



**HAL**  
open science

# Ozone troposphérique : caractérisation de sa dynamique spatiale et temporelle à échelle fine pour une meilleure évaluation de l'exposition de la population humaine dans la perspective du changement climatique

Honorine Gauthier-Manuel

## ► To cite this version:

Honorine Gauthier-Manuel. Ozone troposphérique : caractérisation de sa dynamique spatiale et temporelle à échelle fine pour une meilleure évaluation de l'exposition de la population humaine dans la perspective du changement climatique. Médecine humaine et pathologie. Université Bourgogne Franche-Comté, 2023. Français. NNT : 2023UBFCD009 . tel-04323434

**HAL Id: tel-04323434**

**<https://theses.hal.science/tel-04323434>**

Submitted on 5 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE**  
**PREPAREE A l'Université de Franche-Comté**

Ecole doctorale n°554

Ecole doctorale Environnement Santé

Doctorat de Médecine, santé publique, environnement et société

Par

Mme GAUTHIER-MANUEL Honorine

**Ozone troposphérique : caractérisation de sa dynamique spatiale et temporelle à échelle fine pour une meilleure évaluation de l'exposition de la population humaine dans la perspective du changement climatique**

Thèse présentée et soutenue à Besançon le 14 mars 2023

Composition du Jury :

|                       |   |                      |
|-----------------------|---|----------------------|
| Mme, BLOND, Nadège    | Chargée de recherche CNRS, UMR 7362 Laboratoire image ville environnement   | Rapporteur           |
| Mme, DEGUEN, Séverine | Chercheur Inserm, Université de Bordeaux  | Présidente           |
| Mme, PASCAL, Mathilde | Chargée de projet scientifique air, climat et santé, Santé Publique France  | Examinatrice         |
| Mr, SICARD, Pierre    | Chercheur, ARGANS   | Examineur            |
| Mme, BERNARD, Nadine  | Professeur, CNRS, UMR 6249 Chrono-environnement, UMR 6049 ThéMA   | Directrice de thèse  |
| Mr, MAUNY, Frédéric   | Professeur des Universités-Praticien Hospitalier, CNRS, UMR 6249 Chrono-environnement, Université de Bourgogne - Franche-Comté, CHU de Besançon | Codirecteur de thèse |

## Remerciements

En tout premier lieu, je souhaite remercier mes encadrants, les professeurs Nadine Bernard et Frédéric Mauny, pour la qualité de leur encadrement, leur enseignement, le temps qu'ils ont su prendre (même quand ils n'en avaient pas), et leur bienveillance à toute épreuve. Merci à eux de m'avoir donné l'opportunité de réaliser cette thèse et de m'avoir donné les meilleures conditions possibles pour la mener à bien. J'espère qu'ils verront à travers ces quelques lignes l'expression de ma profonde gratitude, de mon respect et de mon admiration pour leur implication sans faille pour la recherche.

Ensuite, merci au Pr Séverine Deguen et à Nadège Blond d'avoir accepté d'être rapportrices de mes travaux, et les docteurs Pierre Sicard et Mathilde Pascal d'en être les examinateurs. Je souhaite leur exprimer ma gratitude pour le temps qu'ils auront consacré à la lecture du manuscrit et aux commentaires et questions qui permettront d'améliorer mes travaux.

Merci également aux membres de mon comité de suivi scientifique de thèse, Renaud Scheifler et Mathilde Pascal, pour leur temps et leur aide.

Je remercie le Grand Besançon Métropole pour avoir financé ma bourse de thèse. Je remercie le Pôle fédératif de recherche et de formation en santé publique Bourgogne Franche-Comté pour avoir participé au financement d'une partie de ces travaux.

Un grand merci aux collaborateurs de chez ATMO BFC, Mathieu Boilleaut, Marie Ristori, Pablo Campargue et Anaïs Detournay, pour leur grande réactivité, leur efficacité et leur aide tout au long de cette thèse.

Je remercie également Mr Bruno Vermot-Desroches, Directeur du centre Météo France de Besançon, pour son aide dans la définition des données météorologiques à recueillir.

Je remercie les membres de l'uMETH (unité de méthodologie en recherche clinique épidémiologie et santé publique) pour leur accueil et pour avoir créé une ambiance de travail agréable. Merci en particulier à Bérenger Martin (datamanager) pour son aide dans le traitement des données pour certaines étapes clés. Merci aussi à Sophie pour son aide sur divers points techniques et administratifs, ainsi que pour sa

gentillesse. Je souhaite remercier tout particulièrement Anne-Laure Parmentier (pour son soutien, son aide, ses tableaux blancs et schémas, et les conversations et pause qui permettaient d'y voir plus clair), Marc Puyraveau (pour sa bonne humeur, sa gentillesse et ses conseils pour gérer la fin de thèse) et Grégory Tio (pour son soutien et pour m'avoir nourrie à coup de kinder, biscuits, chocolats, mikado et j'en passe ...). Merci à Delphine de m'avoir aidée à rester plus sereine avec ses séances de méditation. Et merci à Sophie avec qui j'ai pu partager les difficultés de la thèse mais également de bons moments.

Je tiens également à remercier mon ancien encadrant de stage, le Dr Frédéric Gimbert, car c'est lui qui m'a permis de découvrir la recherche et qui m'a donné envie de poursuivre dans cette voie.

Merci à Maxime pour son soutien, ses conseils et pour avoir toujours pris des nouvelles même de loin. Merci également à Nina pour son soutien et pour les moments partagés m'aidant à m'éloigner de la thèse.

Je remercie tout particulièrement Esteban, mon conjoint, pour avoir été présent et m'avoir soutenue tout au long de la thèse, avec les hauts et les bas que cela implique. Merci à lui de m'avoir toujours redonné le sourire quand le moral et la motivation flanchaient, bien que le monde de la recherche reste un mystère pour lui.

Un immense merci à Tony qui par sa simple présence, une discussion autour d'un café et quelques mots bien choisis, a toujours su me redonner la joie de vivre. Les mots me manquent pour exprimer toutes les raisons de ma gratitude, mais je mesure la chance que j'ai de l'avoir rencontré sur les bancs de la faculté.

Je termine ces remerciements par mes parents. A eux qui m'ont donné l'opportunité de poursuivre des études dans des conditions très favorables, qui m'ont soutenu et ont cru en moi ... Merci d'avoir rendu tout cela possible.

## Table des matières

|   |    |
|---|----|
| Remerciements .....   | 2  |
| Table des matières .....  | 4  |
| Valorisation au cours de la thèse .....   | 13 |
| Liste des abréviations .....  | 15 |
| Liste des figures.....  | 17 |
| Liste des tableaux .....  | 24 |
| Introduction générale.....  | 25 |
| Chapitre 1 : Ozone troposphérique.....  | 28 |
| Paramètres chimiques intervenant dans la formation de l’ozone troposphérique .....                          | 28 |
| 1. Nature des polluants atmosphériques primaires intervenant dans la formation d’ozone troposphérique ..... | 28 |
| 1.1. Oxydes d’azote .....   | 28 |
| 1.2. Composés organiques volatils .....   | 28 |
| 1.3. Radicaux hydroperoxyde et hydroxyle (HOx) .....  | 29 |
| 2. Réactions photochimiques et régime de formation d’ozone .....  | 30 |
| 2.1. Equilibre photo-stationnaire .....   | 30 |
| 2.2. Régimes de production d’ozone troposphérique : dépendance au rapport [NOx]/[COV]                       | 31 |
| 2.3. Régime saturé en NOx.....  | 33 |
| 2.4. Régime limité en NOx (ou saturé en COV) .....  | 33 |
| 2.4.1. Régime limité par les NOx : quelques dizaines de ppt < [NOx] < quelques ppb..                        | 34 |
| 2.4.2. Régime limité par les NOx : [NOx]<quelques ppt .....   | 35 |

|  |    |
|--|----|
| Paramètres météorologiques intervenant dans les fluctuations des concentrations d’ozone troposphérique.....          | 35 |
| 1. Température et rayonnement solaire.....   | 35 |
| 2. Vagues de chaleur.....  | 36 |
| 3. Régimes de vent.....  | 39 |
| 4. Humidité relative de l’air.....   | 40 |
| 5. Action conjuguée des différents paramètres météorologiques.....   | 40 |
| Paramètres physiques et biologiques de variations des concentrations d’ozone troposphérique.....                     | 41 |
| 1. Topographie.....  | 41 |
| 2. Influence de la végétation sur les concentrations d’O <sub>3</sub> troposphérique.....                            | 43 |
| Dynamiques spatiales de l’ozone troposphérique.....  | 45 |
| 1. Facteurs de variation spatiale.....   | 45 |
| 2. Distribution de l’O <sub>3</sub> et régimes de production d’O <sub>3</sub> en fonction du degré d’urbanisation... | 45 |
| Dynamiques temporelles de l’ozone troposphérique.....  | 47 |
| 1. Cycle nyctéméral.....   | 47 |
| 2. Variations hebdomadaires: l’effet weekend.....  | 49 |
| 3. Variations saisonnières.....  | 49 |
| 4. Evolution au cours des dernières décennies des concentrations d’O <sub>3</sub> et projections à moyen-terme       | 50 |
| Dynamique spatio-temporelle de l’ozone troposphérique et implications en termes d’exposition...                      | 51 |
| 1. Dynamique spatio-temporelle simplifiée de l’O <sub>3</sub> .....  | 51 |
| 2. Importance de la résolution temporelle et spatiale pour caractériser les dynamiques de l’O <sub>3</sub>           |    |

|        |   |    |
|--------|---|----|
| 3.     | Impacts des vagues de chaleur sur la dynamique spatio-temporelle de l’ozone troposphérique et risque de co-exposition à la chaleur et à l’ozone ..... | 54 |
|        | Evaluation des concentrations d’ozone troposphérique : outils de modélisation et évaluation de l’exposition de la population .....                    | 56 |
| 1.     | Outils de modélisation .....  | 56 |
| 1.1.   | Les modèles Land Use Regression (LUR) .....   | 56 |
| 1.2.   | Les Chemical Transport Models (CTMs).....   | 57 |
| 1.3.   | Validation des modèles et besoins potentiels d’optimisation .....   | 59 |
| 2.     | Evaluation de l’exposition de la population.....  | 60 |
| 2.1.   | Outils de mesure.....   | 60 |
| 2.2.   | Outils de modélisation.....   | 60 |
|        | Chapitre 2 : Matériels et Méthodes .....  | 62 |
|        | Evaluation de l’exposition.....   | 62 |
| 1.     | Modélisation brute des concentrations d’ozone.....  | 62 |
| 1.1.   | La plateforme de modélisation PREV’EST.....   | 62 |
| 1.2.   | Valeurs de références : stations de mesure de la qualité de l’air .....   | 63 |
| 2.     | Evaluation de la qualité de la modélisation brute .....   | 64 |
| 2.1.   | Indicateurs de performance moyens .....   | 64 |
| 2.2.   | Indicateurs à l’échelle de la station de mesure : Target plots .....  | 65 |
| 3.     | Méthodologie de correction des concentrations d’ozone.....  | 67 |
| 3.1.   | Modélisation de l’erreur de prédiction comme fonction de variables environnementales  |    |
|        | 67  |    |
| 3.1.1. | Modèles MARS .....  | 67 |

|   |    |
|---|----|
| 3.1.2. Optimisation des hyperparamètres des modèles.....  | 70 |
| 3.2. Différentes approches et corrections développées.....  | 71 |
| 3.3. Validation externe .....   | 72 |
| 3.4. Projection des niveaux d’ozone corrigés sur l’aire urbaine de Besançon et sa périphérie          |    |
| 73  |    |
| Identification de dynamiques spatio-temporelles d’ozone et de leurs déterminants .....                | 74 |
| 1. L’aire urbaine de Besançon.....  | 74 |
| 2. Classification des séries-temporelles d’ozone .....  | 77 |
| 2.1. Dynamic Time Warping .....   | 78 |
| 2.2. Groupement hiérarchique .....  | 80 |
| 3. Identification et caractérisation des phénomènes de vagues de chaleur .....                        | 81 |
| Distribution de la population et affectation des niveaux d’exposition .....                           | 82 |
| 1. Distribution de la population dans les bâtiments .....   | 82 |
| 2. Affectation des niveaux d’exposition au bâtiment.....  | 83 |
| Chapitre 3 : évaluation et amélioration des performances d’un modèle transnational pour une meilleure |    |
| évaluation de l’exposition à l’O <sub>3</sub> .....   | 84 |
| Improvement of downscaled ozone concentrations from the transnational scale to the kilometeric scale: |    |
| Need, interest and new insights .....   | 84 |
| Abstract: .....   | 86 |
| Keywords: Ozone, modeling, improvement, performance, MARS model .....                                 | 87 |
| Highlights: .....   | 87 |
| Graphical abstract:.....  | 87 |
| Study Funding .....   | 87 |



|   |     |
|---|-----|
| 1) Introduction: .....  | 88  |
| 2) Materials and Methods: .....   | 90  |
| 2.1 Assessment of the performance of the uncorrected PREV'EST model .....                           | 90  |
| 2.1.1. Study site and data sources.....   | 90  |
| 2.1.2. Performance indicators.....  | 92  |
| 2.2 Development and performance assessment of the corrections .....                                 | 93  |
| 2.2.1. Statistical relationship between the discrepancy and environmental variables .....           | 93  |
| 2.2.2. Modeling approaches.....   | 94  |
| 2.3.) External validation of the corrections .....  | 95  |
| 3) Results and Discussion.....  | 96  |
| 3.1) Evaluation of uncorrected PREV'EST model performance.....                                      | 96  |
| 3.2) Evaluation of the performance of the corrections .....   | 98  |
| 3.3. External validation.....   | 103 |
| 4 Strengths of the methodology and comparison with existing methodologies .....                     | 107 |
| 5. Conclusion:.....   | 107 |
| Acknowledgments .....   | 108 |
| References .....  | 108 |
| Chapitre 4 : Identification de dynamiques spatiales et temporelles d'ozone et de leurs déterminants | 117 |
| Résumé : .....  | 117 |
| Abstract: .....   | 120 |
| Study Funding .....   | 121 |
| 1) Introduction: .....  | 122 |
| 2) Materials and Methods .....  | 124 |

|  |   |     |
|--|---|-----|
| 2.1)   | Study area .....  | 124 |
| 2.2)   | Ozone concentration modeling .....                                      | 125 |
| 2.3)   | Elevation, urbanization, traffic and vegetation characteristics .....   | 126 |
| 2.4)   | Statistical analysis.....   | 127 |
| 2.4.1  | Dynamic Time Warping.....   | 127 |
| 2.4.2  | Hierarchical clustering .....   | 128 |
| 2.4.3  | Descriptive and analytical statistics.....                              | 129 |
| 3)   | Results .....   | 129 |
| 3.1)   | Classification of the temporal dynamics.....                            | 129 |
| 3.2)   | Characterization and differentiation of the temporal dynamics.....      | 130 |
| 3.3)   | Relationship between O <sub>3</sub> levels and its determinants .....   | 133 |
| 4)   | Discussion: .....   | 136 |
| 4.1)   | Reliability of the data and the method .....                            | 136 |
| 4.2)   | Influence of the determinants on O <sub>3</sub> levels .....            | 137 |
| 4.3)   | Complexity of O <sub>3</sub> dynamics and management implications ..... | 139 |
| 5)   | Conclusion:.....  | 140 |
|  | Acknowledgements: .....   | 141 |
|  | References: .....   | 141 |
| Chapitre 5 : Caractérisation de la contamination de la zone d'étude au regard des recommandations de l'Organisation Mondiale de la Santé et des valeurs cibles réglementaires pour la protection de la santé ..... |   |     |
|  |   | 151 |
| 1)   | Introduction .....  | 152 |
| 2)   | Matériels et Méthodes .....   | 152 |

|   |  |     |
|---|--|-----|
| 2.1   | Zone d'étude : l'aire urbaine de Besançon.....   | 152 |
| 2.2   | Indicateurs réglementaires sanitaires classiquement utilisés .....   | 153 |
| 2.3   | Estimation du maximum journalier de la moyenne sur huit heures de la concentration atmosphérique en ozone (MDA8) ..... | 154 |
| 3)  | Résultats .....  | 156 |
| 4)  | Discussion .....   | 163 |
| 5)  | Conclusion :.....  | 164 |
|   | Références bibliographiques : .....  | 165 |
| Chapitre 6 : Approche descriptive et exploratoire des vagues de chaleur et de leurs relations avec le niveau d'exposition à l'ozone de la population et du degré d'urbanisation ..... |  |     |
|   |  | 166 |
| 1)  | Introduction : .....   | 167 |
| 2)  | Matériels et méthodes : .....  | 167 |
| 1.  | Typologie de territoires caractérisant le degré d'urbanisation .....   | 167 |
| 2.  | Chaîne de traitement de données .....  | 168 |
| 2.1   | Association des différentes sources de données par requête spatiale .....  | 168 |
| 2.2   | Vagues de chaleur (VDC) et concentrations d'O <sub>3</sub> .....   | 168 |
| 2.2.1   | Définition des périodes de vague de chaleur.....   | 168 |
| 2.2.2   | Classification des vagues de chaleur.....  | 170 |
| 2.2.3   | Mise en relation des périodes d'exposition avec les concentrations d'O <sub>3</sub> .....                              | 170 |
| 3.  | Calcul du nombre de dépassements de seuils .....   | 170 |
| 4.  | Analyse statistique .....  | 171 |
| 3)  | Résultats : .....  | 171 |
| 1)  | Description, classification et caractérisation des vagues de chaleur.....  | 171 |

|  |     |
|--|-----|
| 2) Association de la typologie de VDC et leurs caractéristiques aux niveaux d'exposition à l'O <sub>3</sub>  | 175 |
| 2.1) Caractérisation générale des typologies de VDC.....   | 175 |
| 2.2) MDA8 selon les classes de VDC et d'urbanisation .....   | 177 |
| 2.3 Dépassements de seuils selon les classes de VDC et d'urbanisation .....  | 180 |
| 2.4) Distribution du niveau d'exposition au sein de la population de la zone d'étude.....  | 184 |
| 3) Approche comparative des niveaux d'exposition de la population au cours des vagues de chaleurs des deux années les plus contrastées de la période 2008-2019 ..... | 185 |
| 4) Discussion : .....  | 187 |
| 5) Conclusion :.....   | 190 |
| Références bibliographiques : .....  | 190 |
| Chapitre 7 : Discussion générale et perspectives.....  | 193 |
| 1) Synthèse des principaux résultats.....  | 193 |
| 2) Les contrastes d'exposition : concept central de notre démarche .....   | 195 |
| 3) Croisement des différentes échelles spatiales et temporelles des données mobilisées.....  | 195 |
| 4) Forces .....  | 198 |
| 5) Limites.....  | 199 |
| 6) Perspectives .....  | 199 |
| Conclusion.....  | 201 |
| Références bibliographiques .....  | 202 |
| Annexes.....   | 222 |

|   |     |
|---|-----|
| Annexe 1 : Annexes de l'article publié dans Environmental Research intitulé "Improvement of downscaled ozone concentrations from the transnational scale to the kilometric scale: need, interest and new insights" .....            | 222 |
| Annexe 2 : Annexes de l'article publié dans la revue Environment International intitulé "Spatialized temporal dynamics of daily ozone concentrations: identification of its main driving factors and management implications" ..... | 228 |

## **Valorisation au cours de la thèse**

### **Article publié :**

- H., Gauthier-Manuel, F., Mauny, M., Boilleaut, M., Ristori, S., Pujol, F., Vasbien, A.-L., Parmentier, N., Bernard, 2022. Improvement of downscaled ozone concentrations from the transnational scale to the kilometric scale: Need, interest and new insights. Environmental Research 210, 112947. <https://doi.org/10.1016/j.envres.2022.112947>

### **Article soumis :**

- H., Gauthier-Manuel, N., Bernard, M., Boilleaut, P., Giraudoux, S., Pujol, F., Mauny, Spatialized temporal dynamics of daily ozone concentrations: identification of its main driving factors and management implications. Environment International. Soumis le 20 octobre 2022 (en relecture par les reviewers).

### **Communications orales :**

- H., Gauthier-Manuel, F., Mauny, M., Boilleaut, M., Ristori, S., Pujol, A.-L., Parmentier, N., Bernard. 2021. Optimisation d'un modèle atmosphérique transnational de prédiction de la concentration journalière d'ozone : intégration de paramètres météorologiques pour une évaluation de l'exposition à l'échelle kilométrique. 34<sup>ième</sup> colloque de l'Association Internationale de Climatologie, Jul 2021, Casablanca, Maroc.
- H., Gauthier-Manuel, F., Mauny, M., Boilleaut, M., Ristori, S., Pujol, J., Saxer, A.-L., Parmentier, N., Bernard. 2020. Concentrations d'ozone à une échelle spatio-temporelle fine: d'une modélisation transnationale à une application locale. Congrès 2020 de la SFSE. Nov. 2020. Workshop « jeunes chercheurs ».

### **Communications affichées :**

- H. Gauthier-Manuel, F. Mauny, M. Boilleaut, M. Ristori, S. Pujol, F. Vasbien, A.-L. Parmentier, N. Bernard. 2022. Should and can ozone exposure be better assessed? SETAC Europe 32nd Annual Meeting. May 2022.

- H. Gauthier-Manuel., F. Mauny, M. Boilleaut, M. Ristori, S. Pujol, Jill Saxer, A.-L. Parmentier, N. Bernard. 2020. Modélisation des concentrations en ozone, de l'échelle transnationale vers une application locale : ajustement à l'aide de données environnementales Congrès 2020 de la SFSE. Nov. 2020.

**Médiation scientifique :**

- H. Gauthier-Manuel, F. Mauny. 2021. Atelier « Pollutions, vagues de chaleur et vulnérabilités des populations » au cours de la 3<sup>ème</sup> journée annuelle du Pôle fédératif de recherche et de formation en Santé-Publique Bourgogne-Franche-Comté. Déc. 2021.

## Liste des abréviations

AFNOR : Association française de normalisation

ASQAA : Associations Agréées de Surveillance de la Qualité de l'Air

ATMO BFC : ASQAA de Bourgogne-Franche-Comté

CAH : Classification Ascendante Hiérarchique

CLA : Couche limite atmosphérique

COVB : Composé organique volatil biogénique

COVNM : Composé organique volatil non-méthanique

CTM : Chemical Transport Model

DTW : Dynamic Time Warping

FDep : French Deprivation Index

HO<sub>2</sub> : radical hydroperoxyde

HO<sub>x</sub> : radicaux OH et HO<sub>2</sub> (x = 1 ou 2) ainsi que les radicaux peroxydes organiques (RO<sub>2</sub>)

IGN : Institut national de l'information géographique et forestière

INSEE : Institut National de la Statistique et des Etudes Economiques

IRIS : Îlot Regroupé pour l'Information Statistique

ISO : Organisation internationale de normalisation

LAURE : Loi sur l'Air et l'Utilisation Rationnelle de l'Energie

LCM : Local Cost Matrix

LUR : Land Use Regression.

MARS : Multivariate Adaptive Regression Splines

MDA8 : maximum journalier de la moyenne glissante sur 8h (maximum daily 8-hour average)



NO : monoxyde d'azote

NO<sub>2</sub> : dioxyde d'azote

NO<sub>x</sub> : Oxydes d'azote

O<sub>3</sub> : Ozone

OH : radical hydroxyle

ppb : *part per billion* - partie par milliard

ppt : *part per trillion* - partie par billion

PREV'EST : Plateforme de modélisation transnationale gérée par ATMO Grand Est

RO<sub>2</sub> : radicaux peroxydes organiques

VDC : vague de chaleur

## Liste des figures

|   |    |
|---|----|
| Figure 1: Représentation schématique des principaux mécanismes impliqués dans la formation d'ozone troposphérique (Camredon et Aumont, 2007). Les réactions notées R2 à R9 font référence aux réactions décrites dans le corps du texte. ....   | 30 |
| Figure 2: Représentation schématique des réactions impliquées dans la conversion de NO en NO <sub>2</sub> et la formation d'O <sub>3</sub> , d'une part (A) en situation d'équilibre photostationnaire et d'autre part (B) en présence de COV (d'après (Atkinson, 2000)) .....  | 31 |
| Figure 3 : Diagramme isopleth type adapté de Seinfeld et Pandis (1998).....   | 32 |
| Figure 4 : Profil type de l'efficacité de production d'ozone en fonction de la concentration en NO <sub>x</sub> , tiré de Camredon et al., 2007, et adapté de (Liu et al., 1992; McKeen et al., 1991) .....   | 34 |
| Figure 5: Caractérisation d'une vague de chaleur basée sur la température moyenne journalière en France (exemple de série temporelle du 30 juin au 5 août 1983) : durée (start – end), température maximale et intensité globale (aire rouge sur le graphique). Les températures supérieures (inférieures) à la ligne climatologique (période de référence 1981-2010) sont représentées par une zone rose (bleue). D'après Ouzeau et al. (2016). .... | 37 |
| Figure 6: Inventaire et caractérisation des vagues de chaleur (intensité en ordonnées, durée en abscisses, et surface du cercle représentant la sévérité de l'événement) de 1947 à 2020. Les vagues de chaleur des vingt dernières années sont en surbrillance. Illustration produite par Météo France (outil en ligne Climat HD, <a href="https://meteofrance.com/climathd">https://meteofrance.com/climathd</a> ). ....                             | 38 |
| Figure 7: Projections des phénomènes de vagues de chaleur sur la période 2071-2100 selon les scénarios RCP2.6 et RCP8.5, illustration produite par Météo France (outil en ligne Climat HD). ....  | 39 |
| <i>Figure 8: Influence des paramètres météorologiques sur les concentrations d'ozone troposphériques (Nguyen et al., 2022). ....</i>  | 41 |
| Figure 9: Illustration de la chimie atmosphérique de l'O <sub>3</sub> et des NO <sub>x</sub> , ainsi que de la structure de la CLA au cours de la journée et de la nuit (Nguyen et al., 2022). ....   | 42 |

|   |    |
|---|----|
| Figure 10: Concentrations horaires d'O <sub>3</sub> et de NO <sub>x</sub> à l'échelle journalière au niveau d'un site périurbain en 2007, pendant la semaine (WD) et le weekend (WE) (Banja et al., 2009).....  | 48 |
| Figure 11: Schéma synthétisant et simplifiant la dynamique spatio-temporelle de l'O <sub>3</sub> à travers les principaux déterminants influant sur la variation des concentrations d'O <sub>3</sub> spatialement et/ou temporellement.....                         | 52 |
| Figure 12: Domaine de modélisation de la plateforme PREV'EST (grisée), et sous-domaine de modélisation (région Franche-Comté) ayant servi à l'évaluation des performances du modèle. ....   | 63 |
| Figure 13: Exemple de target plot pour 16 stations de mesures (Bonafè, 2017). Les différentes lettres représentent les différentes stations.....  | 65 |
| Figure 14: Exemple de noeuds ou "turning points" identifié avec un modèle MARS dans la relation entre une variable dépendante (Y) et une variable explicative (X),tiré et adapté de (Zhou and Leung, 2007).<br>.....  | 69 |
| Figure 15 : Paire de fonction de base avec $t = 0,5$ . Illustration tirée de Hastie et al., 2009. Le spline de gauche ( $x < t$ , $-x - t + q$ ) est représenté par une ligne en pointillée ; celui de droite ( $x > t$ , $+x - t + q$ ) par une ligne pleine. .... | 69 |
| Figure 16 : Processus de modélisation et de recompilation des prédictions pour les approches globale, annuelle et territoriale.....   | 72 |
| Figure 17: étapes séquentielles de la correction des concentrations d'O <sub>3</sub> à l'aide du modèle MARS "2WF-elevation".....   | 73 |
| Figure 18: Secteur du Grand Besançon Métropole et ses communes environnantes catégorisées d'urbain à rural. Les effectifs indiqués à côté de chaque typologie sont les effectifs de communes appartenant à cette typologie sur le territoire métropolitain. ....    | 74 |
| Figure 19: Densité de population au km <sup>2</sup> par commune pour la zone de Besançon Grand Métropole et les communes environnantes.....   | 75 |
| Figure 20: Répartition des émissions de NO <sub>x</sub> et COVNM par secteurs d'activités agrégées au département du Doubs (2018) selon l'inventaire territorial des émissions effectué par ATMO BFC (Bréhon et al., 2018).....                                     | 76 |

|   |     |
|---|-----|
| Figure 21: Zone d'étude comprenant le maillage de modélisation utilisé pour la modélisation des concentrations journalières d'ozone.....  | 77  |
| Figure 22: Comparaison de deux séries temporelles avec a) la distance Euclidienne et b) le DTW(Giao and Anh, 2016).....   | 78  |
| Figure 23: Calcul de la matrice LCM calculée pour aligner deux séries temporelles x1 et x2 (Tomašev et al., 2015).....  | 79  |
| Figure 24: a. Pour aligner les séries C et Q, un "warping path" P, illustré avec des carrés pleins, est construit. b. La contrainte de la bande de Sakoe-Chiba de largeur r est utilisé comme contrainte globale pour limiter la portée de P (Giao and Anh, 2016). ....   | 80  |
| Figure 25: Caractérisation d'une vague de chaleur basée sur la température moyenne journalière en France (exemple de série temporelle du 30 juin au 5 août 1983) : durée (start – end), température maximale et intensité globale (aire rouge sur le graphique). Les températures supérieures (inférieures) à la ligne climatologique (période de référence 1981-2010) sont représentées par une zone rose (bleue). D'après Ouzeau et al. (2016)..... | 82  |
| Figure 26: PREV'EST modeling area (for the 3 km x 3 km grid mesh, on the top map) and location of the air quality monitoring stations (ATMO BFC) and weather stations (Météo-France). ....  | 91  |
| Figure 27: Target plots of the uncorrected PREV'EST modeling. ....  | 98  |
| Figure 28: Target plots of the uncorrected PREV'EST modeling, global approach 4-WF and 2-WF elevation corrections, annual approach 4-WF and 2-WF elevation corrections, territorial approach 4-WF correction and territorial and annual approach 4-WF correction target .....   | 101 |
| Figure 29: Daily mean ozone concentrations measured, modeled by uncorrected PREV'EST modeling and corrected by the annual 2-WF elevation correction, during 2018 at Station A. ....   | 103 |
| Figure 30: External validation: uncorrected PREV'EST modeling, annual approach 4-WF correction, annual approach 2-WF elevation correction, and annual and territorial approach 4-WF and 2-WF correction target plots.....   | 106 |
| Figure 31: Location of the study site (Besançon), indicated by a red rectangle. The frequency of wind direction and speed is represented by a wind rose in the left top corner of the figure. ....  | 125 |

|  |     |
|--|-----|
| Figure 32: Time-series of the daily average of the mean ozone concentration per class (A, B, or C) from 2008 to 2019 on the top graph and in 2019 on the bottom graph.....   | 130 |
| Figure 33: Boxplots of the average 12-year mean and maximum daily ozone concentrations, mean elevation and proportion of urbanized area, according to the class. The different letters above the boxplots represent significant differences between classes (Kruskalmc).....   | 133 |
| Figure 34: Correlation matrix (Spearman correlation) between the characteristics. The following significance levels (p values): 0, 0.001, 0.01, 0.05, 1 are displayed by the following symbols: "****", "***", "**", "*", " ". The red line in the plots on the bottom left part of the matrix represents the local regression (loess smooth) surrounded by the 95% confidence interval..... | 134 |
| Figure 35: Spatial distribution of the classes (represented by different shapes) through the modeling grid. The color filling the shapes (orange levels) represents the average of the daily mean ozone concentration (2008-2019) per modeling mesh. The background color represents the mean elevation of the modeling mesh. ....   | 136 |
| Figure 36: Couverture spatiale de la zone d'étude et maillage de modélisation de la plateforme PREV'EST (3km x 3km). ....  | 153 |
| Figure 37: Distribution des résidus du modèle linéaire multiple ( $MDA8 \sim [O3]_{\text{maximum}} + ([O3]_{\text{maximum}} - [O3]_{\text{moyenne}})$ ) en fonction des valeurs ajustées. ....   | 155 |
| Figure 38: Nombre de dépassements du seuil OMS (jours) par maille de modélisation en fonction de l'année. Le point rouge correspond à la valeur de la maille sur laquelle se situe la station de suivi de la qualité de l'air. La ligne pointillée bleue représente le nombre de dépassements qui ne doit pas être excédé.....   | 159 |
| Figure 39: Nombre de dépassements (jours) du seuil réglementaire européen par maille de modélisation en fonction de l'année. Le point rouge correspond à la valeur de la maille sur laquelle se situe la station de suivi de la qualité de l'air. La ligne pointillée bleue représente le nombre de dépassements qui ne doit pas être excédé.....  | 160 |

|   |     |
|---|-----|
| Figure 40: Valeur du pic saisonnier ( $\mu\text{g.m}^{-3}$ ), en fonction de l'année. Le point rouge correspond à la valeur de la maille sur laquelle se situe la station de suivi de la qualité de l'air. La ligne pointillée bleue représente la valeur seuil à ne pas excéder. ....  | 161 |
| Figure 41 : Caractérisation d'une vague de chaleur à partir de l'indicateur thermique quotidien sur la France en durée (date de début et de fin), intensité maximum (cercle rouge) et sévérité (aire colorée en marron). D'après Soubeyroux et al. (2016).....  | 169 |
| Figure 42: Profil de température moyenne journalière ( $^{\circ}\text{C}$ ) pour chaque vague de chaleur identifiée à Besançon entre 2008 et 2019.....  | 173 |
| Figure 43: Dendrogramme de la classification hiérarchique ascendante des vagues de chaleur. Les numéros reportés en bas du dendrogramme correspondent au numéro identifiant des vagues de chaleur (voir Tableau 11).....  | 174 |
| Figure 44: Profil du Within-cluster Sum of Squares (WSS) en fonction du nombre de groupes sélectionnés pour la classification hiérarchique ascendante.....  | 174 |
| Figure 45: Durée, sévérité et intensité des vagues de chaleur en fonction de la typologie de vagues de chaleur. ....  | 177 |
| Figure 46: Maximum journalier de la moyenne glissante sur 8 heures de la concentration d'ozone ( $\mu\text{g}/\text{m}^3$ ) en fonction de la typologie de vagues de chaleur et de la typologie de territoire (tr1_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense). ....   | 178 |
| Figure 47: Maximum journalier de la moyenne glissante sur 8 heures de la concentration d'ozone ( $\mu\text{g}/\text{m}^3$ ) en fonction de la typologie de vagues de chaleur.....   | 178 |
| Figure 48: Maximum journalier de la moyenne glissante sur 8 heures de la concentration d'ozone ( $\mu\text{g}/\text{m}^3$ ) pour chaque VDC en fonction de la typologie de territoire (tr1_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense). Le code couleur retranscrit la typologie de vagues de chaleur (A, B et C)..... | 179 |

|  |     |
|--|-----|
| Figure 49: Nombre de dépassements (exprimé en jours) de seuil OMS (100 $\mu\text{g}/\text{m}^3$ ) et réglementaire européen (120 $\mu\text{g}/\text{m}^3$ ) en fonction de la typologie de VDC.....  | 180 |
| Figure 50 : Nombre de dépassements de seuil OMS (100 $\mu\text{g}/\text{m}^3$ ) (exprimé en jours) en fonction de la typologie de VDC et de la typologie de territoire (tr1_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense).<br>.....   | 181 |
| Figure 51 : Nombre de dépassements de seuil règlementaire européen (120 $\mu\text{g}/\text{m}^3$ ) (exprimé en jours) en fonction de la typologie de VDC et de la typologie de territoire (tr1_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense).....   | 181 |
| Figure 52: Nombre de dépassements de seuil OMS (100 $\mu\text{g}/\text{m}^3$ ) (exprimé en jours) pour chaque VDC en fonction de la typologie de territoire (tr1_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense). Le code couleur retranscrit la typologie de vagues de chaleur (A, B et C). ....   | 182 |
| Figure 53: Nombre de dépassements de seuil réglementaire (120 $\mu\text{g}/\text{m}^3$ ) (exprimé en jours) pour chaque VDC en fonction de la typologie de territoire (tr1_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense). Le code couleur retranscrit la typologie de vagues de chaleur (A, B et C). ....   | 183 |
| Figure 54: Histogramme de distribution de la population en fonction du maximum de la moyenne glissante sur 8 heure d'O <sub>3</sub> ( $\mu\text{g}/\text{m}^3$ ), par vague de chaleur. Le code couleur retranscrit la typologie de VDC. ....  | 184 |
| Figure 55: Distribution du niveau d'exposition de la population de la zone d'étude (moyenne du MDA8 ( $\mu\text{g}/\text{m}^3$ ) au cours des VDC d'une même année), années 2008 et 2019 (aucune valeur ne dépasse 110 $\mu\text{g}/\text{m}^3$ pour l'année 2008). La ligne verticale rouge représente la valeur de MDA8 pour la maille sur laquelle se situe la station de mesure de Besançon en 2008. La ligne verticale bleue représente cette même valeur pour l'année 2019. .... | 186 |
| Figure 56: Apport de des quatre chapitres et synthèse des résultats du manuscrit.....  | 194 |

Figure 57 : Synthèse des différentes échelles ou unités temporelles et spatiales mobilisées dans les différents chapitres de la thèse. .... 197



## Liste des tableaux

|   |     |
|---|-----|
| Tableau 1: Forces et faiblesses de deux approches de modélisation : LUR et CTM. ....  | 58  |
| Table 2: MARS modeling approaches and corrections .....   | 95  |
| Table 3: Average performance indicators of the different approaches and corrections .....   | 96  |
| Table 4: Average performance indicators calculated for the external validation. ....  | 104 |
| Table 5: Position and dispersion parameters of the different characterization variables per class. ....   | 131 |
| Table 6: Results of the simple and multiple linear regression and polynomial model of the 12-year mean ozone concentration ( $\mu\text{g}\cdot\text{m}^{-3}$ ) against the main driving factors. ....   | 135 |
| Tableau 7 : Résultat de la régression linéaire multiple ayant pour variable expliquée le MDA8 et pour variables explicative la concentration maximale journalière d'O <sub>3</sub> et l'écart entre la concentration maximum et moyenne journalière. ....                           | 155 |
| Tableau 8: Descriptif du nombre de dépassements du seuil annuel moyen par maille de modélisation pour la totalité de la période 2008-2019, recommandation OMS (100 $\mu\text{g}/\text{m}^3$ ) et seuil réglementaire européen (120 $\mu\text{g}/\text{m}^3$ ).....                  | 156 |
| Tableau 9: Descriptif du nombre de dépassements du seuil OMS (100 $\mu\text{g}/\text{m}^3$ ) et du seuil réglementaire (120 $\mu\text{g}/\text{m}^3$ ) par maille de modélisation et par année, ainsi que la valeur du pic saisonnier par maille de modélisation et par année. .... | 158 |
| Tableau 10: Effectif de population générale, et effectif et pourcentage de population concernée par un dépassement du seuil OMS et réglementaire européen par année.....  | 162 |
| Tableau 11: Inventaire et indicateurs de caractérisation des vagues de chaleur à Besançon entre 2008 et 2019.....   | 171 |
| Tableau 12: Tableau récapitulatif des caractéristiques (MDA8, nombre de dépassements de seuil, durée, intensité, sévérité) des typologies (A, B, C) de vagues de chaleur.....   | 176 |
| Tableau 13: Médianes de la durée, de l'intensité et de la sévérité des vagues de chaleur par année...   | 185 |
| Tableau 14: unité(s) statistique(s) utilisée(s) dans chacun des chapitres de la thèse. ....   | 197 |

## Introduction générale

L'ozone troposphérique ( $O_3$ ) est un polluant photo-oxydant qui est associé à de nombreux impacts sanitaires, environnementaux, et matériels. L'exposition à l' $O_3$  provoque une irritation des voies respiratoires et des muqueuses, des difficultés respiratoires et des dysfonctionnements du système cardiovasculaire (Bejot et al., 2011; Marx et al., 2019; Nuvolone et al., 2018; Zhang et al., 2019). Cette exposition peut également aggraver certaines pathologies comme l'asthme, l'emphysème, et les bronchites chroniques (bronchopneumopathie chronique obstructive). L' $O_3$  serait responsable annuellement, selon l'EEA (2016), de 16000 morts prématurées en Europe. Ces impacts sanitaires s'accompagnent d'impacts sur les écosystèmes. Les propriétés oxydantes de l' $O_3$  induisent chez les végétaux une nécrose foliaire (taches de couleur jaune), une chute prématurée des feuilles, une chlorose, une modification de l'ouverture des stomates<sup>1</sup> entraînant une réduction du taux d'activité photosynthétique et de la croissance (Dalstein et al., 2005; Eghdami et al., 2022; GREC-SUD, 2018). Des pertes conséquentes de rendement de la production agricole sont imputables à ce polluant (Feng et al., 2019). L'ozone, de par ses capacités oxydantes, engendre également une détérioration des matériaux des bâtiments et constitue un risque pour le patrimoine architectural (Screpanti and De Marco, 2009). Il existe donc des enjeux fort pour la gestion du risque lié à l'exposition à l' $O_3$ .

La distribution de ce polluant est très variable dans l'espace et dans le temps. En effet, les concentrations d' $O_3$  troposphérique sont fortement dépendantes des conditions climatiques (température, rayonnement solaire, humidité, vent) et de la concentration des polluants précurseurs de l' $O_3$  (oxydes d'azote-NOx, composés organiques volatiles-COV, monoxyde de carbone-CO). Il est difficile de prédire les concentrations de ce photo-oxydant, et par conséquent d'évaluer ses impacts. Les enjeux liés à la compréhension et à la gestion d'une telle pollution deviennent de plus en plus prégnants car les niveaux augmentent en milieu urbain depuis quelques décennies (Sicard, 2021), ces milieux concentrant une part majeure et croissante de la population humaine.

---

<sup>1</sup> Ensemble de deux cellules de garde riches en chlorophylle, ménageant entre elles une mince ouverture (ostiole) par laquelle s'effectuent les échanges gazeux de la plante pour les fonctions de respiration, transpiration et photosynthèse.

Dans le cadre d'une évaluation quantitative de l'impact sanitaire (EQIS) ou pour le suivi des niveaux d'exposition de la population, les stations de mesures de la qualité de l'air sont classiquement utilisées. A l'échelle d'une unité urbaine, la quantification des niveaux d'exposition de la population repose sur un nombre très restreint de stations de mesure, ne couvrant que très partiellement la diversité des territoires. Le développement de la modélisation et son exploitation sont donc des éléments prometteurs pour l'évaluation de l'exposition des populations. Des outils de modélisation de complexité et de résolution (spatiale et temporelle) très variables ont été développés (Ma et al., 2020). Néanmoins ces approches présentent des limites (Steven Porter et al., 2015). L'enjeu réside dans la production de données couvrant un continuum spatial et temporel tout en garantissant précision et exactitude.

Dans un contexte de changement climatique, la fréquence et l'intensité des vagues de chaleur, augmentent. Les simulations réalisées selon les scénarii d'évolution des émissions de gaz à effet de serre décrits par le GIEC (Representative Concentration Pathways (RCP)), prévoient une hausse des concentrations d'O<sub>3</sub>. L'évolution des émissions de précurseurs d'O<sub>3</sub> sera décisive dans l'évolution des concentrations d'O<sub>3</sub> au cours des vagues de chaleur (Meehl et al., 2018).

Le but de ce travail de recherche est de mieux évaluer l'exposition à l'O<sub>3</sub> troposphérique, sa variabilité spatiale et temporelle et enfin d'en comprendre les implications pour les populations exposées, notamment en période de vague de chaleur.

Quatre objectifs ont été définis, présentés dans les chapitres 3 à 6 :

- 1) Évaluer les performances d'un modèle transnational (PREV'EST) et l'optimiser pour une meilleure évaluation de l'exposition à l'O<sub>3</sub> à l'échelle kilométrique et journalière ;
- 2) Explorer la dynamique spatiale et temporelle de l'ozone à l'échelle journalière et kilométrique de 2008 à 2019, et en identifier les déterminants principaux ;

- 3) Evaluer l'exposition de la population au regard des seuils réglementaires et recommandations relatifs à l'ozone et explorer son hétérogénéité spatiale autour d'une ville de taille moyenne ;
- 4) Décrire les dimensions quantitative et qualitative des vagues de chaleurs observées sur une période récente de douze années et explorer leur relation avec les niveaux d'exposition de la population selon le degré d'urbanisation.

# Chapitre 1 : Ozone troposphérique

## Paramètres chimiques intervenant dans la formation de l'ozone troposphérique

### 1. Nature des polluants atmosphériques primaires intervenant dans la formation d'ozone troposphérique

#### 1.1. Oxydes d'azote

Le terme oxydes d'azotes (NOx) regroupent le monoxyde (NO) et le dioxyde d'azote (NO<sub>2</sub>). Les émissions de NOx proviennent essentiellement des secteurs d'activités anthropiques impliquant des processus de combustion à haute température. Des sources naturelles de NOx telles que les feux de forêts et les éclairs peuvent également être citées. Les NOx sont principalement rejetés sous forme de monoxyde d'azote soit par des sources « naturelles » (éclairs, émissions par les sols) soit par des sources anthropiques (libération d'azote présent dans les combustibles) (Camredon and Aumont, 2007; Lu et al., 2019). Ainsi, le trafic routier, les industries liées à la combustion de divers combustibles (combustibles liquides fossiles, charbon, gaz naturel, biomasses, gaz de procédés...) et des procédés industriels (fabrication de verre, métaux, ciment...) en représentent les principales sources. Au niveau édaphique<sup>2</sup>, les rejets de NO<sub>2</sub> sont quant à eux dus à l'activité microbienne (nitrification et dénitrification) et représentent 10 à 15 % des émissions de NOx (Lu et al., 2019).

#### 1.2. Composés organiques volatils

Les Composés Organiques Volatiles (COV) représentent une famille de composés chimiques d'une grande diversité et correspondant à plusieurs centaines de molécules. L'origine des COVs peut être aussi bien biogénique qu'anthropique. Ces molécules sont émises principalement dans les contextes suivants :

- phénomènes de combustion ;
- évaporation de solvants industriels (peintures, encres, colles, dégraissants, cosmétiques) ;
- évaporation des composés organiques tels que les carburants ;
- réactions biologiques.

---

<sup>2</sup> Relatif au sol.

Parmi les groupes de COV les plus présents en terme d'émission dans l'atmosphère, on peut citer les alcanes (hydrocarbures avec liaisons simples carbone-carbone), les alcènes (hydrocarbures avec une ou plusieurs liaisons doubles carbone-carbone), des composés aromatiques (comprenant au moins un noyau phényle) et des aldéhydes (comprenant un groupe carbonyle, HC=O)(Seigneur, 2018). Les COV sont majoritairement représentés par des hydrocarbures. Afin de préciser la terminologie, nous parlerons ici de COV pour définir les COV non-méthaniques (COVNM) car le méthane, bien que faisant partie des alcanes, n'est que peu réactif et possède une durée de vie relativement longue (10 ans). Ce composé n'est donc pas considéré comme un précurseur de pollution photochimique. Les substances cancérigènes tels que le benzène ou le formaldéhyde font partie des COV (Elichegaray et al., 2010) et font l'objet d'une réglementation plus stricte. Les hydrocarbures représentent une part importante des COV rejetés dans l'environnement, mais environ 90 % de ces molécules sont d'origine biogénique.

En dépit d'un nombre élevé de sources anthropiques, les COVs sont produits en proportion importante par les écosystèmes terrestres, et notamment les forêts (Lu et al., 2019; Weireld, 2020) et seuls 10% des COVs sont issus d'activités anthropiques (Martin, 2008). D'origine anthropique ou biogéniques, leur durée de vie est très variable selon l'espèce de COV considérée, allant de quelques heures à quelques mois (Martin, 2008).

### *1.3. Radicaux hydroperoxy et hydroxyle (HO<sub>x</sub>)*

Cette famille de composés chimiques regroupe les radicaux OH, HO<sub>2</sub> et les peroxyes organiques (RO<sub>2</sub>, où R représente une chaîne organique) (Camredon and Aumont, 2007; Lu et al., 2019). Les HO<sub>x</sub> sont d'une importance primordiale dans les réactions photochimiques de production d'ozone (Sadanaga et al., 2003). D'origine principalement photochimique (réactions de photolyse), leur rôle dans la production d'ozone se joue principalement en période diurne.

Si l'on considère individuellement les composants de cette famille, on constate que le radical OH aura davantage un impact sur la durée de vie des COV, tandis que les radicaux peroxyes (HO<sub>2</sub> et RO<sub>2</sub>) vont intervenir dans la conversion de NO en NO<sub>2</sub> (Figure 1).

Les HO<sub>x</sub> présentent une réactivité élevée, notamment si l'on considère OH (Seigneur, 2018). Leur durée de vie est par conséquent très courte (de l'ordre de la seconde) et leurs concentrations très faibles.

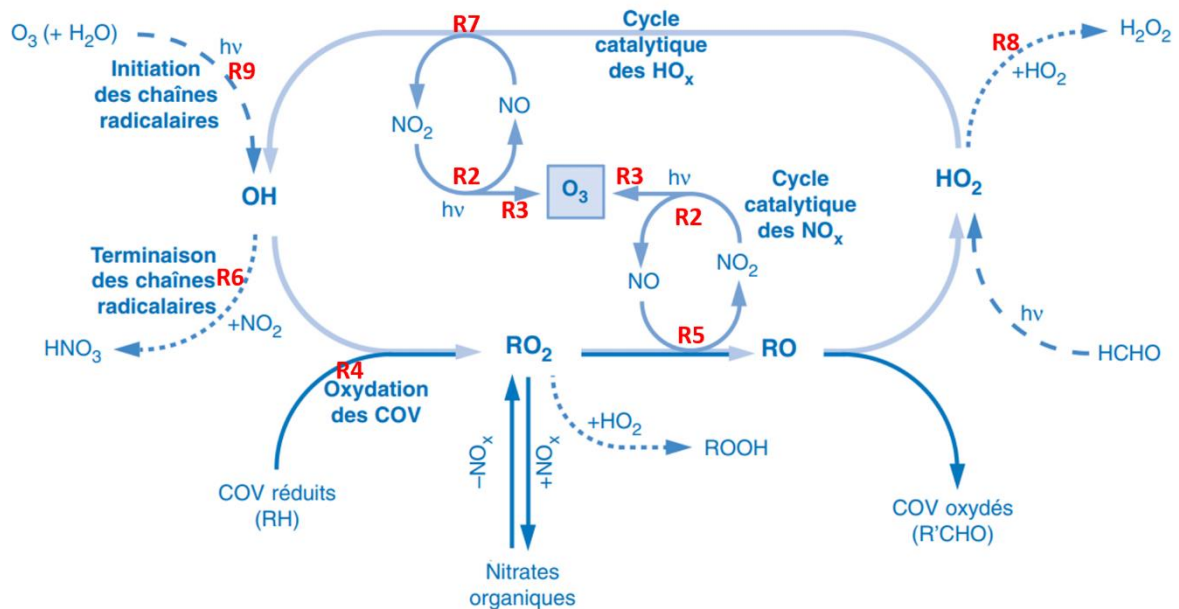
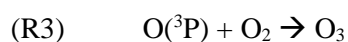
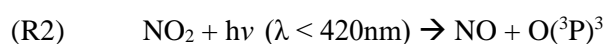
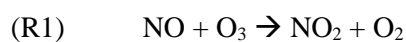


Figure 1: Représentation schématique des principaux mécanismes impliqués dans la formation d'ozone troposphérique (Camredon et Aumont, 2007). Les réactions notées R2 à R9 font référence aux réactions décrites dans le corps du texte.

## 2. Réactions photochimiques et régime de formation d'ozone

### 2.1. Equilibre photo-stationnaire

L'équilibre photo-stationnaire correspond à une boucle de réactions chimiques (Figure 2 A). Les réactions impliquées dans cette boucle sont décrites comme suit :



<sup>3</sup> atome d'oxygène standard.

De manière simplifiée, lorsque le système chimique se trouve à l'équilibre photo-stationnaire, la production d'O<sub>3</sub> (R3) est compensée par sa destruction (R1).

En présence d'oxydant alternatif, comme les COV, NO est converti en NO<sub>2</sub> sans consommation d'O<sub>3</sub> et engendre la production d'une molécule d'O<sub>3</sub> (Figure 2.B). L'équilibre photo-stationnaire est donc rompu et laisse place à une production nette et donc une accumulation d'O<sub>3</sub>. C'est cette perturbation du cycle par les COV qui est à l'origine de l'apparition de concentrations élevées d'O<sub>3</sub> et d'épisodes de smog photochimique.

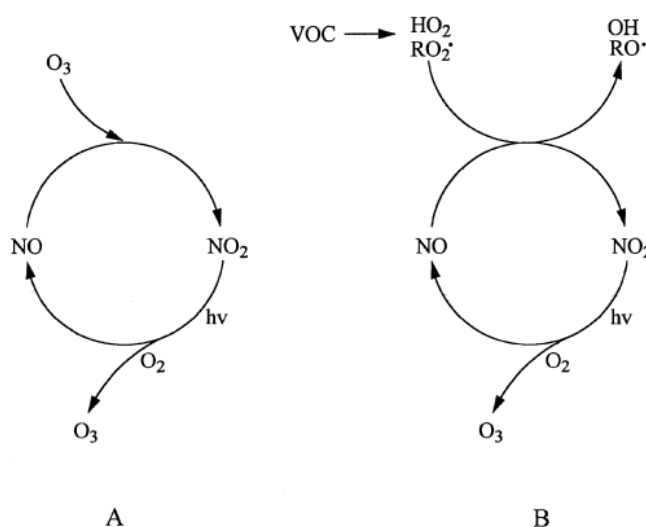


Figure 2: Représentation schématique des réactions impliquées dans la conversion de NO en NO<sub>2</sub> et la formation d'O<sub>3</sub>, d'une part (A) en situation d'équilibre photostationnaire et d'autre part (B) en présence de COV (d'après (Atkinson, 2000))

## 2.2. Régimes de production d'ozone troposphérique : dépendance au rapport [NOx]/[COV]

Parmi les précurseurs d'O<sub>3</sub>, les NO<sub>x</sub> sont ceux ayant le plus d'influence sur la formation d'O<sub>3</sub> (Martin, 2008). Le rapport [COV]/[NO<sub>x</sub>] détermine le type de régime de production d'O<sub>3</sub>. Ces régimes de formation de l'O<sub>3</sub> troposphérique sont décrits par le diagramme isopleth (Figure 3). Le diagramme isopleth (Figure 3) permet de représenter l'évolution des concentrations d'O<sub>3</sub> selon des courbes d'iso-concentrations et selon les concentrations respectives de COV et de NO<sub>x</sub>. Ce type d'outils est de



première importance dans l'orientation des mesures à appliquer pour une réduction efficace des concentrations d'O<sub>3</sub>.

Par convention, sont décrits deux régimes principaux de production d'O<sub>3</sub> troposphérique ainsi qu'un régime intermédiaire :

- **Régime saturé en NOx** (rapport COV/NOx < 4), les concentrations en NOx sont élevées et caractéristiques d'un milieu urbain, les niveaux d'O<sub>3</sub> augmentent lorsque les concentrations en COV augmentent ou lorsque les concentrations en NOx diminuent (jusqu'à un certain seuil) ;
- **Régime limité en NOx** (rapport COV/NOx > 15), ce régime est caractérisé par de faibles concentrations en NOx (représentatif du milieu rural), les concentrations d'O<sub>3</sub> sont sensibles à une hausse des concentrations de NOx mais peu sensibles à une augmentation des concentrations de COV ;
- **Régime intermédiaire**, l'influence des NOx et des COV sur les niveaux d'O<sub>3</sub> est équivalente, une diminution de l'un de ces précurseurs (ou des deux) engendre une diminution des concentrations d'O<sub>3</sub> (zone centrale du diagramme, Figure 3).

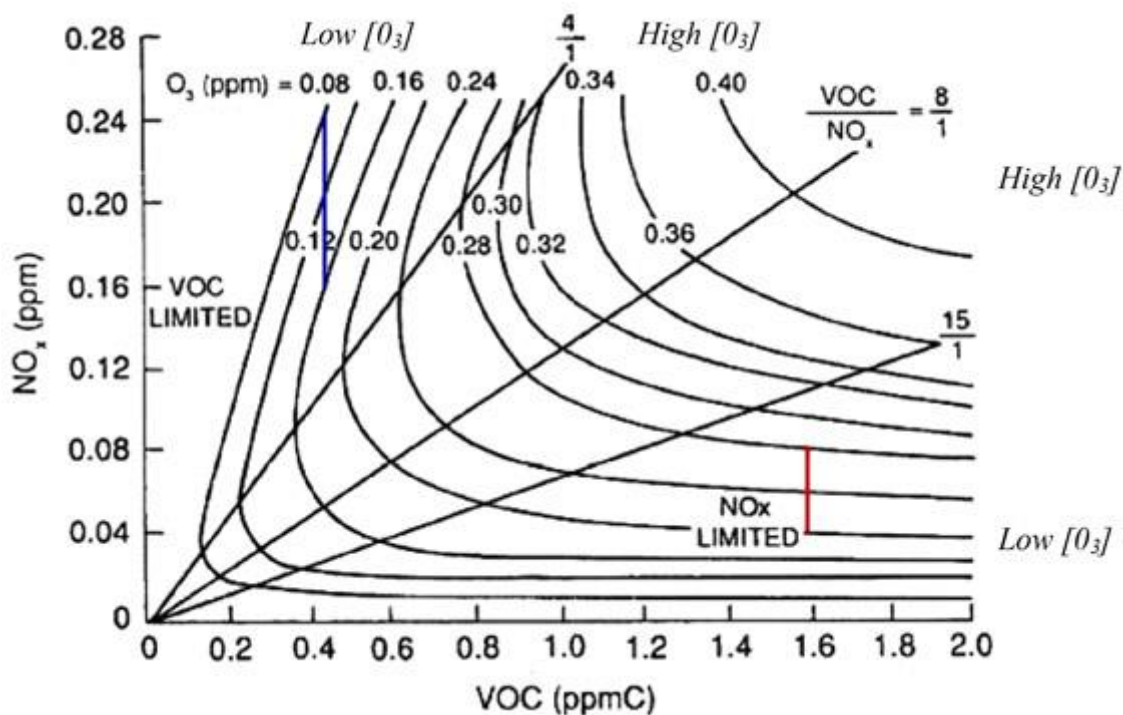
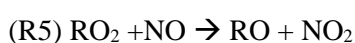
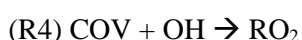


Figure 3 : Diagramme isopleth type adapté de Seinfeld et Pandis (1998).

### 2.3. Régime saturé en NOx

Au-delà d'un certain seuil de concentration en NOx (quelques ppb), on assiste à la régénération complète des radicaux OH consommés au cours de la réaction avec les COV et conduisant à la production de RO<sub>2</sub> (R4) et *in fine* d'ozone troposphérique (Figure 1).

Le système chimique de production de l'O<sub>3</sub> est alors fortement dépendant des réactions suivantes et de leur rapport cinétique :



Les cinétiques des réactions R4 et R6 et le rapport de leurs constantes cinétiques vont très fortement conditionner la production d'O<sub>3</sub>. L'augmentation du rapport des constantes cinétiques des réactions R4/R6 conduit à une production d'ozone car la réaction impliquant les COV (R4) est privilégiée. La réaction R5 est suivie par les réactions R2 et R3 conduisant ainsi à la production d'O<sub>3</sub> (Figure 1). L'efficacité de la production d'O<sub>3</sub> est dépendante de l'équilibre entre la propagation des radicaux libres dans la chaîne de réactions (R4) et les réactions de terminaison<sup>4</sup> de la chaîne de réactions (R5).

Ainsi, sous des conditions caractérisées par un régime de production d'O<sub>3</sub> dit « NOx-saturé », une augmentation des concentrations en COV conduira à une augmentation de la production d'O<sub>3</sub>. La production d'O<sub>3</sub> est limitée par l'intensité des réactions impliquant les COV (Wang et al., 2017).

### 2.4. Régime limité en NOx (ou saturé en COV)

Lorsque la production d'O<sub>3</sub> est limitée par les concentrations en NOx, deux types de régimes « limités par les NOx » co-existent selon les gammes de concentrations de NOx. Un basculement est observé soit en faveur de la production d'O<sub>3</sub> soit en faveur de sa consommation selon le seuil des quelques dizaines de ppb (*part per billion* – parties par milliard) (Figure 4). Ainsi, une concentration de l'ordre de 10 ppt

---

<sup>4</sup> Aussi appelées réactions puits. Réactions entraînant la disparition des intermédiaires responsables de la propagation de la réaction en chaîne. La terminaison de la chaîne des radicaux conduisant à la production d'O<sub>3</sub> (R6) est visible sur la Figure 8-1.

(part per trillion – parties par billion) de NO représente une gamme de concentration pivot entre production et consommation d’ozone (Liu et al., 1992; Seinfeld and Pandis, 2016) (Figure 4).

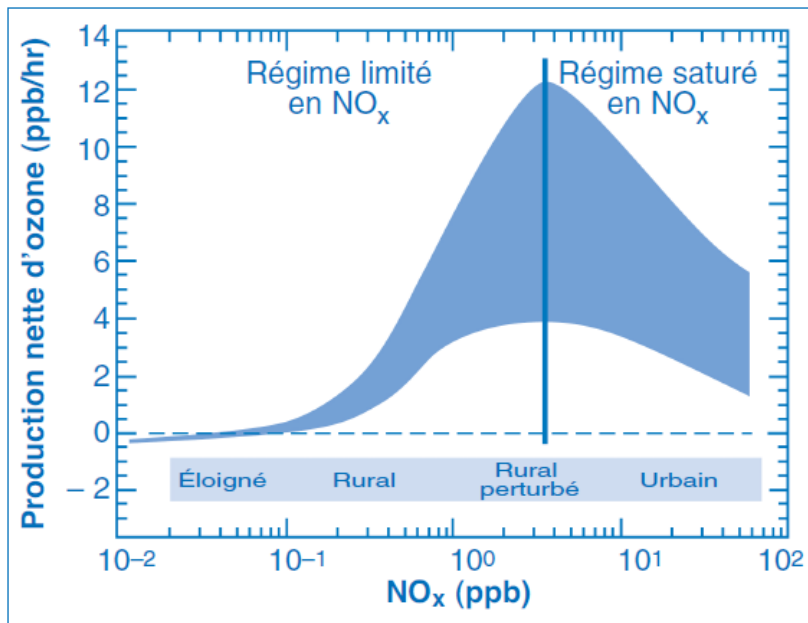


Figure 4 : Profil type de l'efficacité de production d'ozone en fonction de la concentration en NOx, tiré de Camredon et al., 2007, et adapté de (Liu et al., 1992; McKeen et al., 1991)

#### 2.4.1. Régime limité par les NOx : quelques dizaines de ppt < [NOx] < quelques ppb

Lorsque l'on se trouve dans la gamme de concentration inférieure à quelques ppb de NOx, la formation de radicaux peroxydes (RO<sub>2</sub>) est principalement dirigée par la réaction des COV avec des radicaux OH (réaction R4). Cependant les concentrations en NO sont alors insuffisantes pour contrôler complètement l'évolution chimique des radicaux peroxydes qui entrent alors en compétition pour réagir avec NO.

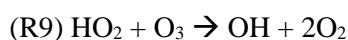
L'évolution du système chimique dépend alors de l'évolution des peroxydes, notamment de HO<sub>2</sub> selon les réactions suivantes :



L'augmentation des concentrations en NO conduit à l'augmentation de la production d'ozone via la réaction R7 suivie des réactions R2 et R3.

#### 2.4.2. Régime limité par les NOx : [NOx] < quelques ppt

Dans un environnement pauvre en NOx ([NOx] < quelques dizaines de ppt), l'O<sub>3</sub> est photolysé et n'est pas complètement régénéré. Le recyclage des radicaux peroxy est en partie pris en charge par O<sub>3</sub> lors de la conversion de HO<sub>2</sub> en OH :



Le bilan des réactions donne lieu à une consommation d'O<sub>3</sub>. Ce régime limité en NOx est notamment observé dans les zones éloignées des sources d'émissions primaires.

#### **Paramètres météorologiques intervenant dans les fluctuations des concentrations d'ozone troposphérique**

Les paramètres météorologiques et leurs variations spatiales et temporelles influencent grandement la formation, la destruction et la distribution de l'O<sub>3</sub>. Dans l'étude de Tarasova and Karpetchko (2003), (menée sur la péninsule de Kola, entre 1999 et 2000), 70 % des variations inter-journalières des niveaux d'O<sub>3</sub> étaient expliquées par les variations de température, de vitesse de vent et d'humidité relative.

Cette seconde partie est consacrée à la description de l'influence de différents paramètres (température, rayonnement solaire, régimes de vent, humidité relative) et phénomènes météorologiques (vagues de chaleur) sur les concentrations d'O<sub>3</sub>. Cette partie se termine par une synthèse des actions conjuguées de chacun de ces paramètres sur les concentrations d'O<sub>3</sub>.

##### 1. Température et rayonnement solaire

L'ozone troposphérique est un polluant secondaire dit photochimique. Sa formation est donc fortement tributaire de l'ensoleillement et donc indirectement de la température. La hausse des températures peut engendrer une hausse des émissions de NOx par le compartiment édaphique, et de COV (par la végétation, ou par évaporation dans le secteur industriel)(Nguyen et al., 2022). La température possède un effet catalyseur sur les réactions de formation de l'O<sub>3</sub> (Houzé, 2006). Les concentrations d'O<sub>3</sub> ambiant ont une saisonnalité marquée ainsi qu'un cycle nyctéméral. Les concentrations les plus élevées sont, comme attendu, observées au printemps et en été à l'échelle annuelle, et en début-milieu d'après-midi à l'échelle journalière (Notario et al., 2013). La variabilité de l'ensoleillement conditionne cette

variabilité temporelle. Les conditions anticycloniques combinant chaleur et stagnation des masses d'air sont généralement associées à des pics d'ozone (causés par une augmentation rapide et intense des concentrations d'O<sub>3</sub>, due à son accumulation dans la basse atmosphère). Les anticyclones sont des systèmes à haute pression atmosphérique associés à des conditions stables, un mélange convectif des masses d'air faible, et une couverture nuageuse faible (favorisant l'ensoleillement et les réactions photochimiques)(Nguyen et al., 2022). Ces phénomènes favorisent donc la formation et l'accumulation de l'O<sub>3</sub> et sont donc associés à des niveaux élevés de ce polluant. La fréquence grandissante de ces épisodes anticycloniques a été mise en relation avec le changement climatique (Nguyen et al., 2022), tout comme les phénomènes de vagues de chaleur.

## 2. Vagues de chaleur

Les phénomènes de vagues de chaleur sont caractérisés par des températures et un ensoleillement élevés favorisant la formation d'O<sub>3</sub>. Météo France définit ces phénomènes selon une méthodologie d'identification et de caractérisation des vagues de chaleur, basée sur des indicateurs thermiques sur une période de référence (1981-2010)(Soubeyroux et al., 2016). Les indicateurs thermiques correspondent aux percentiles 95 (seuil d'interruption, nommé Sint), 97,5 (seuil de début de vague, nommé Sdeb) et 99,5 (seuil de détection, nommé Spic) de l'indicateur thermique sur la période de référence (1981-2010). Une vague de chaleur est identifiée si le seuil Spic est dépassé. Le début de vague de chaleur est défini par le dépassement du seuil Sdeb, et seules les vagues durant au moins trois jours sont prises en compte. La vague est interrompue si la température descend en dessous du seuil Sint (même une seule journée), ou en dessous du seuil Sdeb pendant au moins 3 jours consécutifs (Ouzeau et al., 2016; Soubeyroux et al., 2016). Chaque vague identifiée est ensuite caractérisée par trois indicateurs : durée, intensité et sévérité (Figure 5). L'intensité correspond à la température moyenne journalière maximale observée au cours de la vague. La sévérité est « la somme des écarts positifs entre températures moyennes quotidiennes pendant la vague de chaleur et le seuil de démarrage Sdeb divisée par la différence entre Spic et Sdeb » (Ouzeau et al., 2016).

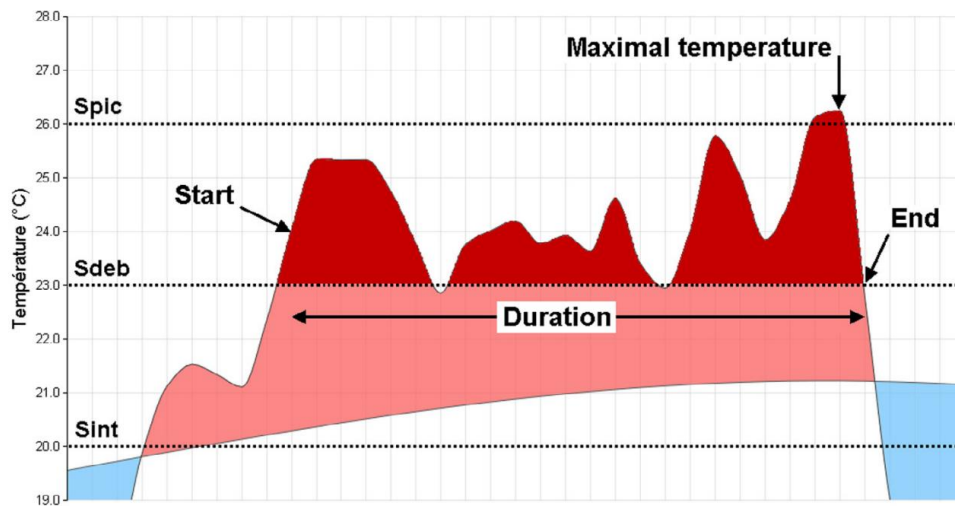


Figure 5: Caractérisation d'une vague de chaleur basée sur la température moyenne journalière en France (exemple de série temporelle du 30 juin au 5 août 1983) : durée (start – end), température maximale et intensité globale (aire rouge sur le graphique). Les températures supérieures (inférieures) à la ligne climatologique (période de référence 1981-2010) sont représentées par une zone rose (bleue). D'après Ouzeau et al. (2016).

Cette méthodologie permet d'avoir un suivi de ces phénomènes et de leur évolution en termes de fréquence, de durée, de sévérité et d'intensité depuis 1947 sur le territoire national. La fréquence des vagues de chaleur sur le territoire national a très nettement augmentée depuis 1947. En effet, au cours des vingt dernières années, les phénomènes observés ont été plus nombreux qu'au cours des cinquante années précédentes (1947-2000) (Figure 6). De plus, l'intensité de ces événements tend à croître.

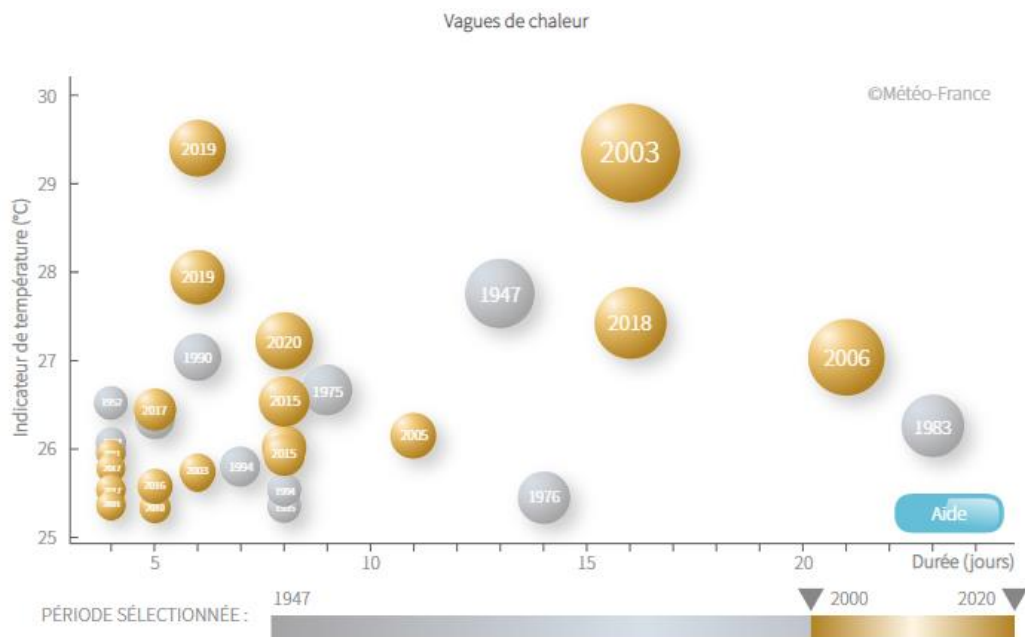


Figure 6: Inventaire et caractérisation des vagues de chaleur (intensité en ordonnées, durée en abscisses, et surface du cercle représentant la sévérité de l'événement) de 1947 à 2020. Les vagues de chaleur des vingt dernières années sont en surbrillance. Illustration produite par Météo France (outil en ligne Climat HD, <https://meteofrance.com/climathd>).

Les vagues de chaleur font partie des extrêmes climatiques auxquels les populations, les organismes, et plus généralement les écosystèmes seront de plus en plus exposés. Le changement climatique déjà amorcé devrait augmenter la fréquence et l'intensité de ces phénomènes (Intergovernmental Panel on Climate Change (IPCC), 2013). Les simulations issues du scénario RCP 8.5<sup>5</sup> développé par l'IPCC prévoient une amplification des vagues de chaleur à l'horizon 2071-2100 tant en termes de durée que d'intensité et de sévérité (Figure 7). Les concentrations d'O<sub>3</sub> lors de ces phénomènes d'ampleur croissante devraient augmenter.

<sup>5</sup> Les scénarii RCP (Representative Concentration Pathways), sont des scénarii d'évolution de concentration de gaz à effet de serre développé par l'IPCC. Le scénario RCP 8.5 correspond au scénario le plus pessimiste, avec une concentration de gaz à effet de serre (ppm) supérieure à 1370 eq-CO2 et un forçage radiatif supérieur à 8,5 W.m<sup>-2</sup> en 2100.

