



Réf : \_\_\_\_\_/UAMOB/FSE/DEP \_\_\_\_\_/2026

# MÉMOIRE DE FIN DE CYCLE

EN VUE DE L'OBTENTION DU DIPLÔME DE MASTER

Domaine : Mathématiques et Informatique

Filière : Mathématiques

Spécialité : RECHERCHE OPÉRATIONNELLE

Présenté par

**TALEB hanane**

Thème

---

**Modélisation statistique et analyse comparative de la  
fiabilité des réseaux ADSL et fibre optique : application  
aux données d'Algérie Télécom (Bouira)**

---

Soutenu le 08/06/2026 devant le jury composé de :

QUIDJA Daya	Président	Univ.BOUIRA
BOUGHANI L'Hadi	Promoteur	Univ.BOUIRA
IFTISSEN El Ghani	Examineur	Univ.BOUIRA
BIROUCHE Madjid	Examineur	Univ.BOUIRA

Année Universitaire (2025/2026)

# Remerciement

Tout d'abord, je remercie **Allah** Tout-Puissant de m'avoir accordé la patience et la force nécessaires pour mener ce travail à bien.

Mes sincères remerciements s'adressent à mon encadreur, Monsieur BOUGHANI L'Hadi, pour avoir accepté de superviser ce travail, ainsi que pour ses conseils, sa disponibilité et ses précieuses orientations tout au long de cette recherche.

Je remercie également les membres du jury d'avoir accepté d'évaluer ce mémoire, en l'occurrence Madame OUIDJA Daya, présidente du jury, ainsi que Messieurs IFTISSEN El Ghani et BIROUCHE Madjid, examinateurs.

J'adresse aussi mes remerciements à Algérie Télécom et à son personnel pour leur aide et la mise à disposition des données nécessaires à la réalisation de cette étude.

Enfin, j'exprime ma profonde gratitude à ma famille ainsi qu'à mes amies, notamment celles de la promotion 2026, pour leur soutien et leurs encouragements.

*Merci infiniment.*

# Dédicace

Je dédie ce modeste travail :

À mes chers parents, qui ont toujours été à mes côtés, même dans les moments les plus difficiles, pour tout leur amour, leurs sacrifices et leur soutien inconditionnel.

À ma sœur et à mon frère, pour leur soutien et leur présence.

À toutes les personnes qui m'ont soutenue tout au long de mon parcours.

*Hanane.T.*

# Résumé

Ce mémoire consiste, à partir de données réelles de pannes et d'utilisation de la bande passante des utilisateurs de Bouira, à construire des modèles de prévision à l'aide des modèles ARIMA sous le logiciel R, ainsi qu'à appliquer des méthodes d'analyse de survie afin d'étudier la fiabilité des réseaux ADSL et fibre optique et de comparer leurs performances.

**Mots-clés** : ADSL, fibre optique, qualité de service, ARIMA, analyse de survie, Kaplan-Meier, modèle de Cox, R, fiabilité des réseaux.

# Abstract

This thesis is based on real data of network failures and bandwidth usage from users in Bouira. It aims to build forecasting models using ARIMA methods in R and to apply survival analysis techniques in order to study the reliability of ADSL and fiber optic networks and compare their performance.

**Keywords** : ADSL, optical fiber, quality of service, ARIMA, survival analysis, Kaplan-Meier, Cox model, R, network reliability.

# Liste des abréviations

LAN	Local Area Network.
MAN	Metropolitan Area Network.
WAN	Wide Area Network.
TCP/IP	Transmission Control Protocol/Internet Protocol.
ADSL	Asymmetric Digital Subscriber Line.
FTTx	Fiber to the x.
FTTH	Fiber To The Home (Fibre optique jusqu'à l'abonné).
QoS	Quality of Service (Qualité de service).
ARIMA	AutoRegressive Integrated Moving Average.
AR	AutoRegressive.
MA	Moving Average (Moyenne mobile).
TS	Trend Stationary.
DS	Difference Stationary.
ADF	Augmented Dickey-Fuller test.
ACF	Autocorrelation Function.
PACF	Partial Autocorrelation Function.
MSE	Mean Squared Error (Erreur quadratique moyenne).
MAE	Mean Absolute Error (Erreur absolue moyenne).
Cox	Modèle de régression de Cox (risques proportionnels).

# Table des matières

<b>Liste des figures</b>	<b>v</b>
<b>Liste des tableaux</b>	<b>vii</b>
<b>Introduction</b>	<b>1</b>
<b>1 La qualité de réseau internet</b>	<b>3</b>
1.1 Généralités sur les réseaux informatiques . . . . .	4
1.1.1 Les réseaux informatiques . . . . .	4
1.1.2 Types de réseaux informatiques . . . . .	4
1.1.3 Les composants de réseau . . . . .	5
1.1.4 Les topologies des réseaux . . . . .	7
1.1.5 Les architectures des réseaux . . . . .	10
1.2 Internet . . . . .	11
1.2.1 Technologies d'accès Internet . . . . .	12
1.3 Qualité de réseau internet . . . . .	14
1.4 Qualité de service (QoS) . . . . .	14
1.4.1 Les critères de QoS . . . . .	15
1.4.2 Concepts : congestion et fiabilité . . . . .	16
1.4.3 Impact de la qualité sur la demande . . . . .	16
1.5 Gestion de l'Internet en Algérie et au niveau de l'ACTEL de Bouira . . . . .	17
1.5.1 Au niveau d'Algérie . . . . .	17
1.5.2 Au niveau de l'ACTEL . . . . .	18
<b>2 Les outils d'analyse de données</b>	<b>19</b>

## Table des matières

---

2.1	Analyse des séries temporelles . . . . .	20
2.1.1	Définition d'une série temporelle . . . . .	20
2.1.2	Les composantes d'une série temporelle . . . . .	20
2.1.3	Décomposition d'une série temporelle . . . . .	21
2.1.4	La stationnarité . . . . .	22
2.1.5	Les modèles ARIMA (p,d,q) . . . . .	23
2.1.6	Lissage exponentiel . . . . .	29
2.2	Analyse de survie . . . . .	32
2.2.1	Les concepts fondamentaux . . . . .	33
2.2.2	Estimateur Kaplan-Meier . . . . .	34
2.2.3	Modèle de Cox . . . . .	34
<b>3</b>	<b>Modélisation de données</b>	<b>36</b>
3.1	Analyse exploratoire : visualisation des séries temporelles . . . . .	36
3.2	Analyse descriptive et comparative . . . . .	38
3.2.1	Analyse descriptive des pannes . . . . .	38
3.2.2	Analyse descriptive de la bande passante . . . . .	39
3.2.3	Test de comparaison . . . . .	41
3.3	Analyse des séries temporelles . . . . .	42
3.3.1	Prévision des pannes ADSL . . . . .	42
3.3.2	Prévision des pannes fibre . . . . .	44
3.3.3	Comparaison des prévisions (pannes) . . . . .	45
3.4	Analyse de survie . . . . .	46
3.4.1	Statistiques descriptives des gaps . . . . .	46
3.4.2	Estimateur Kaplan-Meier . . . . .	48
3.4.3	Modèle de Cox . . . . .	49
	<b>Conclusion</b>	<b>51</b>
	<b>Annexes</b>	<b>52</b>
	<b>Bibliographie</b>	<b>58</b>

# Table des figures

1.1	Schéma d'une topologie en bus. . . . .	8
1.2	Schéma d'une topologie en étoile. . . . .	8
1.3	Schéma d'une topologie en anneau . . . . .	9
1.4	Schéma d'une topologie maillé . . . . .	9
2.1	Schéma d'une modèle additif . . . . .	21
2.2	Schéma d'une modèle multiplicatif . . . . .	22
2.3	Série stationnaire . . . . .	23
3.1	Évolution des données de pannes ADSL . . . . .	37
3.2	Évolution des données de pannes Fibre optique . . . . .	37
3.3	Comparaison graphique des pannes entre l'ADSL et la fibre optique . . . . .	39
3.4	Évolution horaire de l'utilisation moyenne de la bande passante. . . . .	40
3.5	Utilisation moyenne de la bande passante : ADSL vs fibre. . . . .	41
3.6	Résultat du test du khi-deux (ADSL vs fibre optique). . . . .	42
3.7	Paramètres et critères de qualité du modèle ARIMA – ADSL. . . . .	43
3.8	Prévisions des pannes ADSL (10 premiers jours). . . . .	43
3.9	Paramètres et critères de qualité du modèle ARIMA – fibre optique. . . . .	44
3.10	Prévisions des pannes fibre (10 premiers jours). . . . .	44
3.11	Prévisions des pannes (10 premiers jours). . . . .	45
3.12	Comparaison des prévisions ADSL vs fibre . . . . .	45
3.13	Résultats de Kaplan-Meier pour ADSL et Fibre. . . . .	48
3.14	Courbes de survie (ADSL vs fibre optique). . . . .	48
3.15	Résultats de test Log-Rank . . . . .	49

## Table des figures

---

3.16	Résultats de modèle de cox pour ADSL et Fibre. . . . .	49
17	Code source 1 . . . . .	52
18	Code source 2 . . . . .	53
19	Code source 3 . . . . .	54
20	code source 4 . . . . .	55
21	Code source 5 . . . . .	56
22	Prévisions des pannes ADSL sur 30 jours . . . . .	57
23	Prévisions des pannes Fibre sur 30 jours . . . . .	58

# Liste des tableaux

3.1	Statistiques descriptives des pannes quotidiennes . . . . .	38
3.2	Statistiques descriptives de l'utilisation de la bande passante (%). . . . .	40
3.3	Statistiques des pannes par client (ADSL vs Fibre) . . . . .	47

# Introduction

Aujourd'hui, beaucoup de personnes utilisent Internet chaque jour, que ce soit pour les études, le travail, le commerce ou encore la communication. Avec cette utilisation de plus en plus fréquente, avoir une bonne connexion est devenu important pour les utilisateurs. En effet, la qualité des réseaux influence directement la rapidité de connexion ainsi que la stabilité du service, ce qui rend l'amélioration des réseaux de télécommunication de plus en plus nécessaire.

L'ADSL et la fibre optique font partie des technologies les plus utilisées pour accéder à Internet. Cependant, elles ne présentent pas les mêmes performances en termes de débit et de stabilité. L'ADSL a longtemps été la technologie la plus utilisée, mais avec l'augmentation des besoins des utilisateurs, certaines limites sont apparues, notamment au niveau de la vitesse de connexion et de la fiabilité. La fibre optique permet une connexion plus rapide et plus stable. Ces dernières années, Algérie Télécom a commencé à remplacer progressivement l'ADSL par la fibre optique pour améliorer la qualité du service proposé aux utilisateurs.

La fibre optique est surtout connue pour sa rapidité, mais il reste également important d'étudier sa fiabilité. Les coupures et les pannes peuvent perturber l'utilisation d'Internet et créer des difficultés pour les utilisateurs.

C'est dans ce cadre que s'inscrit ce mémoire, qui cherche à répondre à la question suivante : Dans quelle mesure la fibre optique offre-t-elle une meilleure fiabilité et une meilleure qualité de service que l'ADSL ?.

Pour répondre à cette problématique, nous allons d'abord réaliser une analyse descriptive des données afin de comparer les pannes et l'utilisation de la bande passante sur les deux réseaux. Ensuite, des modèles ARIMA seront utilisés pour prévoir les pannes futures à partir des données observées. Enfin, nous appliquerons aussi des méthodes d'analyse de survie, comme l'estimateur de Kaplan-Meier et le modèle de Cox, afin d'étudier le temps entre les pannes et

de comparer la fiabilité des deux réseaux.

Ce mémoire est organisé en trois chapitres :

- Le premier chapitre présente les bases des réseaux informatiques ainsi que les technologies ADSL et fibre optique, avec quelques notions sur la qualité de service.
- Le deuxième chapitre s'intéresse aux méthodes statistiques utilisées dans cette étude, notamment les modèles ARIMA, le lissage exponentiel et l'analyse de survie.
- Le troisième chapitre est consacré à l'application sur des données réelles fournies par Algérie Télécom, ainsi qu'à l'analyse et l'interprétation des résultats obtenus.
- la conclusion résume les principaux résultats et propose quelques suggestions d'amélioration.

# Chapitre 1

## La qualité de réseau internet

### Introduction

Aujourd'hui, les réseaux occupent une place importante dans notre vie quotidienne. Ils relient différents appareils comme les ordinateurs ou les imprimantes qui permet de faciliter l'échange et le transfert d'informations ou de fichiers. En effet, on peut classé plusieurs types de réseaux qui varient selon leur taille par exemple certains se trouvent dans les maisons ou dans les entreprises, tandis que d'autres peuvent couvrir des villes ou des pays. La structure d'un réseau peut influencer la vitesse de transmission des données, ainsi que leur fiabilité et leur sécurité.

Avec le développement des technologies et des communications, en particulier des réseaux internet qui sont devenus essentiels, non seulement pour la communication et les loisirs, mais aussi comme outil fondamental pour améliorer la qualité de vie, grâce à la facilité d'accès à l'information, au commerce électronique, aux services numériques. La performance de ce réseau est essentielle car elle influe directement sur la satisfaction du consommateur. Ainsi, toute défaillance ou défaut du réseau affecte négativement le service, entraînant l'insatisfaction des utilisateurs et, par conséquent, une baisse du nombre de usagers, et il est fondamental de prêter attention à la qualité du réseau internet.

Ce chapitre présente les concepts généraux des réseaux, en expliquant les différents types de réseaux informatiques et leurs composants matériels et logiciels. Il parle aussi d'Internet, en présentant certains services utilisés fréquemment dans le domaine des communications qui

## 1. La qualité de réseau internet

---

permettent aux utilisateurs d'y accéder facilement. Enfin, une section qui traite de la gestion de l'Internet et des réseaux de communication en Algérie, aussi au niveau local avec ACTEL à Bouira.

### 1.1 Généralités sur les réseaux informatiques

#### 1.1.1 Les réseaux informatiques

Le réseau est défini comme un système composé d'un ensemble d'éléments interconnectés. Il constitue également un moyen de communication, permettant à une personne ou à des groupes de partager et de transférer différentes informations ou services.

Il existe plusieurs types de réseaux, parmi lesquels le réseau informatique. Ce dernier est constitué d'un ensemble d'équipements (ordinateurs, routeur, commutateurs, etc) connectés entre eux par des liaisons (filaires ou sans fil) sur le plan matériel. Sur le plan logiciel, le fonctionnement du réseau informatique est assuré par des protocoles, qui représentent un ensemble de règles définissant la manière dont les données sont transférées et traitées.

Le réseau informatique vise principalement à partager les ressources en permettant aux utilisateurs de transférer des données entre différents appareils. Ces ressources sont fichiers, de logiciels et de bases de données. Il permet également la communication entre les personnes par différents moyens, notamment le courrier électronique, ainsi que la réduction des coûts grâce au partage des ressources. Ainsi, au lieu d'attribuer des équipements indépendants à chaque utilisateur, les ressources peuvent être partagées entre plusieurs utilisateurs, notamment les périphériques tels que les imprimantes.

