



NNT : 2014 EMSE 0745

THÈSE

présentée par

Pierre LAURET

pour obtenir le grade de

Docteur de l'École Nationale Supérieure des Mines de Saint-Étienne

Spécialité : Génie de l'Environnement

MODELISATION DE LA DISPERSION ATMOSPHERIQUE SUR UN SITE INDUSTRIEL PAR COMBINAISON D'AUTOMATES CELLULAIRES ET DE RESEAUX DE NEURONES

soutenue à Alès, le 18 juin 2014

Membres du jury

Président :	Didier GRAILLOT	Professeur, Ecole des Mines de Saint Etienne
Rapporteurs :	Richard SAUREL	Professeur, Université d'Aix Marseille
	Claude TOUZET	Maître de conférence, Université d'Aix Marseille
Examineurs :	Nicolas LECYSYN	Docteur, CEA, Gramat
	Jean Marc RICAUD	Docteur, IRSN, Saint Paul lès Durance
	Anne JOHANNET	Maitre de recherche, Ecole des Mines d'Alès
Directeur de thèse :	Gilles DUSSERRE	Directeur de recherche, Ecole des Mines d'Alès
Encadrants :	Laurent APRIN	Docteur, Ecole des Mines d'Alès
	Frédéric HEYMES	Docteur, Ecole des Mines d'Alès

Spécialités doctorales :
SCIENCES ET GENIE DES MATERIAUX
MECANIQUE ET INGENIERIE
GENIE DES PROCÉDES
SCIENCES DE LA TERRE
SCIENCES ET GENIE DE L'ENVIRONNEMENT
MATHÉMATIQUES APPLIQUÉES
INFORMATIQUE
IMAGE, VISION, SIGNAL
GENIE INDUSTRIEL
MICROELECTRONIQUE

Responsables :
K. Wolski Directeur de recherche
S. Drapier, professeur
F. Gruy, Maître de recherche
B. Guy, Directeur de recherche
D. Graillet, Directeur de recherche
O. Roustant, Maître-assistant
O. Boissier, Professeur
JC. Pinoli, Professeur
A. Dolgui, Professeur
S. Dauzere Peres, Professeur

EMSE : Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'État ou d'une HDR)				
ABSI	Nabil	CR		CMP
AVRIL	Stéphane	PR2	Mécanique et ingénierie	CIS
BATTON-HUBERT	Mireille	PR2	Sciences et génie de l'environnement	FAYOL
BERGER DOUCE	Sandrine	PR2		FAYOL
BERNACHE-ASSOLLANT	Didier	PR0	Génie des Procédés	CIS
BIGOT	Jean Pierre	MR(DR2)	Génie des Procédés	SPIN
BILAL	Essaid	DR	Sciences de la Terre	SPIN
BOISSIER	Olivier	PR1	Informatique	FAYOL
BORBELY	Andras	MR(DR2)	Sciences et génie des matériaux	SMS
BOUCHER	Xavier	PR2	Génie Industriel	FAYOL
BRODHAG	Christian	DR	Sciences et génie de l'environnement	FAYOL
BURLAT	Patrick	PR2	Génie Industriel	FAYOL
COURNIL	Michel	PR0	Génie des Procédés	DIR
DARRIEULAT	Michel	IGM	Sciences et génie des matériaux	SMS
DAUZERE-PERES	Stéphane	PR1	Génie Industriel	CMP
DEBAYLE	Johan	CR	Image Vision Signal	CIS
DELAFOSSÉ	David	PR1	Sciences et génie des matériaux	SMS
DESRAYAUD	Christophe	PR2	Mécanique et ingénierie	SMS
DOLGUI	Alexandre	PR0	Génie Industriel	FAYOL
DRAPIER	Sylvain	PR1	Mécanique et ingénierie	SMS
FEILLET	Dominique	PR2	Génie Industriel	CMP
FRACZKIEWICZ	Anna	DR	Sciences et génie des matériaux	SMS
GARCIA	Daniel	MR(DR2)	Génie des Procédés	SPIN
GERINGER	Jean	MA(MDC)	Sciences et génie des matériaux	CIS
GOEURJOT	Dominique	DR	Sciences et génie des matériaux	SMS
GRILLOT	Didier	DR	Sciences et génie de l'environnement	SPIN
GROSSEAU	Philippe	DR	Génie des Procédés	SPIN
GRUY	Frédéric	PR1	Génie des Procédés	SPIN
GUY	Bernard	DR	Sciences de la Terre	SPIN
HAN	Woo-Suck	CR	Mécanique et ingénierie	SMS
HERRI	Jean Michel	PR1	Génie des Procédés	SPIN
KERMOUCHE	Guillaume	PR2	Mécanique et Ingénierie	SMS
KLOCKER	Helmut	DR	Sciences et génie des matériaux	SMS
LAFOREST	Valérie	MR(DR2)	Sciences et génie de l'environnement	FAYOL
LERICHE	Rodolphe	CR	Mécanique et ingénierie	FAYOL
LI	Jean-Michel			CMP
MALLIARAS	Georges	PR1	Microélectronique	CMP
MOLIMARD	Jérôme	PR2	Mécanique et ingénierie	CIS
MONTHEILLET	Frank	DR		SMS
MOUTTE	Jacques	CR		SPIN
NIKOLOVSKI	Jean-Pierre			CMP
PIJOLAT	Christophe	PR0	Génie des Procédés	SPIN
PIJOLAT	Michèle	PR1	Génie des Procédés	SPIN
PINOLI	Jean Charles	PR0	Image Vision Signal	CIS
POURCHEZ	Jérémy	CR	Génie des Procédés	CIS
ROBISSON	Bruno			CMP
ROUSSY	Agnès	MA(MDC)		CMP
ROUSTANT	Olivier	MA(MDC)		FAYOL
ROUX	Christian	PR		CIS
STOLARZ	Jacques	CR	Sciences et génie des matériaux	SMS
TRIA	Assia		Microélectronique	CMP
VALDIVIESO	François	MA(MDC)	Sciences et génie des matériaux	SMS
VIRICELLE	Jean Paul	MR(DR2)	Génie des Procédés	SPIN
WOLSKI	Krzysztof	DR	Sciences et génie des matériaux	SMS
XIE	Xiaolan	PR1	Génie industriel	CIS
YUGMA	Gallian	CR		CMP

ENISE : Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'État ou d'une HDR)				
BERGHEAU	Jean-Michel	PU	Mécanique et Ingénierie	ENISE
BERTRAND	Philippe	MCF	Génie des procédés	ENISE
DUBUJET	Philippe	PU	Mécanique et Ingénierie	ENISE
FEULVARCH	Eric	MCF	Mécanique et Ingénierie	ENISE
FORTUNIER	Roland	PR	Sciences et Génie des matériaux	ENISE
GUSSAROV	Andrey	Enseignant contractuel	Génie des procédés	ENISE
HAMDI	Hédi	MCF	Mécanique et Ingénierie	ENISE
LYONNET	Patrick	PU	Mécanique et Ingénierie	ENISE
RECH	Joël	PU	Mécanique et Ingénierie	ENISE
SMUROV	Igor	PU	Mécanique et Ingénierie	ENISE
TOSCANO	Rosario	PU	Mécanique et Ingénierie	ENISE
ZAHOUANI	Hassan	PU	Mécanique et Ingénierie	ENISE

PR 0	Professeur classe exceptionnelle	Ing.	Ingénieur
PR 1	Professeur 1 ^{ère} classe	MCF	Maître de conférences
PR 2	Professeur 2 ^{ème} classe	MR (DR2)	Maître de recherche
PU	Professeur des Universités	CR	Chargé de recherche
MA (MDC)	Maître assistant	EC	Enseignant-chercheur
DR	Directeur de recherche	IGM	Ingénieur général des mines

SMS	Sciences des Matériaux et des Structures
SPIN	Sciences des Processus Industriels et Naturels
FAYOL	Institut Henri Fayol
CMP	Centre de Microélectronique de Provence
CIS	Centre Ingénierie et Santé

Un scientifique dans son laboratoire est non seulement un technicien : il est aussi un enfant placé devant des phénomènes naturels qui l'impressionnent comme des contes de fées.

Marie Skłodowska-Curie

Remerciements

Je voudrais tout d'abord remercier Didier Graillot d'avoir accepté de présider ce jury de thèse ainsi que Richard Saurel et Claude Touzet, d'avoir évalué et rapporté mon travail. Je remercie également Jean Marc Ricaud pour l'intérêt qu'il a témoigné à ce travail en acceptant de participer au jury de thèse.

Ce travail a été réalisé sur la base d'un contrat de recherche entre l'Ecole des mines d'Alès et le CEA Gramat, je tiens donc à remercier tous les interlocuteurs que j'ai pu rencontrer : Emmanuel Lapébie, Laurent Munier, Antoine Osmont et Nicolas Lecysyn qui m'a fait l'honneur d'être membre du jury.

Je voudrai également remercier le directeur de cette thèse, Gilles Dusserre, tout d'abord pour m'avoir convaincu d'effectuer un doctorat :

- « - Un doctorat ? Je ne suis pas sûr de pouvoir y arriver !
- Mais si, c'est simplement un marathon couru à la vitesse d'un 100 m...
- Dans ce cas-là ! »

Merci également pour la confiance accordée et l'intégration dans l'équipe qui fût facilitée.

Un grand merci à Anne Johannet pour m'avoir initié aux réseaux de neurones, même si quelques-uns ont probablement été perdus dans la bataille, avec sa grande patience, sa gentillesse et son dynamisme communicatif.

Cette thèse n'aurait pas eu la même saveur si l'ambiance n'avait pas été aussi bonne. Pour cela, je remercie infiniment mes encadrants à l'Institut des Sciences des Risques, Laurent Aprin et Frédéric Heymes, d'avoir su m'éviter de me disperser, d'avoir subi mes graphiques psychédéliques et parfois turbulents sans (trop) broncher, d'avoir suscité la créativité qui a permis à cette thèse d'exister.

Laurent, ton optimisme et ta franchise m'ont permis de croire que c'était possible. Aussi, tu as su maintenir ma concentration pour mener à bien ce travail tout en permettant des moments de dispersion nécessaire à la reconstitution de mes neurones. J'ai également beaucoup apprécié t'assister dans les séances de TD que nous avons encadrées ensemble. Really !

Fred, ta rigueur scientifique et ta vision de l'objectif à atteindre ont aidé à mon élévation scientifique malgré les obstacles encombrant le récit. Les nombreuses discussions passées, débordant parfois largement le cadre scientifique, ont favorisé l'ouverture d'esprit nécessaire au chercheur que je tente de devenir.

Merci également à Pierre Slangen d'avoir suivi les différents premiers jets virgule éventuellement négatifs d'avancement de cette thèse sur les produits Inflammables dont la Concentration est Evaluée par Modélisation Analytique des Neurones.

3 ans d'une vie ne sauraient bien se dérouler sans un environnement favorable. Celui dont j'ai bénéficié à l'ISR fût exceptionnel. Ainsi, je souhaiterais typiquement remercier : Aurélia, Jérôme, Flo, Pierre Alain et Sophie avec qui les divers échanges ont apporté des angles de vue différents et constructifs. Par ailleurs, je remercie également Corinne,

Dominique, Andrée, Cathy et Rosario pour leur support dans la jungle administrative et technique.

Merci évidemment aux camarades de classe : Serge qui a montré la voie, Mélanie pour son optimisme, Boris pour les riffs de la mi-journée et les discussions cérébro-percepto-neuronales extrêmement enrichissantes, mais aussi Olivier, Dimitri, Ingrid, Chloé et le petit dernier Philippe.

Un grand merci à Benjamin (et son grand ami) qui m’a supporté (la réciproque est vraie) durant 3 ans, rendant les journées plus agréables, les débats passionnés et les phrases chocs. Merci également aux amis d’en face, Thibault (notre programme est solide Croco !), Guilherme (yes my friend Andrew !) et d’un peu plus loin, l’énergique ☺ Tangi, pour les moments passés à refaire le match/monde.

Je voudrais également remercier l’ensemble des personnes (stagiaires et élèves encadrés notamment) ayant de près ou de loin pris part à ces 3 années.

Un grand merci à l’ensemble du personnel de l’Ecole des mines d’Alès, à Miguel Lopez-Ferber et Yannick Vimont pour m’avoir accueilli au sein du LGEI.

Un immense merci à mes parents. A ma mère qui aurait adoré être à ma place et qui m’a transmis sa rigueur, son dynamisme et son opiniâtreté, à mon père qui a fait germer en moi le goût de la science (construire des montgolfières en CM2, c’est déjà s’attaquer à la mécanique des fluides) et à mon frère qui m’a communiqué sa détermination et appris à lire, écrire et compter.

Enfin, je voudrai remercier Elisa qui sut porter le fardeau de la vie avec un ~~thésard~~ élève chercheur, entre moments d’euphorie et de désespoir (parfois dans la même demi-heure) avec une extrême patience.

Table des matières

REMERCIEMENTS	V
TABLE DES MATIERES	VII
LISTE DES SYMBOLES ET ACRONYMES.....	IX
INTRODUCTION	1
PARTIE I : CONTEXTE ET ETAT DE L'ART	4
I-1. LA DISPERSION ATMOSPHERIQUE	5
I-1.1. Introduction.....	5
I-1.2. Historique	5
I-1.3. Le site industriel : zone d'aléa	11
I-1.4. Gestion des risques industriels.....	14
I-1.5. Conclusion	16
I-2. MECANISME DES ECOULEMENTS ATMOSPHERIQUES	17
I-2.1. L'atmosphère.....	17
I-2.2. La couche limite atmosphérique.	18
I-2.3. Equations fondamentales.....	21
I-2.4. Ecoulement autour d'un stockage.....	30
I-2.5. Conclusion	32
I-3. MODELISATION DE LA DISPERSION ATMOSPHERIQUE.....	34
I-3.1. Différents types de modélisation	35
I-3.2. Modèles de dispersion atmosphérique	37
I-3.3. Conclusion	44
PARTIE II : MODELES ISSUS DE L'INTELLIGENCE ARTIFICIELLE	46
II-1. AUTOMATES CELLULAIRES	47
II-1.1. Historique	47
II-1.2. Principes généraux et définitions	50
II-1.3. Applications à la mécanique des fluides.....	52
II-1.4. Conclusion	54
II-2. RESEAUX DE NEURONES.....	55
II-2.1. Historique et généralités	55
II-2.2. Définitions	55
II-2.3. Applications en dispersion atmosphérique	64
PARTIE III : METHODES ET RESULTATS	68
III-1. STRATEGIE DE L'ETUDE	69
III-1.1. Constitution des bases de données CFD	72
III-1.2. Lien avec la couche limite atmosphérique	73
III-1.3. Principes et évaluation des réseaux de neurones créés.....	76
III-2. MODELISATION DE LA DISPERSION ATMOSPHERIQUE D'UNE SOURCE CONTINUE EN CHAMP LIBRE.....	82
III-2.1. Simulation CFD	82
III-2.2. Base de données et base d'exemples.....	83

III-2.3.	Apprentissage et optimisation	84
III-2.4.	Evaluation du modèle	88
III-2.5.	Temps de calcul	94
III-2.6.	Conclusion	95
III-3.	MODELISATION DE LA DISPERSION ATMOSPHERIQUE D'UNE BOUFFEE EN CHAMP LIBRE	96
III-3.1.	Objectifs.....	96
III-3.2.	Méthode d'utilisation du CA-ANN	99
III-3.3.	Base de données et base d'exemples.....	102
III-3.4.	Réseau de neurones	105
III-3.5.	Evaluation du modèle CA-ANN.....	110
III-3.6.	Gain en temps de calcul	118
III-3.7.	Conclusion	119
III-4.	MODELISATION DE LA DISPERSION ATMOSPHERIQUE AUTOUR D'UN OBSTACLE CYLINDRIQUE	121
III-4.1.	Objectifs.....	121
III-4.2.	De la base de données à la base d'apprentissage	122
III-4.3.	Structure du réseau de neurones.....	126
III-4.4.	Optimisation et sélection des modèles	127
III-4.5.	Evaluation sur des cas non-appris	132
III-4.6.	Application aux cas non-appris.....	145
III-4.7.	Conclusion	151
CONCLUSIONS.....		153
PERSPECTIVES		157
BIBLIOGRAPHIE		159
LISTE DES FIGURES		167
LISTE DES TABLEAUX		175
PARTIE IV : ANNEXES.....		176
IV-1.	RESOLUTION NUMERIQUE DE L'EQUATION D'ADVECTION DIFFUSION (ADE)	177
IV-1.1.	Méthode générale.....	177
IV-1.2.	Résolution de l'advection	177
IV-1.3.	Résolution de la diffusion	178
IV-2.	PRISE EN COMPTE DE DONNEES INITIALES BRUTEES.....	181
IV-3.	ARTICLE PUBLIE DANS CHEMICAL ENGINEERING TRANSACTIONS – VOL. 31, 2013	185
IV-4.	ARTICLE PUBLIE DANS CHEMICAL ENGINEERING TRANSACTIONS – VOL. 36, 2014.....	193
IV-5.	SPECIFICATION DES STATIONS DE TRAVAIL UTILISEES DANS CETTE ETUDE	200

Liste des symboles et acronymes

Acronymes

Symbole	Définition
AC	Automate Cellulaire
ADE	Equation d'Advection Diffusion
ADN	Acide DésoxyriboNucléique
ANN	Artificial Neural Networks
AZF	Usine chimique AZote Fertilisants
CA-ANN	Automate Cellulaire à règle de transition de type réseau de neurones
CFD	Computational Fluid Dynamics
CLA	Couche Limite Atmosphérique
CLS	Couche Limite de Surface
DNS	Direct Numerical Simulation
GPL	Gaz de Pétrole liquéfié
IAEA	International Atomic Energy Agency
ICMESA	Industria Chimica Meda Società Azionaria
ICPE	Installation Classée pour la Protection de l'Environnement
IDLH	Immediately Dangerous to Life or Health
LES	Large Eddy Simulation
LIF	Laser Induced Fluorescence
LII	Limite Inférieure d'Inflammabilité
LSI	Limite Supérieure d'Inflammabilité
MIC	Methyl Isocyanate
OMS	Organisation Mondiale de la Santé
PIV	Particle Image Velocimetry
PNE	Point Normal d'Ebullition
RANS	Reynolds Averaged Navier Stokes equations
RN	Réseau de neurones
SEI	Seuil des Effets Irréversibles
SEL	Seuil des Effets Létaux
TLV-STEL	Threshold Limit Value – Short-Term Exposure Limit

TLV-TWA	Threshold Limit Value – Time Weighted Average
VLE	Valeur Limite d'Exposition
VME	Valeur Moyenne d'exposition

Lettre latines

Symbole	Définition
C	Concentration
$C_{1\epsilon}$	Constante du modèle $k - \epsilon$ standard
$C_{2\epsilon}$	Constante du modèle $k - \epsilon$ standard
C_μ	Constante du modèle $k - \epsilon$ standard
C_p	Chaleur spécifique à pression constante
D	Dimension caractéristique du milieu (diamètre dans le cas d'un cylindre)
D_t	Coefficient de diffusion turbulente
$EQMT$	Erreur Quadratique Moyenne sur l'ensemble de Test
f_i	Résultante des forces extérieures de volume
f	fonction d'activation d'un neurone
$FAC2$	Facteur deux
FB	Fractional Bias
g	Gravité
H_0	Flux d'énergie reçu par le sol
k	Energie cinétique turbulente
K	Tenseur de diffusivité turbulente
K_m	Coefficient d'échange turbulent pour la quantité de mouvement
L_{MO}	Longueur de Monin Obukhov
$N(r)$	Nombre de voisins de la cellule r
$NMSE$	Normalized Mean Squared Error
p	pression
q	Humidité spécifique
R	Rayon du cylindre
R^2	Coefficient de détermination
R_c	Réaction chimique
S_i	Source
S_{VC}	Score de validation croisée
t	Temps
T	Température
T_v	Température virtuelle
U	Vitesse moyenne de l'écoulement
u_*	Vitesse de friction
u_i	Vitesse dans la direction i
\mathbf{w}	Vecteur des paramètres du réseau de neurones
x	distance
\mathbf{x}	Vecteur des variables d'entrée du réseau de neurones
z	Altitude
z_0	Longueur de rugosité

Lettres grecques

Symbole	Définition
α	Diffusion moléculaire

Symbole	Définition
β	Coefficient de flottabilité
δ_{ij}	Symbole de Kronecker
ϵ	Taux de dissipation de l'énergie cinétique turbulente
κ	Constante de Von Karman
μ	Viscosité dynamique
μ_t	Viscosité turbulente
ν	Viscosité cinématique
ν_t	Viscosité cinématique turbulente
Ω	Vitesse angulaire
ρ	Masse volumique
σ_ϵ	Viscosité cinématique turbulente
σ_{ij}	Tenseur des contraintes agissant à la surface de l'élément fluide
σ_k	Constante du modèle $k - \epsilon$ standard
σ_i	Ecart-types de dispersion
θ_p	Température potentielle
θ_v	Température potentielle virtuelle

Nombres adimensionnels

Symbole	Définition
F_r	Nombre de Froude
P_e	Nombre de Peclet
R_e	Nombre de Reynolds
R_i	Nombre de Richardson
R_o	Nombre de Rossby
S_c	Nombre de Schmidt
S_{ct}	Nombre de Schmidt turbulent

Introduction

Généralement, les sites industriels sont des zones où de nombreuses substances chimiques sont présentes. Qu'il s'agisse d'extraction, de transformation, de conditionnement ou de stockage, la présence de ces substances peut représenter des termes sources pour des phénomènes dangereux tels que des surpressions, des flux thermiques ou des concentrations de substances dans le milieu environnant. Si autrefois, les industries en capacité de générer des phénomènes de ce type se situaient dans des zones peu ou pas habitées, ce n'est plus le cas aujourd'hui. L'exemple de l'urbanisation autour de l'usine AZF de Toulouse en est une tragique illustration. Cette usine était située en périphérie de Toulouse lors de sa création. En 2001, la zone impactée par l'explosion était habitée. Cette combinaison de l'aléa et de l'enjeu constitue le risque, industriel et majeur si les phénomènes et/ou les enjeux le sont.

Parmi ces phénomènes, la dispersion atmosphérique n'est pas toujours identifiée directement. En effet, dans le cas de substances inflammables ou explosives, le phénomène majeur est l'inflammation ou l'explosion, la dispersion n'étant qu'une étape. En revanche, dans le cas de substances toxiques, le danger est représenté directement. La catastrophe qui a eu lieu à Bhopal en Inde en 1984, est un exemple de l'impact de la dispersion d'un gaz toxique à partir d'un stockage industriel. La proximité de l'usine d'engrais d'Union Carbide et de la ville de Bhopal associée à la dangerosité du produit dispersé et aux conditions atmosphériques et topographiques défavorables ont causé l'une des pires catastrophes industrielles de tous les temps (Charbonnier, 2004). Un exemple lié à la dispersion de Gaz de Pétrole Liquéfié (GPL) est l'accident de train de Viareggio en Italie survenue en juin 2009. Suite au déraillement d'un wagon citerne, une fuite de GPL a entraîné la formation d'un nuage qui s'est dispersé de manière très spécifique à cause de la configuration des rues à proximité de l'accident. Une source de chaleur a provoqué l'embrasement et l'explosion du nuage. Dans une telle situation, l'échelle de temps avant l'apparition du phénomène est très courte, de l'ordre de 5 minutes dans le cas de Via Reggio (Brambilla, 2011).

Afin de prévenir ce genre d'accidents, les autorités nationales et internationales ont mis en place des règles plus ou moins contraignantes afin d'éviter la répétition de tels accidents. En France, les industries considérées comme étant potentiellement dangereuses sont soumises à la réglementation sur les Installations Classées pour la Protection de l'Environnement (ICPE). Dans le cadre de cette réglementation, l'exploitant doit notamment fournir une étude d'impact qui représente les conséquences des différentes nuisances liées à l'exploitation dans des conditions normales et une étude de danger qui identifie et caractérise les scénarios d'accidents étant susceptibles de se produire. Les industriels mettent alors en place des barrières de défense pour palier à ces effets. Afin de les dimensionner correctement, il est nécessaire de déterminer précisément les phénomènes mis en jeu. Leurs modélisations deviennent subséquemment un élément décisionnel fort sur les moyens de répondre à l'occurrence de tels dangers.

Concernant la modélisation de la dispersion atmosphérique plusieurs stratégies peuvent être mises en place indépendamment ou en complément l'une de l'autre. La

modélisation physique a pour objectif de recréer une situation la plus réaliste possible. Les conditions de dispersion peuvent être simulées dans des souffleries atmosphériques ou des canaux hydrauliques ou en réalisant des essais à échelle réelle. Dans cette dernière perspective, les coûts induits sont importants sans garantie de pouvoir étudier l'ensemble des scénarios susceptibles de se produire. Dans le cas de la modélisation physique à échelle réduite, les contraintes sont imposées par le respect des règles de similitudes. Malheureusement, il est impossible de toutes les satisfaire simultanément, ce qui entraîne une augmentation de l'utilisation de la simulation numérique. Avec l'augmentation des puissances de calculs de manière continue depuis plus de 30 ans, ce type de modélisation devient incontournable.

Afin d'optimiser au mieux les puissances de calcul disponibles, de nombreux modèles cohabitent, possédant chacun des atouts pour différentes situations à évaluer. Celles-ci sont complexes et font appel à des compétences pluridisciplinaires : météorologie, mécanique des fluides, chimie, informatique, mathématiques appliquées, ... Différentes échelles se doivent également d'être étudiées selon que le risque est chronique ou bien accidentel, lié à une substance inflammable, toxique ou radioactive. Les réseaux de surveillance de l'air dans un îlot urbain ne nécessiteront pas les mêmes modèles que ceux permettant de suivre un nuage radioactif à l'échelle du globe tel que celui issu de l'accident de Fukushima. Il est entendu que l'ensemble des échelles ne peut être représenté en utilisant un seul modèle. C'est pourquoi de nombreux modèles existent afin de répondre aux différentes situations rencontrées. S'agissant des sites industriels, la présence d'un nombre important d'obstacles et d'un relief rarement homogène peut perturber grandement un écoulement. Cet environnement sera à prendre en compte afin de caractériser au mieux l'évolution d'un nuage d'une substance autour d'un site.

Dans le cas de la dispersion d'une substance sur un site industriel, l'intérêt se porte sur l'échelle locale, inférieure au kilomètre. L'observation d'un site industriel permet de mettre en évidence des caractéristiques propres à ce type de terrain. Plusieurs types de bâtis, dont les stockages, sont reliés entre eux à travers un réseau de conduites. Ces zones sont tout particulièrement surveillées car d'éventuelles fuites peuvent y être localisées. Les études à réaliser par les industriels se doivent de répondre notamment aux questions suivantes :

- ▶ De quelle manière va évoluer un nuage de substances inflammables ou toxiques à travers le site industriel ?
- ▶ Dans le cas d'une substance inflammable, la concentration sera-t-elle dans la zone d'inflammabilité ? Durant combien de temps ?
- ▶ Existe-t-il des possibilités d'effets dominos si le nuage rencontre une source d'ignition ?
- ▶ Les valeurs en limite du site sont-elles en deçà des seuils de danger ?

Des questions d'ordre technique se posent alors suite à ces interrogations. Elles concernent les objectifs de la modélisation, la caractérisation des scénarios à évaluer et la faisabilité de telles études, notamment en ressources informatiques.

Les modèles permettant de répondre à ces interrogations sont multiples et peuvent différer les uns des autres par divers aspects. Ces modèles peuvent être commerciaux ou libres, modifiables ou non. Les différences fondamentales résident souvent dans la théorie engagée, le degré de complexité et la finalité de l'utilisation. Ces modèles opposent souvent deux critères de sélection : la vitesse et la précision. Les modèles considérés comme simples sont conçus pour être rapides et certaines limitations y sont donc apposées de par les hypothèses de départ, la géométrie du domaine par exemple. Les modèles plus complets tels que les modèles Computational Fluid Dynamics (CFD)

permettent une description généralement plus fine de la réalité mais le temps de calcul associé et l'expertise requise peuvent être rédhibitoires pour certains industriels.

L'objectif de ce travail de recherche est de mettre en évidence des outils permettant de réduire le temps de calcul nécessaire pour arriver à un niveau de détail important. La stratégie développée dans cette thèse est d'utiliser des outils issus de l'intelligence artificielle et de l'apprentissage statistique afin de déporter l'essentiel du calcul en amont de l'utilisation du modèle opérationnel. L'utilisation des réseaux de neurones et des automates cellulaires s'effectue à partir d'une base de données d'écoulements et de dispersions de substances, réalisées grâce à un modèle issu de la CFD. Différentes stratégies sont évaluées selon une complexité croissante : dispersion stationnaire et instationnaire en absence d'obstacles, dispersion stationnaire et instationnaire autour d'un obstacle cylindrique. Pour chaque situation, différentes méthodologies ont été déployées mettant en exergue des aspects scientifiques et techniques :

- ▶ Simulation CFD de la dispersion atmosphérique dans une couche limite atmosphérique
- ▶ Echantillonnage d'une base de données
- ▶ Apprentissage d'un réseau de neurones
- ▶ Sélection des variables d'intérêt, sélection de l'architecture, sensibilité à l'initialisation
- ▶ Développement d'un code de dispersion
- ▶ Outils statistiques d'évaluation des modèles de dispersion

La présentation de ce travail s'articule autour de trois parties principales. La première présente la problématique de dispersion atmosphérique, les différents phénomènes impliqués ainsi que les principaux modèles existants, leurs avantages et leurs limites.

La deuxième partie présentera les outils issus de l'intelligence artificielle utilisés dans ce travail : les réseaux de neurones et les automates cellulaires. Leur origine sera explicitée et leur fonctionnement détaillé.

La troisième partie sera consacrée à l'explication de la méthodologie déployée dans ce travail et les différentes étapes abordées selon la configuration de dispersion atmosphérique.

Partie I : Contexte et état de l'art

"Standing on the Shoulder of Giants"

Noel Gallagher

Cette partie traite du contexte scientifique, technique et réglementaire dans lequel s'inscrit cette étude. La dispersion atmosphérique est abordée de manière générale au chapitre 1, à partir de l'observation des phénomènes dangereux et de la réponse apportée par les scientifiques et les politiques. Les mécanismes des écoulements atmosphériques sont étudiés dans le chapitre 2, à partir du point de vue macroscopique jusqu'aux situations locales qui concernent ce travail. Enfin, la modélisation de la dispersion est détaillée dans le chapitre 3, en distinguant les différentes façons de modéliser, les contraintes et avantages de chaque modèle.

I-1.	LA DISPERSION ATMOSPHERIQUE	5
I-1.1.	Introduction.....	5
I-1.2.	Historique	5
I-1.3.	Le site industriel : zone d'aléa	11
I-1.4.	Gestion des risques industriels.....	14
I-1.5.	Conclusion	16
I-2.	MECANISME DES ECOULEMENTS ATMOSPHERIQUES	17
I-2.1.	L'atmosphère.....	17
I-2.2.	La couche limite atmosphérique.	18
I-2.3.	Equations fondamentales.....	21
I-2.4.	Ecoulement autour d'un stockage.....	30
I-2.5.	Conclusion	32
I-3.	MODELISATION DE LA DISPERSION ATMOSPHERIQUE.....	34
I-3.1.	Différents types de modélisation	35
I-3.2.	Modèles de dispersion atmosphérique	37

I-1. La dispersion atmosphérique

Ce chapitre présente la problématique de la dispersion atmosphérique sur un site industriel. A partir d'un retour d'expérience d'accidents de dispersion atmosphérique, les risques dans ce domaine seront détaillés pour le cas du site industriel. Les mesures mises en place pour prévenir ou gérer ces risques seront alors abordées. Nous verrons enfin dans quelle mesure la modélisation des phénomènes impliqués permet de limiter et prévoir les dangers associés à la dispersion atmosphérique de substances dangereuses.

I-1.1. Introduction

La dispersion atmosphérique consiste en une multitude de phénomènes qui modifient l'émission (terme source) d'une substance dans l'air. Il s'agit d'une pollution atmosphérique lorsque des effets négatifs peuvent être évalués sur les personnes ou l'environnement. Les substances concernées sont multiples. Il peut s'agir d'aérosols (particules) ou de gaz, seuls ou combinés. Par ailleurs, la substance concernée et les conditions de stockage ou d'utilisation au moment de la fuite sont des éléments déterminants pour la caractérisation de ce terme source. On distingue généralement les pollutions chroniques des pollutions accidentelles. Dans le premier cas, les concentrations de produits dans le milieu sont généralement faibles mais continues et peuvent conduire à des maladies chroniques. Dans le second cas, il s'agit de phénomènes où les substances sont présentes en concentrations suffisamment importantes pour s'enflammer ou induire une intoxication immédiate ou à court terme. Ces phénomènes peuvent être fatals. Dans les deux cas, des liens étroits sont tissés entre les scientifiques chargés d'évaluer le danger induit par le phénomène et les scientifiques chargés d'évaluer les limites d'acceptations du danger par l'enjeu (humain ou environnemental). Le risque recouvre l'évaluation des enjeux et des conséquences. La prise en compte de ce risque par les activités industrielles est récente, les accidents passés permettant de faire évoluer la perception du risque et la législation associée.

I-1.2. Historique

Une analyse des accidents industriels majeurs passés donne des indications sur les éléments influants sur le comportement d'une substance dans l'atmosphère.

La toxicité et la pollution s'entendent vis-à-vis de l'Homme. En ce sens, elles étaient déjà présentes au commencement de l'humanité à travers les feux de forêt et l'activité volcanique, émetteurs de gaz et de particules dans l'atmosphère. Néanmoins, depuis l'avènement de l'Homme et le début de l'ère industrielle, l'apparition du progrès technologique a induit l'accélération des émissions et l'apparition d'accidents industriels.

Les premières villes étaient exposées aux pollutions chroniques, issues de l'utilisation du bois de chauffage, des odeurs liées aux activités domestiques ou bien encore des premières activités artisanales. La révolution industrielle et la pollution qui lui est reliée a permis l'observation de liens directs entre mortalité et fortes concentrations de substances dangereuses (notamment en particules fines et en dioxyde de soufre). Le terme smog, contraction des mots smoke (fumée) et fog (brouillard) est apparu suite à des épisodes de grandes pollutions, combinées à des conditions atmosphériques d'une

stabilité exceptionnelle. Les cas de la vallée de la Meuse en 1930 (Allix, 1932), Donora aux Etats-Unis en 1948, Londres en 1952, New York City en novembre 1953, 1963, 1966, Pittsburgh en novembre 1975, Los Angeles en 1954, Mexico City dans les années 1950 et 1960 (Raga et al., 2001) sont notables.

Plus récemment, l'épisode exceptionnel de pollution urbaine de mars 2014 en France a mis en exergue l'accentuation d'un risque chronique. Bien qu'étant différent du risque accidentel, il mérite d'être étudié. A l'occasion de conditions de stabilité atmosphérique exceptionnelles, les niveaux de pollutions aux particules ont atteint des valeurs au-delà des limites fixées par l'Organisation Mondiale de la Santé (OMS). Ces particules sont pour la plupart émises par les activités anthropiques (Sciare, 2014). On distingue ainsi les particules de dimensions inférieures à 10 microns et 2,5 microns, ces dernières étant les plus dangereuses pour la santé.

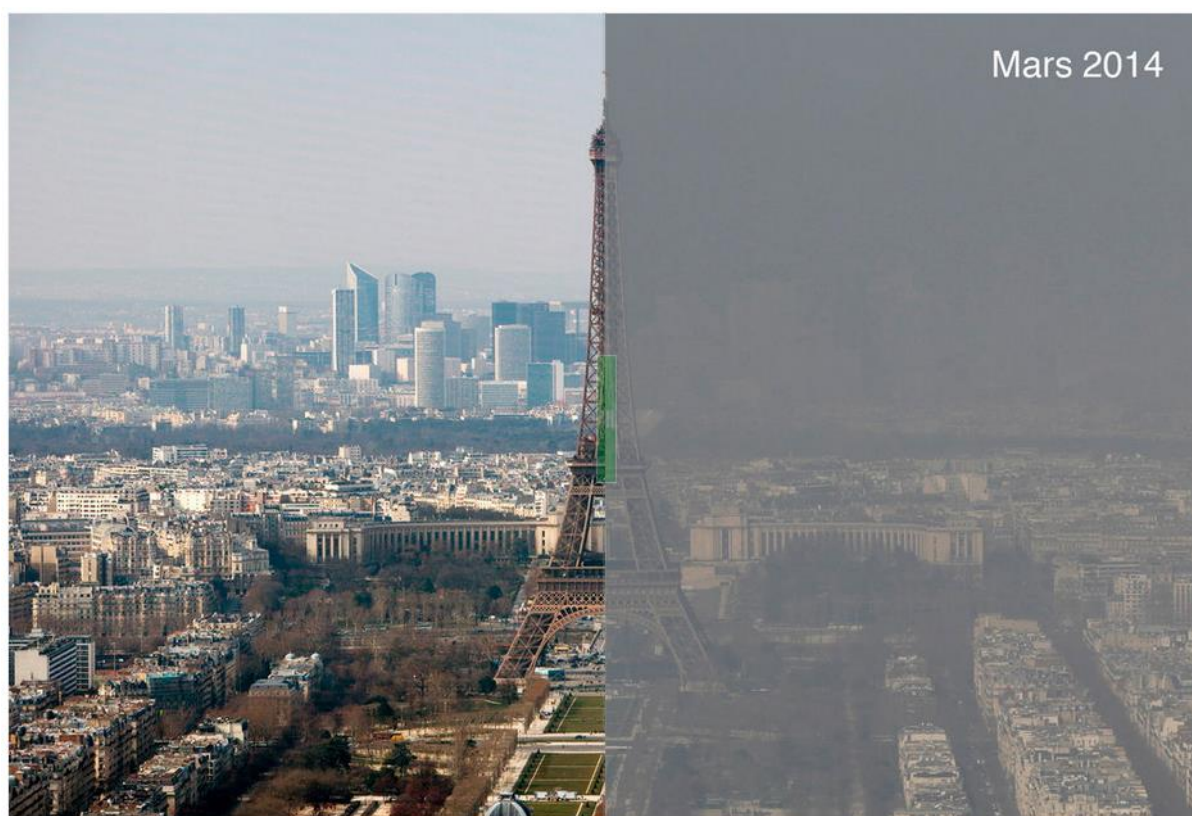


Figure I-1 : Illustration de l'effet de smog sur la ville de Paris (copie d'écran du site lemonde.fr du 14.03.2014)

La photographie ci-dessus montre une comparaison de la pollution atmosphérique entre une journée du 10 mars 2010 (à gauche) et du 13 mars 2014. On constate ainsi la différence entre une journée type et une journée de pollution importante. Le vendredi 14 mars 2014, l'air parisien atteignait un pic de concentration en particules fines PM_{10} de $105 \mu g.m^{-3}$ d'air. En France, un seuil d'information et de recommandation est atteint pour une concentration supérieure à $50 \mu g.m^{-3}$ en moyenne sur 24 heures. Le seuil d'alerte est lui fixé à $80 \mu g.m^{-3}$ en moyenne sur la même période. Afin de protéger d'une part la population et de limiter cette pollution d'autre part, un certain nombre de recommandations ont été faites par les autorités : limitation des activités physiques et sportives intenses, modification éventuelle des traitements pour les personnes vulnérables.

La figure ci-après montre les valeurs des pics de concentrations sur l'ensemble de la France pour la journée du 14 mars 2014 :

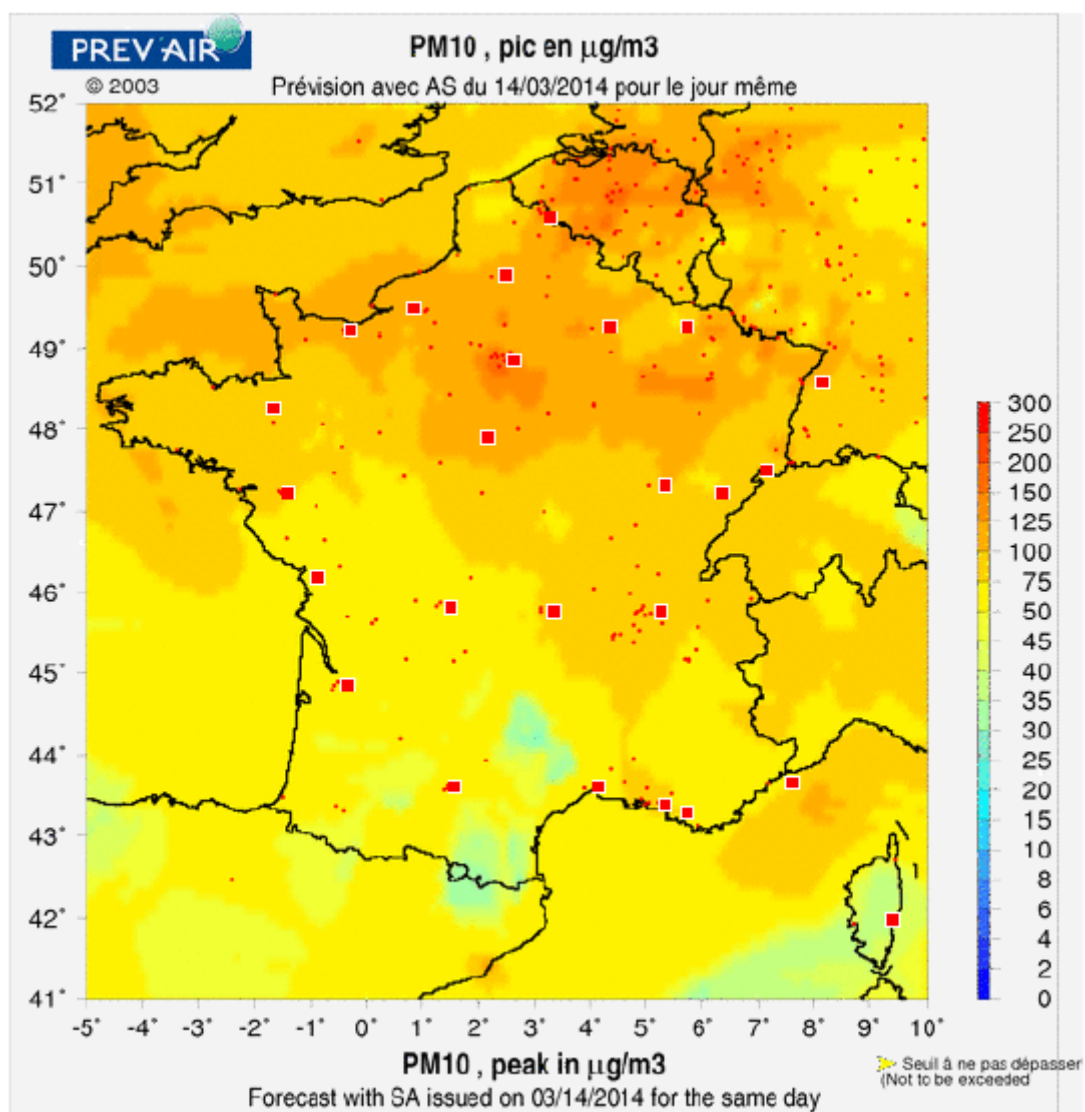


Figure I-2 : Cartographie à l'échelle nationale des pics de concentration en PM10 le vendredi 14 mars 2014 (capture d'écran du site http://www.prevair.org/fr/prevision_pm10.php du 14 mars 2014).

La Figure I-2 met en évidence clairement les zones urbaines à forte densité de population du nord de la France (Ile de France et zones urbaines belges), avec des valeurs des pics supérieures au seuil d'information sur une grande partie de la zone observée.

L'ensemble de ces événements est dû à une accentuation d'une pollution chronique suite à des conditions météorologiques particulières : faible température, vent nul ou faible, plafond atmosphérique bas dû à une inversion de température. Dans ce cas, la pollution chronique se concentre dans la couche ainsi créée (Malek et al., 2006).

Ces catastrophes sont l'amplification d'un phénomène permanent en bruit de fond dont l'échelle de temps et d'espace est macroscopique. D'autres événements sont eux soudains et peuvent être tout aussi néfastes. La dispersion accidentelle de produits chimiques a engendré des accidents majeurs. On les différencie en fonction de l'échelle

géographique et temporelle de l'impact occasionné. Les catastrophes nucléaires de Tchernobyl (1986) et Fukushima (2011) sont des catastrophes dont les effets sont perceptibles à une grande distance du site émetteur durant une période longue. Les catastrophes de Bhopal (1984), Seveso (1976), Via Reggio (2009) sont considérées comme des catastrophes régionales voir locales et de faibles durées. Les conséquences n'en sont pas moins dévastatrices.

En ce qui concerne le cas de la centrale nucléaire de Tchernobyl (Ukraine), l'accident s'est produit durant un exercice programmé pour tester la sécurité de la centrale. La combinaison de plusieurs facteurs (mauvaise organisation, conception spécifique, type de réacteur difficilement pilotable, matériel défaillant) a amené à la libération et l'évaporation d'une grande partie du liquide de refroidissement du réacteur et la formation d'hydrogène. L'explosion de ce gaz a détruit le toit du bâtiment réacteur et enflammé les barres de graphite, modérateur de la réaction nucléaire. Le rejet de matière radioactive s'est élevé à environ 1 000 mètres de hauteur lors de l'explosion. Par la suite, l'élévation dans l'air de particules radioactives a duré plus de dix jours et parcouru des distances supérieures à 2 000 km pour les concentrations significatives en césium et en iode radioactive. Si l'on observe une carte de la dispersion de la radioactivité suite à l'accident de Tchernobyl, on se rend compte des distances parcourues par les nuages contenant de la radioactivité, relevant de l'échelle continentale, durant une semaine.

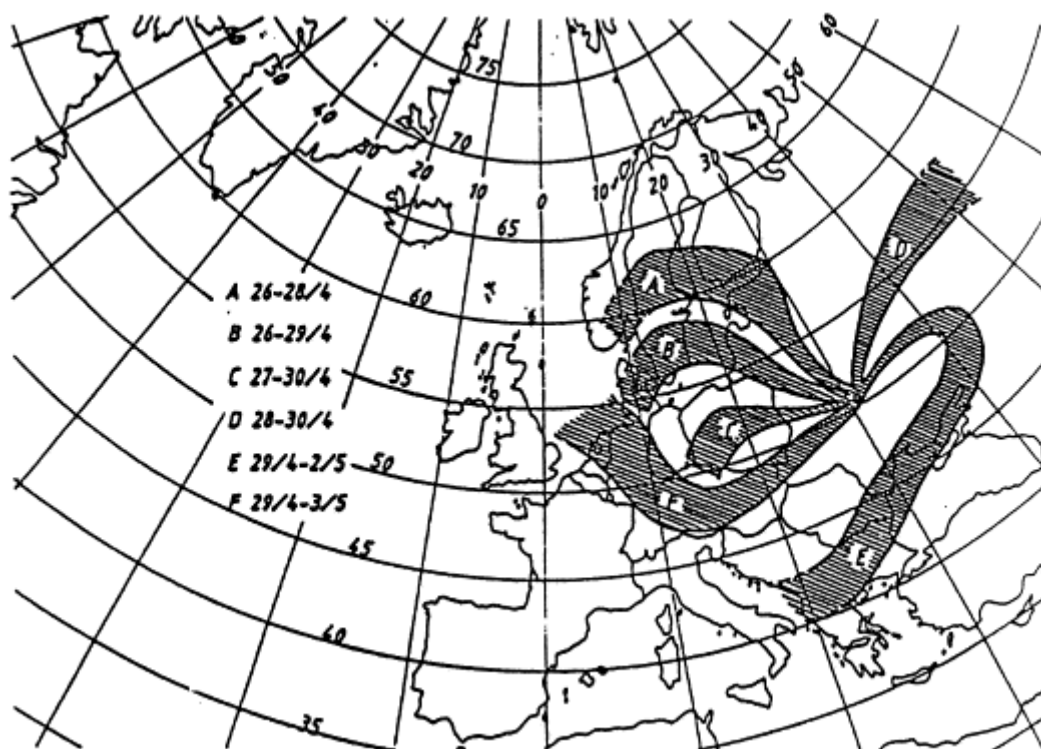


Figure I-3 : Zones de dispersion de la radioactivité entre le 26 avril et le 3 mai 1986 après l'accident de Tchernobyl (document de l'institut royal météorologique de Belgique, issu de Galle et al., 2003)

Ces distances montrent effectivement le caractère global d'une telle dispersion. D'après un rapport récent des Nations Unis (IAEA, 2005), la radio-exposition de la population a causé 4 000 cancers de la thyroïde, essentiellement chez les enfants. La mortalité liée au cancer a augmenté de quelques pour-cent pour les populations les plus exposées (liquidateurs, évacués et résidents de la zone strictement contrôlée). Ces

chiffres souffrent cependant de diverses polémiques liées à l'impact socio-médiatique de l'accident. Le cas d'un accident nucléaire suivi d'une dispersion de radionucléides montre que l'échelle de temps est importante, tout comme celle de distance. Dans le cas de Tchernobyl et d'un incendie de matières radioactives, les particules émises sont susceptibles de retomber à proximité de la centrale. La pollution accidentelle devient alors chronique.

L'accident de Fukushima, Japon (2011) a également émis des radionucléides dans l'atmosphère. Suite à un important tsunami conséquent à un séisme dans la région du Kansaï, la centrale de Fukushima Daichi a perdu les systèmes de refroidissement des réacteurs nucléaires. Ceux-ci n'ont pu être refroidis et la création d'hydrogène et d'oxygène a été responsable de l'explosion de trois des six réacteurs de la centrale. Un nombre important de particules a été relâché dans l'atmosphère. Comme dans le cas de Tchernobyl, les périodes et distances d'effets de cette dispersion de radionucléides dans l'air sont importantes (Korsakissok et al., 2013). Les rejets se sont déroulés entre le 12 et le 25 mars 2011 et des particules ont été mesurées à travers le monde entier, à des niveaux jugés trop faibles pour engendrer des conséquences sur les organismes. Les rejets dangereux pour la santé humaine ont été mesurés dans une zone de 20 à 30 km autour de la centrale. Aucune victime directe n'est à déplorer. Les conséquences sur les vies humaines à proximité de la centrale n'ont pas été évaluées. Une zone d'exclusion a été fixée pour un rayon de 30 km.

Ces deux accidents nucléaires font état de dispersion de particules radioactives. Les différents radionucléides émis dans les rejets sont de différentes natures et de différentes masses volumiques. Les particules les plus lourdes retombent à proximité du site alors que les particules volatiles ou les gaz peuvent franchir de longues distances.

Concernant les deux cas précédents de pollution atmosphérique, il s'agit de transport dans l'atmosphère de particules et leur dépôt consécutif en surface. Il s'agit d'un phénomène ayant une cinétique lente avec une échelle spatiale très large. La prévision de la dispersion atmosphérique permet dans ces cas là d'anticiper les mesures de précaution et de protection en déterminant les zones à risques. La topographie du site industriel n'est que peu déterminante dans l'évolution de la dispersion atmosphérique pour ce type d'accidents.

D'autres accidents industriels, aux conséquences moins étendues géographiquement mais avec des bilans également lourds en terme de vies humaines ont été observés dans la deuxième partie du XX^{ème} siècle.

A une échelle de distance d'effets intermédiaires, l'accident de Seveso en Italie (Juillet 1976) peut être mentionné. Une usine chimique, ICMESA, produisant du Trichlorophénol à travers une réaction exothermique a rejeté par un conduit d'évacuation une grande quantité de matière organique jusqu'à 6 km au sud de l'installation durant environ 1h. Une partie de la zone la plus contaminée fut dépeuplée, représentant environ 750 personnes et 196 cas de chloracné furent déclarés, en grande partie sur des enfants (Pesatori et al., 2003). Par ailleurs, Cavallaro et al. (1982) montrent la limite de certains modèles dans la prévision de la dispersion. Effectivement, une hypothèse d'écoulement homogène ne peut être utilisée qu'en première affirmation. Les effets mécaniques des obstacles tels que les bâtiments et la végétation ainsi que l'effet thermique des terres arables ont modifié largement l'écoulement et la turbulence atmosphérique. A cette échelle de dispersion, on s'aperçoit que les obstacles et la perturbation de l'écoulement ont un impact de première importance sur la dispersion.

Le cas de Bhopal est un exemple tragique de dispersion locale. Considéré comme le pire accident industriel survenu dans la société actuelle, la situation de Bhopal en Inde

concerne le rejet accidentel de méthylisocyanate (MIC). L'entreprise Union Carbide fabriquait des pesticides. Les conditions de sécurité sont mises en défaut par une situation financière compliquée. Ainsi, une opération mal maîtrisée et des dispositifs de secours inefficaces ou tout simplement inexistantes ont engendré le rejet dans l'atmosphère de plus de 32 tonnes de MIC (Havens et al., 2012). La détermination de la densité du mélange de MIC sous forme gaz ou liquide est décisive : dans le cas d'un mélange plus lourd que l'air, le nuage va s'étaler au sol alors que dans le cas d'un gaz passif, c'est l'écoulement qui va diriger la dispersion. Selon le modèle sélectionné, des résultats largement différents peuvent être obtenus (Sharan and Gopalakrishnan, 1997; Singh and Ghosh, 1987). L'importance est mise sur la caractérisation du terme source. Effectivement, des publications antérieures à Havens négligent l'impact de la gravité sur l'écoulement du panache de MIC, et par conséquent réduisent certainement la zone d'impact de cette dispersion. L'étude compare la dispersion de MIC à d'autres produits plus lourds que l'air tels que les GPL (à faible température) ou les jets d'aérosols qui sont des substances qui, de part leur température (GPL) ou leur état (aérosols), ont une densité plus importante que l'air. Dans ces cas là, l'étalement au sol et donc dans la largeur sont bien plus importants que dans le cas d'une dispersion latérale exclusivement passive. L'accident industriel de Bhopal peut être considéré comme l'un des plus graves de notre ère. Selon les différentes sources, le bilan direct varie entre 1 754 et 2 500 morts. Le nombre de personnes affectées est estimé en 2004 à 800 000 personnes (ARIA, 2008).

L'accident de Via Reggio couple les phénomènes de dispersion atmosphérique et d'explosion. Il s'agit d'un accident de train transportant du GPL. Suite à un déraillement, une brèche s'est ouverte sur une des cuves transportées et a entraîné la libération de ce gaz inflammable. La dispersion du gaz s'est effectuée au ras du sol étant donné la densité du produit. Les obstacles et la topographie locale ont conduit à un chemin préférentiel du nuage formé. Celui-ci s'est déplacé à l'intérieur du quartier à proximité de la zone d'accident et le gaz s'est infiltré dans les sous-sols des maisons. La présence de carrefours et de bâtiments a largement participé à la modification du nuage (forme, direction). Les conditions météorologiques de l'événement sont particulières puisque, dans ces conditions précises, la gravité du nuage est prépondérante dans son déplacement. Le vent n'est pas pris en compte dans ce cas là. La dilution du nuage s'est effectuée selon deux phénomènes. En se répandant sur le sol le nuage entraîne de l'air frais en hauteur, ce qui favorise sa dilution. D'autre part, les échanges thermiques entre le sol et le nuage existent mais sont là aussi négligeables devant la topographie du site (Brambilla et al., 2009). L'explosion s'est vraisemblablement déclenchée à partir des maisons. Cet accident a causé la mort de 32 personnes et des dommages matériels importants. La cinétique de la dispersion est rapide dans ce cas : 100 secondes séparent l'accident de l'inflammation. La zone d'impact est restreinte à proximité de la voie ferrée, ce qui reflète le caractère local de l'impact de cet événement.

Peu d'accidents de dispersion atmosphérique à échelle locale sont répertoriés à travers le monde. Cela peut être expliqué de deux manières. Les cas de dispersion sur le long terme et sur de longues distances ont des impacts médiatiques bien plus importants que ceux cantonnés au site industriel (cas du nucléaire). D'autre part, la dispersion de produit inflammable peut être suivie d'une inflammation du gaz pouvant mener à une explosion. Dans ce cas là, l'exposition médiatique est concentrée sur l'explosion plutôt que sur la dispersion.

Afin de pouvoir gérer au mieux ces événements, il est nécessaire de connaître les conditions d'occurrence de tels phénomènes et ainsi de décrire les règles permettant de limiter leur impact.

I-1.3. Le site industriel : zone d'aléa

La dispersion atmosphérique accidentelle de substances dangereuses peut être une étape finale ou intermédiaire d'un accident. La cinétique liée à ce genre d'événements est spécifique. Cette étude s'intègre dans un projet global visant à prévoir les conséquences directes et indirectes de la rupture d'un réservoir de stockage d'un site industriel. Le processus s'exécutant depuis la brèche à la dispersion atmosphérique est complexe. Tout impact sur un réservoir stockant des produits toxiques ou inflammables ne va pas conduire à une dispersion atmosphérique. Pour expliquer le processus amenant à la dispersion atmosphérique, nous nous intéressons à la fois à la substance mise en jeu et à son environnement.

I-1.3.1 Substances et stockages

Les substances stockées sur un site industriel présentent des caractéristiques multiples. Elles peuvent être stables, volatiles, toxiques, corrosives, inflammables, etc. Aux dangers intrinsèques des matières s'ajoutent les dangers liés aux conditions de stockage. Le stockage a pour objectif de garantir trois priorités : la sécurité, un coût d'exploitation raisonnable et une mise en place aisée. Le mode de stockage est ainsi directement lié à la substance à stocker. Son degré de volatilité à la température de stockage permet de définir le type de réservoir utilisé. La tension de vapeur absolue à température ambiante permet de récupérer le Point Normal d'Ebullition (PNE) à pression atmosphérique. On peut alors déterminer les conditions de température et de pression du stockage. Si l'on prend l'exemple du méthane, l'encyclopédie air liquide (AIR LIQUIDE, 2014) donne une valeur de PNE de $-161,48^{\circ}\text{C}$. Cela correspond à un état gazeux à température ambiante. Il est donc nécessaire de refroidir cet alcane afin de le stocker en quantité suffisante à pression atmosphérique.

Différentes structures permettent de valider ces conditions. Chamayou (1997) a répertorié les grandes familles de stockage rencontrées sur les sites industriels. Le cylindre vertical ou horizontal et la sphère sont principalement utilisés. Concernant le méthane, les conditions de stockage requises sont un réservoir cryogénique à double paroi isolées.

Les cylindres verticaux sont les plus représentés sur un site pétrochimique. La figure ci-après montre un site pétrochimique et l'abondance de cylindres verticaux :



Figure I-4 : Vue satellite du site pétrochimique de Lavéra, Martigues, France (copie d'écran Google Map ® 2014)

La présence d'un grand nombre de stockages différents, ainsi que de conduites de gaz liquéfié ou de réfrigérants sur un vaste site industriel multiplie les termes sources possibles ainsi que les conséquences successives éventuelles.

I-1.3.2 Influence de l'environnement

Un autre élément déterminant dans le transport de la matière est l'environnement de la fuite. L'orographie particulière d'un terrain peut influencer grandement le déplacement d'un panache ou d'une bouffée. Des zones de recirculations peuvent apparaître au niveau des changements brusques d'élévations de terrain. La photographie suivante présente ce type de morphologie. Par ailleurs, la présence de cuvettes de rétention, peut également affecter la dispersion. En influant sur l'écoulement, une zone de recirculation peut amplifier des phénomènes d'évaporation de nappes (Forestier, 2011).

La présence d'obstacles (généralement d'autres stockages) peut grandement perturber l'écoulement et le comportement de la substance elle-même (I-2.4).



Figure I-5 : Stockages et conduites de fluides - Site de Lavéra Martigues - Photo Laurent Valencia ©

I-1.3.3 Scénarios d'aléas à prendre en compte

La matière et l'état physico-chimique dans lequel se trouve la substance influencent fortement la chaîne logique des événements potentiels. Un stockage sous pression ou réfrigéré, impacté par un projectile, peut engendrer un jet monophasique, diphasique ou bien une simple vidange gravitaire selon le produit concerné. Dans la continuité de ces événements conduisant à un transport de matière dans l'air, le jet et la vidange amènent à la création d'une flaque de produit, celle-ci peut s'évaporer, notamment par effet de la chaleur fournie par le sol. On mesurera alors une concentration. Il en est de même pour l'évaporation du jet. La figure ci-après illustre les différentes éventualités suite à la brèche sur un stockage de produit liquide ou gazeux.

