Revue des Energies Renouvelables Vol. 17 N°3 (2014) 465 – 473

Modélisation spatio-temporelle de la dispersion du dioxyde de soufre près d'un site industriel: Réseau de neurone & bio-indicateur

D. Taieb * et E. Hamdi †

Unité de Recherche Thermodynamique Appliquée (99/UR/11-21)
Ecole Nationale des Ingénieurs de Gabès, Université de Gabès
Faculté de Sciences, Bizerte, Laboratoire de Bio-surveillance de l'Environnement

(reçu le 10 Février 2014 – accepté le 30 Septembre 2014)

Résumé - *Cet article propose un état de l'art sur la modélisation basée sur l'utilisation des méthodes numériques et géostatiques pour la prévision spatio-temporelle de la pollution atmosphérique en SO_2 dans la région du Gabès (Sud-Est Tunisien). Les outils numériques en mettant l'accent sur un modèle non linéaire, les réseaux de neurones sont applicables aux données horaires (2012) en SO_2 fournies par l'Agence Nationale de Protection de l'Environnement (ANPE) pour faire la paramétrisation de SO_2 en fonction des facteurs externes, ainsi que la prévision des concentrations horaires d'un jour de plus. Le modèle déterministe et probabiliste est applicable à des concentrations de soufre mesurés dans le matériel végétal; le Malva ageptiaca, un bio-indicateur passif présente l'intérêt d'interpoler et de cartographier le soufre dans la zone d'étude et permet de servir, dans les conditions de perturbation, de signal d'alarme sur la zone la plus polluée*.

Abstract – *This article proposes a state of the art modeling based on the use of numerical methods and geostatistical for spatiotemporal prediction of SO_2 air pollution in the region of Gabes (South-East Tunisia). Digital tools focusing on a nonlinear model, neural networks are applicable to hourly data (2012) SO_2 provided by the National Agency for Environmental Protection (ANPE) for parameterization of SO_2 based on external factors, as well as forecasting hourly concentrations of one more day. The deterministic and probabilistic model can be applied to sulfur concentrations measured in plant material; the Malva ageptiaca, passive bio-indicator is of interest to interpolate and map Sulphur in the study area and helps serve in upset conditions, signal alarm on the most polluted area.*

Mots clés: Pollution atmosphérique - Dioxyde de soufre - Modélisation - Réseaux de neurones artificiels – Cartographie.

1. INTRODUCTION

Le dioxyde de soufre est un composé chimique, un gaz toxique provient de combustion de combustibles fossiles, des rejets industriels et du trafic routier. C'est le produit nuisible le plus important qui a un impact direct et indirect sur l'écosystème et sur l'environnement dans la région du Gabès [1, 3]. La surveillance de la qualité de l'air fait appel à plusieurs moyens, parmi lesquels figure la modélisation. Il existe une multitude de méthodes et de modèles pour la modélisation de la dispersion allant de plus simples aux plus compliqués. Dans cet article, on a développé deux types de simulation. La modélisation numérique non linéaire en utilisant le réseau de neurones artificiel (ANN) qui est largement utilisé en sciences de l'atmosphère [4, 5].

* taieb_dalila@yahoo.fr
† essiahamdi@hotmail.fr

L'architecture de l'ANN est constituée de blocs de base appelés neurones, organisés et interconnectés en une ou plusieurs couches [6]. Et la modélisation spatiale à l'aide de Système d'Information Géographique (SIG) en utilisant un bio-indicateur (espèce végétale). Cette modélisation a pour but de connaître en tout point de l'espace les concentrations en SO_2 et d'estimer leur impact sur l'écosystème.

2. MATERIEL ET METHODES

Les éléments de la base de données utilisée dans cette étude, sont les concentrations horaires de SO_2 et les paramètres météorologiques observés à Gabès en 2012 pour y effectuer la modélisation numérique en utilisant logiciel Matlab (Version 2012). Le matériel végétal utilisé comme bio-indicateur de la pollution atmosphérique est le *Malva ageptiaca*. La modélisation spatiale s'effectue par ArcGIS, version 9.3 (niveau ArcMap).

