombre de trains annulés entre 01/2013 et 02/2024

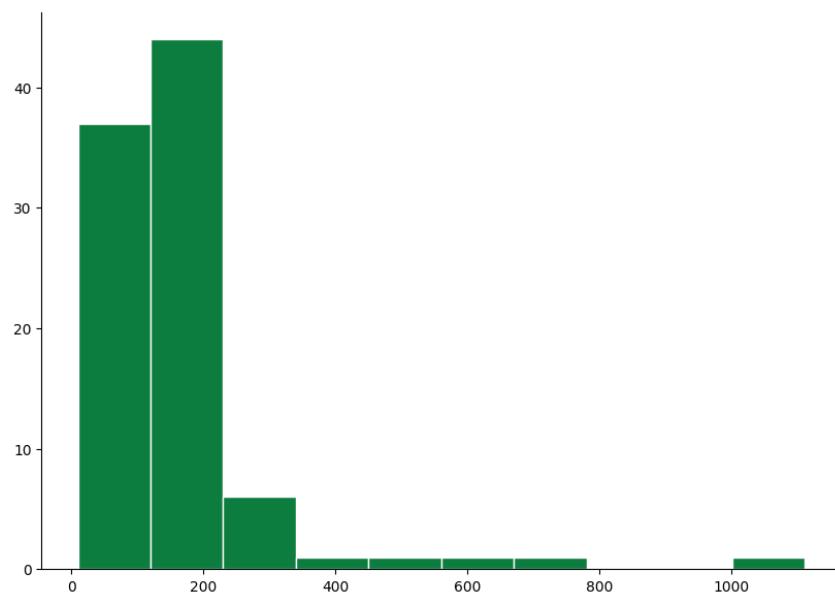


FIGURE 3.32 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de detection de point de rupture et nous obtenons les résultats suivants :

3.3.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2014-06**.

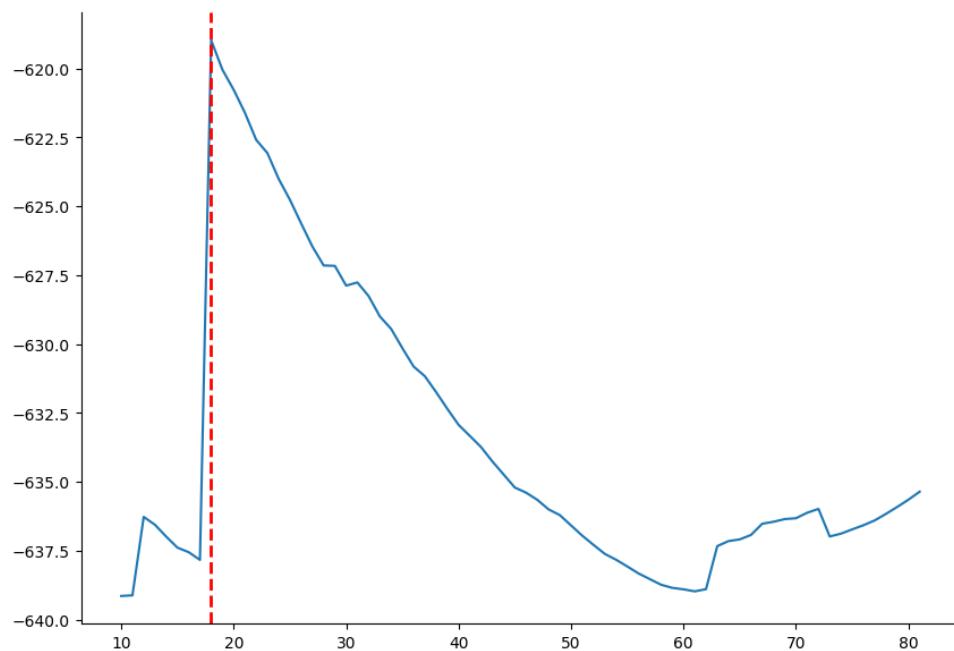


FIGURE 3.33 : Recherche du moment de rupture optimal.

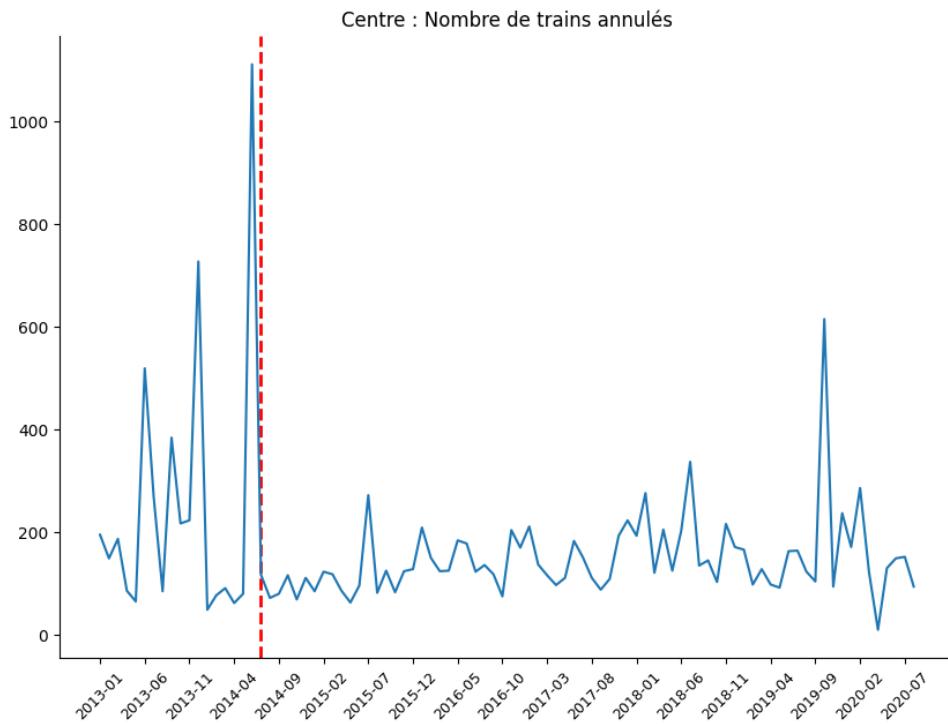


FIGURE 3.34 : La date de reupture.

3.3.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 0.17** : nous ne pouvons pas rejeter H_0 . Le point de rupture ne semble pas pertinent.

3.3.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (9.66, 66.37, 343.50, 114.80, 5.30)$$

Nous observons que la moyenne du nombre de trains annulés dans la région Centre avant le point de rupture ($\hat{\mu}_1^k$) est inférieure à celle observée après le point de rupture ($\hat{\mu}_2^k$), suggérant une augmentation du nombre moyen de trains annulés suite à la rupture. De plus, l'accroissement de la variance après le point de rupture ($\hat{\sigma}_2^k$ comparativement à $\hat{\sigma}_1^k$) indique une plus grande dispersion dans le nombre de trains annulés post-rupture. Le paramètre de skewness positif ($\hat{\theta}^k$) révèle une asymétrie de la distribution, avec une queue s'étendant davantage vers la droite. Cette asymétrie est confirmée visuellement par la Figure 3.34, où le changement notable dans la distribution des annulations de trains

est mis en évidence par la ligne rouge en pointillé indiquant le moment de la rupture.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.35 et 3.36*)

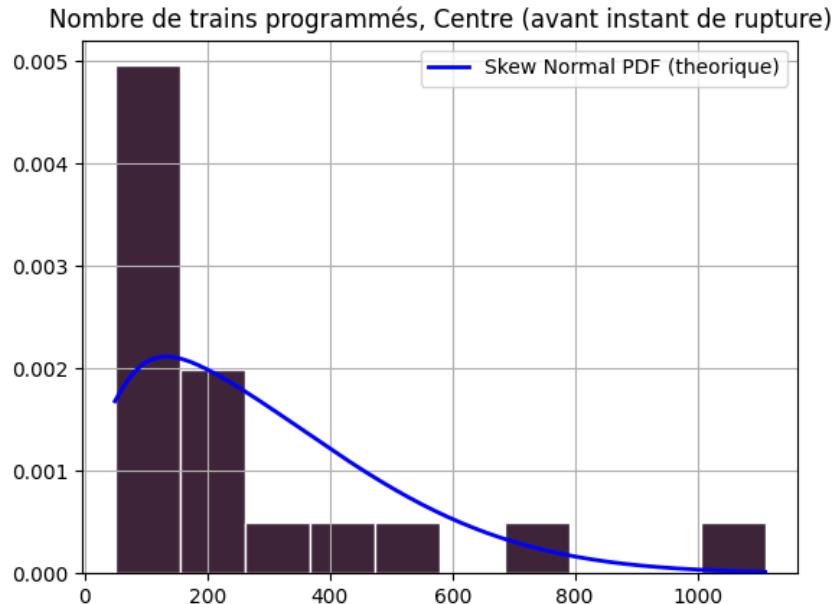


FIGURE 3.35 : Histogramme de l'échantillon avant point de rupture

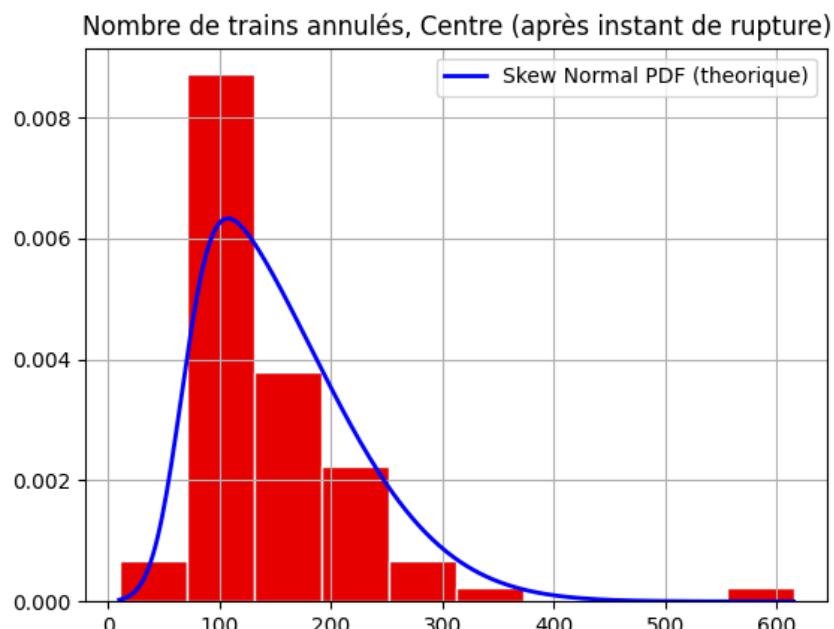


FIGURE 3.36 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ positif, indiquant des distributions normales asymétriques avec une queue de distribution s'étendant plus longuement vers la droite. Cette

caractéristique est confirmée par les histogrammes de la région Centre, tant pour l'échantillon avant qu'après le point de rupture. Les données montrent une prédominance des valeurs excédant la moyenne du côté droit, ce qui est caractéristique d'une skewness positive. Cette asymétrie dans la distribution du nombre de trains annulés est illustrée visuellement par les Figures 3.35 et 3.36, soulignant la présence d'une distribution inclinée post-rupture dans la région étudiée.

3.4 Aquitaine

3.4.1 Trains programmés

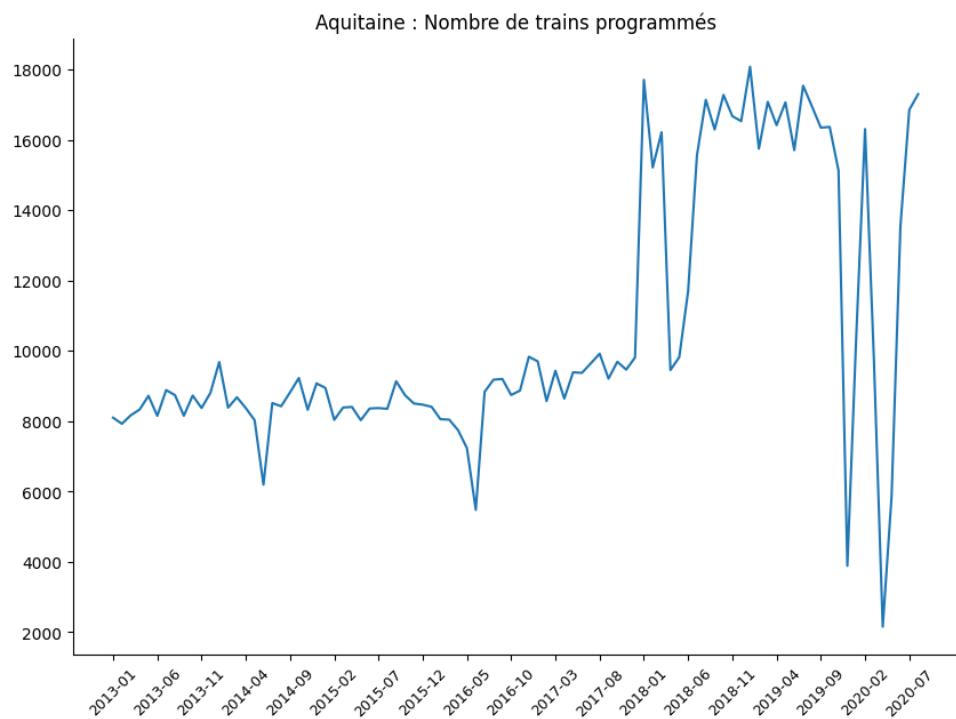


FIGURE 3.37 : Nombre de trains programmés de 2013 à 2024.

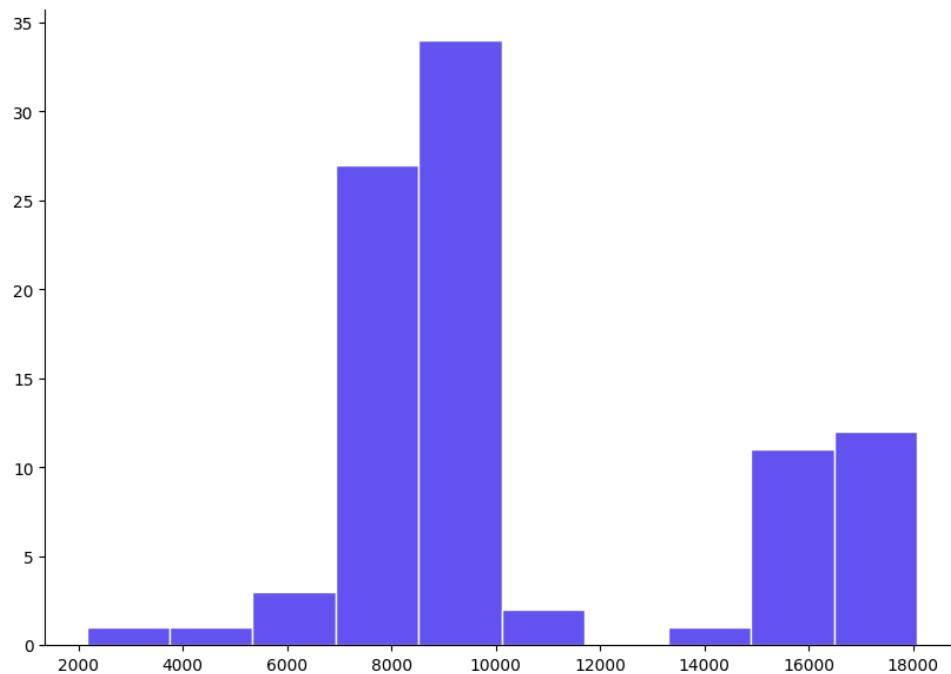


FIGURE 3.38 : Régions selectionnées.

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.4.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2017-11**.