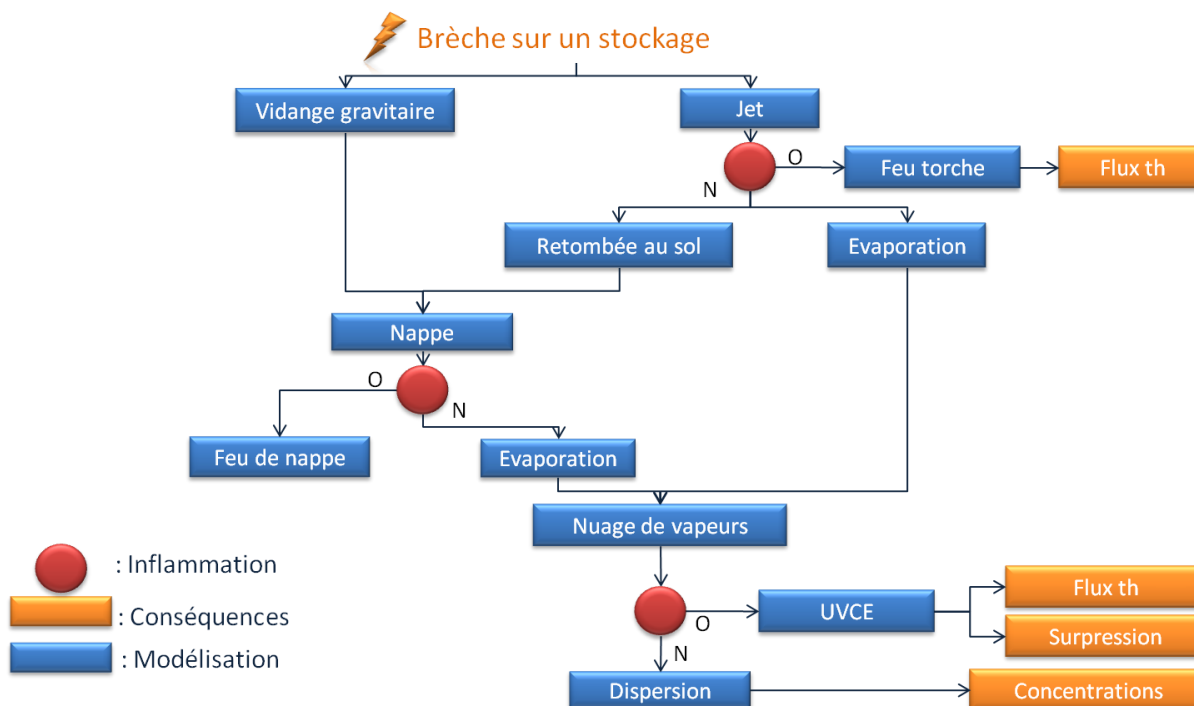


Figure I-6 : Illustration de la chaîne des conséquences de l'impact d'un projectile sur un stockage industriel.

Si ces situations ne sont pas exhaustives, plusieurs scénarios de dispersion atmosphérique se dégagent. Cette dispersion peut notamment être continue, ou sous formes de bouffée, constituée de gaz plus denses que l'air ou bien considérée comme passive.

Afin de pouvoir anticiper au mieux les risques potentiels, des mesures ont été progressivement mises en place au niveau français et européen. Les méthodes d'évaluation de la vulnérabilité ont émergé et contribué à la diminution des risques industriels.

I-1.4. Gestion des risques industriels

Les conséquences, importantes des accidents évoqués précédemment sont en partie dues à un accroissement des zones urbaines jusqu'à la proximité de sites industriels potentiellement dangereux. L'exemple de l'accident de l'usine AZF de Toulouse en France illustre cette problématique (Taveau, 2010). L'existence de tels risques d'accidents dans l'industrie implique une législation adaptée de la part des instances gouvernantes afin de limiter leur occurrence.

La réglementation française s'appuie sur un classement des infrastructures selon le risque associé. Ainsi, les Installations Classées pour la Protection de l'Environnement (ICPE) sont soumises à déclaration, enregistrement ou autorisation pour fonctionner (Ministère de l'écologie, 2014). Ces règles distinguent les événements chroniques des événements accidentels. Le cas accidentel est étudié à travers une étude de danger. Celle-ci permet aux inspecteurs des installations d'attribuer les différents régimes. Elle est effectuée par l'industriel pour prévoir et anticiper les conséquences potentielles d'un accident en évaluant différents scénarios. Chaque phénomène dangereux susceptible d'intervenir est étudié. La dispersion atmosphérique est donc un point important de cette étude, à mettre en relation des autres phénomènes impliqués et la nature du produit.

Afin d'atteindre un niveau de risque aussi bas que possible, des valeurs de référence sont établies.

En ce qui concerne la dispersion atmosphérique, on définit ainsi des valeurs limites correspondant à des concentrations dans l'air de composés chimiques. Selon le danger lié à la substance, inflammable ou toxique, le seuil s'applique différemment. En ce qui concerne l'impact sur une personne, l'indice pris en compte correspond à une concentration que peut respirer une personne pendant un temps déterminé sans risques d'altération de sa santé, même si des modifications physiologiques réversibles sont parfois tolérées. Aucune atteinte organique ou fonctionnelle de caractère irréversible ou prolongée n'est raisonnablement prévisible à ce niveau d'exposition.

- ▶ Un effet toxique instantané se manifeste pendant ou après une exposition très brève (quelques secondes à une ou deux minutes) par des effets aigus fonctionnels et/ou lésionnels pouvant entraîner notamment une perte de connaissance, un coma ou un arrêt cardiaque.
- ▶ Un effet toxique immédiat se manifeste après une exposition de courte durée par une irritation aiguë des muqueuses respiratoires ou de la peau, par une narcose engendrant une inaptitude fonctionnelle, par une atteinte cellulaire irréversible.
- ▶ Un effet toxique à long terme se manifeste après des expositions prolongées, répétées pendant des semaines, voire des années par la survenue de cancers, d'effets toxiques sur la fonction de reproduction, d'atteintes du système nerveux, de réaction d'hypersensibilité retardée...

Nos travaux s'appliquant à l'échelle d'un site industriel et en champ proche, nous nous intéresserons uniquement aux effets toxiques instantanés et immédiats. Pour fixer des limites, il est nécessaire de caractériser une valeur de concentration et une durée d'exposition. La sensibilité des individus rentre également en jeu sur les effets mais n'est pas prise en compte dans le calcul des indices de référence.

Dans le domaine professionnel, la Valeur Limite d'Exposition (VLE) caractérise la concentration maximale que peut respirer une personne pendant 15 minutes sans risques d'altération pour sa santé. Le concept similaire américain est le « Threshold Limit Value – Short-Term Exposure Limit » ou TLV-STEL. Le but d'instaurer une telle valeur pour chaque type de produit est la protection des salariés contre les effets aigus/immédiats ou préjudiciables à court terme d'une exposition momentanée. L'équivalent pour les expositions de longue durée est la Valeur Moyenne d'Exposition (VME) ou « Threshold Limit Value – Time Weighted Average » (TLV-TWA).

Dans le domaine accidentel, qui est celui qui nous intéresse particulièrement, le Seuil des Effets Irréversibles (SEI) caractérise la concentration maximale de substance dangereuse dans l'air pour un temps d'exposition donné en dessous de laquelle, chez la plupart des individus, on n'observe pas d'effets irréversibles. Le Seuil des Effets Létaux (SEL) correspond à la concentration maximale de substance dangereuse dans l'air pour un temps d'exposition donné en dessous de laquelle chez la plupart des individus, on n'observe pas de décès. L'IDLH (Immediately Dangerous to Life or Health) est la valeur maximale pour laquelle un niveau d'exposition en milieu professionnel pour une durée de 30 minutes n'entrave pas l'évacuation des individus, ni n'induit d'effets nocifs irréversibles.

Ces valeurs permettent d'évaluer des zones à risques en cas de dispersion atmosphérique de substances toxiques. Dans le cas de substances inflammables, les valeurs de concentrations à prendre en compte sont les limites inférieures et supérieures d'inflammabilité (LII et LSI). Les zones comprises entre ces limites sont susceptibles de

s'enflammer. Il convient alors d'utiliser des modèles permettant de calculer les surpressions et les flux thermiques engendrés par l'explosion du nuage. Des séquences complexes d'événements peuvent se produire, en effet domino.

L'approche souvent utilisée pour appréhender les dangers liés aux activités de l'industrie du pétrole, de la pétrochimie ou de la chimie croise la dangerosité du phénomène et la vulnérabilité de l'enjeu. Un grand nombre de méthodes d'analyse de risques existent (Tixier et al., 2002). Dans la plupart de ces approches, il est nécessaire de caractériser au mieux le danger. Pour cela, il est nécessaire de s'appuyer sur des modèles. Selon le nombre de scénarios à prendre en compte, plusieurs modélisations sont à effectuer. Aussi, l'intérêt d'utiliser la modélisation numérique réside dans la possibilité de tester un nombre important de paramètres différents. Ce faisant, des contraintes peuvent s'appliquer, notamment en ce qui concerne la vitesse de calcul.

En reliant une carte des zones de danger à une carte de la vulnérabilité des biens et des personnes, les autorités et exploitants peuvent s'entendre dans le dimensionnement optimum, garantissant à la fois la rentabilité de leur activité et la sécurité. La modélisation fait partie intégrante de ce processus. Les choix des modèles utilisés et les scénarios étudiés se doivent d'être motivés afin que l'inspection des installations valide l'étude de danger. L'apport de la modélisation de la dispersion atmosphérique est donc au centre de la réalisation des telles études.

I-1.5. Conclusion

La dispersion atmosphérique accidentelle de substances toxiques ou inflammables a été responsable de conséquences parfois désastreuses sur la population depuis la révolution industrielle. La catastrophe de Bhopal a montré que le danger potentiel d'une installation doit être mis en rapport avec la vulnérabilité de son environnement. L'urbanisation croissante a induit la présence de zones à forts enjeux à proximité d'installations à risques. La réglementation européenne et nationale impose désormais des études de danger sur les activités industrielles, selon leur nature. Ces études permettent de quantifier et localiser les différentes zones à risques en fonction des seuils de concentration définis par la réglementation. Afin d'étudier différents scénarios d'accidents, en fonction de conditions initiales représentatives, la modélisation est utilisée et oriente la solution optimale, compromis d'objectifs économiques et de sécurité des enjeux. Dans le cas de la gestion de crise, la modélisation de la dispersion atmosphérique est utilisée pour déterminer les actions de secours et de lutte contre le danger. Dans ce dernier cas, le modèle utilisé se doit d'être rapide alors que dans une phase de dimensionnement, la précision prend plus de poids. Aussi, différents modèles sont adaptés à différentes situations. Il convient alors de bien caractériser l'environnement et les phénomènes intervenants pour lesquels le modèle va être utilisé.

I-2. Mécanisme des écoulements atmosphériques

"It was one of those days when it's a minute away from snowing. And there's this electricity in the air, you can almost hear it. And this bag was just... dancing with me ... like a little kid begging me to play with it. For fifteen minutes. That's the day I realized that there was this entire life behind things"

American Beauty

Dans ce chapitre, l'environnement atmosphérique du site industriel est décrit afin de présenter les caractéristiques importantes à prendre en compte dans ce travail de modélisation de la dispersion. A partir d'une description générale de l'atmosphère, le focus sera mis sur la couche immédiatement à proximité du sol, la couche limite atmosphérique et notamment la couche de surface. Après avoir décrit le cas du champ libre, la perturbation d'un tel écoulement par un obstacle sera détaillée.

I-2.1. L'atmosphère

L'atmosphère terrestre est l'enveloppe gazeuse de la Terre. Elle est décomposée en plusieurs sous-couches, chacune ayant des particularités qui lui sont propres. La détermination des différentes couches a été effectuée en fonction de l'évolution du profil de température vertical, à savoir :

- ▶ La troposphère. Cette couche correspond à la couche la plus basse. A l'intérieur de celle-ci se produisent les phénomènes météorologiques ainsi que les mouvements atmosphériques horizontaux et verticaux. Elle englobe la partie allant du sol à 8 à 15 km de hauteur. Son profil de température vertical montre une évolution décroissante en fonction de l'altitude.
- ▶ La stratosphère. Il s'agit de la couche juste au dessus, de 15 à 50 km d'altitude. La température suit un profil inverse au précédent, dû à la captation du rayonnement UV par la couche d'ozone.
- ▶ La mésosphère. Le profil de température décroît de nouveau avec l'altitude jusqu'à 50 à 80 km.
- ▶ La thermosphère. Le profil s'inverse de nouveau de 80 à 350-800 km d'altitude.
- ▶ L'exosphère. Il s'agit de la couche la plus haute de l'atmosphère. Elle couvre la zone où les particules sont encore attirées gravitationnellement à la Terre. Elle s'étend jusqu'à 50 000 km d'altitude.

L'illustration suivante reprend les éléments énoncés ci-dessus :

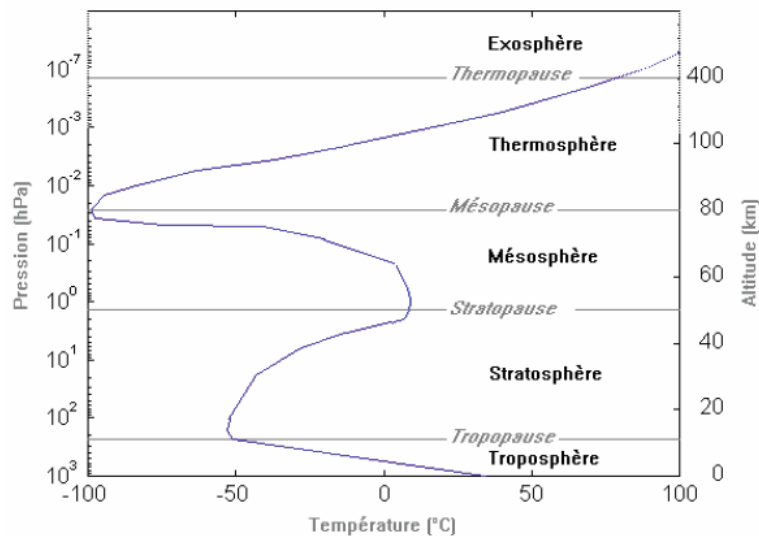


Figure I-7 : Schéma structural de l'atmosphère terrestre et son profil de température (Korsakissok, 2009)

L'énergie nécessaire au mouvement de l'air est fournie par le soleil. Celui-ci chauffe de manière inhomogène la surface terrestre, le rayonnement solaire n'étant pas en tout point perpendiculaire à la surface du globe. La chaleur est redistribuée à la couche d'air entourant la Terre. La différence de température et donc de pression entre deux zones conduit à un déplacement de masses d'air, causant de grands mouvements généraux. Les forces modifiant ce mouvement convectif sont la force de Coriolis et la force de frottement. La première est liée à la rotation de la Terre et induit une déviation des trajectoires. La seconde se manifeste au contact de surface, et ralentit la lame d'air directement à proximité. On définit la zone où l'influence des frottements de l'air est négligeable comme étant l'atmosphère libre. La zone la plus proche de la surface est la Couche Limite Atmosphérique (CLA). Les effets de la force de Coriolis s'y estompent progressivement au profit des effets mécaniques des frottements et des effets thermiques.

Concernant notre problématique, nous ne nous intéressons qu'aux premiers mètres d'altitude ainsi qu'à une zone définie par une largeur maximale d'1 km. Aussi, il n'apparaît pas pertinent de prendre en compte les effets de la force de Coriolis. On se concentre sur l'étude de la zone où les effets des frottements sur la surface et des obstacles sont susceptibles d'apporter des modifications sur l'écoulement.

I-2.2. La couche limite atmosphérique.

Les mouvements à l'intérieur de la CLA sont fortement perturbés par l'environnement. Les frottements au contact de surface présentant des discontinuités participent activement aux mouvements de l'air. Des perturbations sont apportées par le relief et les obstacles. Dans le cas d'obstacles importants tels que les montagnes, les perturbations peuvent atteindre l'atmosphère libre. Dans le cas d'obstacles moins imposants tels que les stockages industriels, l'impact se limite à l'environnement proche. En plus des perturbations mécaniques s'ajoutent les perturbations thermiques. Selon le type de surface impactée par les rayons du soleil, la chaleur fournie va être plus ou moins importante. De ce fait, l'écoulement va être perturbé localement. Un exemple de ce type de phénomène est la brise de mer. Nous détaillerons ci-après le comportement des écoulements dans la couche inférieure de la CLA.

I-2.2.1 Caractéristiques

La couche la plus basse de la troposphère a une épaisseur variant entre 100 m et près de 2 km de hauteur. De fortes interactions entre la surface de la Terre et l'atmosphère, sur des échelles de temps de l'ordre de quelques heures à la journée se traduisent par des effets mécaniques et thermiques. Les frottements mécaniques ralentissent l'écoulement directement en contact avec la surface (vitesse nulle). Par effet de cisaillement, un profil vertical de vitesse est identifiable. D'autre part, les échanges de chaleur entre la surface et l'air sont hétérogènes et participent à la création de tourbillons. Deux sous-couches peuvent ainsi être répertoriées :

- ▶ La couche d'Ekman, où la force de Coriolis influe sur le mouvement des masses d'air.
- ▶ La Couche Limite de Surface (CLS) s'étend du sol jusqu'à environ un dixième de la hauteur de la CLA. Celle-ci est constituée d'une sous-couche rugueuse. Elle est le siège des écoulements dépendant des paramètres locaux de la surface (bâtiments, végétation, ...). De nombreux tourbillons sont générés. La turbulence est intermittente et instationnaire. La canopée est considérée comme la zone de la sous-couche rugueuse englobant les obstacles.

La dimension de ces différentes couches est donnée à titre d'exemple car celles-ci évoluent au cours du temps et en fonction de l'environnement. En l'absence d'entraînement convectif, la stratification thermique dans une atmosphère de stabilité neutre implique une perte de 1°C tous les 100 m d'altitude. Néanmoins, avec le réchauffement du sol par le soleil en journée, on voit apparaître une couche convective qui comporte des caractéristiques différentes. Les températures diminuent plus vite avec l'altitude et un mouvement vertical est observé. A contrario, après le coucher du soleil, le sol contribue moins à ce mouvement convectif et les températures diminuent moins vite avec l'altitude. La hauteur de la couche de surface diminue. L'illustration ci-après montre l'évolution de la couche de surface au cours du temps (Kaimal et Finnigan, 1994):

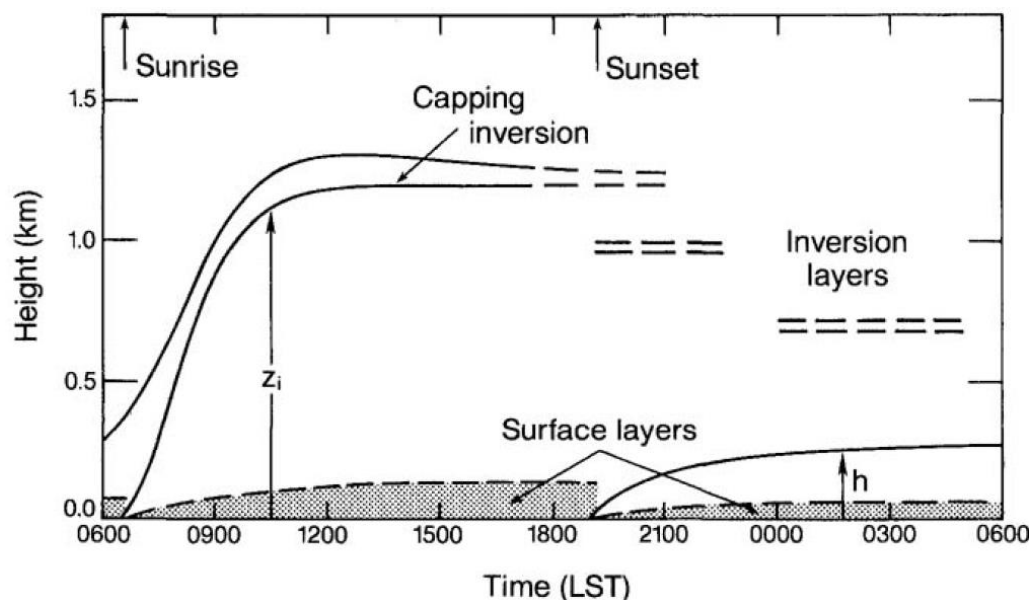


Figure I-8 : Evolution de la couche limite convective et la couche limite stable en fonction du temps et du transfert de chaleur du sol vers l'atmosphère. (Kaimal and Finnigan, 1994)

L'alternance des cycles diurnes et nocturnes crée des flux thermiques variables, conditionnant des couches stables et instables.

Au regard de ces caractéristiques, l'étude de la dispersion atmosphérique d'une substance suite à une fuite sur un site industriel (hors nucléaire) se situe dans le cadre de la sous-couche rugueuse. La turbulence associée à ces caractéristiques devra être prise en compte.

I-2.2.2 Echelles d'observations des mécanismes de l'atmosphère

L'atmosphère peut être séparée en plusieurs échelles, chacune ayant des caractéristiques propres ou des mécanismes spécifiques. Néanmoins, ces mécanismes ne sont pas indépendants. Chacun est influencé par l'ensemble des autres. Les échelles planétaires et synoptiques s'intéressent aux mouvements globaux tels que les anticyclones ou dépressions de grande ampleur. Les caractéristiques de ces échelles sont étudiées en météorologie. Deux autres échelles se doivent d'être mentionnées ici, la méso échelle et la micro échelle. La méso échelle correspond à l'échelle locale. C'est l'échelle de la ville. Aux premiers abords, il semble peu intéressant de prendre en compte les caractéristiques de cette échelle. Pourtant, la présence d'un lac, par exemple, à proximité d'un rejet, pourrait largement influencer les concentrations de polluants au cours du temps et de l'espace. La micro échelle est celle qui nous concerne directement. Son périmètre d'étude est limité à des dimensions inférieures à quelques kilomètres. C'est l'échelle du site industriel susceptible de subir un accident majeur. Les mécanismes agissant à cette échelle seront spécialement étudiés dans la suite.

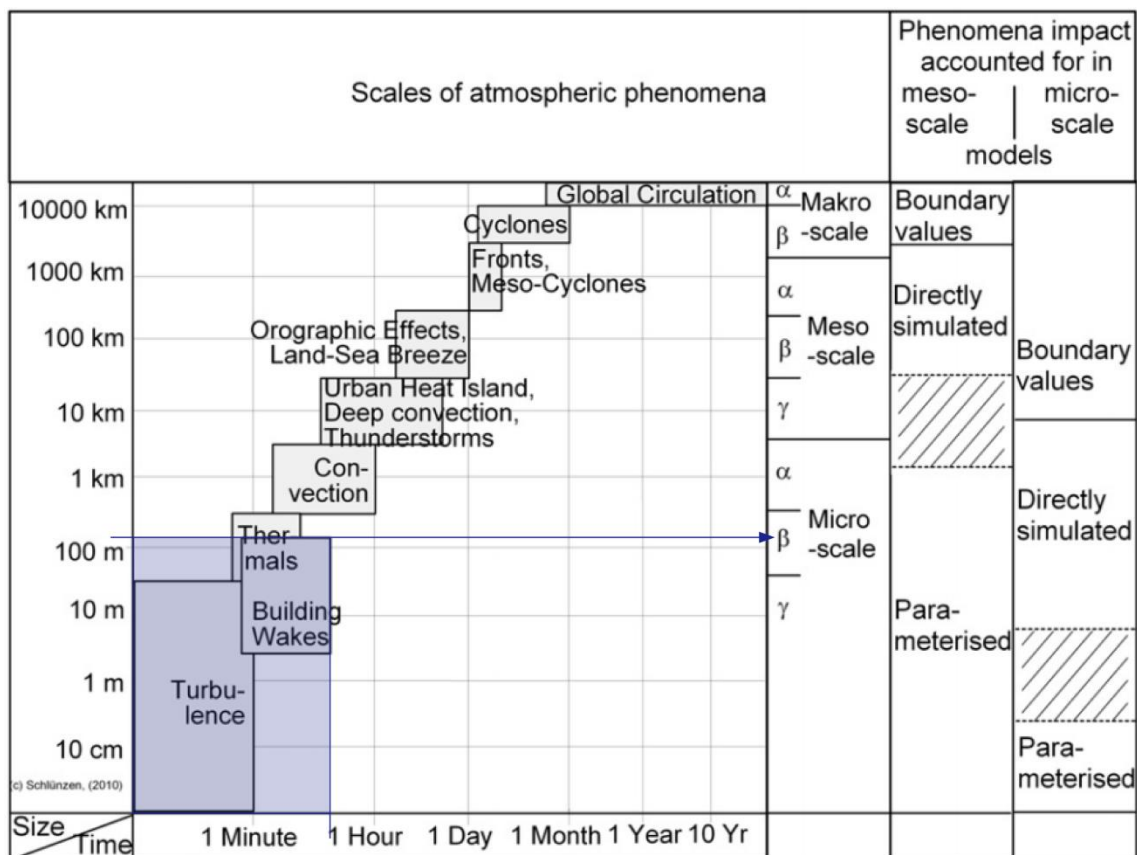


Figure I-9 : Illustration des différentes échelles d'observation des mécanismes de l'atmosphère (Schlünzen et al., 2011) et définition de la zone étudiée dans ce travail de recherche.

Les phénomènes que nous souhaitons représenter dans cette étude concernent notamment l'effet des obstacles sur la dispersion. Les distances à prendre en compte sont de l'ordre de 100 mètres à moins d'un kilomètre de la source. L'échelle temporelle associée se situe entre la minute et l'heure, ces intervalles de temps et d'espace étant représentatifs de la micro échelle.

I-2.3. Equations fondamentales

Afin de décrire la CLA, il convient de détailler les équations caractérisant sa dynamique. Les écoulements atmosphériques sont décrits par les équations de mécanique des fluides de Navier-Stokes, traduisant la conservation des quantités physiques suivantes : la masse, la quantité de mouvement et l'énergie.

I-2.3.1 Représentation du mouvement

Le mouvement d'un fluide peut être envisagé de deux façons différentes. La représentation eulérienne décrit l'évolution d'une grandeur donnée à un point d'observation fixe. La dérivée partielle (ou eulérienne) de cette grandeur est utilisée pour suivre ces variations.

La représentation lagrangienne décrit l'évolution d'une grandeur à partir d'un référentiel correspondant à une particule dans l'écoulement. La dérivée particulaire (ou lagrangienne) de la grandeur considérée est utilisée. Elle décrit les variations locales de cette grandeur ainsi que les modifications liées au déplacement de la particule.

On lie les descriptions lagrangiennes et eulériennes par la décomposition de la dérivée particulaire :

$$\frac{DA}{Dt} = \frac{\partial A}{\partial t} + (\vec{V} \cdot \text{grad})A \quad (I-1)$$

I-2.3.2 Equations de conservation

i. Conservation de la masse

Le bilan de masse du fluide sur un élément de volume (description eulérienne) s'écrit :

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho u_i) = 0 \quad (I-2)$$

ρ	Masse volumique	kg.m ⁻³
t	Temps	s
u_i	Vitesse dans la direction i	m.s ⁻¹

ii. Conservation de la quantité de mouvement

La relation traduisant le principe fondamental de la dynamique s'écrit :

$$\frac{\partial(\rho u_i)}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_i u_j) = \rho f_i + \frac{\partial}{\partial x_j}(\sigma_{ij}) \quad (I-3)$$

\vec{f}	Résultante des forces extérieures de volume
σ_{ij}	Tenseur des contraintes agissant à la surface de l'élément fluide (pression, contraintes visqueuses)

I-2.3.3 Prise en compte de la turbulence dans un écoulement

La résolution des ces équations est possible dans des régimes laminaires, les caractéristiques étant des fonctions déterministes de l'espace et du temps (vitesses, pression, température). Cependant, il est nécessaire de s'intéresser à des aspects particuliers des écoulements atmosphériques. La présence de tourbillons dans un écoulement est la marque de la présence de turbulence. Cette notion très ancienne est encore de nos jours difficiles à appréhender. Leonardo Da Vinci en faisait des croquis, détaillant notamment la cascade de tourbillons des plus gros vers les plus petits jusqu'à leurs dissipations :

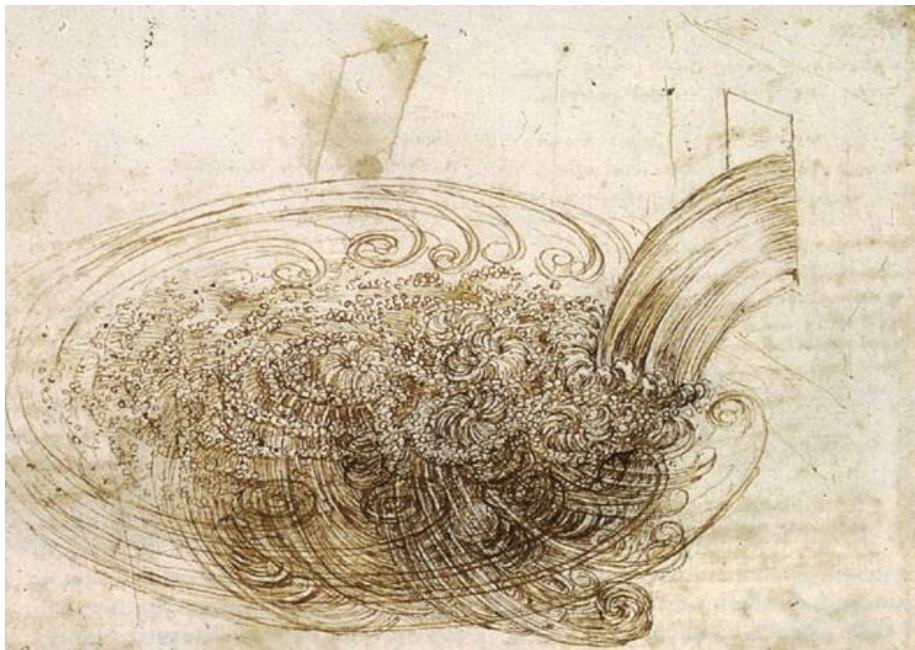


Figure I-10 : Figures de turbulence autour d'un obstacle (Da Vinci, 1508)

Les notions de turbulence et de tourbillons lui sont attribuées, bien que son analyse fût plus descriptive qu'analytique. Reynolds (1883) réalise une expérience permettant de mettre en évidence le changement de régime sur le mélange de colorant. Cette transition est définie par des valeurs d'un nombre adimensionnel portant son nom et décrivant le rapport entre les forces d'inertie et de viscosité :

$$R_e = \frac{\rho D U}{\mu} \quad (I-4)$$

R_e	Nombre de Reynolds	-
ρ	Masse volumique du milieu	kg.m ⁻³
D	Dimension caractéristique du milieu	m
U	Vitesse de l'écoulement	m.s ⁻¹
μ	Viscosité dynamique du milieu	kg.m ⁻¹ .s ⁻¹

La couche limite atmosphérique est considérée comme turbulente, les forces d'inertie étant largement supérieures aux forces de viscosité :

Masse volumique de l'air	1,2 kg.m ⁻³
Hauteur de la couche limite atmosphérique	200 – 2 000 m
Vitesse de l'écoulement	2-20 m.s ⁻¹
Viscosité dynamique	1,8×10 ⁻⁵ kg.m ⁻¹ .s ⁻¹
Nombre de Reynolds	2,67 x 10⁷ – 2,67 x 10⁹

Cette turbulence d'origine mécanique est provoquée par les effets de frottements au niveau du sol. Effectivement la condition d'adhérence au contact d'une paroi amène à une stratification verticale du champ de vitesse horizontal. La valeur du nombre de Reynolds calculée implique toujours la présence de turbulence dynamique dans la CLA. Par ailleurs, la turbulence d'ordre thermique peut s'y ajouter. La turbulence thermique dépend du gradient vertical de température.

La stabilité atmosphérique est principalement fonction des mouvements thermiques liés aux différents transferts de chaleurs du sol à l'air. Le gradient vertical de température permet de caractériser la stabilité de l'atmosphère. Une atmosphère neutre implique l'absence d'effets du gradient thermique sur le mouvement vertical de l'air. Les forces de flottabilité ont tendance à atténuer les mouvements et la turbulence dans l'axe vertical. Une atmosphère instable est caractérisée par une turbulence thermique forte à partir du sol (réchauffement de la surface et échanges importants). Dans de telles conditions, le gradient de température de l'atmosphère est inférieur au gradient thermique de l'adiabatique. Ainsi, la masse d'un volume élémentaire déplacé adiabatiquement est plus petite que l'air environnant et le mouvement d'élévation est amplifié. Dans une atmosphère stable, le gradient thermique de l'atmosphère est supérieur au gradient thermique de l'adiabatique et l'effet inverse se produit, le volume élémentaire tend alors à se déplacer vers le bas.

La prise en compte de la stabilité est effectuée à partir de l'utilisation de la température potentielle θ (Kaimal et Finnigan, 1994). Cette température correspond à la température d'un volume d'air à la température absolue T , et à la pression p , ramené adiabatiquement à la pression de référence au niveau du sol. L'équation s'écrit :

$$\theta_p = T + \left(\frac{g}{c_p} \right) \Delta z \quad (I-5)$$

θ_p	Température potentielle	K
T	Température	K
g	Gravité	m.s ⁻²
C_p	Chaleur spécifique à pression constante	J.kg ⁻¹ .K ⁻¹

z	Altitude	m
-----	----------	-----

On a alors :

$$\frac{\theta_p}{\partial z} = \left(\frac{\partial T}{\partial z} + \frac{g}{c_p} \right) \quad (I-6)$$

Ainsi, à la pression de référence, T et θ sont égaux. Dans une atmosphère adiabatique, $\frac{\partial \theta}{\partial z}$ est nulle et la stabilité est neutre. Une atmosphère stable et une atmosphère instable auront respectivement une pente positive ($\frac{\partial \theta}{\partial z} > 0$) et négative ($\frac{\partial \theta}{\partial z} < 0$). Lorsque de l'humidité est présente de manière significative (influence sur la densité de l'air), la température virtuelle potentielle est incluse dans la définition :

$$\theta_v \equiv \left(\frac{\theta}{T} \right) T_v \quad (I-7)$$

$$T_v = T(1 + 0,61q) \quad (I-8)$$

T_v	Température virtuelle	K
q	Humidité spécifique	-

La température virtuelle est la température à laquelle l'air sec a la même densité que l'air humide à la même pression. En pratique, la différence entre température potentielle et température potentielle virtuelle est faible lorsque l'on se situe au-dessus d'un terrain.

Le lien entre stratification thermique et turbulence dynamique est effectué à travers l'introduction du nombre de Richardson de gradient R_i . Celui-ci permet d'évaluer la compétition entre production mécanique et thermique de l'énergie cinétique k :

$$R_i = \beta \frac{\partial \bar{\theta}_v / \partial z}{\left(\frac{\partial \bar{u}}{\partial z} \right)^2} \quad (I-9)$$

β	Coefficient de flottabilité	$m.s^{-2}.K$
θ_v	Température potentielle virtuelle	K

Le coefficient de flottabilité est défini comme suit :

$$\beta = \frac{g}{\bar{\theta}_v} \quad (I-10)$$

Ce nombre adimensionnel est simple à calculer puisqu'il dépend de gradients de quantités moyennes mesurables. Le tableau ci-après relie les valeurs du nombre de Richardson et la caractérisation de la stabilité de l'atmosphère :

Tableau I-1 : Stabilité de la CLA en fonction du nombre de Richardson

Valeur de R_i Stratification	
$R_i > 0$	Stable
$R_i < 0$	Instable
$R_i = 0$	Neutre

En dessous d'une valeur critique de 0,25, l'écoulement passe de turbulent à laminaire.

Un autre paramètre de stabilité important est la longueur de Monin Obukhov. Celle-ci est déterminée à partir d'une analyse dimensionnelle en appliquant le théorème de Vashy-Buckingham (Buckingham, 1914) à une couche limite de surface établie, en présence d'un flux de chaleur vertical H_0 :

$$\frac{z}{L_{MO}} = \frac{\kappa g z \left(\frac{H_0}{\rho c_p} \right)}{u_*^3 \theta_v} \quad (I-11)$$

H_0	Flux d'énergie reçu par le sol	$W.m^{-2}$
κ	Constante de Von Karman	-
u_*	Vitesse de friction	$m.s^{-1}$
L_{MO}	Longueur de Monin Obukhov	m

La longueur de Monin-Obukhov permet de caractériser la stabilité atmosphérique de la couche convective. La valeur z/L_{MO} traduit la compétition entre effets de cisaillement et flottabilité. La longueur L_{MO} est interprétée comme l'altitude à partir de laquelle la turbulence induite par les effets thermiques prévaut sur la turbulence due au cisaillement. Le signe de la longueur de Monin-Obukhov indique la stabilité de la CLA. Le tableau ci-après relie la longueur de Monin-Obukhov et la stabilité atmosphérique :

Tableau I-2 : Lien entre valeur de la longueur de Monin Obukhov et conditions de stabilité de la Couche Limite Atmosphérique

Valeur de L_{MO}	Stratification
$L_{MO} > 0$	Stable
$L_{MO} < 0$	Instable
$L_{MO} = 0$	Neutre

Comme énoncé précédemment, la turbulence résulte des forces de cisaillement (turbulence mécanique) et des instabilités thermiques (turbulence convective). Les structures tourbillonnaires sont rotationnelles, tridimensionnelles et varient donc en temps et en espace. Une large gamme de tourbillons est transportée par l'atmosphère. La théorie de la turbulence de Richardson établit une succession de différentes échelles de la turbulence. Les gros tourbillons sont instables et se divisent en tourbillons plus petits. Leur énergie cinétique est, elle aussi, divisée progressivement. La cascade d'énergie qui y est liée continue jusqu'à la création de tourbillons de très faible dimension. A ce moment, l'énergie cinétique se transforme en chaleur par dissipation visqueuse.

L'échelle de création des tourbillons les plus imposants est l'échelle intégrale. L'échelle de Kolmogorov (1941) est celle où les tourbillons sont suffisamment petits pour être dissipés en chaleur.

Cette turbulence se traduit par une fluctuation aléatoire des variables d'intérêt :

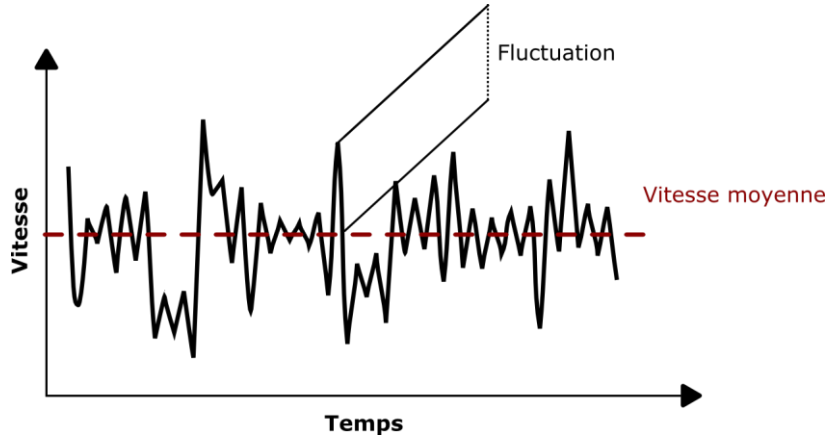


Figure I-11 : Exemple d'un enregistrement de la vitesse d'un écoulement turbulent représenté par les fluctuations autour de la valeur moyenne.

La turbulence et les fluctuations ont un impact sur la dispersion et il convient donc d'intégrer cette influence dans la suite de cette étude.

Dans l'objectif d'extraire la composante turbulente des équations fondamentales de la mécanique des fluides, Reynolds (1894) introduit donc l'opérateur de moyenne. Les grandeurs calculées sont décomposées en une addition d'une partie moyennée traduisant l'intensité et une partie fluctuante, traduisant l'impact de la turbulence. L'opérateur est défini comme suit :

Soit une variable f , sa fluctuation f' et sa moyenne \bar{f} :

$$f = \bar{f} + f' \quad (I-12)$$

$$\bar{f'} = 0 \quad (I-13)$$

$$\overline{f_1 f_2} = \bar{f}_1 \bar{f}_2 + \overline{f'_1 f'_2} \quad (I-14)$$

$$\overline{f_1 + f_2} = \bar{f}_1 + \bar{f}_2 \quad (I-15)$$

$$\overline{gf} = g\bar{f} \quad (I-16)$$

Avec g une constante.

Les variables considérées deviennent :

$$u_i = \bar{u}_i + u'_i \quad (I-17)$$

$$p = \bar{p} + p' \quad (I-18)$$

En considérant qu'à l'intérieur de la couche de surface, le fluide est incompressible, les équations moyennes décrivant l'écoulement turbulent deviennent :

i. Conservation de la masse

$$\frac{\partial u_i}{\partial x_i} = \frac{\partial (\bar{u}_i + u'_i)}{\partial x_i} = 0 \quad (I-19)$$

En prenant la moyenne de l'équation précédente :

$$\frac{\partial \bar{u}_i}{\partial x_i} = \frac{\partial (\bar{u}_i + u'_i)}{\partial x_i} = \frac{\partial \bar{u}_i}{\partial x_i} = 0 \quad (I-20)$$

ii. Conservation de la quantité de mouvement

En ne tenant compte que de la pesanteur pour les forces volumiques, il vient :

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} + \overline{u'_j \frac{\partial u'_i}{\partial x_j}} = g_i - \frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_i} + \nu \frac{\partial^2 \bar{u}_i}{\partial x_j \partial x_j} \quad (I-21)$$

$\nu = \frac{\mu}{\rho}$	Viscosité cinématique	$\text{m}^2.\text{s}^{-1}$
--------------------------	-----------------------	----------------------------

Or, on a :

$$\overline{u'_j \frac{\partial u'_i}{\partial x_j}} = \frac{\partial}{\partial x_j} (\overline{u'_i u'_j}) \quad (I-22)$$

L'équation devient :

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} = g_i - \frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_i} + \frac{1}{\rho} \frac{\partial}{\partial x_j} \left[\mu \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) - \rho \overline{u'_i u'_j} \right] \quad (I-23)$$

Les grandeurs $-\rho \overline{u'_i u'_j}$ sont les contraintes de Reynolds et traduisent le transport de quantité de mouvement par les fluctuations turbulentes de vitesse. Elles rajoutent 6 inconnues supplémentaires au système d'équations.

Le problème d'un écoulement turbulent est défini par :

Quatre équations (une équation de continuité et trois équations bilans de la quantité de mouvement)

10 inconnues :

$\bar{p}, \bar{u}_i \rightarrow$ Quatre inconnues

$-\rho \overline{u'_i u'_j} \rightarrow$ Six inconnues

Le système est donc ouvert. La modélisation de tels écoulements nécessitent donc des équations supplémentaires pour fermer le système.

iii. Fermeture du système d'équations

Différentes méthodes de résolution d'un tel système sont présentées ici. Les modèles de mécanique numériques des fluides sont basés sur la résolution eulérienne des équations de conservation, sur un maillage de l'espace. Selon le problème à modéliser, les équations de continuité, de conservation de la quantité de mouvement et de l'énergie sont résolues.

Plusieurs démarches peuvent être explicitées en fonction du degré de finesse recherché dans les caractéristiques attendues de la turbulence. Selon l'étude, il peut être déterminant de connaître la nature fine des tourbillons. Dans d'autres cas, des informations sur le comportement moyen peuvent suffire. Il convient alors de modéliser cette turbulence.

Différentes méthodes de modélisation de cette turbulence coexistent. Les plus couramment utilisées aboutissent à la résolution des équations moyennées au sens de

Reynolds. Il s'agit des méthodes RANS pour *Reynolds Averaged Navier-Stokes equations*. Une approche basée sur la résolution explicite des grandes structures et la paramétrisation des plus petites permet de simuler plus précisément la turbulence. Cette méthode est la Simulation des Grandes Echelles ou LES pour Large Eddy Simulation. Enfin, il est possible de résoudre de manière explicite et à toutes les échelles les équations de Navier-Stokes. Il s'agit de la Simulation Numérique Directe ou DNS pour Direct Numerical Simulation. Le schéma ci-après illustre ces différents modèles sur le diagramme d'évolution de l'énergie des tourbillons en fonction de leur nombre d'onde :

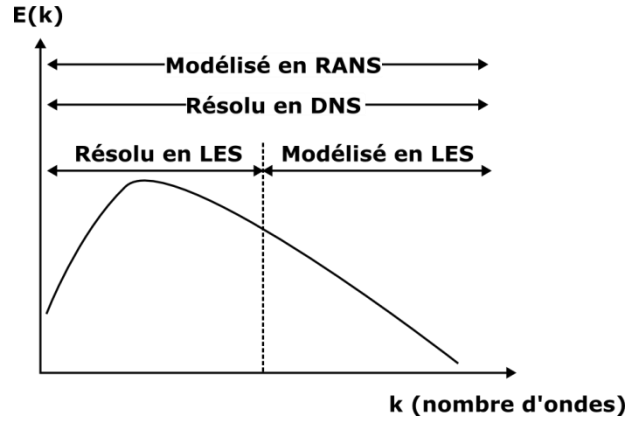


Figure I-12 : Cascade d'énergie des tourbillons et modélisation

La description des équations du mouvement du fluide, dans une approche basée sur la décomposition de Reynolds a fait apparaître une difficulté : la nécessité de fermeture du système d'équations. L'équation de continuité et les équations de Reynolds (passage à la moyenne des équations de Navier-Stokes) forment un système de quatre équations dans le cas tridimensionnel. En plus des valeurs moyennes de pressions et de vitesses, six inconnues sont ajoutées, il s'agit des contraintes de Reynolds : $-\rho \overline{u'_i u'_j}$. Le système étant ouvert, les approches de type RANS proposent de modéliser la turbulence.

Le point central de modélisation de la turbulence a été introduit par Boussinesq (1977). Il postule que le caractère turbulent de l'écoulement induit une modification du champ de contraintes :

$$\sigma_t = \mu_t \frac{\partial u}{\partial y} \quad (I-24)$$

En utilisant l'équation de Reynolds, l'écriture des contraintes selon le concept de viscosité turbulente devient :

$$-\rho \overline{u'v'} = \mu_t \frac{\partial \bar{u}}{\partial y} \quad (I-25)$$

L'écriture tridimensionnelle donne :

$$-\rho \overline{u'_i u'_j} = \mu_t \left[\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right] - \frac{2}{3} \rho k \delta_{ij} \quad (I-26)$$

Cette écriture présente quelques limitations. Elle impose notamment une valeur de contrainte turbulente nulle à l'extremum de profil de vitesse.

L'introduction de la viscosité tourbillonnaire μ_t permet de remplacer les six inconnues des contraintes de Reynolds par une seule, à déterminer. Plusieurs écritures peuvent être

envisagées, entraînant un nombre croissant d'équations supplémentaires à résoudre. Le modèle algébrique à zéro équation est historiquement le premier à être proposé. Prandtl (1925) formule une relation entre la viscosité turbulente et la longueur de mélange, à déterminer en fonction de la situation à résoudre. Le modèle utilisé dans la suite de ce travail est le modèle $k - \epsilon$, qui introduit deux nouvelles équations de transport, sur l'énergie cinétique turbulente k et son taux de dissipation ϵ . La relation entre k et ϵ permet de déterminer implicitement la longueur caractéristique :

$$\epsilon \propto \frac{k^{3/2}}{l} \quad (I-27)$$

Différentes formulations existent pour les modèles $k - \epsilon$. Le modèle utilisé est le modèle $k - \epsilon$ standard. Sa formulation a été décrite par Launder et Spalding (1974). Le système à résoudre est constitué de six inconnues : $\bar{p}, \bar{u}, \bar{v}, \bar{w}, k, \epsilon$ et de six équations : continuité, équations de Reynolds, équations de transport de k et ϵ .

L'équation de transport de l'énergie cinétique turbulente peut être décomposée selon plusieurs termes :

$$\frac{\partial k}{\partial t} + \bar{u}_i \frac{\partial k}{\partial x_i} = Source(k) + Diffusion(k) - Destruction(k) \quad (I-28)$$

Le terme de corrélation double des fluctuations de vitesse moyenne est modélisé à travers le concept de viscosité tourbillonnaire, définie en fonction des variables du système :

$$\mu_t = \rho C_\mu \frac{k^2}{\epsilon} ; \nu_t = C_\mu \frac{k^2}{\epsilon} \quad (I-29)$$

La viscosité tourbillonnaire fait apparaître une constante C_μ . Le terme de diffusion fait apparaître une deuxième constante, σ_k , reliée à la viscosité dynamique turbulente de l'énergie cinétique turbulente, par analogie avec la viscosité tourbillonnaire.

L'équation devient :

$$\frac{\partial k}{\partial t} + \bar{u}_i \frac{\partial k}{\partial x_i} = \nu_t \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) \frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial}{\partial x_i} \left[\left(\nu + \frac{\nu_t}{\sigma_k} \right) \frac{\partial k}{\partial x_i} \right] - \epsilon \quad (I-30)$$

L'équation de transport du taux de dissipation de l'énergie cinétique turbulente est complexe à modéliser. Sa décomposition est faite à travers les termes d'une équation de transport, introduisant trois constantes supplémentaires : $C_{1\epsilon}, C_{2\epsilon}, \sigma_\epsilon$.

L'équation de transport du taux de dissipation de l'énergie cinétique turbulente devient :

$$\frac{\partial \epsilon}{\partial t} + \bar{u}_i \frac{\partial \epsilon}{\partial x_i} = C_{1\epsilon} \frac{\epsilon}{k} \bar{u}_i' \bar{u}_j' \frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial}{\partial x_j} \left[\left(\nu + \frac{\nu_t}{\sigma_\epsilon} \right) \frac{\partial \epsilon}{\partial x_j} \right] - C_{2\epsilon} \frac{\epsilon^2}{k} \quad (I-31)$$

Les constantes de ce modèle sont déterminées en fonction d'une configuration d'écoulement donnée ou bien en comparant les résultats du modèle et les résultats expérimentaux dans différentes configurations. Les constantes du modèle standard ont les valeurs suivantes :

$$C_\mu = 0,09 ; \sigma_k = 1,00 ; \sigma_\epsilon = 1,30 ; C_{1\epsilon} = 1,44 ; C_{2\epsilon} = 1,92 \quad (I-32)$$

Au modèle « standard » s'ajoutent le modèle RNG pour ReNormalization Group et le modèle « realizable ». Ils varient du modèle standard par la formulation des constantes du système (équation I-32).

I-2.4. Écoulement autour d'un stockage

Les obstacles cylindriques verticaux se trouvent en majorité sur les sites industriels. Il a été établi au préalable que ce travail de thèse s'intéresse uniquement aux obstacles cylindriques verticaux. Les écoulements autour de ces stockages sont bien documentés dans la littérature. On observe différentes structures en fonction du type d'écoulements. Le nombre de Reynolds est ici calculé en fonction de la dimension caractéristique de l'obstacle. Selon la valeur de ce nombre, différentes structures sont observées. La figure ci-après répertorie ces régimes :

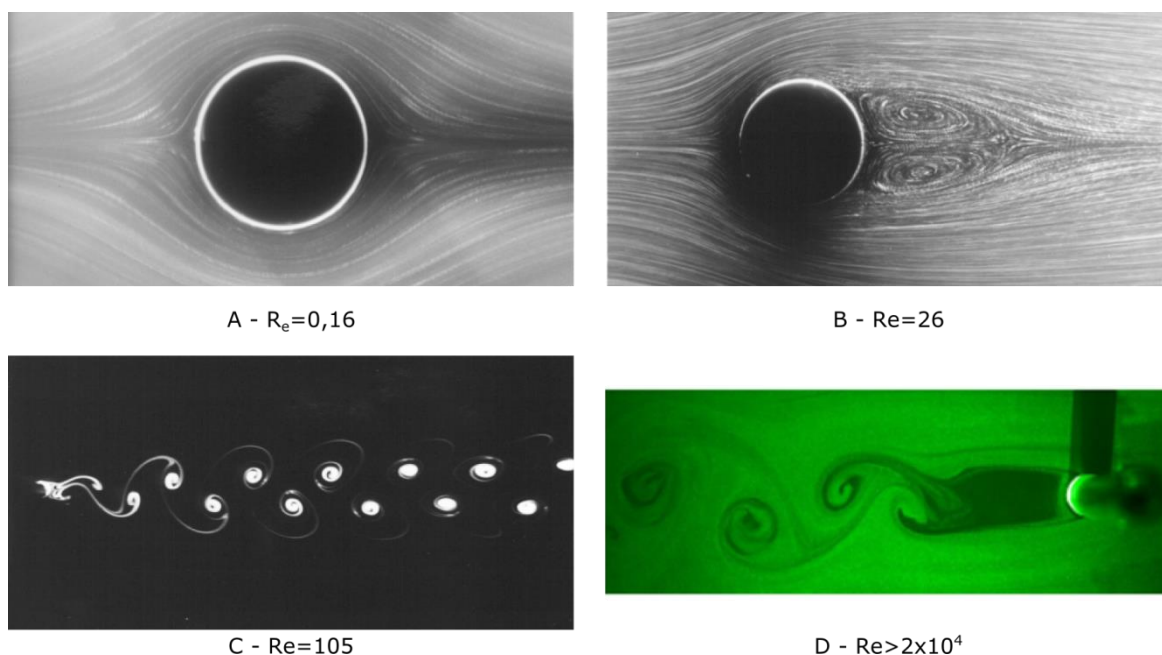


Figure I-13 : Représentation des écoulements autour d'un cylindre pour différents nombre de Reynolds (A, B et C d'après Taneda, 1977 – D réalisé à l'Institut des Sciences des Risques, 2013)

Pour un écoulement laminaire, les lignes de courant contournent l'obstacle et se rejoignent immédiatement derrière celui-ci. A partir de $Re=26$, on observe deux zones de recirculation dans le sillage de l'obstacle, avec la zone de décollement se situant en aval du milieu de l'obstacle. Pour des nombres de Reynolds compris entre 48 et 180, on observe des lâchers de tourbillons. Ce sont des allées de Von Karman. Celles-ci se forment directement en aval de l'obstacle. Pour des nombres de Reynolds supérieurs à 10^4 , des tourbillons sont lâchés en aval de l'obstacle, au niveau d'une zone où les vitesses sont faibles et la turbulence forte. Alors que dans les deux premiers cas, les structures sont stationnaires, les nombres de Reynolds suivants montrent un écoulement instationnaire. Dans la suite des travaux, seule la valeur moyennée du champ de vitesse sera utilisée, de manière à s'affranchir de l'instabilité de tels écoulements.

A titre d'exemple, le nombre de Reynolds obtenu pour un écoulement d'air de 2 m.s^{-1} autour d'un obstacle de diamètre 10 m est :

$$R_e = \frac{\rho u D}{\mu} = \frac{1.2 \cdot 2 \cdot 10}{1.8 \cdot 10^{-5}} = 1,33 \cdot 10^6 \quad (I-33)$$

La valeur obtenue montre clairement un écoulement turbulent. Si l'on considère un cylindre infini, la structure de l'écoulement est caractérisée par les zones suivantes reportées sur la figure ci-après :

- ▶ Une zone de déplacement amont (a) est caractérisée par une diminution de la vitesse de la direction incidente vers la tangente de l'obstacle. Une surpression est à noter sur la face amont de l'obstacle.
- ▶ Les bords de l'obstacle sont le siège de la séparation d'une couche limite de l'écoulement à la paroi de l'obstacle (b).
- ▶ La zone de cavité (c) est la zone immédiatement en aval de l'obstacle. Elle est caractérisée par :
 - ▶ une faible vitesse moyenne,
 - ▶ une grande intensité de turbulence,
 - ▶ une recirculation initiée à partir du point de recollement,
 - ▶ Un temps de résidence des substances relativement long, à cause de la présence de la recirculation,
 - ▶ Une faible pression à valeur quasi-constante,
 - ▶ La zone de génération de tourbillons est à proximité du point de recollement.
- ▶ Le sillage turbulent (d) est caractérisé par la présence de nombreux tourbillons de différentes tailles. La turbulence induite par l'obstacle s'estompe progressivement avec la distance à l'obstacle. 50 à 100 diamètres sont nécessaires pour revenir à des valeurs d'intensité de turbulence et de vitesses approximativement égales à celles de l'écoulement externe, en fonction du nombre de Reynolds de l'écoulement (Hosker Jr, 1985a).

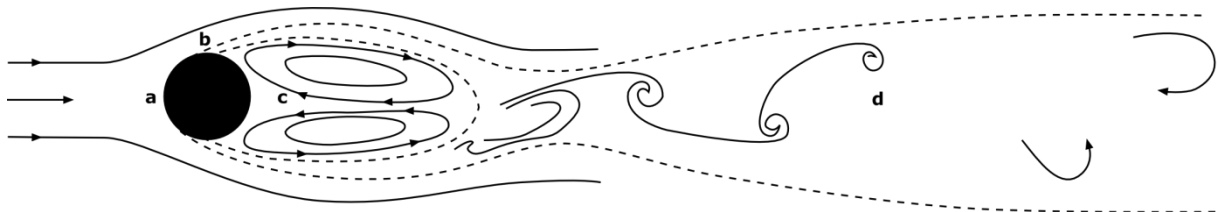


Figure I-14 : Différentes zones caractéristiques d'un écoulement autour d'un obstacle cylindrique (retranscription d'après Hosker Jr, 1985)

Cette illustration présente un écoulement homogène en entrée. La paroi face à l'écoulement est le siège d'une augmentation de la pression. La zone de déplacement est située de part et d'autre de l'obstacle. C'est le siège de l'augmentation de la vitesse selon y . Les courbes en pointillés débutent à la paroi de l'obstacle : c'est le décollement d'une couche limite. En aval de l'obstacle, on distingue deux cavités. Ce sont des zones de recirculation où la turbulence est importante. La pression à la paroi en aval est minimale sur l'axe central. La fin de la zone de recirculation est le point de recollement. C'est à cet endroit que sont générés des tourbillons. Ceux-ci se déplacent puis disparaissent dans le sillage turbulent de l'obstacle.

Les structures tourbillonnaires lâchées en aval de la zone de recirculation présente un caractère instationnaire. La dispersion atmosphérique d'une substance sera affectée par ces lâchers, son orientation dépendant de l'instant où le nuage est déplacé dans cette

zone. Dans la suite de ce travail, l'écoulement est considéré comme étant stationnaire : les variations sont prises en compte à travers la moyenne.

Dans le cas d'un obstacle immergé dans un écoulement libre à couche limite, des tourbillons en fer à cheval accompagnent les arêtes des couches limites de paroi.

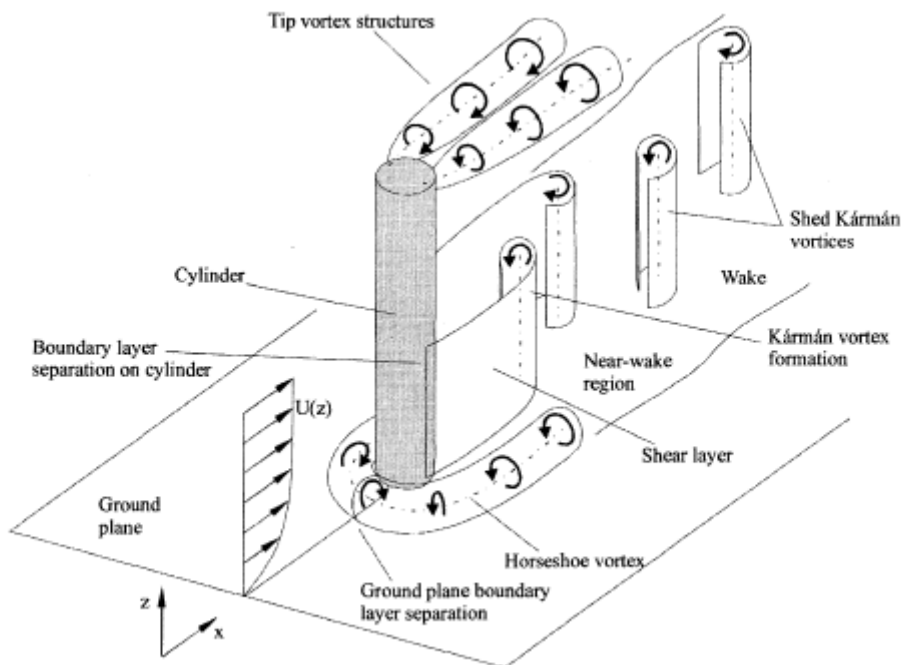


Figure I-15 : Structures tridimensionnelles observées dans un écoulement autour d'un cylindre vertical (Heseltine, 2003)

Si la structure de l'écoulement est modifiée, il en va de même de la turbulence. Ozgoren et al. (2011) montrent une augmentation importante de la turbulence dans le sillage d'un cylindre. Le passage du nombre de Reynolds de 5×10^3 à 10^4 implique par exemple un déplacement plus en amont du point de stagnation du sillage. Par ailleurs, ils soulignent l'instationnarité de l'écoulement à travers la fréquence de lâcher de tourbillons. Celle-ci est directement reliée au Reynolds de l'écoulement. L'influence d'un obstacle sur le sillage de l'écoulement est généralement estimée entre 50 à 100 fois le diamètre de l'obstacle. Ainsi, même si cela dépasse le cadre de cette étude, il est néanmoins primordial d'étudier par la suite l'influence de la perturbation d'un écoulement par un obstacle sur un autre obstacle.

