

# Analyse temporelle et prédictive de la qualité de l'air en Inde (2015–2020)

Projet : Données climatiques et environnementales

Abdoul Bassik IDRISOU

5 novembre 2025

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Contexte et justification	3
1.2	Problématiques	3
1.3	Objectifs de l'étude	3
<b>2</b>	<b>Données et préparation</b>	<b>3</b>
2.1	Source et description du jeu de données	3
2.2	Nettoyage et transformation des données	4
2.3	Évaluation des données par ville	5
2.4	Structuration finale	6
<b>3</b>	<b>Analyse exploratoire des données (EDA)</b>	<b>6</b>
3.1	Statistiques descriptives	6
3.2	Distribution globale de l'AQI	6
3.3	Comparaison inter-villes	7
3.4	Évolution temporelle de l'AQI	8
3.5	Corrélation entre polluants et AQI	9
3.6	Analyse mensuelle et saisonnière	10
<b>4</b>	<b>Visualisation géographique</b>	<b>11</b>
<b>5</b>	<b>Modélisation prédictive : Random Forest</b>	<b>12</b>
5.1	Objectif du modèle	12
5.2	Variables utilisées	13
5.3	Paramètres et entraînement du modèle	13
5.4	Résultats et évaluation	13
5.5	Importance des variables explicatives	13
5.6	Comparaison AQI réel vs AQI prédit	14
<b>6</b>	<b>Prévision temporelle : modèle Prophet</b>	<b>15</b>
6.1	Objectif et motivation	15
6.2	Préparation des données	15
6.3	Entraînement et horizon de prévision	15
6.4	Résultats et visualisations	15
6.5	Projection à long terme (jusqu'en 2028)	17
<b>7</b>	<b>Résultats</b>	<b>17</b>
<b>8</b>	<b>Conclusion et recommandations</b>	<b>18</b>
8.1	Synthèse générale	18
8.2	Apport des modèles prédictifs	18
8.3	Recommandations	19
8.4	Perspectives	19

# 1 Introduction

## 1.1 Contexte et justification

La pollution atmosphérique est devenue l’un des principaux défis environnementaux et sanitaires en Inde, particulièrement dans les zones urbaines densément peuplées. Les concentrations élevées de particules fines ( $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ) et de gaz nocifs ( $\text{NO}_2$ ,  $\text{SO}_2$ ,  $\text{CO}$ ,  $\text{O}_3$ ) dépassent fréquemment les seuils recommandés par l’OMS ([WHO, 2021](#)), provoquant des pathologies respiratoires, cardiovasculaires et des pertes économiques significatives.

Dans ce contexte, comprendre l’évolution de la qualité de l’air et développer des outils de prévision fiables constitue un enjeu majeur pour la planification urbaine et la gestion environnementale.

## 1.2 Problématiques

Cette étude s’articule autour de deux problématiques principales :

- **Problématique 1** : Comment l’évolution quotidienne de la pollution atmosphérique a-t-elle varié dans les grandes villes indiennes entre 2015 et 2020 ?
- **Problématique 2** : Dans quelle mesure peut-on prédire la qualité de l’air (AQI) à court et moyen terme afin d’anticiper les risques sanitaires et environnementaux ?

Ces deux questions orientent la démarche méthodologique et les choix d’analyse présentés dans les sections suivantes.

## 1.3 Objectifs de l’étude

Les objectifs poursuivis sont les suivants :

1. **Analyser les tendances temporelles** de l’Indice de Qualité de l’Air (AQI) dans les principales métropoles indiennes.
2. **Identifier les polluants dominants** ( $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ,  $\text{NO}_2$ , etc.) à l’origine de la dégradation de la qualité de l’air.
3. **Modéliser et prédire l’AQI** à court et moyen terme à l’aide d’approches de machine learning (Random Forest) et de séries temporelles (Prophet).

# 2 Données et préparation

## 2.1 Source et description du jeu de données

L’étude repose sur la base de données `city_day.csv`, extraite du portail public *Central Pollution Control Board (CPCB, India)*. Elle recense les niveaux quotidiens de pollution atmosphérique dans plusieurs grandes villes indiennes entre janvier 2015 et juillet 2020.

Le jeu de données initial comporte **29 531 lignes et 16 colonnes**, correspondant à des mesures quotidiennes pour 26 villes. Chaque enregistrement inclut :

- **City** : nom de la ville (26 au total) ;
- **Date** : date d’enregistrement (format YYYY-MM-DD) ;
- **PM2.5**, **PM10** : particules fines ( $\mu\text{g}/\text{m}^3$ ) ;

- **NO, NO<sub>2</sub>, NO<sub>x</sub>** : oxydes d’azote ( $\mu\text{g}/\text{m}^3$ ) ;
- **NH<sub>3</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>** : autres polluants gazeux ;
- **Benzene, Toluene, Xylene** : composés organiques volatils (COV) ;
- **AQI** : indice global de qualité de l’air ;
- **AQI\_Bucket** : catégorie qualitative (Good, Moderate, Poor, etc.).

TABLE 1 – Structure initiale du jeu de données `city_day.csv`

Variable	Description	Type
City	Nom de la ville	object
Date	Date d’observation	datetime
PM <sub>2.5</sub> , PM <sub>10</sub>	Particules fines et grossières ( $\mu\text{g}/\text{m}^3$ )	float
NO, NO <sub>2</sub> , NO <sub>x</sub>	Oxydes d’azote	float
NH <sub>3</sub> , CO, SO <sub>2</sub> , O <sub>3</sub>	Autres polluants gazeux	float
Benzene, Toluene, Xylene	Composés organiques volatils	float
AQI	Indice de qualité de l’air	float
AQI_Bucket	Catégorie AQI (Good, Moderate, etc.)	object

## 2.2 Nettoyage et transformation des données

a) **Vérification et gestion des valeurs manquantes.** L’analyse initiale du jeu de données a révélé un taux non négligeable de valeurs manquantes pour plusieurs variables. Le tableau 2 présente le pourcentage de valeurs manquantes par colonne avant traitement.

TABLE 2 – Pourcentage de valeurs manquantes par variable avant interpolation.

Variable	% de valeurs manquantes
City	0.00
Date	0.00
PM <sub>2.5</sub>	15.57
PM <sub>10</sub>	37.72
NO	12.13
NO <sub>2</sub>	12.14
NO <sub>x</sub>	14.17
NH <sub>3</sub>	34.97
CO	6.97
SO <sub>2</sub>	13.05
O <sub>3</sub>	13.62
Benzene	19.04
Toluene	27.23
Xylene	61.32
AQI	15.85
AQI_Bucket	15.85

Aucune ligne dupliquée n’a été détectée. Les valeurs manquantes numériques ont été **interpolées linéairement par ville**, afin de préserver la cohérence temporelle et la continuité des séries journalières.

**b) Transformation temporelle.** La colonne `Date` a été convertie au format `datetime`, puis décomposée en trois nouvelles variables : `Year`, `Month` et `Day`. Une variable `Season` a été ajoutée pour distinguer les quatre saisons typiques du climat indien :

- **Winter** : décembre à février ;
- **Summer** : mars à mai ;
- **Monsoon** : juin à septembre ;
- **Autumn** : octobre à novembre.

