



Prévision du NO₂ en utilisant la méthode du réseau de neurones

By/Par | **Soulaima Abdelli, Kaouther Noura, Abdelwahed Trabelsi**

Business and Economic Statistics Modeling Laboratory
Institut Supérieur de Gestion de Tunis

ABSTRACT

This article describes a procedure for modeling the phenomenon of dispersion of the of nitrogen dioxide concentration (NO₂) using the technique of multilayer perceptron (MLP). Our goal is to prove that the concentration of NO₂ is an autoregressive variable explained by meteorological variables. To do this, we proceeded through three stages: in the first stage we used only NO₂ concentration variable, in the second stage we used the meteorological variables only, in the third step we used the concentration of NO₂ in combination with the meteorological variables. The results showed that the third case is better than the other cases that prove our hypothesis.

Keywords: NO₂, forecasting, MLP

RÉSUMÉ

Cet article décrit une procédure de modélisation du phénomène de dispersion de la concentration du dioxyde de nitrogène (NO₂) en utilisant la technique du perceptron multicouche (MLP). Notre objectif est de prouver que la concentration du NO₂ est une variable autorégressive et expliquée par des variables météorologiques. Pour ce faire, nous avons procédé par trois étapes : dans la première étape nous avons utilisé la variable concentration NO₂ uniquement, dans la seconde étape nous avons utilisé les variables météorologiques uniquement et dans la troisième étape nous avons utilisé la concentration du NO₂ combinée avec les variables météorologiques. Les résultats ont montré que le troisième cas est plus performant que les deux autres ce qui prouve notre hypothèse.

Mots-clés: NO₂, prévision, MLP

JEL classification: Q53

INTRODUCTION

Le dioxyde de nitrogène (NO₂) est le résultat de la réaction chimique de l'oxyde de nitrogène (NO_x), provenant de l'émission des véhicules ainsi que d'autres sources industrielles, avec l'ammoniac et l'humidité. Le NO₂ est l'un des polluants d'air qui représente une préoccupation environnementale car il a un impact sur la santé et l'agriculture d'une part, et la production d'autres polluants secondaires tels que l'ozone et les particules en suspension. D'où la nécessité de suivre ce polluant et de prévoir son évolution au cours du temps. La méthode choisie dans ce papier est le réseau de neurones artificiels et plus exactement le perceptron multicouche (en anglais *Multi-Layer Perceptron* ou MLP) qui est la méthode la plus utilisée dans la prévision des polluants d'air. Notre objectif est de prévoir le NO₂ un jour à l'avance et surtout voir si le changement des variables météorologiques a un effet sur sa production. Nous cherchons donc à démontrer que la concentration du NO₂ passée est une série temporelle uni-variée autorégressive expliquée par les variables météorologiques. Pour ce faire, nous avons développé trois MLPs, le premier (appelé MLP1) aura comme input seulement la série temporelle de NO₂, le second (appelé MLP2) aura seulement les variables météorologiques et le dernier (appelé MLP3) aura la série de NO₂ ainsi que les variables météorologiques. Cet article sera structuré comme suit : La première section sera réservée à la revue de la littérature des modèles de prévision des polluants d'air, la deuxième section définira la méthodologie et la troisième décrira les résultats trouvés.

TECHNIQUES DE PRÉVISION DES POLLUANTS D'AIR

Dans la littérature, les réseaux de neurones ont trouvé un grand succès dans la modélisation et la prévision des polluants d'air.

(Perez et Trier, 2001) ont proposé de prévoir la concentration de NO₂ et de l'oxyde nitrique NO à Santiago en se basant sur les variables météorologiques et en utilisant les MLPs, la méthode de régression linéaire et la méthode persistance. Les résultats ont montré que le réseau de neurones est la méthode qui réalise l'erreur de prévision la moins élevée par rapport aux autres méthodes.

(Kukkonen et al., 2003) utilisent 5 modèles de réseaux de neurones (les réseaux de neurones avec un seul bruit gaussien hétéroscédastique (**NN-HeG**), les réseaux de neurones avec une combinaison de deux bruits gaussiens hétéroscédastiques (**NN-2HeG**), les réseaux de neurones avec une combinaison de trois bruits gaussiens hétéroscédastiques (**NN-3HeG**), les réseaux de neurones homoscedastiques avec un bruit gaussien (**NN-HoG**) et les réseaux de neurones avec un bruit de distribution laplacienne (**NN-Lapl**)), un modèle statistique linéaire (**LIN**) et un modèle déterministe (**DET**) pour prévoir la concentration de NO₂ et PM₁₀ (Particules en suspension ou (*Particule Matters*) de diamètre inférieur à 10 microns) dans deux stations différentes et durant la période allant de 1996 à 1999. Il a été démontré que les modèles de réseaux de neurones sont plus performants que le modèle statistique linéaire et le modèle déterministe. Les modèles de réseaux de neurones hétéroscédastiques sont plus performants que les modèles homoscedastiques.