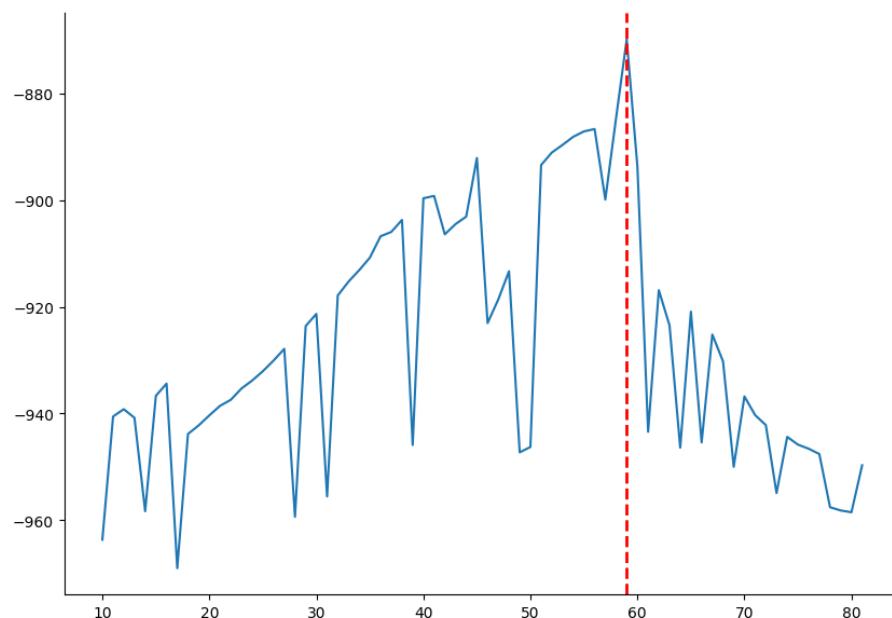


FIGURE 3.39 : Recherche du moment de rupture optimal.

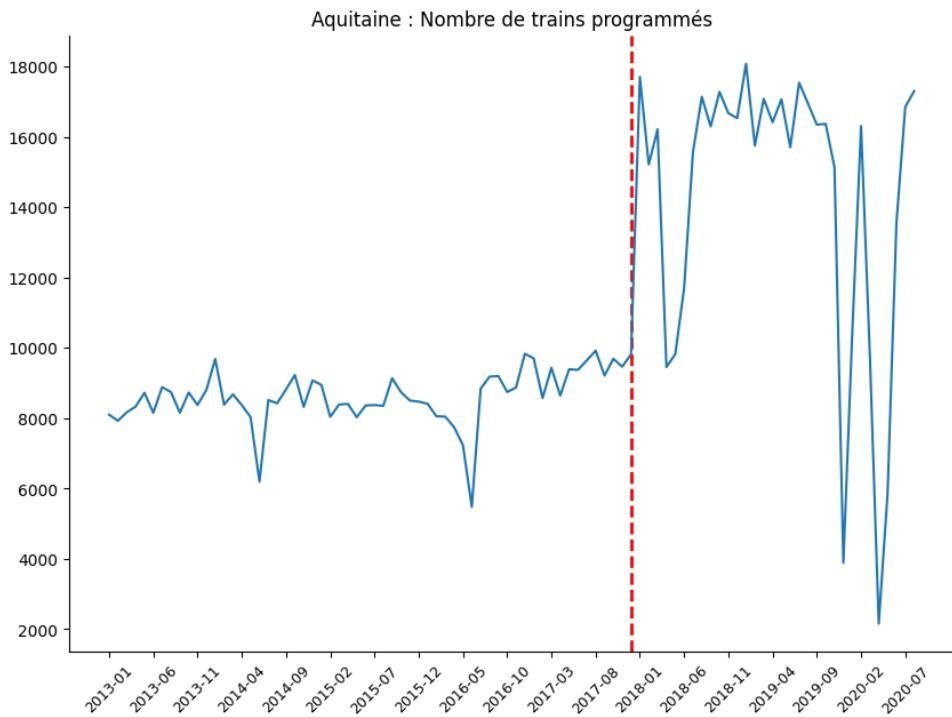


FIGURE 3.40 : La date de reupture.

3.4.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 1.5789521915581405e-15** : nous rejetons H_0 .

3.4.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (9.79476421e+03 \ 1.80417857e+04 \ 1.43618833e+03 \ 5.57450614e+03 - 1.31829067e+01)$$

Pour votre figure de la région Aquitaine et les valeurs des paramètres statistiques données, voici le paragraphe corrigé :

Nous observons que la moyenne du nombre de trains programmés dans la région Aquitaine avant le point de rupture ($\hat{\mu}_1^k$) est considérablement inférieure à celle qui est relevée après le point de rupture ($\hat{\mu}_2^k$), indiquant ainsi une augmentation marquée du nombre moyen de trains programmés post-rupture. En outre, nous notons que la variance après le point de rupture ($\hat{\sigma}_2^k$) surpassé largement celle d'avant la rupture ($\hat{\sigma}_1^k$), ce qui signale une variabilité beaucoup plus grande dans la fréquence des trains programmés suite à cette date. Le skewness négatif très prononcé ($\hat{\theta}^k$) suggère une asymétrie de la distribution, avec

une prédominance de valeurs extrêmement basses. Ces observations sont illustrées par la Figure 3.40, où la ligne rouge en pointillé marque le point de rupture significatif dans la distribution du nombre de trains programmés.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.29 et 3.30*)

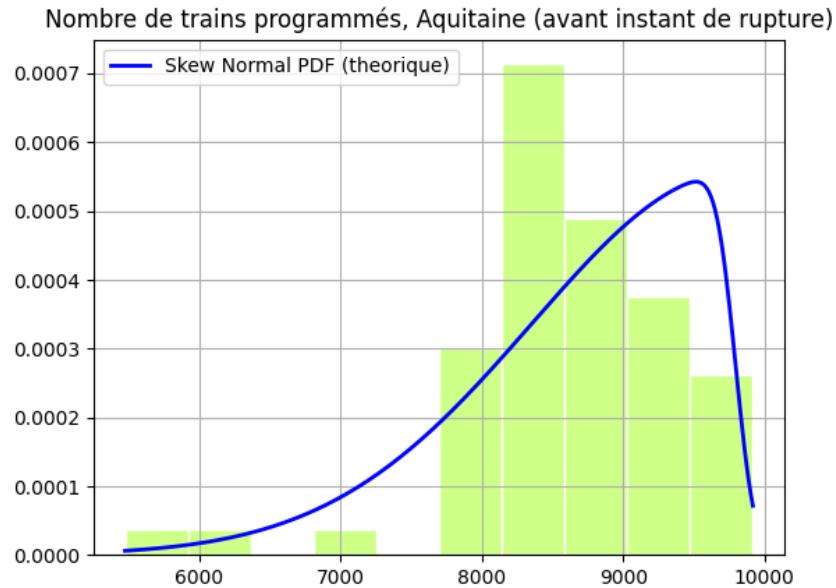


FIGURE 3.41 : Histogramme de l'échantillon avant point de rupture

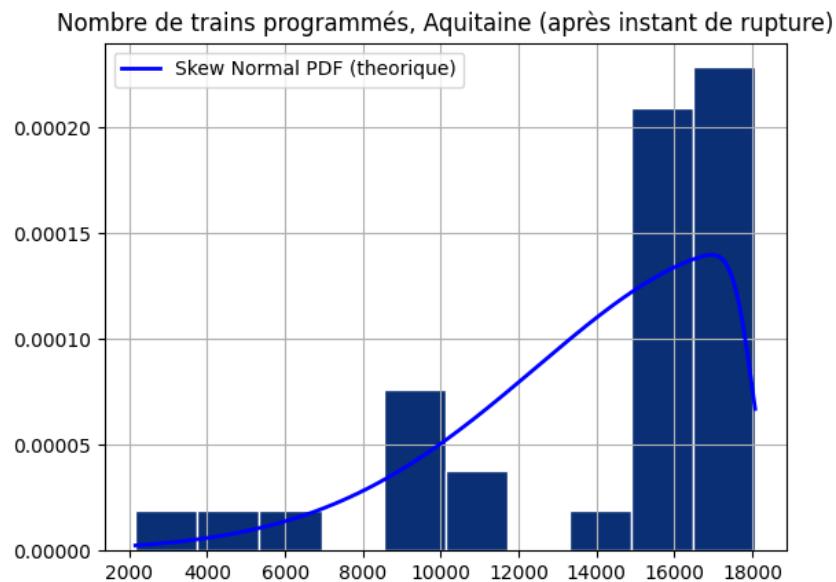


FIGURE 3.42 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ négatif, indiquant des distributions normales

asymétriques inclinées vers la gauche, ce qui se traduit par une queue de distribution prolongée du côté droit. Cette asymétrie est confirmée par les histogrammes de la région Aquitaine, qui montrent un nombre accru de valeurs excédant la moyenne, surtout après le point de rupture. Cela indique que la distribution des trains programmés dans la région Aquitaine après le point de rupture tend vers une fréquence plus élevée de valeurs supérieures, comme démontré par les Figures 3.41 et 3.42, illustrant le changement dans la distribution post-rupture.

3.4.2 Trains annulés

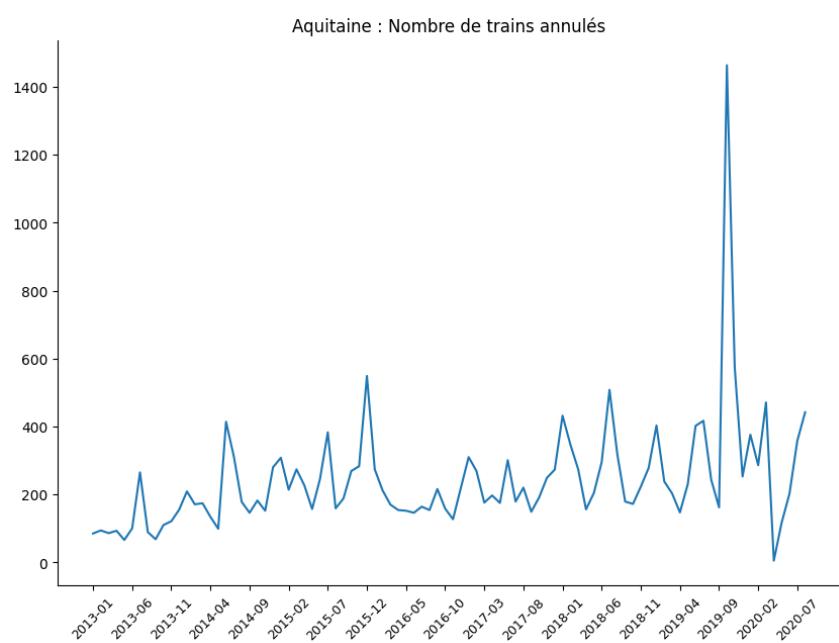


FIGURE 3.43 : Nombre de trains annulés entre 01/2013 et 02/2024

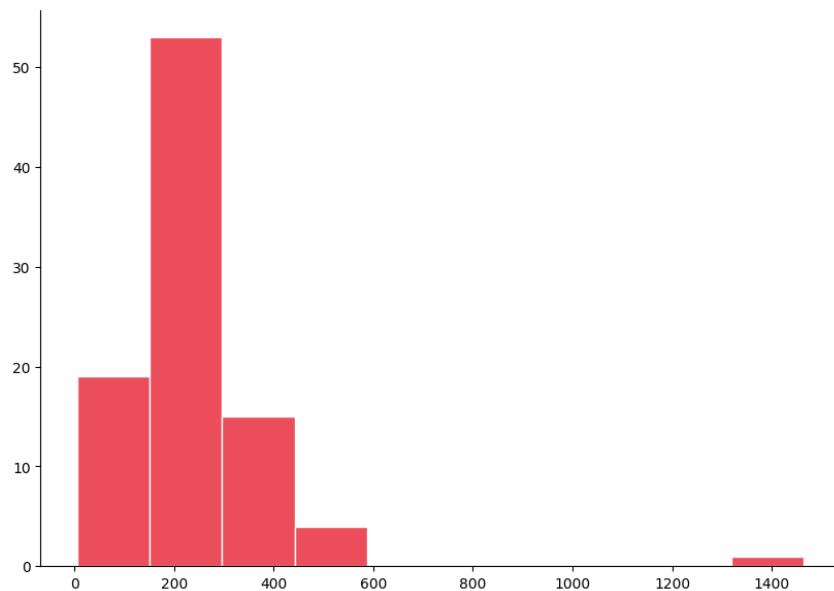


FIGURE 3.44 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.4.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2019-10**.

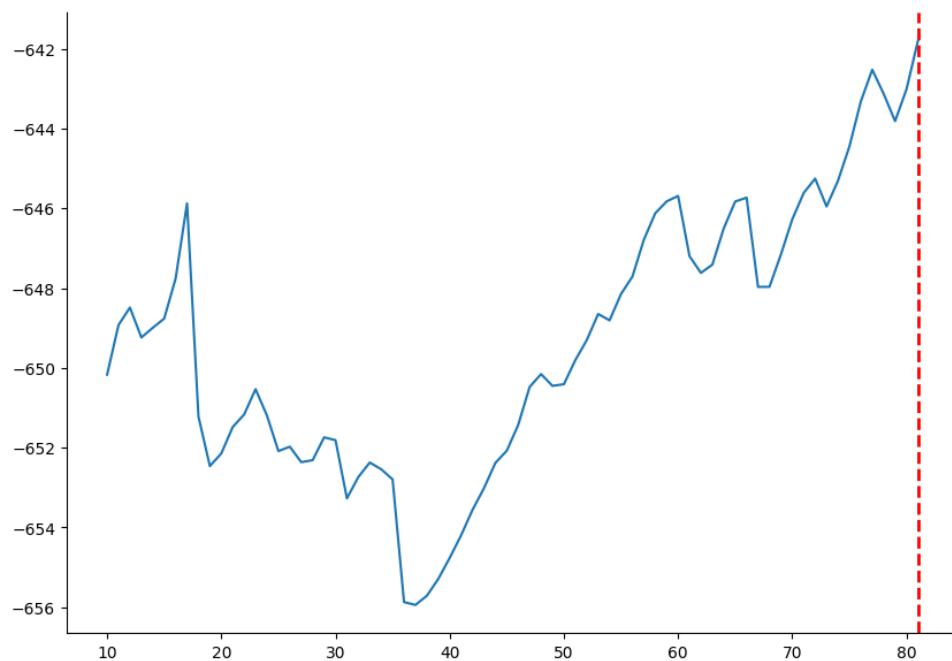


FIGURE 3.45 : Recherche du moment de rupture optimal.

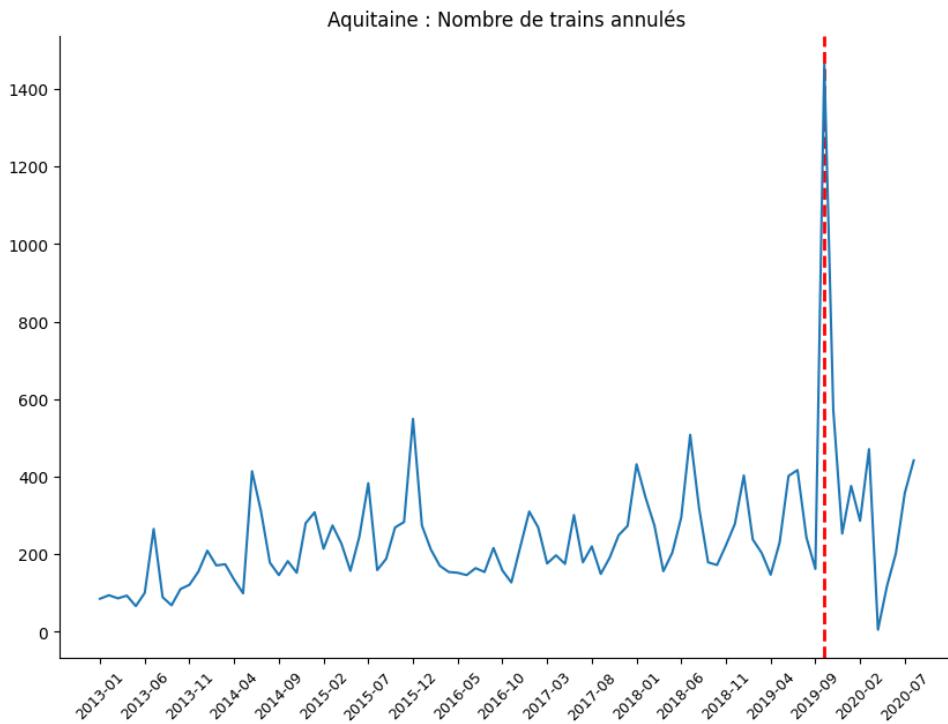


FIGURE 3.46 : La date de reupture.