3. RESULTATS ET DISCUSSIONS

3.1 Modélisation temporelle- Le réseau des neurones

3.1.1 Evolution temporelle des SO_2 observés (2012)

Dans la région de Gabès, le dioxyde de soufre provient essentiellement de l'industrie chimique. La figure 1 représente les graphiques de l'évolution horaire, journalière et mensuelle des concentrations moyennes. Le profil hebdomadaire des concentrations moyennes journalières en SO_2, présente des concentrations plus faibles le week-end, plus particulièrement le dimanche.

Le profil horaire montre que les valeurs les plus élevées ont lieu au voisinage de 18h. A l'échelle mensuelle, Les valeurs les plus dispersées ont été observées pendant les mois de mai, juin, juillet, août et septembre. Certaines conditions météorologiques nuisant à la bonne dispersion des polluants en assurant leur accumulation (l'humidité et vent faible).

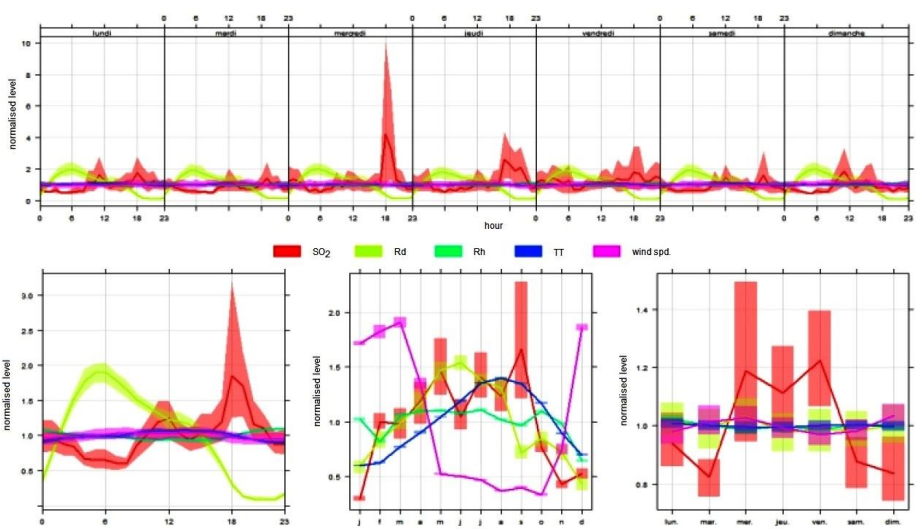

Fig. 1: Evolution temporelle des SO_2 atmosphérique 2012

3.1.2 Paramétrisation de SO_2

La qualité de l'air ne peut être correcte que si certains facteurs de pollution n'interviennent pas. Les conditions atmosphériques telles que l'influence du vent, de la température, de l'ensoleillement, de la pression atmosphérique, de l'humidité ou des précipitations peuvent aggraver les facteurs polluants.

Du fait de leurs propriétés, les réseaux de neurones présentent un intérêt pour établir des paramétrisations de SO_2 à partir des données collectées à la station de mesure (2012) et pour prévoir ces émissions un jour de plus en fonction des paramètres météorologiques prédites un jour de plus (température, vitesse de vent, humidité, précipitations, la pression atmosphérique et rayonnements solaires).

Ici, on travaille avec une architecture de type MPL (Multi-layer Perceptron). Après différents essais, la meilleure MLP est composée de deux couches cachées, avec respectivement 20 et 18 neurones dans chaque couche, le réseau a été formé avec l'algorithme de Levenberg-Marquart, la première unité cachée est activée par la fonction de transfert log-sigmoïde, la seconde par Tang-sigmoïde (Fig. 2).