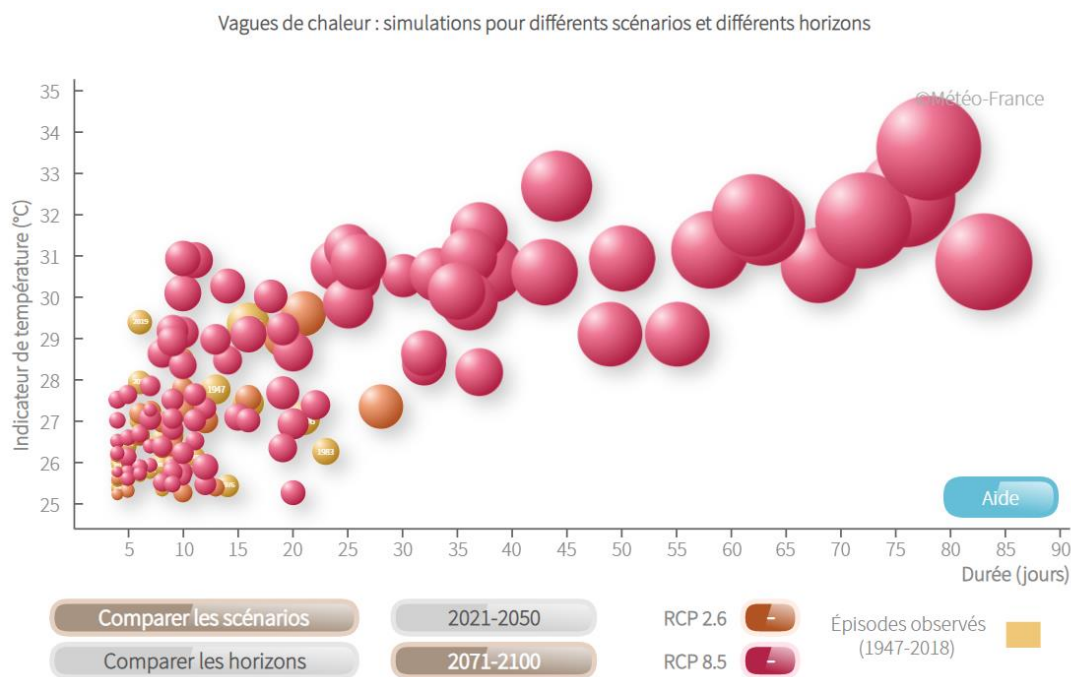


Figure 7: Projections des phénomènes de vagues de chaleur sur la période 2071-2100 selon les scénarios RCP2.6 et RCP8.5, illustration produite par Météo France (outil en ligne Climat HD).

### 3. Régimes de vent

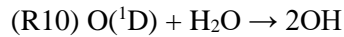
La vitesse et la direction du vent affectent la dispersion, le transport et l'accumulation des précurseurs de l'O<sub>3</sub> et donc la formation de ce composé. Les régimes de vents sont soumis à des variations saisonnières. Ces vents peuvent influencer la formation et l'accumulation d'O<sub>3</sub> des zones sous le vent d'une agglomération. Une zone au vent d'une agglomération (en l'absence d'autres sources de précurseurs) est soumise à des concentrations de précurseurs moindres et donc d'O<sub>3</sub> moins élevées qu'une zone sous le vent de l'agglomération (Nguyen et al., 2022). Les régimes de vents locaux ont donc une influence non négligeable sur les concentrations d'ozone à l'échelle locale, mais dépendent de la direction et de l'intensité des vents tout comme des émissions locales de précurseurs. Les vents synoptiques<sup>6</sup> sont quant à eux responsables de transports à longue distance qui contribuent également à l'apport de précurseurs dans des zones éloignées des sources primaires.

<sup>6</sup> Phénomènes atmosphériques dont l'ordre de grandeur est de quelques milliers de kilomètres pour les dimensions horizontales, de quelques kilomètres pour la dimension verticale et de quelques jours pour la durée ; l'échelle spatio-temporelle ainsi décrite s'appelle précisément l'échelle synoptique. (Définition tirée du glossaire de Météo France)



#### 4. Humidité relative de l'air

L'humidité relative de l'air est associée à la destruction d'O<sub>3</sub>. Cette variable est donc négativement corrélée à la concentration d'O<sub>3</sub>. La vapeur d'eau joue un rôle décisif dans la production d'oxydant OH. En effet, l'atome d'oxygène à l'état excité (O(<sup>1</sup>D)) (issu de la photolyse du NO<sub>2</sub>) réagit avec la vapeur d'eau.



Température et humidité relative modulent la quantité de vapeur d'eau présente dans l'air et par conséquent la disponibilité des molécules d'eau pouvant réagir avec O(<sup>1</sup>D) et régénérer OH. Cette voie réactionnelle est considérée comme un puit (une perte) d'O<sub>3</sub> car la production de radical OH permet d'activer des réactions de terminaison (R9) (Figure 1).

Toutefois, en milieu urbain et en présence de concentrations élevées en NO<sub>x</sub>, le radical OH peut participer à l'oxydation des COV (R4) (Figure 1) et ainsi initier la production d'O<sub>3</sub> (Lu et al., 2019).

#### 5. Action conjuguée des différents paramètres météorologiques

L'influence des paramètres météorologiques sur les concentrations d'ozone troposphériques sont résumées par la Figure 8 (issue de Nguyen et al., 2022). Les différents paramètres agissent simultanément sur les niveaux d'O<sub>3</sub> et sont eux-mêmes corrélés entre eux. Chaque variable impliquée dans ce système présente une variabilité spatiale et temporelle. Ce système d'interaction entre paramètres météorologiques et concentrations d'O<sub>3</sub> est donc complexe car multifactoriel et variable dans l'espace et dans le temps. Par exemple, l'effet du vent varie spatialement. La position sous-le-vent ou au vent d'une source de précurseurs sera respectivement associée à des concentrations plus ou moins élevée (Figure 8).

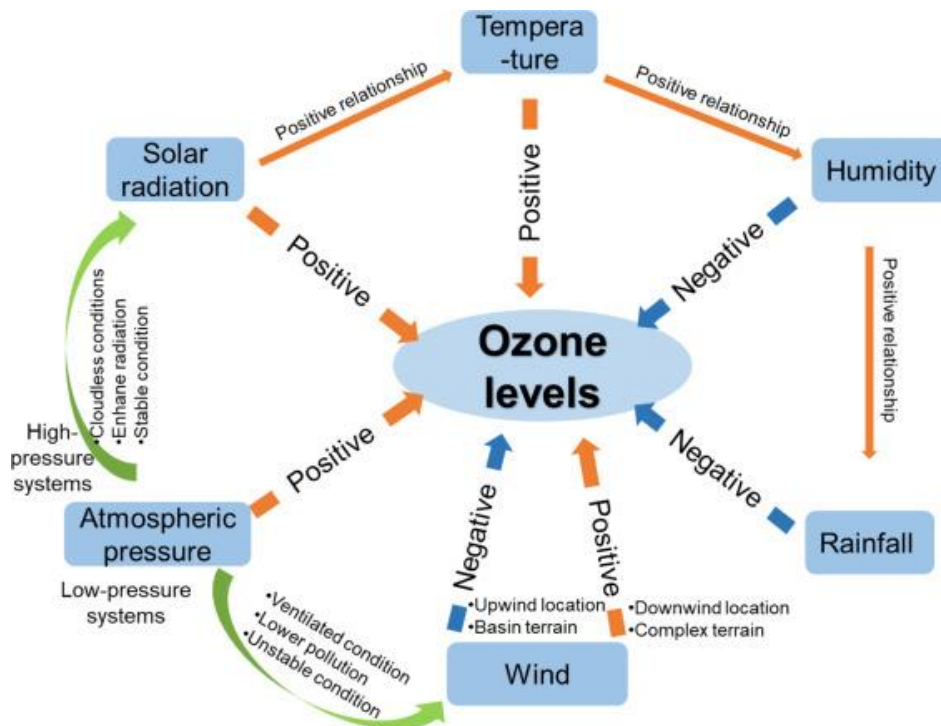


Figure 8: Influence des paramètres météorologiques sur les concentrations d’ozone troposphériques (Nguyen et al., 2022).

### Paramètres physiques et biologiques de variations des concentrations d’ozone troposphérique

Les caractéristiques locales telles que la topographie et la localisation des sources aussi bien anthropiques que biogéniques conditionnent la distribution spatiale de l’O<sub>3</sub>. Dans cette troisième partie, un focus sera fait sur i) l’influence de la topographie sur les concentrations d’O<sub>3</sub> et leur variabilité et ii) l’influence de la végétation sur les concentrations d’O<sub>3</sub>.

#### 1. Topographie

La topographie est un des paramètres clés influençant les concentrations d’O<sub>3</sub> dans la basse atmosphère ainsi que sa variabilité, notamment à l’échelle journalière (Chevalier et al., 2007). Le profil topographique d’une zone influence l’écoulement de l’air, sa circulation et par conséquent le transport des masses d’air et des potentiels polluants qu’elles contiennent. Une agglomération encaissée entre plusieurs massifs montagneux (exemple de la ville de Grenoble) sera sujette à une dispersion moindre des polluants atmosphériques et donc à leur accumulation. Le comportement de l’O<sub>3</sub> est fortement dépendant de la topographie locale qui va contraindre les écoulements d’air (Martin, 2008).

La dynamique, notamment journalière, de l'O<sub>3</sub> est influencée par le comportement de la couche limite atmosphérique (CLA), elle-même fortement dépendante de la topographie locale (Chevalier et al., 2007). La CLA est définie comme étant « la couche de l'atmosphère influencée directement par le sol où les caractéristiques (rugosité, capacité calorifique massique<sup>7</sup>, albédo<sup>8</sup>, teneur en eau, pente...) régissent localement les échanges des différents flux entre le sol et l'atmosphère » (Blein, 2016). La hauteur de la CLA est en moyenne de 1km (Brancher et al., 2017) mais dépend fortement des variations de températures. Au cours de la journée cette hauteur est plus grande car le rayonnement solaire et la température induisent une convection et une élévation des masses d'air. Au cours de la nuit son épaisseur est plus faible (Stull, 1988)(Figure 9).

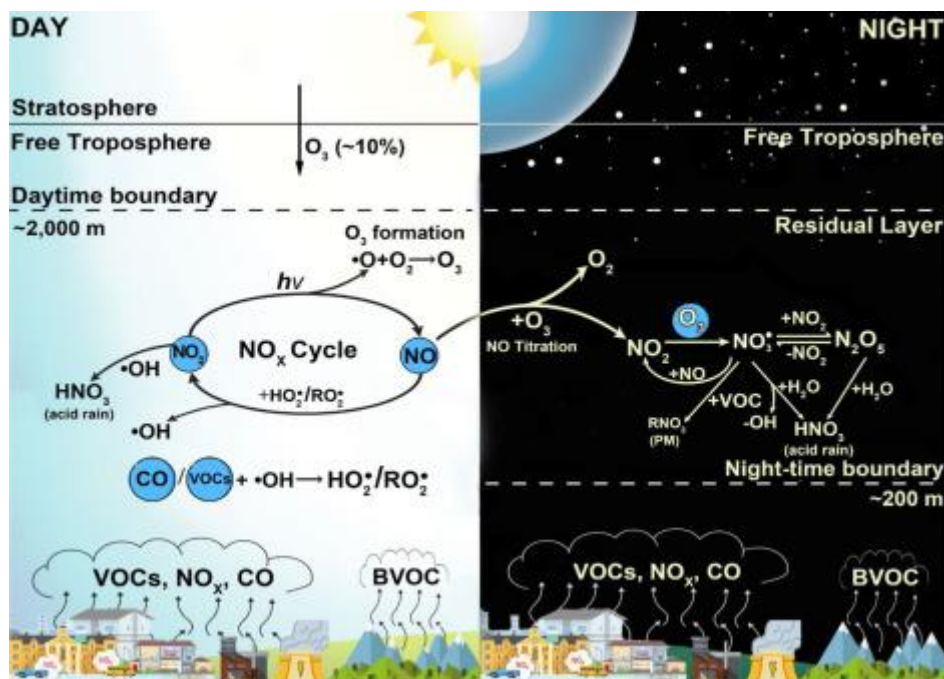


Figure 9: Illustration de la chimie atmosphérique de l'O<sub>3</sub> et des NO<sub>x</sub>, ainsi que de la structure de la CLA au cours de la journée et de la nuit (Nguyen et al., 2022).

Les zones d'altitude, comme les sommets et les cols, se situent généralement dans la couche résiduelle au cours de la nuit (Martin, 2008) (Figure 9). La chute des température au cours de la nuit

<sup>7</sup> Quantité de chaleur qu'un gramme d'une substance doit absorber pour que sa température s'élève d'un degré Celsius ou quantité de chaleur qu'un gramme de cette substance doit perdre pour que sa température baisse d'un degré Celsius.

<sup>8</sup> Fraction de l'énergie de rayonnement incidente qui est réfléchiée ou diffusée par un corps, une surface ou un milieu. De manière générale, l'albédo correspond à la part des rayonnements solaires qui sont renvoyés vers l'atmosphère

permet un brassage important des masses d'air dans la couche résiduelle, ce qui engendre un apport relativement constant d'O<sub>3</sub> (Martin, 2008). Les zones d'altitude sont donc soumises à un brassage des masses d'air plus intense en comparaison à une zone de vallée. L'amplitude des variations des concentrations d'O<sub>3</sub> à l'échelle journalière est donc plus faible en altitude que dans la vallée.

La stabilité des masses d'air dans la vallée permet un dépôt sec et donc une destruction de l'O<sub>3</sub> plus élevée au cours de la phase nocturne. La proximité au trafic, plus importante en vallée qu'en altitude, agit comme source de NO<sub>x</sub> et notamment de NO, et permet la destruction de l'O<sub>3</sub> via la réaction de titrage de l'O<sub>3</sub> par NO (R1) (Gheusi et al., 2011). Au cours de la phase diurne, le phénomène de convection et la disponibilité des précurseurs engendrent un pic journalier qui coïncident avec le maximum d'ensoleillement. L'amplitude journalière des mesures horaires des concentrations d'O<sub>3</sub> réalisées en fond de vallée est donc plus grande.

Les concentrations d'ozone plus élevées sur les reliefs peuvent être dues au rayonnement solaire plus intense en altitude qui stimule donc davantage les réactions photochimiques liées à la production d'O<sub>3</sub>. Les échanges entre troposphère libre et couche limite atmosphérique sont plus intenses à haute altitude. Les émissions de COV biogéniques (COVB) sont également plus élevés en altitude (Steinbacher et al., 2004) car la végétation est présente de manière plus dense sur les reliefs que dans les plaines, souvent occupées par les activités anthropiques. L'ensemble des phénomènes décrits dans cette sous-partie participe à une augmentation des concentrations d'O<sub>3</sub> le long du gradient altitudinal.

## 2. Influence de la végétation sur les concentrations d'O<sub>3</sub> troposphérique

La végétation représente une source biogénique de COV. Néanmoins, l'influence nette de la végétation sur les concentrations d'O<sub>3</sub> troposphérique reste complexe à établir car de nombreux phénomènes aux actions parfois antagonistes agissent simultanément sur la production ou la destruction d'O<sub>3</sub>. L'influence de la végétation sera donc décrite dans cette partie à travers les divers processus impliqués et leur évolution dans un contexte de changement climatique.

La végétation, et en particulier les couverts forestiers, peuvent favoriser la production d'O<sub>3</sub> via l'émission de COVB (Paoletti, 2009). L'intensité de l'émission de ces composés est plus importante

chez les résineux, mais se trouve être espèce-dépendante (Fitzky et al., 2019; Martin, 2008; Sicard et al., 2018).

Les COVBs les plus abondants sont les isoprènes et les monoterpènes (Sindelarova et al., 2014). Les terpènes représentent une part majeure des COVB émis par les résineux (Martin, 2008). Le couvert forestier peut émettre une quinzaine de terpènes différents. Les résineux possèdent un pouvoir d'émission de ces composés en moyenne plus élevé que celui des feuillus. Ces émissions augmentent lors des périodes de sécheresse et de manière plus générale avec l'intensification de l'insolation, la température et la floraison (Calfapietra et al., 2013; Guenther et al., 1993). Ainsi, des concentrations de terpènes habituellement de l'ordre de 70 ppb dans une forêt de pins, peuvent s'élever à 200 voire 800 ppb sous l'influence de fortes chaleurs (Martin, 2008). Ces COVBs agissent sur la production d'O<sub>3</sub> via le cycle présenté en Figure 1. En effet, ces COVB peuvent être oxydés et alimenter le cycle catalytique des NO<sub>x</sub> et ainsi produire de l'O<sub>3</sub> par photolyse du NO<sub>2</sub> (réactions R2 et R3).

Par ailleurs, la végétation peut également être considérée comme un facteur de réduction des niveaux d'O<sub>3</sub> via divers processus. L'ozone peut être détruit par contact sur les feuilles des arbres ou par absorption via leurs stomates (Fitzky et al., 2019). L'ombrage créé par les arbres contribue à tempérer la chaleur qui peut s'installer dans les agglomérations et freine ainsi la formation d'O<sub>3</sub>, dépendante de l'ensoleillement et de la température (Houzé, 2006; Yli-Pelkonen et al., 2018).

L'effet des arbres sur les concentrations d'O<sub>3</sub> est donc dépendant de nombreux facteurs dont : l'espèce d'arbre considérée, son statut physiologique, des facteurs environnementaux dictant les émissions de précurseurs (conditions climatiques, régime de production d'O<sub>3</sub> (niveaux respectifs de COV et de NO<sub>x</sub>)), selon Fitzky et al. (2019). Le bilan net du rôle de la végétation, en particulier en milieu urbain, sur la régulation des niveaux d'O<sub>3</sub> est donc très complexe à établir.

Dans un futur proche et un contexte de changement climatique, l'influence des COVB sur la production d'O<sub>3</sub> devrait s'intensifier. Dans un cadre de gestion de la qualité de l'air et de protection des populations, les politiques publiques tendent à abaisser progressivement les seuils réglementaires d'information et d'alertes des polluants atmosphériques. Les émissions issues du trafic routier, et particulièrement les

NO<sub>x</sub>, font l'objet de mesures de réduction. La diminution des niveaux de NO<sub>x</sub> est susceptible de rendre le régime de production d'O<sub>3</sub> d'autant plus sensible à l'émission de COVB. De plus, dans un contexte de changement climatique déjà amorcé, les phénomènes de vagues de chaleur se feront de plus en plus fréquents et intenses. Ces phénomènes, caractérisés par des températures extrêmes, sont souvent associés à des périodes de sécheresse. Or, les émissions de COVB augmentent avec la hausse des températures et de l'ensoleillement, et les conditions de sécheresse (Baghi, 2013; Peñuelas and Staudt, 2010; Trenberth et al., 2014). Les émissions de COVB sont donc amenées à augmenter à moyen terme, ce qui favorisera une augmentation des concentrations d'O<sub>3</sub> troposphérique.

### **Dynamiques spatiales de l'ozone troposphérique**

#### **1. Facteurs de variation spatiale**

Les variations spatiales des concentrations d'O<sub>3</sub> sont dépendantes des variations spatiales des facteurs influençant la formation, la destruction et le transport de l'O<sub>3</sub> que sont : la topographie (section 1), les conditions climatiques (section 0), la position ou non sous le vent d'une source majeure de précurseurs (même à longue distance) (section 3), l'occupation du sol (degré d'urbanisation, type d'activités anthropiques, la proportion et le type de végétation), le régime de production d'O<sub>3</sub> (section 2.2). L'ensemble des facteurs énoncés précédemment influence simultanément la formation d'O<sub>3</sub> et les concentrations observées. La variabilité spatiale de chacun de ces déterminants contribue à la complexité de la dynamique spatiale de l'O<sub>3</sub>. Les déterminants de l'O<sub>3</sub> peuvent avoir des actions antagonistes ou synergiques qui complexifient la relation entre la concentration d'O<sub>3</sub> observée et chacun des déterminants. Ces phénomènes d'actions antagonistes ou synergiques sont notamment observés entre facteurs météorologiques (section 5).

#### **2. Distribution de l'O<sub>3</sub> et régimes de production d'O<sub>3</sub> en fonction du degré d'urbanisation**

Les régimes de production de l'O<sub>3</sub> décrits en section 2.2 sont spatialement structurés et sont à l'origine des variations spatiales des concentrations d'O<sub>3</sub> (du centre urbain, aux zones périurbaines puis rurales et montagneuses). Cette section traitera des dynamiques spatiales des concentrations d'O<sub>3</sub> mises en relation avec les variations spatiales des régimes de production d'O<sub>3</sub>.

Un pôle urbain est défini selon l'INSEE comme « une unité urbaine<sup>9</sup> offrant au moins 10 000 emplois et qui n'est pas située dans la couronne d'un autre pôle urbain ». Ces pôles, de tailles et de densité très diverses, sont à l'origine d'émissions de précurseurs d'intensité variable. Au sein de ces pôles, les sources d'émissions de NO<sub>x</sub>, via le trafic routier, sont nombreuses et intenses. Les concentrations en NO<sub>x</sub> y sont par conséquent élevées et conduisent à un rapport [COV]/[NO<sub>x</sub>] faible. En conditions saturées en NO<sub>x</sub>, la production d'ozone est peu efficace. Les radicaux OH réagissent avec NO<sub>2</sub> (R6). L'équilibre entre la propagation des radicaux libres dans la chaîne de réactions (R4) et les réactions de terminaison de la chaîne de réactions (R5) tend à être dominé par les réactions de terminaison (Figure 1). Le titrage de l'O<sub>3</sub> par NO (R1) participe également à la modération des concentrations en O<sub>3</sub> dans le milieu urbain (Seigneur, 2018). Ainsi, les teneurs en O<sub>3</sub> en agglomération ne dépassent que ponctuellement la centaine de ppb (soit environ 200 μg.m<sup>-3</sup>) (Camredon and Aumont, 2007).

Avec l'éloignement du centre urbain, une diminution rapide des concentrations de NO<sub>x</sub> est observée. Celle-ci est due à l'éloignement des sources majeures d'émission, à leur dilution dans les masses d'air et à leur temps de vie limité (Kaiser et al., 2007) (en moyenne de quelques heures à quelques jours). L'augmentation des concentrations en COV avec l'éloignement du centre urbain est due à leur temps de vie plus long (que les NO<sub>x</sub>), bien que variable, et à des émissions biogéniques plus conséquentes. Par conséquent, le rapport [COV]/[NO<sub>x</sub>] augmente avec l'éloignement des sources d'émission de NO<sub>x</sub>, et conduit à l'émergence d'un régime limité en NO<sub>x</sub>. La production d'O<sub>3</sub> est donc accélérée quelques dizaines de kilomètres sous le vent des pôles urbains (Camredon and Aumont, 2007). Ainsi, les zones rurales ou forestières sous l'influence de pôles urbains présentent les maximas de concentrations d'O<sub>3</sub> (Figure 4) en raison des réactions entre polluants primaires ayant lieu progressivement au cours de leur advection (Martin, 2008). Les réactions photochimiques à l'origine de la formation d'O<sub>3</sub> nécessitent un temps de latence, tout comme le transport des précurseurs. Ceci génère un décalage temporel et spatial entre l'émission des précurseurs et l'apparition du polluant secondaire, l'O<sub>3</sub> (Reche et al., 2018; Seigneur, 2018).

---

<sup>9</sup> une commune ou un ensemble de communes présentant une zone de bâti continu (pas de coupure de plus de 200 mètres entre deux constructions) qui compte au moins 2 000 habitants. ([INSEE, 2020](#))

## **Dynamiques temporelles de l'ozone troposphérique**

L'ozone troposphérique est caractérisé par des variations temporelles journalières, hebdomadaires (week day/weekend effect), saisonnières et à moyen terme (décennie) (Royal Society (Great Britain), 2008; Zhao et al., 2019). Ces variations à différentes échelles temporelles et leurs origines sont décrites de l'échelle la plus fine à l'échelle la plus large dans les sous-parties 1 à 3. La sous-partie 4 est consacrée à l'évolution des concentrations d'O<sub>3</sub> au cours des dernières décennies et à leur projection à moyen terme.

### **1. Cycle nyctéméral**

A l'échelle journalière, une phase diurne et une phase nocturne peuvent être décrites. Pendant la phase diurne, les concentrations d'O<sub>3</sub> sont principalement liées à l'ensoleillement et aux émissions de précurseurs, en particulier les NO<sub>x</sub>. Les NO<sub>x</sub> sont majoritairement émis aux heures de pointes et coïncident avec la mobilité des travailleurs (en début de matinée et en fin d'après-midi) (Banja et al., 2009). La diminution des concentrations en NO<sub>x</sub> en journée est attribuée à la forte diminution des émissions (par le trafic routier) et à une photolyse plus intense en milieu de journée (Seigneur, 2018).

La photolyse de NO<sub>2</sub> étant maximale au zénith, un pic d'O<sub>3</sub> est observé en milieu de journée avec un léger décalage temporel entre le maximum de rayonnements solaires et le maximum de production d'O<sub>3</sub> (Notario et al., 2013). Les concentrations journalières les plus élevées sont en général observées entre le début et le milieu d'après-midi, comme l'illustre la Figure 10. La production de radicaux (OH et peroxy) et d'O<sub>3</sub> par photolyse diminuent progressivement avec la diminution de l'intensité des rayonnements solaires, et finissent par être dominées par le titrage de l'O<sub>3</sub> par NO.



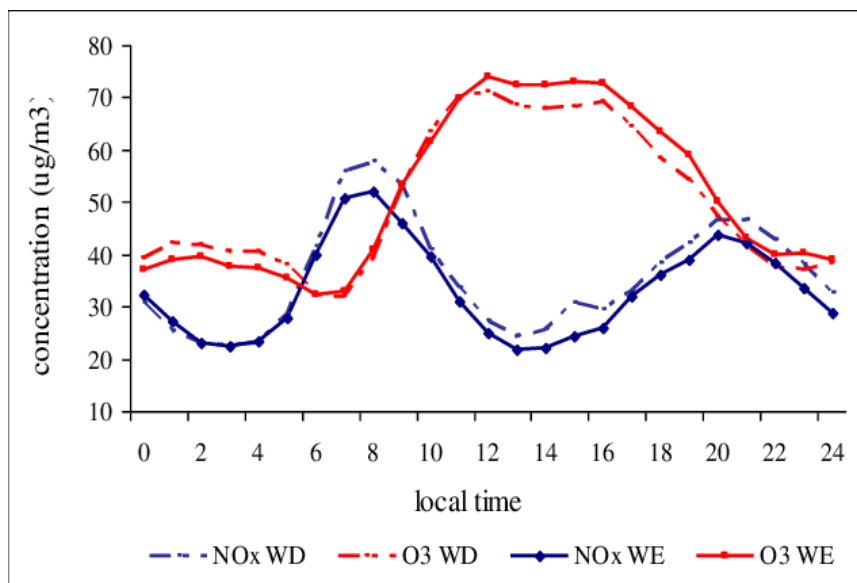


Figure 10: Concentrations horaires d'O<sub>3</sub> et de NO<sub>x</sub> à l'échelle journalière au niveau d'un site périurbain en 2007, pendant la semaine (WD) et le weekend (WE) (Banja et al., 2009).

Au cours de la phase nocturne, la production d'O<sub>3</sub> est progressivement stoppée en absence de rayonnements solaires. La température et la hauteur de la couche limite diminuent. L'O<sub>3</sub> est détruit lorsqu'il réagit au contact des surfaces solides et liquides. On qualifie ces phénomènes de dépôts secs et humides. Cette destruction est le fruit de la décomposition thermodynamique de l'O<sub>3</sub> ( $2O_3 \rightarrow 3O_2$ ) catalysée par les surfaces solides avec lesquelles ce composé entre en contact (Martin, 2008). La vitesse de ces dépôts dépend de la nature de la surface et de l'intensité des turbulences, celles-ci favorisant la rencontre entre les molécules d'O<sub>3</sub> et les surfaces sur lesquelles les molécules peuvent se déposer. Plus une surface sera poreuse, plus le dépôt sera efficace. La période nocturne est particulièrement propice à la formation de rosée et à la conservation d'une certaine humidité ambiante et sur la végétation. L'O<sub>3</sub> est donc également détruit par dépôt humide.

Les zones rurales sont soumises à un cycle nyctéméral présentant une amplitude des concentrations d'O<sub>3</sub> plus élevée qu'en zone urbaine. D'une part, ceci est dû au transport des précurseurs et de l'O<sub>3</sub> des zones urbaines aux zones rurales, engendrant une augmentation progressive des teneurs en O<sub>3</sub>. D'autre part, la nuit est plus fraîche et plus humide qu'en aire urbaine, contribuant à abaisser les concentrations d'O<sub>3</sub> durant la période nocturne. Les niveaux d'O<sub>3</sub> atteignent un minimum en fin de nuit ou très tôt le matin (5h-6h du matin) (Akimoto and Tanimoto, 2022).

## 2. Variations hebdomadaires: l'effet weekend

L'effet weekend correspond à une variation des concentrations d'O<sub>3</sub> entre les jours travaillés de la semaine et le weekend. De manière générale, les concentrations d'O<sub>3</sub> sont plus élevées au cours du weekend que de la semaine. L'amplitude de l'effet weekend est plus grande dans les centres urbains qu'en zone rurale (Zhao et al., 2019) et sous des latitudes où l'ensoleillement est plus élevé.

Ce phénomène peut avoir plusieurs causes (Castell-Balaguer et al., 2012). Tout d'abord, les émissions de NO<sub>x</sub> sont fortement réduites au cours du weekend car les déplacements domiciles-travail sont fortement restreints. Cette diminution des émissions de NO<sub>x</sub> se traduit par une baisse de titrage d'O<sub>3</sub> par NO et donc une augmentation des concentrations d'O<sub>3</sub>. Ensuite, le rayonnement solaire incident peut être augmenté suite à la diminution de la quantité de particules d'origines anthropiques présentes dans l'atmosphère. Enfin un changement de régime de production d'O<sub>3</sub> peut s'opérer suite à la diminution des concentrations de NO<sub>x</sub>, pouvant ainsi passer d'un régime saturé à un régime limité par les NO<sub>x</sub> (section 2.2). Cette relation entre « effet weekend » et régime de production d'ozone permet d'évaluer l'efficacité des mesures de gestion de la qualité de l'air. En effet, la diminution des émissions de NO<sub>x</sub> au cours du weekend constitue une expérimentation « naturelle » des diminutions qui pourraient être attendues dans le cadre d'une réglementation visant à réduire les émissions de NO<sub>x</sub> de manière permanente (Altshuler et al., 1995). L'observation de l'évolution à long terme de l'effet weekend permet d'identifier des changements de régimes de production d'O<sub>3</sub> à l'échelle de décennies (de Foy et al., 2020). L'effet weekend peut également varier en fonction de la saison car le régime de production d'ozone peut être saison-dépendant (Zhao et al., 2019).

## 3. Variations saisonnières

La formation d'O<sub>3</sub> est extrêmement dépendante des conditions météorologiques. Ces paramètres météorologiques présentent une forte saisonnalité. L'augmentation des températures et du rayonnement solaire en période printanière et estivale favorise la formation d'O<sub>3</sub> et conduit à un optimum au cours de cette période. La période estivale est souvent caractérisée par un air plus sec. Cette baisse d'humidité relative favorise une hausse des concentrations d'O<sub>3</sub>. Les épisodes anticycloniques, caractérisé par des

vents plus faibles et un temps de résidence des masses d'air plus long associé à une pression atmosphérique plus élevée (Castell-Balaguer et al., 2012), contribuent à la hausse des concentrations d'O<sub>3</sub>. En période automnale et hivernale, la température, l'ensoleillement diminuent et l'humidité relative de l'air augmente. Les concentrations d'O<sub>3</sub> sont donc plus faibles en automne, et atteignent ensuite leur minimum en période hivernale.

#### 4. Evolution au cours des dernières décennies des concentrations d'O<sub>3</sub> et projections à moyen-terme

Depuis 1990 les concentrations moyennes annuelles d'O<sub>3</sub> suivent une tendance à l'augmentation en milieu urbain et à la diminution en milieu rural (Sicard, 2021; Sicard et al., 2013). Ces tendances globales sont toutefois variables notamment entre pays. Les augmentations différentielles entre les différents pays peuvent être dues aux quantités de précurseurs émis, aux différents niveaux d'industrialisation et aux différentes politiques publiques de gestion de la qualité de l'air mises en place. Les stratégies de réductions des émissions de précurseurs se sont principalement concentrées sur les NO<sub>x</sub>. La réduction de l'émission des NO<sub>x</sub> a été d'une part, trop peu conséquente pour passer d'un régime saturé à un régime limité en NO<sub>x</sub>, d'autre part a conduit à une diminution du titrage de l'O<sub>3</sub> par NO et donc à une augmentation de l'O<sub>3</sub> en milieu urbain. Les zones rurales étant soumises à un régime limité par les NO<sub>x</sub>, une réduction des émissions de NO<sub>x</sub> a permis une diminution des concentrations d'O<sub>3</sub> (Sicard, 2021). Les zones reculées et de haute altitude ont quant à elles subi une augmentation des concentrations d'O<sub>3</sub> (Ordóñez, 2006; Sicard, 2021). Cette augmentation a plusieurs causes associées à l'évolution climatique (bien qu'elles soient interconnectées) : l'augmentation des apports d'O<sub>3</sub> par la stratosphère dans la troposphère libre, la hausse des émissions de CH<sub>4</sub>, une hausse des températures de surface de l'air et des océans, et de l'humidité relative engendrant une augmentation des taux réactionnels, un titrage de l'O<sub>3</sub> par NO plus faible (Lefohn et al., 2012; Myhre et al., 2014; Sicard, 2021).

Ainsi, les concentrations de fond en O<sub>3</sub> continuent à augmenter. Les modèles de projections des concentrations d'O<sub>3</sub> construits selon les scénarii climatiques développés par l'IPCC (IPCC, 2014) prévoient une augmentation à l'échelle mondiale de 4 à 5 ppb d'ici 2100 pour le scénario RCP8.5 et une

diminution de 2 à 10 ppb pour le scénario 2.6 (le plus optimiste) (Sicard et al., 2017). Les effets du changement climatique induisant une augmentation des niveaux d'O<sub>3</sub> sont plus importants et effacent les bénéfices des réductions d'émissions de précurseurs (Hendriks et al., 2016). Toutefois, ces prévisions sont extrêmement difficiles à établir et présentent une certaine incertitude car l'O<sub>3</sub> agit indirectement sur les conditions climatiques. En effet, ce polluant est un gaz à effet de serre qui possède donc des capacités de forçage radiatif et peut altérer le cycle du carbone et la végétation (Lu et al., 2019).

### **Dynamique spatio-temporelle de l'ozone troposphérique et implications en termes d'exposition**

L'effet de chacun des facteurs énoncés précédemment dans ce Chapitre 1 est bien connu mais leur interaction dans une dimension temporelle et dans une dimension spatiale rend la résultante en termes de concentration d'O<sub>3</sub> complexe à appréhender et à prédire. Cette partie est subdivisée en trois sous-parties. La partie 1 synthétise la dynamique spatio-temporelle de l'ozone, la partie 2 est consacrée aux résolutions temporelles et spatiales pour étudier cette dynamique. Et enfin la partie 3 décrit l'influence des vagues de chaleur sur la dynamique spatio-temporelle de l'ozone et les manques bibliographiques identifiés dans ce domaine.

#### 1. Dynamique spatio-temporelle simplifiée de l'O<sub>3</sub>

La Figure 11 synthétise de manière simplifiée la dynamique spatio-temporelle de l'O<sub>3</sub> et illustre les différents facteurs agissant simultanément sur la formation, la destruction et le transport de l'O<sub>3</sub>. Chacun de ces facteurs (chimiques, météorologiques, physiques) possède une variabilité propre, spatiale et/ou temporelle. Certains facteurs interagissent entre eux, telles que la vitesse et la direction du vent et les émissions de précurseurs. La variabilité temporelle et spatiale de chaque facteur (variables météorologiques par exemple, présentant une forte saisonnalité et une structuration spatiale en fonction des climats, de la topographie...), combinée aux multiples interactions entre facteurs (Figure 8), rend extrêmement difficile la projection d'une concentration d'O<sub>3</sub> à un instant donné et de manière continue dans l'espace. Il est par ailleurs complexe d'identifier les déterminants agissant sur les concentrations d'O<sub>3</sub> et l'amplitude de cet effet sur un continuum spatial. Les caractéristiques locales (type et quantité d'émissions, topographie, climat, vents dominants) rendent chaque site d'étude unique.

La Figure 11 permet de schématiser les principaux mécanismes impliqués dans la dynamique spatio-temporelle de l'O<sub>3</sub>. Ce schéma illustre néanmoins des situations archétypales. Ainsi, il est admis qu'il existe un gradient croissant de concentration d'O<sub>3</sub> du milieu urbain au milieu rural, et le long du gradient altitudinal (Agathokleous et al., 2022a; Brodin et al., 2010; Brönnimann et al., 2000). Il est également communément admis que les maximums de concentrations d'O<sub>3</sub> à l'échelle journalière et annuelle, sont observées respectivement en début d'après-midi et en période estivale. Ces éléments restent toutefois des généralités. Les situations intermédiaires entre ces situations archétypales sont plus difficiles à identifier et à caractériser. La résolution, aussi bien spatiale que temporelle, des données disponibles conditionnent grandement les variations observables de concentrations d'O<sub>3</sub> et donc l'identification des phénomènes sous-jacents.

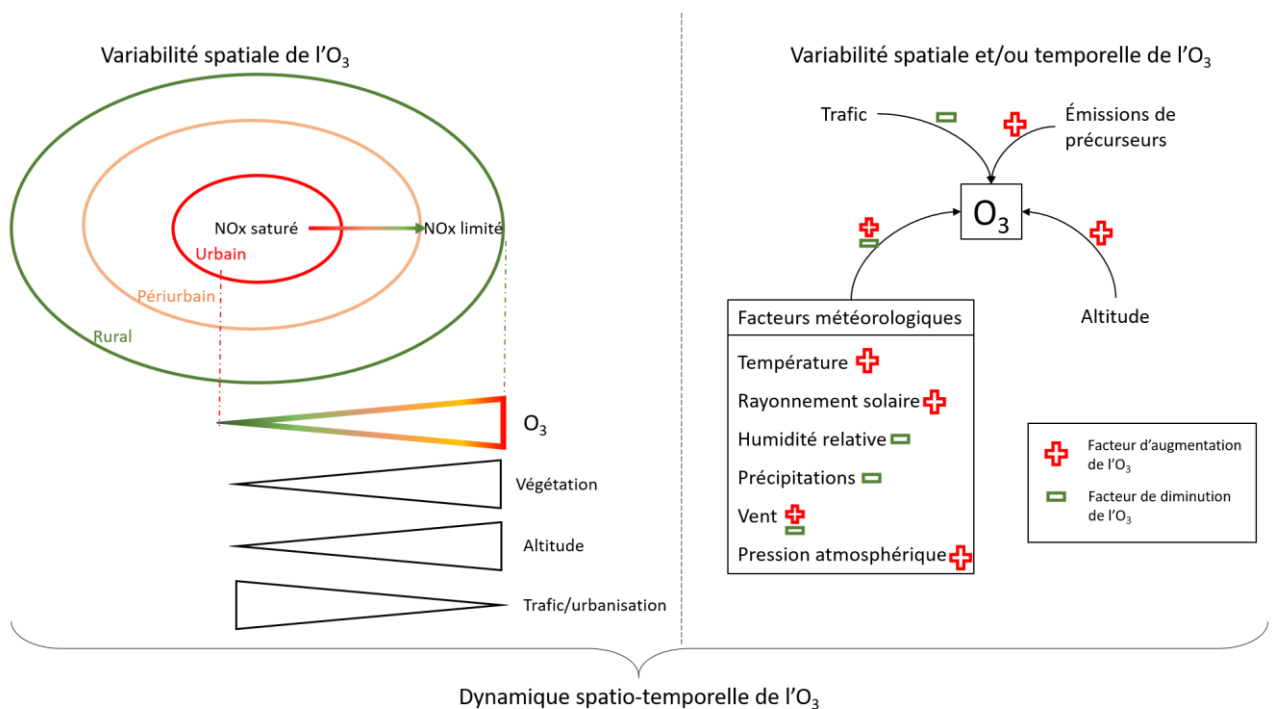


Figure 11: Schéma synthétisant et simplifiant la dynamique spatio-temporelle de l'O<sub>3</sub> à travers les principaux déterminants influant sur la variation des concentrations d'O<sub>3</sub> spatialement et/ou temporellement.

## 2. Importance de la résolution temporelle et spatiale pour caractériser les dynamiques de l'O<sub>3</sub>

L'étude de la dynamique spatio-temporelle de l'O<sub>3</sub> est réalisée à l'aide de mesures ou d'outils de modélisation (détaillés en partie [suivante](#)). Quelques études se sont attachées à décrypter cette dynamique en intégrant sa dimension temporelle et spatiale, le plus souvent à partir de données mesurées (continues dans le temps mais ponctuelles dans l'espace) (Duarte et al., 2022; Fan et al., 2020; Ma et al., 2021; Sicard et al., 2016) et plus rarement à partir d'outils de modélisation à résolution temporelle ou spatiale fine (Gao et al., 2020; Wang et al., 2018). La précision, aussi bien temporelle que spatiale, des outils utilisés dans ces études est un élément clé qui conditionnera les résultats ainsi obtenus.

La saisonnalité de l'ozone, son cycle nyctéméral et son évolution à long terme font appel à des mécanismes différents. La résolution temporelle des données induit la prise en compte de processus à des échelles différentes. Une résolution horaire permettra d'identifier des profils journaliers avec un potentiel décalage en fonction de la typologie de milieux étudiés (urbain, périurbain, rural). Une résolution journalière permettra quant à elle d'observer un « effet weekend » (partie 2). Les moyennes annuelles limitent donc considérablement l'information pouvant être extraite de telles données. De la même manière la résolution spatiale d'un modèle atmosphérique conditionnera l'échelle spatiale de travail et les contrastes d'exposition explorés ainsi que l'identification de leurs déterminants. Les méthodes permettant d'intégrer les dimensions temporelle et spatiale de la variabilité de l'O<sub>3</sub> et en retirer les mécanismes sous-jacents expliquant cette dynamique sont encore en développement et nécessitent d'être approfondies.

L'identification des zones les plus exposées et des déterminants des niveaux d'O<sub>3</sub> élevés permettent la mise en place de levier d'action et de prévention. Les zones rurales sous l'influence de centres urbains sont soumises aux niveaux les plus élevés. Or les populations vivant dans ces zones ne sont que peu sensibilisées aux risques liés à l'exposition à l'ozone et peu intégrées dans les études épidémiologiques (Bell, 2013; Madrigano et al., 2015). Les moyens de suivi et de prévention sont en effet concentrés sur les zones urbaines et périurbaines, concentrant une part plus importante de la population. L'intérêt grandissant pour l'étude de l'urbanisation et de son influence sur l'exposition environnementale révèle

un besoin de recherche et de développement de méthodes pour obtenir une meilleure compréhension de la distribution et de l'effet de polluants comme l'O<sub>3</sub> au sein des espaces urbains et à diverses échelles spatiales et temporelles (Monks et al., 2015).

Des actions de prévention permettraient de réduire l'exposition en période de pics d'ozone, en particulier dans les zones présentant une forte couverture végétale et qui plus est forestière (2). Les outils de mesure ne permettent pas l'identification des zones de vulnérabilités ; ils ont une faible couverture spatiale, une distribution spatiale hétérogène, et le coût financier que représenterait l'implantation de stations de mesure supplémentaire est élevé. La majorité des recherches explorant la dynamique spatiale et temporelle de l'ozone utilisent des données de mesures ne permettant pas une exploration de manière continue dans l'espace. Les outils de modélisation pourraient donc fournir une meilleure compréhension de la dynamique spatio-temporelle de l'ozone et de ses déterminants. Une étude récente menée en Chine à l'aide d'un outil de modélisation à résolution spatiale et temporelle fine (données journalières au km<sup>2</sup> de 2013 à 2019) illustre le développement aussi récent que rapide de ces outils (Meng et al., 2022). Cette résolution croissante permet de considérer les contrastes spatiaux d'exposition de la population et les pics d'exposition estivaux.

### 3. Impacts des vagues de chaleur sur la dynamique spatio-temporelle de l'ozone troposphérique et risque de co-exposition à la chaleur et à l'ozone

Les vagues de chaleur, caractérisée par des températures extrêmes, engendrent une augmentation des concentrations d'O<sub>3</sub> par une accélération des cinétiques chimiques des réactions de formation d'O<sub>3</sub> (Pyrgou et al., 2018). Les niveaux d'ozone peuvent également être impactés de manière indirecte par ces phénomènes climatiques. En effet, la hausse des températures entraîne une augmentation des émissions de COVB et une diminution de leur absorption par voie stomatique par le couvert végétal (Chapitre 1 - 2)(Solberg et al., 2008). De plus, l'humidité relative de l'air est souvent réduite au cours de ces épisodes. L'humidité ambiante étant considérée comme un facteur de réduction des concentrations d'O<sub>3</sub> (section 4), l'assèchement de l'air participe également à la hausse des niveaux

d'ozone. Les conditions anticycloniques, souvent associées aux vagues de chaleur, participent à la stagnation des masses d'air et à l'accumulation des polluants.

L'interaction entre les dynamiques spatio-temporelles des vagues de chaleur et des concentrations d'O<sub>3</sub> reste encore très incertaine (Kalisa et al., 2018). De nombreuses études se sont attachées à évaluer l'impact de l'exposition conjointe à l'O<sub>3</sub> et aux vagues de chaleurs sur la mortalité (Filleul et al., 2006; Krug et al., 2020; Madrigano et al., 2015; Reid et al., 2012). Néanmoins, la relation temporelle, spatiale, et spatio-temporelle, entre dynamique d'ozone et de température est peu caractérisée (Kalisa et al., 2018; Struzewska and Jefimow, 2016). Les délais d'induction d'un pic d'ozone à partir d'un pic de chaleur sont extrêmement peu décrits et peuvent dépendre des caractéristiques locales (Khomsy et al., 2022). Les caractéristiques des vagues de chaleur (intensité, durée, sévérité) semblent moduler l'intensité et la durée des pics d'ozone (Schnell and Prather, 2017; Struzewska and Jefimow, 2016). Cette compréhension de l'interaction entre ozone et vagues de chaleur apparaît cependant comme un élément clé pour une meilleure gestion et prévention face à ce double facteur de risque. En effet, le risque de mortalité associé à l'exposition à l'ozone semble plus élevé aux cours des vagues de chaleur (Madrigano et al., 2015). Les vagues de chaleur pourraient perturber la distribution de l'O<sub>3</sub> entre zones urbaines, périurbaines et rurales. Le milieu urbain abrite une part majeure de la population mais est exposé en moyenne à des niveaux moindres d'O<sub>3</sub>. Si le rapport d'exposition entre zones urbaines et rurales venait à s'inverser lors de phénomènes d'exposition aiguë, le risque sanitaire associé à cette exposition pourrait s'avérer préoccupant. Une étude récente menée par Zhang et al. (2017) a mis en évidence que lorsque la vitesse du vent est très faible les zones périurbaines et rurales sont moins soumises à des concentrations élevées d'O<sub>3</sub>. Dans ces mêmes conditions, les zones urbaines sont davantage soumises à des concentrations élevées d'O<sub>3</sub>. Ceci s'explique, d'une part par l'apport de précurseurs moindre des zones urbaines vers les zones périurbaines et rurales, et d'autre part par la stagnation des masses d'air et un brassage vertical moindre diminuant la vitesse des dépôts d'O<sub>3</sub> en milieu urbain.

Selon Ryu et al. (2013) l'augmentation des concentrations en période de vague de chaleur en milieu urbain serait plus élevée au cours de la phase nocturne (5,3 ppb) que diurne (2,4 ppb). Les surfaces artificialisées du milieu urbain (asphalte, béton) redistribuent au cours de la nuit la chaleur accumulée



au cours de la journée (Ulpiani, 2021). Les environnements fortement artificialisés favorisent donc les températures nocturnes élevées. Les niveaux d'O<sub>3</sub> au cours de la nuit restent élevés et un phénomène d'accumulation au cours de la vague de chaleur peut apparaître. Ceci pourrait expliquer pourquoi la durée des vagues de chaleur semble moduler l'intensité et la durée des pics d'ozone (Schnell et al., 2015; Struzewska and Jefimow, 2016).

## **Evaluation des concentrations d'ozone troposphérique : outils de modélisation et évaluation de l'exposition de la population**

### 1. Outils de modélisation

Il existe une grande diversité d'outils de modélisation utilisés pour l'évaluation de l'exposition à la pollution atmosphérique. Ces outils permettent une évaluation des niveaux de pollution atmosphérique de manière continue dans l'espace, contrairement aux mesures. Ces modèles peuvent être divisés en deux types d'approches : statistiques (dont les Land Use Regression (LUR)) ou déterministes (dont les Chemical Transport Models (CTM)). Les approches statistiques se subdivisent entre les modèles purement statistiques et les modèles géostatistiques (De Marco et al., 2022). Dans un souci de concision, un focus sera ici fait sur deux types de modèles parmi les plus largement utilisés : les modèles LUR (approche géostatistique) et les CTMs (approche déterministe).

#### *1.1. Les modèles Land Use Regression (LUR)*

Les LUR constituent l'une des méthodes géostatistiques les plus utilisées et efficaces pour estimer les concentrations d'O<sub>3</sub> troposphérique (De Marco et al., 2022). Les LUR combinent des valeurs de concentrations de polluants atmosphériques mesurées au niveau de stations de mesure et des variables spatialisées permettant d'en expliquer la distribution et ses variations (Hoek et al., 2008). Le modèle ainsi construit est ensuite appliqué aux zones non-échantillonnées pour prédire les concentrations de manière continue dans l'espace. Les variables explicatives classiquement intégrées sont : des indicateurs de trafic routier (densité du trafic, linéaire de route...), l'occupation du sol, un indice de végétation, la distance au littoral, les émissions de précurseurs, l'altitude (Hoek et al., 2008; Sicard et al., 2016). Les LUR sont généralement appliqués à petites échelles spatiales. Leur applicabilité à l'échelle nationale a

été démontrée dans de récentes études (Sicard et al., 2016). Toutefois, le nombre de stations de mesure de la qualité de l'air disponibles ainsi que l'homogénéité de leur distribution sur le territoire étudié sont des éléments clés pour obtenir des performances élevées (Basagaña et al., 2012; Wang et al., 2012). La variabilité spatiale des concentrations d'O<sub>3</sub> est bien représentée par le LUR, ce qui n'est pas le cas pour la variabilité temporelle de ce composé (Ma et al., 2020).

### *1.2. Les Chemical Transport Models (CTMs)*

Les modèles de chimie-transport (CTMs) permettent de modéliser la formation, la destruction et le transport de polluants atmosphériques (Jerrett et al., 2005). Ces modèles reposent sur des équations déterministes qui simulent le comportement des composés chimiques dans l'atmosphère. Ces équations nécessitent des données d'entrée complexes et volumineuses que sont les émissions de polluants primaires, la topographie, les conditions météorologiques. Cette simulation dynamique des concentrations de polluants atmosphériques nécessite des ressources conséquentes en termes de calculs, de données d'entrée et d'expertise quant à la chimie atmosphérique. Toutefois, ce type d'approche a été largement utilisée au cours des dernières décennies afin d'obtenir un suivi des concentrations de polluants atmosphériques sur le long-terme (de l'ordre de la décennie) et sur des surfaces très vastes (territoires régionaux, nationaux ou internationaux), notamment dans le cadre de l'évaluation de l'exposition de la population (De Marco et al., 2022; Reche et al., 2018). Les CTMs sont en effet capables de reproduire de manière assez fidèle les contrastes spatiaux d'exposition (Wang et al., 2016). La résolution spatiale de ces modèles est généralement inversement proportionnelle à la surface couverte. Selon Schaap et al. (2015), en Europe, la résolution varie de 12 à 25 km pour des applications à l'échelle européenne (Mues et al., 2014; Pay et al., 2010; Zhang et al., 2012), de 4 à 10 km sur territoires nationaux (Hendriks et al., 2013; Vieno et al., 2010) et enfin peut descendre à 1 km pour certaines zones spécifiques (Pay et al., 2014).

Les forces et faiblesses des LUR et des CTMs sont synthétisées dans le Tableau 1. Certaines études se sont déjà attachées à combiner LUR et CTM afin de retirer les avantages respectifs de ces deux méthodes (Wang et al., 2016). En effet, ces deux approches sont complémentaires.

Tableau 1: Forces et faiblesses de deux approches de modélisation : LUR et CTM.

|     |            |   |
|-----|------------|---|
| LUR | Faiblesses | <ul style="list-style-type: none"> <li>• Performances fortement dépendantes du nombre de stations de mesure et de leur répartition spatiale (Sicard et al., 2016)</li> <li>• Validation du modèle utilisant une « <i>cross-validation</i> » (validation interne), en particulier la « <i>Leave One Out Cross-Validation (LOOCV)</i> » (Ma et al., 2020), pouvant surestimer les capacités prédictives du modèle (Wang et al., 2013)</li> <li>• Intégration rare des paramètres météorologiques clés comme la température, le rayonnement solaire, l'humidité relative... (Wang et al., 2016)</li> <li>• Capacité limitée à représenter la variabilité temporelle (pics d'ozone par exemple) et possibilité de passer à côté de patterns régionaux si les paramètres du LUR présentent peu de variabilité (Ma et al., 2020)</li> </ul> |
|     | Forces     | <ul style="list-style-type: none"> <li>• Caractérisation de la variabilité de l'ozone à des échelles spatiales fines</li> <li>• Technique de modélisation aboutie avec une faible complexité de calcul, une facilité d'utilisation permettant d'obtenir des performances satisfaisantes à petites échelles (Ma et al., 2020)</li> <li>• Besoins de ressources en termes de données et de calculs qui sont moindres que pour le CTM (Braun et al., 2017)</li> </ul>  |
| CTM | Faiblesses | <ul style="list-style-type: none"> <li>• Résolution spatiale en moyenne plus grossière (<math>\geq 4\text{km}</math>) que les LUR</li> <li>• Inventaires des émissions incertains</li> <li>• Besoin d'une capacité de calcul conséquente, de données d'entrées volumineuses et d'une bonne connaissance des processus atmosphériques et des sources de pollution (Braun et al., 2017; Tong et al., 2011)</li> </ul>   |
|     | Forces     | <ul style="list-style-type: none"> <li>• Large couverture spatiale (zones régionales ou (trans)nationales) (Wang et al., 2016)</li> <li>• Bonne représentation de la variabilité spatio-temporelle de l'ozone (Wang et al., 2016) grâce à la prise en compte des processus atmosphériques et météorologiques</li> <li>• Prise en compte de la topographie, des émissions et des conditions météorologiques (processus physico-chimiques, transport et chimie atmosphérique)</li> </ul>  |

Des approches de modèles d'ensemble (« *ensemble models* ») sont actuellement en développement et leur utilisation est encouragée pour de futures études (De Marco et al., 2022). Ces modélisations

d'ensemble combinent les prédictions issues de plusieurs modèles basés sur des algorithmes ou sur des jeux de données différents. Ce type d'approche permet de généraliser des modèles et de réduire les erreurs de prédiction lors de cette phase de généralisation (Kotu and Deshpande, 2019). Cependant, si les ressources nécessaires pour développer un modèle CTM unique sont conséquentes, celles nécessaires pour mettre en place une approche de modélisation d'ensemble le sont encore bien davantage. Ce type d'approche sont plus performantes et seront peut-être utilisées en routine à l'avenir mais reste actuellement peu accessible.

### *1.3. Validation des modèles et besoins potentiels d'optimisation*

Ces modèles, bien qu'étant de puissants outils d'évaluation de l'exposition à large échelle, peuvent présenter des biais et un manque de précision, notamment lorsque l'on souhaite travailler à l'échelle locale. Or, la précision, l'exactitude et le contraste dans les données d'exposition sont indispensables pour évaluer correctement la relation entre une exposition et son potentiel effet sanitaire ou environnemental (Crooks and Özkaynak, 2014). Ceci est particulièrement le cas dans un contexte épidémiologique où de faibles effets sont investigués. De plus, l'exposition aiguë correspondant à des pics d'exposition, par exemple des pics estivaux d'ozone, ne peut être correctement prise en compte que si le modèle représente correctement cette variabilité temporelle. Or les modèles peuvent lisser les courbes d'exposition, bien qu'étant en moyenne exacts. Les pics estivaux d'O<sub>3</sub> représentent un enjeu épidémiologique car ils sont souvent associés à des températures extrêmes voire à des phénomènes de vagues de chaleur. L'utilisation des données de modélisation pour l'évaluation de l'exposition doit donc faire l'objet d'une phase de validation préalable, avec une attention particulière pour les pics estivaux. Cette validation dépend d'approches métrologiques primordiales permettant un comparatif entre les valeurs modélisées et mesurées. Un manque d'exactitude et de précision, ainsi qu'un manque de variabilité des valeurs par rapport aux mesures peuvent être détectés suite à cette évaluation des performances du modèle (Simon et al., 2012; Steven Porter et al., 2015). Les mesures restent donc incontournables pour valider les modèles de prédiction des concentrations d'O<sub>3</sub> troposphériques.

## 2. Evaluation de l'exposition de la population

### 2.1. Outils de mesure

Le suivi des niveaux d'exposition de la population (réalisé par mesures) est centré sur les pôles urbains. Cette focalisation sur les zones urbaines s'explique par deux éléments. D'une part, la majorité de la population réside dans ces zones, ce qui définit un enjeu sanitaire prioritaire. D'autre part, la législation encadrant le suivi de la qualité de l'air sur le territoire national (loi LAURE, n°96-1236 du 30 décembre 1996) était initialement centrée sur les centres urbains très denses puis a été élargie aux villes de 100 000 habitants et enfin l'ensemble du territoire. Or, la pollution à l'O<sub>3</sub> trouve son paroxysme en périphérie des pôles urbains ainsi que dans les zones rurales perturbée (Figure 4).

Le suivi par mesure, bien qu'indispensable, présente certaines limites : couverture spatiale limitée, distribution spatiale hétérogène et très faible couverture des zones rurales, coût élevé. Il est possible d'équiper les individus avec des capteurs portatifs (sur eux ou placé à leur domicile) afin de prendre en compte leur mobilité et leur environnement intérieur (Dessimond, 2021). Ce type d'appareils permet d'obtenir à un coût réduit des données d'exposition individuelle prenant en compte la mobilité des individus. Ces instruments présentent néanmoins des manques de précision et de fiabilités en comparaison aux stations de référence (Dessimond, 2021). Cette démarche n'est pas adaptée à une évaluation de l'exposition en population générale, car cela représenterait des contraintes logistiques et financières inenvisageables.

### 2.2. Outils de modélisation

Les outils de modélisation permettent de répondre aux enjeux d'évaluation de l'exposition à large échelle (avec une résolution temporelle et spatiale fine) à coût réduit comparativement aux mesures.

Les méthodes d'évaluation de l'exposition à l'échelle individuelle ou populationnelle se centrent donc désormais de plus en plus sur la modélisation. Parmi les modèles évoqués précédemment, les CTMs sont recommandés pour l'évaluation du risque sanitaire et environnemental à large échelle (De Marco et al., 2022). Des améliorations métrologiques sont toujours nécessaires pour atteindre une précision à la fois spatiale et temporelle afin d'identifier aussi bien une exposition chronique qu'aiguë. Ceci

représente un défi technique pour le développement d'outils à large échelle spatiale (Colette et al., 2014). Cet enjeu est particulièrement prégnant lorsqu'il s'agit d'un polluant extrêmement variable dans l'espace et dans le temps, tel que l'O<sub>3</sub> (Hoek, 2017).

L'évaluation de l'exposition des individus repose bien souvent sur la localisation de leur domicile. Toutefois, les individus passent la majeure partie de leur temps dans les lieux d'habitation (donc exposés à la pollution de l'air intérieur) et dans leur environnement de travail (ne correspondant pas le plus souvent à leur lieu de résidence). Ceci représente une limite, partagée par les outils de modélisation et les stations de mesure fixes, car seule la contamination de l'air ambiant est évaluée par ces outils. De plus, la mobilité des individus n'est généralement pas prise en compte dans les études en épidémiologie car elle représente un frein technique important. Toutefois, l'utilisation de la géolocalisation des individus, à l'aide de leur téléphone portable ou d'autres appareils, peut permettre la prise en compte de la mobilité des individus au cours d'une journée (déplacement domicile-travail, lieu de travail...). Simoncic et al. (2022) ont récemment proposé une procédure permettant de calculer l'exposition individuelle quotidienne aux polluants atmosphériques en la pondérant par le temps passé dans chacun des lieux les plus fréquentés (travail, achats, adresse résidentielle, etc.) et dans les trajets domicile-travail. La disponibilité des données relatives à la mobilité des individus est extrêmement réduite, en particulier à l'échelle de la population générale, mais pourrait s'accroître via l'utilisation de la géolocalisation. Un outil de modélisation couplé à une prise en compte de la mobilité par géolocalisation des individus constitue un outil puissant d'évaluation de l'exposition individuelle.

## Chapitre 2 : Matériels et Méthodes

Ce chapitre décrivant les matériels et méthodes utilisés dans le cadre de ces travaux de thèse est subdivisé en trois parties séquentielles. Dans un premier temps, la plateforme de modélisation (PREV'EST) utilisée comme outils d'évaluation de l'exposition sera présentée, ainsi que les méthodes d'évaluation de ses performances puis de correction de ses données. Dans un second temps, la méthodologie permettant d'étudier la dynamique spatio-temporelle de l'O<sub>3</sub> est décrite. Et dans un troisième temps, les données populationnelles et la méthodologie de mise en relation avec les dynamiques spatio-temporelles d'O<sub>3</sub> et de vagues de chaleur seront présentées.

### Evaluation de l'exposition

#### 1. Modélisation brute des concentrations d'ozone

##### *1.1. La plateforme de modélisation PREV'EST*

La plateforme de modélisation PREV'EST, gérée par ATMO Grand-Est (ASQAA régionale), fournit quotidiennement une valeur moyenne et maximum de concentration d'O<sub>3</sub> sur un maillage de 3km par 3 km qui s'étend sur le quart Nord-Est de la France ainsi qu'une partie des pays limitrophes (Luxembourg, Allemagne, Suisse) (Figure 12). Ce maillage est issu de domaines imbriqués : 45 km par 45km, puis 15km par 15km et enfin 3km par 3km. Cette plateforme combine un modèle météorologique (WRF (Weather Research and Forecasting)) et le modèle de chimie-transport CHIMERE (Mailler et al., 2017; Menut et al., 2013; Valari et al., 2011). Le processus de modélisation est présenté dans le Chapitre 3 -

2.1.1. Study site and data sources La zone rouge sur la Figure 12 correspond au sous-domaine de modélisation utilisé pour les travaux présentés en Chapitre 3.

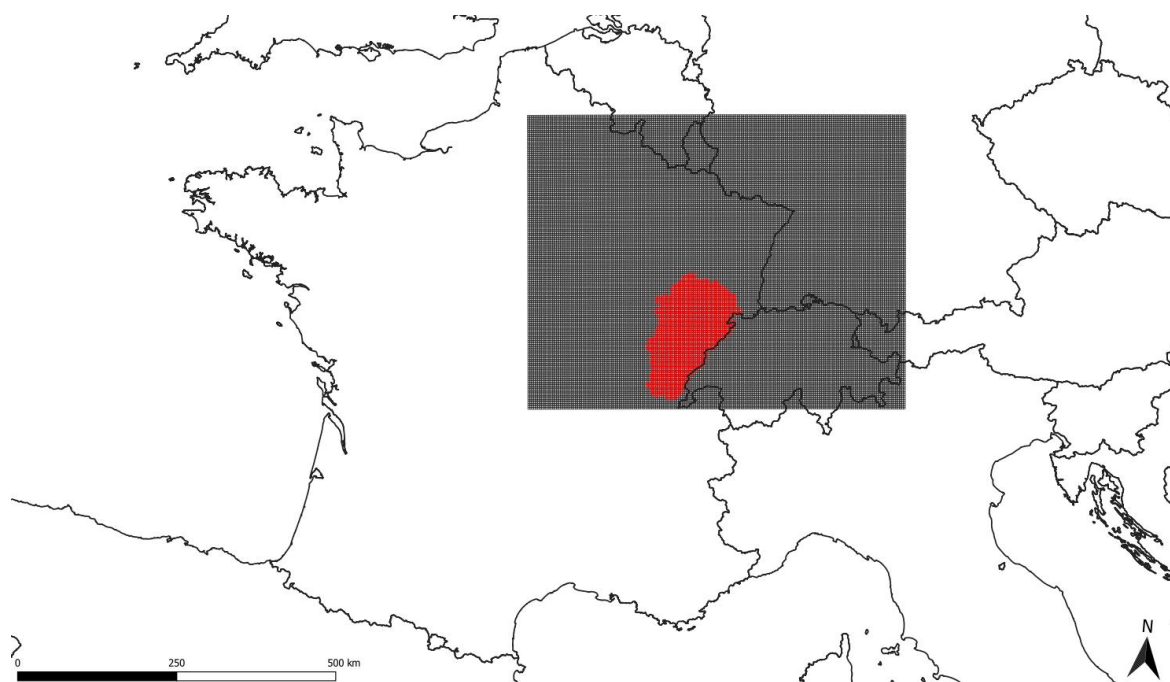


Figure 12: Domaine de modélisation de la plateforme PREV'EST (grisée), et sous-domaine de modélisation (région Franche-Comté) ayant servi à l'évaluation des performances du modèle.

### 1.2. Valeurs de références : stations de mesure de la qualité de l'air

De manière complémentaire à la modélisation, les mesures effectuées par le réseau des stations de surveillance de la qualité de l'air ont été fournies par ATMO BFC. La moyenne ainsi que le maximum journalier de concentration d'O<sub>3</sub> ont été enregistrés quotidiennement. Ces mesures, soumises à des procédures normalisées (normes ISO ou AFNOR selon l'échelle internationale ou nationale) (Dessimond, 2021), sont utilisées comme valeurs de référence afin d'évaluer les performances des données de modélisation issues de PREV'EST. Ces stations sont permanentes (enregistrement toute l'année) ou saisonnières (enregistrement d'avril à septembre), et relèvent de différentes typologies d'urbanisation (urbaine, périurbaine, rurale). Chaque station utilisée a été associée à la maille de modélisation sur laquelle elle se trouve afin d'effectuer un comparatif modèle-mesure et de calculer des indicateurs de performances du modèle. Ces comparatifs sont généralement conduits à l'aide de calculs d'indicateurs de performance, présentés dans la partie suivante.



## 2. Evaluation de la qualité de la modélisation brute

L'évaluation des performances du modèle (PREV'EST) est réalisée par comparaison des données modélisées brutes aux valeurs mesurées. Dans un premier temps, des indicateurs de performance moyens sont calculés sur l'ensemble du jeu de données. Dans un second temps, ces indicateurs sont calculés à l'échelle de la station de mesure sous la forme de « target plots » intégrant plusieurs indicateurs de performances.

### 2.1. Indicateurs de performance moyens

Les indicateurs de performances moyens utilisés sont : le biais moyen (Mean Bias Error (MBE)), l'erreur absolue moyenne (Mean Absolute Error (MAE)), le Root Mean Square Error (RMSE) et le coefficient de corrélation de Pearson ( $r$ ).

Le MBE permet de quantifier la sur- ou sous-estimation générée par le modèle.

$$MBE = \frac{1}{n} \sum_{i=1}^n (M_i - O_i) \quad (1)$$

Le MAE quantifie l'erreur de modélisation absolue moyenne (Abdulah Al-Sudani et al., 2019; Adnan et al., 2020; Kisi and Parmar, 2016; Martin, 2011; Mehdizadeh et al., 2018) selon l'équation suivante :

$$MAE = \frac{1}{n} \sum_{i=1}^n |M_i - O_i| \quad (2)$$

Le RMSE est un indicateur de précision (Abdulah Al-Sudani et al., 2019; Lesmeister, 2019; Sekulic and Kowalski, 1992; Yilmaz et al., 2018) qui donne plus de poids que le MAE aux écarts modèle-mesure importants (Bouthevillain and Mathis, 1995). Il est calculé comme suit :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - O_i)^2} \quad (3)$$

Enfin le coefficient de corrélation de Pearson est calculé selon la formule suivante :

$$r = \frac{\sum_{i=1}^n (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (4)$$

Avec  $O_i$  étant la  $i^{\text{ème}}$  mesure,  $M_i$  la  $i^{\text{ème}}$  valeur modélisée,  $\bar{O}$  la moyenne des concentrations journalières mesurées, et  $\bar{M}$  la moyenne des concentrations journalières modélisées.

## 2.2. Indicateurs à l'échelle de la station de mesure : Target plots

Les indicateurs de performances moyens (décrits dans la section précédente) ne permettent pas d'identifier un potentiel biais ou manque de précision différentiel entre stations. Pour obtenir des indicateurs de performances déclinés à l'échelle de la station nous utilisons un outil synthétique sous forme de cible directement inspiré du DELTA Tool développé par le JRC (Joint Research Center) FAIRMODE (Forum for Air Modelling in Europe) (Janssen et al., 2017; Kushta et al., 2019; Viaene et al., 2016). Un exemple de « target plot » est illustré en Figure 13.

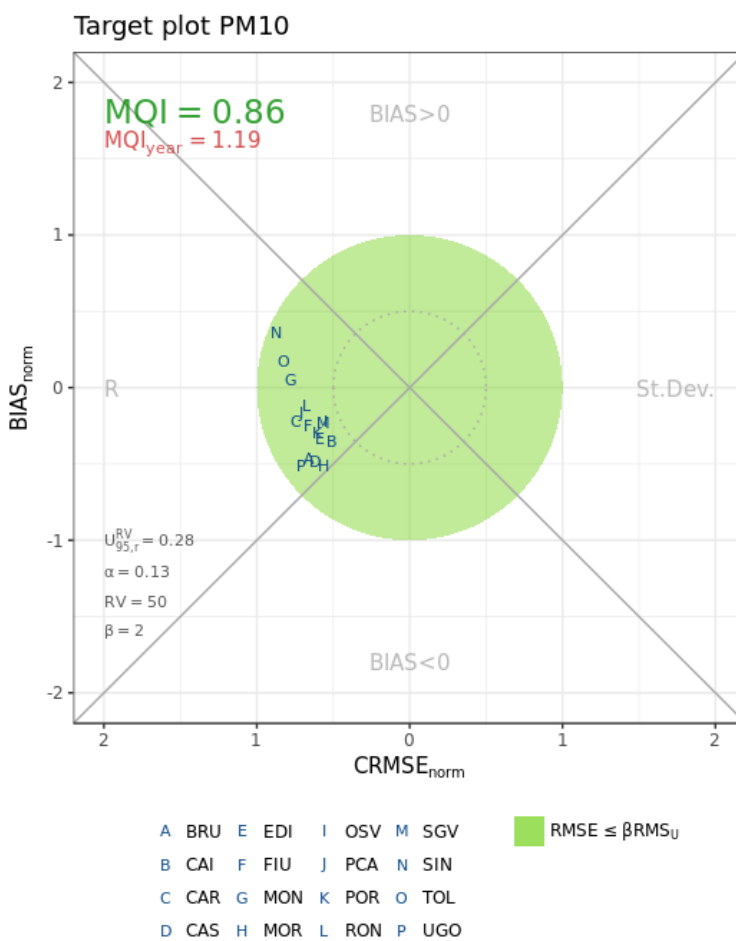


Figure 13: Exemple de target plot pour 16 stations de mesures (Bonafè, 2017). Les différentes lettres représentent les différentes stations.

Cette « cible » combine les indicateurs suivants : CRMSE (Centred Root Mean Square Error),  $r$  (coefficient de corrélation (Pearson)), SD (Standard Deviation), BIAS<sub>Norm</sub> (biais moyen normalisé). Ils sont calculés selon les équations suivantes :

$$CRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [(M_i - \bar{M}) - (O_i - \bar{O})]^2} \quad (5)$$

$$CRMSE_{Norm} = \frac{CRMSE}{\beta \times RMSU} \quad (6)$$

$$BIAS_{Norm} = \frac{MBE}{\beta \times RMSU} \quad (7)$$

Avec  $O_i$  et  $M_i$  étant respectivement la  $i^{ème}$  concentration observée et modélisée,  $\beta$  étant une constante dont la valeur est fixée à 2 et RMSU un indicateur d'incertitude autour des observations (Bonafè, 2017; Riviere et al., 2019).

Le CRMSE reflète la capacité ( $CRMSE \approx 0$ ) ou l'incapacité ( $CRMSE > 1$  ou  $CRMSE < -1$ ) du modèle à reproduire les oscillations journalières des concentrations mesurées. Cet indicateur quantifie la dispersion des données modélisées autour de leur moyenne comparativement à la dispersion des données mesurées autour de leur moyenne.

Chaque point projeté sur le « target plot » représente une station de mesure et sa position sur la cible décrit les performances du modèle à cette station. Si les points sont localisés dans la partie gauche de la cible, l'erreur du CRMSE est principalement due à un manque de corrélation entre données modélisées et mesurées. S'ils sont localisés dans la partie droite, cette erreur de prédiction sera plutôt due à un manque de variabilité dans les données modélisées. Enfin les parties supérieure et inférieure de la cible caractérisent respectivement une sur- ou une sous-estimation (Kushta et al., 2019; Thunis and Cuvelier, 2018).

Les points se trouvant à l'intérieur de la cible respecte un indicateur de qualité appelé MQO (Modelling Quality Objective). Le MQO est satisfait si le MQI (Modelling Quality Indicator) est inférieur ou égal à un. Le MQI quantifie l'écart entre données modélisées et mesurées normalisé par l'incertitude de mesure (Janssen et al., 2017) :

$$MQI = \frac{|O_i - M_i|}{\beta U_{95}(O_i)}, \beta = 2 \quad (8)$$

Avec:

U95: l'incertitude de mesure du 95<sup>ème</sup> percentile du niveau de concentration observé

$\beta$ : coefficient qui pondère la proportionnalité du biais à l'incertitude de mesure.

La valeur de  $\beta$  détermine la rigueur du critère du MQO. La valeur de  $\beta$  est arbitrairement fixée à 2, ce qui permet d'obtenir un écart entre les concentrations modélisées et mesurées égal à deux fois l'incertitude de mesure (Janssen et al., 2017).