#### 1.1.2 Types de réseaux informatiques

Dans différents environnements, plusieurs types de réseaux informatiques sont utilisés, chaque type étant dédié à un usage spécifique afin de répondre aux besoins de chacun. Ils peuvent être classés en fonction de la taille de réseau, de la capacité de transmission et de l'étendue géographique.

## 1. La qualité de réseau internet

---

- **Le réseau local (LAN) :**

Un réseau local (en anglais Local Area Network), est un réseau qui relie principalement des équipements tels que des ordinateurs ou des serveurs, dans une zone géographique limitée telle qu'un bâtiments, une école ou une université, dans le but de partager les ressources notamment des périphériques ou des données, assurant plus de sécurité et de communications privées, utilisant la même technologie la plus répandue étant Ethernet.[7]

La taille d'un réseau local peut atteindre 100 voire 1000 machines.[13]

- **Le réseau métropolitain (MAN) :**

Un réseau métropolitain (en anglais Metropolitan Area Network), également appelé réseau fédérateur, est une interconnexion de plusieurs réseaux locaux dans une zone géographique plus étendue qu'un LAN, par exemple une ville. dans le but de partager des informations et de faciliter la communication entre les personnes sur de plus longues distances, tout en assurant la vitesse de transmission adaptée aux besoins de ces usagers.[14]

- **Le réseau étendu (WAN) :** [14]

Un réseau étendu (en anglais Wide Area Network) regroupe plusieurs types de réseaux, dont les réseaux locaux (LAN) et métropolitains (MAN), et est considéré comme le réseau le plus étendu en termes de zone géographique, couvrant plusieurs villes ou même pays. Le WAN utilisent certains équipements, tels que les routeurs, pour déterminer le chemin le plus approprié et le plus efficace pour atteindre un du réseau.

L'Internet est l'un des WAN connues.

### 1.1.3 Les composants de réseau

Le réseau informatique n'est pas seulement un ensemble d'ordinateurs ou de serveurs, mais aussi un ensemble de composants matériels et logiciels, qu'il est important d'aborder pour comprendre le fonctionnement de réseau informatique en matière de partage et de transfert

## 1. La qualité de réseau internet

---

les données, pour garantir des performances optimales et une connexion stable pour les usagers.

### 1.1.3.1 Les équipements d'interconnexion

Il existe de nombreux équipements d'interconnexion qui permet de connecter plusieurs machines entre elles, parmi lesquels :

- **La carte réseau (Network Interface Card)**

La carte réseau est une interface matérielle qui permet de connecter un ordinateur à un réseau informatique. Elle assure l'envoi, la réception et le contrôle des données transmises sous forme de paquets. Elle est généralement intégrée à la carte mère.[14]

- **Le répéteur (Repeater)**

Le répéteur est l'un des composants matériels les plus simples d'un réseau informatique. Il permet d'étendre la taille du réseau en régénérant le signal entre deux nœuds du réseau. Il assure également de connecter deux supports matériels différents.

- **Le concentrateur (Hub)**

Le concentrateur est un matériel qui permet de régénérer le signal au sein de réseau. Il comporte plusieurs ports permettant de relier des ordinateurs et des périphériques au niveau d'un réseau. Son rôle consiste à diffuser les données transmises vers l'ensemble des ports, il est parfois appelé répéteur multi-ports.[13]

- **Le commutateur (Switch)**

Le commutateur est un équipement réseau qui permet de connecter plusieurs ordinateurs entre eux, à l'instar d'un concentrateur, sa seule différence est que le commutateur est capable de déterminer le port vers lequel envoyer les données en fonction de l'adresse MAC (Media Access Control c'est l'identification de l'appareil) de l'équipement concerné.[14]

- **Le routeur (Router)**

Le routeur est un appareil qui permet de connecter des réseaux similaires ou diffé-

## 1. La qualité de réseau internet

---

rents, il assure le routage des paquets entre les équipements et les réseaux en transférant les données d'un expéditeur à un destinataire, tout en déterminant le meilleur chemin pour les acheminer, en fonction des informations contenues dans la table de routage.

[7]

### 1.1.3.2 Les supports de transmission

Les supports de transmission sont des éléments permettant de transférer les données de la source vers la destination sous forme de signaux. On distingue principalement deux types de supports : les supports filaires, tels que la paire torsadée, le câble coaxial et la fibre optique, et les supports sans fil, tels que le Wi-Fi et le Bluetooth.[19]

### 1.1.3.3 Les protocoles de communication

Les protocoles sont un ensemble de règles et de procédures qui permettent l'échange fiable de données entre les équipements d'un réseau. Ils jouent un rôle essentiel dans les réseaux informatiques en garantissant la communication et le transfert des informations. la communication sur internet est basée une suite de protocoles appelées TCP/IP (Transmission Control Protocol/Internet Protocol), qui constitue plusieurs protocoles, tels que TCP et UDP pour le transfert de données, HTTP pour les services web, IP pour l'adressage, et d'autre comme FTP pour le transfert des fichiers ou Telnet..etc.[13],

## 1.1.4 Les topologies des réseaux

La topologie est la structure selon laquelle les équipements d'un réseau sont interconnectés et organisés, on distingue deux types :

### 1.1.4.1 Les topologies physiques

Les topologies physiques constituent l'arrangement physique des éléments matériels d'un réseau. Parmi celles-ci, on trouve :

- **Topologie en bus :**

## 1. La qualité de réseau internet

---

Cette topologie repose principalement sur un seul support de transmission filaire (un câble coaxial), qui relie les différents ordinateurs et périphériques entre eux. C'est la structure de réseau la plus simple selon leur fonctionnement et leur organisation, mais elle présente un défaut : si l'un des appareils tombe en panne, cela peut entraîner l'arrêt du réseau.[13]

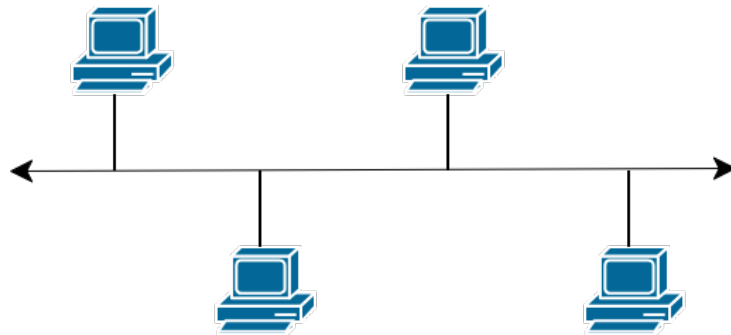


FIGURE 1.1 – Schéma d'une topologie en bus.

- **Topologie en étoile :**

On parle de la topologie la plus courante, dans laquelle tous les équipements de réseau sont connectés en étoile à un nœud central comme un commutateur (Switch) ou un concentrateur (Hub) qui permet d'assurer la communication entre les machines. [13]

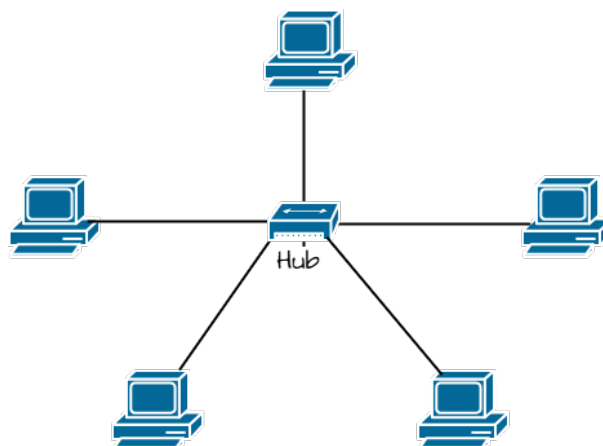


FIGURE 1.2 – Schéma d'une topologie en étoile.

## 1. La qualité de réseau internet

---

- **Topologie en anneau :**

Cette topologie repose sur une boucle fermée, où chaque appareil est relié à deux autres appareils un de chaque côté à l'aide d'un câble circulaire unique, où les données passent d'un appareil à l'autre pour atteindre leur destination dans un seul sens à travers la boucle.

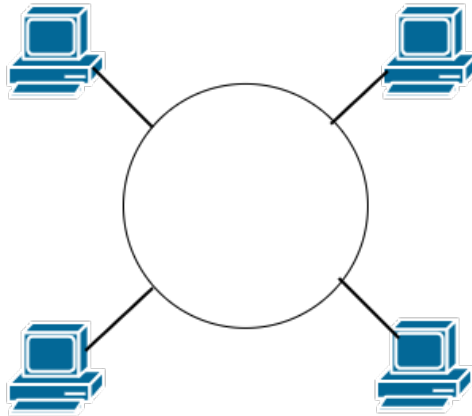


FIGURE 1.3 – Schéma d'une topologie en anneau

- **Topologie maillée :**

Dans ce type, chaque appareil est relié directement avec d'autres appareils via des câbles. Cette topologie est fiable, notamment en cas de panne de câble, les données peuvent être passées par d'autres chemins vers le destinataire. En effet, elle présente aussi quelques défauts, comme le coût élevé en raison du grand nombre de câbles.[3]

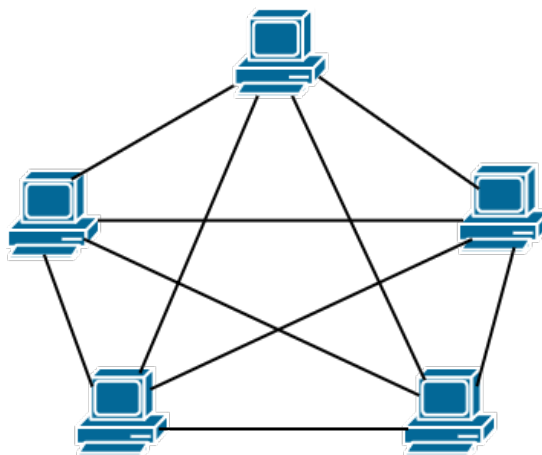


FIGURE 1.4 – Schéma d'une topologie maillée

## 1. La qualité de réseau internet

---

### 1.1.4.2 Les topologies logiques

Contrairement à la topologie physique, topologies logiques représente la manière dont les données sont transmises via les supports de communication. Les topologies logiques Les plus populaires sont Ethernet, Token Ring et FDDI.[13]

### 1.1.5 Les architectures des réseaux

L'architecture d'un réseau peut classer deux types principales selon leur fonctionnement :

#### 1.1.5.1 L'architecture poste à poste

Est également appelé en anglais « peer to peer (P2P) », tous les ordinateurs sont connectés entre eux et il n'existe pas d'appareil central chargé de les contrôler. Chaque ordinateur peut ainsi partager différentes ressources avec les autres appareils du réseau par exemple des fichiers. Cette architecture est définie par sa simplicité et par son faible coût, notamment selon des équipements ou des câbles utilisés. Cependant, elle présente aussi certains inconvénients. En effet, l'absence de centralisation peut rendre la gestion et le contrôle du réseau plus difficiles, et le niveau de sécurité est relativement faible. Ainsi, cette architecture est généralement utilisée dans les réseaux de petite taille ou lorsqu'il y a un nombre limité d'appareils.[13]

#### 1.1.5.2 L'architecture client/serveur

Dans cette architecture, les ordinateurs des utilisateurs sont tous reliés à un ordinateur central, appelé serveur, est une appareil très puissante selon de leur capacité d'entrée et de sortie. Il permet de contrôler le fonctionnement de réseau et de partage des données entre les différents utilisateurs facilement. Grâce à cette position centrale, il assure une meilleure sécurité élevée comparativement à l'architecture poste à poste. En effet, ce type d'architecture a aussi quelques problèmes, notamment le coût élevée à cause de la technologie avancée du serveur et en cas de panne de serveur peut arrêter le fonctionnement des autres appareils.[13]

### 1.2 Internet

Le mot « Internet » est un terme américaine apparu avec le réseau ARPANET, créé par le ministère de la Défense des États-Unis en 1968. Par la suite, des chercheurs et des professeurs se sont intéressés à ce réseau et ont contribué à son développement. Internet est considéré comme le plus grand réseau informatique au monde en termes d'utilisation et de diffusion. Il est également perçu comme une vaste source d'informations et un puissant moyen de communication.[32]

Aujourd'hui, Internet est devenu un outil indispensable dans la vie quotidienne des particuliers et des entreprises en raison de son développement rapide. Il est considéré comme un réseau de réseaux qui comprend de nombreux systèmes connectés entre eux, et également regroupe des millions d'ordinateurs à travers de monde. Il permet aux utilisateurs de rechercher et d'échanger facilement diverses informations et connaissances, de communiquer à distance avec d'autre personnes dans le monde et de découvrir différents services numériques tels que la messagerie, les réseaux sociaux, l'apprentissage à distance et le commerce électronique.

Le fonctionnement d'Internet repose essentiellement sur un ensemble de protocoles qui permettent un échange fiable des données, notamment la suite TCP/IP. Il existe plusieurs services Internet, parmi lesquels : Le world Wide Web, également appelé Web, est l'un des services les plus connues, il permet à les utilisateurs de naviguer sur des pages web à l'aide de programmes tels que le navigateur ou browser. Le courrier électronique (e-mail) est l'un des services les plus utilisés sur Internet, il permet à l'expéditeur d'envoyer des messages à un ou plusieurs destinataires à la fois de manière sécurisée et rapide.[13]

Il existe plusieurs technologies pour accéder à Internet qui diffèrent selon plusieurs indicateurs, notamment le débit. Les plus connues et les plus utilisées étant l'ADSL et la fibre optique, qui permettent aux usagers de se connecter au réseau Internet et d'obtenir une qualité de service élevée, et se distinguent par des caractéristiques de qualité de service comme le débit, latence et le délai de transmission.

### 1.2.1 Technologies d'accès Internet

L'accès à Internet et son utilisation dans plusieurs domaines par les particuliers et les institutions se font grâce à différents technologies modernes qui facilitent cette opération, parmi lesquelles les plus connues sont l'ADSL et la fibre optique.

#### 1.2.1.1 L'ADSL

xDSL est un ensemble de technologies qui permettent d'accéder à Internet via une ligne téléphonique en cuivre. Il existe plusieurs types de cette technologie qui se distinguent par le débit et la qualité de service.[20]

ADSL est l'un des principaux types de xDSL, le ADSL signifie «Asymmetric Digital Subscriber Line »[8], en français, on l'appelle ligne d'abonné numérique asymétrique, est une technologie d'accès à Internet à haut débit. Elle permet à l'utilisateur de rester connecté à Internet et de transférer différents types de données via les fils de cuivre du téléphone fixe. Il est ainsi possible d'utiliser Internet et de passer des appels téléphoniques simultanément, sans que l'un n'affecte l'autre, car cette technologie utilise des bandes de fréquences différentes pour chaque service, les fréquences basses attribuées aux appels, tandis que les fréquences élevées sont utilisées pour le transfert de données.

