

Apprentissages automatiques pour le monitoring énergétique et environnemental dans le cadre d'une chaudière de régénération (procédé papetier Kraft)

Matthieu Sainlez

CRISIA - Centre de Recherche de l'Institut Supérieur Industriel d'Arlon
Haute Ecole Robert Schuman - Chemin de Weyler 2, B-6700 Arlon, Belgique
matthieu.sainlez@hers.be

L'objectif de cette recherche est de fournir aux industriels confrontés à une gestion énergétique complexe des outils performants d'aide à la décision. En partant des données collectées sur les sites de production, nous mettons en oeuvre des outils innovants de modélisation et d'analyse prédictive. Ces historiques de mesures constituent une source d'information hautement valorisable, particulièrement dans le cadre de la construction de modèles prédictifs de performances énergétiques et/ou de monitoring des polluants atmosphériques. Le marché du papier étant particulièrement préoccupé par le coût de l'énergie et par les contraintes environnementales, le développement d'un prototype de ces solutions se fait en collaboration avec une papeterie installée en région wallonne.

Nous introduisons brièvement ci-après diverses méthodologies appliquées dans le cadre de ce monitoring de polluants atmosphériques (i.e., les émissions d'oxyde d'azote) et d'indicateurs énergétiques (i.e., le tonnage vapeur HP). Les principales applications de la recherche sont détaillées dans cinq publications liées au projet (reprises ci-après).

1 Préambule

Le procédé Kraft est un procédé chimique de fabrication de pâtes à papier. Il est basé sur la dissolution de la lignine et son extraction des parois des fibres du bois par une solution aqueuse de $NaOH$ et de Na_2S (la liqueur blanche). La régénération consiste alors à retransformer la liqueur noire issue de la cuisson des copeaux en liqueur blanche. Ces liqueurs noires sont concentrées par évaporation et calcinées sous atmosphère réductrice dans la chaudière de régénération. Celle-ci constitue généralement la source la plus importante d'émissions atmosphériques polluantes: NO_x , SO_2 , CO , poussières et TRS (soufre réduit total). Le corps de la chaudière étudiée est composé de deux parties principales: la partie basse, en atmosphère réductrice, régénère les minéraux de la liqueur noire alors que l'oxydation complète des différents composés organiques s'effectue dans la partie supérieure (selon quatre étages d'admission d'air). La liqueur noire est injectée dans la partie inférieure, ainsi que du fuel lourd et des gaz non condensables (NCG).

A ce jour, le développement d'outils performants d'analyse de données connaît un essor considérable dans le milieu industriel, particulièrement en ce qui concerne la gestion énergétique et environnementale via l'analyse des historiques de données. Par exemple, la modélisation des NO_x ou du tonnage vapeur HP y est particulièrement complexe. Parmi les méthodes de modélisation existantes, les méthodes par arbres et les réseaux de neurones artificiels peuvent être considérés par les ingénieurs comme des alternatives intéressantes aux méthodes classiques de modélisation dans le cadre des processus de combustion. Elles ont une capacité d'apprentissage automatique, une gestion efficace des données bruitées et une aptitude à traiter des problèmes de modélisation non linéaire.

2 Modélisation par forêts aléatoires

Une covariable est une variable qui joue un rôle explicatif dans un modèle. L'algorithme Random Forests (Forêts Aléatoires) fait partie des méthodes d'ensemble qui cherchent à améliorer les prédictions obtenues à l'aide de différents modèles par combinaison ou pondération. Une propriété intéressante de ce modèle est qu'il peut classer les covariables selon leur importance dans la prédiction en sortie du modèle.

2.1 Segmentation par arbres: principes généraux

La technique de l'arbre de décision est l'une des plus populaires du data mining. En effet, elle est intuitive et fournit des règles explicites de classement. De plus, elle supporte tant les données hétérogènes, les valeurs manquantes et les effets non linéaires. D'abord, on choisit la variable qui sépare le mieux les individus de chaque classe, de façon à avoir des sous-populations (que l'on appelle *noeuds*), contenant chacune le plus possible d'individus d'une seule classe. Puis, on réitère l'opération sur chaque nouveau noeud obtenu jusqu'à ce que la séparation des individus ne soit plus possible.