### 3. Méthodologie de correction des concentrations d'ozone

La méthodologie de correction des concentrations d'O<sub>3</sub> proposée dans ces travaux sera décrite de manière séquentielle selon les étapes suivantes : 1) modélisation et prédiction de l'écart modèle-mesure, 2) description des différents types de correction proposés, 3) évaluation des performances des corrections proposées, 4) validation externe des corrections proposées, et enfin 5) application et projection de la correction sélectionnée.

#### *3.1. Modélisation de l'erreur de prédiction comme fonction de variables environnementales*

Une première phase de modélisation de l'erreur de prédiction du modèle PREV'EST brut est nécessaire. Cette phase, décrite dans la partie matériels et méthodes du Chapitre 3, permettra ensuite d'effectuer un correctif *a posteriori* des concentrations modélisées brutes de PREV'EST. Cette sous-section 1.3.1 s'attachera à décrire le fonctionnement du modèle MARS en partie 1.3.1.1 puis l'optimisation de ses hyperparamètres sera détaillée en partie 1.3.1.2.

##### *3.1.1. Modèles MARS*

Le modèle MARS (Multivariate Adaptive Regression Splines) est une méthode de régression non paramétrique proposée par Friedman en 1991 (Friedman, 1991; Friedman and Roosen, 1995). Cette méthode est capable de générer des régressions linéaires par « morceaux » en identifiant des points d'inflexion (« turning points » ou « knots ») dans la relation entre Y (variable dépendante expliquée) et chacune des variables explicatives (Figure 14). Ce type de modèle permet de modéliser des relations complexes entre variable dépendante et indépendante(s).

Pour chaque nœud identifié, une paire de fonctions de base est produite de part et d'autre du nœud. Ces fonctions étudiées par de nombreux auteurs (De Cos Juez et al., 2009; Friedman, 1991; Friedman and Roosen, 1995; Hastie et al., 2009; Sekulic and Kowalski, 1992) peuvent être décrites avec les équations suivantes :

$$[-(x - t)]_+^q = \begin{cases} (t - x)^q & \text{if } x < t \\ 0 & \text{sinon} \end{cases} \quad (9)$$

$$[+(x - t)]_+^q = \begin{cases} (t - x)^q & \text{if } x \geq t \\ 0 & \text{sinon} \end{cases} \quad (10)$$

Avec  $q (\geq 0)$  étant un exposant du spline qui détermine le degré de lissage de la fonction. Une valeur de  $q=1$  permet uniquement de construire un spline linéaire (García-Nieto et al., 2016), ce qui est le cas pour nos modèles. Une paire de splines avec  $q=1$  et  $t=0,5$  est illustrée en (Figure 15). Ces fonctions de base décrites (Alonso Fernández et al., 2014; Álvarez Antón et al., 2013; De Cos Juez et al., 2009; Friedman, 1991; Friedman and Roosen, 1995) sont les éléments de base qui constituent le modèle qui peut être traduit par l'équation suivante:

$$y = f(X) + e \quad (11)$$

$$\hat{y} = \hat{f}_M(x) = c_0 + \sum_{m=1}^M c_m B_m(X) \quad (12)$$

Avec :

$\hat{y}$ : variable dépendante modélisée et prédite par le modèle

$f(X)$ : somme pondérée des fonctions de bases ( $B_m(X)$ ) dépendant de  $X$  (groupe de variables explicatives indépendantes) et  $e$  (résidus)

$\hat{f}_M(x)$  : fonction de prédiction de  $\hat{y}$  selon les valeurs prises par  $x$  et selon les différentes fonctions de bases  $B_m(X)$

$M$ : nombre de fonctions de bases

$c_0$  : constante

$B_m(X)$  : la  $m^{\text{ième}}$  fonction de base

$c_m$ : coefficient de la  $m^{\text{ième}}$  fonction de base

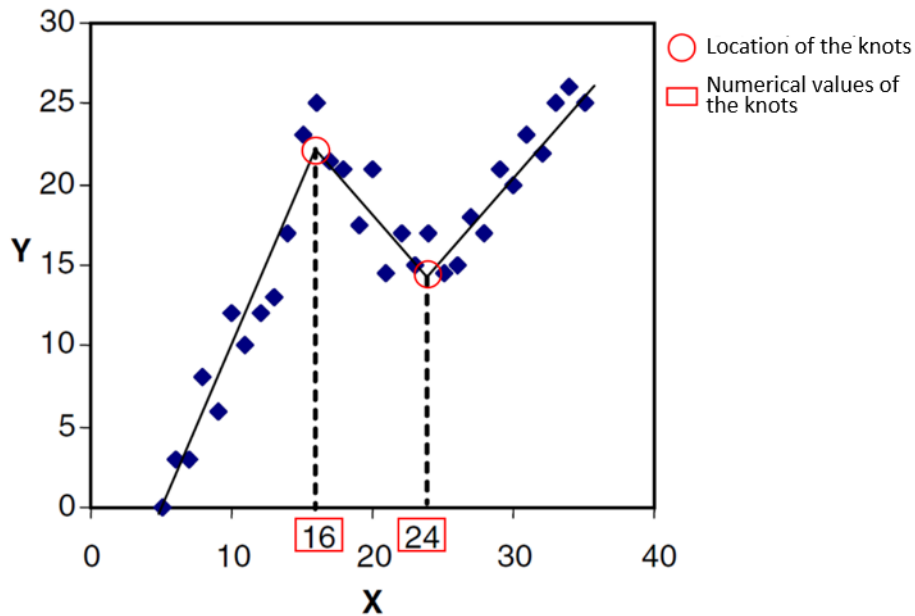


Figure 14: Exemple de noeuds ou “turning points” identifié avec un modèle MARS dans la relation entre une variable dépendante ( $Y$ ) et une variable explicative ( $X$ ), tiré et adapté de (Zhou and Leung, 2007).

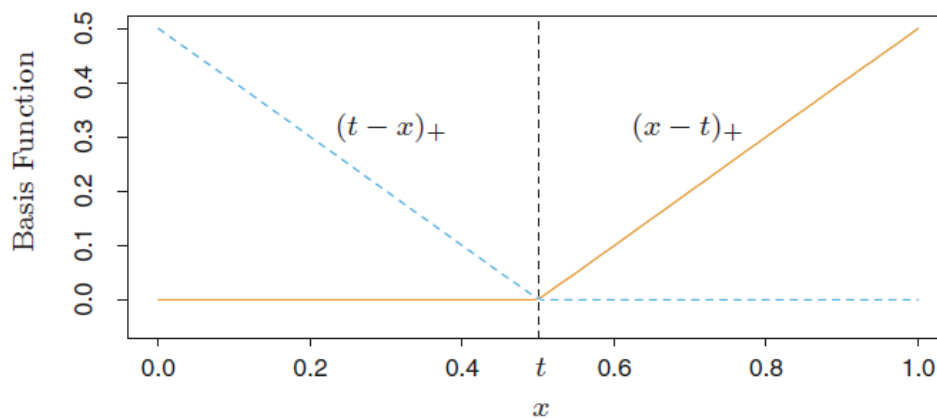


Figure 15 : Paire de fonction de base avec  $t = 0,5$ . Illustration tirée de Hastie et al., 2009. Le spline de gauche ( $x < t$ ,  $[-(x - t)]_+^q$ ) est représenté par une ligne en pointillée ; celui de droite ( $x > t$ ,  $[(x - t)]_+^q$ ) par une ligne pleine.

La sélection des fonctions des base ( $B_m(X)$ ) intégrées dans le modèle a été réalisée via une procédure en deux étapes appelées «forward-backward» (Milborrow, 2019). La première étape (forward) consiste à ajouter les termes au modèles de manière séquentielle (Friedman, 1991; Friedman and Roosen, 1995; García Nieto et al., 2012; Sekulic and Kowalski, 1992). Le modèle alors obtenu est surajusté et présente

de faibles capacités prédictives sur de nouvelles données. La deuxième phase (backward) permet d'identifier, parmi toutes les fonctions de base du modèle en sortie de la première phase (forward), celles qui minimiseront le GCV (Generalized Cross-Validation). Le GCV est un indicateur combinant l'évaluation de la capacité prédictive du modèle et sa parcimonie. Il est résumé par la formule suivante :

$$GCV(M) = \frac{\sum_{i=1}^n (y_i - f_M(x_i))^2}{(1 - \frac{C(M)}{n})^2} \quad (13)$$

où  $C(M)$  est une pénalité de complexité du modèle (parcimonie), croissante avec le nombre de fonctions de base introduites dans le modèle. Le pénalité de complexité  $C(M)$  (De Cos Juez et al., 2009; Friedman and Roosen, 1995; Garcia Nieto et al., 2011; García Nieto et al., 2012, p.; Xu et al., 2004) est calculée par l'expression suivante:

$$C(M) = (M + 1) + dM \quad (14)$$

$M$  étant le nombre de fonctions de base ( $B_m(X)$ ) et  $d$  étant une pénalité appliquée à chaque BF ajoutée au modèle. La lettre  $d$  représente une contribution supplémentaire de chaque fonction de base à la complexité globale du modèle résultant de l'ajustement (non linéaire) des paramètres de la fonction de base aux données à chaque étape itérative (Friedman and Roosen, 1995).

### 3.1.2. Optimisation des hyperparamètres des modèles

Certains paramètres du modèle MARS peuvent être optimisés pour atteindre des performances prédictives maximisées (Yagli et al., 2019). Les deux paramètres influençant principalement les performances du modèle sont le nombre maximal de fonctions de base autorisées dans le modèle et le degré d'interaction entre ces fonctions (Boehmke and Greenwell, 2019). Une valeur de degré d'interaction de 1 correspondrait à la construction d'un modèle additif (sans interaction entre les termes du modèle). Un algorithme automatique (fonction *train*, package *caret* de R) a permis de déterminer la combinaison optimale des valeurs de ces deux paramètres pour chaque modèle. Cet algorithme repose sur un rééchantillonnage de type *k-fold cross-validation* (où  $k=10$ ), selon une pratique courante en machine learning (James et al., 2014; Yagli et al., 2019).

### 3.2. Différentes approches et corrections développées

Les approches de modélisation et corrections sont décrites dans le Chapitre 3 : évaluation et amélioration des performances d'un modèle transnational pour une meilleure évaluation de l'exposition à l'O<sub>3</sub>). Leur description dans cette sous-section sera donc plus succincte et générale.

Les variables météorologiques suivantes, enregistrées quotidiennement par les stations du réseau Météo France, ont été utilisées : température moyenne (°C), hauteur de précipitations (mm), humidité relative moyenne (%), vitesse de vent moyenne à 10m (m.s<sup>-1</sup>), vitesse maximale de vent instantané (m.s<sup>-1</sup>), et la vitesse de vent maximale mesurée sur 10 minutes (m.s<sup>-1</sup>). Chaque station de mesure de la qualité de l'air a ainsi été associée à une station météorologique (la plus proche avec un différentiel d'altitude minimisé) et une maille de modélisation (sur laquelle elle se trouve).

La relation entre, l'écart quotidien entre donnée modélisée et donnée mesurée, et les variables météorologiques, la date et l'altitude (à la station) est analysée avec le modèle MARS :

$$\text{Écart quotidien modèle-mesure} = f(T, RA, RH, WS, \text{altitude, date}) \quad (15)$$

avec T : température moyenne, RA : hauteur de précipitations, RH : humidité relative moyenne, et WS : variables de vitesse de vent.

Le modèle MARS permet ensuite de prédire l'erreur de modélisation du modèle PREV'EST brut. Cette erreur prédite est ensuite retranchée à la concentration modélisée brute afin de la corriger :

$$[O_3]_{\text{corrigée journalière}} = [O_3]_{\text{modélisée journalière}} - \text{écart modèle-mesure journalier prédit} \quad (16)$$

Trois différents groupes de variables, correspondant à 3 corrections, sont utilisés afin d'établir l'intérêt ou non d'intégrer des variables enregistrées par un nombre restreint de station météorologiques (RH, WS) : i) 4-WF (toutes les variables météorologiques et la date), ii) 2WF-elevation (T, RA, date et altitude) et iii) 2-WF (T, RA et date). Ces 3 corrections sont déclinées selon quatre approches : globale, annuelle (un modèle par année), territoriale (un modèle par station), territoriale et annuelle (un modèle par station et par année). Le jeu de données est donc divisé pour construire un modèle spécifique (par année, par station...). Puis les prédictions sont recompilées pour effectuer une évaluation des



performances sur le même nombre d'unités statistiques (nombre d'écart modèle-mesure) et pour pouvoir ainsi comparer les approches et les corrections entre elles (Figure 16).

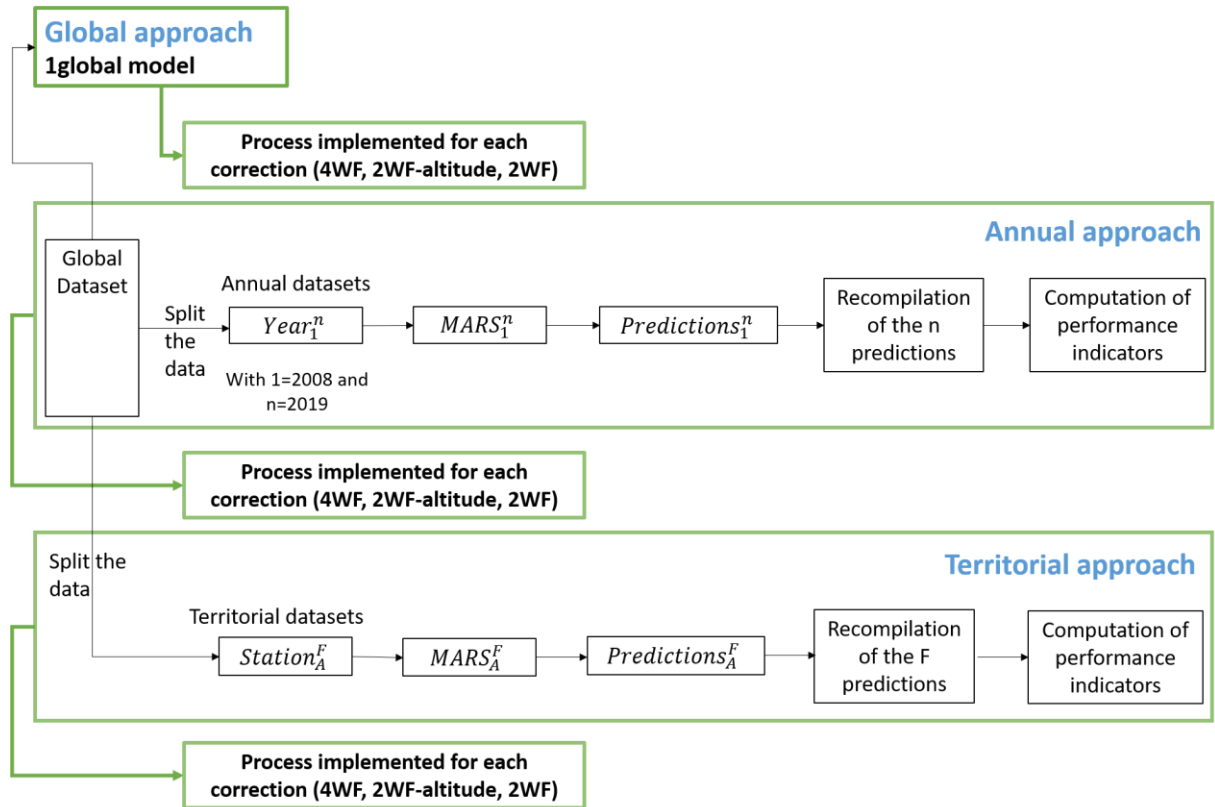


Figure 16 : Processus de modélisation et de recompilation des prédictions pour les approches globale, annuelle et territoriale.

Les performances des corrections sont évaluées sur la base des indicateurs décrit en 2.1 et 2.2. La procédure est également décrite dans la partie 3.2) Evaluation of the performance of the corrections de la partie matériel et méthodes présentée en Chapitre 3.

### 3.3. Validation externe

Une phase de validation externe a été conduite pour évaluer les capacités prédictives des modèles sur un territoire plus large (régional) que celui couvert par les six stations utilisées lors de la phase de validation précédente. Les données de six stations supplémentaires, ayant des périodes d'enregistrement plus courtes (2 à 4 ans), ont été utilisées.

Les modèles construits (3.2) ont été utilisés pour prédire l'erreur de modélisation puis corriger les données modélisées sur la maille de modélisation où se situe chacune des 6 nouvelles stations. Les performances des modèles sur ces stations supplémentaires sont ensuite analysées à l'aide des mêmes indicateurs cités précédemment 2.1 et 2.2.

#### 3.4. Projection des niveaux d'ozone corrigés sur l'aire urbaine de Besançon et sa périphérie

La correction « 2WF-elevation » décrite en partie 3.2 et dont les performances sont présentées dans la partie 3) Results and Discussion du Chapitre 3, a été appliquée à l'aire urbaine de Besançon, ainsi qu'à sa périphérie. Les données météorologiques de la station la plus proche ont été associées à chaque maille de modélisation, tout en veillant à minimiser l'écart d'altitude entre la station météorologique et les mailles de modélisation corrigées. L'altitude moyenne de chaque maille de modélisation est extraite d'un modèle numérique de terrain (BDALTI®, résolution 25m) fourni par l'Institut national de l'information géographique et forestière (IGN). La température moyenne journalière, la hauteur de précipitations journalière, la date ainsi que l'altitude moyenne de la maille sont intégrées en tant que variables explicatives du modèle MARS 2WF-elevation. Une prédiction de l'erreur de modélisation est réalisée à l'aide de ce modèle. Cette erreur prédite est ensuite retranchée à la concentration modélisée brute, pour obtenir une concentration corrigée (équation 16). Ces étapes qui permettent de corriger les concentrations d'O<sub>3</sub> sont présentées en Figure 17.

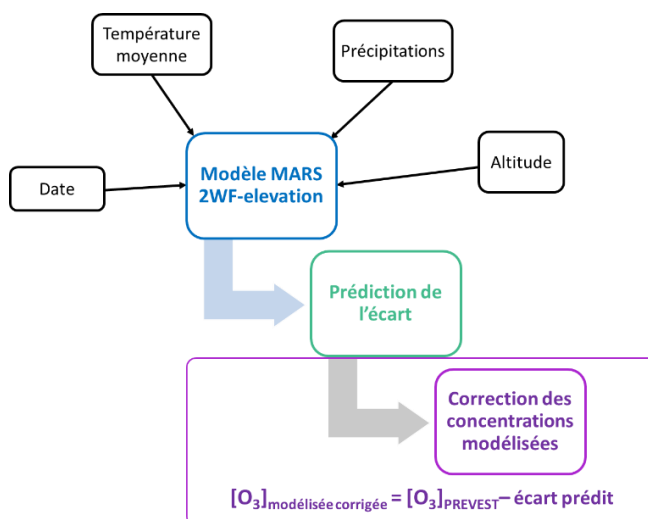


Figure 17: étapes séquentielles de la correction des concentrations d'O<sub>3</sub> à l'aide du modèle MARS "2WF-elevation".

## Identification de dynamiques spatio-temporelles d'ozone et de leurs déterminants

### 1. L'aire urbaine de Besançon

La territoire étudié (Figure 18), est centré sur Besançon, et comprend une diversité élevée de territoires avec un gradient urbain à rural complet (Figure 18). Le pourcentage de surface artificialisée varie de 0% à 64% dans les communes de cette zone. La part de surface de forêts et milieux semi-naturels, ainsi que celle de surfaces agricoles varient quant à elles de 0 à 87 %, et de 6 à 98% respectivement. Les zones rurales autour de Besançon sont fortement influencées par les activités de cette capitale régionale de 117912 habitants.

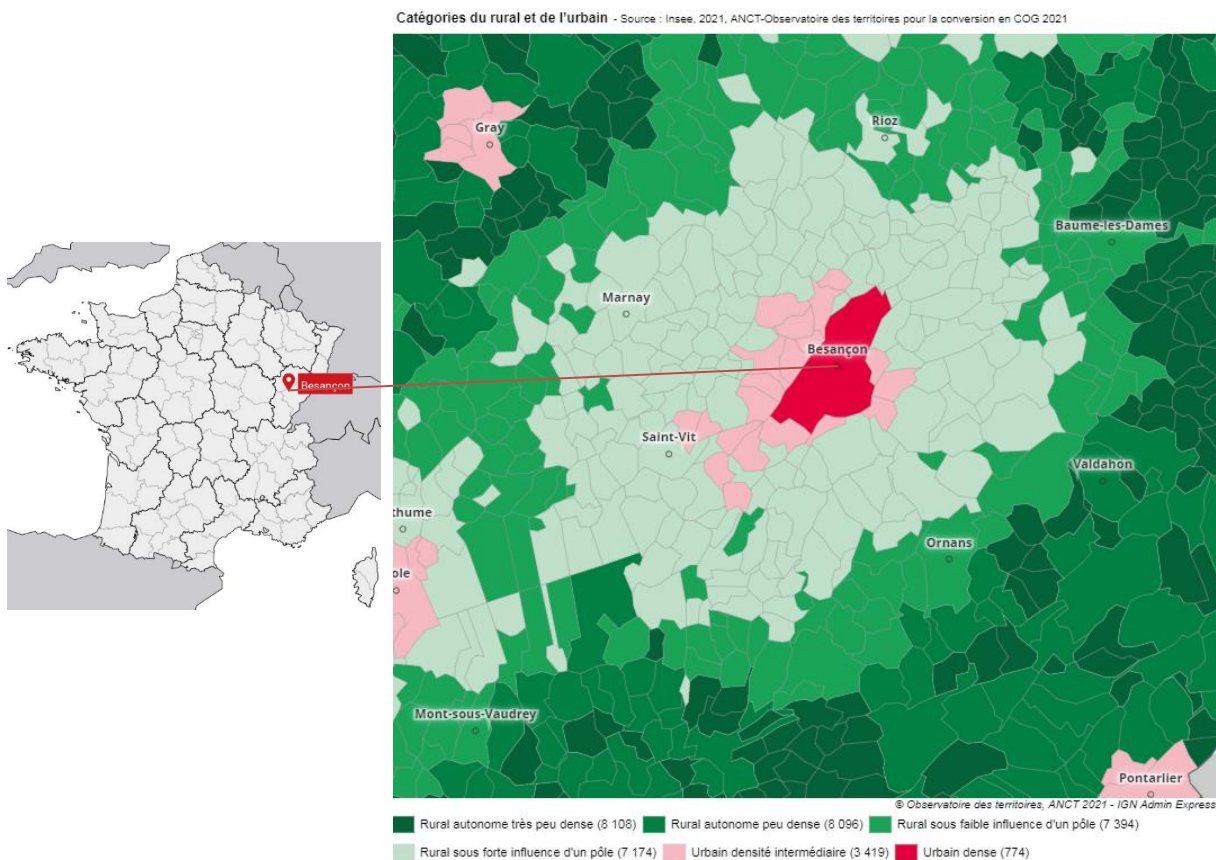


Figure 18: Secteur du Grand Besançon Métropole et ses communes environnantes catégorisées d'urbain à rural. Les effectifs indiqués à côté de chaque typologie sont les effectifs de communes appartenant à cette typologie sur le territoire métropolitain.

La densité de population et sa distribution spatiale (Figure 19), illustre l'influence du Grand Besançon Métropole sur les communes environnantes avec l'établissement d'une zone périurbaine et rurale hébergeant une part non négligeable des personnes travaillant dans l'agglomération de Besançon.

Densité de population, 2018 (habitants au km<sup>2</sup>) - Source : Insee, RP 1968-2018

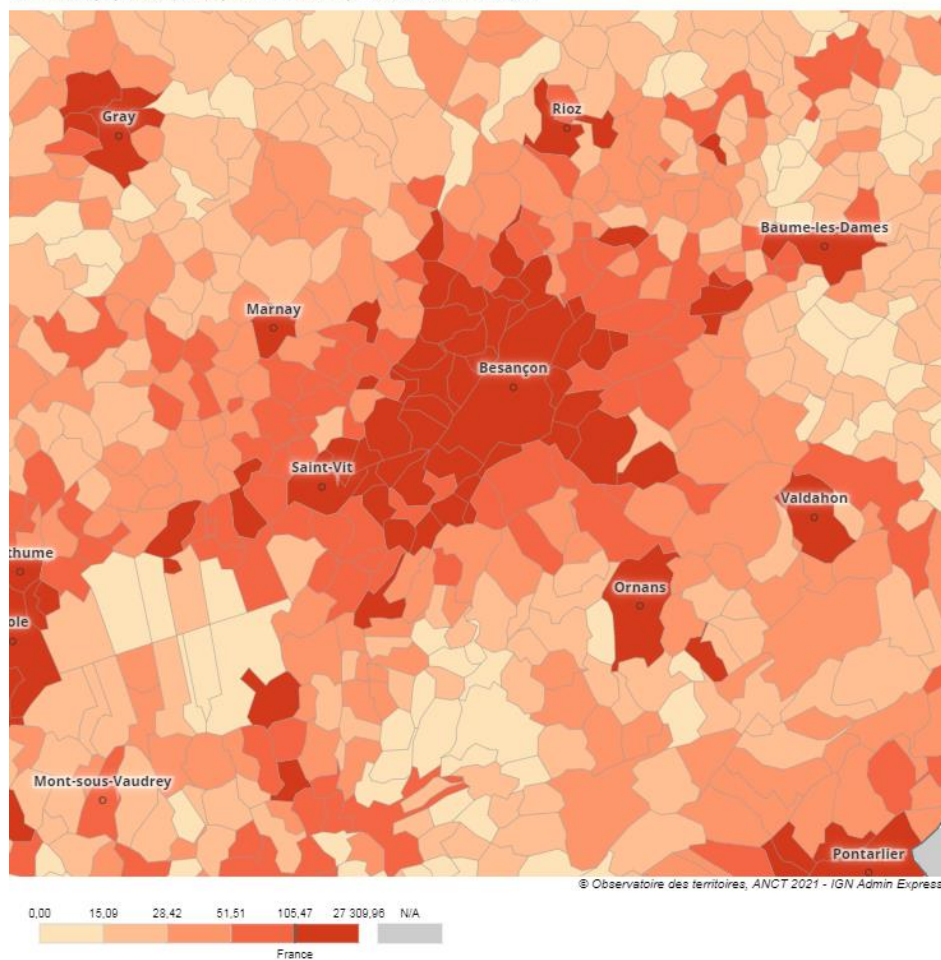


Figure 19: Densité de population au km<sup>2</sup> par commune pour la zone de Besançon Grand Métropole et les communes environnantes.

Les activités anthropiques du département émettent des précurseurs d'O<sub>3</sub> (NO<sub>x</sub>, COVNM) selon des parts variables. Les NO<sub>x</sub> sont principalement émis par le transport routier (70%) puis l'industrie manufacturière et l'agriculture (11% et 7% respectivement) (Figure 20). Les COVNM sont émis en majorité par l'industrie manufacturière (47%) et le secteur résidentiel (41%). Ces inventaires sont réalisés par ATMO BFC (ASQAA régionale) selon la procédure en vigueur à l'échelle nationale (Bréhon et al., 2018).

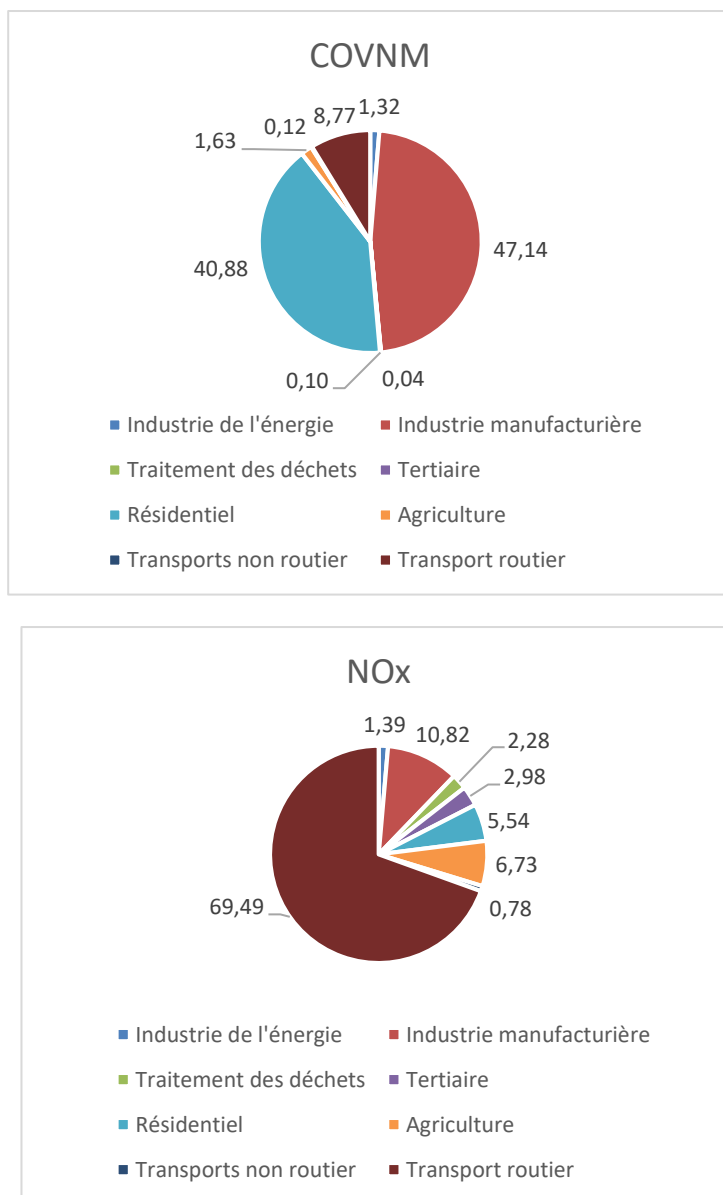


Figure 20: Répartition des émissions de NOx et COVNM par secteurs d'activités agrégées au département du Doubs (2018) selon l'inventaire territorial des émissions effectué par ATMO BFC (Bréhon et al., 2018).

La zone d'étude choisie permet de délimiter l'aire urbaine de Besançon (Urban Atlas 2018) et de s'affranchir de l'influence des autres pôles urbains majeurs du secteur géographique (Figure 21). Le maillage de modélisation de la zone d'étude, défini selon une résolution de 3km x 3 km sur une surface de 1134 km<sup>2</sup>, est visible en Figure 21. Chaque maille de modélisation est associée à une concentration journalière moyenne et maximale d'O<sub>3</sub> pour la période de 2008 à 2019. Au cours de cette période, 3 maillages différents se sont succédés. Un maillage de 4 km par 4 km était appliqué de 2008 à 2011, puis

ensuite de 3 km par 3 km mais selon un calage différent entre les deux périodes 2011- 2014 et 2015-2019. Ce maillage est présenté Figure 21 (3 km par 3 km). Le centroïde de chaque maille du maillage le plus récent (2015-2019) a été associé aux mailles de modélisation plus anciennes afin d’obtenir une série temporelle complète pour chaque maille. Les concentrations utilisées par la suite ont été corrigées à l’aide de la correction dite « 2WF-altitude » (voir 3.2).

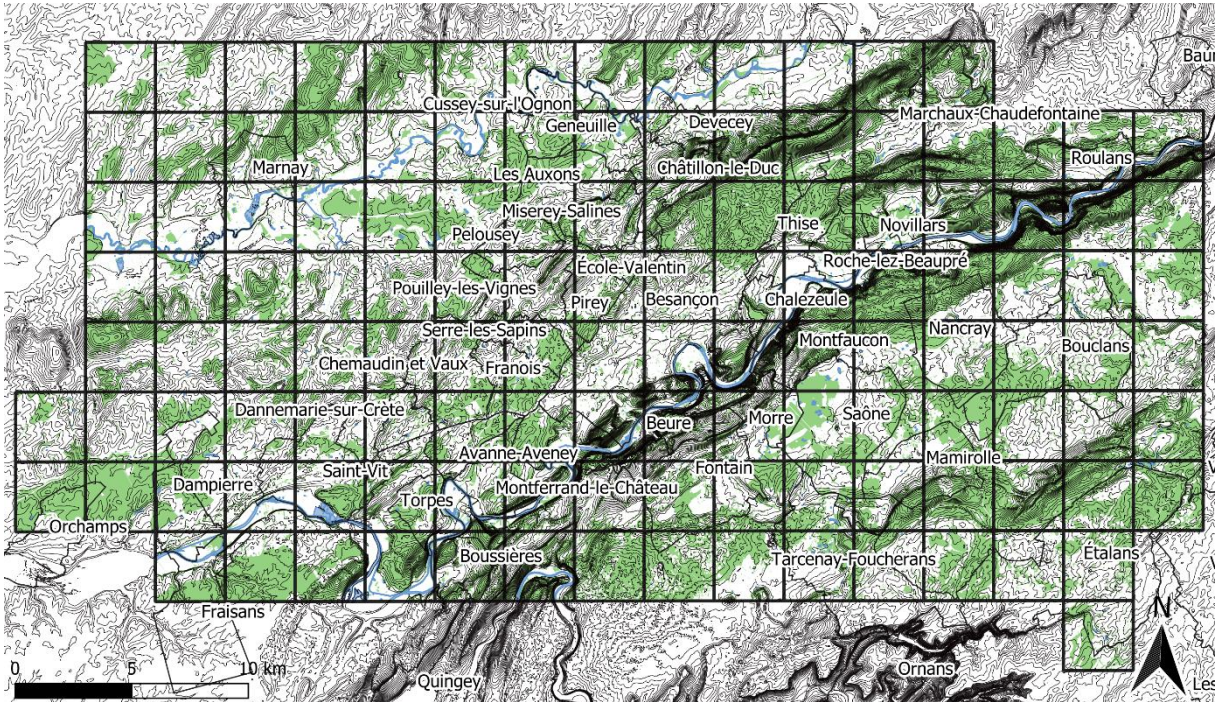


Figure 21: Zone d'étude comprenant le maillage de modélisation utilisé pour la modélisation des concentrations journalières d'ozone.

Les mailles de modélisation sont ensuite caractérisées selon leur altitude moyenne, la densité d’urbanisation et de trafic, ainsi que leur proportion de surface en végétation. La méthodologie de caractérisation des mailles est détaillée en Chapitre 4 - 2.3).

## 2. Classification des séries-temporelles d’ozone

Chaque maille de modélisation illustrée en Figure 21 est associée à une série temporelle de concentrations d’O<sub>3</sub> journalières (moyennes et maximales) sur 12 années (2008-2019). Ces 126 séries temporelles sont ensuite soumises à un processus de classification en deux étapes : 1) calcul de la (dis)similarité entre les séries avec le *Dynamic Time Warping* (DTW) puis 2) l’utilisation d’une classification hiérarchique ascendante sur la base de la matrice de distance calculée à l’aide du DTW.

## 2.1. Dynamic Time Warping

La distance utilisée afin d'évaluer la (dis)similarité des séries temporelle d'ozone est le *Dynamic Time Warping* (DTW). Cette distance dite « élastique » permet de détecter des motifs similaires dans les séries temporelles tout en tolérant un décalage temporel. Ainsi les valeurs ne sont pas nécessairement comparées une à une pour chaque journée, contrairement à la distance Euclidienne (Figure 22). Cette méthode permet d'obtenir de meilleures performances et une meilleure précision que la distance Euclidienne (Giao and Anh, 2016). Le DTW permet de prendre en compte à la fois la proximité en termes de valeur et la similarité dans la forme de la série temporelle.

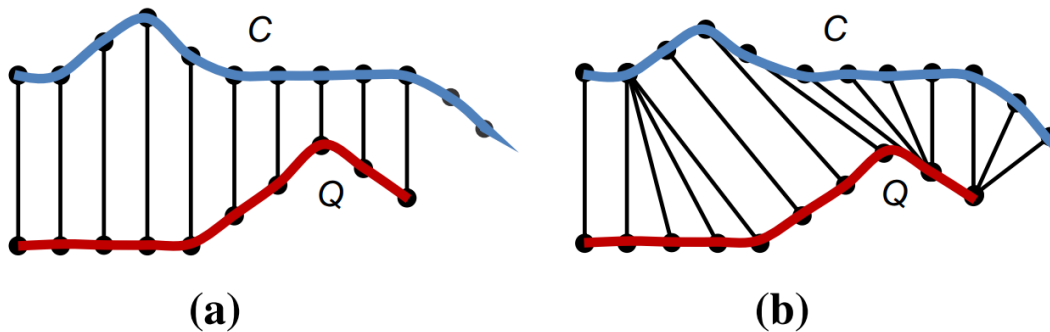


Figure 22: Comparaison de deux séries temporelles avec a) la distance Euclidienne et b) le DTW (Giao and Anh, 2016)

L'alignement entre les séries temporelles  $Q(1,n)$  et  $C(1,m)$  ( $n$  et  $m$  étant les longueurs respectives des séries temporelles  $Q$  et  $C$ ), aussi appelé « warping path » et noté  $W = p_1, p_2, \dots, p_K$  se construit grâce à une matrice appelée LCM (*Local Cost Matrix*) de dimensions  $n \times m$  (Figure 24). Cette matrice est construite selon 3 contraintes globales (Roelofsen, 2018) :

1. contrainte de condition limite : le chemin d'alignement entre les séries doit commencer et se terminer dans les coins de la diagonale de la matrice :

$$w_1 = (1, 1) \text{ et } w_K = (n, m)$$

2. contrainte de continuité : seuls les éléments adjacents dans la matrice peuvent permettre d'avancer dans cette matrice, y compris en diagonale. Si  $w_p = (i, j)$ , alors l'élément :

$$w_{q+1} \text{ est } (i+1, j), (i, j+1) \text{ ou } (i+1, j+1)$$

$$\text{pour } q = 1, \dots, K-1 \text{ et } i = 1, \dots, n-1 \text{ et } j = 1, \dots, m-1$$

3. contrainte de monotonicit  : l' l ment suivant dans la matrice doit  tre espac  de mani re monotone dans le temps (le chemin d'alignement ne peut pas revenir en arri re).

Pour avancer dans la matrice en minimisant le coup d'alignement, l'algorithme suit la proc dure suivante, en consid rant que les  $i^{\text{ me}}$  et  $j^{\text{ me}}$  cellule de la matrice contiennent les valeurs  $C_i$  et  $Q_j$  des s ries C et Q :

$$d(i, j) = d(x_i, y_j) + \min \{d(i - 1, j - 1), d(i - 1, j), d(i, j - 1)\} \quad (17)$$

La distance entre deux points  $i$  et  $j$  des s ries temporelles ( $d(i,j)$ ,  $i$  allant de 1    $n$  et  $j$  allant de 1    $m$ ) est  gale   la diff rence absolue entre les valeurs    $i$  et  $j$  somm e avec la valeur minimale des cellules de la matrice :  $d(i - 1, j - 1)$ ,  $d(i - 1, j)$ ,  $d(i, j - 1)$ . Ce calcul de la matrice (LCM) est illustr  en Figure 23.

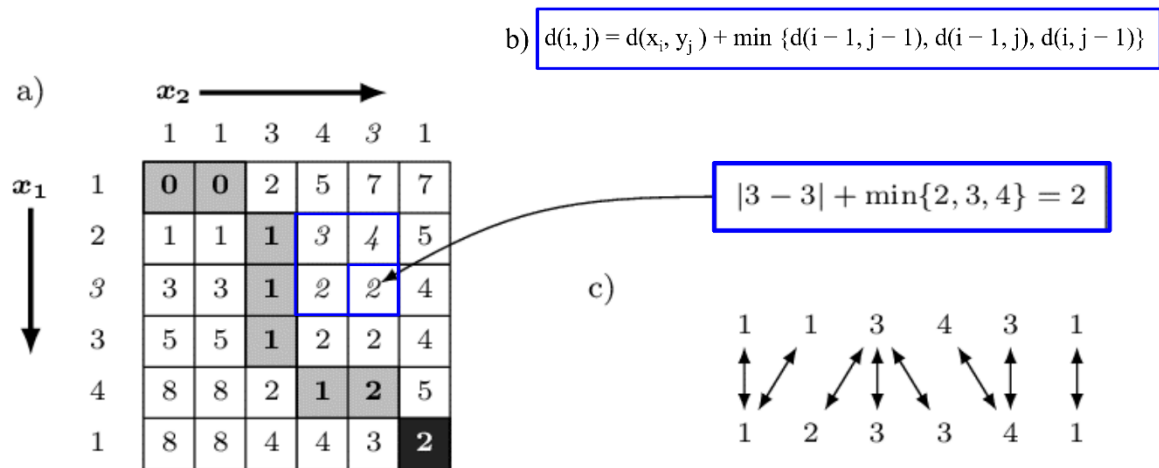


Figure 23: Calcul de la matrice LCM calcul e pour aligner deux s ries temporelles  $x_1$  et  $x_2$  (Tomašev et al., 2015).

Ensuite, l'alignement entre les s ries temporelles est r alis  en choisissant le chemin pris dans la matrice qui minimise le c t d'alignement :

$$DTW(C, Q) = \min (\sum_1^K d(w_K)) \quad (18)$$

Une contrainte suppl mentaire peut  tre appliqu e   l'algorithme afin de le contraindre   n'explorer qu'une partie de la matrice. Cette contrainte additionnelle, appel e bande de Sakoe-Chiba (Sakoe and



Chiba, 1978), permet d'éviter une association d'un point à une période temporelle excessivement longue sur la seconde série temporelle (C. Ratanamahatana and Keogh, 2004) (Figure 24).

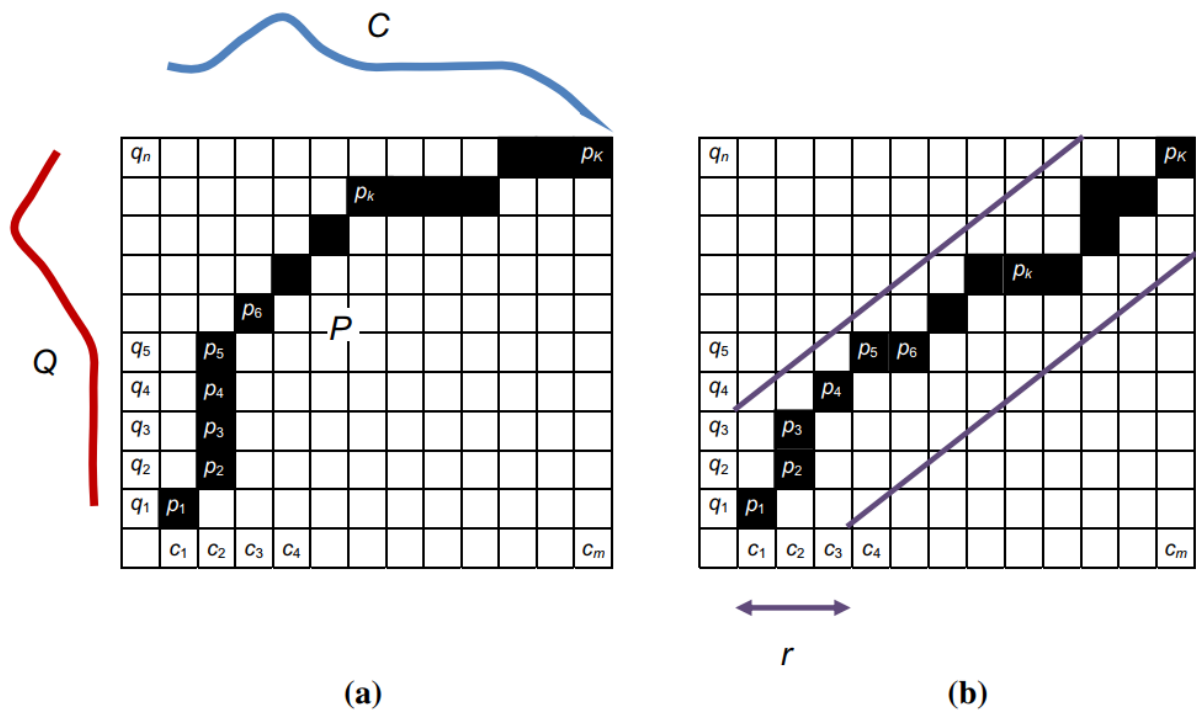


Figure 24: a. Pour aligner les séries  $C$  et  $Q$ , un "warping path"  $P$ , illustré avec des carrés pleins, est construit. b. La contrainte de la bande de Sakoe-Chiba de largeur  $r$  est utilisé comme contrainte globale pour limiter la portée de  $P$  (Giao and Anh, 2016).

## 2.2. Groupement hiérarchique

Les séries-temporelles sont ensuite regroupées selon un processus de classification hiérarchique ascendante (lien complet) sur la base de la matrice de distance construite en 2.1. Les entités (séries temporelles) sont tout d'abord considérées comme toutes distinctes, puis sont réunies dans des groupes de taille croissante de manière ascendante. Le résultat de cette classification est illustré par un dendrogramme. La méthode utilisée pour identifier le nombre optimal de groupes est le WSS (*Within Sum of Squares*) total. Des détails complémentaires sont apportés en Chapitre 4.

### 3. Identification et caractérisation des phénomènes de vagues de chaleur

Les vagues de chaleur sont inventoriées selon la définition donnée par Météo France (Soubeyrou et al., 2016), déjà évoquée en partie 2 du Chapitre 1. Cette définition est basée sur des seuils de température moyenne journalière correspondant aux percentiles 95 (seuil d'interruption, nommé Sint), 97,5 (seuil de début de vague, nommé Sdeb) et 99,5 (seuil de détection, nommé Spic) de l'indicateur thermique sur la période de référence (1981-2010). Nous utiliserons dans le cadre de ces travaux des seuils départementaux (Pascal et al., 2019) afin d'identifier des épisodes de manière plus fine et spécifique au vu des températures observées sur le territoire considéré. L'inventaire est réalisé à l'échelle de la station météorologique de Besançon.

Une vague de chaleur est identifiée si le seuil Spic est dépassé. Le début de vague de chaleur est défini par la date de dépassement du seuil Sdeb, et seules les vagues durant au moins trois jours sont prises en compte. La vague est interrompue si la température descend en dessous du seuil Sint (même une seule journée), ou en dessous du seuil Sdeb pendant au moins 3 jours consécutifs (Ouzeau et al., 2016; Soubeyrou et al., 2016).

Chaque vague identifiée est ensuite caractérisée par trois indicateurs : durée, intensité et sévérité (Figure 25). L'intensité correspond à la température moyenne journalière maximale observée au cours de la vague. La sévérité est « la somme des écarts positifs entre températures moyennes quotidiennes pendant la vague de chaleur et le seuil de démarrage Sdeb divisée par la différence entre Spic et Sdeb » (Ouzeau et al., 2016).

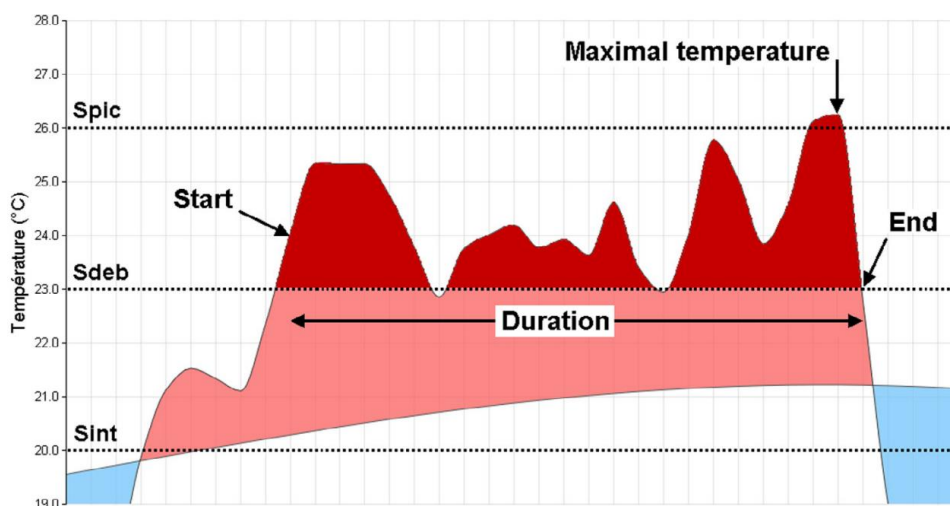


Figure 25: Caractérisation d'une vague de chaleur basée sur la température moyenne journalière en France (exemple de série temporelle du 30 juin au 5 août 1983) : durée (start – end), température maximale et intensité globale (aire rouge sur le graphique). Les températures supérieures (inférieures) à la ligne climatologique (période de référence 1981-2010) sont représentées par une zone rose (bleue). D'après Ouzeau et al. (2016).

## Distribution de la population et affectation des niveaux d'exposition

### 1. Distribution de la population dans les bâtiments

Les recensements de la population décrivant les effectifs de population (globale et par tranches d'âges) sont délivrés par l'INSEE à l'échelle de l'ilot regroupé pour l'information statistique (IRIS<sup>10</sup>). Ces recensements sont mis à jour annuellement. La position des bâtiments d'habitation est extraite de la BDTOPO fournit par l'IGN (édition 2008 et 2018 selon les années considérées). Les bâtiments appartenant uniquement à la catégorie « bâti indifférencié » ont été sélectionnés pour ne conserver que les bâtiments susceptibles d'héberger la population. Chaque bâtiment a ensuite été associé à l'IRIS sur lequel son centroïde était localisé. Les effectifs de population ont été ensuite distribués en multipliant

<sup>10</sup> brique de base en matière de diffusion de données infra-communales, qui regroupe environ 2000 habitants. L'IRIS doit respecter des critères géographiques et démographiques et avoir des contours identifiables sans ambiguïté et stables dans le temps. Les communes d'au moins 10 000 habitants et une forte proportion des communes de 5 000 à 10 000 habitants sont découpées en IRIS. Ce découpage constitue une partition de leur territoire. Par extension, afin de couvrir l'ensemble du territoire, on assimile à un IRIS chacune des communes non découpées en IRIS. ([INSEE, 2016](#))

l'effectif total de l'IRIS par le rapport de la surface du dit bâtiment sur la surface total de bâti de l'IRIS contenue dans la zone de modélisation de l'O<sub>3</sub>.

## 2. Affectation des niveaux d'exposition au bâtiment

Comme décrit dans la sous-partie précédente (1) chaque bâtiment a été associé à un IRIS mais également à une maille de modélisation selon la position du centroïde du bâtiment. Cette procédure est effectuée annuellement. Chaque bâtiment est donc associé annuellement à un effectif de population et une maille de modélisation. L'identifiant de la maille de modélisation permet ensuite par requête dans les bases de données d'affecter un niveau d'exposition d'O<sub>3</sub> journalier (moyen et maximum) à chaque bâtiment et à la population qu'il abrite.

## Chapitre 3 : évaluation et amélioration des performances d'un modèle transnational pour une meilleure évaluation de l'exposition à l'O<sub>3</sub>

### Improvement of downscaled ozone concentrations from the transnational scale to the kilometric scale: Need, interest and new insights

Article publié dans : Environmental Research. 2022 Jul; 210: 112947. doi: 10.1016/j.envres.2022.112947

#### **Résumé :**

##### **Contexte**

La pollution à l'ozone (O<sub>3</sub>) engendre de nombreux effets néfastes sur la santé humaine et celle des écosystèmes. Afin de quantifier ces impacts, il est nécessaire, dans un premier temps, d'évaluer avec précision l'exposition des organismes et des populations. Pour travailler à cette échelle populationnelle, des modèles de prédiction des niveaux de pollution atmosphérique sont utilisés. Malgré le développement et l'amélioration rapide de ces outils, la fiabilité des valeurs projetées à échelle locale peut être faible et nécessiter une correction.

L'objectif de l'étude décrite dans ce chapitre se décline en trois sous-objectifs : i) évaluer si les concentrations d'O<sub>3</sub> estimées par une plateforme de modélisation transnationale à l'échelle kilométrique (9 km<sup>2</sup>) doivent être corrigées, ii) proposer une potentielle correction de ces concentrations d'ozone et iii) évaluer l'efficacité et l'applicabilité d'une telle correction.

##### **Matériel et méthodes**

La méthodologie utilisée dans cette étude est déclinée en trois phases pour répondre aux 3 sous-objectifs précédents. Tout d'abord, les performances de la plateforme de modélisation (PREV'EST) sont évaluées en confrontant les données journalières modélisées de 2008 à 2019 et des données de mesures. Des indicateurs de performances généraux (biais moyen, *mean absolute error* (MAE), *root mean square error* (RMSE), coefficient de corrélation de Pearson) puis à la station (*target plots*) sont calculés. Dans un second temps, différentes approches (générale, annuelle, par station, par année et par station) ainsi que différentes corrections sont déclinées. Les différentes corrections sont issues de la modélisation de l'erreur de prédiction de PREV'EST en fonction de différentes sélections de variables explicatives (température moyenne, humidité relative moyenne, hauteur de précipitations, vitesse de vent, altitude et date) à l'aide du modèle MARS (*Multivariate Adaptive Regression Splines*). Les performances des corrections proposées sont évaluées grâce aux indicateurs présentés précédemment. Enfin une validation externe à l'aide de données de 6 stations supplémentaires est effectuée.

##### **Résultats**

Les données de modélisations brutes présentaient un biais moyen considérable (surestimation moyenne de 14,67 µg/m<sup>3</sup> mais différentielle entre les stations), et un manque de précision, se traduisant par un lissage des courbes d'exposition. Les contrastes d'exposition, et en particulier, les pics d'ozone étaient mal représentés. Les corrections proposées permettaient toutes d'améliorer les performances moyennes des données modélisées. Toutefois, la correction 2WF-elevation (prenant en compte température moyenne, la hauteur de précipitations, l'altitude et la date) apparaissait comme la plus performante, flexible et facilement applicable sur de vastes territoires.

##### **Conclusion**

L'intégration de données locales a permis de restituer une précision locale à un modèle délivrant des prédictions à l'échelle transnationale. La méthodologie décrite dans cette étude est efficace, validée à l'aide de données externes et reproductible dans d'autres régions. Cette amélioration permet notamment de détecter les pics d'ozone qui sont des éléments d'exposition aiguë, et qui tendent à s'amplifier dans un contexte de changement climatique.

## **Title page**

Original Research Article

### **Title**

Improvement of downscaled ozone concentrations from the transnational scale to the kilometeric scale: need, interest and new insights

### **Authors**

Honorine GAUTHIER-MANUEL<sup>a,b\*</sup>, Frédéric MAUNY<sup>a,b</sup>, Mathieu BOILLEAUT<sup>c</sup>, Marie RISTORI<sup>c</sup>,  
Sophie PUJOL<sup>a,b</sup>, Florent VASBIEN<sup>d</sup>, Anne-Laure PARMENTIER<sup>a,b</sup>, Nadine BERNARD<sup>a,e</sup>

### **Affiliation of the authors**

a: UMR 6249, Laboratoire Chrono-environnement, Université de Bourgogne Franche-Comté, CNRS,  
25000 Besançon, France

b: Unité de méthodologie en recherche clinique, épidémiologie et santé publique (uMETH), Inserm  
CIC 1431, CHU, 25030 Besançon cedex, France

c: ATMO Bourgogne-Franche-Comté, 25000 Besançon, France

d: ATMO Grand Est, 67300 Schiltigheim, France

e: Centre National de la Recherche Scientifique, UMR 6049, Laboratoire ThéMA,  
Université de Bourgogne Franche-Comté, 25000 Besançon, France

\*Corresponding author.

E-mail address: [honorine.gauthier-manuel@univ-fcomte.fr](mailto:honorine.gauthier-manuel@univ-fcomte.fr) (H. Gauthier-Manuel)

**Abstract:**

**Background:** Ground-level ozone is a major public health issue worldwide. An accurate assessment of ozone exposure is necessary. Modeling tools have been developed to tackle this issue in large areas. However, these models could present inaccuracies at the local scale.

**Objectives:** The objective of this study was i) to assess whether O<sub>3</sub> concentrations estimated by transnational modeling at the kilometric scale (9 km<sup>2</sup>) could be improved, ii) to propose a potential correction of these downscaled ozone concentrations and iii) to evaluate the efficiency and applicability of such a correction.

**Method:** The present work was carried out in three phases. First, the performance of a transnational modeling platform (PREV'EST) was assessed at 6 geographic points by comparison with data from 6 air quality monitoring stations. Performance indicators were used for this purpose (MBE (mean bias error), MAE (mean absolute error), RMSE (root mean square error), r (Pearson correlation coefficient), and target plots). Second, several corrections were developed using MARS (multivariate adaptive regression splines) and integrating different sets of variables (mean temperature, relative humidity, rainfall amount, wind speed, elevation, and date). Their performance was evaluated. Third, external validation of the corrections was conducted using the data from six additional air quality monitoring stations.

**Results:** The uncorrected PREV'EST model presented a lack of exactitude and precision. These concentrations did not reproduce the interday variability of the measurements, leading to a lack of temporal contrast in exposure data. For the best performance enhancement, the correction applied improved MBE, MAE, RMSE and r from 14.67, 19.23, 23.18 and 0.67 to 0.00, 8.00, 10.19 and 0.91, respectively. External validation confirmed the efficiency of the corrections at the regional scale.

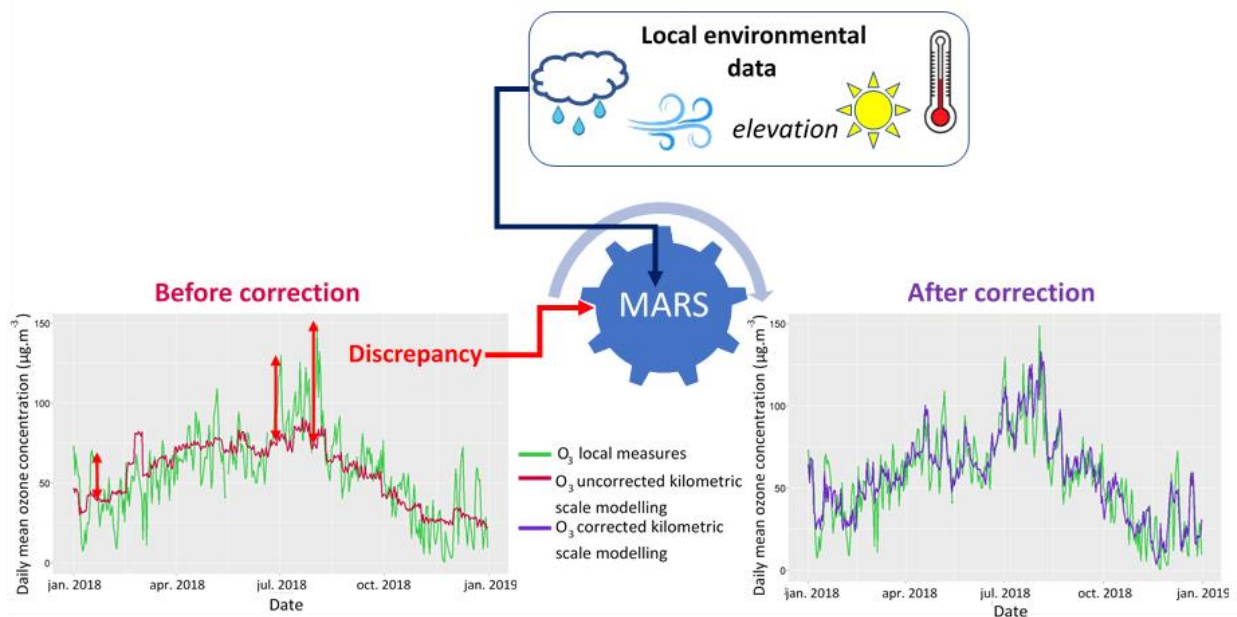
**Conclusions:** We propose a validated and efficient methodology integrating local environmental variables. The methodology is adaptable according to the context, needs and data available.

**Keywords: Ozone, modeling, improvement, performance, MARS model**

**Highlights:**

- 1) The transnational model needed improvements to accurately assess O<sub>3</sub> exposure.
- 2) Local environmental variables and the MARS model were used to correct O<sub>3</sub> values.
- 3) We proposed a validated method that improved both exactitude and precision.
- 4) This method was adaptable according to the context, needs and data available.

**Graphical abstract:**



**Study Funding**

This work was supported by the agglomeration community of Grand Besançon Métropole (PhD scholarship, 2019) and the “Pôle fédératif de recherche et de formation en santé publique Bourgogne-Franche-Comté” (2019 and 2020 research project calls).



## 1) Introduction:

Outdoor air pollution is a major global public health issue involving both primary and secondary pollutants (Lelieveld., 2015; Cohen et al., 2017). Ground-level ozone ( $O_3$ ) is a secondary air pollutant which causes adverse human health effects (Brunekreef and Holgate, 2002; Zhang et al., 2019) that can range from eye and nose irritation to pulmonary and cardiovascular system disorders (Henrotin et al., 2010; Marx et al., 2019; Nuvolone et al., 2018; Zhang et al., 2019). Therefore, it is necessary to assess its spread in the environment and human exposure to this pollutant. Tropospheric ozone is particularly challenging to characterize in expology because it highly fluctuates in space and time. From 2000 to 2017, ozone rose in urbanized areas (with increasing annual ozone-related premature deaths) and decreased in rural areas in the EU-28 (Sicard et al., 2021). However, rural areas still remain more exposed. In response to climate change, ozone exposure is prone to be exacerbated in polluted regions (Lu et al., 2019). Indeed, this photooxidant is formed by photochemical reactions initiated in the presence of solar radiation and primary pollutant precursors ( $NO_x$ , volatile organic compounds (VOCs)) (Royal Society (Great Britain), 2008; Wang et al., 2017b). Its distribution depends on numerous environmental parameters, such as solar radiation (and, indirectly, temperature), wind speed, relative humidity, and elevation (Ramos et al., 2018; Royal Society, 2008). It also depends on ozone formation regime, strongly influenced by the VOCs-to- $NO_x$  ratio (Li et al., 2021). Ozone concentration evaluation is complex due to its extreme variability in space and time (diurnal cycle and seasonality) (Castell-Balaguer et al., 2012).

Networks of air quality monitoring stations provide an assessment of population exposure and monitor regulatory threshold exceedances. These data are nevertheless insufficient because they can offer limited spatial and temporal coverage (Chen et al., 2014; Garcia et al., 2010). Co-kriging approaches can be applied to interpolate background stations data (Sicard et al., 2013) and provide a spatially continuous estimation of ozone levels. Approaches combining local regression and residuals kriging can reach high spatial resolution (250 x 250m) (Sicard et al., 2016). However, these approaches are tributary to the number and spatial distribution of air quality monitoring stations (Wang et al., 2016).

Land-Use Regression (LUR) and chemical transport models (CTMs) are the most commonly used modeling approaches in the long-term assessment of population exposure (Joly et al., 2021; Wang et al., 2016). CTMs such as CHIMERE (Mailler et al., 2017b; Menut et al., 2013), CMAQ or MOCAGE, provide grid meshes that commonly range from 4 to 12km (U.S. EPA, 2018). However, the grid cell size is proportional to the spatial coverage of the model. CTMs covering Europe generally handle resolutions from 12 to 25 km, those covering a single country usually use 4 to 10 km resolutions and more rarely can reach ~1 km on some complex terrains regions (Schaap et al., 2015). CHIMERE is one of the most popular CTM (Thunis et al., 2016) and exploits the outputs of a meteorological model (Weather Research and Forecasting (WRF) model, for example) by modeling platforms delivering European, national or regional air quality forecasts (Honore et al., 2005; Riviere et al., 2019; Rouil et al., 2009). PREV'AIR is a modeling platform providing air quality forecasts at European and national scales (France) and integrates WRF and CHIMERE models. The data stemming from these forecasting platforms are used in exposure risk assessment (Riviere et al., 2019). However, regional air quality modeling tools can present biases and inaccuracies (Steven Porter et al., 2015). Improvements could be needed prior to using these data for: i) exposure assessment (Crooks and Özkaynak, 2014) and ii) better depict both spatial and temporal variability of ozone.

Modeling platforms, as PREV'AIR, require large computational capacity and extensive input data (Braun et al., 2017; Wang et al., 2016), but they constantly generate predictions to prevent and assess population exposure. We propose a correction of the outputs of a CTM predicted at the finest spatial scale, already available, to develop a tool which is more accessible, faster-processing, with lower computational cost than *ex nihilo* implementing LUR or CTM, and with better performance.

The objective of this study was i) to assess whether downscaled O<sub>3</sub> concentrations estimated by transnational modeling (initial resolution 45 km x 45 km, i.e., 2025 km<sup>2</sup>) to the kilometric scale (9 km<sup>2</sup>) should be improved, ii) to propose a potential correction of these downscaled ozone concentrations and iii) to evaluate the efficiency and applicability of such a correction.

## 2) Materials and Methods:

The present work was conducted in three phases. First, the performance of the transnational model was assessed by comparison with air quality monitoring data at 6 geographic points. Second, several corrections were added to the model, and their performance were re-evaluated. Third, external validation of the corrections was conducted using the data from six additional air quality monitoring stations.

### 2.1 Assessment of the performance of the uncorrected PREV'EST model

#### 2.1.1. Study site and data sources

The mesoscale meteorological model, WRF, is already used in numerous air quality modeling approaches (Thunis et al., 2016). The CHIMERE model, an Eulerian chemistry-transport model, integrated WRF outputs, emission inventories and chemical boundary conditions to forecast air quality (Mailler et al., 2017b; Menut et al., 2013). The PREV'EST modeling platform used in this study is a regional/transnational platform developed and managed by ATMO Grand Est (regional air quality monitoring association). Its functioning is analogous to PREV'AIR, a modeling platform used at the national and European scales (Honore et al., 2005; Riviere et al., 2019; Rouil et al., 2009), and integrates the WRF and CHIMERE models. PREV'EST integrates Europe PREVAIR's forecasts into the forecasting mode as boundary conditions to improve the modeling by limiting the edge effect. The PREV'EST and PREV'AIR modeling platforms are both subjected to daily performance assessments, comparing modeled concentrations with measurements (bias, percentage of errors lower than a certain level and correlation) (Honore et al., 2005; Rouil et al., 2009). ATMO Grand Est also applies corrections on yearly aggregated data.

The PREV'EST modeling platform generated outputs on three grid meshes of different sizes: 45 km x 45 km (over Europe), 15 km x 15 km (over France and adjacent countries), and 3 km x 3 km (eastern France and the boundary area of France, Germany, Switzerland and Luxembourg). The finest grid mesh used in this study covered 274,428 km<sup>2</sup>, or a 194 x 154 pixel grid of 9 km<sup>2</sup>. The modeling and nesting processes were conducted by ATMO Grand Est, who delivered the outputs of the finest grid mesh (3 km x 3 km). We conducted the optimization processes (described below) on this finest grid mesh.

CHIMERE’s parametrization was adapted to this area thanks to the expertise and field knowledge of the team who developed and managed this platform (ATMO Grand Est). Each pixel represented a basic spatial unit associated with a daily average modeled ozone concentration (called  $[O_3]_{\text{modelled}}$ ).

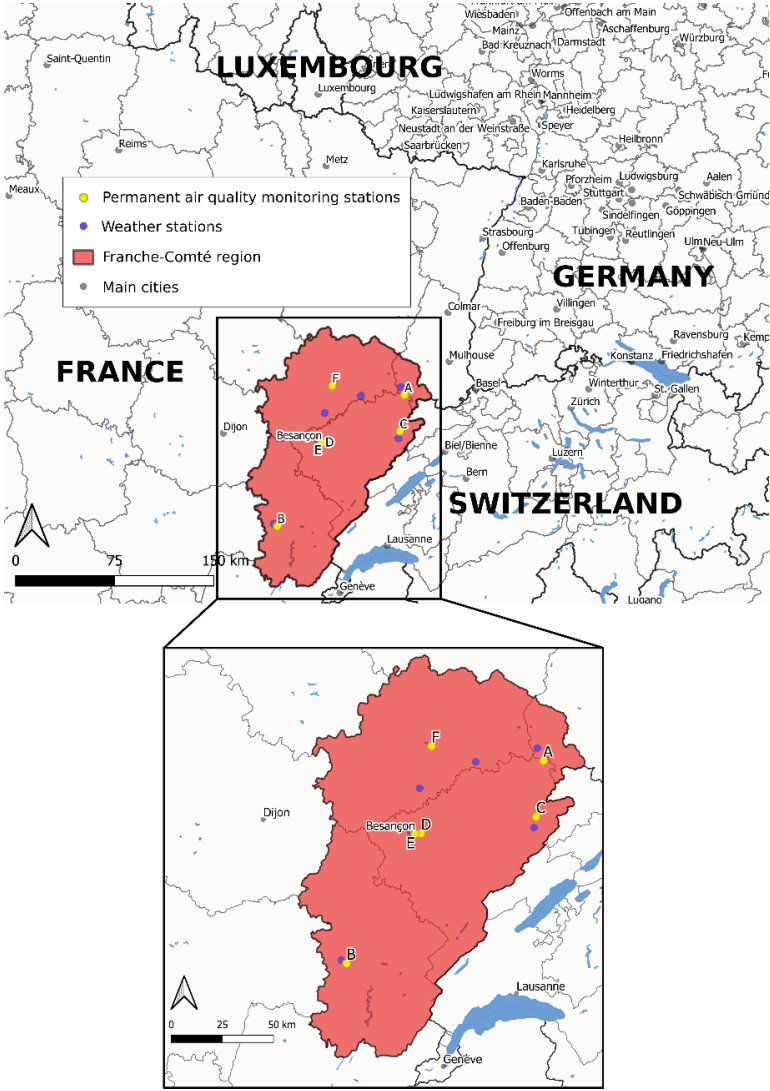


Figure 26: PREV’EST modeling area (for the 3 km x 3 km grid mesh, on the top map) and location of the air quality monitoring stations (ATMO BFC) and weather stations (Météo-France).

Six air quality monitoring stations (three urban, two suburban and one rural station) measured daily average  $O_3$  concentrations (called  $[O_3]_{\text{measured}}$ ) in the Franche-Comté region (France) from 2008 to 2019 (Figure 26). Each monitoring station was related to the pixel of the PREV’EST model on which it is located. These daily measured concentrations were considered the gold standard for the  $O_3$

concentrations at the local place of the monitoring station (Thunis et al., 2016). ATMO Bourgogne-Franche-Comté (ATMO BFC) provided the measured data.

For each of the six stations, the daily discrepancy between modeled and measured concentrations was quantified by the differences between daily average O<sub>3</sub> concentrations:

$$\text{Daily Discrepancy} = [\text{O}_3]_{\text{modeled}} - [\text{O}_3]_{\text{measured}} \quad (1)$$

### 2.1.2. Performance indicators

The following performance indicators were retained and are described in Appendix 1: mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE) and Pearson correlation coefficient (r). MBE quantifies the mean model-measure discrepancy and a potential overestimation (MBE>0) or underestimation (MBE<0) of the model (Kato, 2016; Martin, 2011). MAE is a precision indicator measuring the mean absolute model-measure discrepancy. MAE is higher when the model-measure discrepancies are large and scattered. RMSE is also a precision indicator (Abdulelah Al-Sudani et al., 2019; Lesmeister, 2019; Sekulic and Kowalski, 1992; Yilmaz et al., 2018) that gives more weight than MAE to high model-measure discrepancies (Bouthevillain and Mathis, 1995). RMSE is higher for large discrepancies (regardless of the sign, positive or negative) with high variability. The Pearson correlation coefficient (r) assesses the linear correlation between modeled and measured concentrations.

In addition to these average performance indicators, target plots were used to represent the performance of the model at the station scale. The localization of the station in the circle inside the target plot indicates respect to these quality criteria: the vertical axis displays the normalized bias (BIAS<sub>Norm</sub>), while the horizontal axis displays the CRMSE<sub>Norm</sub> (normalized centered RMSE) value, both normalized by the measurement uncertainty associated with the pollutant. An extended description of both the quality criterion and target plots is provided in Appendix 1. Target plots were built using R software ('dartle' package).

## 2.2 Development and performance assessment of the corrections

The statistical associations between model-measure discrepancies and environmental variables (meteorological variables and elevation) were analyzed using MARS (multivariate adaptive regression splines) models (Friedman, 1991; Friedman and Roosen, 1995).

### 2.2.1. *Statistical relationship between the discrepancy and environmental variables*

Daily meteorological data were collected from 2008 to 2019 and provided by the French meteorological institute (Météo-France). The meteorological parameters were mean temperature (Celsius degrees (°C)), rainfall amount (mm), mean relative humidity (expressed in percentage (%)), average wind speed at a height of 10 meters ( $\text{m}\cdot\text{s}^{-1}$ ), live maximum wind speed ( $\text{m}\cdot\text{s}^{-1}$ ), and maximum wind speed measured over 10 minutes ( $\text{m}\cdot\text{s}^{-1}$ ). The data from the nearest weather station were assigned to each air quality monitoring station, minimizing the difference in elevation between the meteorological station and air quality monitoring stations. The elevation of each air quality monitoring station was recorded by ATMO-BFC.

The relationships between observed discrepancies and environmental variables were analyzed using MARS models (Eq. (2)). MARS models generate a piecewise linear model by finding knots (or “turning points”) in the relationship between Y (the variable we want to predict) and each of the explanatory variables (Figures A.1 and A.2).

These models were used as the prediction function of the daily discrepancy according to environmental variables (Eq. (2)). The predicted daily discrepancy was then used to correct the raw (called “uncorrected”) data. The performance of the corrected  $\text{O}_3$  concentrations was assessed by the process described in Section 2.1.2.

$$\text{Daily observed discrepancy} = f(T^\circ, \text{RA}, \text{RH}, \text{WS}, \text{elevation}, \text{date}) \quad (2)$$

where  $T^\circ$  is the mean temperature, RA is the rainfall amount, RH is the mean relative humidity, and WS is the wind speed variables.

Different sets of explanatory variables were iteratively used to assess the relevance of integrating some variables less frequently recorded (*e.g.*, RH and WS) (Table 2): four weather factors (4-WF correction,

every meteorological variable and the date), two weather factors and elevation (2-WF elevation correction, mean temperature, rainfall amount, elevation and the date) and two weather factors (2-WF correction, mean temperature, rainfall amount and the date).