De nombreuses études ont permis de dégager des caractéristiques de perturbation de l'écoulement pour d'autres formes d'obstacles (Mavroidis et al., 2003). Par ailleurs, des études expérimentales existent également pour le traitement des groupes d'obstacles (Gailis, 2004). Elles ne sont pas détaillées ici, cette étude s'intéressant prioritairement aux obstacles cylindriques.

I-2.5. Conclusion

La caractérisation des écoulements atmosphériques a permis de mettre en avant des caractéristiques essentielles pour la suite de ce travail en s'intéressant donc à la couche de surface, qui est directement à proximité du sol. Les caractéristiques de la CLA imposent la prise en compte de la turbulence. Celle-ci est d'autant plus présente lorsqu'un obstacle perturbe l'écoulement. Dans de tels cas, l'écoulement ne peut être

considéré comme stationnaire car des structures tourbillonnaires se forment en aval de l'obstacle et sont relâchées selon une fréquence spécifique.

D'autre part, un certain nombre d'hypothèses seront à prendre en compte afin de pouvoir modéliser correctement un écoulement atmosphérique dans des conditions de stabilité données.

I-3. Modélisation de la dispersion atmosphérique

Dans ce chapitre, l'intérêt de la modélisation de la dispersion atmosphérique est décrit. La modélisation physique est détaillée, tout comme la modélisation numérique. Les liens constructifs existants entre l'expérimental et le numérique seront abordés. Enfin, les différentes méthodes de modélisation, des plus rapides aux plus complexes, seront explicitées. Des modèles alternatifs seront également présentés, impliquant des approches différentes, dans le but de concilier les deux aspects antagonistes de la modélisation numérique : précision et vitesse.

Effectuer des essais pour évaluer les conséquences d'un phénomène dangereux ne permet pas d'être exhaustif. La modélisation de la dispersion atmosphérique permet d'éviter des simulations physiques à échelle réelle et donc des essais coûteux. D'autre part, cela permet de répondre aux demandes de la réglementation en termes de nombre de scénarios pouvant être testés.

Les avancées importantes de la simulation numérique permettent dorénavant de coupler les approches. Une méthodologie hybride s'appuie sur trois fondamentaux : l'expérimental, la théorie et l'informatisation (Schlesinger, 1979). Le développement, la vérification et la validation d'un modèle est progressive, de l'observation et la mesure de la réalité à la modélisation, de l'utilisation du modèle pour simuler des écoulements et enfin comparer ces simulations à la réalité. Cette méthodologie est illustrée dans la figure ci-après :

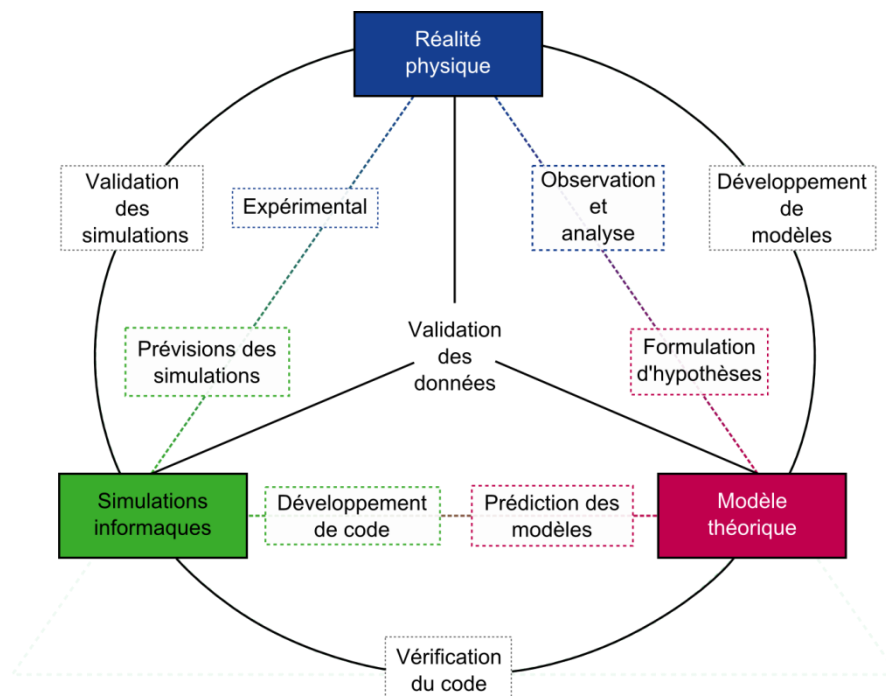


Figure I-16 : Représentation schématique des liens entre réalité physique, développement théorique et simulations informatiques - adapté de Meroney (2004)

A partir de ce schéma, Oberkampf et Trucano (2002) décrivent deux types de modèles : les modèles conceptuels, créés en analysant et en observant les systèmes physiques et les modèles informatisés qui sont les programmes opérationnels qui implémentent les modèles conceptuels. Ainsi, les modèles conceptuels peuvent être vérifiés à partir de la comparaison entre les solutions fournies par le modèle informatisé et d'éventuelles solutions analytiques ou des solutions fournies par les modèles de référence. La validation est effectuée à partir de la comparaison avec des données issues d'expérimentations.

Différentes approches présentant chacune des avantages et des inconvénients sont détaillés ci-après.

I-3.1. Différents types de modélisation

La modélisation comprend à la fois la modélisation physique et la modélisation numérique. Ces deux types d'études sont complémentaires (Mavroidis et al., 2003).

I-3.1.1 Expérimentation à échelle réduite

L'expérimentation à échelle réduite est largement utilisée pour modéliser un grand nombre de scénarios tenant en compte différents écoulements. Effectivement, les avantages sont multiples : la modélisation à échelle réduite dans une soufflerie ou un canal hydraulique permet d'obtenir des données réelles dans la limite des erreurs de mesures. Le fluide utilisé conserve les mêmes propriétés qu'en champ libre. La réalisation d'essais à échelle réduite permet d'effectuer des simulations de dispersion atmosphérique, visant à reproduire le comportement réel. La limitation se situe dans les appareils de mesure, qui peuvent induire des erreurs ou bien perturber l'écoulement. D'autre part, la difficulté est de reproduire l'ensemble des phénomènes intervenants dans la perturbation de l'écoulement.

La comparaison entre des données obtenues à l'échelle réelle et à l'échelle réduite n'est possible qu'à la condition du respect des conditions de similitude. L'étude des similitudes s'effectue en adimensionnant les équations de la mécanique des fluides : conservation de la quantité de mouvement, continuité, énergie. Les tentatives de reproduire les conditions réelles de la CLA imposent d'utiliser des variables de référence : distance L_R , vitesse U_R , densité ρ_R , pression P , écart de température δT_R , vitesse angulaire Ω_R . Snyder (1981) utilise les variables adimensionnées suivantes :

$$\begin{aligned} x'_i &= \frac{x_i}{L} & U'_i &= \frac{U_i}{U_R} & t' &= \frac{U_R}{L} t & \rho' &= \frac{\rho_0}{\rho_R} \\ \delta P' &= \frac{\delta P}{\rho_R U_R^2} & \delta T' &= \frac{\delta T}{\delta T_R} & \Omega'_j &= \frac{\Omega_j}{\Omega_R} \end{aligned}$$

En utilisant ces expressions dans les équations de la mécanique des fluides, on peut caractériser un écoulement à partir des nombres adimensionnels qui le composent : Rossby, Froude, Reynolds et Peclet.

$$\frac{\partial U'_i}{\partial t'} + U'_j \frac{\partial U'_i}{\partial x'_j} + \frac{2}{Ro} \epsilon_{ijk} U'_k \Omega'_j = -\frac{1}{\rho'} \frac{\partial \delta P'}{\partial x'_i} + \frac{1}{Fr^2} \delta T' \delta_{3i} + \frac{1}{Re} \frac{\partial^2 U'_i}{\partial x'_j \partial x'_j} \quad (I-34)$$

$$\frac{\partial U'_i}{\partial x'_j} = 0 \quad (I-35)$$

$$\frac{\partial \delta T'}{\partial t'} + U_i' \frac{\partial \delta T'}{\partial x_i'} = \frac{1}{Pe} \frac{\partial^2 \delta T'}{\partial x_i' \partial x_i'} \quad (I-36)$$

R_o	Nombre de Rossby
F_r	Nombre de Froude
Re	Nombre de Reynolds
Pe	Nombre de Peclet

Ces nombres sont définis comme suit :

$$Ro = \frac{U_R}{L_{\Omega R}} \quad Fr = \frac{U_R}{\left(gL \frac{\delta T_R}{T_0}\right)} \quad Re = \frac{U_R L}{\nu} \quad Pe = \frac{U_R L}{\kappa}$$

Ainsi, l'utilisation des équations 34 à 36 et l'utilisation de conditions aux limites permettent de déterminer complètement l'écoulement. Il est à noter que le nombre de Rossby présente peut d'intérêt lorsque l'écoulement est effectué pour une échelle de distance inférieure à cinq kilomètres. L'atteinte du nombre de Froude dans la maquette à échelle réduite permet de simuler correctement la stratification de l'atmosphère. Pour cela, il est nécessaire de contrôler la température. Le nombre de Reynolds est rarement strictement identique dans la réalité et dans la modélisation à échelle réduite. Cependant, en tenant compte de certaines hypothèses (absence d'effets de la force de Coriolis, ou d'effets thermiques), la structure de l'écoulement turbulent est similaire à partir d'un nombre de Reynolds suffisamment élevé. Le nombre de Peclet est également exprimé comme le produit du nombre de Reynolds par le nombre de Prandtl :

$$Pe = \frac{U_R L}{\kappa} = \frac{U_R L}{\nu} \frac{\nu}{\kappa} = Re \cdot Pr$$

Ce dernier représente le rapport entre la diffusivité de la quantité de mouvement et la diffusivité thermique. Pour pouvoir prendre en compte la dispersion dans un essai à échelle réduite, il est nécessaire d'ajouter l'équation d'advection diffusion adimensionnée :

$$\frac{\partial c'}{\partial t'} + U_i' \frac{\partial c'}{\partial x_i'} = \frac{1}{ReSc} \frac{\partial^2 c'}{\partial x_i' \partial x_i'} \quad (I-37)$$

$Sc = \frac{\nu}{\alpha}$	Nombre de Schmidt	-
C	Concentration instantanée	kmol.m ⁻³
α	Diffusion moléculaire	m ² .s ⁻¹

Le nombre de Schmidt est le rapport entre la viscosité et la diffusivité massique. Dans ce cas là, la diffusion moléculaire est prise en compte. Elle représente la réalité de la dispersion. La turbulence n'est pas modélisée, elle est incluse dans le mouvement de l'écoulement. Ainsi, deux écoulements aux dimensions différentes où les nombres adimensionnels seraient égaux sont semblables, les résultats obtenus pour l'écoulement à échelle réduite caractérisant également l'écoulement à échelle réelle.

L'un des inconvénients de la modélisation physique est l'impossibilité d'obtenir simultanément l'égalité de tous les nombres adimensionnels (réalité et modèle) lorsque le rapport des longueurs est supérieur à dix (Snyder, 1981). Par conséquent, des méthodes ont été développées, notamment pour augmenter la turbulence sans changer la vitesse. Ainsi, la condition à la limite représentant le sol est réalisée par l'ajout

d'éléments rugueux, la difficulté étant de déterminer la taille correcte et l'espacement entre ces différents éléments.

Ces techniques montrent une des limites de la modélisation physique. D'autres difficultés existent également au niveau de la mesure. Selon l'objectif de l'étude, il peut être impératif d'éviter une perturbation de l'écoulement par le capteur. Récemment, avec l'augmentation des résolutions des cameras numériques, ainsi que des fréquences de tirs des lasers, des visualisations précises peuvent être accomplies. L'utilisation de telles données à travers différents algorithmes permet d'obtenir des informations sur les vitesses et la turbulence de l'écoulement (Particle Image Velocimetry ou PIV) ou bien directement des concentrations dans le cas de modélisations de dispersion atmosphérique (Laser Induced Fluorescence ou LIF) sans perturbation de l'écoulement. Les limitations de ces modèles sont donc liées aux incertitudes sur les conditions initiales, l'équipement de mesure, les observations du prototype. Il est donc difficile d'évaluer simplement la fiabilité des résultats.

I-3.1.2 Simulation numérique

Un grand nombre de modèles différents existent et se complètent selon les phénomènes d'intérêt et la finesse requise. Généralement le calcul du champ de vent est découplé du calcul de la dispersion atmosphérique. Ce dernier est principalement gouverné par le caractère turbulent de l'écoulement. La plupart des modèles existants calculent séparément ces deux phénomènes.

Le champ de vent peut être obtenu à partir de profils analytiques. Il s'agit alors d'utiliser la théorie de Monin-Obukhov, qui permet d'obtenir l'évolution verticale de la vitesse et de la turbulence. Aussi, le calcul eulérien à partir des équations de la mécanique des fluides permet la détermination de l'écoulement en tout point du domaine de calcul.

La modélisation de la dispersion atmosphérique peut alors s'effectuer selon deux approches. L'approche eulérienne se fait par le calcul des variables d'intérêt, en tout point et à chaque instant. Plusieurs niveaux de complexité existent, selon les objectifs de l'étude. L'étude lagrangienne de la dispersion s'intéresse au suivi des trajectoires de particules fluides ou de polluants. Le champ moyen est utilisé pour calculer la vitesse d'une particule. La turbulence est modélisée à travers l'ajout d'une composante aléatoire à la vitesse moyenne.

I-3.2. Modèles de dispersion atmosphérique

La dispersion atmosphérique d'une substance est directement affectée par la turbulence d'un écoulement. Le système d'équation présenté au I-2.3.3 permet de déterminer le champ de vitesse d'un écoulement d'air. Le suivi d'une espèce supplémentaire peut être considéré de trois manières :

- La résolution analytique de l'équation d'advection diffusion sous certaines hypothèses
- L'approche lâcher lagrangien stochastique
- La résolution de l'équation d'advection diffusion selon une approche CFD.

La dispersion atmosphérique caractérise l'évolution dans le temps et l'espace d'un nuage. Deux phénomènes associés à la vitesse moyenne du mouvement, ainsi qu'à son caractère aléatoire caractérisent la dispersion atmosphérique :

- ▶ L'advection est le transport de substance par l'écoulement moyen. Le volume du nuage est conservé, sa forme peut changer si le champ de vent n'est pas uniforme.
- ▶ La diffusion ne conserve pas le volume ni la concentration. La diffusion moléculaire et la diffusion turbulente apportent chacune leur contribution. Cependant, dans le cas de la dispersion atmosphérique, la valeur de la diffusion turbulente est d'un ordre de grandeur suffisamment supérieur à la valeur de la diffusion moléculaire pour négliger ce dernier terme. Ainsi, ce phénomène correspond à l'étalement du nuage de polluant par action des fluctuations de vitesse, c'est-à-dire la composante turbulente de la vitesse du vent.

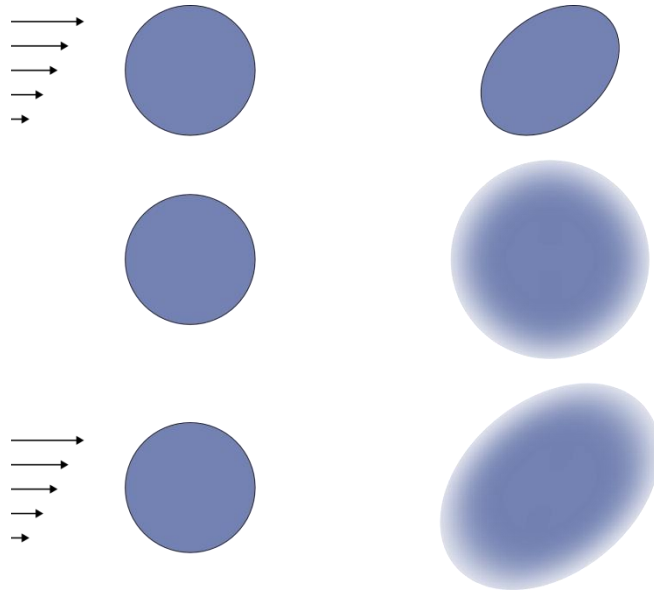


Figure I-17 : Illustration des mécanismes d'advection - diffusion et combinaison des deux phénomènes.

L'équation traduisant le principe de conservation de la quantité scalaire concentration sous l'hypothèse d'incompressibilité est l'équation d'advection diffusion. C'est une équation de transport :

$$\frac{\partial c}{\partial t} + u_i \frac{\partial c}{\partial x_j} = \frac{\partial}{\partial x_j} \left(D_i \cdot \frac{\partial c}{\partial x_j} \right) + S_i + R_c \quad (I-38)$$

D	Coefficient de diffusion de la substance	$\text{m}^2.\text{s}^{-1}$
$\frac{\partial c}{\partial t}$	Terme instationnaire	$\text{kmol}.\text{m}^{-3}.\text{s}^{-1}$
$u_{xi} \frac{\partial c}{\partial x_j}$	Terme advectif	$\text{kmol}.\text{m}^{-3}.\text{s}^{-1}$
$\frac{\partial}{\partial x_j} \left(D_i \cdot \frac{\partial c}{\partial x_j} \right)$	Terme diffusif	$\text{kmol}.\text{m}^{-3}.\text{s}^{-1}$
S_i	Source	$\text{kmol}.\text{m}^{-3}.\text{s}^{-1}$
R_c	Réaction chimique	$\text{kmol}.\text{m}^{-3}.\text{s}^{-1}$

S'agissant d'un écoulement turbulent, la décomposition de Reynolds s'applique également à la concentration :

$$c = \bar{c} + c' \quad (I-39)$$

En appliquant l'opérateur de moyenne de Reynolds, l'équation devient :

$$\frac{\partial \bar{c}}{\partial t} + \bar{u}_i \frac{\partial \bar{c}}{\partial x_i} = \frac{\partial}{\partial x_j} \left(D \cdot \frac{\partial \bar{c}}{\partial x_j} - \overline{u'_i c'} \right) + \bar{S}_i + \bar{R}_c \quad (I-40)$$

$$\overline{u'_i c'} \quad \text{Flux turbulent du scalaire considéré}$$

Cette équation possède quatre inconnues, \bar{c} et les trois composantes du vecteur $\overline{u'_i c'}$. Il est donc nécessaire d'exprimer les composantes de ce flux par des variables connues.

I-3.2.1 Les modèles gaussiens

Les modèles gaussiens sont basés sur une solution analytique de l'équation d'advection-diffusion. Les hypothèses conduisant à une solution analytique sont nombreuses. Il convient ici de fermer le système d'équations décrit au paragraphe I-2.3. Pour cela, par analogie avec la loi de diffusion moléculaire de Fick, on introduit la notion de coefficient de diffusivité turbulent K relié au flux turbulent $\overline{u'_i c'}$ de la manière suivante :

$$\overline{u'_i c'} = -K \nabla \bar{c} \quad (I-41)$$

Le cas général conduit à un tenseur de diffusivité turbulente K fonction du temps et de l'espace. En admettant un certain nombre d'hypothèses simplificatrices :

- ▶ K est admis comme constant dans le temps et l'espace
- ▶ Le flux turbulent dans une direction est uniquement proportionnel au gradient de concentration moyen dans cette direction
- ▶ La substance suivie est considérée comme passive
- ▶ Le vent est considéré comme stationnaire et homogène
- ▶ La turbulence est considérée comme stationnaire
- ▶ La diffusion moléculaire est considérée comme négligeable par rapport à la diffusion turbulente
- ▶ On se place dans le sens de l'écoulement
- ▶ Le rejet est ponctuel et instantané

L'équation d'advection diffusion devient alors :

$$\frac{\partial \bar{c}}{\partial t} + u_x \frac{\partial \bar{c}}{\partial x} = K_y \frac{\partial^2 \bar{c}}{\partial y^2} + K_z \frac{\partial^2 \bar{c}}{\partial z^2} + S_i \quad (I-42)$$

$\frac{\partial \bar{c}}{\partial t}$	Terme instationnaire
$u_x \frac{\partial \bar{c}}{\partial x}$	Terme advectif
$K_y \frac{\partial^2 \bar{c}}{\partial y^2} + K_z \frac{\partial^2 \bar{c}}{\partial z^2}$	Terme diffusif
S_i	Source (rejet)

La solution analytique de cette équation est :

$$c(x) = \frac{q}{4\pi(K_y K_z)^{1/2} x} e^{\left(-\frac{u_x(y^2}{K_y} + \frac{z^2}{K_z})\right)} \quad (I-43)$$

Cette répartition gaussienne donne son nom au modèle. Grâce à la linéarité de l'équation de Fick, on peut intégrer cette solution en fonction des conditions du problème : bouffée, panache, rejet ponctuel, linéique.

Les limitations liées à cette méthode de résolution concernent les temps courts. Il est nécessaire d'évaluer les écarts-types de cette distribution de concentration :

$$\sigma_i = \sqrt{2K_i t} \quad (I-44)$$

Ces écarts-types dépendent donc du temps écoulé depuis l'émission ou temps de transport ou bien de la distance à la source. La détermination de ces écarts-types est réalisée de manière empirique, à partir d'essais expérimentaux. Les méthodes les plus usitées sont celles de Doury (1976) et de Pasquill (1961).

i. Classes de Pasquill

Les classes de Pasquill sont définies en fonction de la stabilité de l'atmosphère, de A à G, d'extrêmement instable à extrêmement stable. Cette stabilité dépend de la période de la journée (nuit ou jour), de l'ensoleillement ou nébulosité et de la vitesse du vent. La formule permettant le calcul de l'écart type de dispersion dépend là de la distance à la source :

$$\sigma_i = ax^b + c \quad (I-45)$$

Les coefficients a , b et c sont déterminés à partir de l'abaque combinant classe de stabilité et distance à la source.

Des limitations sont associées à ces modèles gaussiens, outre les hypothèses contraignantes détaillées précédemment, les modèles gaussiens ne prennent pas en compte la topographie, la rugosité du terrain et la présence de bâtiments et d'obstacles. Des améliorations de ces modèles (Carruthers et al., 1994 ; Riddle et al., 2004) permettent de prendre en compte certaines de ces considérations en utilisant notamment le principe de conservation de la masse et la modification du champ de vitesse.

ii. Classes de Doury

Doury définit deux classes de diffusion, faible et normale, en fonction de la vitesse du vent et de la période de la journée. La formule suivante permet de retrouver, à partir d'abaques issus d'observations expérimentales, la valeur des écarts-types à appliquer :

$$\sigma_i = (2K_i t)^{1/2} = (A_i t)^{k_i} \quad (I-46)$$

Les coefficients A_i et k_i sont déterminées à partir d'un abaque spécifique, du type de diffusion et du temps de transport :

$$t = \frac{x}{u} \quad (I-47)$$

x	Distance	m
u	Vitesse	m.s ⁻¹

Ces modèles présentent un intérêt important dans la modélisation puisque le temps de calcul est extrêmement rapide. Ils sont également utilisés en situation de crise. De plus, de par leur utilisation massive, il existe beaucoup de retour d'expériences sur ce type de modèles et l'incertitude associée. Certaines formulations permettent de prendre en compte des surhauteurs de cheminées dans le cas de rejets chroniques. D'autre part, des recherches sont faites pour adapter ce type de modèle à des environnements urbains (Brook et al., 2001).

Les modèles intégraux ont été développés pour prendre en compte les phénomènes en amont de la dispersion passive. Ceux-ci intègrent différentes phases de calcul où différents modèles sont appliqués. Les différentes phases sont :

- ▶ Le rejet. La phase de rejet est caractérisée par une vitesse initiale, la plupart du temps différente de la vitesse de l'écoulement dans lequel s'effectue le rejet.
- ▶ L'étalement au sol. Dans le cas des gaz plus lourds que l'air, il est nécessaire de prendre en compte la phase de rampement au sol.
- ▶ La dispersion passive. Une fois que le gaz a perdu sa quantité de mouvement initiale due au rejet, sa dispersion s'effectue en fonction de l'écoulement dans lequel il est immergé. Cette phase de calcul est effectuée par un modèle gaussien.

I-3.2.2 Les modèles Lagrangiens

Les modèles lagrangiens reposent sur le calcul de trajectoires de particules fluides. Ces modèles utilisent des caractéristiques statistiques des écoulements afin de pondérer le champ moyen de vent en tout point par une composante fluctuante. Ceux-ci sont en général adaptés à l'étude du champ proche à la zone de rejet. Ils sont indiqués dans les écoulements complexes présentant par exemple une turbulence inhomogène. Il est possible de ramener une densité de particules à une concentration à partir de la caractérisation des conditions initiales.

Ces modèles lagrangiens s'appuient donc sur un calcul eulérien préalable de l'écoulement. La description de la position d'une particule revient à la résolution du système suivant :

$$\begin{cases} dX = U_x dt = u_x(x, y, z, t)dt \\ dY = U_y dt = u_y(x, y, z, t)dt \\ dZ = U_z dt = u_z(x, y, z, t)dt \end{cases} \quad (I-31)$$

X, Y, Z	Coordonnées lagrangiennes de la particule dans le référentiel de l'écoulement
$u_x, u_y \text{ et } u_z$	Composantes de la vitesse de l'écoulement au sens eulérien
$U_x, U_y \text{ et } U_z$	composantes du champ de vitesse lagrangien

En appliquant une décomposition de Reynolds à ce système, celui-ci devient :

$$\begin{cases} dX = (\overline{U_x} + U_x')dt = (\overline{u_x} + u_x')dt \\ dY = (\overline{U_y} + U_y')dt = (\overline{u_y} + u_y')dt \\ dZ = (\overline{U_z} + U_z')dt = (\overline{u_z} + u_z')dt \end{cases} \quad (I-48)$$

La fluctuation de la vitesse lagrangienne est la base de cette approche. Celle-ci est décrite par l'équation de Langevin :

$$dU'_i = \frac{1}{\tau} U'_i dt + \mu dt \quad (I-49)$$

τ	Taux de relaxation de la particule
μ	Action aléatoire homogène à une accélération de moyenne nulle

Le principe est de représenter le signal étudié comme la somme d'un signal auto corrélé à l'échelle du pas de temps retenu et d'un signal décorrélé (le terme aléatoire μ). Cette décomposition de Reynolds permet de distinguer un terme déterministe et un terme stochastique.

La difficulté de ce type de modèle est de caractériser le terme stochastique, directement relié à la turbulence de l'écoulement. D'autre part, il est nécessaire pour utiliser ce genre de modèle de disposer au préalable le champ de vitesse de l'écoulement. En général, cette opération est effectuée à travers l'utilisation d'un modèle eulérien CFD.

I-3.2.3 La CFD

La mécanique des fluides numériques consiste à résoudre de manière eulérienne les équations de conservation, sur un maillage de l'espace. Selon le problème à modéliser, les équations de continuité, de conservation de la quantité de mouvement et de l'énergie sont résolues.

Ce système d'équations permet de calculer l'écoulement du milieu dans lequel s'effectue la dispersion. Pour résoudre cette dernière, il est nécessaire d'ajouter l'équation d'advection diffusion :

$$\frac{\partial C}{\partial t} + u_i \frac{\partial C}{\partial x_i} = D_t \frac{\partial^2 C}{\partial x_i^2} + S_C \frac{\partial C}{\partial t} + \frac{\partial U_i C}{\partial x_i} = D \frac{\partial^2 C}{\partial x_i^2} + S_C \frac{\partial C}{\partial t} + U_{x_i} \frac{\partial C}{\partial x_i} = \frac{\partial}{\partial x_i} \left(D_t \frac{\partial C}{\partial x_i} \right) + S_i \quad (I-50)$$

U_{x_i}	Vitesse dans la direction x_i	$m.s^{-1}$
t	Temps	s
C	Concentration	$kmol.m^{-3}$
S_i	Source d'émission	$kmol.m^{-3}.s^{-1}$
D_t	Coefficient de diffusion turbulente	$m^2.s^{-1}$

Le modèle $k - \epsilon$ standard est utilisé dans la suite de ce travail. La diffusion moléculaire est également négligée, puisqu'elle est inférieure de plusieurs ordres de grandeur comparée à la diffusion turbulente. Pour ce modèle, le coefficient de diffusion turbulente D_t est calculé en utilisant la définition de la viscosité turbulente μ_t :

$$\mu_t = \rho C_\mu \frac{k^2}{\epsilon} \quad (I-51)$$

C_μ	Constante du modèle	0,09
ρ	Masse volumique	$kg.m^{-3}$

Le coefficient de diffusion turbulente D_t est directement relié à la viscosité turbulente à travers le nombre de Schmidt :

$$S_{ct} = \frac{\mu_t}{\rho \cdot D_t} = 0,7$$

(I-52)

La valeur de 0,7 est classiquement utilisée. Néanmoins, ce choix est discuté et n'est pas toujours respectée expérimentalement (Koeltzsch, 2000).

Certaines limitations sont liées à l'utilisation de tels modèles. Les temps de calcul dépendent grandement de la puissance du processeur utilisé et bien que des progrès immenses aient été réalisés dans ce domaine, la durée d'un calcul de dispersion atmosphérique sur un site industriel est à minima de l'ordre de l'heure et augmente avec la prise en compte des échelles de plus en plus petites de la turbulence. Par ailleurs, des difficultés liées à la modélisation de la couche limite de surface existent. La difficulté principale est le maintien à la valeur d'entrée des profils d'énergie cinétique turbulente et de son taux de dissipation (Hargreaves and Wright, 2007 ; Pontiggia et al., 2009). Par ailleurs, ces modèles se basent sur des constantes déterminées expérimentalement, notamment lors d'essais en soufflerie (Launder and Spalding, 1974). Plusieurs valeurs différentes sont reportées dans la littérature, avec de grandes disparités selon la constante considérée.

I-3.2.4 Autres approches

A côté de ces modèles classiques cohabitent des modèles alternatifs dans leur approche. Souvent, ces modèles sont issus de projets ayant pour but d'améliorer des caractéristiques des modèles classiques (par exemple la vitesse de calcul) ou bien d'adapter ceux-ci à des configurations précises. On présente ici un modèle utilisant une méthodologie combinant ces deux aspects : Flow'Air-3D/SLAM. Ce modèle hybride a été développé dans le but de réduire le temps de calcul d'une simulation CFD. Vendel (2011) a développé ce code de calcul à destination d'un site industriel pétrochimique, pour étudier la dispersion d'une substance rejetée, pour le suivi temps réel et en champ proche de la pollution atmosphérique. Pour ce faire, le calcul du champ de vent est effectué séparément du calcul de la dispersion, à l'aide d'un modèle eulérien de type RANS $k - \varepsilon$. Une base de données est créée en amont de l'utilisation du modèle. Elle intègre un grand nombre de conditions atmosphériques et de profils de vent de la couche limite de surface. Ce champ de vent autour des bâtiments composant le site industriel est sauvegardé dans une base de données. Lors de l'utilisation du modèle, la base de données de champ de vent est interpolée à partir des conditions météorologiques du moment. Cela permet d'obtenir très rapidement un écoulement représentatif de la réalité. Un modèle lagrangien SLAM (Safety Lagrangian Atmospheric Model) est appliqué à partir de la détermination du terme source de dispersion. Le schéma ci-après illustre la préparation amont et le fonctionnement en direct du modèle :

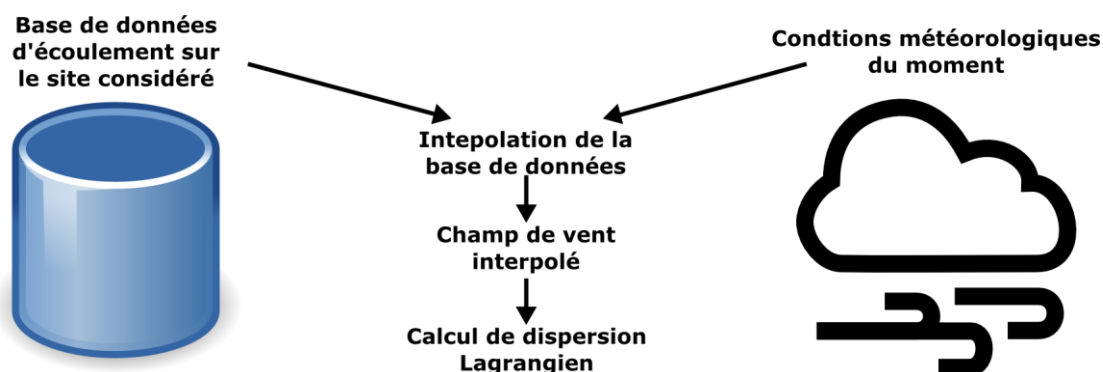


Figure I-18 : Schéma de fonctionnement du modèle de Vendel (2011)

Néanmoins, l'inconvénient majeur de ce modèle est l'adaptation à un site unique dû à la nécessité de disposer d'une base de données de champ de vent préexistante sur le site.

D'autres modèles alternatifs existent et font appel à un paradigme totalement différent. Par exemple, l'approche systémique peut être utilisée grâce à des outils mathématiques issus de l'intelligence artificielle : les automates cellulaires et les réseaux de neurones. Ces approches utilisées en modélisation de la dispersion sont décrits à la suite des chapitres II-1 et II-2.

I-3.3. Conclusion

La modélisation de la dispersion atmosphérique est possible de plusieurs façons : modélisation physique à échelle réduite, simulation numérique ou statistique. La première méthode est difficile à mettre en œuvre et ne permet pas d'être exhaustif dans les situations à modéliser. Avec l'augmentation des puissances de calculs, la simulation numérique permet de modéliser des situations de plus en plus variées et complexes. Chaque modèle est adapté à une situation. La fiabilité des prédictions va souvent de pair avec un temps de calcul important. Si les modèles gaussiens les plus simples sont soumis à un grand nombre de limitations, ils permettent néanmoins des calculs extrêmement rapides et adaptés à la situation de crise ou aux études de danger nécessitant un nombre important de scénarios à évaluer. Les modèles les plus fins issus de la CFD permettent de prendre en compte des situations plus réelles avec la modélisation de l'impact des obstacles sur l'écoulement.

Dans chacun des modèles, la turbulence joue un rôle majeur dans la dispersion. Selon le modèle utilisé, cette turbulence est gérée de manière différente. La figure ci-après reprend les éléments énoncés plus haut dans ce chapitre :

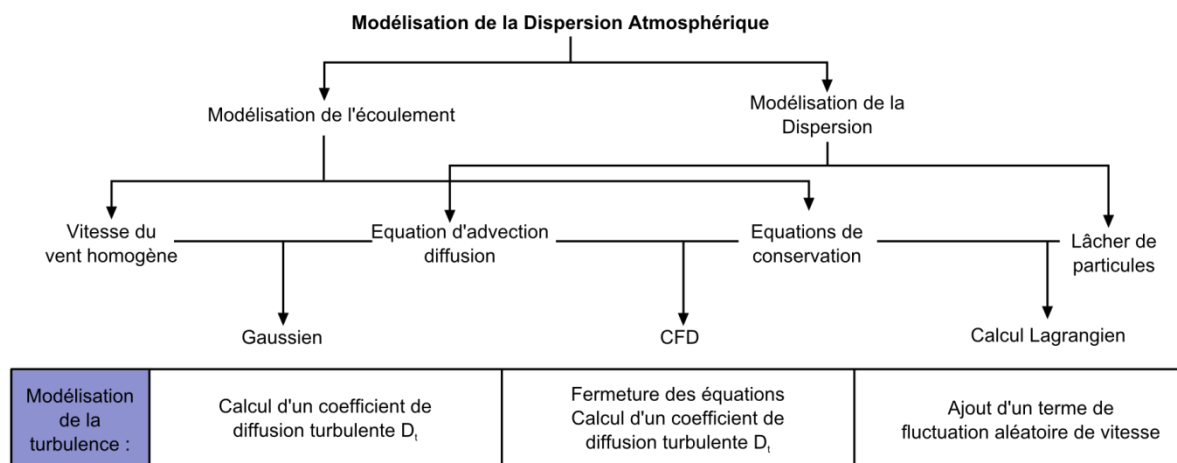


Figure I-19 : Modélisation de la turbulence pour différents modèles

Les modèles gaussiens utilisent un coefficient de diffusion turbulente qui est récupéré à partir d'abaques. Ces abaques ont été déterminés à partir d'expérimentations sur un terrain plat pour différentes conditions atmosphériques. Les modèles CFD modélisent à différents degrés la turbulence atmosphérique. Si les méthodes DNS résolvent directement la turbulence jusqu'aux plus petits tourbillons, les méthodes LES résolvent les équations de l'écoulement pour les grands tourbillons et modélisent les petits. En ce qui concerne les méthodes RANS, l'ensemble de la turbulence est modélisée. On récupère un coefficient de dispersion turbulente, qui est fonction des grandeurs utilisées en tant qu'équations de fermeture du système. Dans les méthodes RANS les plus usitées, les équations de transport d'énergie cinétique turbulente et de son taux de dissipation

permettent de calculer le coefficient de diffusion turbulente dans l'espace. Celui-ci est utilisé par l'ADE pour calculer la dispersion atmosphérique.

Nous verrons dans les chapitres suivants différentes méthodes et outils utilisant la précision des méthodes CFD tout en maintenant une rapidité de calcul permettant l'opérabilité d'un tel modèle.

Partie II : Modèles issus de l'intelligence artificielle

Dans ces deux chapitres sont décrits les outils issus de l'intelligence artificielle et utilisés dans cette thèse : les automates cellulaires et les réseaux de neurones. A chaque fois, les fondements historiques puis techniques seront abordés. Par ailleurs, des applications à la mécanique des fluides et à la dispersion atmosphérique seront détaillées.

II-1.	AUTOMATES CELLULAIRES	47
II-1.1.	Historique	47
II-1.2.	Principes généraux et définitions	50
II-1.3.	Applications à la mécanique des fluides.....	52
II-1.4.	Conclusion	54
II-2.	RESEAUX DE NEURONES.....	55
II-2.1.	Historique et généralités	55
II-2.2.	Définitions	55
II-2.3.	Applications en dispersion atmosphérique	64

II-1. Automates cellulaires

Ce chapitre présente les automates cellulaires, l'un des outils utilisés dans le cadre de cette thèse. Ceux-ci font partie des systèmes dynamiques discrets, permettant la modélisation de phénomènes spatiotemporels. Les fondements historiques, théoriques et les applications en mécanique des fluides y sont décrits.

II-1.1. Historique

Les automates cellulaires ont été inventés dans les années 1940, à l'occasion du projet Manhattan, visant à fabriquer la bombe atomique. Initialement, Von Neumann était intéressé par la formalisation logique du mécanisme d'autoreproduction identifié dans de nombreux procédés biologiques. L'intérêt de tels travaux était d'appliquer le principe de la machine de Turing et d'y intégrer une information permettant à la machine de se « reproduire » elle-même. Cette description correspond à la description actuelle du fonctionnement de la molécule d'ADN. Sur cette base, Von Neumann (1966) conçut un automate cellulaire de quelques 200 000 cellules à 29 états contenant un copieur universel, une description de lui-même et une machine de Turing pour la supervision. Chaque cellule est connectée localement avec les cellules qui l'entourent. L'évolution des états s'effectue selon une règle de mise à jour fonction des états voisins. Cette règle s'applique de manière synchrone sur l'ensemble des cellules. Ce fonctionnement est représentatif d'un certain nombre de systèmes physiques ou biologique. Kari (2005) en propose une revue. John Conway popularise cette vision avec son automate cellulaire « game of life », rapporté par Gardner (1970). Le principe de fonctionnement est très simple. Chaque cellule peut prendre deux états symbolisés par des valeurs, 1 représentant l'état « vivant » et 0 l'état « vide ». En fonction du voisinage immédiat de la cellule, composé des huit cellules adjacentes, l'état varie ou reste inchangé. On distingue plusieurs cas, représentant des processus biologiques :

- ▶ Naissance : chaque cellule à l'état 0 (vide) adjacente à exactement trois cellules vivantes sera au pas de temps suivant à l'état « vivant ».
- ▶ Survie : chaque cellule à l'état 1 (vivant) avec deux ou trois voisins à l'état 1 reste à l'état 1 au pas de temps suivant.
- ▶ Mort : chaque cellule à l'état 1 avec quatre voisins ou plus à l'état 1 passe à l'état 0 au pas de temps suivant. Si le nombre de voisins à l'état 1 est égal ou inférieur à 1, la cellule passe également à l'état 0.

Les règles de transition d'états dictées par Conway sont extrêmement simples. Cependant, il n'en reste pas moins que le comportement d'un état quelconque est difficilement prévisible, notamment sur le devenir de la « colonie » initiale. Aussi, certains motifs initiaux conduisent à un état 0 pour l'ensemble des cellules, d'autres conduisent à des états 1, stables. Enfin, certains motifs sont modifiés à travers les pas de temps en tenant compte d'une période. Le jeu de la vie inventé par Conway a rencontré un intérêt important, guidé par la recherche de motifs identifiables : stables, oscillants, en translation ou autoreproducteurs (figure ci-après) :

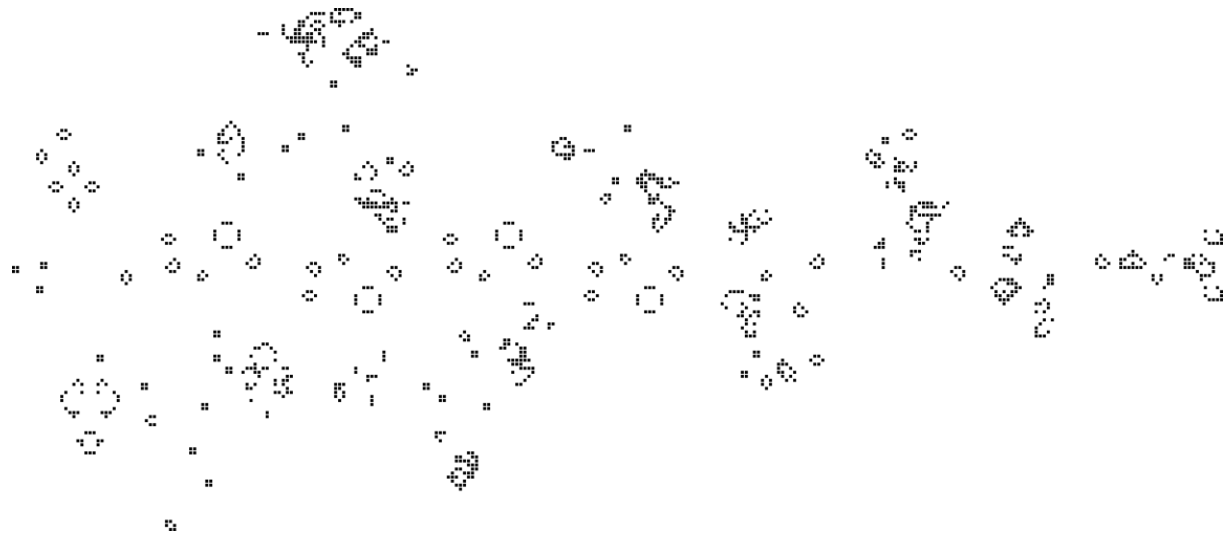


Figure II-1 : Échantillon de configuration de l'automate cellulaire "game of Life" (source : http://commons.wikimedia.org/wiki/File:Puffer_train.svg)

Wolfram (1983) a étudié de manière systématique les automates cellulaires unidimensionnels et à deux états. Il est initialisé de manière aléatoire, tout comme dans le jeu de la vie de Conway, avec les valeurs « 0 » et « 1 ». Le voisinage est réduit puisque les voisins directs, à gauche et à droite, sont les seuls pris en compte par la règle de transition. Huit configurations différentes sont possibles. A chaque configuration, deux règles de transition sont possibles. Il existe donc 256 automates différents. Wolfram évalue alors l'évolution temporelle de chaque automate unidimensionnel à travers des observations sur les graphiques générés.

Un exemple de règle est illustré par la figure suivante :

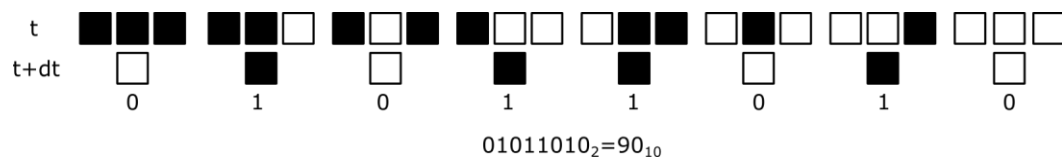


Figure II-2 : Règle de transition de l'automate cellulaire n°90

Le numéro de la règle est assigné entre 0 et 255. Chaque configuration possible pour le voisinage en binaire est écrit dans l'ordre décroissant (111, 110, 101, ...). La valeur de la cellule centrale est à évaluer au pas de temps suivant. Aussi, son état ainsi que celui de ces deux voisins constituent une combinaison. La figure II.2 montre, pour chaque configuration, l'état de la cellule au pas de temps suivant. Le résultat pour l'ensemble de la règle est interprété comme le numéro de la règle. Ici, 01011010 en binaire donne 90 en base 10. Le calcul synchrone de chaque cellule pour un automate cellulaire unidimensionnel pour un nombre de pas de temps donné est illustré sur la figure ci-après :

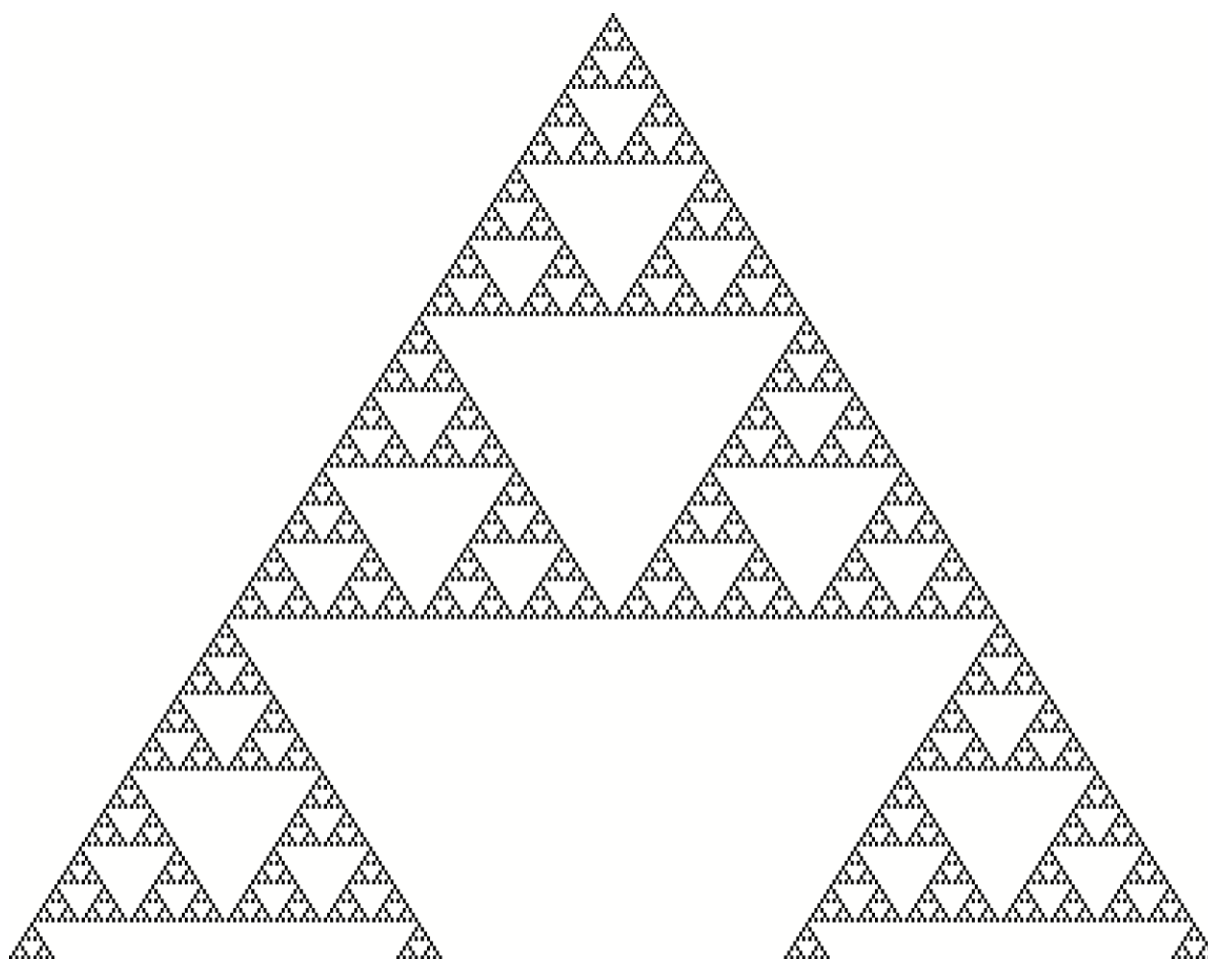


Figure II-3 : Visualisation de l'automate cellulaire de Wolfram - règle 90 - 200 pas de temps

Sur cet exemple, une configuration initiale simple composée d'une cellule dans un état et l'ensemble des autres dans l'état opposé amène à l'observation de structures complexes, répétitives et ordonnées.

Wolfram identifie ainsi différents comportements des automates cellulaires générés par les 256 règles possibles. La classe I conduit toute configuration initiale à un état homogène. La classe II fait apparaître des structures stables ou périodiques. Le comportement des règles de la classe III montre un comportement chaotique avec des motifs apériodiques. La classe IV montre l'émergence de structures complexes pouvant osciller, se translater et s'auto organiser malgré des perturbations structurelles. Les conclusions de son travail sont qu'à partir de simples règles locales, il est possible d'observer des phénomènes très complexes tels que l'évolution biologique des systèmes ou la structure de croissance des organismes. Il a démontré que les automates cellulaires peuvent émuler le comportement spécifique d'un phénomène biologique ou physique observé en réalité. Néanmoins, cette vision de Wolfram fait l'objet de débat (Schleicher, 2013). La pertinence de la modélisation réside dans le choix de la règle à appliquer.

Par la suite, les automates cellulaires ont notamment été étudiés à travers l'utilisation en modélisation de systèmes complexes dans des domaines variés : chimie, physique, biologie, écologie, économie, sociologie, ...

II-1.2. Principes généraux et définitions

Les automates cellulaires (Cellular Automata ou CA) sont des modèles utilisés notamment pour la modélisation de phénomènes physiques selon une discrétisation spatiale et temporelle. Itami (1994) a donné une formalisation des automates cellulaires. Celle-ci est complétée par Ollinger (2002) et Hutzler (2006). L'espace de l'automate cellulaire est une grille de cellules avec une géométrie régulière, notée L . Les unités spatiales ou cellules sont de formes spécifiques. Les cellules carrées sont les plus utilisées. Les états discrets possibles pour l'automate cellulaire sont notés S . Dans cette définition, il existe un nombre fini d'états possibles. Le voisinage des cellules fournissant des valeurs d'entrée pour les règles de transition est noté N . Le temps est discrétisé. Le changement d'état se fait de manière synchrone, en utilisant la même règle de mise à jour f (ou fonction de transition) pour toutes les cellules composant l'automate cellulaire. Cette règle de transition dépend uniquement de relations locales avec les voisins de la cellule concernée. L'automate cellulaire est donc défini par l'ensemble (L, S, N, f) . Une configuration est une fonction qui associe un état à chacune des cellules de la grille :

$$C_t: L \rightarrow S \quad (II-1)$$

Ainsi, à partir d'une configuration initiale C_0 , la fonction de transition modifie C_t en C_{t+1} selon la relation :

$$C_{t+1}(r) = f(\{C_t(i) | i \in N(r)\}) \quad (II-2)$$

C_t	Configuration au pas de temps t
$N(r)$	Nombre de voisins de la cellule r

$$N(r) = \{i \in L | r - i \in N\} \quad (II-3)$$

Chaque paramètre doit ainsi être caractérisé en fonction de l'application envisagée. Plusieurs formes peuvent être utilisées pour les cellules de l'automate cellulaire :

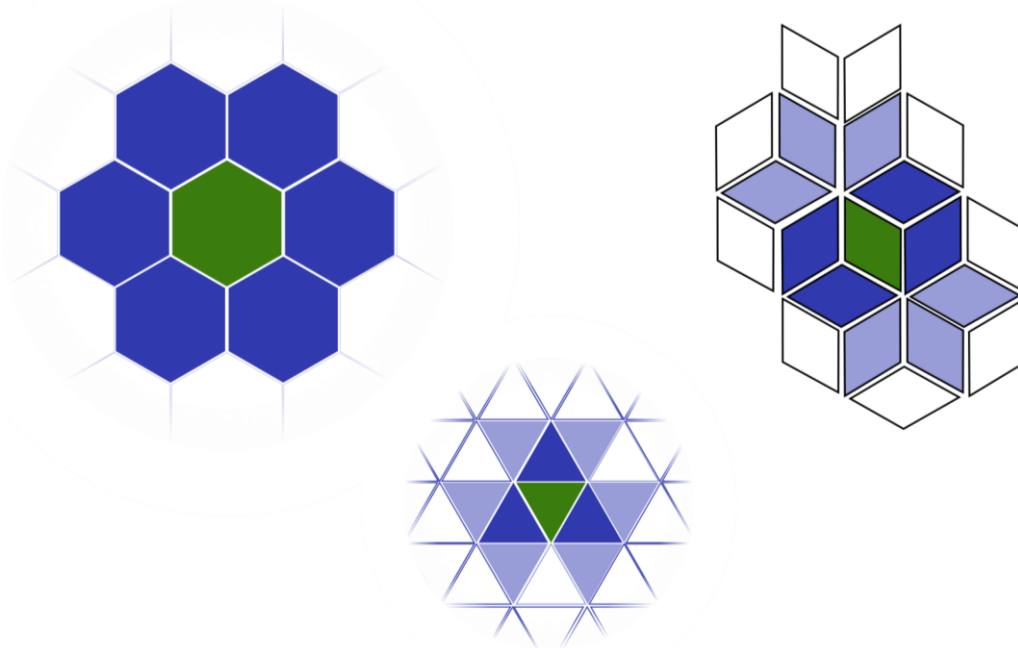


Figure II-4 : Différentes formes d'automates cellulaires

En vert figure la cellule concernée et en bleu foncé le voisinage immédiat, avec un rayon de 1. Selon la règle de transition, plusieurs voisinages peuvent être envisagés. Les automates cellulaires ayant une grille régulière composée de cellules carrées utilisent le voisinage de Moore ou le voisinage de Von Neumann pour entrée de la règle de transition.

Le voisinage de Moore au rayon 1 est composé des cellules partageant un point avec la cellule centrale. Le voisinage au rayon 2 est composé des cellules partageant un point avec le voisinage au rayon 1. Ainsi, le nombre de voisins inclus dans le voisinage de Moore est égal à $(2r + 1)^2 - 1$. Le voisinage de Von Neumann de rayon 1 est lui composé uniquement des cellules partageant une arête. Le voisinage au rayon 2 est composé des cellules partageant une arête avec le voisinage au rayon 1. Ainsi, le nombre de voisins inclus dans le voisinage de Von Neumann est égal à $2r(r + 1) + 1$. La figure ci-après représente différents rayons pour ces 2 types de voisinages :

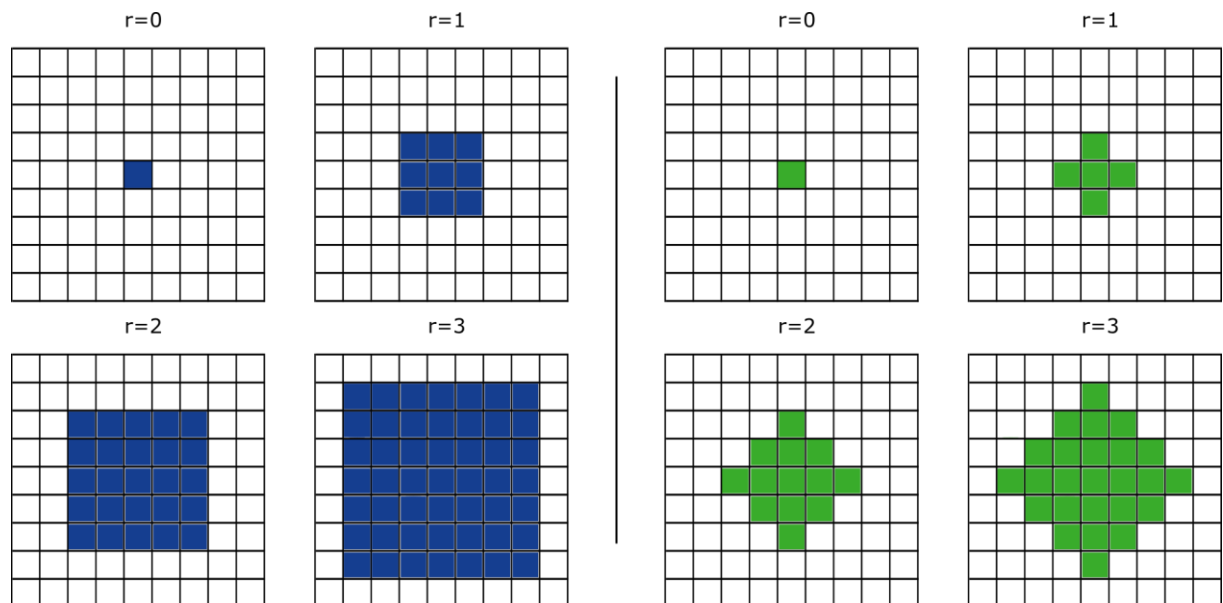


Figure II-5 : Voisinage de Moore (gauche) et Von Neumann (droite) pour différents rayons

L'utilisation d'un rayon plus ou moins grand peut être perçue comme une manière de rendre plus ou moins locale une règle de transition.

Pour rendre fonctionnel un automate cellulaire grâce à un algorithme numérique, le nombre de cellules ne peut être infini. Un domaine doit être défini, impliquant la présence de conditions aux limites. Celles-ci varient en fonction du phénomène à modéliser. Les conditions possibles sont la périodicité, la réflexion ou l'imposition d'une valeur fixe. Les conditions de type périodiques sont utilisées afin de supprimer les limites. Les conditions réfléchissives, au contraire permettent de simuler un système avec des limites. Elles correspondent à des conditions aux limites de sorties des applications CFD. Fixer des limites de type valeurs fixes permet de représenter des conditions physiques de types puits ou source.

Comme dans une approche d'un problème d'éléments finis, il est nécessaire d'initialiser le domaine avant d'initier le calcul ou simulation. Cette étape est à la discrétion du modélisateur, en fonction du phénomène physique et de la configuration à tester.

Le nombre d'états possibles dépend lui aussi du phénomène à modéliser. Les automates cellulaires en tant qu'objets de recherche possèdent en général un nombre

restreint d'états possibles. Ceci permet d'évaluer systématiquement le comportement de chaque automate. Augmenter ce nombre d'états permet en revanche de gagner en précision.

La règle de transition T définit comment l'état d'une cellule est modifié d'un pas de temps à un autre. La règle de transition peut être définie explicitement en détaillant chaque configuration possible de voisinage d'une cellule et l'état résultant pour chaque configuration. Cette option est utilisée par exemple par Wolfram (2002). Il est possible de définir cette règle en utilisant une formule donnant le nouvel état de la cellule, en fonction de son état et de ceux de ses voisins. Un grand nombre de techniques mathématiques quantitatives peut être utilisé : réseaux de neurones artificiels, algorithme génétique, chaîne de Markov, simulations de Monte Carlo, logique floue, etc....