(Nagendra et Khare, 2006) utilisent le modèle de réseaux de neurones stochastique déterministe qui s'appelle le ANN-VEE (*Véhicule Exhaust Emission*) pour modéliser le phénomène de dispersion de la série temporelle journalière NO₂ à partir de l'émission des

véhicules dans deux stations différentes à Delhi dans les routes urbaines. Les modèles de NO₂ basés sur les ANN ont été formulés en utilisant trois choix d'inputs : le premier choix consiste à choisir comme inputs les variables trafiques et météorologiques ; le deuxième choix consiste à choisir seulement les variables météorologiques, et le dernier ne considère que les variables trafiques. Les résultats montrent que le modèle formé par les variables météorologiques et trafiques est plus performant dans les deux stations en question.

(Al-Alawi et al., 2008) proposent une nouvelle méthode pour prévoir l'ozone, il s'agit de la méthode qui utilise le réseau de neurones ANN combiné avec la méthode de régression de composantes principales PCR. D'après les coefficients de corrélation, il a été démontré que la prévision combinée (ANN et PCR) améliore la précision et fournit une meilleure prévision que les modèles ANN et PCR utilisés séparément.

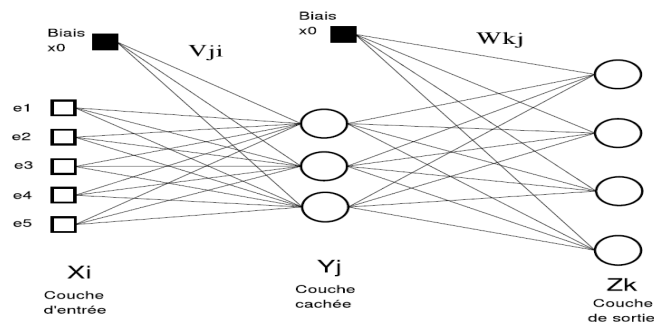
(Ping Shi et Harisson, 1997) ont développé un modèle de moindre carrée ordinaire (en anglais *OLS : ordinary least square*) et un modèle autorégressif de premier ordre (en anglais autorégressif) (AR(1)) pour analyser la régression et la prévision du NO_x et la concentration du NO₂. Les résultats ont montré que le AR (1) est plus efficace que la méthode (OLS). En nous basant sur la littérature nous allons essayer de démontrer que notre série temporelle NO₂ est une série autorégressive, univariée expliquée par des variables météorologiques.

MÉTHODOLOGIE

Le réseau de neurones est un modèle de calcul dont la conception est schématiquement inspirée du fonctionnement des neurones biologiques du cerveau humain vu la grande capacité de stockage d'information et la connexion inter-neurone appelée poids synaptique pour stocker les connaissances (Haykin, 1998). Les modèles de réseau de neurones artificiels sont des fonctions d'approximations universelles entre les valeurs prédites et les valeurs observées utilisées pour schématiser une fonction non linéaire entre les vecteurs d'inputs et les vecteurs d'outputs qui ont une relation inconnue donc difficile à adapter durant la phase d'apprentissage (Darbellay et Slama, 2000) sans les hypothèses prioritaires sur les données d'origine (Boznar et al., 1993). Le perceptron est une couche de neurone où les poids et les biais peuvent être formés pour produire un vecteur cible (*Target*) correct c'est-à-dire plus proche de l'output (Rosenblatt, 1957). Les poids et les signaux d'output assurent cette interconnexion. Sur chaque couche il y a un différent nombre de nœuds ou de neurones. Tous les neurones dans chaque couche sont joints à d'autres neurones dans la couche suivante. (Stanley, 1990) montre l'importance de la connexion de neurones sur la performance du réseau. Le réseau de neurones est constitué de plusieurs couches. La première couche ou la couche la plus basse est la couche d'input où l'information externe est reçue puis transmise aux neurones des couches cachées sans calcul. Les couches cachées (couches intermédiaires) qui séparent entre la couche d'entrée et la couche de sorties reçoivent des signaux de tous les neurones de la couche précédente. La dernière couche est la couche de sortie où la solution du problème (valeur prédite) est obtenue. Les neurones de sorties renvoient les informations à l'environnement extérieur. Dans notre cas nous allons utiliser les MLPs dont la propagation des données se fait des inputs vers les outputs, il s'agit de la « propagation avant ».