**c) Suppression des variables trop incomplètes.** La variable `Xylene`, comportant plus de 60% de données manquantes, a été retirée du jeu de données.

**d) Ajout de variables dérivées.** Des **moyennes mobiles sur 7 jours** ont été créées pour les colonnes `PM2.5` et `AQI`, afin de lisser les fluctuations quotidiennes.

## 2.3 Évaluation des données par ville

Un calcul du pourcentage de valeurs manquantes par ville a permis d'identifier celles disposant de mesures les plus complètes. La figure 1 illustre la proportion de valeurs absentes pour la variable `AQI` selon la ville.

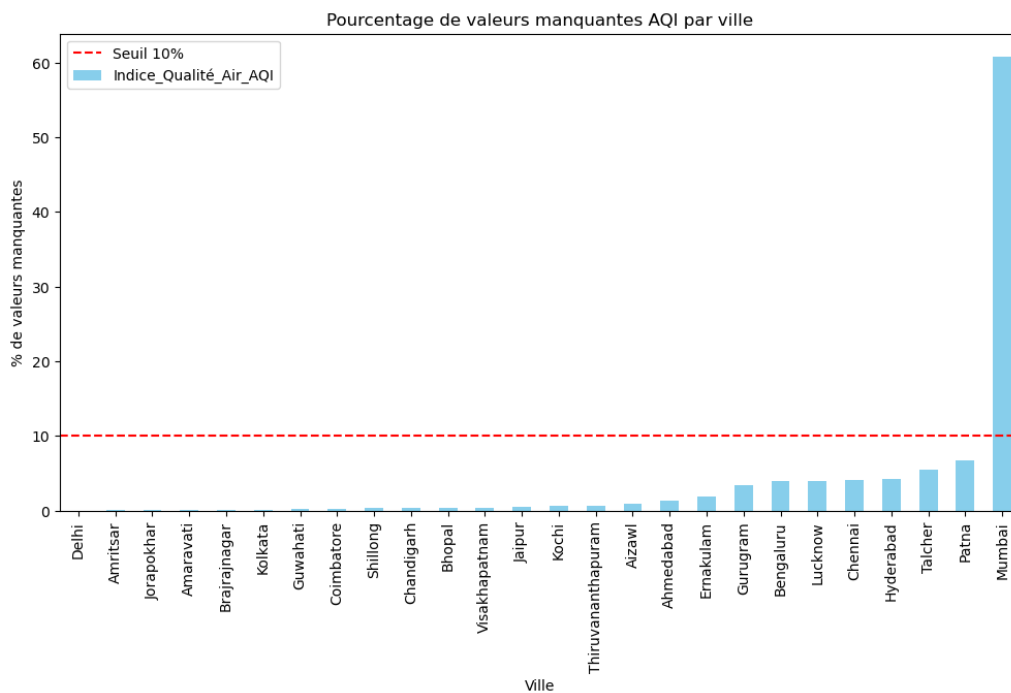


FIGURE 1 – Pourcentage de valeurs manquantes de l'AQI par ville (2015–2020).

Les dix villes présentant le plus faible taux de données manquantes sur l'AQI ont été retenues pour les analyses :

Delhi, Amritsar, Jorapokhar, Amaravati, Brajrajnagar, Kolkata, Guwahati, Coimbatore, Shillong, Cha

TABLE 3 – Taux de valeurs manquantes de l'AQI par ville (extrait des 10 meilleures)

Ville	% de valeurs manquantes (AQI)
Delhi	0.00
Amritsar	0.08
Jorapokhar	0.09
Amaravati	0.11
Brajrajnagar	0.11
Kolkata	0.12
Guwahati	0.20
Coimbatore	0.26
Shillong	0.32
Chandigarh	0.33

## 2.4 Structuration finale

Après nettoyage, le nouveau jeu `df_top10` comprend :

- **8 604 lignes** (observations quotidiennes) ;
- **19 variables** ;
- Une répartition équilibrée entre 10 grandes villes.

### Insight - Qualité des données

Le nettoyage et la sélection ont permis d'obtenir un échantillon robuste couvrant dix villes majeures, avec moins de 1% de valeurs manquantes sur la variable clé AQI. La structure temporelle est cohérente et permet des analyses multi-échelles (jour, mois, saison, année).

### Limitation

Certaines variables secondaires (COV, NH<sub>3</sub>) restent partiellement incomplètes. De plus, le nombre d'observations varie selon les villes : Delhi (2 009 jours) contre seulement 304 pour Chandigarh, ce qui peut biaiser certaines comparaisons inter-villes.

## 3 Analyse exploratoire des données (EDA)

### 3.1 Statistiques descriptives

Le jeu de données nettoyé `df_top10` décrit la pollution de l'air dans dix grandes villes indiennes. Les statistiques descriptives permettent d'obtenir une première idée de la distribution des polluants et de l'Indice de Qualité de l'Air (AQI).

### 3.2 Distribution globale de l'AQI

L'histogramme de la figure 2 montre la distribution de l'AQI pour les dix villes sélectionnées. La majorité des observations se situent entre 50 et 200, correspondant à une

TABLE 4 – Résumé statistique des principales variables du jeu de données (2015–2020).

Variable	Moyenne	Écart-type	Min	Max
PM <sub>2.5</sub>	67.8	65.5	2.0	917.8
PM <sub>10</sub>	136.5	108.3	1.1	1000.0
NO <sub>2</sub>	26.4	23.7	0.02	237.3
SO <sub>2</sub>	15.6	16.8	0.48	158.3
CO	1.09	1.51	0.00	30.4
AQI	154.2	108.9	14.0	956.0

qualité de l'air *Satisfaisante à Modérée*, mais une longue queue à droite indique des épisodes extrêmes de pollution.

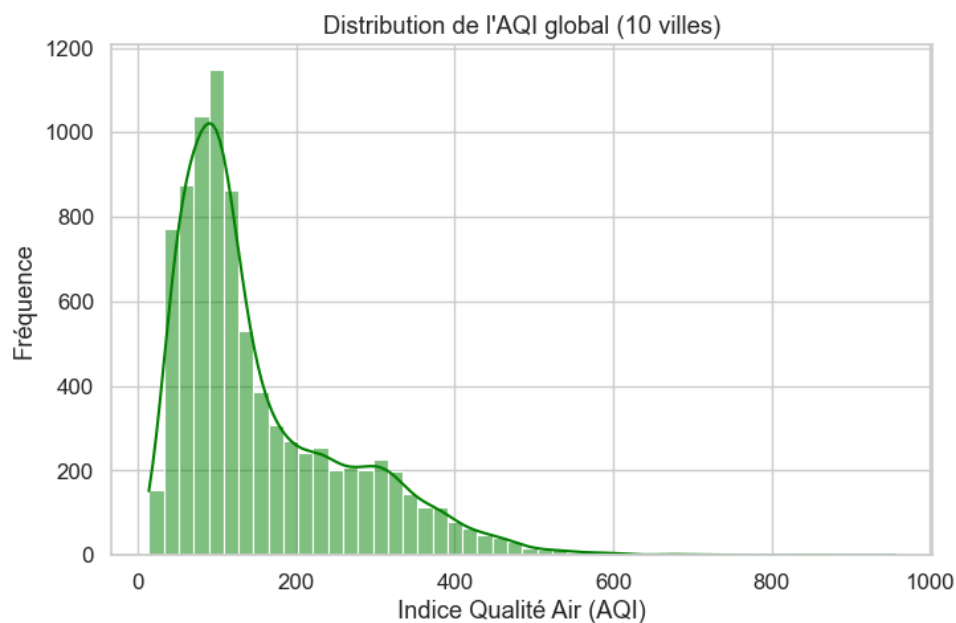


FIGURE 2 – Distribution de l'Indice de Qualité de l'Air (AQI) dans les dix villes sélectionnées.