De plus, la technologie ADSL permet de transférer la voix et différents types de données via les paires de fils de cuivre des lignes téléphoniques de la Boucle Local.[20]

La technologie ADSL est dite asymétrique afin de distinguer deux types de débit : débit montant, correspondant à l'envoi de données, et le débit descendant, utilisé lors de la navigation sur des sites Internet ou du téléchargement de fichiers, le débit descendant est généralement supérieur au débit montant.[8]

Plusieurs facteurs influencent directement de débit et la qualité de la technologie ADSL, notamment la distance entre le centre téléphonique et le domicile ou l'établissement de l'utilisateur. En effet, le débit est affecté par cette distance : plus celle-ci est large, plus le débit diminue. En plus de la qualité des équipements, parmi lesquelles : le modem, le routeur ou les câbles en cuivres.

La technologie ADSL s'est largement répandue et est l'une des plus utilisées en raison de

## 1. La qualité de réseau internet

---

son faible coût et de ses débits élevés. Il existe plusieurs types d'ADSL, notamment l'ADSL G.lite, l'ADSL2 et l'ADSL 2+ [20]. Ces dernières années, cette technologie a commencé à disparaître et à reculer progressivement, en raison du développement de technologies offrant des hauts débits, notamment la fibre optique.

### 1.2.1.2 La fibre optique

antes et les plus utilisées. Elles se composent d'un fil central en verre ou en plastique très fin appelé «cœur », entouré d'une gaine optique, elle-même recouverte d'une gaine protectrice en plastique. Le cœur et la gaine optique sont constitués de matériaux transparents et isolants électriquement, généralement du verre, ayant des indices de réfraction différents, afin de protéger la fibre contre les facteurs externes et d'assurer la diffusion de la lumière à l'intérieur du cœur.[22]

En effet, ces fibres optiques se caractérisent par leur capacité de conduire la lumière, qui permet la transmission des données sous forme de signaux lumineux sur de longues distances et à des vitesses très élevées comparativement à d'autres supports de transmission basés sur les câbles en cuivre tels que l'ADSL.[22]

Les fibres optiques peuvent être classées en deux types différents selon diamètre de cœur et domaines d'utilisation. : les fibres monomodes et les fibres multimodes. Les fibres monomodes se caractérisent par un diamètre de cœur très petit (environ 10 microns). Elles utilisent généralement une source laser, qui permet la transmission de la lumière dans un seul mode, ce qui réduit la dispersion modale et améliore la qualité du signal et favorise les communications à longue distance à très haut débit, notamment pour relier des villes ou des pays.[20],[22]

Le second type est la fibre multimode, qui se distingue par un diamètre de cœur plus grand que la fibre monomode, généralement compris entre 50 et 62.5 microns. Un faisceau lumineux traverse le cœur de la fibre d'une extrémité à l'autre, permettant la transmission de données dans plusieurs modes à la fois. Cette structure peut entraîner une dispersion modale. Ce type de fibre est utilisé dans les réseaux locaux et les communications à courte distance.[20],[22]

De plus, les technologies d'accès basées sur la fibre optique peuvent être désignées par l'acronyme FTTx (Fiber to the x), où le x signifie le point d'accès de la fibre. Par exemple,

## 1. La qualité de réseau internet

---

FTTH indique « fibre jusqu'au domicile », tandis que FTTO désigne l'accès pour les entreprises ou les institutions.

La fibre optique présente plusieurs avantages, dont les plus importantes : le très haute débit et la capacité à transférer de grandes quantités de données simultanément. Elle présente toutefois certains inconvénients, notamment son coût élevé et la fragilité des fibres.

### 1.3 Qualité de réseau internet

La qualité est une notion importante et a plusieurs sens, la qualité est défini par le dictionnaire de l'Académie Française comme une chose qui est bonne ou mauvaise, grande ou petite, chaude ou froide, blanche ou noire.etc [4].

D'autre manière, la qualité est un ensemble des spécifications relatives à un produit ou un service, qui répondent aux besoins et attentes des consommateurs et sont exemptes de défauts. Elle est également considérée comme une notion importante et fondamentale dans les réseaux informatiques modernes, notamment internet. La qualité de réseau internet est une mesure de la performance du réseau et de sa capacité à répondre aux besoins des utilisateurs, ne se limite pas à la vitesse de connexion, mais englobe plusieurs indicateurs techniques, telle que la quantité des données et sa vitesse de transfert.

La qualité du réseau internet peut être mesurée grâce à des indicateurs de qualité de service (QoS).

### 1.4 Qualité de service (QoS)

Les réseaux modernes se caractérisent par de nombreuses propriétés, dont la plus importante est la qualité de service. celle-ci se traduit par la capacité du réseau à transférer les données de l'expéditeur au destinataire avec un bon niveau de performance qui garantit la satisfaction des utilisateurs, en offrant une vitesse de transfert adéquate tout en réduisant les retards et le taux de perte de données, en particulier pour les applications et les services sensibles tels que les vidéos et les appels téléphoniques.

## 1. La qualité de réseau internet

---

La qualité de service (QoS : quality of service en anglais ) est un ensemble de mesures et de technologies modernes qui régulent le trafic et gèrent la capacité du réseau afin de garantir le bon fonctionnement de tous les services sans interruption, ce qui aide le consommateur à déterminer le meilleurs services.[18]

Dans le cadre de la gestion efficace des services, la qualité de service représente un grand défi pour les fournisseurs. Elle repose sur un ensemble de paramètres à la fois qualitatifs et quantitatifs. Les paramètres qualitatifs permettent de décrire la qualité perçue par l'utilisateur final comme les coupures, la clarté de voix et la fluidité de la vidéo. Au contraire, Les paramètres quantitatifs, quant à eux permettent de mesurer le niveau de qualité relatif au réseau, notamment la latence des paquets .[28]

### 1.4.1 Les critères de QoS

Pour mieux comprendre le terme «qualité de service »et son fonctionnement, il est important ses paramètres et ses instruments de mesure, qui permettant de décrire et donne une description des performances de réseau, et aussi d'améliorer celles des applications.

Les critères de QoS peuvent généralement être utilisées pour évaluer la qualité du réseau internet, et sont définies comme suivant :

#### 1. **Le débit** : ( throughput en anglais)

Est la quantité de données pouvant être transférées par une unité de temps, et affectée par des facteurs tels que la perte de paquets et les retards. Il est également représenté par le taux de transmission (vitesse de transfert) des données d'une source vers la destination sur un réseau. Généralement, on mesure le débit en bits et leurs multiples par seconde.[31]

#### 2. **La perte de paquets** : (packet loss en anglais)

Est un phénomène réseau qui se produit lorsque les paquets de données transférés n'atteignent pas leur destination pour plusieurs raisons, notamment lorsque le volume de données est dépasse la capacité de réseau, ce qui entraîne une congestion, ainsi qu'en raison de qualité et de fiabilité des connexions dans les réseaux internet.[18]

## 1. La qualité de réseau internet

---

### 3. **Le délai** : ( delay en anglais)

Appelé aussi la latence ou le temps de réponse, représente l'intervalle de temps nécessaire à la transmission d'un paquet de données d'une source vers une destination, il est exprimé en fonction de la capacité des liens de connexion et de la charge de service, lorsque cette dernière augmente, le délai augmente également.[31]

### 4. **La gigue** : (jitter en anglais)

Est un critère important pour les applications sensibles telles que l'audio et la vidéo, elle représente aussi une variation de latence (délai) des paquets de données transmis de l'émetteur vers le récepteur, et sa valeur doit être très faible pour garantir une bonne qualité de service.[18]

La bande passante désigne la capacité maximale de transmission d'un canal de communication, généralement mesurée en méga bits par seconde (Mbps). Elle ne doit pas être confondue avec le débit, qui correspond à la quantité réelle de données transmises.

## 1.4.2 Concepts : congestion et fiabilité

### **Définition 1.4.1.** (*La congestion*)

*La congestion est une situation où le réseau est trop chargé par les données en même temps. Cela peut réduire la vitesse de transmission des informations ou provoquer la perte de certains paquets.*

### **Définition 1.4.2.** (*La fiabilité*)

*La fiabilité est la capacité d'un réseau à transmettre des données et des informations de manière efficace et sans interruption pendant une période donnée.*

## 1.4.3 Impact de la qualité sur la demande

Dans les réseaux modernes, dont fait partie Internet, plusieurs éléments interviennent, dont le plus important est la qualité, considérée comme une mesure du niveau de performance du réseau. Elle peut influencer le comportement du client à travers un ensemble de critères liés à

## 1. La qualité de réseau internet

---

la qualité de service (QoS) utilisés aujourd'hui par les entreprises, notamment le débit, dont toute dégradation dans ce dernier peut affecter la capacité du client à utiliser le réseau de manière régulière. En outre, les retards et les interruptions fréquentes peuvent entraîner une insatisfaction des besoins de l'utilisateur, le conduisant éventuellement à changer d'offre ou à ne pas renouveler son abonnement.

Il existe une relation étroite entre la qualité et la demande : lorsque la qualité du service est élevée, la demande et le nombre d'abonnés augmentent, tandis que lorsqu'elle est faible, la demande stagne ou diminue, ce qui peut porter atteinte à l'image de l'entreprise. La qualité de service est donc un facteur important qui influe directement sur l'évolution de la demande.

La prévision de la demande future ne dépend pas uniquement du nombre d'abonnés ou du facteur temps, mais également des paramètres de qualité du service qui contribuent à l'augmentation ou à la diminution de la demande, ainsi, une amélioration de la qualité à un moment donné entraîne une augmentation de la demande à l'avenir, et inversement.

### **1.5 Gestion de l'Internet en Algérie et au niveau de l'AC-TEL de Bouira**

#### **1.5.1 Au niveau d'Algérie**

Aujourd'hui, avec le développement rapide des réseaux de communication, Internet est devenu le réseau le plus utilisé au monde. En Algérie, sa gestion s'effectue à deux niveaux distincts : au niveau national, principalement par l'État, qui met en place des plans et des lois afin de garantir la couverture d'Internet sur l'ensemble du territoire, d'assurer la protection des données sous toutes leurs formes ainsi que leur transfert à travers des réseaux sécurisés, et de définir les normes de qualité de service, notamment en matière de débit et de temps de latence.

Algérie Télécom est le principal fournisseur de ce réseau. Elle applique principalement ces orientations sur le terrain pour assurer le bon fonctionnement du service. L'entreprise prend en charge l'installation des infrastructures adaptées, qu'il s'agisse de câbles en cuivre ou de fibre optique, et veille à l'amélioration continue des performances du réseau ainsi qu'au suivi

## 1. La qualité de réseau internet

---

de sa maintenance, afin de garantir la satisfaction des utilisateurs et la continuité d'un service de qualité. Par ailleurs, l'administration nationale définit également des mécanismes et des procédures destinés à traiter les pannes et à assurer leur réparation dans les meilleurs délais.

### 1.5.2 Au niveau de l'ACTEL

En Algérie, l'Internet est géré au niveau local par les agences commerciales d'Algérie Télécom, appelées Actel. Ces agences représentent un point de contact direct entre l'opérateur et les utilisateurs. Elles sont présentes dans toutes les wilayas d'Algérie et assurent l'application des décisions prises au niveau national.

Au niveau local, Actel assure plusieurs missions. Elle reçoit les demandes d'abonnement aux différentes technologies d'accès à Internet et prend en charge les nouvelles inscriptions. Elle intervient également dans le suivi des réclamations liées aux pannes du réseau et participe aux opérations de maintenance pour garantir la continuité du service et donc la satisfaction des usagers.

Par ailleurs, ces agences contrôlent également les indicateurs de la qualité de service (QoS), notamment le débit et le latence. Ces dernières années, on constate une évolution importante de la technologie de la fibre optique, qui permet d'offrir des vitesses de transmission plus élevées et mieux adaptées aux besoins croissants des utilisateurs.

## Conclusion

La qualité d'un réseau Internet est liée fortement à la qualité de service. Elle dépend de plusieurs paramètres comme le débit, le délai, la perte de paquets, la gigue et la stabilité des communications. En effet, ces paramètres sont essentiels pour la satisfaction des utilisateurs ce qui peut entraîner une augmentation ou une baisse de la demande, il faut les prendre en considération.

# Chapitre 2

## Les outils d'analyse de données

D'une certaine manière, l'analyse des données peut être considérée comme un domaine vaste, qui fait également partie des disciplines récentes s'appuyant sur plusieurs méthodes statistiques multidimensionnelles utilisant de grands tableaux, qu'elle transforme en graphiques et en diagrammes, pour mieux représenter et comprendre les données. En effet, ces méthodes ont commencé à se développer au XXe siècle notamment pour répondre aux besoins de la biométrie et de la psychométrie, et ont atteint un bon niveau de développement parallèlement aux statistiques mathématiques. Parmi leurs fondateurs figurent K. Pearson et C. Spearman (1990), ainsi que ceux qui ont poursuivi leur développement : R. Fisher et H. Hotelling (1930). En France, on peut également citer J.P. Benzécri (1960). [30]

Par la suite, après le développement général des méthodes statistiques, une première école exploratoire est apparue dans les années 70 sous le nom d' « analyse des données à la française », où elle s'est largement répandue, suscitant l'intérêt de certains statisticiens et le rejet d'autres. Elle a été fondée par le mathématicien français Jean-Paul Benzécri, qui a suivi une approche particulière en accord avec sa propre vision et qui était un élève de Cartan. L'idée de cette école portait sur les données, à la différence de la statistique classique qui s'appuie d'abord sur un modèle mathématique, ce qui a conduit à la découverte de l'analyse des correspondances, une méthode d'analyse factorielle des tableaux. Grâce à cette évolution et à l'introduction de différentes méthodes, un exemple a été appliqué pour la première fois sur un ordinateur IBM 1620. Cette école a réussi à combiner statistiques et structures, elle ne s'intéresse pas uniquement aux calculs, mais aussi à l'observation et à la compréhension des relations entre les données, et se caractérise par son approche empirique. [1]

## 2. Les outils d'analyse de données

---

Actuellement, le domaine de l'analyse des données continue d'évoluer fortement, donnant naissance à la science des données (Data science), qui englobe les données, leur analyse et les statistiques. Elle a également conduit à l'utilisation d'outils modernes de prévision des séries chronologiques à l'aide des statistiques descriptives et de la programmation, notamment via des logiciels tels que R ou Python.

Ce chapitre présente une généralité sur les séries temporelles et quelques outils d'analyse des données, comme les modèles ARIMA et le lissage exponentiel pour la prévision. Il présente aussi l'analyse de survie avec l'estimateur de Kaplan-Meier et le modèle de Cox.