Différents auteurs s'intéressent aux automates cellulaires pour représenter la physique de phénomènes en considérant exclusivement des interactions locales. Ces méthodes s'appuient généralement sur une description microscopique de la physique. De tels modèles discrétisent en temps et en espace des phases discrètes, pouvant être vue comme des « molécules booléennes » (Frisch et al., 1986). Elles constituent une des applications des automates cellulaires en mécanique des fluides, sous l'appellation de gaz sur réseau ou *lattice-gas automata*.

II-1.3. Applications à la mécanique des fluides

Les automates cellulaires sont utilisés en mécanique des fluides de deux manières différentes selon le niveau d'observation. On distingue ainsi une vision microscopique représentée par l'utilisation des automates cellulaires de type gaz sur réseau ou *lattice-gas cellular automata* et les méthodes utilisant les lois de conservation au niveau macroscopique.

II-1.3.1 Gaz sur réseau et méthode de Lattice Boltzmann

Les modèles de gaz sur réseau sont des automates cellulaires spécifiques qui sont spécialisés dans la résolution des équations de Navier-Stokes au niveau particulaire. Le modèle HPP (Hardy et al., 1973) est le premier modèle de gaz sur réseau. Son fonctionnement est le suivant :

Le réseau est constitué de particules de même masse, distribuées selon l'initialisation. Ces particules représentent le fluide. Chaque particule est caractérisée par sa position sur l'automate. L'état de chaque particule représente sa direction. Les modèles de gaz sur réseau utilisent le principe de Pauli : 2 particules ne peuvent se trouver sur le même nœud, dans le même état, excepté si cet état est le repos. Deux phases se dégagent ainsi :

- ▶ Une phase de déplacement libre : les particules se déplacent en suivant les mailles du réseau, selon la valeur de la vitesse.
- ▶ Une phase de collision : si plusieurs particules arrivent sur le même nœud, il y a collision : chaque particule va changer de direction. Les règles sont établies en amont, à partir des principes généraux de conservation de la quantité de mouvement et l'énergie.

Plusieurs modèles existent et diffèrent, notamment par la géométrie du maillage utilisé.

Les méthodes de Lattice Boltzmann ou réseau de Boltzmann utilisent les automates cellulaires en combinaison de la physique statistique (Chopard et al., 2002). L'équation de Boltzmann représente l'évolution spatio-temporelle d'une fonction $f = f(x, c, t)$

représentant la distribution de particules ayant une vitesse donnée à une position et un temps donné :

$$\frac{\partial f}{\partial t} + u_i \frac{\partial f}{\partial x_i} + F_i \frac{\partial f}{\partial u_i} = \left(\frac{\partial f}{\partial t} \right)_{collision} \quad (II-4)$$

u_i	Vitesse selon x_i	m.s^{-1}
F_i	Force extérieure	N
x_i	Distance	m

Lorsque les particules ne sont pas en collision, elles évoluent grâce à l'action de la force F . Lorsqu'elles se rencontrent, l'évolution du système dépend de la forme de l'opérateur de collision.

La discrétisation de l'équation différentielle de Boltzmann, à travers l'usage des LGCA, le passage de valeurs booléennes à des valeurs réelles et des améliorations de l'opérateur de collisions ont permis d'utiliser cette technique pour simuler les écoulements de la mécanique des fluides. Divers modèles et applications sont disponibles dans la littérature. On peut notamment citer les travaux de Margolus (1984) sur les concepts d'entropie et d'énergie dans les automates cellulaires, ceux de Toffoli (1984) dans la modélisation de la physique par les automates cellulaires plutôt que par la résolution d'équations différentielles.

II-1.3.2 Application en dispersion atmosphérique

L'utilisation des automates cellulaires pour des applications en mécanique des fluides en restant dans une approche macroscopique est également possible.

Vick (2007) s'est intéressé à la modélisation de phénomènes multi physique par automates cellulaires. Chaque phénomène est décomposé et mis en équation en utilisant les différences finies de manière explicite : advection, diffusion, réaction. Le phénomène global est ensuite assemblé en appliquant chaque règle élémentaire en cascade, de manière séquentielle. La difficulté réside dans les conditions de stabilité des équations. Ainsi, il insiste sur l'augmentation de la précision avec la diminution des pas de temps et sur la nécessité de bien déterminer les paramètres de l'automate cellulaire.

Des études complémentaires sur la modélisation de la dispersion de polluants ont été menées selon une approche sensiblement similaire par Guariso et Maniezzo (1992), et Sarkar et Abbasi (2006). Les auteurs s'intéressent à l'influence de l'environnement sur les zones d'impact suite à une explosion. Le point de départ est la décomposition des phénomènes liés à la dispersion de polluant : la gravité (dans le cas d'un gaz à la densité différente de l'air), le transport advectif (déterminé par la vitesse du vent) et la diffusion turbulente. Ces méthodes sont tridimensionnelles.

Aussi, le voisinage de Von Neumann est adapté à cette géométrie (figure ci-après) :

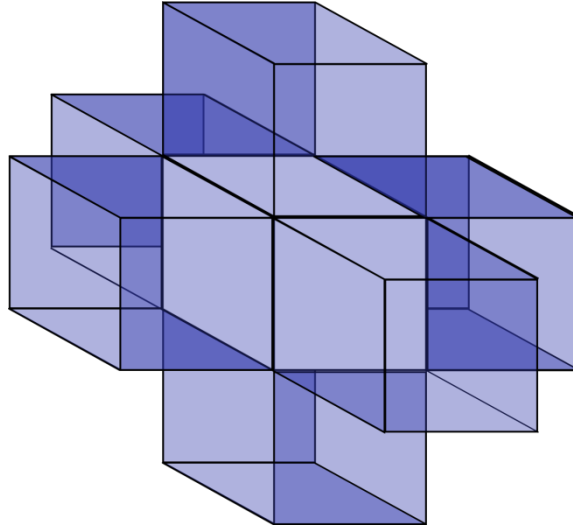


Figure II-6 : Structure tridimensionnelle du voisinage de Von Neumann

L'équation globale de modification de la concentration dans la cellule est :

$$\Delta C_{ijk} = \alpha C_{i,j,k} + \beta C_{i,j,k+1} + \gamma C_{i,j-1,k} + \varepsilon C_{i-1,j,k} + d_x C_{i+1,j,k} + d_y C_{i,j+1,k} + d_z C_{i,j,k-1} \quad (II-5)$$

La difficulté réside dans la détermination de ces coefficients. Ceux-ci dépendent d'un grand nombre de paramètres, notamment : l'environnement, la discrétisation des équations, le nombre de voisins pris en compte. Marin et al. (2000) utilise une méthode similaire pour modéliser la dispersion de polluant à partir de quatre cheminées industrielles distribuées spatialement. La calibration des coefficients est ici effectuée pour correspondre à chaque situation par essais et erreurs.

II-1.4. Conclusion

L'utilisation des automates cellulaires peut être intéressante en mécanique des fluides et donc en dispersion atmosphérique. La description à un niveau macroscopique de l'évolution d'un scalaire tel que la concentration est différente d'une description à partir d'équations différentielles continues, discrétisées. Les règles peuvent être de plusieurs natures : empirique ou bien physique. A chaque fois, le choix du nombre de voisins et des dimensions du maillage est déterminant. De plus, des paramètres sont à ajuster. La difficulté majeure réside en ce point. Généralement, cet ajustement est effectué empiriquement à partir d'expérimentations. Les paramètres sont ajustés afin que la modélisation corresponde à l'observation. D'autres méthodes peuvent être envisagées. Une méthode d'apprentissage statistique est développée dans le chapitre ci-après.

II-2. Réseaux de neurones

Ce chapitre présente un outil d'apprentissage statistique de phénomènes non-linéaires que sont les réseaux de neurones. Les fondements historiques, théoriques et les applications en mécanique des fluides y sont décrits.

II-2.1. Historique et généralités

Le développement des réseaux de neurones artificiels (Artificial Neural Networks ou ANN) provient de l'idée initiale d'imiter la structure et le comportement du cerveau humain (Minsky and Papert, 1969). En cela, la recherche de la modélisation du cerveau biologique se rapproche du neuromimétisme. Néanmoins, l'utilisation que nous ferons des réseaux de neurones dans la suite de cette thèse ne le sera pas sous cette lecture. Cet outil sera utilisé ici pour ces capacités d'approximation de systèmes non linéaires. L'objet initial est donc abandonné au profit des analyses mathématiques qu'il permet d'effectuer.

Les réseaux de neurones sont des outils de modélisation de phénomènes non-linéaires. Ces modèles empiriques, sont de types « boîtes noires ». Il s'agit, pour calibrer les paramètres du ANN, d'effectuer un apprentissage sur une base de données du phénomène. On fait correspondre cette étape au calcul d'une régression non-linéaire. Les principes importants sont détaillés dans la suite de ce chapitre.

II-2.2. Définitions

II-2.2.1 Neurone formel

Dreyfus (2004) présente un neurone comme un opérateur mathématique qui modifie une ou plusieurs variables d'entrée en une ou plusieurs variables de sortie. Les variables d'entrée et de sortie sont stockées respectivement dans un vecteur x et y . Dans la plupart des applications et dans la suite des définitions, on considèrera le vecteur de sortie comme un singleton, donc un scalaire y . L'opération réalisée par le neurone est l'application d'une fonction non-linéaire des variables d'entrée et de paramètres spécifiques au neurone w :

$$y = f(x, w) \quad (II-6)$$

Le calcul de la sortie du neurone s'effectue en deux étapes. Le calcul préalable du potentiel du neurone est effectué à partir de la somme pondérée des variables d'entrée (vecteur x) et des paramètres spécifiques du neurone (vecteur w). La fonction f ou fonction d'activation est alors appliquée au potentiel du neurone. Selon l'objectif d'utilisation du neurone, plusieurs choix sont possibles :

Fonction à seuil telle que la fonction de Heaviside définie comme suit :

$$\forall x \in \mathbb{R}, H(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \quad (II-7)$$

- Fonction sigmoïde : la fonction tangente hyperbolique est la plus couramment rencontrée (figure ci-après). Elle est continue et dérivable.

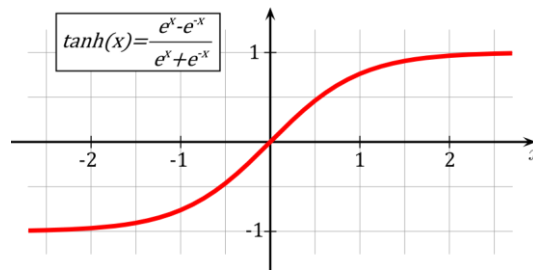


Figure II-7 : représentation graphique d'une fonction sigmoïde (source : commons.wikimedia.org/wiki/File:Hyperbolic_Tangent.svg)

► Fonction linéaire : Il s'agit de la fonction identité $y = x$.

L'équation du neurone devient alors :

$$y = f\left(\sum_{j=0}^n (w_j x_j)\right) \quad (II-8)$$

f	Fonction d'activation
w_j	Valeurs des vecteurs des paramètres
x_j	Valeurs du vecteur des variables d'entrée

Avec f la fonction d'activation et w_j et x_j respectivement les valeurs des vecteurs des paramètres et des variables d'entrée. On représente généralement un neurone formel par l'illustration suivante :

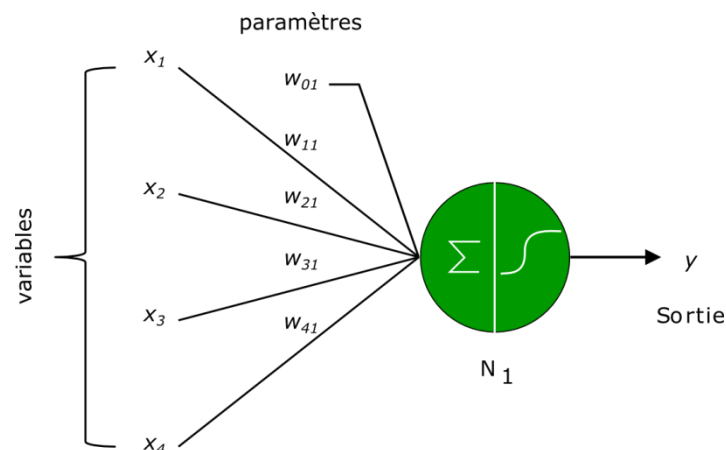


Figure II-8 : Neurone mathématique - Représentation avec une fonction d'activation de type sigmoïde

Un prétraitement est nécessaire afin que les différentes variables soient comprises dans le même intervalle afin qu'elles aient la même importance a priori. Ce prétraitement de base des variables d'entrée consiste à les centrer et réduire. D'autre part, il est important que la combinaison linéaire des variables d'entrée et des paramètres ne sature pas la sigmoïde, de manière à éviter l'obtention d'une dérivée nulle, empêchant l'apprentissage (II-2.2.4).

II-2.2.2 Réseau de neurones

La connexion de plusieurs neurones constitue un réseau de neurones. Le terme de connexion provient lui aussi des similitudes avec l'approche biologique. Cette expression se traduit mathématiquement par des équations et un algorithme. L'intérêt de cette mise en relation réside dans la composition des fonctions non-linéaires réalisées par chacun des neurones. L'architecture de tels réseaux est particulière. On distingue ainsi différentes couches : la ou les couches cachées et la couche de sortie. La valeur de sortie des neurones des couches cachées correspond à une variable d'entrée pour un ou plusieurs neurones d'une autre couche cachée ou de la couche de sortie. La valeur de sortie des neurones d'une couche cachée n'est pas interprétée en général. Sa mesure ne présente pas d'intérêt. La valeur de sortie d'un neurone de la couche de sortie est elle connue. C'est l'objet de l'utilisation des réseaux de neurones. Dans ce travail, cette valeur représente une grandeur physique (concentration, vitesse, ...).

L'architecture d'un réseau de neurones est aussi qualifiée par sa prise en compte éventuelle de l'aspect temporel. On distingue les réseaux de neurones statiques des réseaux de neurones dynamiques.

Dans les réseaux statiques, l'information circule de l'entrée vers la sortie, sans bouclage. De ce fait, un grand nombre de possibilités existent. Néanmoins, la majorité des applications des réseaux de neurones utilisent des réseaux à couches dont la figure ci-après est un exemple :

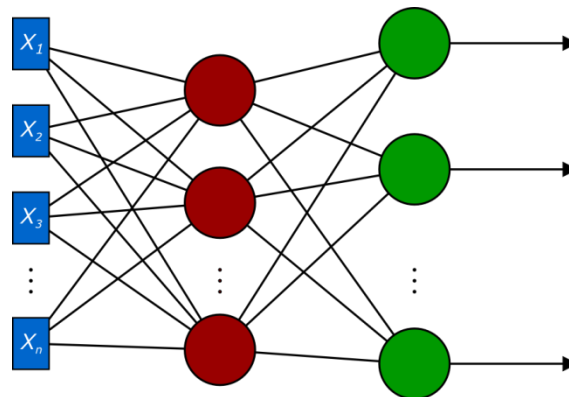


Figure II-9 : Réseau de neurones à n variables (bleu), N_c neurones cachés (rouge), N_s neurones de sortie (vert)

Chacune des sorties est calculée à partir de la fonction réalisée par le neurone de sortie, elle-même issue des fonctions non-linéaires réalisées par les neurones précédents. Pour chaque calcul de la fonction réalisée par le neurone, le temps de calcul est suffisamment faible pour le considérer comme négligeable.

Un réseau de neurones dynamique ou bouclé ou récurrent est un réseau de neurones dont la ou les sorties sont connectées avec une ou plusieurs entrées. (Figure ci-après).

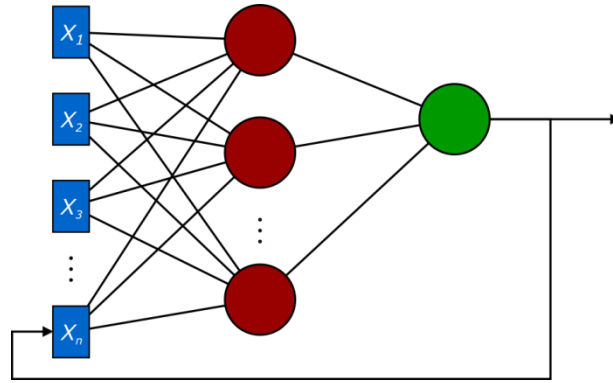


Figure II-10 : Illustration d'un réseau de neurones récurrent

II-2.2.3 Le Perceptron Multicouches

La plupart du temps, et c'est le cas dans ce travail, les réseaux de neurones sont utilisés avec une couche de neurones cachés à fonction d'activation sigmoïde et une couche de sortie linéaire. On utilisera ici uniquement des réseaux de neurones statiques. Les caractéristiques précédentes sont celles du Perceptron Multicouches (*MultiLayer Perceptron* ou *MLP*). Un tel modèle est représenté dans la figure ci-après :

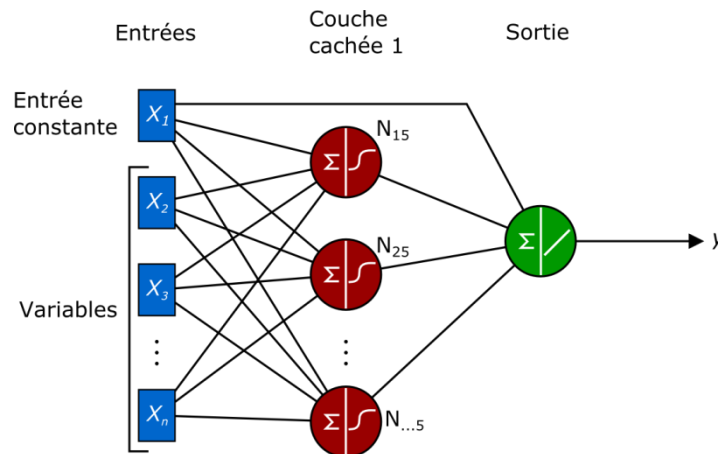


Figure II-11 : Schéma d'un Perceptron Multicouches

L'expression mathématique de ce modèle de réseau de neurones est :

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}_2 \cdot \mathbf{f}(\mathbf{W}_1 \mathbf{x}) = \sum_{i=1}^{N_c} [W_{N_c+1,i} \tanh(\sum_{j=1}^n w_{ij} x_j + w_{i0})] + w_{N_c+1,0} \quad (II-9)$$

\mathbf{x}	Vecteur des variables
\mathbf{w}_2	Vecteur des paramètres de la couche de sortie
\mathbf{W}_1	Matrice des connexions de la première couche (entre les variables d'entrée et les neurones de la couche cachée)
\mathbf{f}	Vecteur des sorties des neurones cachés

De plus, un biais est ajouté au calcul. Il s'agit d'une entrée constante qui confère au modèle un degré de liberté supplémentaire en lui permettant de délivrer une sortie non nulle lorsque toutes les variables d'entrée sont nulles.

Un tel modèle est très intéressant de par les propriétés qui lui ont associées, notamment l'approximation universelle et la parcimonie. Hornik, Stinchcombe, & White (1989) ont établi que toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire.

Il s'agit d'un théorème d'existence de l'approximation universelle, qui ne fournit pas de méthode pour déterminer les paramètres adéquats.

Selon Barron (1993), si l'approximation dépend des paramètres ajustables de manière non linéaire, elle est plus parcimonieuse que si elle dépend linéairement des paramètres. En d'autres termes, avec un nombre croissant de variables, le nombre de paramètres nécessaires à l'approximation est plus important dans le cas de modèles linéaires en leurs paramètres que dans le cas des réseaux de neurones non linéaires en leurs paramètres. Ainsi, la complexité d'un réseau de neurones par rapport à un modèle linéaire sera plus faible. Le phénomène de surajustement, qui est détaillé dans la suite, sera alors plus facilement contrôlable.

Ces deux propriétés expliquent que le choix se soit porté sur ce type de modèle pour la suite de l'étude.

II-2.2.4 Apprentissage

L'apprentissage est le processus permettant d'ajuster les paramètres des différentes couches du réseau de neurones afin que l'erreur sur la sortie soit la plus faible possible. Il est possible de définir une fonction de coût, traduisant la « distance » entre les sorties du modèle $y(\mathbf{x}_k, \mathbf{w})$ et les mesures y_k^p issues de la base d'exemples de l'apprentissage (de dimensions N). Plusieurs fonctions existent, la plus usitée étant la fonction de coût des moindres carrés :

$$J(\mathbf{w}) = \sum_{k=1}^N \left(y_k^p - y(\mathbf{x}_k, \mathbf{w}) \right)^2 \quad (II-10)$$

y^p	Valeurs cibles
$y(\mathbf{x}, \mathbf{w})$	Valeur estimée par le modèle
\mathbf{x}	Vecteur des entrées
\mathbf{w}	Paramètres du réseau

Le modèle présenté n'est pas linéaire en ses paramètres, aussi, il est nécessaire d'utiliser des techniques d'optimisation pour minimiser cette fonction de coût. Ces méthodes itératives utilisent la modification des paramètres en fonction du gradient de la fonction de coût par rapport à ses paramètres. A chaque itération, on évalue le gradient de la fonction puis on modifie les paramètres afin d'approcher un minimum de la fonction de coût.

Le calcul du gradient de la fonction de coût est effectué en utilisant l'algorithme de rétropropagation (Rumelhart et al., 1986). Une fois cette évaluation faite, on modifie les paramètres des neurones. Il existe plusieurs méthodes. Celle du gradient simple consiste à modifier les paramètres proportionnellement au gradient de la fonction de coût :

$$\partial \mathbf{w} = -\mu \nabla J(\mathbf{w}) \quad (II-11)$$

On calcule en fait la direction de la descente, elle est opposée au gradient. Le coefficient μ permet d'ajuster la vitesse de descente. C'est le pas du gradient ou pas d'apprentissage.

Les méthodes de second ordre sont dérivées de la méthode de Newton. Le gradient d'une fonction de coût $f(x)$ est nul au minimum. A partir d'un développement de Taylor d'une fonction $f(x)$ d'une seule variable x au voisinage d'un minimum x^* :

$$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^2 \left(\frac{\partial^2 f}{\partial x^2} \right)_{x=x^*} + o(x^3) \quad (II-12)$$

On obtient une approximation du gradient de la fonction de coût :

$$\frac{df}{dx} \approx (x - x^*) \left(\frac{\partial^2 f}{\partial x^2} \right)_{x=x^*} \quad (II-13)$$

La connaissance de la dérivée seconde de la fonction permet d'atteindre le minimum ; en modifiant la variable w de la quantité :

$$\Delta x = \frac{\left(\frac{\partial f}{\partial x} \right)}{\left(\frac{\partial^2 f}{\partial x^2} \right)_{x=x^*}}$$

Si on prend le cas où la fonction de coût est calculée pour plusieurs variables, la matrice hessienne remplace la dérivée seconde. Elle est de terme général $\frac{\partial^2 f}{\partial x_i \partial x_j}$. La modification suivante permet d'atteindre le minimum de la fonction de poids :

$$\Delta x = -H(x^*)^{-1} \nabla f(x) \quad (II-14)$$

En pratique, on ne connaît pas le vecteur x^* . Néanmoins, cette expression sert de base à des techniques qui utilisent l'approximation itérative de la matrice hessienne. L'algorithme utilisé dans la suite est celui de Levenberg-Marquardt, adapté aux réseaux de neurones par Hagan & Menhaj (1994). Il s'agit de la méthode actuelle la plus efficace.

La calibration des paramètres est faite en utilisant l'application d'un algorithme utilisant la base d'apprentissage et conçu pour diminuer l'erreur du modèle :

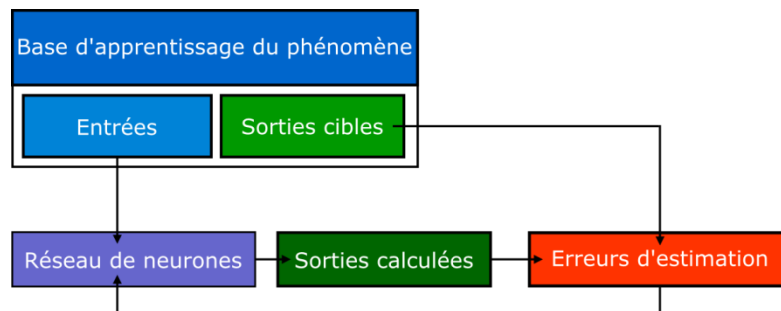


Figure II-12 : Principe de l'apprentissage d'un ANN.

Il convient au préalable d'initialiser les valeurs des paramètres. Les paramètres associés aux biais doivent être initialisés à zéro afin de s'assurer que les sigmoïdes des neurones sont initialement utilisées dans leur zone linéaire. Les valeurs des paramètres ne doivent pas permettre d'obtenir des valeurs de sortie des neurones cachées proche de 1 et -1. Effectivement, la dérivée des fonctions sigmoïdes au voisinage de 1 et -1 est voisine de zéro. Le gradient de la fonction de coût est lui-aussi proche de zéro et le

processus de minimisation ne peut pas commencer. Concrètement, cela se traduit par la sélection des valeurs d'initialisation des paramètres dans une distribution centrée de variance $1/(n - 1)$. D'autres formulations existent (Hernández-espinosa, 2001). Elles font intervenir pour la plupart le nombre de variables d'entrée dans le calcul de la distribution de valeurs.

II-2.2.5 Configuration d'un modèle

A partir des données établies précédemment, plusieurs modèles peuvent être créés. Les différentes configurations dépendent du choix de l'échantillonnage des données, de l'architecture et de l'initialisation.

i. Echantillonnage des données

Dans le cas où l'on dispose de suffisamment de données, il est nécessaire de les échantillonner pour créer la base d'apprentissage du phénomène. Cet échantillonnage dépend de la représentativité recherchée. Différents échantillonnages peuvent être utilisés. Ce problème sera abordé au III-1.3.1iii

ii. Architecture

L'architecture du réseau est sa définition. Aussi, il convient d'établir le nombre optimal de couches cachées du réseau de neurones, le nombre de neurones dans la couche cachée, les fonctions d'activation.

iii. Initialisation

L'initialisation des paramètres influe sur la qualité de l'apprentissage. Afin d'obtenir le meilleur modèle possible, il est nécessaire de tester suffisamment d'initialisations pour évaluer cette influence.

II-2.2.6 Sélection de modèles

Afin d'obtenir le meilleur modèle possible, il est nécessaire de juger de la qualité d'un apprentissage. Afin d'évaluer les capacités de généralisation d'un modèle sur des données non-inclues dans la base d'apprentissage, trois ensembles disjoints composent la base d'exemples du phénomène :

- ▶ Les exemples de l'ensemble d'apprentissage sont utilisés lors de la phase d'apprentissage du réseau de neurones. C'est à partir de ces exemples que s'effectue la modification des paramètres.
- ▶ L'ensemble d'arrêt permet de contrôler les performances de généralisation du modèle.
- ▶ L'ensemble de test sert à évaluer la performance du modèle.

Avoir la meilleure estimation des paramètres pour une base d'exemples donnée n'est, dans ce cas, pas l'unique objectif. Il convient d'éviter le phénomène de surajustement, pour garder une capacité à généraliser à des données extérieures à l'ensemble d'apprentissage.

Cette difficulté est résumée dans le dilemme biais-variance. Lorsqu'un modèle a une complexité élevée, il a plus de facilité à apprendre le bruit des données de la base d'exemple et a des difficultés à représenter le phénomène sous-jacent. A contrario, un modèle trop peu complexe n'arrive pas à s'ajuster suffisamment aux données (Gardner et Dorling, 1998). Le principe est illustré dans la figure ci-après :

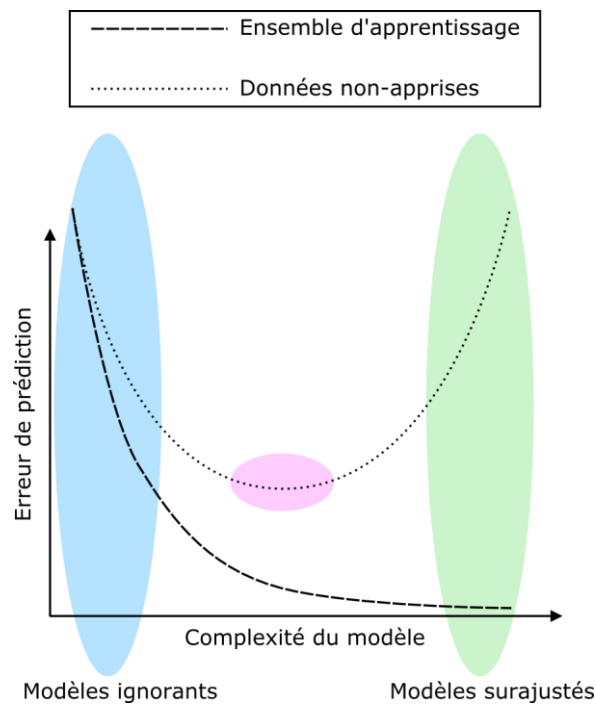


Figure II-13 : illustration du dilemme biais-variance

L'ensemble d'arrêt comprend des données non-apprises. L'erreur quadratique moyenne (EQM) nous renseigne sur la qualité du modèle.

$$EQMT = \sqrt{\frac{1}{N_T} \sum_{k=1}^{N_T} (y_k^p - g(\mathbf{x}_k, \mathbf{w}))^2} \quad (II-15)$$

y^p	Valeur cible
g	Valeur prédite par le modèle
k	Exemple de la base d'apprentissage
\mathbf{x}	Vecteur des entrées
\mathbf{w}	Paramètres du réseau

Aussi, lorsque la valeur de l'EQM sur l'ensemble de validation cesse de diminuer, le modèle se spécialise sur les données de l'ensemble d'apprentissage. La capacité de généralisation est diminuée. Afin de sélectionner le meilleur modèle disponible, on compare les erreurs quadratiques moyennes calculées sur l'ensemble de test. Cela permet d'évaluer la capacité du réseau de neurones à généraliser ses estimations sur des données disjointes de l'ensemble d'apprentissage.

Plusieurs méthodes de régularisation existent afin d'éviter le surajustement d'un réseau de neurones lors d'un apprentissage. Les méthodes actives opèrent directement lors de cette phase tandis que les méthodes passives permettent de sélectionner le meilleur modèle une fois l'apprentissage effectué.

i. Arrêt précoce

L'arrêt précoce est une technique permettant de sélectionner les modèles situés dans la zone optimale (en rose sur la figure précédente). Cette méthode active de régularisation s'effectue durant l'apprentissage. Le principe est d'arrêter l'apprentissage

avant le surajustement. On utilise alors l'ensemble de validation en tant qu'ensemble d'arrêt. En calculant à chaque itération la valeur de la fonction de coût, on stoppe l'apprentissage lorsque cette valeur sur l'ensemble d'arrêt commence à augmenter. Cela traduit le fait que l'apprentissage de la fonction devient spécifique à l'ensemble d'apprentissage et que la généralisation pourrait être moins bonne si le processus se poursuivait.

ii. Validation

Les méthodes de validation sont des méthodes passives. Elles permettent d'évaluer et de sélectionner le meilleur modèle parmi les différentes configurations établies.

En utilisant la validation simple, un seul apprentissage est effectué. La qualité du modèle est évaluée à l'aune de la qualité de son estimation sur l'ensemble de validation.

Une autre méthode possible est la validation croisée (Stone, 1974). En utilisant celle-ci, les modèles ne sont pas surajustés. A partir de l'ensemble des données, trois sous-ensembles sont constitués : un ensemble d'apprentissage, un ensemble d'arrêt et un ensemble de test. Le premier ensemble est divisé en D sous-ensembles disjoints. L'apprentissage est effectué sur $D-1$ sous-ensembles, l'ensemble restant étant utilisé en tant qu'ensemble d'arrêt. On calcule alors la somme des carrés des erreurs sur le sous-ensemble des données restantes :

$$S_i = \sum_{k \in \text{sous-ensemble de validation } i} (y_k^p - y(\mathbf{x}_k, \mathbf{w}))^2 \quad (II-16)$$

L'opération est répétée afin que chaque exemple ait été présent une seule fois dans un sous-ensemble de validation. Le Score de Validation Croisée ou S_{VC} est évalué et le modèle sélectionné est celui qui donne le S_{VC} le plus faible. Une fois le bon modèle déterminé, l'apprentissage est effectué sur l'ensemble apprentissage-arrêt. La performance est évaluée sur l'ensemble de test. Le schéma d'utilisation est visible sur la figure suivante :

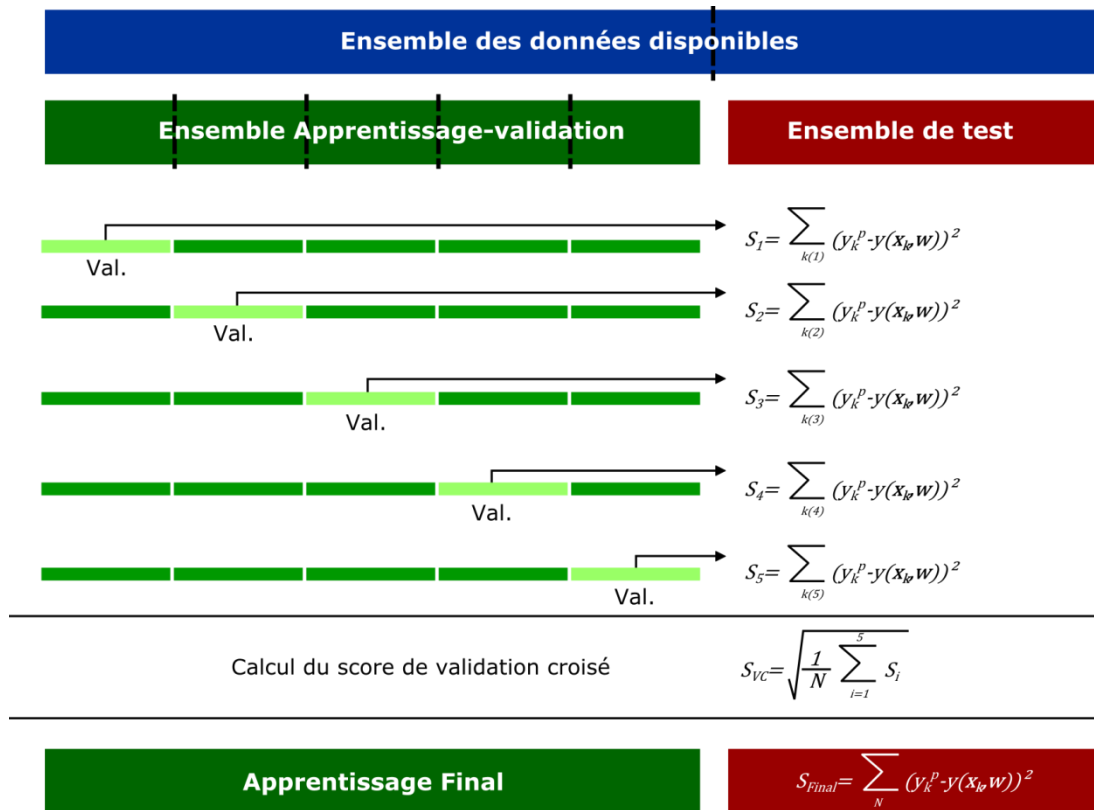


Figure II-14 : Utilisation de la validation croisée

Lorsque le sous-ensemble est constitué d'un unique exemple, la validation croisée est nommée *Leave-one-out*. Il s'agit de la limite en nombre de sous-ensemble D .

II-2.2.7 Résumé des étapes

Les différentes étapes de la définition du problème à la sélection du modèle le plus performant sont résumées sur la figure ci-après.

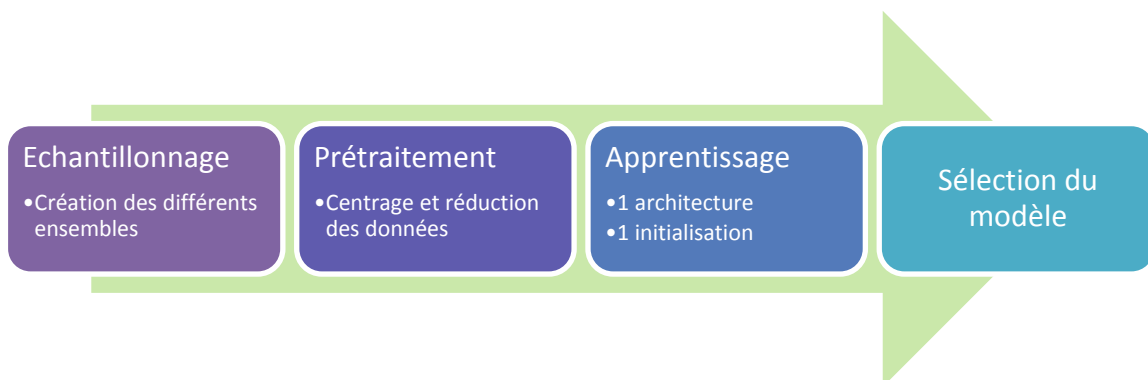


Figure II-15 : Schéma des différentes étapes de la définition du problème à la sélection du modèle le plus performant.

II-2.3. Applications en dispersion atmosphérique

Les réseaux de neurones sont utilisés en dispersion atmosphérique essentiellement afin de suivre la pollution chronique (Gardner et Dorling, 1998). Effectivement, les différentes législations sur la qualité de l'air imposent un suivi permanent et donc, des

données quantitatives importantes sont enregistrées. Différentes initiatives ont permis d'appliquer les réseaux de neurones à la dispersion atmosphérique accidentelle. Néanmoins, les exemples sont peu nombreux. Cette problématique peut se heurter à la difficulté de prédire l'évolution d'un système spatial avec les réseaux de neurones qui n'ont en général que peu de sorties. Aussi, la plupart des applications des ANN en dispersion atmosphérique ou en météorologie sont limitées à un phénomène en particulier : concentration en un point, valeur maximale journalière d'une concentration (Yi and Prybutok, 1996), prédiction de phénomènes violents (tornades), ...

II-2.3.1 Contrôle de la qualité de l'air

La présence de bases de données conséquentes pour la pollution atmosphérique, ainsi que la difficulté de prévoir l'évolution de concentrations sur de larges distances, en conditions réelles (instationnaire, présence d'obstacles, ...) font des réseaux de neurones un outil de choix pour de telles études. Différentes méthodes ont été déclinées.

Boznar, Lesjak, & Mlakar, (1993) proposent une méthode de prédiction à court terme de concentrations dans l'environnement de sites industriels. Ils utilisent un maillage de stations de mesures du polluant (SO_2) à intervalle de temps donné, ainsi que d'autres paramètres reliés aux conditions météorologiques. La sortie du réseau de neurones est la prédiction de concentration en un point, pour l'intervalle de temps suivant. L'apprentissage est effectué sur l'historique des données des stations météorologiques. Les résultats en prédiction montrent une bonne prévision des pics de polluants. L'intérêt de cette méthode est la précision de la détermination de la concentration en un point (0D). Par contre, cela ne donne pas d'informations sur les concentrations dans l'espace.

II-2.3.2 Dispersion accidentelle

Pelliccioni & Tirabassi, (2006), combinent l'apport des modèles gaussiens à ceux des réseaux de neurones statiques. Ayant observée des erreurs systématiques du modèle gaussien selon des conditions particulières, la méthode développée est l'utilisation des réseaux de neurones en tant que filtre. Les entrées du réseau de neurones concernent les conditions de stabilité atmosphérique (la hauteur de mélange, la longueur de Monin-Obukhov, vitesse du vent, vitesse de friction), des données sur la localisation du point dont on veut connaître la concentration (distance à la source) et la sortie d'un modèle gaussien. De cette manière, on observe une amélioration des performances du modèle couplé gaussien-ANN par rapport au modèle gaussien seul. On notera ici que le modèle développé permet d'obtenir une prévision selon l'axe principal de dispersion, de manière localisée (1D).

Il est également possible, en ajoutant des entrées géographiques sur la position de la concentration souhaitée, d'obtenir une prévision distribuée géographiquement (Wahid et al., 2013).

Parallèlement à ces méthodes, des recherches ont été effectuées afin de déterminer des coefficients de modèles de dispersion atmosphérique existant. Cette approche permet, une fois la détermination du coefficient effectuée, d'utiliser un outil issu d'une modélisation physique. On parle alors de modélisation semi-physique ou boîte grise (Oussar and Dreyfus, 2001). Ainsi, Devito, Andrews, & Roy, (2000) propose un modèle de détermination des coefficients de dispersion des modèles gaussiens. Les modèles gaussiens utilisent des coefficients de dispersion calculés en fonction de la stabilité atmosphérique. Une des limites de ces modèles est la suivante : pour une même distance à la source, les valeurs des coefficients de dispersion sont discrètes pour

différentes classes de stabilité. L'approche proposée ici est de déterminer ces coefficients à partir des réseaux de neurones, de manière à prendre en compte une variété de conditions météorologiques plus importantes. Le modèle ainsi créé est adapté à l'objectif de l'étude de Devito, à savoir la prévision de la distribution de concentration en champ proche. D'autres coefficients peuvent également être déterminés. Un exemple est celui de Bitsuamlak et al. (2004) qui propose de caractériser un coefficient d'accélération du vent dans le contexte de terrains complexes.

Enfin, Cao (2007) utilise les réseaux de neurones afin de déterminer le déplacement d'un nuage d'aérosols dans les trois dimensions de l'espace. La prévision est effectuée directement à partir des entrées du réseau de neurones. On distingue trois catégories dans les entrées proposées. Les éléments relatifs à la source, aux conditions météorologiques et à la sortie désirée. Dans le cadre de cette étude, la base de données du phénomène est réalisée à partir de données expérimentales. Les prévisions de concentrations de ce modèle sont de l'ordre de grandeur de celles des modèles gaussiens.

Partie III : Méthodes et Résultats

« Cet effort consenti pour comprendre l'Univers est l'une des rares choses qui élèvent la vie humaine au-dessus de la farce en lui en conférant un peu de dignité dans la tragédie »

Steven Weinberg

Il s'agit dans cette partie d'évaluer les possibilités d'utilisation des réseaux de neurones et des automates cellulaires à travers la modélisation de la dispersion atmosphérique, en champ libre et en champ obstrué, d'une source continue ou d'une bouffée.

III-1.	STRATEGIE DE L'ETUDE	69
III-1.1.	Constitution des bases de données CFD	72
III-1.2.	Lien avec la couche limite atmosphérique	73
III-1.3.	Principes et évaluation des réseaux de neurones créés.....	76
III-2.	MODELISATION DE LA DISPERSION ATMOSPHERIQUE D'UNE SOURCE CONTINUE EN CHAMP LIBRE.....	82
III-2.1.	Simulation CFD	82
III-2.2.	Base de données et base d'exemples.....	83
III-2.3.	Apprentissage et optimisation	84
III-2.4.	Evaluation du modèle.....	88
III-2.5.	Temps de calcul	94
III-2.6.	Conclusion	95
III-3.	MODELISATION DE LA DISPERSION ATMOSPHERIQUE D'UNE BOUFFEE EN CHAMP LIBRE.....	96
III-3.1.	Objectifs.....	96
III-3.2.	Méthode d'utilisation du CA-ANN	99
III-3.3.	Base de données et base d'exemples.....	102
III-3.4.	Réseau de neurones	105
III-3.5.	Evaluation du modèle CA-ANN.....	110
III-3.6.	Gain en temps de calcul	118
III-3.7.	Conclusion	119
III-4.	MODELISATION DE LA DISPERSION ATMOSPHERIQUE AUTOUR D'UN OBSTACLE CYLINDRIQUE	121
III-4.1.	Objectifs.....	121
III-4.2.	De la base de données à la base d'apprentissage	122
III-4.3.	Structure du réseau de neurones	126
III-4.4.	Optimisation et sélection des modèles	127
III-4.5.	Evaluation sur des cas non-appris	132
III-4.6.	Application aux cas non-appris.....	145
III-4.7.	Conclusion	151

III-1. Stratégie de l'étude

Ce chapitre présente la démarche développée dans cette étude. Les méthodes et outils utilisés sont détaillés. Une base de données CFD est utilisée pour les apprentissages des différents réseaux de neurones. Une méthodologie de simulation d'une couche limite atmosphérique sera explicitée afin de garantir une cohérence des résultats. Elle sera implémentée pour chaque base de données utilisée dans la suite de ce travail.

Comme vu dans les chapitres précédents, la dispersion atmosphérique en présence d'obstacles dans le champ proche est tributaire d'un grand nombre de variables. La modélisation de tels phénomènes de manière précise se révèle être une difficulté de premier plan pour les industriels lors des études d'impacts, de dangers et en situation d'urgence.

D'une part, les modèles simples et rapides ne sont pas utilisables en champ proche et obstrué. D'autre part, les approches plus précises telles que la CFD sont beaucoup plus lentes. Un certain nombre d'initiatives ont été développées afin de réduire le temps de calcul d'un côté et augmenter la précision de l'autre. Par ailleurs, certaines des méthodes ainsi créées sont spécifiques à un site d'étude ce qui limite leur utilisation.

Ce constat permet de construire un cahier des charges du modèle adapté à la problématique de la dispersion atmosphérique accidentelle sur un site industriel, dans un cadre opérationnel. Ce modèle doit permettre un calcul rapide et fiable, de manière à se rapprocher du temps de calcul des modèles gaussiens et de la précision des modèles CFD. Pour pouvoir le comparer aux modèles CFD, celui-ci doit prendre en compte la présence d'obstacles ainsi que la turbulence. Il ne doit pas être spécifique à un site en particulier mais doit au contraire pouvoir être adapté à n'importe quelle configuration.

Dans le cadre de ce travail visant à évaluer le potentiel des réseaux de neurones pour prédire la dispersion atmosphérique, le système a été réduit à deux dimensions (plan horizontal) afin d'alléger les temps de calculs. Le passage 2D vers la 3D sera discuté dans les perspectives de ce document.

La figure ci-après représente le cahier des charges du modèle développé dans ce travail :

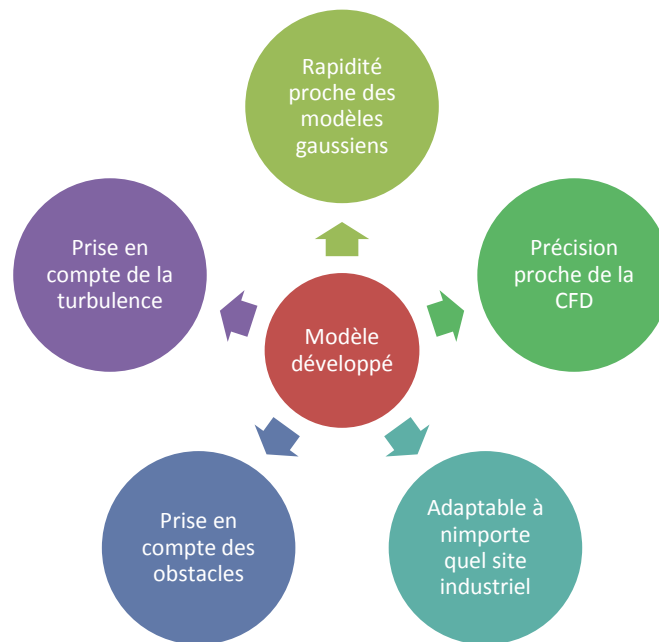


Figure III-1 : Cahier des charges du modèle

Ces travaux proposent d'évaluer les possibilités d'apprentissage de réseaux de neurones pour prédire la dispersion atmosphérique.

L'idée fondamentale repose sur la capacité des réseaux de neurones à pouvoir prédire tout état si le système considéré est correctement décrit. L'état hydrodynamique d'un système étant parfaitement décrit par les équations de Navier Stokes, on pourra donc considérer le réseau de neurone développé comme un émulateur de ces équations. Dans le cas de systèmes dynamiques, le réseau de neurones sera associé aux automates cellulaires. Ce dernier point sera décrit dans la partie correspondante.

La précision des résultats fournis par un réseau de neurones est fonction de la qualité de la base de données utilisée pour son apprentissage. Afin de rendre les prévisions les plus précises possibles, tout en limitant les essais, et donc les coûts engendrés, les bases d'exemples destinées à l'apprentissage des réseaux de neurones seront créées en utilisant le logiciel de CFD ANSYS FLUENT. Les automates cellulaires sont utilisés pour leur faculté à représenter les phénomènes spatiotemporels.

Plusieurs stratégies, démontrant une prise en compte graduelle des difficultés sont établies à partir des contraintes du cahier des charges et des opportunités fournies par les réseaux de neurones et les automates cellulaires. Chaque stratégie est implémentée de manière à respecter les étapes du développement d'un modèle et de son évaluation énoncées par Blocken et Gualtieri (2012) :

- ▶ Définition des objectifs du modèle
- ▶ Spécification du contexte de modélisation
- ▶ Conceptualisation du modèle
- ▶ Sélection des modèles et fonctionnalités
- ▶ Sélection des paramètres à utiliser
- ▶ Sélection des critères de performance
- ▶ Sélection du modèle
- ▶ Analyse de la sensibilité du système aux bruits des données d'entrée
- ▶ Quantification des erreurs
- ▶ Evaluation du modèle et comparaison

On peut distinguer les différentes situations à modéliser en fonction des objectifs. La présence ou non d'obstacles, la prise en compte d'une dispersion continue ou discontinue (de type bouffée) imposent différentes méthodes d'utilisation des outils à disposition.

Le cas du champ libre stationnaire est modélisé par un réseau de neurones seul. La base de données CFD à constituer représente un rejet de méthane en phase gazeuse dans des conditions définies au paragraphe correspondant. Les entrées du réseau de neurones utilisent des indications sur la localisation de la prévision à effectuer ainsi que des informations concernant le terme source et l'écoulement. Les entrées du modèle sont à la fois globales et locales.

L'ajout de la variable temps à ce cas d'étude a des conséquences sur la stratégie de modélisation. Effectivement, l'utilisation d'un réseau de neurones seul impose l'apprentissage de l'évolution de chaque bouffée. Par ailleurs, la présence de deux bouffées qui pourraient se succéder dans le milieu ne peut pas être calculée. Un automate cellulaire est introduit de manière à s'affranchir de cette contrainte. Le réseau de neurones est utilisé en tant que règle de transition. Ce type de modèle est dénommé CA-ANN pour Cellular Automata Artificial Neural Network ruled. Les entrées de ce réseau de neurones sont liées aux conditions locales de l'écoulement et de dispersion. La base d'apprentissage est alors constituée des entrées possibles de cellules, avec l'ensemble des concentrations possibles.

La présence d'un obstacle dans l'écoulement est la situation étudiée ensuite. L'utilisation des réseaux de neurones seuls est là aussi compromise par la multiplicité des configurations possibles (combinaisons des valeurs de vitesse selon x et y et concentration extrêmement nombreuses). Il est clair qu'une telle base n'est pas envisageable car trop importante. Ainsi, il est difficile d'envisager l'utilisation du CA-ANN dans cette situation. Face à cette difficulté, la stratégie déployée ici est de calculer les paramètres déterminants (vitesses selon x et y et coefficient de diffusion turbulente) pour la résolution de l'équation d'advection diffusion par un réseau de neurones. Ainsi, cette étape constituant 99% du temps de calcul des modèles CFD, il est possible d'utiliser différentes méthodes numériques pour résoudre cette équation. L'approche utilisée ici est la méthode des différences finies.

Les différentes méthodes utilisées sont répertoriées dans la figure ci-après :

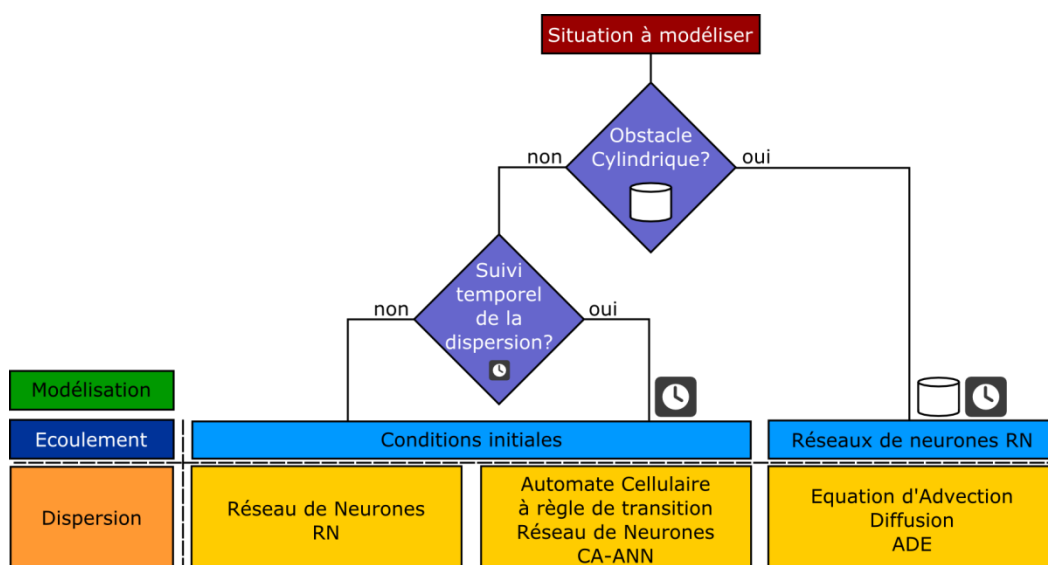


Figure III-2 : Proposition de différentes méthodes de modélisation selon les cas étudiés.

Dans la suite de ce travail, les différentes stratégies seront présentées, ainsi que les résultats obtenus. Pour chacune des méthodes, les simulations CFD ANSYS FLUENT constituent une base d'apprentissage pour le réseau de neurones. Chaque simulation effectuée utilise une modélisation de la couche limite de surface pour les valeurs de turbulence. L'apprentissage de l'ANN est effectué en faisant varier plusieurs paramètres, de manière à déterminer le meilleur modèle possible. Ces éléments sont détaillés dans la suite.

III-1.1. Constitution des bases de données CFD

Toute simulation CFD comporte des imperfections, erreurs d'interpolation ou bruit numérique. L'objet de cette étude n'est pas l'évaluation des différents modèles issus de la CFD. On se propose ici d'utiliser le modèle $k - \epsilon$ standard décrit au I-2.3.3. Celui-ci est largement utilisé et documenté et sa validité pour modéliser la dispersion atmosphérique ne sera pas discutée.

Ce modèle correspond aux équations de Navier-Stokes incompressibles turbulentes. Les équations d'Euler auraient été aussi bien adaptées, sauf en proximité des parois, où une corrélation de frottement est nécessaire, quelle que soit la formulation. Dans un souci d'application des modèles développés à des environnements encombrés de type sites industriels, ce travail se place directement dans le cas d'un fluide réel dans un écoulement turbulent.

L'utilisation de ce modèle pour la création de différentes bases de données à destination des réseaux de neurones doit respecter un certain nombre de règles d'usage :

- ▶ Création de la géométrie
- ▶ Maillage de la géométrie
- ▶ Sélection du modèle
- ▶ Détermination des conditions aux limites
- ▶ Initialisation du domaine
- ▶ Calcul de la configuration par le modèle

Ce travail de thèse s'intéresse uniquement au cas de dispersions bidimensionnelles, selon le plan horizontal. Pour déterminer l'impact sur les personnes, les simulations sont réalisées à une altitude de 2 m. Les géométries utilisées ci-après sont rectangulaires. Le maillage varie d'une approche à l'autre. Néanmoins, les règles d'usage sont utilisées afin de faciliter la convergence du solveur. Le maillage est effectué uniquement par la juxtaposition d'éléments de type quadrilatères. Sa validité est évaluée en utilisant des indicateurs comme l'indice d'asymétrie des éléments (skewness). Celui-ci utilise l'angle minimum et maximum entre deux droites passant par le milieu des côtés opposés de l'élément ainsi que l'angle correspondant pour un élément idéal. Ce critère varie entre 0 et 1 et est défini comme suit :

$$\text{Normalized Equiangular Skewness} = \max\left(\frac{\theta_{\max} - \theta_i}{180 - \theta_i}, \frac{\theta_i - \theta_{\min}}{\theta_i}\right) \quad (III-1)$$

Avec θ_{\max} l'angle maximum, θ_{\min} l'angle minimum et θ_i l'angle pour un élément idéal (90° pour un rectangle et un carré). La figure ci-après illustre la mesure de ces angles :

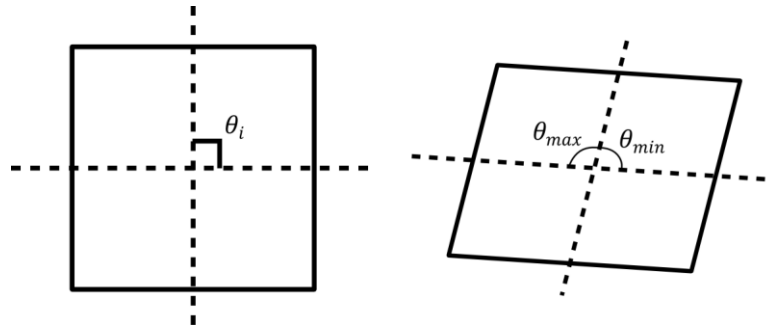


Figure III-3 : Mesure des angles θ_i , θ_{max} et θ_{min}

Un maillage correct comprend des éléments adaptés aux dimensions du domaine et des valeurs d'asymétrie pour chaque cellule inférieures à 0,25, 0 étant la valeur pour un parallélépipède (ANSYS, 2014). Le maillage est créé en respectant ces recommandations. Avant d'utiliser le solveur $k - \epsilon$ standard, il convient de fixer les conditions aux limites.

On s'intéresse à un écoulement en entrée homogène en vitesse et valeurs de turbulence. Les côtés haut et bas représentent le champ lointain dans le sens perpendiculaire à l'écoulement. On utilise classiquement des conditions de symétrie pour représenter cette réalité physique (Gousseau et al., 2011). En sortie du domaine on utilise une condition de pression relative constante égale à 0. En entrée du domaine, on utilisera des conditions de vitesse et turbulence constante, à fixer selon l'objectif de la base de données (figure ci-après).

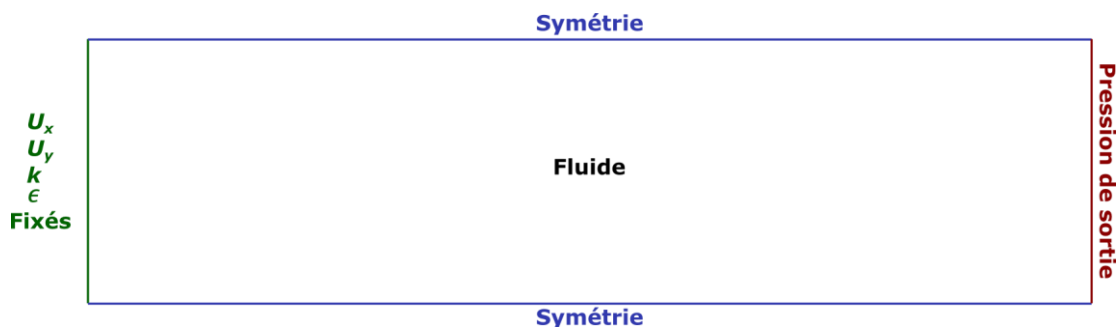


Figure III-4 : Schéma du cas CFD simulé et les conditions aux limites associées

L'initialisation des éléments du domaine est effectuée en utilisant les valeurs de la condition limite d'entrée.

L'objectif est de représenter une couche de surface à une hauteur donnée, en considérant la turbulence comme horizontalement homogène. On décrit dans la suite la méthode classiquement utilisée pour représenter la turbulence dans une couche limite atmosphérique.