### 3.3 Comparaison inter-villes

La figure 3 présente la distribution de l'AQI par ville à l'aide d'un boxplot. Delhi se distingue comme la ville la plus polluée avec une médiane autour de 200–250 et plusieurs épisodes dépassant 400. À l'inverse, Shillong et Coimbatore montrent des niveaux de pollution nettement inférieurs (médiane  $\approx$  60–70).

#### Insight - Variabilité spatiale

Delhi présente les niveaux de pollution les plus élevés et les plus variables, alors que Shillong et Coimbatore bénéficient d'une qualité d'air nettement meilleure. Cette disparité traduit l'influence combinée du trafic, de la densité urbaine et des conditions météorologiques locales.

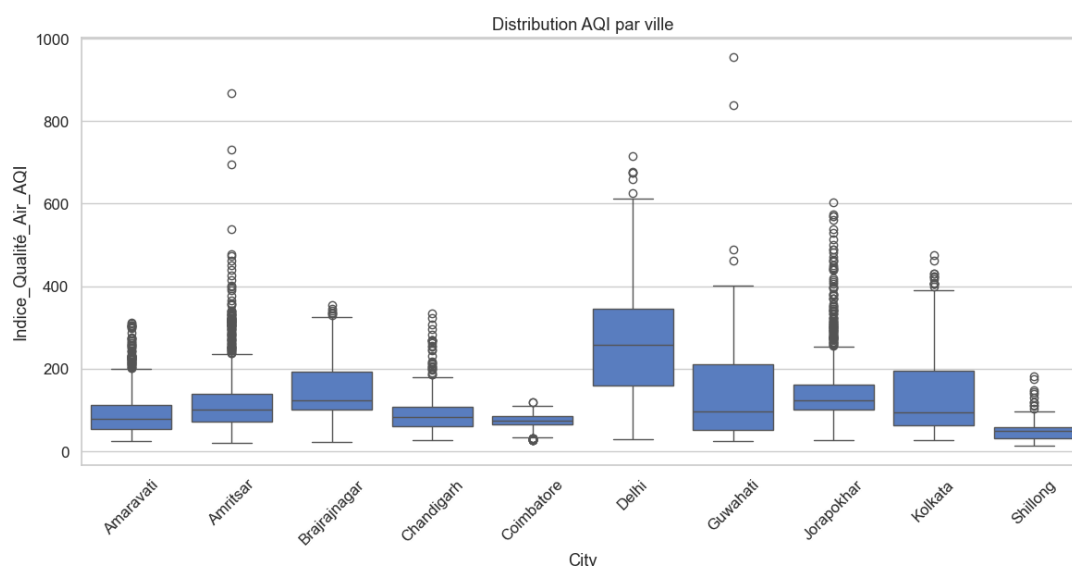


FIGURE 3 – Distribution de l'AQI par ville (2015–2020).

### 3.4 Évolution temporelle de l'AQI

L'évolution temporelle de la qualité de l'air est étudiée à partir de la moyenne mobile sur 30 jours de l'AQI. Afin d'obtenir une visualisation plus claire, seules trois villes représentatives du groupe à pollution extrême ont été conservées : **Delhi**, **Kolkata** et **Guwahati**.

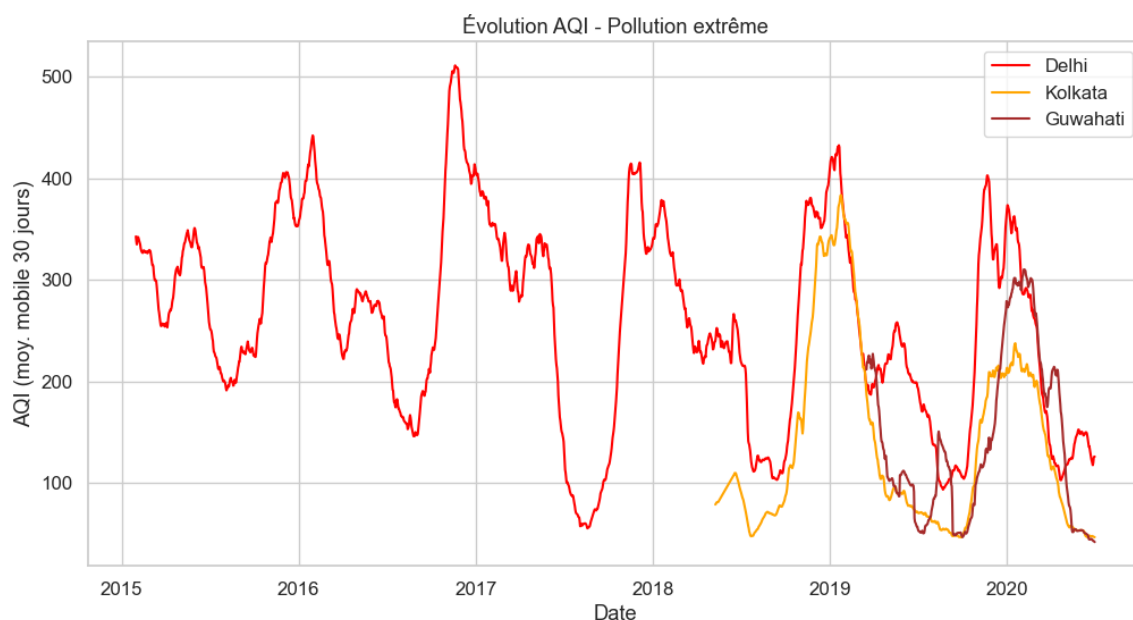


FIGURE 4 – Évolution de l'AQI moyen (moyenne mobile 30 jours) pour trois villes à forte pollution.

Sur la période 2015–2020, on observe :

- des **pics récurrents** de pollution durant les mois d'hiver (novembre à janvier), particulièrement marqués à Delhi ;



- une **baisse nette pendant la mousson** (juin à septembre) où les précipitations nettoient l’atmosphère ;
- une forte variabilité interannuelle liée aux épisodes locaux de brûlage agricole, aux inversions thermiques et au trafic routier.

### Insight – Évolution temporelle des villes les plus polluées

Delhi reste la ville la plus touchée, avec des pics supérieurs à 400 AQI presque chaque hiver, traduisant une pollution chronique sévère. Kolkata et Guwahati montrent des profils similaires mais d’intensité moindre, confirmant que la région nord-est de l’Inde subit également des épisodes critiques. La tendance générale indique une légère amélioration après 2019, probablement liée à des politiques environnementales renforcées et à la baisse d’activité de 2020 (confinements).

## 3.5 Corrélation entre polluants et AQI

L’analyse de corrélation (figure 5) révèle une forte association positive entre l’AQI et les particules fines  $PM_{2.5}$  ( $r = 0.92$ ) et  $PM_{10}$  ( $r = 0.88$ ). Les gaz comme  $NO_2$  ou  $CO$  présentent également une corrélation modérée, tandis que l’ $O_3$  est faiblement corrélée négativement.