3.4.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 0.03042677774966097** : nous rejettions H_0 .

3.4.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (98.60025567 \ 39.88619881 \ 155.60519799 \ 515.51660342 \ 4.87537057)$$

Nous observons que la moyenne du nombre de trains annulés dans la région Aquitaine avant le point de rupture ($\hat{\mu}_1^k$) est supérieure à celle observée après le point de rupture ($\hat{\mu}_2^k$), révélant ainsi une diminution du nombre moyen de trains annulés post-rupture. Par ailleurs, l'augmentation considérable de la variance après le point de rupture ($\hat{\sigma}_2^k$ en comparaison à $\hat{\sigma}_1^k$) indique une variabilité plus importante dans le nombre de trains annulés après cette date. Le paramètre de skewness positif ($\hat{\theta}^k$) suggère une asymétrie dans la distribution, avec une queue s'étendant plus significativement vers la droite. Cette asymétrie est illustrée par la Figure 3.46, où le trait rouge en pointillé marque un changement significatif dans la distribution des annulations de trains à l'instant de la rupture.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.35 et 3.36*)

Nombre de trains programmés, Aquitaine (avant instant de rupture)

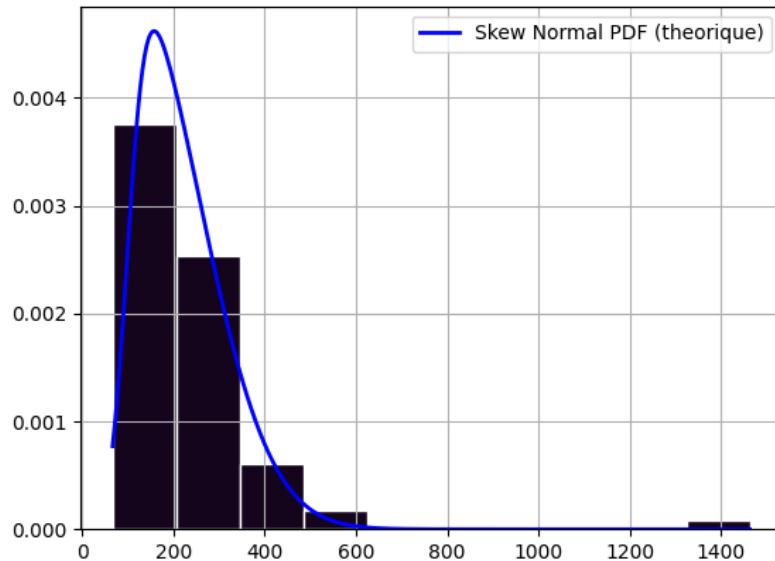


FIGURE 3.47 : Histogramme de l'échantillon avant point de rupture

Nombre de trains annulés, Aquitaine (après instant de rupture)

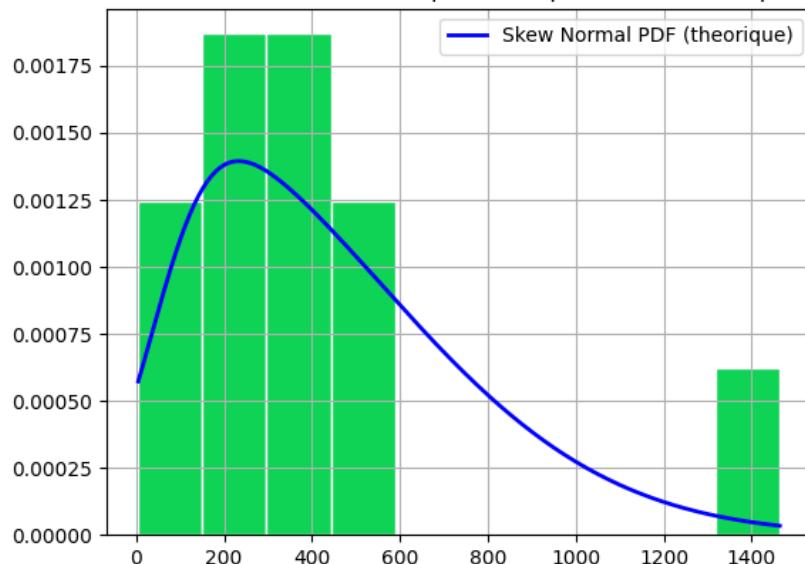


FIGURE 3.48 : Histogramme de l'échantillon après point de rupture

Nous avons obtenu un paramètre $\hat{\theta}^k$ positif, ce qui indique des distributions normales asymétriques avec une queue plus étendue vers la droite pour la région Aquitaine. Cette caractéristique est clairement observée dans les histogrammes, tant pour l'échantillon

avant qu'après le point de rupture. Les histogrammes révèlent une prévalence de valeurs excédant la moyenne sur le côté droit de la distribution, cohérente avec une skewness positive. Cette asymétrie dans la distribution du nombre de trains annulés est confirmée visuellement par les Figures 3.47 et 3.48, mettant en évidence une distribution qui penche davantage vers des valeurs supérieures après le point de rupture.

3.5 Provence Alpes Côte d'Azur

3.5.1 Trains programmés

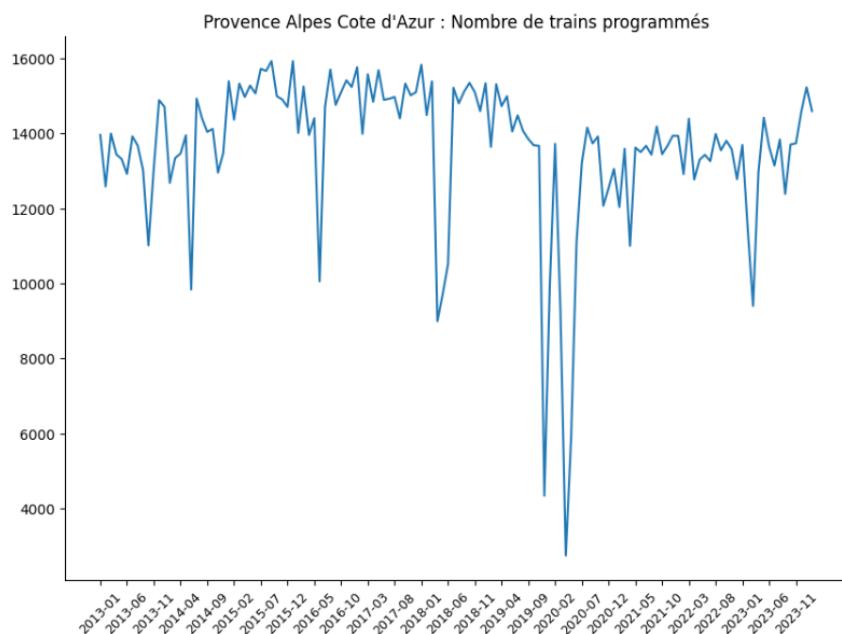


FIGURE 3.49 : Nombre de trains programmés entre 01/2013 et 02/2024

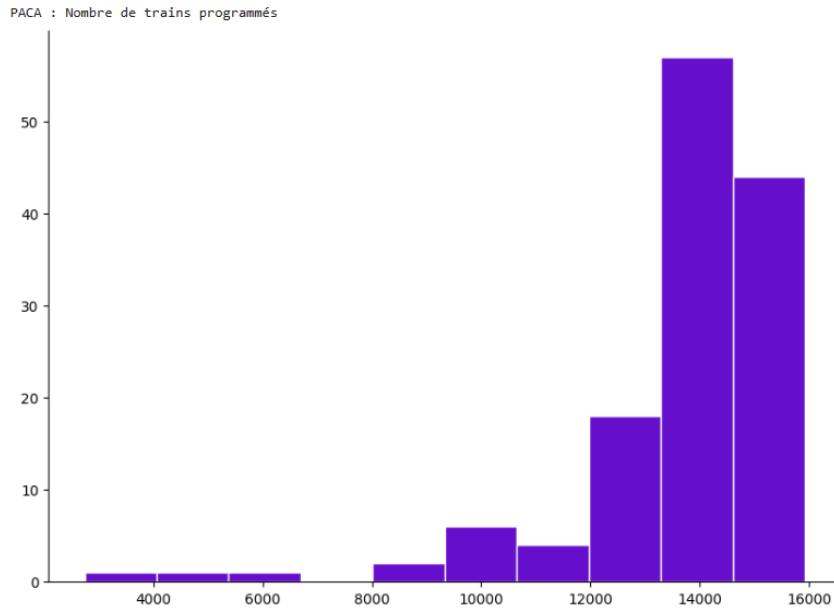


FIGURE 3.50 : Histogramme du nombre de trains programmés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.5.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2018-02**.

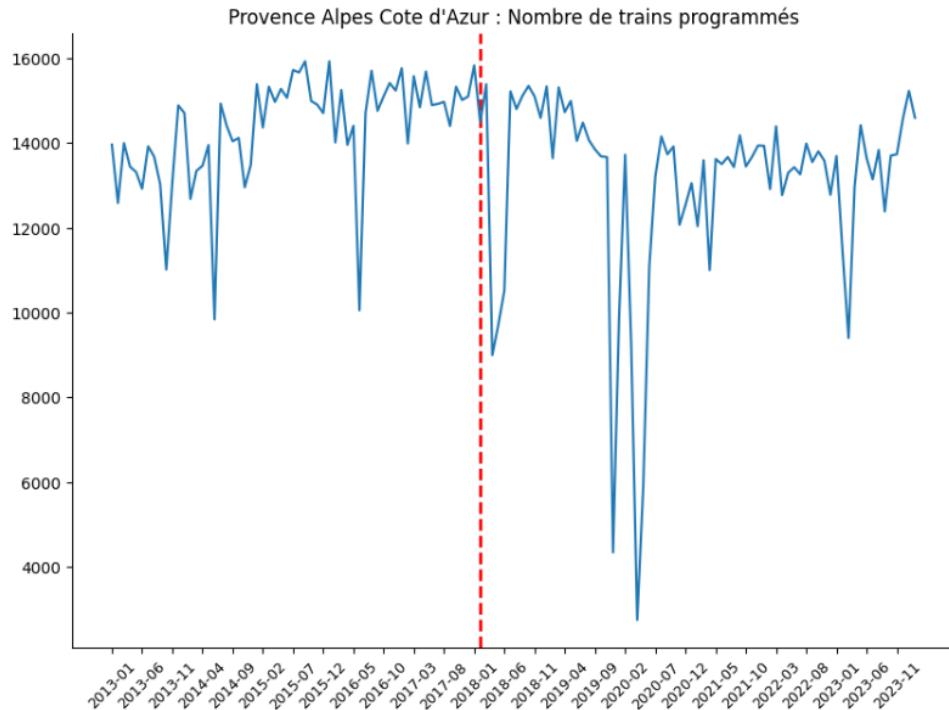


FIGURE 3.51 : Date de rupture

3.5.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 2.30e-06** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

3.5.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (15925.16, 15384.91, 2025.92, 3341.31, -4657.06)$$

Nous constatons que la moyenne de la loi normale asymétrique avant point de rupture est légèrement plus élevée que pour celle après point de rupture. Mais surtout, nous observons une augmentation significante de la variance après point de rupture. Ces observations sont cohérentes avec ce que nous observons à la *Figure 3.51*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.52 et 3.53*)

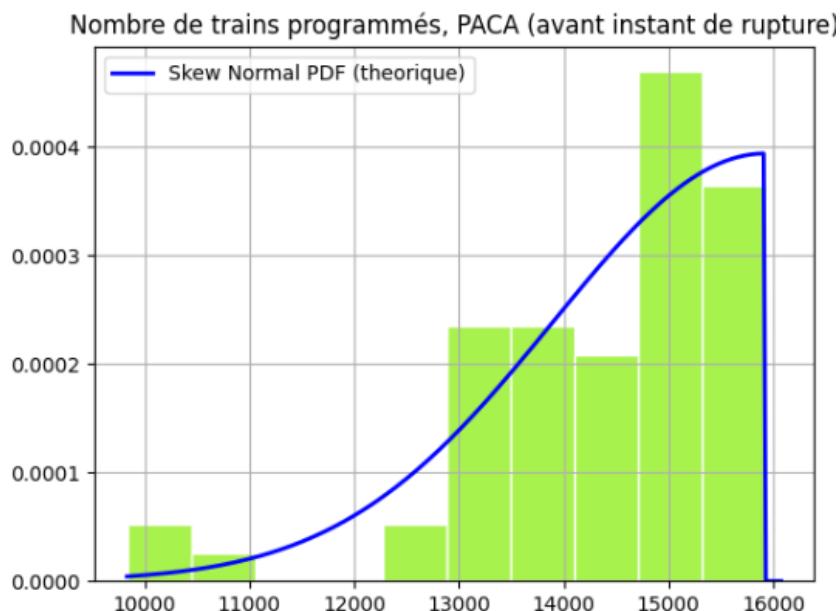


FIGURE 3.52 : Histogramme de l'échantillon avant point de rupture

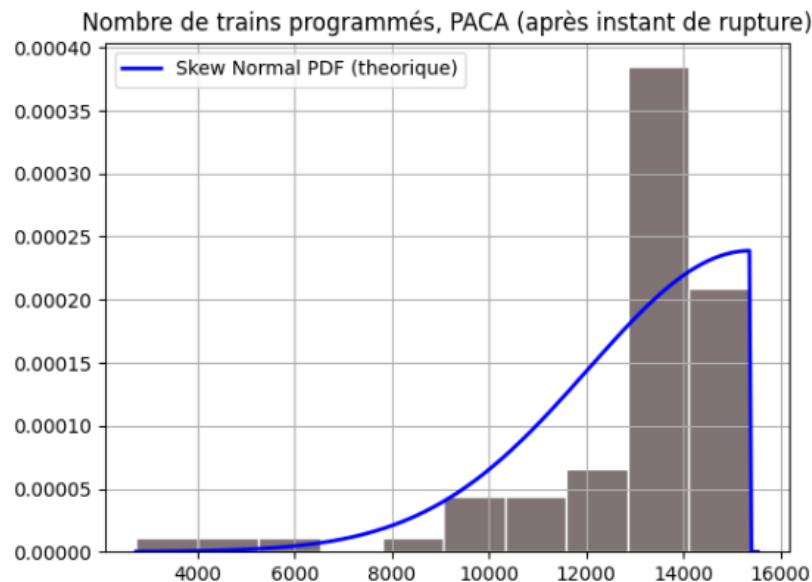


FIGURE 3.53 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ négatif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la droite, avec une queue de distribution plus importante sur la gauche. Cela paraît cohérent avec les histogrammes obtenus. En revanche, la correspondance entre droite théorique et histogramme est un peu décevante, en particulier pour la Figure 3.9.