III-1.2. Lien avec la couche limite atmosphérique

Dans la suite de ce travail, l'approximation de Boussinesq est utilisée en considérant des fluctuations de la masse volumique faible dans la CLA. L'écoulement en entrée est donc défini comme incompressible à vitesse constante selon x et nulle selon y et z . Un profil de couche limite est établi à travers la détermination des variables liées à la turbulence, fixées à une valeur constante en entrée du domaine. Il est nécessaire de déterminer les valeurs de k et ϵ en entrée du domaine. Les variations horizontales des différentes grandeurs sont négligées. Par analogie avec les lois régissant la diffusion

moléculaire dans les écoulements laminaires, il est possible d'exprimer le flux de quantité de mouvement τ en fonction du gradient vertical de vitesse moyenne :

$$\tau = K_m \rho \frac{\partial \bar{u}}{\partial z} \quad (III-2)$$

\bar{u}	Vitesse moyenne selon x	m.s^{-1}
K_m	Coefficient d'échange turbulent pour la quantité de mouvement	$\text{kg.m}^{-1}.\text{s}^{-2}$
z	Altitude	m

L'extension de ce concept d'échange turbulent dans l'atmosphère est connue comme les K-théories (Kaimal and Finnigan, 1994). L'écoulement étant horizontalement homogène, on admet que \bar{u} varie uniquement selon l'axe vertical. On définit une vitesse de friction u_* pour représenter l'effet des contraintes de cisaillement sur le sol :

$$\tau_0 = \rho u_*^2 \quad (III-3)$$

Cette valeur varie avec la nature du sol rencontré. Dans le cas d'une atmosphère neutre, on obtient une formulation simple du profil de vent en remarquant que K_m a les dimensions d'une distance multipliée par une vitesse. On peut le représenter par le produit des paramètres d'échelle de la couche de surface : z et u_* .

$$K_m = \kappa u_* z \quad (III-4)$$

κ	Constante de Von Karman
----------	-------------------------

En remplaçant maintenant τ et K_m dans l'équation I.19 et en intégrant, on obtient un profil de vent logarithmique en condition de stabilité neutre :

$$\bar{u}(z) = \frac{u_*}{\kappa} \ln\left(\frac{z}{z_0}\right) \quad (III-5)$$

z_0	Longueur de rugosité	m
-------	----------------------	---

La longueur de rugosité, qui est une constante d'intégration, correspond à la hauteur à laquelle s'applique la condition d'adhérence : la vitesse moyenne est nulle (Macdonald et Hall, 1998).

Pour être représentatif de la CLA, un profil de vent logarithmique est utilisé en entrée du domaine de calcul. La modélisation des turbulences observées dans la couche limite atmosphérique, utilise les valeurs proposées par Richards & Hoxey (1993), qui sont largement répandues dans les simulations de la CLA en conditions de stabilité neutre :

$$\frac{U(z)}{u_*} = \frac{1}{\kappa} \ln\left(\frac{z}{z_0}\right) ; k = \frac{u_*^2}{\sqrt{C_\mu}} ; \varepsilon = \frac{u_*^3}{\kappa.(z+z_0)} \quad (III-6)$$

$U(z)$	Vitesse selon x en fonction de l'altitude z	m.s^{-1}
z	Altitude	m
u_*	Vitesse de friction	m.s^{-1}
κ	Constante de Von Karman	-
k	Energie cinétique turbulente	$\text{m}^2.\text{s}^{-2}$
ε	Taux de dissipation de k	$\text{m}^2.\text{s}^{-3}$

C_μ	Constante du modèle k- ϵ	0,009
z_0	Longueur de rugosité	m

Il est ainsi possible d'obtenir les valeurs de k et ϵ , en fonction des valeurs de longueur de rugosité, de vitesse de vent en entrée du domaine et de la vitesse de friction à l'altitude requise ($z = 2$ m). L'objectif est d'obtenir les valeurs de turbulence pour un plan horizontal (figure ci-après) :

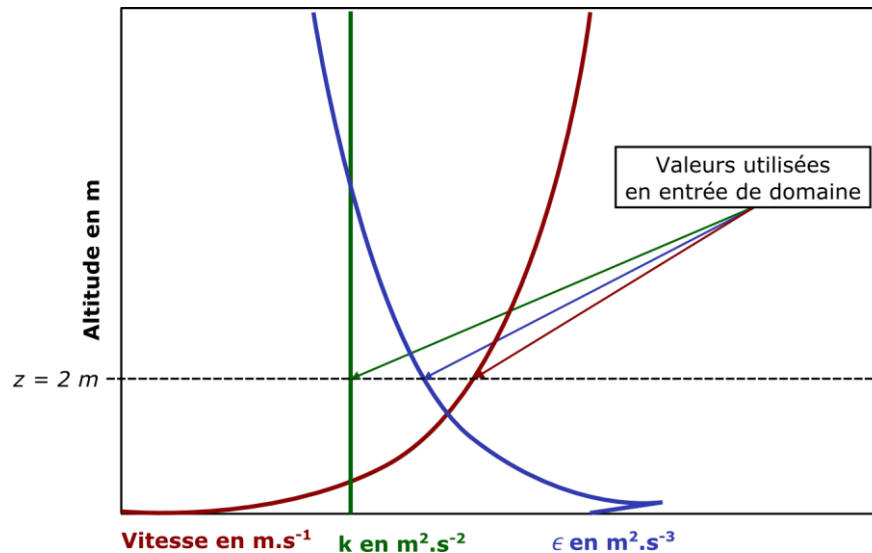


Figure III-5 : Schéma des profils de vitesse, d'énergie cinétique turbulente et de son taux de dissipation utilisés en tant que conditions aux limites en entrée d'un domaine tridimensionnel et plan horizontal associé.

La longueur de rugosité est déterminée en fonction du terrain, à partir de la classification de Davenport (Wieringa, 1993). Dans la suite de ce travail, la valeur de la rugosité est fixée à 0,03, ce qui correspond à un terrain plat avec un revêtement de bitume. Il s'agit de la valeur la plus représentative d'un site industriel en absence d'obstacles dans la description de Davenport :

Tableau III-1 : Longueur de rugosité en fonction du terrain.

Catégorie du terrain	Classe	Surface	Description	Longueur de rugosité z_0 en m
1	1	Mer	Mer, aires asphaltées, plaines couvertes de neige, déserts sans orographie.	0,0002
1	2	Lisse	Plages, champs couverts de neige.	0,005
2	3	Sans obstacles (ouvert)	Prairies herbacées et champs cultivés, aéroports	0,03
3	4	Peu d'obstacles	Champs cultivés avec quelques obstacles	0,1
3	5	Quelques obstacles	Cultures de hauteur variable, vignobles	0,25
3	6	Beaucoup d'obstacles	Fermes avec arbres fruitiers et d'autres arbres ainsi que quelques bâtisses	0,5

Catégorie du terrain	Classe	Surface	Description	Longueur de rugosité z_0 en m
4	7	Fermé	Villages, forêts matures, banlieue	1
4	8	Chaotique	Forêts avec éclaircies, grandes villes (gratte-ciels, maisons d'habitation, parcs)	>2

Dans le cadre de notre méthodologie, différentes valeurs de vitesse d'entrée sont utilisées pour constituer une base de données. Elles permettent de déduire la vitesse de friction :

$$u_* = \frac{U(z) \cdot \kappa}{\ln\left(\frac{z}{z_0}\right)} \quad (III-7)$$

Il est ainsi possible de calculer l'énergie cinétique turbulente et son taux de dissipation en entrée du domaine à une altitude donnée.

Les propriétés physico-chimiques du milieu ambiant sont répertoriées dans le tableau suivant.

Tableau III-2 : Propriétés physico-chimiques du milieu ambiant

Propriété	Valeur	Unité
Température du milieu	20	°C
Viscosité dynamique de l'air	1,8e-5	kg.m ⁻¹ .s ⁻¹
Viscosité dynamique du méthane	1,34e-5	kg.m ⁻¹ .s ⁻¹
Densité de l'air	1,225	-
Densité méthane	0,6709	-

Le méthane a été choisi car c'est un composé très inflammable et très répandu dans l'industrie. C'est un gaz léger qui perturbe l'écoulement dans la réalité ; mais dans notre cas la restriction du système au cas 2D neutralise cet aspect particulier.

Ces principes généraux s'appliquent pour la création des bases de données CFD et sont utilisés pour créer les bases d'apprentissage des réseaux de neurones. La suite de l'étude détaillera la mise en forme de ces données.

III-1.3. Principes et évaluation des réseaux de neurones créés

III-1.3.1 Définitions

Afin de clarifier et faciliter la compréhension de la suite, les définitions de cinq termes sont explicitées telles que considérées dans ce travail.

i. Simulation

Une simulation est définie comme un calcul par le solveur CFD sans changement dans les conditions aux limites ou aux conditions initiales. Les champs de vents sont considérés comme stationnaires. Une simulation peut être statique ou dynamique selon le cas considéré. Dans cette étude, les simulations dynamiques concernent uniquement la dispersion de méthane. On parlera alors de champ de concentration pour caractériser une distribution géographiquement distribuée de concentration de méthane à un instant donné.

ii. Base de données du phénomène

La base de données du phénomène correspond à l'ensemble des simulations effectuées en faisant varier les paramètres influant sur la grandeur cible dans une configuration donnée (stationnaire/instationnaire, champ libre/obstacle).

iii. Echantillonnage

L'échantillonnage de la base de données du phénomène permet la création de l'ensemble d'apprentissage. Chaque maillage comprenant un nombre important de valeurs, souvent identiques, il est impossible et inutile d'effectuer un apprentissage sur la totalité des données. La capacité informatique nécessaire n'est pas suffisante pour effectuer un apprentissage sur un nombre trop important de données. D'autre part, un réseau de neurones est conçu pour permettre la généralisation de la prédiction à des cas non-appris, fournir des données en trop grand nombre sur des exemples semblables peut nuire à cette capacité de généralisation à des cas rares (concentrations élevées). L'échantillonnage doit permettre de diminuer le nombre de données en restant représentatif du phénomène.

La méthode appliquée ici se décompose de la façon suivante : chaque simulation contient une distribution de valeurs de la grandeur à prédire. Un nombre d'intervalles I_c est défini. Un nombre de points E_c est sélectionné dans chaque intervalle. Ainsi, pour chaque simulation, une distribution uniforme de la grandeur considérée est obtenue. L'échantillonnage est ainsi défini par un couple $\{I_c, E_c\}$.

iv. Mise en forme des données

Avant l'apprentissage du réseau de neurones, il est nécessaire de créer les tableaux comportant les variables d'entrée et les sorties cibles. Les variables d'entrée sont définies en fonction de l'application. D'autre part, afin d'éviter qu'une variable avec une distribution très large ait une influence plus importante qu'une autre lors de l'apprentissage, chaque distribution est centrée et normalisée.

v. Base d'exemples

La base d'exemples comprend les données échantillonnées et mises en forme. Elle se décompose en trois sous-ensembles : apprentissage, arrêt et test comme détaillé au paragraphe II-2.2.4.

Le schéma ci-après remet ces définitions dans le processus d'apprentissage :

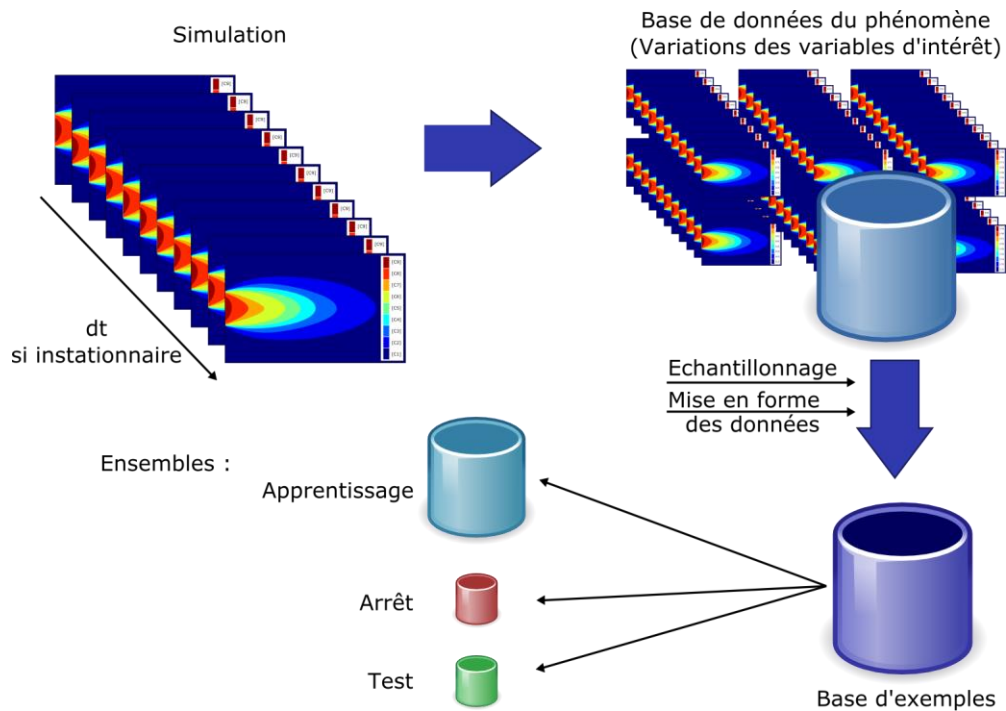


Figure III-6 : Représentation des éléments utilisés pour l'apprentissage

III-1.3.2 Optimisation de l'apprentissage

La qualité d'un apprentissage d'un réseau de neurones est évaluée grâce au calcul de l'Erreur Quadratique Moyenne sur l'ensemble de Test. Un seul apprentissage ne permet pas de garantir la meilleure performance en termes de prévision. Il est donc indispensable d'effectuer une phase d'optimisation et de sélection de l'apprentissage du réseau de neurones. Des variations sont effectuées sur les paramètres influant la qualité de l'apprentissage de manière à identifier leur influence. Ces paramètres sont : la taille de la base d'exemples qui est reliée aux couples $\{I_c, E_c\}$, l'initialisation des paramètres des neurones et l'architecture de ce même réseau (nombre de neurones dans la couche cachée, type de fonction d'activation, nombre de couches cachées). Chaque ensemble regroupant un échantillonnage, une initialisation et une architecture représente une configuration d'apprentissage. Dans cette phase d'optimisation, il convient de déterminer la configuration optimale.

III-1.3.3 Evaluation d'un modèle

L'évaluation d'un apprentissage est effectuée sur l'ensemble de test. Celui-ci comprend des valeurs de chaque simulation de la base de données n'ayant pas servi à l'apprentissage. Cela permet une évaluation sur un échantillon représentatif. D'autre part, il est également important d'évaluer le modèle créé sur des simulations tests et de les comparer aux résultats issus de la CFD. Ainsi, pour chaque méthode développée dans le cadre de ce travail de thèse, la performance des modèles sera évaluée grâce à des indicateurs utilisés spécifiquement en dispersion atmosphérique.

Certains critères de performance initialement répertoriés par Chang et Hanna (2004), seront appliqués pour ces comparaisons. Ces indicateurs sont : le biais fractionnel ou

Fractional Bias (FB), l'erreur quadratique moyenne normalisée ou Normalized Mean Square Error (NMSE) et le facteur deux ou factor of two (FAC2).

La seule valeur d'un indicateur ne permet pas de conclure sur la performance d'un modèle, il est nécessaire de mettre en perspective l'ensemble de ces indicateurs. On distingue ainsi la détermination de l'erreur systématique de l'erreur aléatoire. L'erreur systématique est mesurée par le biais du modèle. Elle conduit à une surestimation ou une sous-estimation systématique des valeurs prédites.

On utilise le calcul de FB afin de mesurer cette erreur systématique :

$$FB = 2 \frac{(\bar{c}_o - \bar{c}_p)}{(\bar{c}_o + \bar{c}_p)} \quad (III-8)$$

\bar{c}_o	Valeur observée
\bar{c}_p	Valeur prédite

Dans notre cas, la valeur observée est la valeur représentant la réalité CFD ou réalité simulée. Un modèle parfait aura une valeur de FB proche de 0 et une valeur de MG proche de 1. Une surestimation implique une valeur de FB négative alors qu'une sous-estimation donne une valeur de FB positive.

L'erreur totale est mesurée à partir de l'écart entre la mesure et la prédiction. Les valeurs se situent autour de la valeur vraie, la moyenne arithmétique tend vers la valeur vraie avec le nombre d'exemples. La valeur des erreurs totales est définie par le calcul de NMSE :

$$NMSE = \frac{(\bar{c}_o - \bar{c}_p)^2}{\bar{c}_o \bar{c}_p} \quad (III-9)$$

Un modèle parfait aura une valeur de NMSE proche de 0. Par analogie avec la caractérisation de la mesure d'un capteur, on définit la fidélité d'un modèle par sa capacité à prévoir des valeurs très proches de la valeur moyenne des prévisions. La justesse est définie par la capacité de ce modèle à prédire une valeur moyenne très proche de la valeur vraie. Lorsqu'un modèle est à la fois juste et fidèle, on dit qu'il est précis. La figure ci-après présente ces différents types d'erreur :

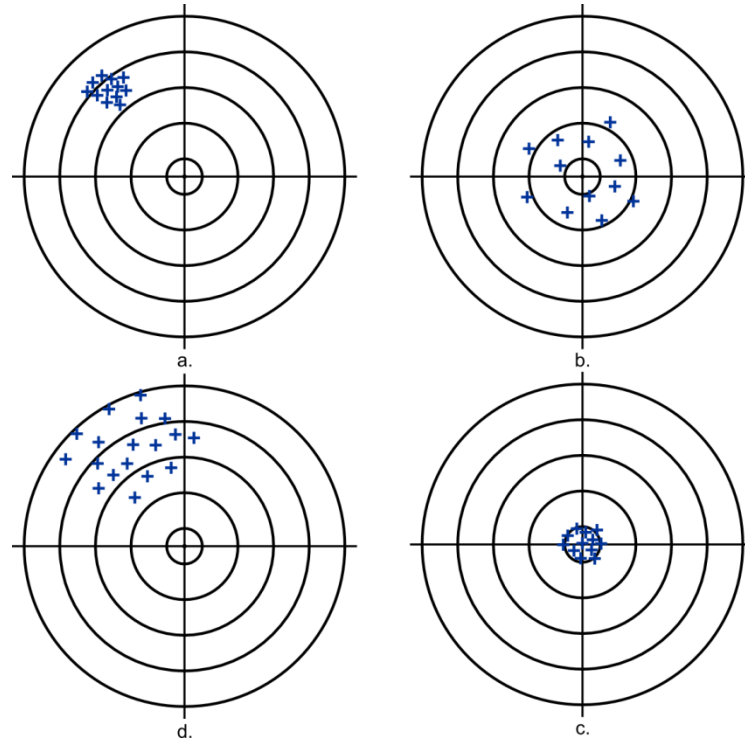


Figure III-7 : Représentation schématique des différents types d'erreurs : a. Modèle fidèle mais non juste (erreurs systématiques) b. Modèle juste mais non fidèle (erreurs aléatoires) c. modèle ni juste ni fidèle (erreurs systématiques et aléatoires) d. modèle juste et fidèle donc précis (peu d'erreurs). Adapté de Asch (2010).

Par ailleurs les valeurs de FB et NMSE peuvent être influencées par des valeurs à prédire importantes et non fréquentes, ce qui est le cas lorsque l'on observe une distribution à allure log-normale, fréquente pour le cas de la dispersion atmosphérique.

En plus de ces indicateurs, le coefficient de détermination R^2 sera utilisé :

$$R^2 = 1 - \frac{\sum_i^N (c_{oi} - c_{pi})^2}{\sum_i^N (c_{oi} - \bar{c}_o)^2} \quad (III-10)$$

Ce critère permet la comparaison des valeurs de simulation à un modèle calculant la moyenne des prévisions. S'il est positif, la prévision est meilleure qu'une prévision de la valeur moyenne. Une valeur négative indique l'inadéquation du modèle utilisé. Un modèle parfait aura une valeur de R^2 proche de 1. Le numérateur du coefficient de détermination correspond à l'erreur quadratique ce qui signifie que lorsque celle-ci tend vers 0, le coefficient de détermination tend vers 1.

Enfin, le FAC2 représente la fraction des prédictions comprises dans un intervalle borné par la moitié et le double de la valeur observée :

$$FAC2 = \text{fraction des données comprises entre } 0,5 \text{ et } 2 : 0,5 \leq \frac{c_p}{c_o} \leq 2 \quad (III-11)$$

Il s'agit du critère le plus robuste puisqu'il n'est pas influencé par l'étendue des valeurs. Ainsi, une multitude de critères de performance est nécessaire à l'évaluation d'un modèle. Dans l'évaluation des méthodes décrites ci-après, ces indicateurs seront utilisés en fonction des données prédites et mesurées. Dans le cas des simulations statiques, ces indicateurs seront comparés dans l'espace et pour le cas des simulations dynamiques, ils seront comparés dans le temps et l'espace.

III-2. Modélisation de la dispersion atmosphérique d'une source continue en champ libre

Ce chapitre présente le cas simplifié de la modélisation de la dispersion atmosphérique en régime continu dans un écoulement stationnaire, en champ libre. Ce travail a fait l'objet d'un article publié dans Chemical Engineering Transactions (Lauret et al., 2013) disponible en annexe.

III-2.1. Simulation CFD

Dans ce travail préliminaire à une modélisation complète de la dispersion atmosphérique, une situation simplifiée est étudiée. Une source continue de méthane en phase gazeuse est considérée, caractérisée par son débit massique. Le méthane est alors mélangé à l'écoulement turbulent, caractéristique d'une couche limite atmosphérique horizontalement homogène. Les valeurs de turbulence reportées en entrée de l'écoulement sont donc celles calculées conformément aux recommandations du paragraphe III-1.2.

Le domaine de calcul mesure 40 mètres de long pour 20 mètres de large. On utilise la symétrie du système afin de diminuer le temps de calcul. Le maillage est constitué de 401 301 quadrilatères. Celui-ci présente des mailles aux dimensions plus faibles à proximité du rejet, de manière à faciliter le calcul aux endroits où les gradients des différentes grandeurs sont importants comme illustré sur la figure suivante :

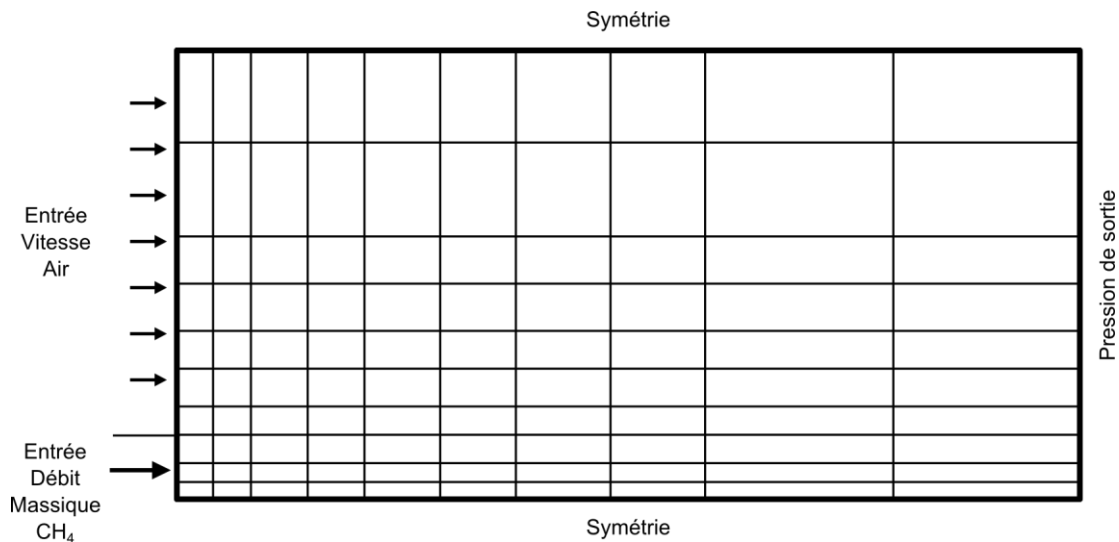


Figure III-8 : Définition du maillage et des conditions aux limites de la simulation CFD

Les mailles de calcul sont initialisées à partir des valeurs d'entrée de vitesse de vent. Chaque simulation est ensuite calculée par l'utilisation du solveur ANSYS FLUENT en utilisant des schémas de discrétisation du second ordre pour les équations de conservation de la matière, de la quantité de mouvement, l'équation de transport du méthane, et du premier ordre pour les équations de transport d'énergie cinétique turbulente, son taux de dissipation. Le calcul est itéré jusqu'à atteindre les critères de

convergence sur les résidus des équations. Celui-ci est fixé à 10^{-6} . Les paramètres de simulation étant fixés, il convient de créer la base de données du phénomène.

III-2.2. Base de données et base d'exemples

La base de données du phénomène est obtenue en faisant varier le débit massique entre $0,5$ et 5 kg.s^{-1} et la vitesse de l'écoulement atmosphérique entre 2 et 10 m.s^{-1} . Ces paramètres influent directement sur les concentrations en aval du rejet. Par conséquent, ils seront utilisés en tant que variable d'entrée du réseau de neurones. On dispose donc de grandeurs associées au terme source (débit massique) et aux conditions météorologiques (vitesse du vent). La sortie d'un réseau de neurones choisi est unique et non localisée a priori. Il est donc nécessaire de fournir au réseau de neurones des informations sur la localisation de la concentration à prédire. Dans ce travail, les coordonnées polaires donnent une information reliée directement au phénomène et seront donc utilisées. Le tableau ci-après résume la constitution de la base de données avant échantillonnage.

Tableau III-3 : Variables d'entrée du réseau de neurones

	Terme source	Conditions atmosphériques	Localisation	
Grandeur	Débit Massique CH_4	Vitesse de l'écoulement d'air	Coordonnées polaires	
			θ	ρ
Unité	kg.s^{-1}	m.s^{-1}	$^\circ$	m
Intervalle	[0,5 ; 5]	[2 ; 10]	[0 ; 90]	[0 ; 44,7]
Pas	0,5	1	Fonction du maillage	

La sortie du réseau de neurones est donc la concentration au point déterminé par les coordonnées polaires. La base de données contient 90 simulations avec des valeurs de concentration pour chaque nœud soit 401 301 valeurs par simulation. Cela représente donc plus de 36 millions d'exemples potentiels pour le réseau de neurones.

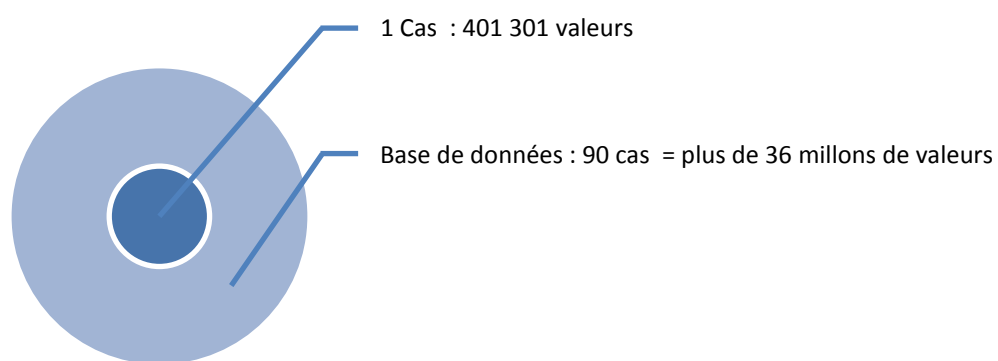


Figure III-9 : Représentation du contenu de la base de données CFD

Effectuer l'apprentissage du réseau de neurones sur l'ensemble des données n'est pas réaliste, notamment du point de vue des capacités informatiques. Il est nécessaire de la réduire (figure ci-après) :

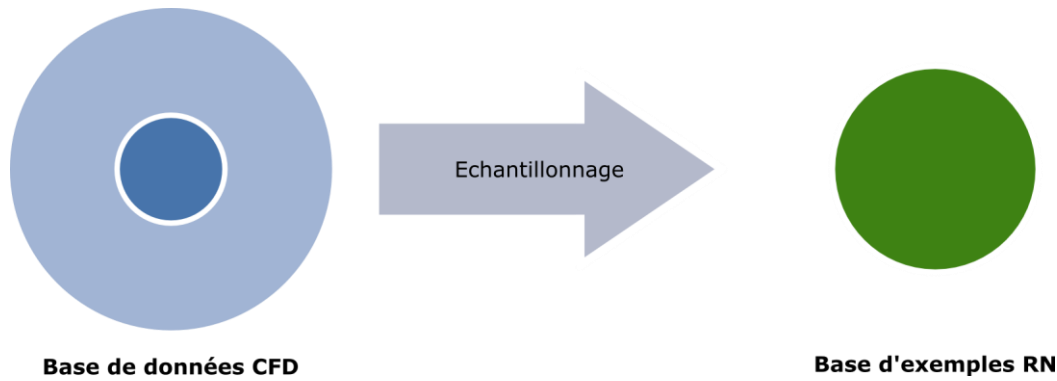


Figure III-10 : Représentation de l'échantillonnage de la base de données CFD pour la création de la base de données ANN.

L'échantillonnage est effectué en prélevant E_c exemples dans I_c intervalles de concentration pour chaque simulation.

III-2.3. Apprentissage et optimisation

Un perceptron à deux couches est utilisé. La couche cachée comporte des neurones avec une fonction d'activation tangente hyperbolique. La couche de sortie a une fonction d'activation linéaire ($y=x$). Le réseau de neurones peut être schématiquement représenté comme sur la figure suivante :

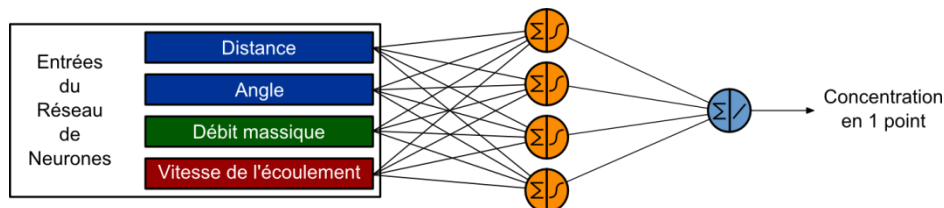


Figure III-11 : Entrées et sortie du réseau de neurones.

La recherche du réseau de neurones optimal se fait en étudiant l'influence des différents paramètres de l'apprentissage : architecture du réseau, échantillonnage et initialisation. On mesure la qualité du modèle en généralisation en utilisant l'Erreur Quadratique Moyenne de Test dont on rappelle l'expression :

$$EQMT = \sqrt{\frac{1}{N_T} \sum_{k=1}^{N_T} (y_k^p - g(x_k, w))^2} \quad (III-12)$$

Celle-ci est calculée sur un ensemble disjoint de l'ensemble d'apprentissage et de l'ensemble d'arrêt. Cet ensemble de test est représentatif de l'ensemble de la base de données, les points sélectionnés étant issus de chacun des cas de la base de données CFD. Il comprend 19 258 points répartis aléatoirement selon les variables d'entrée du réseau de neurones. C'est la valeur de l'EQMT qui permet de déterminer quel apprentissage est le plus à même de généraliser de manière précise les prédictions sur une configuration non-apprise.

III-2.3.1 Influence de l'initialisation

Avant d'évaluer l'influence de l'échantillonnage et du nombre de neurones cachés, il est important de quantifier le rôle de l'initialisation dans l'apprentissage. Pour cela, 20 apprentissages sont effectués à partir d'un échantillonnage issu d'un couple $\{I_c; E_c\}$: $\{30; 30\}$. Cela représente 40 300 exemples dans l'ensemble d'apprentissage. 20 neurones sont utilisés dans la couche cachée. La figure ci-après représente l'EQMT pour chacun de ces apprentissages.

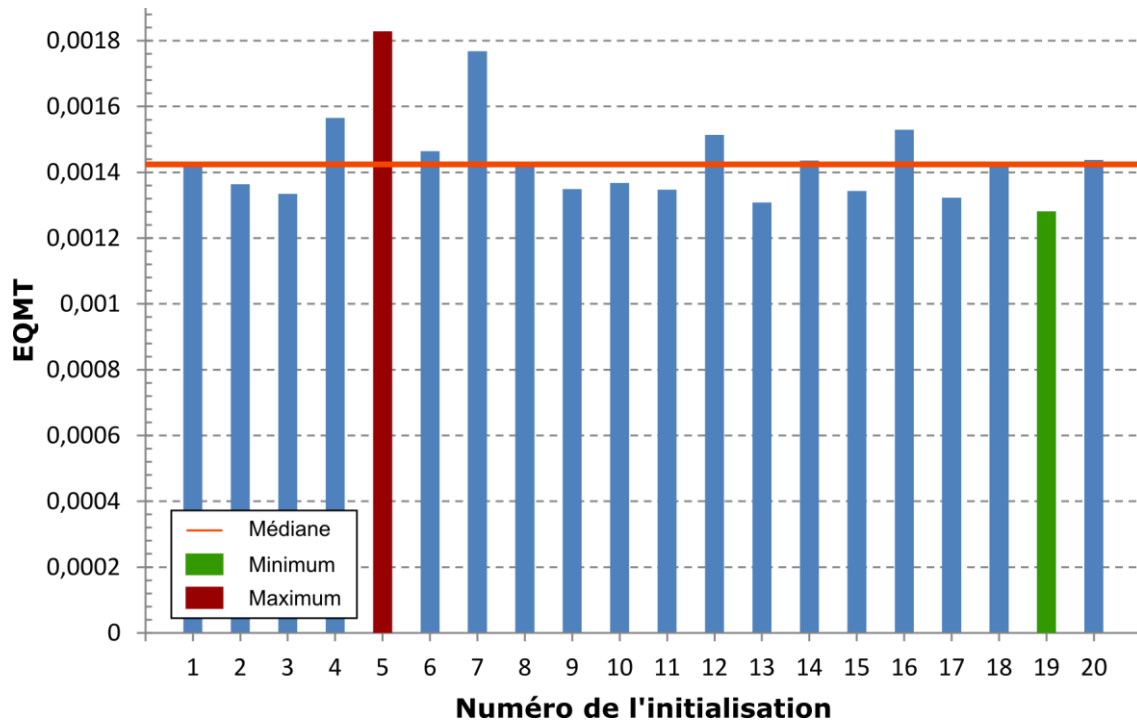


Figure III-12 : Influence de l'initialisation sur l'Erreur Quadratique Moyenne sur l'ensemble de Test

Sur ces 20 apprentissages, excepté les initialisations 5 et 7, l'écart à la médiane est faible. Certaines initialisations mènent à un meilleur modèle. Des initialisations différentes peuvent conduire à trouver, dans l'espace des paramètres, des minima différents (Oussar, 1998). Il est cependant impossible de connaître au préalable si la valeur d'un vecteur des paramètres va permettre d'atteindre un minimum plus faible que les autres. Afin de sélectionner le meilleur modèle une fois les autres paramètres fixés (échantillonnage et nombre de neurones dans la couche cachée), il est important d'effectuer un nombre suffisant d'initialisations pour s'assurer de pouvoir sélectionner un des meilleurs modèles.

III-2.3.2 Influence du nombre de neurones dans la couche cachée

De la même façon que précédemment, on évalue l'influence du nombre de neurones dans la couche cachée sur la généralisation en effectuant 20 apprentissages pour chaque valeur, avec une initialisation différente. Afin d'éviter de biaiser la comparaison, les apprentissages sont effectués sur un seul échantillonnage, considéré comme étant un bon compromis entre durée d'apprentissage et représentativité du phénomène. On prend un échantillonnage de couple $\{20; 20\}$, contenant 15 440 exemples dans la base d'apprentissage.

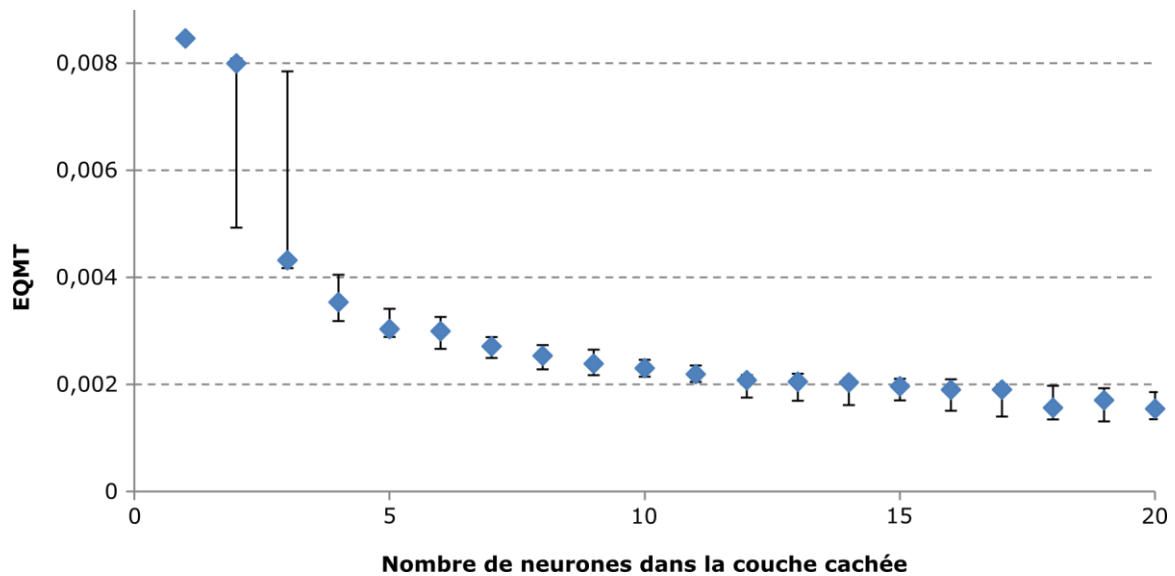


Figure III-13 : Influence du nombre de neurones dans la couche cachée sur la qualité de la généralisation (Valeurs médianes pour 20 initialisations – barres d'erreurs : minimum et maximum des 20 initialisations)

Sur la figure précédente, une diminution nette de l'EQMT avec le nombre de neurones dans la couche cachée est identifiée. La dispersion des extremums autour des valeurs médianes obtenues pour les apprentissages concernant un nombre de neurones dans la couche cachée de 19 et 20 montre qu'il est impératif de tester plusieurs initialisations avant de sélectionner un modèle. Les valeurs de la médiane permettent d'obtenir l'influence du nombre de neurones dans la couche cachée alors que le minimum permet de sélectionner le meilleur modèle. Ainsi, la valeur médiane pour 20 neurones dans la couche cachée est inférieure à celle pour 19 neurones. Cependant, le minimum de l'ensemble de ces apprentissages est obtenu avec 19 neurones dans la couche cachée. L'augmentation du nombre de neurones dans la couche cachée améliore la qualité du modèle. Du fait du dilemme biais-variance, ce comportement est spécifique d'un modèle dont les données ne comportent pas de bruit. Ceci montre que le bruit numérique dans la base de données CFD est faible et n'a que peu d'effets dans la capacité de généralisation du réseau de neurones. Il serait possible de rajouter des neurones dans la couche cachée de manière importante mais cela n'aurait que peu d'impact sur le résultat final, la courbe s'approchant d'une asymptote. Par contre, cela augmenterait considérablement le temps de calcul de l'apprentissage et des différents essais que nous devons réaliser pour ajuster au mieux l'architecture du modèle.

III-2.3.3 Influence de l'échantillonnage

Plusieurs échantillonnages sont effectués en faisant varier les valeurs des couples $\{I_c, E_c\}$ entre 10 et 30, par pas de cinq. Pour chacun de ces échantillonnages, 20 apprentissages sont effectués avec différents vecteurs d'initialisation des paramètres. La valeur médiane de ces apprentissages est alors comparée afin d'observer l'influence de la dimension de la base d'exemples, indépendamment des initialisations. Ces apprentissages sont effectués pour un même nombre de neurones dans la couche cachée : 20.

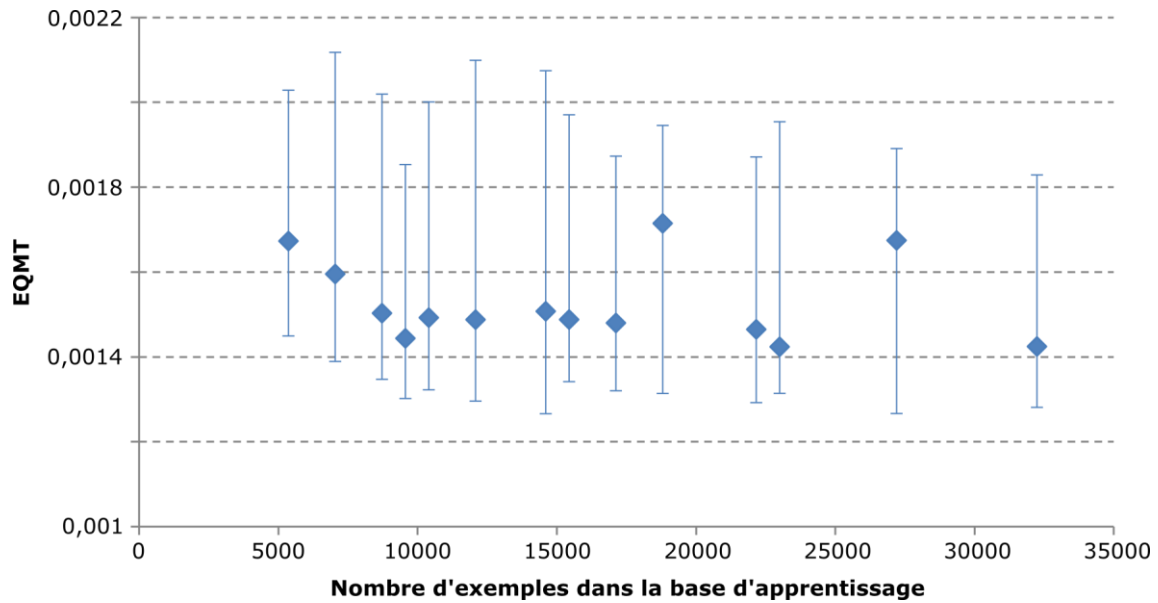


Figure III-14 : Influence du nombre d'exemples sur la qualité du modèle. (Valeurs médianes pour 20 initialisations – barres d'erreurs : minimum et maximum des 20 initialisations)

Une faible diminution, peu notable et difficile à évaluer, de l'EQMT avec l'augmentation du nombre d'exemples dans la base d'apprentissage est observée. Les barres d'erreurs représentent le minimum et le maximum obtenus pour différentes initialisations. Ainsi, la comparaison entre valeurs minimales montre une influence faible de la taille de la base d'apprentissage :

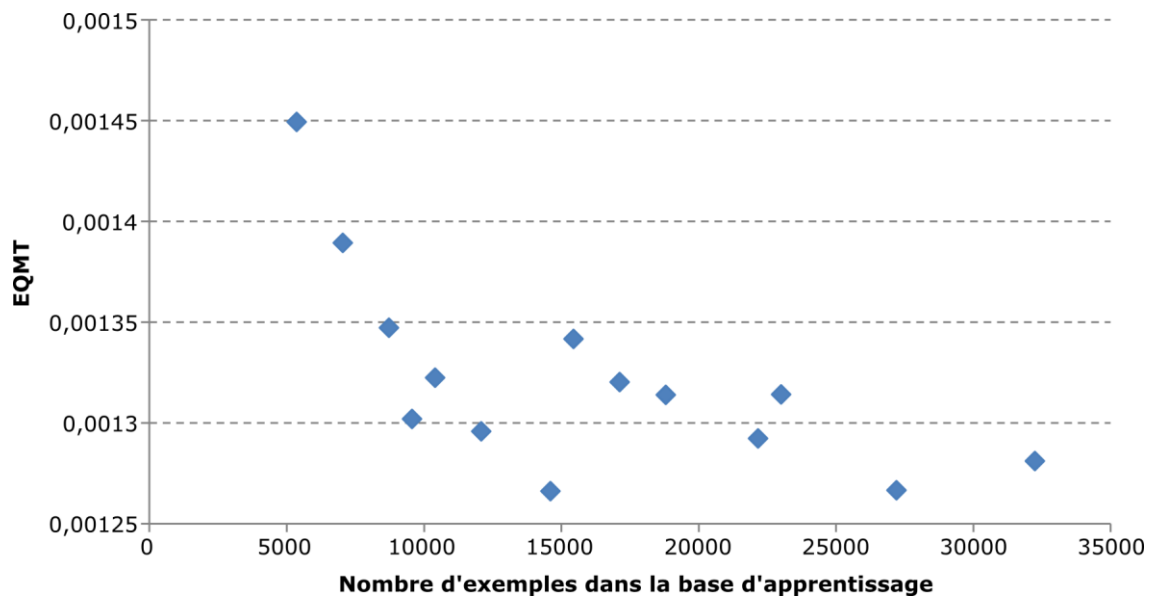


Figure III-15 : Influence du nombre d'exemples sur la qualité du modèle. Minimum de 20 apprentissages.

A partir d'un nombre d'exemples supérieur à 10 000, les valeurs de l'EQMT varient peu. Néanmoins, le minimum de l'erreur quadratique moyenne est obtenu pour un échantillonnage à 14 600 exemples dans la base d'apprentissage.

III-2.3.4 Sélection du modèle

Les paragraphes précédents ont permis d'évaluer quantitativement et qualitativement l'influence de chaque paramètre. Il est désormais possible de définir l'apprentissage souhaité :

- ▶ La base d'exemples doit être suffisamment importante et variée pour être représentative du phénomène dans tout l'espace des états prévu pour son fonctionnement. Une base d'exemples constituée de 14 600 points est sélectionnée, conformément à ce qu'indique la figure III.10.
- ▶ Le nombre de neurones dans la couche cachée doit être compris entre 15 et 20 afin de pouvoir rendre compte de la complexité du phénomène.
- ▶ Il est impératif de tester plusieurs initialisations afin de s'assurer d'obtenir le meilleur modèle.

Finalement, les contraintes ci-dessus ont conduit à sélectionner un réseau de neurones avec les caractéristiques suivantes :

Tableau III-4 : Caractéristiques du modèle sélectionné et de sa conception.

Caractéristiques	Valeur
Nombre d'intervalles de concentrations	25
Nombre d'exemples par intervalle	15
Nombre de neurones dans la couche cachée	20
Numéro de l'initialisation	19
Valeur de l'EQMT en fin d'apprentissage	$1,27.10^{-3}$
Valeur du coefficient de détermination	0,989
Durée d'apprentissage	675 s

III-2.4. Evaluation du modèle

La qualité de ce modèle à réseau de neurones est évaluée sur un ensemble de test, choisi de manière à être représentatif des comportements représentés par la base de données CFD. Afin d'évaluer le modèle ANN dans des conditions variées d'utilisation, celui-ci est évalué sur neuf cas tests. Ceux-ci sont composés de trois débits massiques : $\{1 \text{ kg.s}^{-1} ; 3 \text{ kg.s}^{-1} ; 5 \text{ kg.s}^{-1}\}$ et trois valeurs de vitesse de l'écoulement : $\{2 \text{ m.s}^{-1} ; 4 \text{ m.s}^{-1} ; 6 \text{ m.s}^{-1}\}$.

Pour éviter de prendre des valeurs de concentrations de méthane non-significatives, les valeurs CFD pouvant être extrêmement faibles et donc non détectables, le calcul des indicateurs statistiques est limité aux valeurs supérieures à 100 ppm, la limite inférieure d'inflammabilité du méthane étant égale à 44 000ppm. Cela correspond à une limite de quantification par les appareils de mesures pouvant être utilisés *in situ* (Milliez and Carissimo, 2006).

Le tableau ci-après présente les résultats concernant le facteur deux (FAC2) et le coefficient de détermination (R^2) entre la modélisation par réseau de neurones et la simulation CFD :

Tableau III-5 : Critères de performance - Comparaison ANN/CFD

Cas	Débit (kg.s ⁻¹)	Vitesse (m.s ⁻¹)	R ²	FAC2
1	1	2	0,99	0,80
2	1	4	0,99	0,93
3	1	6	0,99	0,89
4	3	2	0,99	0,78
5	3	4	0,99	0,96
6	3	6	0,99	0,90
7	5	2	0,99	0,84
8	5	4	0,99	0,96
9	5	6	0,99	0,96

Les résultats de ces critères montrent une bonne performance dans la prévision des cas tests. Le coefficient de détermination est proche de sa valeur maximale quel que soit le cas test envisagé. En ce qui concerne le FAC2, les valeurs sont supérieures à 0,8, excepté pour le cas test 4. Le modèle semble avoir plus de difficultés à prévoir les concentrations dans le cas de vitesses faibles (n° 1, 4 et 7) comme l'illustre la figure ci-après :

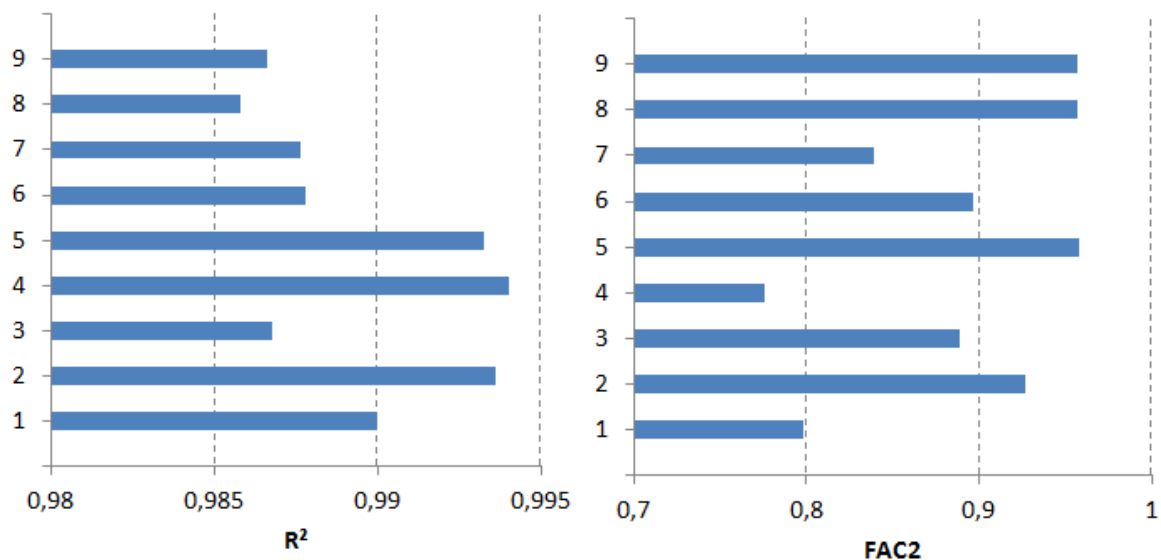


Figure III-16 : Valeurs du R² et du FAC2 pour les neuf cas tests – Comparaison entre les résultats de la simulation CFD et la modélisation ANN.

Les résultats du FAC2 pour les vitesses faibles sont entre 11% et 15 % moins performants que les vitesses moyennes et entre 14 et 23% en ce qui concerne les vitesses importantes. Le débit ne semble pas influencer de manière significative le résultat de la prévision.

Afin de juger au mieux la qualité du modèle, on s'intéresse aux valeurs du biais fractionnel (FB) et à la valeur de l'erreur quadratique moyenne normalisée (NMSE) obtenue. On rappelle que le FB traduit les erreurs systématiques et le NMSE l'erreur systématique et l'erreur aléatoire. Il est possible de tracer un graphe combinant ces deux informations, permettant une analyse complète de la performance du modèle.

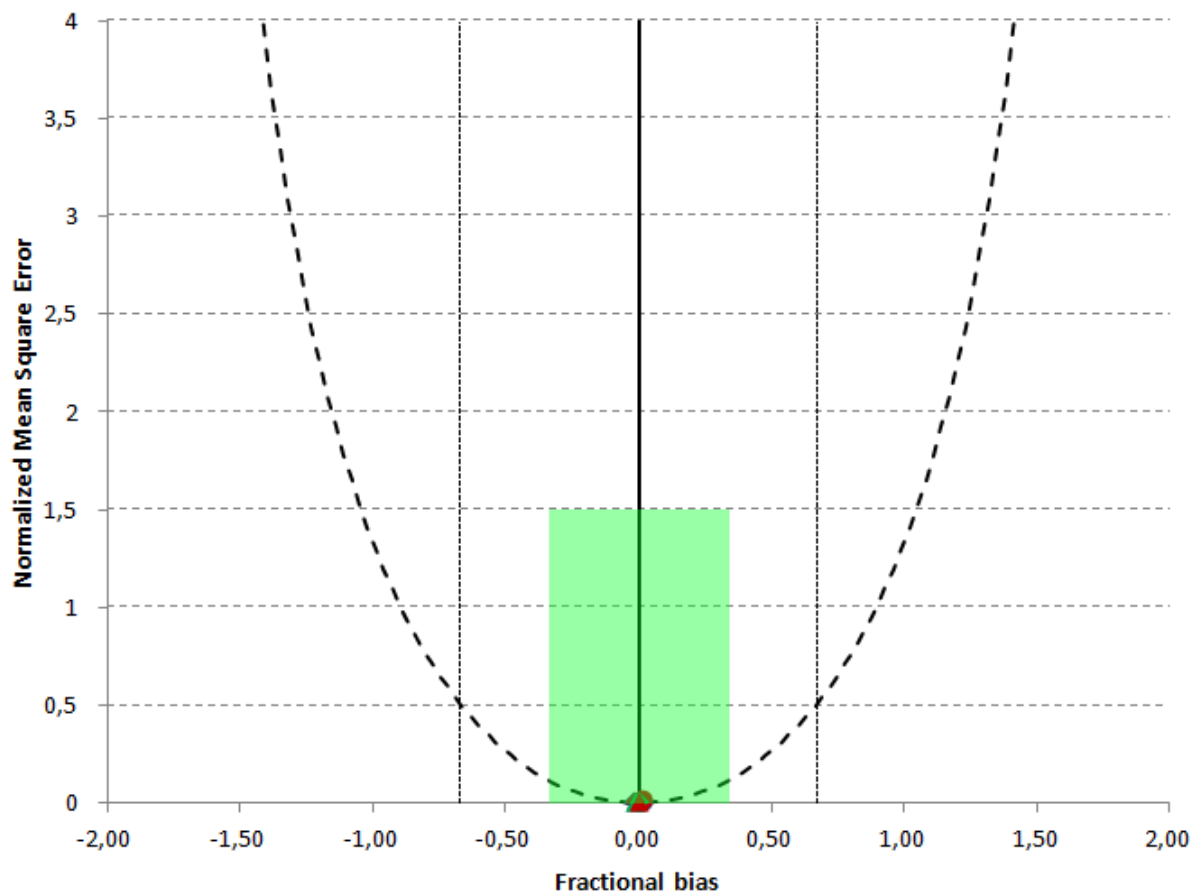


Figure III-17 : Performance du modèle ANN sur les neuf cas tests

La parabole représente la valeur minimale de l'erreur quadratique moyenne normalisée en fonction de la valeur du biais fractionnel. Son équation est :

$$NMSE_{minimum} = \frac{4FB^2}{4-FB^2} \quad (III-13)$$

Cela traduit la part de l'erreur systématique présente dans le NMSE. La zone colorée correspond aux recommandations de Chang and Hanna (2004). On remarque que l'on se situe au cœur de cette zone. Les modélisations ANN sont précises. Il est nécessaire de zoomer afin de qualifier les tendances du modèle selon les cas tests modélisés (figure ci-après). Le modèle ne semble pas surestimer ou sous estimer un cas en particulier. Les cas tests au débit important et à vitesse moyenne semblent être légèrement mieux modélisés.

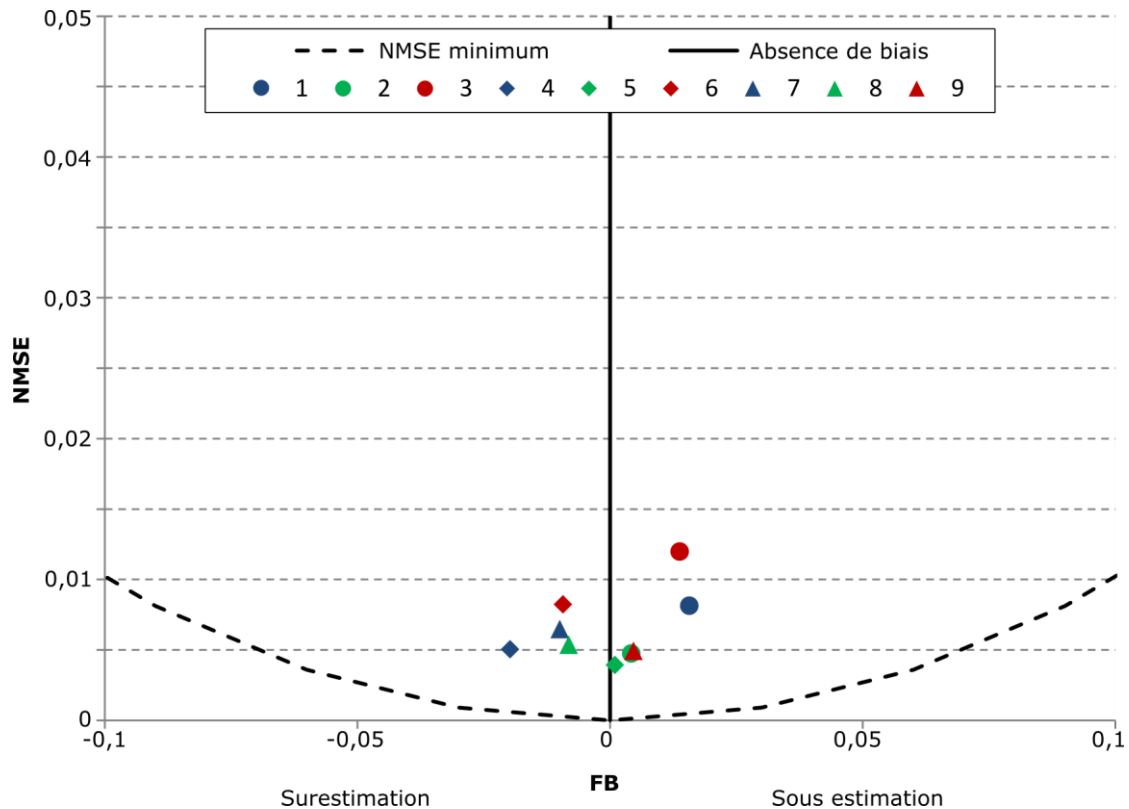


Figure III-18 : Erreur systématique et aléatoire pour l'ensemble des cas tests

Afin d'identifier les points forts et faibles de la modélisation, les résultats sont comparés visuellement sur deux cas tests. Il est ainsi possible de visualiser les isocontours de concentrations modélisées simultanément aux concentrations représentant la réalité CFD (figure ci-après). La meilleure prédiction (cas test 5) et la moins bonne (cas test 3) sont sélectionnées.

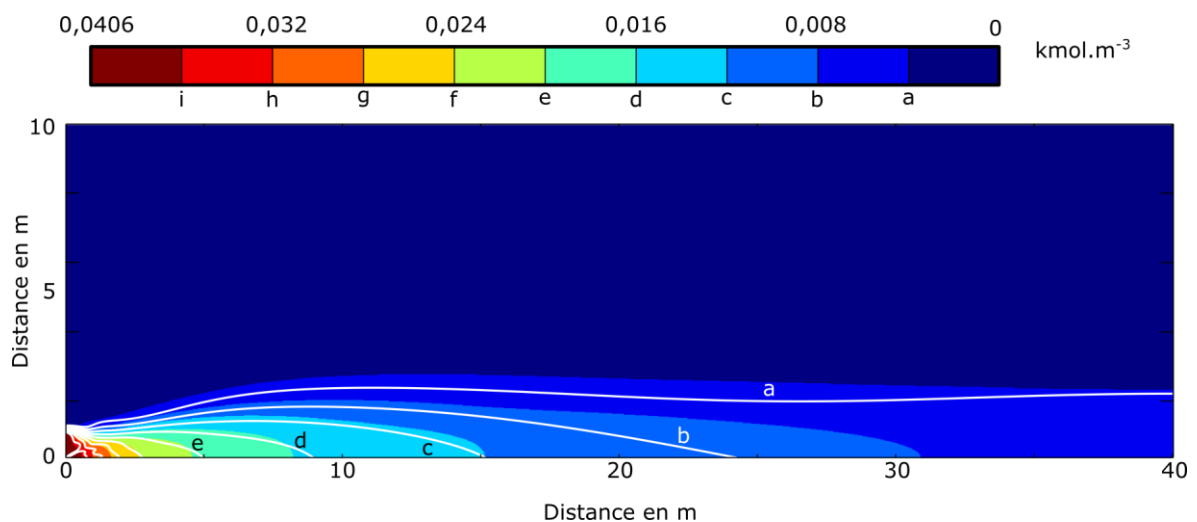


Figure III-19 : Cas 3 (moins bonne prédiction)- iso-surfaces CFD et iso-contours ANN.