### 2.1 Analyse des séries temporelles

#### 2.1.1 Définition d'une série temporelle

##### Définition 2.1.1.

*Une série chronologique n'est pas simplement une suite de chiffres dans le temps, c'est l'observation des valeurs réalisées par un processus aléatoire ( $Y_t$ ) sur une période donnée (minute, jour, mois ou année). Son objectif est d'expliquer, de décrire et de prévoir un phénomène futur, et elle peut être appliquée en médecine, en économie, voire en démographie.*

#### 2.1.2 Les composantes d'une série temporelle

La série chronologique se caractérise par trois composantes fondamentaux, à savoir :

##### 1. La tendance :

Elle représente l'évolution des données d'une série au fil du temps. Elle permet d'observer la variation du phénomène : lorsque la variation moyenne de ( $Y_t$ ) augmente, on parle d'une tendance croissante ; lorsqu'elle diminue, on parle d'une tendance décroissante. Elle peut être linéaire, exponentielle ou parabolique et est généralement notée par ( $T_t$ ).

## 2. Les outils d'analyse de données

---

### 2. La saisonnalité :

Désigne les fluctuations observées dans la série sous la forme de pics récurrents sur le graphique, qui se manifestent de manière périodique et régulière à des intervalles de temps, généralement mensuels, trimestriels ou annuels. On notée par  $(S_t)$ .

### 3. Composante irrégulière :

Est la partie aléatoire d'une série, qui reflète les variations soudaines et imprévisibles et correspond également à l'erreur. Elle peut résulter de crises économiques ou d'autres facteurs. Elle représente tout ce qui reste une fois que la tendance et la saisonnalité ont été retirées. On notée par  $(\varepsilon_t)$ .

### 2.1.3 Décomposition d'une série temporelle

Pour décomposer  $(Y_t)$  on distingue deux schéma :

#### 1. Schéma additif :

Le modèle additif est utilisé lorsque les variations ne dépendent pas du niveau de la tendance et restent constantes dans le temps, et s'écrit sous la forme de la somme des trois composantes précédentes :

$$Y_t = T_t + S_t + \varepsilon_t.$$

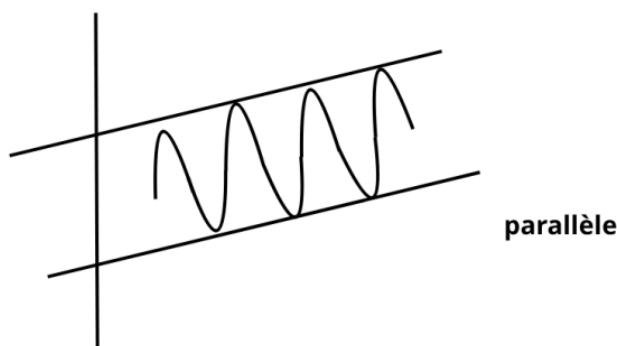


FIGURE 2.1 – Schéma d'une modèle additif

### 2. Schéma multiplicatif :

Le modèle multiplicatif est utilisé lorsque les variations dépendent du niveau de la tendance et changent avec celui-ci, et s'écrit sous la forme du produit des trois composantes précédentes :

$$Y_t = T_t \times S_t \times \varepsilon_t.$$

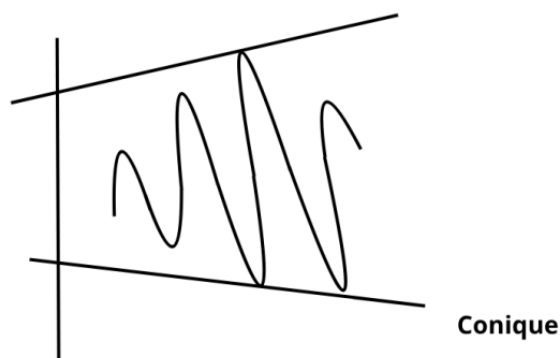


FIGURE 2.2 – Schéma d'une modèle multiplicatif

**Remarque 2.1.1.** *Il existe un schéma dit « mixte », qui combine les schémas additif et multiplicatif, dans lequel  $\varepsilon_t$  est indépendant et ne dépend ni de la tendance ni de la saisonnalité et s'écrit sous la forme suivante :*

$$Y_t = T_t \times S_t + \varepsilon_t.$$

### 2.1.4 La stationnarité

**Définition 2.1.2.** *On dit qu' une série  $Y_t$  est stationnaire, lorsqu'elle ne comporte ni de tendance ni de saisonnalité.*

#### Définition 2.1.3.

*Une série chronologique est dite stationnaire si ses fluctuations restent constantes dans*

## 2. Les outils d'analyse de données

---

le temps, en d'autres termes, son graphique indique que les valeurs tournent autour d'une moyenne stable.

Mathématiquement, on peut dire qu'une série stationnaire  $Y_t$  est une réalisation d'un processus stationnaire, ce qui signifie que :

$$E(Y_t) = \mu, \quad \text{Var}(Y_t) = \sigma^2.$$

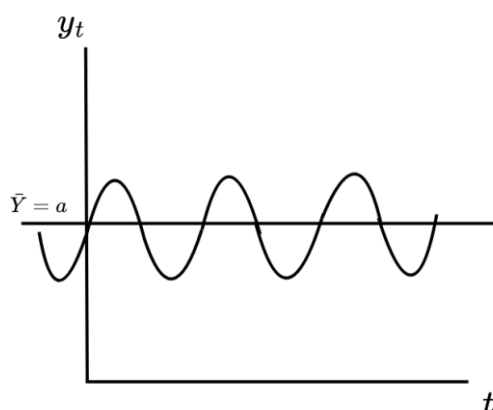


FIGURE 2.3 – Série stationnaire

**Exemple 2.1.1.** La série définie par  $Y_t = \varepsilon_t$ , est appelé série bruit blanc, et si  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$  alors la série est un bruit blanc gaussien.

Un bruit blanc est un série stationnaire.

### 2.1.5 Les modèles ARIMA (p,d,q)

#### Définition 2.1.4.

Les modèles ARIMA (Auto-Regressive Integrated Moving Average), comptent parmi les principales méthodes d'analyse des données et de prévision des séries chronologiques. En effet, ils ont été développés et popularisés par Box et Jenkins en 1976, ce sont des modèles statistiques stationnaires utilisés pour analyser et étudier des phénomènes ainsi que pour établir des prévisions, à partir de ces séries qui s'appuient sur le passé et sont liées au présent.

## 2. Les outils d'analyse de données

---

### 2.1.5.1 Les composantes

Les modèles ARIMA se composent de trois éléments fondamentaux, à savoir le processus auto-régressif (AR), la différenciation (I) et la moyenne mobile (MA). Chacun de ces éléments joue un rôle spécifique dans la modélisation des séries chronologiques.

#### Auto-régressif AR(p) [27]

Le processus auto-régressif d'ordre  $p$  est stationnaire et s'appuie sur la même série, c'est-à-dire qu'il utilise les valeurs précédentes de la variable avec un bruit blanc  $\varepsilon_t$ . Ainsi, l'effet de l'erreur diminue progressivement, mais ses traces persistent et s'écrit sous la forme suivant :

$$Y_t = \varepsilon_t + \sum_{i=1}^p \varphi_i Y_{t-i}.$$

avec  $\varphi_i$  sont des coefficients auto-régressif ( $\varphi_i \neq 0$ ).

#### Exemple 2.1.2.

1. Le modèle Auto-Régressif d'ordre 1 ( $p=1$ ), AR(1) s'écrit sous la forme suivant :

$$Y_t = \varepsilon_t + \varphi_1 Y_{t-1}.$$

2. Le modèle Auto-Régressif d'ordre 3 est s'écrit par :

$$Y_t = \varepsilon_t + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3}.$$

*Dans le modèle AR(3), la valeur actuelle dépend de trois valeurs précédentes.*

#### Différenciation I(d) [9]

Les modèles ARIMA sont utilisés pour prédire les valeurs futures d'un phénomène. La condition préalable est que la série chronologique soit stationnaire, c'est-à-dire qu'elle ne change pas dans le temps, ce qui permet d'obtenir de bonnes prévisions en se basant sur le passé.

## 2. Les outils d'analyse de données

---

En effet, lorsque la série est non stationnaire, on applique la méthode des différences pour la rendre stationnaire. De plus, le processus obtenu est alors dit intégré, ce qui correspond au terme « integrated » ou à la composante « I » du modèle ARIMA.

Le processus intégré (Integrated) d'ordre  $d$  est noté par  $I(d)$ , avec  $d$  est le nombre de différences.

### Exemple 2.1.3.

1. Le processus intégré d'ordre 1 est défini par la marche aléatoire, s'écrit comme :

$$Y_t = Y_{t-1} + \varepsilon_t$$

Après la première différenciation on obtient un bruit blanc stationnaire. Ainsi, ce processus peut être représenté par un modèle ARIMA(0,1,0).

$$Y_t - Y_{t-1} = \varepsilon_t.$$

2. Le processus  $I(2)$  signifie que la série doit être différenciée deux fois pour devenir stationnaire, c'est-à-dire la différence de la différence.

la seconde différence écrit par :

$$(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}).$$

Après la deuxième différenciation on obtient un processus stationnaire, ce processus peut être représenté par un modèle ARIMA(0,2,0).

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t.$$

### Remarque 2.1.2.

Si l'ordre de différenciation est nul ( $d=0$ ), on obtient un modèle ARMA( $p,q$ ), noté ARIMA( $p,0,q$ ).

### Moyenne mobile MA( $q$ ) [27]

Le processus moyenne mobile (Moving Average) noté par  $MA(q)$ , est un mélange d'erreurs

## 2. Les outils d'analyse de données

---

actuelles et passées calculé sous forme d'une somme pondérée des erreurs récentes. Les erreurs aléatoires influencent la série, mais leur effet reste limité au-delà d'un ordre  $q$ , correspondant à l'ordre de la moyenne mobile, elles n'ont plus d'impact.

$$Y_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}.$$

avec les  $\theta_i$  sont des coefficients de moyenne mobile ( $\theta_i \neq 0$ ) et  $\varepsilon_t$  un bruit blanc.

### Exemple 2.1.4.

1. Un processus de moyenne mobile d'ordre 1, c'est-à-dire  $MA(1)$  est écrit comme suit :

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

2. Dans le processus  $MA(3)$ , la valeur actuelle dépend de trois erreurs précédentes et écrit par la forme :

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}.$$

### 2.1.5.2 Le processus ARIMA

Le modèle ARIMA combine les modèles AR et MA pour obtenir un modèle intégré, généralement est noté par  $ARIMA(p,d,q)$ , qui s'écrit comme suit :

$$Y_t = AR + MA$$

$$\Delta^d Y_t = \sum_{i=1}^p \varphi_i \Delta^d Y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}.$$

Le terme  $\Delta^d Y_t$  indique que la série est devenue stationnaire après la différenciation.

## 2. Les outils d'analyse de données

---

### 2.1.5.3 La non stationnarité

Selon [12]. Les séries chronologiques présentent plusieurs caractéristiques, dont la stationnarité. Selon Nelson et Plosser (1982), les séries non stationnaires se divisent en deux types : les séries TS (Trend Stationary) et les séries DS (Difference Stationary).

Les séries TS sont stationnaires autour d'une tendance déterministe. Elles peuvent être représentées comme une fonction déterministe du temps, à laquelle s'ajoute un processus stationnaire ( $E(Y_t) = \mu$ ,  $Var(Y_t) = \sigma^2$ ). Ce type de série peut être rendu stationnaire en supprimant la tendance déterministe.

Le deuxième type correspond aux séries DS, dites stationnaires en différences, qui se caractérisent par une tendance stochastique. Ces séries deviennent stationnaires après application de la différenciation d'ordre  $d$ .

La différence entre les processus TS et DS réside dans la nature de la tendance : une tendance déterministe dans le cas TS, et une tendance stochastique dans le cas DS. Cette dernière est associée à la présence d'une racine unitaire, comme dans le cas d'une marche aléatoire.

Les tests de racine unitaire permettent de déterminer si une série chronologique est stationnaire ou non, ainsi que de préciser le type de non stationnarité (TS ou DS). Parmi ceux-ci, on trouve le test Dickey-Fuller Augmenté 1981 (ADF) et le test de Phillips et Perron 1988.

Le test ADF repose sur les hypothèses suivantes :

$$\begin{cases} H_0 : \text{La présence d'une racine unitaire (la série n'est pas stationnaire).} \\ H_1 : \text{absence de racine unitaire (la série est stationnaire)} \end{cases}$$

Si la  $p$ -value est supérieure au seuil de signification  $\alpha$ , généralement égale à 5%, on accepte  $H_0$  et la série n'est pas stationnaire ; sinon, on rejette  $H_0$  et la série est stationnaire.

### 2.1.5.4 La méthodologie de Box-Jenkins

Est une approche d'analyse des séries temporelles Box et Jenkins qui permet de construire le modèle ARIMA le plus approprié pour la prévision. Elle comprend quatre étapes fondamen-

## 2. Les outils d'analyse de données

---

tales :[26]

### 1. Identification :

d'abord, on teste la stationnarité de la série temporelle avec ADF ou PP. Si elle n'est pas stationnaire, on applique la différenciation pour la rendre stationnaire et trouver l'ordre  $d$ . Par la suite, on détermine les paramètres  $p$  et  $q$  grâce à l'ACF (fonction d'autocorrélation) et la PACF (fonction d'autocorrélation partielle). L'ACF permet d'identifier l'ordre de la moyenne mobile  $q$ , tandis que la PACF permet de déterminer l'ordre autorégressif  $p$ .

À partir des graphiques ou des corrélogrammes, on peut identifier ces paramètres facilement. En effet, pour un modèle  $MA(q)$ , l'ACF s'arrête au rang  $q$ , tandis que la PACF décroît progressivement vers zéro. Mais pour un modèle  $AR(p)$ , la PACF s'arrête au rang  $p$  alors que l'ACF décroît progressivement vers zéro.

### 2. Estimation :

Estimation : Une fois le modèle défini lors de la première étape, cette phase consiste à estimer les coefficients auto-régressifs ( $\varphi_i$ ) et de la moyenne mobile ( $\theta_i$ ) à l'aide de différentes méthodes, notamment la méthode du maximum de vraisemblance ou celle des moindres carrés.

### 3. Diagnostic :

À cette étape, on s'assure que le modèle choisi est valide et adapté aux données avant de l'utiliser pour la prévision. La validité du modèle est vérifiée à travers deux conditions : les résidus aléatoires ne sont pas corrélés, c'est-à-dire qu'il s'agit d'un bruit blanc à l'aide du test de Ljung-Box. Le deuxième point est de vérifier la significativité des coefficients à l'aide du test de Student (t-test).

### 4. Prévision :

C'est la dernière étape de la méthodologie Box-Jenkins, au cours de laquelle on obtient un modèle valide et adapté grâce aux étapes précédentes, il est donc possible de l'utiliser pour obtenir une prévision fiable.

### 2.1.6 Lissage exponentiel

Les méthodes de lissage exponentiel (Exponential Smoothing) constituent une méthode de prévision des séries chronologiques à court terme. Elles ont été introduites par Holt en 1958 et Winters en 1960, puis popularisées par Brown en 1963. [25]

En effet, ces méthodes se basent sur les valeurs passées, en accord avec une importance relative aux observations les plus récentes et moins aux plus anciennes, avec la décroissance des poids étant exponentielle. On trouve trois types de lissage : simple, double (Holt) et Holt-Winters, chacun dépendant d'un paramètre compris entre 0 et 1.