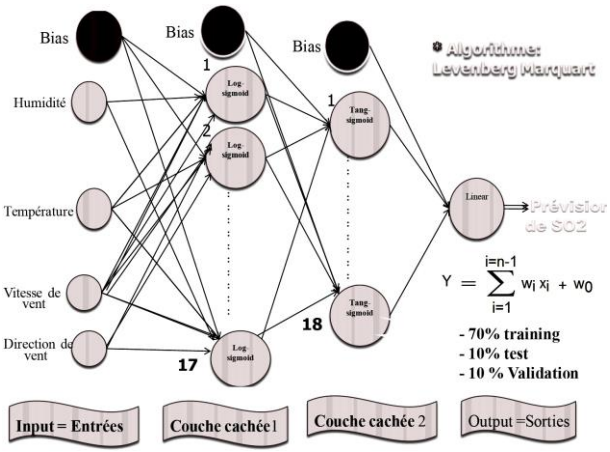

Fig. 2: Architecture de réseau de neurones pour la prévision de SO_2

Pour un réseau reliant les variables d'entrées aux sorties, on déduit que le réseau de neurone dont le coefficient de corrélation linéaire est R= 0.930, est plus performant (Fig. 3).

- Le nuage de point de la régression linéaire (Fig. 3) est relativement dispersé autour de la droite de régression. Le poids de la relation de chacune des variables déterminée par régression linéaire avec les concentrations journalières de SO_2 permet de tirer les remarques suivantes:

- Les températures ambiantes, la vitesse de vent, l'humidité, la radiation solaire fournissent l'essentiel de la variance expliquée.
- La relation entre la vitesse du vent et la concentration de SO_2 est positive. Or un vent fort est un agent efficace de dispersion. Deux éléments peuvent enrayer cette apparente contradiction: premièrement, SO_2 est relativement lourd par rapport à l'air normal, d'où la nécessité d'une certaine turbulence pour empêcher son dépôt au sol. Deuxièmement, les vitesses du vent concernées par cette régression linéaire sont comprises entre 2 et 8 m/s.

- La relation entre la température et le dioxyde de soufre SO_2 de l'air est positive. Il est tout à fait normal qu'à des températures plus élevées correspondent des concentrations de SO_2 plus importantes. Les températures de l'air représentent une expression indirecte de rayonnements solaires donc de l'absence de précipitations.
- De même, il existe une corrélation positive entre l'humidité et le dioxyde de soufre SO_2. En effet, l'humidité influence la transformation des polluants primaires émis, sous l'action de l'humidité (sulfates se forment à partir de dioxyde de soufre).

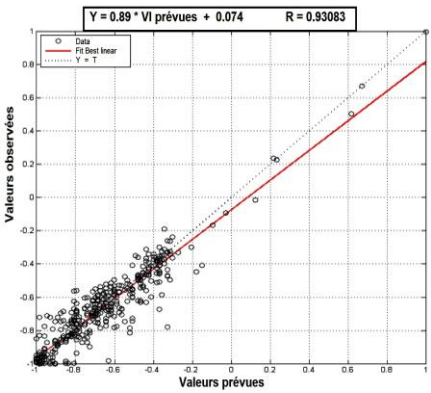

Fig. 3: Nuage des points de la régression linéaire multiple de SO_2 à Gabès au cours de l'année 2012

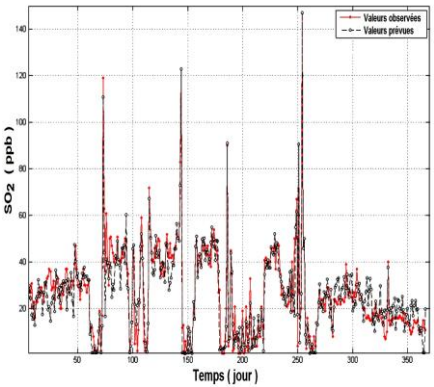

Fig. 4: Valeurs journalières mesurées et prévues de la concentration de SO_2

Bien que préliminaire, cette étude a montré la pertinence de l'utilisation de la prédiction neuronale pour appréhender la pollution atmosphérique en SO_2 dans la région du Gabès. Compte tenu du nombre et de la qualité des données dont nous disposions, les résultats obtenus nous semblent prometteurs.