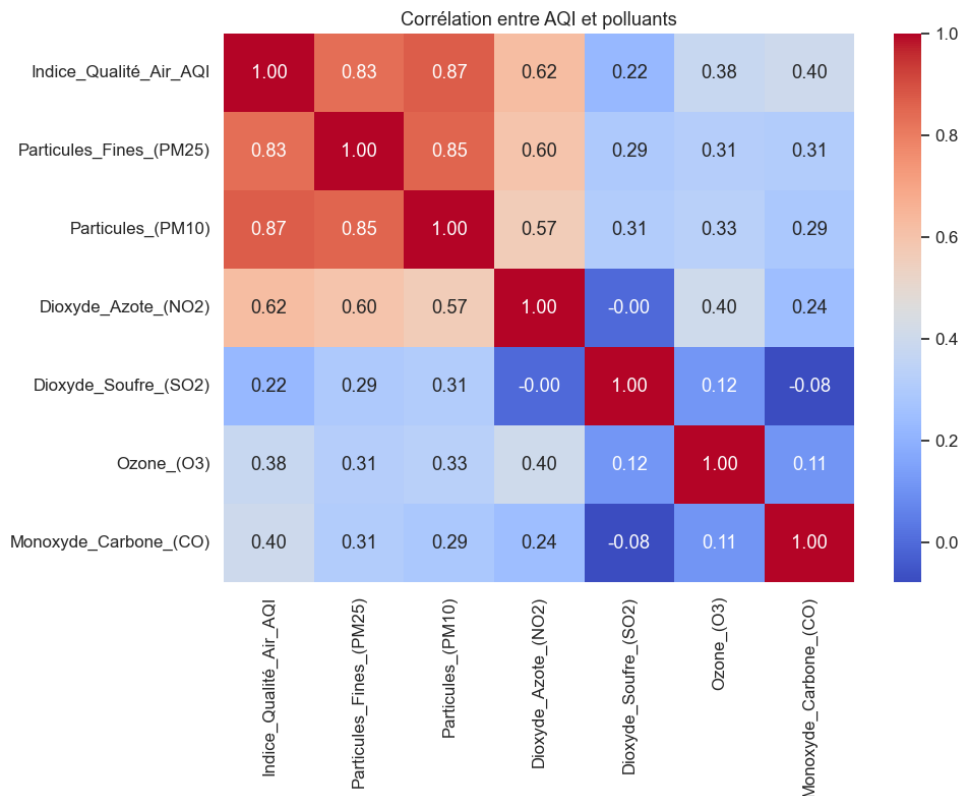


FIGURE 5 – Matrice de corrélation entre l’AQI et les principaux polluants.

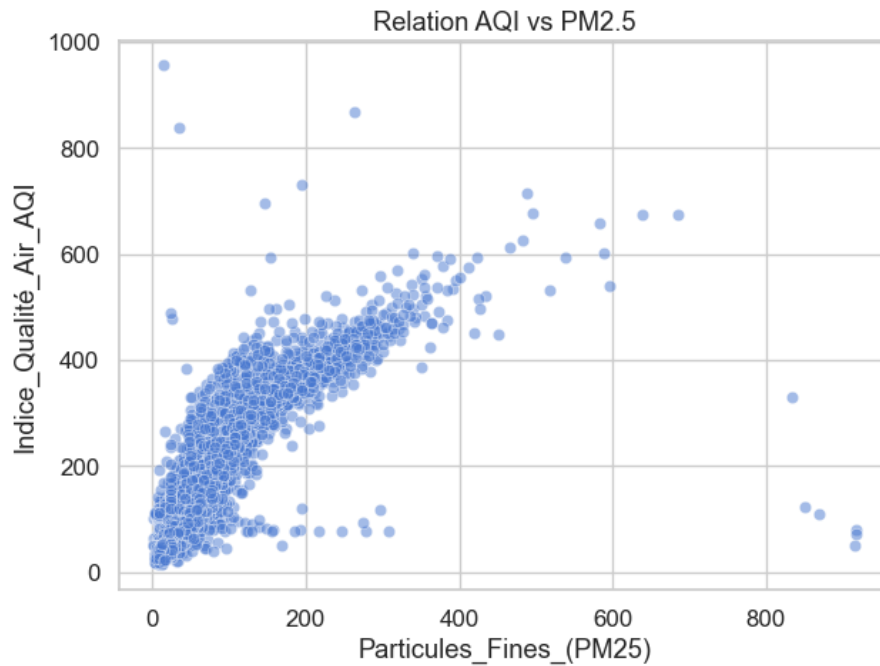


FIGURE 6 – Relation entre AQI et  $PM_{2.5}$  (scatterplot).

### 3.6 Analyse mensuelle et saisonnière

Les variations mensuelles moyennes de l'AQI (figure 7) montrent que les pics se situent entre novembre et janvier, tandis que les niveaux les plus bas apparaissent entre juin et septembre (mousson).

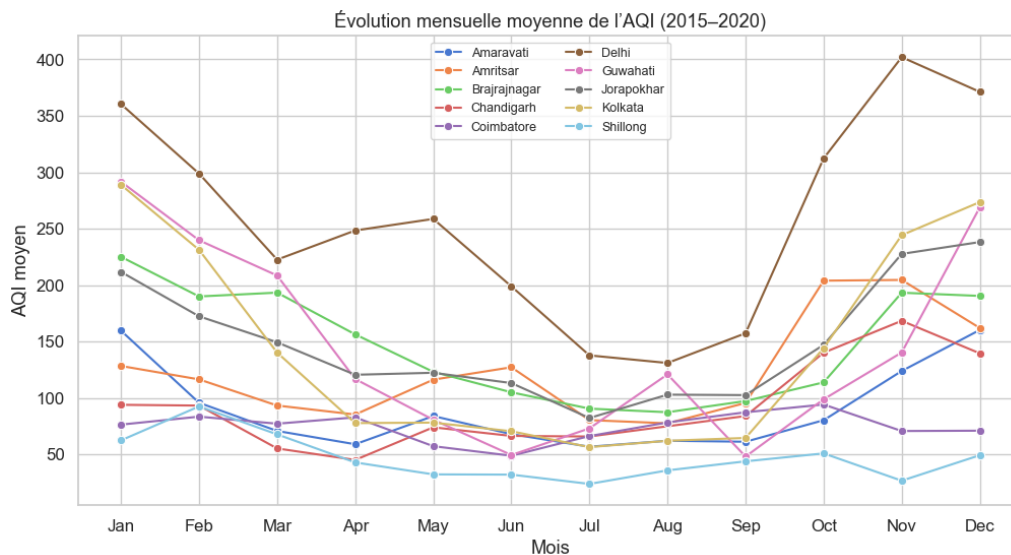


FIGURE 7 – Évolution mensuelle moyenne de l'AQI (2015–2020).

La distribution saisonnière (figure 8) confirme que la période hivernale (*Winter*) concentre les pires niveaux de pollution, suivie de l'automne.