Le schéma ci-après (Fig.1) illustre le partitionnement récursif (et binaire) de l'espace des variables d'entrées x_1 et x_2 . Premièrement, on divise l'espace des variables de départ en deux régions selon que $x_1 \leq t_1$ ou $x_1 \geq t_1$, où t_1 est un paramètre du modèle. Ce qui crée deux sous-régions dont chacune peut être (re)divisée indépendamment l'une de l'autre. Cette division récursive est alors représentée explicitement par le parcours d'un arbre binaire. A partir d'une nouvelle entrée \mathbf{x} , on peut alors déterminer dans quelle région R_i , celle-ci sera prédite et ce via une règle de décision fournie par l'arbre binaire.

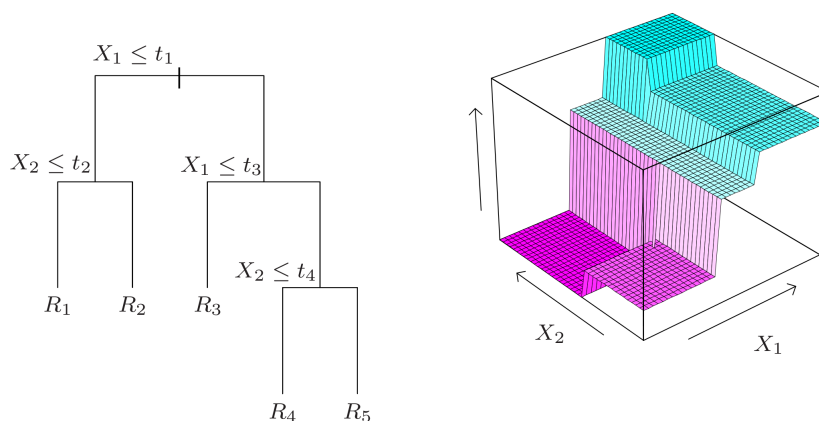


Figure 1: Partitionnement récursif de l'espace des variables d'entrées et arbre correspondant (Hastie, 2010)

Leur principal défaut réside donc dans leur manque de robustesse, mais ceci peut être surmonté avec un rééchantillonnage: on construit alors des arbres sur plusieurs échantillons successifs et on les agrège par un score ou une moyenne. On parle alors de bagging ou de boosting.

2.2 Agrégation de modèles: Bagging et Random Forests

Un principe appliqué dans le cadre du Bagging (Bootstrap Aggregating) recourt à la notion de bootstrap qui consiste à tirer aléatoirement avec remise m échantillons de taille n dans l'ensemble des n données de départ. On obtient alors m modèles différents: appliqués à une observation \mathbf{x} , ils fournissent m prédictions différentes de la sortie y . Le bagging consiste à prendre la moyenne des m prédictions (ou le vote majoritaire si on est en classification), celle-ci a donc une variance inférieure à celle de la prédiction initiale. La méthode Random Forests (Breiman, 2001) part de ce principe de bagging mais en tentant de l'améliorer. Ici, il s'agit de tirer un échantillon bootstrap puis de construire un arbre sur cet échantillon avec randomisation des variables: la recherche de chaque noeud optimal est précédée d'un tirage aléatoire d'un sous-ensemble de k prédicteurs (d'où le terme $TREE_k$ sur Fig.2).