3.2 Modélisation spatiale

La cartographie permet également de montrer les variations spatiales de soufre dans la région du Gabès en utilisant un bio-indicateur passif le Malva. Cet analyse géostatistique permet de créer des modèles continus à partir de données échantillonnées et de prédire ainsi les valeurs sur des régions où le prélèvement n'a pas pu être effectué [7-9], d'une part et d'autre part, par l'étude d'impact.

L'utilisation des réponses à tous les niveaux d'organisation biologique (moléculaire, biochimique, cellulaire, physiologique, tissulaire, morphologique, écologique) d'un organisme ou d'un ensemble d'organismes pour prévoir et/ou révéler une altération de l'environnement. Dans notre cas, on a choisi 30 échantillons de Malva qui ont été utilisé comme bio-indicateur passif.

Malva *ageptiaca* a été choisie en raison de sa présence dans la région d'étude (présente dans tous les sites d'échantillonnage). Cette espèce peut servir d'indicateur d'effets et d'exposition. En effet, c'est une espèce spontanée (herbe), plante robuste, poussant le long des chemins et dans les friches dont le type biologique défini par Raunkiaer (1937) et Thérophyte (annuelle).

La stratégie d'échantillonnage mise en œuvre pour le prélèvement des échantillons 'in situ', est basée sur l'éloignement de la zone industrielle. Les sites d'étude sont répartis sur des cercles concentriques, ayant pour centre la zone industrielle de Ghannouche (Fig. 5).

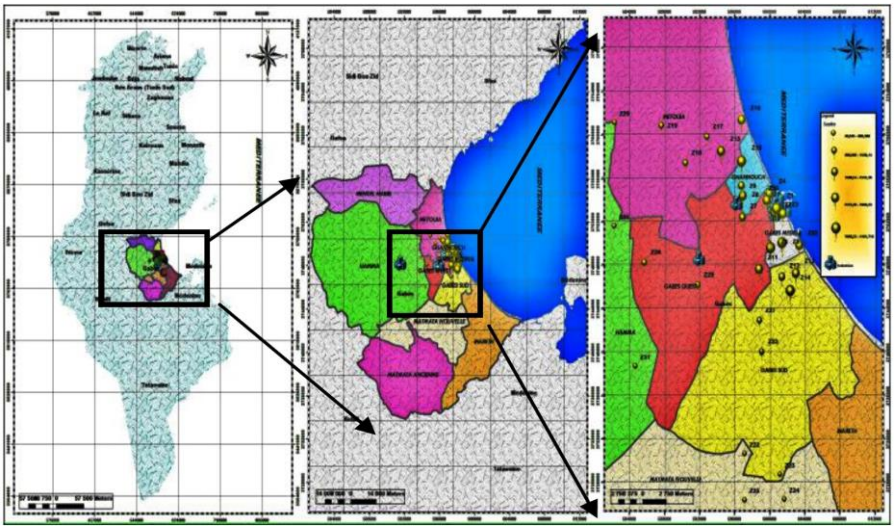

Fig. 5: Représentation spatiale de zone d'étude

Fig. 6: Aperçu des feuilles, près de la zone industrielle

Fig. 7: Aperçu des feuilles, loin de la zone industrielle

En rapprochant des sources d'émissions de SO_2 (Station 1, 2, 3, 6), on a remarqué sur les feuilles des taches foliaires brunes claires à fauve nécrosant les bords du limbe qui se dessèche, les feuilles sont très atteintes présentant un aspect 'grillé'. On parle dans ce cas de la nécrose, (Fig. 6).

En comparant les échantillons près des sources d'émission à ceux les plus loin, on constate plus qu'on s'éloigne de la source de la pollution plus que notre échantillon présent un bon état de santé, (Fig. 7).