Même si ce cas test est le moins bien estimé par le modèle ANN, la tendance générale montre une bonne adéquation entre la prévision par réseau de neurones et la réalité CFD. Les isocontours sont presque superposés aux zones représentant les intervalles de concentration CFD, même si des erreurs apparaissent au niveau de l'axe de symétrie notamment. Il est nécessaire d'évaluer la zone proche de la source pour évaluer l'erreur

au niveau des hauts gradients (figure ci-après). Les isocontours et les isosurfaces sont très proches. Un léger décalage existe pour ce qui est du cas test 3. Le champ lointain est bien représenté. Il est nécessaire de zoomer proche de la source pour identifier des erreurs plus importantes.

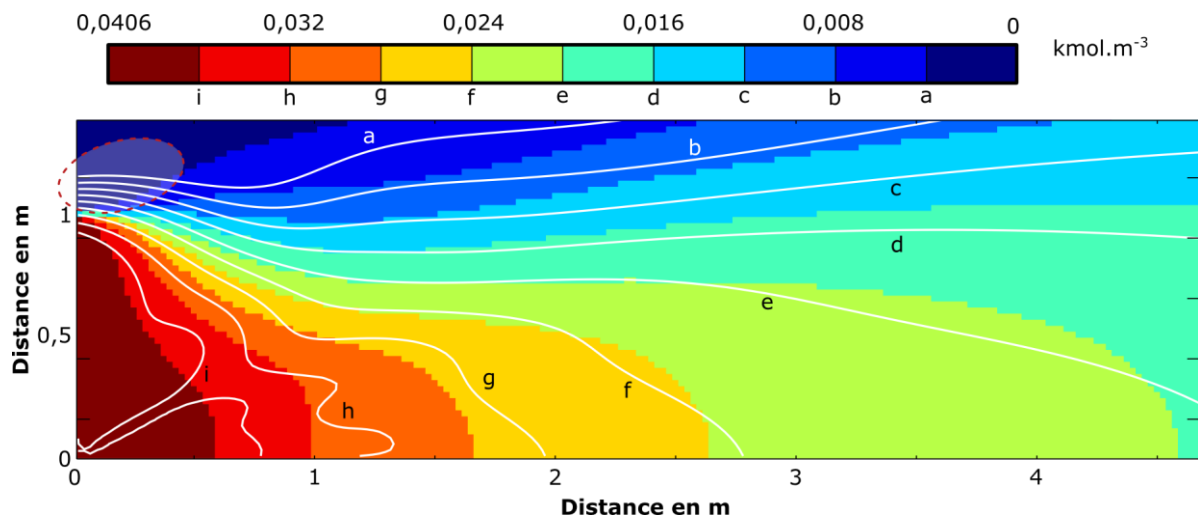


Figure III-20 : Cas 3 – Agrandissement de la zone à plus hauts gradients de concentration

L'ANN prédit correctement les concentrations. Dans la solution représentant la réalité CFD, des concentrations nulles sont obtenues au-delà d'une distance supérieure à 1 mètre selon y pour une distance nulle selon x (zone en pointillés rouges). Cela est normal puisque la source est simulée pour une dimension d'un mètre. L'ANN modélise néanmoins des valeurs de concentrations non-nulles dans cette zone au-delà de 1 mètre.

Si l'on regarde maintenant les résultats pour le cas test le mieux prédit, on s'aperçoit que la prédiction de l'ANN est satisfaisante, notamment en ce qui concerne les distances importantes (figure ci-après). Pour avoir plus de détails sur le comportement au niveau des hauts gradients, il est nécessaire d'effectuer un agrandissement.

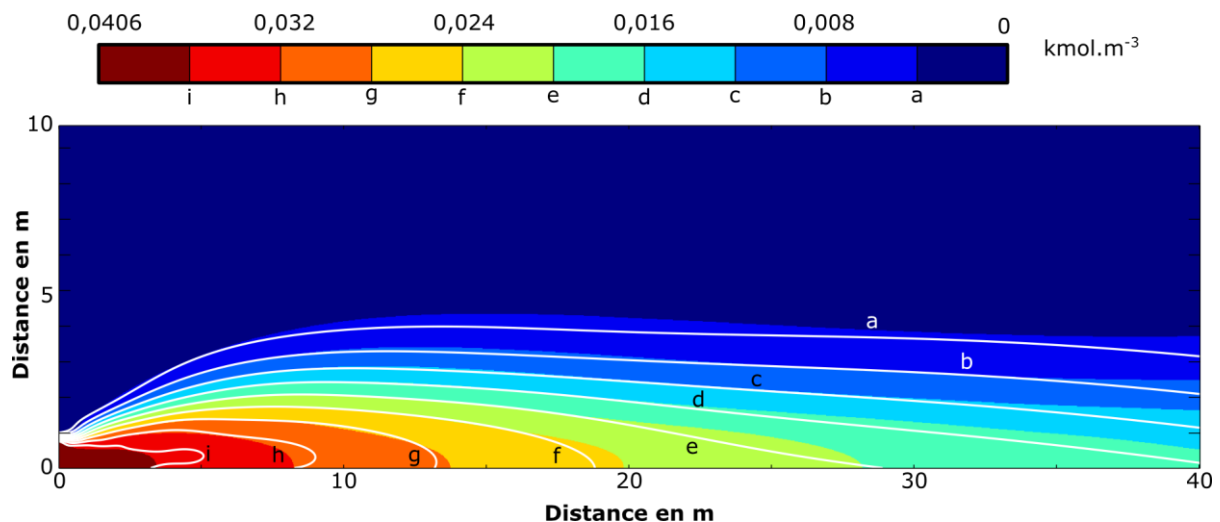


Figure III-21 : cas test 5 - iso-surfaces CFD et iso-contours ANN.

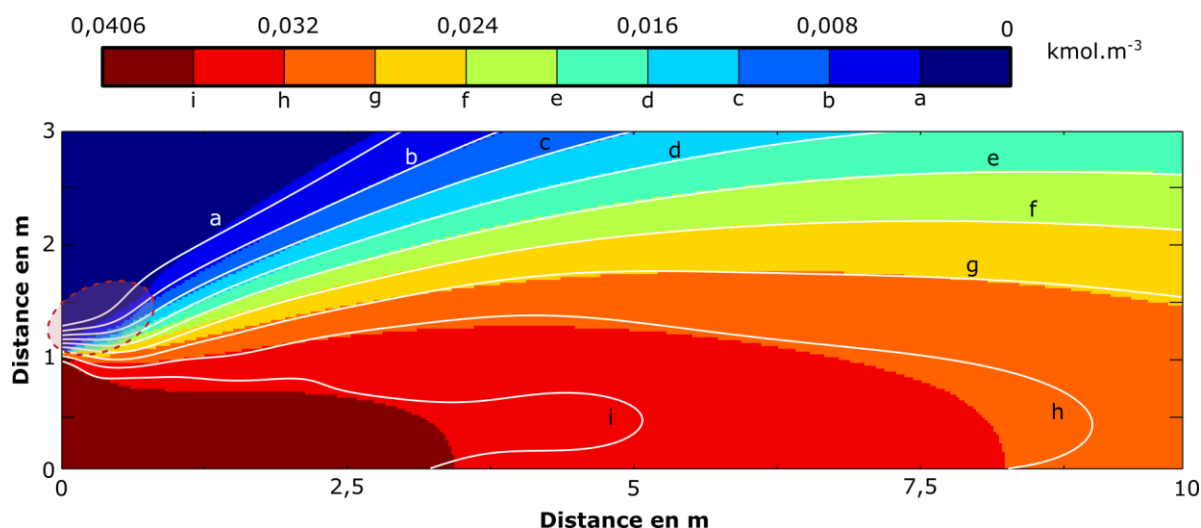


Figure III-22 : Cas 5 – Agrandissement de la zone à plus haut gradients de concentration

La zone où l'ANN prédit des concentrations au-delà de la dimension de la source est également identifiée (en pointillés rouges sur la figure). Les isocontours h et i surestiment légèrement leur zone respective alors que les autres isocontours sont confondus avec les délimitations issues de la CFD. Il est nécessaire d'utiliser des moyens de visualisations des résultats supplémentaires.

Sur la figure ci-après, le diagramme de dispersion (à gauche) affiche les points observés et prédits, appairés en espace. Les droites rouges indiquent une surestimation ou une sous-estimation d'un facteur deux. La droite en vert est la première bissectrice. La figure de droite est un diagramme quantile-quantile : les points sont appairés selon le rang de la valeur de concentration. Les zones identifiées précédemment sont également relevées ici. Ainsi, la surestimation des valeurs faibles observées correspond à la zone proche de la source, mais dans l'écoulement où aucune valeur de concentration ne devrait être observée. Par ailleurs, la majorité des points se situent à proximité de la première bissectrice. Les valeurs fortes de concentrations ont une modélisation caractérisée par des disparités : certains points sont surestimés alors que d'autres sont sous-estimés. Enfin, des concentrations faibles (inférieures à 10 % du maximum relevé) sont modélisées avec une valeur nulle par le modèle.

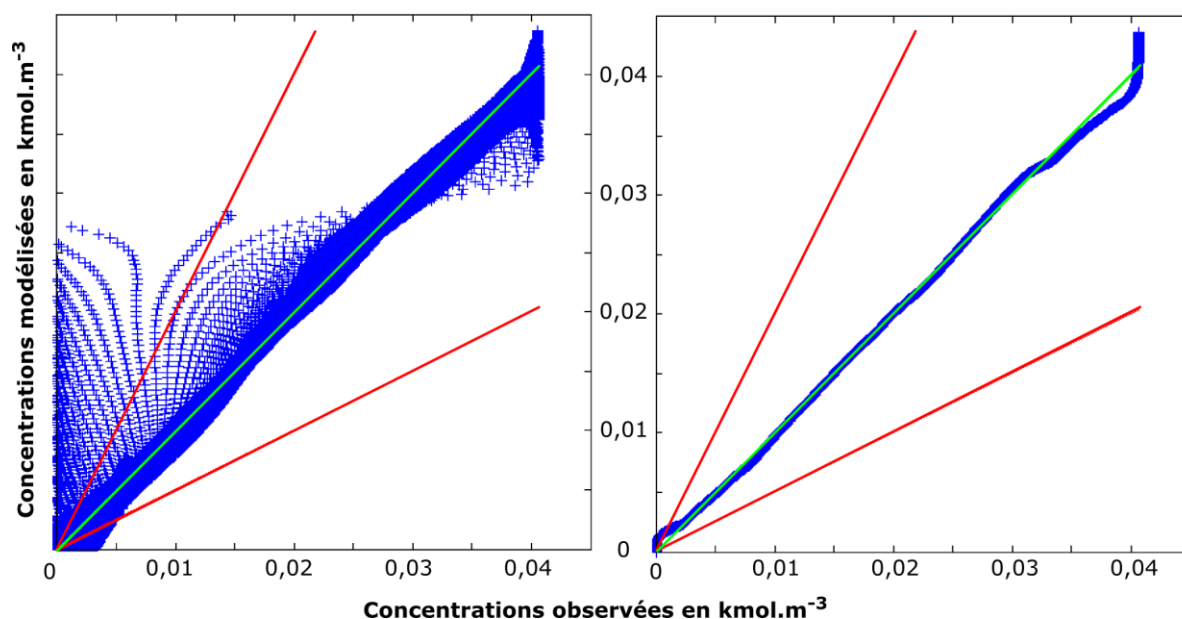


Figure III-23 : Diagramme de dispersion des valeurs observées et modélisées pour le cas défavorable 3 (appairées dans l'espace sur la figure de gauche et par rang dans la figure de droite).

L'observation de la dispersion des points (figure ci-après) montre de bons résultats pour une grande majorité de points. On note une légère surestimation de la concentration maximale, d'approximativement 7%. Cette surestimation est certainement liée à la zone observée précédemment. Cependant, les valeurs fortes ont également des points modélisés par des valeurs sous-estimées. Par ailleurs, les surestimations observées dans les faibles et moyennes concentrations sont l'illustration du décalage par rapport à la source de la même manière que pour le cas défavorable. Enfin, le modèle donne de meilleurs résultats pour les valeurs faibles, la zone de valeurs nulles du modèle vue dans le graphique précédent n'est pas observée sur ce cas.

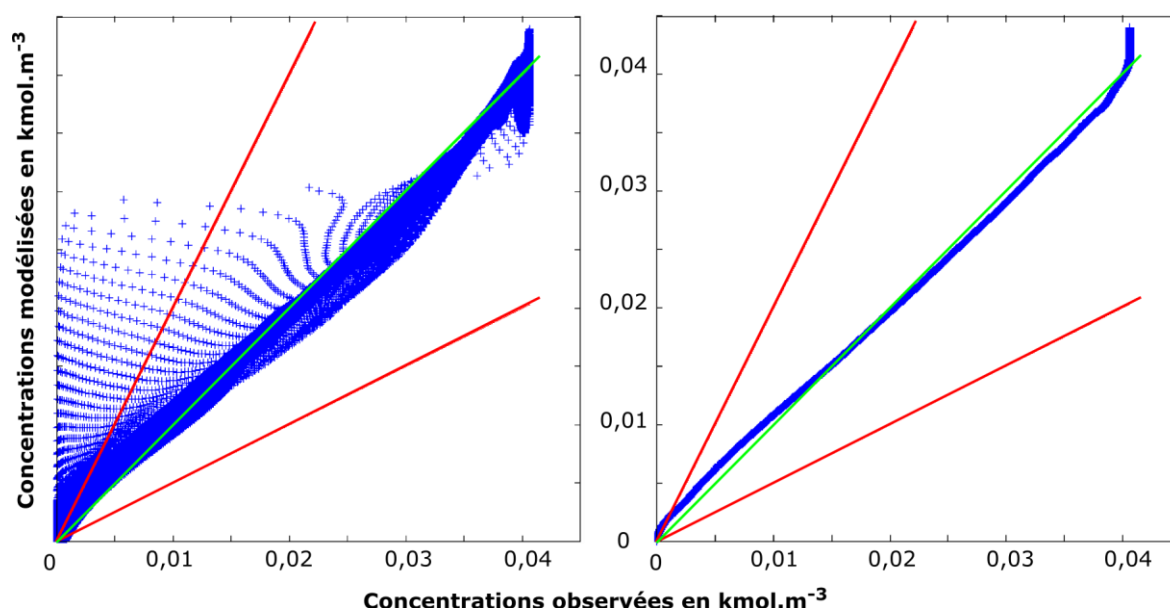


Figure III-24 : Diagramme de dispersion des valeurs observées et modélisées pour le cas favorable 5 (appairées dans l'espace sur la figure de gauche et par rang dans la figure de droite).

III-2.5. Temps de calcul

La constitution de la base de données CFD est longue. Effectivement, chaque simulation nécessite des itérations de convergence du système d'équations. Selon le cas, le temps de calcul est d'environ 10 minutes. Pour l'apprentissage du réseau de neurones, il est nécessaire de créer une base de données représentative du phénomène à modéliser. Une fois cette base de données obtenue, une base d'exemples est créée à partir de la base de données. Celle-ci est présentée au réseau de neurones lors de l'apprentissage. L'optimisation de l'architecture implique un nombre d'apprentissages important. Toutes ces étapes sont effectuées en amont de l'utilisation du modèle. La durée d'apprentissage peut être longue (675 secondes pour le modèle sélectionné) sans relation avec le temps de calcul lors de l'utilisation. En effet, la simulation d'un cas test par le réseau de neurones est extrêmement rapide. Le réseau de neurones simule des milliers de points de manière quasi-instantanée (durée inférieure à la seconde) réduisant drastiquement le temps de calcul.

III-2.6. Conclusion

Les possibilités d'utilisation des réseaux de neurones ont été présentées pour l'émulation d'un modèle CFD dans le cas simple de la dispersion atmosphérique stationnaire en champ libre. Les critères permettant de comparer ces modèles ont été définis comme la vitesse d'exécution et la précision.

Le modèle par réseau de neurones nécessite une base de données afin d'effectuer un apprentissage. Celle-ci est issue d'un modèle CFD de type RANS $k-\epsilon$ permettant d'obtenir la précision recherchée. La dispersion de méthane a été simulée en faisant varier deux paramètres, la vitesse de l'écoulement et le débit massique. La base de données utilisée par les réseaux de neurones a été échantillonnée de manière à être représentative du phénomène.

L'influence de l'échantillonnage sur la qualité de généralisation du réseau de neurones a été étudiée. Il a été mis en évidence qu'elle était moins déterminante que le nombre de neurones dans la couche cachée, ou que l'initialisation du modèle neuronal avant l'apprentissage. L'influence de cette initialisation des paramètres de l'ANN a été étudiée spécifiquement.

Par ailleurs, l'apprentissage d'un réseau de neurones est généralement effectué sur des données expérimentales. Ce type de données n'est généralement pas exempt de bruit de mesure. Dans le cadre de cette étude, les données d'apprentissage proviennent d'un modèle numérique pour lequel il n'y a pas de bruit. La présence de diffusion numérique éventuelle n'est pas un facteur pénalisant dans la mesure où l'on compare le modèle ANN au modèle CFD. Par ailleurs, il peut subsister des approximations dues à l'utilisation d'un maillage. Ces éléments permettent d'expliquer l'amélioration continue de l'apprentissage avec l'augmentation du nombre de neurones dans la couche cachée, moyennant un nombre suffisant d'initialisations testées. Ce type de comportement ne serait sûrement pas observé en utilisant une base de données issue de mesures expérimentales.

Le meilleur modèle a été extrait de manière à procéder à une évaluation du modèle en phase de généralisation sur des cas tests. Différentes valeurs de vitesse et de débit ont été simulées. Les résultats montrent une bonne restitution de la majorité des concentrations. Le réseau de neurones a plus de difficultés à modéliser les zones de hauts gradients, notamment à proximité de la source. D'autre part, la concentration maximale est surestimée dans les différents cas tests. Cette surestimation reste cependant contenue à moins de 7% de la valeur CFD.

Ce chapitre a fait l'objet de la rédaction d'un article disponible en annexe. Ces travaux ont été présentés au 14th International Symposium on Loss Prevention and Safety Promotion in the Process Industries à Florence, Italie, le 14 mai 2013.

III-3. Modélisation de la dispersion atmosphérique d'une bouffée en champ libre

Ce chapitre présente l'évolution de la dispersion atmosphérique d'une bouffée de méthane dans un écoulement stationnaire, en champ libre. Afin de prendre en compte l'aspect temporel, le réseau de neurones est couplé à un automate cellulaire.

III-3.1. Objectifs

L'objectif de cette partie est d'évaluer les possibilités d'utilisation des réseaux de neurones pour modéliser la dispersion atmosphérique instationnaire. La difficulté majeure ici est de fournir une estimation de la concentration à la fois localisée dans le temps et dans l'espace. Dans le paragraphe III-2, les réseaux de neurones ont été utilisés afin de fournir une sortie localisée dans l'espace. Cela a été rendu possible par l'utilisation de données de localisation en tant que données d'entrée. La difficulté supplémentaire du présent chapitre réside dans l'ajout de la composante temporelle. Si celle-ci est prise en tant que variable d'entrée, alors la base de données doit contenir toutes les dimensions de nuages envisageables, ainsi que la totalité des durées que l'on souhaitera modéliser par la suite et les différentes valeurs du champ de vitesse de l'écoulement. La figure ci-après schématise un exemple de ce fonctionnement :

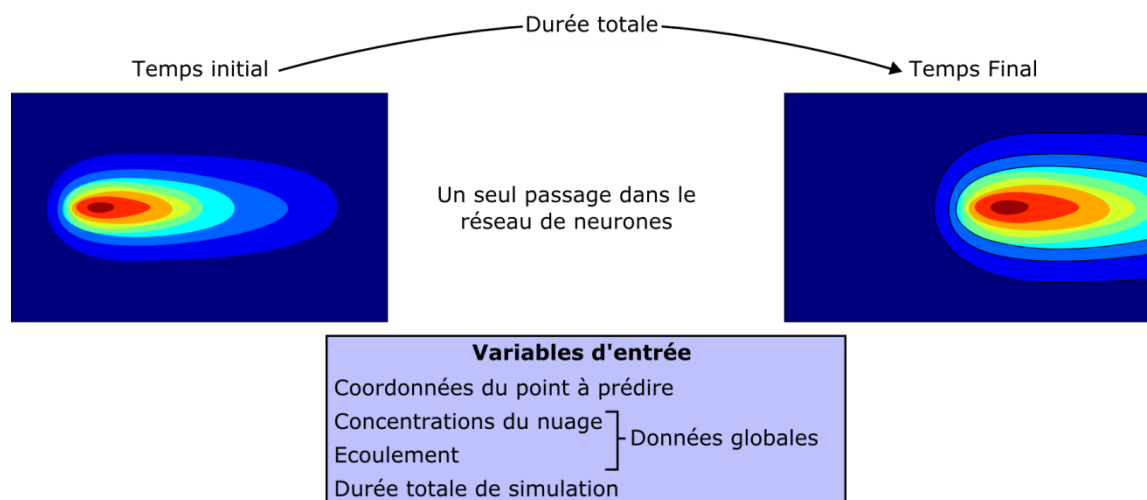


Figure III-25 : Schématisation du fonctionnement d'un réseau de neurones statique tel que celui développé au III-2 avec ajout de la variable temporelle.

Une telle base de données n'est pas réalisable en pratique. L'approche envisagée dans ce chapitre est donc d'utiliser les automates cellulaires pour discrétiser spatialement et temporellement le problème. Ainsi, la règle de transition est réalisée par un réseau de neurones en fonction des interactions locales. Un tel modèle est schématisé par la figure ci-après :

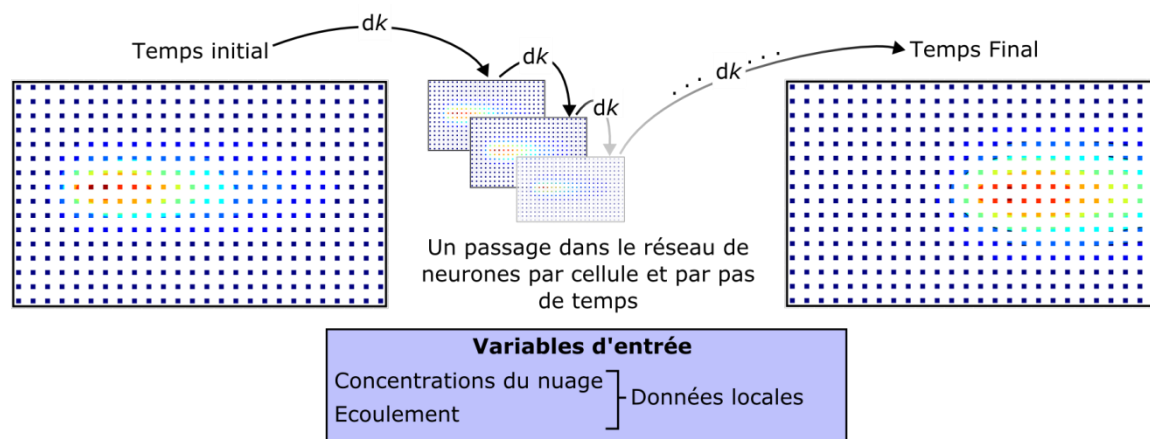


Figure III-26 : Schématisation du modèle développé : les interactions locales sont prises en compte à travers un maillage régulier. La règle de transition de type réseau de neurones permet d'estimer la concentration dans chaque cellule au pas de temps suivant.

III-3.1.1 Conditions globales du système

Cette utilisation combinée des automates cellulaires comportant une règle de transition implémentée par un réseau de neurones, appelée CA-ANN (Cellular Automata Artificial Neural Network ruled) conditionne des choix sur le modèle neuronal. Chaque automate cellulaire doit représenter dans le temps l'évolution du phénomène dynamique de dispersion : à chaque pas de temps de l'automate cellulaire, le réseau de neurones calcule la concentration au pas de temps suivant. La modélisation de phénomènes dynamiques impose la définition d'hypothèses concernant la manière dont le bruit intervient dans le processus. Schématiquement, ainsi que proposé par Dreyfus (2004), le bruit peut affecter l'état du processus postulé pour représenter le processus physique (hypothèse « bruit d'état » dû à une erreur sur la fonction du processus ou sur la valeur des variables) ; ou bien affecter la sortie mesurée de ce dernier (hypothèse « bruit de sortie » dû à une erreur de mesure). Dans le premier cas, l'information de concentration pertinente à appliquer en entrée est la concentration mesurée, car l'état est bruité (le modèle est donc non-récurrent, même s'il représente un phénomène dynamique); dans le second cas la concentration à appliquer au modèle est la concentration estimée par lui-même puisque la mesure est bruitée (le modèle est donc récurrent). Dans la majorité des cas réels, les deux types de bruit affectent les processus ou les bases de données, l'analyse du bruit le plus déterminant permet de choisir le modèle adéquat. En ce qui concerne notre application, comme il n'y a pas de bruit sur la base de données CFD, à l'exception du bruit de quantification, le modèle non récurrent (comportant la concentration mesurée au pas de temps précédent en entrée) est le plus adéquat. Néanmoins, en situation réelle de dispersion atmosphérique, nous ne disposons pas de la valeur mesurée de la concentration en chaque point et à chaque pas de temps ; le modèle purement récurrent, dans lequel la concentration estimée par le modèle à l'instant k est ré-injectée en entrée à l'instant $k+dk$ s'impose donc à nous. Nous pouvons donc dès à présent noter que ce modèle n'est pas optimal par rapport à l'utilisation que nous en faisons dans ce travail, mais qu'il prendra tout son sens lors d'une utilisation opérationnelle. Le modèle récurrent permet de représenter l'évolution locale de la concentration en effectuant une prévision à un pas de temps en fonction des variables locales. Pour la première itération, ce bruit provient d'un état mesuré, ou de l'initialisation, alors que pour les pas de temps suivants, l'état est non mesuré, le bruit étant alors introduit par l'erreur d'estimation du réseau de neurones. Ce type de modèle

permet d'utiliser le prédicteur à un pas en tant que simulateur. La figure ci-après présente l'utilisation de ce modèle dit récurrent :

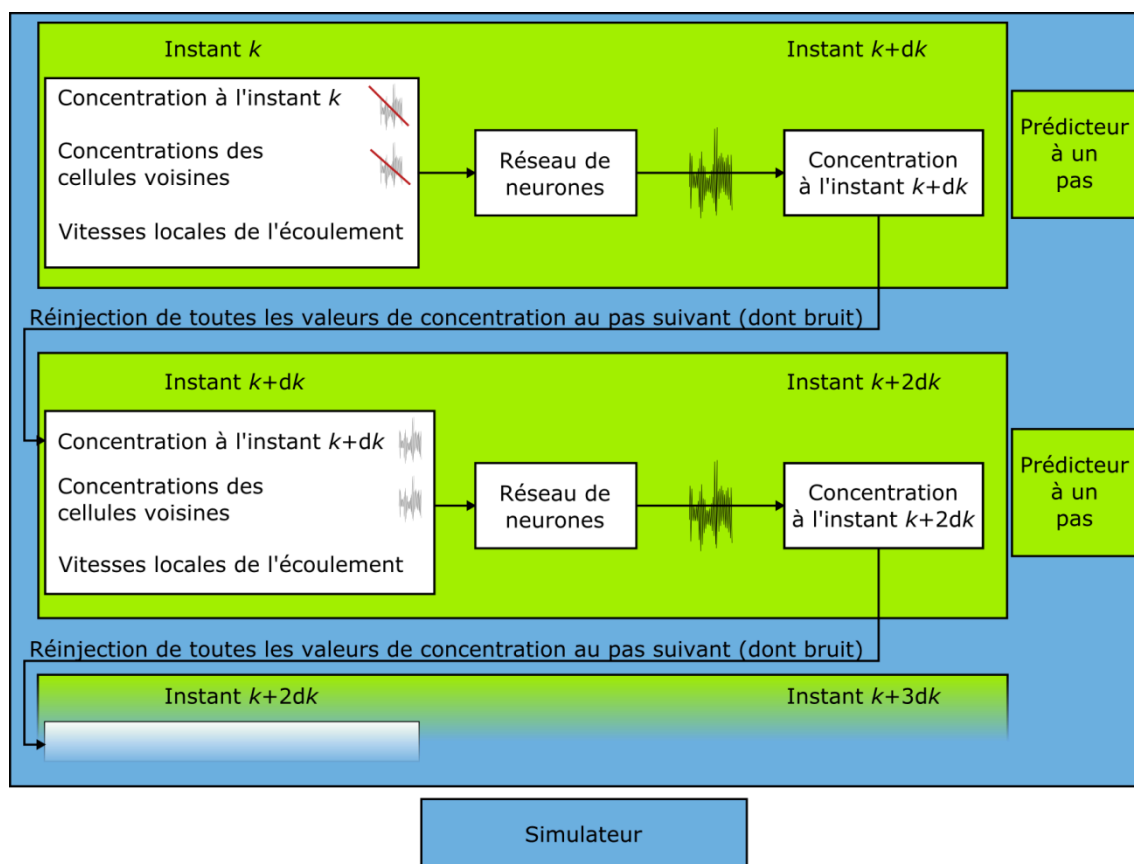


Figure III-27 : Représentation schématique du fonctionnement du prédicteur neuronal associé au modèle postulé « bruit d'état ». Evolution sur plusieurs pas de temps dk .

Les concentrations initiales de la cellule considérée ainsi que celles des cellules voisines ne contiennent aucun bruit à l'instant initial, elles sont initialisées aux valeurs issues de la CFD. L'estimation du réseau de neurones introduit une erreur due essentiellement au fait qu'il modélise imparfaitement la fonction souhaitée. Cette erreur est propagée au cours des pas de temps.

III-3.1.2 Interactions locales

L'évolution de la concentration est la conséquence de plusieurs phénomènes (I-3). Le transport (advection et diffusion) pilote l'évolution de la concentration au cours du temps. Les phénomènes permettant la prédiction de transport d'un scalaire tel que la concentration sont le vent et la turbulence. Le vent dirige l'advection tandis que la diffusion est la conséquence de la turbulence. L'équation d'advection-diffusion sous sa forme eulérienne retranscrit cette dépendance locale. Il est donc nécessaire de prendre en compte ces phénomènes. Les automates cellulaires sont des outils conçus pour représenter les phénomènes spatiotemporels où l'interaction locale est prépondérante pour la prévision du comportement global. C'est pourquoi cet outil est envisagé ici. Sa combinaison avec les capacités d'identification de fonction non-linéaire des réseaux de neurones permet d'envisager leur utilisation.

III-3.2. Méthode d'utilisation du CA-ANN

Le modèle présenté dans le cadre de ce chapitre doit permettre de modéliser l'évolution au cours du temps d'un nuage de méthane en champ libre. On considère un champ de vent homogène et une turbulence liée à des conditions de stabilité atmosphérique neutre (III-1.2).

La règle de transition de cet automate cellulaire est réalisée par un réseau de neurones. Les entrées de ce réseau sont composées des paramètres locaux intervenant sur la dispersion. Le fonctionnement de cette combinaison est détaillé dans la suite.

III-3.2.1 Paramètres de la modélisation

L'automate cellulaire pave le domaine de calcul. Celui-ci est orienté dans le sens de l'écoulement.

Cet automate est composé de cellules qui ont chacune une règle de transition afin de déterminer leur état à chaque pas de temps. Cette règle de transition utilise les valeurs des cellules adjacentes. Chaque cellule voit son état mis à jour de manière synchrone pour l'ensemble des cellules en utilisant la règle de transition :

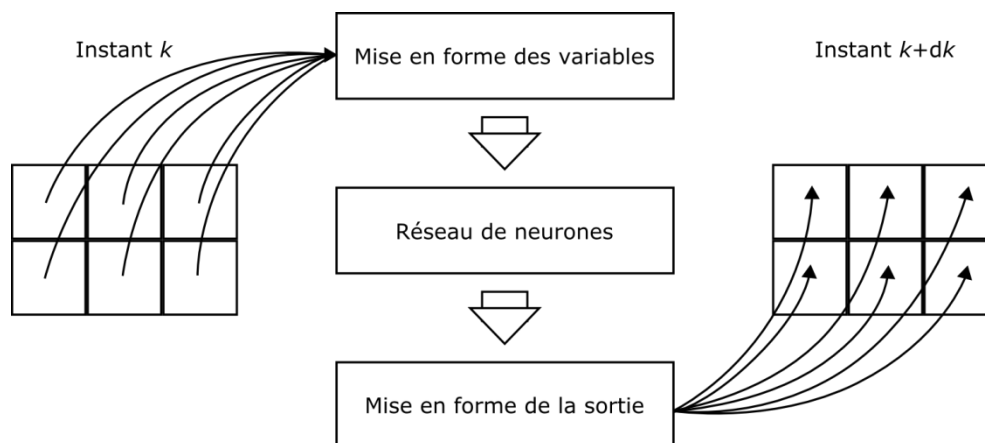


Figure III-28 : Schématisation de la mise à jour de l'état de l'automate cellulaire de manière synchrone au cours d'un pas de temps dk .

La mise en forme des variables d'entrée et de la sortie consiste en la normalisation des variables d'entrée et en la dénormalisation de la valeur de sortie fournie par le réseau de neurones (II-2.2.4).

L'automate cellulaire est défini grâce aux éléments suivants :

- ▶ Définition du domaine de calcul
- ▶ Maillage du domaine
- ▶ Détermination des variables d'entrée de la règle de transition
- ▶ Détermination des conditions aux limites
- ▶ Initialisation des cellules du domaine

Le domaine de calcul et le maillage sont des caractéristiques intrinsèques de l'automate cellulaire. La dimension des cellules est à fixer. Les cellules sont de forme carrée, de manière à faciliter la détermination des variables d'entrée de la règle de transition.

III-3.2.2 Détermination des variables d'entrée

Les entrées de la règle de transition de chaque cellule sont déterminées à partir des états des voisins. La sélection des variables pertinentes pour le réseau de neurones est une opération importante pour obtenir une complexité la plus faible possible afin d'éviter le surapprentissage (Doursat et al., 1992). Bien que les variables d'intérêt puissent être sélectionnées par essais successifs, il est possible et plus approprié dans notre cas d'utiliser les connaissances des équations régissant la dispersion atmosphérique. Le transport de substances par simulation CA-ANN s'inspire donc des équations résolues par le modèle CFD. Dans ce cas là, l'équation d'advection-diffusion (ADE voir I-3) est à émuler.

Cette équation est résolue à travers un solveur en passant par la discrétisation des différents termes la composant. On distingue ainsi les termes d'advection $U_{x_i} \frac{\partial C}{\partial x_i}$ et les termes de diffusion $\frac{\partial}{\partial x_i} \left(D_t \frac{\partial C}{\partial x_i} \right)$. Dans le cas du champ libre, on considère une turbulence homogène dans le plan horizontal. Ainsi, le terme de diffusion se simplifie : $D_t \frac{\partial^2 C}{\partial x_i^2}$. Leur discrétisation fait apparaître les valeurs de concentrations des cellules adjacentes, ainsi que des valeurs de vitesse de vent. Le calcul des termes d'advection et de diffusion à travers une discrétisation permet d'obtenir les variables d'entrée du réseau de neurones qui implémente la règle de transition.

Les entrées du réseau de neurones sont calculées en fonction d'une discrétisation explicite par différences finies, décentrée amont en ce qui concerne les termes d'advection et centrée pour les termes de diffusion. Ainsi, la figure ci-après représente la création des variables d'entrée :

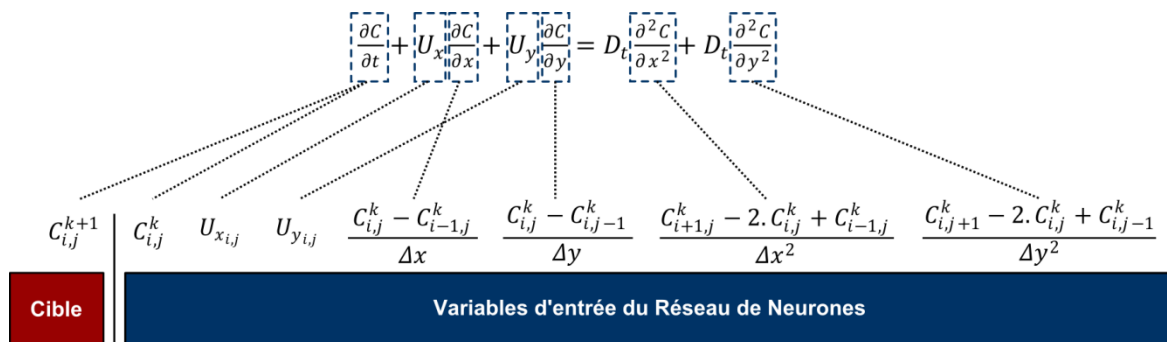


Figure III-29 : Variables d'entrée et cible du Réseau de Neurones

La valeur cible est la concentration dans la cellule considérée au pas de temps suivant. Les cellules sont définies par leurs indices : i pour la direction selon x et j pour la direction selon y . k représente le temps discret. Les variables d'entrée et la cible étant déterminées, la construction de la base de données numérique peut commencer.

III-3.2.3 Conditions aux limites

En limite du domaine, il n'est pas possible d'effectuer le calcul par manque de données. Il est nécessaire d'appliquer des conditions aux limites afin de fournir des valeurs permettant de mettre à jour l'état de la cellule considérée. Elles sont définies comme suit :

- ▶ Entrée : concentrations fixées à zéro ;
- ▶ Côté du domaine : de la même façon que pour les méthodes CFD, on utilise des conditions de symétries pour le calcul ;

- Sortie du domaine : les termes de diffusion selon x ne peuvent être calculés. On fixe à zéro la valeur de la concentration dans la cellule adjacente.

Le domaine de calcul maillé et avec les conditions aux limites est illustré sur la figure ci-après :

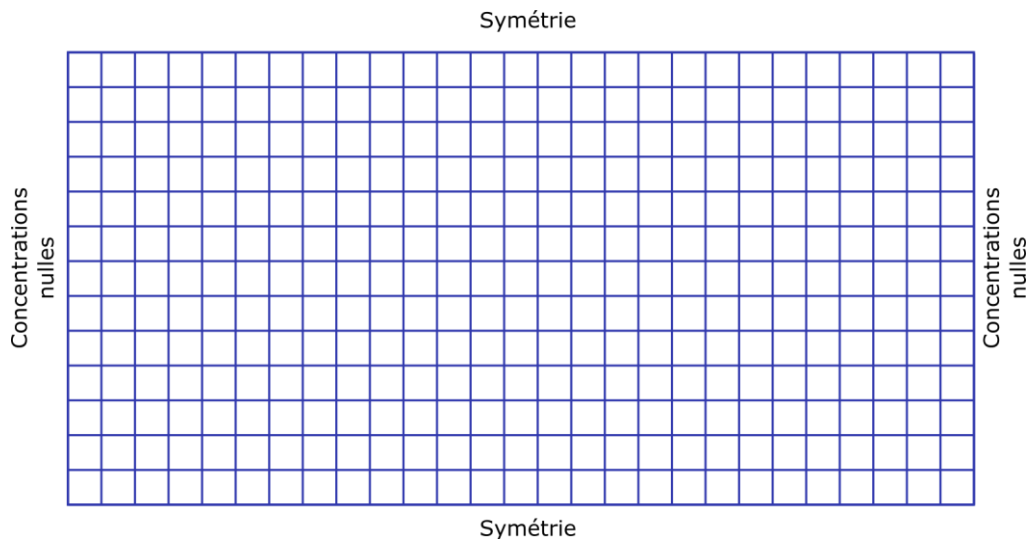


Figure III-30 : Maillage et conditions aux limites du domaine de calcul de l'automate cellulaire

III-3.2.4 Utilisation du CA-ANN

Le CA-ANN modélise un processus transitoire en calculant l'évolution d'une distribution spatiale de concentrations à chaque pas de temps. Ce pas de temps est défini en fonction du nombre de Courant Friedrich Lewy (Courant et al., 1967). Ce nombre est un critère de convergence utilisé en calcul par différences finies représentant une condition nécessaire mais non-suffisante. Il conditionne les dimensions des cellules au pas temporel :

$$C_{FL} = \frac{v \Delta t}{\Delta x} \quad (III-14)$$

v	Plus grande vitesse observée	m.s^{-1}
Δx	Intervalle spatial	m
Δt	Intervalle temporel	s

Ce nombre permet de fixer le seuil au-delà duquel une instabilité de calcul est observée et grandit avec le nombre de pas de temps. Cette instabilité apparaît lorsque la distance parcourue par une particule portée par la vitesse maximale dans le domaine devient supérieure à la dimension de la maille de calcul. Dans ce travail, la vitesse maximale dans la base de données est de 20 m.s^{-1} . En fixant la maille de calcul à une largeur de 0,2 m et en limitant la valeur de C_{FL} à 1, le pas de temps doit être pris inférieur ou égal à 0,01 seconde. Ces paramètres sont à assurer durant la phase d'apprentissage comme dans la phase finale d'utilisation du modèle.

Une fois ces paramètres fixés, il est possible de déterminer une séquence d'utilisation de l'automate cellulaire à règle de transition neuronale à partir du fonctionnement d'une cellule :

- Lecture de l'état des voisins : concentrations et valeurs de vitesse de vent ;

- ▶ Calcul des dérivées discrètes des termes d'advection et de diffusion de l'ADE ;
- ▶ Prétraitement des variables d'entrée (normalisation) ;
- ▶ Calcul de l'état suivant : application de la règle de transition neuronale ;
- ▶ Récupération de la valeur de la concentration ;
- ▶ Le calcul se termine en fonction du nombre de pas de temps spécifié par l'utilisateur.

La figure ci-après illustre ce procédé :

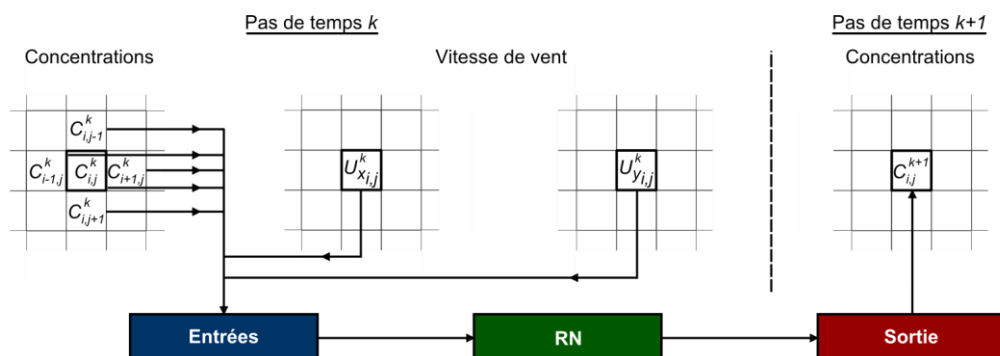


Figure III-31 : Procédure de mise à jour de l'état d'une cellule

Ce fonctionnement, décrit pour une cellule est effectué simultanément sur l'ensemble des cellules constituant l'automate cellulaire. Le nombre d'itérations est fonction de la durée d'observation désirée. Pour pouvoir utiliser ces réseaux de neurones en tant que règle de transition, il est nécessaire d'effectuer un apprentissage des différentes configurations possibles. Il est donc nécessaire de créer une nouvelle base de données représentative des situations à modéliser.

III-3.3. Base de données et base d'exemples

III-3.3.1 Création de la base de données

Comme précédemment, on utilisera une base de données issue de calculs numériques à partir d'un modèle $k - \epsilon$ standard, représentant une réalité « simulée ». L'étude est donc basée sur des simulations de la dispersion atmosphérique de méthane en deux dimensions selon le plan horizontal. L'objectif est de modéliser un rejet de méthane au bord du domaine de calcul et de suivre l'évolution de cette bouffée au cours du temps.

La géométrie du domaine est rectangulaire, de 20 m de large et 30 m de long. Les effets de la gravité ne sont pas pris en compte. Le plan horizontal est situé à une altitude de deux mètres de hauteur.

Le maillage est composé de 240 000 nœuds. Les mailles sont d'autant plus fines que l'on se rapproche de l'axe de symétrie du domaine afin de faciliter le calcul des plus hauts gradients.

Les conditions aux limites sont multiples :

- ▶ Les côtés du domaine sont représentés par des conditions de symétries ;
- ▶ La sortie du domaine est fixée d'après la pression atmosphérique ;
- ▶ Les entrées du domaine sont de plusieurs types :
 - ▶ Vitesse du vent sur les côtés
 - ▶ Vitesse de rejet de mélange méthane/air au centre

La figure ci-après schématise ce fonctionnement :

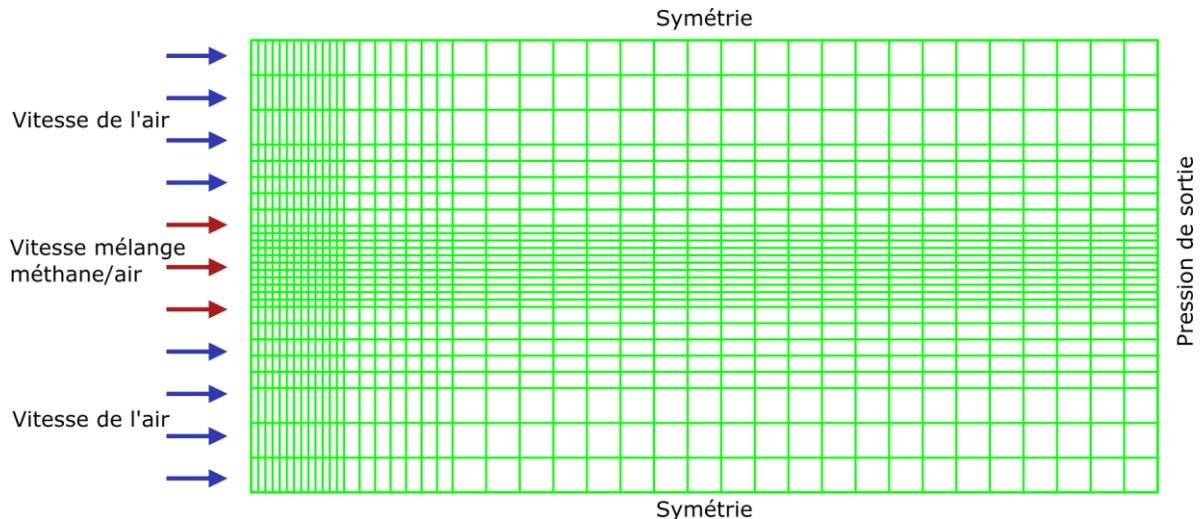


Figure III-32 : Schéma du maillage CFD utilisé

Le méthane est injecté durant 0,3 à 3 secondes en fonction de la vitesse de l'écoulement considéré, de manière à créer une bouffée. La distribution spatiale de concentration est enregistrée à chaque pas de temps aux nœuds du maillage CFD. Les variables d'intérêt sont la vitesse de l'écoulement et la fraction massique de méthane libérée durant le rejet. L'illustration suivante représente à un instant donné la dispersion du nuage généré :

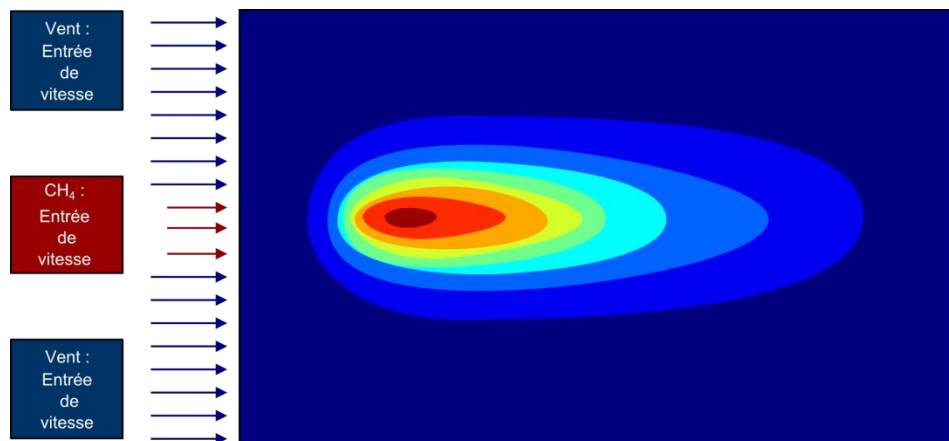


Figure III-33 : Configuration du cas général pour la constitution de la base de données CFD

La base de données CFD est donc composée de 95 configurations différentes résultant de variations de vitesse de 2 à 20 m.s^{-1} par pas de 1 m.s^{-1} et de fraction massique de 0,2 à 1 par pas de 0,2. Pour chaque configuration, 20 pas de temps sont enregistrés, séparés de 0,01 seconde comme défini par la relation entre le pas spatial, la vitesse maximale de l'écoulement et le nombre de Courant CFL. Chaque enregistrement est effectué lorsque le résiduel des itérations de convergence atteint une valeur inférieure à 10^{-5} . Dans la base de données, il y a donc 95 configurations différentes, 20 pas de temps (dt) pour chaque configuration et 240 000 valeurs par enregistrement (figure ci-après) :

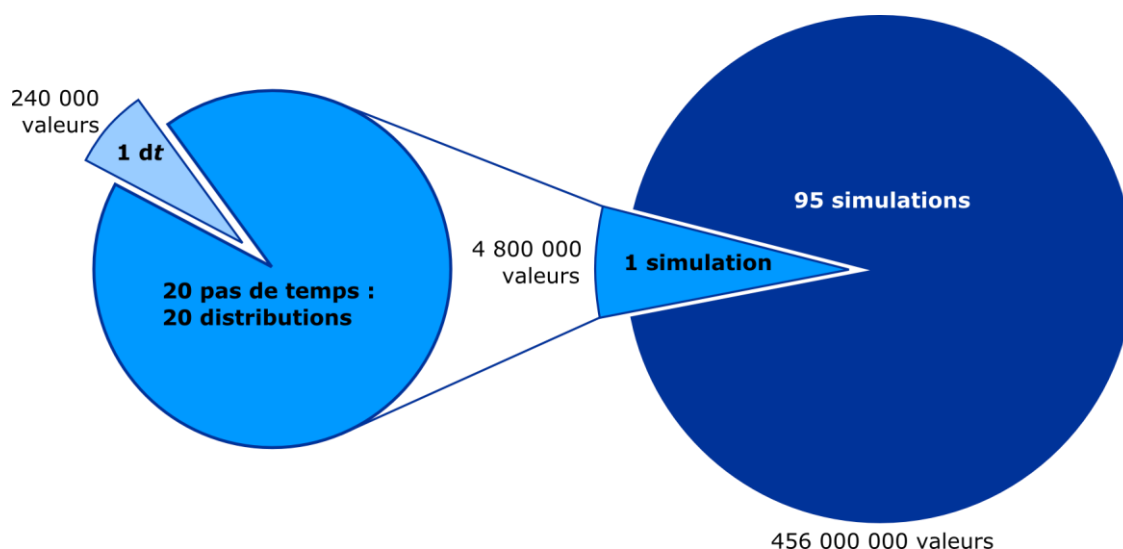


Figure III-34 : Représentation schématique de la base de données CFD.

L'ensemble des valeurs de la totalité des simulations CFD représente environ 456 millions de valeurs. Il n'est pas envisageable d'effectuer l'apprentissage du réseau de neurones avec autant de points. Le passage de la base de données à la base d'exemples est détaillé dans la suite.

III-3.3.2 De la base de données à la base d'exemples

Comme énoncé précédemment, l'automate cellulaire a un pas spatial de 0,2 m. Les cellules étant carrées, le domaine de calcul est composé de 151 cellules en longueur et 101 en largeur soit 15 251 cellules au total. Afin de pouvoir effectuer l'apprentissage, on ramène le nombre de valeurs issues du calcul de la dispersion aux 15 251 cellules de l'automate cellulaire. On utilise pour cela une interpolation bicubique des valeurs des points CFD sur le maillage AC (figure ci-après) :

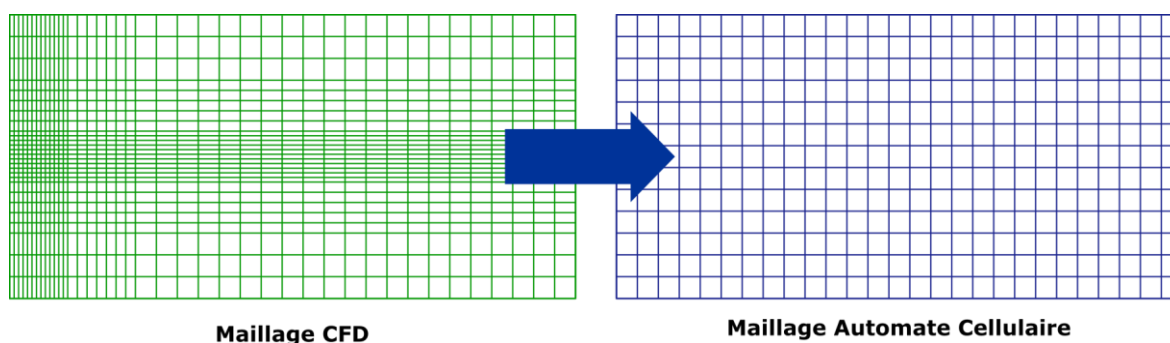


Figure III-35 : Représentation schématique de l'interpolation du maillage CFD (vert) sur le maillage de l'automate cellulaire (bleu)

La procédure décrite au III-1.3.1iii est utilisée pour effectuer l'échantillonnage et ainsi créer la base d'exemples. La variable à prédire étant la concentration, les fortes concentrations doivent à minima avoir le même poids et donc la même représentativité que les valeurs faibles. On utilise donc des couples nombre d'intervalles de concentration (I_c) et nombre de valeurs par intervalles (E_c) pour définir un échantillonnage. En effectuant cette manipulation sur l'ensemble des enregistrements, et en observant la distribution des valeurs de concentrations obtenues sur 10 classes, on note une surreprésentation des valeurs faibles par rapport aux valeurs fortes.

Effectivement, si des cellules à faible concentration existent dans chaque configuration, ce n'est pas le cas pour les fortes concentrations. Pour rétablir la représentativité des fortes concentrations, un seuil est fixé à partir du nombre d'éléments de la classe de concentrations fortes. Chaque classe contient ainsi le même nombre de valeurs. La figure ci-après illustre le principe sur une distribution issue d'un cas de la base de données :

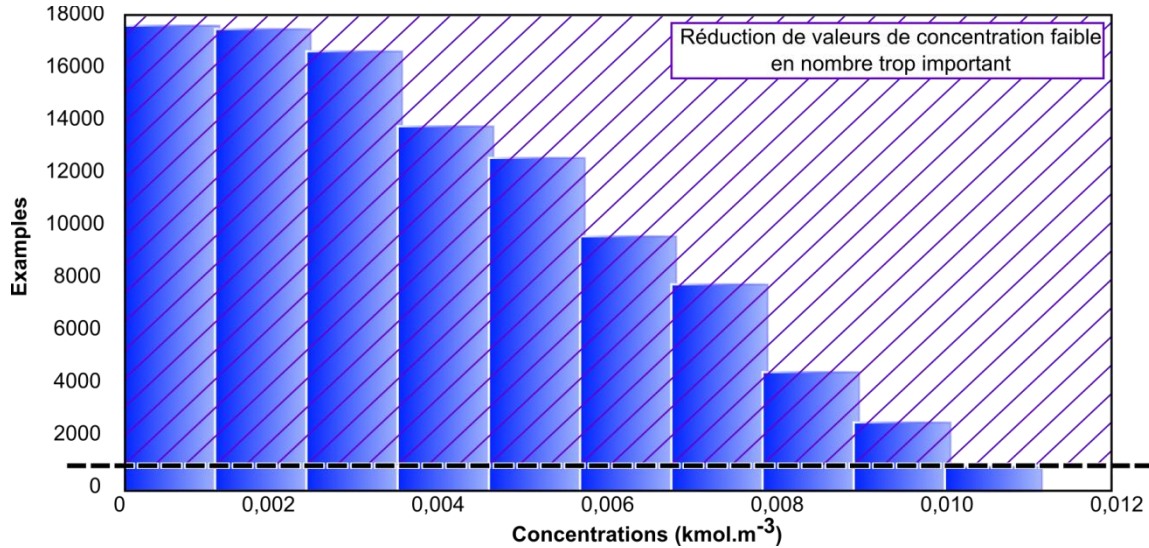


Figure III-36 : Représentation du seuil définissant le nombre d'éléments extraits de chaque classe de concentration

Les différentes valeurs de I_c et E_c choisies vont déterminer le nombre final d'exemples dans la base comme illustré dans la figure ci-après :

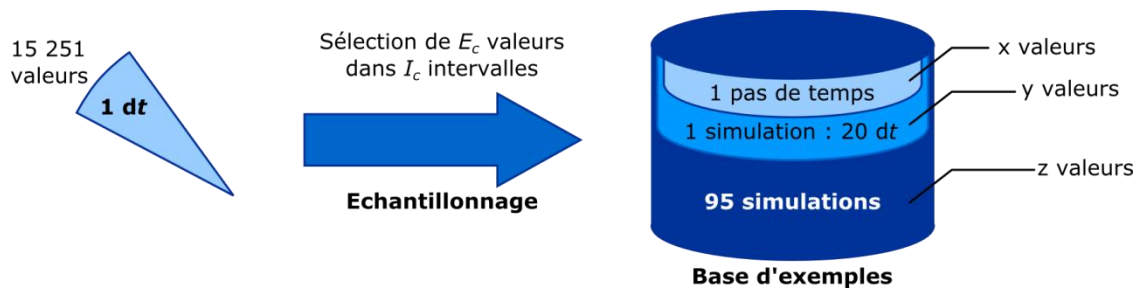


Figure III-37 : Représentation schématique de la constitution de la base d'exemples à partir de l'échantillonnage de la base de données CFD.

Il s'agit à présent d'effectuer différents échantillonnages afin de déterminer le nombre optimal d'exemples dans la base d'exemples.

III-3.4. Réseau de neurones

III-3.4.1 Architecture

Le réseau de neurones utilisé est un perceptron à une couche cachée. Les fonctions d'activations sont des tangentes hyperboliques pour les neurones de la couche cachée. Le neurone de sortie a une fonction d'activation linéaire. L'influence du nombre de neurones dans la couche cachée est étudiée en faisant varier cette valeur de 1 à 20.

Les entrées et cibles du réseau de neurones sont centrées et réduites dans l'intervalle $[-0,9 ; 0,9]$, excepté pour le cas de la concentration, centrée et réduite dans l'intervalle $[0 ; 0,9]$, de manière à éviter des concentrations négatives en sortie. Puisque l'influence de la valeur des paramètres initiaux a été montrée au III-2.3.1, on effectuera 20 initialisations pour chaque configuration de l'ANN étudiée.

Afin de sélectionner le meilleur modèle, on calcule le score de validation croisée S_{vc} , détaillé au II-2.2.6. Pour chaque échantillonnage réalisé, l'ensemble d'apprentissage comprend 80 % des valeurs. 10 % des valeurs sont allouées à l'ensemble d'arrêt et 10 % à l'ensemble de test. L'ensemble d'arrêt est utilisé pour effectuer l'arrêt précoce. La valeur de l'erreur quadratique moyenne sur l'ensemble de test permet de comparer les différents modèles et de sélectionner le plus performant.

III-3.4.2 Méthode d'évaluation de l'apprentissage

Les performances du modèle sont optimisées en étudiant les influences de l'échantillonnage de la base de données, du nombre de neurones dans la couche cachée et de l'initialisation des paramètres. Afin de sélectionner le meilleur modèle, la validation croisée est utilisée à travers le calcul du score de validation croisée S_{vc} . La séquence d'optimisation est illustrée dans la figure ci-après. L'ensemble des configurations des modèles est évalué en regard du score S_{vc} le plus bas (erreur quadratique).

