



LES MÉTHODES D'ÉVALUATION DE LA QUALITÉ DE L'AIR

1. Méthodes développées pour l'exploitation des concentrations mesurées

1.1. Paramètres statistiques

Cette section décrit les principaux paramètres statistiques utilisés dans le cadre des évaluations de qualité de l'air. Ces paramètres se réfèrent en grande partie aux valeurs normatives fixées dans les directives européennes.

Considérons un échantillon représentatif de valeurs mesurées d'une variable météorologique telle que la température ou la vitesse du vent. Le caractère représentatif de l'échantillon est assorti d'un grand nombre de données et permet de s'affranchir de l'influence d'éventuels événements extrêmes. Si l'on reporte les fréquences cumulées sur un diagramme, on obtient une courbe symétrique en forme de cloche (figure IV.1). On parle dans ce cas d'une distribution normale ou gaussienne. Si l'on considère une variable aléatoire x représentée par une série de n données mesurées x_i , deux paramètres statistiques suffisent à la décrire :

- la moyenne arithmétique : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- l'écart-type (exprime la dispersion des mesures) : $\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$

Dans une distribution normale, la moyenne arithmétique et la médiane sont égales. Sur base des deux paramètres indiqués ci-dessus, la densité de probabilité de la distribution normale est décrite par la fonction suivante :

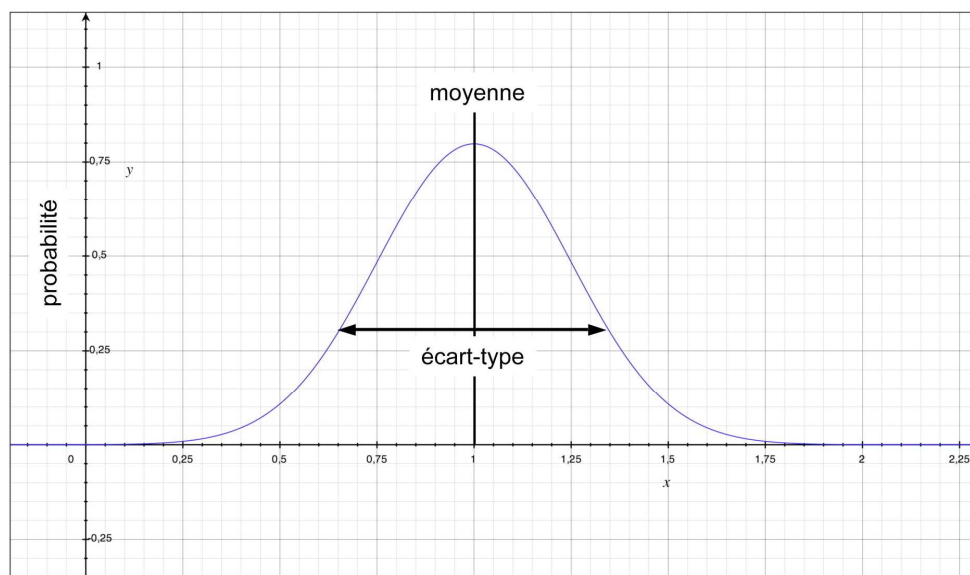
$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}$$

La surface sous une courbe de distribution normale représente la probabilité qu'un phénomène se produise dans cet intervalle, la surface totale sous la courbe valant toujours 1 (figure IV.1).



Figure IV.1 : Distribution gaussienne d'une population donnée.

Source : IBGE.



La distribution normale est particulièrement intéressante parce que la très grande majorité des phénomènes naturels tendent vers cette distribution quand on prend un grand nombre de mesures. Les mesures de concentrations de polluants constituent cependant une exception à cette règle. En effet, en traçant les fréquences cumulées de concentrations d'un polluant donné sur un diagramme, on obtient une courbe non symétrique qui s'étire du côté des grandes valeurs de concentrations, avec une médiane généralement inférieure à la moyenne arithmétique (figure IV.2, courbe a). A priori, une telle courbe ne présente pas de propriété remarquable. Mais si l'on représente plutôt le logarithme des fréquences cumulées, la courbe prend la forme d'une distribution gaussienne (figure IV.2, courbe b). La distribution des teneurs en polluants atmosphériques suit ainsi une loi log-normale. Cette dernière est caractérisée par les deux paramètres statistiques suivants :

- la moyenne géométrique ou médiane : $\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i} = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln(x_i)\right)$
- la déviation standard géométrique : $\sigma_g = \exp\left(\sqrt{\frac{(\ln(x) - \ln(\bar{x}_g))^2}{n}}\right)$

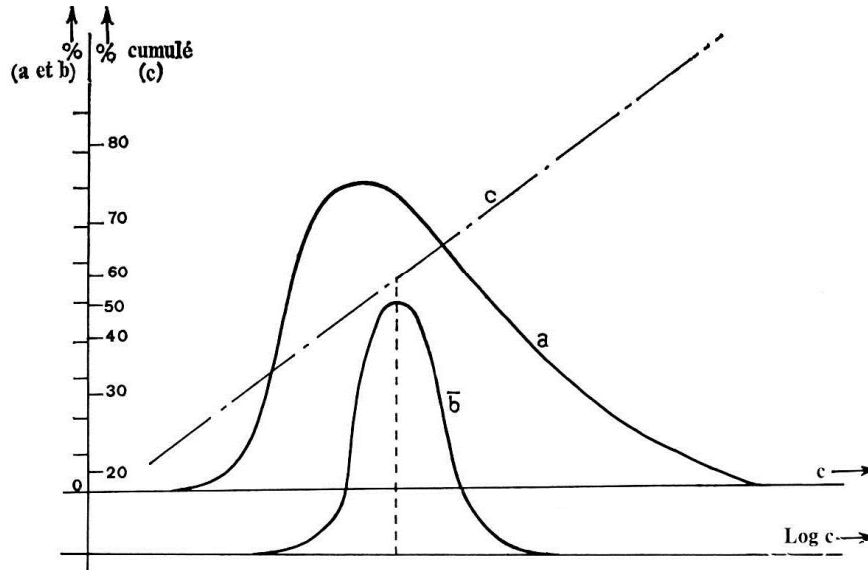
La densité de probabilité de la loi log-normale est donnée par :

$$p(x) = \frac{\exp\left(\frac{-(\ln(x) - \ln(\bar{x}_g))^2}{2\sigma_g^2}\right)}{x\sigma_g\sqrt{2\pi}}$$



Figure IV.2 : Courbes de fréquence de répartition de données de la pollution atmosphérique. a : ordonnées en %, abscisses en concentrations ; b : ordonnées en %, abscisses en logarithme des concentrations ; c : ordonnées en fréquences cumulées, abscisses en logarithme des concentrations.

Source : Chovin et Roussel, 1973.



Outre les médianes et moyennes arithmétiques, d'autres paramètres statistiques tels que les centiles sont régulièrement utilisés dans le cadre de l'évaluation de la qualité de l'air. Par définition, les centiles (percentiles en anglais) $P\alpha$ correspondent aux valeurs telles que α % des mesures leur soient inférieures (et par conséquent $(100 - \alpha)$ les dépassent). Par exemple, la première directive européenne 80/779/CE relative à la qualité de l'air fixait une valeur limite de $250 \mu\text{g}/\text{m}^3$ sur le centile P98 des concentrations journalières de SO_2 : cela signifie que le seuil journalier de $250 \mu\text{g}/\text{m}^3$ sur les concentrations de SO_2 ne pouvait être dépassé pendant plus de 7 jours par an. Concrètement, le calcul des centiles est réalisé en deux étapes :

- les n valeurs disponibles sont classées par ordre croissant : $x_1 < x_2 < \dots < x_k < \dots < x_{n-1} < x_n$;
- le centile α est la valeur du $k^{\text{ème}}$ élément où $k = \alpha \cdot n / 100$, la valeur de k étant arrondie au nombre entier le plus proche.

Les centiles P90, P95 et P98 sont les plus couramment utilisés dans les évaluations. Il faut également remarquer que la médiane correspond au centile P50.

La notion de centile était mentionnée explicitement dans les premières directives de qualité de l'air. Actuellement, elle n'est plus reprise explicitement : on parlera plutôt de valeur limite à ne pas dépasser plus d'un certain nombre de fois par an. Par exemple, la directive 2008/50/CE fixe, à partir du 1 janvier 2010, une valeur limite de $200 \mu\text{g}/\text{m}^3$ sur les concentrations horaires de NO_2 qui ne peut être dépassée plus de 18 fois par an. Ce nombre autorisé de dépassements revient à fixer une valeur limite de $200 \mu\text{g}/\text{m}^3$ sur le centile P99.8 des concentrations horaires de NO_2 .

1.2. Modèle de LARSEN

La fréquence des mesures de qualité de l'air dépend étroitement des contraintes techniques liées à l'appareillage utilisé. Les mesures semi-automatiques, qui nécessitent une analyse des prélèvements en laboratoire, n'offrent généralement pas une résolution supérieure à 1 jour.



Pourtant, les concentrations moyennes sur 24 heures ne donnent pas nécessairement une indication précise des valeurs qui peuvent être atteintes sur des échelles de temps plus courtes. C'est notamment pour résoudre cette difficulté que Larsen a établi les lois de distribution des teneurs des principaux polluants.

Le modèle Larsen (Chovin et Roussel, 1973 ; Derouane et Verduyn, 1980) a été mis au point principalement pour résoudre les problèmes suivants :

- calculer la fréquence de dépassement d'un seuil fixé sur les concentrations ;
- comparer deux séries de données caractérisées par des fréquences d'échantillonnage différentes.

Le modèle de Larsen suppose que les concentrations de polluants sont fonction de la fréquence et de la durée d'échantillonnage. Plus précisément, il repose sur les hypothèses suivantes :

- la distribution des concentrations de polluants suit une loi log-normale, quelle que soit la fréquence d'échantillonnage ;
- la médiane et la concentration maximale sont une fonction puissance de la durée d'échantillonnage ; la concentration minimale est inversement proportionnelle à une puissance de la durée d'échantillonnage ;
- la moyenne arithmétique est indépendante de la durée d'échantillonnage ;
- pour la durée d'échantillonnage la plus longue, la moyenne arithmétique, la médiane, les concentration minimale et maximale sont égales.

L'élaboration du modèle de Larsen repose sur le constat que la concentration maximale d'un polluant que l'on détermine à partir de prélèvements successifs pendant un laps de temps déterminé est d'autant plus faible que la durée d'échantillonnage, que l'on note Δt , est plus longue (figure IV.3). Inversement, les valeurs minimales croissent lorsque Δt augmente. Partant de ce constat, Larsen a établi une relation entre la valeur maximale des concentrations et la durée d'échantillonnage Δt :

$$\text{médiane}^{\Delta t} = a.(\Delta t)^b$$

$$C_{\max}^{\Delta t} = a'.(\Delta t)^{b'}$$

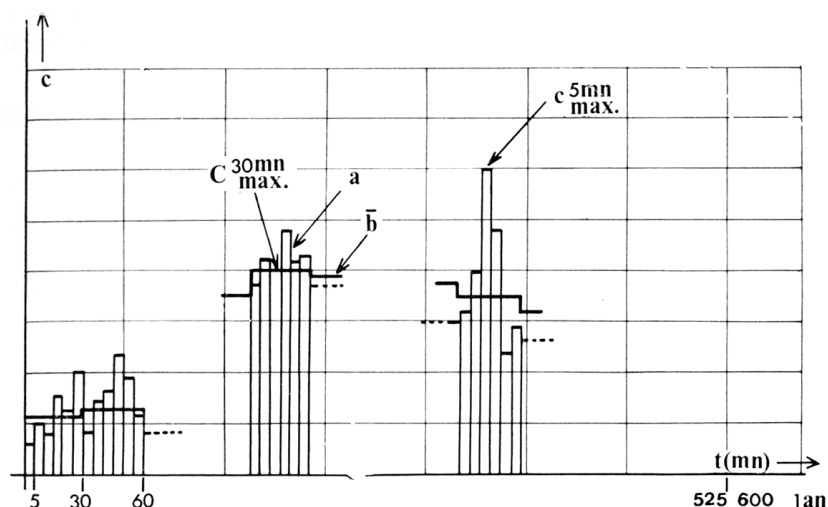
$$C_{\min}^{\Delta t} = a''.(\Delta t)^{b''}$$

Les constantes a et b sont ensuite déterminées en appliquant les relations à deux durées d'échantillonnage pour lesquelles les concentrations sont connues (figure IV.4).



Figure IV.3 : Variation de la concentration moyenne d'un polluant, a) pendant des intervalles de 5 minutes, b) pendant des intervalles de 30 minutes.

Source : Chovin et Roussel, 1973.



Les applications de ce modèle sont toutefois pas illimitées. Si une série de concentrations mesurées n'est pas distribuée selon une loi log-normale, les concentrations maximales prévues par le modèle pour des durées d'échantillonnage intermédiaires seront imprécises.

Le modèle est généralement correct pour des durées de prélèvement courtes (inférieures à quelques jours) et longues (plus d'un an). Pour des durées de prélèvement inférieures à une heure, le maximum observé est souvent surestimé par le modèle. Les écarts entre le modèle et les mesures apparaissent surtout pour les polluants secondaires, ce qui s'explique par le temps caractéristique pour certaines réactions chimiques qui peut être supérieur au temps de prélèvement. Une autre explication réside dans le temps de réponse des appareillages de mesure.

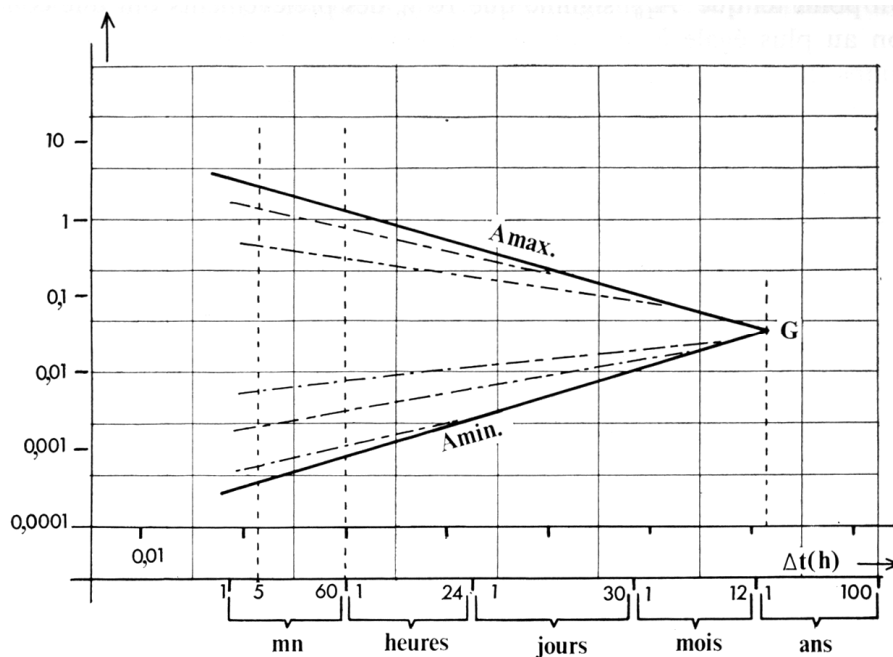
Malgré ces limitations, le modèle mathématique de Larsen présente un grand intérêt pour :

- comparer des données de pollution acquises en des endroits différents et éventuellement avec durées d'échantillonnage différentes ;
- déduire des valeurs de pointe de pollution pour des durées d'échantillonnage plus courtes que celles du prélèvement, et ceci dans le but de déterminer l'exposition de la population aux concentrations élevées ;
- estimer le pourcentage du temps pendant lequel une valeur limite a pu être dépassée.