3.5.2 Trains annulés

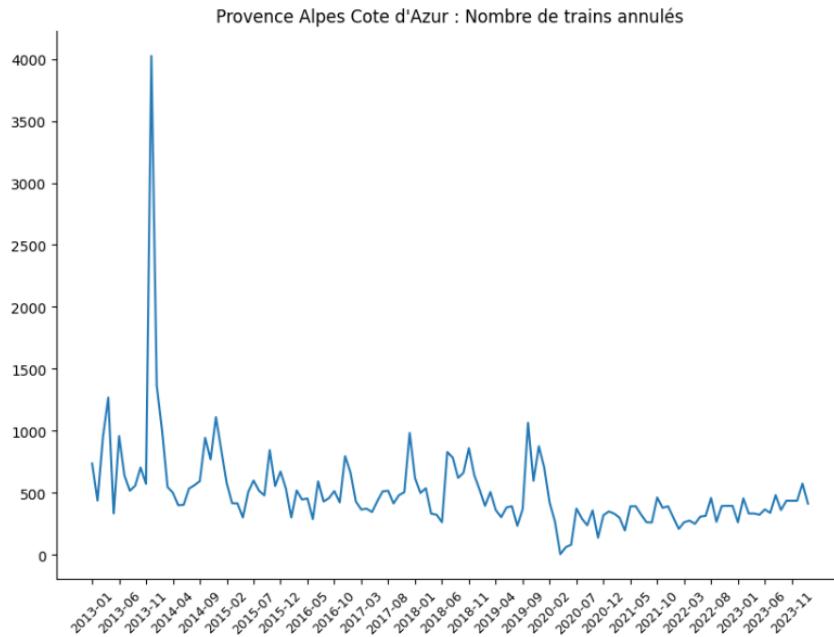


FIGURE 3.54 : Nombre de trains annulés entre 01/2013 et 02/2024

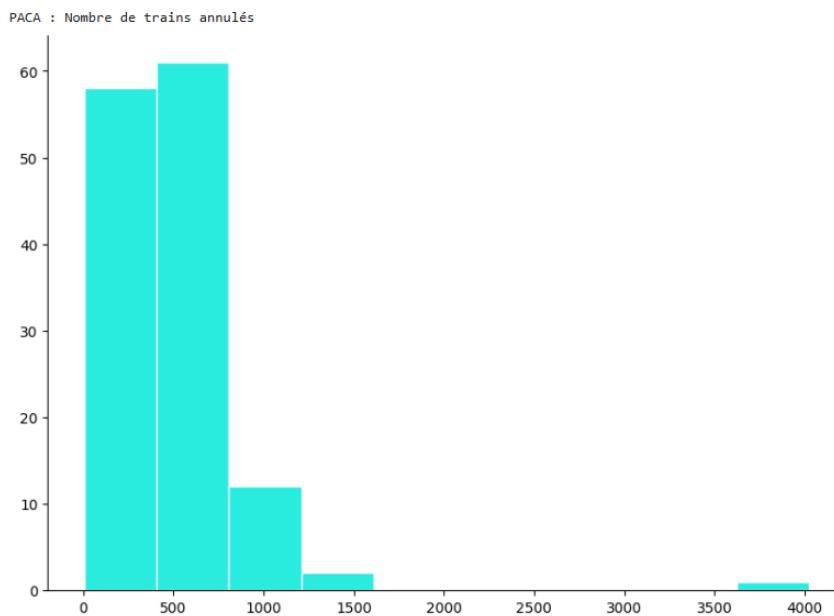


FIGURE 3.55 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.5.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2014-02**.

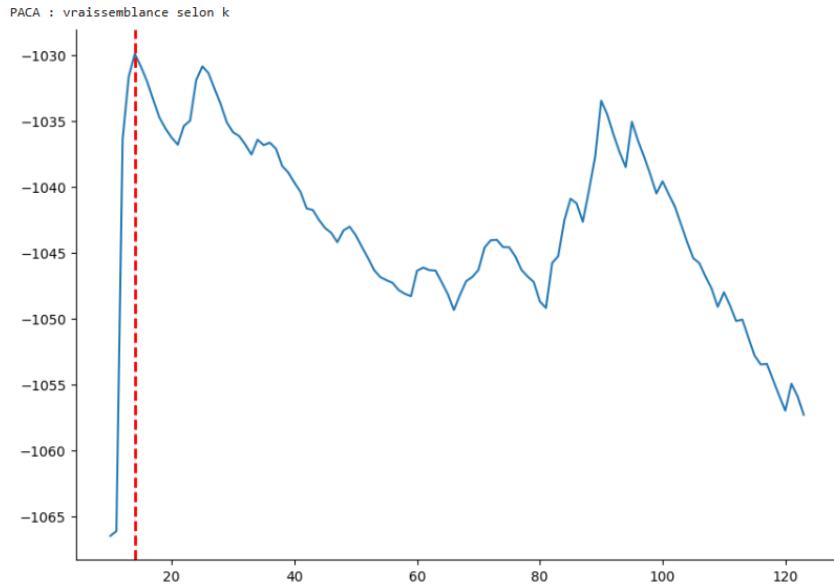


FIGURE 3.56 : Recherche du moment de rupture optimal

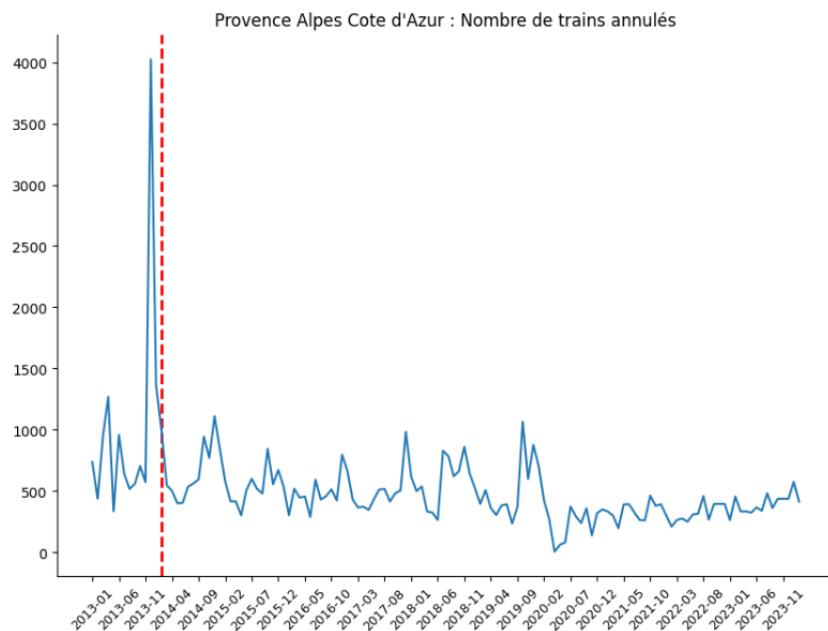


FIGURE 3.57 : Date de rupture

3.5.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value = 9.50e-05** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

3.5.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (285.35, 238.41, 1052.41, 294.60, 3.03)$$

Comme précédemment la moyenne de la loi normale asymétrique avant point de rupture est légèrement plus élevée que celle après point de rupture. Le changement le plus important concerne les variances : la variance de la loi normale assymétrique avant rupture est bien plus élevée comparée à la variance de la loi normale assymétrique après rupture. Ces observations sont cohérentes avec ce que nous observons à la *Figure 3.57*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.58 et 3.59*)

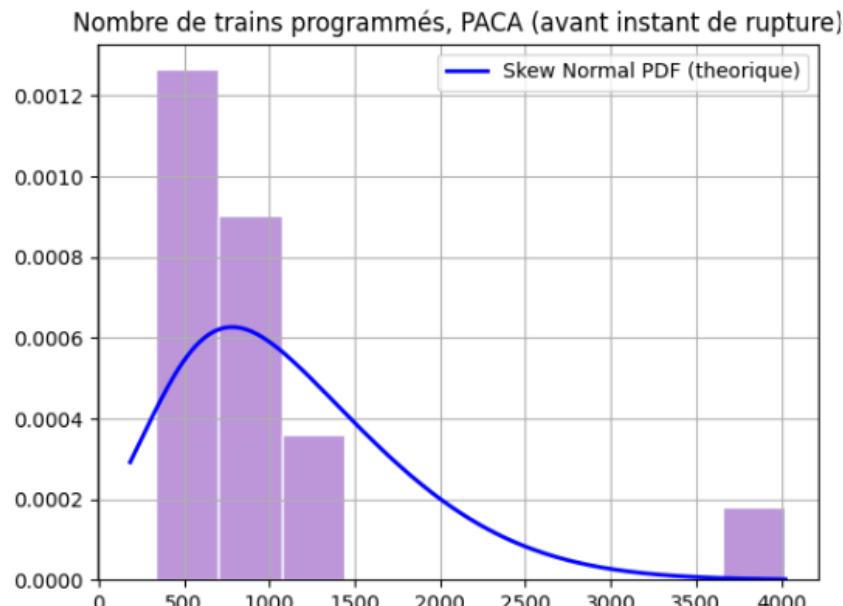


FIGURE 3.58 : Histogramme de l'échantillon avant point de rupture

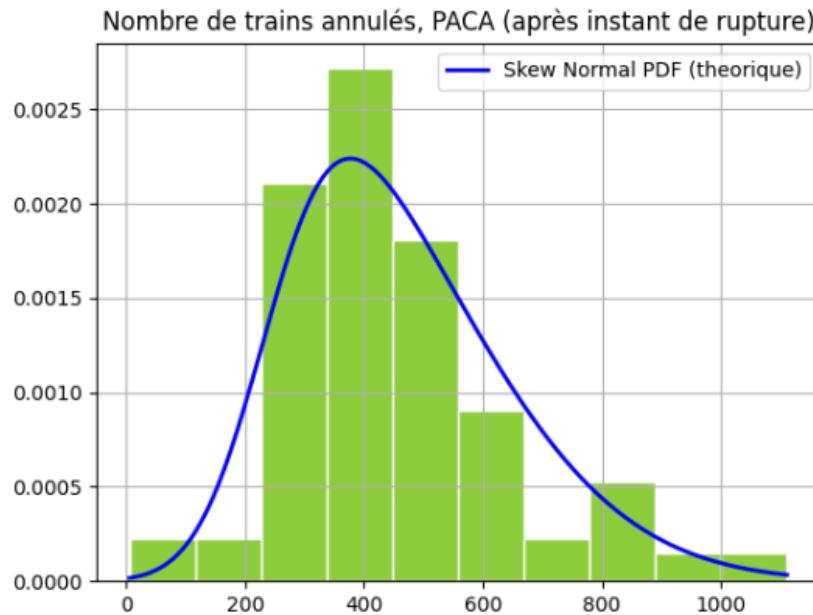


FIGURE 3.59 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ positif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la gauche, avec une queue de distribution plus importante sur la droite. Cela paraît cohérent avec les histogrammes obtenus. Puisque $\hat{\theta}^k$ est petit en valeur absolue, cela signifie que l'assymétrie est légère. C'est cohérent avec l'histogramme de l'échantillon après rupture mais pas très cohérent avec l'histogramme de l'échantillon avant rupture pour lequel l'assymétrie semble beaucoup plus prononcée. Dans ce cas, un modèle de rupture pour lequel les deux lois normales assymétriques ont chacune un paramètre θ différent aurait pu être plus adapté.

La correspondance entre droite théorique et histogramme est un peu décevante pour le figure avant rupture (*cf Figure 3.58*). En revanche elle est satisfaisante pour la figure après rupture (*cf Figure 3.59*).

3.6 Normandie

3.6.1 Trains programmés

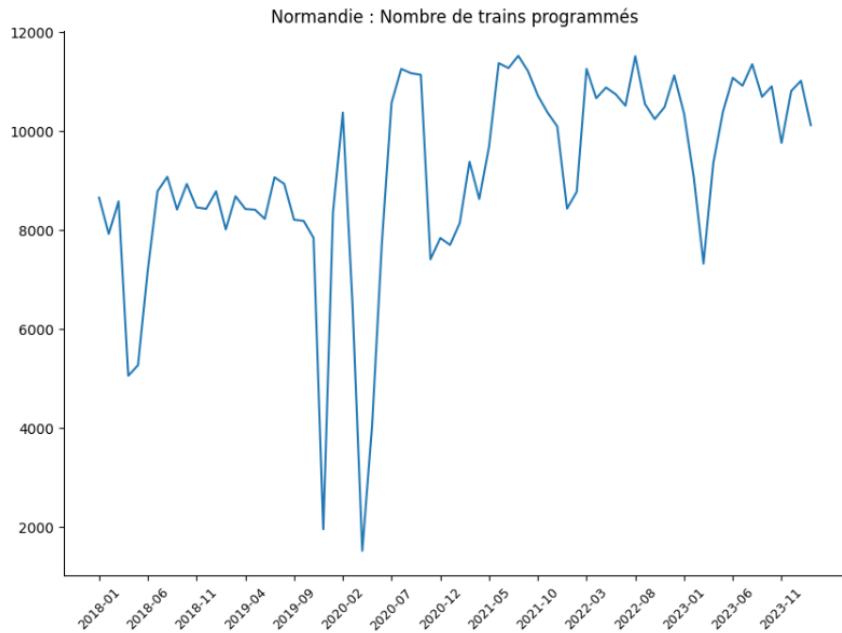


FIGURE 3.60 : Nombre de trains programmés entre 01/2018 et 02/2024

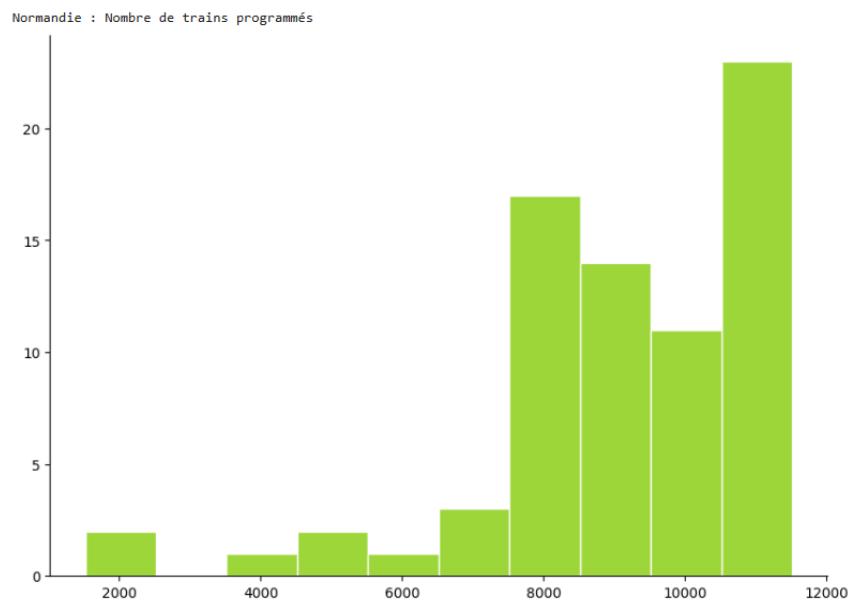


FIGURE 3.61 : Histogramme du nombre de trains programmés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.6.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2019-11**.