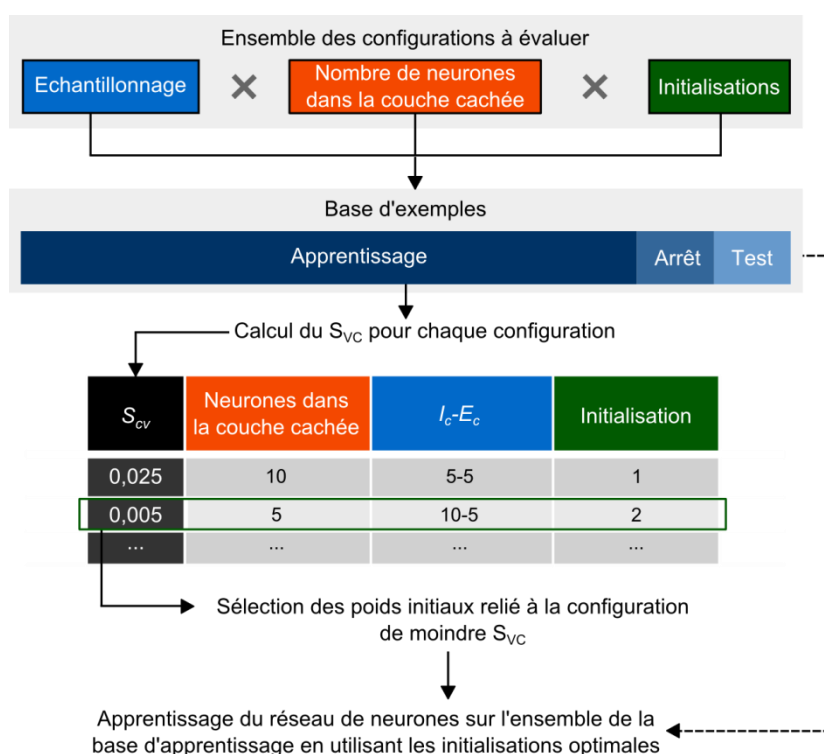


Figure III-38 : Sélection et apprentissage du meilleur modèle possible

Une fois le modèle de fonction de transition optimisé, il sera nécessaire d'évaluer le modèle complet d'automates cellulaires utilisant cette règle de transition sur des cas tests.

III-3.4.3 Apprentissage et optimisation

La conception du modèle de fonction de transition est l'objet de ce paragraphe. Afin de sélectionner le meilleur modèle, plusieurs configurations ont été testées. 10 initialisations de paramètres, avec 20 architectures différentes (le nombre de neurones dans la couche cachée varie de 1 à 20) et trois échantillonnages différents de la base de données ont été réalisés. Trois bases d'exemples de différentes dimensions ont été créées à partir des couples $\{I_c ; E_c\}$ suivants : $\{5 ; 7\}$, $\{15 ; 20\}$ et $\{25 ; 10\}$ contenant respectivement 18 060, 125 300 et 418 160 exemples.

Regardons tout d'abord l'influence de l'initialisation sur la qualité de l'apprentissage. Le score de validation croisée S_{vc} est évalué sur l'échantillon comportant 18 060 exemples, avec 20 neurones dans la couche cachée. L'évaluation de l'influence des différents paramètres est évaluée sur la base de données contenant le moins d'exemples pour des considérations de temps de calcul, discutés dans la suite de ce travail. La figure ci-après représente les valeurs de S_{vc} pour 10 initialisations différentes.

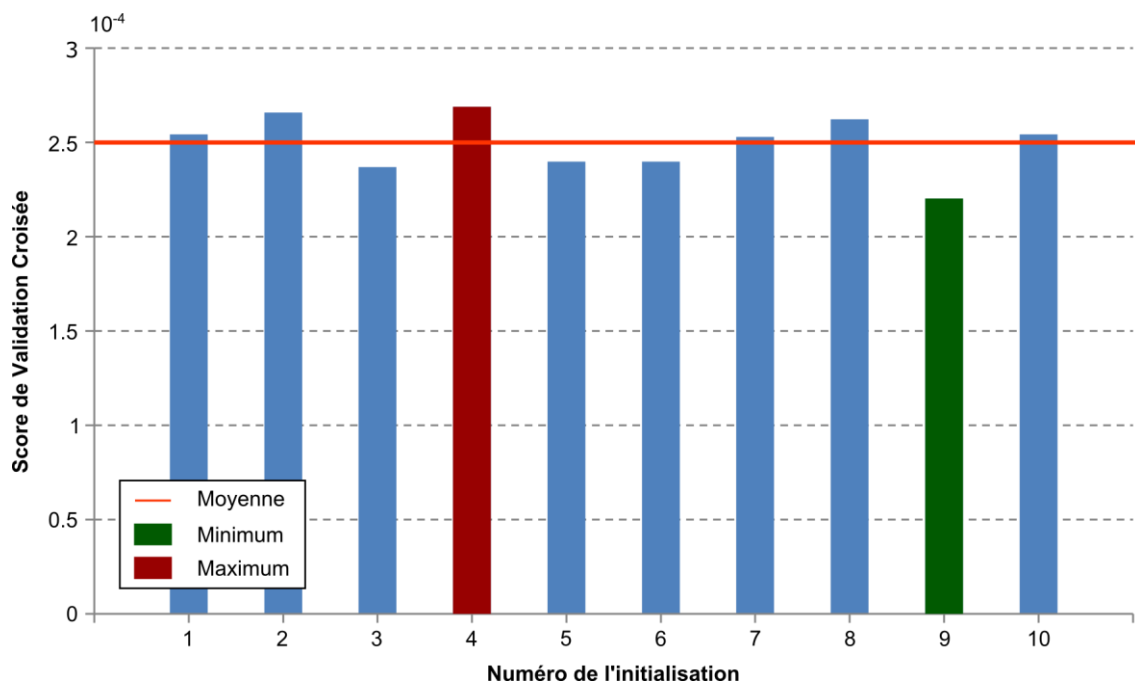


Figure III-39 : Influence de l'initialisation sur la qualité la généralisation

Les valeurs du score de validation croisée sont peu écartées de la moyenne. L'écart type de cette distribution est de moins de 10 % de la valeur moyenne. L'initialisation a donc un impact sur la qualité de l'apprentissage. Néanmoins, cet impact n'est pas prépondérant et cette valeur de 10 initialisations sera conservée.

L'influence du nombre de neurones dans la couche cachée est maintenant évaluée. A partir des mêmes bases d'apprentissage que précédemment, le nombre de neurones dans la couche cachée est modifié : de 1 à 20. Pour chaque apprentissage, 10 initialisations différentes sont effectuées. La figure ci-après reporte la valeur médiane de ces 10 valeurs. Les barres d'erreurs représentent les valeurs maximum et minimum obtenues pour le score de validation croisée durant ces apprentissages. Afin de visualiser au mieux les différences entre les configurations, l'échelle associée au score de validation croisée est logarithmique.

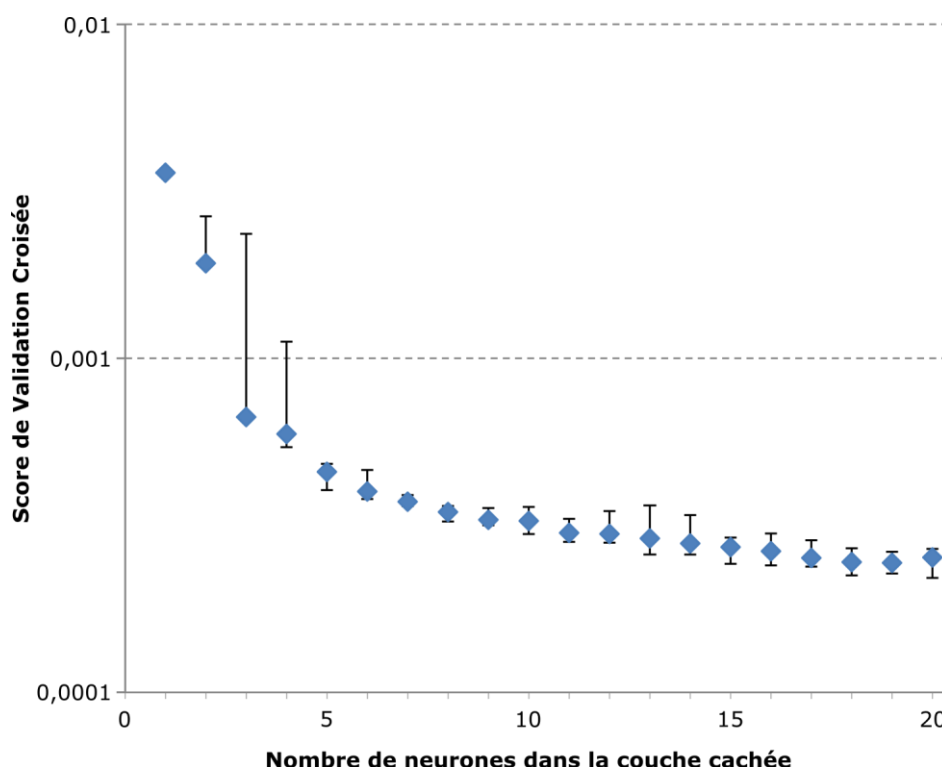


Figure III-40 : Influence du nombre de neurones dans la couche cachée pour des apprentissages sur une base d'exemples de 18 060 exemples (Valeurs médianes pour 10 initialisations – barres d'erreurs : minimum et maximum des 20 initialisations).

Une diminution du score de validation croisée avec le nombre de neurones dans la couche cachée est notée. Cette diminution est très prononcée lorsque moins de 5 neurones composent la couche cachée ce qui signifie qu'une complexité minimale est nécessaire. Lorsque le nombre de neurones dans la couche cachée continue d'augmenter, le score diminue plus faiblement puis oscille. La valeur minimale atteinte par le score de validation croisée pour l'ensemble des configurations testées est de $2,2 \times 10^{-4}$. Elle est obtenue pour 20 neurones dans la couche cachée. Cette diminution en fonction du nombre de neurones dans la couche cachée montre que le bruit contenu dans la base d'exemples, dû à l'interpolation des données sur le maillage régulier de l'automate cellulaire est faible. Dans l'absolu, on pourrait essayer de rajouter des neurones dans la couche cachée mais le gain induit comparé à la complexité accrue et donc au temps de calcul croissant n'est pas une priorité dans cette première approche. Cette piste est néanmoins une voie d'amélioration du modèle.

Il est nécessaire d'évaluer l'influence du nombre d'exemples sur le score de validation croisée. Pour cela, les trois bases sont utilisées sur des apprentissages avec un nombre de neurones dans la couche cachée de 1 à 20 afin de visualiser une tendance. La valeur médiane de 10 initialisations est utilisée pour la base la moins fournie afin de renforcer la robustesse du résultat. Pour les deux autres, une seule initialisation est testée car les durées d'apprentissages n'ont malheureusement pas permis une étude statistique. La figure ci-après présente ces résultats :

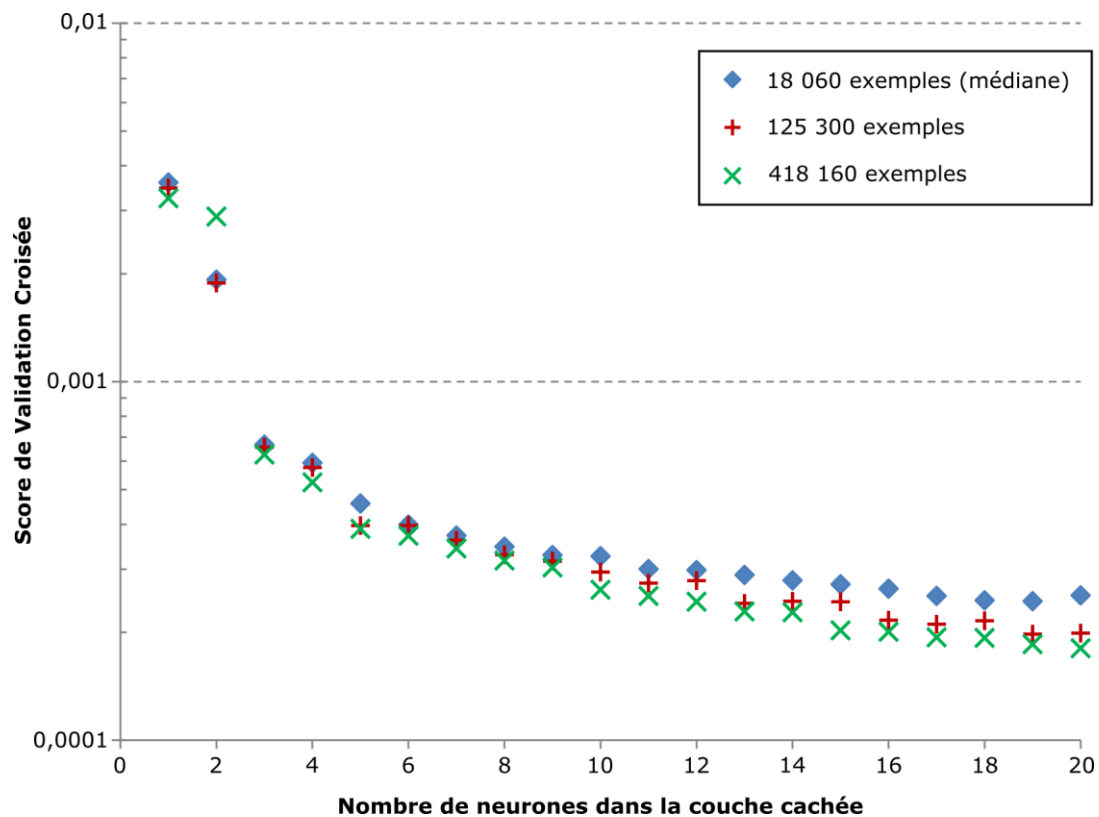


Figure III-41 : Scores de validation croisée pour des échantillonnages différents de la base de données

Les différents échantillonnages montrent une tendance similaire. L'augmentation du nombre d'exemples dans la base d'apprentissage est associée à une diminution du score de validation croisée dès que la complexité du modèle est suffisante (au-delà de 10 neurones dans la couche cachée). Cette augmentation de la qualité de l'apprentissage est de plus en plus importante avec l'augmentation du nombre de neurones dans la couche cachée. Cependant, selon l'initialisation, les valeurs du score de validation croisée pour la base la moins fournie sont proches du S_{VC} de la base moyenne (tableau ci-dessous). Néanmoins, d'autres initialisations auraient pu diminuer le score de validation croisée pour la base de données moyenne et faible.

Tableau III-6 : Base d'apprentissage et score de validation croisée minimum

Base d'apprentissage	Valeur du score de validation croisée minimum
18 060	$2,2 \times 10^{-4}$
125 300	$2,0 \times 10^{-4}$
418 160	$1,8 \times 10^{-4}$

Il est important de souligner que le gain en valeur du score de validation croisée est faible par rapport à l'augmentation du nombre d'exemples dans la base d'apprentissage. Comme, la durée d'apprentissage est directement reliée au nombre d'exemples, la figure ci-après représente les durées moyennes d'apprentissage pour différents échantillonnages :

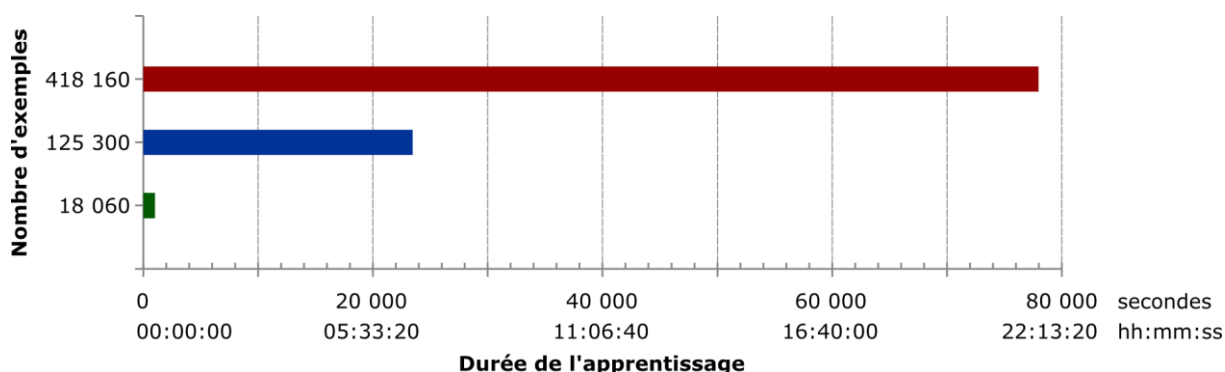


Figure III-42 : Durée d'apprentissage en fonction de l'échantillonnage

La durée d'apprentissage d'un réseau de neurones pour la base d'exemples la plus grande est supérieure à 21 heures. Pour la base moyenne, l'apprentissage est inférieur à 7 heures et pour une base plus petite, l'apprentissage est inférieur à 20 minutes.

Dans le cadre de l'utilisation du modèle sur des cas tests non-appris, deux apprentissages seront comparés, respectivement issus d'un échantillonnage à 418 160 exemples (ANN₁) et 18 060 exemples (ANN₂), 20 neurones dans la couche cachée et l'initialisation conduisant au plus faible score de validation croisée parmi 20 initialisations pour la base la moins fournie. Ces réseaux de neurones sont utilisés en combinaison avec l'automate cellulaire afin de modéliser les cas tests.

III-3.5. Evaluation du modèle CA-ANN

Afin d'obtenir une estimation globale de la performance du modèle, différents critères de performance (III-1.3.3) sont calculés sur l'ensemble des données fournies par l'automate cellulaire, à chaque pas de temps. Les données sont ainsi appairées en temps et en espace.

Les cas tests modélisés pour valider le modèle CA-ANN sont sélectionnés sur une gamme complète de valeurs. Neuf cas tests sont ainsi créés en prenant trois valeurs de vitesse d'écoulement : {3,2 m.s⁻¹ ; 10,2 m.s⁻¹ ; 18,8 m.s⁻¹} et trois valeurs de fraction massique initiale dans le rejet : {0,26 ; 0,5 ; 0,89}.

Dans un premier temps, l'évolution des différentes bouffées entre le modèle CA-ANN et le modèle CFD est évaluée, pour chacun des cas tests, à un même pas de temps : le 46^{ème}. Cela correspond à 90% de la masse encore contenue dans le domaine pour le cas test à la vitesse la plus élevée.

Le tableau ci-après présente les résultats concernant le facteur deux (FAC2) et le coefficient de détermination (R^2) pour les deux réseaux de neurones précédemment cités.

Tableau III-7 : Critères de performance – comparaison CA-ANN/CFD – valeurs pour deux réseaux de neurones différents (46 itérations)

	Vitesse (m.s ⁻¹)	Fraction massique	R^2		FAC2	
			ANN ₁	ANN ₂	ANN ₁	ANN ₂
1	3,2	0,26	0,99	0,98	0,87	0,83

	Vitesse (m.s ⁻¹)	Fraction massique	R ²		FAC2	
			ANN ₁	ANN ₂	ANN ₁	ANN ₂
2	3,2	0,5	0,99	0,99	0,88	0,85
3	3,2	0,89	0,97	0,99	0,88	0,86
4	10,2	0,26	0,97	0,97	0,80	0,72
5	10,2	0,5	0,97	0,98	0,78	0,74
6	10,2	0,89	0,96	0,99	0,77	0,76
7	18,8	0,5	0,92	0,96	0,74	0,70
8	18,8	0,26	0,87	0,94	0,74	0,72
9	18,8	0,89	0,93	0,97	0,73	0,69

Ces résultats globaux ont des valeurs de R^2 et de FAC2 supérieures respectivement à 0,8 et 0,7. Cela montre que le modèle CA-ANN effectue une bonne prédiction à 46 pas de temps. Néanmoins, les résultats sont inégaux en fonction des cas observés. Les dispersions associées aux faibles vitesses de vent sont mieux modélisées que celles associées aux fortes vitesses de vent. Cette tendance est observée aussi bien sur le coefficient de détermination que sur le facteur deux. Comme l'illustre la figure ci-après, il est difficile d'interpréter une tendance en ce qui concerne l'influence de la concentration maximale initiale dans le domaine.

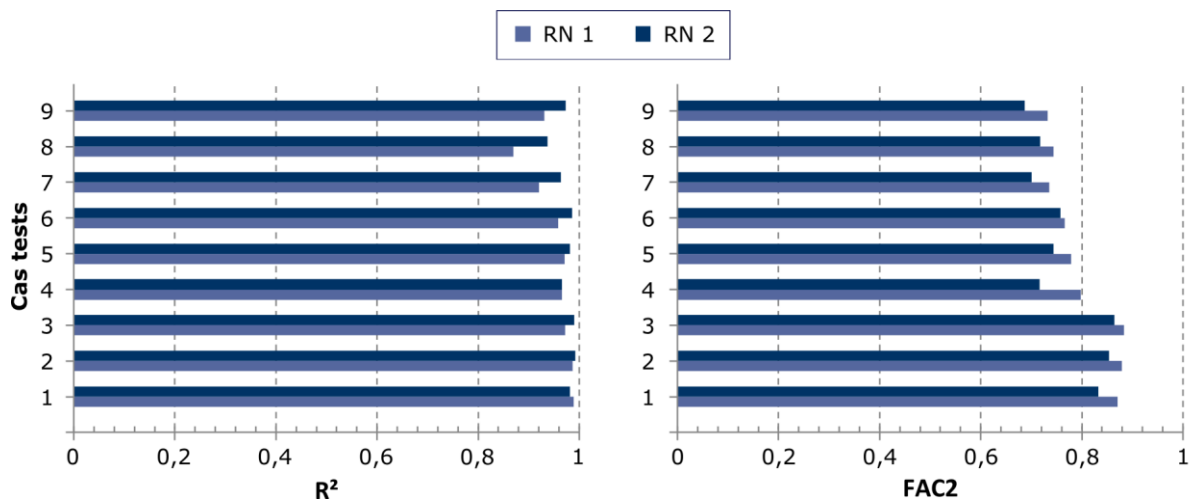


Figure III-43 : Valeurs du R^2 et du FAC2 pour les neuf cas tests à 46 pas de temps pour les deux règles de transition ANN testées.

En se référant au coefficient de détermination, le CA-ANN utilisant la règle de transition issue de l'ANN₂ modélise mieux le phénomène. En prenant en compte le FAC2, le résultat est inverse et c'est le CA-ANN avec règle de transition ANN₁ qui modélise le mieux le phénomène. Comme énoncé au III-1.3.3, il est nécessaire d'utiliser plusieurs indicateurs afin de pouvoir juger la performance d'un modèle. Les valeurs du biais fractionnel (FB) et de l'erreur quadratique moyenne normalisée (NMSE) permettent cette analyse. Un graphe permettant de visualiser à la fois l'erreur systématique et l'erreur aléatoire commise par le modèle est tracé ci-après :

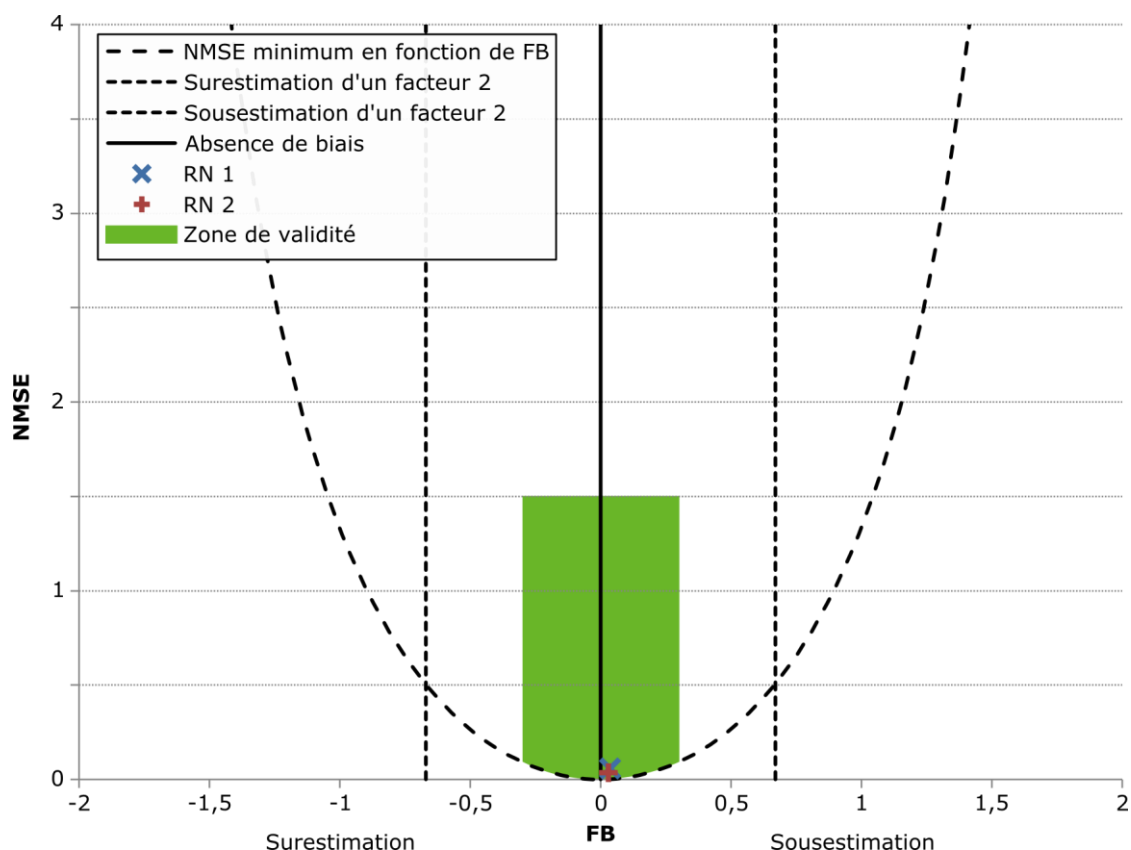


Figure III-44 : Performance des deux modèles selon les critères FB et NMSE.

La parabole représente la valeur minimale de l'erreur quadratique moyenne normalisée en fonction de la valeur du biais fractionnel puisque le NMSE comprend une part aléatoire et une part systématique. Un point situé au centre de l'axe des abscisses montre une erreur systématique nulle. Un modèle parfait se situera donc au minimum de la parabole. On rappelle qu'un FB négatif traduit une surestimation. Deux limites verticales représentent les valeurs de surestimation et de sous-estimation d'un facteur 2. Enfin, la zone verte représente la zone dans laquelle un modèle est considéré comme valable d'après Chang et Hanna (2004).

Sur ce graphique avec une vue générale, les deux modèles semblent se comporter correctement. Les erreurs systématiques et aléatoires sont faibles. Il est nécessaire d'effectuer un agrandissement d'échelle pour différencier ces modèles.

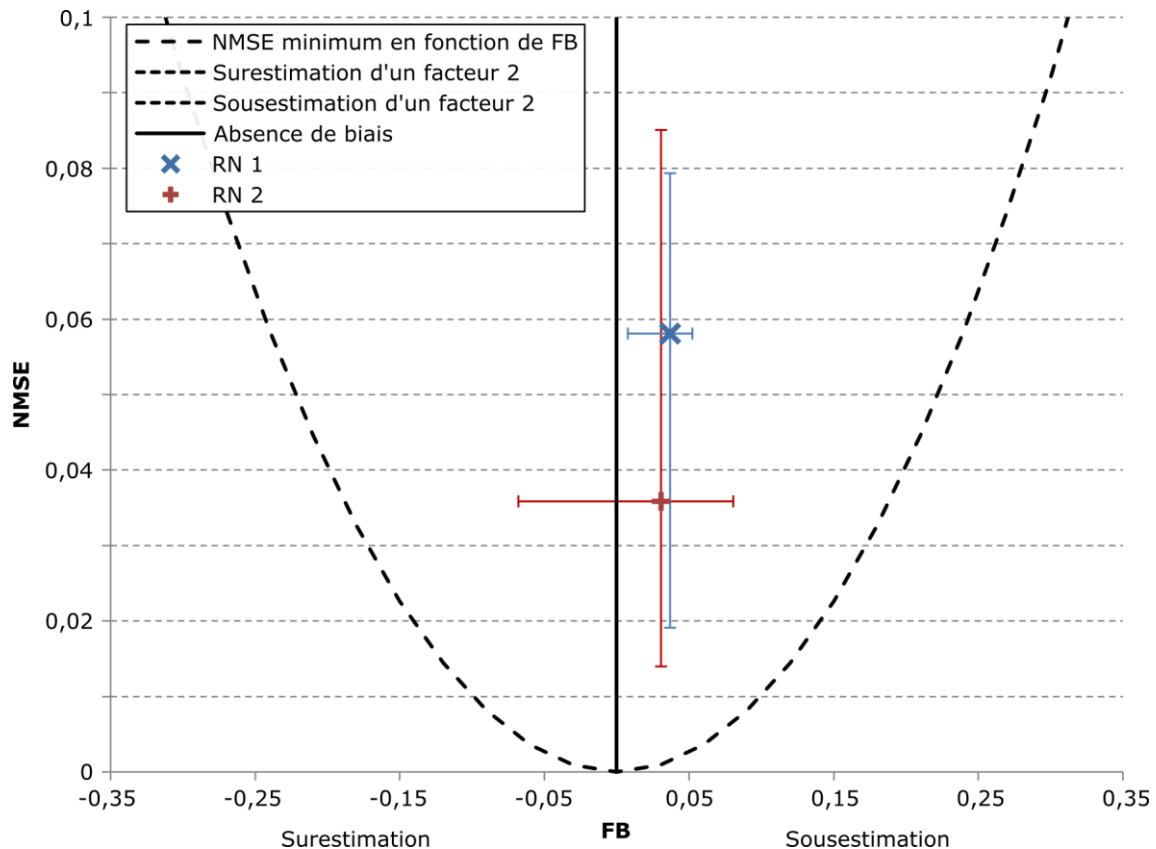


Figure III-45 : Comparaison des deux modèles selon FB et NMSE

Sur ce graphique figurent les valeurs médianes de FB et NMSE pour l'ensemble des neuf cas tests. Cette valeur médiane est en faveur du modèle CA-ANN avec l'ANN2. Néanmoins, on s'aperçoit que les valeurs minimales et maximales représentées par les barres d'erreur fluctuent plus, notamment en ce qui concerne l'erreur systématique. Alors que le CA-ANN₁ sous-estime les valeurs de concentration pour chaque exemple, le comportement du CA-ANN₂ surestime quelques cas.

Ces comparaisons sont faites à 46 pas de temps. Afin de suivre l'évolution de l'erreur commise par les modèles, on calcule les critères de performance pour 90 % de la masse restant dans le domaine de calcul pour chaque cas. Pour des vitesses de 3,2, 10,2 et 18,8 m.s⁻¹, cela correspond respectivement à 272, 85 et 46 itérations. Les valeurs du coefficient de détermination et du facteur 2 sont reportées dans le tableau ci-après :

Tableau III-8 : R^2 et FAC2 pour neuf cas modélisés, 90% de la masse restant dans le domaine de calcul

Vitesse (m.s ⁻¹)	Fraction massique	Pas de temps	R^2		FAC2	
			ANN ₁	ANN ₂	ANN ₁	ANN ₂
3,2	0,26	272	0,60	-0,05	0,71	0,64
3,2	0,5	272	0,78	0,64	0,75	0,71
3,2	0,89	272	0,21	0,86	0,59	0,74
10,2	0,26	85	0,92	0,91	0,72	0,64

Vitesse (m.s ⁻¹)	Fraction massique	Pas de temps	R ²		FAC2	
			ANN ₁	ANN ₂	ANN ₁	ANN ₂
10,2	0,5	85	0,94	0,96	0,70	0,67
10,2	0,89	85	0,93	0,97	0,70	0,69
18,8	0,5	46	0,92	0,96	0,74	0,70
18,8	0,26	46	0,87	0,94	0,74	0,72
18,8	0,89	46	0,93	0,97	0,73	0,69

Une diminution des valeurs de R^2 avec l'augmentation du nombre de pas de temps est observée. Une valeur négative est atteinte par le CA-ANN₂ pour un cas à vitesse faible et donc un nombre de pas de temps plus grand. Le CA-ANN₁ et CA-ANN₂ ont une erreur qui augmente avec le nombre de pas de temps. En ce qui concerne le FAC2, les valeurs sont également en baisse. Celle-ci est moins importante que pour le coefficient de détermination. Ceci suggère qu'il y a une propagation d'erreur qui s'effectue lors de chaque nouvelle itération du CA-ANN, vraisemblablement due au fait que l'architecture du modèle effectuant la règle de transition est récurrente. Le CA-ANN₁ semble, de manière globale, mieux prédire l'évolution de la concentration d'après le FAC2. Le graphique ci-après montre cette tendance :

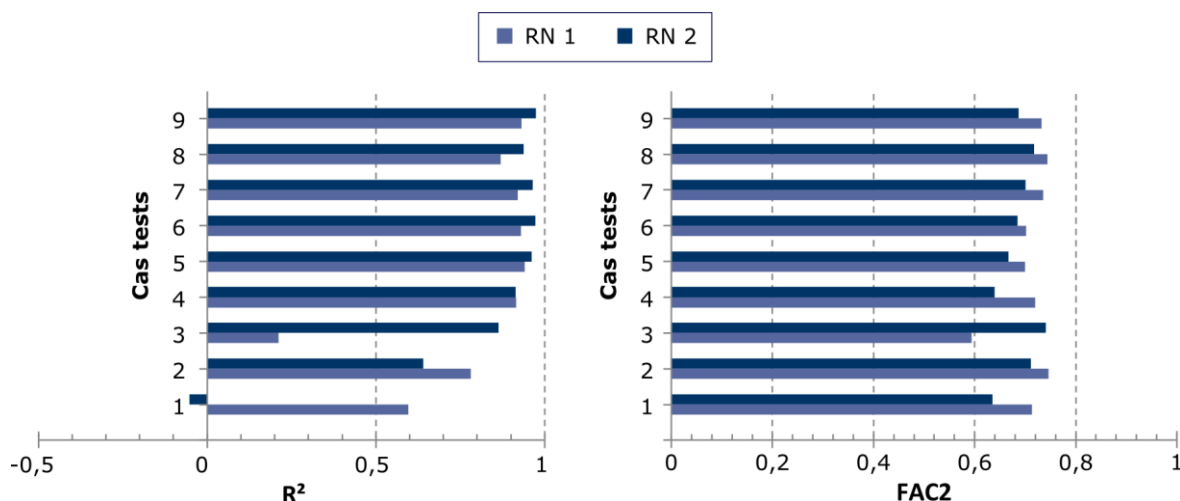


Figure III-46 : Valeurs du R^2 et du FAC2 pour les 9 cas tests à 90% de la masse restant dans le domaine.

Pour analyser la valeur négative du R^2 pour le cas à vitesse et fraction massique initiale faible, on trace la distribution spatiale de concentrations représentant la réalité CFD et les iso-contours de la modélisation CA-ANN :

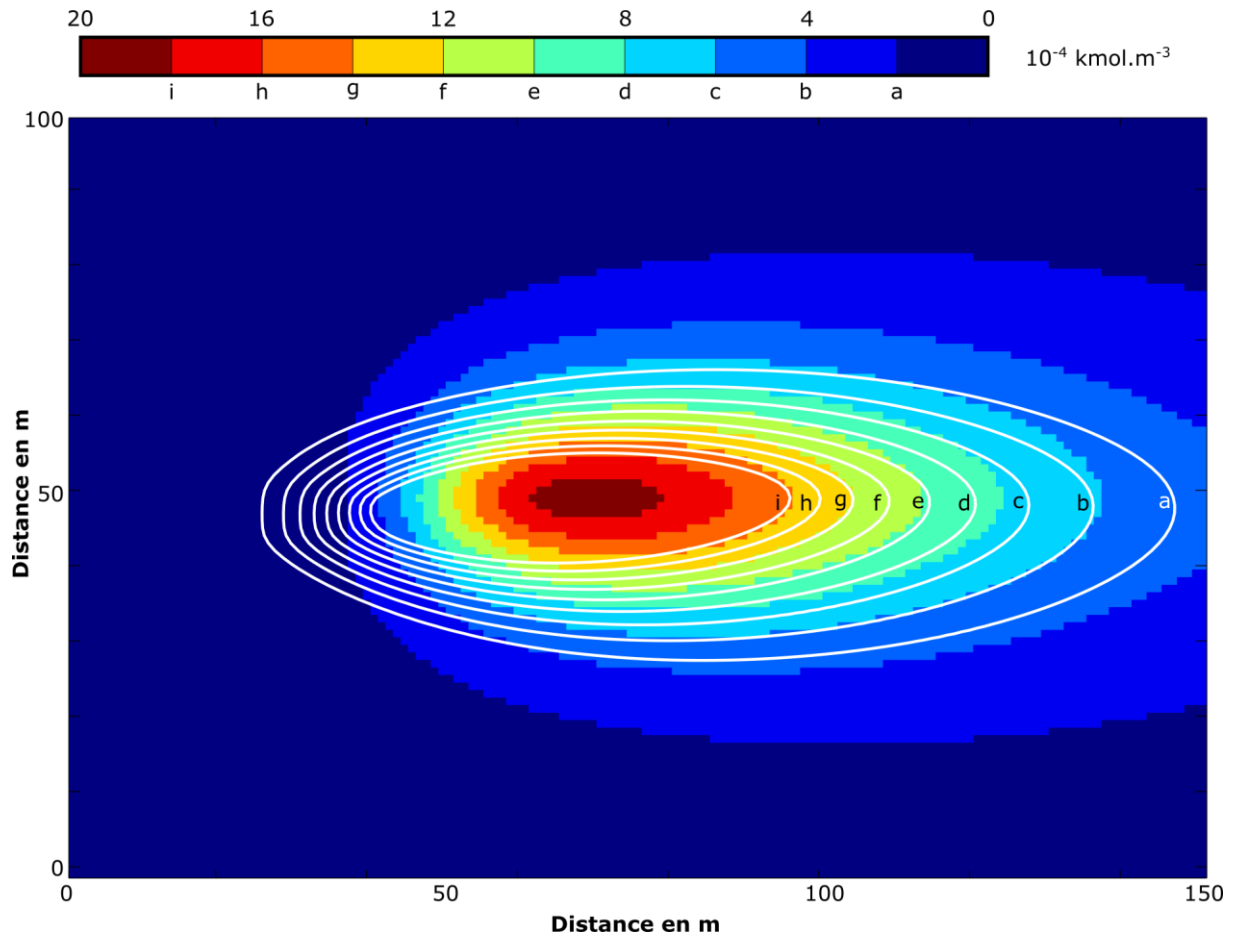


Figure III-47 : Comparaison des isosurfaces de concentrations CFD et isocontours de concentrations CA-ANN₂ pour le cas 1.

Il existe un décalage dans les zones de concentrations maximales, dans l'axe de l'écoulement, le modèle CA-ANN semble être en retard sur la réalité CFD. D'autre part une diffusion plus faible est observée dans la largeur. En ce qui concerne les valeurs maximales de concentrations, des divergences importantes apparaissent également : $0,0020 \text{ kmol.m}^{-3}$ pour la réalité CFD et $0,0035 \text{ kmol.m}^{-3}$ pour le modèle CA-ANN, restant dans un intervalle de facteur deux. En traçant la dispersion des points entre modèle et simulation, une tendance à la surestimation des concentrations fortes est identifiée. Par contre les concentrations faibles sont sous-estimées, ce qui peut expliquer la déformation appliquée au nuage par le CA-ANN. Cette dispersion peut être tracée en appairant les points en espace ou alors en réalisant un tri des valeurs les plus élevées aux plus faibles (diagramme quantile-quantile). Dans les deux cas, une tendance à la surestimation des concentrations fortes et une sous-estimation des concentrations faibles est visualisée. Le cas des valeurs faibles est particulier. Des concentrations nulles sont observées dans le CA-ANN alors que la réalité CFD en possède beaucoup moins.

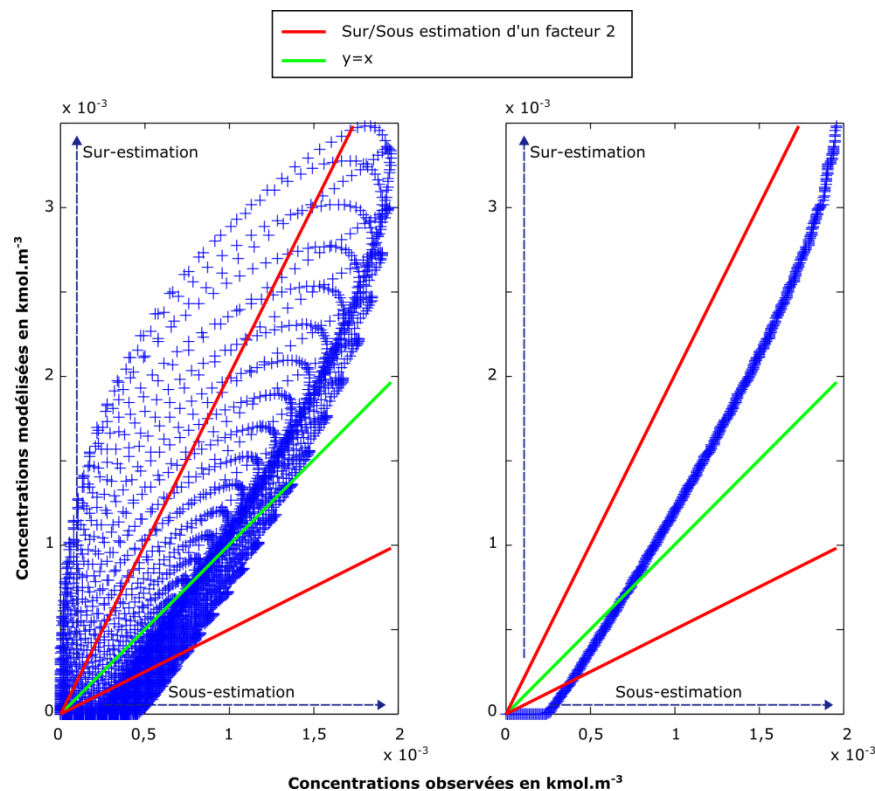


Figure III-48 : Dispersion des points observés (abscisses) et modélisés (ordonnées). Les points sont appairés en espace dans la figure de gauche (diagramme de dispersion). Dans la figure de droite, un tri des valeurs de concentration est effectué au préalable (diagramme quantile-quantile).

En analysant les valeurs de biais et d'erreurs aléatoires, on note des disparités dans les performances de prédiction.

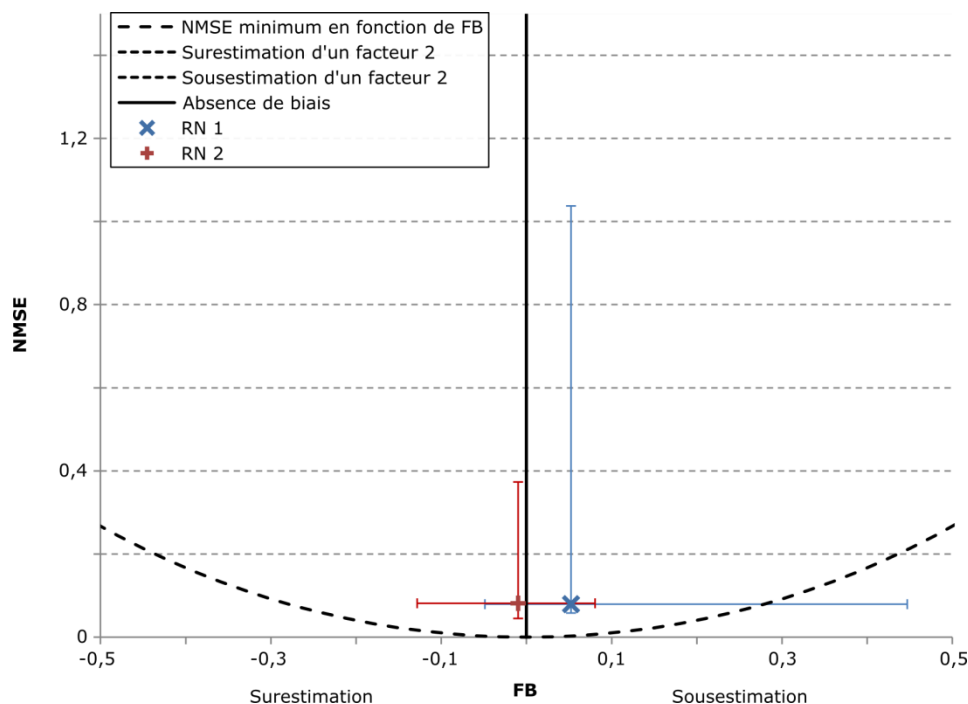


Figure III-49 : Erreurs aléatoires et systématiques pour les deux CA-ANN et pour 90 % de la masse restante dans le système

Dans cette situation, le CA-ANN₂ prédit mieux que le CA-ANN₁. L'erreur aléatoire est moins importante. L'erreur systématique est également plus proche de la valeur nulle.

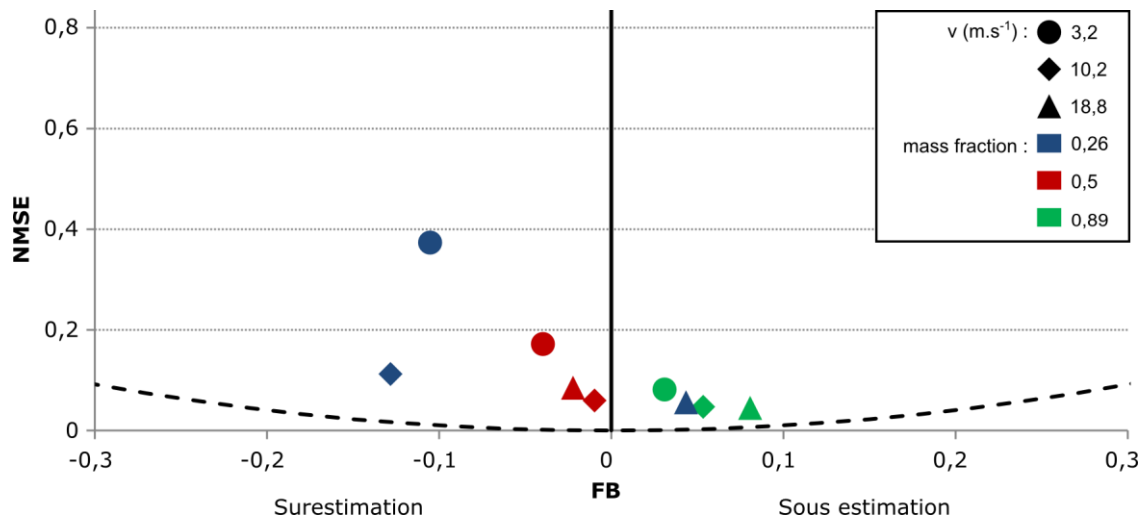


Figure III-50 : Différentes erreurs selon les cas testés pour le CA-ANN₂.

Par ailleurs, les fluctuations d'un cas à l'autre sont moins prononcées que lorsque le temps de simulation n'était pas ajusté en fonction de la vitesse de vent. Une diminution de la vitesse (et donc une augmentation du nombre de pas de temps) implique une augmentation de l'erreur. Comme nous l'avons souligné précédemment, une erreur se propage au cours des itérations successives. Une augmentation de la part aléatoire de l'erreur de manière plus rapide que la part de l'erreur systématique est observée sur le cas favorable :

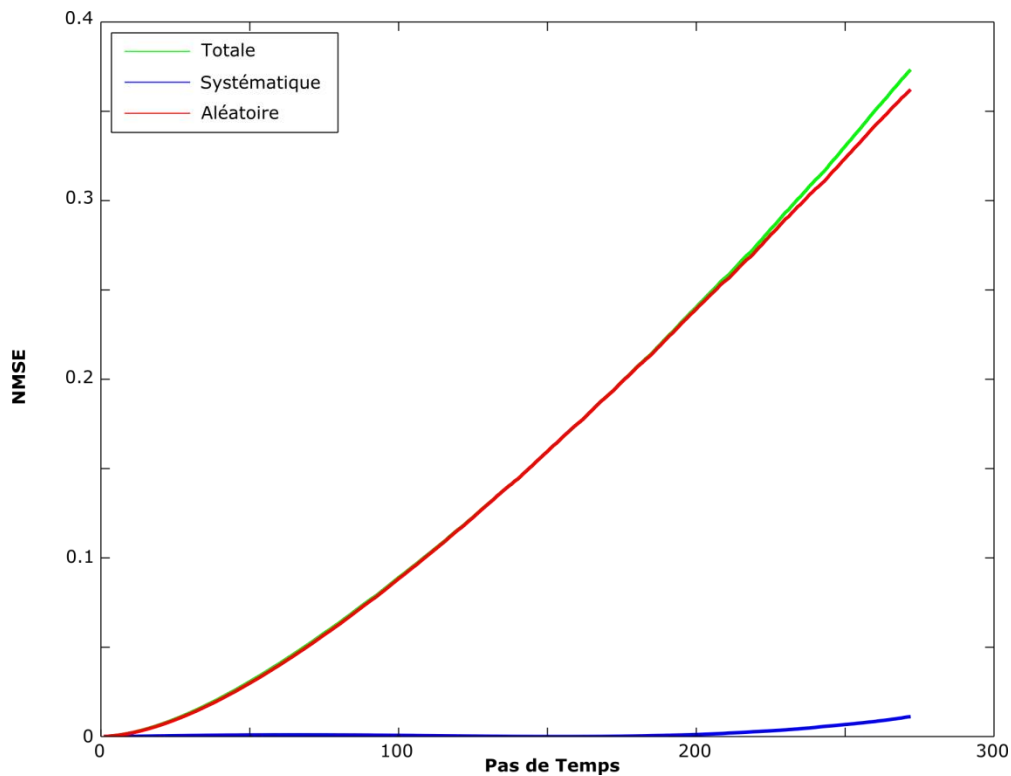


Figure III-51 : Propagation de l'erreur aléatoire, systématique et totale au cours des pas de temps pour le cas favorable.

Il s'agit ici du cas le mieux prédit. Néanmoins, cette tendance se vérifie sur chacun des cas. La prépondérance de cette erreur aléatoire sur l'erreur systématique peut être interprétée comme une illustration du dilemme biais variance. En effet, compte tenu des données disponibles, une erreur systématique faible signifie que le biais est minimum. L'erreur totale est portée par l'erreur systématique qui est proportionnelle à la variance.

Dans ce travail, l'évolution de la concentration s'est effectuée sur un certain nombre de pas de temps sans qu'aucun bilan de matière n'ait été effectué. Il est possible d'évaluer ce fonctionnement sur la quantité de matière présente dans le système (figure ci-après) :

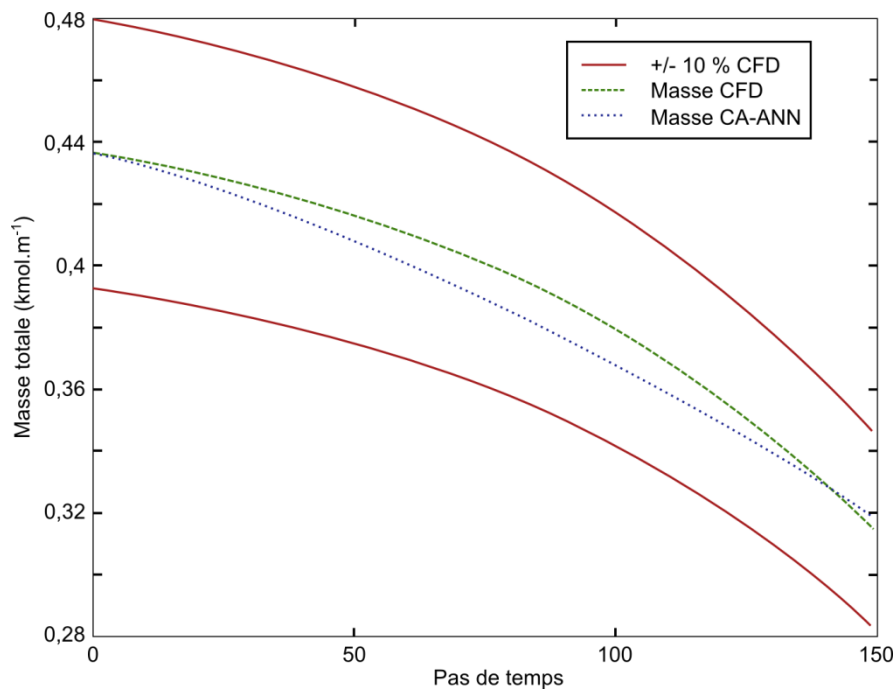


Figure III-52 : Evolution de la masse restante dans le domaine pour la simulation CFD et la simulation CA-ANN₂.

La courbe concernant le modèle CA-ANN suit la courbe CFD avec une erreur inférieure à 10 %. Le modèle développé, bien qu'aucun bilan de matière ne soit effectué, traduit correctement le comportement du nuage à travers les pas de temps.

III-3.6. Gain en temps de calcul

Comme vu précédemment au III-2, le calcul d'une sortie par réseau de neurones est extrêmement rapide. Dans le cas de cette méthode, il y a autant de passages dans le réseau de neurones que de cellules d'automates déployées en parallèle. La présente configuration donne 15 251 cellules. Par ailleurs, ce calcul est effectué à chaque pas de temps. Ce dernier est fixé au préalable par le respect de la condition CFL. Enfin, il est nécessaire d'intégrer dans ce calcul la mise en forme des données : acquisition des valeurs et calcul des dérivées discrètes des différents termes.

Les modèles CFD ont une formulation implicite de la discrétisation temporelle. Ainsi, cela leur permet de ne pas calculer l'évolution du nuage à chaque pas de temps. Comparer les modèles CFD et CA-ANN en fonction du temps de calcul peut être donc effectué de deux manières :

- Comparer le temps de calcul pour le passage d'un temps à un autre séparé de l'intervalle utilisé par l'automate cellulaire

- Comparer le temps de calcul pour une durée de transport déterminée au préalable.

La première comparaison est largement favorable au CA-ANN. Le calcul CFD s'effectue environ en une minute pour un pas de temps. Le calcul des états des cellules de l'automate cellulaire est extrêmement rapide, inférieur à la demi-seconde sur le même processeur, soit un gain en temps de calcul de plus de 120. Cette durée comprend la mise en forme des données ainsi que la règle de transition. Dans ce cas là, l'utilisation de CA-ANN est extrêmement intéressante pour une situation opérationnelle en adaptant les pas spatiaux et donc temporels.

Pour le calcul de la deuxième méthode, le gain en temps de calcul est moindre, puisque le CA-ANN doit calculer chaque pas de temps, au contraire d'un modèle CFD à discrétisation implicite. Ainsi, pour le même domaine et en utilisant la même station de travail, le modèle CA-ANN est 1,5 fois plus rapide que le calcul CFD.

III-3.7. Conclusion

La prévision de la dispersion atmosphérique en champ libre par l'utilisation conjointe d'un automate cellulaire et d'un réseau de neurones effectuant le calcul de la règle de transition a été réalisée et évaluée. Une des difficultés majeures de cette méthode est la propagation puis l'amplification des erreurs au cours du temps. Ces erreurs peuvent conduire, dans certains cas et pour un nombre de pas de temps important (supérieurs à 100), vers ce qui pourrait être une divergence du modèle si l'automate cellulaire était plus étendu. Néanmoins, l'application de cette méthode à des cas tests a démontré son efficacité dans le domaine limité de nos expérimentations.

L'utilisation de cette méthode permet une prise en compte implicite de la turbulence. Effectivement, tous les termes de l'équation d'advection diffusion ne sont pas explicitement utilisés en tant qu'entrée pour le réseau de neurones. En particulier, le coefficient de diffusion turbulente permet de quantifier l'intensité de la diffusion dans le phénomène de transport turbulent de substance. De plus, ce coefficient est considéré comme constant selon un plan horizontal dans le cas d'une couche limite atmosphérique. Dans la suite de ce travail, et dans la perspective de modéliser la dispersion atmosphérique en milieu obstrué, il est nécessaire de prendre en compte la perturbation d'un écoulement par un obstacle et ainsi les modifications induites dans les valeurs du coefficient de diffusion turbulente et du champ de vent. C'est le sujet du chapitre suivant.

III-4. Modélisation de la dispersion atmosphérique autour d'un obstacle cylindrique

Il s'agit dans ce chapitre de déterminer la méthode permettant d'évaluer les vitesses d'écoulement autour d'un obstacle cylindrique. Le coefficient de diffusion turbulente sera également déterminé dans un cas de stabilité atmosphérique neutre. La méthode utilise les réseaux de neurones seuls pour le calcul de l'écoulement. Afin de simuler la dispersion atmosphérique, la discrétisation de l'équation d'advection diffusion par la méthode des différences finies est utilisée.

III-4.1. Objectifs

Comme abordé précédemment, la difficulté d'utiliser un réseau de neurones statique en tant que règle de transition est l'augmentation de l'erreur à chaque pas de temps. Par ailleurs, l'utilisation d'un automate cellulaire limite soit la résolution spatiale, soit la résolution temporelle à travers le nombre de Courant. Des modifications peuvent être effectuées afin de passer outre ces limitations en ce qui concerne le champ libre. Dans ce chapitre, la dispersion atmosphérique autour d'un obstacle pose des problèmes liés à la création d'une base de données suffisamment grande pour être représentative de l'ensemble des situations potentielles. Par ailleurs, la turbulence est prise en compte de manière implicite dans la méthode précédente. Il est néanmoins possible d'utiliser les réseaux de neurones pour déterminer le coefficient de diffusion turbulente, qui traduit les mécanismes de diffusion dans le transport de matières.

Aussi, la méthode développée dans ce chapitre diffère de la méthode précédente par la séparation entre calcul des paramètres influants sur la dispersion et le calcul de la dispersion atmosphérique elle-même. En termes d'outils mis en place, les réseaux de neurones vont permettre la prévision du champ de vent et du coefficient de diffusion turbulente. La résolution de l'équation d'advection diffusion peut être effectuée par différentes méthodes numériques. Nous prendrons ici l'exemple des différences finies.

La forme de l'équation d'advection diffusion est rappelée :

$$\frac{\partial c}{\partial t} + u_i \frac{\partial c}{\partial x_j} = \frac{\partial}{\partial x_j} \left(D_t \cdot \frac{\partial c}{\partial x_j} \right) + S_i + R_c \quad (III-15)$$

La connaissance des vitesses d'écoulement u_i et du coefficient de diffusion turbulente D_t permettent de calculer le transport de substance, en l'absence de réaction chimique dans le milieu. Dans les codes CFD, ce champ de vent est calculé à partir du couplage entre équations de conservation de la masse et équations de conservation de la quantité de mouvement. De plus, la modélisation de la turbulence vient compléter le système lorsque le régime d'écoulement l'oblige. Le coefficient de diffusion turbulente est directement relié aux valeurs de l'énergie cinétique turbulente et de son taux de dissipation. Dans un écoulement perturbé par un obstacle, la turbulence n'est pas homogène selon un plan horizontal. L'hypothèse postulée au chapitre précédent n'est plus valable. Afin de résoudre numériquement l'ADE, il est donc également nécessaire de déterminer ce coefficient.

La prise en compte de la dispersion atmosphérique pourra s'effectuer de manière stationnaire ou instationnaire, simplement en modifiant les termes de résolution de l'ADE.

III-4.2. De la base de données à la base d'apprentissage

L'utilisation des réseaux de neurones implique la création d'une base de données représentative de la situation à prédire.

III-4.2.1 Simulation CFD

Un modèle $k-\epsilon$ standard est utilisé afin de générer une base de données d'écoulements horizontaux bidimensionnels autour d'un cylindre vertical à l'altitude $z = 2 \text{ m}$. Cet écoulement est effectué dans des conditions de stabilité neutre. La turbulence dans la couche limite atmosphérique peut donc être représentée par la théorie décrite au III-1.2. Cela va permettre de déterminer les conditions d'entrée dans le domaine CFD rectangulaire. Ces dimensions varient en fonction des dimensions de l'obstacle. Effectivement, la turbulence est largement modifiée dans le sillage de l'obstacle et il est nécessaire de bénéficier d'une distance suffisamment importante en aval de celui-ci, pour caractériser la totalité de la perturbation. Les conditions d'entrée du domaine sont donc représentées par des valeurs constantes d'énergie cinétique turbulente k et de son taux de dissipation ϵ . La vitesse d'entrée dans le domaine est homogène, de direction normale à la limite. Des conditions de symétries sur les bords du domaine ainsi qu'une pression de sortie équivalente à la pression atmosphérique seront utilisées. Le périmètre du cylindre est représenté par une condition de type « wall » ce qui implique notamment des valeurs de vitesse d'écoulement nulles au contact à la paroi.