Figure IV.4 : Variation de la concentration maximale (ou minimale) avec la durée d'échantillonnage.

Source : Chovin et Roussel, 1973.



1.3. Evaluation de l'exposition aux polluants – Interpolation RIO

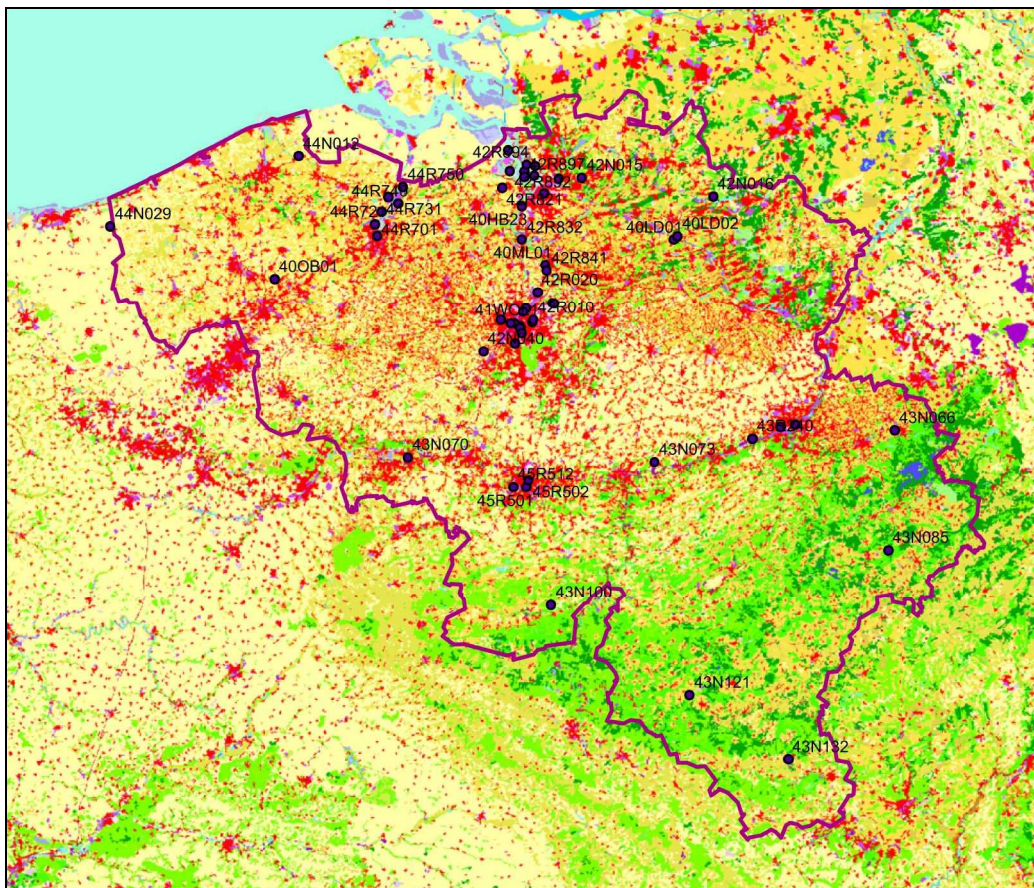
En plus de la mise à disposition au public des données de qualité de l'air mesurées en temps réel, les Etats membres doivent évaluer la qualité de l'air dans toutes les agglomérations et les zones définies sur leur territoire. Depuis la directive cadre 96/62/CE et ses 4 directives filles associées, le rapportage officiel de la qualité de l'air est réalisé dans le cadre du Questionnaire défini par la directive 2004/461/CE. Ce dernier comprend les concentrations et les dépassements de normes relevés aux stations de mesures afin de vérifier le respect des valeurs limites et valeurs guides européennes. La nouvelle directive de qualité de l'air 2008/50/CE conserve ce principe mais, en accord avec la directive INSPIRE 2007/2/CE, elle soutient l'utilisation d'outils d'évaluation de la qualité de l'air à orientation géographique. Plus précisément, elle vise à étendre et vérifier le respect des normes en tout endroit des zones définies par les Etats membres, plutôt que se limiter seulement aux postes de mesure qui ne sont pas toujours représentatifs de la zone en question. Dans cette perspective, il est essentiel de disposer d'outils permettant d'exploiter et d'interpoler au mieux les données du réseau de mesures en vue d'établir l'exposition de la population aux polluants et aux dépassements de normes.

Il existe différentes méthodes pour interpoler des données disposées de façon irrégulière. Dans le domaine de la qualité de l'air, les plus connues sont l'interpolation IDW (Inverse Distance Weighting) et le krigeage ordinaire. Quelle que soit l'approche choisie, la représentativité spatiale des données à interpoler est supposée identique en tout point du domaine considéré. En pratique, cette hypothèse est loin d'être satisfaite : les concentrations mesurées à proximité immédiate d'un émetteur de polluants (par exemple, une industrie ou encore un axe routier dense) possèdent une représentativité limitée et généralement de l'ordre de quelques centaines de mètres, alors qu'une station située en forêt sera représentative de la zone forestière en question et, par conséquent, d'une surface beaucoup plus étendue. Dans ces conditions, l'application d'une méthode d'interpolation aura pour conséquence d'exagérer la portée spatiale des mesures effectuées dans les stations à représentativité spatiale limitée.



Figure IV.5 : Base de données d'occupation de surface "Corine Land Cover" à 100 m de résolution, utilisée pour la détermination du coefficient de pollution.

Source : CELINE.



Pour résoudre cette difficulté et satisfaire l'hypothèse d'homogénéité spatiale de représentativité des données à interpoler, la méthode d'interpolation RIO (Janssen et al., 2008), développée dans le cadre d'une collaboration entre CELINE et le VITO, propose une approche originale et novatrice. Celle-ci consiste à retirer l'influence locale de sources d'émission sur les valeurs mesurées avant de procéder à l'interpolation, et de réintroduire les effets locaux après interpolation.

L'influence locale sur les concentrations de polluants est définie comme le supplément dû à la présence de sources polluantes locales et correspond à la différence que l'on mesurerait par rapport à une station située en zone rurale. La détermination de cette contribution locale repose sur l'association de chaque type d'occupation de surface à un coefficient de pollution : en fonction du polluant considéré, le coefficient de pollution permettra de caractériser et différencier les niveaux de pollution rencontrés dans des environnements aussi différents que les zones rurales, urbaines et industrielles. Concrètement, les coefficients de pollution relatifs à chaque classe d'occupation sont calculés sur base d'une analyse statistique reliant les séries de mesures avec les occupations de surface. Les classes d'occupation de sol sont obtenues à partir de la base de données Corine Land Cover CLC2000 (EEA 1995) à 100 m de résolution spatiale (figure IV.5). A partir des coefficients de pollution attribués à chaque classe d'occupation de surface, un coefficient de pollution moyen peut être estimé à l'endroit de chaque station de mesure : ce dernier est obtenu en prenant en compte toutes les occupations de surface dans un rayon de 2 km autour du point considéré (figure IV.6), de façon à tenir compte de son environnement proche.



Figure IV.6 : Environnement pris en considération autour d'une station de mesure (dans cet exemple, la station 43R240 à Engis) pour la détermination du coefficient de pollution sur base des classes d'occupation de surface de "Corine Land Cover".

Source : CELINE.

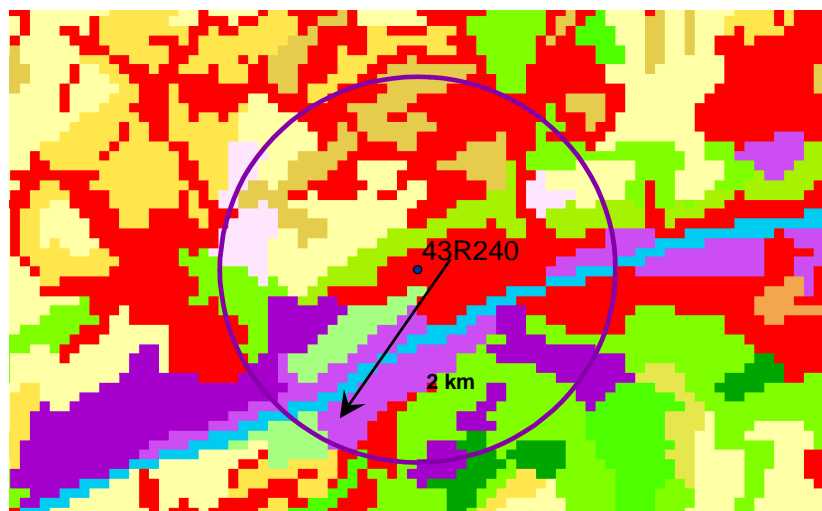
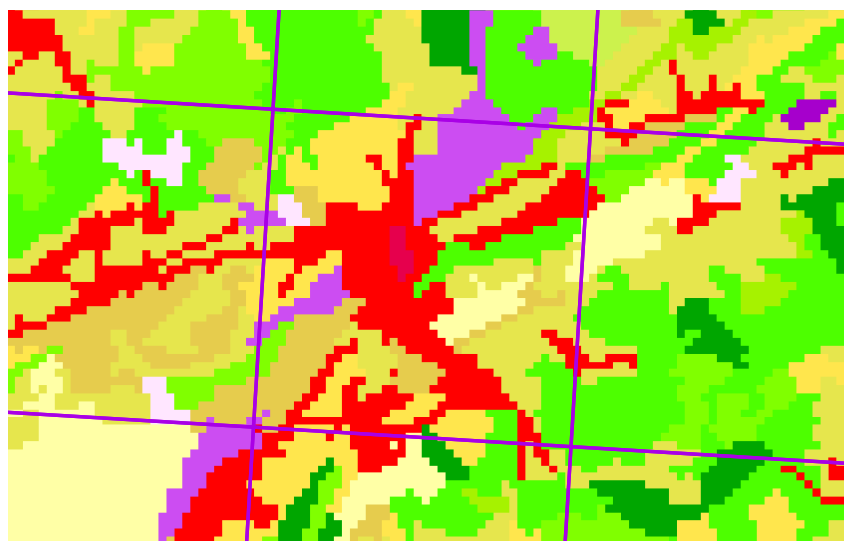


Figure IV.7 : Sur base des classes d'occupation de surface de "Corine Land Cover", détermination du coefficient de pollution dans chaque maille de la grille d'interpolation.

Source : CELINE.



La méthode RIO comporte trois étapes principales :

- Adaptation des données à interpoler en retirant les effets locaux, ce qui revient à ramener les concentrations au niveau d'une zone rurale. La correction à appliquer s'effectue à l'aide du coefficient de pollution.
- Interpolation des données adaptées. Dans le cas de l'interpolation RIO, la méthode d'interpolation choisie est le krigeage ordinaire. Sur un plan conceptuel, le krigeage peut être appliquée aux données adaptées puisque celles-ci satisfont l'hypothèse d'homogénéité spatiale de représentativité. La grille d'interpolation est habituellement choisie à une résolution de 4 km, même si des tests récents ont démontré la possibilité de descendre à des résolutions plus fines de l'ordre du kilomètre.



- Adaptation des résultats de l'interpolation pour ramener ceux-ci du niveau d'une zone rurale vers l'environnement (ville, zone industrielle, etc) qui leur correspond. Il s'agit de l'opération inverse de l'étape 1, si ce n'est qu'elle est appliquée aux points de la grille interpolée et non plus aux données de départ. Cette adaptation nécessite la détermination préalable des coefficients de pollution pour chaque maille de la grille d'interpolation (figure IV.7).

Les figures IV.8 et IV.9 présentent les cartes d'interpolation obtenues pour les moyennes annuelles de NO_2 en 2002 respectivement à l'aide des interpolations IDW et RIO. Les résultats obtenus par l'interpolation IDW reflètent d'exagération de la portée spatiale de certaines mesures effectuées dans des stations à représentativité spatiale limitée. L'interpolation RIO prend en compte la représentativité spatiale de chaque poste de mesure et produit, de ce fait, une distribution spatiale plus réaliste.

La figure IV.10 montre un exemple d'application de l'interpolation aux concentrations annuelles de SO_2 mesurées en 2008.

Figure IV.8 : Carte des concentrations annuelles de NO_2 pour l'année 2002 obtenue à l'aide de l'interpolation IDW.

Source : CELINE.

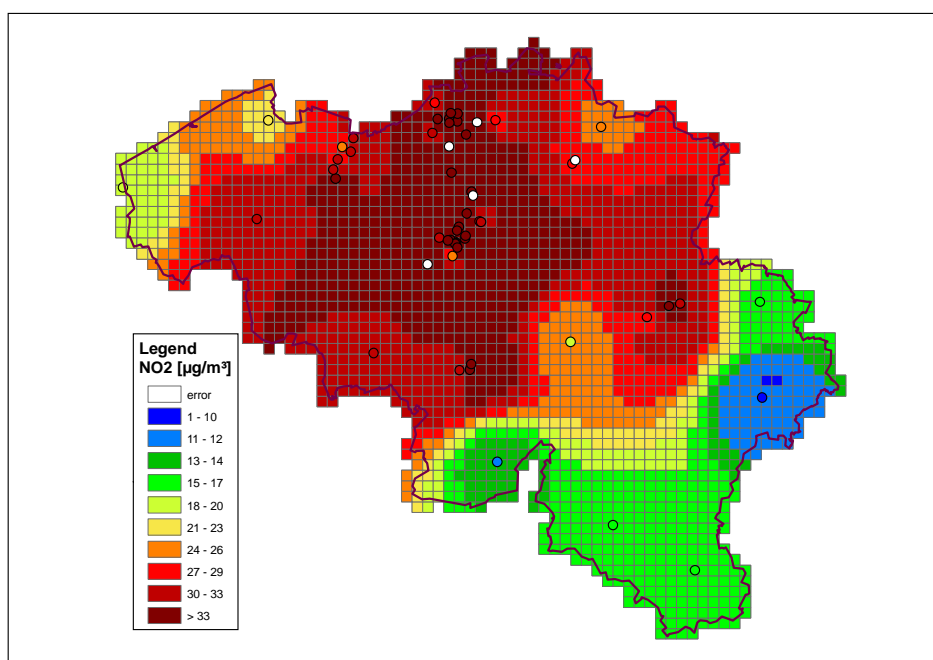




Figure IV.9 : Carte des concentrations annuelles de NO₂ pour l'année 2002 obtenue à l'aide de l'interpolation RIO.