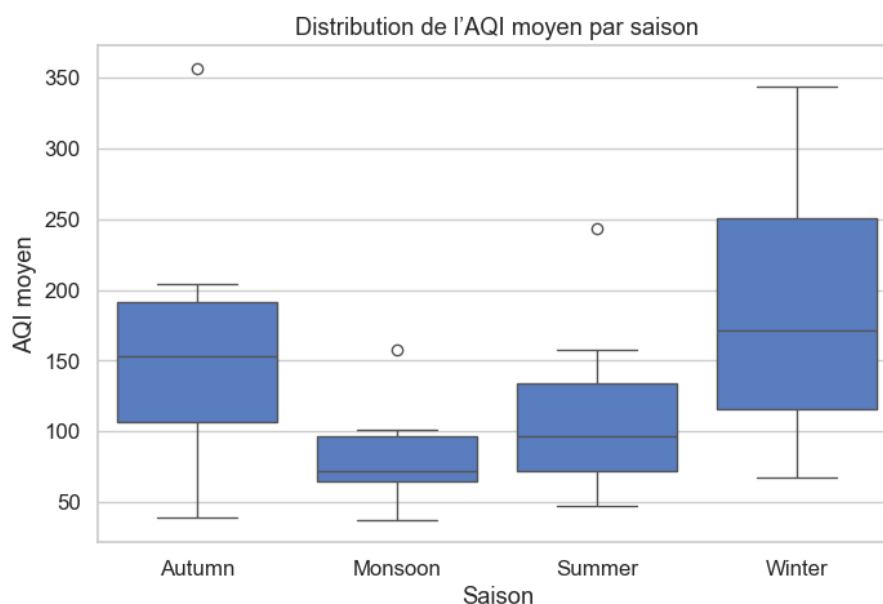


FIGURE 8 – Distribution moyenne de l'AQI par saison.

#### Insight - Dynamique temporelle

L'analyse temporelle met en évidence un schéma récurrent : la qualité de l'air se dégrade nettement en hiver en raison des conditions météorologiques défavorables (stagnation de l'air, émissions accrues, brûlage agricole), puis s'améliore pendant la mousson grâce aux précipitations.

#### Limitation

Certaines villes possèdent des séries temporelles incomplètes (moins de 2 ans de mesures continues), ce qui peut réduire la robustesse des comparaisons saisonnières. De plus, les effets météorologiques (vent, humidité, température) ne sont pas encore pris en compte à ce stade de l'analyse.

## 4 Visualisation géographique

Pour compléter l'analyse temporelle, une représentation spatiale de la qualité de l'air a été réalisée à l'aide de la bibliothèque *Folium*. Chaque cercle sur la carte correspond à l'une des dix villes étudiées, et sa couleur traduit la moyenne de l'Indice de Qualité de l'Air (AQI) sur la période 2015–2020 :

- vert : AQI < 100 → qualité d'air *bonne à satisfaisante*,
- orange : AQI entre 100 et 200 → pollution *modérée*,
- rouge : AQI > 200 → pollution *élevée à sévère*.

La figure 9 met en évidence une nette hétérogénéité spatiale :

- les villes du nord et du nord-est (Delhi, Amritsar, Kolkata, Guwahati) présentent les AQI moyens les plus élevés (> 200) ;
- les villes du sud (Coimbatore, Amaravati) affichent une qualité d'air globalement meilleure (AQI < 100) ;

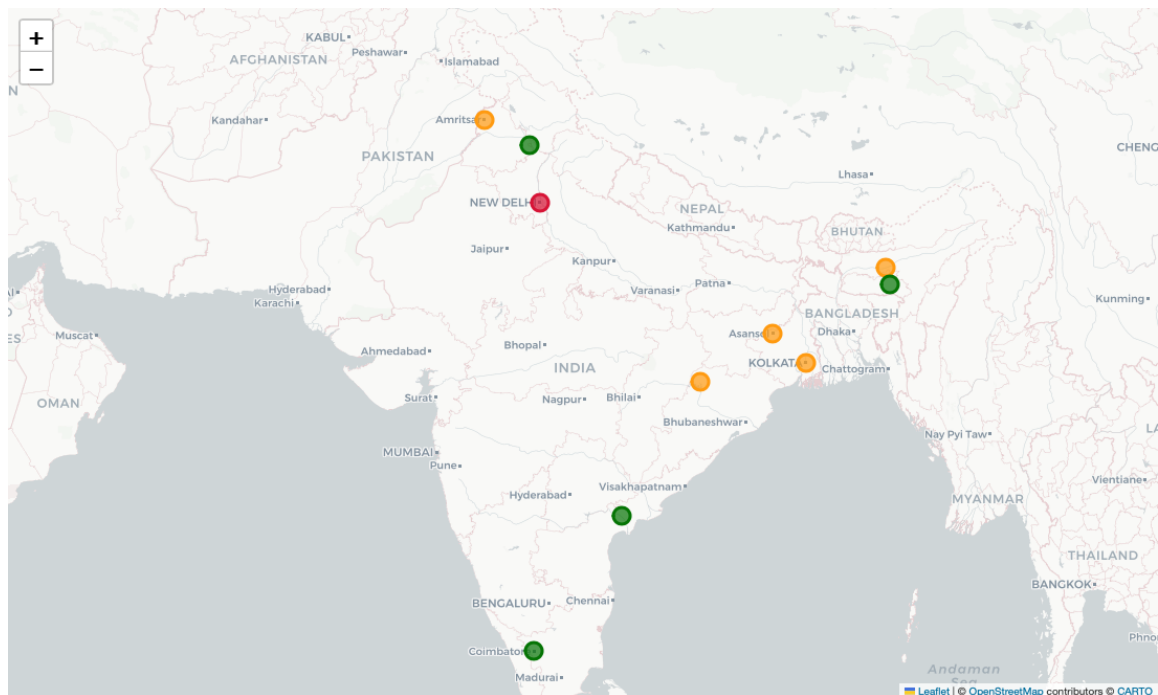


FIGURE 9 – Carte moyenne de la qualité de l’air (AQI moyen 2015–2020) dans les dix villes indiennes.

- les régions industrielles centrales (Brajrajnagar, Jorapokhar) se situent dans une zone intermédiaire (AQI 150–180).

### Insight – Lecture spatiale de la pollution

La visualisation géographique confirme la prédominance de la pollution dans la région nord de l’Inde, particulièrement autour de Delhi et Kolkata. La combinaison d’une forte densité urbaine, d’émissions industrielles et de conditions météorologiques défavorables explique ces concentrations élevées. Les villes du sud et du littoral bénéficient d’une meilleure dispersion atmosphérique et d’émissions plus faibles.

### Limitation

La carte statique représente des moyennes annuelles et ne tient pas compte des variations saisonnières. De plus, les coordonnées géographiques ont été saisies manuellement : de légers décalages peuvent exister entre les positions réelles des stations de mesure et leur représentation sur la carte.

## 5 Modélisation prédictive : Random Forest

### 5.1 Objectif du modèle

L’objectif de cette étape est de prédire la valeur de l’Indice de Qualité de l’Air (AQI) à partir des concentrations mesurées de polluants atmosphériques. La méthode de **ré-**

gression par forêts aléatoires (**Random Forest Regressor**) a été retenue pour sa robustesse face aux données bruitées et sa capacité à modéliser des relations non linéaires.