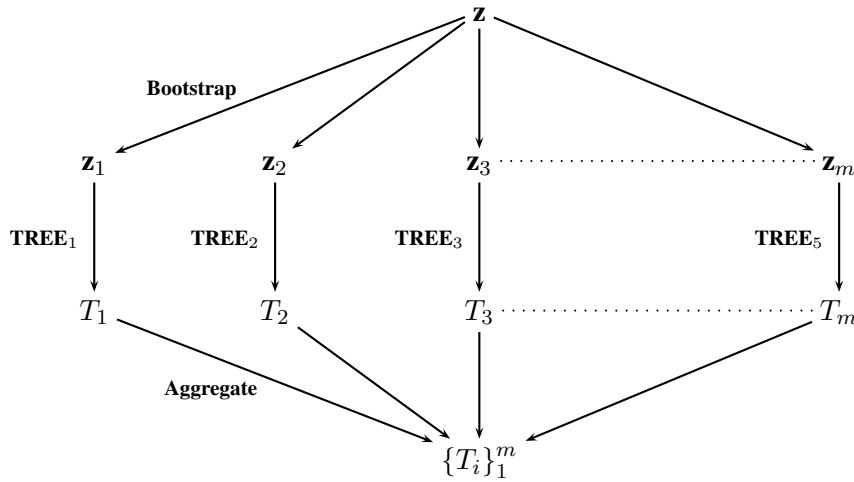


Figure 2: Random Forests: construction de l'ensemble d'arbres et agrégation.

Pour chaque observation $z_i = (x_i, y_i)$, on construit son prédicteur RF en moyennant seulement les arbres correspondants à des échantillons bootstrap dans lesquels z_i n'apparaît pas. A chaque coupure nodale dans un arbre, l'amélioration du critère de coupure est équivalente à la mesure d'importance attribuée à la variable de coupure. Cette mesure est accumulée à travers tous les arbres de la forêt aléatoire, et ce, séparément pour chaque variable. L'algorithme RF va donc passer en revue toutes les variables et permettre de les classer en fonction de leur importance. A titre d'exemple, la figure ci-après (Fig.3) reprend la hiérarchie des covariables établie pour la prédiction des NO_x en partant d'une base de données relative à la chaudière de régénération comprenant 56 covariables candidates.

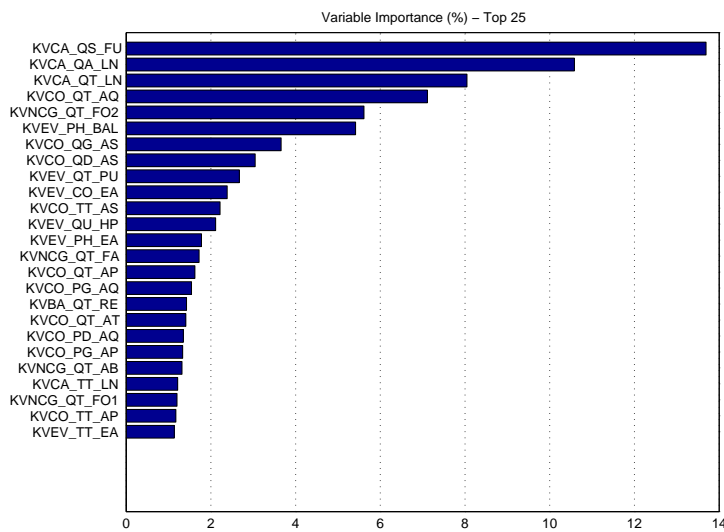


Figure 3: Scores d'importance des variables d'entrées

Le fuel injecté au niveau de la sole (KVCA-QS-FU) se détache par rapport au débit de liqueur noire (débit aller KVCA-QA-LN et débit total KVCA-QT-LN), le mode de fonctionnement en fuel induit donc logiquement davantage de NO_x . Ensuite, le débit d'air quaternaire (KVCO-QT-AQ) à environ 7% est induit par la relation de consigne le liant au débit de liqueur noire; notons que l'air quaternaire a été introduit dans la construction des générations récentes de chaudière de régénération en vue de contrer les NO_x . Les NCG (gaz non condensables) sont des gaz que l'on brûle ultérieurement par une buse spécifique de la chaudière, leur influence sur les NO_x (via la variable KVNCG-QT-FO2) semble bien présente. Ce type d'analyses (développées dans [4] et [5]) sont très utiles dans le cadre d'un monitoring dédié et de la recherche des causes de variations principales de la sortie du modèle.