Source : CELINE.

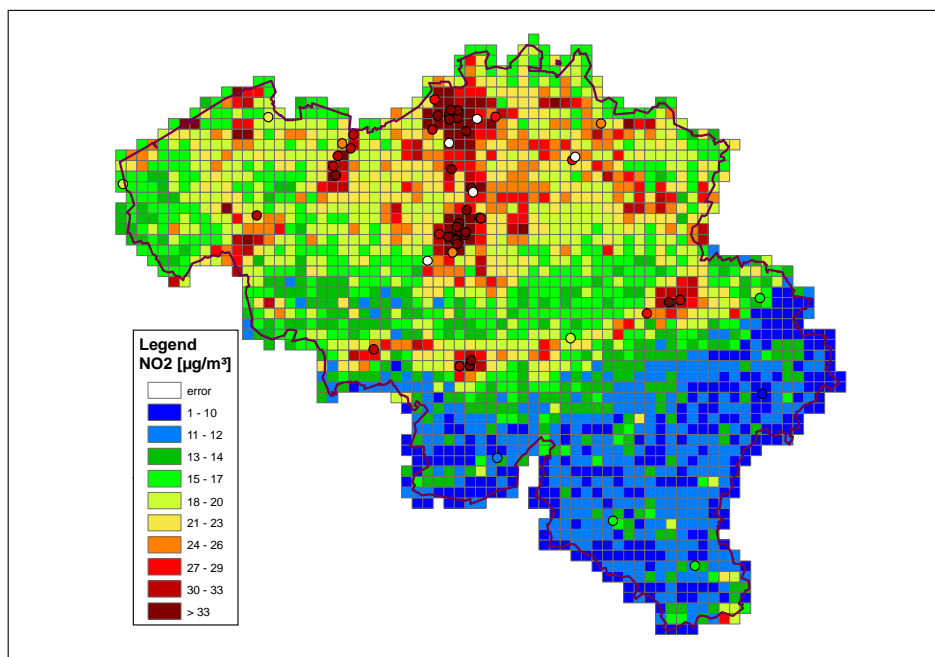
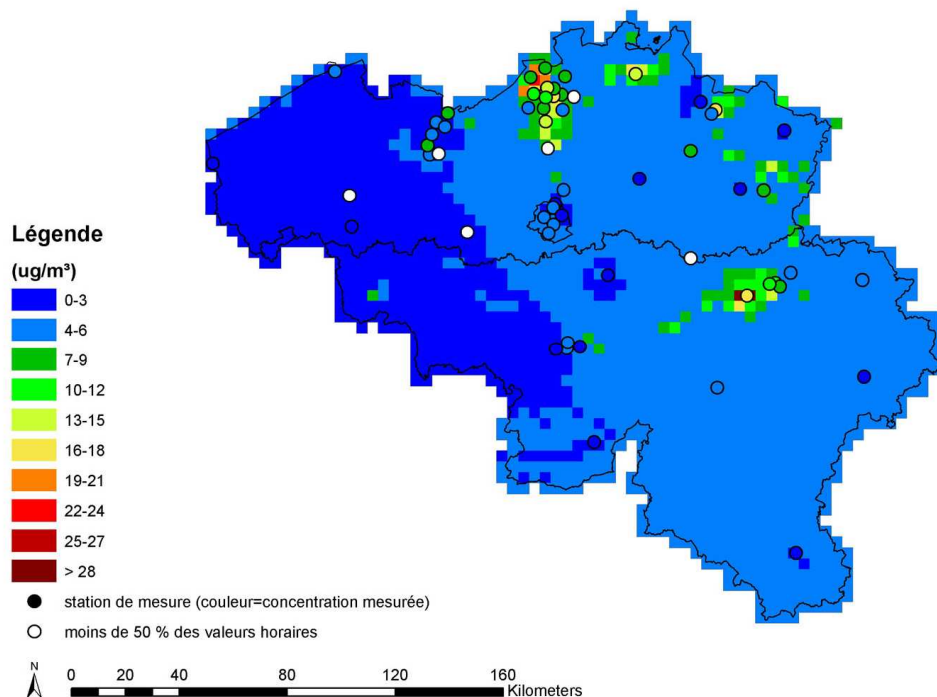


Figure IV.10 : Moyenne annuelle des concentrations de SO₂ en 2008

Source : CELINE.





2. Exploitation des variables météorologiques

Sur le court terme, l'évolution des concentrations de polluants dans l'air ambiant est régie, au premier ordre, par la capacité de la basse atmosphère à disperser les polluants. On peut notamment remarquer que la plupart des épisodes de pollution sont caractérisés par une très faible dispersion des polluants : la quasi-absence de vent et la présence d'une inversion thermique persistante à basse altitude sont les principales causes de l'accumulation des polluants dans les premières dizaines de mètres de l'atmosphère. L'analyse des variables météorologiques apparaît donc primordiale pour analyser et interpréter l'évolution de la pollution de l'air.

2.1. Procédure d'alerte basée sur la longueur de transport

A la demande de l'IBGE en 2001, l'IRM a développé une procédure de prévision des conditions météorologiques défavorables à la dispersion des polluants (Termonia et Quinet, 2004). Celle-ci s'appuie sur la notion de longueur de transport qui caractérise l'intensité de la dispersion dans la couche limite de surface. Elle est définie comme suit :

$$l = \frac{U}{\nu} \quad (1)$$

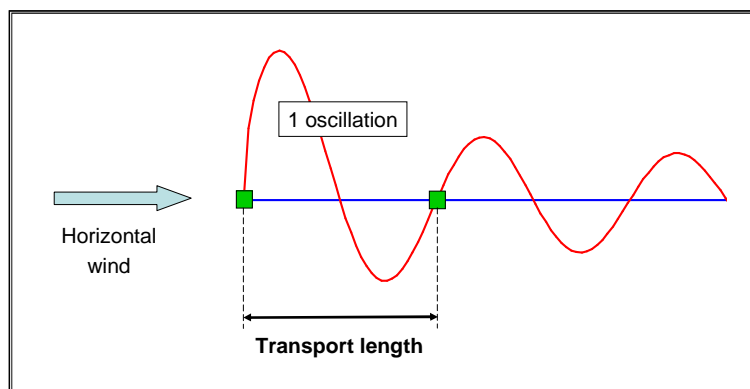
où U est la vitesse du vent et ν est la fréquence de Brunt-Väisälä :

$$\nu = \sqrt{\frac{g}{\theta} \frac{d\theta}{dz}} \quad (2)$$

Dans la relation (2), g est la gravité et θ représente la température potentielle. La longueur de transport est donc une grandeur qui représente la distance parcourue par une parcelle d'air emportée par le vent moyen au cours d'une oscillation de Brunt-Väisälä, la fréquence des oscillations étant étroitement liée à l'état de stabilité de l'atmosphère. Par conséquent, la longueur de transport prend en compte les processus mécaniques et thermiques qui sont susceptibles de générer de la turbulence. D'une certaine manière, cette grandeur caractérise les capacités dispersives de l'atmosphère.

Figure IV.11 : Interprétation physique de la notion de « longueur de transport », en fonction de la vitesse moyenne du vent et des oscillations de Brunt-Väisälä.

Source : IBGE.





De faibles valeurs de l traduisent la présence de conditions météorologiques défavorables à la dispersion des polluants. Si ces conditions persistent pendant une durée suffisamment importante, généralement de l'ordre de 12 à 20 heures, elles peuvent donner lieu à une augmentation significative des concentrations de polluants.

Une première procédure de prévision a été testée et évaluée au cours de l'hiver 2002-2003 sur la Région bruxelloise. Les prévisions de longueur de transport étaient calculées à partir des sorties du modèle ALADIN de l'IRM. L'analyse des situations rencontrées au cours de cette période ont contribué à améliorer encore les capacités de détection des situations à risque, en affinant notamment certains seuils relatifs à la persistance des conditions défavorables à la dispersion. Le critère de détection des situations défavorables à la dispersion des polluants et pouvant donner lieu à un pic de pollution se résumait comme suit :

Pendant une durée d'au moins 17 heures, la longueur de transport (l) doit être inférieure à 100 m en tout point d'une couche de 100 m d'épaisseur (à partir de la surface).

Les conditions météorologiques seront d'autant plus défavorables que la durée de l'événement sera élevée et que la couche dans laquelle $l < 100$ m sera épaisse.

Le critère de détection d'alerte à la pollution énoncé ci-dessus inclut implicitement les effets liés à la présence des conditions anticycloniques et d'une inversion thermique. En effet, l'inversion thermique, qui se développe préférentiellement en situation anticyclonique, se traduit par une importante stabilité des couches qui est prise en compte à travers la fréquence de Brunt-Väisälä (dénominateur de l). Ce critère traduit donc la faible intensité des mouvements à la fois horizontaux et verticaux.

A Bruxelles, la prévision des épisodes de pollution basée sur la détection des conditions météorologiques défavorables à la dispersion des polluants est entrée en vigueur le 1 décembre 2003. Quelques jours après l'opérationnalisation de la procédure, les prévisions du modèle ALADIN initialisé le 8 décembre à 0 h TU (figure IV.12) ont permis de mettre en évidence des conditions météorologiques très défavorables à la dispersion des polluants pour les 9 et 10 décembre. Le caractère exceptionnel de cette situation était lié à des vitesses de vent particulièrement faibles et à la présence d'une forte inversion thermique. Sur base des prévisions fournies par l'IRM et en concertation avec CELINE (Cellule Interrégionale de l'Environnement) et l'IBGE (Institut Bruxellois de la Gestion de l'Environnement), les autorités bruxelloises ont décidé de diffuser un bulletin d'alerte à la pollution auprès des médias afin de sensibiliser la population et de limiter l'utilisation des véhicules.

L'analyse à posteriori de cet épisode a permis de vérifier la pertinence de cette prévision. Comme le montre la figure IV.13, les concentrations horaires de NO_2 – principalement influencées par les émissions issues des véhicules – ont atteint des valeurs élevées. Le seuil de $200 \mu\text{g}/\text{m}^3$ en moyenne horaire (18 dépassements annuels autorisés) a été dépassé à quatre stations : Molenbeek ($251 \mu\text{g}/\text{m}^3$), Arts-Loi ($309 \mu\text{g}/\text{m}^3$), Haren ($217 \mu\text{g}/\text{m}^3$) et Parc Meudon ($224 \mu\text{g}/\text{m}^3$). Les concentrations atteintes en NO_2 à cette période étaient alors les plus importantes enregistrées en 2003 et confirment que la situation observée entre le 9 et le 10 décembre était une des plus graves sur la période 1989-2003, surtout en terme de durée de l'événement.



Figure 12: Prévisions initialisées le 8 décembre 2003 à 0 h (GMT). Ce graphique présente l'évolution temporelle de la longueur de transport (L). Il montre, à l'aide de barettes bleues, les instants auxquels le critère d'alerte est satisfait.
Source : IRM.

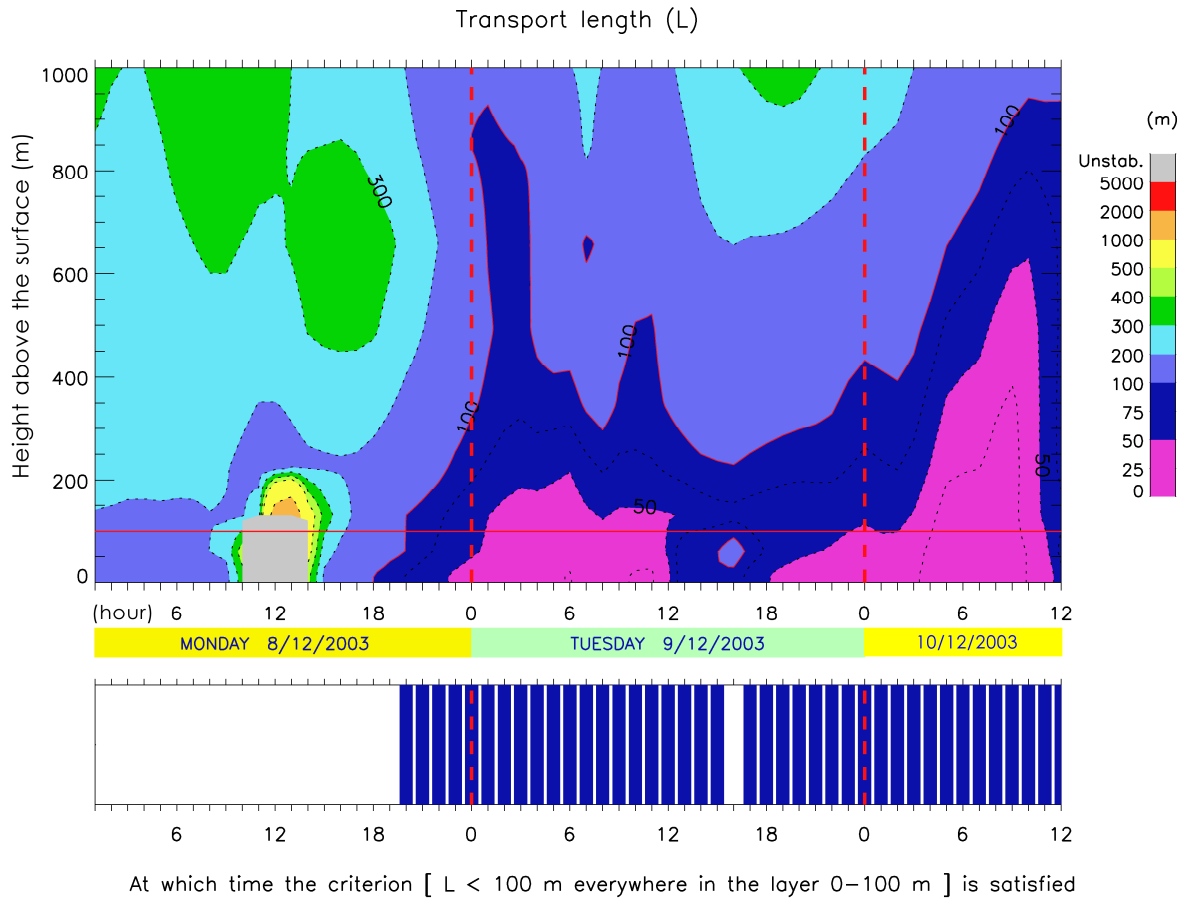
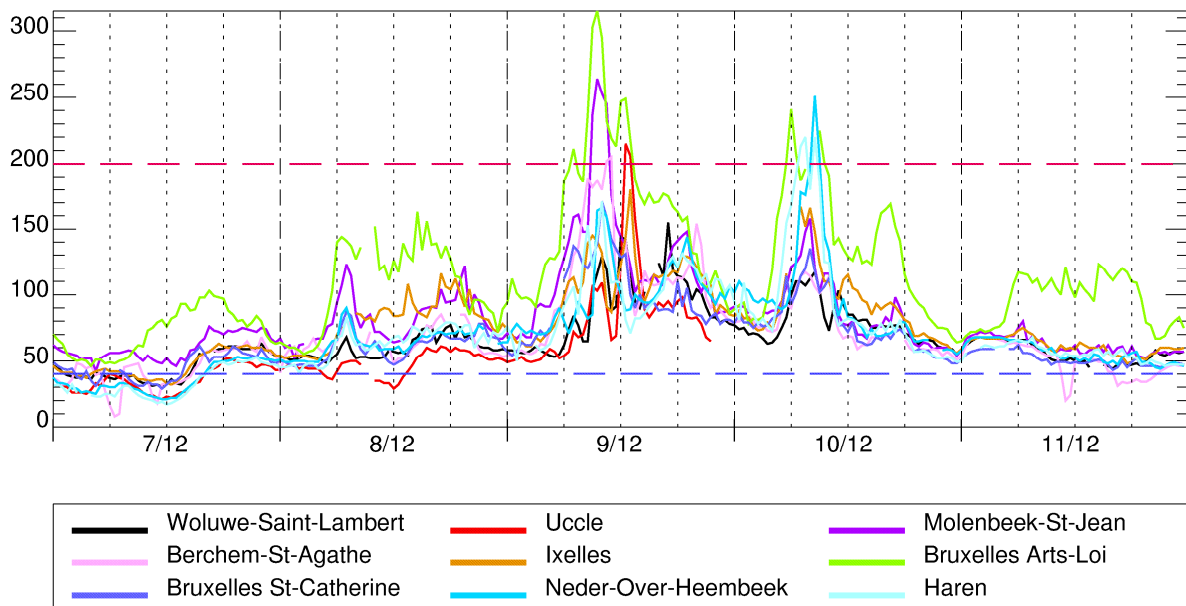