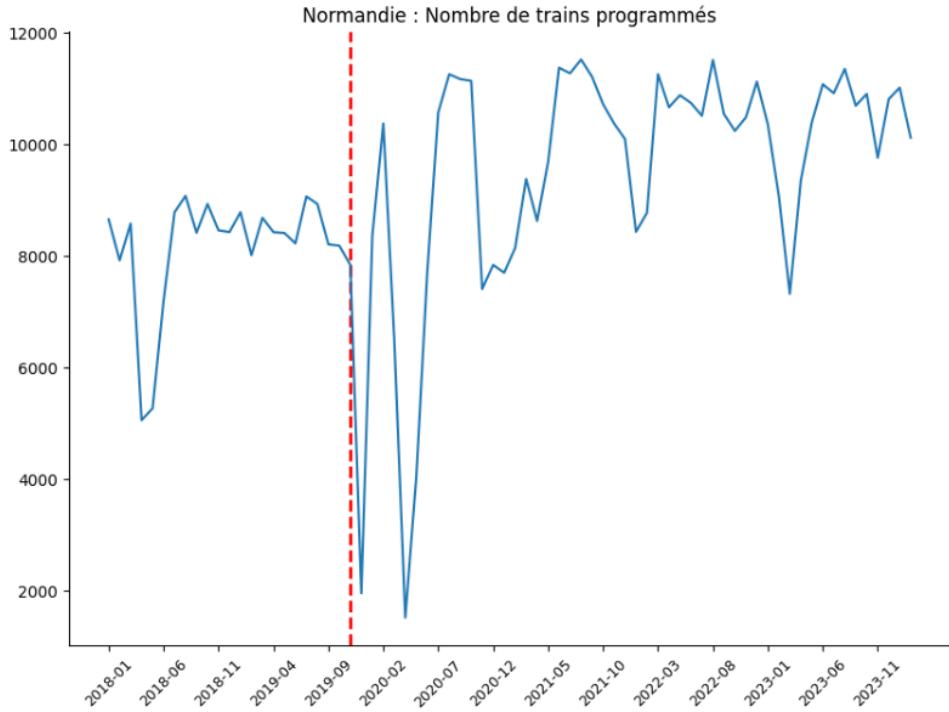


FIGURE 3.62 : Date de rupture

3.6.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 9.50e-05** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

3.6.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (9076.73, 11520.44, 1373.40, 2933.11, -5965.37)$$

Nous constatons une moyenne plus élevée pour la loi normale assymétrique après le point de rupture comparé à celle avant le point de rupture. Nous constatons aussi une augmentation importante de la variance après le point de rupture. Ces observations sont cohérentes avec ce que nous observons à la *Figure 3.62*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.63 et 3.64*)

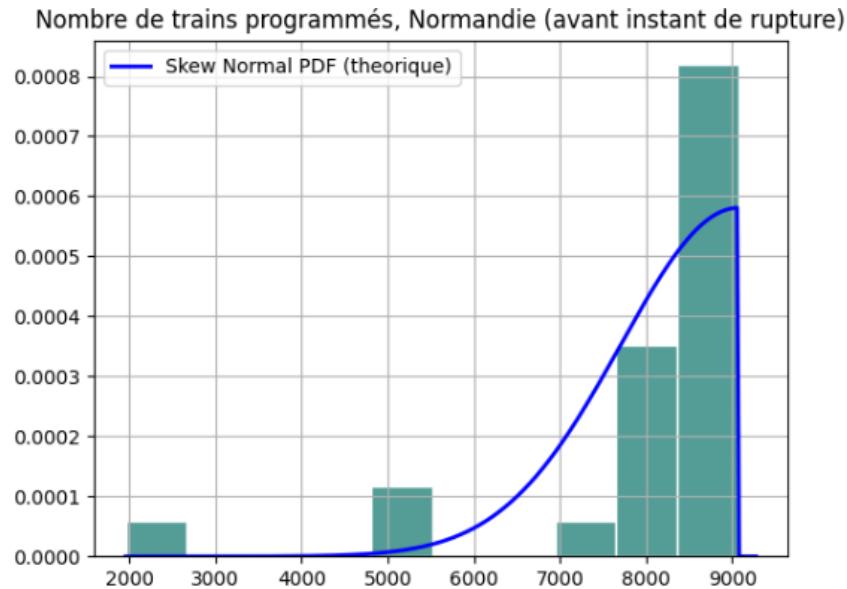


FIGURE 3.63 : Histogramme de l'échantillon avant point de rupture

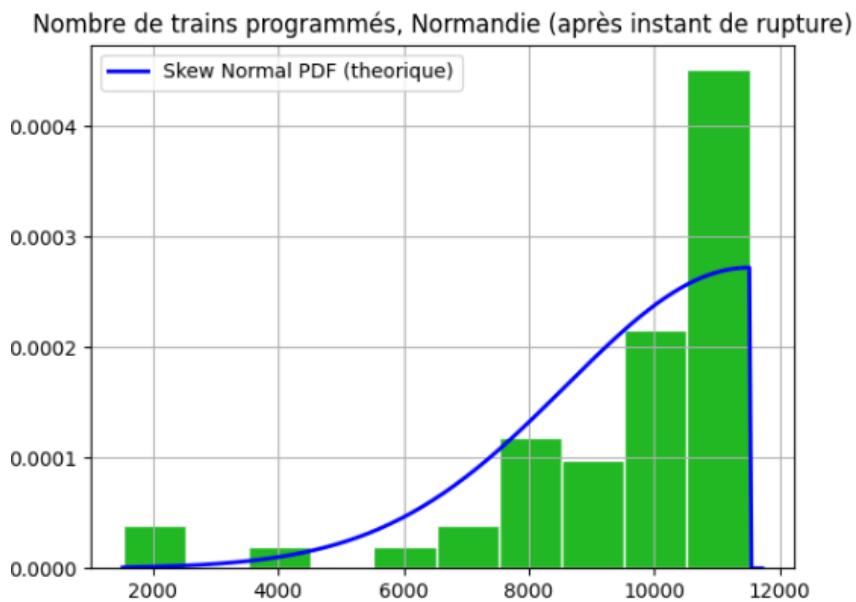


FIGURE 3.64 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ négatif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la droite, avec une queue de distribution plus importante sur la gauche. Puisque $\hat{\theta}^k$ est grand en valeur absolue, cela signifie que l'assymétrie est importante. Cela paraît cohérent avec les histogrammes obtenus. En revanche, la correspondance entre droite théorique et histogramme est décevante.

3.6.2 Trains annulés

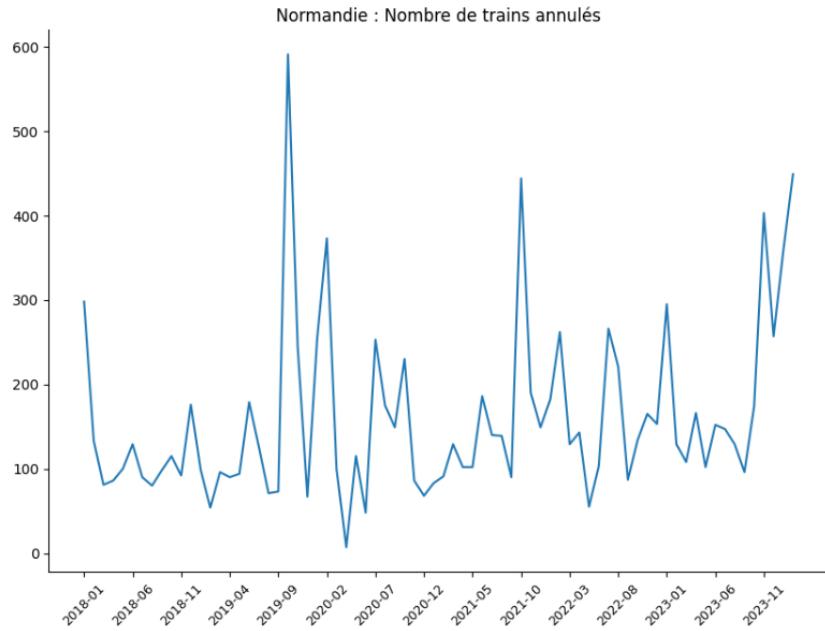


FIGURE 3.65 : Nombre de trains annulés entre 01/2018 et 02/2024

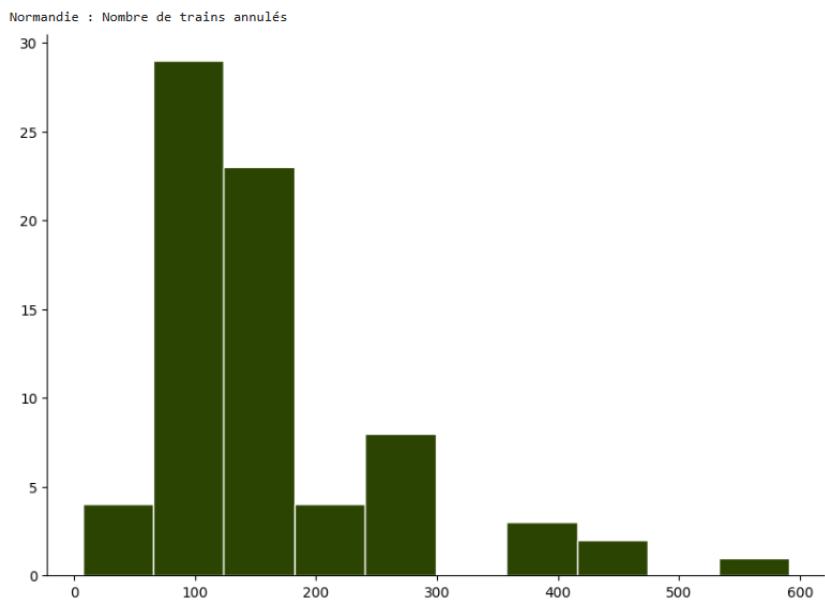


FIGURE 3.66 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.6.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2019-09**.

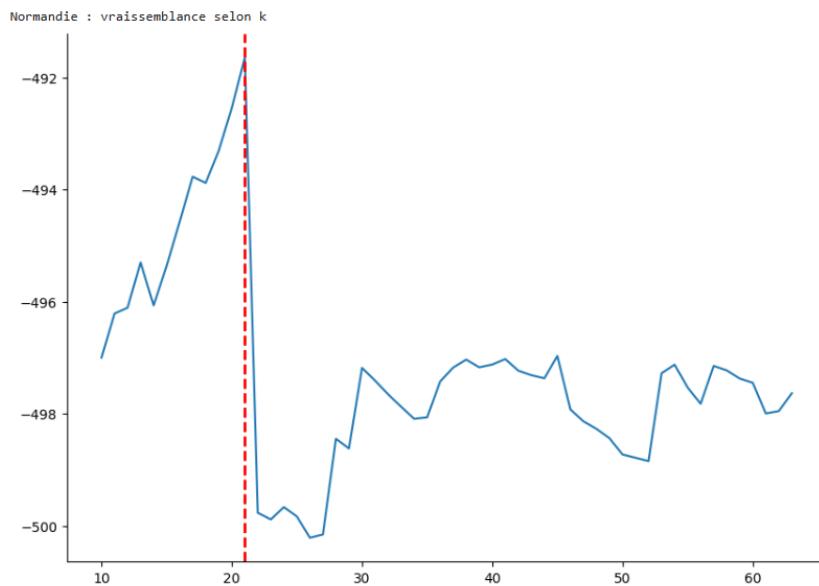


FIGURE 3.67 : Recherche du moment de rupture optimal

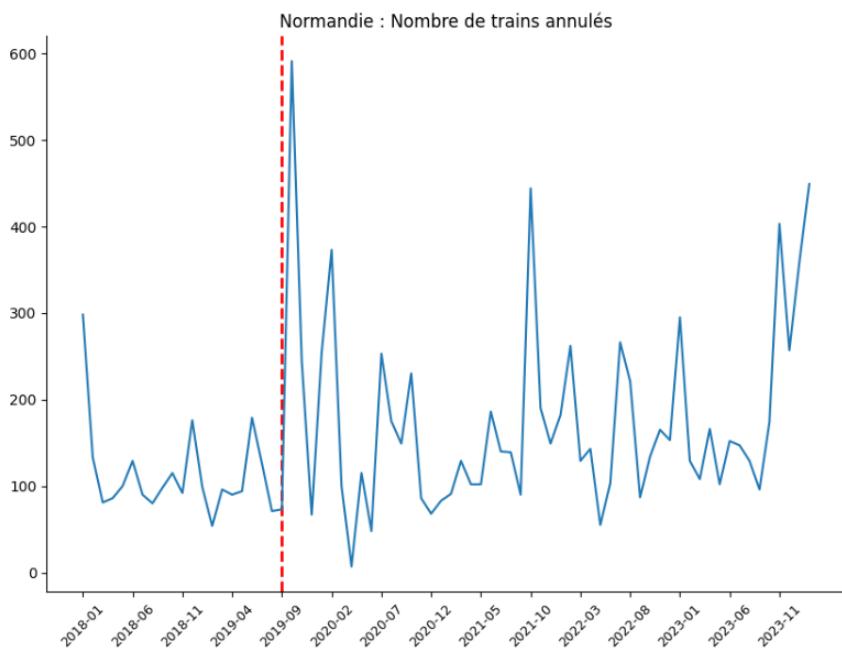


FIGURE 3.68 : Date de rupture

3.6.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value = 0.007** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

3.6.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (62.36, 52.55, 71.21, 170.25, 6.13)$$

La moyenne de la loi normale asymétrique avant point de rupture est plus élevée que celle après point de rupture. Le changement le plus important concerne les variances : la variance de la loi normale assymétrique après rupture est bien plus élevée que la variance de la loi normale assymétrique avant point rupture. Ces observations sont cohérentes avec ce que nous observons à la *Figure 3.68*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.69 et 3.70*)

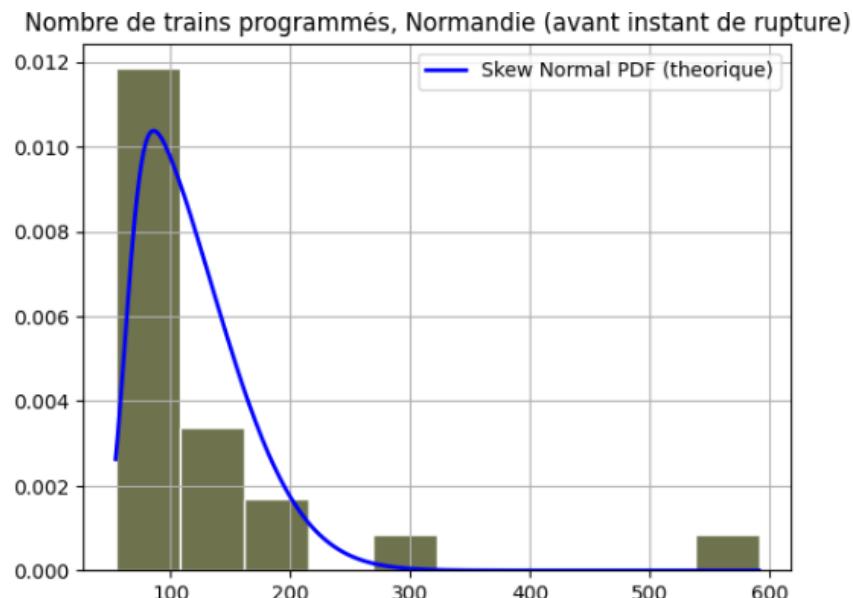


FIGURE 3.69 : Histogramme de l'échantillon avant point de rupture

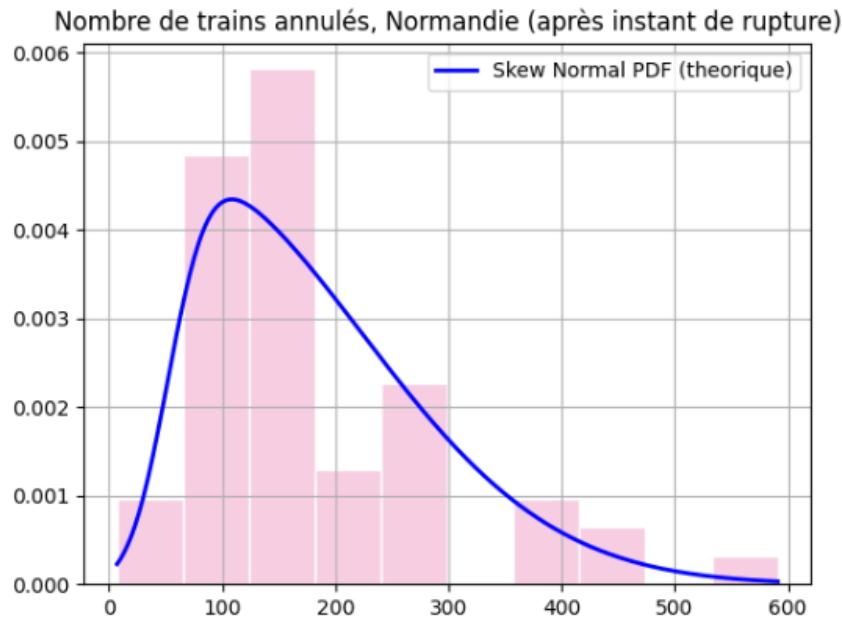


FIGURE 3.70 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ positif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la gauche, avec une queue de distribution plus importante sur la droite. Puisque $\hat{\theta}^k$ est petit en valeur absolue, cela signifie que l'assymétrie est légère. Cela parait cohérent avec les histogrammes obtenus.