3 Modélisation par réseaux de neurones

Un réseau neuronal est l'association, en un graphe plus ou moins complexe, d'objets élémentaires, les neurones formels. Les principaux réseaux se distinguent par l'organisation du graphe, c'est-à-dire leur architecture, son niveau de complexité (le nombre de neurones) et par le type des neurones utilisés.

De manière générale, on distingue les réseaux de neurones statiques (p.ex. réseau MLP) où la sortie est calculée directement à partir des entrées par connexions directes dans le réseau et les réseaux de neurones dynamiques où la sortie ne dépend plus uniquement des entrées actuelles du réseau mais aussi des entrées précédentes, des sorties ou des états du réseau. On distingue également les architectures dynamiques à connexions directes uniquement (le graphe ne contient pas de cycle, p.ex. NARX) et celles qui contiennent des connexions récurrentes dans le réseau (p.ex. le réseau d'Elman).

3.1 Multilayer Perceptron

Le Multilayer Perceptron (MLP) est la structure de réseau de neurones la plus répandue (Fig.4), particulièrement la structure à 2 couches dans laquelle les unités d'entrée et la couche de sortie sont interconnectées avec une couche cachée intermédiaire. Le modèle de chaque neurone dans le réseau inclut une fonction d'activation non linéaire qui est différentiable. Ce réseau peut réaliser une fonction in/out entre un espace d'entrée et un espace de sortie. Pour un réseau MLP avec d entrées (indices k), h neurones cachés (indices i) et un noeud de sortie; le mapping correspondant au temps t peut s'écrire:

$$\begin{cases} y(t) = \sum_{i=1}^h h_i(t)v_i + b \\ h_i(t) = \tanh\left(\sum_{k=1}^d u_k(t)w_{ik} + b_i\right) \end{cases}$$

On dénote la matrice des poids des connexions entre entrées et couche cachée $W_{h \times d}$, les poids v_i pour les h poids entre la couche cachée et le noeud de sortie, et les termes de biais b_i et b pour les neurones cachés et de sortie. Les fonctions de transfert utilisées sont des tangentes hyperboliques pour la couche cachée et une fonction linéaire pour le neurone de sortie.