Figure 13: Concentrations de NO₂ (µg/m³) mesurées au cours du 7 au 11 décembre 2003 en Région bruxelloise.
Source : IBGE.



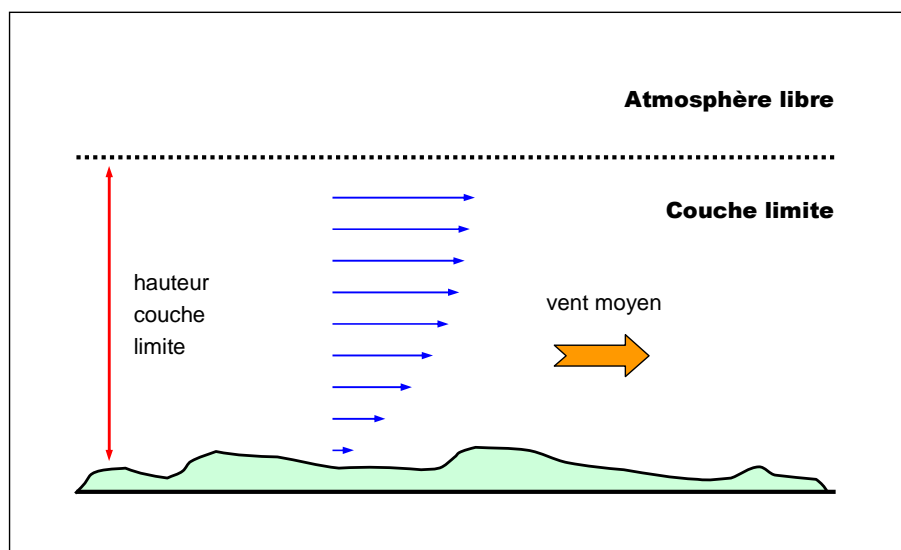


2.2. Transport horizontal dans la couche limite

Un autre paramètre météorologique peut apporter un complément d'analyse fort utile : le transport horizontal dans la couche limite. Ce dernier est calculé par le produit de la hauteur de la couche limite et du vent moyen dans cette couche (figure IV.14).

Figure IV.14 : Transport horizontal dans la couche limite, obtenu par le produit de la vitesse moyenne du vent dans la couche limite et la hauteur de la couche limite.

Source : CELINE.



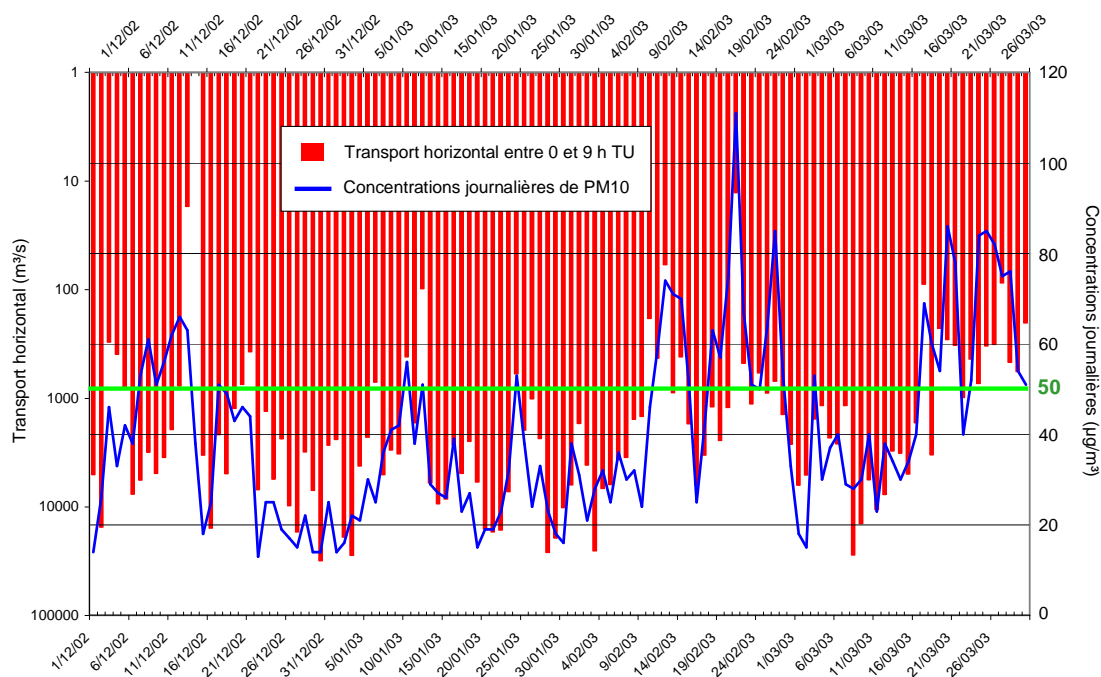
Le transport horizontal dans la couche limite est une variable qui peut être mise avantageusement en relation avec l'évolution des concentrations de PM10. La figure IV.15 montre en effet une intéressante anti-corrélation entre le logarithme du transport horizontal moyenné journalièrement entre 0 et 9 h et les concentrations moyennes journalières de PM10. Le transport horizontal moyenné sur la journée entière n'a pas montré un accord aussi marqué avec les concentrations journalières de PM10. Ceci souligne l'importance des conditions météorologiques en début de journée sur les concentrations mesurées, plus précisément la conjonction de conditions de faible dispersion avec le pic de pollution matinal (heures de pointe, relance des chauffages, etc). Le pic de pollution qui a lieu en soirée semble jouer un rôle moins important sur les concentrations journalières, probablement en raison des conditions de dispersion qui demeurent généralement plus favorables que le matin.

L'analyse du transport horizontal dans la couche limite au cours des hivers de 1999-2003 montre que des valeurs horaires inférieures à 20 m³/s coïncident avec des conditions de dispersion très faibles. A l'échelle journalière, ce seuil se traduit approximativement par des valeurs du transport horizontal inférieur à 100 m³/s.



Figure IV.15 : Représentation du transport horizontal dans la couche limite, en relation avec les concentrations journalières de PM10 observées à Bruxelles. Les valeurs journalières du transport horizontal sont établies à partir du transport moyen entre 0 et 9 h. Les valeurs du transport sont représentées suivant un axe logarithmique inversé, de façon à faciliter la comparaison avec l'évolution des PM10.

Source : CELINE.



2.3. Autres variables météorologiques explicatives de la dispersion des polluants

L'analyse météorologique d'une situation donnée ne se limite pas à la longueur de transport ou encore au transport horizontal dans la couche limite. D'autres variables plus élémentaires peuvent apporter des informations essentielles pour mieux comprendre les facteurs météorologiques qui vont moduler l'intensité de la dispersion.

L'analyse du champ de pression permet de situer l'influence des masses d'air dans le contexte de la circulation synoptique. En règle générale, des masses d'air d'origine maritimes seront moins chargées en polluants que des courants continentaux qui passent au-dessus de régions plus ou moins polluées. Des courants instables favorisent une dispersion efficace des polluants. Le passage de zone frontale (front chaud, froid ou occlus) est souvent à l'origine d'un renforcement de la turbulence qui s'avère bénéfique pour la dispersion. Au contraire, la présence de zone de haute pression (anticyclone) tend à stabiliser la baisse d'air et à repousser les polluants près du sol.

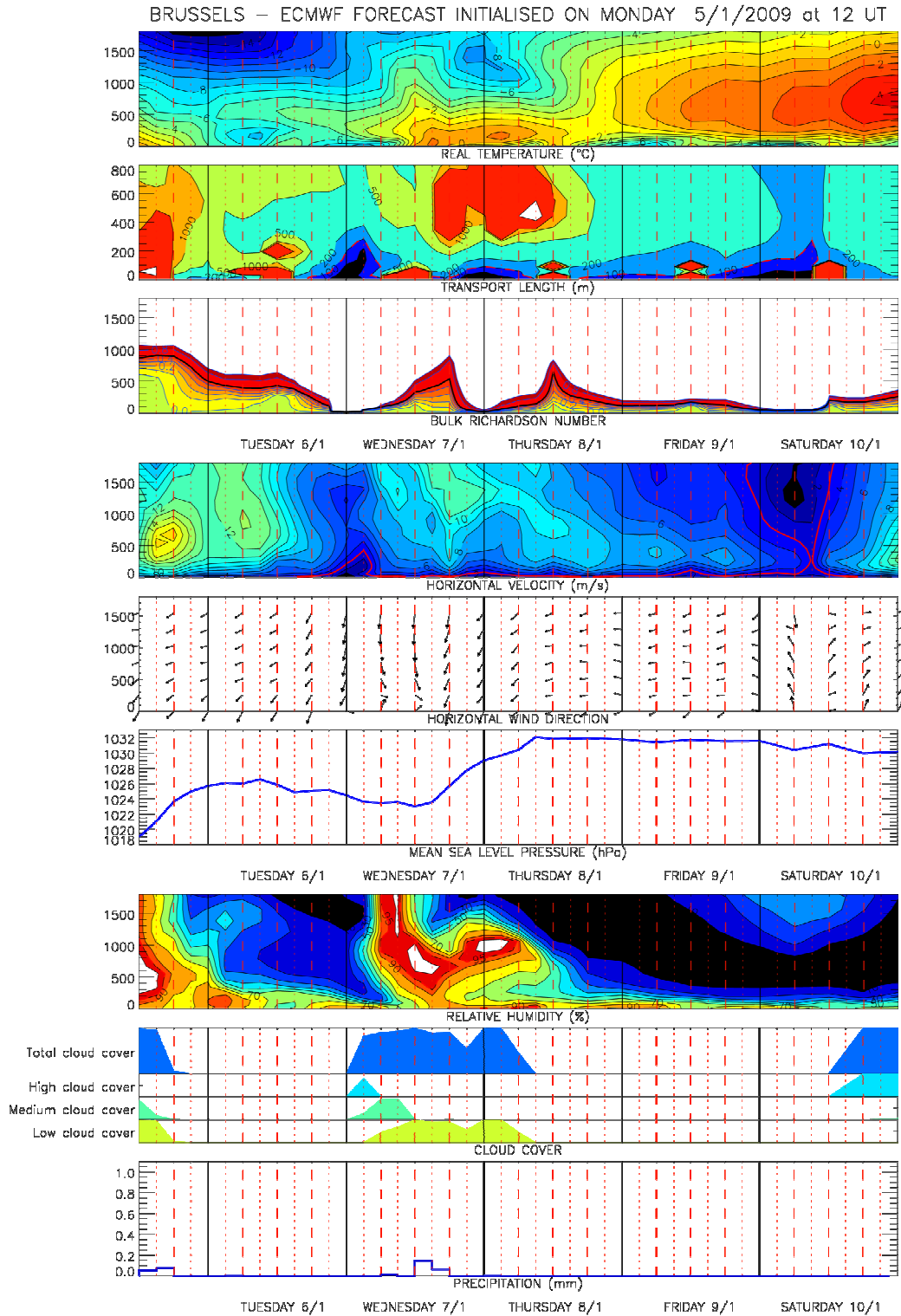
L'examen de l'évolution temporelle du profil vertical de température (exemple, figure IV.16) apporte des informations essentielles sur la stabilité de l'atmosphère. Il permet notamment de se rendre compte de :

- l'éventuelle présence d'une inversion thermique de subsidence, c'est-à-dire une inversion thermique formée dans les courants descendants associés à une masse d'air anticyclonique ;
- l'intensité de l'inversion thermique, matérialisée par la différence entre la température de l'inversion et la température en surface ;
- la persistance de l'inversion thermique en cours de journée, et en particulier durant l'après-midi, lorsque l'échauffement de la surface tend à diminuer son intensité.



Figure IV.16 : Exemple de météogramme établi à partir des prévisions ECMWF initialisée le 5 janvier 2009 à 12h. Les variables suivantes sont représentées (de haut en bas) : température, longueur de transport, nombre de Richardson avec indication de la hauteur de couche limite, vitesse du vent, direction du vent, pression en surface, humidité relative, nébulosité aux étages bas, moyen et élevé, précipitations.

Source : CELINE.





A cette liste pourraient aussi s'ajouter la vitesse du vent, l'état d'humidité de l'atmosphère et les précipitations (exemple, figure IV.16) :

- L'analyse du profil de vent permet de caractériser la turbulence d'origine mécanique. De faibles vitesses de vent dans les premières centaines de mètres de l'atmosphère sont souvent synonymes de conditions défavorables à la dispersion des polluants, même en l'absence d'inversion thermique.
- L'humidité de l'atmosphère peut interférer sur la masse des particules. Certaines catégories de particules sont hydrophiles et peuvent augmenter leur masse par capture d'eau en atmosphère très humide.
- Les précipitations, lorsqu'elles sont suffisamment intenses, contribuent au nettoyage de l'atmosphère par lessivage des polluants.