The model-measure discrepancy predicted by MARS was then used to calculate corrected modeled O<sub>3</sub> concentrations as follows:

$$\text{Daily [O}_3\text{]}_{\text{corrected}} = \text{daily [O}_3\text{]}_{\text{modelled}} - \text{predicted daily discrepancy} \quad (3)$$

A positive discrepancy reflected an overestimation by the uncorrected PREV'EST model, and conversely, a negative discrepancy corresponded to an underestimation.

The efficiency of each correction was assessed using the performance indicators described in Section 2.1.2. (MBE, MAE, RMSE, r, and target plots) on the daily [O<sub>3</sub>]<sub>corrected</sub> values.

### 2.2.2. *Modeling approaches*

Four modeling approaches of increasing complexity were developed (Table 2). The global approach included the data of the six stations during a 12-year period. The annual approach included the 6 stations together but was separately conducted for each of the 12 years (one independent model per year, n = 12). The territorial approach included the data of the 12 years but was separately conducted for each of the six stations (one independent model per station, n = 6). In the territorial and annual approach, one independent model was performed per year and per station (n = 6 x 12) (Table 2). The predictions of all these models were recompiled to calculate the performance indicators and to keep these modeling approaches comparable (Figure A.3). Indeed, although we split the dataset (per year and/or per station), we kept the same number of statistical units (22,640 “discrepancies” used to train the models).

Table 2: MARS modeling approaches and corrections

| Type of approach                    | Kind of correction applied <sup>2</sup> | Number of models per approach | Number of models per kind of correction |
|-------------------------------------|---|-------------------------------|---|
| Global                              | 4WF                                     | 3                             | 1                                       |
|                                     | 2WF-elevation                           |                               | 1                                       |
|                                     | 2WF                                     |                               | 1                                       |
| Annual                              | 4WF                                     | 36                            | 12                                      |
|                                     | 2WF-elevation                           |                               | 12                                      |
|                                     | 2WF                                     |                               | 12                                      |
| Territorial <sup>1</sup>            | 4WF                                     | 12                            | 6                                       |
|                                     | 2WF                                     |                               | 6                                       |
| Territorial and annual <sup>1</sup> | 4WF                                     | 130 <sup>3</sup>              | 65 <sup>3</sup>                         |
|                                     | 2WF                                     |                               | 65 <sup>3</sup>                         |
| Total                               |   | 181                           |   |

<sup>1</sup>The territorial and, the territorial and annual approaches did not include a 2WF-elevation correction because the elevation was steady (only one station considered per model).

<sup>2</sup> 4WF: correction including 4 weather factors (Temperature (T°), Relative Humidity (RH), Rainfall Amounts (RA), Wind speed parameters (WS)) and the date; 2WF – elevation: correction including 2 weather factors (T°, RA), the date and the elevation; 2WF: correction including 2 weather factors (T°, RA), and the date.

<sup>3</sup> Theoretically we should have 72 models but the station E opened only in 2013 and the station B opened only in 2010.

### 2.3.) External validation of the corrections

External validation was conducted to ensure that this methodology could be applied to a regional territory and not only to the six stations considered to train the model. Six additional air quality monitoring stations (location presented in Figure A.4) were used in this phase. These stations have temporarily measured O<sub>3</sub> concentrations. However, the additional stations were subject to the same quality approach as the six previous stations. The processes described in Sections 2.2.1 and 2.2.2 were applied for the periods when the additional stations were active. Finally, the efficiency of each correction (and for each approach) was assessed using the performance indicators described in Section 2.1.2. The



quality of the correction was assessed by the difference between the corrected O<sub>3</sub> concentration and measured O<sub>3</sub> concentration (Eq. (4)).

$$([\text{O}_3]_{\text{modeled}} - \text{predicted daily discrepancy using the function in Eq. (2)}) - [\text{O}_3]_{\text{measured}} \quad (4)$$

The corrected O<sub>3</sub> concentration was defined by:

$$([\text{O}_3]_{\text{modeled}} - \text{predicted discrepancy using the function in Eq. (2)}) \quad (5)$$

All statistical treatments were performed with R software (version 4.0.3), and the MARS models were implemented with the ‘earth’ package.

### 3) Results and Discussion

#### 3.1) Evaluation of uncorrected PREV’EST model performance

The discrepancy between the uncorrected kilometric scale modeling and the local measurements was substantial (+14.67 μg m<sup>-3</sup> (SD = 17.9)). This discrepancy was nonlinearly structured according to some environmental parameters, particularly temperature (Figure A.5). Table 3 presents the average performance indicators reflecting the general performance averaged among all stations. The average performance of the uncorrected PREV’EST model (Table 3) suggested a need for improvement of both exactitude (MBE = +14.67 μg m<sup>-3</sup>) and precision (RMSE = 23.18 μg m<sup>-3</sup>, MAE = 19.23 μg m<sup>-3</sup>) to be used in exposure assessment at the local scale. According to the literature, the national model PREV’AIR had a very comparable performance to the study PREV’EST model (Meleux and Ung, 2020).

Table 3: Average performance indicators of the different approaches and corrections

| Kind of correction applied      | MBE <sup>a</sup> | MAE <sup>a</sup> | RMSE <sup>a</sup> | r <sup>b</sup> |
|---------------------------------|------------------|------------------|-------------------|----------------|
| Uncorrected PREV’EST            | 14.67            | 19.23            | 23.18             | 0.67           |
| Global 4WF correction           | 0.00             | 10.78            | 13.64             | 0.84           |
| Global 2WF-elevation correction | 0.00             | 11.37            | 14.32             | 0.81           |
| Global 2WF correction           | 0.03             | 12.21            | 15.34             | 0.78           |
| Annual 4WF correction           | 0.02             | 9.40             | 11.92             | 0.87           |

|                                       |      |       |       |      |
|---------------------------------------|------|-------|-------|------|
| Annual 2WF-elevation correction       | 0.01 | 10.04 | 12.67 | 0.85 |
| Annual 2WF correction                 | 0.04 | 10.81 | 13.60 | 0.83 |
| Territorial 4WF correction            | 0.00 | 9.41  | 11.96 | 0.88 |
| Territorial 2WF correction            | 0.00 | 10.92 | 13.76 | 0.83 |
| Territorial and annual 4WF correction | 0.01 | 8.00  | 10.19 | 0.91 |
| Territorial and annual 2WF correction | 0.02 | 9.52  | 12.05 | 0.87 |

a Mean bias error (MBE), mean absolute error (MAE) and root mean square error (RMSE) should be minimized to reflect better a performance.

b r: Pearson correlation coefficient should be maximized to reach better a performance.

The performance of the uncorrected PREV'EST model was also separately estimated for the six air quality monitoring stations (Figure 27). The stations are not centered on the horizontal axis (Figure 27). They are also positioned in the left part of the target plot, which indicates a lack of correlation between uncorrected PREV'EST data and measurements (Janssen et al., 2017b). The PREV'EST model also overestimated the ozone concentrations, as seen previously (Table 3, MBE = 14.67  $\mu\text{g m}^{-3}$ ). The stations are differently distributed on the vertical axis (bias axis) (Figure 27): the PREV'EST model overestimated the global data but more (Station C) or less (Stations D and E) according to the station.

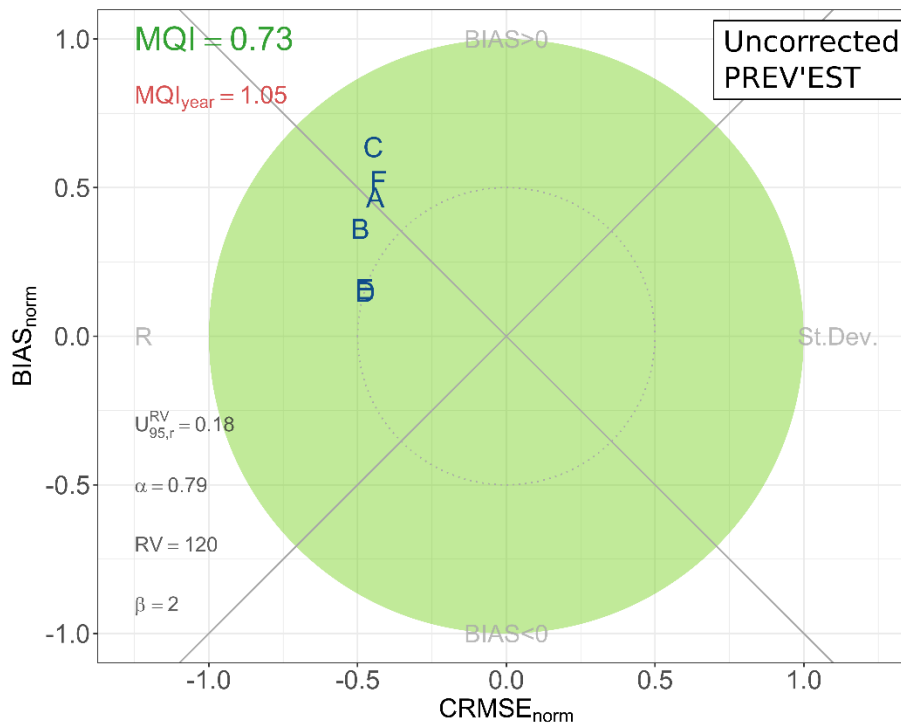


Figure 27: Target plots of the uncorrected PREV'EST modeling.

Different letters represent the different air quality monitoring stations.

According to the current guidelines, the performance of the uncorrected PREV'EST model could be qualified as good. However, this model, already well calibrated, presented faults: a lack of precision and exactitude (a tendency to overestimate differentially according to the station).

### 3.2) Evaluation of the performance of the corrections

The average performance indicators presented in Table 3 were calculated by compiling the predictions of the models according to the different corrections proposed. The mean bias error (MBE) was drastically reduced regardless of the correction applied (initial value: 14.67; range of MBE after correction: 0.00 to 0.04  $\mu\text{g m}^{-3}$ ) (Table 3). This indicated an improvement in the exactitude of the modeled ozone concentrations. The MBE values varied very little between the corrections and were thus not discriminant to identifying the best correction.

According to the other indicators (MAE, RMSE, and r), the 4-WF correction could be considered the most efficient, followed by the 2-WF elevation correction and then the 2-WF correction, regardless of the approach considered. The refinement of the type of approach led to a gradual improvement in

performance indicators (*i.e.*, decrease of MBE, MAE and RMSE and increase of  $r$ ). The territorial and annual approach coupled to the 4-WF correction could be seen as the most efficient (MAE =  $8.00 \mu\text{g m}^{-3}$ , RMSE =  $10.19 \mu\text{g m}^{-3}$ , and  $r = 0.91$ ). Taking into account the specificities of the site of each station provided a more precise correction.

Figure 28 allowed us to explore the potential difference in the performance of the corrections between the six stations. We observed an improvement in the performance of the modeled ozone concentrations (Figure 28) compared with uncorrected PREV'EST (Figure 27). All stations displayed a null bias and were closer to the target center after correction. A differential bias persisted with the 4-WF correction in the global and annual approaches (Figure 28). The differential bias could be damaging when the exposure assessment is simultaneously conducted in different areas. The 2-WF elevation correction allowed for the suppression of this differential bias. The integration of the elevation in the global model with the 2-WF elevation correction offers the opportunity to consider the local characteristics of the sites of each station. Moreover, elevation is well known in the literature to be a key factor in ozone distribution (Brönnimann et al., 2000; Chevalier et al., 2007; Ezcurra et al., 2013). A study conducted by Martin (2010) revealed some lack of precision in a comparable model (a regional application of the PREV'AIR modeling platform, AIRES). Martin (2010) suggested refining the spatial resolution of the model to better consider topography. This lack of local precision could be offset by the integration of local meteorological data and elevation. Moreover, the elevation is easily available over large territories thanks to digital elevation models. Other modeling methods, as LUR or (co)-kriging, integrate the digital elevation models, land use, population density, etc., and provide finer resolution. These methods present some advantages: the ability to characterize small spatial scale variability, ease to use, and require far fewer computational resources than CTMs (Braun et al., 2017; Wang et al., 2016). However, they use internal data to validate the model through a cross-validation process (Ma et al., 2020) which can overestimate the predictive ability of the model (Wang et al., 2013). LUR focuses on spatial variation (Ren et al., 2020) and can fail to depict the temporal variation of ozone (Ma et al., 2020). CTMs rely on topography, emissions, meteorology, and physicochemical processes. They provide a well captured spatio-temporal ozone variability and a large spatial coverage (Wang et al., 2016). Nevertheless, they

present coarser grid mesh and require large computational capacity and extensive input data (Braun et al., 2017), reducing their accessibility.

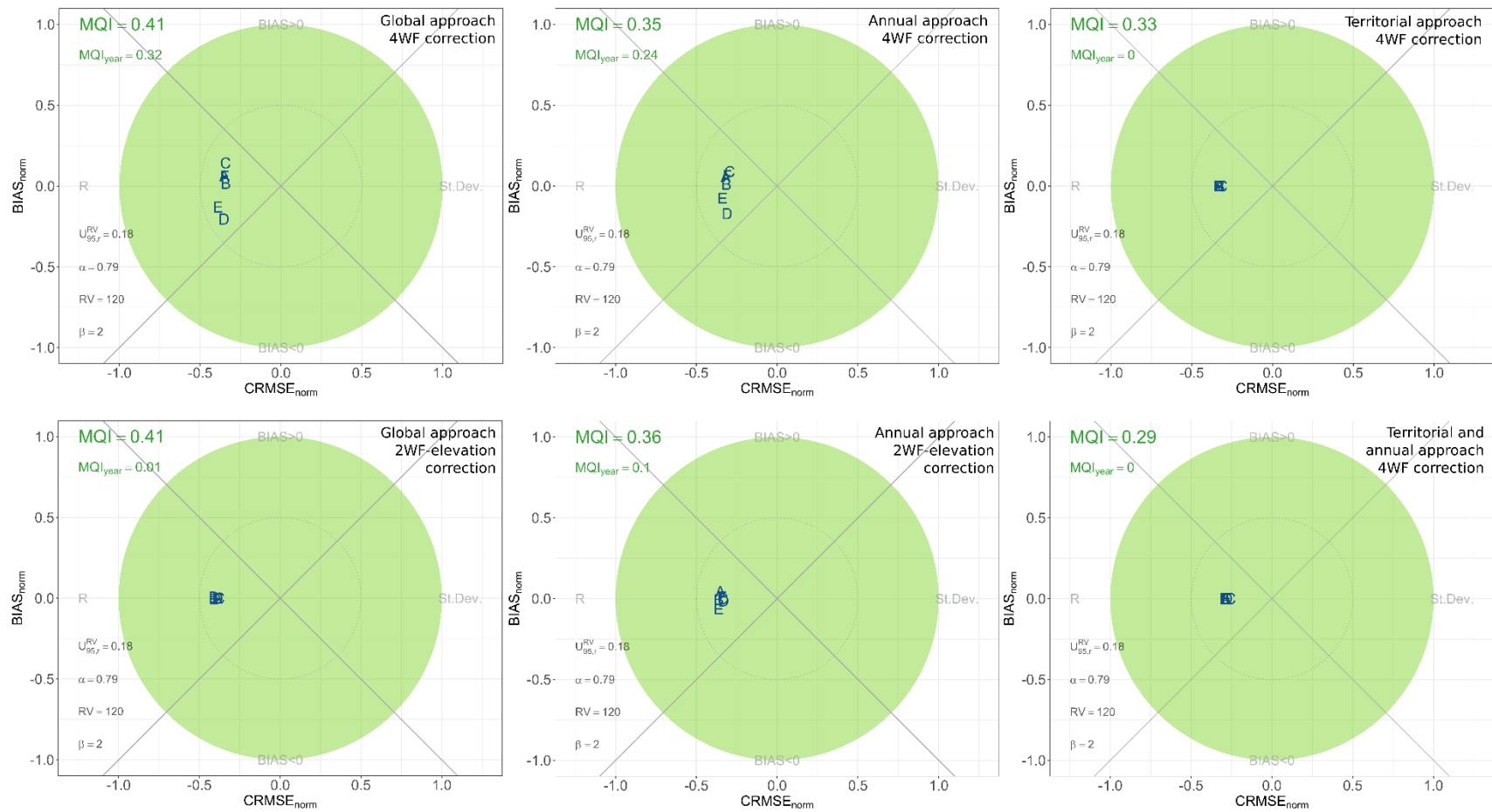


Figure 28: Target plots of the uncorrected PREV'EST modeling, global approach 4-WF and 2-WF elevation corrections, annual approach 4-WF and 2-WF elevation corrections, territorial approach 4-WF correction and territorial and annual approach 4-WF correction target

Different letters represent the different air quality monitoring stations.

In terms of application, if we consider a correction dedicated to numerous stations or a regional projection, the 2-WF elevation correction in the annual approach should be favored. This correction suppressed the differential bias between stations and is easily applicable. Some meteorological variables (such as wind speed) were not recorded by most of the weather stations. Our 2WF-elevation correction used common and few variables.

The territorial and territorial and annual approaches had the best performance (Figure 28). From an operational point of view, these approaches could be used to locally correct the exposure data for a few stations and their surrounding areas. The implementation of the territorial and annual approach is very efficient but time-consuming. The territorial approach could be chosen instead of the territorial and annual approach to reach a very satisfactory performance faster. In any case, the territorial (and annual) approach should be applied with caution because the area around a station on which the model is reliable is unknown and hardly definable. External validation will allow the determination of whether these territorial corrections could be extended to other areas while remaining reliable or not.

Figure 29 shows the time series of the concentrations measured, uncorrected and corrected, using the annual 2-WF elevation correction. These corrected concentrations provided better interday variability than uncorrected PREV'EST data. The latter did not reproduce the interday variability of the measurements and smoothed the curve of exposure assessment data. This led to a lack of contrast in exposure data and could make the identification of a harmful effect of acute O<sub>3</sub> exposure difficult. The day-to-day contrast is essential to identify ozone peaks (Ren et al., 2020) which appear as a key element in the occurrence of health effects (Di et al., 2017). Such a phenomenon has already been observed in air quality modeling (Hogrefe et al., 2004). Hogrefe and his collaborators (2004) suggested that a regional scale resolution was insufficient to well represent ozone variations in the short term (daily, interday, and intraday variations). The methodology proposed in this study offers a potential solution to this lack of precision and variability in the short-term assessment of ozone levels at the regional scale.

Considering the spatial variability of ozone, by definition, a grid cell displays a homogenous exposure level. The assessed level is the same for all individuals living in the same grid cell. This could lead to

under or overestimation at the individual scale. However, we could apply the methodology presented in this article to models with finer spatial resolution, to better capture the spatial variability of ozone.

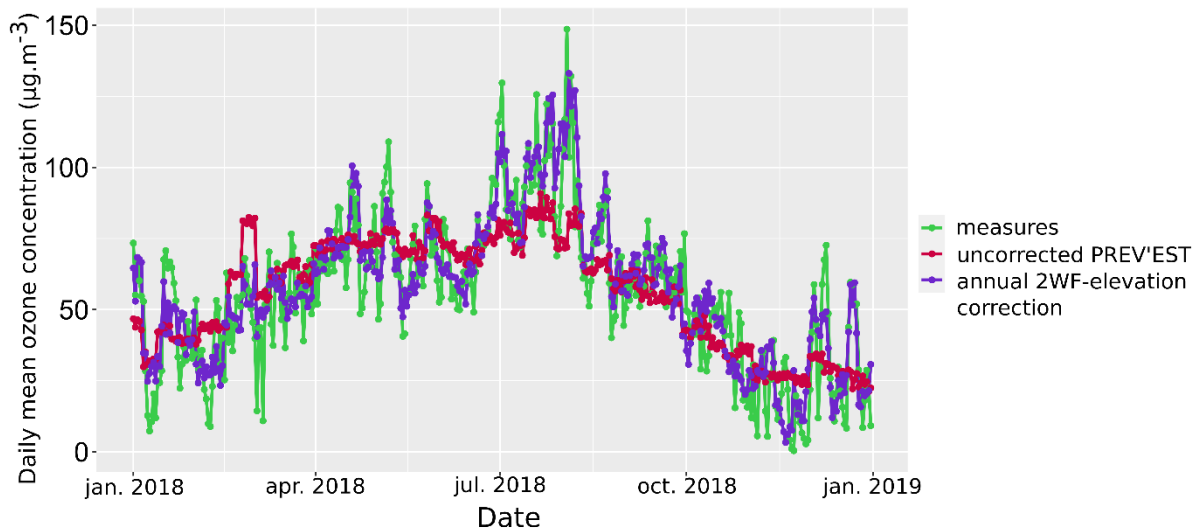


Figure 29: Daily mean ozone concentrations measured, modeled by uncorrected PREV'EST modeling and corrected by the annual 2-WF elevation correction, during 2018 at Station A.

### 3.3. External validation

In this validation, 14,846 statistical units (number of discrepancies modeled) were implied. The corrections of ozone concentrations led to an improvement of every average indicator (Table 4). The initial value of MBE, MAE, RMSE and r were 14.83, 17.63, 21.35 and 0.74, respectively; and their ranges after correction, according to the correction, were 0.04 to 4.87, 9.48 to 12.76, 12.11 to 16.11 and 0.75 to 0.85, respectively (Table 4). All average indicators converged to identify the annual approach with the 4-WF correction as the most efficient. However, the territorial and territorial and annual approaches were only applied to the models of two stations (A and D among the six available stations) because of the locations of these stations (Figure A.4).



Table 4: Average performance indicators calculated for the external validation.

|                                       | <b>MBE</b> | <b>MAE</b> | <b>RMSE</b> | <b>r</b> |
|---------------------------------------|------------|------------|-------------|----------|
| Uncorrected PREV'EST                  | 14.83      | 17.63      | 21.35       | 0.74     |
| Global 4WF correction                 | 0.54       | 10.24      | 13.00       | 0.85     |
| Global 2WF-elevation correction       | -0.96      | 11.64      | 14.70       | 0.77     |
| Global 2WF correction                 | -0.23      | 11.55      | 14.56       | 0.78     |
| Annual 4WF correction                 | 0.51       | 9.48       | 12.11       | 0.85     |
| Annual 2WF-elevation correction       | -2.01      | 10.74      | 13.59       | 0.81     |
| Annual 2WF correction                 | 0.04       | 10.76      | 13.61       | 0.81     |
| Territorial 4WF correction            | 4.87       | 12.26      | 15.61       | 0.79     |
| Territorial 2WF correction            | 3.67       | 12.76      | 16.11       | 0.75     |
| Territorial and annual 4WF correction | 3.45       | 11.11      | 14.29       | 0.81     |
| Territorial and annual 2WF correction | 3.42       | 12.31      | 15.67       | 0.76     |

Number of statistical units (number of discrepancies predicted): 14 846

The prediction was performed on data of the air quality monitoring station which were not integrated into the MARS models for its training (additional stations, n = 6).

As already observed in the internal validation, the target plots presented in Figure 30 displayed a differential bias among stations in the 4-WF correction, which was reduced by the 2-WF elevation correction. This latter correction could be seen as the most efficient and flexible in this external validation.

This external validation confirmed the main points of the observed performance of the different corrections assessed. If the user desires to apply a correction with fewer spatial constraints, the annual approach will give the best performance. The correction to be favored could be the 4-WF correction because it presented the best average performance (Table 4). However, the 2-WF elevation correction presented a lower differential bias among stations (Figure 30) and very satisfactory average performance (Table 4). This correction was more flexible and required fewer variables that were commonly recorded

by most of the weather stations. The 2-WF correction showed an equivalent average performance but lower performance at the station scale.

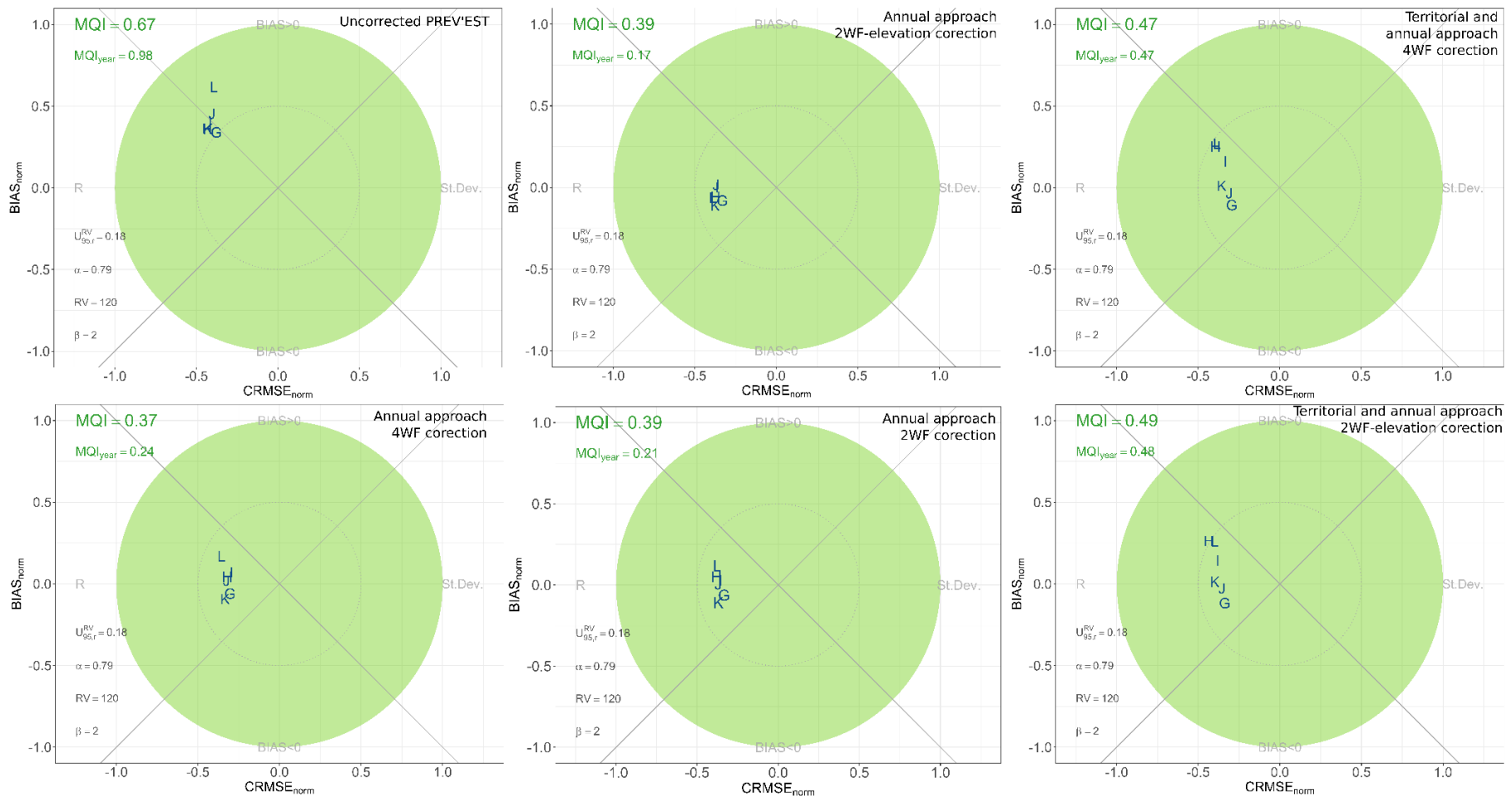


Figure 30: External validation: uncorrected PREV'EST modeling, annual approach 4-WF correction, annual approach 2-WF elevation correction, and annual and territorial approach 4-WF and 2-WF correction target plots.

Different letters represent the different air quality monitoring stations.

#### **4 Strengths of the methodology and comparison with existing methodologies**

The main strengths of this methodology are its efficiency in improving the precision and exactitude of exposure data, its ability to recover a good contrast in these data, and its transposability in other regions. Furthermore, this methodology is adaptable according to the context, needs and data available. The quantification of the benefit of the correction especially using external data represents a complete and powerful validation process. Indeed, kriging methods or LUR models commonly use cross-validation which can overestimate the predictive ability of the model (Wang et al., 2013). Our approach allows also to better estimate the ozone concentration peaks, which are of great importance in daily exposure assessment and could be underestimated by other modeling approaches.

MARS models have been used instead of other classical models (such as generalized additive models (GAMs) and artificial neural networks (ANNs)) because they present a better performance, the possibility to introduce interactions between variables and the ability to capture complex nonlinear relationships between variables (Bordignon et al., 2002; Leathwick et al., 2006; Martin, 2011). MARS models are simple to apply and interpret, and require a short calculation time.

MARS models have already been efficaciously used to locally model atmospheric pollutant concentrations (García Nieto et al., 2015; García Nieto and Álvarez Antón, 2014; Roy et al., 2018). To our knowledge, they have never been used to adjust a pre-existing and well-established air quality model. Guillas et al. (2008) used model diagnostic and correction (MDC) integrating variables associated with wind speed, precipitation amounts and the diurnal cycle to predict the discrepancy and correct ozone concentrations. However, this approach only used linear regressions. This was not suitable for our data because of the nonlinear relationship between our discrepancy and our environmental variables.

#### **5. Conclusion:**

We identified a lack of exactitude and precision of transnational modeling downscaled at the kilometric scale compared with local measurements. We proposed a validated and efficient methodology,

integrating local environmental variables, to remedy this and demonstrated its broader applicability. This methodology can be transferable to other areas, applicable to variable ranges of concentration and periods, and requires only a few environmental variables. This procedure is also flexible according to the context, needs and data available. Our approach allows us to improve the ozone exposure assessment. This is a more accessible tool, faster-processing, lower-cost than computing LUR or CTM, externally validated, and which better captures ozone peaks that can be associated with health effects (Di et al., 2017). The quantification of the benefit of the correction, especially using external data, constitutes a complete and powerful validation process. In the context of climate change, such an improvement will be of paramount importance for impact assessment of ozone exposure, including on wildlife (Sanderfoot and Holloway, 2017) or vegetation (Proietti et al., 2021).

### **Acknowledgments**

The authors wish to thank everyone who contributed to the development of the project and would like to acknowledge Météo-France, especially Mr. Vermot-Desroches, for providing the meteorological data.

### **References**

- Abdulelah Al-Sudani, Z., Salih, S.Q., sharafati, A., Yaseen, Z.M., 2019. Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *J. Hydrol.* 573, 1–12. <https://doi.org/10.1016/j.jhydrol.2019.03.004>
- Bordignon, S., Gaetan, C., Lisi, F., 2002. Nonlinear models for ground-level ozone forecasting. *Stat. Methods Appl.* 11, 227–245. <https://doi.org/10.1007/BF02511489>
- Bouthevillain, C., Mathis, A., 1995. Prévisions : mesures, erreurs et principaux résultats. *Econ. Stat.* 285, 89–100. <https://doi.org/10.3406/estat.1995.5982>
- Braun, S., Achermann, B., De Marco, A., Pleijel, H., Karlsson, P.E., Rihm, B., Schindler, C., Paoletti, E., 2017. Epidemiological analysis of ozone and nitrogen impacts on vegetation – Critical evaluation and recommendations. *Science of The Total Environment* 603–604, 785–792. <https://doi.org/10.1016/j.scitotenv.2017.02.225>

- Brönnimann, S., Schuepbach, E., Zanis, P., Buchmann, B., Wanner, H., 2000. A climatology of regional background ozone at different elevations in Switzerland (1992–1998). *Atmos. Environ.*, Sixth Scientific Conference of the International Global Atmospheric 34, 5191–5198. [https://doi.org/10.1016/S1352-2310\(00\)00193-X](https://doi.org/10.1016/S1352-2310(00)00193-X)
- Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. *The Lancet* 360, 1233–1242. [https://doi.org/10.1016/S0140-6736\(02\)11274-8](https://doi.org/10.1016/S0140-6736(02)11274-8)
- Castell-Balaguer, N., Téllez, L., Mantilla, E., 2012. Daily, seasonal and monthly variations in ozone levels recorded at the Turia river basin in Valencia (Eastern Spain). *Environ. Sci. Pollut. Res.* 19, 3461–3480. <https://doi.org/10.1007/s11356-012-0881-5>
- Chen, G., Li, J., Ying, Q., Sherman, S., Perkins, N., Sundaram, R., Mendola, P., 2014. Evaluation of observation-fused regional air quality model results for population air pollution exposure estimation. *Sci. Total Environ.* 485–486, 563–574. <https://doi.org/10.1016/j.scitotenv.2014.03.107>
- Chevalier, A., Gheusi, F., Delmas, R., Ordóñez, C., Sarrat, C., Zbinden, R., Thouret, V., Athier, G., Cousin, J.-M., 2007. Influence of altitude on ozone levels and variability in the lower troposphere: a ground-based study for western Europe over the period 2001–2004. *Atmospheric Chem. Phys.* 7, 4326.
- Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C.A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C.J.L., Forouzanfar, M.H., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet* 389, 1907–1918. [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6)
- Crooks, J.L., Özkaynak, H., 2014. Simultaneous statistical bias correction of multiple PM<sub>2.5</sub> species from a regional photochemical grid model. *Atmos. Environ.* 95, 126–141. <https://doi.org/10.1016/j.atmosenv.2014.06.024>

- Di, Q., Dai, L., Wang, Y., Zanobetti, A., Choirat, C., Schwartz, J.D., Dominici, F., 2017. Association of Short-term Exposure to Air Pollution With Mortality in Older Adults. *JAMA* 318, 2446–2456. <https://doi.org/10.1001/jama.2017.17923>
- Ezcurra, A., Benech, B., Echelecou, A., Santamaría, J.M., Herrero, I., Zulueta, E., 2013. Influence of local air flow regimes on the ozone content of two Pyrenean valleys. *Atmos. Environ.* 74, 367–377. <https://doi.org/10.1016/j.atmosenv.2013.03.051>
- Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. *Ann. Stat.* 19, 1–67. <https://doi.org/10.1214/aos/1176347963>
- Friedman, J.H., Roosen, C.B., 1995. An introduction to multivariate adaptive regression splines. *Stat. Methods Med. Res.* 4, 197–217. <https://doi.org/10.1177/096228029500400303>
- García Nieto, P.J., Álvarez Antón, J.C., 2014. Nonlinear air quality modeling using multivariate adaptive regression splines in Gijón urban area (Northern Spain) at local scale. *Appl. Math. Comput.* 235, 50–65. <https://doi.org/10.1016/j.amc.2014.02.096>
- García Nieto, P.J., Antón, J.C.Á., Vilán, J.A.V., García-Gonzalo, E., 2015. Air quality modeling in the Oviedo urban area (NW Spain) by using multivariate adaptive regression splines. *Environ. Sci. Pollut. Res.* 22, 6642–6659. <https://doi.org/10.1007/s11356-014-3800-0>
- Garcia, V.C., Foley, K.M., Gego, E., Holland, D.M., Rao, S.T., 2010. A Comparison of Statistical Techniques for Combining Modeled and Observed Concentrations to Create High-Resolution Ozone Air Quality Surfaces. *J. Air Waste Manag. Assoc.* 60, 586–595. <https://doi.org/10.3155/1047-3289.60.5.586>
- Guillas, S., Bao, J., Choi, Y., Wang, Y., 2008. Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta. *Atmos. Environ.* 42, 1338–1348. <https://doi.org/10.1016/j.atmosenv.2007.10.027>
- Henrotin, J.-B., Zeller, M., Lorgis, L., Cottin, Y., Giroud, M., Béjot, Y., 2010. Evidence of the role of short-term exposure to ozone on ischaemic cerebral and cardiac events: the Dijon Vascular Project (DIVA). *Heart Br. Card. Soc.* 96, 1990–1996. <https://doi.org/10.1136/hrt.2010.200337>

- Hogrefe, C., Biswas, J., Lynn, B., Civerolo, K., Ku, J.-Y., Rosenthal, J., Rosenzweig, C., Goldberg, R., Kinney, P.L., 2004. Simulating regional-scale ozone climatology over the eastern United States: model evaluation results. *Atmos. Environ.* 38, 2627–2638. <https://doi.org/10.1016/j.atmosenv.2004.02.033>
- Honore, C., Rouil, L., Meleux, F., Malherbe, L., Bessagnet, B., Vautard, R., Poisson, N., Peuch, V.-H., Dufour, A., 2005. The PREV’AIR system, an operational system for large scale air quality forecasts over Europe; applications at the local scale. Presented at the International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO 10), Crete, Greece, pp. 123–127.
- Janssen, S., Guerreiro, C., Viaene, P., Georgieva, E., Thunis, P., 2017. Guidance Document on Modelling Quality Objectives and Benchmarking.
- Joly, D., Gilbert, D., Diaz-de-Quijano, M., Hilal, M., Joly, M., Bernard, N., 2021. Enhancing air quality forecasts by geomatic downscaling: an application to daily PM10 concentrations in France. *Theor. Appl. Climatol.* 143, 327–339. <https://doi.org/10.1007/s00704-020-03418-7>
- Kato, T., 2016. Chapter 4 - Prediction of photovoltaic power generation output and network operation, in: Funabashi, T. (Ed.), *Integration of Distributed Energy Resources in Power Systems*. Academic Press, pp. 77–108. <https://doi.org/10.1016/B978-0-12-803212-1.00004-0>
- Leathwick, J.R., Elith, J., Hastie, T., 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Model., Predicting Species Distributions* 199, 188–196. <https://doi.org/10.1016/j.ecolmodel.2006.05.022>
- Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* 525, 367–371. <https://doi.org/10.1038/nature15371>
- Lesmeister, C., 2019. *Mastering Machine Learning with R* Ed. 3. Packt Publishing.
- Li, R., Xu, M., Li, M., Chen, Z., Zhao, N., Gao, B., Yao, Q., 2021. Identifying the spatiotemporal variations in ozone formation regimes across China from 2005 to 2019 based on polynomial



- simulation and causality analysis. *Atmospheric Chemistry and Physics* 21, 15631–15646.  
<https://doi.org/10.5194/acp-21-15631-2021>
- Lu, X., Zhang, L., Shen, L., 2019. Meteorology and Climate Influences on Tropospheric Ozone: a Review of Natural Sources, Chemistry, and Transport Patterns. *Curr. Pollut. Rep.* 5, 238–260.  
<https://doi.org/10.1007/s40726-019-00118-3>
- Ma, R., Ban, J., Wang, Q., Li, T., 2020. Statistical spatial-temporal modeling of ambient ozone exposure for environmental epidemiology studies: A review. *Science of The Total Environment* 701, 134463. <https://doi.org/10.1016/j.scitotenv.2019.134463>
- Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G., Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., Meleux, F., 2017. CHIMERE-2017: from urban to hemispheric chemistry-transport modeling. *Geosci. Model Dev.* 10, 2397–2423. <https://doi.org/10.5194/gmd-10-2397-2017>
- Martin, E.P., 2011. Comparative Performance of Different Statistical Models for Predicting Ground-Level Ozone (O<sub>3</sub>) and Fine Particulate Matter (PM<sub>2.5</sub>) Concentrations in Montréal, Canada (masters). Concordia University.
- Martin, N., 2010. Performances des modélisations déterministes d’ozone à méso-échelle et à micro-échelle dans les Alpes-Maritimes. *Cybergeo Rev. Eur. Géographie Eur. J. Geogr.*  
<http://cybergeo.revues.org/index23183.html>.
- Marx, T., Bernard, N., Parmentier, A.-L., Puyraveau, M., Martin, B., Gantelet, M., Pretalli, J.-B., Dalphin, J.-C., Mauny, F., Desmettre, T., 2019. Does air pollution really impact the onset of spontaneous pneumothorax? A French case-crossover study. *Environ. Int.* 127, 317–323.  
<https://doi.org/10.1016/j.envint.2019.01.056>
- Meleux, F., Ung, A., 2020. Performances Prev’Air en 2017 et 2018. LCSQA.
- Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., Vivanco, M.G., 2013. CHIMERE 2013: a model for regional

- atmospheric composition modelling. *Geosci. Model Dev.* 6, 981–1028.  
<https://doi.org/10.5194/gmd-6-981-2013>
- Nuvolone, D., Petri, D., Voller, F., 2018. The effects of ozone on human health. *Environ. Sci. Pollut. Res.* 25, 8074–8088. <https://doi.org/10.1007/s11356-017-9239-3>
- Proietti, C., Fornasier, M.F., Sicard, P., Anav, A., Paoletti, E., De Marco, A., 2021. Trends in tropospheric ozone concentrations and forest impact metrics in Europe over the time period 2000–2014. *J. For. Res.* 32, 543–551. <https://doi.org/10.1007/s11676-020-01226-3>
- Ramos, Y., Requia, W.J., St-Onge, B., Blanchet, J.-P., Kestens, Y., Smargiassi, A., 2018. Spatial modeling of daily concentrations of ground-level ozone in Montreal, Canada: A comparison of geostatistical approaches. *Environ. Res.* 166, 487–496.  
<https://doi.org/10.1016/j.envres.2018.06.036>
- Ren, X., Mi, Z., Georgopoulos, P.G., 2020. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. *Environment International* 142, 105827.  
<https://doi.org/10.1016/j.envint.2020.105827>
- Riviere, E., Bernard, J., Hulin, A., Virga, J., Dugay, F., Charles, M.-A., Cheminat, M., Cortinovis, J., Ducroz, F., Laborie, A., Malherbe, L., Piga, D., Real, E., Robic, P.-Y., Zaros, C., Seyve, E., Lepeule, J., 2019. Air pollution modeling and exposure assessment during pregnancy in the French Longitudinal Study of Children (ELFE). *Atmos. Environ.* 205, 103–114.  
<https://doi.org/10.1016/j.atmosenv.2019.02.032>
- Rouil, L., Honoré, C., Vautard, R., Beekmann, M., Bessagnet, B., Malherbe, L., Meleux, F., Dufour, A., Elichegaray, C., Flaud, J.-M., Menut, L., Martin, D., Peuch, A., Peuch, V.-H., Poisson, N., 2009. Prev'air: An Operational Forecasting and Mapping System for Air Quality in Europe. *Bull. Am. Meteorol. Soc.* 90, 73–84. <https://doi.org/10.1175/2008BAMS2390.1>
- Roy, S.S., Pratyush, C., Barna, C., 2018. Predicting Ozone Layer Concentration Using Multivariate Adaptive Regression Splines, Random Forest and Classification and Regression Tree, in: Balas, V.E., Jain, L.C., Balas, M.M. (Eds.), *Soft Computing Applications, Advances in Intelligent*

- Systems and Computing. Springer International Publishing, Cham, pp. 140–152.  
[https://doi.org/10.1007/978-3-319-62524-9\\_11](https://doi.org/10.1007/978-3-319-62524-9_11)
- Royal Society (Great Britain), 2008. Ground-level ozone in the 21st century: future trends, impacts and policy implications. The Royal Society, London.
- Sanderfoot, O.V., Holloway, T., 2017. Air pollution impacts on avian species via inhalation exposure and associated outcomes. *Environ. Res. Lett.* 12, 083002. <https://doi.org/10.1088/1748-9326/aa8051>
- Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J.M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M.T., Rouil, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., Wind, P., 2015. Performance of European chemistry transport models as function of horizontal resolution. *Atmospheric Environment* 112, 90–105.  
<https://doi.org/10.1016/j.atmosenv.2015.04.003>
- Sekulic, S., Kowalski, B.R., 1992. MARS: A tutorial. *J. Chemom.* 6, 199–216.  
<https://doi.org/10.1002/cem.1180060405>
- Sicard, P., Agathokleous, E., De Marco, A., Paoletti, E., Calatayud, V., 2021. Urban population exposure to air pollution in Europe over the last decades. *Environ Sci Eur* 33, 28.  
<https://doi.org/10.1186/s12302-020-00450-2>
- Sicard, P., De Marco, A., Troussier, F., Renou, C., Vas, N., Paoletti, E., 2013. Decrease in surface ozone concentrations at Mediterranean remote sites and increase in the cities. *Atmospheric Environment* 79, 705–715. <https://doi.org/10.1016/j.atmosenv.2013.07.042>
- Sicard, P., Serra, R., Rossello, P., 2016. Spatiotemporal trends in ground-level ozone concentrations and metrics in France over the time period 1999–2012. *Environmental Research* 149, 122–144.  
<https://doi.org/10.1016/j.envres.2016.05.014>
- Steven Porter, P., Rao, S.T., Hogrefe, C., Gego, E., Mathur, R., 2015. Methods for reducing biases and errors in regional photochemical model outputs for use in emission reduction and exposure assessments. *Atmos. Environ.* 112, 178–188. <https://doi.org/10.1016/j.atmosenv.2015.04.039>

- Thunis, P., Miranda, A., Baldasano, J.M., Blond, N., Douros, J., Graff, A., Janssen, S., Juda-Rezler, K., Karvosenoja, N., Maffei, G., Martilli, A., Rasoloharimahefa, M., Real, E., Viaene, P., Volta, M., White, L., 2016. Overview of current regional and local scale air quality modelling practices: Assessment and planning tools in the EU. *Environ. Sci. Policy, Multidisciplinary research findings in support to the EU air quality policy: experiences from the APPRAISAL, SEFIRA and ACCENT-Plus EU FP7 projects* 65, 13–21. <https://doi.org/10.1016/j.envsci.2016.03.013>
- U.S. EPA, 2018. Modeling Guidance for Demonstrating Air Quality Goals for Ozone, PM<sub>2.5</sub>, and Regional Haze (No. EPA 454/R-18-009). U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC, USA.
- Wang, M., Beelen, R., Basagana, X., Becker, T., Cesaroni, G., de Hoogh, K., Dedele, A., Declercq, C., Dimakopoulou, K., Eeftens, M., Forastiere, F., Galassi, C., Gražulevičienė, R., Hoffmann, B., Heinrich, J., Iakovides, M., Künzli, N., Korek, M., Lindley, S., Mölter, A., Mosler, G., Madsen, C., Nieuwenhuijsen, M., Phuleria, H., Pedeli, X., Raaschou-Nielsen, O., Ranzi, A., Stephanou, E., Sugiri, D., Stempfelet, M., Tsai, M.-Y., Lanki, T., Udvardy, O., Varró, M.J., Wolf, K., Weinmayr, G., Yli-Tuomi, T., Hoek, G., Brunekreef, B., 2013. Evaluation of Land Use Regression Models for NO<sub>2</sub> and Particulate Matter in 20 European Study Areas: The ESCAPE Project. *Environ. Sci. Technol.* 47, 4357–4364. <https://doi.org/10.1021/es305129t>
- Wang, M., Sampson, P.D., Hu, J., Kleeman, M., Keller, J.P., Olives, C., Szpiro, A.A., Vedal, S., Kaufman, J.D., 2016. Combining Land-Use Regression and Chemical Transport Modeling in a Spatiotemporal Geostatistical Model for Ozone and PM<sub>2.5</sub>. *Environ. Sci. Technol.* 50, 5111–5118. <https://doi.org/10.1021/acs.est.5b06001>
- Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L., 2017. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total Environ.* 575, 1582–1596. <https://doi.org/10.1016/j.scitotenv.2016.10.081>

- Yilmaz, B., Aras, E., Nacar, S., Kankal, M., 2018. Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. *Sci. Total Environ.* 639, 826–840. <https://doi.org/10.1016/j.scitotenv.2018.05.153>
- Zhang, J., Wei, Y., Fang, Z., 2019. Ozone Pollution: A Major Health Hazard Worldwide. *Front. Immunol.* 10, 2518. <https://doi.org/10.3389/fimmu.2019.02518>

## Chapitre 4 : Identification de dynamiques spatiales et temporelles d'ozone et de leurs déterminants

### Spatialized temporal dynamics of daily ozone concentrations: identification of the main spatial differences

Article accepté le 27 février 2023 dans : Environment International.

#### **Résumé :**

##### **Contexte :**

L'ozone troposphérique est l'un des polluants atmosphériques les plus préoccupants compte tenu de ses impacts sanitaires et environnementaux. Les enjeux liés à la gestion de ce polluant secondaire sont prégnants, et tendent à s'intensifier dans un contexte de changement climatique où les concentrations d'O<sub>3</sub> sont prévues à la hausse dans les décennies à venir. Une meilleure gestion des concentrations d'O<sub>3</sub> requiert une compréhension approfondie de la dynamique spatio-temporelle de l'O<sub>3</sub> et de ses déterminants. Des outils de modélisation sont nécessaires afin de disposer des valeurs de concentrations d'O<sub>3</sub> de manière continue dans l'espace et dans le temps, ainsi qu'une résolution fine. La concentration modélisée en un point donné est la résultante de l'action conjuguée des différents déterminants de l'O<sub>3</sub>.

L'objectif des travaux décrits dans ce chapitre est i) d'explorer la dynamique spatio-temporelle de l'O<sub>3</sub> à l'échelle journalière et kilométrique (9 km<sup>2</sup>), sur 12 ans et environ 1000 km<sup>2</sup>, et ii) d'identifier les potentiels déterminants de cette dynamique.

##### **Méthode**

La zone d'étude est centrée sur Besançon (Franche-Comté, France) et s'étend sur 1134 km<sup>2</sup>. La plateforme de modélisation PREV'EST (présentée en Chapitre 2 - 1.1) génère une valeur de concentration d'O<sub>3</sub> (moyenne et maximum) de manière journalière sur chaque maille (3 km x 3 km) de 2008 à 2019. Les concentrations sont corrigées à l'aide de la correction 2WF-elevation décrite en Chapitre 3 -2). Une série temporelle de concentrations corrigées d'O<sub>3</sub> moyennes journalières de 2008 à 2019 est donc associée à chaque maille de modélisation. Ces séries temporelles sont soumises à un processus de classification en deux phases : i) calcul de la matrice de distance entre les séries à l'aide du DTW et, ii) classification hiérarchique ascendante.

Chaque maille de modélisation est ensuite caractérisée par son degré d'urbanisation et de trafic, sa proportion de surface végétalisée ainsi que son altitude moyenne. Ces caractéristiques sont mises en relation avec les concentrations d'O<sub>3</sub> des mailles (moyennée sur 12 années) à l'aide de régressions linéaire (simple ou multiple) et de fonctions polynomiales.

##### **Résultats**

Trois dynamiques temporelles de concentrations journalières d'O<sub>3</sub> ont été identifiées. Ces dynamiques coïncident avec les milieux urbain, périurbain et rural. Un gradient croissant de concentration d'ozone est observé du milieu urbain au milieu rural, et s'avère renforcé sous l'influence de l'altitude. L'altitude et la proportion de surface végétalisée sont associées positivement à la concentration d'O<sub>3</sub>, contrairement à l'urbanisation. Les différentes dynamiques temporelles d'O<sub>3</sub> sont issues de l'action conjuguée de ces déterminants de manière simultanée. Les zones rurales sont soumises aux concentrations les plus élevées en moyenne mais ne font que peu l'objet de surveillance de la qualité de l'air. Les séries temporelles de cette zone présentent

également une prédictibilité moindre. Ces zones peuvent être considérées comme vulnérables face à l'exposition à l'O<sub>3</sub>.

### ***Conclusion***

La méthodologie employée permet de passer d'une modélisation des concentrations d'O<sub>3</sub> à une projection des différentes dynamiques temporelles d'O<sub>3</sub> identifiées ainsi qu'à l'identification de leurs déterminants. L'action conjuguée des déterminants retenus a été illustrée de manière synthétique. Cette méthode permet également d'identifier des zones de vulnérabilité de manière systématique et reproductible. Les résultats ainsi obtenus pourraient être exploités dans le cadre d'un processus de gestion du risque environnemental et/ou sanitaire.

**Title page**

Original Research Article

**Title**

Spatialized temporal dynamics of daily ozone concentrations: identification of the main spatial differences

**Authors:**

Honorine Gauthier-Manuel<sup>a,b,\*</sup>, Nadine Bernard<sup>a,c</sup>, Mathieu Boilleaut<sup>d</sup>, Patrick Giraudoux<sup>a</sup>, Sophie Pujol<sup>a,b</sup>, Frédéric Mauny<sup>a,b</sup>

**Affiliation** of the authors:

a Chrono-environnement UMR 6249, CNRS, Université de Franche-Comté, F-25000, Besançon, France

b Unité de Méthodologie en Recherche Clinique, épidémiologie et Santé Publique (uMETH), Inserm CIC 1431, CHU, 25030, Besançon Cedex, France

c Centre National de La Recherche Scientifique, UMR 6049, Laboratoire ThéMA, Université de Bourgogne Franche-Comté, 25000, Besançon, France

d ATMO Bourgogne-Franche-Comté, 25000, Besançon, France

\*Corresponding author.

E-mail address: honorine.gauthier-manuel@univ-fcomte.fr (H. Gauthier-Manuel)



**Abstract:**

Ground-level ozone (O<sub>3</sub>) is one of the most worrisome air pollutants regarding environmental and health impacts. There is a need for a deeper understanding of its spatial and temporal dynamics. Models are needed to provide continuous temporal and spatial coverage in ozone concentration data with a fine resolution. However, the simultaneous influence of each determinant of ozone dynamics, their spatial and temporal variations, and their interaction make the resulting dynamics of O<sub>3</sub> concentrations difficult to understand.

This study aimed to i) identify different classes of temporal dynamics of O<sub>3</sub> at daily and 9 km<sup>2</sup> resolution over a long-term period of 12 years, ii) identify the potential determinants of these dynamics and, iii) explore the spatial distribution of the potential classes of temporal dynamics on a spatial continuum and over about 1000km<sup>2</sup>.

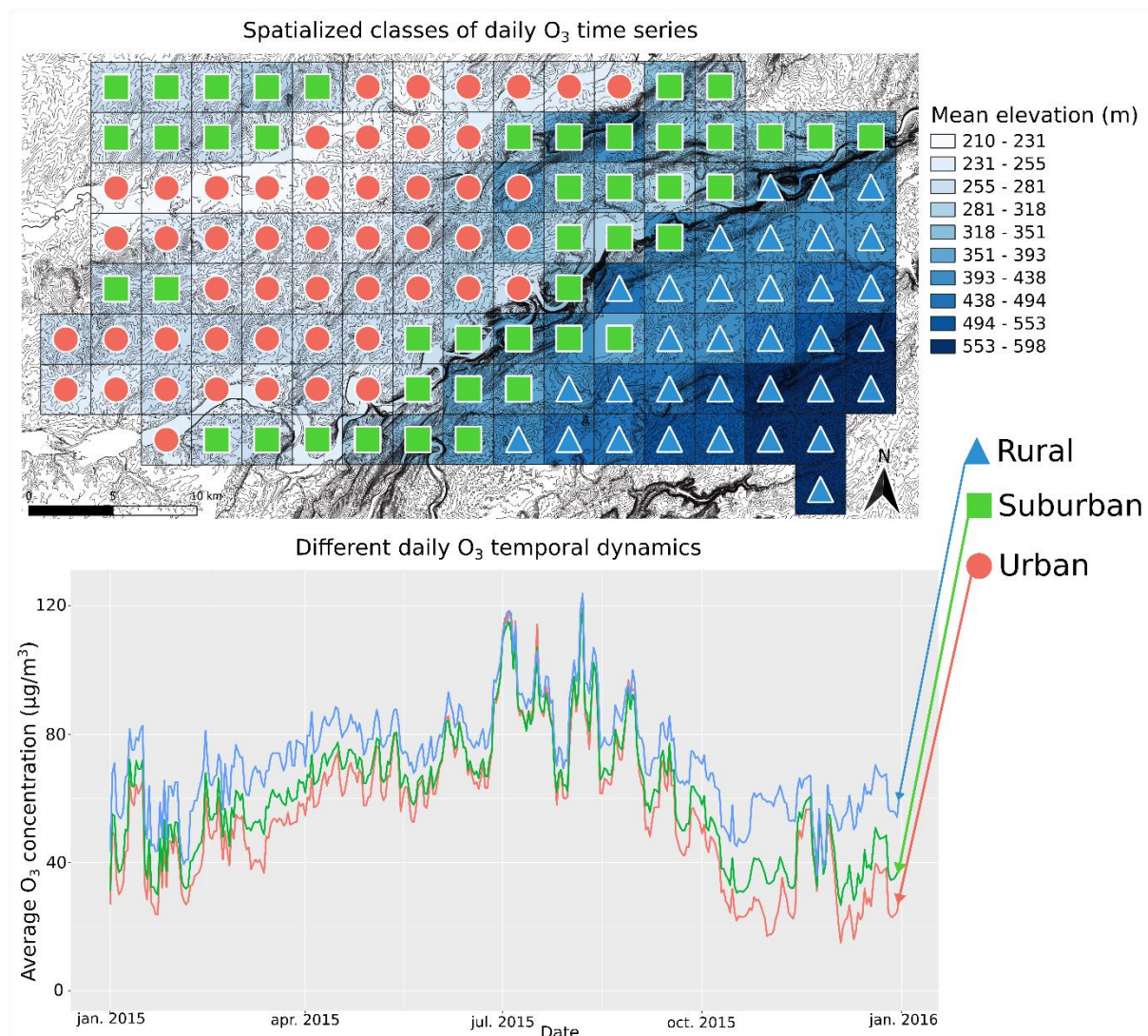
Thus, 126 time series of 12-year daily ozone concentrations were classified using dynamic time warping (DTW) and hierarchical clustering (study area centered on Besançon, eastern France). The different temporal dynamics obtained differed on elevation, ozone levels, proportions of urbanized and vegetated surfaces.

We identified different daily ozone temporal dynamics, spatially structured, that overlapped areas called urban, suburban and rural. Urbanization, elevation and vegetation acted as determinants simultaneously. Individually, elevation and vegetated surface were positively correlated with O<sub>3</sub> concentrations ( $r = 0.84$  and  $r = 0.41$ , respectively), while the proportion of urbanized area was negatively correlated with O<sub>3</sub> ( $r = -0.39$ ). An increasing ozone concentration gradient was observed from urban to rural areas and was reinforced by the elevation gradient. Rural areas were both subject to higher ozone levels ( $p < 0.001$ ), least monitoring and lower predictability.

We identified main determinants of the temporal dynamics of ozone concentrations. The joint influence of determinants was also synthesized. This study proposed a systematic, and reproducible way to build exposure area mapping.

Keywords: ozone, spatiotemporal dynamics, land use, vulnerability areas, management

## Graphical abstract:



## Study Funding

This work was supported by the agglomeration community of Grand Besançon Métropole (PhD scholarship, 2019) and the “Pôle fédératif de recherche et de formation en santé publique Bourgogne-Franche-Comté” (2019 and 2020 research project calls).

## 1) Introduction:

Ozone is a worrisome air pollutant that has diverse adverse effects on both environmental compartments and human health (Zhang et al., 2019). The impact of ozone (O<sub>3</sub>) on crop yields, forest productivity, human health and cultural heritage building triggers sizable financial costs (Feng et al., 2019; Screpanti and De Marco, 2009; Vlachokostas et al., 2010). These impacts are increased worldwide by the recent increasing trends of O<sub>3</sub> levels, especially in urban areas (Yan et al., 2019), and will be strengthened by the future O<sub>3</sub> increase predicted under the RCP8.5 scenario developed by the Intergovernmental Panel on Climate Change (IPCC, 2014; Sicard et al., 2017; Yang et al., 2019).

This secondary air pollutant is produced in the presence of sunlight and precursors (NO<sub>x</sub>, VOC). Due to the strong dependence of O<sub>3</sub> production on sunlight (Royal Society (Great Britain), 2008), this pollutant is characterized by high temporal variability at different temporal scales. A daily cycle is characterized by the highest levels at the beginning of the afternoon (Notario et al., 2013). A seasonal cycle is well known, with the highest O<sub>3</sub> levels observed during spring and summer (at midlatitude regions of the Northern Hemisphere). The short-term influence of meteorological parameters, such as solar radiation, temperature, humidity, wind speed and direction, is well-established in the literature (Nguyen et al., 2022; Otero et al., 2016; Upadhaya et al., 2020). Our approach focused on a long-term based temporal behavior (12 years of daily mean ozone content e.g. 12 seasonal cycles) and the exploration of the factors of a potential spatial heterogeneity. Most studies tend to focus on acute exposure during the summer period (Carro-Calvo et al., 2017; Liu et al., 2021; Querol et al., 2018) or on heat waves phenomena (Lacour et al., 2006; Pascal et al., 2021; Pyrgou et al., 2018). However, chronic exposure has also been associated with adverse health effects (Liu et al., 2022; Wang et al., 2022) and crop yield losses (Roberts et al., 2022), and deserves to be further investigated.

Anthropogenic sources of precursors of O<sub>3</sub> are concentrated in urban areas, leading to a (well-established) structuration of O<sub>3</sub> concentrations associated with the degree of urbanization (Zhang et al., 2020). NO<sub>x</sub> are mainly emitted by road traffic. The titration reaction occurring between NO and O<sub>3</sub> ( $\text{NO} + \text{O}_3 \rightarrow \text{NO}_2 + \text{O}_2$ ) leads to local O<sub>3</sub> depletion/destruction near the main traffic axes and thus in urban areas (Huang et al., 2018; Reche et al., 2018). The highest concentrations of O<sub>3</sub> observed in suburban

and rural areas are due to; i) the time-delay needed to perform chemical reactions responsible for its production, and ii) the transport of its precursors from urban sites to these areas (Derstroff et al., 2017; Huang et al., 2018; Monks et al., 2015; Sicard et al., 2020; Yan et al., 2019). Elevation has also been shown to be a determinant of O<sub>3</sub> levels that can spatially structure it (Agathokleous et al., 2022).

At the European scale, air quality monitoring relies on a network of stations focused on the most populated areas, which concentrate sources of (primary) pollutants targeted by health risk management guidelines (EEA, 2015; Madrigano et al., 2015). Considering other primary air pollutants (especially NO<sub>2</sub> and PM), the areas the most monitored are both the most polluted and the most populated. However, it may seem paradoxical that the most ozone-polluted areas are the least monitored.

The quantification of the concentration levels relies on two main approaches: measurements and modeling. Measures have the advantage of being very reliable thanks to a robust validation process. However, they present limited spatial coverage (Jerrett et al., 2005) and do not allow information to be continuously obtained in space (Garcia et al., 2010). Modeling tools handle this issue, providing estimation of O<sub>3</sub> levels through large areas without sizable costs compared to the measurements required to cover the same area (regional, national, or even European) continuously in space. However, the data stemming from regional and global nested models do not reproduce the temporal variability of O<sub>3</sub> levels (Derwent et al., 2015; Yan et al., 2018, 2017). Optimization processes can suppress this lack of temporal precision (Gauthier-Manuel et al., 2022). Modeling approaches allow for crossing spatial and temporal dimensions.

Considering only the temporal or the spatial dimension of O<sub>3</sub> dynamics delivers only a partial representation of the O<sub>3</sub> dynamics. These two dimensions are necessarily very related and need to be combined to obtain a more complete depiction and a better understanding of it. Few studies have attempted to address this issue but have used a restricted approach: either measurement (Duarte et al., 2022; Fan et al., 2020; Ma et al., 2021; Sicard et al., 2016), either modeling tools with a coarse grid mesh (ranging from 10 km x 10 km to 25 km x 25 km), or a low temporal precision (Gao et al., 2020; Yapeng Wang et al., 2018) or only focused on urban areas and/or summer periods (Reche et al., 2018;

Yuanyuan Wang et al., 2018). To our knowledge, only one study combined high spatial and temporal resolution over a long period of time and a large spatial range (Meng et al., 2022).

To better explore the distribution of tropospheric O<sub>3</sub>, we need to simultaneously consider the spatial and temporal dimensions at a fine resolution (~ some km<sup>2</sup>) and over a large range (~ thousands of km<sup>2</sup>), in order to get a deep understanding of its spatial and temporal dynamics. The relationship between O<sub>3</sub> dynamics and each determinant is well described and understood. However, the combination of the influence of each of these determinants, their spatial and/or temporal variations, and their interaction (synergistic, antagonistic) make the resulting temporal dynamics of O<sub>3</sub> concentrations extremely difficult to predict across specific territory areas.

The objective of this study was i) identify different classes of temporal dynamics of O<sub>3</sub> at daily and 9 km<sup>2</sup> resolution over a long-term period of 12 years, ii) identify the potential determinants of these dynamics and, iii) explore the spatial distribution of the potential classes of temporal dynamics on a spatial continuum and over about 1000km<sup>2</sup>. The temporal and spatial dimensions of the O<sub>3</sub> dynamics were dissociated and iteratively analyzed.

## **2) Materials and Methods**

### **2.1) Study area**

The study area extends over a total surface of 1134 km<sup>2</sup> located in Eastern France in the Burgundy–Franche-Comté region (Figure 31) and presents a complete urbanization gradient (from dense urban areas to rural and agricultural areas). The land cover is characterized by a low mean percentage of artificialized area (9.5%) and 41.3% of forests and seminatural environments. The elevation extends from 210 to 598 m over the total study area. The climate is continental with dominant winds coming from the northeast and southwest (Figure 31). The population of the area is 247,280 inhabitants for a mean population density of 167 inhabitants per km<sup>2</sup>. On the administrative boundary of the French *department* scale (according to the national process of emissions inventory), NO<sub>x</sub> are emitted by traffic at 70%, by the manufacturing industry at 10%, and by agricultural and residential activities at 7% and

6%, respectively. The main lines of business emitting nonmethane COV are the manufacturing industry (47%), residential activities (41%) and traffic (9%).

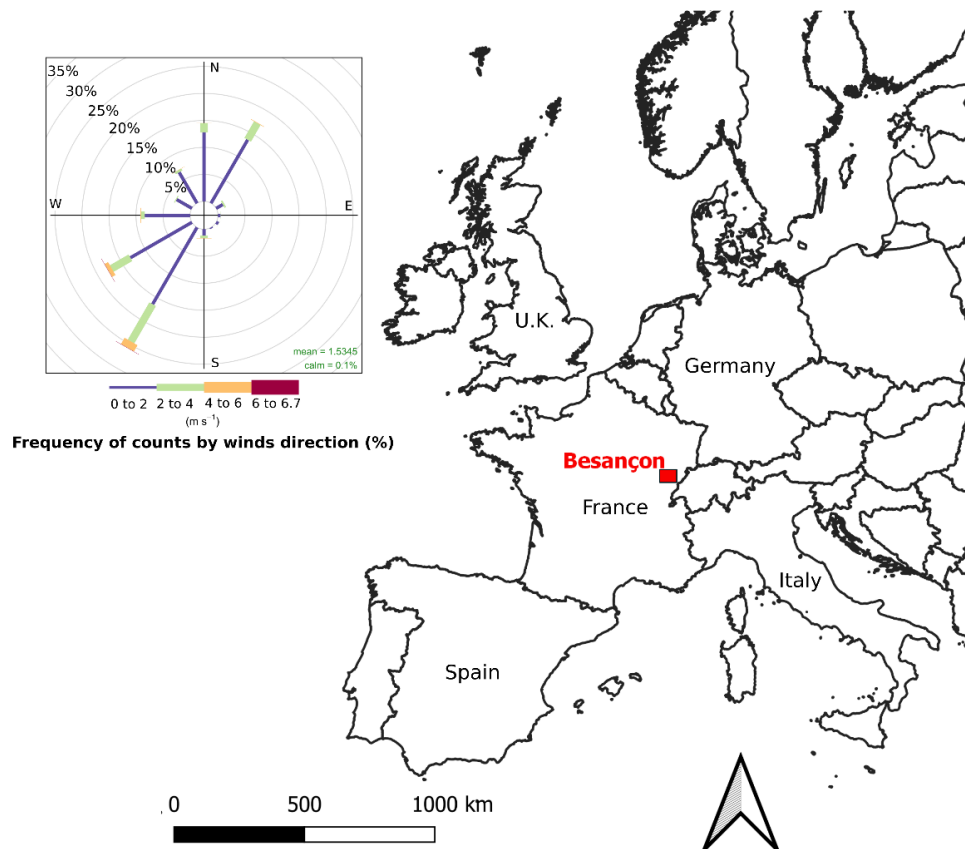


Figure 31: Location of the study site (Besançon), indicated by a red rectangle. The frequency of wind direction and speed is represented by a wind rose in the left top corner of the figure.

This study focused on Besançon city (117,912 inhabitants in 2019), a medium-sized city (Dijkstra and Poelman, 2012) without any airport or sizable industrial pole. The city is crossed by the Doubs River and surrounded by several hills. This topography constrains the development of the city.

## 2.2) Ozone concentration modeling

We used O<sub>3</sub> concentration modeling to provide daily mean and maximum values over a 12-year period (2008 to 2019) based on a 3 km x 3 km grid mesh over the whole study area. Each modeling mesh included 4283 daily values for both the mean and maximum concentrations.

The PREV'EST modeling platform used natively in this study integrates the meteorological model WRF (Powers et al., 2017) and the CHIMERE chemical transport model (CTM) (Mailler et al., 2017). WRF is a tridimensional non-hydrostatic model, taking into account the vertical acceleration in the atmosphere relying on equations of fluid mechanics and thermodynamics. The wind speed and direction, rainfalls, air temperature (2 meters above the ground) and the cloud cover fraction are considered in the WRF model. PREV'EST was developed and managed by ATMO Grand Est (regional air quality monitoring association), and applied at a transnational scale (eastern France and the boundary area of France, Germany, Switzerland, and Luxembourg).

An annual correction process proposed by Gauthier-Manuel et al. (2022) (2 WF-elevation correction) has been applied to the raw modeling data to improve its precision and exactitude and the representation of O<sub>3</sub> peaks. The correction model was based on 22,640 statistical units. The validation of the correction was based on the comparison between measurements and modeled data. The improvement of the modeled values was quantified by multiple performance indicators (MAE, RMSE, MBE, r, target plots). MBE allows quantifying an over- or underestimation. RMSE and MAE are precision indicators. The r is the Pearson correlation coefficient between modeled and measured data. The correction improved the performance of the model. The MBE, MAE, RMSE, and r values were, for the corrected models 0.01, 10.04, 12.67, and 0.85, respectively, and were 14.67, 19.23, 23.18, and 0.67, respectively, for the uncorrected modeled data. Appendix 1 provides more detailed information (description of the performance indicators and target plots) and the target plots before and after correction (Figure A.1). The (internal) validation process (performed on data used to build the correction) was followed by an external validation (6 additional air quality monitoring stations, 14,846 statistical units) whose results are presented in Figure A.2.

### 2.3) Elevation, urbanization, traffic and vegetation characteristics

The determinants used to study the O<sub>3</sub> dynamics and characterize the classes were: the mean elevation, the proportion of urbanized surface, the length of linear major road axis, and the proportion of vegetated

surface of the modeling grid meshes. These determinants are defined as many factors combined together to affect the ozone concentrations and their dynamics.

The mean elevation of each modeling mesh was extracted using a digital elevation model (with a 25 m resolution) named BDALTI® and provided by the French geographical Institute (IGN). The urbanized surface was computed selecting the following codes-categories from the urban atlas (Copernicus Land Monitoring Service, 2022) to delineate the urbanized area: 11100-Continuous Urban Fabric (S.L. > 80%), 11210-Discontinuous Dense Urban Fabric (S.L.: 50% - 80%), 11220-Discontinuous Medium Density Urban Fabric (S.L.: 30% - 50%), 11230-Discontinuous Low Density Urban Fabric (S.L.: 10% - 30%), 11240-Discontinuous Very Low Density Urban Fabric (S.L. < 10%), 11300-Isolated Structures, 12100-Industrial, commercial, public, military and private units, 12210-Fast transit roads and associated land, 12220-Other roads and associated land, 12230-Railways and associated land, 12300-Port areas, 14100-Green urban areas, 14200-Sports and leisure facilities, 50000-Water bodies. The length of linear major road axis was computed using the Route500® (IGN) vector layer describing the road networks and their importance to traffic.

The IGN also provided the BD Topo® (2017) vector layer describing the total vegetation coverage. The proportions of urbanized and vegetated surface, and the linear of major road axis per modeling mesh were calculated using GIS software (QGIS 3.8).

#### 2.4) Statistical analysis

The O<sub>3</sub> dynamics classification of the meshes was conducted by a two-step process: construction of the time-series distance matrix followed by a hierarchical clustering.

##### 2.4.1 *Dynamic Time Warping*

First, we calculated a distance matrix between the time-series of each modeling mesh to assess their relative proximity using the dynamic time warping (DTW). No preprocessing (normalization (Offset Translation, Amplitude Scaling, Trend Removal, and Smoothing) (Ma, 2020)) was done to keep the characteristics and structures of the time-series. This allowed to identify interday variations and ozone peaks, and then identify different temporal dynamics. Moreover, our time-series had the same length,



nature ( $\text{O}_3$  concentration) and unit ( $\mu\text{g}\cdot\text{m}^{-3}$ ), reducing the potential impact of the absence of normalization. DTW is based on three classical constraints: boundary condition, continuity and monotonicity (Roelofsen, 2018). DTW requires the computation of the local cost matrix (LCM). The LCM was calculated using the R package called “TSclust” (Montero and Vilar, 2015). The over-alignment between time series that can occur using DTW has been handled by applying an additional constraint (Sakoe-Chiba band) to the algorithm (Sakoe and Chiba, 1978). It constrains the warping path of the cost matrix (the region that can be explored by the algorithm) and avoids pathological warping (*e.g.*, a single point matching an extremely large sequence) (Ratanamahatana and Keogh, 2004). An illustration of an LCM with a Sakoe-Chiba band is provided in Figure A.4. We used a window size of 10 cells around the diagonal of the LCM (Giorgino, 2009). The value of the band was tested on a range of 10 to 40 cells in the LCM. Usually, the recommended value of the band is about 5% of the total length of the time series (Javed et al., 2020). However, considering our data, its seasonality and our purpose, 10 to 40 cells in the LCM seemed more appropriate.

#### 2.4.2 Hierarchical clustering

Agglomerative hierarchical clustering was used to classify the modeling meshes based on the distance matrix computed using DTW. In this process, each object (each modeling mesh) is first considered a distinct cluster. The clusters are then gradually merged together in a bottom-up way (Aghabozorgi et al., 2015). The result of this process is summarized by a dendrogram. We used the complete linkage to build the dendrogram. The number of classes was determined by visualizing the dendrogram and using the “elbow method”. This method relies on the visualization of the total within-cluster sum of squares (WSS) as a function of the number of clusters. The break or “elbow” displayed in this relationship describes the “optimal” number of clusters, leading to a rapid decrease in WSS for the lowest number of clusters. This means that adding more clusters does not improve the partition process.

### 2.4.3 Descriptive and analytical statistics

Elevation, urbanization, traffic and vegetation characteristics were compared across the identified classes of O<sub>3</sub> temporal dynamics by performing Kruskal–Wallis tests followed by post hoc pairwise comparison tests (kruskalmc, pgirmess package) (Siegel and Castellan, 1988).