La correspondance entre droite théorique et histogramme est satisfaisante pour les deux échantillons.

3.7 Hauts-de-France

3.7.1 Trains programmés

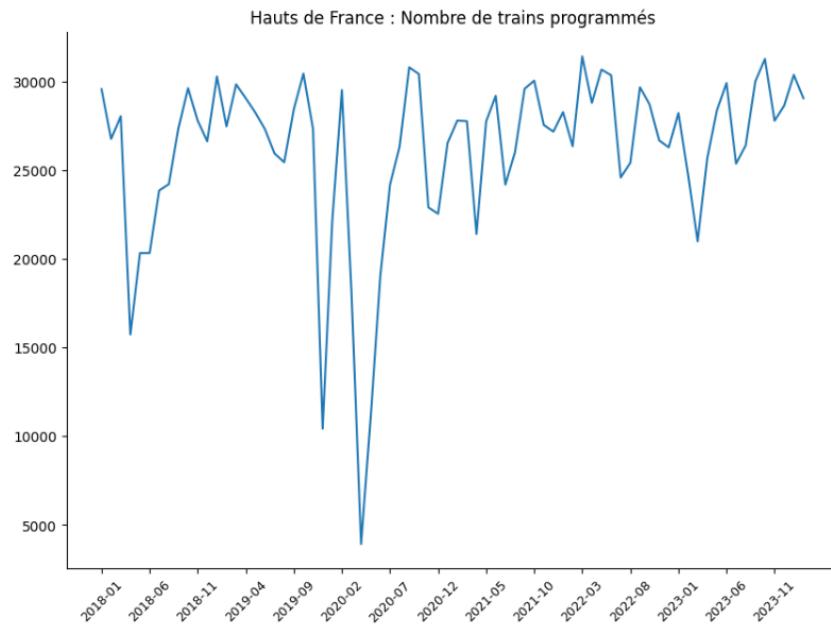


FIGURE 3.71 : Nombre de trains programmés entre 01/2018 et 02/2024

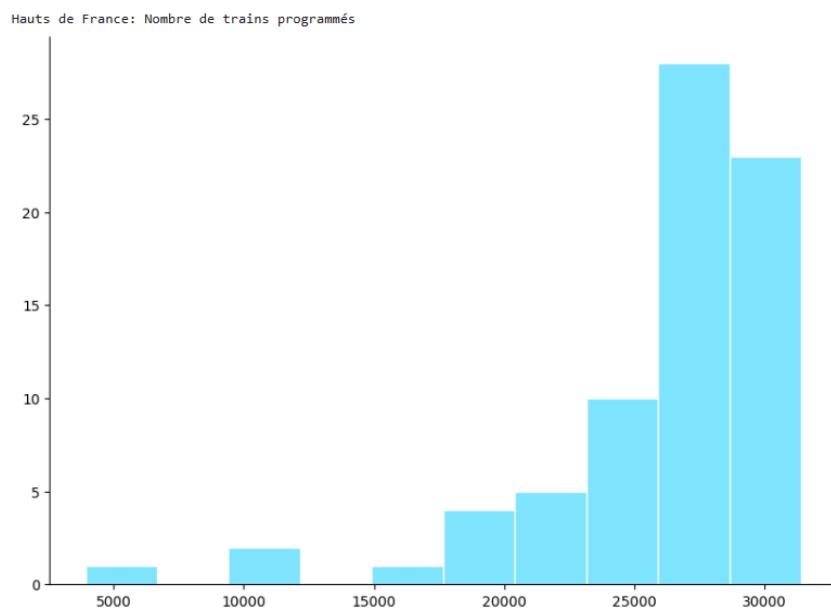


FIGURE 3.72 : Histogramme du nombre de trains programmés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.7.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2021-08**.

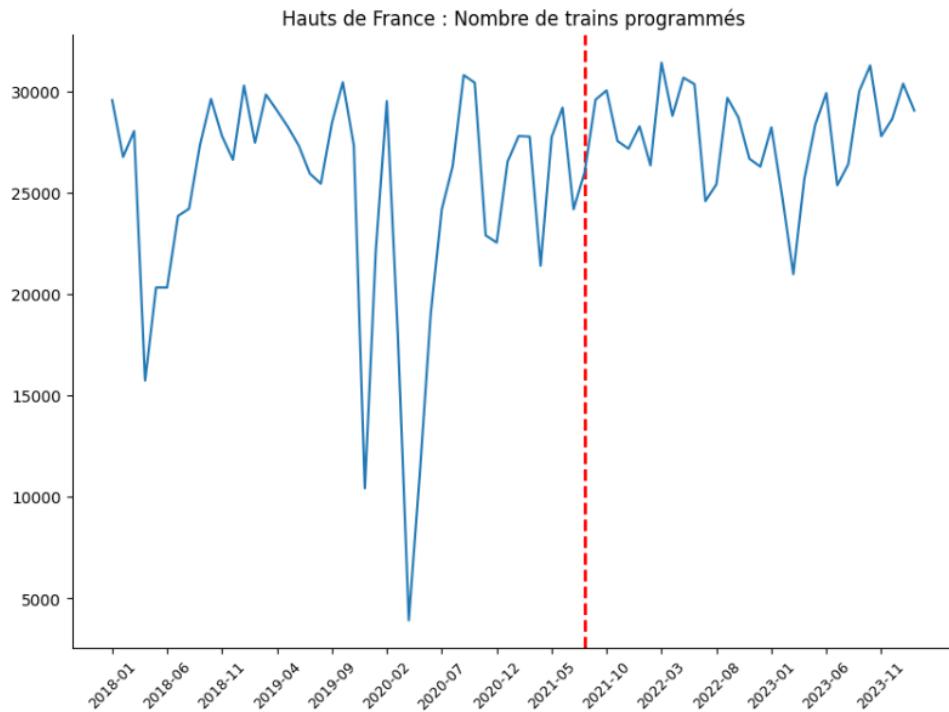


FIGURE 3.73 : Date de rupture

3.7.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 0.04** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

Cependant, nous remarquons que notre p-value est plutôt élevée comparé à ce que nous avions pu obtenir précédemment. Avec un niveau de confiance plus important (par exemple $\alpha=0,01$) nous n'aurions pas assez de preuve pour rejeter H_0 .

3.7.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (30788.67, 31398.90, 8249.20, 4146.03, -6679.83)$$

Nous constatons une hausse légère de la moyenne après le point de rupture. Mais de manière plus significative, nous observons que la variance de la loi avant point de rupture est le double de celle de la loi après point de rupture. Ces observations sont cohérentes

avec ce que nous observons à la *Figure 3.73*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.74 et 3.75*)

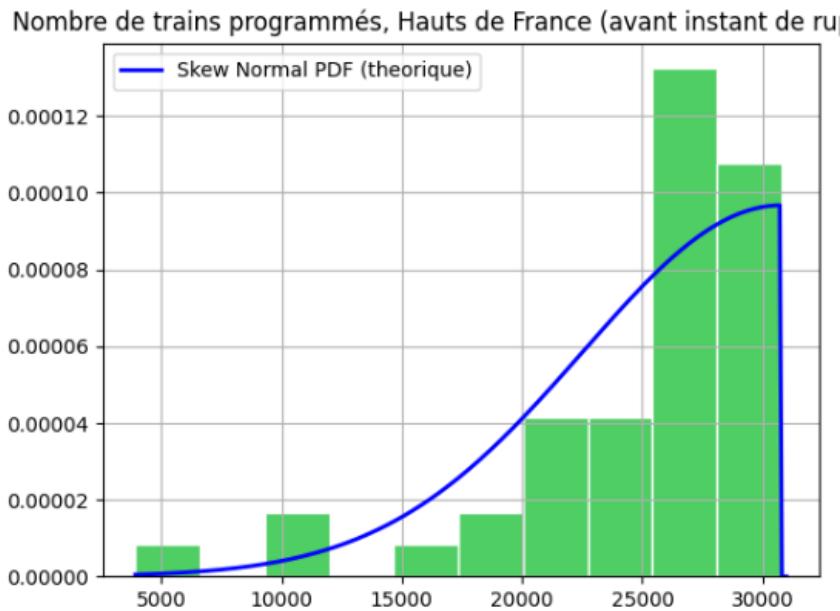


FIGURE 3.74 : Histogramme de l'échantillon avant point de rupture

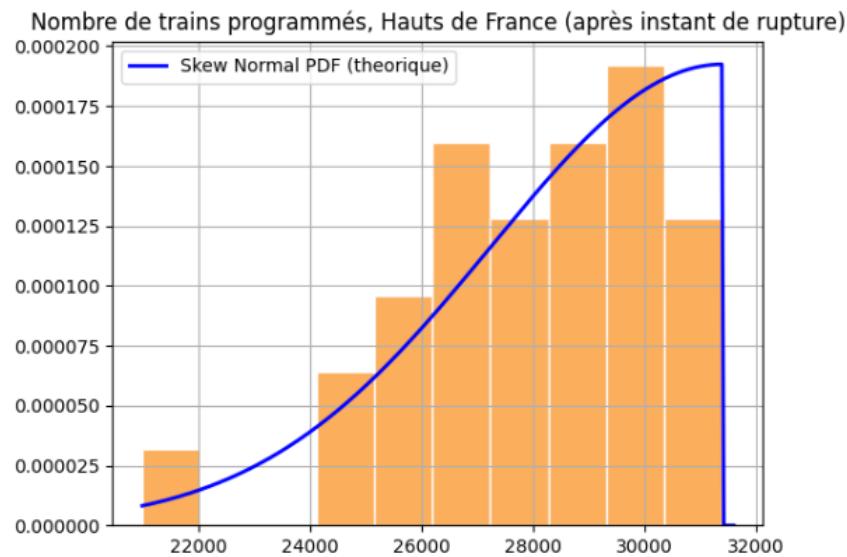


FIGURE 3.75 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ négatif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la droite, avec une queue de distribution

plus importante sur la gauche. Cela paraît cohérent avec les histogrammes obtenus. En revanche, la correspondance entre droite théorique et histogramme est un peu décevante, particulièrement pour la Figure 3.74.

3.7.2 Trains annulés

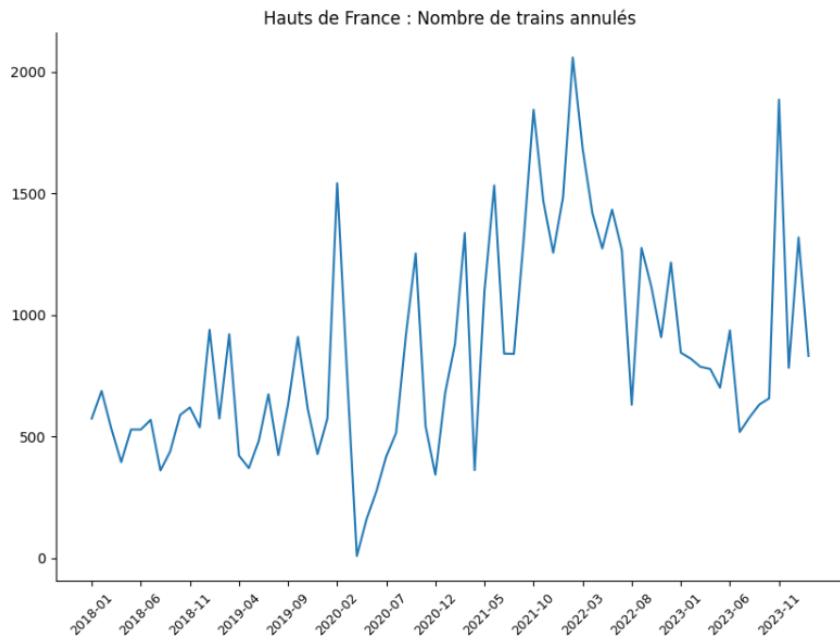


FIGURE 3.76 : Nombre de trains annulés entre 01/2018 et 02/2024

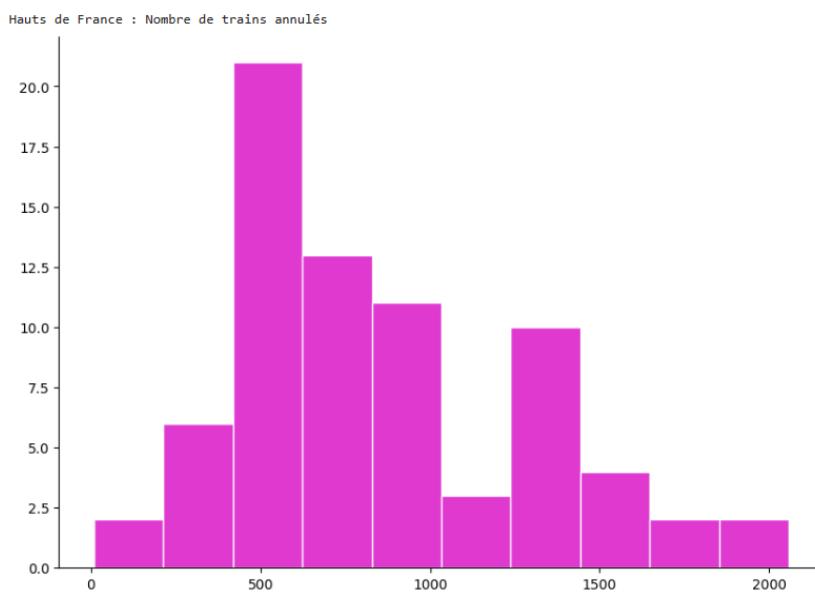


FIGURE 3.77 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.7.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2020-01**.