3.2.1 Méthode IDW

La méthode d'interpolation inversement proportionnelle à la distance (Inverse Distance Weighting) a été ici testée. Cette méthode se déroule en deux étapes, la première étape est préparer les sites qui vont intervenir dans l'estimation. Dans un deuxième temps, on attribue à chaque site (S_i) retenu un poids inversement proportionnel à distance entre ce site et le point à estimer x_0.

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i \times z(x_i)$$

Où:

$Z(x_i)$ est la valeur mesurée à l'emplacement S_i.

λ_i est le poids de la valeur mesurée dans le $i^{ème}$ emplacement .

$Z(x_0)$ est la valeur estimée à l'emplacement x_0 .

n est le nombre de valeurs mesurées.

3.2.2 Méthode Krigeage (Kriging) ordinaire

Le Krigeage se passe en deux étapes, on calcule dans un premier temps les paramètres de l'entourage et la dépendance statistique des mesures. Ceci détermine la pondération de chaque point pour chaque maille de la carte. On calcule ensuite une prévision pour chaque maille.

Cette méthode utilise donc deux fois les données. Le modèle de base de cette méthode s'énonce comme suit:

$$Z(s) = \mu(s) + \delta(s), \qquad s \in D$$

Avec μ quasi-constante inconnue et $\delta(.)$, fonction aléatoire stationnaire intrinsèque d'espérance nulle et de structure de dépendance connue.

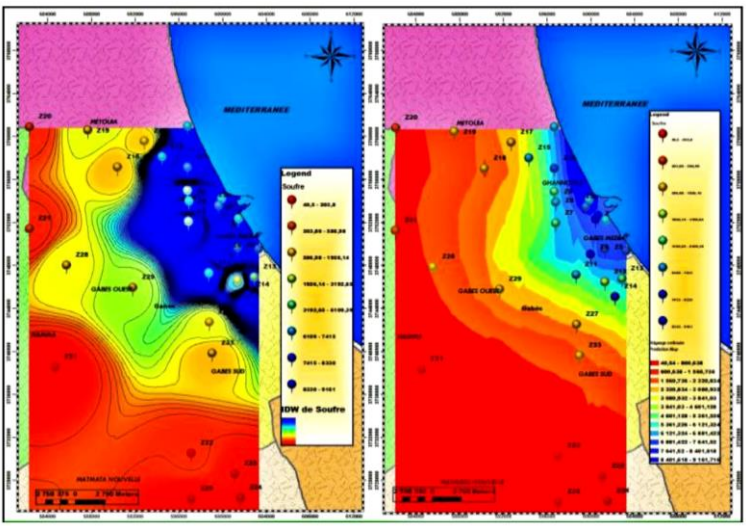

Fig. 8: Comparaison de deux méthodes d'interpolation

Tableau 1: Comparaison entre les deux méthodes d'interpolation

Méthode d'interpolation	IDW Déterministe	Krigeage ordinaire Probabilistique
Pourcentage d'erreur (%)	4.48	1.52

En discutant les deux méthodes d'interpolation, on constate que la dispersion garde à peu près la même répartition spatiale avec les deux méthodes.

D'après les cartes de la prévision, on remarque que SO_2 est plus concentré dans le nord de Gabès Ghannouch Métouia, ce qui ci demande une réflexion. En effet, ces deux villes sont caractérisées par une activité industrielle intense, dans lesquelles se localisent les grandes sources d'émission en particulier le groupe chimique.

Il faut toutefois noter que la direction du vent a tendance à intervenir dans la répartition spatiale de ces éléments. Les vents dominants sont de direction Est / Nord-Est, provenant donc du secteur maritime, ce qui peut expliquer la concentration des polluants dans la direction Est / Nord-Est de Gabès.

L'erreur de l'interpolation par la méthode de Krigeage ordinaire est plus faible que celle d'inverse distance. Donc on peut dire que l'interpolation par le krigeage ordinaire pour notre cas est plus efficace que l'IDW (**Tableau 1**).