Les processus mentionnés ci-dessus ne sont pas exhaustifs, mais ils doivent tous être examinés pour caractériser et expliquer le plus précisément possible la dispersion des polluants.



3. Modélisation de la qualité de l'air

3.1. Apports de la modélisation en matière d'évaluation de la qualité de l'air

Comme pour la météorologie depuis plusieurs décennies, la modélisation numérique est devenu un outil important pour la recherche et les applications dans le domaine de la physico-chimie de l'atmosphère. Dans le sens le plus général, un modèle numérique rassemble les connaissances théoriques acquises dans un domaine donné et les formule à l'aide d'équations mathématiques qui ne peuvent généralement pas être résolues par voie analytique et, par voie de conséquence, nécessitent un traitement numérique. Il sera toujours dépendant des hypothèses sous-jacentes, et donnera une image plus ou moins proche de la réalité en fonction de la validité des hypothèses et des lois physiques ou chimiques. Les modèles peuvent donc être compris comme des outils d'intégration des connaissances, capables de traiter plusieurs processus interdépendants. Ils contribuent ainsi à faire progresser les connaissances, par exemple en identifiant les principales incertitudes.

Les modèles numériques existent depuis longtemps, mais ils ont réellement pris leur essor au cours de ces 30 dernières années avec le développement spectaculaire des capacités informatiques. La modélisation des propriétés physico-chimiques de l'atmosphère est, par essence, étroitement liée au développement des modèles météorologiques. La dynamique de l'atmosphère est en effet un des processus qui affecte significativement l'évolution des polluants atmosphériques. Mais d'autres aspects entrent aussi en compte, par exemple les interactions et réactions chimiques entre les différents constituants présents dans l'air ambiant. La complexité des différents processus connus peut être intégrée dans un modèle. Toutefois, les ressources informatiques n'étant pas illimitées, la grande majorité des modèles est bâtie sur des compromis. Par exemple, il s'agit d'identifier les processus les plus importants et ceux qui peuvent être négligés. Le compromis porte aussi sur le choix des méthodes numériques, le choix du domaine de simulation et de la résolution spatiale, le recours aux techniques d'imbrications multiples de domaines, etc.

Actuellement, la modélisation est impliquée dans différentes problématiques :

- *les prévisions de qualité de l'air* : prévoir les concentrations de polluants permet, par exemple, de prévenir les épisodes de pollution et de mettre en œuvre les mesures adéquates visant à réduire les émissions polluantes ;
- *les études de cas* : nombre de situations caractérisées par des concentrations élevées en polluants présentent des difficultés d'interprétation. Les modèles peuvent contribuer à mieux comprendre l'importance relative et la connaissance des différents processus physico-chimiques impliqués ;
- *l'assimilation de données* : les concentrations relevées aux stations de mesures possèdent une représentativité spatiale qui dépend du polluant considéré et de la proximité de sources polluantes. Un réseau de mesure ne possède généralement pas une densité suffisante pour donner une représentation complète de l'exposition aux polluants. Aussi les modèles peuvent contribuer à améliorer la représentation de la distribution spatiale des polluants. En particulier, les méthodes d'assimilation consistent à combiner de façon optimale les champs simulés par un modèle avec les données d'observation.
- *les scénarios d'émissions* : les directives européennes fixent des valeurs limites qui ne sont pas respectées pour certains polluants. Le respect des normes passe par une réduction, parfois très importante, des émissions des polluants concernés. L'impact des mesures concrètes qui seront mises en œuvre peut être préalablement évalué à l'aide des modèles.



3.2. Hiérarchie de modèles

Il existe différentes catégories modèles visant à décrire l'évolution des concentrations d'un ensemble d'espèces chimiques présentes dans l'air ambiant.

L'approche la plus complète consiste à simuler l'évolution du système couplé, c'est-à-dire à le fluide et les espèces transportées. Ceci revient à résoudre les équations de Navier-Stokes complétées d'une équation pour les concentrations. En pratique, la plupart des modèles sont basés sur un découplage entre la dynamique du fluide et les concentrations des espèces chimiques. Concrètement, ce découplage revient, d'une part, à négliger la contribution due aux réactions chimiques dans l'équation d'évolution de la température et, d'autre part, à figer l'interaction entre la matière et le rayonnement. Pour des études de qualité de l'air ou pour la prévision opérationnelle, ces deux approximations sont acceptables. Par contre, l'interaction entre matière et rayonnement est un processus qui ne peut être écarté dans le cadre d'études climatiques : dans ce cas de figure, on préférera un modèle couplant les processus dynamiques et chimiques.

Parmi les modèles exploitant l'hypothèse de découplage la dynamique du fluide et la chimie des espèces, on peut citer la hiérarchie suivante :

Type de modèle	Applications	Période
Gaussien	Risque industriel	
Bouffées	Risque industriel	
Lagrangien particulaire	Risque industriel	
Boîte eulérienne	Prévision	1980 – 2000
Boîte lagrangienne	Etude d'impact	1980 – 2000
Eulérien 3D, module chimique pour la phase gazeuse	Prévision, étude d'impact, étude de scénarios d'émissions	Depuis 1990
Eulérien 3D, module chimique multiphasique	Prévision, étude d'impact, étude de scénarios d'émissions	Depuis 2000
Modèle couplé (dynamique et module chimique)	Recherche	Depuis 2000
Réseaux neuronaux (perceptron)	Prévision	Depuis 1957

3.2.1. Modèles gaussiens

Les modèles gaussiens (U.S. Env. Protection Agency, 1970 ; Benarie, 1980 ; Sportisse, 2008) sont les plus anciens (vers 1936) et les plus simples. Ces modèles, dits de dispersion passive, supposent que la répartition du polluant suit une loi gaussienne dont les écarts-types dépendent de la distance à la source ou du temps écoulé depuis l'émission, ainsi que des caractéristiques de la structure de l'atmosphère (classes de stabilité).

Ils sont basés sur la solution analytique de l'équation d'advection-diffusion en tenant compte des hypothèses suivantes :

- polluant passif ou pouvant être considéré comme tel sur la période étudiée ;
- pas d'influence de l'émission de polluants sur l'écoulement d'air ;
- champ de vent constant et de vitesse non nulle, aligné sur une direction ;
- pas de relief ni d'obstacles.



Dans le cas d'un champ de vent u aligné sur la direction de l'axe des x , l'équation de dispersion pour le rapport de mélange C du polluant considéré s'écrit :

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} = K_x \frac{\partial^2 C}{\partial x^2} + K_y \frac{\partial^2 C}{\partial y^2} + K_z \frac{\partial^2 C}{\partial z^2}$$

Pour une émission ponctuelle à l'origine en $(x,y,z)=0$ et en $t=0$, la condition initiale est donnée par $C(x,y,z,0) = S$. La solution gaussienne de l'équation de dispersion est alors donnée par :

$$C(x,y,z,t) = \frac{S}{(2\pi)^{3/2} \sqrt{K_x K_y K_z}} \exp \left[-\frac{(x-ut)^2}{4K_x t} - \frac{y^2}{4K_y t} - \frac{z^2}{4K_z t} \right]$$

Cette solution analytique montre que le panache de polluants présentera, dans chaque direction, un étalement donné par la variance $2Kt$, souvent notée σ^2 . Elle constitue la base des modèles gaussiens. Compte tenu des hypothèses du modèle, le domaine de validité se limite à quelques kilomètres autour de la source.

Pour une source continue émettant à l'origine ($x=y=0$), à la hauteur h , la solution de l'équation de dispersion devient :

$$C(x,y,z) = \frac{S}{2\pi u \sigma_y \sigma_z} \exp \left(-\frac{y^2}{2\sigma_y^2} \right) \left[\exp \left(-\frac{(z-h)^2}{2\sigma_z^2} \right) + \exp \left(-\frac{(z+h)^2}{2\sigma_z^2} \right) \right]$$

Des paramétrisations ont été développées pour estimer les coefficients de dispersion σ^2 en fonction des champs météorologiques et de la distance à la source. Elles peuvent être discrètes, notamment lorsqu'elles sont fondées sur des classes de stabilité de l'atmosphère, ou calculées à partir de grandeurs météorologiques telles que la longueur de Monin-Obukhov ou la hauteur de mélange.

Le modèle gaussien pour une source continue peut encore être étendu au cas d'une source caractérisée par une émission variable dans le temps. On parle alors de modèle gaussien à bouffées. Concrètement, la variation temporelle des émissions sera représentée par des bouffées ponctuelles successives et de durée Δt . La concentration sera donnée par la somme des concentrations issues des bouffées individuelles.

Malgré leurs grandes limitations, ces modèles présentent l'avantage d'être rapides et, par conséquent, d'être capables d'étudier ainsi un grand nombre de cas. Ils fournissent des résultats acceptables dans les situations pour lesquelles ils ont été paramétrés. Leurs principales utilisations concernent les études et/ou applications liées à la dispersion de traceurs accidentels passifs dans une relative proximité (de l'ordre de quelques kilomètres) par rapport à la source.

3.2.2. Modèles lagrangiens particuliers

Le principe des modèles lagrangiens (Sportisse, 2008) repose sur le suivi de particules numériques. Pour un champ de vitesse du vent $\vec{V} = (u,v,w)$ et un champ de diffusion $K = (K_x, K_y, K_z)$, une particule numérique donnée se trouvant en $(x(t_n), y(t_n), z(t_n))$ au temps t_n sera déplacée comme suit au temps t_{n+1} :

$$\begin{aligned} x(t_{n+1}) &= x(t_n) + \left(u + \frac{\partial K_x}{\partial x} \right) \Delta t + \sqrt{2K_x} \Delta W_x \\ y(t_{n+1}) &= y(t_n) + \left(v + \frac{\partial K_y}{\partial x} \right) \Delta t + \sqrt{2K_y} \Delta W_y \\ z(t_{n+1}) &= z(t_n) + \left(w + \frac{\partial K_z}{\partial x} \right) \Delta t + \sqrt{2K_z} \Delta W_z \end{aligned}$$



ΔW_* est un processus stochastique qui représente un bruitage et qui suit une loi normale de moyenne nulle et de variance Δt .

Ce type de modèle permet de suivre, avec une relative finesse, une émission ponctuelle. Il présente par contre les inconvénients de nécessiter un grand nombre de particules pour converger et de ne considérer que des traceurs passifs.

3.2.3. Modèles de boîtes

Le principe des modèles de boîtes (Benarie, 1980 ; Sportisse, 2008) consiste à définir des sous-domaines (appelés « boîtes ») de grande dimension par rapport à une maille classique, et dans lesquels les concentrations sont considérées comme homogènes. Plusieurs boîtes peuvent être connectées entre elles en introduisant des termes de flux entre boîtes. Par exemple, une boîte peut représenter la couche de mélange, une autre la couche résiduelle, etc. Dans chaque boîte, les équations de la cinétique chimique, du dépôt et de lessivage sont résolues. Par contre, la turbulence n'est pas modélisée.

En raison de leur coût calcul plus limité qu'un modèle recourant à un maillage complet, ces modèles étaient utilisés principalement dans les années 1990. Ils servaient notamment à estimer les relations entre sources et récepteurs pour la pollution transfrontière. Actuellement, ils sont pratiquement délaissés.

3.2.4. Modèles de chimie-transport

Les modèles de chimie-transport (Sportisse, 2008) résolvent explicitement les équations d'advection et diffusion qui déterminent le transport des polluants. Ils comportent des paramétrisations physiques pour les différents processus, ainsi qu'un schéma réactionnel régissant l'évolution des constituants présents dans l'air ambiant.

Les aspects essentiels pris en compte par les modèles de chimie-transport peuvent être résumés comme suit (figure IV.11) :

a) Emissions

Les émissions constituent une des données d'entrée essentielle aux modèles de chimie-transport. Elles sont en général issues des inventaires annuels d'émissions. Pour chaque polluant concerné, ceux-ci consistent en un total annuel réparti en fonction des différents secteurs d'activités. Comme les modèles requièrent généralement la spécification des émissions à l'échelle horaire, des profils types sont utilisés pour désagréger – au mieux – les totaux annuels. Un autre aspect sensible concerne la redistribution de composés génériques donnés globalement, tels que les COV (composés organiques volatils), selon les différentes espèces réelles. Cette difficulté se pose également pour la redistribution des émissions de particules par classes de diamètre.

b) Champs météorologiques

Les champs météorologiques sont issus de modèles de prévision et sont interpolés sur la grille du modèle de chimie-transport. Ils concernent non seulement les variables tridimensionnelles classiques, telles que le vent, la température et l'humidité relative, mais aussi des champs bidimensionnels comme les flux de surface, la couverture nuageuse et les précipitations. La spécification des champs météorologiques conditionne directement la représentation des processus de transport et de dispersion des polluants.

c) Schéma réactionnel

La représentation des mécanismes chimiques est également un élément essentiel d'un tel modèle. La pollution photochimique met typiquement en jeu une centaine d'espèces et implique plusieurs centaines de réactions pour la phase gazeuse. La plupart des schémas réactionnels incluent aussi un certain nombre de réactions pour la prise en compte des aérosols.



d) Elimination des polluants par lessivage et dépôt sec

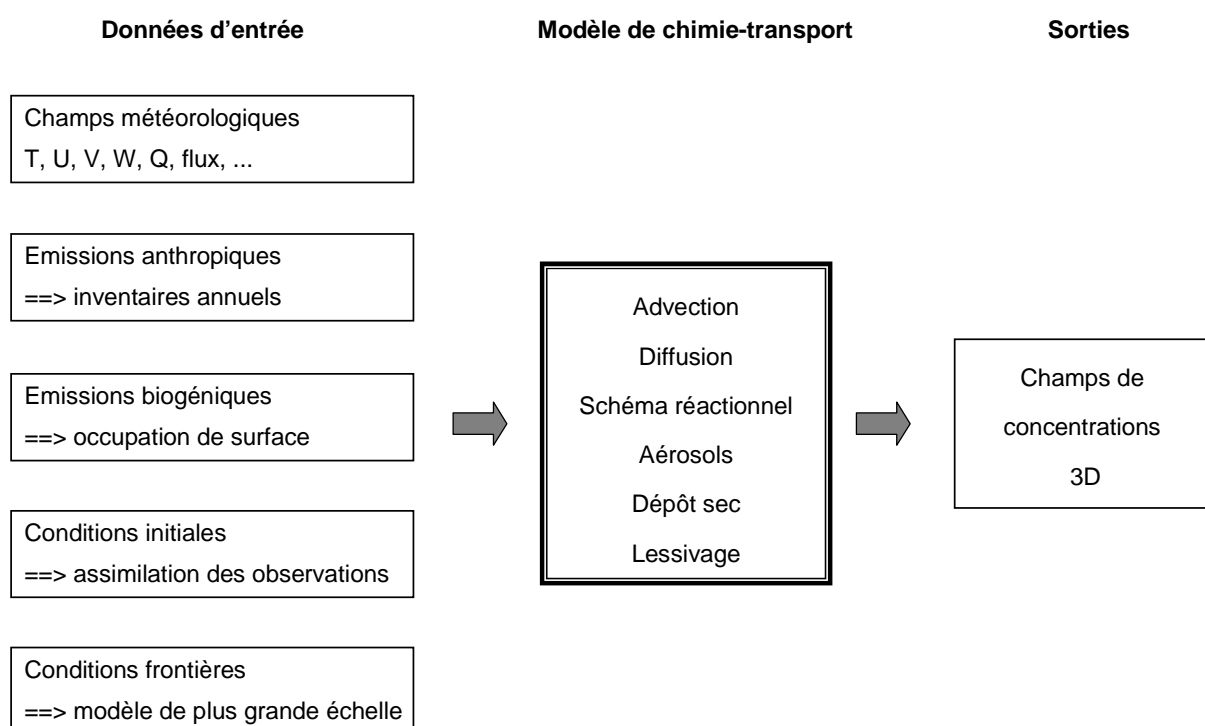
Le lessivage et le dépôt sec conditionnent le temps de résidence des espèces gazeuses solubles et des aérosols dans l'air ambiant. La représentation de ces processus est cependant entachée d'une forte incertitude.