La fonction d'activation appelée aussi fonction de transfert s'intéresse à la relation entre les entrées et les sorties des nœuds du réseau. Les choix les plus habituels pour la fonction d'activation sont les fonctions binaires (les fonctions d'activation à seuil), les fonctions linéaires, la fonction d'identité, les fonctions hyperboliques ou sigmoïdes. La figure 1 montre un exemple de réseau de neurones MLP à une couche cachée avec 5 entrées, 3 neurones dans la couche cachée et 4 sorties.

Figure 1. Exemple de réseau de MLP à une couche cachée



Source : [CHATELAIN C., 2003]

Propagation des entrées : Les entrées e_i sont présentées à la couche d'entrée de neurones x_i ensuite ils sont propagés vers la couche cachée où ils subissent le traitement selon l'équation (1):

$$Y_j = f\left(\sum_{i=1}^m x_i v_{ij} + x_0\right) \quad (1)$$

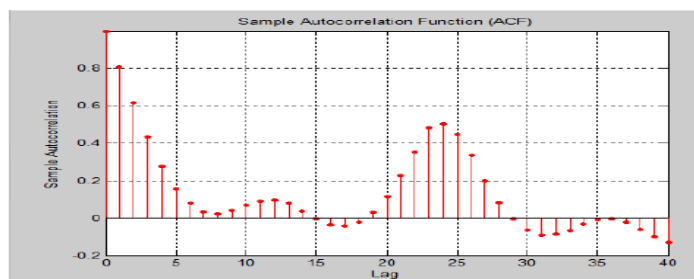
Où f est la fonction de transfert et v_{ij} les poids des synapses (x_i, y_j) . Puis de la couche cachée vers la couche de sortie selon l'équation (2):

$$Z_k = f\left(\sum_{j=1}^n y_j w_{kj} + y_0\right) \quad (2)$$

Où w_{kj} sont les poids des synapses (y_j, z_k) et les valeurs x_0 et y_0 sont les biais.

Dans le but de démontrer que notre série est une série autorégressive et expliquée par les variables météorologiques, nous avons développé trois MLPs : MLP1 basé sur la série temporelle NO₂, MLP2 basée sur les variables météorologiques et MLP3 sur la régression de la série NO₂ ainsi que les variables météorologiques. En calculant la fonction d'autocorrélation (voir figure 3), nous avons constaté que la série NO₂ est une série saisonnière (de saisonnalité $s = 24$) et autorégressive de retard $P = 1$. Pour déterminer le meilleur MLP, nous avons procédé par une variation du nombre de couches cachées, du nombre de neurones par couche cachée, des fonctions de transfert (aussi bien pour les neurones cachés que les neurones de sortie) ainsi que des algorithmes d'apprentissage.

Figure 2. Fonction d'autocorrélation pour le NO₂



DESCRIPTION DES DONNÉES

Notre base de données provient du Réseau National de surveillance de la Qualité de l'Air (RNSQA) et elle est composée des :

Mesures horaires de la concentration du NO₂, dans la région de Bab Alioua, durant la période du 1/1/2008 au 31/12/2008.

Mesures horaires des variables météorologiques, de la même région, telles que la vitesse du vent, la direction du vent, l'humidité relative et la température durant la période du 2/1/2008 au 1/1/2009.

Les valeurs manquantes sont calculées par la méthode de prédécesseur, elles seront donc remplacées par les valeurs disponibles qui les précèdent. Les entrées et les sorties sont normalisés à un intervalle de [0 1] grâce à une transformation linéaire (Lapedes et Farber, 1988) selon l'équation (3) :

$$X_n = (x_0 - x_{\min}) / (x_{\max} - x_{\min}) \quad (3)$$

RÉSULTATS ET DISCUSSION

La base d'apprentissage et la base de test

La première option est reliée à la répartition de la base de données en deux bases : base d'apprentissage et base de test. D'après le calcul de l'erreur quadratique moyenne

(MSE :mean square error) dans le tableau 1, nous avons remarqué que le meilleur pourcentage pour la base d'apprentissage est de 70% pour les MLP1 et MLP2 et de 80% pour MLP3.