Simple and multiple linear regressions were performed to explain the O<sub>3</sub> concentration variations. A polynomial function was introduced to model the nonlinear relationship between O<sub>3</sub> concentration and elevation. The spatial autocorrelation of the classes was assessed by computing the ratio of neighboring meshes belonging to the same class as the mesh considered. The mean ratio was compared to a random distribution (1000 random draws). The spatial autocorrelation of the variables of the linear regressions was evaluated using Moran's index.

The (Shannon) spectral entropy was computed using the R package called “feasts”. This indicator quantifies the predictability of a time series. A series presenting strong temporal structuration (*e.g.*, seasonality) will be easier to forecast and will be associated with an entropy value close to 0, contrary to a noisy series (*e.g.*, difficult to forecast) that would be associated with an entropy value close to 1 (Hyndman and Athanasopoulos, 2021; Zhou et al., 2021).

## 3) Results

### 3.1) Classification of the temporal dynamics

The study area was divided into 126 grid meshes associated with both 4283 daily mean and maximum levels of O<sub>3</sub> concentrations, for a total of 2 \* 539,658 concentration values. Based on the time series classification process, we obtained three classes of temporal dynamics of the daily O<sub>3</sub> concentrations. The temporal dynamics of the daily mean value of these three classes are presented in Figure 32, highlighting a marked gradient between areas called rural, suburban and urban. The gap between curves tends to get slighter during summer ozone peaks.

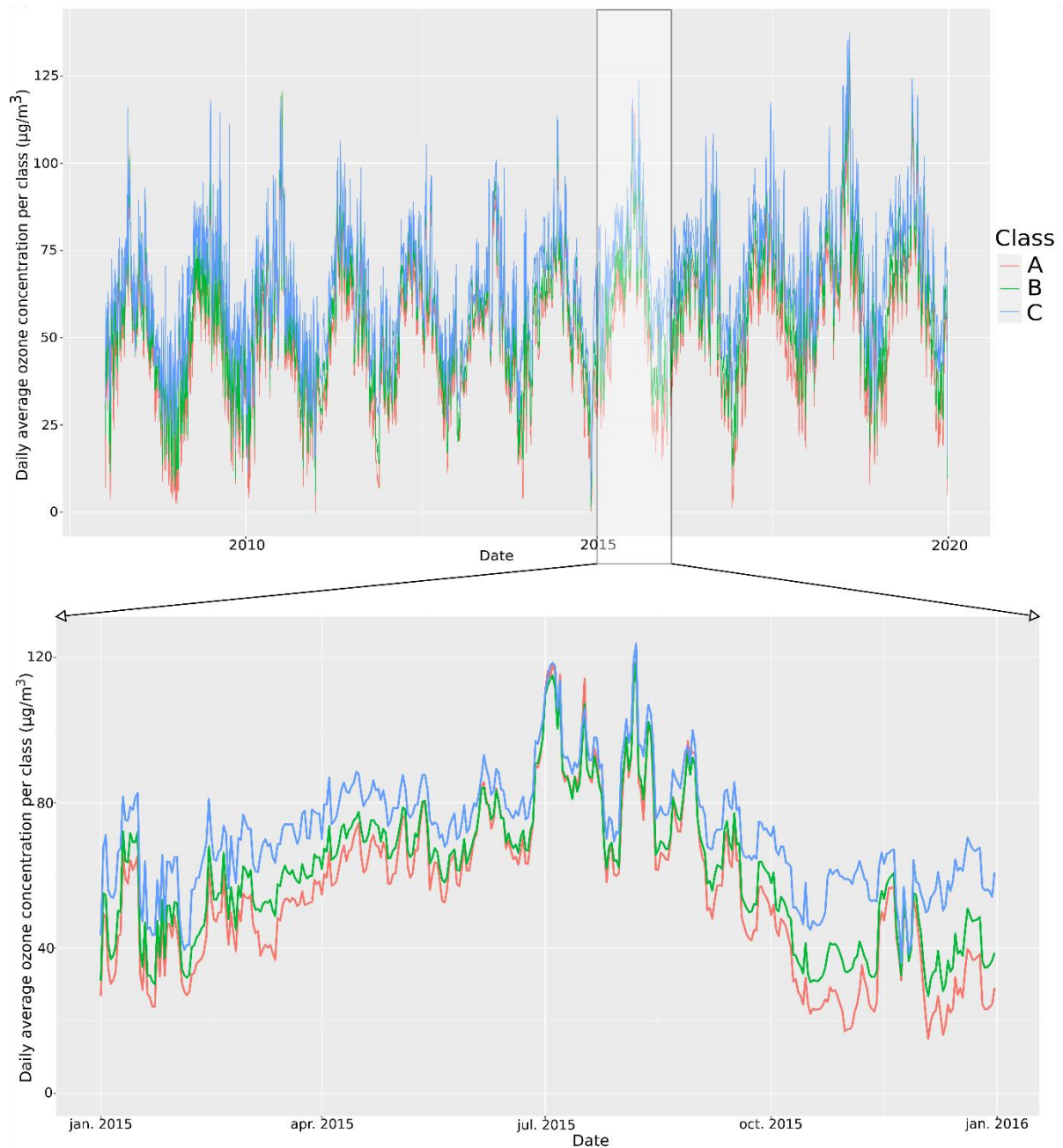


Figure 32: Time-series of the daily average of the mean ozone concentration per class (A, B, or C) from 2008 to 2019 on the top graph and in 2019 on the bottom graph.

### 3.2) Characterization and differentiation of the temporal dynamics

The classes differed significantly in 12-year mean O<sub>3</sub> levels and number of threshold exceedances, land use (proportion of urbanized and vegetated surface, traffic indicators), elevation and temporal dynamics (Table 5) (all p values < 0.001). The spectral entropy of class C ( $0.705 \pm 0.01$  (mean  $\pm$  SD)) was significantly higher than that of classes A ( $0.695 \pm 0.009$ ) and B ( $0.699 \pm 0.014$ ). This higher entropy

reflects its lower predictability. Class A, characterized by the highest proportion of urbanization (18%) and the lowest vegetation coverage (35%), was termed “urban”. Class C, which presented a less urbanized surface (urbanized surface = 7%, length of linear major road axis = 2.27 km), was termed rural. The third class (B), with intermediary urbanization and the highest vegetation cover, was termed the suburban class. The highest mean O<sub>3</sub> concentrations were observed in the rural class and then in the suburban class (Figure 33). In the study area, an increasing elevation was observed from urban, suburban to rural classes. The determinants of O<sub>3</sub> are strongly correlated. The independent effect of each determinant is addressed in the next section of this paper.

Table 5: Position and dispersion parameters of the different characterization variables per class.

| Characteristic  | All classes,<br>N = 126 | Class A,<br>N = 50 | Class B,<br>N = 43 | Class C,<br>N = 33 | Kruskal-Wallis<br>p-value |
|---|-------------------------|--------------------|--------------------|--------------------|---------------------------|
| <b>Land use</b>   |                         |                    |                    |                    |                           |
| <b>Percentage of vegetated surface (%)</b>                            |                         |                    |                    |                    |                           |
| Mean (SD)   | 45 (19)                 | 35 (16)            | 54 (18)            | 51 (16)            | <0.001                    |
| Range   | 10 - 92                 | 13 - 80            | 10 - 92            | 21 - 83            |                           |
| <b>Percentage of urbanized surface (%)</b>                            |                         |                    |                    |                    |                           |
| Mean (SD)   | 14 (14)                 | 18 (18)            | 13 (13)            | 7 (5)              | <0.001                    |
| Range   | 1 - 87                  | 2 - 87             | 2 - 55             | 1 - 22             |                           |
| <b>Total length of linear major road axis (km)</b>                    |                         |                    |                    |                    |                           |
| Mean (SD)   | 4.03 (3.25)             | 5.21 (3.68)        | 4.02 (2.73)        | 2.27 (2.28)        | <0.001                    |
| Range   | 0.00 - 20.37            | 0.00 - 20.37       | 0.00 - 10.18       | 0.00 - 9.01        |                           |
| <b>Cumulative index of road traffic and urban surface<sup>a</sup></b> |                         |                    |                    |                    |                           |
| Mean (SD)   | 22 (19)                 | 29 (23)            | 21(15)             | 12 (8)             | <0.001                    |
| Range   | 1 - 127                 | 4 - 127            | 3 - 69             | 1 - 27             |                           |
| <b>Elevation</b>  |                         |                    |                    |                    |                           |
| <b>Mean elevation (m)</b>   |                         |                    |                    |                    |                           |
| Mean (SD)   | 334 (103)               | 252 (29)           | 319 (49)           | 479 (68)           | <0.001                    |
| Range   | 210 - 598               | 210 - 362          | 246 - 409          | 385 - 598          |                           |

| Characteristic  | All classes,<br>N = 126 | Class A,<br>N = 50 | Class B,<br>N = 43 | Class C,<br>N = 33 | Kruskal-Wallis<br>p-value |
|---|-------------------------|--------------------|--------------------|--------------------|---------------------------|
| <b>Ozone levels</b>   |                         |                    |                    |                    |                           |
| <b>Average of the 12-year mean daily concentrations (<math>\mu\text{g}\cdot\text{m}^{-3}</math>)</b>                                    |                         |                    |                    |                    |                           |
| Mean (SD)   | 56.9 (5.7)              | 51.6 (1.4)         | 56.7 (2.2)         | 65.2 (2.0)         | <0.001                    |
| Range   | 48.1 - 69.1             | 48.1 - 54.8        | 53.5 - 61.3        | 61.0 - 69.1        |                           |
| <b>95<sup>th</sup> percentile of the 12-year mean daily concentrations (<math>\mu\text{g}\cdot\text{m}^{-3}</math>)</b>                 |                         |                    |                    |                    |                           |
| Mean (SD)   | 89.3 (4.3)              | 85.4 (1.4)         | 89.3 (1.9)         | 95.1 (2.4)         | <0.001                    |
| Range   | 82.5 - 99.0             | 82.5 - 89.7        | 84.5 - 92.5        | 89.8 - 99.0        |                           |
| <b>Number of days with threshold (<math>120 \mu\text{g}\cdot\text{m}^{-3}</math>) exceedances for mean daily ozone concentrations</b>   |                         |                    |                    |                    |                           |
| Mean (SD)   | 13 (8)                  | 6 (3)              | 12 (4)             | 24 (8)             | <0.001                    |
| Range   | 2 - 38                  | 2 - 12             | 5 - 21             | 12 - 38            |                           |
| <b>Average of the 12-year maximum daily concentrations (<math>\mu\text{g}\cdot\text{m}^{-3}</math>)</b>                                 |                         |                    |                    |                    |                           |
| Mean (SD)   | 82.12 (2.03)            | 80.37 (1.54)       | 83.40 (1.56)       | 83.11 (1.17)       | <0.001                    |
| Range   | 76.97 - 86.56           | 76.97 - 82.78      | 80.41 - 86.56      | 81.71 - 86.22      |                           |
| <b>95<sup>th</sup> percentile of the 12-year maximum daily concentrations (<math>\mu\text{g}\cdot\text{m}^{-3}</math>)</b>              |                         |                    |                    |                    |                           |
| Mean (SD)   | 130.2 (3.6)             | 130.2 (2.8)        | 132.7 (2.9)        | 127.2 (3.1)        | <0.001                    |
| Range   | 122.6 - 137.8           | 124.4 - 135.1      | 127.8 - 137.8      | 122.6 - 135.2      |                           |
| <b>Number of days with threshold (<math>180 \mu\text{g}\cdot\text{m}^{-3}</math>) exceedances for daily maximum ozone concentration</b> |                         |                    |                    |                    |                           |
| Mean (SD)   | 3 (2)                   | 2 (2)              | 4 (2)              | 2 (2)              | <0.001                    |
| Range   | 0 - 9                   | 1 - 7              | 1 - 9              | 0 - 8              |                           |
| <b>Number of statistical units (number of days x number of modeling meshes)</b>   | 539,658                 | 214,150            | 184,169            | 141,339            |                           |

a: defined by the sum of, the length of linear major road axis (km) weighted by their importance to the road traffic, and the percentage of urbanized area.

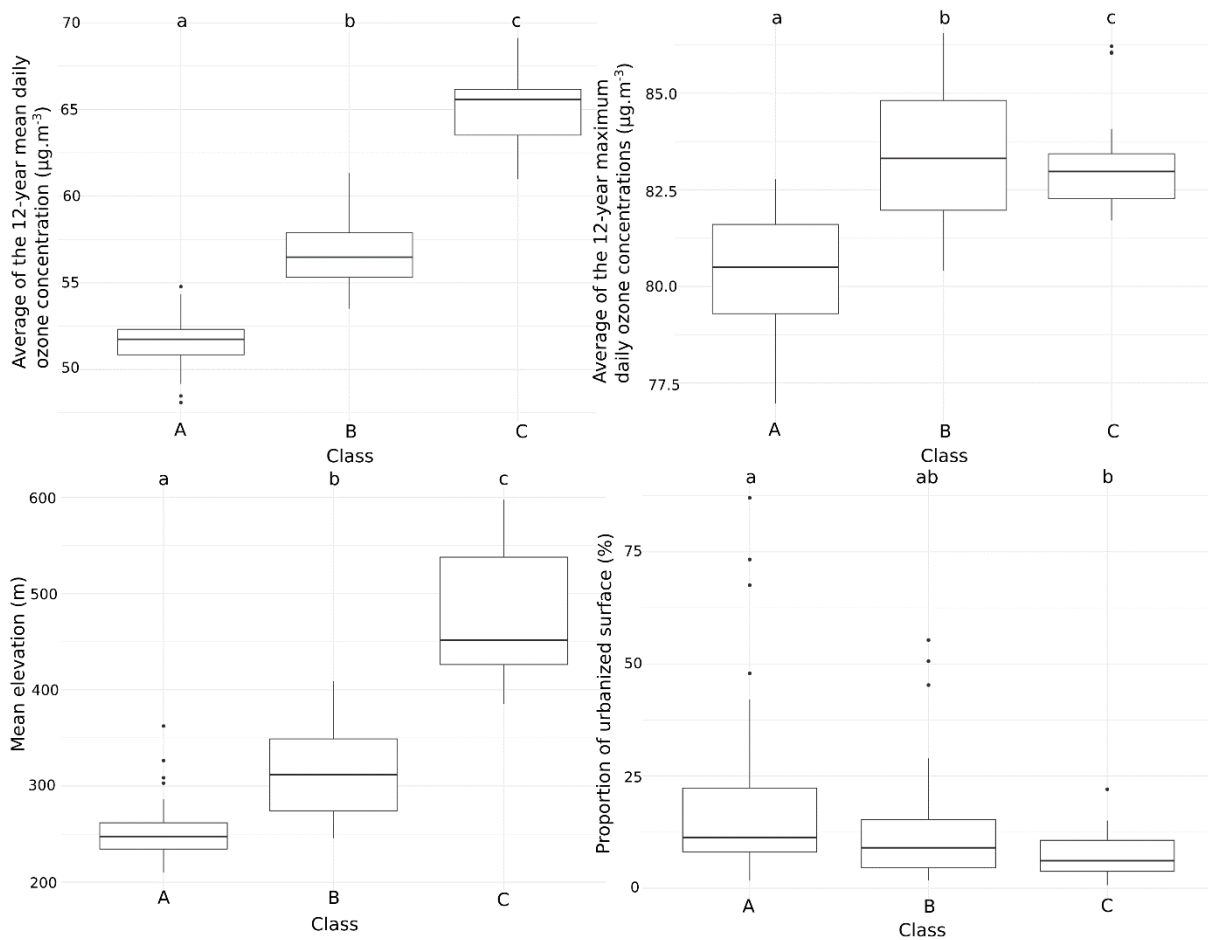


Figure 33: Boxplots of the average 12-year mean and maximum daily ozone concentrations, mean elevation and proportion of urbanized area, according to the class. The different letters above the boxplots represent significant differences between classes (Kruskal-Wallis test).

### 3.3) Relationship between O<sub>3</sub> levels and its determinants

A multivariable approach was also conducted on the 12-year mean O<sub>3</sub> concentrations. Not surprisingly, the proportion of urbanized land and the traffic indicators were negatively correlated with the O<sub>3</sub> concentration levels; the proportion of vegetated surface was positively correlated with O<sub>3</sub> (Figure 34 and Table 6). Elevation was strongly and positively correlated with O<sub>3</sub> concentrations. After adjusting for elevation (multivariate linear model taking the elevation into account, then adding land cover variables), urbanization remained significantly associated with O<sub>3</sub> levels, in contrast to the proportion of vegetated area (Table 6). Therefore, the relationship between elevation and 12-year mean O<sub>3</sub> appeared to be nonlinear (Figure 34), with a turning point (located at 484 m) above which the slope becomes negative. Beyond this threshold (484 m), the O<sub>3</sub> concentrations decrease.

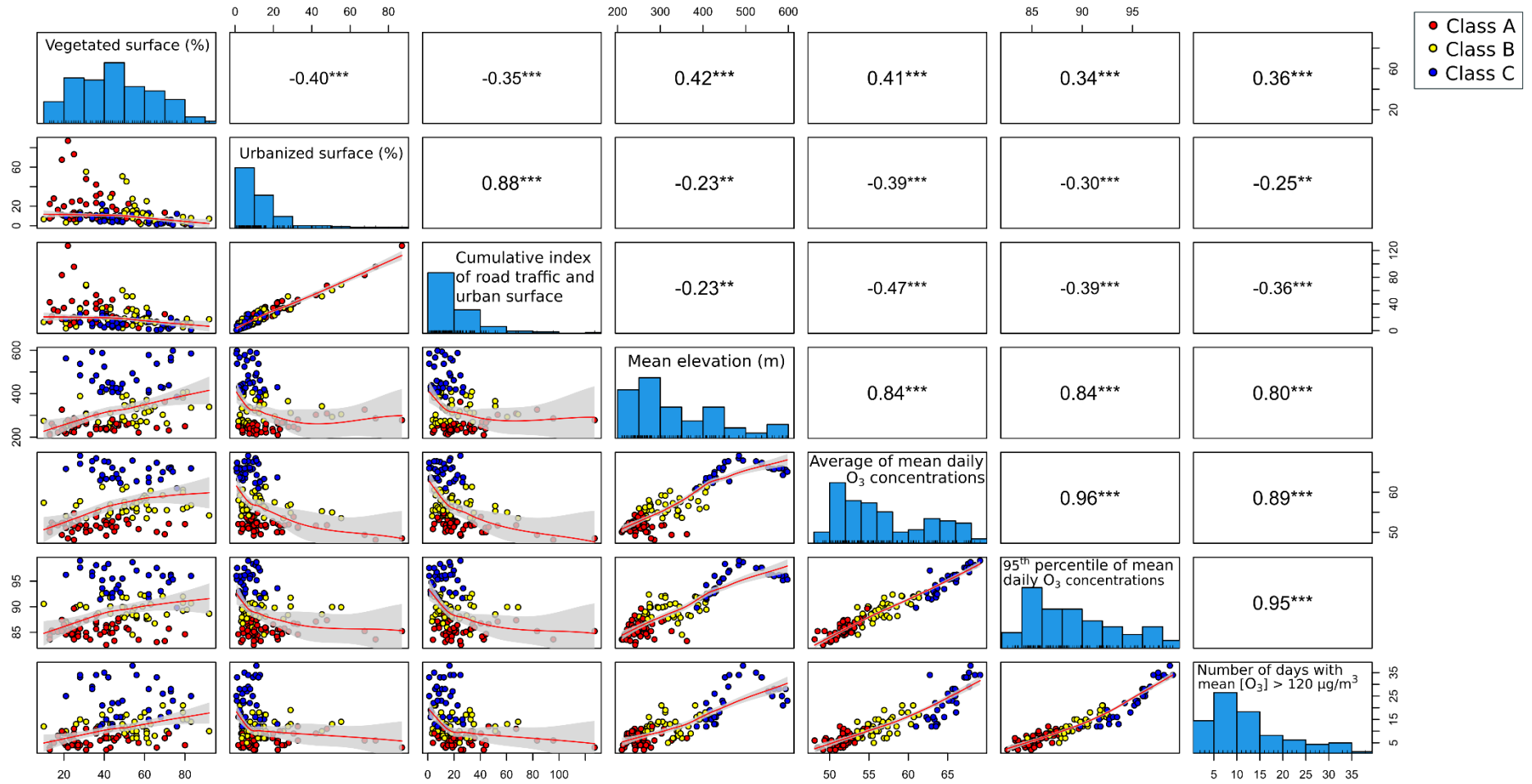


Figure 34: Correlation matrix (Spearman correlation) between the characteristics. The following significance levels (p values): 0, 0.001, 0.01, 0.05, 1 are displayed by the following symbols: "\*\*\*\*", "\*\*\*", "\*\*", "\*". The red line in the plots on the bottom left part of the matrix represents the local regression (loess smooth) surrounded by the 95% confidence interval.

Table 6: Results of the simple and multiple linear regression and polynomial model of the 12-year mean ozone concentration ( $\mu\text{g.m}^{-3}$ ) against the main driving factors.

| Model   | Coefficient(s)   | Standard Error(s)  | p-value(s)   | Adjusted R <sup>2</sup> |
|---|--|--|--|-------------------------|
| <b>Simple linear regression</b>   |  |  |  |                         |
| ~ elevation   | 0.049  | 0.002  | < 2.2e-16 ***  | 0.789                   |
| ~ percentage of urbanized area  | -0.156   | 0.033  | 6.67e-06 ***   | 0.145                   |
| ~ total length of linear major road axis  | -0.725   | 0.144  | 1.78e-06 ***   | 0.162                   |
| ~ cumulative index of road traffic and urban surface  | -0.134   | 0.024  | 1.23e-07 ***   | 0.196                   |
| ~ percentage of surface covered by vegetation   | 0.103  | 0.026  | 0.000125 ***   | 0.105                   |
| <b>Multiple linear regression</b>   |  |  |  |                         |
| ~ elevation + percentage of urbanized area  | elevation 0.047<br>urbanized area -0.083                           | elevation 0.002<br>urbanized area 0.015                          | elevation. < 2e-16***<br>urbanized area. 2.29e-07 ***                  | 0.829                   |
| ~ elevation + total length of linear major road axis  | elevation 0.047<br>linear of roads -0.390                          | elevation 0.002<br>linear of roads 0.066                         | elevation < 2e-16 ***<br>linear of roads 3.12e-08 ***                  | 0.834                   |
| ~ elevation + cumulative index of road traffic and urban surface  | elevation 0.046<br>cumulative index -0.07                          | elevation 0.002<br>cumulative index 0.011                        | elevation < 2e-16 ***<br>cumulative index 1.51e-10 ***                 | 0.848                   |
| ~ elevation + percentage of surface covered by vegetation   | elevation 0.049<br>vegetation 0.004                                | elevation 0.002<br>vegetation 0.014                              | elevation < 2e-16 ***<br>vegetation 0.743                              | 0.788                   |
| ~ elevation + percentage of urbanized area + total length of linear major road axis                                       | elevation 0.046<br>urbanized area -0.046<br>linear of roads -0.264 | elevation 0.002<br>urbanized area 0.019<br>linear of roads 0.082 | elevation < 2e-16 ***<br>urbanized area 0.014<br>linear of roads 0.002 | 0.841                   |
| <b>Polynomial</b>   |  |  |  |                         |
| ~ centered elevation + (centered elevation) <sup>2</sup>  | 9.733e-02<br>-6.263e-05  | 1.636e-02<br>2.117e-05   | 2.620e-08<br>0.004   | 0.801                   |
| ~ centered elevation + (centered elevation) <sup>2</sup> + centered elevation <sup>3</sup>                                | -3.934e-01<br>1.238e-03<br>-1.089e-06                              | 9.498e-02<br>2.494e-04<br>2.083e-07                              | 6.390e-05<br>2.270e-06<br>7.120e-07                                    | 0.837                   |
| ~ centered elevation + (centered elevation) <sup>2</sup> + centered elevation <sup>3</sup> + percentage of urbanized area | -2.635e-01<br>9.017e-04<br>-8.194e-07<br>-7.652e-02                | 8.848e-02<br>2.321e-04<br>1.933e-07<br>1.389e-02                 | 0.004<br>1.670e-04<br>4.430e-05<br>2.080e-07                           | 0.868                   |



The complex relationship between the O<sub>3</sub> concentration and the different determinants is illustrated in Figure 35. Ozone levels and, not surprisingly, the classes were spatially structured (p value < 0.05 (Figure A.6) and p-value < 0.001, respectively). The elevation was also highly spatially structured (p value < 0.05, Figure A.6) and presented a gradual increase in the bottom right corner of the map Figure 35). This increase was associated with an increase in O<sub>3</sub> levels, to an optimum at 484 m elevation and then a slight decrease.

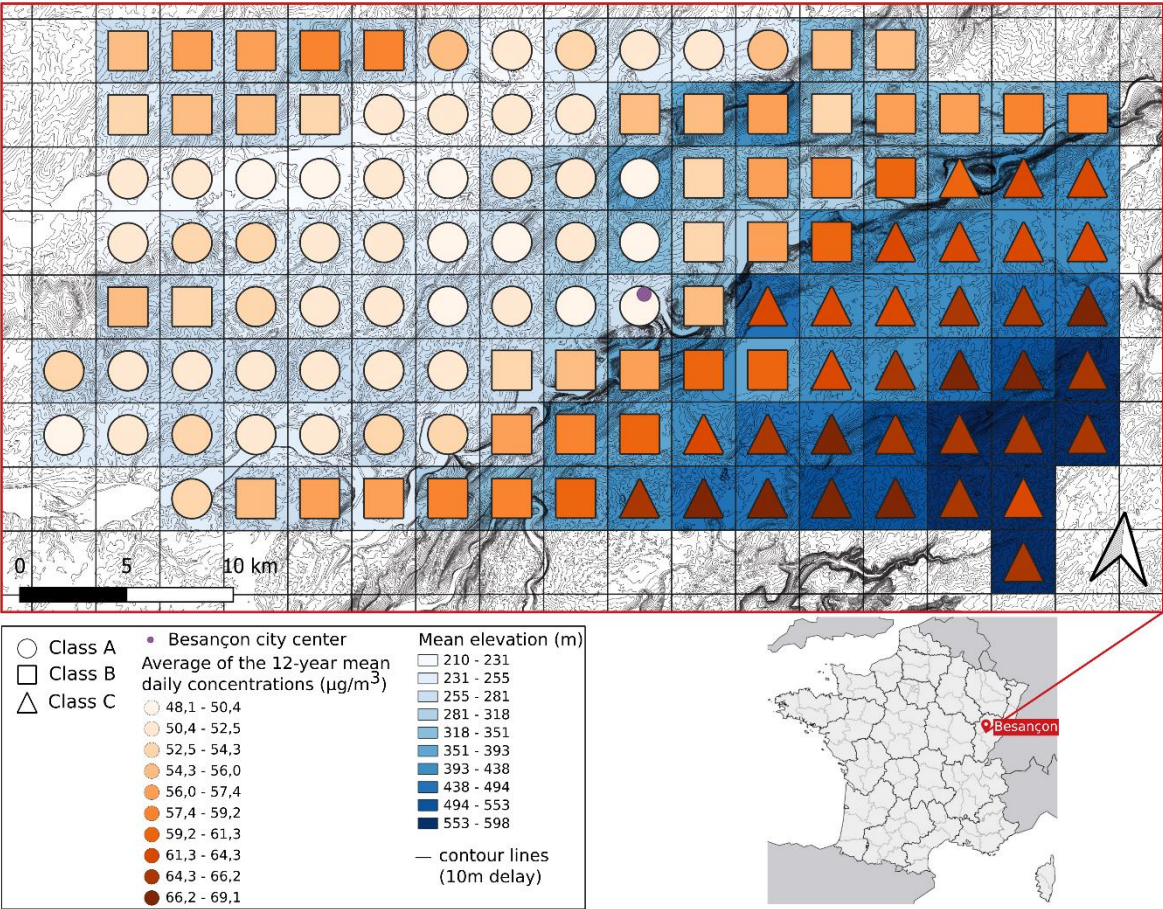


Figure 35: Spatial distribution of the classes (represented by different shapes) through the modeling grid. The color filling the shapes (orange levels) represents the average of the daily mean ozone concentration (2008-2019) per modeling mesh. The background color represents the mean elevation of the modeling mesh.

**4) Discussion:**

4.1) Reliability of the data and the method

Based on the temporal dynamics of daily O<sub>3</sub> concentrations modeled at a 9 km<sup>2</sup> resolution, 3 classes of different temporal dynamics were identified thanks to DTW-based clustering. In addition to the O<sub>3</sub>

levels, these classes also differed in land use and elevation. The spatial distribution of urbanization and elevation led to a strong spatial structuration of O<sub>3</sub> levels and therefore of classes.

The modeling platform provided 12 years of daily concentrations continuously through a large territory ( $\approx 1000 \text{ km}^2$ ). These concentrations benefited from great reliability (precision and exactitude), thanks to a prior correction process (Gauthier-Manuel et al., 2022). Then, the time series classification, without any preconceptions, based on the DTW measure revealed the temporal dynamics of each modeling mesh based on a 12-year study period (comprising 12 seasonal cycles). It captured the general dynamics of each mesh based on 12-year daily data, not reflecting the effect of punctual or seasonal phenomena. DTW can be sensitive to noisy data and outliers. However, our approach benefited from a large dataset (12 years of daily data on 126 modeling grid meshes). As suggested by Aghabozorgi et al. (2015) the size of the dataset impacts greatly the effectiveness of the measure computation. The size of our dataset should have reduced the influence of noise and outliers on alignment accuracy (Alqahtani et al., 2021). The characterization of the modeling meshes (vegetation, road networks, elevation, urbanized surface) was based on standardized and updated spatialized databases (BDALTI®, BD Topo®, Route500®, Urban atlas). The availability on a spatial continuum and the reliability of these data conditioned their selection to be considered as determinants of O<sub>3</sub> dynamics in this study. These determinants are strongly correlated and we cannot exclude the simultaneous influence of meteorological parameters on O<sub>3</sub> concentrations. However, the spatial heterogeneity of solar radiation or prevailing wind direction and speed could be considered low across our study area, and very low across the modeling meshes. It is unlikely that a confounding effect between land use and weather factors could have strongly influenced our results.

#### 4.2) Influence of the determinants on O<sub>3</sub> levels

As illustrated by our results, land use has multiple implications regarding the spatial distribution of O<sub>3</sub> concentrations. Anthropogenic sources of NO<sub>x</sub> and VOC are concentrated in urban areas due to traffic and industrial activities. In a simplified way, in the absence of VOCs, the photoequilibrium is respected, and the net production of O<sub>3</sub> is null. This equilibrium is broken in the presence of VOCs

which convert NO to NO<sub>2</sub> and hence fuel O<sub>3</sub> production. The interconversion between NO and NO<sub>2</sub> is much faster than that producing O<sub>3</sub> (O<sub>2</sub> + O = O<sub>3</sub>). The transport of air masses charged with pollutants from city centers to peripheral areas and the time needed to produce O<sub>3</sub> lead to higher O<sub>3</sub> concentrations in suburban and surrounding rural areas.

The increasing distance to anthropogenic sources is also associated with an increase in vegetated surfaces. Vegetation is a biogenic source of VOCs and can favor O<sub>3</sub> concentration increases (Paoletti, 2009). Additionally, trees moderate temperature by providing shade and have the ability to reduce O<sub>3</sub> concentrations by O<sub>3</sub> deposition on leaves and through stomatal absorption (Fitzky et al., 2019). The net impact of vegetation on O<sub>3</sub> concentrations remains complex to quantify and depends on the type of vegetation (Sicard et al., 2018) and the climatic conditions (Peñuelas and Staudt, 2010; Trenberth et al., 2014). In this study, the influence of vegetation appeared lower than other determinants. This could partially be explained both by an inverse correlation with urbanization and by the complexity of the processes by which vegetation impacts O<sub>3</sub> levels.

The well-known positive association between elevation and O<sub>3</sub> concentrations is related to higher solar radiation and biogenic volatile organic compounds (BVOCs) emissions (Kulkarni et al., 2011; Lefohn et al., 2012; Sicard et al., 2017). The vegetation and especially forest cover is generally denser on the reliefs leading to higher BVOC emissions.

The local characteristics of our area, especially topography, reinforced the O<sub>3</sub> concentration increase from urban to rural areas. The bottom-right corner of the study area illustrates the result of such a complex relationship between determinants. We observed an inflection point at 484 meters of elevation, above which O<sub>3</sub> concentrations slightly decrease, although elevation keeps increasing; and three potential explanations could be proposed. First, the distance from the main NO<sub>x</sub> and anthropogenic VOC sources (centered on Besançon), although low compared to megacities emissions, is increasing leading to less influence of O<sub>3</sub> precursors. Second, there is a hill followed by a valley whose orientation is the same as the prevailing wind, at the same location as the observed turning point in O<sub>3</sub> concentrations. This relief could influence the dispersion of O<sub>3</sub>. The hill presents a dense forest cover,

while the valley is mainly occupied by agricultural activities (pastures). Third, the mean elevation of the modeling meshes might not reflect exactly this complexity and the variability in the local topography.

#### 4.3) Complexity of O<sub>3</sub> dynamics and management implications

All the determinants cited above act simultaneously on O<sub>3</sub> levels through complex processes. Although the chemical mechanisms at the origin of O<sub>3</sub> formation are independently well understood, the understanding of their joint effects on O<sub>3</sub> levels remains challenging (Monks et al., 2015). Our approach allowed us to synthesize and visualize the resultant of all these processes and decompose the respective influence of each determinant. The conjunction of all these influences of different amplitudes leads to the picture captured in Figure 35. To our knowledge, the spatial and temporal dynamics of the O<sub>3</sub> concentration are poorly described continuously in space. Meng et al. (2022) provided ozone spatial and temporal analysis over China using high spatial and temporal precision over 7 years. However, they analyzed the O<sub>3</sub> temporal variability at monthly and annual scales. We integrated the temporal variations in O<sub>3</sub> at the daily scale and identified its main determinants. This daily scale was essential to detect short-term variation in O<sub>3</sub> concentrations and potential O<sub>3</sub> peaks, and to better differentiate the temporal dynamics. To our knowledge, such an approach based on a spatial continuum has not already been reported. Existing studies relied on measurements and aimed at understanding the relationship between O<sub>3</sub> and atmospheric dynamics (Beaver and Palazoğlu, 2006; Di Bernardino et al., 2023). Using modeled data allowed us to define our clusters on a spatial continuum while using measurements would have been challenging to delineate the territory over which a station's data are valid.

The predictability of ozone time series was lower in rural areas than in urban and suburban areas, characterizing the higher difficulty of forecasting O<sub>3</sub> levels in rural areas. Our results also highlighted the O<sub>3</sub> paradox of the air pollution assessment: suburban and rural areas are both the more exposed and, the least monitored and subject to prevention acts (Madrigano et al., 2015). In this study, rural areas presented the highest number of threshold exceedances for the mean ozone concentration (24 days above 120 µg.m<sup>-3</sup>). This O<sub>3</sub> paradox combined with the higher difficulty in forecasting O<sub>3</sub> concentrations illustrates the additional difficulty in managing rural ozone exposure. Our results highlight a potential

vulnerability of these rural areas. The vulnerability of the rural areas regarding ozone exposure arises from four factors. First, the rural population experiences higher exposure levels. Second, these areas are subject to less monitoring of air quality implying less identification of the exceedance of the alert threshold. Third, there is a potentially higher proportion of elders, more vulnerable to ozone exposure, in rural areas. At the national scale (in France), the rural areas host 25% of people aged 65 years and older, compared to 18% in urban areas (D'Alessandro et al., 2021). Fourth, people spend potentially more time outside in rural areas (Matz et al., 2015). This multifactorial vulnerability of rural areas warrants the need for better monitoring of these areas.

The urban–suburban–rural typology is not always consistent with different ozone dynamics or levels because of the direction and strength of dominant winds that can transport air pollutants and affect differently areas belonging to the same typology. Considering functional exposure areas identified using our approach, rather than an urban–suburban–rural typology, would be relevant. Among these functional areas, we identified rural areas as the most exposed and potentially vulnerable. Our approach provides new insight concerning the way to objectify different daily O<sub>3</sub> dynamics and then different exposure areas. This could improve the environmental and health risk assessment and land-use management. The spatial determination of exposure areas has strong implications in terms of land-use management for decision-makers and planners. Prevention acts should be implemented to inform rural populations during O<sub>3</sub> peaks and adapt forest and crop management planning. According to climate scenario RCP 8.5, described by the IPCC (Yang et al., 2019), ozone concentrations will continue to increase in the future. In this context, it appears essential to anticipate the distribution of this risk and its potential consequences.

## **5) Conclusion:**

Different daily dynamics of O<sub>3</sub> concentrations have been identified within the same geographical area, showing a strong spatial structuration. Local characteristics played a key role in the O<sub>3</sub> spatial and temporal dynamics, but the actual levels remain hardly predictable, especially in rural areas. This study proposed a pragmatic, systematic, and reproducible way to build exposure area mapping and could improve mitigation strategies in all fields concerned with O<sub>3</sub> impacts. These impacts are and will be

amplified worldwide under climate change influence. It would be relevant to identify to which extent heat waves can and will modify the O<sub>3</sub> dynamics and gradients observed in this paper because this could have strong implications for population exposure management.

### **Acknowledgements:**

The authors wish to thank everyone who contributed to the development of the project and, the agglomeration community of Grand Besançon Métropole (PhD scholarship, 2019) and the “Pôle fédératif de recherche et de formation en santé publique Bourgogne-Franche-Comté” for providing financial support.

### **References:**

- Agathokleous, S., Saitanis, C.J., Savvides, C., Sicard, P., Agathokleous, E., De Marco, A., 2022. Spatiotemporal variations of ozone exposure and its risks to vegetation and human health in Cyprus: an analysis across a gradient of altitudes. *J. For. Res.* <https://doi.org/10.1007/s11676-022-01520-2>
- Aghabozorgi, S., Seyed Shirخورshidi, A., Ying Wah, T., 2015. Time-series clustering – A decade review. *Information Systems* 53, 16–38. <https://doi.org/10.1016/j.is.2015.04.007>
- Alqahtani, A., Ali, M., Xie, X., Jones, M.W., 2021. Deep Time-Series Clustering: A Review. *Electronics* 10, 3001. <https://doi.org/10.3390/electronics10233001>
- Beaver, S., Palazoğlu, A., 2006. A cluster aggregation scheme for ozone episode selection in the San Francisco, CA Bay Area. *Atmospheric Environment* 40, 713–725. <https://doi.org/10.1016/j.atmosenv.2005.10.003>
- Carro-Calvo, L., Ordóñez, C., García-Herrera, R., Schnell, J.L., 2017. Spatial clustering and meteorological drivers of summer ozone in Europe. *Atmospheric Environment* 167, 496–510. <https://doi.org/10.1016/j.atmosenv.2017.08.050>
- Copernicus Land Monitoring Service, 2022. Urban Atlas [WWW Document]. URL <https://land.copernicus.eu/local/urban-atlas> (accessed 8.1.22).

D'Alessandro, C., Levy, D., Regnier, T., 2021. Une nouvelle définition du rural pour mieux rendre compte des réalités des territoires et de leurs transformations – La France et ses territoires | Insee. Insee.

Derstroff, B., Hüser, I., Bourtsoukidis, E., Crowley, J.N., Fischer, H., Gromov, S., Harder, H., Janssen, R.H.H., Kesselmeier, J., Lelieveld, J., Mallik, C., Martinez, M., Novelli, A., Parchatka, U., Phillips, G.J., Sander, R., Sauvage, C., Schuladen, J., Stöner, C., Tomsche, L., Williams, J., 2017. Volatile organic compounds (VOCs) in photochemically aged air from the eastern and western Mediterranean. *Atmospheric Chemistry and Physics* 17, 9547–9566. <https://doi.org/10.5194/acp-17-9547-2017>

Derwent, R.G., Utembe, S.R., Jenkin, M.E., Shallcross, D.E., 2015. Tropospheric ozone production regions and the intercontinental origins of surface ozone over Europe. *Atmospheric Environment* 112, 216–224. <https://doi.org/10.1016/j.atmosenv.2015.04.049>

Di Bernardino, A., Iannarelli, A.M., Casadio, S., Pisacane, G., Siani, A.M., 2023. Spatial-temporal assessment of air quality in Rome (Italy) based on anemological clustering. *Atmospheric Pollution Research* 14, 101670. <https://doi.org/10.1016/j.apr.2023.101670>

Dijkstra, L., Poelman, H., 2012. CITIES IN EUROPE THE NEW OECD-EC DEFINITION (No. RF 01/2012), Regional and Urban Policy. OECD-EC.

Duarte, A.L., Schneider, I.L., Artaxo, P., Oliveira, M.L.S., 2022. Spatiotemporal assessment of particulate matter (PM<sub>10</sub> and PM<sub>2.5</sub>) and ozone in a Caribbean urban coastal city. *Geoscience Frontiers* 13, 101168. <https://doi.org/10.1016/j.gsf.2021.101168>

EEA, 2015. Air quality in Europe: 2015 report. European Environmental Agency. (No. 5/2015). Publications Office, LU. <https://doi.org/10.2800/62459>

Fan, H., Zhao, C., Yang, Y., 2020. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. *Atmospheric Environment* 220, 117066. <https://doi.org/10.1016/j.atmosenv.2019.117066>

Feng, Z., De Marco, A., Anav, A., Gualtieri, M., Sicard, P., Tian, H., Fornasier, F., Tao, F., Guo, A., Paoletti, E., 2019. Economic losses due to ozone impacts on human health, forest productivity and crop

yield across China. *Environment International* 131, 104966.  
<https://doi.org/10.1016/j.envint.2019.104966>

Fitzky, A.C., Sandén, H., Karl, T., Fares, S., Calfapietra, C., Grote, R., Saunier, A., Rewald, B., 2019. The Interplay Between Ozone and Urban Vegetation—BVOC Emissions, Ozone Deposition, and Tree Ecophysiology. *Frontiers in Forests and Global Change* 2, 50. <https://doi.org/10.3389/ffgc.2019.00050>

Gao, C., Xiu, A., Zhang, X., Chen, W., Liu, Y., Zhao, H., Zhang, S., 2020. Spatiotemporal characteristics of ozone pollution and policy implications in Northeast China. *Atmospheric Pollution Research* 11, 357–369. <https://doi.org/10.1016/j.apr.2019.11.008>

Garcia, V.C., Foley, K.M., Gego, E., Holland, D.M., Rao, S.T., 2010. A Comparison of Statistical Techniques for Combining Modeled and Observed Concentrations to Create High-Resolution Ozone Air Quality Surfaces. *Journal of the Air & Waste Management Association* 60, 586–595. <https://doi.org/10.3155/1047-3289.60.5.586>

Gauthier-Manuel, H., Mauny, F., Boilleaut, M., Ristori, M., Pujol, S., Vasbien, F., Parmentier, A.-L., Bernard, N., 2022. Improvement of downscaled ozone concentrations from the transnational scale to the kilometric scale: Need, interest and new insights. *Environmental Research* 210, 112947. <https://doi.org/10.1016/j.envres.2022.112947>

Giorgino, T., 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software* 31, 1–24. <https://doi.org/10.18637/jss.v031.i07>

Huang, D., Li, Q., Wang, X., Li, G., Sun, L., He, B., Zhang, L., Zhang, C., 2018. Characteristics and Trends of Ambient Ozone and Nitrogen Oxides at Urban, Suburban, and Rural Sites from 2011 to 2017 in Shenzhen, China. *Sustainability* 10, 4530. <https://doi.org/10.3390/su10124530>

Hyndman, R.J., Athanasopoulos, G., 2021. *Forecasting: principles and practice*.

IPCC (Ed.), 2014. *Climate change 2014: impacts, adaptation, and vulnerability: Working Group II contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, United Kingdom and New York, NY.



- Javed, A., Lee, B.S., Rizzo, D.M., 2020. A benchmark study on time series clustering. *Machine Learning with Applications* 1, 100001. <https://doi.org/10.1016/j.mlwa.2020.100001>
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. *J Expo Sci Environ Epidemiol* 15, 185–204. <https://doi.org/10.1038/sj.jea.7500388>
- Kulkarni, P.S., Bortoli, D., Salgado, R., Antón, M., Costa, M.J., Silva, A.M., 2011. Tropospheric ozone variability over the Iberian Peninsula. *Atmospheric Environment* 45, 174–182. <https://doi.org/10.1016/j.atmosenv.2010.09.029>
- Lacour, S.A., de Monte, M., Diot, P., Brocca, J., Veron, N., Colin, P., Leblond, V., 2006. Relationship between ozone and temperature during the 2003 heat wave in France: consequences for health data analysis. *BMC Public Health* 6, 261. <https://doi.org/10.1186/1471-2458-6-261>
- Lefohn, A.S., Wernli, H., Shadwick, D., Oltmans, S.J., Shapiro, M., 2012. Quantifying the importance of stratospheric-tropospheric transport on surface ozone concentrations at high- and low-elevation monitoring sites in the United States. *Atmospheric Environment* 62, 646–656. <https://doi.org/10.1016/j.atmosenv.2012.09.004>
- Liu, H., Liu, J., Liu, Y., Yi, K., Yang, H., Xiang, S., Ma, J., Tao, S., 2021. Spatiotemporal variability and driving factors of ground-level summertime ozone pollution over eastern China. *Atmospheric Environment* 265, 118686. <https://doi.org/10.1016/j.atmosenv.2021.118686>
- Liu, Z., Xu, P., Gong, F., Tan, Y., Han, J., Tian, L., Yan, J., Li, K., Xi, Z., Liu, X., 2022. Altered lipidomic profiles in lung and serum of rat after sub-chronic exposure to ozone. *Science of The Total Environment* 806, 150630. <https://doi.org/10.1016/j.scitotenv.2021.150630>
- Ma, M., Yao, G., Guo, J., Bai, K., 2021. Distinct spatiotemporal variation patterns of surface ozone in China due to diverse influential factors. *Journal of Environmental Management* 288, 112368. <https://doi.org/10.1016/j.jenvman.2021.112368>

- Ma, R., 2020. Clustering of Time Series Data: Measures, Methods, and Applications. Georgia State University. <https://doi.org/10.57709/20253148>
- Madrigano, J., Jack, D., Anderson, G.B., Bell, M.L., Kinney, P.L., 2015. Temperature, ozone, and mortality in urban and non-urban counties in the northeastern United States. *Environmental Health* 14, 3. <https://doi.org/10.1186/1476-069X-14-3>
- Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G., Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., Meleux, F., 2017. CHIMERE-2017: from urban to hemispheric chemistry-transport modeling. *Geoscientific Model Development* 10, 2397–2423. <https://doi.org/10.5194/gmd-10-2397-2017>
- Matz, C.J., Stieb, D.M., Brion, O., 2015. Urban-rural differences in daily time-activity patterns, occupational activity and housing characteristics. *Environ Health* 14, 88. <https://doi.org/10.1186/s12940-015-0075-y>
- Meng, X., Wang, W., Shi, S., Zhu, S., Wang, P., Chen, R., Xiao, Q., Xue, T., Geng, G., Zhang, Q., Kan, H., Zhang, H., 2022. Evaluating the spatiotemporal ozone characteristics with high-resolution predictions in mainland China, 2013–2019. *Environmental Pollution* 299, 118865. <https://doi.org/10.1016/j.envpol.2022.118865>
- Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., Williams, M.L., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics* 15, 8889–8973. <https://doi.org/10.5194/acp-15-8889-2015>
- Montero, P., Vilar, J.A., 2015. TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software* 62, 1–43. <https://doi.org/10.18637/jss.v062.i01>
- Nguyen, D.-H., Lin, C., Vu, C.-T., Cheruiyot, N.K., Nguyen, M.K., Le, T.H., Lukkhasorn, W., Vo, T.-D.-H., Bui, X.-T., 2022. Tropospheric ozone and NO<sub>x</sub>: A review of worldwide variation and

meteorological influences. *Environmental Technology & Innovation* 28, 102809.  
<https://doi.org/10.1016/j.eti.2022.102809>

Notario, A., Bravo, I., Adame, J.A., Díaz-de-Mera, Y., Aranda, A., Rodríguez, A., Rodríguez, D., 2013. Variability of oxidants (OX=O<sub>3</sub>+NO<sub>2</sub>), and preliminary study on ambient levels of ultrafine particles and VOCs, in an important ecological area in Spain. *Atmospheric Research* 128, 35–45.  
<https://doi.org/10.1016/j.atmosres.2013.03.008>

Otero, N., Sillmann, J., Schnell, J.L., Rust, H.W., Butler, T., 2016. Synoptic and meteorological drivers of extreme ozone concentrations over Europe. *Environ. Res. Lett.* 11, 024005.  
<https://doi.org/10.1088/1748-9326/11/2/024005>

Paoletti, E., 2009. Ozone and urban forests in Italy. *Environmental Pollution, Special Issue Section: Ozone and Mediterranean Ecology: Plants, People, Problems* 157, 1506–1512.  
<https://doi.org/10.1016/j.envpol.2008.09.019>

Pascal, M., Wagner, V., Alari, A., Corso, M., Le Tertre, A., 2021. Extreme heat and acute air pollution episodes: A need for joint public health warnings? *Atmospheric Environment* 249, 118249.  
<https://doi.org/10.1016/j.atmosenv.2021.118249>

Peñuelas, J., Staudt, M., 2010. BVOCs and global change. *Trends in Plant Science* 15, 133–144.  
<https://doi.org/10.1016/j.tplants.2009.12.005>

Powers, J.G., Klemp, J.B., Skamarock, W.C., Davis, C.A., Dudhia, J., Gill, D.O., Coen, J.L., Gochis, D.J., Ahmadov, R., Peckham, S.E., Grell, G.A., Michalakes, J., Trahan, S., Benjamin, S.G., Alexander, C.R., Dimego, G.J., Wang, W., Schwartz, C.S., Romine, G.S., Liu, Z., Snyder, C., Chen, F., Barlage, M.J., Yu, W., Duda, M.G., 2017. The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *Bulletin of the American Meteorological Society* 98, 1717–1737.  
<https://doi.org/10.1175/BAMS-D-15-00308.1>

Pyrgou, A., Hadjinicolaou, P., Santamouris, M., 2018. Enhanced near-surface ozone under heatwave conditions in a Mediterranean island. *Sci Rep* 8, 9191. <https://doi.org/10.1038/s41598-018-27590-z>

Querol, X., Alastuey, A., Gangoiti, G., Perez, N., Lee, H.K., Eun, H.R., Park, Y., Mantilla, E., Escudero, M., Titos, G., Alonso, L., Temime-Roussel, B., Marchand, N., Moreta, J.R., Revuelta, M.A., Salvador, P., Artíñano, B., García dos Santos, S., Anguas, M., Notario, A., Saiz-Lopez, A., Harrison, R.M., Millán, M., Ahn, K.-H., 2018. Phenomenology of summer ozone episodes over the Madrid Metropolitan Area, central Spain. *Atmospheric Chemistry and Physics* 18, 6511–6533. <https://doi.org/10.5194/acp-18-6511-2018>

Ratanamahatana, C.A., Keogh, E., 2004. Making Time-series Classification More Accurate Using Learned Constraints, in: *Proceedings of the 2004 SIAM International Conference on Data Mining. Presented at the Proceedings of the 2004 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics*, pp. 11–22. <https://doi.org/10.1137/1.9781611972740.2>

Reche, C., Moreno, T., Amato, F., Pandolfi, M., Pérez, J., de la Paz, D., Diaz, E., Gómez-Moreno, F.J., Pujadas, M., Artíñano, B., Reina, F., Orio, A., Pallarés, M., Escudero, M., Tapia, O., Crespo, E., Vargas, R., Alastuey, A., Querol, X., 2018. Spatio-temporal patterns of high summer ozone events in the Madrid Basin, Central Spain. *Atmospheric Environment* 185, 207–220. <https://doi.org/10.1016/j.atmosenv.2018.05.002>

Roberts, H.R., Dodd, I.C., Hayes, F., Ashworth, K., 2022. Chronic tropospheric ozone exposure reduces seed yield and quality in spring and winter oilseed rape. *Agricultural and Forest Meteorology* 316, 108859. <https://doi.org/10.1016/j.agrformet.2022.108859>

Roelofsen, P., 2018. Time series clustering (Master thesis Business Analytics). Vrije Universiteit Amsterdam Faculty of Science; PricewaterhouseCoopers Advisory N.V. Consulting - Data Analytics.

Royal Society (Great Britain), 2008. Ground-level ozone in the 21st century: future trends, impacts and policy implications. The Royal Society, London.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>

- Screpanti, A., De Marco, A., 2009. Corrosion on cultural heritage buildings in Italy: A role for ozone? *Environmental Pollution* 157, 1513–1520. <https://doi.org/10.1016/j.envpol.2008.09.046>
- Sicard, P., Agathokleous, E., Araminiene, V., Carrari, E., Hoshika, Y., De Marco, A., Paoletti, E., 2018. Should we see urban trees as effective solutions to reduce increasing ozone levels in cities? *Environmental Pollution* 243, 163–176. <https://doi.org/10.1016/j.envpol.2018.08.049>
- Sicard, P., Anav, A., De Marco, A., Paoletti, E., 2017. Projected global ground-level ozone impacts on vegetation under different emission and climate scenarios. *Atmos. Chem. Phys.* 17, 12177–12196. <https://doi.org/10.5194/acp-17-12177-2017>
- Sicard, P., De Marco, A., Carrari, E., Dalstein-Richier, L., Hoshika, Y., Badea, O., Pitar, D., Fares, S., Conte, A., Popa, I., Paoletti, E., 2020. Epidemiological derivation of flux-based critical levels for visible ozone injury in European forests. *J. For. Res.* 31, 1509–1519. <https://doi.org/10.1007/s11676-020-01191-x>
- Sicard, P., Serra, R., Rossello, P., 2016. Spatiotemporal trends in ground-level ozone concentrations and metrics in France over the time period 1999–2012. *Environmental Research* 149, 122–144. <https://doi.org/10.1016/j.envres.2016.05.014>
- Siegel, S., Castellan, N.J., 1988. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill.
- Trenberth, K.E., Dai, A., van der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., Sheffield, J., 2014. Global warming and changes in drought. *Nature Clim Change* 4, 17–22. <https://doi.org/10.1038/nclimate2067>
- Upadhaya, P., Du, H., Kommalapati, R.R., 2020. Meteorological Detrending of Ozone at Three Sites in the Dallas-Fort Worth Area: Application of KZ Filter Method. *Atmosphere* 11, 1226. <https://doi.org/10.3390/atmos11111226>
- Vlachokostas, Ch., Nastis, S.A., Achillas, Ch., Kalogeropoulos, K., Karmiris, I., Moussiopoulos, N., Chourdakis, E., Baniyas, G., Limperi, N., 2010. Economic damages of ozone air pollution to crops using

- combined air quality and GIS modelling. *Atmospheric Environment* 44, 3352–3361. <https://doi.org/10.1016/j.atmosenv.2010.06.023>
- Wang, Y., Cao, R., Xu, Z., Jin, J., Wang, J., Yang, T., Wei, J., Huang, J., Li, G., 2022. Long-term exposure to ozone and diabetes incidence: A longitudinal cohort study in China. *Science of The Total Environment* 816, 151634. <https://doi.org/10.1016/j.scitotenv.2021.151634>
- Wang, Yuanyuan, Du, H., Xu, Y., Lu, D., Wang, X., Guo, Z., 2018. Temporal and spatial variation relationship and influence factors on surface urban heat island and ozone pollution in the Yangtze River Delta, China. *Science of The Total Environment* 631–632, 921–933. <https://doi.org/10.1016/j.scitotenv.2018.03.050>
- Wang, Yapeng, Yu, C., Tao, J., Wang, Z., Si, Y., Cheng, L., Wang, H., Zhu, S., Chen, L., 2018. Spatio-Temporal Characteristics of Tropospheric Ozone and Its Precursors in Guangxi, South China. *Atmosphere* 9, 355. <https://doi.org/10.3390/atmos9090355>
- Yan, Y., Lin, J., He, C., 2018. Ozone trends over the United States at different times of day. *Atmospheric Chemistry and Physics* 18, 1185–1202. <https://doi.org/10.5194/acp-18-1185-2018>
- Yan, Y., Lin, J., Pozzer, A., Kong, S., Lelieveld, J., 2019. Trend reversal from high-to-low and from rural-to-urban ozone concentrations over Europe. *Atmospheric Environment* 213, 25–36. <https://doi.org/10.1016/j.atmosenv.2019.05.067>
- Yan, Y., Pozzer, A., Ojha, N., Lin, J., Lelieveld, J., 2017. Analysis of European ozone trends in the period 1995–2014 (preprint). *Gases/Atmospheric Modelling/Troposphere/Chemistry (chemical composition and reactions)*. <https://doi.org/10.5194/acp-2017-1077>
- Yang, P., Zhang, Y., Wang, K., Doraiswamy, P., Cho, S.-H., 2019. Health impacts and cost-benefit analyses of surface O<sub>3</sub> and PM<sub>2.5</sub> over the U.S. under future climate and emission scenarios. *Environmental Research* 178, 108687. <https://doi.org/10.1016/j.envres.2019.108687>
- Zhang, J., Wei, Y., Fang, Z., 2019. Ozone Pollution: A Major Health Hazard Worldwide. *Front. Immunol.* 10, 2518. <https://doi.org/10.3389/fimmu.2019.02518>

Zhang, X., Zhao, L., Cheng, M., Wu, X., Chen, D., 2020. Urban ozone sink inferred from surface measurements in China. *Journal of Cleaner Production* 253, 119881. <https://doi.org/10.1016/j.jclepro.2019.119881>

Zhou, K., Wang, W., Huang, L., Liu, B., 2021. Comparative study on the time series forecasting of web traffic based on statistical model and Generative Adversarial model. *Knowledge-Based Systems* 213, 106467. <https://doi.org/10.1016/j.knosys.2020.106467>

## Chapitre 5 : Caractérisation de la contamination de la zone d'étude au regard des recommandations de l'Organisation Mondiale de la Santé et des valeurs cibles réglementaires pour la protection de la santé

### Résumé :

#### Contexte

L'évaluation du risque sanitaire associé à l'exposition de la population à l'O<sub>3</sub> est classiquement basée sur les valeurs du maximum journalier de la moyenne glissante sur 8 heures des concentrations d'O<sub>3</sub> (*Maximum daily 8-hour mean* - MDA8). Les seuils d'exposition pour la protection de la santé humaine sont définis à partir des valeurs de MDA8. Dans le cadre des travaux menés dans cette thèse, les valeurs maximums ou moyennes journalières ont été délivrées par la plateforme de modélisation PREV'EST. La caractérisation des niveaux d'exposition de notre zone d'étude au regard des seuils en vigueur n'est donc pas applicable avec l'utilisation des données disponibles. L'objectif du présent chapitre est de : i) développer une méthode d'estimation du MDA8 et ii) caractériser les niveaux d'exposition de la zone d'étude au regard des seuils (réglementaire et définis par l'OMS).

#### Méthode

A partir des données mesurées à l'échelle horaire par les stations localisées en Bourgogne – Franche-Comté entre 2007 et 2022, ont été calculés : le MDA8, les valeurs journalières maximale et moyenne sur 24 heures. Un modèle de régression linéaire multiple ayant pour prédicteur la concentration maximale journalière et l'écart journalier entre la concentration maximale et moyenne journalière (nombre d'unités statistiques = 2 722 822) a été implémenté ( $R^2 = 0,97$ ). Ce modèle a ensuite permis d'estimer les valeurs de MDA8 sur chacune des mailles de modélisation de la zone d'étude. Ces valeurs ont permis de quantifier le nombre de dépassements du seuil réglementaire européen (120  $\mu\text{g}/\text{m}^3$ ) et du seuil recommandé par l'OMS (100  $\mu\text{g}/\text{m}^3$ , 60  $\mu\text{g}/\text{m}^3$  pour le pic saisonnier) par année et par maille. Les recensements de l'INSEE à l'échelle de l'IRIS ont permis par requête spatiale de distribuer les individus recensés dans les bâtiments issus de la BDTPO de l'IGN. La population, localisée dans chaque maille de modélisation et soumise aux niveaux d'exposition calculés précédemment, a été quantifiée annuellement.

#### Résultats

L'ensemble de la population est donc soumis à des dépassements de seuils non conformes aux recommandations de l'OMS. Le seuil réglementaire, moins protecteur, est dépassé plus de 25 fois par an pour 20% de la population en moyenne entre 2008 et 2019. Le nombre de dépassements du seuil réglementaire est toutefois très variables d'une année à l'autre, probablement en lien avec la variabilité des conditions climatiques. Quel que soit le seuil considéré, le nombre de dépassement par an et par maille tend à augmenter de 2008 et 2019. La variabilité du nombre de dépassement de seuil entre les mailles de modélisation pour une même année est élevée. A titre illustratif, la valeur estimée pour la maille de modélisation sur laquelle est située la station de suivi de la qualité de l'air de Besançon est reporté dans les figures. La variabilité des indicateurs calculée illustre les contrastes d'exposition qui ne peuvent être pris en compte avec l'utilisation de données mesurées à l'échelle d'une station de mesure. Le seuil OMS est systématiquement dépassé pour toutes les mailles.

#### Conclusions

L'exposition de la population au-delà des seuils en vigueur présente une hétérogénéité conséquente, qui n'aurait pas pu être prise en compte avec une station de suivi de la qualité de l'air. L'augmentation du nombre de dépassements de seuils induit une exposition de la population au-delà du seuil réglementaire grandissante. Cette augmentation illustre l'enjeu de compréhension de l'évolution des niveaux d'exposition au cours des vagues de chaleur, dans un contexte où ces phénomènes se sont d'ores et déjà amplifiés ces dernières décennies.



## **1) Introduction**

Les chapitres 3 et 4 ont permis d'analyser la dynamique spatiale et temporelle de l'O<sub>3</sub>. Ces concentrations modélisées de manière journalière et projetées sur le territoire d'étude correspondent aux valeurs journalières moyennes et maximales. Or, le seuil réglementaire européen ou les recommandations de l'OMS sont définis en fonction de la valeur du maximum journalier calculée sur une moyenne glissante. Une adaptation des données modélisées apparaissait donc comme un pré requis pour conduire une démarche d'évaluation de l'exposition humaine à l'ozone dans la perspective d'une évaluation quantitative du risque sanitaire.

L'objectif du présent chapitre est de : i) développer une méthode d'estimation du maximum journalier de la moyenne sur huit heures de la concentration atmosphérique en ozone (MDA8) et ii) caractériser les niveaux d'exposition de la zone d'étude au regard des seuils (réglementaire et définis par l'OMS).

## **2) Matériels et Méthodes**

### **2.1 Zone d'étude : l'aire urbaine de Besançon**

Le territoire concerné par cette phase descriptive est centré sur l'aire urbaine de Besançon et coïncide avec le territoire étudié dans le chapitre 4. La zone d'étude choisie (Figure 36) permet de délimiter l'aire urbaine de Besançon et de s'affranchir de l'influence des autres pôles urbains majeurs du secteur géographique.

Le maillage visible sur la Figure 36 correspond au maillage appliqué à la période 2015-2019 (3 km x 3 km) et comprend 142 mailles de modélisation. La surface du territoire étudié est de 1278 km<sup>2</sup>. Chaque maille est associée quotidiennement aux valeurs journalières moyenne et maximale, corrigées à l'aide de la méthodologie décrite en Chapitre 3.

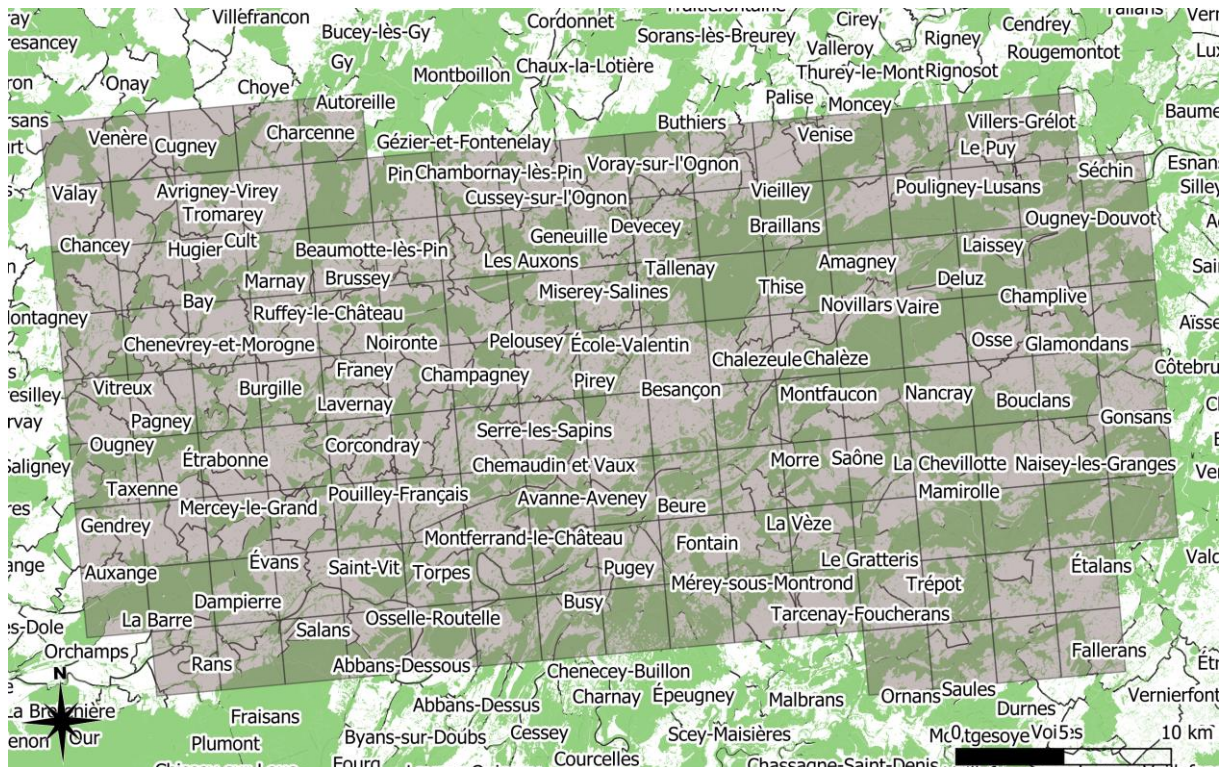


Figure 36: Couverture spatiale de la zone d'étude et maillage de modélisation de la plateforme PREVEST (3km x 3km).

## 2.2 Indicateurs réglementaires sanitaires classiquement utilisés

Les trois indicateurs ci-dessous reposent sur le calcul du maximum journalier des moyennes glissantes sur 8 h des concentrations d'O<sub>3</sub> (Maximum daily 8-hour mean - MDA8). Ils diffèrent sur la fréquence de dépassement ou la longueur de la période considérée.

Le seuil réglementaire pour la protection de la santé en Europe est fixé à 120 µg/m<sup>3</sup>. Ce seuil correspond à une valeur cible à ne pas dépasser plus de 25 fois par an en moyenne sur 3 ans pour éviter, prévenir ou réduire les effets nocifs sur la santé humaine (European Environment Agency, 2020).

Le seuil de référence défini par l'OMS se situe à 100 µg/m<sup>3</sup>. Ce seuil ne doit pas être dépassé plus de 3 fois par an selon les recommandations de l'OMS (World Health Organization, 2021).

Un nouvel indicateur a été défini dans le rapport 2021 de l'OMS (World Health Organization, 2021) afin de caractériser le pic saisonnier d'exposition à l'ozone. Cet indicateur correspond à la valeur maximale de la moyenne du MDA8 calculée sur une fenêtre glissante de 6 mois consécutifs. Le seuil recommandé par l'OMS pour le pic saisonnier est de 60 µg/m<sup>3</sup>.

### 2.3 Estimation du maximum journalier de la moyenne sur huit heures de la concentration atmosphérique en ozone (MDA8)

Les données de modélisation fournies par la plateforme PREV'EST étaient les valeurs maximales et moyennes sur 24 heures (chapitre 3 et 4). Aucune donnée horaire n'était disponible aux échelles spatiales retenues pour ce travail. Une estimation du MDA8 était nécessaire pour conduire une approche comparative de nos données avec les valeurs seuils réglementaires et les recommandations de l'OMS.

Une analyse statistique de la relation entre le MDA8 et les valeurs journalières disponibles (valeurs moyenne et maximale) a été conduite incluant une phase descriptive, une analyse de corrélation linéaire simple et l'utilisation de modèles de régression linéaire multiples. Cette approche a été menée sur les données de mesures horaires de toutes les stations de suivi de la qualité de l'air de Bourgogne – Franche-Comté ( $n = 29$ ) de 2007 à 2022 (soit 138 767 données horaires possibles pour chaque station). Le nombre de données utilisées toutes stations confondues, une fois les données manquantes prises en compte, est de 2 722 822.

A partir de ces données, par station et par jour, ont été calculés : le MDA8, les valeurs journalières maximales et moyennes sur 24 heures. Un modèle de régression linéaire multiple ayant pour prédicteur la concentration maximale journalière et l'écart journalier entre la concentration maximale et moyenne journalière a été implémenté.

Les résultats de l'analyse (coefficients et paramètres d'adéquation du modèle) sont présentés dans le Tableau 7. La distribution des résidus de ce modèle en fonction des valeurs ajustées est présentée en Figure 37. Ce modèle a ensuite été utilisé pour prédire, pour chaque maille, le MDA8 à partir des valeurs modélisées de concentration maximale et moyenne journalières. Le traitement des bases de données et l'analyse statistique ont été menés avec le logiciel R (version 4.2.2).

Tableau 7 : Résultat de la régression linéaire multiple ayant pour variable expliquée le MDA8 et pour variables explicative la concentration maximale journalière d'O<sub>3</sub> et l'écart entre la concentration maximum et moyenne journalière.

|   | Coefficient | Erreur standard         | Degré de significativité | R <sup>2</sup> ajusté |
|---|-------------|-------------------------|--------------------------|-----------------------|
| Intercept   | - 3,74      | 0,04                    | < 2 x 10 <sup>-16</sup>  | 0.97                  |
| Concentration maximale journalière                                      | 1,06        | 6,00 x 10 <sup>-4</sup> | < 2 x 10 <sup>-16</sup>  |                       |
| Ecart journalier entre la concentration maximale et moyenne journalière | - 0,31      | 1,19 x 10 <sup>-3</sup> | < 2 x 10 <sup>-16</sup>  |                       |

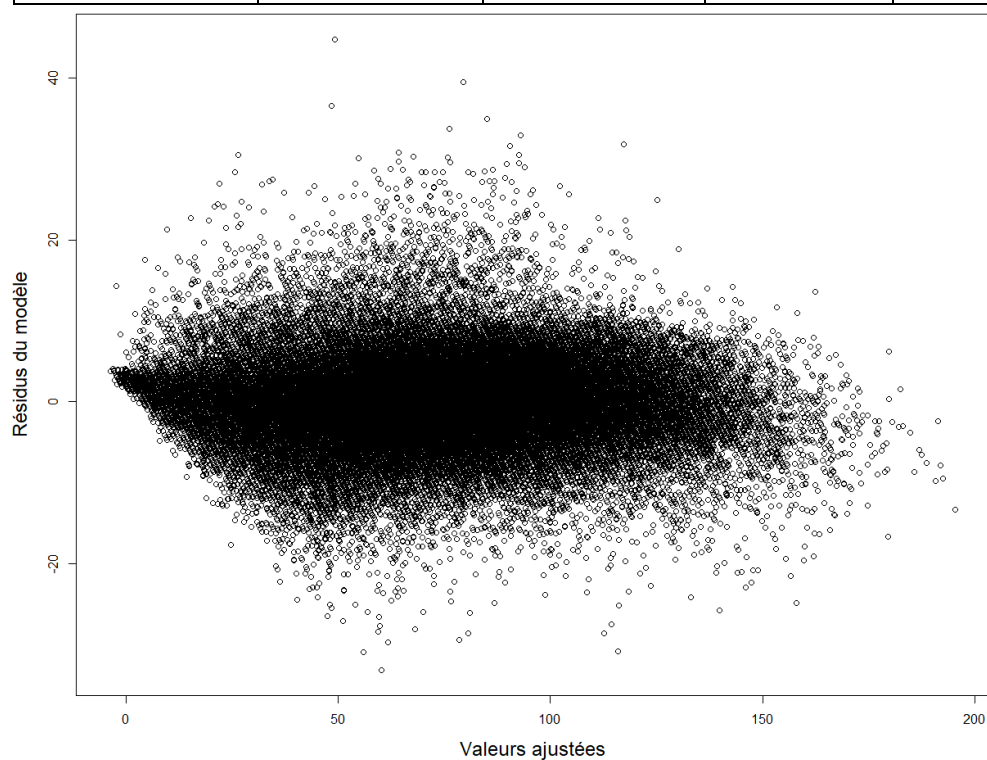


Figure 37: Distribution des résidus du modèle linéaire multiple ( $MDA8 \sim [O3]_{maximum} + ([O3]_{maximum} - [O3]_{moyenne})$ ) en fonction des valeurs ajustées.

Une fois les données de MDA8 estimées, les indices suivants ont été calculés pour chaque maille :

- nombre de dépassements du seuil de référence OMS (100 µg/m<sup>3</sup>) ;
- nombre de dépassements du seuil réglementaire européen (120 µg/m<sup>3</sup>) ;
- valeur du pic saisonnier, soit la valeur maximale du MDA8 calculée sur 6 mois consécutifs.

La moyenne glissante sur trois ans du nombre de dépassements du seuil  $120 \mu\text{g}/\text{m}^3$  par maille n'a pas pu être calculée car les trois maillages successivement utilisés lors de la modélisation PREV'EST ne coïncidaient pas. A titre illustratif, la valeur de chaque indicateur pour la maille sur laquelle est située la station de suivi de la qualité de l'air de Besançon (de typologie urbaine) a été reporté dans le tableau décrivant les indicateurs par année et par maille de modélisation.