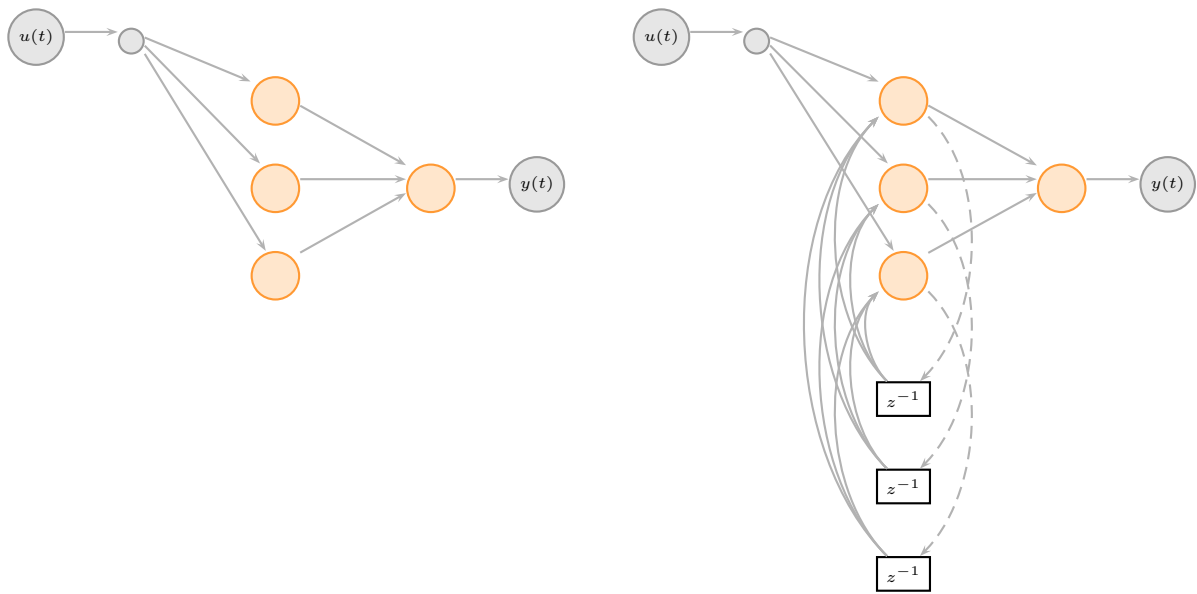
3.2 Elman Recurrent Neural Network

Les réseaux de neurones récurrents (RNN) ont été développés pour une meilleure compréhension et analyse des systèmes dynamiques. Les états d'un système dynamique varient avec le temps et ce fait joue un rôle critique dans l'apprentissage d'un réseau de neurones. Dans le cas du RNN d'Elman (Fig.4), la dimension temporelle est introduite dans la topologie du réseau par l'utilisation de connexions récurrentes (feedback). Il incorpore une couche additionnelle (couche de contexte) dont les noeuds sont des éléments à délais unitaires imbriqués dans une récurrence locale.

Pour un réseau à deux couches avec n noeuds d'entrées (index k), m noeuds cachés (index i), m unités de contexte (index u), et enfin p noeuds de sortie (index j); la fonction entrées/sortie(s) correspondante au temps t peut être développée comme suit:

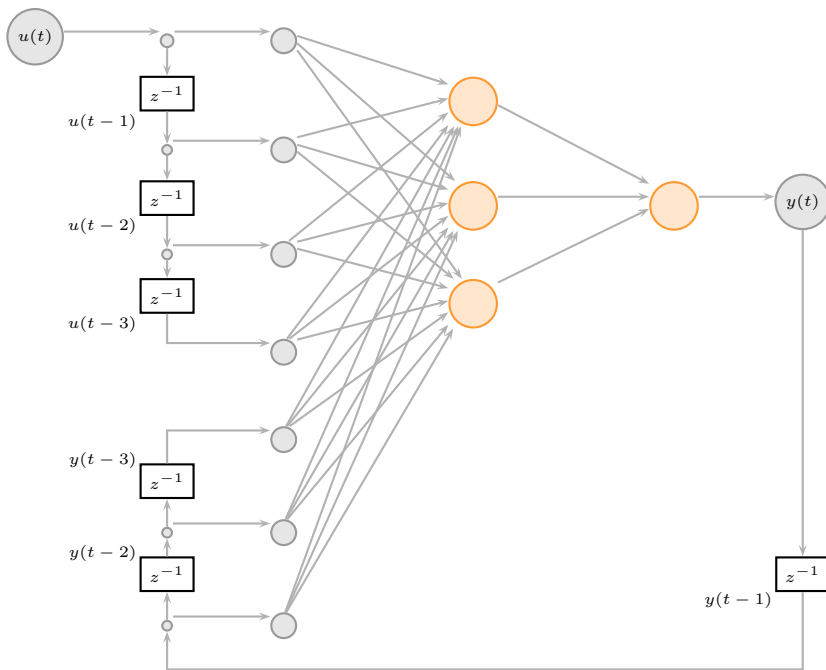
$$\begin{cases} y_j(t) = f\left(\sum_{i=1}^m h_i(t)v_{ji} + b_j\right) \\ h_i(t) = g\left(\sum_{k=1}^n x_k(t)w_{ik} + \sum_{u=1}^m h_i(t-1)c_{iu} + b_i\right) \end{cases}$$

Nous remarquons que l'activation d'état $h_i(t)$ au temps t est combinée avec celle au temps précédent, $h_i(t-1)$, via les unités de contexte. Ce type de réseau est appliqué et comparé au MLP dans [3] pour la prédiction du tonnage de vapeur HP lié à la chaudière.