Tableau 1. La répartition de l'ensemble des données

	MSE1	MSE2	MSE3
Base d'apprentissage : 90% Base de test : 10%	0.0140	0.01383	0.01099
Base d'apprentissage : 80% Base de test : 20%	0.0141	0.01377	<u>0.01074</u>
Base d'apprentissage : 70% Base de test : 30%	<u>0.0132</u>	<u>0.01237</u>	0.01087

Le nombre de couches cachées

D'après le calcul de MSE, il s'est avéré que le meilleur choix est d'utiliser **une seule couche cachée** par rapport à l'utilisation de plusieurs. Dans le tableau 2 nous avons présenté les calculs des MSE pour une, deux, trois et quatre couches cachées et nous avons remarqué que l'augmentation du nombre des couches cachées augmente la charge des calculs sans aucun gain de performance.

Tableau 2. Le nombre de couches cachées (station Bab Alioua)

Nombre de couches cachées	MSE 1	MSE2	MSE3
1	<u>0.0132</u>	<u>0.0130927</u>	<u>0.0107</u>
2	0.0137	0.01556	0.0908
3	0.09045	0.0097	0.0896
4	0.0143	0.0088	0.091

Sélection de la meilleure fonction d'activation

Il y a plusieurs fonctions d'activation dont les plus importantes sont :

- La fonction d'activation **linéaire**.
- la fonction **sigmoïde** ou la fonction log-sigmoïde.
- La fonction **tangente hyperbolique** ou bipolaire.

Pour connaître la fonction correspondante pour chaque couche, nous avons testé différentes combinaisons dans la couche cachée et la couche de sortie du réseau en utilisant la meilleure

répartition de la base de données avec une seule couche cachée dans chacun des cas (voir tableau 3).

Tableau 3. La fonction d'activation dans la couche cachée et dans la couche de sortie

<i>Couche cachée</i>	<i>Couche de sortie</i>	<i>MSE 1</i>	<i>MSE2</i>	<i>MSE3</i>
Sigmoïde	Sigmoïde	0.0138742	0.558133	0.558133
Sigmoïde	Linéaire	0.0138253	0.0139153	0.0113
Sigmoïde	Hyperbolique	0.09015	0.01299	0.0104
Hyperbolique	Sigmoïde	0.556685	0.557177	0.588147
Hyperbolique	Linéaire	0.0139132	0.0127	0.0108
Hyperbolique	Hyperbolique	0.0138711	0.000757	0.0103
Linéaire	Sigmoïde	0.5601	0.5574	0.56083
Linéaire	Hyperbolique	0.0159	0.0178	0.0150
Linéaire	Linéaire	0.0147	0.0186	0.0140

Les résultats ont montré que les meilleures combinaisons pour les fonctions d'activation dans respectivement la couche cachée et la couche de sortie sont :

- **Sigmoïde-linéaire** pour MLP1.
- **Hyperbolique-hyperbolique** pour MLP2.
- **Hyperbolique-Hyperbolique** pour MLP3.

Le nombre de neurones dans la couche cachée

Pour définir le meilleur nombre de neurones dans la couche cachée, nous avons fait varier ce nombre entre un et vingt en choisissant la meilleure combinaison de couche ainsi que la meilleure répartition dans chaque cas et nous avons trouvé que le meilleur nombre de neurones dans la couche cachée est de (voir tableau 4):

- 17 neurones pour MLP1.
- 20 neurones pour MLP2.
- 18 neurones pour MLP3.

Tableau 4. Le nombre de neurone dans la couche cachée

Nombre de neurone dans la couche caché	MSE1	MSE2	MSE3
1	0.0140325	0.0166098	0.0134571
2	0.0140328	0.0171616	0.0129733
3	0.0140226	0.0146401	0.0118572
4	0.0907487	0.0907487	0.0116398
5	0.0139674	0.0139819	0.0115531
6	0.0139683	0.0148949	0.011558
7	0.0139275	0.0138017	0.0111436
8	0.0139275	0.013762	0.0115156
9	0.013969	0.0139714	0.0109489
10	0.0139388	0.0136969	0.0109604
11	0.0139397	0.0143774	0.0107569
12	0.0139545	0.0135508	0.0114806
13	0.0139478	0.0136772	0.010724
14	0.0139266	0.0133577	0.0109763
15	0.0139376	0.014456	0.0105406
16	0.0139155	0.0132524	0.0105367
17	0.0139151	0.013732	0.0108216
18	0.0139264	0.0129423	0.0103721
19	0.0139787	0.0133382	0.0103395
20	0.013916	0.0127146	0.01043