#### 2.1.6.1 Lissage exponentiel simple

D'après [25]. Le lissage exponentiel simple est défini comme un modèle de prévision des séries temporelles ne présentant ni tendance ni saisonnalité, et qui s'écrit sous la forme suivante :

$$\hat{Y}_T = \alpha \cdot \sum_{i=0}^{T-1} (1 - \alpha)^i \cdot X_{T-1}.$$

Avec  $\alpha$  est une constante de lissage tel que  $0 < \alpha < 1$ .

$\hat{Y}_T(h)$  désigne la prévision de la valeur future à l'instant  $T + h$ , où  $T$  représente la date et  $h$  l'horizon. Dans ce modèle, la série ne comporte ni tendance ni saisonnalité, ainsi, le futur ne change pas avec l'horizon et toutes les prévisions restent constantes, quelle que soit la valeur de  $h$ . Ainsi :

$$\hat{Y}_T(h) = \hat{Y}_T.$$

#### La formule de mise à jour :

Lorsqu'une nouvelle observation est disponible, cette formule permet de calculer une nouvelle prévision à la date  $T$  à partir de la prévision précédente  $\hat{Y}_{T-1}$ . Elle s'écrit sous les deux formes suivantes :

## 2. Les outils d'analyse de données

---

$$\hat{Y}_T = \alpha \cdot Y_T + (1 - \alpha) \cdot \hat{Y}_{T-1}.$$

$$\begin{aligned}\hat{Y}_T &= \hat{Y}_{T-1} + \alpha(Y_T - \hat{Y}_{T-1}). \\ &= \hat{Y}_{T-1} + \alpha \cdot \varepsilon_T.\end{aligned}$$

Avec  $\varepsilon_T$  est l'erreur de prévision à l'instant T,  $\varepsilon_T = Y_T - \hat{Y}_{T-1}$ .

### Choix de la constante :

Le choix de  $\alpha$  est très important pour obtenir une prévision correcte, car il permet de donner plus de poids aux valeurs récentes de la série, qui sont les plus proches de l'avenir.

En effet, on peut minimiser l'erreur quadratique moyenne (MSE) ou l'erreur absolue moyenne (MAE) pour déterminer le paramètre  $\alpha$  et on peut également préférer la MAE, car elle accorde moins d'importance aux grandes erreurs que la MSE.

$$MSE = \frac{1}{n} \sum_{t=T-n+1}^T \varepsilon_t^2.$$

$$MAE = \frac{1}{n} \sum_{t=T-n+1}^T |\varepsilon_t|.$$

Où n représente le nombre de valeurs récentes.

### 2.1.6.2 Lissage exponentiel double

[16]. Le lissage exponentiel double, également connu sous le nom de méthode de Holt, est utilisée pour la prévision des séries présentant une tendance mais pas de saisonnalité, à l'instant t, on peut approximer la série par une droite d'équation :

$$Y_t = a + b(t - T).$$

Où  $a$  représente l'ordonnée de la droite de prévision à un instant T, appelée le niveau, et  $b$  représente la tendance (ou pente selon certaines sources). Ils écrivent comme suit :

## 2. Les outils d'analyse de données

---

$$a_T = \alpha Y_t + (1 - \alpha)(a_{T-1} + b_{T-1}).$$

$$b_T = \beta(a_T - a_{T-1}) + (1 - \beta)b_{T-1}.$$

Avec  $\alpha$  et  $\beta$  sont des paramètres de lissage tels que  $0 < \alpha, \beta < 1$ .

L'équation de prévision à l'horizon  $h$  par la méthode de Holt est s'écrit :

$$\hat{Y}_T(h) = a(T) + h \times b(T).$$

### 2.1.6.3 Méthode Holt-Winters

La méthode de Holt-Winters est un modèle de prévision des séries temporelles présentant à la fois une tendance et une saisonnalité, en l'absence de saisonnalité, elle se réduit à la méthode de Holt, c'est-à-dire au lissage exponentiel double.

Il existe deux variantes de cette méthode selon le type de composante saisonnière : le modèle additif et le modèle multiplicatif.

**le modèle additif :** [16]

Dans le modèle additif, les variations saisonnières ne dépendent pas du niveau de la série et restent constantes. Cette composante correspond à une valeur absolue dans la série observée.

$$a_T = \alpha(Y_T - s_{T-m} + (1 - \alpha)(a_{T-1} + b_{T-1})).$$

$$b_T = \beta(a_T - a_{T-1} + (1 - \beta)b_{T-1}).$$

$$s_t = \gamma(Y_T - a_T) + (1 - \gamma)s_{T-m}.$$

La prévision à l'horizon  $h$  par ce modèle est s'écrit :

$$\hat{Y}_T(h) = a_T + h \times b_T + s_{T+h-m(k+1)}.$$

Où  $k = \frac{h-1}{m}$ ,  $m$  représente la période saisonnière (par exemple  $m = 12$  pour les données

## 2. Les outils d'analyse de données

---

mensuelles),  $s_T$  la composante saisonnière et  $\gamma$  son paramètre, avec  $0 < \alpha, \beta, \gamma < 1$ .

**le modèle multiplicatif :** [16]

Dans le modèle multiplicatif, les variations saisonnières dépendent du niveau de la série. En effet, cette composante correspond à un pourcentage par rapport à la série observée.

$$a_T = \alpha \frac{Y_T}{s_{T-m}} + (1 - \alpha)(a_{T-1} + b_{T-1}).$$

$$b_T = \beta(a_T - a_{T-1}) + (1 - \beta)b_{T-1}.$$

$$s_T = \gamma \frac{Y_T}{a_{T-1} + b_{T-1}} + (1 - \gamma)s_{T-m}.$$

La prévision à l'horizon  $h$  par ce modèle est s'écrit :

$$\hat{Y}_T(h) = (a_T + h \times b_T) \times s_{T+h-m(k+1)}.$$

## 2.2 Analyse de survie

L'analyse de survie constitue une branche des statistiques initialement appliqué en démographie, où l'on se basait sur les registres de mortalité pour estimer l'espérance de vie et les probabilités de décès. Avec le temps, elle s'est élargie à divers secteurs comme l'assurance, la fiabilité et la santé.

En effet, dans le domaine de la fiabilité, Weibull a développé un modèle paramétrique basé sur la loi de Weibull devenue importante dans l'analyse de survie.

Plus tard, Kaplan-Meier a proposé un modèle non paramétrique, alors que Cox a développé un modèle semi-paramétrique. [23]

L'analyse de survie consiste à étudier la durée avant la survenue d'un événement, appelée durée de vie. Dans le domaine médical, elle est utilisée pour étudier le temps de survie des patients. Elle est également appliquée dans le domaine de la fiabilité pour étudier la durée de panne d'une machine, notamment dans ce mémoire où elle est appliquée à l'étude des pannes des réseaux ADSL et fibre optique.

### 2.2.1 Les concepts fondamentaux

Pour bien comprendre l'analyse de survie, il est important de connaître plusieurs concepts :

#### 2.2.1.1 La durée de survie

##### Définition 2.2.1. [10]

*La durée de survie est représentée par une variable aléatoire continue positive, noté  $T$ , qui mesure le temps écoulé entre un instant initial et la survenue d'un événement.*

*La fonction de répartition de  $T$  à l'instant  $t$  s'écrit :*

$$F(t) = \mathbb{P}(T \leq t), \quad t \in \mathbb{R}^+.$$

#### 2.2.1.2 La fonction de survie

##### Définition 2.2.2. [10]

*C'est la probabilité qu'un événement donné ne se soit pas produit jusqu'à instant  $t$ , elle est noté par  $S(t)$  et s'écrit :*

$$\begin{aligned} S(t) &= \mathbb{P}(T > t). \\ &= 1 - F(t). \end{aligned}$$

*La fonction  $S(t)$  est décroissante et monotone, avec :*

$$S(0) = 1, \quad \lim_{t \rightarrow \infty} S(t) = 0.$$

#### 2.2.1.3 La fonction de risque

On appelle aussi la fonction de risque instantané, représente la probabilité que l'événement se produise à l'instant  $t$ , sachant qu'aucune défaillance ne s'est produite avant cet instant. Elle est noté par  $h(t)$  et s'écrit : [10]

$$\begin{aligned}h(t) &= \lim_{\Delta_t \rightarrow 0} \frac{\mathbb{P}(t \leq T < t + \Delta_t | T \geq t)}{\Delta_t} \\ &= \frac{f(t)}{S(t)}.\end{aligned}$$

Où  $f(t)$  représente la densité de probabilité.

### 2.2.2 Estimateur Kaplan-Meier

Selon [15]. Estimateur Kaplan-Meier est un estimateur non paramétrique, il développé par Kaplan et Meier en 1958 pour estimer la fonction de survie. Son principe repose sur l'idée que, pour être en vie après un instant  $t$ , il faut être en vie juste avant  $t$  et le rester à cet instant.

L'estimateur Kaplan-Meier s'écrit :

$$\hat{S}(t) = \prod_{t_i < t} 1 - \frac{d_i}{n_i}$$

où :

- $t_i$  représente le temps d'observation.
- $n_i$  est le nombre d'individus encore à risque de panne à l'instant  $t_i$ .
- $d_i$  est le nombre de pannes observées au temps  $t_i$ .

### 2.2.3 Modèle de Cox

Selon [15]. Le modèle de Cox est l'un des modèles à risques proportionnels. Il s'agit d'une méthode de régression multivariée qui prend en compte plusieurs covariables à la fois. Développé par Cox en 1972, ce modèle est dit semi-paramétrique, car il ne relie pas le risque de base à l'effet des covariables. Il s'écrit :

$$h(t | X) = h_0(t) \times \exp(\beta^T X).$$

où :

- $h_0(t)$  est la fonction de risque de base.
- $\beta$  est le vecteur des coefficients de régression.

## 2. Les outils d'analyse de données

---

—  $X$  est le vecteur des covariables.

### **Conclusion**

Plusieurs méthodes statistiques sont utilisées pour analyser les données. Par exemple, les modèles ARIMA et le lissage exponentiel servent surtout aux prévisions, tandis que l'analyse de survie est utilisée pour étudier le temps avant la survenue d'un événement. Ainsi, le choix de la méthode dépend principalement de l'objectif de l'étude

# Chapitre 3

## Modélisation de données

Ce chapitre présente les résultats obtenus à partir de l'analyse des données des pannes et de l'utilisation de la bande passante des réseaux ADSL et fibre optique. Nous commençons par une analyse exploratoire pour visualiser l'évolution des pannes. Ensuite, nous présentons l'analyse descriptive et comparative, pour montrer les différences entre les deux réseaux. Par la suite, nous appliquons des modèles de prévision des pannes, comme ARIMA. Enfin, l'analyse de survie, à l'aide de l'estimateur de Kaplan-Meier et du modèle de Cox, est utilisée pour étudier le temps entre les pannes et comparer la fiabilité des deux réseaux.

Les données utilisées dans cette étude proviennent d'Algérie Télécom à Bouira. Les données sur les pannes sont enregistrées quotidiennement sur une période de deux ans, tandis que celles de l'utilisation de la bande passante sont relevées chaque heure pendant une semaine complète pour plusieurs équipements d'accès au réseau. Ces données sont ensuite analysées à l'aide du logiciel R.

### 3.1 Analyse exploratoire : visualisation des séries temporelles

L'analyse exploratoire des séries temporelles permet d'observer les fluctuations quotidiennes des pannes et de vérifier l'absence de tendance marquée ou de saisonnalité évidente. Les figures ci-dessous présentent cette évolution pour l'ADSL et la fibre optique.

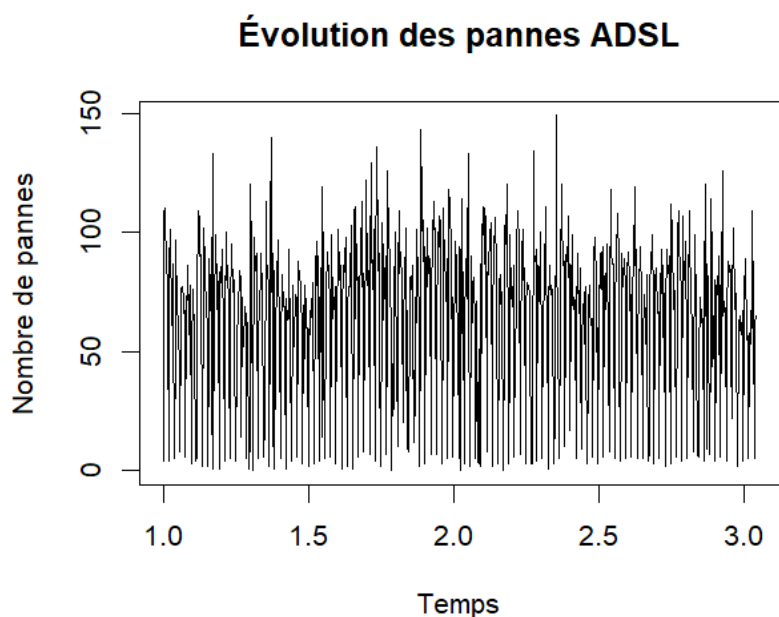


FIGURE 3.1 – Évolution des données de pannes ADSL

Les données de pannes ADSL fluctuent fortement d'un jour à l'autre, sans tendance claire à la hausse ou à la baisse. Des pics élevés apparaissent de façon irrégulière, sans cycle saisonnier évident. Cette forte variabilité indique que les pannes sont plutôt aléatoires et difficiles à prévoir sans modèle statistique.

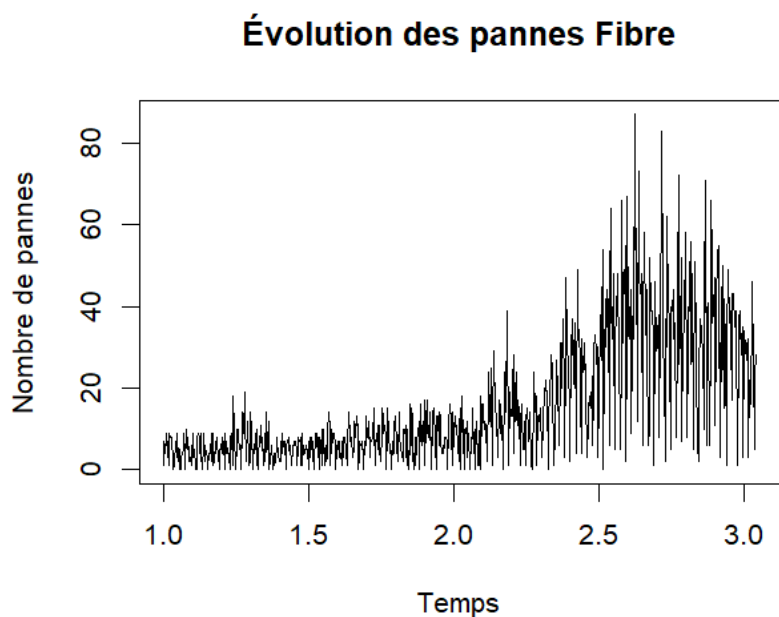


FIGURE 3.2 – Évolution des données de pannes Fibre optique

### 3. Modélisation de données

---

Les données de pannes de la fibre optique restent très faibles pendant presque toute la période, avec quelques pics isolés. Aucune tendance ni saisonnalité n'est perceptible. Cette stabilité confirme la meilleure fiabilité de la fibre optique par rapport à l'ADSL.