En ce qui concerne le lessivage par les nuages ou par les précipitations, cette incertitude provient tant de la paramétrisation proprement dite que des paramètres météorologiques dont elle dépend (microphysique nuageuse, précipitations).

Le dépôt sec dépend quant à lui des composants considérés, des conditions météorologiques dans la couche limite de surface et de l'occupation des sols. Il est habituellement paramétré en fonction d'une vitesse de déposition qui est estimée par un modèle de résistances. Son évaluation est néanmoins incertaine.

Figure IV.11 : Schéma de principe d'un modèle de chimie-transport

Source : IBGE.



Comme tout modèle, les modèles de chimie-transport sont entâchés d'incertitudes. Les principales sources sont :

- la faible qualité des données d'émission dérivées à partir des inventaires annuels disponibles pour quelques polluants ;
- les incertitudes au niveau des champs météorologiques qui se répercutent directement sur le modèle de dispersion ;
- les paramétrisations de sous-maille qui représentent implicitement un processus lorsque la résolution spatiale n'est pas suffisante pour le représenter explicitement ;
- la prise en compte d'un nombre limité d'espèces et de réactions chimiques pour limiter les besoins en temps de calcul ;
- la résolution spatiale du modèle limitée par la précision des données d'entrée (émissions et champs météorologiques), ainsi que par les ressources informatiques.



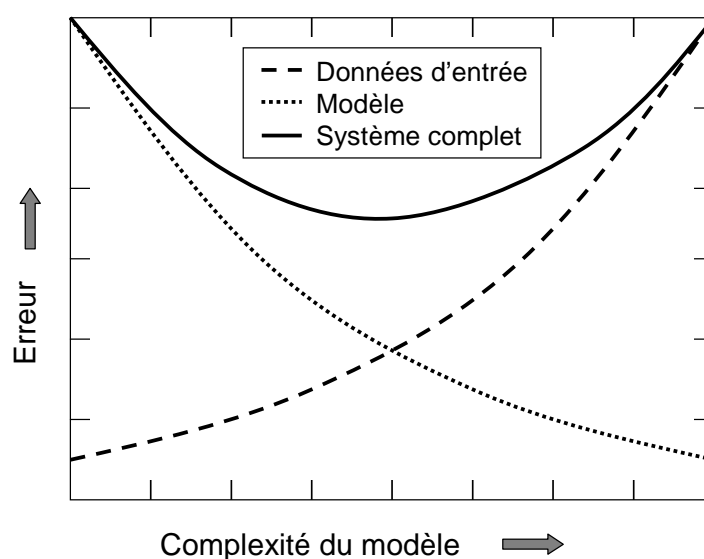
Pour fixer les idées, le tableau ci-dessous donne une estimation des incertitudes typiques sur les données d'entrée d'un modèle de chimie-transport :

Données d'entrée	Incertitude relative
Emissions anthropiques	+/- 50%
Emissions biogéniques	+/- 100%
Couverture nuageuse	+/- 30%
Vitesse du vent	+/- 20%
Epaisseur de la couche limite	+/- 20%
Précipitations	+/- 50%
Conditions aux limites pour l'ozone	+/- 20%
Constante photolytique	+/- 30%
Vitesse de dépôt sec	+/- 30%

Il est donc nécessaire de garder un équilibre entre la complexité des modèles et la qualité des données d'entrée. Comme le représente la figure IV.12, un modèle complexe est censé mieux représenter la réalité, mais il nécessite aussi des données ou paramètres d'entrée plus détaillés et parfois difficiles à estimer. Un modèle simple sera plus approximatif, mais il fera appel à des données d'entrée qui seront plus aisées à obtenir et, par conséquent, moins incertaines. L'erreur du système complet – i.e. modèle et données d'entrée – correspond à la somme des erreurs associées au modèle et aux entrées. L'erreur minimale sera atteinte pour une complexité raisonnable du modèle.

Figure IV.12 : Erreur sur les sorties d'un modèle exprimée en fonction de sa complexité et de la qualité des données d'entrée.

Source : IBGE.



Le modèle CHIMERE (Vautard et al., 2003 ; Bessagnet et al., 2004 ; Honoré et al., 2008), utilisé pour les prévisions opérationnelles de qualité de l'air à CELINE, est un exemple de modèle eulérien de chimie-transport. Les processus physico-chimiques pris en compte dans ce modèle sont :

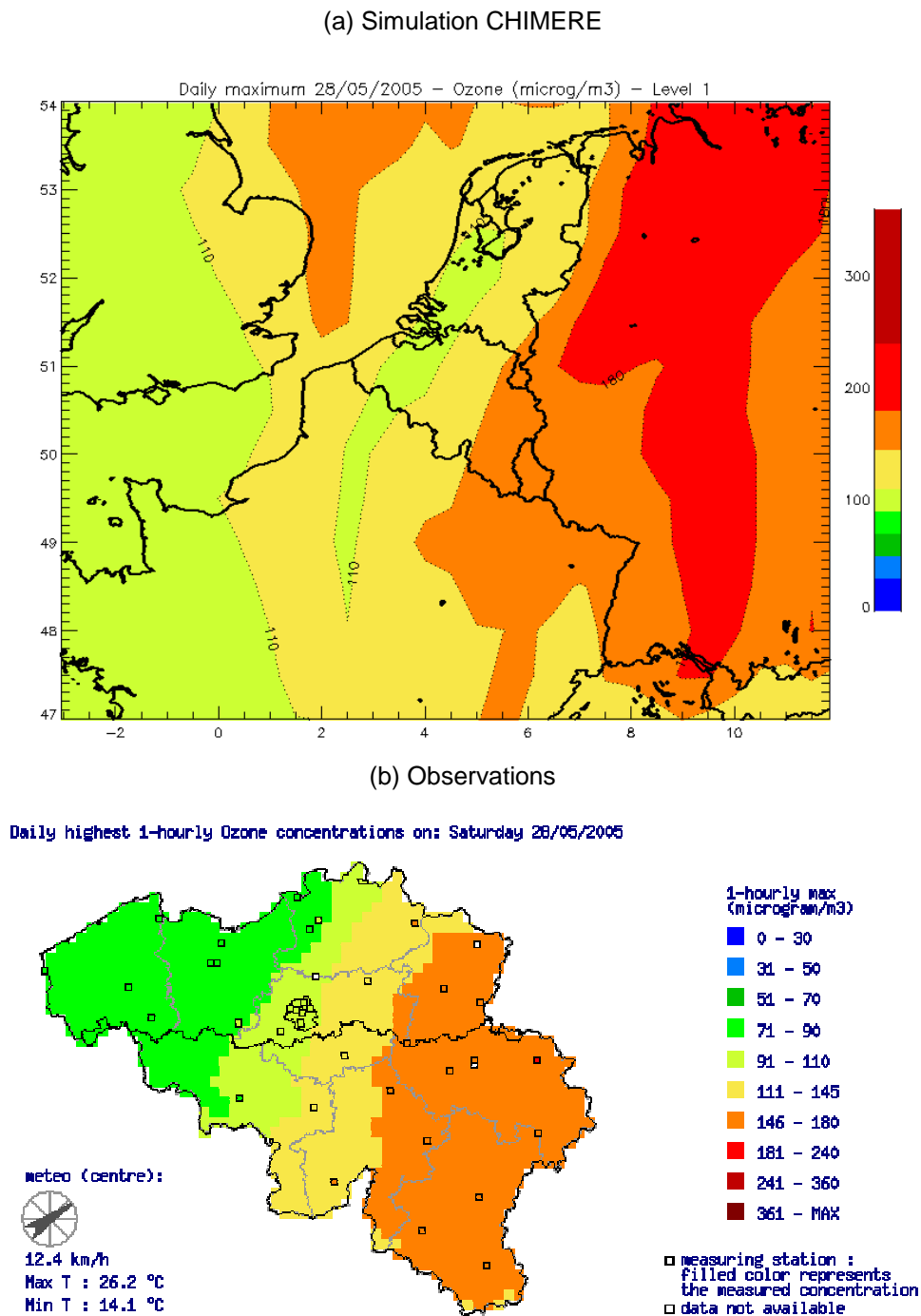
- la dispersion horizontale des espèces chimiques par la composante horizontale du vent ;
- la dispersion verticale des espèces chimiques par la composante verticale du vent et par la convection verticale ;
- le dépôt sec sur le sol des espèces chimiques ;
- les réactions chimiques entre les espèces présentes dans l'atmosphère ;
- la microphysique et chimie des aérosols.



Afin d'estimer les flux de production ou de perte en polluants, liés aux processus physico-chimiques, un certain nombre de données d'entrée doit être fournis aux modèles de chimie transport, en particulier les prévisions météorologiques et les inventaires d'émissions de polluants.

Depuis 2005, CHIMERE produit quotidiennement les cartes de prévision qui sont mises à disposition sur le site internet de CELINE, pour les différents polluants faisant l'objet d'un suivi continu (ozone, PM10, NO₂, SO₂). Il constitue aussi un des éléments essentiels dans la prévision des situations de pollution et, par conséquent, dans la prise de décision conduisant à la mise en œuvre des plans d'actions régionaux. Un exemple de prévision d'ozone pour la journée du 28 mai 2005 est présenté en figure 13.

Figure IV. 13 : Concentrations maximales d'ozone (a) simulées par CHIMERE et (b) observées le 28 mai 2005.
Source : CELINE.





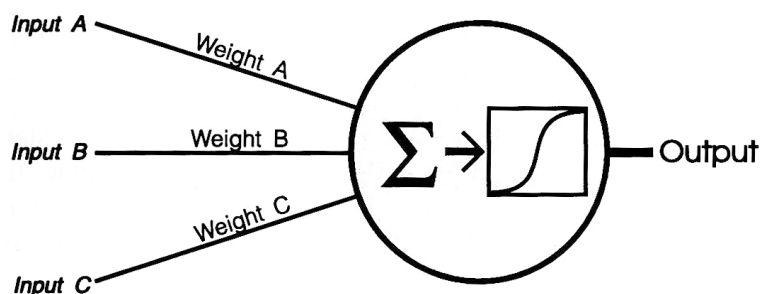
3.2.5. Modèles neuronaux

Un réseau de neurones artificiels (Hewitson et Crane, 1994) est un modèle conceptuellement inspiré du fonctionnement des neurones biologiques. Il possède en commun avec le cerveau les mêmes macrostructures de base, et est composé d'éléments interconnectés et assimilés aux neurones. Trois éléments principaux caractérisent les réseaux neuronaux :

- *Les fonctions de transfert associées à chaque neurone.* Chaque neurone peut être compris comme un automate doté d'une fonction de transfert qui transforme ses entrées en sorties selon des règles précises. Par exemple, un neurone peut sommer ses entrées, comparer la résultante à un seuil, et donner un signal si la somme dépasse le seuil en question (figure IV.13).
- *La topologie des interconnexions entre neurones est variable ;* elle peut être proactive, récurrente, etc.
- *L'efficacité de la transmission d'un neurone à l'autre peut varier.* Celle-ci est réglée par des poids attribués à chaque neurone qui peuvent être modulés en fonction des règles d'apprentissage.

Figure IV. 13 : Exemple de nœud (neurone) dans un réseau neuronal

Source : Bruce and Crane, 1994.



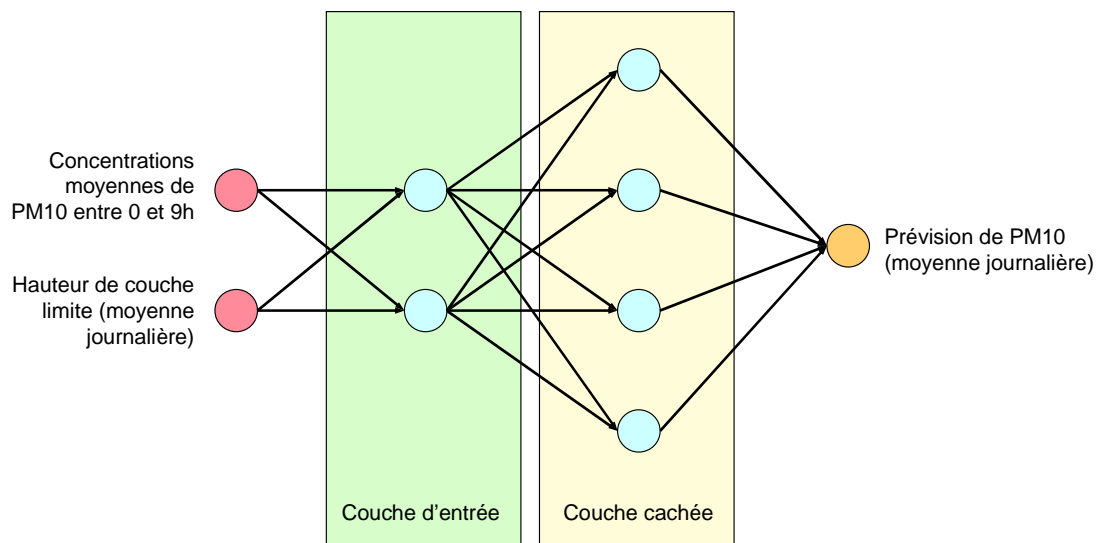
Les réseaux neuronaux sont donc capables d'opérer des classifications et d'apprendre à les améliorer. Contrairement aux modèles de nature déterministe, l'approche neuronale ne requiert pas de compréhension préalable des processus régissant la problématique étudiée. La qualité de ce type de modèle dépendra principalement de la pertinence des données d'entrée et de l'algorithme d'apprentissage qui exploitera ces données pour ajuster au mieux les différents paramètres du modèle.