## 5.2 Variables utilisées

Les variables explicatives sélectionnées sont les suivantes :

$$\text{PM}_{2.5}, \text{PM}_{10}, \text{NO}_2, \text{SO}_2, \text{CO}, \text{O}_3, \text{Year}, \text{Month}$$

La variable cible est l'**AQI**.

Les données ont été divisées en deux sous-ensembles : **80 %** pour l'entraînement et **20 %** pour le test. Les lignes comportant des valeurs manquantes dans les variables clés ont été supprimées.

## 5.3 Paramètres et entraînement du modèle

Le modèle Random Forest a été entraîné avec les paramètres suivants :

- nombre d'arbres : 200 ;
- profondeur maximale : 10 ;
- aléa contrôlé : 42 ;

afin d'assurer une bonne généralisation sans sur-apprentissage.

## 5.4 Résultats et évaluation

Les performances du modèle sur l'échantillon test sont les suivantes :

$$\text{MAE} = 18.96, \quad \text{RMSE} = 33.78, \quad R^2 = 0.903$$

TABLE 5 – Indicateurs de performance du modèle Random Forest.

Métrique	Valeur	Interprétation
MAE	18.96	Erreur moyenne absolue (faible)
RMSE	33.78	Écart-type de l'erreur (faible dispersion)
$R^2$	0.903	90 % de la variance de l'AQI expliquée

Ces résultats démontrent que le modèle explique la majeure partie de la variabilité de l'AQI, tout en conservant une erreur moyenne inférieure à 20 points d'indice.

## 5.5 Importance des variables explicatives

La figure 10 présente la contribution relative de chaque variable à la prédiction de l'AQI. Les particules fines  $\text{PM}_{2.5}$  et  $\text{PM}_{10}$  sont de loin les plus influentes, suivies du dioxyde d'azote ( $\text{NO}_2$ ) et du monoxyde de carbone (CO).

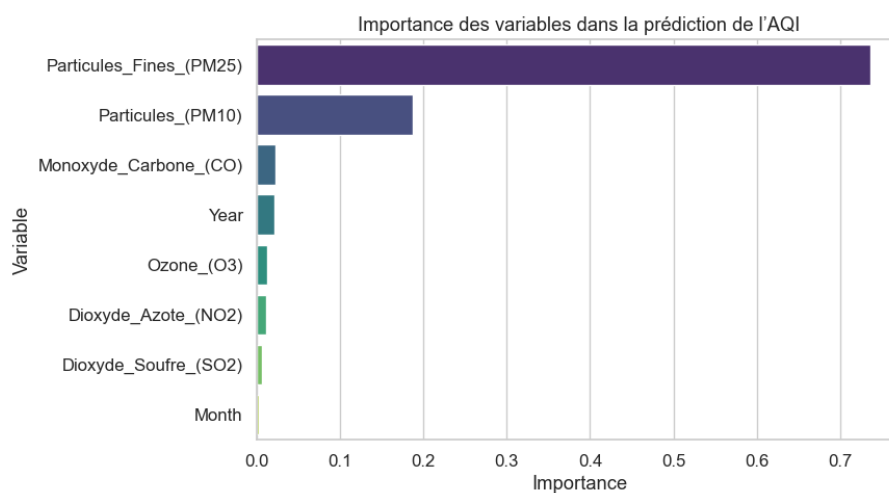


FIGURE 10 – Importance des variables dans la prédiction de l'AQI (Random Forest).

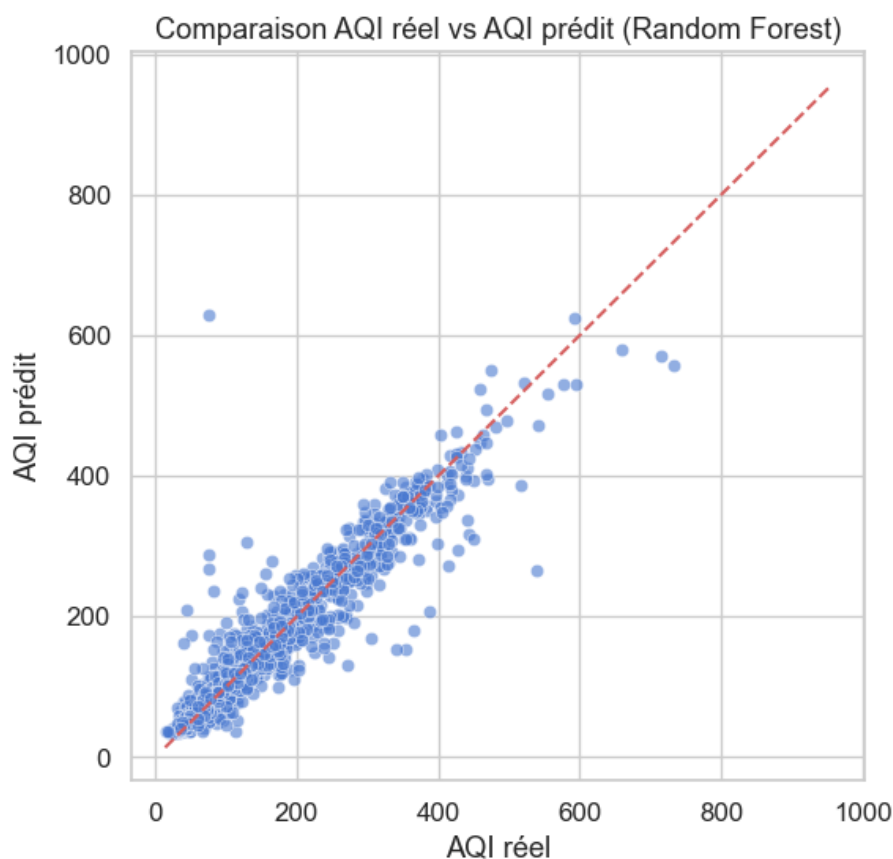


FIGURE 11 – Comparaison entre AQI réel et AQI prédit par le modèle Random Forest.

## 5.6 Comparaison AQI réel vs AQI prédit

La figure 11 illustre la relation entre les valeurs réelles et celles prédites par le modèle sur le jeu de test. La majorité des points se situent près de la diagonale, signe d'une excellente concordance.

### Insight – Prédiction de la qualité de l’air

Le modèle Random Forest atteint un coefficient de détermination  $R^2 \approx 0.90$ , ce qui signifie qu’il explique 90 % de la variabilité observée de l’AQI. Les variables  $PM_{2.5}$  et  $PM_{10}$  apparaissent comme les principaux déterminants de la pollution atmosphérique. Le modèle reproduit fidèlement les valeurs observées, avec une légère sous-estimation lors des pics extrêmes de pollution.

### Limitation

Les variables météorologiques (vent, température, humidité) ne sont pas intégrées ; elles pourraient améliorer la précision du modèle. Les performances élevées doivent également être interprétées avec prudence, car le modèle ne capture pas encore les dynamiques temporelles (inertie saisonnière) de la pollution.

## 6 Prévion temporelle : modèle Prophet

### 6.1 Objectif et motivation

Après l’analyse descriptive et la modélisation prédictive, cette étape vise à prévoir l’évolution future de la qualité de l’air (AQI) sur la base des tendances passées. Le modèle **Facebook Prophet** a été choisi car il détecte automatiquement les composantes temporelles principales : *tendance*, *saisonnalité annuelle* et *effets calendaires éventuels*. Il est particulièrement adapté aux séries longues avec saisonnalités marquées, comme celles que nous observons à Delhi.