### 3) Résultats

Un résumé des nombres de dépassement de seuil au cours de la période d'étude 2008-2019 est présenté dans le Tableau 8. Un nombre élevé de dépassements à la fois du seuil recommandé par l'OMS ( $100 \mu\text{g}/\text{m}^3$ ) et du seuil réglementaire européen ( $120 \mu\text{g}/\text{m}^3$ ) est constaté (Tableau 8).

*Tableau 8: Descriptif du nombre de dépassements du seuil annuel moyen par maille de modélisation pour la totalité de la période 2008-2019, recommandation OMS ( $100 \mu\text{g}/\text{m}^3$ ) et seuil réglementaire européen ( $120 \mu\text{g}/\text{m}^3$ )*

| <b>Nombre de dépassements annuel moyen du seuil OMS (<math>100\mu\text{g}/\text{m}^3</math>) par maille (jours)</b>                    |              |
|--|--------------|
| Médiane (Q1 – Q3)  | 59 (48 - 78) |
| Moyenne (ET)   | 64 (23)      |
| Étendue  | 12 – 152     |
| <b>Nombre de dépassements annuel moyen du seuil réglementaire européen (<math>120\mu\text{g}/\text{m}^3</math>) par maille (jours)</b> |              |
| Médiane (Q1 – Q3)  | 18 (10 - 26) |
| Moyenne (ET)   | 20 (13)      |
| Étendue  | 0 – 74       |

Les résultats de l'analyse stratifiée par année des trois indicateurs sont présentés dans le Tableau 9 et dans les Figure 38 à Figure 40. Les concentrations d'ozone sont amplement supérieures aux recommandations de l'OMS sur les 12 années, le même constat est fait pour le pic saisonnier. Le seuil réglementaire européen est plus fréquemment dépassé en fin de période. Bien qu'une certaine variabilité interannuelle soit présente, une tendance à l'augmentation est observée pour ces trois indicateurs de 2008 à 2019. La variabilité intra-annuelle du nombre de dépassements de seuils et du pic saisonnier apparaît élevée, illustrant la variabilité spatiale des niveaux d'exposition sur le territoire de l'étude.

Tableau 9: Descriptif du nombre de dépassements du seuil OMS (100 µg/m<sup>3</sup>) et du seuil réglementaire (120µg/m<sup>3</sup>) par maille de modélisation et par année, ainsi que la valeur du pic saisonnier par maille de modélisation et par année.

| Année                   | Nombre de dépassements (jours) du seuil OMS (100µg/m <sup>3</sup> ) |              |          |                                | Nombre de dépassements (jours) du seuil réglementaire (120µg/m <sup>3</sup> ) |              |         |                                | Pic saisonnier (µg/m <sup>3</sup> ) |              |           |                                |
|-------------------------|---|--------------|----------|--------------------------------|---|--------------|---------|--------------------------------|-------------------------------------|--------------|-----------|--------------------------------|
|                         | Médiane (Q1 – Q3)   | Moyenne (ET) | Étendue  | Maille de la station de mesure | Médiane (Q1 – Q3)   | Moyenne (ET) | Étendue | Maille de la station de mesure | Médiane (Q1 – Q3)                   | Moyenne (ET) | Étendue   | Maille de la station de mesure |
| <b>2008,</b><br>n = 73  | 33 (31 - 35)  | 33 (3)       | 26 - 40  | 29                             | 7 (4 - 8)   | 6 (3)        | 0 - 12  | 5                              | 86 (86 - 87)                        | 86 (1)       | 80 - 88   | 83                             |
| <b>2009,</b><br>n = 73  | 53 (48 - 59)  | 54 (11)      | 29 - 78  | 37                             | 10 (9 - 11)   | 10 (2)       | 5 - 15  | 6                              | 91 (89 - 93)                        | 91 (4)       | 83 - 97   | 86                             |
| <b>2010,</b><br>n = 73  | 53 (47 - 55)  | 50 (6)       | 32 - 59  | 49                             | 24 (17 - 25)  | 21 (6)       | 6 - 28  | 20                             | 90 (88 - 92)                        | 90 (2)       | 84 - 92   | 87                             |
| <b>2011,</b><br>n = 140 | 68 (61 - 76)  | 67 (11)      | 37 - 89  | 37                             | 13 (8 - 19)   | 13 (6)       | 2 - 27  | 4                              | 93 (91 - 95)                        | 93 (3)       | 84 - 99   | 86                             |
| <b>2012,</b><br>n = 140 | 55 (35 - 67)  | 51 (19)      | 12 - 84  | 50                             | 11 (9 - 14)   | 11 (4)       | 2 - 21  | 11                             | 92 (88 - 94)                        | 91 (4)       | 80 - 98   | 90                             |
| <b>2013,</b><br>n = 140 | 49 (46 - 51)  | 49 (5)       | 38 - 60  | 43                             | 22 (20 - 24)  | 22 (4)       | 12 - 32 | 20                             | 88 (87 - 89)                        | 88 (2)       | 83 - 94   | 84                             |
| <b>2014,</b><br>n = 140 | 52 (45 - 58)  | 52 (13)      | 24 - 94  | 41                             | 8 (6 - 9)   | 8 (2)        | 2 - 13  | 7                              | 91 (90 - 93)                        | 91 (3)       | 85 - 98   | 89                             |
| <b>2015,</b><br>n = 142 | 78 (74 - 86)  | 79 (10)      | 53 - 99  | 74                             | 34 (27 - 44)  | 36 (10)      | 19 - 56 | 40                             | 102 (100 - 104)                     | 102 (3)      | 92 - 109  | 99                             |
| <b>2016,</b><br>n = 142 | 47 (41 - 52)  | 47 (8)       | 26 - 65  | 36                             | 16 (12 - 20)  | 15 (6)       | 1 - 27  | 7                              | 91 (88 - 92)                        | 90 (3)       | 84 - 97   | 84                             |
| <b>2017,</b><br>n = 142 | 68 (59 - 78)  | 69 (16)      | 33 - 118 | 46                             | 16 (13 - 22)  | 18 (7)       | 8 - 37  | 11                             | 97 (95 - 99)                        | 97 (4)       | 87 - 108  | 92                             |
| <b>2018,</b><br>n = 142 | 110 (97 - 122)  | 110 (16)     | 76 - 152 | 76                             | 48 (42 - 54)  | 48 (10)      | 26 - 74 | 28                             | 108 (106 - 111)                     | 108 (4)      | 100 - 118 | 100                            |
| <b>2019,</b><br>n = 142 | 78 (69 - 86)  | 77 (13)      | 44 - 106 | 69                             | 26 (21 - 30)  | 25 (6)       | 10 - 37 | 23                             | 102 (99 - 104)                      | 101 (4)      | 89 - 108  | 99                             |

Les effectifs indiqués pour chaque année correspondent au nombre de mailles de modélisation couvrant la zone.

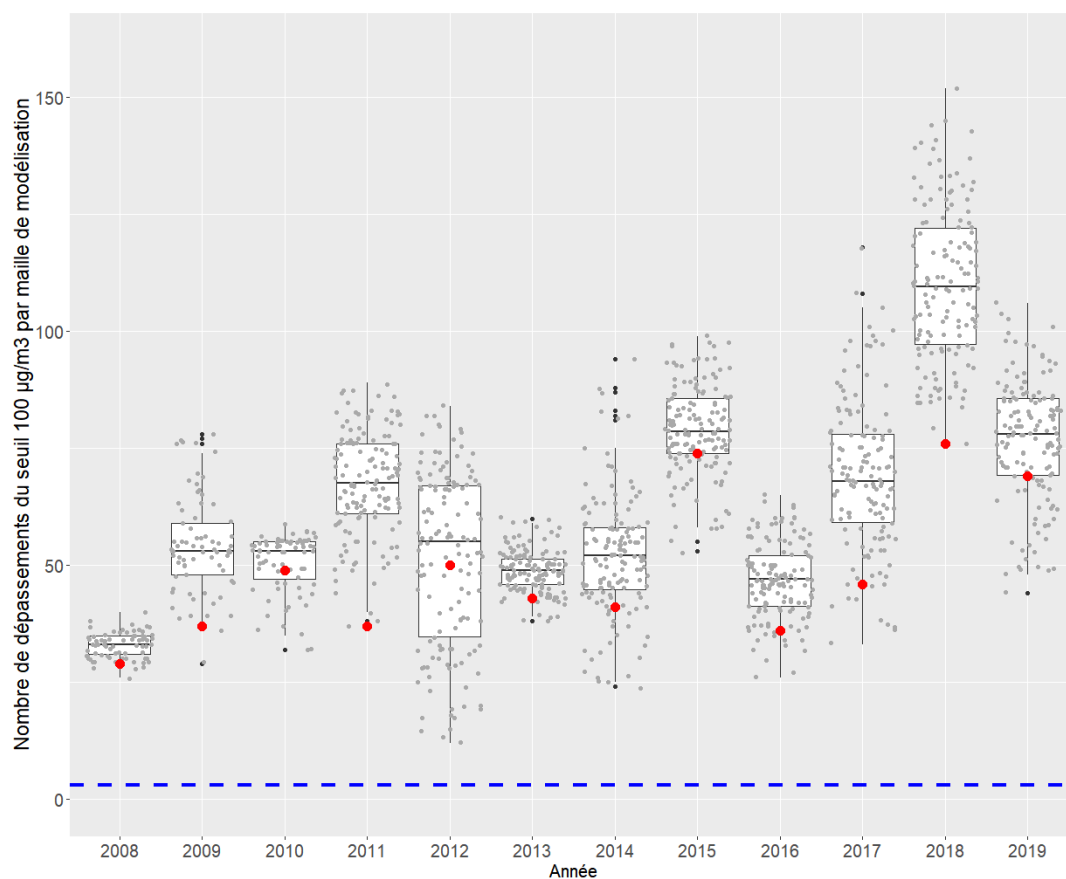


Figure 38: Nombre de dépassements du seuil OMS (jours) par maille de modélisation en fonction de l'année. Le point rouge correspond à la valeur de la maille sur laquelle se situe la station de suivi de la qualité de l'air. La ligne pointillée bleue représente le nombre de dépassements qui ne doit pas être excédé.



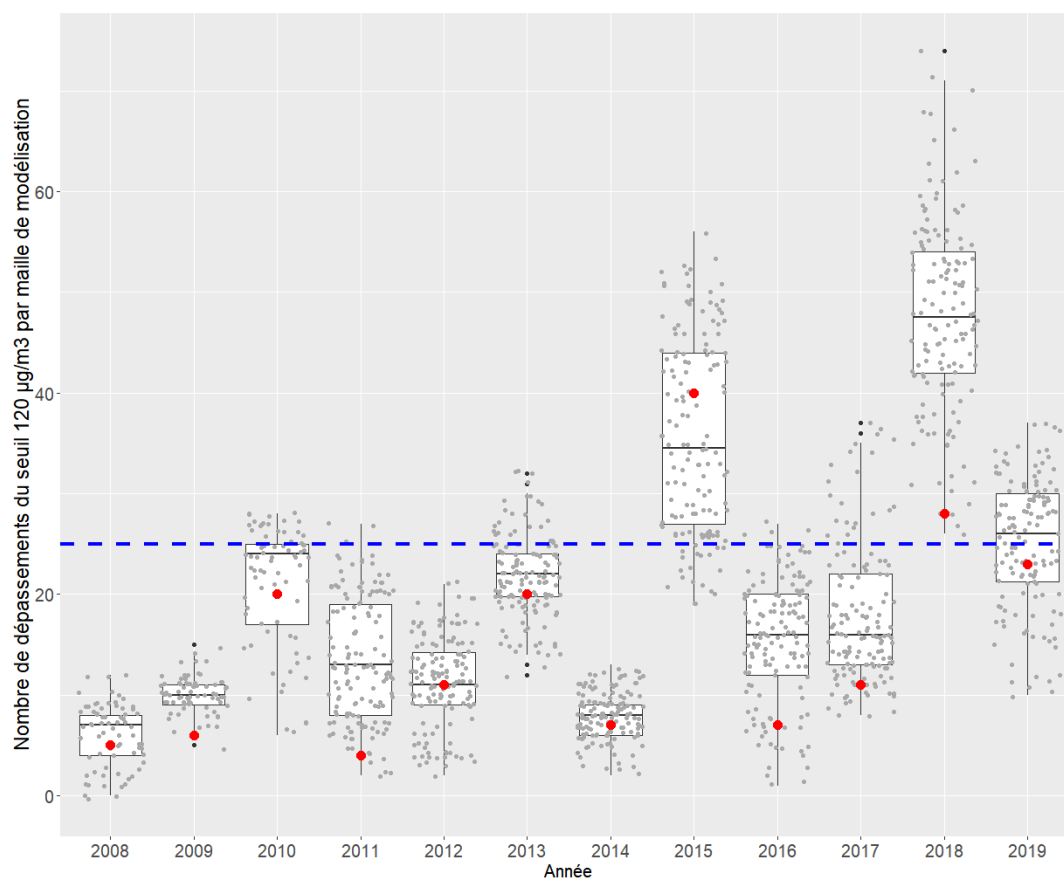


Figure 39: Nombre de dépassements (jours) du seuil réglementaire européen par maille de modélisation en fonction de l'année. Le point rouge correspond à la valeur de la maille sur laquelle se situe la station de suivi de la qualité de l'air. La ligne pointillée bleue représente le nombre de dépassements qui ne doit pas être excédé.

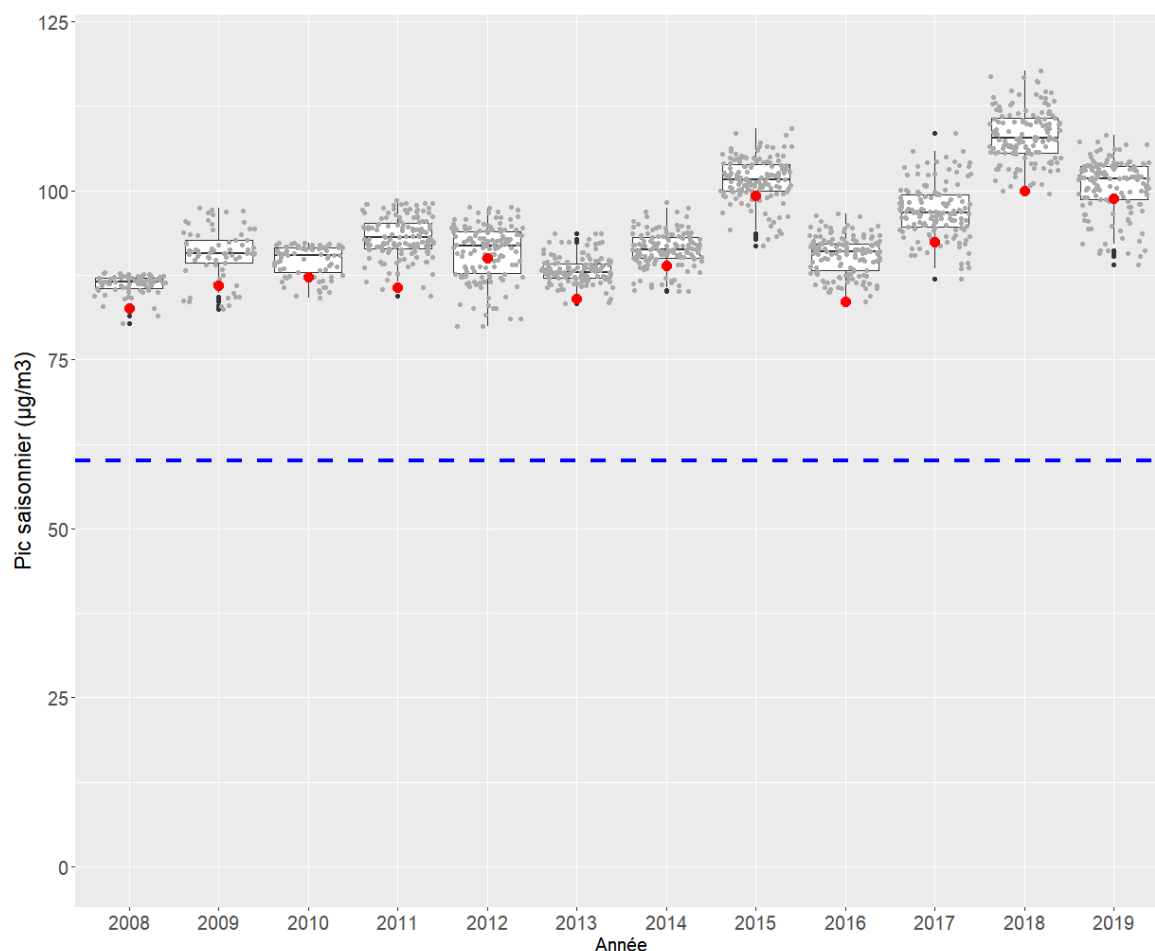


Figure 40: Valeur du pic saisonnier ( $\mu\text{g}\cdot\text{m}^{-3}$ ), en fonction de l'année. Le point rouge correspond à la valeur de la maille sur laquelle se situe la station de suivi de la qualité de l'air. La ligne pointillée bleue représente la valeur seuil à ne pas excéder.

Dans le Tableau 9 et dans les Figure 38 à Figure 40, sont également présentées les valeurs estimées de la maille sur laquelle est localisée la station de suivi de la qualité de l'air de Besançon (maille de la station de mesure). Pour les indicateurs de l'OMS, la valeur de la maille de la station de mesure est systématiquement inférieure à la médiane, et inférieure ou égale à la valeur du premier quartile pour 85 % des années. Concernant le seuil réglementaire européen, cette valeur est supérieure ou égale à la médiane pour deux années, et inférieure ou égale au premier quartile pour 50 % des années étudiées.

La fluctuation des niveaux d’ozone interannuelle, ainsi que sa tendance à l’augmentation, se reflètent également dans la proportion de population soumise à des dépassements de seuil réglementaire européen (Tableau 10).

Tableau 10: Effectif de population générale, et effectif et pourcentage de population concernée par un dépassement du seuil OMS et réglementaire européen par année.

| Année | Population totale | Population soumise au dépassement de seuil |             |                                     |             |
|-------|-------------------|--|-------------|-------------------------------------|-------------|
|       |                   | OMS <sup>a</sup><br>et pic saisonnier      |             | Réglementaire européen <sup>b</sup> |             |
|       |                   | Effectif                                   | Pourcentage | Effectif                            | Pourcentage |
| 2008  | 224 497           | 224 497                                    | 100         | 0                                   | 0           |
| 2009  | 225 292           | 225 292                                    | 100         | 0                                   | 0           |
| 2010  | 225 681           | 225 681                                    | 100         | 24 440                              | 10,8        |
| 2011  | 225 375           | 225 375                                    | 100         | 839                                 | 0,4         |
| 2012  | 226 752           | 226 752                                    | 100         | 0                                   | 0           |
| 2013  | 228 557           | 228 557                                    | 100         | 15 838                              | 6,9         |
| 2014  | 229 149           | 229 149                                    | 100         | 0                                   | 0           |
| 2015  | 229 272           | 229 272                                    | 100         | 217 285                             | 94,8        |
| 2016  | 230 185           | 230 185                                    | 100         | 748                                 | 0,3         |
| 2017  | 231 271           | 231 271                                    | 100         | 12 564                              | 5,4         |
| 2018  | 232 459           | 232 459                                    | 100         | 232 459                             | 100,0       |
| 2019  | 232 459           | 232 459                                    | 100         | 62 881                              | 27,1        |

a : plus de 3 fois par an

b : plus de 25 fois par an

#### 4) Discussion

Ce chapitre, incluant une approche méthodologique pour l'estimation du MDA8, a abouti à une quantification des niveaux d'exposition en population sur une échelle fine et une comparaison avec les indicateurs réglementaires.

Les indicateurs d'exposition développés par les instances gouvernementales, par l'OMS et utilisés dans le cadre de l'évaluation du risque sanitaire reposent sur les valeurs seuil de MDA8 (EEA, 2015; Pascal et al., 2019; World Health Organization, 2021). Une caractérisation des niveaux d'exposition de la population nécessitait donc de recueillir ou recalculer ce type de données. La méthode d'estimation du MDA8 développée, basée sur une régression linéaire multiple, porte sur un nombre élevé d'unités statistiques et permet d'expliquer la quasi-totalité de la variabilité du MDA8. Cette approche offre une estimation robuste de cet indicateur.

La démarche conduite dans ce chapitre a permis de calculer le MDA8 pour chaque maille de modélisation (effectifs de mailles variant de 73 à 142) et ainsi d'identifier l'hétérogénéité spatiale des niveaux d'exposition à l'ozone et du nombre de dépassements de seuil qui en découle. La visualisation de la valeur de la maille sur laquelle se situe la station de mesure permet d'illustrer l'écart entre le niveau évalué ponctuellement à une station de mesure et les niveaux sur le reste d'un territoire considéré. Or, les dépassements de seuils sont généralement calculés à l'échelle d'une station (ou d'un nombre restreint de station) de mesure de la qualité de l'air (EEA, 2015). Les résultats de ce calcul réalisé sur un nombre restreint de stations sont utilisés pour quantifier les niveaux d'exposition de la totalité de la population habitant un pôle urbain. Cette approche repose donc sur une hypothèse forte d'homogénéité spatiale des niveaux d'exposition (Pascal et al., 2019). Nos résultats illustrent l'écart entre la réalité de l'exposition et cette hypothèse d'homogénéité. Projeter sur l'ensemble du territoire la valeur de la station de mesure risque d'engendrer une sous-estimation pour de nombreuses parties du territoire et pour la population qu'elles abritent. Si la sous-estimation est systématiquement associée à certaines zones, par exemple loin des centres urbains, cela impliquerait un biais dans l'évaluation de l'exposition de la population et dans l'évaluation des impacts sanitaires associés.

La prise en compte de la variabilité spatiale des niveaux d'exposition, et notamment selon le gradient d'urbanisation, devrait permettre une meilleure affectation à la population des niveaux d'exposition. Cela permettrait ainsi une meilleure analyse de la relation exposition - effet sanitaire en approche analytique et une meilleure estimation du risque en évaluation quantitative de l'impact sanitaire.

Les seuils définis par l'OMS, plus protecteurs que le seuil réglementaire, sont systématiquement dépassés sur notre territoire d'étude. Les indicateurs calculés à partir du MDA8 prédit ont permis de mettre en évidence une non-conformité des niveaux d'exposition vis-à-vis du seuil réglementaire européen pour la protection de la santé humaine et des seuils définis par l'OMS. L'augmentation du nombre de dépassements de seuil et l'augmentation du pic saisonnier, fait écho à une augmentation de fréquence, d'intensité et de durée des vagues de chaleur qui sont des périodes propices au dépassement de seuil (conditions anticycloniques prolongées, fortes températures, fort ensoleillement) (Lin et al., 2001; Struzewska and Jefimow, 2016; Varotsos et al., 2019). La durée des vagues de chaleur est notamment corrélée à l'exposition de la population au-delà du seuil de 120  $\mu\text{g}/\text{m}^3$  (Struzewska and Jefimow, 2016).

## **5) Conclusion :**

Le développement d'une méthode d'estimation du MDA8 a permis de mener une approche comparative entre les niveaux d'exposition de la population et les seuils en vigueur (réglementaire et définis par l'OMS) de manière continue dans l'espace. Les contrastes d'exposition présents dans la zone d'étude et illustrés en chapitre 4, sont également identifiables dans ce chapitre à travers la variabilité du nombre de dépassements de seuils entre les mailles de modélisation.

La tendance à la hausse observée dans le nombre de dépassements de seuils fait écho à l'augmentation des phénomènes de vagues de chaleur. Une analyse ciblée sur ce type de phénomènes serait particulièrement pertinente pour caractériser l'évolution des niveaux d'exposition, en termes de valeur et de distribution spatiale entre les milieux urbains, périurbains et ruraux.

### **Références bibliographiques :**

- EEA, 2015. Air quality in Europe: 2015 report. European Environmental Agency. (No. 5/2015). Publications Office, LU. <https://doi.org/10.2800/62459>
- European Environment Agency., 2020. Air quality in Europe: 2020 report. (No. 09/2020). Publications Office, LU.
- Lin, C.-Y.C., Jacob, D.J., Fiore, A.M., 2001. Trends in exceedances of the ozone air quality standard in the continental United States, 1980–1998. *Atmospheric Environment* 35, 3217–3228. [https://doi.org/10.1016/S1352-2310\(01\)00152-2](https://doi.org/10.1016/S1352-2310(01)00152-2)
- Pascal, M., Yvon, J.-M., Medina, S., 2019. Pollution atmosphérique. Guide pour la réalisation d'une évaluation quantitative des impacts sur la santé (EQIS). EQIS avec une exposition modélisée. Santé publique France.
- Struzewska, J., Jefimow, M., 2016. A 15-Year Analysis of Surface Ozone Pollution in the Context of Hot Spells Episodes over Poland. *Acta Geophys.* 64, 1875–1902. <https://doi.org/10.1515/acgeo-2016-0067>
- Varotsos, K.V., Giannakopoulos, C., Tombrou, M., 2019. Ozone-temperature relationship during the 2003 and 2014 heatwaves in Europe. *Reg Environ Change* 19, 1653–1665. <https://doi.org/10.1007/s10113-019-01498-4>
- World Health Organization, 2021. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.

## **Chapitre 6 : Approche descriptive et exploratoire des vagues de chaleur et de leurs relations avec le niveau d'exposition à l'ozone de la population et du degré d'urbanisation**

### **Résumé :**

#### ***Contexte***

Les vagues de chaleur sont des périodes d'exposition de la population à des températures extrêmes et sont associées à un impact sanitaire reconnu. Il s'agit également d'épisodes favorables à l'apparition de pics d'O<sub>3</sub>. Ces épisodes de double exposition représentent un enjeu sanitaire majeur et qui s'intensifie dans un contexte de changement climatique. Les éléments de compréhension sur l'influence des vagues de chaleur et leurs caractéristiques (durée, intensité...) sur la dynamique spatiale et temporelle de l'O<sub>3</sub> sont encore peu nombreux. Ce chapitre, proposé sous la forme d'une approche exploratoire, a pour objectifs de i) quantifier et qualifier les vagues de chaleur survenues sur la zone d'étude pour une période de 12 ans, ii) identifier une éventuelle typologie de ces vagues de chaleur, iii) explorer la relation entre ces vagues de chaleur et le niveau d'exposition à l'O<sub>3</sub> de la population de la zone d'étude, et, iv) explorer l'influence du degré d'urbanisation sur la relation précédemment décrite.

#### ***Méthode***

Les périodes de vagues de chaleur ont été identifiées et caractérisées (durée, intensité, sévérité) selon la définition de Météo France à l'échelle de la station météorologique de Besançon de 2008 à 2019. Les vagues de chaleur recensées sont soumises à une classification ascendante hiérarchique et trois typologies sont identifiées. Les niveaux d'exposition à l'O<sub>3</sub> (MDA8) sont projetés à l'échelle du bâtiment pour chaque journée de vague de chaleur. Le nombre de dépassements du seuil réglementaire européen (120 µg/m<sup>3</sup>) et du seuil défini par l'OMS (100 µg/m<sup>3</sup>) est calculé pour chaque bâtiment et pour chaque vague de chaleur. Chaque bâtiment est associé à un effectif de population et à une typologie de territoires définie par l'INSEE (de l'urbain dense au rural autonome). Les typologies de vagues de chaleur sont caractérisées sur la base de leur durée, intensité, sévérité, de la valeur moyenne de MDA8 au cours de la vague de chaleur et du nombre de dépassements de seuils au cours de ces vagues de chaleur.

#### ***Résultats***

Les trois typologies de vagues de chaleur présentent des durées, intensités et sévérité croissantes de la typologie A à C. Ces typologies présentent également des niveaux d'exposition (MDA8) et un nombre de dépassements de seuils croissants. L'augmentation de la durée, de l'intensité et de la sévérité des vagues de chaleur influencent les niveaux d'exposition et le nombre de dépassements de seuils. Les caractéristiques de vagues de chaleur expliquent partiellement les niveaux d'exposition de la population à l'O<sub>3</sub>. Les années les plus contrastées vis-à-vis des caractéristiques des vagues de chaleur présentent un différentiel dans les niveaux d'exposition de la population. L'utilisation sur l'ensemble du territoire étudié de la valeur d'exposition de la maille sur laquelle la station de mesure se trouve, pourrait conduire à une sous-estimation des niveaux d'exposition pour 86 % et 49 % de la population, en 2008 et 2019 respectivement.

#### ***Conclusion***

L'augmentation des caractéristiques des vagues de chaleur (durée, intensité, sévérité) est associée à une augmentation des niveaux d'exposition de la population. Les milieux périurbains semblent plus exposés à l'O<sub>3</sub> et aux dépassements de seuils réglementaires. Cette approche exploratoire ouvre des perspectives sur une analyse plus poussée et notamment la prise en compte de la variabilité spatiale de la température et donc de la définition de vague de chaleur à l'échelle d'un territoire.

## **1) Introduction :**

L'augmentation des concentrations d'O<sub>3</sub> est concomitante d'une augmentation de la fréquence et de l'intensité des phénomènes de vagues de chaleur (VDC) dans un contexte de changement climatique. Les VDC touchent particulièrement la population vivant en milieu urbain, où l'artificialisation intense et la nature des matériaux utilisés sont responsables de températures nocturnes élevées (Ulpiani, 2021). Les conditions météorologiques à l'origine des vagues de chaleur pourraient également modifier la distribution spatiale de ce photo-oxydant. Ainsi, les zones urbaines pourraient devenir un espace à forte densité de population humaine exposée à des concentrations d'O<sub>3</sub> élevées et à des températures extrêmes. Un effet synergique de l'exposition à l'O<sub>3</sub> et aux vagues de chaleur a été identifié pour la prématurité (Sun et al., 2020) et la mortalité respiratoire (Fu et al., 2022). L'augmentation de l'occurrence de cette co-exposition est préoccupante et représente un potentiel enjeu sanitaire.

L'approche exploratoire, présentée dans ce présent chapitre, a pour objectifs de :

- i) quantifier et qualifier les vagues de chaleur survenues sur la zone d'étude pour une période de 12 ans
- ii) identifier une éventuelle typologie de ces vagues de chaleur
- iii) explorer la relation entre ces vagues de chaleur et le niveau d'exposition à l'O<sub>3</sub> de la population de la zone d'étude, et,
- iv) explorer l'influence du degré d'urbanisation sur la relation précédemment décrite.

## **2) Matériels et méthodes :**

### **1. Typologie de territoires caractérisant le degré d'urbanisation**

Le caractère urbain et rural est défini à l'échelle des communes selon une classification établie par l'INSEE. La définition de cette typologie de territoires, selon leur caractère urbain et rural, a été établie en 2017. Cette typologie est ici utilisée pour l'ensemble de la période d'étude (2008-2019).

Les communes rurales :

- sous forte influence d'un pôle d'emploi appartiennent à une aire d'attraction des villes de plus de 50000 habitants, et plus de 30 % de leurs actifs occupés travaillent dans le pôle de cette aire.



- sous faible influence d'un pôle d'emploi appartiennent à une aire d'attraction des villes de plus de 50000 habitants et moins de 30 % de leurs actifs occupés travaillent dans le pôle de cette aire.

- autonomes ne sont pas soumises à l'influence d'un pôle d'emploi ou appartiennent à une aire de moins de 50 000 habitants. Ces communes qualifiées de rurales autonomes, fonctionnent sans l'influence d'un pôle, ou sous l'influence d'un petit pôle qui structure peu son environnement. Parmi ces communes on distingue les communes peu denses des communes très peu denses (D'Alessandro et al., 2021).

La distinction entre les communes à caractère urbain et celles à caractère rural est réalisée selon la grille communale de densité de population (INSEE, 2022).

## 2. Chaîne de traitement de données

### *2.1 Association des différentes sources de données par requête spatiale*

En France, les données du recensement de la population humaine sont disponibles à l'échelle de l'IRIS (îlot regroupé d'informations statistiques). Les données de recensement des années 2014 à 2019 ont été utilisées dans cette étude (Insee, 2021). Les bâtiments de type « bâti indifférencié » issus de la BDTOPO de l'IGN ont été utilisés pour identifier les bâtiments d'habitation de la zone d'étude. A l'échelle de l'IRIS, chaque bâtiment d'habitation s'est vu affecter un effectif de sujets en utilisant le rapport de proportionnalité entre la surface du bâtiment et la totalité des surfaces des bâtiments d'habitations de l'IRIS. Le géoréférencement des centroïdes des bâtiments, des contours des IRIS et des mailles de modélisation de la plateforme de modélisation PREV'EST ont permis de relier ces différents éléments par requête spatiale à l'aide d'un SIG (logiciel QGis). La même procédure a été répétée pour chacune des 12 années de la période d'étude et une base de données relationnelle, intégrant les données relatives à ces entités, a ainsi été constituée.

### *2.2 Vagues de chaleur (VDC) et concentrations d'O<sub>3</sub>*

#### *2.2.1 Définition des périodes de vague de chaleur*

La définition des vagues de chaleur de Météo France (Soubeyroux et al., 2016) a été utilisée. Celle-ci est basée sur des seuils de température moyenne journalière correspondant aux percentiles 95 (seuil

d'interruption, nommé Sint), 97,5 (seuil de début de vague, nommé Sdeb) et 99,5 (seuil de détection, nommé Spic) de l'indicateur thermique sur la période de référence (1981-2010).

Une vague de chaleur est identifiée si le seuil Spic est dépassé. Le début de vague de chaleur est défini par le dépassement du seuil Sdeb, et seules les vagues durant au moins trois jours sont prises en compte. La vague est interrompue si la température descend en dessous du seuil Sint (même une seule journée), ou en dessous du seuil Sdeb pendant au moins 3 jours consécutifs (Ouzeau et al., 2016; Soubeyroux et al., 2016).

Chaque vague identifiée est ensuite caractérisée par trois indicateurs : durée, intensité et sévérité (Figure 41). L'intensité correspond à la température moyenne journalière maximale observée au cours de la vague. La sévérité est « la somme des écarts positifs entre températures moyennes quotidiennes pendant la vague de chaleur et le seuil de démarrage Sdeb divisée par la différence entre Spic et Sdeb » (Ouzeau et al., 2016).

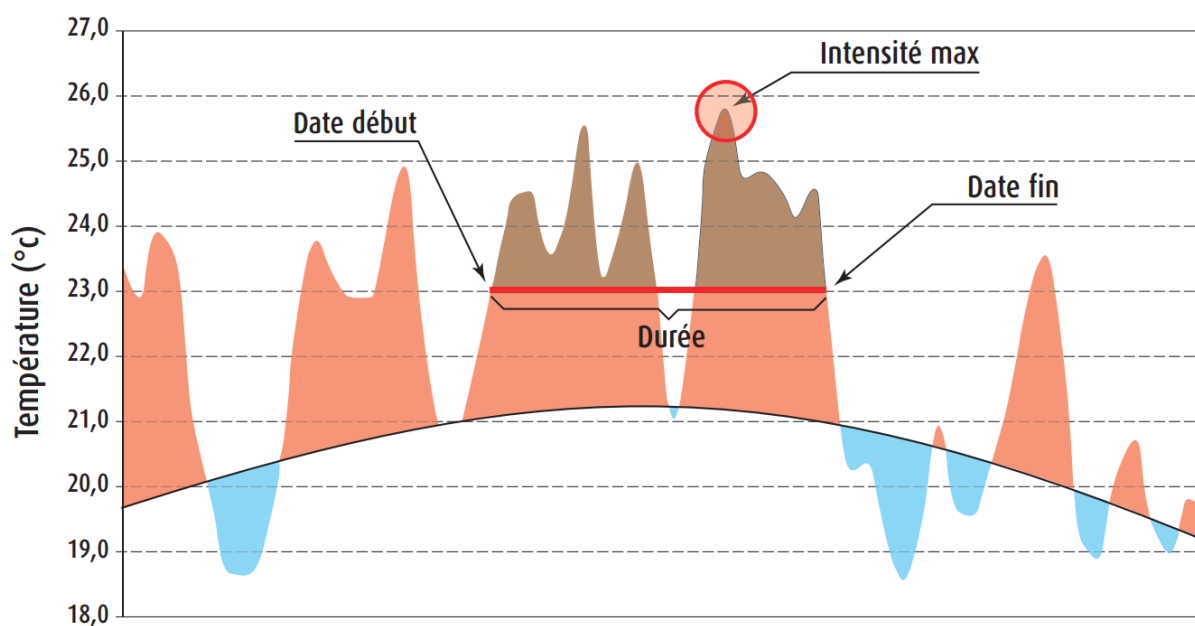


Figure 41 : Caractérisation d'une vague de chaleur à partir de l'indicateur thermique quotidien sur la France en durée (date de début et de fin), intensité maximum (cercle rouge) et sévérité (aire colorée en marron). D'après Soubeyroux et al. (2016).

Les seuils départementaux sont utilisés afin d'identifier des épisodes de manière plus fine et spécifique au vu des températures observées sur le territoire considéré. L'inventaire est réalisé à l'échelle de la station météorologique de Besançon.

### *2.2.2 Classification des vagues de chaleur*

La sévérité prenant indirectement en compte la durée et l'intensité des VDC, cet indicateur est celui le plus intégrateur des caractéristiques des VDC (Figure 41). Une classification hiérarchique ascendante basée sur la sévérité a été réalisée afin d'établir une typologie de vagues de chaleur et d'identifier d'éventuels profils. La construction du dendrogramme a été réalisée en utilisant « l'association moyenne (Unweighted Pair-Group Method using Arithmetic averages (UPGMA)) ». Le nombre de classes a été déterminé en visualisant le dendrogramme et en utilisant la méthode dite du « coude » (« *elbow method* »). Cette méthode repose sur la visualisation de la valeur du WSS (« Within-cluster Sum of Squares ») en fonction du nombre de classes. La rupture ou le "coude" visible dans cette relation décrit le nombre "optimal" de groupes, conduisant à une diminution rapide du WSS pour le plus petit nombre de groupes (Figure 44). Cela signifie que l'ajout de groupes supplémentaires n'améliore pas le processus de partition.

### *2.2.3 Mise en relation des périodes d'exposition avec les concentrations d'O<sub>3</sub>*

Les périodes de vague de chaleur décrites dans le paragraphe 2.2.1 sont ensuite associées à des niveaux d'exposition. Pour chaque vague de chaleur identifiée, le maximum journalier de la moyenne glissante sur 8 heures des concentrations d'O<sub>3</sub> (MDA8) a été affecté à chacun des bâtiments de la zone d'étude. Chaque bâtiment est aussi associé à une valeur moyenne de MDA8 par vague de chaleur.

## *3. Calcul du nombre de dépassements de seuils*

Le nombre de dépassements de seuil est calculé au cours de chaque vague de chaleur et pour chaque bâtiment. Ce calcul est réalisé pour les deux seuils utilisés dans le chapitre 5 : le seuil réglementaire européen (fixé à 120 µg/m<sup>3</sup>, à ne pas dépasser plus de 25 fois par an, en moyenne sur 3 ans (European Environment Agency, 2020)) et le seuil OMS (fixé à 100 µg/m<sup>3</sup>, à ne pas dépasser plus de 3 fois par an (World Health Organization, 2021)). Pour chaque journée de vague de chaleur et pour chaque bâtiment, les dépassements de seuils sont identifiés et sommés.

#### 4. Analyse statistique

Les éléments calculés sont présentés sous forme de moyenne, médiane, premier et troisième quartile. Les classes de VDC sont comparées à l'aide du test non paramétrique de Kruskal-Wallis suivi par un test de comparaison multiple (fonction `kruskalmc`, package `pgirmess`). Les analyses statistiques sont conduites avec le logiciel R (version 4.2.2).

### 3) Résultats :

#### 1) Description, classification et caractérisation des vagues de chaleur

Les vagues de chaleur identifiées entre 2008 et 2019 à l'échelle de la station météorologique de Besançon sont présentées en Tableau 11. En moyenne, 3,3 VDC ont été observées annuellement. Le même numéro identifiant est employé pour la présentation de l'ensemble des résultats. La Figure 42 présente les profils de température moyenne journalière des 39 vagues de chaleur (VDC) et illustre une variabilité importante tant en termes de durée que de profil (présence de plusieurs pics, montée progressive de la température puis redescente progressive...). La VDC numéro 34 est particulièrement longue (24 jours), intense (29.1 °C) et sévère (33,9).

Tableau 11: Inventaire et indicateurs de caractérisation des vagues de chaleur à Besançon entre 2008 et 2019.

| Numéro identifiant | Date de début | Date de fin | Durée (jours) | Intensité (°C) | Sévérité |
|--------------------|---------------|-------------|---------------|----------------|----------|
| 1                  | 2008-06-22    | 2008-06-26  | 5             | 24,1           | 3,3      |
| 2                  | 2008-06-29    | 2008-07-02  | 4             | 25,6           | 3,2      |
| 3                  | 2008-07-26    | 2008-07-31  | 6             | 24,0           | 3,6      |
| 4                  | 2009-06-29    | 2009-07-06  | 8             | 24,1           | 3,6      |
| 5                  | 2009-08-15    | 2009-08-21  | 7             | 27,8           | 8,3      |
| 6                  | 2010-06-27    | 2010-07-06  | 10            | 25,3           | 6,1      |
| 7                  | 2010-07-08    | 2010-07-16  | 9             | 26,3           | 10,4     |
| 8                  | 2010-07-19    | 2010-07-21  | 3             | 24,5           | 2,0      |
| 9                  | 2011-06-26    | 2011-06-28  | 3             | 26,5           | 4,0      |
| 10                 | 2011-08-18    | 2011-08-25  | 8             | 26,8           | 8,8      |
| 11                 | 2012-06-27    | 2012-06-30  | 4             | 25,4           | 4,3      |

|    |            |            |    |      |      |
|----|------------|------------|----|------|------|
| 12 | 2012-07-25 | 2012-07-28 | 4  | 24,8 | 3,2  |
| 13 | 2012-08-18 | 2012-08-24 | 7  | 26,7 | 6,8  |
| 14 | 2013-06-17 | 2013-06-19 | 3  | 25,5 | 4,6  |
| 15 | 2013-07-16 | 2013-07-28 | 13 | 27,1 | 14,0 |
| 16 | 2013-07-31 | 2013-08-06 | 7  | 26,0 | 7,5  |
| 17 | 2014-06-07 | 2014-06-13 | 7  | 25,6 | 6,2  |
| 18 | 2014-07-17 | 2014-07-19 | 3  | 24,8 | 3,1  |
| 19 | 2015-06-04 | 2015-06-07 | 4  | 24,7 | 2,1  |
| 20 | 2015-06-29 | 2015-07-08 | 10 | 28,6 | 17,7 |
| 21 | 2015-07-13 | 2015-07-24 | 12 | 28,8 | 17,2 |
| 22 | 2015-08-03 | 2015-08-09 | 7  | 28,2 | 7,3  |
| 23 | 2015-08-11 | 2015-08-14 | 4  | 25,6 | 4,2  |
| 24 | 2015-08-27 | 2015-08-31 | 5  | 25,2 | 6,9  |
| 25 | 2016-06-22 | 2016-06-24 | 3  | 25,9 | 3,4  |
| 26 | 2016-07-18 | 2016-07-21 | 4  | 24,9 | 4,3  |
| 27 | 2016-08-23 | 2016-08-29 | 7  | 25,1 | 6,7  |
| 28 | 2017-06-18 | 2017-06-28 | 11 | 26,3 | 10,6 |
| 29 | 2017-07-05 | 2017-07-10 | 6  | 27,4 | 6,4  |
| 30 | 2017-07-17 | 2017-07-22 | 6  | 25,3 | 5,0  |
| 31 | 2017-07-30 | 2017-08-05 | 7  | 26,5 | 6,4  |
| 32 | 2017-08-23 | 2017-08-30 | 8  | 26,2 | 8,9  |
| 33 | 2018-06-29 | 2018-07-04 | 6  | 25,8 | 6,5  |
| 34 | 2018-07-17 | 2018-08-09 | 24 | 29,1 | 33,9 |
| 35 | 2018-08-19 | 2018-08-23 | 5  | 25,0 | 3,8  |
| 36 | 2019-06-24 | 2019-07-08 | 15 | 28,2 | 20,6 |
| 37 | 2019-07-19 | 2019-07-26 | 8  | 30,0 | 15,9 |
| 38 | 2019-08-04 | 2019-08-06 | 3  | 24,7 | 2,6  |
| 39 | 2019-08-25 | 2019-08-31 | 7  | 25,1 | 4,9  |

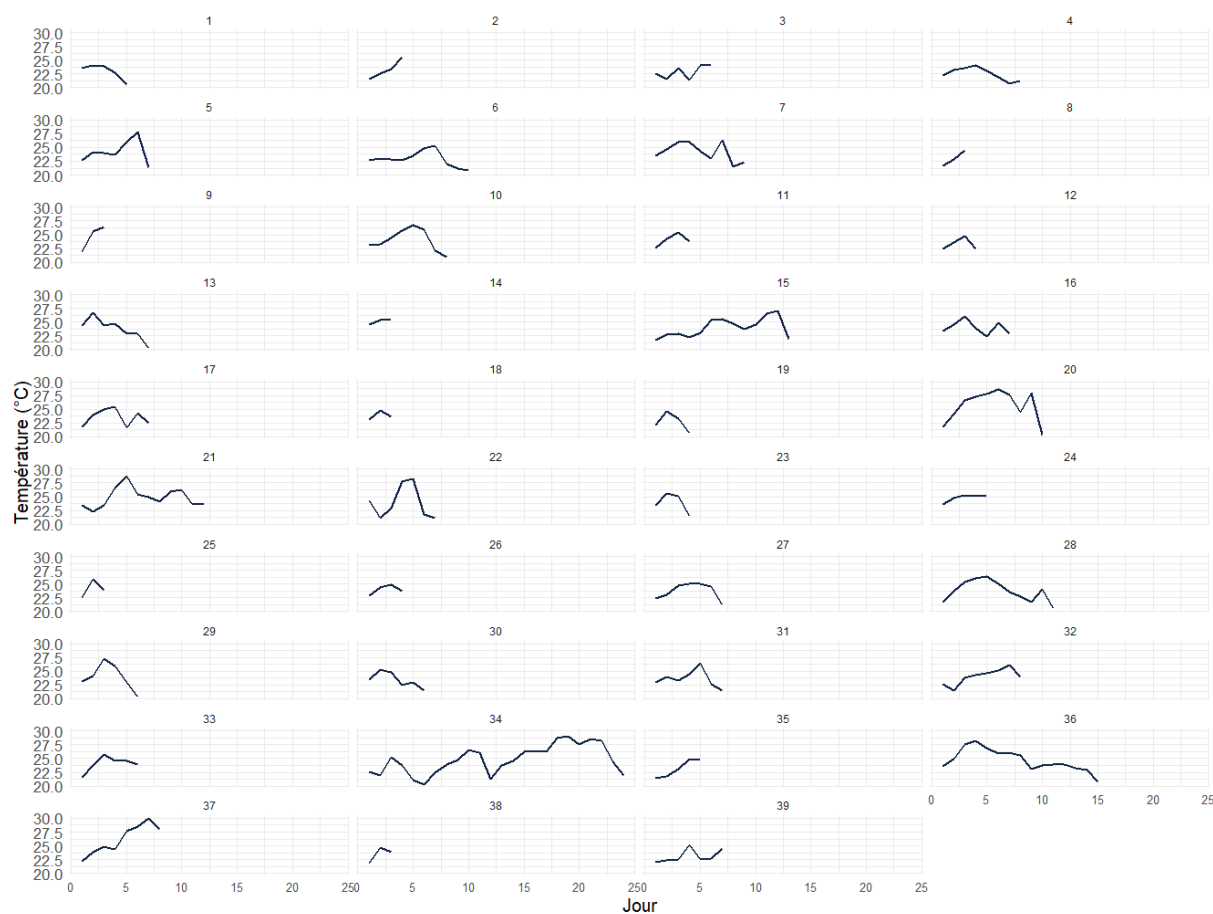


Figure 42: Profil de température moyenne journalière (°C) pour chaque vague de chaleur identifiée à Besançon entre 2008 et 2019.

Le dendrogramme de la classification hiérarchique ascendante basée sur la valeur de la sévérité est observable en Figure 43 et permet d'identifier trois classes (délimitées par les rectangles rouges) ainsi que les VDC les composant. La Figure 44 permet de visualiser la diminution rapide du WSS qui nous permet de sélectionner trois groupes (« *elbow method* »). Les effectifs des trois typologies (classes) identifiées A, B et C sont respectivement de 33, 5 et 1 VDC.

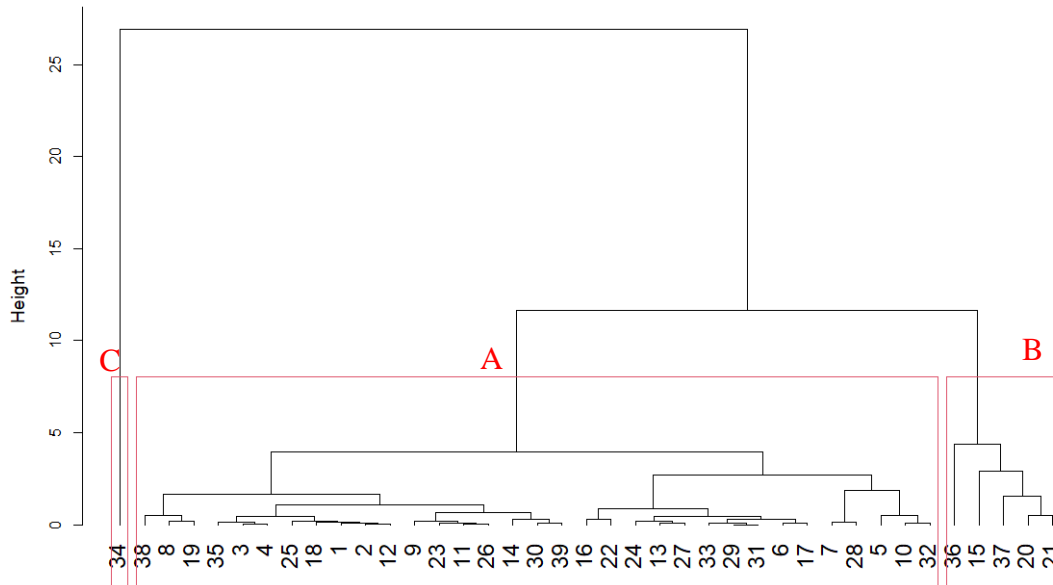


Figure 43: Dendrogramme de la classification hiérarchique ascendante des vagues de chaleur. Les numéros reportés en bas du dendrogramme correspondent au numéro identifiant des vagues de chaleur (voir Tableau 11).

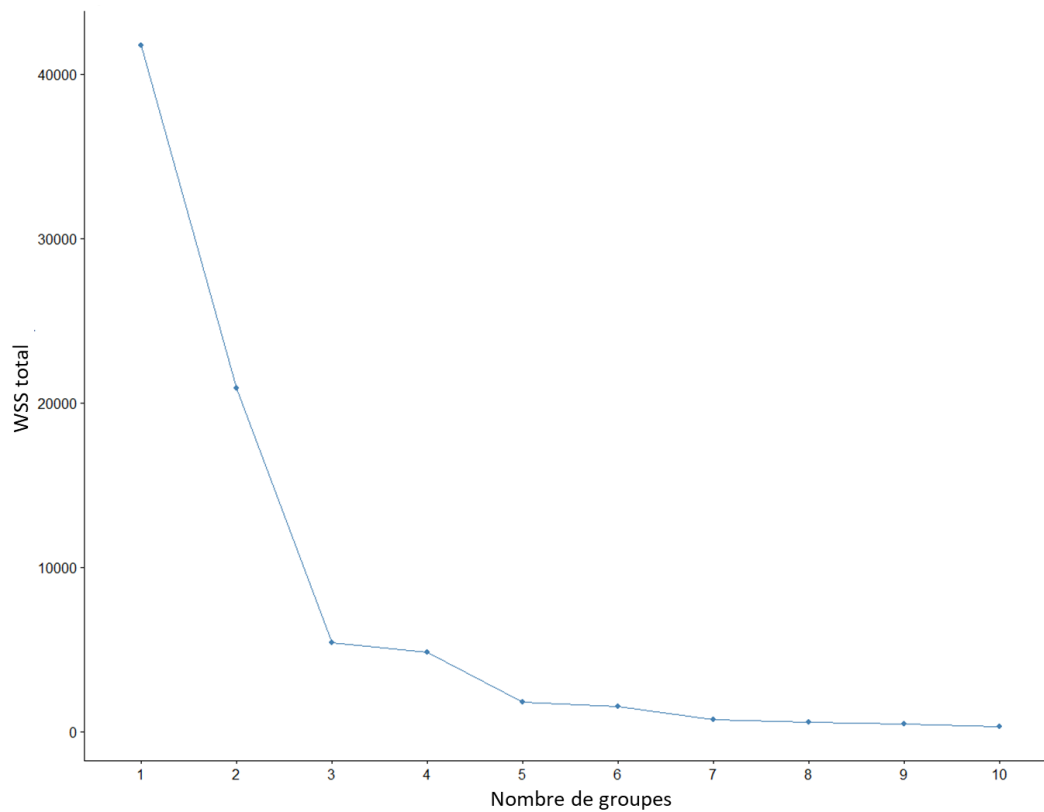


Figure 44: Profil du Within-cluster Sum of Squares (WSS) en fonction du nombre de groupes sélectionnés pour la classification hiérarchique ascendante.

## 2) Association de la typologie de VDC et leurs caractéristiques aux niveaux d'exposition à l'O<sub>3</sub>

### 2.1) *Caractérisation générale des typologies de VDC*

Un récapitulatif synthétique des caractéristiques des VDC est présenté en Tableau 12. Les données de ce tableau concernent les valeurs agrégées à la maille de modélisation. Les effectifs (n) correspondent au nombre de mailles de modélisation (pour l'année concernée) multiplié par le nombre d'épisodes de vagues de chaleur (unité statistique = maille - VDC). Ainsi la typologie C, qui ne comprend que la VDC numéro 34 apparue en 2018, comprend un effectif de 142 (soit le nombre de mailles pris en compte pour l'année 2018).

Les typologies diffèrent selon leurs caractéristiques intrinsèques (durée, intensité et sévérité) visibles en Tableau 12 et Figure 45, avec une augmentation des valeurs des indicateurs de la classe A à la classe C. La classe C contient une seule VDC particulièrement longue et sévère comparativement aux autres VDC recensées. Les niveaux d'exposition, la durée, l'intensité et la sévérité augmentent de la typologie A à la typologie C.



Tableau 12: Tableau récapitulatif des caractéristiques (MDA8, nombre de dépassements de seuil, durée, intensité, sévérité) des typologies (A, B, C) de vagues de chaleur.

|  | <b>A,<br/>n = 4 116</b> | <b>B,<br/>n = 708</b> | <b>C,<br/>n = 142</b> | <b>p<sup>b</sup></b>      |
|--|-------------------------|-----------------------|-----------------------|---------------------------|
| <i>Paramètres intrinsèques VDC<sup>a</sup></i>                                 |                         |                       |                       |                           |
| Durée (jours)  | 6 (4 - 7)               | 12 (10 - 13)          | 24 (24 - 24)          | < 2,2 x 10 <sup>-16</sup> |
| Intensité (°C)   | 25,45 (24,95 - 26,30)   | 28,60 (28,20 - 28,75) | 29,05 (29,05 - 29,05) | < 2,2 x 10 <sup>-16</sup> |
| Sévérité   | 4,9 (3,6 - 6,8)         | 17,2 (15,9 - 17,7)    | 33,9 (33,9 - 33,9)    | < 2,2 x 10 <sup>-16</sup> |
| <i>Niveaux d'exposition à l'O<sub>3</sub> (basés sur MDA8)<sup>a</sup></i>     |                         |                       |                       |                           |
| Valeur moyenne du MDA8 au cours de la VDC (µg/m <sup>3</sup> )                 | 118 (109 - 128)         | 133 (125 - 139)       | 143 (140 - 147)       | < 2,2 x 10 <sup>-16</sup> |
| Nombre de jours de dépassement du seuil OMS (100 µg/m <sup>3</sup> )           | 4,5 (3,0 - 7,0)         | 12,0 (10,0 - 12,0)    | 24,0 (24,0 - 24,0)    | < 2,2 x 10 <sup>-16</sup> |
| Nombre de jours de dépassement du seuil réglementaire (120 µg/m <sup>3</sup> ) | 2,0 (1,0 - 4,2)         | 9,0 (7,0 - 11,0)      | 22,0 (21,0 - 23,0)    | < 2,2 x 10 <sup>-16</sup> |

a : Médiane (Q1-Q3)

b : degré de significativité du test de Kruskal-Wallis

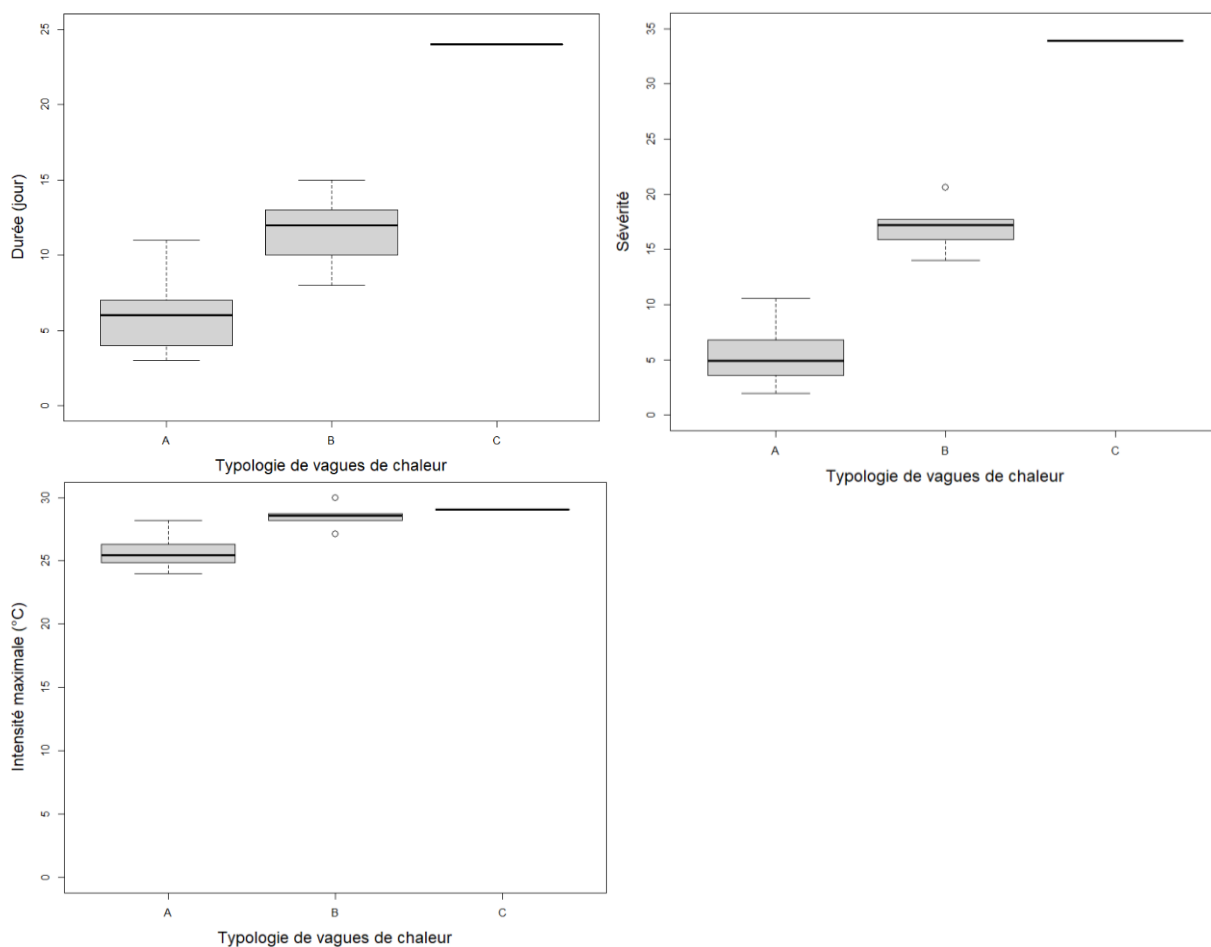


Figure 45: Durée, sévérité et intensité des vagues de chaleur en fonction de la typologie de vagues de chaleur.

## 2.2) MDA8 selon les classes de VDC et d'urbanisation

Les valeurs de MDA8 présentent un gradient croissant de la classe A à la classe C (Figure 47). La Figure 46 permet de visualiser ces valeurs d'exposition à l'échelle de chaque typologie de territoire. Les niveaux d'ozone semblent s'homogénéiser entre les typologies rurales autonome, sous faible et sous forte influence d'un pôle ainsi que l'urbain de densité intermédiaire pour la typologie C. Cette typologie ne comprend qu'une seule VDC qui est très différente des autres car sa durée, sa sévérité et son intensité sont également très élevées. Les concentrations d'O<sub>3</sub> semblent donc s'homogénéiser lorsque la vague de chaleur s'intensifie. Toutefois, seule une tendance peut être observée car un seul événement ne permet pas de tester cette hypothèse.

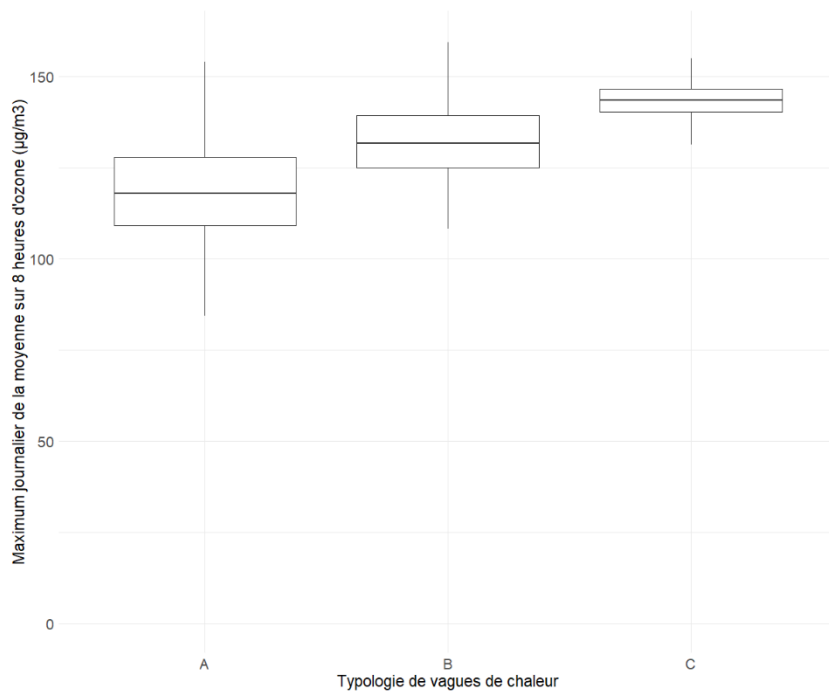


Figure 47: Maximum journalier de la moyenne glissante sur 8 heures de la concentration d'ozone ( $\mu\text{g}/\text{m}^3$ ) en fonction de la typologie de vagues de chaleur.

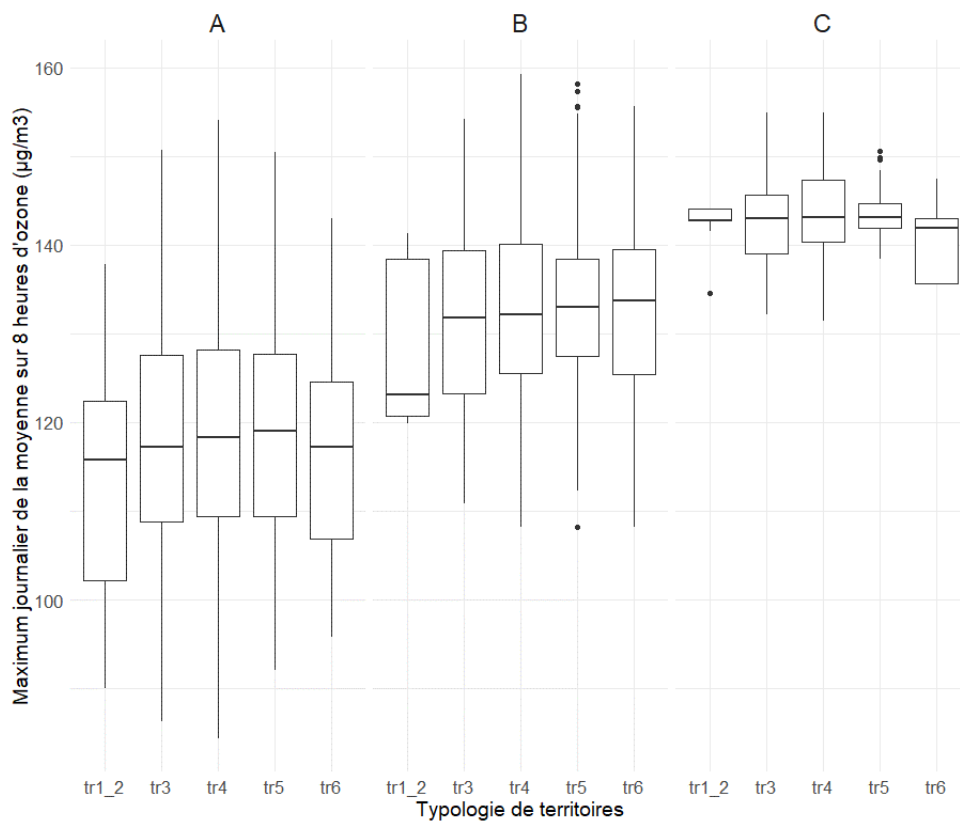


Figure 46: Maximum journalier de la moyenne glissante sur 8 heures de la concentration d'ozone ( $\mu\text{g}/\text{m}^3$ ) en fonction de la typologie de vagues de chaleur et de la typologie de territoire (tr1\_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense).

La variabilité intra typologie de VDC peut ensuite être observée à l'aide de la Figure 48. Au sein d'une même typologie de VDC, les concentrations d'ozone en fonction de la typologie de territoire se distribuent de manière très variable. La classe A, qui présente l'effectif le plus élevé (33), présente une forte variabilité des concentrations d'O<sub>3</sub>. Les VDC n°4, 5, 10, 26, 27 et 35 illustrent une possible corrélation négative entre la valeur du MDA8 et le gradient d'urbanisation, à l'inverse des VDC n°11, 12, 13, 32 qui elles illustrent une corrélation positive. Toutes ces VDC font partie de la classe A. Pour les classes B et C, les niveaux de MDA8 ne semblent pas différer en fonction du gradient d'urbanisation (Figure 48).

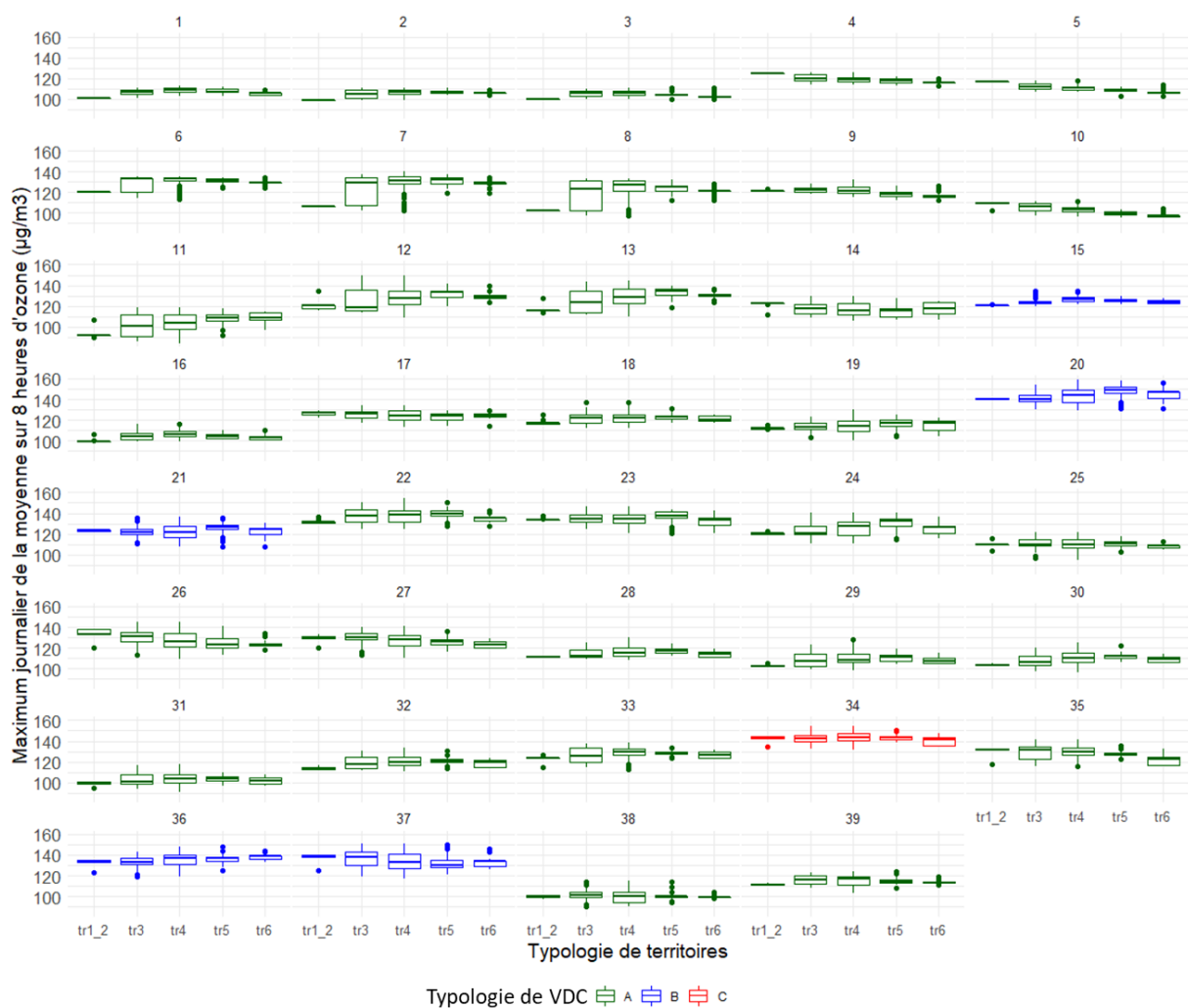


Figure 48: Maximum journalier de la moyenne glissante sur 8 heures de la concentration d'ozone ( $\mu\text{g}/\text{m}^3$ ) pour chaque VDC en fonction de la typologie de territoire (tr1\_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense). Le code couleur retranscrit la typologie de vagues de chaleur (A, B et C).

### 2.3 Dépassements de seuils selon les classes de VDC et d'urbanisation

Le nombre de dépassements du seuil OMS et du seuil réglementaire européen augmentent de la classe A à la classe C (Tableau 12, Figure 49). Le nombre de dépassements de seuils semble donc associé positivement aux caractéristiques des VDC. Le seuil OMS était dépassé sur la totalité des jours de la VDC de la classe C (n = 24 jours, Figure 42).

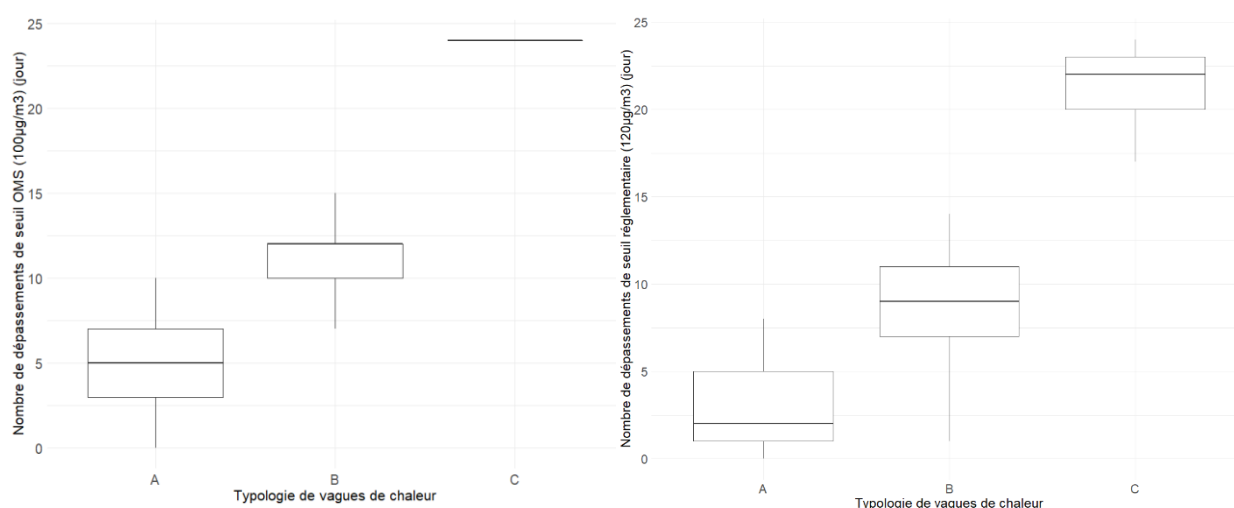


Figure 49: Nombre de dépassements (exprimé en jours) de seuil OMS ( $100 \mu\text{g}/\text{m}^3$ ) et réglementaire européen ( $120 \mu\text{g}/\text{m}^3$ ) en fonction de la typologie de VDC.

Le nombre de dépassements de seuils selon les classes VDC et d'urbanisation sont présentés pour le seuil OMS (Figure 50) et le seuil européen (Figure 51). Par exemple, 50 % des bâtiments localisés en zone rurale sous faible influence d'un pôle (tr3) ont présenté plus de 5 jours de dépassement du seuil OMS par VDC pour la typologie A. Le nombre de jours de dépassement de seuil OMS ne semble pas différer en fonction du gradient d'urbanisation, contrairement à celui du seuil européen, qui présente un optimum en milieu périurbain (typologies tr3 à tr5 : milieux ruraux sous influence (forte ou faible) d'un pôle et urbain de densité intermédiaire).