Le modèle OVL est un exemple d'application de ce type d'approche à la prévision opérationnelle de la qualité de l'air. Développé par Hooyberghs et al. (2005), il est utilisé pour prévoir les concentrations de PM10 à l'endroit des stations de mesure. La topologie du réseau neuronal est de type perceptron proactif à 1 couche cachée de 4 nœuds (figure IV.14).



Figure IV. 14 : Topologie du modèle OVL, en tant que réseau neuronal de type perceptron à une couche cachée.

Source : IBGE.

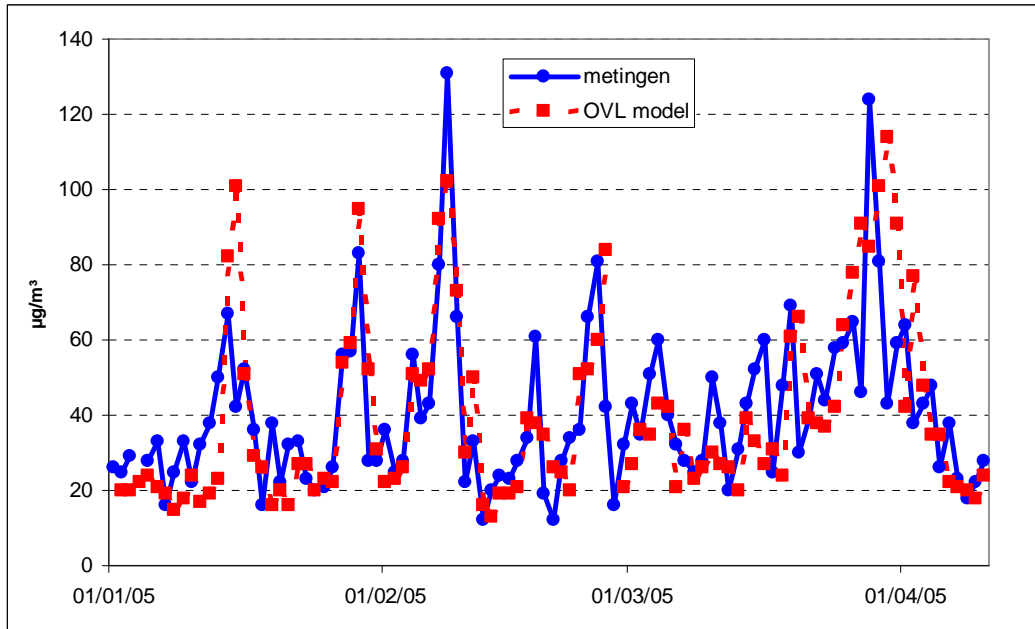


Sa mise en oeuvre a nécessité la recherche des variables d'entrée les plus pertinentes pour décrire l'évolution des concentrations journalières de PM10. Les principales variables étudiées étaient la vitesse et la direction du vent à différentes hauteurs, la hauteur de la couche limite, la longueur de transport moyenne dans différentes couches, le nombre de Richardson, la température, l'humidité relative et la couverture nuageuse. Parmi celles-ci, l'analyse des concentrations mesurées a montré que la hauteur de la couche limite était le paramètre le plus représentatif pour décrire la dispersion des particules fines à l'échelle journalière (figure IV.15). Cette grandeur est donc devenue une des entrées du modèles OVL. Les concentrations moyennes entre 0 et 9h ont été ajoutées comme seconde entrée, de façon à indiquer au modèle l'état des concentrations de PM10 au moment où la prévision est effectuée.



Figure IV. 15 : Comparaison entre les concentrations journalières de PM10 prévues par le modèle OVL (rouge) et les observations (bleu) à la station de Borgerhout (42R801) et en considérant la période du 1 janvier au 15 avril 2005.

Source : CELINE

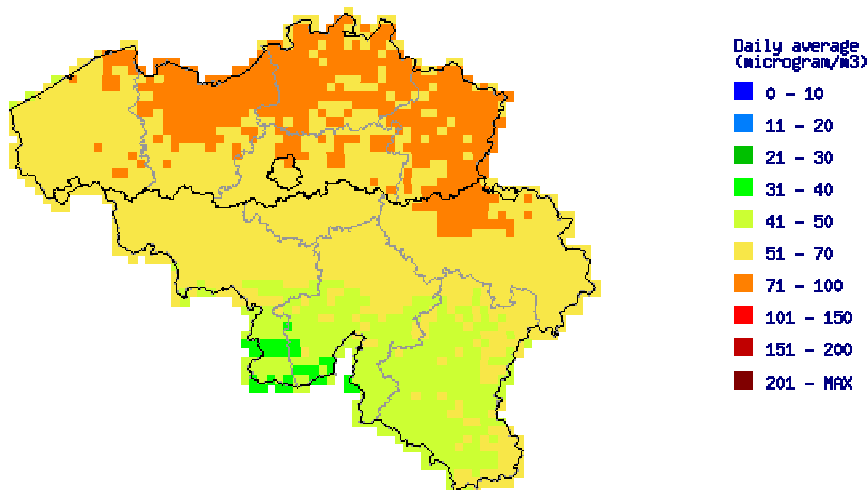


La figure IV.15 présente un exemple de validation du modèle OVL par comparaison avec les moyennes journalières observées en PM10. Les prévisions sont calculées en l’endroit des stations pour lesquelles il existe un historique de mesures portant sur une période d’au moins 3 ans. En exploitant la méthode d’interpolation RIO, il est possible de dresser une carte de concentrations de PM10 à partir des valeurs prévues aux stations. Un exemple est présenté en figure IV.16.

Figure IV. 16 : Exemple de carte de prévision de concentrations journalières de PM10, obtenue en appliquant l’interpolation RIO aux valeurs prévues aux stations par le modèle OVL.

Source : CELINE.

Forecasted daily mean Particulate Matter (PM10) concentrations for Sunday 07/02/2010





3.3. Optimisation des sorties de modèles dans le cadre de la prévision opérationnelle

Différentes raisons expliquent les erreurs constatées dans les sorties d'un modèle de prévision : les données d'entrée, la formulation intrinsèque du modèle, la résolution spatiale, les schémas numériques, etc. Afin de mieux cerner leur importance relative sur les résultats de simulation, plusieurs techniques ont été développées pour pallier aux incertitudes sur les données d'entrée des modèles.

3.3.1. Prévision d'ensemble

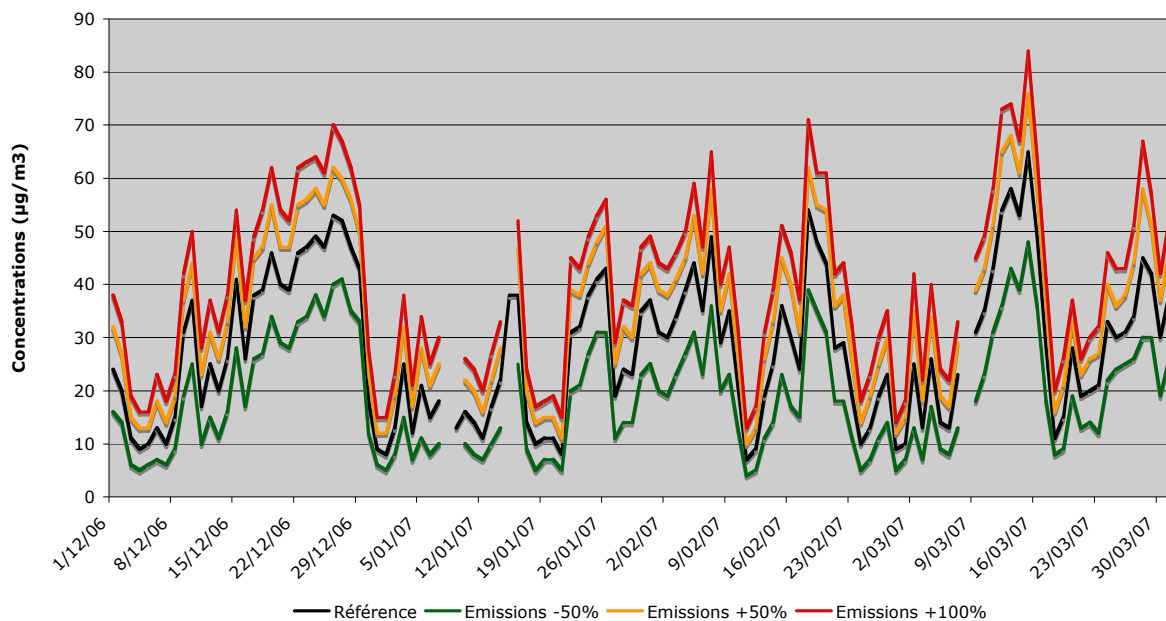
La prévision d'ensemble est une technique de prévision dans laquelle un modèle numérique de prévision est exécuté plusieurs fois, pour une même situation à prévoir, à partir de conditions initiales différant entre elles par de petites quantités compatibles avec les incertitudes existant sur la connaissance de l'état initial du système. La prévision d'ensemble est conçue pour capturer la probabilité d'événements et la zone d'incertitude inhérente à chaque situation prévue.

Dans le cas des modèles de chimie-transport, les incertitudes ne se limitent pas seulement aux champs initiaux de concentrations de polluants, mais il convient aussi de considérer les données d'entrées telles que les émissions et les champs météorologiques qui sont généralement entachés de fortes incertitudes. Pour illustrer le propos, on peut estimer qu'une erreur de 1 m/s sur le champ de vent est généralement sans incidence significative pour des situations météorologiques caractérisées par de bonnes conditions de dispersion ; au contraire, cette même erreur peut affecter significativement l'évolution des concentrations d'oxydes d'azote en cas de pic de pollution.

Pour estimer l'importance relative des erreurs ou incertitudes sur les différentes données d'entrée fournies au modèle, un ensemble de simulations perturbées peut être réalisé afin de mieux cerner la réponse du modèle. Un exemple de perturbations sur les données d'émission est présenté en figure IV.13.

Figure IV.13 : Evaluation de l'impact des émissions sur les concentrations moyennes journalières de PM10 simulées par le modèle CHIMERE (prévisions pour le jour J+1) durant l'hiver 2006-2007, en moyenne spatiale sur la Belgique. Ce graphique prend en compte les perturbations suivantes sur les émissions (pour tous les polluants) à raison de -50, +50 et +100%. La simulation de référence (non perturbée) du modèle CHIMERE est indiquée en trait noir.

Source : CELINE.





Lorsque l'on dispose d'un ensemble de simulations obtenues en appliquant des perturbations sur des champs considérés comme sensibles ou critiques, la meilleure prévision peut être déduite par combinaison des différentes prévisions. La méthode la plus simple consiste prendre la moyenne d'ensemble ou, pour des applications liées aux risques accidentels, la valeur médiane. Une autre approche serait de pondérer l'importance de chaque simulation en fonction de critères objectifs, par exemple en prenant en compte les observations récentes et en estimant les poids pour minimiser l'erreur de la prévision d'ensemble.

3.3.2. Assimilation de données

Le système d'assimilation des données a pour objectif de réduire au minimum les erreurs sur les champs initiaux des modèles de prévision en utilisant les différentes données d'observation disponibles. Elle permet ainsi de coupler les modèles aux observations. L'interprétation de cette technique peut être vue sous deux angles : soit l'application aux sorties des modèles d'une contrainte par les données d'observation, soit une interpolation des mesures contrainte par les sorties des modèles.

La méthode d'assimilation la plus simple est l'interpolation optimale. Pour en décrire brièvement le principe, considérons y_{mod} et y_{obs} comme étant respectivement les valeurs modélisée et observée d'une concentration de polluant. Ces deux valeurs possèdent leurs propres incertitudes, de sorte que l'on désigne par σ_{mod}^2 et σ_{obs}^2 leurs variances respectives. L'inverse des variances peut être vu comme la confiance attribuée à la variable correspondante. L'interpolation optimale va produire une valeur analysée qui sera une combinaison des valeurs modélisée et observée. Celle-ci sera obtenue en minimisant la fonction de coût J :

$$J = \frac{1}{2\sigma_{\text{mod}}^2} (y - y_{\text{mod}})^2 + \frac{1}{2\sigma_{\text{obs}}^2} (y - y_{\text{obs}})^2$$

La fonctionnelle J étant une fonction convexe, son minimum est obtenu en annulant la dérivée première. La valeur analysée est donnée par :

$$y_{\text{analyse}} = \left(\frac{y_{\text{mod}}}{\sigma_{\text{mod}}^2} + \frac{y_{\text{obs}}}{\sigma_{\text{obs}}^2} \right) / \left(\frac{1}{\sigma_{\text{mod}}^2} + \frac{1}{\sigma_{\text{obs}}^2} \right)$$

3.3.3. Modélisation inverse

La modélisation inverse (Sportisse, 2008 ; ECMWF, 2010) s'appuie également sur une approche semblable à l'assimilation de données. Comme indiqué précédemment, les données d'entrée des modèles de chimie-transport sont affectées par des incertitudes qui nuisent à la qualité des simulations. Pour des paramètres d'entrée difficiles à estimer à partir des seules observations, la modélisation inverse peut se révéler particulièrement utile.

Considérons le cas simple donné par le modèle linéaire $y = Ax$, où x est un vecteur à p dimensions correspondant aux entrées du modèle que l'on cherche à contrôler. On définit un opérateur linéaire d'observation H qui lie les observations aux sorties du modèle via la relation $y_{\text{obs}} = H \cdot y$. y_{obs} est un vecteur à m éléments, correspondant aux observations effectuées. Si l'observation était complète (pour tous les points visés par le modèle) et sans erreur, il suffirait d'inverser le système linéaire $y_{\text{obs}} = H \cdot A \cdot x$ pour déterminer x . En pratique, ce système est généralement sous-déterminé ($m < p$) et les observations comportent une incertitude. Pour une valeur estimée a priori de x , notée x_b , la recherche de x passera par la minimisation de la fonction de coût suivante :

$$J(x) = \frac{1}{2} (y_{\text{obs}} - HAx)^T R^{-1} (y_{\text{obs}} - HAx) + \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b)$$

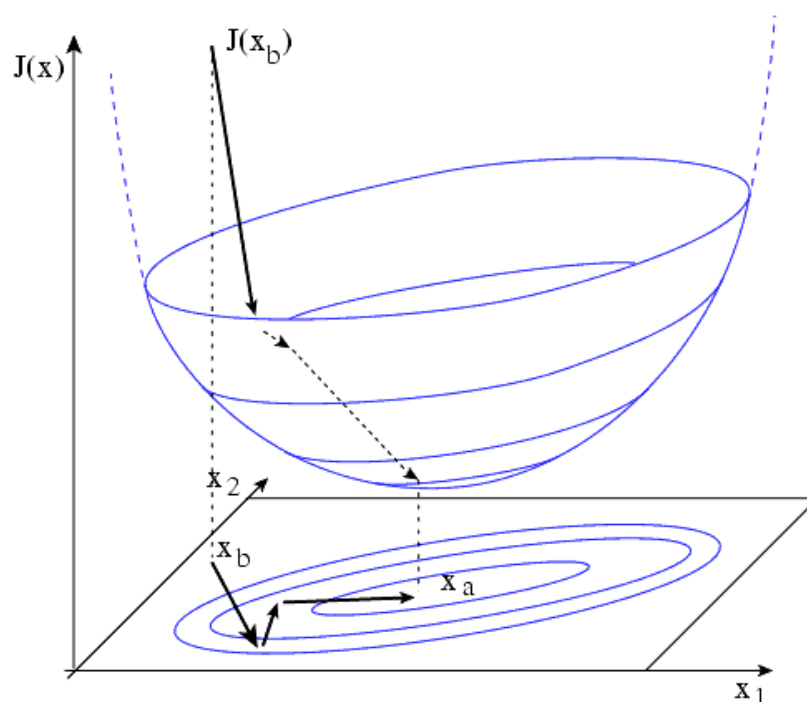
où R et B représentent respectivement les matrices de covariances d'erreur sur l'observation et l'estimation x_b . L'exemple présenté ci-dessus correspond à une assimilation variationnelle de type 3D-VAR, le « 3D » désignant la seule prise en compte de l'espace.