(a) MultiLayer Perceptron

(b) Elman Recurrent NN



(c) Narx Neural Network

Figure 4: Trois modèles de réseaux de neurones: MLP (a), Elman (b), Narx (c)

3.3 Narx Neural Network

L'appellation NARX (Nonlinear AutoRegressive model with eXogenous inputs) regroupe une importante classe de systèmes non linéaires à temps discrets; ces modèles sont capables de représenter une variété importante de comportements dynamiques non linéaires. Dans le cas d'un système MISO, on a la formulation suivante:

$$\begin{cases} y(i) = f[\mathbf{x}(i)] \\ \mathbf{x}(i) = [y(i-1) \dots y(i-m) u_1(i-1) \dots u_1(i-n) \dots u_d(i-1) \dots u_d(i-n)]^T \end{cases}$$

Pour $i = 1, \dots, N$ avec N la taille de l'échantillon, une entrée $\mathbf{x}(i)$ du modèle à l'instant i est un vecteur colonne de dimension $(m + d \cdot n)$ regroupant les m valeurs précédentes de la variable de sortie y et les n valeurs précédentes des d variables d'entrées u_j . Les paramètres n et m sont nommés respectivement les ordres d'entrées et de sortie du modèle NARX. La fonction $f[\cdot]$ est une fonction non linéaire qui peut être approximée par un réseau de neurones standard (MLP), il en résulte alors un réseau neuronal récurrent NARX (Fig.4) introduit par Narendra et Parthasarathy (1990).

3.4 Etude de cas industriels

Par exemple, nous évaluons ici une approche neuronale dynamique pour la prédiction des émissions d'oxydes d'azote (NO_x) en sortie de la chaudière de régénération des liqueurs. Nous disposons pour cela d'une base de données de production (Fig.5) et désirons, notamment, prédire deux pics critiques de NO_x . L'approche développée inclut une détermination des ordres d'entrées et de sortie du réseau NARX par la méthode des quotients de Lipschitz et un apprentissage du réseau par algorithme de rétropropagation des erreurs avec moment et taux d'apprentissage adaptatif. Nous comparons ces résultats au réseau de neurones statique MLP (Multilayer Perceptron) afin de mieux cerner les bénéfices de l'approche dynamique dans cette étude.

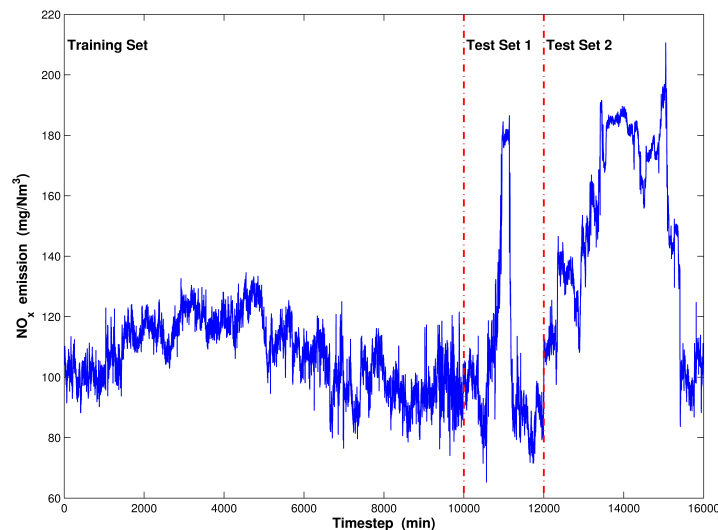


Figure 5: Echantillonnage en trois ensembles distincts: un ensemble d'apprentissage pour construire le modèle et deux ensembles de test (avec deux pics critiques) pour en évaluer la performance.

Dans une récente étude, détaillée dans [1], nous comparons deux modélisations neuronales (MLP statique et NARX dynamique) en vue de prédire des pics critiques de NO_x . Après avoir sélectionné les entrées du modèle (débits de fuel lourd, de liqueur noire, de vapeurs auxiliaires, d'admissions d'airs), nous avons comparé deux architectures neuronales selon leur complexité respective (i.e., le nombre de neurones cachés). Mais, du fait de la présence de bruit de mesure ou de mauvaise initialisation des poids, un réseau de neurone pris seul peut souffrir d'un manque de robustesse face aux changements significatifs dans le procédé à modéliser. Pour éviter cela, on agrège des ensembles de réseaux de neurones.