### 6.2 Préparation des données

Seule la série temporelle de la ville de **Delhi** a été retenue, car elle présente la couverture temporelle la plus complète (2015–2020). Les variables ont été formatées selon les exigences de Prophet :

`ds = Date, y = AQI`

### 6.3 Entraînement et horizon de prévision

Le modèle Prophet a été entraîné sur les données historiques de 2015 à 2020, puis utilisé pour prédire l’évolution de l’AQI sur un horizon de **6 mois**, puis étendu jusqu’à **2028** afin d’observer les tendances à long terme.

### 6.4 Résultats et visualisations

La figure 12 montre la courbe des prévisions (en bleu) superposée aux observations historiques (points noirs). La zone bleue pâle représente l’intervalle de confiance à 95 %. La figure 13 présente les composantes temporelles extraites par le modèle :

- **Trend (tendance)** : une baisse progressive à partir de 2017–2018 ;
- **Yearly seasonality (saisonnalité annuelle)** : pics récurrents entre octobre et janvier ;

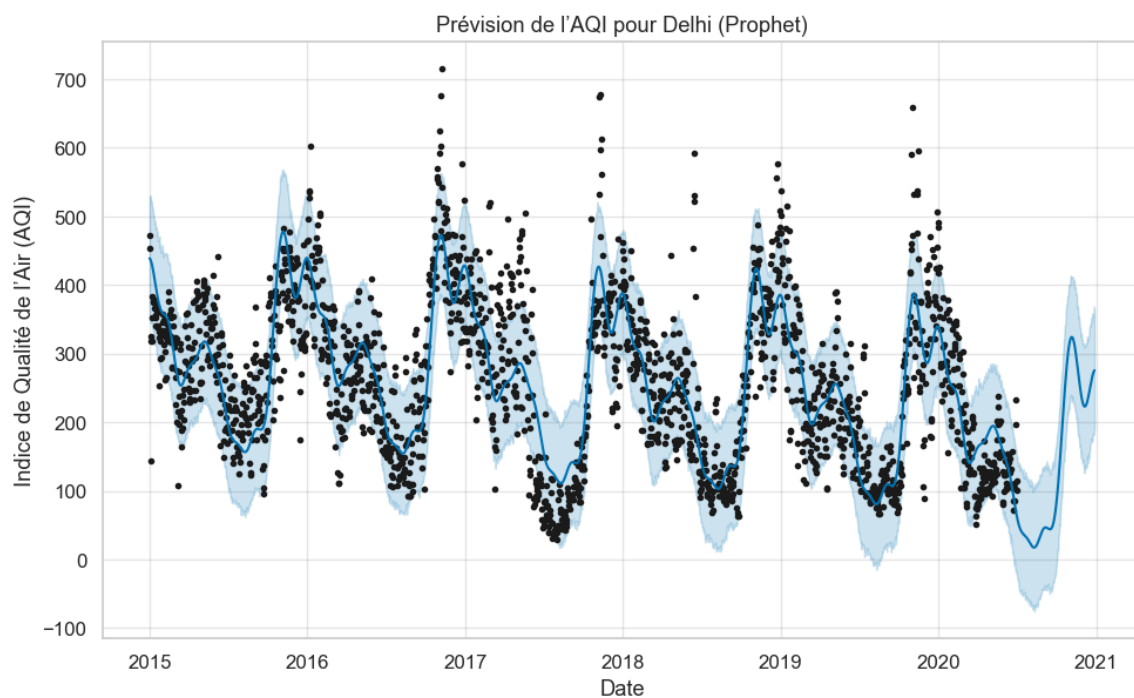


FIGURE 12 – Prévision de l'AQI pour Delhi à l'aide du modèle Prophet (horizon 6 mois).

— **Reste** : fluctuations irrégulières liées à des épisodes de pollution ponctuels.

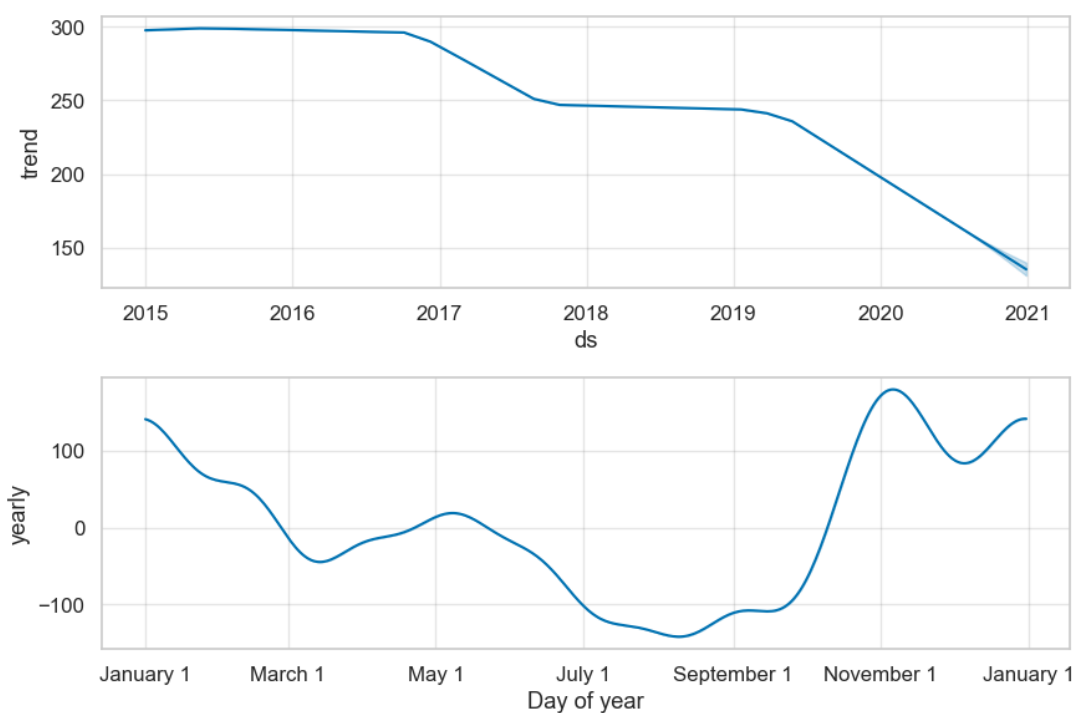


FIGURE 13 – Composantes temporelles détectées par Prophet (Delhi, 2015–2020).



## 6.5 Projection à long terme (jusqu'en 2028)

Le modèle a ensuite été étendu à un horizon de huit ans (figure 14). Prophet prévoit une tendance légèrement décroissante de l'AQI moyen, mais avec une saisonnalité persistante, marquée par des hausses hivernales et des baisses estivales.