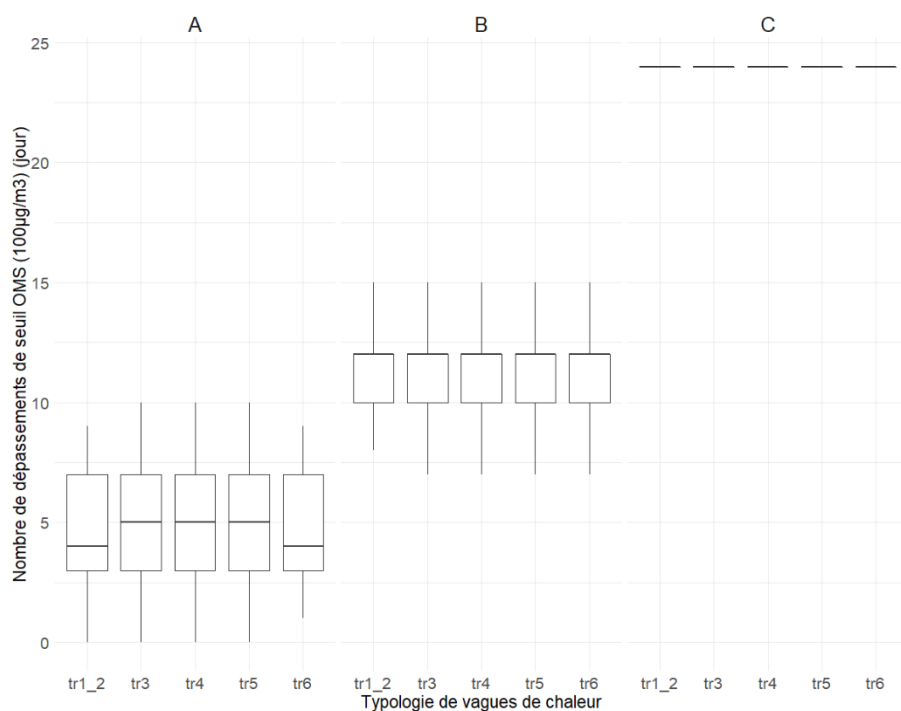


Figure 50 : Nombre de dépassements de seuil OMS (100µg/m³) (exprimé en jours) en fonction de la typologie de VDC et de la typologie de territoire (tr1\_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense).

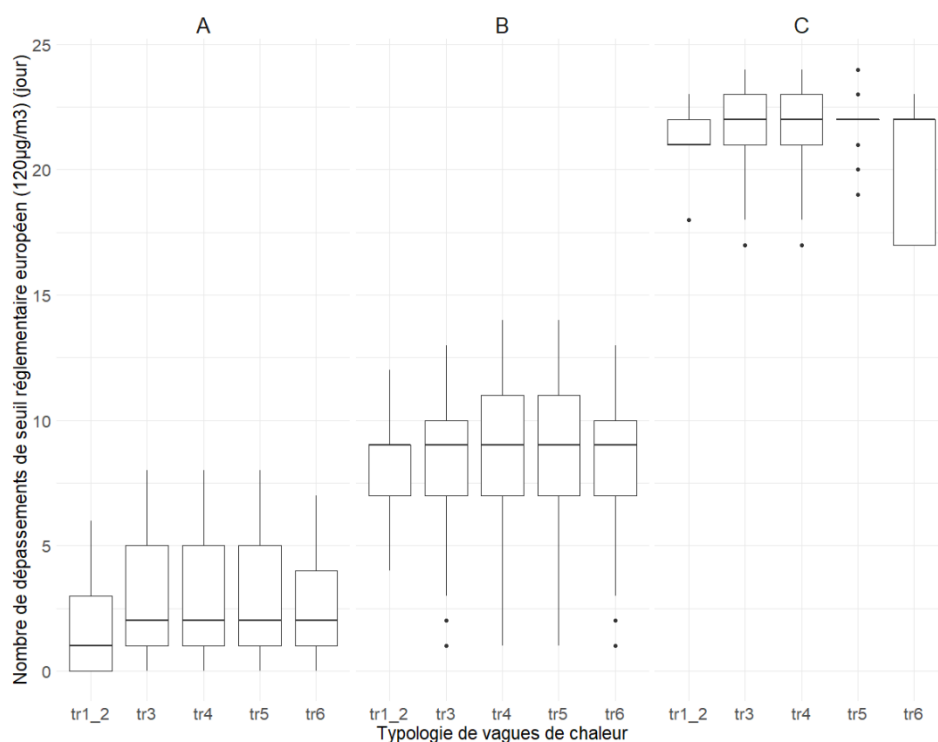


Figure 51 : Nombre de dépassements de seuil règlementaire européen (120µg/m³) (exprimé en jours) en fonction de la typologie de VDC et de la typologie de territoire (tr1\_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense).

A l'échelle de chaque VDC, le nombre de dépassements de seuils OMS semble peu ou pas influencé par le gradient d'urbanisation et la typologie de VDC (Figure 52), contrairement à celui concernant le seuil européen (Figure 53).

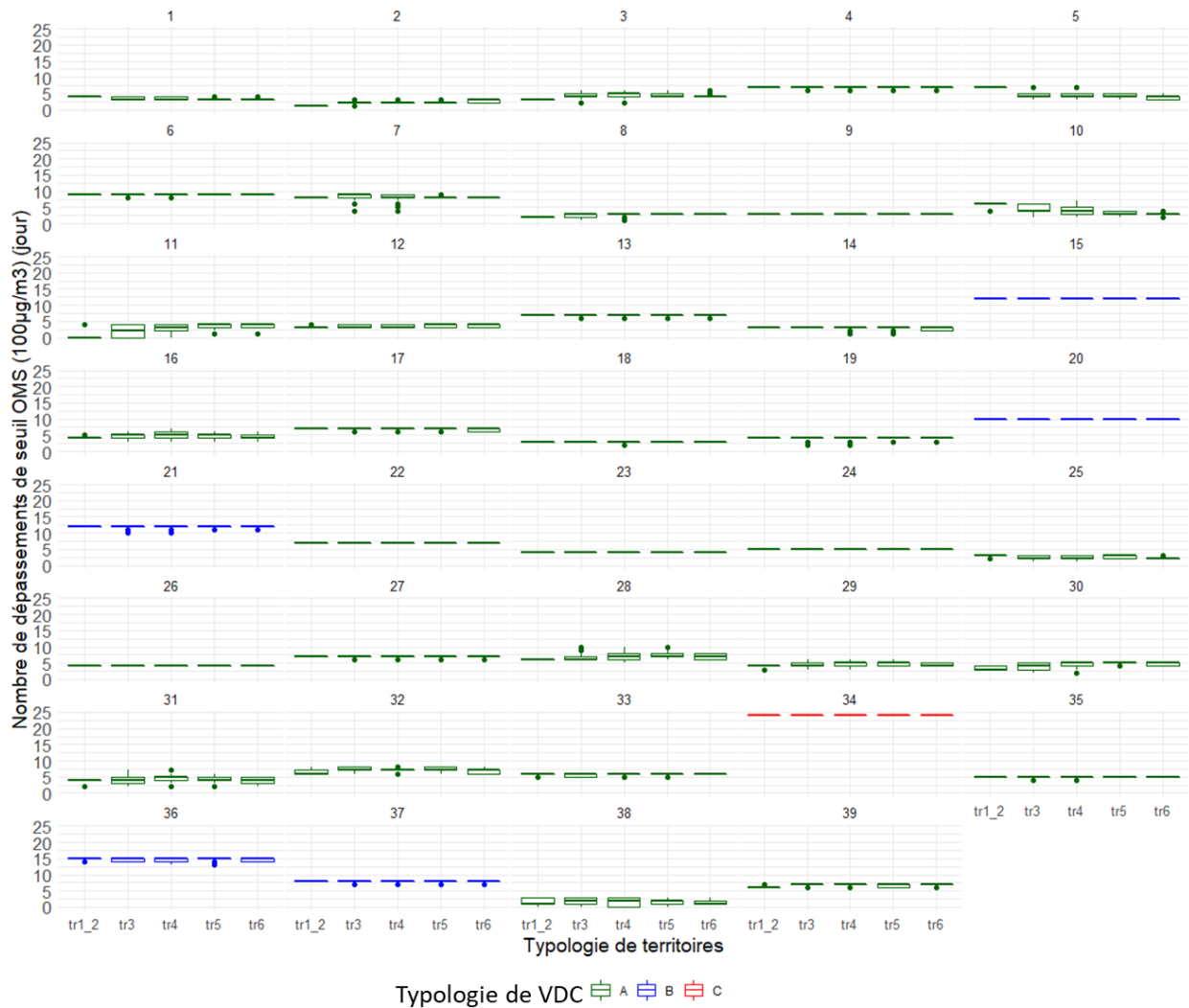


Figure 52: Nombre de dépassements de seuil OMS ( $100\mu\text{g}/\text{m}^3$ ) (exprimé en jours) pour chaque VDC en fonction de la typologie de territoire (tr1\_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense). Le code couleur retranscrit la typologie de vagues de chaleur (A, B et C).

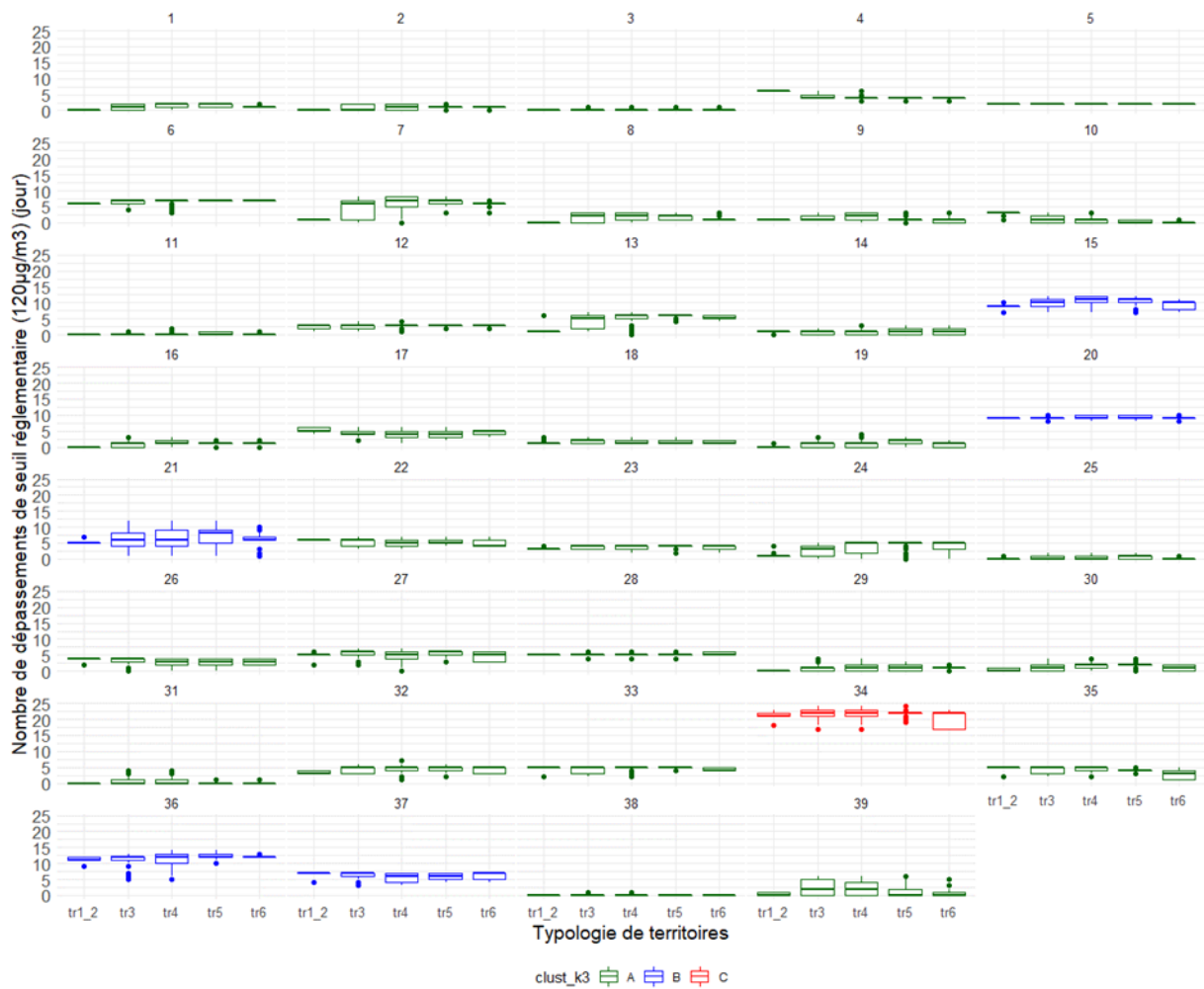


Figure 53: Nombre de dépassements de seuil réglementaire ( $120\mu\text{g}/\text{m}^3$ ) (exprimé en jours) pour chaque VDC en fonction de la typologie de territoire (tr1\_2 : rural autonome, tr3 : rural sous faible influence d'un pôle, tr4 : rural sous forte influence d'un pôle, tr5 : urbain densité intermédiaire, tr6 : urbain dense). Le code couleur retranscrit la typologie de vagues de chaleur (A, B et C).



#### 2.4) Distribution du niveau d'exposition au sein de la population de la zone d'étude

La distribution du niveau d'exposition de la population de la zone d'étude (moyenne du MDA8 au cours la VDC) est présentée pour chaque VDC en Figure 54. La forme des distributions (symétrique, dissymétrique à gauche ou à droite), l'étendue et les valeurs modales semblent très variables.

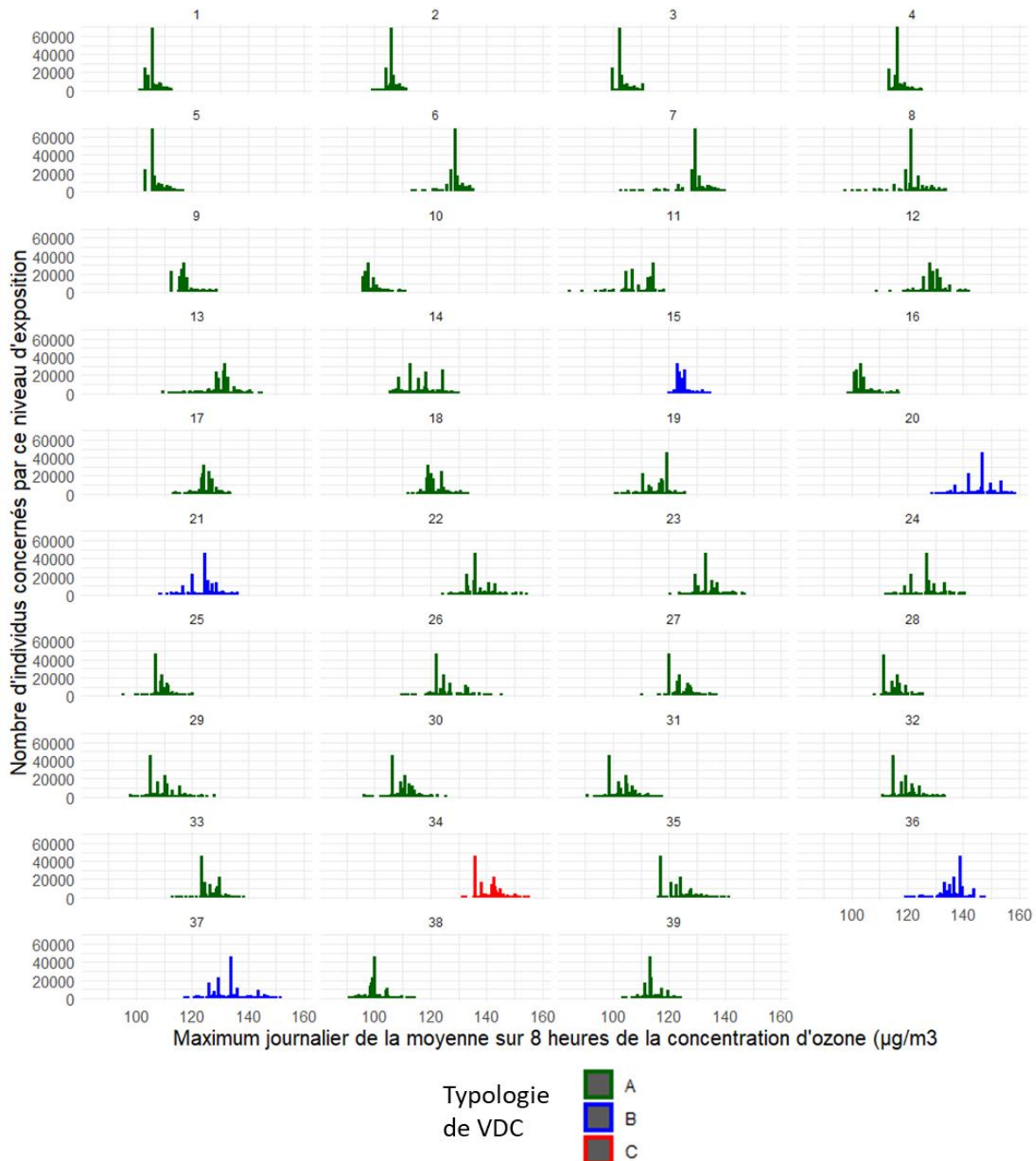


Figure 54: Histogramme de distribution de la population en fonction du maximum de la moyenne glissante sur 8 heures d' $\text{O}_3$  ( $\mu\text{g}/\text{m}^3$ ), par vague de chaleur. Le code couleur retranscrit la typologie de VDC.

3) Approche comparative des niveaux d'exposition de la population au cours des vagues de chaleurs des deux années les plus contrastées de la période 2008-2019

Le Tableau 13 présente les valeurs médianes annuelles de la durée, de l'intensité et de la sévérité des vagues de chaleur par année. Les valeurs pour ces trois indicateurs permettent d'identifier l'année 2008 et l'année 2019 comme celles présentant le plus de contraste. L'année 2008 est celle présentant les phénomènes de vague de chaleur les moins longs, intenses et sévères. *A contrario*, l'année 2019 est celle considérée comme présentant les VDC les plus intenses et sévères.

Tableau 13: Médianes de la durée, de l'intensité et de la sévérité des vagues de chaleur par année.

| Année | Durée (jours) | Intensité (°C) | Sévérité |
|-------|---------------|----------------|----------|
| 2008  | 5,0           | 24,1           | 3,3      |
| 2009  | 7,5           | 25,9           | 6,0      |
| 2010  | 9,0           | 25,3           | 6,1      |
| 2011  | 5,5           | 26,6           | 6,4      |
| 2012  | 4,0           | 25,4           | 4,3      |
| 2013  | 7,0           | 26,0           | 7,5      |
| 2014  | 5,0           | 25,2           | 4,7      |
| 2015  | 6,0           | 26,9           | 7,1      |
| 2016  | 4,0           | 25,1           | 4,3      |
| 2017  | 7,0           | 26,3           | 6,4      |
| 2018  | 6,0           | 25,8           | 6,5      |
| 2019  | 7,5           | 26,7           | 10,4     |

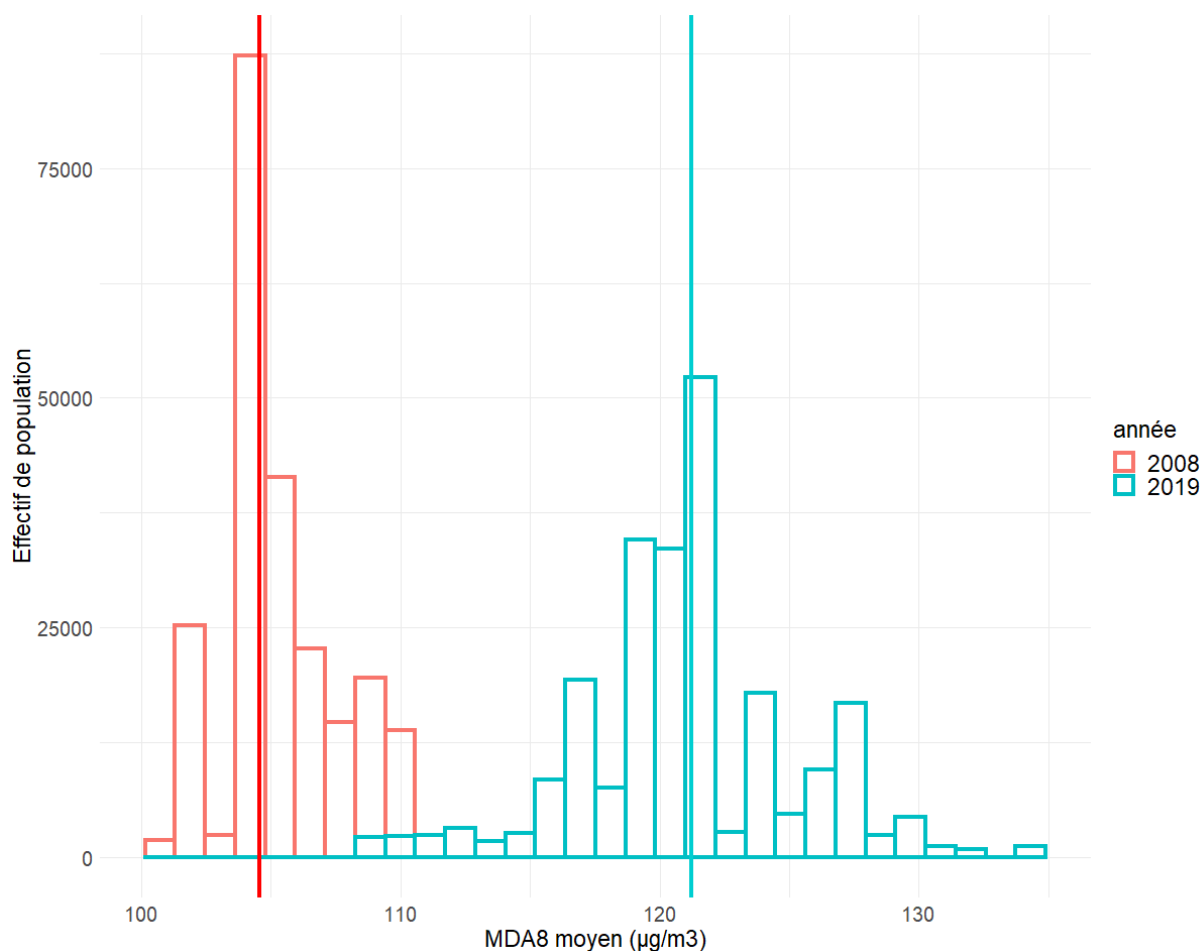


Figure 55: Distribution du niveau d'exposition de la population de la zone d'étude (moyenne du MDA8 ( $\mu\text{g}/\text{m}^3$ ) au cours des VDC d'une même année), années 2008 et 2019 (aucune valeur ne dépasse  $110 \mu\text{g}/\text{m}^3$  pour l'année 2008). La ligne verticale rouge représente la valeur de MDA8 pour la maille sur laquelle se situe la station de mesure de Besançon en 2008. La ligne verticale bleue représente cette même valeur pour l'année 2019.

La distribution du niveau d'exposition de la population de la zone d'étude (moyenne du MDA8 au cours des VDC d'une même année) est représentée sur le même graphique pour les années 2008 et de 2019 (Figure 55). Ce graphique confirme la faible superposition des niveaux d'exposition de la population entre ces deux années et déjà observée à l'échelle de la VDC en Figure 54. Les lignes verticales rouges et bleues représentent les valeurs de MDA8 moyen sur la maille de modélisation sur laquelle se situe la station de mesure de Besançon, pour 2008 et 2019 respectivement. La population soumise au-delà de ces valeurs représentaient 86% et 49% de la population, en 2008 et 2019 respectivement.

#### 4) Discussion :

La définition des périodes de vagues de chaleur est basée sur les données de température moyenne journalière mesurées par la station météorologique de Besançon. Cette définition repose sur l'hypothèse d'homogénéité de la valeur de la température sur l'ensemble du territoire étudié. Cette démarche de définition uniforme des mêmes périodes de VDC sur l'ensemble du territoire a permis une première approche de comparaison des niveaux d'exposition à l'O<sub>3</sub> sur l'ensemble du territoire étudié. Cependant, le gradient d'urbanisation et d'altitude du territoire est susceptible de fragiliser cette hypothèse, en particulier au Sud-Est de la zone d'étude (gradient altitudinal plus marqué, cf. chapitre 4). Une approche complémentaire considérant la dynamique spatiale de la température permettrait d'analyser plus finement l'influence de la dynamique de la température sur celle de l'O<sub>3</sub> selon le gradient d'urbanisation, le différentiel de température entre typologie urbaine et rurale autonome étant probablement élevé. Lors de la démarche de classification des vagues de chaleur, trois groupes ont été définis à partir du dendrogramme et selon le profil du « Within-cluster Sum of Squares » (WSS) total (Figure 44). Une autre sélection de la hauteur sur le dendrogramme aboutirait à retenir quatre ou cinq classes. La pertinence d'un tel choix reste à établir mais nécessiterait la prise en compte d'un nombre de vague de chaleur plus élevé.

Lors de l'année 2019, la plus sévère en termes de VDC, pour les périodes de dépassements des seuils d'alerte canicule et les départements concernés en Bourgogne-Franche-Comté, la surmortalité était estimée à +7 % (Ciccardini and Retel, 2019). Lors de l'année 2022, pendant les périodes de dépassements des seuils d'alerte canicule et pour les départements concernés en Bourgogne-Franche-Comté, la surmortalité était estimée à +13 % (Retel et al., 2022). La sévérité des périodes de VDC était moindre en 2022 par rapport à 2019 mais le nombre de jours de VDC en 2022 a été particulièrement élevé. Une exposition durable et répétée à la température a pu affecter l'état de santé de la population, seule ou en synergie avec l'exposition à l'O<sub>3</sub>. Ces éléments illustrent à la fois la complexité des phénomènes observés et la nécessité de mieux prévenir les impacts sanitaires associés à ces périodes d'exposition aiguë.

L'une des forces de cette approche repose sur la prise en compte de l'hétérogénéité des niveaux d'exposition sur la zone d'étude, et en particulier selon le degré d'urbanisation. La Figure 55 illustre parfaitement l'importance de la prise en compte de cette hétérogénéité car les niveaux d'exposition de la population seraient sous-estimés pour 86% et 49 % de la population en 2008 et 2019 respectivement, si la seule valeur de la maille de la station de mesure était intégrée. L'écart entre la valeur sur la maille de la station et les différentes mailles du territoire pourrait être modulée par les caractéristiques des vagues de chaleur et/ou être spatialement structurée. La distribution du MDA8 selon la typologie de territoire semble variable mais non structurée par la typologie de vague de chaleur. Le milieu périurbain (rural sous forte et sous faible influence d'un pôle, urbain densité intermédiaire) semble généralement plus exposé à l'O<sub>3</sub> et aux dépassements de seuil réglementaire au cours des VDC. Les niveaux d'exposition semblent s'homogénéiser pour la vague de chaleur n°34, seule dans la typologie C, et correspondant à la VDC la plus longue et la plus sévère. Une absence de réponse univoque des niveaux d'exposition selon la typologie de territoire est observée. Toutefois, cette absence de signal est potentiellement due à la typologie de territoire utilisée. La typologie de territoire utilisée est basée sur les effectifs et de populations et la part d'actifs travaillant dans le pôle urbain considéré. Cette approche ne prend pas en compte l'occupation du sol et n'est peut-être pas totalement adaptée dans ce contexte.

La difficulté la plus prégnante rencontrée dans la mise en place de ce chapitre est l'association de différentes échelles temporelles (jour, année, VDC) et spatiales (bâtiment, maille de modélisation, IRIS) avec une évolution des entités à spatialiser d'une année à l'autre. Le temps nécessaire à la conceptualisation des bases de données et, au croisement et la mise en cohérence des différentes échelles spatiales et temporelles, a été conséquent. Certaines manipulations n'étaient pas réalisables faute de compatibilité des bases de données. Par exemple, l'évolution du bâti (en découpe, en effectif, identifiants instables) n'a pas permis de travailler à l'échelle du bâtiment et d'agréger les données d'exposition sur plusieurs années.

La distribution de la population recensée à l'échelle de l'IRIS dans les bâtiments a permis de quantifier l'exposition de la population, et ce en fonction du gradient d'urbanisation. Cependant, le nombre de bâtiments diffère fortement entre les typologies (très faible dans le milieu rural autonome) ce qui limite

la comparabilité des données entre typologies de territoire. Les niveaux d'exposition sont calculés à partir de la maille de modélisation sur laquelle est localisé chaque bâtiment. Un nombre élevé de bâtiments situés sur une même maille a pour conséquence potentielle une structuration non indépendante des données. Une analyse statistique prenant en compte cette structuration (modèles multiniveaux par exemple) pourra être menée par la suite pour tester cette hypothèse.

Le nombre de dépassements de seuils, calculés à l'échelle des VDC pour chaque bâtiment, permet d'estimer la sensibilité du MDA8 à dépasser ces seuils au cours de ces périodes particulières. Les résultats obtenus montrent une saturation rapide du seuil recommandé par l'OMS mais également du seuil réglementaire européen. La durée des VDC conditionne nécessairement le nombre de jours de dépassements de seuil possibles. La durée des VDC s'allongeant de la classe A à la classe C, le nombre de dépassements possible augmente mécaniquement. Il est donc difficile de discriminer totalement l'influence des différentes caractéristiques de VDC sur le nombre de dépassement de seuil. L'augmentation du nombre de dépassements de seuils observés au chapitre précédent entre 2008 et 2019 coïncide avec à une l'augmentation des caractéristiques de VDC (durée, intensité, sévérité). Le risque sanitaire associé à l'exposition à ces niveaux d'O<sub>3</sub> et à la température au cours des VDC devrait donc également croître.

Il existe un contraste d'exposition de la population entre les années les plus contrastées en termes de VDC. L'augmentation de la durée, de l'intensité et de la sévérité des VDC est associé à une augmentation de l'exposition de la population. Cependant, les caractéristiques des VDC n'expliquent que partiellement les niveaux d'exposition à l'O<sub>3</sub>. La relation observée entre les caractéristiques des VDC et le MDA8 semble complexe et d'autres paramètres météorologiques pourraient être pris en compte. La prise en compte de l'ensoleillement, des régimes de vent, de l'humidité relative pourrait aider à caractériser les conditions météorologiques dans lesquelles s'installent les VDC. Le niveau de sécheresse préalable à l'apparition d'une vague de chaleur, résultant des conditions météorologiques citées précédemment, possède également une influence notable sur l'intensité de la vague de chaleur. En effet, il existe une boucle rétroactive positive entre la température de l'air et la sécheresse du sol (Perkins, 2015). Certaines études ont caractérisé des types de vagues de chaleur en relation avec des

conditions atmosphériques responsables localement de pics d'O<sub>3</sub>. Il semblerait que toutes les vagues de chaleur ne soient pas responsable d'une amplification des concentrations d'O<sub>3</sub>. La capacité d'une vague de chaleur à influencer ces concentrations serait dépendante des conditions météorologiques et de circulation atmosphérique (Yang et al., 2022). Notre approche exploratoire a permis de décrire l'influence partielle des caractéristiques des VDC sur les niveaux d'exposition à l'O<sub>3</sub>. La prochaine phase de ces travaux, permettra potentiellement de rechercher plus finement les paramètres qui modulent ces contrastes entre VDC.

### **5) Conclusion :**

Ce chapitre a confirmé une influence des vagues de chaleur (VDC) et de leurs caractéristiques sur les niveaux d'exposition à l'O<sub>3</sub>. Les VDC soumettent la population à des dépassements de seuils (sanitaire et réglementaire) de plus en plus nombreux. Les territoires périurbains pourraient être soumis à des niveaux d'exposition et à un nombre de dépassements de seuils réglementaire plus élevés.

Toutefois, cette première approche exploratoire a également suscité de nombreuses interrogations tant techniques que conceptuelles qui pourront être poursuivies par la suite. L'utilisation d'une typologie de territoire reposant sur l'occupation du sol pourra être privilégiée dans un second temps. Enfin, une définition des vagues de chaleur basée sur des valeurs de la température définie à une échelle spatiale plus fine (par exemple kilométrique) pourrait participer à une meilleure connaissance des vagues de chaleur, de leurs caractéristiques, et de leur influence sur les niveaux d'exposition de la population à l'O<sub>3</sub>.

### **Références bibliographiques :**

Ciccardini, M., Retel, O., 2019. Bulletin de santé publique canicule en Bourgogne et Franche-Comté.

Septembre 2019. Santé publique France.

D'Alessandro, C., Levy, D., Regnier, T., 2021. Une nouvelle définition du rural pour mieux rendre

compte des réalités des territoires et de leurs transformations – La France et ses territoires |

Insee. Insee.

- European Environment Agency., 2020. Air quality in Europe: 2020 report. (No. 09/2020). Publications Office, LU.
- Fu, S., Zhou, Y., Peng, L., Ye, X., Yang, D., Yang, S., Zhou, J., Luo, B., 2022. Interactive effects of high temperature and ozone on COPD deaths in Shanghai. *Atmospheric Environment* 278, 119092. <https://doi.org/10.1016/j.atmosenv.2022.119092>
- INSEE, 2022. Méthode de construction de la grille communale de densité [WWW Document]. URL <https://www.insee.fr/fr/statistiques/fichier/2114627/methode-constitution.pdf> (accessed 11.17.22).
- Insee, 2021. Population en 2018 - Recensement de la population - Base infracommunale (IRIS) [WWW Document]. URL <https://www.insee.fr/fr/statistiques/5650720#consulter> (accessed 1.19.23).
- Ouzeau, G., Soubeyroux, J.-M., Schneider, M., Vautard, R., Planton, S., 2016. Heat waves analysis over France in present and future climate: Application of a new method on the EURO-CORDEX ensemble. *Climate Services* 4, 1–12. <https://doi.org/10.1016/j.cliser.2016.09.002>
- Perkins, S.E., 2015. A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research* 164–165, 242–267. <https://doi.org/10.1016/j.atmosres.2015.05.014>
- Retel, O., François, D., Chêne, S., 2022. Bulletin de santé publique canicule en Bourgogne et Franche-Comté. Octobre 2022. [WWW Document]. URL <https://www.santepubliquefrance.fr/regions/bourgogne-franche-comte/documents/bulletin-regional/2022/bulletin-de-sante-publique-canicule-en-bourgogne-franche-comte.-bilan-ete-2022> (accessed 1.14.23).
- Soubeyroux, J.-M., Ouzeau, G., Schneider, M., Cabanes, O., Kounkou-Arnaud, R., 2016. Les vagues de chaleur en France : analyse de l'été 2015 et évolutions attendues en climat futur. *Météorologie* 8, 45. <https://doi.org/10.4267/2042/60704>
- Sun, Y., Ilango, S.D., Schwarz, L., Wang, Q., Chen, J.-C., Lawrence, J.M., Wu, J., Benmarhnia, T., 2020. Examining the joint effects of heatwaves, air pollution, and green space on the risk of



preterm birth in California. *Environ Res Lett* 15, 104099. <https://doi.org/10.1088/1748-9326/abb8a3>

Ulpiani, G., 2021. On the linkage between urban heat island and urban pollution island: Three-decade literature review towards a conceptual framework. *Science of The Total Environment* 751, 141727. <https://doi.org/10.1016/j.scitotenv.2020.141727>

World Health Organization, 2021. WHO global air quality guidelines: particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.

Yang, X., Zeng, G., Iyakaremye, V., Zhu, B., 2022. Effects of different types of heat wave days on ozone pollution over Beijing-Tianjin-Hebei and its future projection. *Science of The Total Environment* 837, 155762. <https://doi.org/10.1016/j.scitotenv.2022.155762>

## Chapitre 7 : Discussion générale et perspectives

### 1) Synthèse des principaux résultats

Nous avons développé la première approche de ces travaux en réponse à des manques de précision et d'exactitude dans les données modélisées à notre disposition. Une méthodologie à la fois accessible, souple, transposable et robuste a ainsi été proposée pour corriger les données modélisées. Cette méthodologie nous a permis de mieux représenter le contraste temporel fortement atténué par le modèle, permettant ainsi d'analyser les pics d'ozone estivaux initialement effacés par la modélisation brute. Le chapitre 3 constitue la première étape des travaux futurs en vue d'évaluer l'exposition de la population à l'O<sub>3</sub>, en particulier en période de vague de chaleur. Les contrastes temporels et spatiaux mieux représentés en Chapitre 3 ont permis de conduire les travaux présentés en chapitres 4 à 6 (Figure 56). En effet, l'étude de la dynamique spatiale et temporelle de l'O<sub>3</sub>, la quantification sur un continuum spatial des dépassements de seuils et le focus sur les pics d'O<sub>3</sub> en périodes de vague de chaleur n'auraient pas été possibles sans cette première étape (Chapitre 3).

Le chapitre 4 nous a ensuite permis de caractériser les déterminants de l'O<sub>3</sub> et les contrastes spatiaux dans les concentrations d'O<sub>3</sub> sous l'influence simultanée de ces déterminants. Nous avons ainsi pu distinguer différentes dynamiques temporelles se superposant sur des territoires caractérisés comme urbains, périurbains et ruraux. Les déterminants de l'O<sub>3</sub> étaient individuellement identifiés (Agathokleous et al., 2022b; Gao et al., 2020; Liu et al., 2021). Cette approche a permis d'illustrer la résultante de ces influences simultanées sur la dynamique spatiale et temporelle de l'O<sub>3</sub>.

L'objectif visé du Chapitre 5 était d'évaluer de manière spatialement continue l'exposition de la population et sa traduction au regard du seuil réglementaire européen et des recommandations de l'OMS. La construction d'une estimation du MDA8 (maximum journalier de la moyenne glissante sur 8h) nous a permis de mettre en évidence une variabilité élevée du nombre de dépassements de seuils entre mailles de modélisation, reflet d'une hétérogénéité spatiale dans l'exposition de la population. Nous avons également constaté une augmentation sensible des dépassements de seuils de 2008 à 2019. Cette hétérogénéité spatiale, mise en relation avec la valeur qui aurait été projetée sur toute la zone si une

station avait été utilisée, nous a permis d'identifier un risque de sous-estimation des niveaux d'exposition de la population.

Renforcé par le constat des niveaux croissants d'O<sub>3</sub> au cours de la période étudiée, l'approche conduite dans le Chapitre 6 visait à décrire l'influence des vagues de chaleur et de leurs caractéristiques sur les niveaux d'exposition de la population de notre zone d'étude. Lors de cette approche exploratoire, nous avons pu identifier une influence des caractéristiques des vagues de chaleur sur les niveaux d'exposition et le nombre de dépassements. Cette augmentation du nombre des dépassements de seuils, en particulier avec l'augmentation de la durée des vagues de chaleur, est en accord avec la bibliographie (Lin et al., 2001; Struzewska and Jefimow, 2016; Varotsos et al., 2019). Des années contrastées vis-à-vis des caractéristiques des vagues de chaleur, montraient un différentiel de niveau d'exposition de la population et une augmentation de la variabilité du MDA8 pour l'année la plus fortement soumise aux vagues de chaleur.

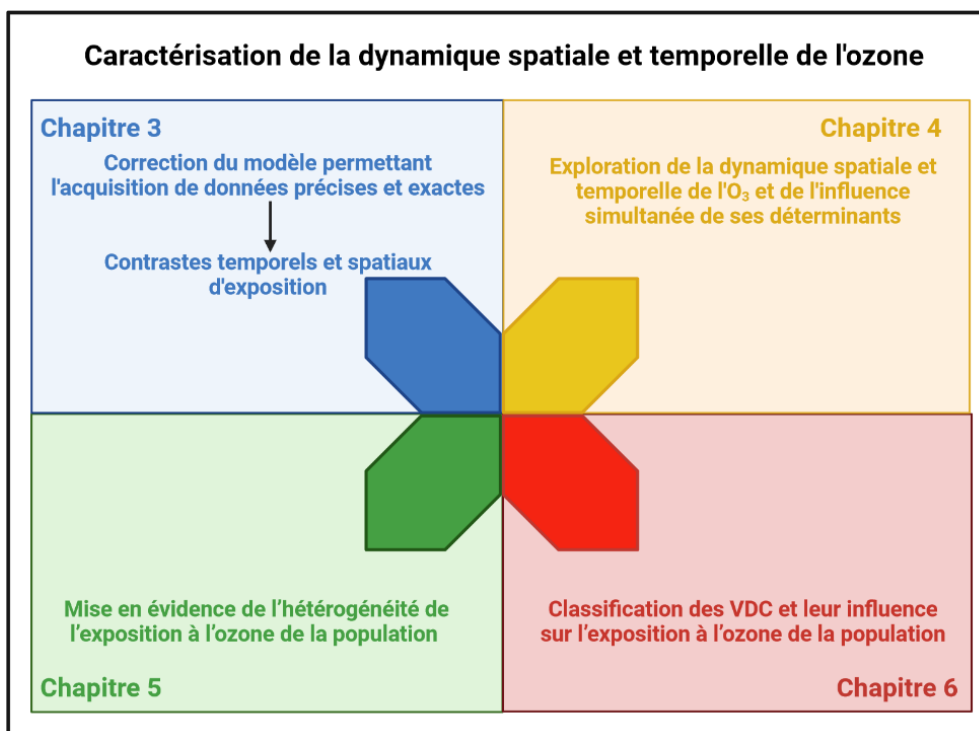


Figure 56: Apport de des quatre chapitres et synthèse des résultats du manuscrit.

## **2) Les contrastes d'exposition : concept central de notre démarche**

Le concept central de nos travaux repose sur l'identification des contrastes spatiaux et temporels d'exposition. Dans chacun des chapitres nous souhaitons mieux représenter ces contrastes ou illustrer comment l'absence de contraste pouvait affecter grandement l'évaluation de l'exposition voire son risque associé. Ainsi, en chapitre 3 l'enjeu était précisément de représenter ces contrastes en particulier dans leur dimension temporelle pour représenter correctement les pics d'exposition qui constituent un enjeu sanitaire. Dans le chapitre 4, nous avons distingué et caractérisé trois zones d'exposition qui correspondaient à trois typologies de dynamiques temporelles. Dans les chapitres 5 et 6, nous avons illustré le différentiel possible entre une unique valeur fournie à la station de mesure et un continuum de valeurs estimées sur une échelle fine projetée sur le territoire. D'après nos résultats, l'utilisation de la donnée à la station pour quantifier l'exposition globale de la population à l'ozone aurait systématiquement conduit à une sous-estimation des valeurs : du nombre de dépassement de seuils annuels pour au moins 75% du territoire étudié et de la concentration d'O<sub>3</sub> pour 86 % de la population au cours des vagues de chaleur de l'année 2008, considérée comme l'année la moins touchée par les vagues de chaleur de notre période d'étude. Cette erreur dans l'estimation semble de plus différentielle, 1) dans le temps (l'écart estimé étant variable au cours des 12 années de l'étude), 2) en fonction de l'intensité des vagues de chaleurs (l'écart estimé étant variable entre les périodes de vague de chaleur de la première et celles de la dernière année de l'étude) et 3) vraisemblablement dans l'espace (cet aspect n'ayant pas été exploré dans le cadre de ce travail). Ces résultats renforcent la nécessité d'estimer le niveau d'exposition de la population à une échelle fine. Dans ce but, il apparaît indispensable de favoriser la construction et la validation temporelle et spatiale de modèles de prédiction de polluants atmosphériques définis sur une échelle fine (kilométrique).

## **3) Croisement des différentes échelles spatiales et temporelles des données mobilisées**

La complexité du travail réalisé a résidé dans la multiplication des échelles spatiales et temporelles et de la nécessité de les faire coïncider à des niveaux variables selon les objectifs visés dans des différents chapitres (3 à 6) (Figure 57). Le maillage de modélisation était l'unité spatiale sur laquelle ont été définis et projetés quotidiennement les niveaux d'exposition à l'O<sub>3</sub>. Ce maillage était également l'élément pivot

entre les différentes échelles et unités spatiales et temporelles employées dans les différents chapitres. L'unité spatiale du bâtiment a par exemple été utilisée à l'échelle de la vague de chaleur et de l'année. De la même manière, l'échelle temporelle de la vague de chaleur a été utilisée à l'échelle du bâtiment, de la maille de modélisation et de la station météorologique, et ce uniquement pour le chapitre 6 (code couleur rouge). La Figure 57 illustre donc le fait que les diverses approches menées ont impliqué le croisement de diverses sources, échelles et/ou unités de données. Un travail de data management conséquent a dû être mis en œuvre pour croiser, rendre compatibles (dans l'espace et dans le temps) et moduler le niveau d'agrégation des données selon les objectifs fixés. La compatibilité immédiate des différentes sources de données était limitée mais a été finalement possible grâce à des requêtes spatiales et à la mise en place d'une base de données intégrant des tables définies annuellement. Toutefois, la stabilité et la compatibilité des données au sein d'une même source de données est également problématique. Par exemple, la couche de bâti indifférencié issue de la BDTPO de 2008 (50707 bâtiments sur le maillage de modélisation) n'était pas compatible avec celle de 2018 (79086 sur cette même zone). En effet, la méthodologie employée dans la découpe des bâtiments ayant évolué, une part non-négligeable des bâtiments ont été subdivisés en plusieurs entités. Les bâtiments et leurs identifiants (pour un même corps de bâtiment) n'étaient donc pas traçables, rendant impossible l'agrégation de certains indicateurs sur plusieurs années.

| Échelles ou unités spatiales / Échelles temporelles | Bâtiment | IRIS  | Commune (typologie de territoires) | Maille de modélisation (3 km x 3 km) | Modèle numérique de terrain (25m) | Station météorologique | Station de suivi de la qualité de l'air | Végétation | Réseau routier | Atlas urbain |         |
|---|----------|-------|------------------------------------|--------------------------------------|-----------------------------------|------------------------|---|------------|----------------|--------------|---------|
|   | 1 heure  |       |                                    |                                      |                                   |                        |   | ■          |                |              |         |
| 8 heures  |          |       |                                    | ■ ■                                  |                                   |                        | ■                                       |            |                |              | ■ ■     |
| 24 heures   |          |       |                                    | ■ ■ ■                                |                                   | ■ ■                    | ■                                       |            |                |              | ■ ■ ■ ■ |
| Vague de chaleur*                                   | ■        |       |                                    | ■                                    |                                   | ■                      |   |            |                |              | ■       |
| Saison  |          |       |                                    |                                      |                                   |                        |   |            |                |              |         |
| Année   | ■ ■      | ■ ■   |                                    | ■ ■                                  |                                   |                        | ■                                       |            |                |              | ■ ■ ■   |
| 12 années (2008-2019)                               |          |       | ■                                  | ■ ■ ■                                | ■ ■                               |                        | ■                                       | ■          | ■              | ■            | ■ ■ ■ ■ |
| Horizon 2020-2050                                   |          |       |                                    |                                      |                                   |                        |   |            |                |              |         |
|   | ■ ■      | ■ ■ ■ | ■                                  | ■ ■ ■ ■                              | ■ ■                               | ■ ■                    | ■ ■                                     | ■          | ■              | ■            |         |

\* 3 jours ou plus

■ Chapitre 3 ■ Chapitre 4 ■ Chapitre 5 ■ Chapitre 6

Figure 57 : Synthèse des différentes échelles ou unités temporelles et spatiales mobilisées dans les différents chapitres de la thèse.

Le Tableau 14 illustre lui aussi le statut charnière du maillage de modélisation entre les différentes entités mobilisées mais également l'évolution de l'unité statistique modulée par les objectifs des chapitres.

Tableau 14: unité(s) statistique(s) utilisée(s) dans chacun des chapitres de la thèse.

| Chapitre     | Unité statistique principale |
|--------------|------------------------------|
| ■ Chapitre 3 | Maille - jour                |
| ■ Chapitre 4 | Maille – 12 années           |
| ■ Chapitre 5 | Maille - année               |
| ■ Chapitre 6 | Bâtiment<br>Vague de chaleur |

#### 4) Forces

L'un des atouts de notre méthodologie est la résolution journalière et kilométrique des données sur 12 années. Cette échelle temporelle et spatiale fine nous a permis d'adapter l'unité statistique sur laquelle nous souhaitons nous appuyer en fonction de l'objectif fixé. Cette flexibilité n'est possible qu'en ayant des données à échelle fine pouvant ensuite être agrégées (ou non) aux échelles choisies.

La période longue d'intégration des données nous a permis de stabiliser nos analyses et de bénéficier d'une puissance statistique élevée (stabilisation de la classification des dynamiques temporelles (Chapitre 4) et identification de 39 vagues de chaleur (Chapitre 6)). Ce continuum temporel a également été accompagné d'un continuum spatial qui a permis de travailler sur les contrastes aussi bien spatiaux que temporels d'exposition. Le contraste temporel mieux représentés après correction (Chapitre 3) est également un atout permettant d'identifier des épisodes d'exposition aiguë et d'éventuelles dynamiques temporelles différentes (Chapitre 4). Cette correction a également augmenté la précision et l'exactitude des données.

L'ensemble des approches a été mené sur un même territoire (Besançon), ce qui nous a permis de mettre au point de multiples indicateurs (caractérisation des mailles et de typologies dans le Chapitre 4, nombre de dépassements de seuils dans le Chapitre 5...) sur la même zone, de les croiser, et de mieux appréhender les spécificités du territoire.

La correction des données modélisées a été projetée sur un territoire de plus de 1000 km<sup>2</sup>. Bien que la dynamique spatio-temporelle de l'O<sub>3</sub> soit spécifique à notre territoire d'étude, toutes les méthodes utilisées (la correction, la classification des séries temporelles, le calcul des indicateurs de dépassements de seuils ...) sont reproductibles et transposables dans d'autres zones. Enfin, l'utilisation de référentiels extérieurs et standardisés (couches de bâti, de végétation, axes routiers, recensement de la population, typologies de territoires, définition des vagues de chaleur...) pour le calcul des indicateurs permet d'assurer une démarche objective dans la définition des entités que nous avons croisées. La typologie de territoire (employée pour caractériser le degré d'urbanisation en Chapitre 6) est également définie sur l'ensemble du territoire métropolitain. La répartition de la population entre territoires urbains et

ruraux sur notre territoire était de 65 % et 35 % respectivement, ce qui est tout à fait comparables aux valeurs observées à l'échelle nationale (67 % et 33 % respectivement) (D'Alessandro et al., 2021). Notre territoire présente donc une répartition de population représentative de celle observée à l'échelle nationale et une transposabilité de cette méthode à d'autres pôles urbains français présentant cette même distribution et dynamique de population.

## **5) Limites**

La disponibilité de certains types de données météorologiques et socio-économiques a conditionné l'approche méthodologique utilisée. Dans le cadre du Chapitre 3, le rayonnement solaire (intensité, durée) n'a pas été intégré aux corrections proposées. Néanmoins, ce type de variable n'était que très peu disponible et n'aurait pas rendu possible une projection spatiale de la correction.

Les bases de données spatialisées (couches de bâti, de végétation, axes routiers, recensement de la population, typologies de territoires ...) sont mises à jour de manière asynchrone. L'approche menée sur une période relativement longue rendait nos travaux dépendants de la fréquence des mises à jour réalisées par les fournisseurs de ces données.

De plus, les vagues de chaleur ont été définies à partir des valeurs relevées auprès de la station météorologique de Besançon. A l'échelle du territoire étudié, la représentativité des données de cette station doit être discutée. Les zones rurales présentent probablement un différentiel de température élevé avec le milieu urbain. Ce phénomène peut également s'intensifier avec le différentiel altitudinal. Les vagues de chaleur en milieu urbain dense seront par conséquent probablement différentes (en durée et en intensité) dans les zones rurales mais également périurbaines.

## **6) Perspectives**

Les perspectives immédiates de ces travaux sont d'analyser l'éventuelle structure spatiale 1) de l'exposition de la population, et 2) des écarts d'estimation liés à l'utilisation d'une valeur unique de quantification du polluant plutôt qu'une modélisation à échelle fine. Cette approche pourra être stratifiée sur plusieurs déterminants de la santé, en particulier l'âge et le niveau socio-économiques des habitants. Enfin, il apparaît cohérent d'élargir l'emprise spatiale du territoire étudié au périmètre de la région



Bourgogne Franche-Comté et d'évaluer la stabilité des résultats précédemment obtenus sur le territoire d'étude de ce travail de thèse. La projection spatiale et temporelle des niveaux d'exposition à l'O<sub>3</sub> en périodes de vagues de chaleur permettrait de construire des cartes d'impact sanitaire associé à cette exposition. Dans une démarche d'évaluation quantitative de l'impact sanitaire, le nombre de décès attribuables à l'augmentation des niveaux d'exposition due au changement climatique, pourrait être calculé et spatialisé afin d'identifier des zones de vulnérabilité. Une telle approche, projetée sur un large territoire, délivrerait des informations clés. Ceci pourrait constituer un outil d'aide à la décision publique, au service des acteurs régionaux et locaux pour une meilleure gestion et anticipation de l'évolution des niveaux d'exposition.

Les perspectives à court terme des présents travaux pourraient être d'analyser l'effet de la co-exposition à l'O<sub>3</sub> et aux vagues de chaleur sur l'incidence de la prématurité et du retard de croissance chez le fœtus. La relation entre la chaleur extrême et les issues défavorables de la grossesse est bien décrite mais peu d'études se sont attachées à analyser l'influence conjointe de l'exposition à l'O<sub>3</sub> et aux vagues de chaleur sur le retard de croissance du fœtus et sur la prématurité (Sun et al., 2020; Wang et al., 2020).

Les perspectives à moyen terme seraient de projeter les niveaux d'O<sub>3</sub> et les vagues de chaleur, attendus à un horizon proche (2021-2050) et lointain (2071-2100), selon différents scénarii d'émissions de gaz à effet de serre. Cette projection permettrait d'anticiper les niveaux d'exposition à l'O<sub>3</sub> et à la chaleur ainsi que les conséquences sanitaires attribuables à l'exposition à l'O<sub>3</sub> sous l'influence des vagues de chaleur.

## **Conclusion**

Les travaux de cette thèse, basés sur l'analyse à échelle fine et sur un temps long (douze années) de la pollution à l'ozone troposphérique, apporte un éclairage nouveau sur la dynamique spatiale et temporelle de ce polluant et sur l'exposition de la population, notamment en période de vague de chaleur. Une hétérogénéité spatiale marquée a été objectivée à l'échelle d'une zone restreinte pour l'ensemble des indicateurs utilisés : concentration moyenne journalière, MDA8, dépassement de seuils réglementaires et distribution de l'exposition au sein de la population. Nos résultats ont objectivé un risque systématique de sous-estimation des niveaux d'exposition de la population. Cette erreur dans l'estimation semble de plus différentielle, dans le temps, en fonction de l'intensité des vagues de chaleurs, et dans l'espace. Ces résultats renforcent la nécessité d'estimer le niveau d'exposition de la population à une échelle fine. Dans ce but, il apparaît indispensable de favoriser la construction et la validation temporelle et spatiale de modèles de prédiction de polluants atmosphériques définis sur une échelle fine.

## Références bibliographiques

- Abdulelah Al-Sudani, Z., Salih, S.Q., sharafati, A., Yaseen, Z.M., 2019. Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *Journal of Hydrology* 573, 1–12. <https://doi.org/10.1016/j.jhydrol.2019.03.004>
- Adnan, R.M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O., Li, B., 2020. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *Journal of Hydrology* 586, 124371. <https://doi.org/10.1016/j.jhydrol.2019.124371>
- Agathokleous, S., Saitanis, C.J., Savvides, C., Sicard, P., Agathokleous, E., De Marco, A., 2022a. Spatiotemporal variations of ozone exposure and its risks to vegetation and human health in Cyprus: an analysis across a gradient of altitudes. *J. For. Res.* <https://doi.org/10.1007/s11676-022-01520-2>
- Agathokleous, S., Saitanis, C.J., Savvides, C., Sicard, P., Agathokleous, E., De Marco, A., 2022b. Spatiotemporal variations of ozone exposure and its risks to vegetation and human health in Cyprus: an analysis across a gradient of altitudes. *J. For. Res.* <https://doi.org/10.1007/s11676-022-01520-2>
- Akimoto, H., Tanimoto, H., 2022. Rethinking of the adverse effects of NO<sub>x</sub>-control on the reduction of methane and tropospheric ozone – Challenges toward a denitrified society. *Atmospheric Environment* 277, 119033. <https://doi.org/10.1016/j.atmosenv.2022.119033>
- Alonso Fernández, J.R., García Nieto, P.J., Díaz Muñoz, C., Álvarez Antón, J.C., 2014. Modeling eutrophication and risk prevention in a reservoir in the Northwest of Spain by using multivariate adaptive regression splines analysis. *Ecological Engineering* 68, 80–89. <https://doi.org/10.1016/j.ecoleng.2014.03.094>
- Altshuler, S.L., Arcado, T.D., Lawson, D.R., 1995. Weekday vs. Weekend Ambient Ozone Concentrations: Discussion and Hypotheses with Focus on Northern California. *Journal of the*

- Air & Waste Management Association 45, 967–972.  
<https://doi.org/10.1080/10473289.1995.10467428>
- Álvarez Antón, J.C., García Nieto, P.J., de Cos Juez, F.J., Sánchez Lasheras, F.S., Viejo, C.B., Roqueñí Gutiérrez, N., 2013. Battery State-of-Charge Estimator Using the MARS Technique. *IEEE Transactions on Power Electronics* 28, 3798–3805.  
<https://doi.org/10.1109/TPEL.2012.2230026>
- Atkinson, R., 2000. Atmospheric chemistry of VOCs and NOV. *Atmospheric Environment* 39.
- Baghi, R., 2013. Emissions biogéniques de composés organiques volatils en région méditerranéenne - développement instrumental, mesures et modélisation (Océan, Atmosphère.). Université Paul Sabatier - Toulouse III.
- Banja, M., Papanastasiou, D.K., Melas, D., 2009. Comparing weekend and weekday ozone levels in Tirana area. Presented at the 7th International Conference on Air Quality-Science and Application.
- Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., Künzli, N., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmospheric Environment* 54, 634–642. <https://doi.org/10.1016/j.atmosenv.2012.01.064>
- Bejot, Y., Henrotin, J.B., Osseby, G.V., Zeller, M., Cottin, Y., Giroud, M., 2011. Impact of ozone air pollution on ischemic cerebral and cardiac events in Dijon, France. *Archives of Cardiovascular Diseases Supplements, XXIIes journées européennes de la société française de cardiologie* 3, 93.  
[https://doi.org/10.1016/S1878-6480\(11\)70282-8](https://doi.org/10.1016/S1878-6480(11)70282-8)
- Bell, E.J., 2013. Climate change and health research: Has it served rural communities? *Rural and Remote Health* 13, 237–252. <https://doi.org/10.3316/informit.409012370715200>
- Blein, S., 2016. Observation et modélisation de couche limite atmosphérique stable en relief complexe: le processus turbulent d'écoulement catabatique (Météorologie). Université Grenoble Alpes.
- Boehmke, B., Greenwell, B.M., 2019. *Hands-On Machine Learning with R*. Chapman and Hall/CRC.  
<https://doi.org/10.1201/9780367816377>

- Bonafè, G., 2017. RPubS - Air quality model benchmarking with dartle [WWW Document]. URL <https://rpubs.com/gbonafe/dartle-intro> (accessed 9.24.20).
- Bouthevillain, C., Mathis, A., 1995. Prévisions : mesures, erreurs et principaux résultats. *Economie et Statistique* 285, 89–100. <https://doi.org/10.3406/estat.1995.5982>
- Brancher, M., Griffiths, K.D., Franco, D., de Melo Lisboa, H., 2017. A review of odour impact criteria in selected countries around the world. *Chemosphere* 168, 1531–1570. <https://doi.org/10.1016/j.chemosphere.2016.11.160>
- Braun, S., Achermann, B., De Marco, A., Pleijel, H., Karlsson, P.E., Rihm, B., Schindler, C., Paoletti, E., 2017. Epidemiological analysis of ozone and nitrogen impacts on vegetation – Critical evaluation and recommendations. *Science of The Total Environment* 603–604, 785–792. <https://doi.org/10.1016/j.scitotenv.2017.02.225>
- Bréhon, D., Clair, P., Gueguen, C., Luneau, G., Mazurais, S., Malacarne, R., Perrussel, O., Piga, D., Racher, A., Real, E., 2018. Guide méthodologique pour l'élaboration des inventaires territoriaux des émissions atmosphériques (polluants de l'air et gaz à effet de serre). Direction générale de l'Énergie et du climat, Bureau de la qualité de l'air.
- Brodin, M., Helmig, D., Oltmans, S., 2010. Seasonal ozone behavior along an elevation gradient in the Colorado Front Range Mountains. *Atmospheric Environment* 44, 5305–5315. <https://doi.org/10.1016/j.atmosenv.2010.06.033>
- Brönnimann, S., Schuepbach, E., Zanis, P., Buchmann, B., Wanner, H., 2000. A climatology of regional background ozone at different elevations in Switzerland (1992–1998). *Atmospheric Environment, Sixth Scientific Conference of the International Global Atmospheric* 34, 5191–5198. [https://doi.org/10.1016/S1352-2310\(00\)00193-X](https://doi.org/10.1016/S1352-2310(00)00193-X)
- Calfapietra, C., Fares, S., Manes, F., Morani, A., Sgrigna, G., Loreto, F., 2013. Role of Biogenic Volatile Organic Compounds (BVOC) emitted by urban trees on ozone concentration in cities: A review. *Environmental Pollution, Selected Papers from Urban Environmental Pollution 2012* 183, 71–80. <https://doi.org/10.1016/j.envpol.2013.03.012>

- Camredon, M., Aumont, B., 2007. Modélisation chimique de l’ozone et des oxydants gazeux Chemical modelling of ozone and gaseous oxidants. *Pollution atmosphérique* 10.
- Castell-Balaguer, N., Téllez, L., Mantilla, E., 2012. Daily, seasonal and monthly variations in ozone levels recorded at the Turia river basin in Valencia (Eastern Spain). *Environ Sci Pollut Res* 19, 3461–3480. <https://doi.org/10.1007/s11356-012-0881-5>
- Chevalier, A., Gheusi, F., Delmas, R., Ordóñez, C., Sarrat, C., Zbinden, R., Thouret, V., Athier, G., Cousin, J.-M., 2007. Influence of altitude on ozone levels and variability in the lower troposphere: a ground-based study for western Europe over the period 2001–2004. *Atmospheric Chemistry and Physics* 7, 4326.
- Colette, A., Bessagnet, B., Meleux, F., Terrenoire, E., Rouïl, L., 2014. Frontiers in air quality modelling. *Geoscientific Model Development* 7, 203–210. <https://doi.org/10.5194/gmd-7-203-2014>
- Crooks, J.L., Özkaynak, H., 2014. Simultaneous statistical bias correction of multiple PM2.5 species from a regional photochemical grid model. *Atmospheric Environment* 95, 126–141. <https://doi.org/10.1016/j.atmosenv.2014.06.024>
- D’Alessandro, C., Levy, D., Regnier, T., 2021. Une nouvelle définition du rural pour mieux rendre compte des réalités des territoires et de leurs transformations – La France et ses territoires | Insee. Insee.
- Dalstein, L., Vas, N., Tagliaferro, F., Ferrara, Am., Spaziani, F., 2005. Effets de l’ozone sur la forêt et la végétation dans les Alpes franco-italiennes. *Forêt Méditerranéenne* XXVI, 149–156.
- De Cos Juez, F.J., Lasheras, F.S., García Nieto, P.J., Suárez, M.A.S., 2009. A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women. *International Journal of Computer Mathematics* 86, 1878–1887. <https://doi.org/10.1080/00207160902783557>
- de Foy, B., Brune, W.H., Schauer, J.J., 2020. Changes in ozone photochemical regime in Fresno, California from 1994 to 2018 deduced from changes in the weekend effect. *Environmental Pollution* 263, 114380. <https://doi.org/10.1016/j.envpol.2020.114380>

- De Marco, A., Garcia-Gomez, H., Collalti, A., Khaniabadi, Y.O., Feng, Z., Proietti, C., Sicard, P., Vitale, M., Anav, A., Paoletti, E., 2022. Ozone modelling and mapping for risk assessment: An overview of different approaches for human and ecosystems health. *Environmental Research* 211, 113048. <https://doi.org/10.1016/j.envres.2022.113048>
- Dessimond, B., 2021. Exposition individuelle à la pollution de l'air : mesure par capteurs miniatures, modélisation et évaluation des risques sanitaires associés. Sorbonne Université.
- Duarte, A.L., Schneider, I.L., Artaxo, P., Oliveira, M.L.S., 2022. Spatiotemporal assessment of particulate matter (PM10 and PM2.5) and ozone in a Caribbean urban coastal city. *Geoscience Frontiers* 13, 101168. <https://doi.org/10.1016/j.gsf.2021.101168>
- EEA, 2016. Air quality in Europe (Publication). European Environment Agency.
- Eghdami, H., Werner, W., Büker, P., Sicard, P., 2022. Assessment of ozone risk to Central European forests: Time series indicates perennial exceedance of ozone critical levels. *Environmental Research* 203, 111798. <https://doi.org/10.1016/j.envres.2021.111798>
- Elichegaray, C., Bouallala, S., Maitre, A., Ba, M., 2010. État et évolution de la pollution atmosphérique. *Revue Française d'Allergologie* 50, 381–393. <https://doi.org/10.1016/j.reval.2009.08.003>
- Fan, H., Zhao, C., Yang, Y., 2020. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. *Atmospheric Environment* 220, 117066. <https://doi.org/10.1016/j.atmosenv.2019.117066>
- Feng, Z., De Marco, A., Anav, A., Gualtieri, M., Sicard, P., Tian, H., Fornasier, F., Tao, F., Guo, A., Paoletti, E., 2019. Economic losses due to ozone impacts on human health, forest productivity and crop yield across China. *Environment International* 131, 104966. <https://doi.org/10.1016/j.envint.2019.104966>
- Filleul, L., Cassadou, S., Médina, S., Fabres, P., Lefranc, A., Eilstein, D., Le Tertre, A., Pascal, L., Chardon, B., Blanchard, M., Declercq, C., Jusot, J.-F., Prouvost, H., Ledrans, M., 2006. The Relation Between Temperature, Ozone, and Mortality in Nine French Cities During the Heat Wave of 2003. *Environ Health Perspect* 114, 1344–1347. <https://doi.org/10.1289/ehp.8328>

- Fitzky, A.C., Sandén, H., Karl, T., Fares, S., Calfapietra, C., Grote, R., Saunier, A., Rewald, B., 2019. The Interplay Between Ozone and Urban Vegetation—BVOC Emissions, Ozone Deposition, and Tree Ecophysiology. *Frontiers in Forests and Global Change* 2, 50. <https://doi.org/10.3389/ffgc.2019.00050>
- Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. *Ann. Statist.* 19, 1–67. <https://doi.org/10.1214/aos/1176347963>
- Friedman, J.H., Roosen, C.B., 1995. An introduction to multivariate adaptive regression splines. *Stat Methods Med Res* 4, 197–217. <https://doi.org/10.1177/096228029500400303>
- Gao, C., Xiu, A., Zhang, X., Chen, W., Liu, Y., Zhao, H., Zhang, S., 2020. Spatiotemporal characteristics of ozone pollution and policy implications in Northeast China. *Atmospheric Pollution Research* 11, 357–369. <https://doi.org/10.1016/j.apr.2019.11.008>
- García Nieto, P.J., Alonso Fernández, J.R., Sánchez Lasheras, F., de Cos Juez, F.J., Díaz Muñoz, C., 2012. A new improved study of cyanotoxins presence from experimental cyanobacteria concentrations in the Trasona reservoir (Northern Spain) using the MARS technique. *Science of The Total Environment* 430, 88–92. <https://doi.org/10.1016/j.scitotenv.2012.04.068>
- García Nieto, P.J., Sánchez Lasheras, F., de Cos Juez, F.J., Alonso Fernández, J.R., 2011. Study of cyanotoxins presence from experimental cyanobacteria concentrations using a new data mining methodology based on multivariate adaptive regression splines in Trasona reservoir (Northern Spain). *Journal of Hazardous Materials* 195, 414–421. <https://doi.org/10.1016/j.jhazmat.2011.08.061>
- García-Nieto, P.J., García-Gonzalo, E., Alonso Fernández, J.R., Díaz Muñoz, C., 2016. Using evolutionary multivariate adaptive regression splines approach to evaluate the eutrophication in the Pozón de la Dolores lake (Northern Spain). *Ecological Engineering* 94, 136–151. <https://doi.org/10.1016/j.ecoleng.2016.05.047>
- Gheusi, F., Ravetta, F., Delbarre, H., Tsamalis, C., Chevalier-Rosso, A., Leroy, C., Augustin, P., Delmas, R., Ancellet, G., Athier, G., Bouchou, P., Campistron, B., Cousin, J.-M., Fourmentin, M., Meyerfeld, Y., 2011. Pic 2005, a field campaign to investigate low-tropospheric ozone



- variability in the Pyrenees. *Atmospheric Research, International Conference on Nucleation and Atmospheric Aerosols (Part 2)* 101, 640–665. <https://doi.org/10.1016/j.atmosres.2011.04.014>
- Giao, B.C., Anh, D.T., 2016. Similarity search for numerous patterns over multiple time series streams under dynamic time warping which supports data normalization. *Vietnam J Comput Sci* 3, 181–196. <https://doi.org/10.1007/s40595-016-0062-4>
- Giorgino, T., 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software* 31, 1–24. <https://doi.org/10.18637/jss.v031.i07>
- GREC-SUD, 2018. GREC-SUD – Impact de l’ozone sur la forêt méditerranéenne [WWW Document]. GREC-SUD. URL <http://www.grec-sud.fr/article-cahier/les-articles-du-cahier-agriculture-et-foret/le-changement-climatique-une-nouvelle-contrainte-pour-la-foret-mediterraneenne/impact-de-lozone-sur-la-foret-mediterraneenne/> (accessed 7.20.22).
- Guenther, A.B., Zimmerman, P.R., Harley, P.C., Monson, R.K., Fall, R., 1993. Isoprene and monoterpene emission rate variability: model evaluations and sensitivity analyses. *Journal of Geophysical Research* 98, 12–617. <https://doi.org/10.1029/93jd00527>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Science & Business Media.
- Hendriks, C., Forsell, N., Kieseewetter, G., Schaap, M., Schöpp, W., 2016. Ozone concentrations and damage for realistic future European climate and air quality scenarios. *Atmospheric Environment* 144, 208–219. <https://doi.org/10.1016/j.atmosenv.2016.08.026>
- Hendriks, C., Kranenburg, R., Kuenen, J., van Gijlswijk, R., Wichink Kruit, R., Segers, A., Denier van der Gon, H., Schaap, M., 2013. The origin of ambient particulate matter concentrations in the Netherlands. *Atmospheric Environment* 69, 289–303. <https://doi.org/10.1016/j.atmosenv.2012.12.017>
- Hoek, G., 2017. Methods for Assessing Long-Term Exposures to Outdoor Air Pollutants. *Curr Envir Health Rpt* 4, 450–462. <https://doi.org/10.1007/s40572-017-0169-5>

- Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* 42, 7561–7578. <https://doi.org/10.1016/j.atmosenv.2008.05.057>
- Houzé, M.-L., 2006. Prévission et spatialisation des concentrations en ozone troposphérique en Bourgogne (phdthesis). Université de Bourgogne.
- Intergovernmental Panel on Climate Change (IPCC) (Ed.), 2013. *Climate change 2013: the physical science basis ; summary for policymakers, a report of Working Group I of the IPCC, technical summary, a report accepted by Working Group I of the IPCC but not approved in detail and frequently asked questions ; part of the Working Group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Intergovernmental Panel on Climate Change, New York.
- IPCC (Ed.), 2014. *Climate change 2014: impacts, adaptation, and vulnerability: Working Group II contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, United Kingdom and New York, NY.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2014. *An Introduction to Statistical Learning: with Applications in R*. Springer Publishing Company, Incorporated.
- Janssen, S., Guerreiro, C., Viaene, P., Georgieva, E., Thunis, P., 2017. *Guidance Document on Modelling Quality Objectives and Benchmarking* 58.
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. *J Expo Sci Environ Epidemiol* 15, 185–204. <https://doi.org/10.1038/sj.jea.7500388>
- Kaiser, A., Scheifinger, H., Spangl, W., Weiss, A., Gilge, S., Fricke, W., Ries, L., Cemas, D., Jesenovec, B., 2007. Transport of nitrogen oxides, carbon monoxide and ozone to the Alpine Global Atmosphere Watch stations Jungfraujoch (Switzerland), Zugspitze and Hohenpeissenberg (Germany), Sonnblick (Austria) and Mt. Kravavec (Slovenia). *Atmospheric Environment* 41, 9273–9287. <https://doi.org/10.1016/j.atmosenv.2007.09.027>

- Kalisa, E., Fadlallah, S., Amani, M., Nahayo, L., Habiyaremye, G., 2018. Temperature and air pollution relationship during heatwaves in Birmingham, UK. *Sustainable Cities and Society* 43, 111–120. <https://doi.org/10.1016/j.scs.2018.08.033>
- Kato, T., 2016. Chapter 4 - Prediction of photovoltaic power generation output and network operation, in: Funabashi, T. (Ed.), *Integration of Distributed Energy Resources in Power Systems*. Academic Press, pp. 77–108. <https://doi.org/10.1016/B978-0-12-803212-1.00004-0>
- Khomsi, K., Chelhaoui, Y., Alilou, S., Souri, R., Najmi, H., Souhaili, Z., 2022. Concurrent Heat Waves and Extreme Ozone (O<sub>3</sub>) Episodes: Combined Atmospheric Patterns and Impact on Human Health. *Int J Environ Res Public Health* 19, 2770. <https://doi.org/10.3390/ijerph19052770>
- Kisi, O., Parmar, K.S., 2016. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology* 534, 104–112. <https://doi.org/10.1016/j.jhydrol.2015.12.014>
- Kotu, V., Deshpande, B., 2019. Chapter 2 - Data Science Process, in: Kotu, V., Deshpande, B. (Eds.), *Data Science (Second Edition)*. Morgan Kaufmann, pp. 19–37. <https://doi.org/10.1016/B978-0-12-814761-0.00002-2>
- Krug, A., Fenner, D., Mücke, H.-G., Scherer, D., 2020. The contribution of air temperature and ozone to mortality rates during hot weather episodes in eight German cities during the years 2000 and 2017 (preprint). *Atmospheric, Meteorological and Climatological Hazards*. <https://doi.org/10.5194/nhess-2020-91>
- Kushta, J., Georgiou, G.K., Proestos, Y., Christoudias, T., Thunis, P., Savvides, C., Papadopoulos, C., Lelieveld, J., 2019. Evaluation of EU air quality standards through modeling and the FAIRMODE benchmarking methodology. *Air Qual Atmos Health* 12, 73–86. <https://doi.org/10.1007/s11869-018-0631-z>
- Lefohn, A.S., Wernli, H., Shadwick, D., Oltmans, S.J., Shapiro, M., 2012. Quantifying the importance of stratospheric-tropospheric transport on surface ozone concentrations at high- and low-elevation monitoring sites in the United States. *Atmospheric Environment* 62, 646–656. <https://doi.org/10.1016/j.atmosenv.2012.09.004>

- Lesmeister, C., 2019. *Mastering Machine Learning with R* Ed. 3. Packt Publishing.
- Lin, C.-Y.C., Jacob, D.J., Fiore, A.M., 2001. Trends in exceedances of the ozone air quality standard in the continental United States, 1980–1998. *Atmospheric Environment* 35, 3217–3228. [https://doi.org/10.1016/S1352-2310\(01\)00152-2](https://doi.org/10.1016/S1352-2310(01)00152-2)
- Liu, H., Liu, J., Liu, Y., Yi, K., Yang, H., Xiang, S., Ma, J., Tao, S., 2021. Spatiotemporal variability and driving factors of ground-level summertime ozone pollution over eastern China. *Atmospheric Environment* 265, 118686. <https://doi.org/10.1016/j.atmosenv.2021.118686>
- Liu, S.C., Trainer, M., Carroll, M.A., Hübler, G., Montzka, D.D., Norton, R.B., Ridley, B.A., Walega, J.G., Atlas, E.L., Heikes, B.G., Huebert, B.J., Warren, W., 1992. A study of the photochemistry and ozone budget during the Mauna Loa Observatory Photochemistry Experiment. *J. Geophys. Res.* 97, 10463. <https://doi.org/10.1029/91JD02298>
- Lu, X., Zhang, L., Shen, L., 2019. Meteorology and Climate Influences on Tropospheric Ozone: a Review of Natural Sources, Chemistry, and Transport Patterns. *Curr Pollution Rep* 5, 238–260. <https://doi.org/10.1007/s40726-019-00118-3>
- Ma, M., Yao, G., Guo, J., Bai, K., 2021. Distinct spatiotemporal variation patterns of surface ozone in China due to diverse influential factors. *Journal of Environmental Management* 288, 112368. <https://doi.org/10.1016/j.jenvman.2021.112368>
- Ma, R., Ban, J., Wang, Q., Li, T., 2020. Statistical spatial-temporal modeling of ambient ozone exposure for environmental epidemiology studies: A review. *Science of The Total Environment* 701, 134463. <https://doi.org/10.1016/j.scitotenv.2019.134463>
- Madrigano, J., Jack, D., Anderson, G.B., Bell, M.L., Kinney, P.L., 2015. Temperature, ozone, and mortality in urban and non-urban counties in the northeastern United States. *Environmental Health* 14, 3. <https://doi.org/10.1186/1476-069X-14-3>
- Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G., Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., Meleux, F., 2017. CHIMERE-2017: from urban to hemispheric chemistry-transport modeling. *Geosci. Model Dev.* 10, 2397–2423. <https://doi.org/10.5194/gmd-10-2397-2017>

- Martin, E.P., 2011. Comparative Performance of Different Statistical Models for Predicting Ground-Level Ozone (O<sub>3</sub>) and Fine Particulate Matter (PM<sub>2.5</sub>) Concentrations in Montréal, Canada (masters). Concordia University.
- Martin, N., 2008. La pollution par l’ozone et la climatologie dans un espace méditerranéen : les Alpes-Maritimes.
- Marx, T., Bernard, N., Parmentier, A.-L., Puyraveau, M., Martin, B., Gantelet, M., Pretalli, J.-B., Dalphin, J.-C., Mauny, F., Desmettre, T., 2019. Does air pollution really impact the onset of spontaneous pneumothorax? A French case-crossover study. *Environment International* 127, 317–323. <https://doi.org/10.1016/j.envint.2019.01.056>
- McKeen, S.A., Hsie, E.-Y., Trainer, M., Tallamraju, R., Liu, S.C., 1991. A regional model study of the ozone budget in the eastern United States. *Journal of Geophysical Research: Atmospheres* 96, 10809–10845. <https://doi.org/10.1029/91JD00052>
- Meehl, G.A., Tebaldi, C., Tilmes, S., Lamarque, J.-F., Bates, S., Pendergrass, A., Lombardozzi, D., 2018. Future heat waves and surface ozone. *Environ. Res. Lett.* 13, 064004. <https://doi.org/10.1088/1748-9326/aabdc>
- Mehdizadeh, S., Behmanesh, J., Khalili, K., 2018. Comprehensive modeling of monthly mean soil temperature using multivariate adaptive regression splines and support vector machine. *Theor Appl Climatol* 133, 911–924. <https://doi.org/10.1007/s00704-017-2227-1>
- Meng, X., Wang, W., Shi, S., Zhu, S., Wang, P., Chen, R., Xiao, Q., Xue, T., Geng, G., Zhang, Q., Kan, H., Zhang, H., 2022. Evaluating the spatiotemporal ozone characteristics with high-resolution predictions in mainland China, 2013–2019. *Environmental Pollution* 299, 118865. <https://doi.org/10.1016/j.envpol.2022.118865>
- Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., Vivanco, M.G., 2013. CHIMERE 2013: a model for regional atmospheric composition modelling. *Geoscientific Model Development* 6, 981–1028. <https://doi.org/10.5194/gmd-6-981-2013>

- Milborrow, S., 2019. Notes on the earth package 68. <http://www.milbo.org/doc/earth-notes.pdf>
- Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., Williams, M.L., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmos. Chem. Phys.* 15, 8889–8973. <https://doi.org/10.5194/acp-15-8889-2015>
- Montero, P., Vilar, J.A., 2015. TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software* 62, 1–43. <https://doi.org/10.18637/jss.v062.i01>
- Mues, A., Kuenen, J., Hendriks, C., Manders, A., Segers, A., Scholz, Y., Hueglin, C., Builtjes, P., Schaap, M., 2014. Sensitivity of air pollution simulations with LOTOS-EUROS to the temporal distribution of anthropogenic emissions. *Atmospheric Chemistry and Physics* 14, 939–955. <https://doi.org/10.5194/acp-14-939-2014>
- Myhre, G., Shindell, D., Pongratz, J., 2014. Anthropogenic and Natural Radiative Forcing, in: Stocker, T. (Ed.), . Ludwig-Maximilians-Universität München, Cambridge, pp. 659–740. <https://doi.org/10.1017/CBO9781107415324.018>
- Nguyen, D.-H., Lin, C., Vu, C.-T., Cheruiyot, N.K., Nguyen, M.K., Le, T.H., Lukkhasorn, W., Vo, T.-D.-H., Bui, X.-T., 2022. Tropospheric ozone and NO<sub>x</sub>: A review of worldwide variation and meteorological influences. *Environmental Technology & Innovation* 28, 102809. <https://doi.org/10.1016/j.eti.2022.102809>
- Notario, A., Bravo, I., Adame, J.A., Díaz-de-Mera, Y., Aranda, A., Rodríguez, A., Rodríguez, D., 2013. Variability of oxidants (OX=O<sub>3</sub>+NO<sub>2</sub>), and preliminary study on ambient levels of ultrafine particles and VOCs, in an important ecological area in Spain. *Atmospheric Research* 128, 35–45. <https://doi.org/10.1016/j.atmosres.2013.03.008>
- Nuvolone, D., Petri, D., Voller, F., 2018. The effects of ozone on human health. *Environ Sci Pollut Res* 25, 8074–8088. <https://doi.org/10.1007/s11356-017-9239-3>

- Ordóñez, C.G., 2006. Trend analysis of ozone and evaluation of nitrogen dioxide satellite data in the troposphere over Europe (Doctoral Thesis). ETH Zurich. <https://doi.org/10.3929/ethz-a-005168151>
- Ouzeau, G., Soubeyroux, J.-M., Schneider, M., Vautard, R., Planton, S., 2016. Heat waves analysis over France in present and future climate: Application of a new method on the EURO-CORDEX ensemble. *Climate Services* 4, 1–12. <https://doi.org/10.1016/j.cliser.2016.09.002>
- Paoletti, E., 2009. Ozone and urban forests in Italy. *Environmental Pollution, Special Issue Section: Ozone and Mediterranean Ecology: Plants, People, Problems* 157, 1506–1512. <https://doi.org/10.1016/j.envpol.2008.09.019>
- Pascal, M., Wagner, V., Corso, M., Laaidi, K., Le Tertre, A., 2019. Évolutions de l'exposition aux canicules et de la mortalité associée en France métropolitaine entre 1970 et 2013. *Santé publique France*.
- Pay, M.T., Martínez, F., Guevara, M., Baldasano, J.M., 2014. Air quality forecasts on a kilometer-scale grid over complex Spanish terrains. *Geoscientific Model Development* 7, 1979–1999. <https://doi.org/10.5194/gmd-7-1979-2014>
- Pay, M.T., Piot, M., Jorba, O., Gassó, S., Gonçalves, M., Basart, S., Dabdub, D., Jiménez-Guerrero, P., Baldasano, J.M., 2010. A full year evaluation of the CALIOPE-EU air quality modeling system over Europe for 2004. *Atmospheric Environment* 44, 3322–3342. <https://doi.org/10.1016/j.atmosenv.2010.05.040>
- Peñuelas, J., Staudt, M., 2010. BVOCs and global change. *Trends in Plant Science* 15, 133–144. <https://doi.org/10.1016/j.tplants.2009.12.005>
- Pyrgou, A., Hadjinicolaou, P., Santamouris, M., 2018. Enhanced near-surface ozone under heatwave conditions in a Mediterranean island. *Sci Rep* 8, 9191. <https://doi.org/10.1038/s41598-018-27590-z>
- Ratanamahatana, C., Keogh, E., 2004. Everything you know about dynamic time warping is wrong. Conference paper. Third Workshop on Mining Temporal and Sequential Data.