Ci-après, nous comparons les deux architectures statique (MLP à 3 neurones cachés) et dynamique (NARX à 4 neurones cachés) qui mènent à l'erreur moyenne quadratique minimale sur l'ensemble de test. Nous remarquons (Fig.6) que la modélisation NARX est la plus appropriée en termes de justesse du monitoring et de suivi dynamique des changements significatifs dans le procédé.

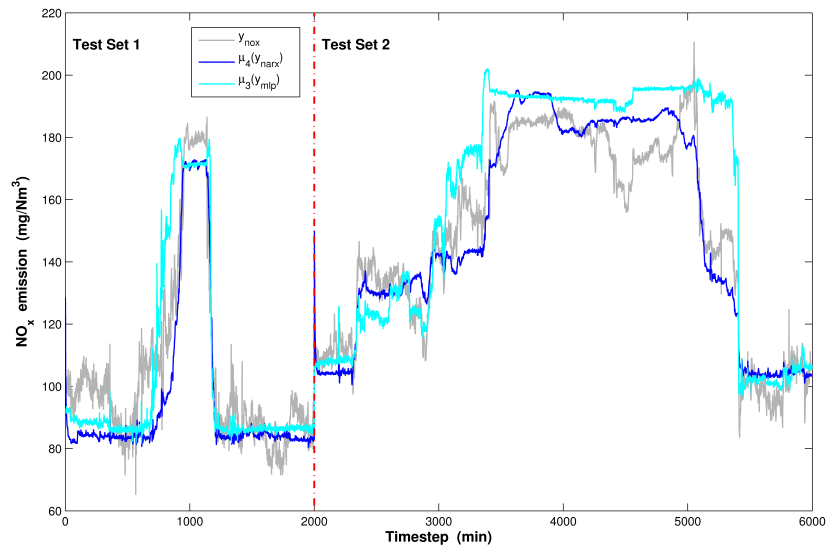


Figure 6: Prédictions neuronales par ensembles de réseaux dynamiques (NARX (μ_4)) et statiques (MLP (μ_3)) par rapport aux émissions de NO_x cibles dans les ensembles de test.

Partenaires du projet

Ce projet est supporté par la Région Wallonne dans le cadre d'un programme First. Les deux partenaires industriels sont BURGO Ardennes S.A. (Virton) et PEPITE S.A. (Liège); le LASSC (Laboratoire d'Analyse et Synthèse des Systèmes Chimiques) de l'université de Liège est partenaire scientifique.



Publications liées au projet

- [1] Matthieu Sainlez and Georges Heyen (2011) *Comparison of Machine Learning techniques for atmospheric pollutant monitoring in a Kraft pulp mill*, Advanced Computational Methods in Engineering, Acomen 2011.
- [2] Matthieu Sainlez, Georges Heyen, and Philippe Lumen (2011) *Approche neuronale dynamique pour la prédiction de polluants atmosphériques: application à l'industrie papetière*, 43^e édition des Journées de Statistique, SFdS 2011.
- [3] Matthieu Sainlez and Georges Heyen (2011) *Recurrent neural network prediction of steam production in a Kraft recovery boiler*, 21st European Symposium on Computer Aided Process Engineering, E.N. Pistikopoulos, M.C. Georgiadis & A.C. Kokossis (Eds.), pp 1784-1788, Elsevier 2011.
- [4] Matthieu Sainlez, Georges Heyen, and Sébastien Lafourcade (2010) *Supervised learning for a Kraft recovery boiler: a data mining approach with random forests*, 23rd International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, ECOS 2010 Volume IV (Power plants and Industrial processes), D. Favrat & F. Maréchal (Eds.), pp. 235-241.
- [5] Matthieu Sainlez, Georges Heyen (2010) *Performance monitoring of an industrial boiler: classification of relevant variables with Random Forests*, 20th European Symposium on Computer Aided Process Engineering, S. Pierucci & G. B. Ferraris (Eds.), pp. 403-408, Elsevier 2010.