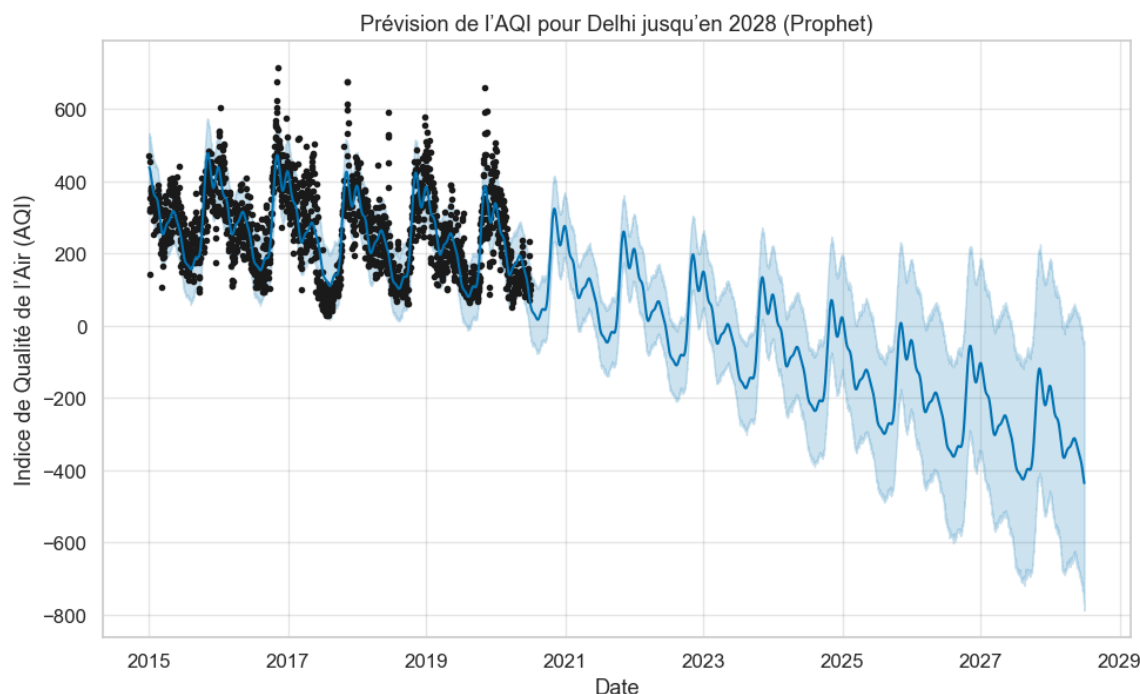


FIGURE 14 – Projection de l'AQI à Delhi jusqu'en 2028 (modèle Prophet).

### Insight – Tendance et saisonnalité de la qualité de l'air à Delhi

Prophet met en évidence une tendance légèrement **baissière** de l'AQI après 2017. Ce qui traduit les premiers effets des politiques environnementales et de la baisse d'activité de 2020. Cependant, la **saisonnalité reste très marquée** : chaque hiver, l'AQI dépasse largement les seuils critiques ( $> 300$ ), tandis que la mousson entraîne une amélioration temporaire de la qualité de l'air. À long terme, la pollution devrait diminuer lentement, sans disparition des pics saisonniers.

### Limitation

Les prévisions Prophet reposent uniquement sur les données historiques d'AQI et ne tiennent pas compte des variables explicatives (polluants, météo, politiques publiques). De plus, l'incertitude croît fortement après 2023, rendant les projections au-delà de 2025 indicatives plutôt que prédictives.

## 7 Résultats

Nos principales conclusions sont regroupées dans les encadrés ci-dessous.

## Insight global

### Résultats clés :

- $R_{RF}^2 \approx 0.90$  le modèle Random Forest explique 90 % de la variabilité observée de l'AQI ;
- $PM_{2.5}$  et  $PM_{10}$  sont les principaux déterminants de la pollution atmosphérique ;
- l'hiver (novembre–janvier) est la période la plus critique, avec des pics récurrents de pollution ;
- la tendance long terme est légèrement baissière, mais la saisonnalité reste persistante selon Prophet ;
- les villes du nord (Delhi, Kolkata, Guwahati) concentrent les plus hauts niveaux de pollution.

## Limitation globale

- Les données de certaines villes démarrent plus tard (séries temporelles incomplètes).
- Les valeurs manquantes ont été interpolées, ce qui peut lisser artificiellement certaines variations locales.
- Le modèle Prophet repose uniquement sur l'historique de l'AQI, sans intégrer les conditions météorologiques ni les émissions anthropiques réelles.
- Les prévisions à long terme (post-2023) comportent une incertitude importante.

## 8 Conclusion et recommandations

### 8.1 Synthèse générale

Cette étude a permis d'évaluer la dynamique spatio-temporelle de la qualité de l'air dans dix grandes villes indiennes entre 2015 et 2020. Les analyses révèlent que :

- la pollution atmosphérique reste un problème majeur, avec des niveaux d'AQI dépassant régulièrement les seuils recommandés par l'OMS ( $PM_{2.5}$  :  $15 \mu\text{g}/\text{m}^3/24\text{h}$ ,  $PM_{10}$  :  $45 \mu\text{g}/\text{m}^3/24\text{h}$ ,  $\text{NO}_2$  :  $25 \mu\text{g}/\text{m}^3/24\text{h}$ ,  $\text{O}_3$  :  $100 \mu\text{g}/\text{m}^3/8\text{h}$ ) (WHO, 2021)
- les particules fines  $PM_{2.5}$  et  $PM_{10}$  constituent les principaux agents de dégradation de l'air
- la saison hivernale demeure la plus problématique, en raison des inversions thermiques et du brûlage agricole
- les régions méridionales bénéficient d'une meilleure qualité de l'air grâce à une ventilation atmosphérique plus favorable.

### 8.2 Apport des modèles prédictifs

Les approches de modélisation mises en œuvre ont montré leur pertinence :

- le modèle Random Forest offre une estimation fiable de l'AQI à partir des concentrations de polluants mesurés ;

- le modèle Prophet a permis de capter la tendance et la saisonnalité, et d’anticiper les cycles de pollution ;
- combinés, ces outils fournissent une base solide pour la **prévision et la gestion proactive de la qualité de l’air**.

### 8.3 Recommandations

Pour améliorer la qualité de l’air et la précision des prévisions, il est recommandé de :

1. **Renforcer la collecte et la transparence des données** : inclure les variables météorologiques (vent, humidité, précipitations) et les données d’émission.
2. **Développer des systèmes de prévision intégrés** combinant intelligence artificielle et données satellitaires.
3. **Mettre en œuvre des politiques ciblées** pour réduire les émissions industrielles et le trafic routier dans les zones les plus critiques.
4. **Encourager les initiatives locales** (transports propres, urbanisme durable, sensibilisation citoyenne).

### 8.4 Perspectives

À l’avenir, cette étude pourrait être enrichie par :

- l’intégration de données météorologiques et socio-économiques pour améliorer la robustesse des modèles prédictifs ;
- l’application de réseaux de neurones récurrents (LSTM) pour la modélisation séquentielle des séries temporelles ;
- l’extension de l’analyse à d’autres pays d’Asie du Sud confrontés à des problématiques similaires.

#### Conclusion finale

Les résultats obtenus soulignent la pertinence de l’analyse combinée entre *statistiques exploratoires*, *modèles de machine learning* et *prévisions temporelles* pour comprendre et anticiper la pollution de l’air. Ce travail fournit une base solide pour des politiques publiques de surveillance et d’amélioration durable de la qualité de l’air en Inde.