Les méthodes variationnelles ont notamment été utilisées pour estimer les émissions de méthane et de composés organiques volatils. Elle peut aussi être exploitée pour localiser des sources polluantes, par exemple lors d'émissions accidentelles.

Figure IV.14 : Représentation schématique de la minimisation de la fonction de coût pour l'assimilation de type 3D-VAR. La fonction de coût est représentée par un paraboloïde dont le minimum est l'optimum à atteindre. Partant d'une solution à priori, la minimisation consiste à progresser, en plusieurs étapes, vers cet optimum. Pour accélérer la convergence, les techniques numériques exploitent l'analyse du gradient local sur le paraboloïde.

Source : <http://www.ecmwf.int>.



3.3.4. Correction temporelle des prévisions

Les modèles de prévisions simulent l'évolution temporelle des concentrations de polluants. En fonction des sources d'incertitudes et des processus physico-chimiques les plus pertinents, les modèles peuvent présenter des biais variables dans le temps. Le recours aux données d'observation peut à nouveau s'avérer bénéfique pour réduire l'ampleur des biais sur les valeurs simulées. Plusieurs approches existent.

i) Filtre de Kalman :

Les méthodes séquentielles sont basées sur la théorie du filtre de Kalman (Welch and Bishop, 2006). Le filtre de Kalman est un estimateur récursif d'ordre 1 : pour estimer l'état courant, seuls l'état précédent et les observations à l'instant même sont nécessaires. Cela signifie que l'historique des observations et des estimations n'est pas requis. Il est possible d'utiliser un filtre de Kalman d'ordre supérieur en tenant compte de plusieurs étapes antérieures, ce qui le rend plus performant, mais aussi d'autant plus complexe.

L'intérêt du filtre de Kalman réside dans le contrôle de la prédiction : il estime l'état du système, puis améliore son résultat avec les données d'observation. Il comporte donc deux étapes distinctes : la prédiction et la correction. La phase de prédiction utilise l'état estimé de l'instant précédent pour produire une estimation de l'état courant. Dans l'étape de mise à jour, les observations de l'instant courant sont utilisées pour corriger l'état prédit dans le but d'obtenir une estimation plus précise (figure IV.15).

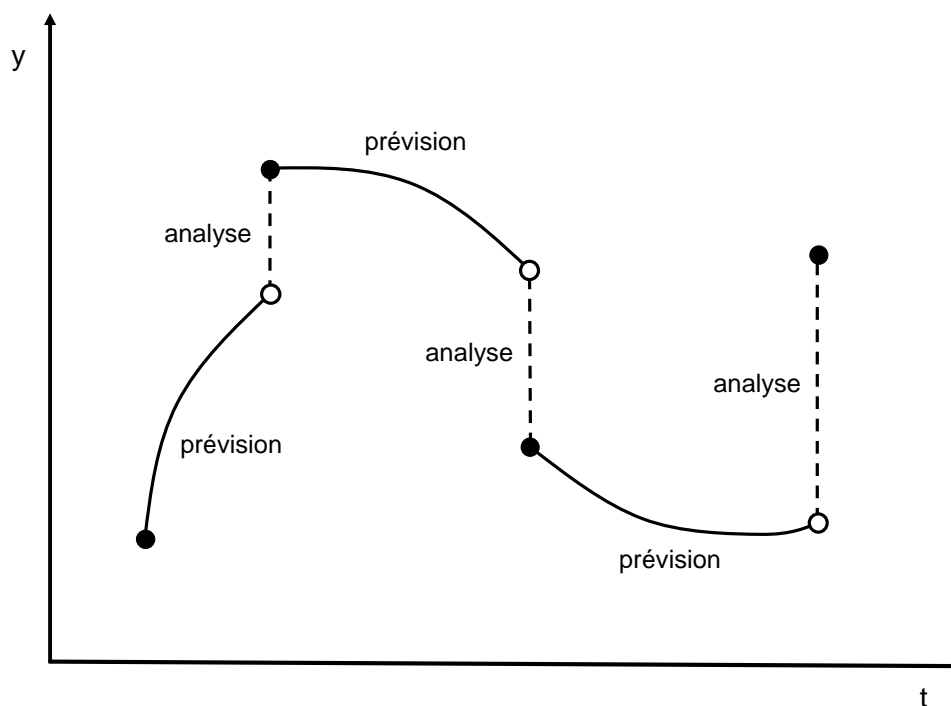


Le filtre de Kalman est fréquemment utilisé pour analyser le comportement d'un système linéaire qui fonctionne sous des conditions de bruit gaussien. Cependant, la plupart des systèmes physiques sont non linéaires : la non-linéarité peut provenir du modèle d'évolution, du modèle d'observation ou bien des deux. Dans le filtre de Kalman étendu, les modèles d'évolution et d'observation n'ont pas besoin d'être des fonctions linéaires de l'état, mais deviennent des fonctions différentiables. Dans ce cas, la convergence n'est cependant pas assurée et dépend en grande partie des valeurs de l'état initial que l'on choisit de façon plus ou moins empirique.

En résumé, le filtre de Kalman est une technique très efficace pour obtenir la solution optimale d'un problème dont la connaissance n'est que partielle. Sa principale force vient de l'exploitation des données d'observation qui sont utilisées pour corriger les estimations à partir du modèle d'évolution.

Figure IV.15 : Assimilation de données par une méthode séquentielle appliquée à la prévision. A chaque instant auquel les données d'observations sont disponibles, une analyse par interpolation optimale est effectuée dans le but de corriger l'état prévu par le modèle. Le nouvel état analysé est ensuite utilisé comme condition initiale du modèle de prévision.

Source : IBGE.



ii) Méthodes variationnelles 4D-VAR

Les méthodes variationnelles 4D-VAR (Delmas et al., 2005 ; Benedetti et al., 2008 ; ECMWF, 2010) sont une extension de l'approche 3D-VAR décrite précédemment, dans le sens où elles prennent en compte le temps en plus de l'espace tridimensionnel (figure IV.16).

Soit x le paramètre à contrôler. Le modèle utilisé pour calculer l'évolution des concentrations de polluants dépend à la fois des prévisions au pas précédent et des paramètres à contrôler :

$$y_{\text{mod}}(t_{i+1}) = F(y_{\text{mod}}(t_i), x)$$

La formulation variationnelle 4D-VAR consiste à minimiser la fonction de coût suivante :

$$J(x) = \frac{1}{2} \sum_{i=1}^n \left[(y_{\text{obs},t_i} - H \cdot y(t_i))^T R_i^{-1} (y_{\text{obs},t_i} - H \cdot y(t_i)) \right] + \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b)$$

La suite des valeurs simulées $y(t_i)$ est contrainte à la fois par le modèle et par x . La minimisation de $J(x)$ est réalisée par une méthode de type gradient.

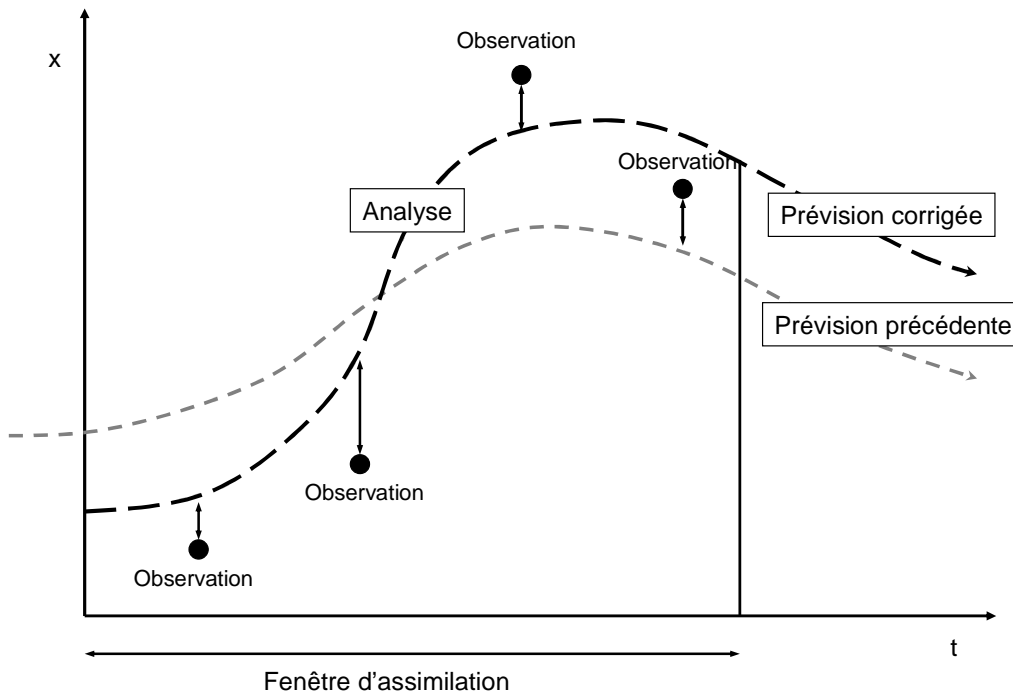


Toutefois, les non-linéarités des modèles numériques atmosphériques peuvent conduire à des fonctions de coût complexes. Les minimiseurs utilisés sur ces fonctions n'aboutissent pas forcément à une minimisation fiable et présentent, de surcroît, un coût élevé.

En regard de ces contraintes, l'utilisation de méthodes variationnelles de type 4D-VAR doit être analysée en regard des ressources informatiques disponibles et des gains réels obtenus par rapport à d'autres méthodes d'assimilation comparativement plus simples à mettre en œuvre.

Figure IV.16 : Illustration du principe de l'assimilation variationnelle 4D-VAR. La prévision du modèle est corrigée en prenant en compte les observations disponibles. La fenêtre d'assimilation différencie l'approche 4D-VAR de la méthode 3D-VAR en prenant en compte le temps en plus de l'espace tridimensionnel.

Source : IBGE.





Sources

1. BENARIE M., 1980 : « Urban air pollution modelling », The Macmillan Press Ltd, 405 pp.
2. BENEDETTI, A., J.-J. MORCRETTE, O. BOUCHER, A. DETHOF, R.J. ENGELEN, M. FISHER, H. FLENTJE, N. HUNEEUS, L. JONES, J.W. KAISER, S. KINNE, A. MANGOLD, M. RAZINGER, A.J. SIMMONS, M. SUTTIE, and the GEMS-AER team, 2008 : « Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Data Assimilation », *J. Geophys. Res.*, 114, D13205, doi:10.1020/2008JD011115.
3. BESSAGNET B., A.HODZIC, R.VAUTARD, M.BEEKMANN, S.CHEINET, C.HONORÉ, C.LIOUSSE and L.ROUIL, 2004 : « Aerosol modeling with CHIMERE - preliminary evaluation at the continental scale », *Atmospheric Environment*, 38, 2803-2817.
4. BRUCE C.H., and R.G. Crane, 1994 : "Neural Nets: Applications in Geography", Kluwer Academic Publishers.
5. CHOVIN P., et A. ROUSSEL, 1973, « Physicochimie et physicopathologie des polluants atmosphériques », Ed. Masson et Cie Editeur, 299 pp.
6. DELMAS R., MÉGIE G., et V.-H. PEUCH, 2005 : « Physique et chimie de l'atmosphère », Ed. Belin, 639 pp.
7. DEROUANE A., et G. VERDUYN, 1980 : « Méthodologie des enquêtes en matière de pollution de l'air », rapport de l'Institut d'Hygiène et d'Epidémiologie (IHE).
8. ECMWF, 2010 : « Data assimilation : IFS documentation front page », <http://www.ecmwf.int/research/ifsdocs/ASSIMILATION/>
9. HEWITSON B. C. and R. G. CRANE, 1994 : « Neural Nets : Applications in Geography », The GeoJournal Library, volume 29, 194 pp., Kluwer Academic Publishers.
10. HONORÉ C., L. ROUIL, R. VAUTARD, M. BEEKMANN, B. BESSAGNET, A. DUFOUR, C. ELICHEGARAY, J.-M. FLAUD, L. MALHERBE, F. MELEUX, L. MENUT, D. MARTIN, A. PEUCH, V.-H. PEUCH, N. POISSON, 2008 : « Predictability of European air quality: The assessment of three years of operational forecasts and analyses by the PREV'AIR system », *J. Geophys. Res.*, Atmospheres , 113, D04301, <http://dx.doi.org/10.1029/2007JD008761>
11. HOOYBERGHS J., C. MENSINCK, G. DUMONT, F. FIERENS and O. BRASSEUR, 2005 : « A neural network forecast for daily average PM10 concentrations in Belgium », *Atm. Env.*, 39, 3279-3289.
12. JANSSEN S., DUMONT G., FIERENS F., and C. MENSINK, 2008 : « Spatial interpolation of air pollution measurements using CORINE land cover data », *Atm. Env.*, 42, 4884-4903.
13. SPORTISSE B., 2008 : « Pollution atmosphérique », Ed. Springer, 345 pp.
14. TERMONIA, P., and A. QUINET, 2004: « A New Transport Index for Predicting Episodes of Extreme Air Pollution », *J. Appl. Meteorol.*, 43, 631-640.
15. U.S. Environmental Protection Agency, 1970 : « Proceedings of symposium on multiple-source urban diffusion models », Air Pollution Control Office Publication No. AP-86.
16. VAUTARD, R., C. HONORE, M. BEEKMANN, L. ROUIL, 2005: « Simulation of ozone during the August 2003 heat wave and emission control scenarios », *Atmospheric Environment*, Volume 39, Issue 16, May 2005, Pages 2957-2967.
17. WELCH G., and G. BISHOP, 2006 : « An introduction to the Kalman Filter », Department of Computer Science, http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf, TR 95-041.



Autres fiches à consulter

Fiches documentées de la thématique « air »

Auteur(s) de la fiche

BRASSEUR Olivier

Date de mise à jour : Janvier 2011