Sélection de l'algorithme d'apprentissage optimum

Les algorithmes d'apprentissage les plus utilisés sont :

- L'algorithme à descente la plus forte par paquet (*the batch steepest descent training function*).
- L'algorithme à descente la plus forte à un taux d'apprentissage adaptatif (*The steepest descent with adaptive learning rate*).
- L'algorithme à descente la plus forte avec un moment (*the steepest descent with momentum*).
- L'algorithme à plus forte pente avec taux d'apprentissage adaptatif et un moment (*the steepest descent with adaptive learning rate and momentum*).
- L'algorithme levenberg-Marquardt (*The levenberg-Marquardt algorithm*).

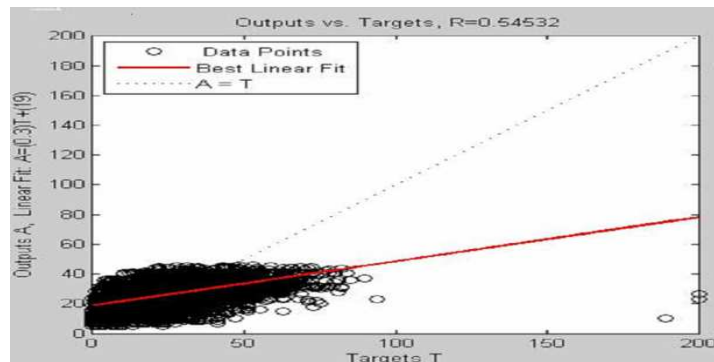
Tableau 5. Le choix de l'algorithme

Algorithme	Paramètres / Valeurs		MSE1	MSE2	MSE3
	Lr				
A descente la plus raide par paquet	0.01		0.099	0.146	0.0472
	0.1		0.0191	0.0249	0.032
A descente la plus raide à un taux d'apprentissage adaptatif	0.01		0.0257	0.0429	0.033
	0.1		0.0191	0.0324	0.0227
	Lr	Mc			
A descente la plus raide avec un moment	0.01	0.9	1.0921	0.309	0.050
	0.1	0.9	2.075	0.950	0.0869
	0.01	0.3	0.042	0.1712	0.0606
	0.1	0.3	0.0215	0.03756	0.0747
	0.01	0.6	0.050	0.123	0.0826
	0.1	0.6	0.0219	0.041	0.0495
A plus forte pente avec taux d'apprentissage adaptatif et un moment.	0.01	0.9	0.0173	0.58	0.057
	0.1	0.9	0.0160	0.0212	0.0176
	0.01	0.6	0.0193	0.036	0.025
	0.1	0.6	0.021	0.0273	0.07758
	0.01	0.3	0.02149	0.0273	0.0775
	0.1	0.3	0.02158	0.0259	0.0400
Levenberg-Marquardt			<u>0.01384</u>	<u>0.0132</u>	<u>0.0113</u>

Les trois cas (MLP1 ,2 et 3) montrent que l'algorithme **levenberg-marquardt** est le meilleur algorithme avec les erreurs quadratiques moyennes minimales.

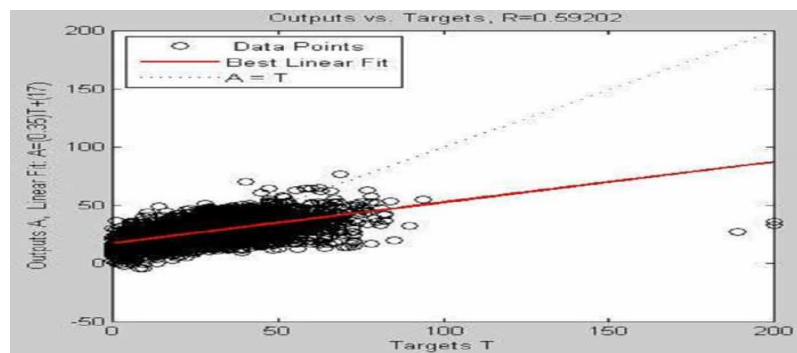
Nuage de points et concentrations des séries chronologiques des valeurs mesurées et prédites