La méthode de krigeage a l'avantage de prévoir construire le variogramme expérimental, qui permet de décrire la variabilité moyenne du phénomène dans l'espace (corrélation spatiale). Le krigeage est donc un interpolateur qui prend en compte la structure spatiale du phénomène permettant ainsi d'affiner l'estimation et réduire l'incertitude dans les zones sous-échantillonnées (Fig. 9, 10).

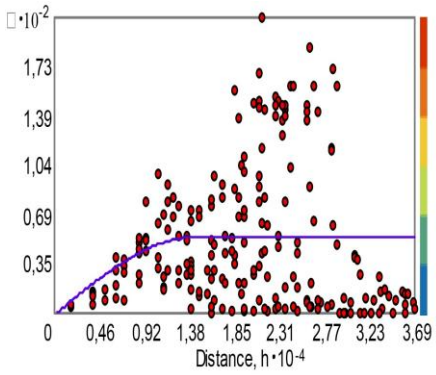

Fig. 9: Covariance

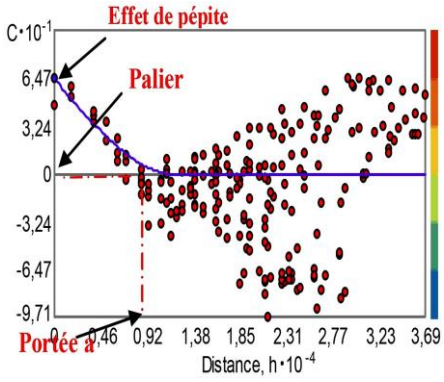

Fig. 10: Semi-variogramme

Le semi- variogramme, est une fonction de h, croissante. C'est un outil principal permettant l'évolution de la semi-variance en fonction de la distance entre les mesures et permet ainsi d'étudier le lien spatial entre les données (Fig. 10). Il est défini de la manière suivante:

$$Y(h) = 1/2 \times \mathrm{Var}\big(Z(s+h) - Z(s)\big) \qquad \forall s \in DZ(.)$$

- Y est la variable régionalisée étudiée.
- s est le vecteur des coordonnées.
- h est le vecteur distance.

- D est le domaine géographique considéré.

Le semi-variogramme est souvent caractérisé par trois paramètres:

- **L'effet de pépite**, correspond à la limite du variogramme en zéro. Elle représente la variation entre deux mesures très proches, qui est dans notre cas égale à zéro. Cela indique une présence de corrélation entre les valeurs prises en deux sites très proches. C'est-à-dire qu'il y a une ressemblance entre les valeurs régionalisées très voisines.

- **Le palier**, valeur du variogramme pour la distance égale à la portée.

- **La portée**, distance où deux observations ne se ressemblent plus du tout. Leur covariance est nulle.

Chaque point rouge de la figure représente deux emplacements. L'axe (y), appelé la valeur du semi-variogramme, est la différence au carré entre les mesures de deux emplacements. L'axe (x) est la distance entre ces deux emplacements.

Le principe de l'auto-corrélation spatiale nous dit qu'une paire d'emplacements proche l'un de l'autre doit avoir des valeurs proches. On remarque que lorsque la distance entre les points augmente, la valeur du semi-variogramme aussi augmente (donc si X augmente, alors Y augmente).

A partir d'une certaine distance, on constate que le nuage de points s'aplatit. Cela signifie que les emplacements ne sont plus corrélés. L'atteinte de la portée indique qu'à partir de 0.91×10^{-4}, il n'y a plus de dépendance spatiale entre les concentrations en soufre dans la région d'étude.

La comparaison entre semi-variogramme et covariance nous permet de constater que Le vecteur h croît pour le semi-variogramme et décroit pour la variance, le palier pour la covariance, est égal à: $\mathrm{cov}(h = 0)\,\mathrm{cov}\ (h = 0)$, hors ce n'est pas le cas pour semi–variogramme.