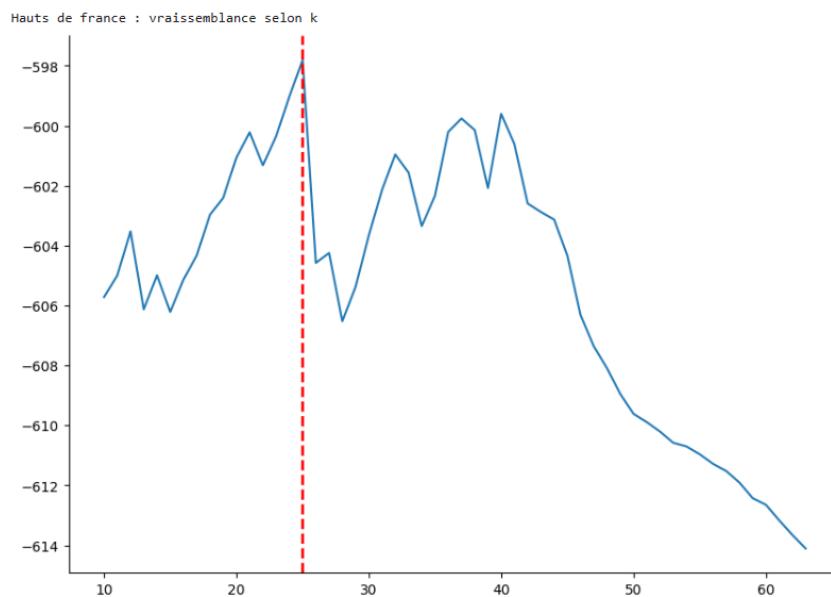


FIGURE 3.78 : Recherche du moment de rupture optimal

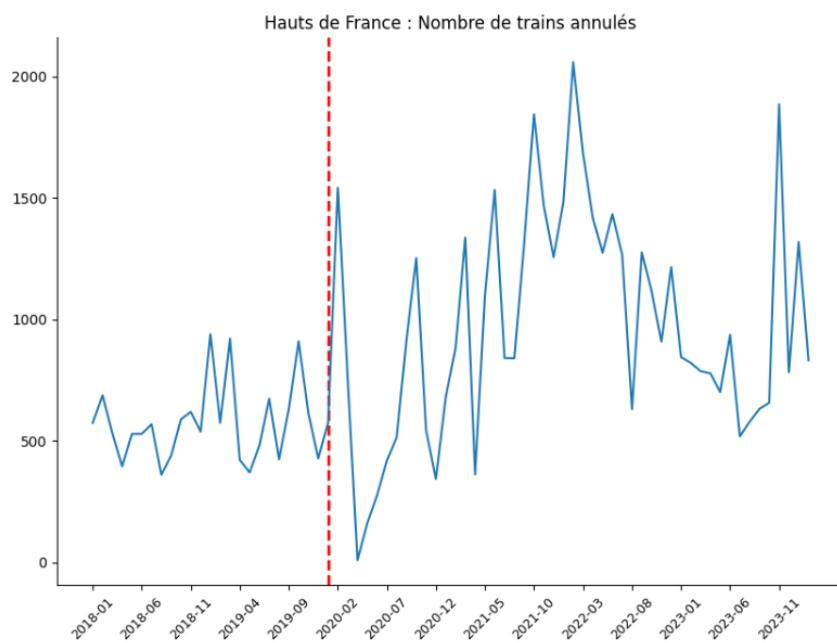


FIGURE 3.79 : Date de rupture

3.7.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 1.90e-05** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

3.7.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (405.98, 438.87, 223.43, 718.55, 2.68)$$

Nous constatons une augmentation de la moyenne après le moment de rupture. De même, la variance augmente après la rupture mais de manière plus importante. Ces observations sont cohérentes avec ce que nous observons à la *Figure 3.79*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.80 et 3.81*)

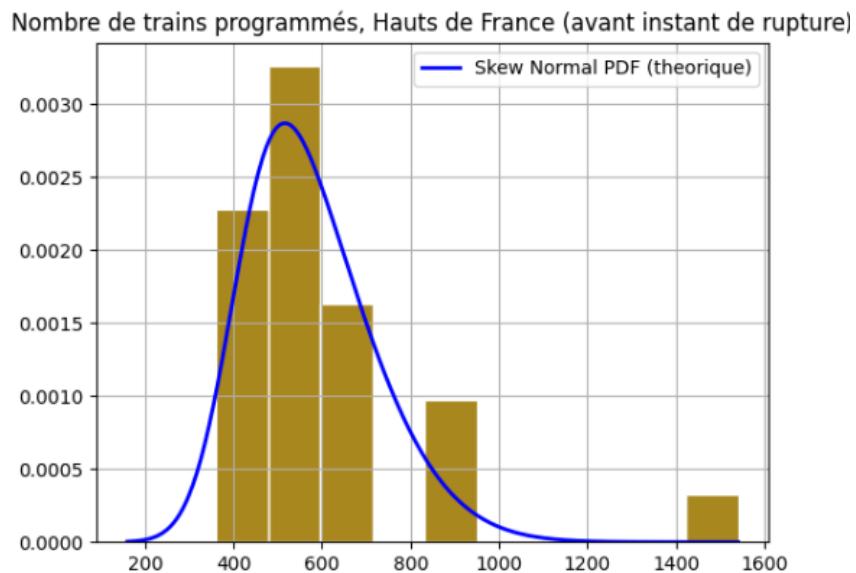


FIGURE 3.80 : Histogramme de l'échantillon avant point de rupture

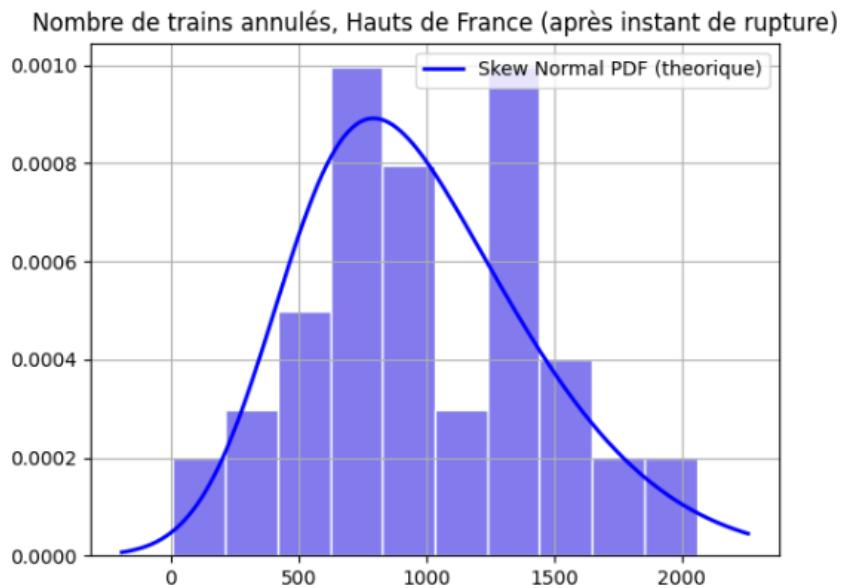


FIGURE 3.81 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ positif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la gauche, avec une queue de distribution plus importante sur la droite. Cela paraît cohérent avec les histogrammes obtenus.

En revanche, la correspondance entre droite théorique et histogramme est un peu décevante.

3.8 Auvergne-Rhône-Alpes

3.8.1 Trains programmés

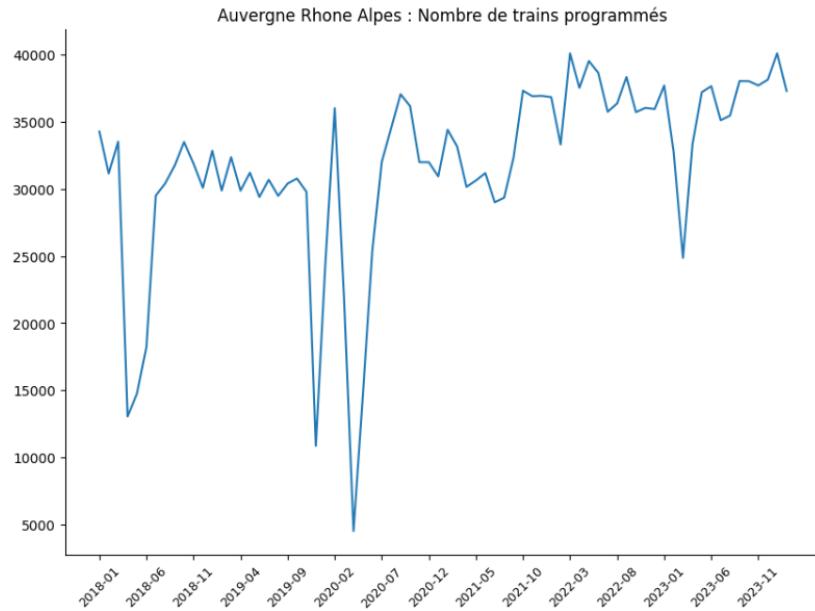


FIGURE 3.82 : Nombre de trains programmés entre 01/2018 et 02/2024

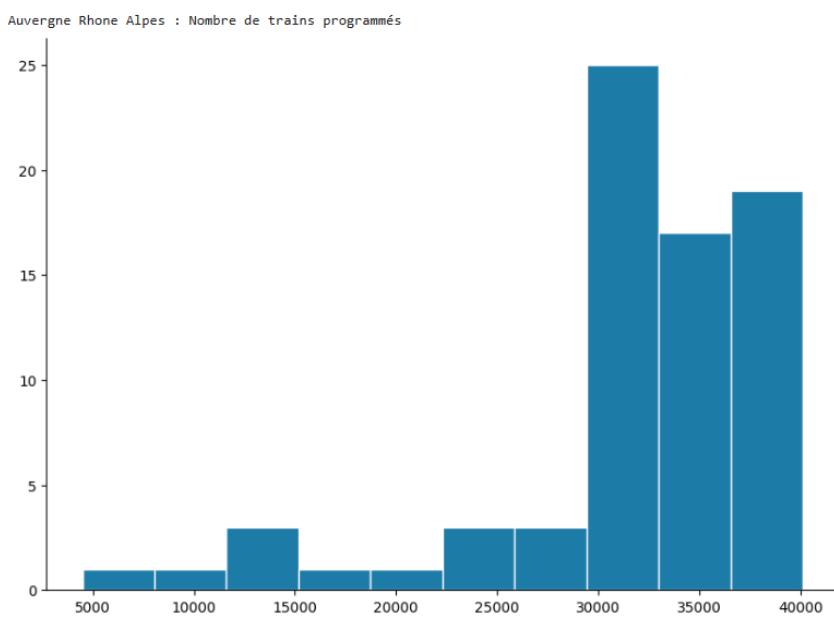


FIGURE 3.83 : Histogramme du nombre de trains programmés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.8.1.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2021-09**.

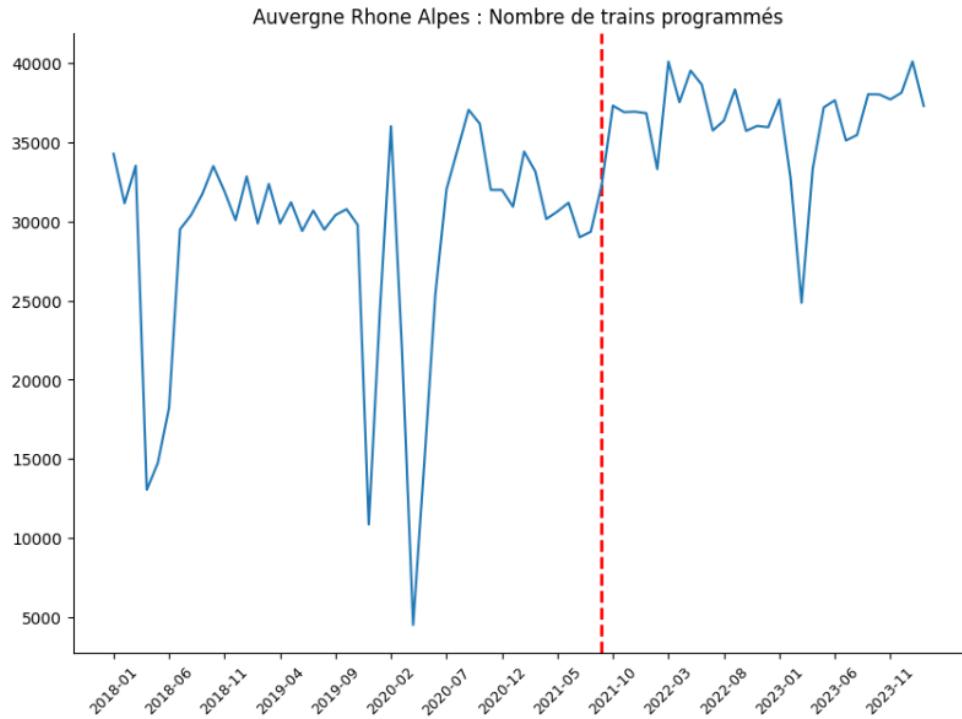


FIGURE 3.84 : Date de rupture

3.8.1.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value= 4.36-11** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

3.8.1.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (35702.70, 39669.58, 9849.48, 4294.28, -6.04)$$

Nous constatons que la moyenne de la loi normale asymétrique après le point de rupture est plus élevée que pour celle avant point de rupture. De plus, nous observons une baisse importante de la variance après point de rupture. Ces observations sont cohérentes avec ce que nous observons à la *Figure 3.84*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.85 et 3.86*)

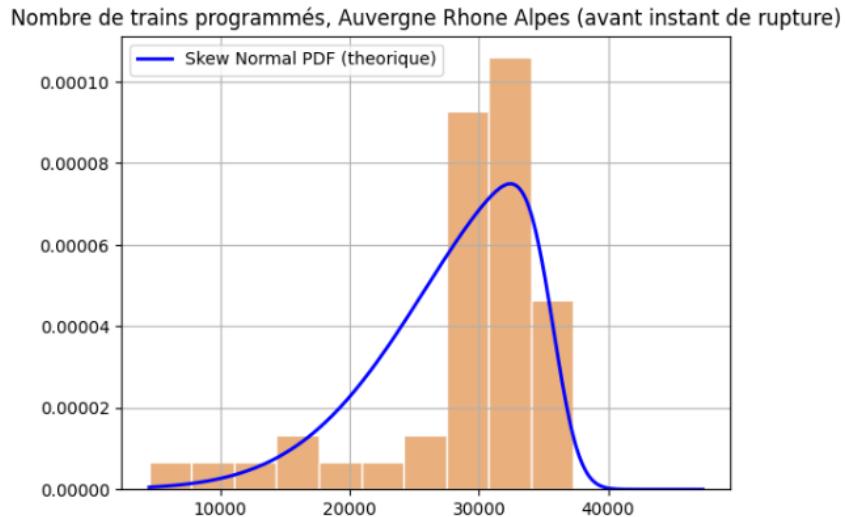


FIGURE 3.85 : Histogramme de l'échantillon avant point de rupture

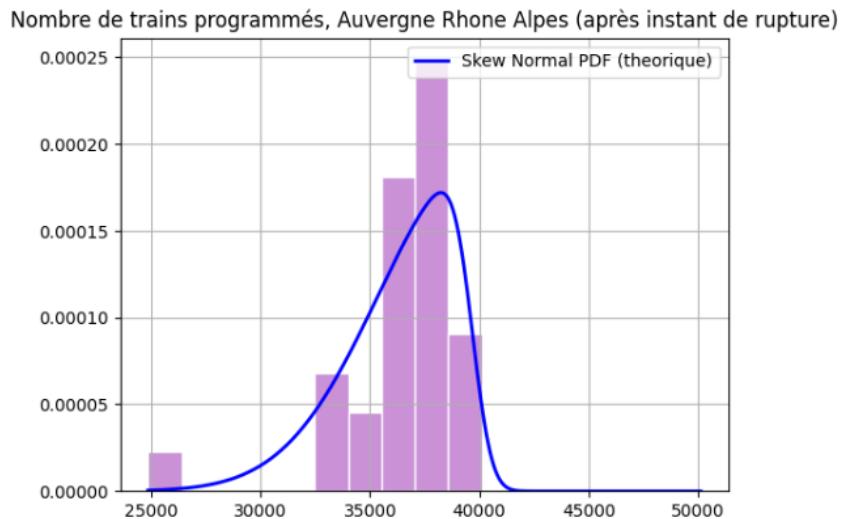


FIGURE 3.86 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ négatif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la droite, avec une queue de distribution plus importante sur la gauche. Cela paraît cohérent avec les histogrammes obtenus.

En revanche, la correspondance entre droite théorique et histogramme est un peu décevante.