## 3.2 Analyse descriptive et comparative

Dans cette section, nous abordons l'analyse descriptive des données relatives aux pannes et à l'utilisation de la bande passante, à l'aide de statistiques (moyenne, écart-type, médiane, etc.), avec des graphiques permettant de comparer les performances des deux réseaux et des tests visant à déterminer si la différence est statistiquement significative.

### 3.2.1 Analyse descriptive des pannes

Le tableau 3.1 présente les statistiques descriptives des pannes sur les réseaux ADSL et fibre optique.

Mesure	ADSL	Fibre
Moyenne	66.32	15.73
Médiane	73	9
Écart-type	32.52	15.43
Minimum	0	0
Maximum	149	87

TABLE 3.1 – Statistiques descriptives des pannes quotidiennes

On observe que le nombre moyen de pannes sur le réseau ADSL est de 66,32 par jour, contre 15,73 pour la fibre optique. La médiane atteint 73 pour l'ADSL alors qu'elle est de seulement 9 pour la fibre, ce qui montre que les pannes sont moins fréquentes sur ce réseau. L'écart-type est aussi plus élevé pour l'ADSL (32,52), ce qui signifie que les pannes y sont plus

### 3. Modélisation de données

---

irrégulières que sur la fibre (15,43). La même différence apparaît avec les valeurs maximales : 149 pour l'ADSL contre 87 pour la fibre.

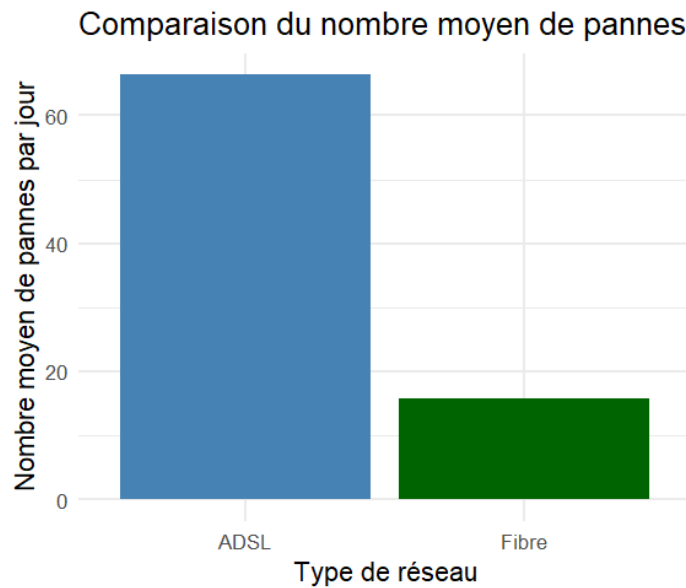


FIGURE 3.3 – Comparaison graphique des pannes entre l'ADSL et la fibre optique

La figure 3.3 confirme visuellement que la moyenne de l'ADSL (66.32) est supérieure à celle de la fibre (15.73).

#### 3.2.2 Analyse descriptive de la bande passante

Les paramètres classiques de la QoS (délai, gigue, perte de paquets) ne sont pas disponibles dans les données qui nous ont été fournies. Nous avons donc utilisé le taux d'utilisation de la bande passante (Rx Bandwidth) comme indicateur substitut pour évaluer indirectement la qualité de service.

Le tableau 3.2 présente les statistiques descriptives de l'utilisation de la bande passante sur les réseaux ADSL et fibre optique.

### 3. Modélisation de données

Mesure	ADSL	Fibre
Moyenne	13.9	10.3
Médiane	12.1	3.59
Écart-type	9.65	13.3
Minimum	0	0.01
Maximum	46.1	54

TABLE 3.2 – Statistiques descriptives de l'utilisation de la bande passante (%).

L'utilisation moyenne de la fibre (10,3 %) reste plus faible que celle de l'ADSL (13,9 %), probablement grâce à sa rapidité. Sa médiane est aussi plus basse (3,59 % contre 12,1 % pour l'ADSL), ce qui veut dire qu'elle reste peu utilisée pendant une grande partie du temps. L'écart-type de la fibre est plus élevé (13,3 % contre 9,65%), donc son utilisation change selon les périodes : elle peut être faible à certains moments et plus élevée à d'autres. On retrouve aussi cette différence dans les valeurs maximales, avec 54 % pour la fibre contre 46,1 % pour l'ADSL, surtout le soir.

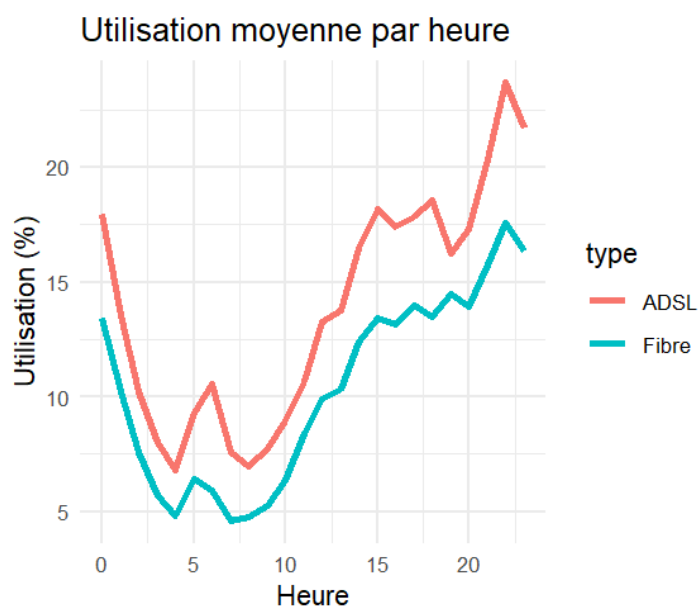


FIGURE 3.4 – Évolution horaire de l'utilisation moyenne de la bande passante.

### 3. Modélisation de données

---

La figure 3.4 montre que l'utilisation de l'ADSL est supérieure à celle de la fibre optique à toutes les heures de la journée, ce qui s'explique par la vitesse de la fibre, avec des pics communs en soirée à partir de 20 h.

La figure 3.5 confirme visuellement que la moyenne de l'ADSL (13,9 %) est supérieure à celle de la fibre (10,3 %).

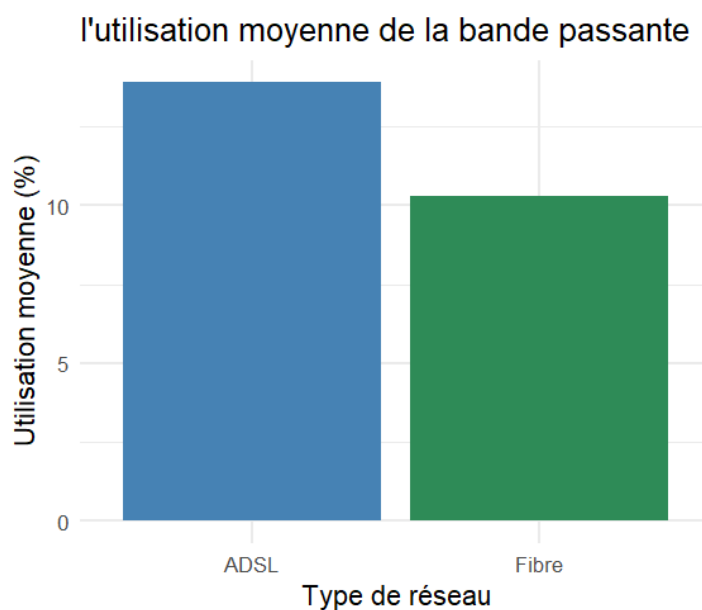


FIGURE 3.5 – Utilisation moyenne de la bande passante : ADSL vs fibre.

#### 3.2.3 Test de comparaison

Pour vérifier si les différences observées entre l'ADSL et la fibre optique sont statistiquement significatives, nous avons appliqué le test du Khi-deux ( $\chi^2$ ). Les hypothèses du test sont les suivantes :

$$\begin{cases} H_0 : \text{Le type de réseau n'a pas d'influence sur la présence de pannes.} \\ H_1 : \text{Le type de réseau influence la présence de pannes.} \end{cases}$$

### 3. Modélisation de données

---

```
Pearson's Chi-squared test with Yates' continuity correction  
data: table_khi  
X-squared = 29.697, df = 1, p-value = 5.052e-08
```

FIGURE 3.6 – Résultat du test du khi-deux (ADSL vs fibre optique).

Le résultat du test du Khi-deux montre que la différence entre l'ADSL et la fibre optique est statistiquement significative, car la valeur de  $p$  ( $5.05 \times 10^{-8}$ ) est inférieure à 0,05. Alors on rejette l'hypothèse  $H_0$ , ce qui indique que la présence de pannes varie selon le type de réseau. Donc on conclut que la fibre optique est plus fiable que l'ADSL.

## 3.3 Analyse des séries temporelles

Dans cette section, nous utilisons les modèles ARIMA pour prévoir l'évolution future des pannes et de l'utilisation de la bande passante pour les réseaux ADSL et fibre optique. Les prévisions sont établies sur un horizon de 30 jours pour les pannes et de 24 heures pour la bande passante.

### 3.3.1 Prévision des pannes ADSL

#### 3.3.1.1 Modélisation (ARIMA)

Pour modéliser les pannes du réseau ADSL, les modèles ARIMA sont utilisés à l'aide de la fonction `auto.arima()` du logiciel R, qui permet de sélectionner automatiquement le modèle le plus adapté. Le modèle obtenu est ARIMA(2,0,3), dont les paramètres estimés sont présentés ci-dessous :

### 3. Modélisation de données

---

```
Series: adsl_ts
ARIMA(2,0,3) with non-zero mean

Coefficients:
      ar1      ar2      ma1      ma2      ma3      mean
    -0.4058 -0.2566  0.5417  0.1503 -0.3206  66.3236
s.e.   0.1068   0.1067  0.1052  0.1226   0.0524   0.9322

sigma^2 = 957.5:  log likelihood = -3611.33
AIC=7236.65  AICc=7236.8  BIC=7268.94

Training set error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set 0.003677897 30.81926 23.73365 -Inf  Inf  0.6835522 0.01655533
```

FIGURE 3.7 – Paramètres et critères de qualité du modèle ARIMA – ADSL.

Les critères AIC et BIC sont proches, indiquant un modèle stable et acceptable pour la prévision des pannes de l'ADSL. La valeur MASE (0.68), inférieure à 1, confirme que le modèle ARIMA(2,0,3) est plus performant.

#### 3.3.1.2 Prévisions (ARIMA)

Jour	Prévision	Lo_80	Hi_80
1	73.05800	33.40152	112.71448
2	57.86538	17.84473	97.88604
3	74.64605	34.11635	115.17576
4	65.11672	22.98786	107.24558
5	64.67774	22.07906	107.27641
6	67.30134	24.70088	109.90180
7	66.34925	23.71173	108.98677
8	66.06236	23.42031	108.70441
9	66.42312	23.78058	109.06565
10	66.35033	23.70711	108.99355

FIGURE 3.8 – Prévisions des pannes ADSL (10 premiers jours).

La figure 3.8 présente les prévisions des pannes du réseau ADSL pour les 10 premiers jours avec un niveau de confiance de 80 %. Les 20 jours restants sont présentés dans l'annexe B (voir 22). On observe que le nombre moyen de pannes sur ce réseau varie entre 57 et 75 par jour.

### 3. Modélisation de données

#### 3.3.2 Prédiction des pannes fibre

##### 3.3.2.1 Modélisation (ARIMA)

Pour la fibre optique, le modèle obtenu après les mêmes étapes est ARIMA(4,1,4), dont les paramètres estimés sont présentés ci-dessous :

```
Series: fttx_ts
ARIMA(4,1,4)

Coefficients:
      ar1      ar2      ar3      ar4      ma1      ma2      ma3      ma4
-0.9536 -0.0927  0.0574 -0.3603  0.1570 -0.9282 -0.4916  0.5593
s.e.    0.0692  0.1058  0.0754  0.0410  0.0649  0.0458  0.0534  0.0658

sigma^2 = 89.45:  log likelihood = -2725.19
AIC=5468.39  AICC=5468.63  BIC=5509.89

Training set error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set 0.2169828 9.400297 6.363435 -Inf  Inf  0.3179208 0.007624404
```

FIGURE 3.9 – Paramètres et critères de qualité du modèle ARIMA – fibre optique.

Les critères AIC et BIC sont proches, indiquant un modèle stable et acceptable pour la prédiction des pannes de fibre. La valeur MASE (0.31), inférieure à 1, confirme que le modèle ARIMA(4,1,4) est plus performant.

##### 3.3.2.2 Prévisions (ARIMA)

Jour	Prédiction	Lo_80	Hi_80
1	25.91216	33.40152	112.71448
2	27.82767	17.84473	97.88604
3	22.15308	34.11635	115.17576
4	25.57279	22.98786	107.24558
5	23.69988	22.07906	107.27641
6	24.15308	24.70088	109.90180
7	26.13522	23.71173	108.98677
8	22.86352	23.42031	108.70441
9	26.50045	23.78058	109.06565
10	23.28604	23.70711	108.99355

FIGURE 3.10 – Prévisions des pannes fibre (10 premiers jours).

La figure 3.10 présente les prévisions des pannes du réseau fibre pour les 10 premiers jours avec un niveau de confiance de 80 %. Les 20 jours restants sont présentés dans l'annexe B

### 3. Modélisation de données

---

(voir 23). On observe que le nombre moyen de pannes sur ce réseau varie entre 23 et 28 par jour, donc plus faible que l'ADSL.

#### 3.3.3 Comparaison des prévisions (pannes)

Pour comparer les pannes des réseaux ADSL et fibre optique, le tableau suivant regroupe les prévisions obtenues pour les 10 premiers jours pour les deux réseaux dans la section précédente.

Jour	ADSL	Fibre
1	73.05800	25.91216
2	57.86538	27.82767
3	74.64605	22.15308
4	65.11672	25.57279
5	64.67774	23.69988
6	67.30134	24.15308
7	66.34925	26.13522
8	66.06236	22.86352
9	66.42312	26.50045
10	66.35033	23.28604

FIGURE 3.11 – Prévisions des pannes (10 premiers jours).

On constate que le nombre de pannes est plus élevé sur le réseau ADSL que sur la fibre optique, ce qui indique une meilleure fiabilité de cette dernière.