4. CONCLUSION

En ce qui concerne la prévision temporelle de SO_2 par les réseaux de neurones artificiels (ANN), la performance dépend d'une part du nombre de couches cachées et du nombre de neurones dans chaque couche cachée, ainsi que la fonction d'activation. Et d'autre part des paramètres d'entrée (paramètres météo, NO_2, NO et PM_{10}).

Cette méthode numérique est capable de prévoir la concentration de SO_2 un jour d'avance. Mais elle ne renseigne pas à elle seule sur les effets du/des polluants ni de l'imprégnation de l'environnement.

Ici, le SIG permet à l'aide des modèles des simulations (ArcMap), la modélisation et la cartographie des rejets atmosphériques en SO_2. Dans notre cas, l'utilisation de modèle probabiliste krigeage donne la cartographie la plus performante.

Donc la modélisation neuronale et la modélisation spatiale du dioxyde de soufre SO_2 mesuré est d'une importance particulière pour bien comprendre la distribution spatio-temporelle de ces émissions par rapport à la source d'émission, et en quelque sorte les conditions qui contrôlent leurs répartitions (vent, localisation par rapport à la source de la pollution).

Ces informations sont fondamentales du fait qu'elles jouent un rôle essentiel dans le contrôle et la surveillance de la santé de l'environnement.

Remerciements- A l'issue de ce travail, nous tenons tout particulièrement à remercier l'Agence Nationale de Protection de l'Environnement, Département Contrôle et Suivi de la Pollution et le Laboratoire de traitement de déchets du Centre International des Technologies de l'Environnement de Tunis (CITET).

REFERENCES

[1] D. Taïeb and A. Ben Brahim, '*An Eco-Friendly Process for Cleaner Hydrogen Production: Description and Evaluation'*, Global Journal on Technology and Optimization, Vol. 3, N°2, pp. 122 – 131, 2012.

[2] D. Taïeb and A. Ben Brahim, '*Methodology for Developing an Air Quality Index (AQI) for Tunisia*', International Journal of Renewable Energy Technology, Vol. 4, N°1, pp. 88 – 100, 2013.

[3] D. Taïeb and A. Ben Brahim, 'E*lectrochemical Method for Sulphur Dioxide Removal from Flue Gases: Application on Sulphuric Acid Plant in Tunisia*', Comptes Rendus Chimie, Vol. 16, N°1, pp. 39 – 50, 2013.

[4] M.W. Gardner and S.R. Dorling, '*Artificial Neural Networks (the Multilayer Perceptron) - A Review of Applications in the Atmospheric Sciences*', Atmospheric Environment, Vol. 32, N°14-15, pp. 2627 – 2636, 1998.

[5] W. Gonzalez-Manteiga, J.M. Prada-Sanchez, R. Cao, I. Garcia-Jurado, M. Febrero-Bande and T. Lucas-Dominguez, '*Time-Series Analysis for Ambient Concentrations*', Atmospheric Environment, Vol. 27, N°2, pp. 153 – 158, 1993.

[6] D.E. Rumelhart, J.L. McClelland and the PDP Research Group, '*Parallel Distributed Processing, Exploration in the Microstructure of Cognition*', Vol. 1, MIT Press, Cambridge, 1986.

[7] M. Arnaud et X. Emery, '*Estimation et Interpolation Spatiale – Méthodes Déterministes et Méthodes Géostatistiques*', Hermès Science, 221 p., 2000.

[8] E. Isaaks and R. Srivastava, 'An Introduction to Applied Geostatistics', Oxford University Press, New York, 1990.

[9] B. Denby, J. Horálek, S. Walker, K. Eben and J. Fiala, '*Interpolation and Assimilation Methods for European Scale Air Quality Assessment and Mapping. Part I. Review and Recommendations*', Tech. Rep. 2005/7. The European Topic Centre on Air and Climate Change, 2005.

[10] EPA, '*Developing Spatially Interpolated Surfaces and Estimating Uncertainty*', Tech. Rep. EPA-454/R-04-004. US-EPA, 2004.