3.8.2 Trains annulés

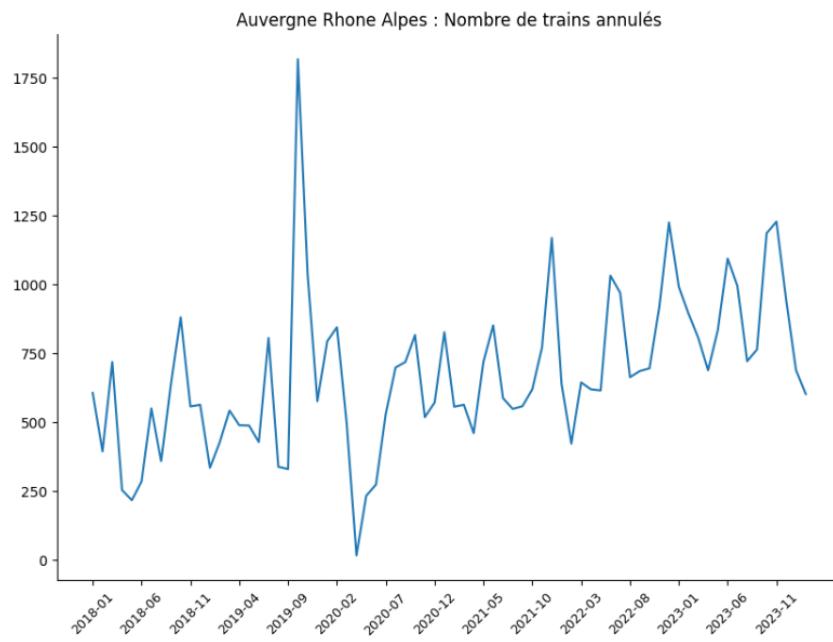


FIGURE 3.87 : Nombre de trains annulés entre 01/2018 et 02/2024

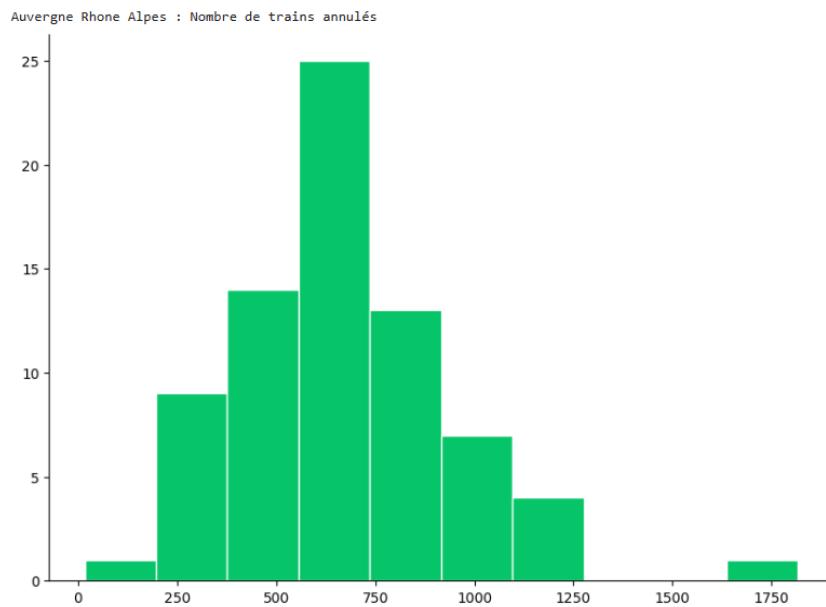


FIGURE 3.88 : Histogramme du nombre de trains annulés

Nous appliquons notre méthode de détection de point de rupture et nous obtenons les résultats suivants :

3.8.2.1 Moment de rupture \hat{k}

La date de rupture obtenue est : **2020-06**.

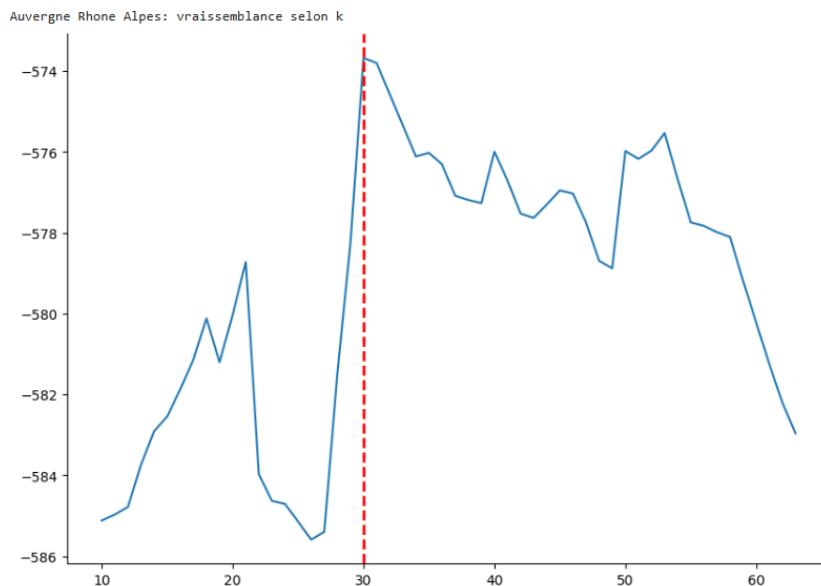


FIGURE 3.89 : Recherche du moment de rupture optimal

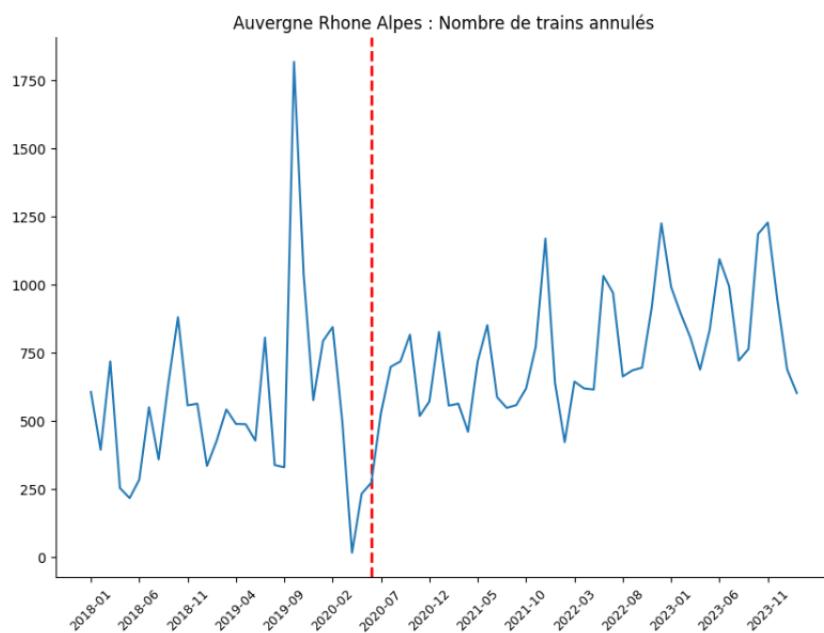


FIGURE 3.90 : Date de rupture

3.8.2.2 Test de Kolmogorov-Smirnov

Nous appliquons le test de Kolmogorov-Smirnov aux deux échantillons. Nous obtenons **p-value = 7.61e-05** : nous rejetons H_0 , ainsi le point de rupture est pertinent.

3.8.2.3 Évaluation des paramètres des lois

$$\hat{\mu}_1^k, \hat{\mu}_2^k, \hat{\sigma}_1^k, \hat{\sigma}_2^k, \hat{\theta}^k = (180.63, 508.59, 490.83, 327.96, 4.58)$$

Nous constatons que la moyenne de la loi normale asymétrique après le point de rupture augmente tandis que la variance baisse. Cela paraît cohérent avec ce que nous observons à la *Figure 3.90*.

Ci-dessous, nous comparons les histogrammes des deux échantillons aux lois théoriques (*cf Figures 3.91 et 3.92*)

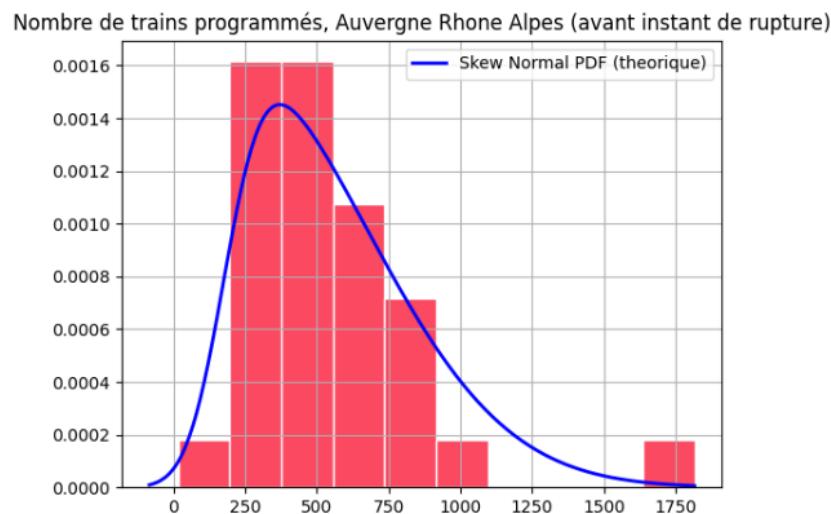


FIGURE 3.91 : Histogramme de l'échantillon avant point de rupture

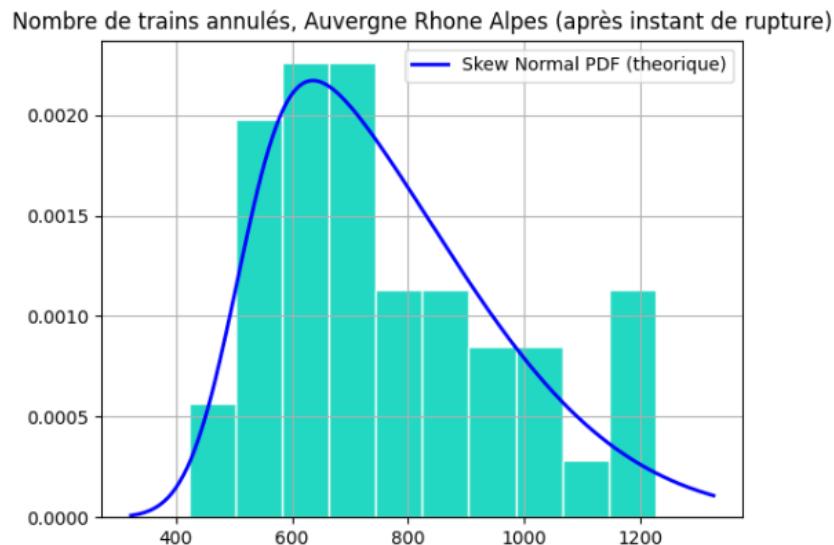


FIGURE 3.92 : Histogramme de l'échantillon après point de rupture

Nous avions obtenu un paramètre $\hat{\theta}^k$ positif, c'est à dire des lois normales asymétriques pour lesquelles la distribution est inclinée vers la gauche, avec une queue de distribution plus importante sur la droite. Cela paraît cohérent avec les histogrammes obtenus.

En outre, la correspondance entre droite théorique et histogramme est satisfaisante pour les deux échantillons.

Conclusion et ANNEXE

3.9 Conclusion

En conclusion, ce rapport a décrit une méthode de détection de point de rupture pour des lois normales asymétriques, permettant d'obtenir le point de rupture optimal ainsi que les paramètres des lois de part et d'autre du point de rupture. Cette méthode a ensuite été appliquée à des données mensuelles de TER pour différentes régions.

L'application s'est révélée plutôt concluante. En effet nous avons obtenus des points de rupture ainsi que des estimations de paramètres de lois cohérents aux données observées.

Concernant la correspondance entre droites théoriques et histogrammes nous avons obtenus des résultats plus ou moins satisfaisants selon la série temporelle et la région donnée. Les exemples décevants pourrait avoir plusieurs causes. D'une part, le modèle n'était peut être pas le mieux adapté aux données. D'autre part, l'utilisation de la fonction `scipy.optimize.minimize` peut parfois nous faire atteindre des minimums locaux plutôt que globaux. Il aurait par exemple pu être intéressant de tester la réponse pour différentes valeurs initiales. Mais malheureusement, les temps d'exécution longs ont été un frein.

3.10 ANNEXE

Imports

```
import numpy as np
import math
import scipy
import scipy.optimize
from scipy.stats import norm
from scipy.special import ndtr
import matplotlib.pyplot as plt
from scipy.stats import skewnorm
import scipy.stats as stats
```

3.10.1 Code permettant d'estimer les paramètres des lois pour un k donné

D'abord nous avons défini une fonction retournant la vraisemblance en fonction de l'échantillon, des paramètres des lois et du moment de rupture choisi

```
#Fonction qui en fonction de l'échantillon, les paramètres de loi
# et le moment de rupture, retourne la vraisemblance

def log_lik(X,phi,k):
    n=len(X)

    mu_1=phi[0]
    mu_2=phi[1]
    sigma_1=phi[2]
    sigma_2=phi[3]
    theta=phi[4]

    sum_sq_1=sum([(x-mu_1)**2 for x in X[:k]])
    sum_sq_2=sum([(x-mu_2)**2 for x in X[k:]])
```

list_norm_cdf_1=[norm.cdf(theta*((x-mu_1)/sigma_1)) for x in X[:k]]
list_norm_cdf_2=[norm.cdf(theta*((x-mu_2)/sigma_2)) for x in X[k:]]

#Lorsque x est negatif et grand en valeur absolue, parfois le résultat de norm.cdf
#est approximé par zero, ce qui pose problème puisque l'on prend ensuite son log
#Puisque nous savons qu'en réalité cette quantité est non nulle nous l'approximons
#par 1e-99 plutôt que 0 pour éviter le pb avec le log
sum_log_1=sum([math.log(1e-99) if x == 0 else math.log(x) for x in list_norm_cdf_1])
sum_log_2=sum([math.log(1e-99) if x == 0 else math.log(x) for x in list_norm_cdf_2])

log_lik=n*math.log(2)-n*math.log(2*math.pi)-k*math.log(sigma_1)-(n-k)*math.log(sigma_2)
log_lik=log_lik-(1/(2*sigma_1**2))*sum_sq_1-(1/(2*sigma_2**2))*sum_sq_2+sum_log_1+sum_log_2

return log_lik

Nous avons ensuite pu définir la fonction renvoyant l'estimation des paramètres par maximisation de la vraisemblance en fonction du point de rupture choisi

```
#On suppose que k est le moment de rupture et on estime les paramètres
#des lois normales asymétriques par estimation par maximisation de la vraisemblance

def reg_scipy(k,X):
    def fonction_objective(phi): #fonction à minimiser ie -(log likelihood)
        log_lik=log_lik(X,phi,k)
        return -log_lik

    bounds = [(None, None),(None, None), (1e-09, None), (1e-09, None),(None, None)]
    #Les paramètres sigma 1 et sigma 2 doivent être positifs strictement

    phi_0=numpy.random.rand(5)
    res1 = scipy.optimize.minimize(fonction_objective, phi_0 ,method='L-BFGS-B', bounds=bounds,
    ||||| | | | | options={'maxiter': 2500, 'ftol': 1e-09, 'disp': False})

    return res1.x
```

3.10.2 Code permettant de trouver le point de rupture optimal

```
#Etant donne un echantillon, fonction qui les renvoie les vraisemblances,
#parametres estimes pour chaque moment de rupture candidats ainsi
#qu'une liste des instants candidats

def candidate_rupt(X):
    index_list=[]
    log_liq_list=[]
    param_list=[]

    n=len(X)

    for k in range(10,n-10):

        #Pour le moment de rupture k,
        #nous determinons les parametres de lois les plus plausibles
        param=reg_scipy(k,X)

        #Avec ces parametres estimes, nous estimons la vraisemblance de l'echantillon
        log_liq_est=log_liq(X,param,k)

        index_list.append(k)
        param_list.append(param)
        log_liq_list.append(log_liq_est)

    return index_list, param_list, log_liq_list

#Par exemple pour la region PACA et la serie temporelle des trains programmés

PACA=df[df['Région']=="Provence Alpes Côte d'Azur"]
PACA=PACA.reset_index(inplace=False, drop=True)

index_list, param_list, log_liq_list=candidate_rupt(PACA['Nombre de trains programmés'])

max_value = max(log_liq_list)
max_index=log_liq_list.index(max_value)

#Le moment de rupture optimal est alors
k_hat=index_list[max_index]-1
```

3.10.3 Test de Kolmogorov-Smirnov

```
data1 = PACA.loc[:k_hat+1,'Nombre de trains programmés']
data2 = PACA.loc[k_hat+1:,'Nombre de trains programmés']

statistic, pvalue = stats.kstest(data1, data2)

print("p-value:", pvalue)

# Interprétation (alpha=0.05)
if pvalue < 0.05:
    print("on rejette H0")
else:
    print("pas assez de preuves pour rejeter H0")
```

Le détail du reste du code se trouve dans le fichier code joint au rapport.