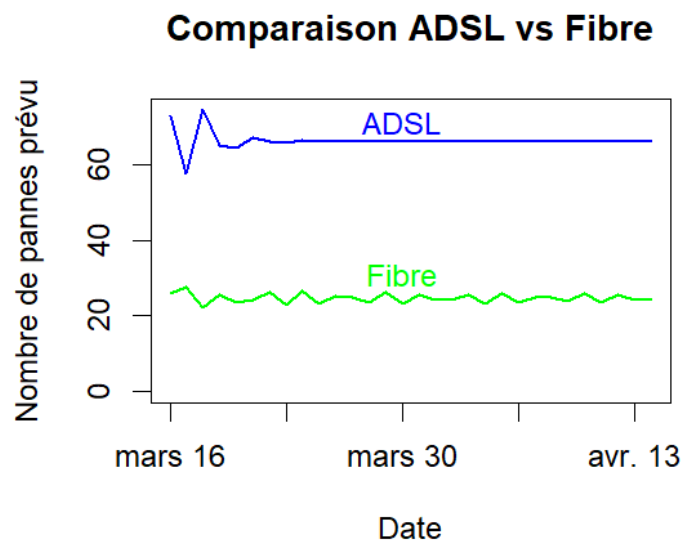


FIGURE 3.12 – Comparaison des prévisions ADSL vs fibre

### 3. Modélisation de données

---

La figure 3.12 montre que les prévisions des pannes pour la fibre optique restent plus faibles que celles de l'ADSL pendant les 30 jours. Cela montre que le réseau fibre optique est plus fiable.

## 3.4 Analyse de survie

Dans cette section, nous présentons une analyse de survie des pannes pour les réseaux ADSL et fibre optique. L'objectif est d'étudier la durée entre les pannes successives (gap) et de comparer la fiabilité des deux technologies. Nous utilisons pour cela l'estimateur de Kaplan-Meier et le modèle de Cox. Les analyses sont basées sur les données de 20 clients (10 ADSL et 10 fibre optique) présentant au moins deux pannes chacun.

### 3.4.1 Statistiques descriptives des gaps

Le tableau 3.3 présente les statistiques des gaps entre les pannes (moyenne, minimum et maximum) pour chaque client. On remarque d'abord que le nombre de pannes est plus faible pour la fibre optique que pour l'ADSL. Les gaps sont aussi plus élevés pour la fibre, ce qui veut dire que les pannes y sont moins fréquentes et que ce réseau est plus fiable.

### 3. Modélisation de données

---

<b>Client</b>	<b>Réseau</b>	<b>Nb pannes</b>	<b>Moyenne gap (jours)</b>	<b>Min gap</b>	<b>Max gap</b>
1	ADSL	6	134.0	38	275
2	ADSL	10	64.1	5	160
3	ADSL	6	133.0	1	269
4	Fibre	3	257.0	101	413
5	ADSL	4	175.0	97	237
6	ADSL	6	72.0	1	129
7	ADSL	4	221.0	55	333
8	ADSL	3	120.0	49	190
9	Fibre	2	246.0	246	246
10	ADSL	5	150.0	15	284
11	Fibre	2	49.0	49	49
12	Fibre	2	166.0	166	166
13	Fibre	2	76.0	76	76
14	Fibre	3	228.0	121	334
15	Fibre	2	328.0	328	328
16	ADSL	3	30.0	27	33
17	ADSL	3	224.0	188	260
18	Fibre	2	273.0	273	273
19	Fibre	2	1.0	1	1
20	Fibre	2	301.0	301	301

TABLE 3.3 – Statistiques des pannes par client (ADSL vs Fibre)

#### 3.4.2 Estimateur Kaplan-Meier

```
Call: survfit(formula = surv(time, event) ~ network, data = surv_data)

      n events median 0.95LCL 0.95UCL
network=ADSL 65     65    42     31     76
network=Fibre 12     12   173    105     NA
```

FIGURE 3.13 – Résultats de Kaplan-Meier pour ADSL et Fibre.

Dans la figure 3.13, le temps médian sans panne pour la fibre optique est de 173 jours, contre seulement 42 jours pour l'ADSL. La valeur NA pour la borne supérieure de la fibre indique que plus de 50% des clients fibre n'ont pas encore eu de nouvelle panne à la fin de la période d'observation.

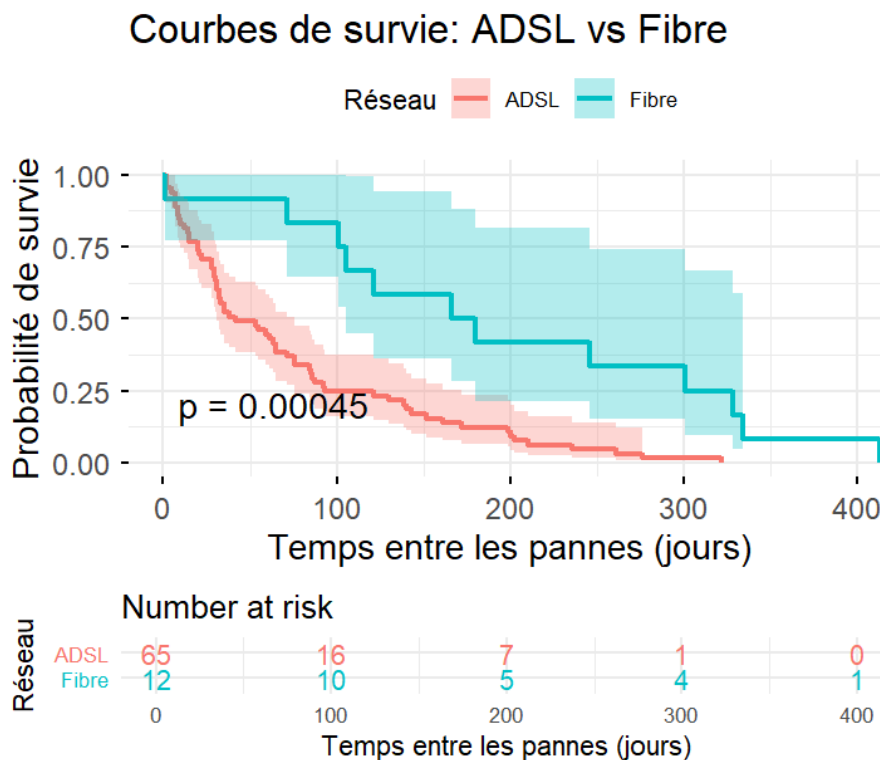


FIGURE 3.14 – Courbes de survie (ADSL vs fibre optique).

La figure 3.14 présente les courbes de survie estimées à partir du modèle de Kaplan-Meier. La courbe de la fibre optique reste au-dessus de celle de l'ADSL sur toute la période d'observation. Cela montre que la probabilité de fonctionner sans panne est plus élevée pour

### 3. Modélisation de données

---

la fibre. Le tableau (Number at risk) indique le nombre de clients encore exposés au risque, c'est-à-dire ceux qui n'ont pas encore eu de panne. On observe une forte baisse pour l'ADSL, ce qui met en évidence la meilleure fiabilité de la fibre.

On applique le test du log-rank (voir figure 3.15), et il confirme que cette différence est statistiquement significative ( $p = 5 \times 10^{-4}$ ). Cette valeur apparaît également sur la figure précédente ( $p = 0.00045$ ).

```
survdif(formula = Surv(time, event) ~ network, data = surv_data)

              N Observed Expected (O-E)^2/E (O-E)^2/V
network=ADSL  65         65     51.9       3.28      12.3
network=Fibre 12         12     25.1       6.81      12.3

Chisq= 12.3  on 1 degrees of freedom, p= 5e-04
```

FIGURE 3.15 – Résultats de test Log-Rank

#### 3.4.3 Modèle de Cox

La figure ci-dessous présente les résultats du modèle de Cox pour les réseaux fibre optique et ADSL. Le coefficient de régression de la fibre est négatif (coef = -1,23). Le hazard ratio (HR) est estimé à 0,29, ce qui correspond à environ 29% du risque observé pour l'ADSL. La valeur  $p=0,00086$  est inférieure à 5%, la différence observée est donc statistiquement significative. Ces résultats montrent une meilleure fiabilité de la fibre optique.

```
coxph(formula = surv(time, event) ~ network, data = surv_data)

              coef exp(coef) se(coef)      z      p
networkFibre -1.2328    0.2915  0.3700 -3.332 0.000863

Likelihood ratio test=14.43  on 1 df, p=0.0001458
n= 77, number of events= 77
```

FIGURE 3.16 – Résultats de modèle de cox pour ADSL et Fibre.

## Conclusion

Ce chapitre présente les résultats obtenus à partir des données réelles de pannes et l'utilisation de la bande passante. L'analyse descriptive montre que la fibre optique enregistre moins

### 3. Modélisation de données

---

de pannes que l'ADSL, avec une moyenne de 15,73 contre 66,32 et une médiane de 9 contre 73. Pour la bande passante, la fibre est moins utilisée en moyenne en raison de sa rapidité, mais elle présente des pics plus élevés en soirée (54 % contre 46,1 %).

Les modèles ARIMA confirment cette tendance avec des prévisions de pannes plus faibles pour la fibre optique sur les 30 prochains jours. L'analyse de survie (Kaplan-Meier et Cox) montre aussi une différence importante, avec un temps médian sans panne plus long pour la fibre (173 jours contre 42 jours) et une réduction du risque d'environ 71% (HR = 0,2915,  $p < 5\%$ ).

La fibre optique apparaît plus fiable et plus performante que l'ADSL.

# Conclusion

Cette étude a permis de comparer la fiabilité des réseaux ADSL et fibre optique. Les résultats obtenus montrent que la fibre optique enregistre globalement moins de pannes que l'ADSL, avec des périodes de fonctionnement sans incident plus longues. Ce résultat est confirmé par les modèles ARIMA ainsi que par l'analyse de survie.

Il est donc recommandé de poursuivre et d'accélérer le déploiement de la fibre optique afin d'améliorer davantage la qualité de service. Les modèles ARIMA peuvent aussi être utilisés pour anticiper les pannes et mieux organiser la maintenance.

La principale limite de ce travail est l'absence de données détaillées sur la qualité de service (QoS) par utilisateur. L'analyse s'est donc concentrée principalement sur les pannes et l'utilisation de la bande passante.

Globalement, les résultats montrent que la fibre optique offre une meilleure fiabilité et une meilleure qualité de service que l'ADSL.

# Annexes

## Annexe A : codes R

Cette section regroupe l'intégralité des codes R utilisés dans ce mémoire. Ils sont présentés sous forme de captures d'écran et couvrent le nettoyage des données, les modèles ARIMA ainsi que l'analyse de survie.

```
library(readxl)
library(dplyr)
library(tidyr)
library(stringr)
library(lubridate)
library(ggplot2)
library(forecast)
# extraire tableau de panne pour ADSL et fibre
df <- read_excel("C:/Users/lenovo/Downloads/Z.xlsx")
colnames(df)[3] <- "ticket_no"
colnames(df)[6] <- "service"

df$date_str <- str_extract(df$ticket_no, "\\d{8}")
df$failure_date <- as.Date(df$date_str, format = "%Y%m%d")
df <- df[!is.na(df$failure_date), ]
daily <- df %>%
  group_by(failure_date, service) %>%
  summarise(count = n(), .groups = "drop") %>%
  pivot_wider(names_from = service, values_from = count, values_fill = 0)

all_dates <- seq.Date(from = min(daily$failure_date),
                     to = max(daily$failure_date),
                     by = "day")

complete_daily <- data.frame(date = all_dates) %>%
  left_join(daily, by = c("date" = "failure_date")) %>%
  mutate(across(where(is.numeric), ~replace_na(.x, 0)))

#tableau finale
View(complete_daily)

#Analyse exploratoire des données|
# ADSL
ts_ads1 <- ts(complete_daily$ADSL, frequency = 365)
plot(ts_ads1, main = "Évolution des pannes ADSL",
     ylab = "Nombre de pannes", xlab = "Temps")

# Fibre
ts_fibre <- ts(complete_daily$FTTX, frequency = 365)
plot(ts_fibre, main = "Évolution des pannes Fibre",
     ylab = "Nombre de pannes", xlab = "Temps")
```

FIGURE 17 – Code source 1

```

# Analyse descriptive et comparatives
stats_adsl_pannes <- complete_daily %>%
  summarise(
    Moyenne = mean(ADSL),
    Mediane = median(ADSL),
    Ecart_Type = sd(ADSL),
    Minimum = min(ADSL),
    Maximum = max(ADSL) )

stats_fibre_pannes <- complete_daily %>%
  summarise(
    Moyenne = mean(FTTx),
    Mediane = median(FTTx),
    Ecart_Type = sd(FTTx),
    Minimum = min(FTTx),
    Maximum = max(FTTx))

print(stats_adsl_pannes)
print(stats_fibre_pannes)

#test de khi_deux
complete_daily$panne_adsl <- ifelse(complete_daily$ADSL > 0, 1, 0)
complete_daily$panne_fibre <- ifelse(complete_daily$FTTx > 0, 1, 0)
table_khi <- matrix(c(
  sum(complete_daily$panne_adsl), sum(complete_daily$panne_adsl == 0),
  sum(complete_daily$panne_fibre), sum(complete_daily$panne_fibre == 0)
), nrow = 2, byrow = TRUE)
colnames(table_khi) <- c("Avec pannes", "Sans pannes")
rownames(table_khi) <- c("ADSL", "Fibre")
chisq.test(table_khi)

#Barplot pour la comparaison visuelle
df_moyennes <- data.frame(
  reseau = c("ADSL", "Fibre"),
  moyenne_pannes = c(66.32215, 15.7302))
ggplot(df_moyennes, aes(x = reseau, y = moyenne_pannes, fill = reseau)) +
  geom_col() +
  labs(title = "Comparaison du nombre moyen de pannes par jour",
    x = "Type de réseau",
    y = "Nombre moyen de pannes par jour") +
  theme_minimal() +
  theme(legend.position = "none") +
  scale_fill_manual(values = c("ADSL" = "steelblue", "Fibre" = "darkgreen"))

#serie temporelle
adsl_ts <- ts(complete_daily$ADSL, frequency = 365)
fttx_ts <- ts(complete_daily$FTTx, frequency = 365)

#les modèles ARIMA
model_adsl <- auto.arima(adsl_ts, seasonal = FALSE)
model_fttx <- auto.arima(fttx_ts, seasonal = FALSE)
summary(model_adsl)
summary(model_fttx)

#prévision de panne ADSL et fibre
forecast_adsl <- forecast(model_adsl, h = 30)
forecast_fttx <- forecast(model_fttx, h = 30)

#comparaison entre adsl et fibre
comparaison_simple <- data.frame(
  Date = forecast_adsl_df$Date,
  ADSL = forecast_adsl_df$Forecast_ADSL,
  Fibre = forecast_fttx_df$Forecast_Fibre)
print(comparaison_simple)