- Ratanamahatana, C.A., Keogh, E., 2004. Making Time-series Classification More Accurate Using Learned Constraints, in: Proceedings of the 2004 SIAM International Conference on Data Mining. Presented at the Proceedings of the 2004 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, pp. 11–22.  
<https://doi.org/10.1137/1.9781611972740.2>
- Reche, C., Moreno, T., Amato, F., Pandolfi, M., Pérez, J., de la Paz, D., Diaz, E., Gómez-Moreno, F.J., Pujadas, M., Artíñano, B., Reina, F., Orio, A., Pallarés, M., Escudero, M., Tapia, O., Crespo, E., Vargas, R., Alastuey, A., Querol, X., 2018. Spatio-temporal patterns of high summer ozone events in the Madrid Basin, Central Spain. *Atmospheric Environment* 185, 207–220.  
<https://doi.org/10.1016/j.atmosenv.2018.05.002>
- Reid, C.E., Snowden, J.M., Kontgis, C., Tager, I.B., 2012. The Role of Ambient Ozone in Epidemiologic Studies of Heat-Related Mortality. *Environmental Health Perspectives* 120, 1627–1630.  
<https://doi.org/10.1289/ehp.1205251>
- Riviere, E., Bernard, J., Hulin, A., Virga, J., Dugay, F., Charles, M.-A., Cheminat, M., Cortinovia, J., Ducroz, F., Laborie, A., Malherbe, L., Piga, D., Real, E., Robic, P.-Y., Zaros, C., Seyve, E., Lepeule, J., 2019. Air pollution modeling and exposure assessment during pregnancy in the French Longitudinal Study of Children (ELFE). *Atmospheric Environment* 205, 103–114.  
<https://doi.org/10.1016/j.atmosenv.2019.02.032>
- Roelofsen, P., 2018. Time series clustering (Master thesis Business Analytics). Vrije Universiteit Amsterdam Faculty of Science; PricewaterhouseCoopers Advisory N.V. Consulting - Data Analytics.
- Royal Society (Great Britain), 2008. Ground-level ozone in the 21st century: future trends, impacts and policy implications. The Royal Society, London.
- Ryu, Y.-H., Baik, J.-J., Lee, S.-H., 2013. Effects of anthropogenic heat on ozone air quality in a megacity. *Atmospheric Environment* 80, 20–30.  
<https://doi.org/10.1016/j.atmosenv.2013.07.053>



- Sadanaga, Y., Matsumoto, J., Kajii, Y., 2003. Photochemical reactions in the urban air: Recent understandings of radical chemistry. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* 4, 85–104. [https://doi.org/10.1016/S1389-5567\(03\)00006-6](https://doi.org/10.1016/S1389-5567(03)00006-6)
- Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>
- Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J.M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M.T., Rouil, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., Wind, P., 2015. Performance of European chemistry transport models as function of horizontal resolution. *Atmospheric Environment* 112, 90–105. <https://doi.org/10.1016/j.atmosenv.2015.04.003>
- Schnell, J.L., Prather, M.J., 2017. Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America. *Proceedings of the National Academy of Sciences of the United States of America* 114, 2854–2859. <https://doi.org/10.1073/pnas.1614453114>
- Schnell, J.L., Prather, M.J., Josse, B., Naik, V., Horowitz, L.W., Cameron-Smith, P., Bergmann, D., Zeng, G., Plummer, D.A., Sudo, K., Nagashima, T., Shindell, D.T., Faluvegi, G., Strode, S.A., 2015. Use of North American and European air quality networks to evaluate global chemistry–climate modeling of surface ozone. *Atmos. Chem. Phys.* 15, 10581–10596. <https://doi.org/10.5194/acp-15-10581-2015>
- Screpanti, A., De Marco, A., 2009. Corrosion on cultural heritage buildings in Italy: A role for ozone? *Environmental Pollution* 157, 1513–1520. <https://doi.org/10.1016/j.envpol.2008.09.046>
- Seigneur, C., 2018. *Pollution atmosphérique: Concepts, théorie et applications*. Belin.
- Seinfeld, J.H., Pandis, S.N., 2016. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons.
- Sekulic, S., Kowalski, B.R., 1992. MARS: A tutorial. *J. Chemometrics* 6, 199–216. <https://doi.org/10.1002/cem.1180060405>

- Sicard, P., 2021. Ground-level ozone over time: An observation-based global overview. *Current Opinion in Environmental Science & Health* 19, 100226. <https://doi.org/10.1016/j.coesh.2020.100226>
- Sicard, P., Agathokleous, E., Araminiene, V., Carrari, E., Hoshika, Y., De Marco, A., Paoletti, E., 2018. Should we see urban trees as effective solutions to reduce increasing ozone levels in cities? *Environmental Pollution* 243, 163–176. <https://doi.org/10.1016/j.envpol.2018.08.049>
- Sicard, P., Anav, A., De Marco, A., Paoletti, E., 2017. Projected global ground-level ozone impacts on vegetation under different emission and climate scenarios. *Atmos. Chem. Phys.* 17, 12177–12196. <https://doi.org/10.5194/acp-17-12177-2017>
- Sicard, P., De Marco, A., Troussier, F., Renou, C., Vas, N., Paoletti, E., 2013. Decrease in surface ozone concentrations at Mediterranean remote sites and increase in the cities. *Atmospheric Environment* 79, 705–715. <https://doi.org/10.1016/j.atmosenv.2013.07.042>
- Sicard, P., Serra, R., Rossello, P., 2016. Spatiotemporal trends in ground-level ozone concentrations and metrics in France over the time period 1999–2012. *Environmental Research* 149, 122–144. <https://doi.org/10.1016/j.envres.2016.05.014>
- Simon, H., Baker, K.R., Phillips, S., 2012. Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012. *Atmospheric Environment* 61, 124–139. <https://doi.org/10.1016/j.atmosenv.2012.07.012>
- Simonic, V., Pozzar, M., Enaux, C., Deguen, S., Kihal-Talantikite, W., 2022. Optimization and Process Design Tools for Estimation of Weekly Exposure to Air Pollution Integrating Travel Patterns during Pregnancy. *Open Journal of Statistics* 12, 408–432. <https://doi.org/10.4236/ojs.2022.123026>
- Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J.-F., Kuhn, U., Stefani, P., Knorr, W., 2014. Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years. *Atmos. Chem. Phys.* 14, 9317–9341. <https://doi.org/10.5194/acp-14-9317-2014>

- Solberg, S., Hov, Ø., Søvde, A., Isaksen, I.S.A., Coddeville, P., De Backer, H., Forster, C., Orsolini, Y., Uhse, K., 2008. European surface ozone in the extreme summer 2003. *J. Geophys. Res.* 113, D07307. <https://doi.org/10.1029/2007JD009098>
- Soubeyrou, J.-M., Ouzeau, G., Schneider, M., Cabanes, O., Kounkou-Arnaud, R., 2016. Les vagues de chaleur en France : analyse de l'été 2015 et évolutions attendues en climat futur. *Météorologie* 8, 45. <https://doi.org/10.4267/2042/60704>
- Steinbacher, M., Henne, S., Dommen, J., Wiesen, P., Prevot, A.S.H., 2004. Nocturnal trans-alpine transport of ozone and its effects on air quality on the Swiss Plateau. *Atmospheric Environment* 38, 4539–4550. <https://doi.org/10.1016/j.atmosenv.2004.05.008>
- Steven Porter, P., Rao, S.T., Hogrefe, C., Gego, E., Mathur, R., 2015. Methods for reducing biases and errors in regional photochemical model outputs for use in emission reduction and exposure assessments. *Atmospheric Environment* 112, 178–188. <https://doi.org/10.1016/j.atmosenv.2015.04.039>
- Struzewska, J., Jefimow, M., 2016. A 15-Year Analysis of Surface Ozone Pollution in the Context of Hot Spells Episodes over Poland. *Acta Geophys.* 64, 1875–1902. <https://doi.org/10.1515/acgeo-2016-0067>
- Stull, R.B., 1988. *An Introduction to Boundary Layer Meteorology*. Springer Science & Business Media.
- Sun, Y., Ilango, S.D., Schwarz, L., Wang, Q., Chen, J.-C., Lawrence, J.M., Wu, J., Benmarhnia, T., 2020. Examining the joint effects of heatwaves, air pollution, and green space on the risk of preterm birth in California. *Environ Res Lett* 15, 104099. <https://doi.org/10.1088/1748-9326/abb8a3>
- Tarasova, O.A., Karpetchko, A.Yu., 2003. Accounting for local meteorological effects in the ozone time-series of Lovozero (Kola Peninsula). *Atmospheric Chemistry and Physics* 3, 941–949.
- Thunis, P., Cuvelier, C., 2018. Delta Version 5.6: Concepts/User's Guide/Diagrams [WWW Document]. URL <http://atmosys.eu/download/Delta.pdf> (accessed 10.2.20).

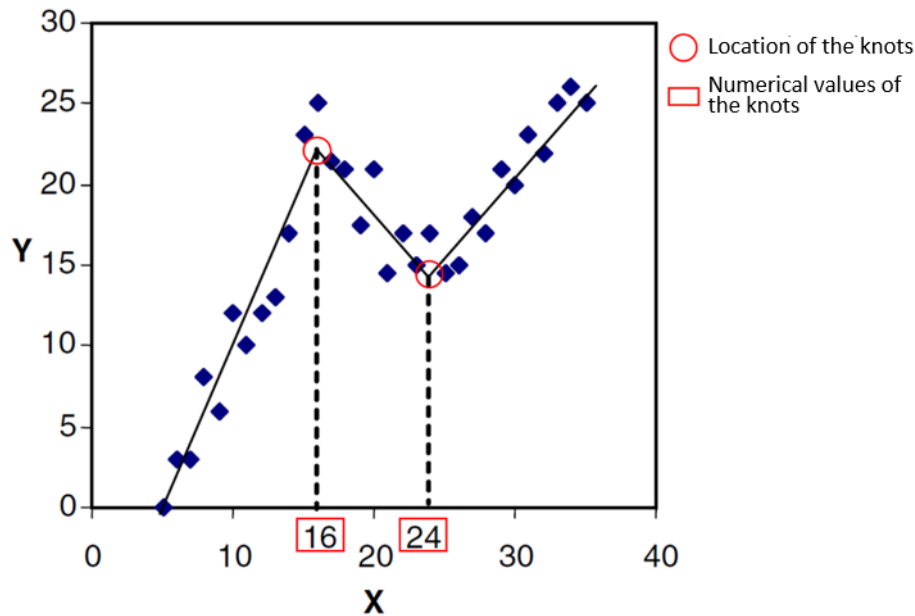
- Tong, N.Y.O., Leung, D.Y.C., Liu, C.-H., 2011. A Review on Ozone Evolution and Its Relationship with Boundary Layer Characteristics in Urban Environments. *Water Air Soil Pollut* 214, 13–36. <https://doi.org/10.1007/s11270-010-0438-5>
- Trenberth, K.E., Dai, A., van der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., Sheffield, J., 2014. Global warming and changes in drought. *Nature Clim Change* 4, 17–22. <https://doi.org/10.1038/nclimate2067>
- Ulpiani, G., 2021. On the linkage between urban heat island and urban pollution island: Three-decade literature review towards a conceptual framework. *Science of The Total Environment* 751, 141727. <https://doi.org/10.1016/j.scitotenv.2020.141727>
- Valari, M., Menut, L., Chatignoux, E., 2011. Using a Chemistry Transport Model to Account for the Spatial Variability of Exposure Concentrations in Epidemiologic Air Pollution Studies. *Journal of the Air & Waste Management Association* 61, 164–179. <https://doi.org/10.3155/1047-3289.61.2.164>
- Varotsos, K.V., Giannakopoulos, C., Tombrou, M., 2019. Ozone-temperature relationship during the 2003 and 2014 heatwaves in Europe. *Reg Environ Change* 19, 1653–1665. <https://doi.org/10.1007/s10113-019-01498-4>
- Viaene, P., Janssen, S., Thunis, P., Guerreiro, C., Cuvelier, K., Trimpeneers, E., Wesseling, J., Montero, A., Miranda, A., Stocker, J., Olesen, H.R., Santos, S., Vincent, K., Carnevale, C., Stortini, M., Bonafè, G., Minguzzi, E., Georgieva, E., Malherbe, L., Deserti, M., 2016. Guidance Document on Modelling Quality Objectives and Benchmarking 74.
- Vieno, M., Dore, A.J., Stevenson, D.S., Doherty, R., Heal, M.R., Reis, S., Hallsworth, S., Tarrason, L., Wind, P., Fowler, D., Simpson, D., Sutton, M.A., 2010. Modelling surface ozone during the 2003 heat-wave in the UK. *Atmospheric Chemistry and Physics* 10, 7963–7978. <https://doi.org/10.5194/acp-10-7963-2010>
- Wang, M., Beelen, R., Basagana, X., Becker, T., Cesaroni, G., de Hoogh, K., Dedele, A., Declercq, C., Dimakopoulou, K., Eeftens, M., Forastiere, F., Galassi, C., Gražulevičienė, R., Hoffmann, B., Heinrich, J., Iakovides, M., Künzli, N., Korek, M., Lindley, S., Mölter, A., Mosler, G., Madsen,

- C., Nieuwenhuijsen, M., Phuleria, H., Pedeli, X., Raaschou-Nielsen, O., Ranzi, A., Stephanou, E., Sugiri, D., Stempfelet, M., Tsai, M.-Y., Lanki, T., Udvardy, O., Varró, M.J., Wolf, K., Weinmayr, G., Yli-Tuomi, T., Hoek, G., Brunekreef, B., 2013. Evaluation of Land Use Regression Models for NO<sub>2</sub> and Particulate Matter in 20 European Study Areas: The ESCAPE Project. *Environ. Sci. Technol.* 47, 4357–4364. <https://doi.org/10.1021/es305129t>
- Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekreef, B., 2012. Systematic Evaluation of Land Use Regression Models for NO<sub>2</sub>. *Environ. Sci. Technol.* 46, 4481–4489. <https://doi.org/10.1021/es204183v>
- Wang, M., Sampson, P.D., Hu, J., Kleeman, M., Keller, J.P., Olives, C., Szpiro, A.A., Vedal, S., Kaufman, J.D., 2016. Combining Land-Use Regression and Chemical Transport Modeling in a Spatiotemporal Geostatistical Model for Ozone and PM<sub>2.5</sub>. *Environ Sci Technol* 50, 5111–5118. <https://doi.org/10.1021/acs.est.5b06001>
- Wang, Q., Li, B., Benmarhnia, T., Hajat, S., Ren, M., Liu, T., Knibbs, L.D., Zhang, H., Bao, J., Zhang, Y., Zhao, Q., Huang, C., 2020. Independent and Combined Effects of Heatwaves and PM<sub>2.5</sub> on Preterm Birth in Guangzhou, China: A Survival Analysis. *Environ Health Perspect* 128, 17006. <https://doi.org/10.1289/EHP5117>
- Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L., 2017. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *Science of The Total Environment* 575, 1582–1596. <https://doi.org/10.1016/j.scitotenv.2016.10.081>
- Wang, Y., Yu, C., Tao, J., Wang, Z., Si, Y., Cheng, L., Wang, H., Zhu, S., Chen, L., 2018. Spatio-Temporal Characteristics of Tropospheric Ozone and Its Precursors in Guangxi, South China. *Atmosphere* 9, 355. <https://doi.org/10.3390/atmos9090355>
- Weireld, D., 2020. Emissions de COV [WWW Document]. Agence wallonne de l'air et du climat (AwAC). URL <http://www.awac.be/index.php/de/thematiques/inventaires-d-emission/par-polluants/cov#> (accessed 1.20.20).
- Xu, Q.-S., Daszykowski, M., Walczak, B., Daeyaert, F., de Jonge, M.R., Heeres, J., Koymans, L.M.H., Lewi, P.J., Vinkers, H.M., Janssen, P.A., Massart, D.L., 2004. Multivariate adaptive regression

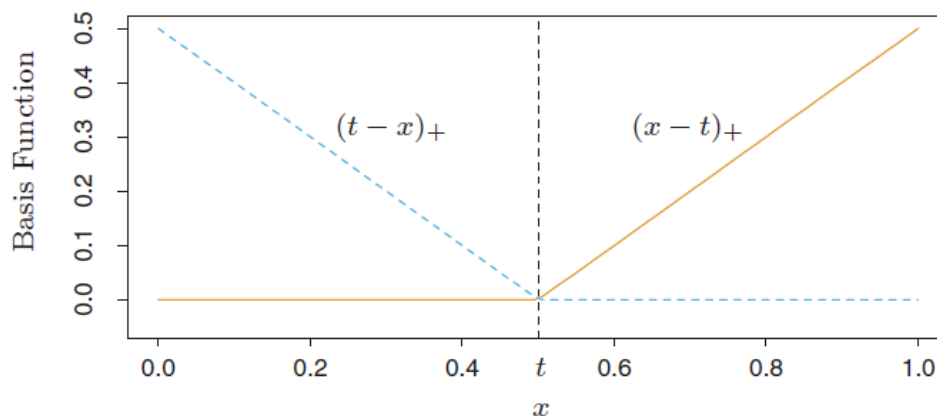
- splines—studies of HIV reverse transcriptase inhibitors. *Chemometrics and Intelligent Laboratory Systems* 72, 27–34. <https://doi.org/10.1016/j.chemolab.2004.02.007>
- Yagli, G.M., Yang, D., Srinivasan, D., 2019. Automatic hourly solar forecasting using machine learning models. *Renewable and Sustainable Energy Reviews* 105, 487–498. <https://doi.org/10.1016/j.rser.2019.02.006>
- Yilmaz, B., Aras, E., Nacar, S., Kankal, M., 2018. Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. *Science of The Total Environment* 639, 826–840. <https://doi.org/10.1016/j.scitotenv.2018.05.153>
- Yli-Pelkonen, V., Viippola, V., Rantalainen, A.-L., Zheng, J., Setälä, H., 2018. The impact of urban trees on concentrations of PAHs and other gaseous air pollutants in Yanji, northeast China. *Atmospheric Environment* 192, 151–159. <https://doi.org/10.1016/j.atmosenv.2018.08.061>
- Zhang, H., Wang, Y., Park, T.-W., Deng, Y., 2017. Quantifying the relationship between extreme air pollution events and extreme weather events. *Atmospheric Research* 188, 64–79. <https://doi.org/10.1016/j.atmosres.2016.11.010>
- Zhang, J., Wei, Y., Fang, Z., 2019. Ozone Pollution: A Major Health Hazard Worldwide. *Front. Immunol.* 10, 2518. <https://doi.org/10.3389/fimmu.2019.02518>
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012. Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment* 60, 632–655. <https://doi.org/10.1016/j.atmosenv.2012.06.031>
- Zhao, X., Zhou, W., Han, L., 2019. Human activities and urban air pollution in Chinese mega city: An insight of ozone weekend effect in Beijing. *Physics and Chemistry of the Earth, Parts A/B/C* 110, 109–116. <https://doi.org/10.1016/j.pce.2018.11.005>
- Zhou, Y., Leung, H., 2007. Predicting object-oriented software maintainability using multivariate adaptive regression splines. *Journal of Systems and Software, The Impact of Barry Boehm's Work on Software Engineering Education and Training* 80, 1349–1361. <https://doi.org/10.1016/j.jss.2006.10.049>

## Annexes

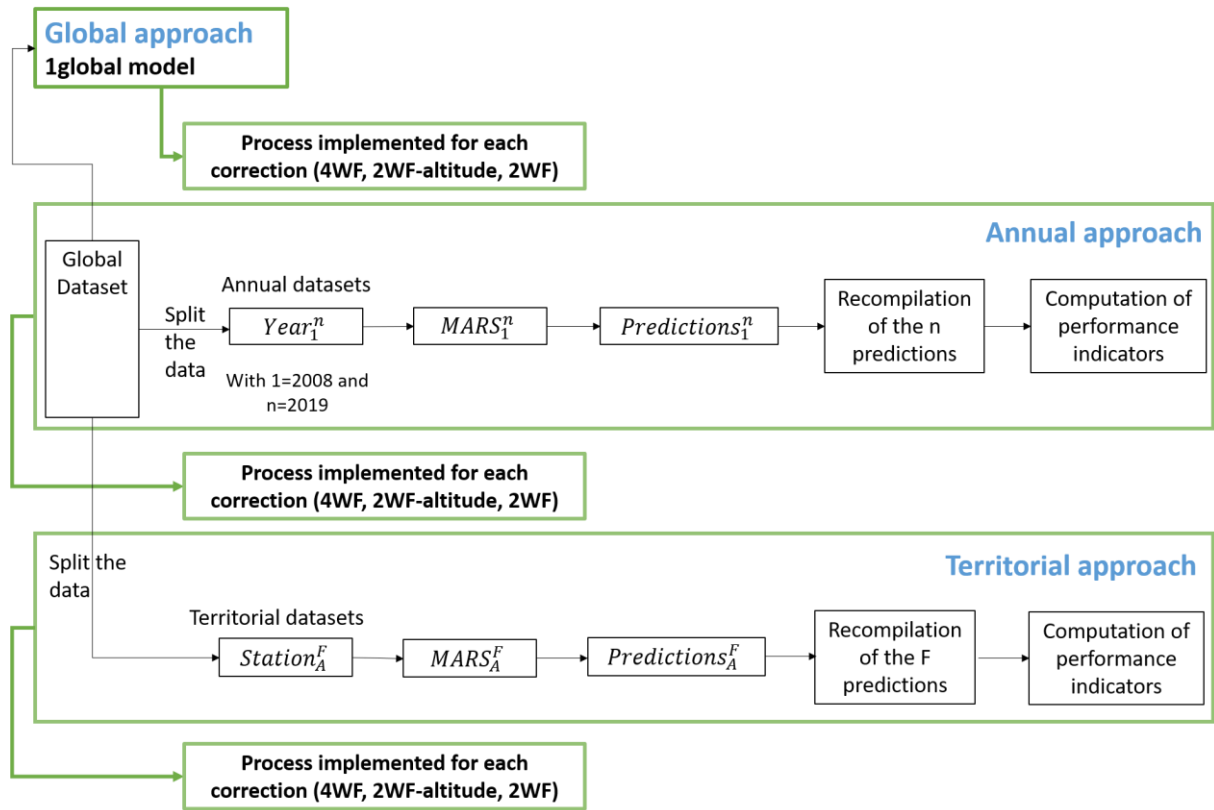
Annexe 1 : Annexes de l'article publié dans *Environmental Research* intitulé "Improvement of downscaled ozone concentrations from the transnational scale to the kilometric scale: need, interest and new insights"



**Figure A.1:** Illustration of knots location after MARS modelling of a relationship between a dependent variable (Y) and an independent variables (X), from Zhou and Leung (2007).



**Figure A.2:** Graphical representation of a pair of basis functions with  $t=0.5$ . Illustration from Hastie et al., 2009. The left spline ( $x < t$ ,  $[-(x - t)]_+^q$ ) is shown as a dashed line; and the right spline ( $x > t$ ,  $[(x - t)]_+^q$ ) as a solidline.



**Figure A.3:** Modeling process and recompilation of the predictions for the global, annual and territorial approaches.

#### Annexe 1: Description des indicateurs de performance

The performance indicators used in this study are: the Mean Bias Error (MBE), the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the correlation coefficient of Pearson ( $r$ ).

The MBE (Kato, 2016; Martin, 2011) is a bias indicator (overestimation or underestimation) described by the following equation:

$$MBE = \frac{1}{n} \sum_{i=1}^n (M_i - O_i) \quad (A.1)$$

The MAE indicating the mean order of magnitude of the forecasting errors (Abdulelah Al-Sudani et al., 2019; Adnan et al., 2020; Kisi and Parmar, 2016; Martin, 2011; Mehdizadeh et al., 2018) described by the following equation:

$$MAE = \frac{1}{n} \sum_{i=1}^n |M_i - O_i| \quad (A.2)$$

The RMSE is a precision indicator (Abdulelah Al-Sudani et al., 2019; Lesmeister, 2019; Sekulic and Kowalski, 1992; Yilmaz et al., 2018) described by the following equation:



$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - O_i)^2} \quad (A.3)$$

The correlation coefficient  $r$  is described by the following equation:

$$r = \frac{\sum_{i=1}^n (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (A.4)$$

Where  $O_i$  is the  $i^{\text{th}}$  observation,  $M_i$  the  $i^{\text{th}}$  forecast,  $\bar{O}$  is the mean of observed concentrations, and  $\bar{M}$  is the mean of modelled concentrations.

We also used synthetical benchmarking tools, contained in the ‘dartle’ R package (Bonafè, 2017). This package is directly inspired from the DELTA Tool developed by JRC (Joint Research Center) FAIRMODE (Forum for Air Modelling in Europe) (Janssen et al., 2017; Kushta et al., 2019; Viaene et al., 2016) and provides target plots combining different performance indicators: CRMSE (Centred Root Mean Square Error),  $r$  (correlation coefficient), SD (Standard Deviation),  $BIAS_{\text{Norm}}$  (normalized mean bias). The target plot’s horizontal axis gives the value of the CRMSE normalized by the measurement uncertainty of the considered pollutant, while its vertical axis displays the  $BIAS_{\text{Norm}}$  value.

The CRMSE is defined as:

$$CRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [(M_i - \bar{M}) - (O_i - \bar{O})]^2} \quad (A.5)$$

$$CRMSE_{\text{Norm}} = \frac{CRMSE}{\beta \times RMSU} \quad (A.6)$$

The  $BIAS_{\text{Norm}}$  is defined as:

$$BIAS_{\text{Norm}} = \frac{MBE}{\beta \times RMSU} \quad (A.7)$$

With  $O_i$  and  $M_i$  being respectively the  $i^{\text{th}}$  observed and modelled concentrations,  $\beta$  being a constant and  $RMSU$  being an indicator of observation uncertainty (Riviere et al., 2019).

The CRMSE reflects the ability ( $CRMSE \approx 0$ ) or inability ( $CRMSE \gg 0$ ) of the model to reproduce the inter-day oscillation of the measurements. CRMSE quantify the dispersion of the modelled data around its mean (comparatively to the dispersion of observation around its mean).

Then, if the dots are located on the left part of the target plot the CRMSE error is mainly due to a lack of correlation between modelling and measures. If the dots are located on the right part of the target plot

the CRMSE error is mainly due to a lack of variability in modelling data. And the top/bottom areas exhibit positive/negative bias (over/under-estimation) (Kushta et al., 2019; Thunis and Cuvelier, 2018).

The MQO (Modelling Quality Objective) is fulfilled if the MQI is less or equal to one. The MQI and the  $MQI_{year}$  (visible in the upper left corner of target plots) are respectively the value of the model quality indicator for both the daily (MQI) and the year average values. The dots located in the circle inside the target plot respect this quality criterion.

The target plots provide the MQI (Modelling Quality Indicator) which describes the discrepancy between modelling and measurements (Janssen et al., 2017) normalized by the measurement uncertainty:

$$MQI = \frac{|O_i - M_i|}{\beta U_{95}(O_i)}, \beta = 2 \quad (A.8)$$

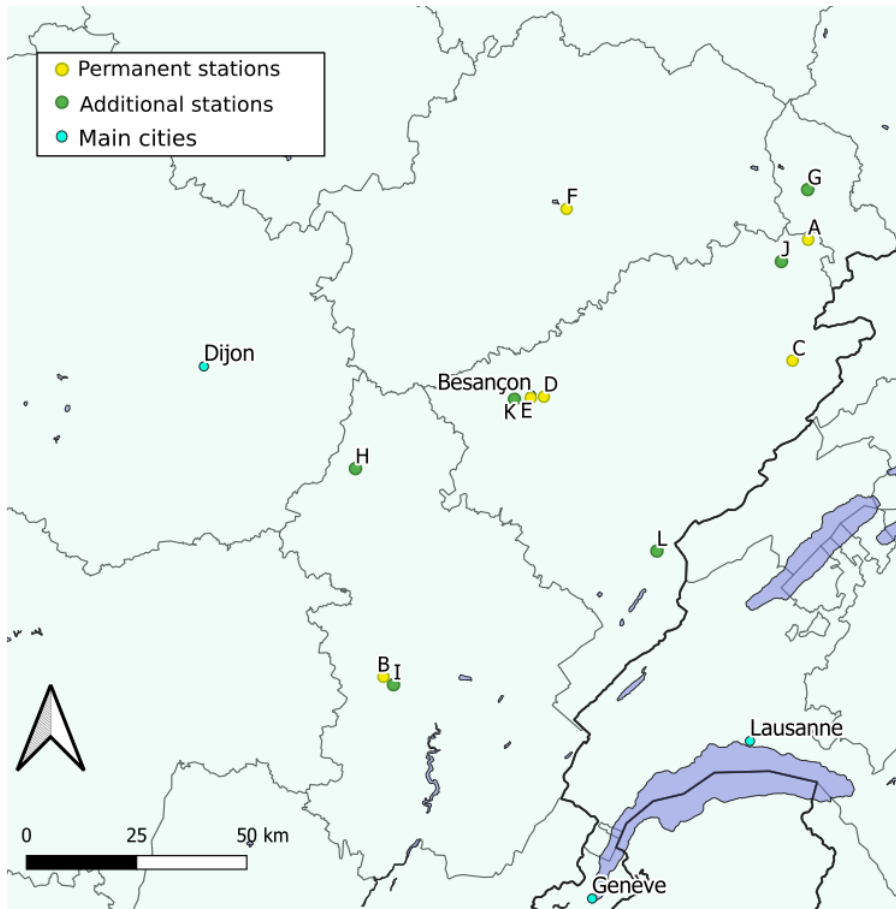
With:

$U_{95}$ : the 95<sup>th</sup> percentile measurement uncertainty of the observed concentration level

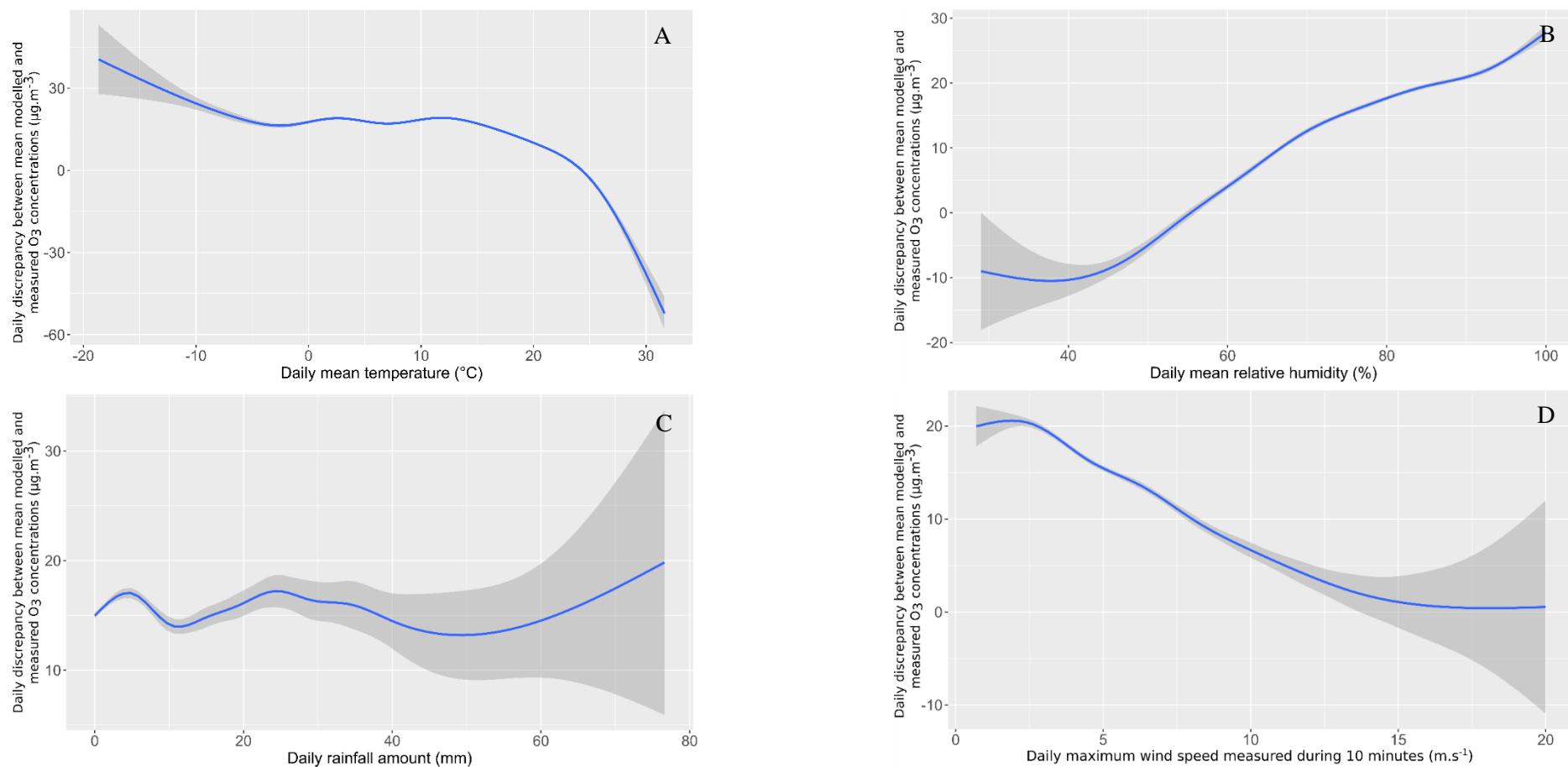
$\beta$ : a coefficient that scales the proportionality of the bias to the measurement uncertainty

$$MQI_{year} = \frac{|\bar{O} - \bar{M}|}{\beta U_{95}(\bar{O})} \quad (A.9)$$

With  $U_{95}(\bar{O})$  being the 95<sup>th</sup> percentile measurement uncertainty of the mean concentration.



**Figure A.4:** Location of the permanent and supplementary air quality monitoring stations (ATMO BFC).



**Figure A.5:** Daily discrepancy between uncorrected modelled and measured ozone concentrations according to A) daily mean temperature, B) daily relative humidity, C) daily rainfall amount and D) daily maximum wind speed measured during 10 minutes. The solid blue line represents the smoothing function values of the discrepancy (discrepancy =  $s(x, \text{bs} = \text{“cs”})$ ) using GAM. The shaded area around the solid blue line represents the confidence interval (lower and upper limits).

**Annexe 2 : Annexes de l'article publié dans la revue Environment International intitulé "Spatialized temporal dynamics of daily ozone concentrations: identification of its main driving factors and management implications"**

**Appendix A.1: Performance indicators description**

The performance indicators used in this study were based on the comparison of modeled and measured ozone concentrations. The performance indicators used in this study were: the Mean Bias Error (MBE), the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the correlation coefficient of Pearson (r).

The MBE (Kato, 2016; Martin, 2011) is a bias indicator (overestimation or underestimation) described by the following equation:

$$MBE = \frac{1}{n} \sum_{i=1}^n (M_i - O_i) \quad (A.1)$$

The MAE indicating the mean order of magnitude of the forecasting errors (Abdulelah Al-Sudani et al., 2019; Adnan et al., 2020; Kisi and Parmar, 2016; Martin, 2011; Mehdizadeh et al., 2018) described by the following equation:

$$MAE = \frac{1}{n} \sum_{i=1}^n |M_i - O_i| \quad (A.2)$$

The RMSE is a precision indicator (Abdulelah Al-Sudani et al., 2019; Lesmeister, 2019; Sekulic and Kowalski, 1992; Yilmaz et al., 2018) described by the following equation:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - O_i)^2} \quad (A.3)$$

The correlation coefficient r is described by the following equation:

$$r = \frac{\sum_{i=1}^n (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (A.4)$$

Where  $O_i$  is the  $i^{\text{th}}$  observation,  $M_i$  is the  $i^{\text{th}}$  forecast,  $\bar{O}$  is the mean of observed concentrations, and  $\bar{M}$  is the mean of modeled concentrations.

We also used synthetical benchmarking tools, contained in the 'dartle' R package (Bonafè, 2017). This package is directly inspired by the DELTA Tool developed by JRC (Joint Research Center) FAIRMODE (Forum for Air Modelling in Europe) (Janssen et al., 2017b; Kushta et al., 2019; Viaene

et al., 2016) and provides target plots combining different performance indicators: CRMSE (Centred Root Mean Square Error),  $r$  (correlation coefficient), SD (Standard Deviation),  $BIAS_{Norm}$  (normalized mean bias). The target plot's horizontal axis gives the value of the CRMSE normalized by the measurement uncertainty of the considered pollutant, while its vertical axis displays the  $BIAS_{Norm}$  value.

The CRMSE is defined as:

$$CRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [(M_i - \bar{M}) - (O_i - \bar{O})]^2} \quad (A.5)$$

$$CRMSE_{Norm} = \frac{CRMSE}{\beta \times RMSU} \quad (A.6)$$

The  $BIAS_{Norm}$  is defined as:

$$BIAS_{Norm} = \frac{MBE}{\beta \times RMSU} \quad (A.7)$$

With  $O_i$  and  $M_i$  being respectively the  $i^{th}$  observed and modeled concentrations,  $\beta$  being a constant and  $RMSU$  being an indicator of observation uncertainty (Riviere et al., 2019).

The CRMSE reflects the ability ( $CRMSE \approx 0$ ) or inability ( $CRMSE \gg 0$ ) of the model to reproduce the inter-day oscillation of the measurements. CRMSE quantifies the dispersion of the modeled data around its mean (comparatively to the dispersion of observation around its mean).

Then, if the dots are located on the left part of the target plot the CRMSE error is mainly due to a lack of correlation between modeling and measures. If the dots are located on the right part of the target plot the CRMSE error is mainly due to a lack of variability in modeling data. And the top/bottom areas exhibit positive/negative bias (over/under-estimation) (Kushta et al., 2019; Thunis and Cuvelier, 2018).

The MQO (Modelling Quality Objective) is fulfilled if the MQI is less or equal to one. The MQI and the  $MQI_{year}$  (visible in the upper left corner of target plots) are respectively the value of the model quality indicator for both the daily (MQI) and the year average values. The dots located in the circle inside the target plot respect this quality criterion.

The target plots provide the MQI (Modelling Quality Indicator) which describes the discrepancy between modeling and measurements (Janssen et al., 2017b) normalized by the measurement uncertainty:

$$MQI = \frac{|O_i - M_i|}{\beta U_{95}(O_i)}, \beta = 2 \quad (A.8)$$

With:

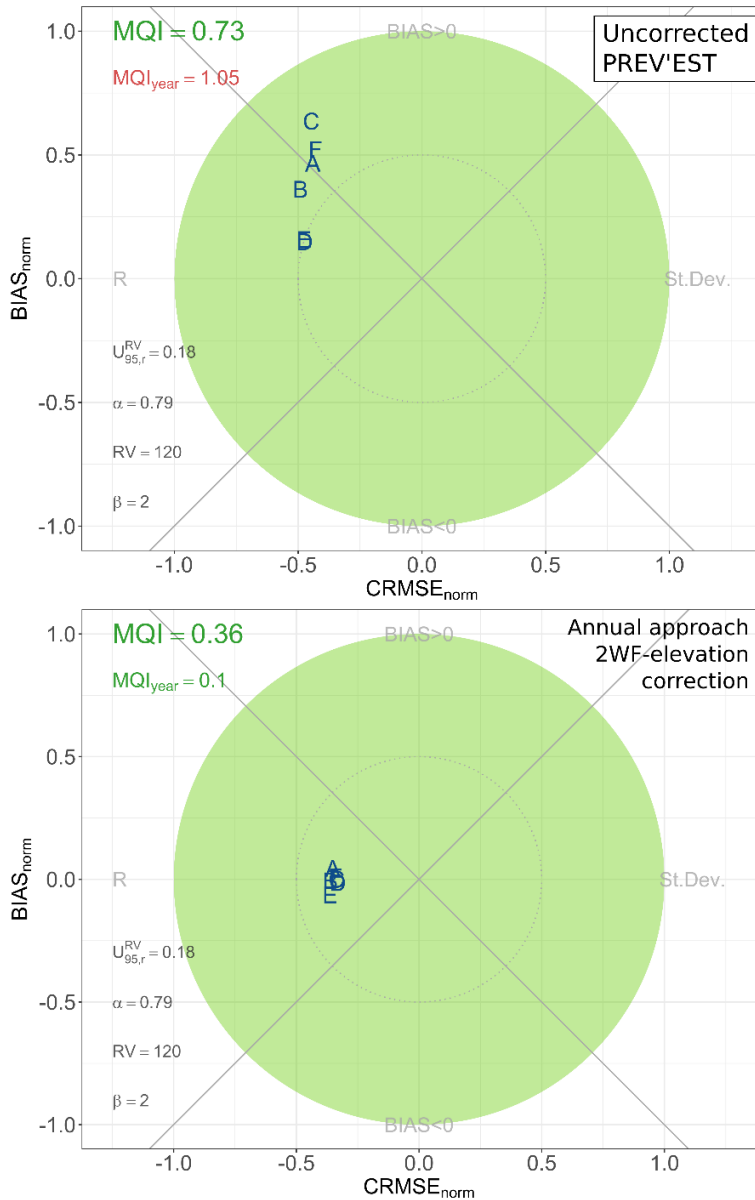
$U_{95}$ : the 95<sup>th</sup> percentile measurement uncertainty of the observed concentration level

$\beta$ : a coefficient that scales the proportionality of the bias to the measurement uncertainty

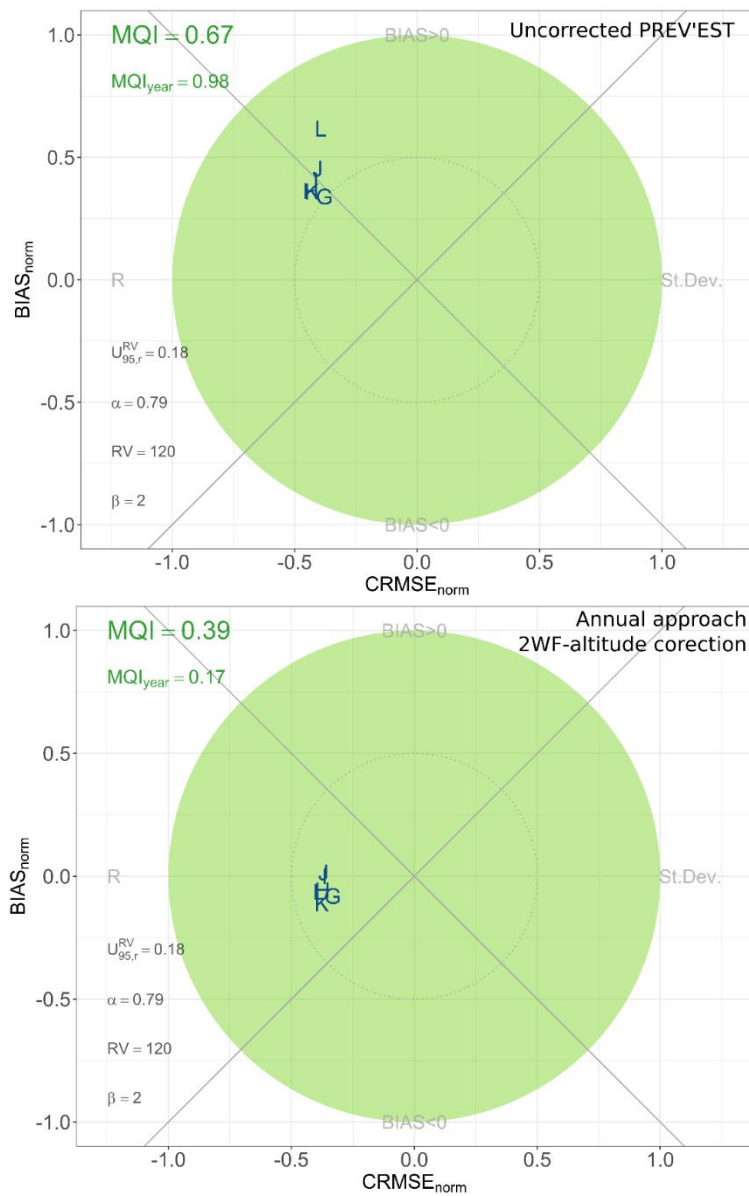
$$MQI_{year} = \frac{|\bar{O} - \bar{M}|}{\beta U_{95}(\bar{O})} \quad (A.9)$$

With  $U_{95}(\bar{O})$  being the 95th percentile measurement uncertainty of the mean concentration.

**Figure A.1:** Target plots of the uncorrected PREV'EST modeling and the annual 2WF-elevation correction. Different letters represent the different air quality monitoring stations.



**Figure A.2:** Target plots of the uncorrected PREV'EST modeling and the annual 2WF-elevation correction for the external validation. Different letters represent the different air quality monitoring stations.





**Appendix A.2:** Description of the local cost matrix (LCM) used to compute the dynamic time warping (DTW)

The LCM of dimensions  $n \times m$  was calculated using the TSclust R package (Montero and Vilar, 2015). Then, the warping path to align the two time series  $W = w_1, w_2, \dots, w_K$  is found, with  $\max(n, m) \leq K \leq m+n-1$ . This warping path has to respect three constraints (Roelofsen, 2018):

1. Boundary condition: The path must start and end in the diagonal corners of the LCM:

$$w_1 = (1, 1) \text{ and } w_K = (n, m).$$

2. Continuity: Only adjacent elements in the matrix are allowed for steps in the path. This includes diagonal adjacent elements. So, if  $w_q = (i, j)$ , then  $w_{q+1}$  is either element  $(i + 1, j)$ ,  $(i, j + 1)$  or  $(i + 1, j + 1)$  for  $q = 1, \dots, K - 1$  and  $i = 1, \dots, n - 1$  and  $j = 1, \dots, m - 1$ .

3. Monotonicity: Subsequent steps in the path must be monotonically spaced in time.

The total distance for path  $W$  is summed and minimized.

The following algorithm is used to find the lowest cost path  $W$  and obtain the summed DTW distance:

$$d_{cum}(i, j) = d(x_i, y_j) + \min \{d_{cum}(i - 1, j - 1), d_{cum}(i - 1, j), d_{cum}(i, j - 1)\} \quad (\text{A.1})$$

An additional constraint (warping window), the Sakoe-Chiba band (Sakoe and Chiba, 1978) was applied to constrain the warping path of the cost matrix (the region that can be explored by the algorithm) and to avoid pathological warping (*e.g.* a single point matching with an extremely large sequence)(Ratanamahatana and Keogh, 2004). We used a window size of 10 cells around the diagonal of the LCM (Giorgino, 2009).

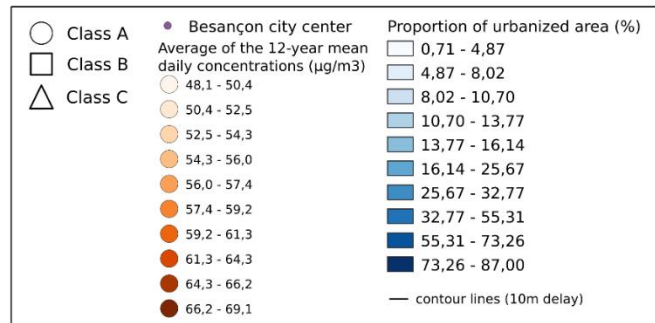
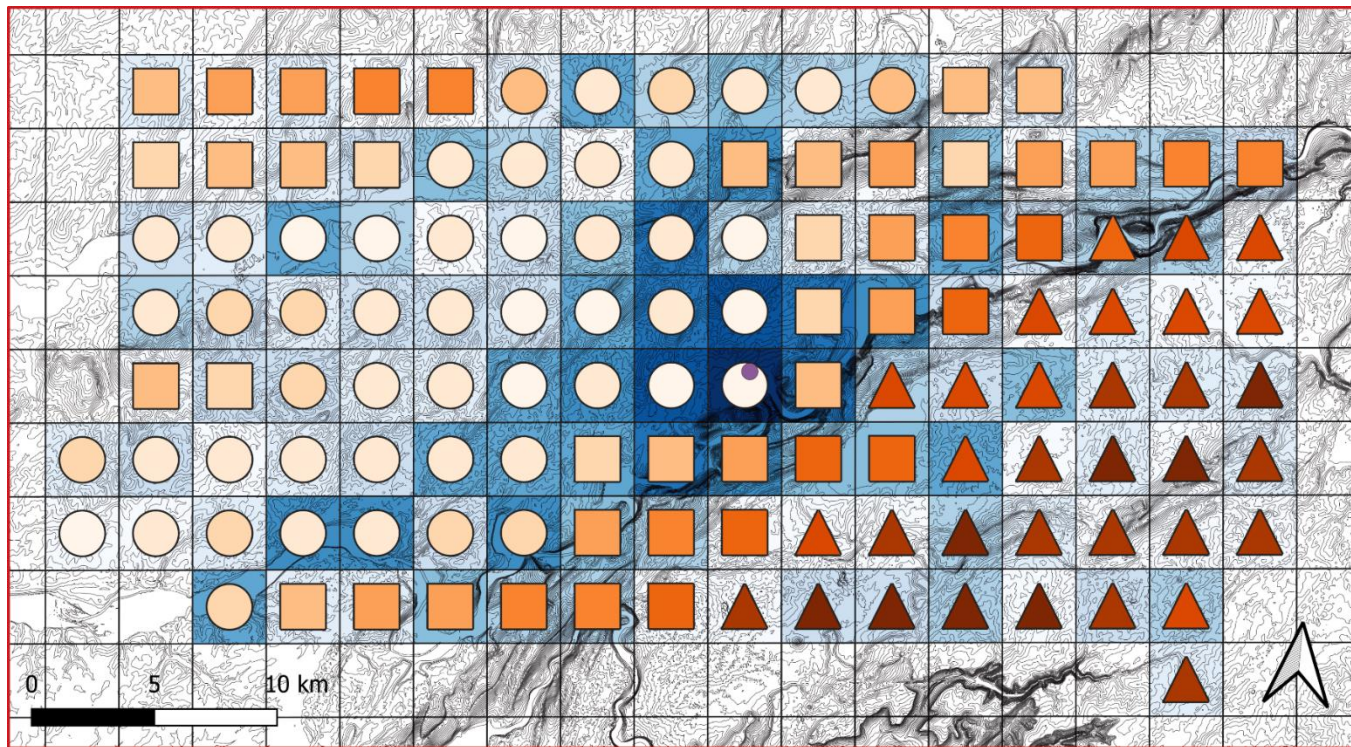


**Table A.1:** Difference between classes for the different characteristics

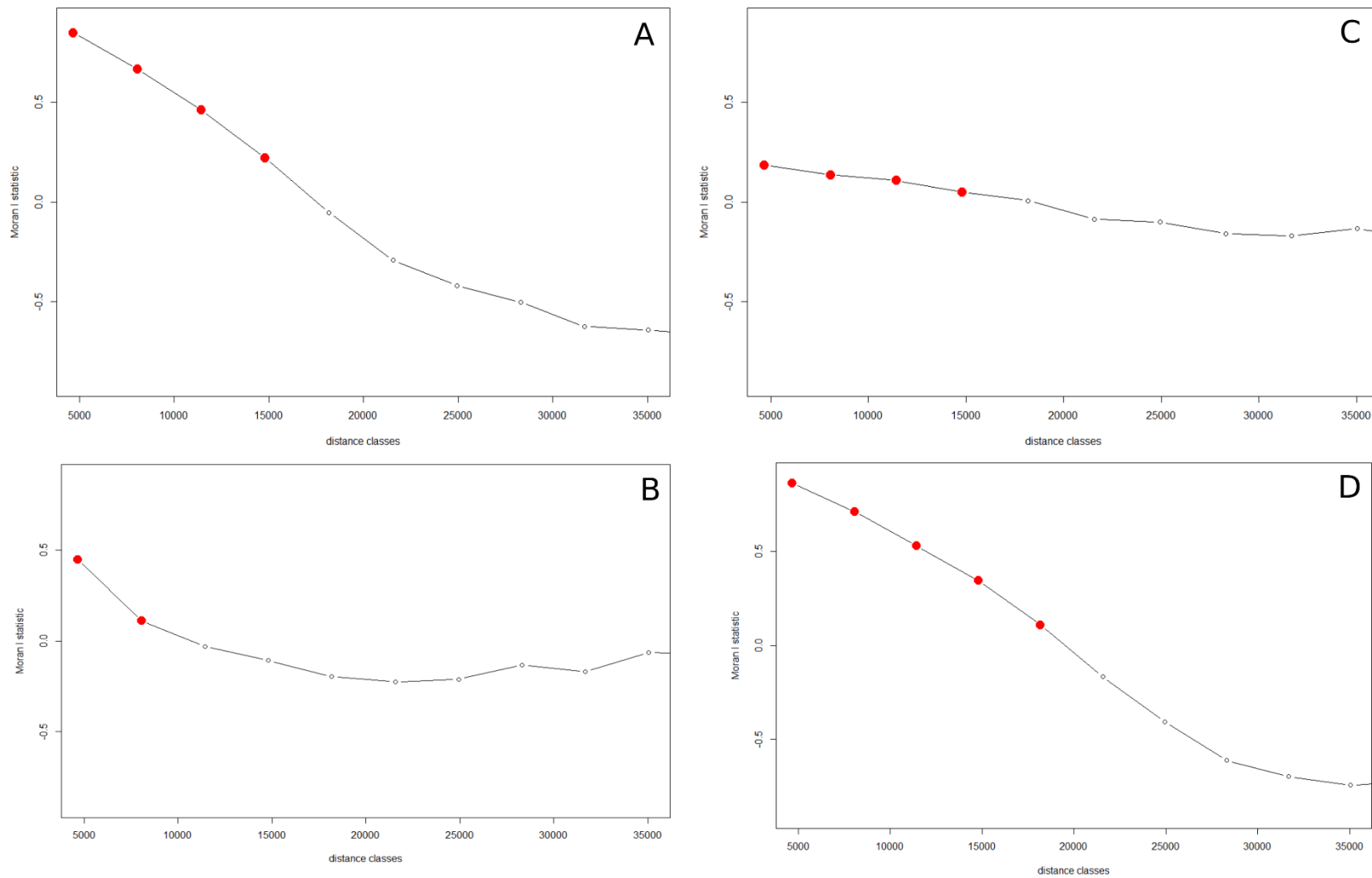
| <b>Variable (aggregated at the grid mesh scale)</b>   | <b>Difference between classes A and B</b> | <b>Difference between classes A and C</b> | <b>Difference between classes B and C</b> |
|---|---|---|---|
| Average of the 12-year mean daily concentrations  | TRUE                                      | TRUE                                      | TRUE                                      |
| 95th percentile of the 12-year mean daily concentrations  | TRUE                                      | TRUE                                      | TRUE                                      |
| Number of days with threshold (120 $\mu\text{g.m}^{-3}$ ) exceedances                                       | TRUE                                      | TRUE                                      | TRUE                                      |
| Average of the 12-year maximum daily concentrations   | TRUE                                      | TRUE                                      | FALSE                                     |
| 95 <sup>th</sup> percentile of the 12-year maximum daily concentrations                                     | TRUE                                      | TRUE                                      | TRUE                                      |
| Number of days with threshold (180 $\mu\text{g.m}^{-3}$ ) exceedances for daily maximum ozone concentration | TRUE                                      | FALSE                                     | TRUE                                      |
| Linear length of major road axis (km)   | FALSE                                     | TRUE                                      | TRUE                                      |
| Mean elevation (m)  | TRUE                                      | TRUE                                      | TRUE                                      |
| Percentage of urbanized surface (%)   | FALSE                                     | TRUE                                      | FALSE                                     |
| Cumulative index of road traffic and urban surface <sup>a</sup>   | FALSE                                     | TRUE                                      | TRUE                                      |
| Percentage of vegetated surface (%)   | TRUE                                      | TRUE                                      | FALSE                                     |

TRUE: means that the difference between two classes is significant ( $p < 0.05$ )

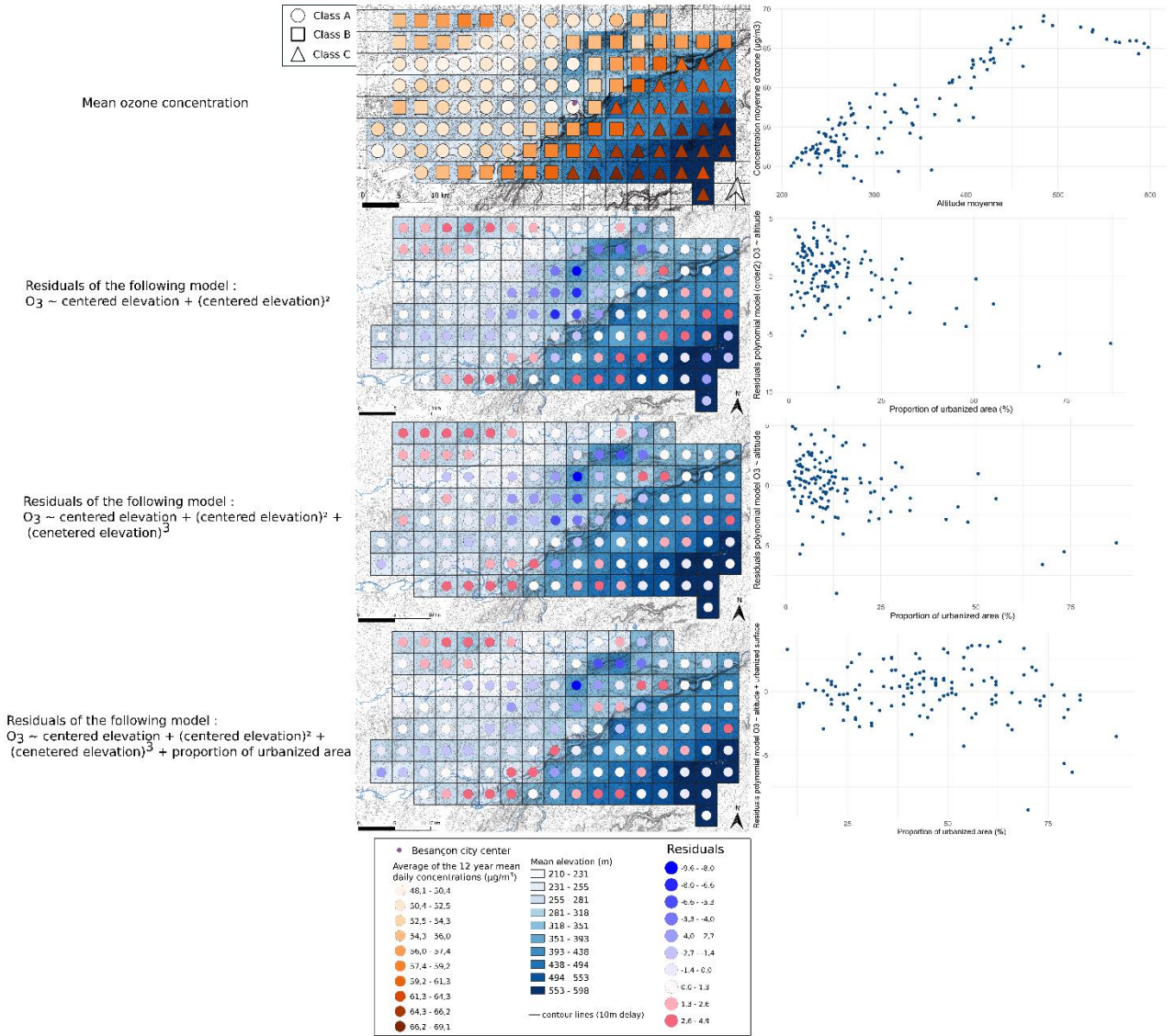
a: defined by the sum of, the length of linear major road axis (km) weighted by their importance to the road traffic, and the percentage of urbanized are



**Figure A.5:** Spatial distribution of the classes (represented by different shapes) through the modeling grid. The color filling the shapes (orange levels) represents the average of the daily mean ozone concentration (2008-2019) per modeling mesh. The background color represents the proportion of urbanized area of the modeling mesh.



**Figure A.6:** Correlograms representing Moran's coefficient according to distance classes (distance in meters) for A) mean ozone concentrations, B) percentage of urbanized area, C) percentage of vegetated surface, and D) mean elevation. The red points highlight the statistically significant values ( $p < 0.05$ ).



**Figure A.7:** Decomposition of the modelled relationship between ozone concentration and elevation, and spatialization of the residuals of the regressions.







**Titre :** Ozone troposphérique : caractérisation de sa dynamique spatiale et temporelle à échelle fine pour une meilleure évaluation de l'exposition de la population humaine dans la perspective du changement climatique

**Mots clés :** Ozone troposphérique, évaluation de l'exposition humaine, dynamique spatiale, dynamique temporelle, modélisation, vague de chaleur

**Résumé :** L'ozone troposphérique ( $O_3$ ) est un polluant atmosphérique dit secondaire dont la dynamique spatiale et temporelle est difficile à appréhender. L'exposition à ce polluant est associée à des impacts sur la santé humaine et les écosystèmes. L'évolution de la dynamique de l' $O_3$  au cours des vagues de chaleur, toujours plus fréquentes et intenses, est encore peu décrite. Ces travaux de thèse ont pour objectif d'analyser à échelle fine et sur un temps long (douze années) la dynamique spatiale et temporelle de l' $O_3$  et ses implications pour l'exposition de la population, notamment en période de vague de chaleur. Ce travail a été décomposé en quatre parties. La première a permis d'apporter une correction des concentrations d' $O_3$  modélisées et ainsi d'améliorer la précision, l'exactitude et la représentation des pics d' $O_3$  grâce à un meilleur contraste temporel. Lors de la deuxième partie, la dynamique spatiale et temporelle de l' $O_3$  et l'influence de ses déterminants (altitude, degré d'urbanisation, végétation) ont été analysées sur une zone d'étude autour de la ville de Besançon (région Bourgogne Franche-Comté – France). L'exposition de la population au regard des seuils réglementaires et recommandation relatifs à l' $O_3$  a été quantifiée ainsi que son hétérogénéité spatiale, en particulier l'écart aux valeurs fournies par la station de référence (troisième partie). Enfin, l'approche exploratoire conduite dans la dernière partie a permis de décrire les dimensions quantitatives et qualitatives des vagues de chaleur et d'explorer leur relation avec les niveaux d'exposition de la population selon le degré d'urbanisation. Ces travaux illustrent la nécessité d'estimer le niveau d'exposition de la population à une échelle fine. Il apparaît indispensable de favoriser l'utilisation de modèles de prédiction de polluants atmosphériques définis sur une échelle fine (kilométrique).

**Title:** Tropospheric ozone: characterization of its spatial and temporal dynamics on a fine scale for a better assessment of the exposure of the human population in the perspective of climate change

**Keywords:** Tropospheric ozone, human exposure assessment, spatial dynamics, temporal dynamics, modeling, heat wave

**Abstract:** Tropospheric ozone ( $O_3$ ) is a secondary air pollutant whose spatial and temporal dynamics are challenging to grasp. Exposure to this pollutant is associated with impacts on human health and ecosystems. The evolution of  $O_3$  dynamics during heat waves, which are becoming more frequent and intense, is still poorly described. The objective of this thesis work is to analyze on a fine scale and over a long period of time (twelve years) the spatial and temporal dynamics of  $O_3$  and its implications for population exposure, especially during heat waves. This work was split into four parts. The first part allowed for correcting the modeled  $O_3$  concentrations and thus improving its precision, accuracy, and representation of  $O_3$  peaks thanks to better temporal contrast. In the second part, the spatial and temporal dynamics of  $O_3$  and the influence of its determinants (elevation, degree of urbanization, vegetation) have been studied in a study area around the city of Besançon. The exposure of the population with respect to the regulatory and World Health Organization (WHO) thresholds for  $O_3$  was quantified as well as its spatial heterogeneity, in particular, the deviation from the values provided by the air quality station (third part). Finally, the exploratory approach conducted in the last part allowed us to describe the quantitative and qualitative dimensions of heat waves and to explore their relationship with the levels of population exposure according to the degree of urbanization. This work illustrates the need to assess population exposure at a fine scale. It seems essential to favor the use of prediction models of air pollutants defined on a fine scale (kilometric scale).