Figure 3. Coefficient de corrélation linéaire au premier cas MLP1



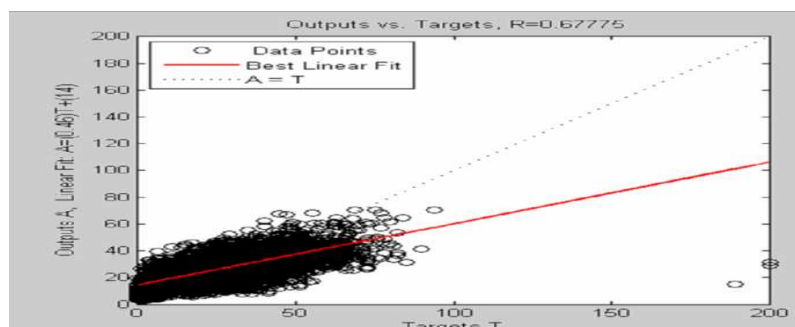
Le coefficient linéaire de corrélation qui décrit le rapprochement de l'output (valeur prédite) du target (valeur observée) pour le premier cas MLP1 est $R_1 = 0.54$ (voir figure 3). La valeur de ce coefficient montre que ce modèle n'est pas performant.

Figure 4. Coefficient de corrélation linéaire au deuxième cas MLP2



Le coefficient linéaire de corrélation pour le deuxième cas MLP2 est $R_2 = 0.59$. Il est donc plus performant que le premier modèle (voir figure 4).

Figure 5. Le coefficient de corrélation au troisième cas MLP 3



Le coefficient de corrélation linéaire pour le dernier cas MLP3 est $R3 = 0.67$. On déduit que le MLP3 qui utilise comme input la concentration passée du NO₂ et les variables météorologiques est plus performant que les deux MLPs qui utilisent seulement la concentration du NO₂ passée ou bien les variables météorologiques.

CONCLUSION

Dans ce travail nous avons utilisé les réseaux de neurones pour démontrer que la concentration du NO₂ est une variable qui n'agit pas seule mais est expliquée par d'autres variables telles que la vitesse du vent, la direction du vent, la température et l'humidité. Pour ce faire, nous avons développé trois perceptrons multicouches: le MLP1 qui ne prend comme input que la concentration passée du NO₂, le MLP2 qui ne prend que les variables météorologiques et le MLP3 qui prend la combinaison des variables météorologiques et la concentration passée de NO₂. Nous avons remarqué d'après les résultats que le coefficient de corrélation de MLP3 est le meilleur ($R3 = 0.67$). Bien que nous sommes persuadés que les variables de trafic peuvent expliquer aussi cette concentration et par conséquent améliorer la performance des MLPs.

RÉFÉRENCES

- Al Alawi, S.M, Abdul-Wahab, S.A, Bakheit, C. S. (2008). 'Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone', *Environmental Modelling & Software*, 23: 396-403.
- Bonn. (2003). 'Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide', Report on a WHO Working Group, Germany, pp.13–15.
- Boznar, M., Lesjak, M. et Mlakar, P. (1993). 'A neural network based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain', *Atmospheric Environment*, 27(2): 221 -230.
- Chatelain, C. (2003). *Les Multi-layer perceptrons*.
- Darbelay, A., Slama, M. (2000). 'Do neural networks stand a better chance?', *International Journal of Forecasting*, 16: 71–83.

- Haykin, S. (1998). 'Feedforward neural networks: An introduction'
- Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G. (2003). 'Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations compared with a deterministic modelling system and measurements in central Helsinki', *Atmospheric Environment*, 37: 4539–4550.
- Lapedes, A., Farber, R. (1988). 'How neural nets work'. In: Anderson, D.Z, *Neural Information Processing systems*, American Institute of physics, p.442-456.
- Moutarde, F. (2008). *Introduction aux réseaux de neurones*
- Nagendra, S.M., Khare, M. (2006). 'Artificial neural network approach for modeling nitrogen dioxide dispersion from vehicular exhaust emissions', *Ecological Modelling*, 190: 99-115.
- Perez, P., Trier, A. (2001). Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago, *Atmospheric Environment*, 35: 1783-1789.
- Shi, JP. et Harrison, R.M. (1997). 'Regression modelling of hourly NO_x and NO₂ concentrations in urban air in London', 31(24): 4081-4094.
- Stanley, C., Ahalt, Ashok, K., Krishnamurthy, P., Chen, Douglas, E., Melton (1990). 'Competitive learning algorithms for vector quantization', *Neural Networks*, 3: 277-290.
- Wang, D., Lu, W. (2006). 'Ground-level ozone prediction using multilayer perception trained with an innovative hybrid approach', *ecological modelling*, 41(16): 3502-3513.