```

FIGURE 18 – Code source 2

```

plot(comparaison_simple$Date, comparaison_simple$ADSL,
     type = "l", col = "blue", lwd = 2,
     ylim = c(0, max(comparaison_simple$ADSL, comparaison_simple$Fibre)),
     main = "Comparaison ADSL vs Fibre",
     xlab = "Date", ylab = "Nombre de pannes prévu")
lines(comparaison_simple$Date, comparaison_simple$Fibre, col = "green", lwd = 2)

text(x = comparaison_simple$Date[15],
     y = comparaison_simple$ADSL[15] + 5,
     labels = "ADSL", col = "blue", font = 1)

text(x = comparaison_simple$Date[15],
     y = comparaison_simple$Fibre[15] + 8,
     labels = "Fibre", col = "green", font = 1)

# données de l'utilisation de la bande passante (Rx bandwidth %)
#Regrouper les 2 fichiers
adsl <- read_excel("C:/Users/lenovo/Downloads/QoS_ADSL.xlsx", skip=1)
names(adsl) <- c("datetime", paste0("eqa_adsl_", 1:10))

fibre <- read_excel("C:/Users/lenovo/Downloads/QoS_Fibre.xlsx", skip = 1)
names(fibre) <- c("datetime", paste0("eqa_fibre_", 1:10))

adsl <- adsl %>%
  mutate(across(-datetime, ~ as.numeric(gsub(",", "\\.", .))))
fibre <- fibre %>%
  mutate(across(-datetime, ~ as.numeric(gsub(",", "\\.", .))))
adsl_long <- adsl %>%
  pivot_longer(cols = -datetime,
               names_to = "equipement",
               values_to = "utilisation") %>%
  mutate(type = "ADSL")
fibre_long <- fibre %>%
  pivot_longer(cols = -datetime,
               names_to = "equipement",
               values_to = "utilisation") %>%
  mutate(type = "Fibre")
df <- bind_rows(adsl_long, fibre_long)
df$datetime <- as.POSIXct(df$datetime, format = "%Y-%m-%d, %H:%M")
df <- df %>%
  mutate(datetime = if_else(is.na(datetime),
                           as.POSIXct("2026-02-26 01:00:00"),
                           datetime))

View(df)

# Analyse descriptive et comparative

stats_bw_adsl <- df %>%
  filter(type == "ADSL") %>%
  summarise(
    Moyenne = mean(utilisation, na.rm = TRUE),
    Mediane = median(utilisation, na.rm = TRUE),
    Ecart_Type = sd(utilisation, na.rm = TRUE),
    Minimum = min(utilisation, na.rm = TRUE),
    Maximum = max(utilisation, na.rm = TRUE) )

stats_bw_fibre <- df %>%
  filter(type == "Fibre") %>%
  summarise(
    Moyenne = mean(utilisation, na.rm = TRUE),
    Mediane = median(utilisation, na.rm = TRUE),
    Ecart_Type = sd(utilisation, na.rm = TRUE),
    Minimum = min(utilisation, na.rm = TRUE),
    Maximum = max(utilisation, na.rm = TRUE) )

print(stats_bw_adsl)
print(stats_bw_fibre)

```

FIGURE 19 – Code source 3

```

#Barplot pour la comparaison visuelle

df_bandwidth <- data.frame(
  reseau = c("ADSL", "Fibre"),
  utilisation_moyenne = c(13.9, 10.3))

ggplot(df_bandwidth, aes(x = reseau, y = utilisation_moyenne, fill = reseau)) +
  geom_col() +
  labs(title = "l'utilisation moyenne de la bande passante",
       x = "Type de réseau",
       y = "Utilisation moyenne (%)") +
  theme_minimal() +
  theme(legend.position = "none") +
  scale_fill_manual(values = c("ADSL" = "steelblue", "Fibre" = "seagreen"))

#Graphe d'utilisation de la bande passante
library(lubridate)
peak_hours <- df %>%
  mutate(heure = hour(datetime)) %>%
  group_by(heure, type) %>%
  summarise(moyenne = mean(utilisation, na.rm = TRUE))

# رسم المنحنى
ggplot(peak_hours, aes(x = heure, y = moyenne, color = type)) +
  geom_line(size = 1.2) +
  labs(title = "Utilisation moyenne par heure",
       x = "Heure", y = "Utilisation (%)") +
  theme_minimal()

#Analyse de survie
# Préparation des données: dates de pannes par client et calcul des gaps(intervalles)
colnames(df)[3] <- "ticket_no"
colnames(df)[6] <- "service"
df$service <- ifelse(df$service == "FTTx", "Fibre", df$service)
df$date_failure <- as.Date(str_extract(df$ticket_no, "\\d{8}"), format = "%Y%m%d")
df$client_id <- str_extract(df$ticket_no, "\\d{5}$")
df <- df %>% filter(!is.na(client_id), !is.na(date_failure))
set.seed(128)

clients_adsl <- df %>%
  group_by(client_id) %>%
  summarise(
    has_adsl = any(service == "ADSL"),
    has_fibre = any(service == "Fibre"),
    nb_adsl = sum(service == "ADSL")
  ) %>%
  filter(has_adsl == TRUE, has_fibre == FALSE, nb_adsl >= 3) %>%
  pull(client_id)

clients_fibre <- df %>%
  group_by(client_id) %>%
  summarise(
    has_fibre = any(service == "Fibre"),
    has_adsl = any(service == "ADSL"),
    nb_fibre = sum(service == "Fibre")
  ) %>%
  filter(has_fibre == TRUE, has_adsl == FALSE, nb_fibre >= 2) %>%
  pull(client_id)

clients_adsl <- sample(clients_adsl, 10)
clients_fibre <- sample(clients_fibre, 10)
df %>%
  filter(service == "ADSL", client_id %in% clients_adsl) %>%
  group_by(client_id) %>%
  summarise(nb_pannes = n()) %>%
  arrange(desc(nb_pannes)) %>%
  print()

```

FIGURE 20 – code source 4

```

df %>%
  filter(service == "Fibre", client_id %in% clients_fibre) %>%
  group_by(client_id) %>%
  summarise(nb_pannes = n()) %>%
  arrange(desc(nb_pannes)) %>%
  print()
all_clients <- c(clients_adsl, clients_fibre)

table_finale <- data.frame()

for(client in all_clients) {

  client_data <- df %>%
    filter(client_id == client) %>%
    arrange(date_failure) %>%
    select(client_id, service, date_failure)

  if(nrow(client_data) > 0) {
    client_data$gap <- c(NA, diff(client_data$date_failure))
    table_finale <- bind_rows(table_finale, client_data)} }

print(table_finale)
#statistique descriptive de gaps
stats <- table_finale %>%
  group_by(client_id, service) %>%
  summarise(
    nb_pannes = n(),
    moyenne_gap = round(mean(gap, na.rm = TRUE), 1),
    min_gap = min(gap, na.rm = TRUE),
    max_gap = max(gap, na.rm = TRUE),
    .groups = "drop")

print(stats)

library(survival)
library(survminer)

surv_data <- table_finale %>%
  filter(!is.na(gap)) %>%
  mutate(
    time = gap,
    event = 1,
    network = service )

# Kaplan-Meier

km_fit <- survfit(Surv(time, event) ~ network, data = surv_data)
print(km_fit)

# graphe de survie
p1 <- ggsurvplot(km_fit,
  data = surv_data,
  title = "Courbes de survie: ADSL vs Fibre",
  font.title = 16, font.x = 14, font.y = 14, font.tickslab = 12,
  legend.font.size = 12, risk.table.fontsize = 4,
  xlab = "Temps entre les pannes (jours)",
  ylab = "Probabilité de survie", legend.title = "Réseau",
  legend.labs = c("ADSL", "Fibre"), pval = TRUE,
  conf.int = TRUE, risk.table = TRUE,
  risk.table.col = "strata", ggtheme = theme_minimal() )

print(p1)
# test Log-rank

logrank_test <- survdiff(Surv(time, event) ~ network, data = surv_data)
print(logrank_test)

#modèle de Cox
cox_model <- coxph(Surv(time, event) ~ network, data = surv_data)
cox_model

```

FIGURE 21 – Code source 5

## Annexe B

Cette section présente les résultats détaillés qui n'ont pas été inclus dans le corps du mémoire pour éviter de surcharger la lecture. On y trouve les prévisions complètes des pannes sur 30 jours pour les réseaux ADSL et fibre optique.

Jour	Prévision	Lo_80	Hi_80
1	73.05800	33.40152	112.71448
2	57.86538	17.84473	97.88604
3	74.64605	34.11635	115.17576
4	65.11672	22.98786	107.24558
5	64.67774	22.07906	107.27641
6	67.30134	24.70088	109.90180
7	66.34925	23.71173	108.98677
8	66.06236	23.42031	108.70441
9	66.42312	23.78058	109.06565
10	66.35033	23.70711	108.99355
11	66.28729	23.64405	108.93054
12	66.33155	23.68829	108.97482
13	66.32977	23.68649	108.97305
14	66.31913	23.67586	108.96241
15	66.32391	23.68063	108.96719
16	66.32470	23.68142	108.96798
17	66.32315	23.67987	108.96643
18	66.32358	23.68030	108.96686
19	66.32380	23.68052	108.96708
20	66.32360	23.68032	108.96688
21	66.32363	23.68035	108.96690
22	66.32367	23.68039	108.96695
23	66.32364	23.68037	108.96692
24	66.32364	23.68036	108.96692
25	66.32365	23.68037	108.96693
26	66.32365	23.68037	108.96693
27	66.32365	23.68037	108.96693
28	66.32365	23.68037	108.96693
29	66.32365	23.68037	108.96693
30	66.32365	23.68037	108.96693

FIGURE 22 – Prévisions des pannes ADSL sur 30 jours

Jour	Prévision	Lo_80	Hi_80
1	25.91216	33.40152	112.71448
2	27.82767	17.84473	97.88604
3	22.15308	34.11635	115.17576
4	25.57279	22.98786	107.24558
5	23.69988	22.07906	107.27641
6	24.15308	24.70088	109.90180
7	26.13522	23.71173	108.98677
8	22.86352	23.42031	108.70441
9	26.50045	23.78058	109.06565
10	23.28604	23.70711	108.99355
11	25.11227	23.64405	108.93054
12	25.05620	23.68829	108.97482
13	23.44559	23.68649	108.97305
14	26.24955	23.67586	108.96241
15	23.06383	23.68063	108.96719
16	25.76957	23.68142	108.96798
17	24.22588	23.67987	108.96643
18	24.25409	23.68030	108.96686
19	25.67332	23.68052	108.96708
20	23.25392	23.68032	108.96688
21	25.98727	23.68035	108.96690
22	23.67631	23.68039	108.96695
23	24.97649	23.68037	108.96692
24	24.97939	23.68036	108.96692
25	23.73871	23.68037	108.96693
26	25.82876	23.68037	108.96693
27	23.48244	23.68037	108.96693
28	25.45389	23.68037	108.96693
29	24.35836	23.68037	108.96693
30	24.33265	23.68037	108.96693

FIGURE 23 – Prévisions des pannes Fibre sur 30 jours

# Bibliographie

- [1] Armatte, M. (2008). *Histoire et préhistoire de l'analyse des données*. JEHPS, 4(2).
- [2] Atelin, P. (2014). *Réseaux informatiques : notions fondamentales*. ENI Éditions.
- [3] Benakil, S. (2014). *Conception et réalisation d'une application de gestion des frais de missions d'une direction technique de wilaya*. Université de Tizi-Ouzou.
- [4] Beuran, R. (2004). *Mesure de la qualité dans les réseaux informatiques*. POLITEHNICA Bucarest ; Jean Monnet Saint-Étienne.
- [5] Bourbonnais, R., & Terraza, V. (2022). *Analyse des séries temporelles*. Dunod.
- [6] Boutahar, M., & Royer-Carenzi, M. (2019). *Méthodes en séries temporelles et applications avec R*. Ellipses.
- [7] Bouteldja, A., & Bakdi, K. (2019). *Gestion des Dossiers des Œuvres Sociales de la Direction de l'éducation*. Université de Tiaret.
- [8] Bouygues Telecom Pro. ADSL : Définition.  
<https://www.bouyguestelecom-pro.fr/mag-business/lexique/adsl/>.
- [9] Desbois, D. (2005). *Une introduction à la méthodologie de Box et Jenkins : l'utilisation de modèles ARIMA avec SPSS*. Revue MODULAD, 33.
- [10] Ducrot, L. (2024). *Réseaux bayésiens et analyse de survie*. Sorbonne Université.
- [11] ENSAI. *Polycopié de cours : Séries temporelles*.  
<https://ensai.fr/wp-content/uploads/2019/06/Polyseriestemp.pdf>.
- [12] Ertur, C. (1998). *Méthodologies de test de la racine unitaire*. LATEC, Université de Bourgogne, n° 9813.
- [13] Francois, J., & Lemainque, F. (2015). *Tout sur les réseaux et Internet*. Dunod.
- [14] Gharbi, M. (2019). *La mise en place et la gestion d'un réseau informatique au sein de la FSEI*. Université de Mostaganem.
- [15] Graffeo, N. (2014). *Méthodes d'analyse de la survie nette*. Aix-Marseille Université.

- [16] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting : Principles and Practice*. OTexts.
- [17] Jacques, J. *Introduction aux séries temporelles*. Université de Lyon.
- [18] Kherbache, Z., & Laribi, A. *Étude de la QoS dans les réseaux WiFi*. Université de Tlemcen.
- [19] Khatir, N., & Belhadri, A. (2021). *Support de cours sur les réseaux informatiques*. ESG2E Oran.
- [20] Khenniche, M. O. (2020). *Étude comparative entre l'ADSL et la fibre optique*. Université de Biskra.
- [21] Lagnoux, A. *Séries chronologiques*. Université de Toulouse.
- [22] Lemkak, T. (2018). *Conception d'une interface graphique en vue d'un choix optimisé d'un support de transmission*. Université de Biskra.
- [23] Lorino, T. (2002). *Modèles statistiques pour des données de survie corrélées*. INRA Paris-Grignon.
- [24] Mekkakia, M. Z., Selka, S., Berbaoui, A., & Metadjer, Y. (2004). Vers une qualité de service dans Internet. *Communication Science & Technologie*, 3.
- [25] Pardoux, C., & Goldfarb, B. (2013). *Prévision à court terme : méthodes de lissage exponentiel*. Université Paris-Dauphine.
- [26] Rabenantenaina, T., Bemarisika, P., & Totohasina, A. (2020). *Deux approches efficaces pour estimer un modèle de série temporelle*. Université d'Antsiranana.
- [27] Rubenthaler, S., & Vasilieiadis, A. (2024). *Séries chronologiques (avec R) – Cours et exercices*.
- [28] Seddiki, M. S. (2015). *Allocation dynamique des ressources et gestion de la qualité de service dans la virtualisation des réseaux*. Université de Lorraine.
- [29] Taleb, I. (2006). *Qualité de service dans les services Web*. Université de Montréal.
- [30] Vallet, M. Analyse des données. *Bulletin de l'Union des Physiciens*, 671, 595–600.
- [31] Yankam, Y. F. (2021). *Gestion autonome de la QoS dans les réseaux virtuels*. Cameroun.
- [32] Yende Raphael, G. (2018). *Support de cours de Télématique et Protocoles de Communication*. Butembo.