Le schéma ci-après représente la modélisation CFD :

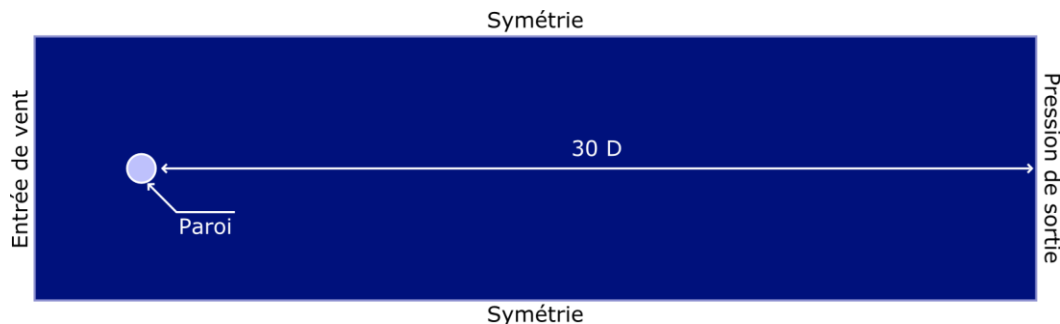


Figure III-53 : Schématisation du cas Fluent

Il est nécessaire d'effectuer un maillage fin sur ce type de simulation, de façon à permettre le calcul de la turbulence, notamment à proximité des zones où les gradients des différentes grandeurs physiques vont être importants. Il n'est pas possible de conserver uniquement des mailles carrées ou rectangles dans ce type de configuration. Aussi, une légère distorsion des angles des mailles est nécessaire. Les valeurs de skewness relevées garantissent cependant la convergence d'une telle simulation. Le maillage proposé initialement est donc composé de 112 000 nœuds, répartis de manière progressive à proximité de l'obstacle comme illustré dans la figure ci-après.

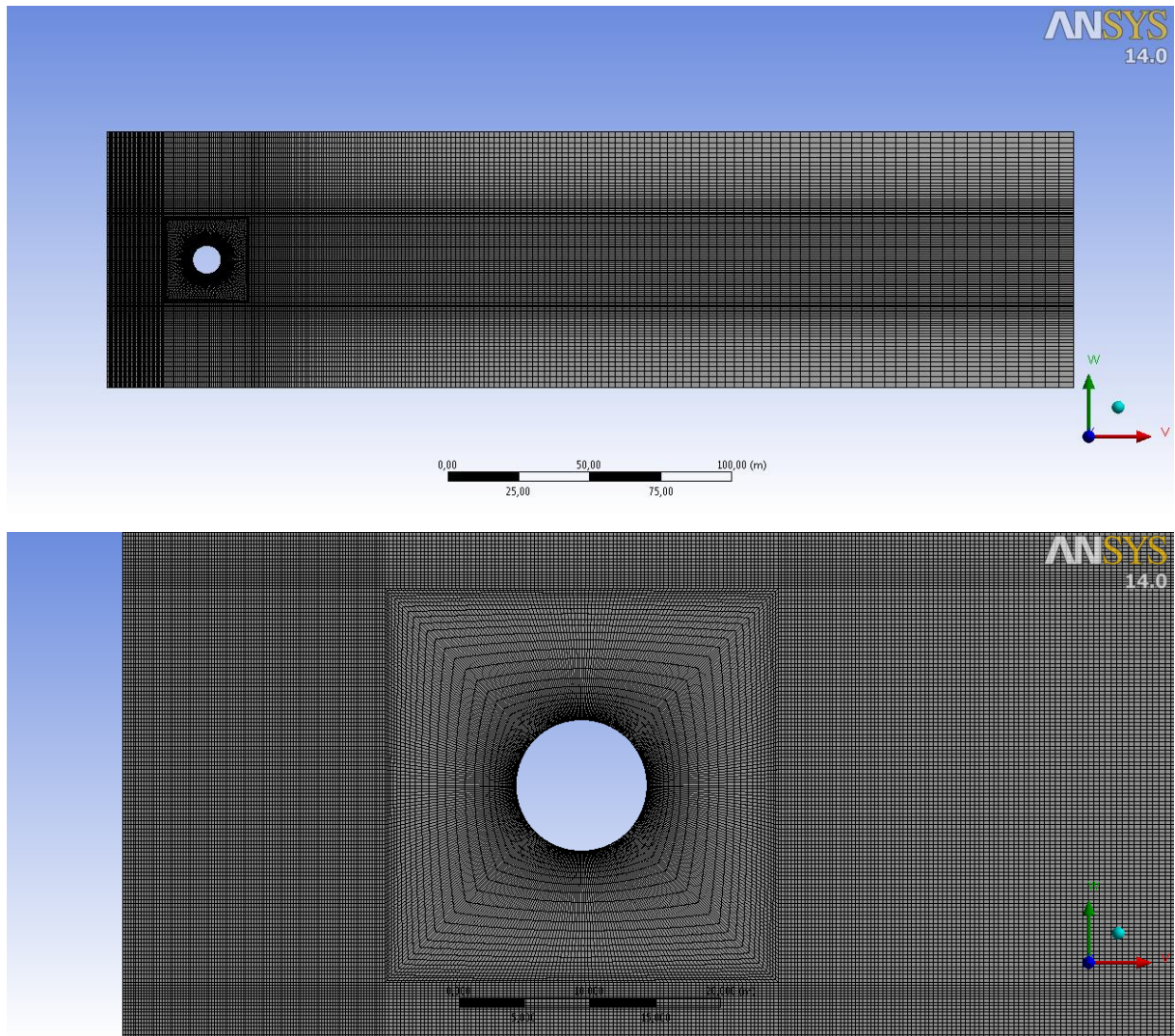


Figure III-54 : Maillage du cas CFD

Ce maillage initial permet le calcul de la turbulence pour une majeure partie des simulations composant la base de données. Néanmoins, des limitations peuvent apparaître lorsque le domaine de calcul s'étire avec l'augmentation du diamètre du cylindre. Aussi, les mailles initiales à faibles dimensions au niveau des hauts gradients deviennent beaucoup plus grandes. Il est nécessaire d'augmenter la taille du maillage pour les diamètres importants. Comme précédemment, la simulation CFD est stoppée lorsque le résiduel de convergence est stabilisé et inférieur à 10^{-6} pour l'ensemble des équations.

Comme vu au I-2.4, l'obstacle perturbe l'écoulement en termes de vitesse et de turbulence. La figure ci-après présente les résultats d'un calcul CFD autour d'un obstacle de 52 m de diamètre et une vitesse incidente de 10 m.s^{-1} :

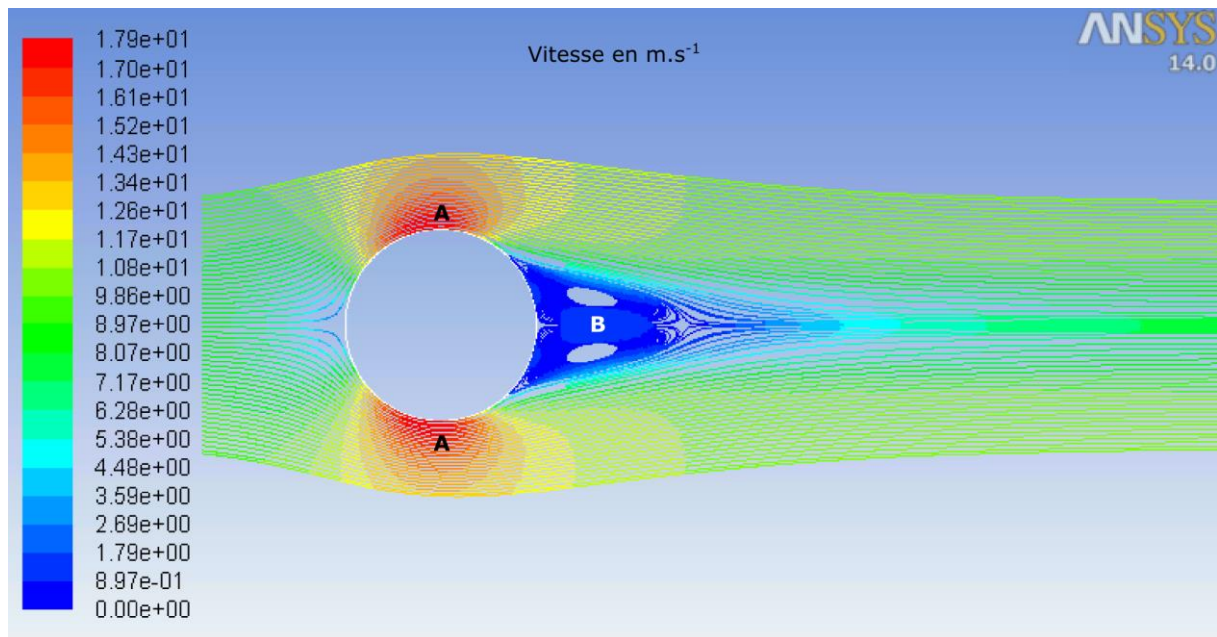


Figure III-55 : Lignes de courant obtenues autour d'un obstacle cylindrique de 52 m de diamètre pour une vitesse d'écoulement en entrée de 10 m.s^{-1} - Simulation par modèle $k - \epsilon$ standard.

Une zone de recirculation est identifiée en aval de l'obstacle. C'est une zone où une substance relâchée est susceptible de stagner. Par ailleurs, une augmentation de la vitesse est observée sur les côtés du cylindre (A) et des vitesses extrêmement faibles dans la zone de recirculation (B). La figure III-56 montre les perturbations dans le sillage de l'obstacle. L'impact de celui-ci sur l'écoulement se traduit par une augmentation de turbulence illustrée par les contours spatiaux du coefficient de diffusion turbulente.

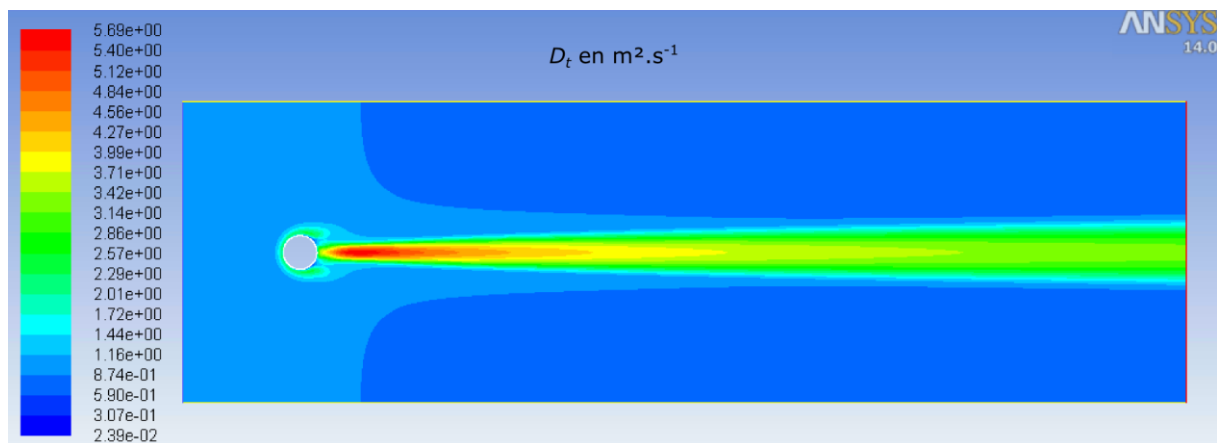


Figure III-56 : Contours du coefficient de diffusion turbulente autour d'un obstacle cylindrique de 52 m de diamètre pour une vitesse d'écoulement en entrée de 10 m.s^{-1} - Simulation par un modèle $k - \epsilon$ standard.

Il est important de noter sur la figure précédente que la zone d'influence de l'obstacle est supérieure au domaine de calcul. Néanmoins, l'étude de la perturbation de l'obstacle sera limitée dans une distance atteignant 30 fois le diamètre de l'obstacle car la prise en compte de la totalité de la zone de perturbation conduirait à des simulations CFD expansives en temps de calcul. D'autre part, la zone d'intérêt particulier se situe dans la proximité de l'obstacle.

La base de données est créée en faisant varier les dimensions de l'obstacle ainsi que la vitesse initiale en entrée à partir de cette modélisation CFD.

III-4.2.2 Représentativité de la base de données

Comme évoqué précédemment, des variations des différents paramètres permettent d'obtenir une base de données représentative des cas à envisager sur un site industriel. Les paramètres importants ainsi que leurs variations sont donc énoncés dans le tableau suivant :

Tableau III-9 : Variation des paramètres dans la base de données d'écoulements autour d'un obstacle

Paramètre	Unité	Intervalle	Pas
Vitesse initiale	m.s ⁻¹	[2 ; 10]	1
Diamètre du cylindre	m	[10 ; 52]	6
Reynolds	-	[1,3 × 10 ⁶ ; 3,5 × 10 ⁷]	-

Les valeurs de Reynolds associées à ces différents cas sont notées dans le tableau précédent. Les valeurs relevées situent les simulations dans le domaine turbulent. Cela représente 72 simulations CFD différentes. Pour chaque simulation, un minima de 112 000 valeurs de vitesse selon x et y et de coefficient de diffusion turbulente sont sauvegardées. La base de données CFD possèdent donc plus de 8 064 000 valeurs de vitesse et de coefficient de diffusion turbulente (figure ci-après) :

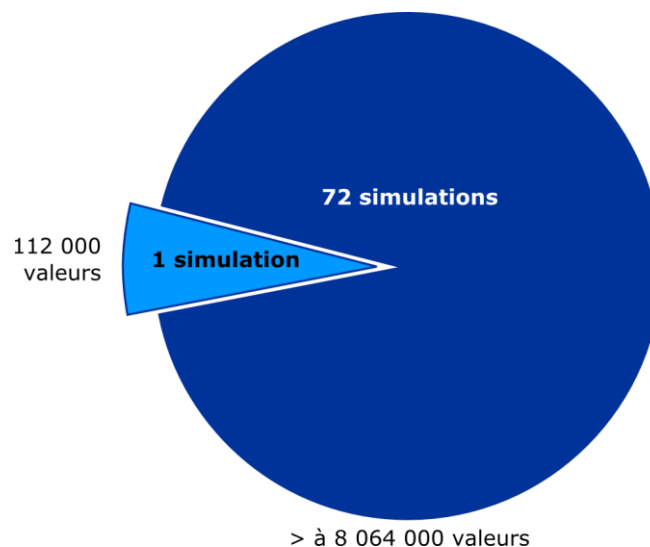


Figure III-57 : Représentation schématique du contenu de la base de données CFD des simulations d'écoulements autour d'obstacles

La symétrie du domaine permet de n'utiliser que les points issus de la partie supérieure de ce dernier. Pour effectuer l'apprentissage du réseau de neurones, il est nécessaire de réduire cette base de données. L'échantillonnage est effectué en prélevant, pour chaque simulation, E_c exemples dans I_c intervalles de vitesse/coefficient de diffusion turbulente comme illustré sur la figure ci-après :

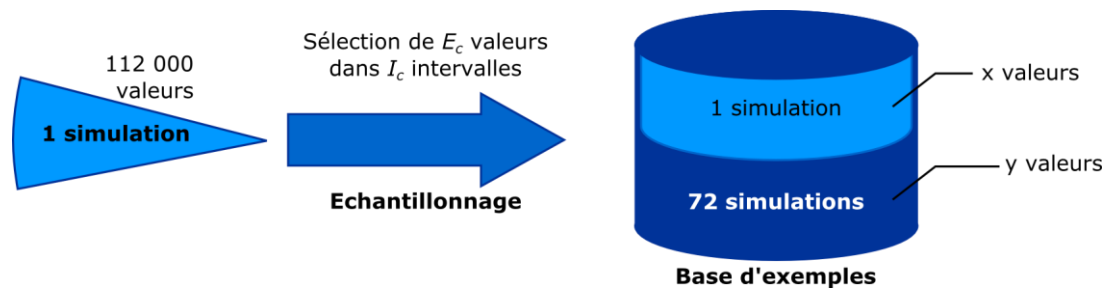


Figure III-58 : Echantillonnage de la base de données CFD des simulations d'écoulements autour d'obstacles pour créer la base d'exemples du réseau de neurones.

Ces points sont répartis de manière irrégulière dans le domaine. Les résultats du chapitre sur la dispersion stationnaire en champ libre (III-2) sont utilisés pour caractériser les entrées du réseau de neurones correspondants à la localisation du point à déterminer : les coordonnées polaires du point à estimer sont utilisées en tant que variables d'entrée.

III-4.3. Structure du réseau de neurones

III-4.3.1 Détermination des variables d'entrée

Les paramètres influents sur la détermination de la vitesse selon x et y et le coefficient de diffusion turbulente sont de trois types :

- ▶ Conditions atmosphériques locales. La vitesse d'écoulement dans le domaine est évidemment à inclure dans les variables d'entrée. La turbulence atmosphérique est à prendre en compte. Sa prise en compte est effectuée lors de la détermination du coefficient de diffusion turbulente.
- ▶ Obstacle. Dans le cas du cylindre, le diamètre ou le rayon permettent de caractériser au mieux la géométrie de l'obstacle.
- ▶ La localisation. Comme vu au chapitre de la dispersion atmosphérique continue, il est nécessaire d'évaluer la vitesse et le coefficient de diffusion turbulente en chaque point de l'espace.

Le nombre de Reynolds caractérise la turbulence dans un milieu en fonction d'une dimension caractéristique. Dans notre travail, il est possible d'utiliser une variable proportionnelle au Reynolds en tant qu'entrée du réseau de neurones. Cela permet d'adimensionnaliser les distances en entrées. Par ailleurs, la sortie peut également être adimensionnée. Dans le cas de la prédiction de vitesses d'écoulement, le rapport entre la sortie désirée et la vitesse en entrée sera utilisé alors que pour la prédiction du coefficient de diffusion turbulente, la variable sera utilisée sans modification. Les variables suivantes :

Tableau III-10 : Variables d'entrée et cible des réseaux de neurones

Entrée	Angle	Distance adimensionnée	Nombre caractéristique de l'écoulement	Cible		
				U_x/U_{ini}	U_y/U_{ini}	D_t
Intervalle initial	$[0 ; \pi]$	$[1 ; 60]$	$[10 ; 260]$	$[-0,5 ; 1,8]$	$[-0,95 ; 0,95]$	$[0 ; 7]$
Prétraitement des données	$[-0,9 ; 0,9]$					

Il est donc nécessaire d'utiliser trois réseaux de neurones différents afin de déterminer le champ des vitesses selon x , le champ des vitesses selon y et les variations du coefficient de diffusion turbulente.

Les données seront normalisées avant apprentissage et utilisation par le réseau de neurones afin d'éviter la saturation des fonctions de transition.

III-4.3.2 Architecture du réseau de neurones

Dans la suite de ce travail, un réseau de neurones de type perceptron multicouche sera utilisé :

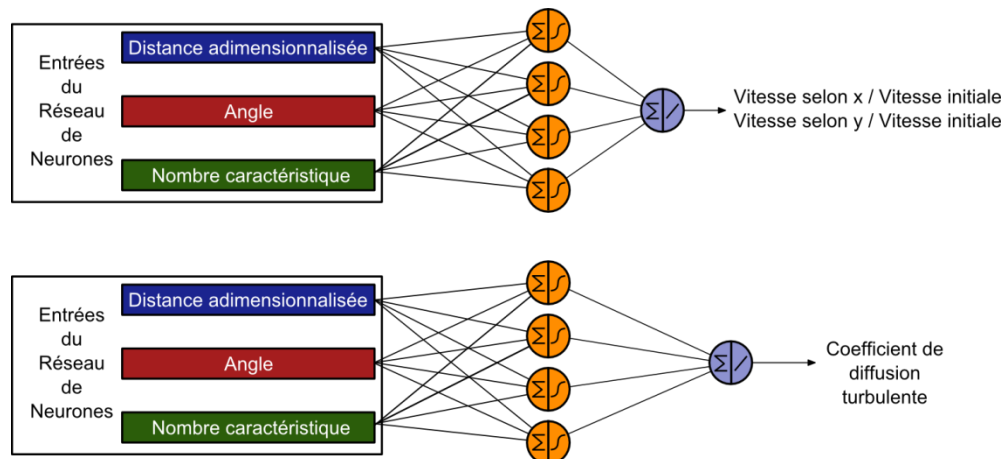


Figure III-59 : Structure des réseaux de neurones utilisée dans la détermination des champs des vitesses selon x et y et des variations spatiales du coefficient de diffusion turbulente.

L'évaluation de l'influence du nombre de neurones dans la couche cachée et l'initialisation de l'échantillonnage est étudiée dans la suite du manuscrit de façon à sélectionner le meilleur modèle possible.

III-4.4. Optimisation et sélection des modèles

III-4.4.1 Influence des paramètres de l'apprentissage

Les deux chapitres précédents ont permis de mettre en avant des particularités lors de l'apprentissage d'un réseau de neurones à partir de données issues d'équations représentant la physique d'un phénomène. L'augmentation du nombre de neurones dans la couche cachée permet un meilleur apprentissage, mais il est difficile d'estimer la limite pour laquelle il n'est plus nécessaire de rajouter des neurones. Cependant, il a été convenu précédemment que le nombre de neurones dans cette couche ne doit pas excéder 20. Effectivement, une augmentation du nombre de neurones cachés au-delà de cette valeur impliquera des difficultés à généraliser lorsque ce modèle sera utilisé sur des données réelles (comportant du bruit), du fait du dilemme biais-variance. Par ailleurs, il a été démontré que l'initialisation des paramètres influe sur la qualité de l'apprentissage. Il est nécessaire d'effectuer plusieurs tentatives avant de sélectionner le modèle sub-optimal. Enfin, l'échantillonnage de la base de données en base d'exemples doit être étudié pour conclure sur le nombre suffisant d'exemples dans la base d'apprentissage.

i. Influence de l'initialisation

L'influence de l'initialisation est testée sur chaque variable de sortie en effectuant 20 initialisations pour un même échantillonnage avec un perceptron à une couche cachée et 20 neurones dans la couche cachée. Pour chaque apprentissage la valeur de l'EQMT est calculée. La figure ci-après illustre les résultats obtenus :

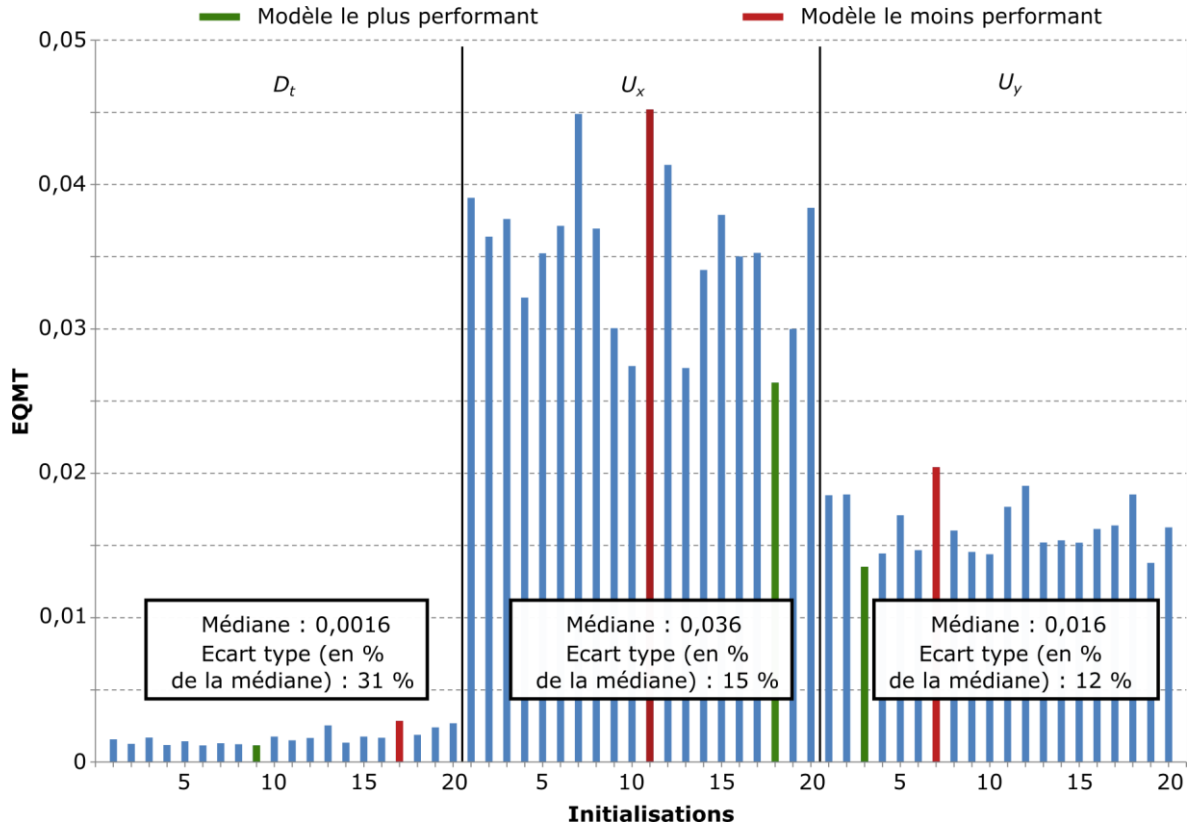


Figure III-60 : Influence de l'initialisation sur la qualité des modèles réseaux de neurones (D_t , U_x , U_y), rappelons que le vent est orienté selon x .

Si l'on compare les valeurs de l'écart type, on observe de plus grandes variations pour les apprentissages concernant le coefficient de diffusion turbulente que pour la vitesse. Les valeurs de l'erreur quadratique moyenne sur l'ensemble de test sont plus faibles pour la détermination du coefficient de diffusion turbulente. Ce résultat attendu est à prendre en compte pour l'apprentissage final.

ii. Influence du nombre de neurones dans la couche cachée

Afin de quantifier l'influence du nombre de neurones dans la couche cachée, il est nécessaire d'effectuer plusieurs initialisations à chaque ajout de neurone. Aussi, pour le même échantillonnage, on testera un intervalle de 1 à 20 neurones pour 20 initialisations à chaque fois. On reporte la valeur minimale pour chaque ensemble d'initialisations pour les différentes variables à prédire sur la figure ci-après :

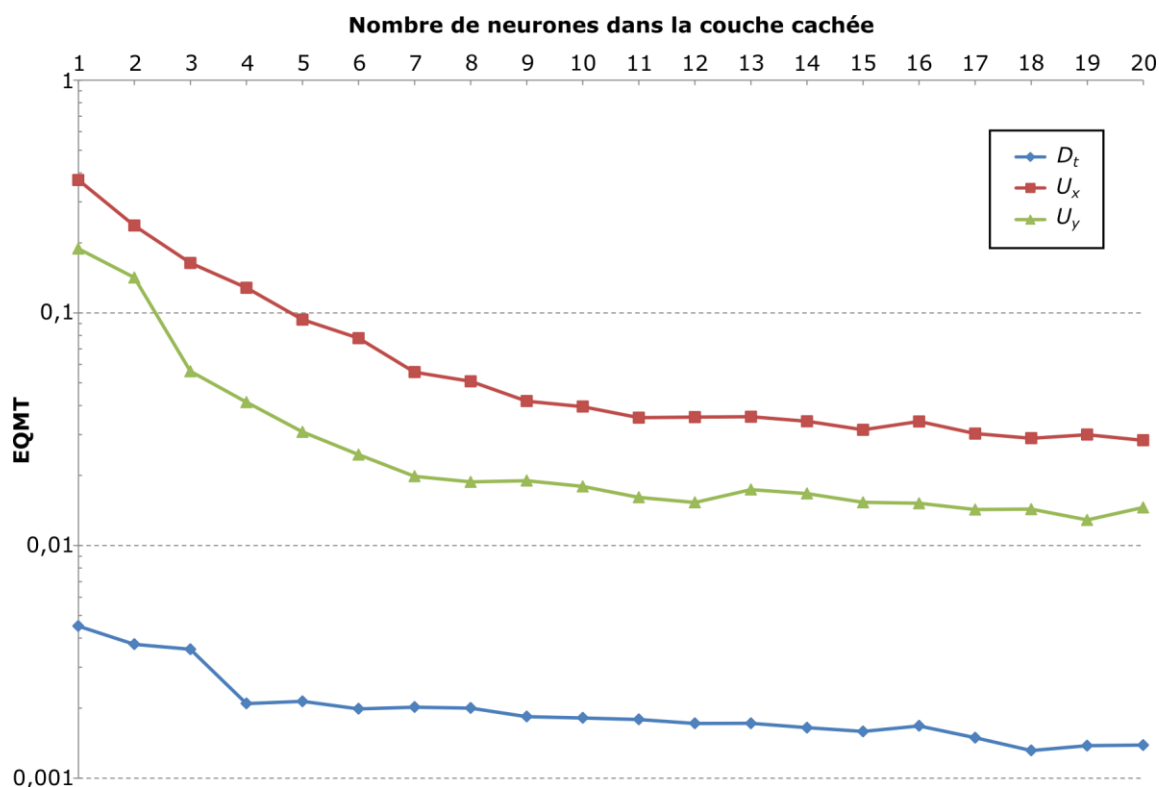


Figure III-61 : Influence du nombre de neurones dans la couche cachée pour les différentes variables de sortie.

Pour chaque variable d'intérêt, on observe la même tendance : l'augmentation du nombre de neurones dans la couche cachée diminue la valeur de l'erreur quadratique moyenne sur l'ensemble de test. Un nombre de neurones dans la couche cachée supérieur à 15 sera utilisé pour s'assurer l'atteinte d'une erreur quadratique moyenne minimale.

iii. Influence de l'échantillonnage

L'échantillonnage est effectué selon la méthode décrite au II-2.2.5i. On s'assure que l'ensemble d'apprentissage contient des exemples issus de chaque cas de la base de données. Le graphe ci-après présente la qualité de l'apprentissage selon l'EQMT en fonction du nombre de valeurs dans la base d'exemple. Il n'y a pas de tendance nette, la qualité de l'apprentissage ne semblant pas être liée directement au nombre de valeurs disponibles dans la base d'exemples. L'EQMT varie entre 0,026 et 0,031. L'écart-type de cette distribution est égale à 4% de la valeur médiane. Les variations sont donc relativement faibles. Les disparités observées dans la valeur de l'EQMT sont à mettre en perspective des valeurs obtenues par le coefficient de détermination. Celles-ci sont comprises entre 0,996 et 0,997 ce qui traduit un très bon niveau d'apprentissage. L'apprentissage à l'EQMT le plus faible est sélectionné sans raffiner l'échantillonnage.

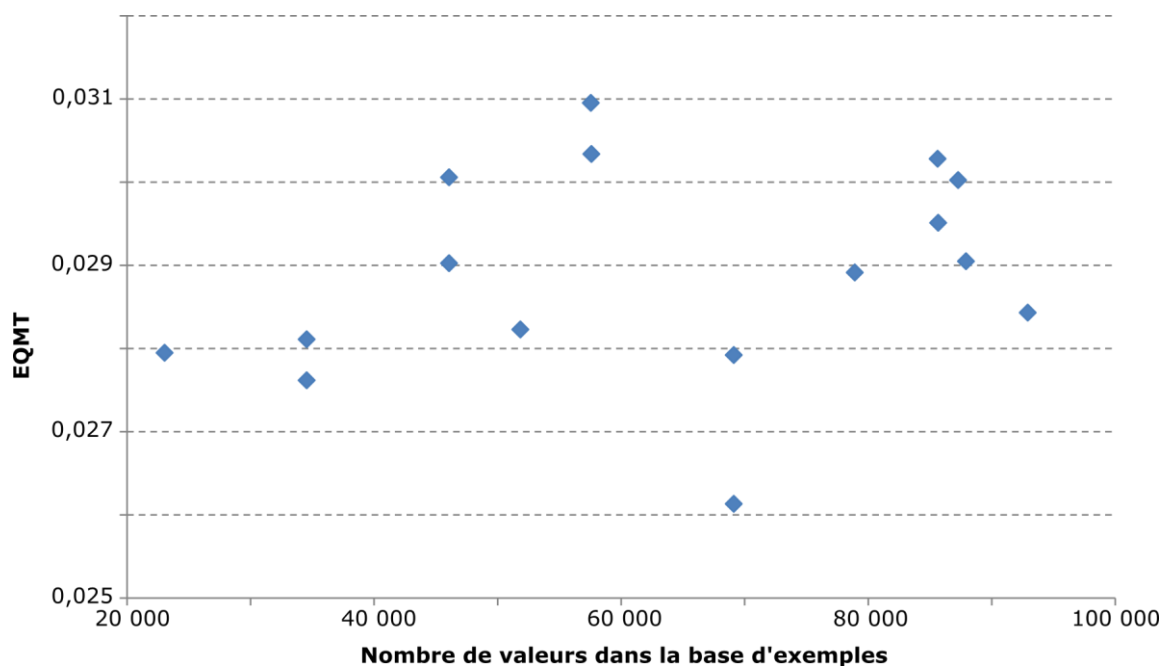


Figure III-62 : Erreur quadratique moyenne sur l'ensemble de test et nombre de valeurs dans la base d'exemples pour la détermination de la vitesse selon la direction x .

En ce qui concerne l'apprentissage de la vitesse selon y , l'EQMT diminue avec le nombre total de valeurs :

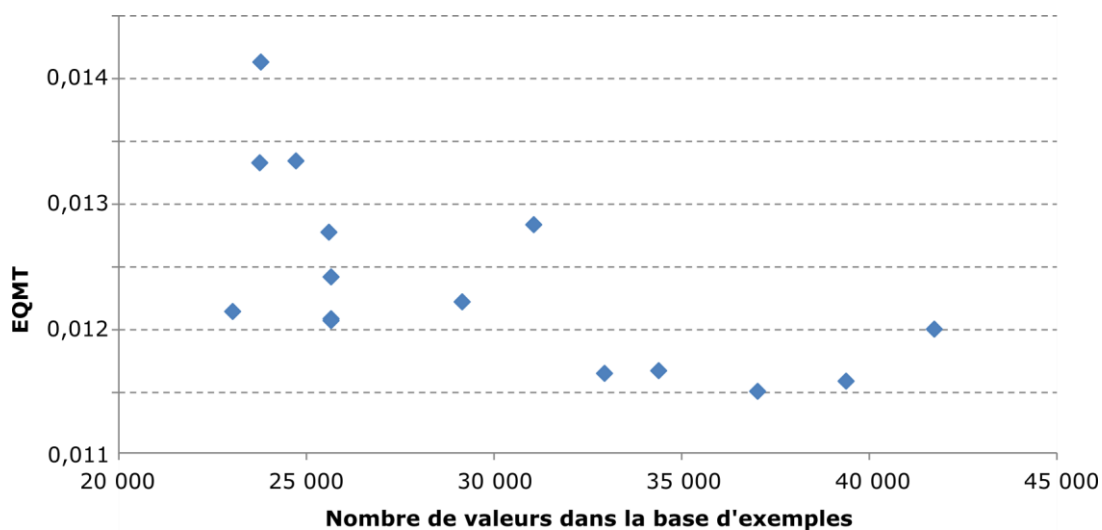


Figure III-63 : Qualité de l'apprentissage en fonction du nombre de valeurs dans la base d'exemples pour la détermination de la vitesse selon y .

L'EQMT varie entre 0,0115 et 0,014. L'écart-type de cette distribution est égale à 6% de la valeur médiane. Les écarts sont également faibles et les valeurs du coefficient de détermination compris entre 0,997 et 0,999. La base d'exemples finale comportera plus de 30 000 valeurs afin de s'assurer de la sélection du meilleur modèle.

En s'intéressant à cette même problématique sur la qualité de l'apprentissage pour la détermination du coefficient de diffusion turbulente, il est possible d'extraire une tendance sur l'influence du nombre d'intervalles :

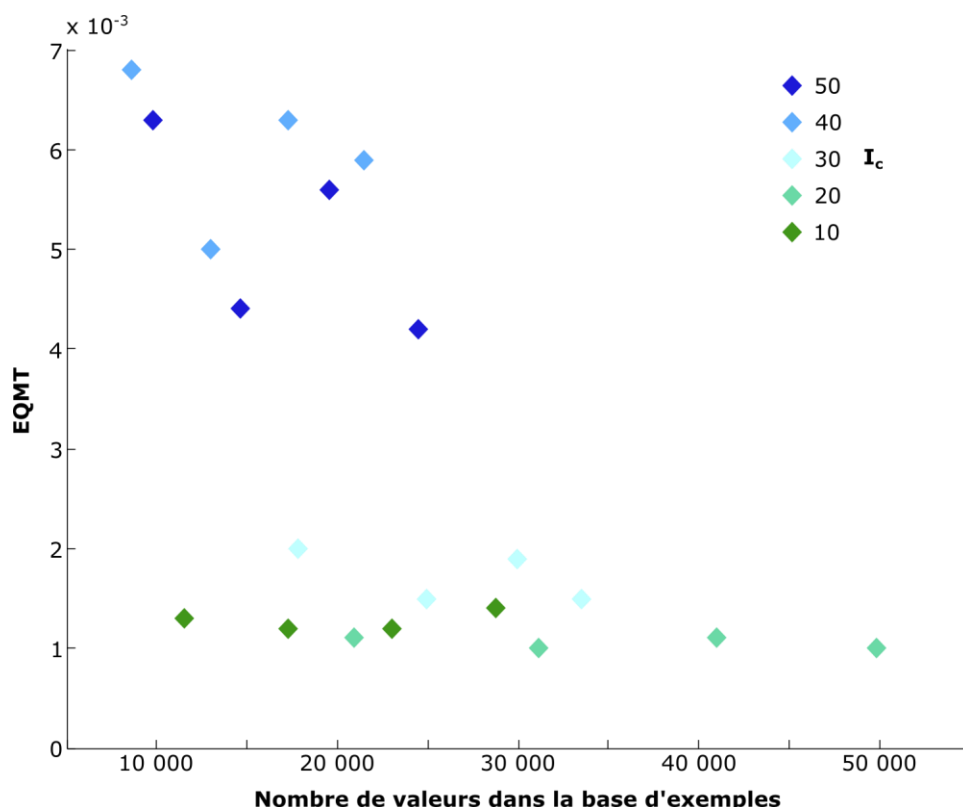


Figure III-64 : Qualité de l'apprentissage du réseau de neurones pour la détermination du coefficient de diffusion turbulente en fonction du nombre total d'exemples dans la base d'apprentissage et du nombre d'intervalles sélectionnés par cas.

Des préconisations peuvent être effectuées à partir d'observations sur ce graphique : avec l'augmentation du nombre de valeurs dans la base d'exemple, l'EQMT diminue. Par ailleurs, pour un nombre de valeurs similaire dans la base d'exemples, le modèle est plus performant lorsqu'il est appris à partir d'une base créée à partir d'un nombre d'intervalles compris entre 10 et 30. Pour la sélection du meilleur modèle possible, il est nécessaire de se limiter à un nombre d'intervalles inférieur à 30.

iv. Sélection des meilleurs modèles

Finalement, l'évaluation des paramètres influant sur l'apprentissage permet de sélectionner le meilleur réseau de neurones pour chaque paramètre à déterminer. Le tableau ci-après répertorie cette sélection :

Tableau III-11 : Caractéristiques de l'apprentissage optimal sélectionné pour chaque paramètre.

Paramètre déterminé	D_t	U_x	U_y
Nombre d'intervalles de concentrations	20	40	20
Nombre d'exemples par intervalle	30	20	20
Nombre de neurones dans la couche cachée	20	20	20
Numéro de l'initialisation	4	6	13
Valeur de l'EQMT en fin d'apprentissage	$9,8 \times 10^{-4}$	0,026	0,0113
Valeur du coefficient de détermination	0,985	0,997	0,999
Durée d'apprentissage (s)	906	3731	2775

III-4.5. Evaluation sur des cas non-appris

Des cas non-appris sont simulés par méthode CFD pour les comparer à la méthode développée ici et ainsi évaluer les performances de généralisation des réseaux de neurones utilisés. Ainsi, trois diamètres (12, 26 et 50 mètres de diamètre) et trois vitesses initiales (2,5 ; 5,5 et 9,5 m.s⁻¹) ont été sélectionnés. A chaque fois, on sélectionne des valeurs représentant des valeurs faibles, moyennes et fortes de l'intervalle initial. Le domaine de calcul dans lequel le réseau de neurones a été appris est celui défini au III-4.2.1. En tenant compte de la symétrie du problème, il est possible de modéliser un domaine rectangle ayant les caractéristiques suivantes :

- ▶ Trois diamètres en amont,
- ▶ 30 diamètres en aval,
- ▶ Trois diamètres de largeur.

Au-delà de ces valeurs, la validité du réseau de neurones, évaluée par la qualité de son apprentissage, n'est pas garantie. Pour autant, il n'est pas exclu que le réseau de neurones fournissent des résultats cohérents.

III-4.5.1 Vitesse selon x

L'évaluation sur des cas non-appris de la composante selon x de la vitesse fournit de bons résultats. Ceux-ci sont répertoriés dans le tableau ci-après :

Tableau III-12 : Prédiction de la vitesse selon x pour des cas non-appris

Cas	Diamètre (en m)	Vitesse (m.s ⁻¹)	R ²	FAC2	FB	NMSE
1	12	2,5	0,98	0,99	10 ⁻³	10 ⁻³
2	12	5,5	0,98	0,99	10 ⁻⁴	
3	12	9,5	0,97	0,99	-10 ⁻⁴	
4	26	2,5	0,96	0,99	10 ⁻³	
5	26	5,5	0,96	0,99	10 ⁻³	
6	26	9,5	0,96	0,99	10 ⁻³	
7	50	2,5	0,94	0,99	10 ⁻³	
8	50	5,5	0,95	0,99	10 ⁻³	
9	50	9,5	0,96	0,99	10 ⁻³	

Le critère (FAC2) est respecté dans 99% des mesures pour chacun des cas. La vitesse selon x étant une composante essentielle de la dispersion, ce résultat est une condition nécessaire pour l'application de l'équation d'advection diffusion. Concernant le coefficient de détermination, les valeurs obtenues sont également satisfaisantes. De légères disparités peuvent cependant être notées. Pour les diamètres importants, le réseau de neurones semble avoir plus de difficultés à prédire les valeurs correctes. Dans la prévision de ces cas, la comparaison nécessite plus de points d'observations afin de garantir un quadrillage de la totalité de la zone à modéliser par la suite. Les différences de vitesse ne semblent pas affecter la qualité de la modélisation. De manière générale, l'ordre de grandeur de l'erreur quadratique normalisée (NMSE) est faible. Par ailleurs, le biais fractionnel (FB) est lui aussi relativement faible. Excepté pour le cas trois, pour lequel FB est négatif, les autres modélisations ont des valeurs de vitesse légèrement sous-estimées (FB positif).

Ces critères généraux ne permettent pas d'identifier les zones où les prévisions sont performantes et celles où elles ne le sont pas. Ainsi, une première indication est donnée

par la visualisation des champs de vitesses selon x pour le cas CFD et le cas ANN. Tout d'abord, on regarde pour le meilleur cas prédit, le cas un :

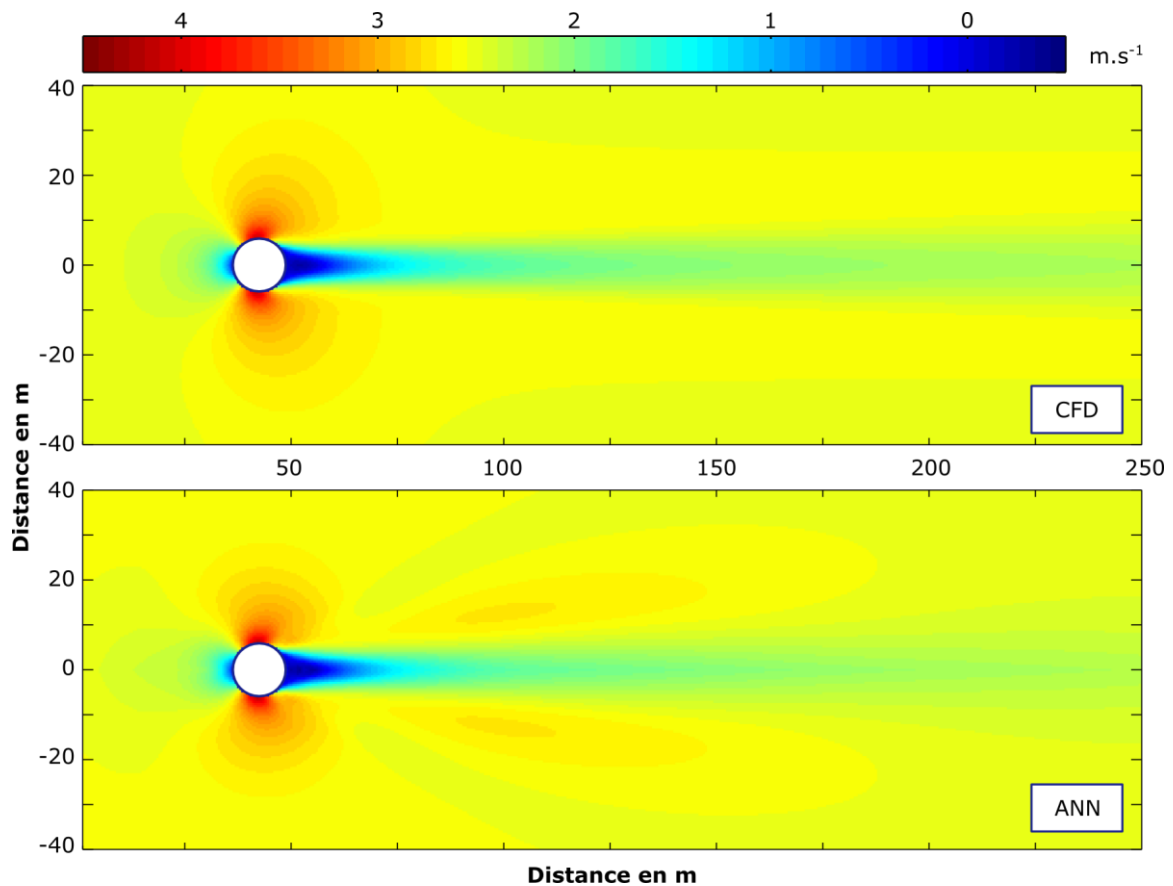


Figure III-65 : Visualisation de la réalité CFD et de la modélisation ANN de la vitesse selon la direction x pour le meilleur cas non-appris (Diamètre de l'obstacle : 12 m - Vitesse en entrée : $2,5 \text{ m.s}^{-1}$)

L'ensemble du domaine est correctement modélisé par le réseau de neurones. Les zones d'accélération sur les côtés de l'obstacle sont un peu moins larges dans le cas de la modélisation ANN. La zone directement en aval de l'obstacle présente des valeurs de vitesse négatives correspondant à la zone de recirculation. Des singularités apparaissent néanmoins en aval de l'obstacle, sur les côtés de la zone de perturbation. Elles entraînent une surestimation des valeurs de vitesse.

En ce qui concerne la visualisation du cas défavorable, les difficultés de modélisation sont également identifiables. Elles sont plus présentes dans ce cas. Des zones présentent une surestimation : en amont de l'obstacle, au-delà de sa zone d'influence et en aval, sur les bords de la zone de perturbation. Dans ces deux situations, le réseau de neurones prédit mal les zones où il n'y a pas de variations majeures de l'écoulement. Par ailleurs, les zones de recirculation aval, de ralentissement amont et d'accélération sur les bords de l'obstacle sont correctement modélisées.

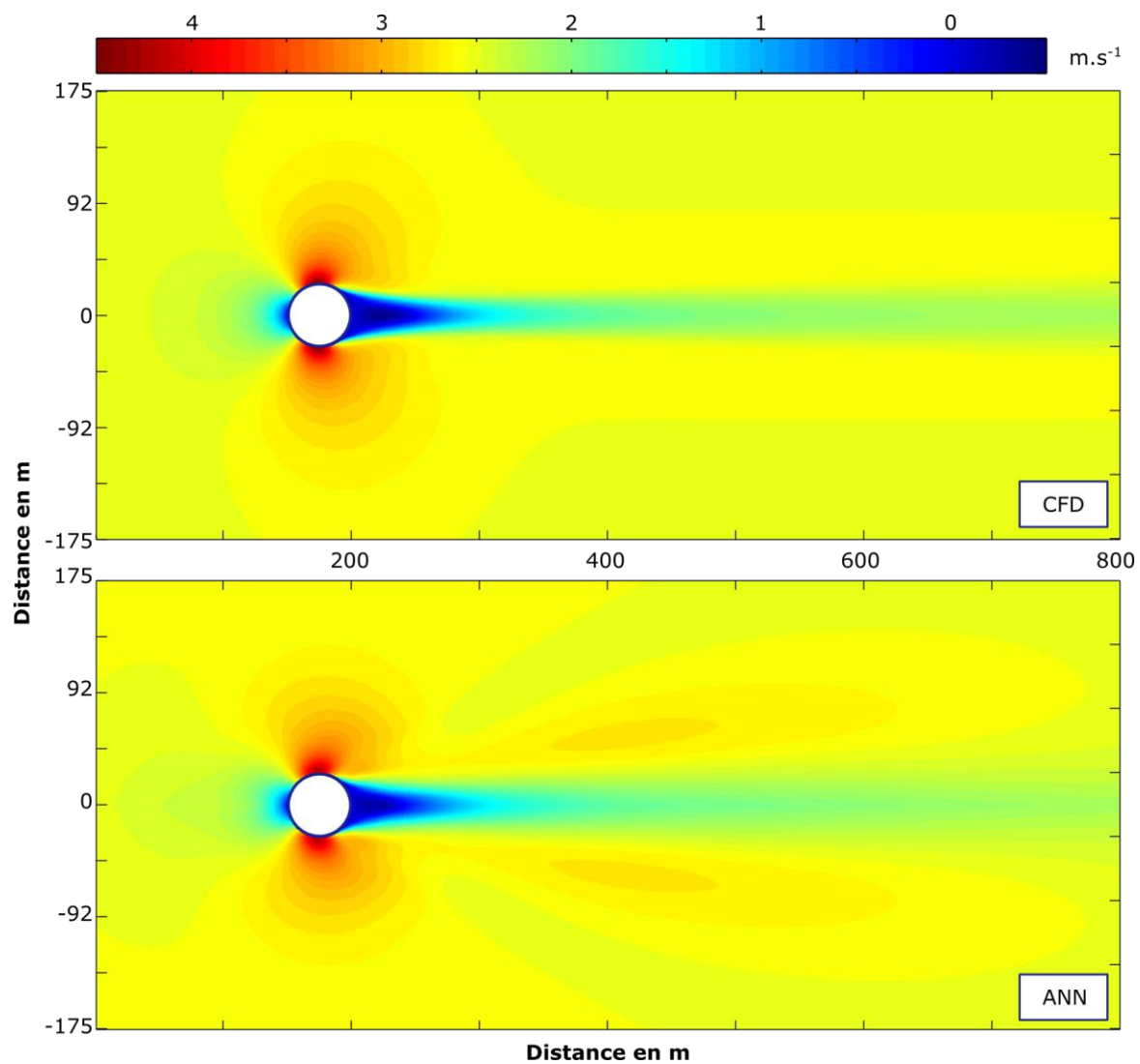


Figure III-66 : Visualisation de la réalité CFD et de la modélisation ANN de la vitesse selon la direction x pour le cas non-appris défavorable (Diamètre de l'obstacle : 50 m - Vitesse en entrée : 2,5 m.s⁻¹)

Ces observations visuelles permettent la critique qualitative des prédictions du modèle neuronal. Obtenir des informations quantitatives est possible en traçant notamment la dispersion des valeurs entre la réalité CFD et la modélisation ANN.

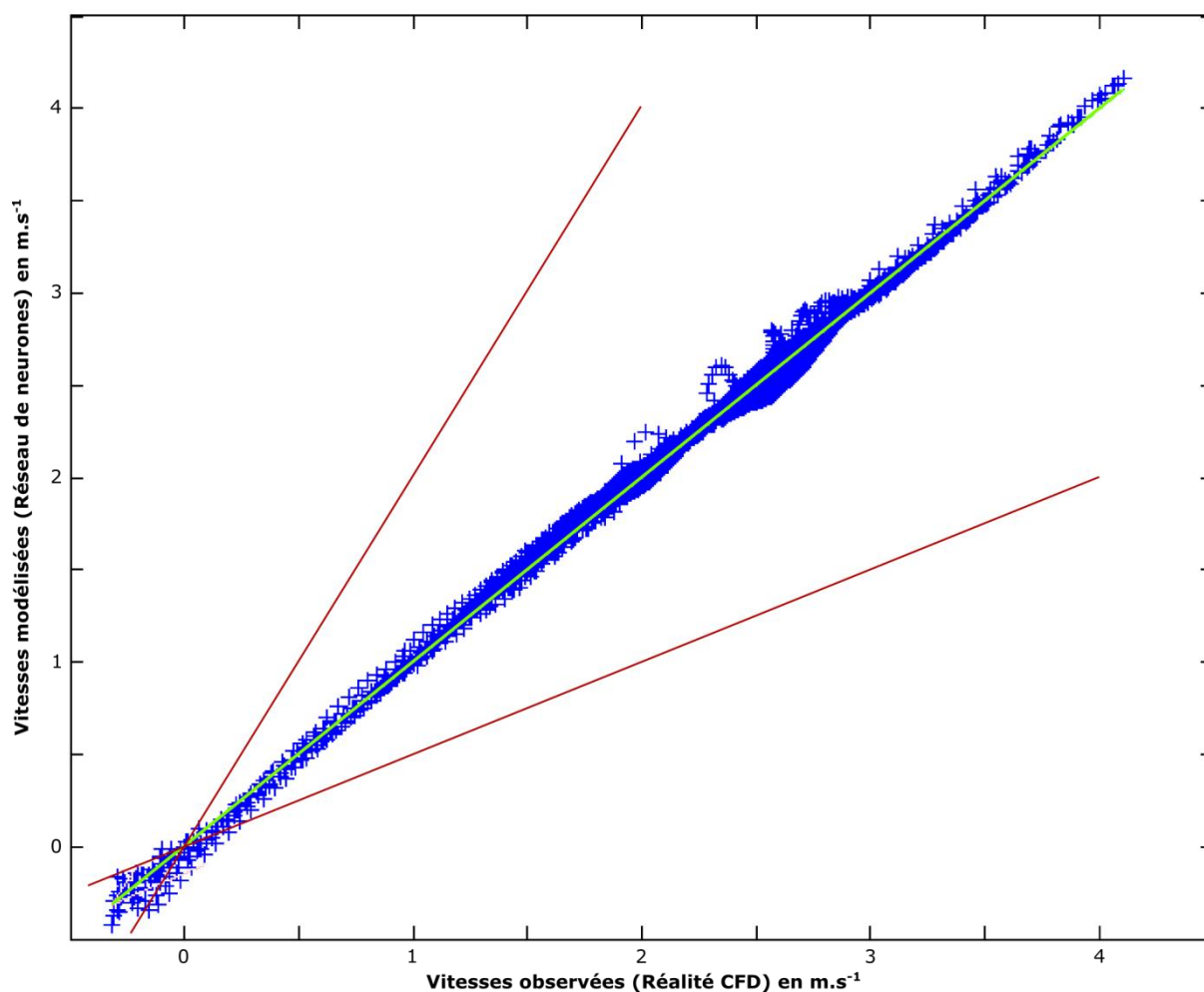


Figure III-67 : Diagramme de dispersion entre vitesses observées et vitesses modélisées par ANN. Les droites rouges signalent une sur/sous estimation d'un facteur 2. La droite verte est la fonction $y=x$. Représentation du cas favorable (diamètre : 12m – vitesse du vent : $2,5 \text{ m.s}^{-1}$)

Les 87 373 points modélisés sont répartis de différentes façons sur l'intervalle des vitesses. La tendance générale est la proximité à la droite $y=x$, confirmant une prévision satisfaisante. Beaucoup de valeurs sont concentrées autour de la vitesse d'entrée, présentant une légère surestimation ou sous-estimation selon les points considérés. 99% des valeurs estimées sont comprises entre les droites représentant le facteur deux. Néanmoins, en considérant les valeurs proches de zéro, certains points excèdent ces valeurs. L'utilisation du facteur deux est dans cette situation est discutable.

Les mêmes observations sont possibles sur le cas défavorable :

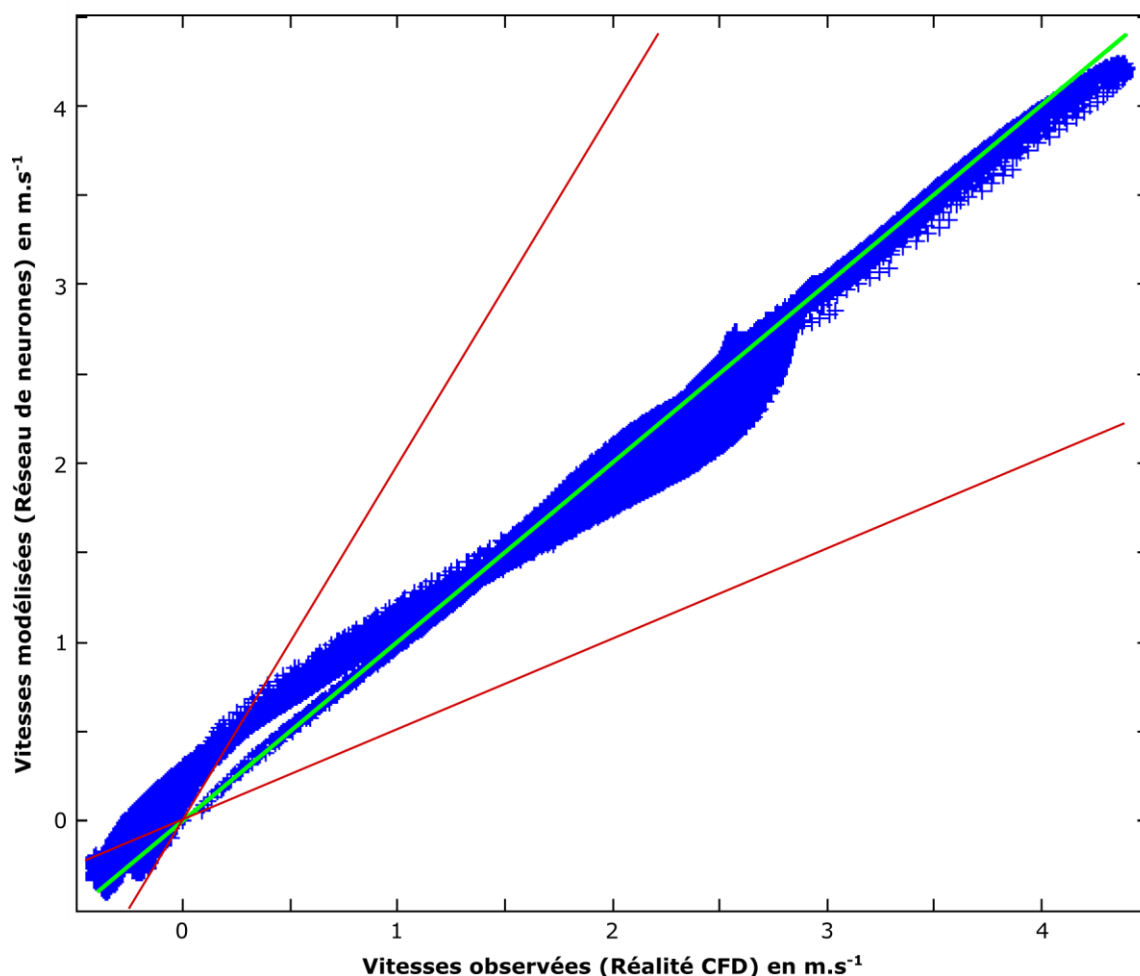


Figure III-68 : Diagramme de dispersion entre vitesses observées et vitesses modélisées par l'ANN. Les droites rouges signalent une sur/sous estimation d'un facteur 2. La droite verte est la fonction $y=x$. Représentation du cas favorable (diamètre : 50 m – vitesse de vent : $2,5 \text{ m.s}^{-1}$)

Les 377 676 points du domaine dans le cas d'un large diamètre présentent des valeurs de vitesse proche de la vitesse d'entrée en grande quantité. La vitesse maximale est légèrement sous-estimée par le réseau de neurones. Par ailleurs, les vitesses faibles non-négatives sont surestimées. La grande majorité des valeurs est correctement modélisée par le réseau de neurones : 99% des valeurs estimées sont comprises dans un facteur deux, même dans ce cas qui est le plus défavorable.

Il est possible d'approfondir l'analyse en observant les valeurs de vitesse dans l'axe central du domaine :

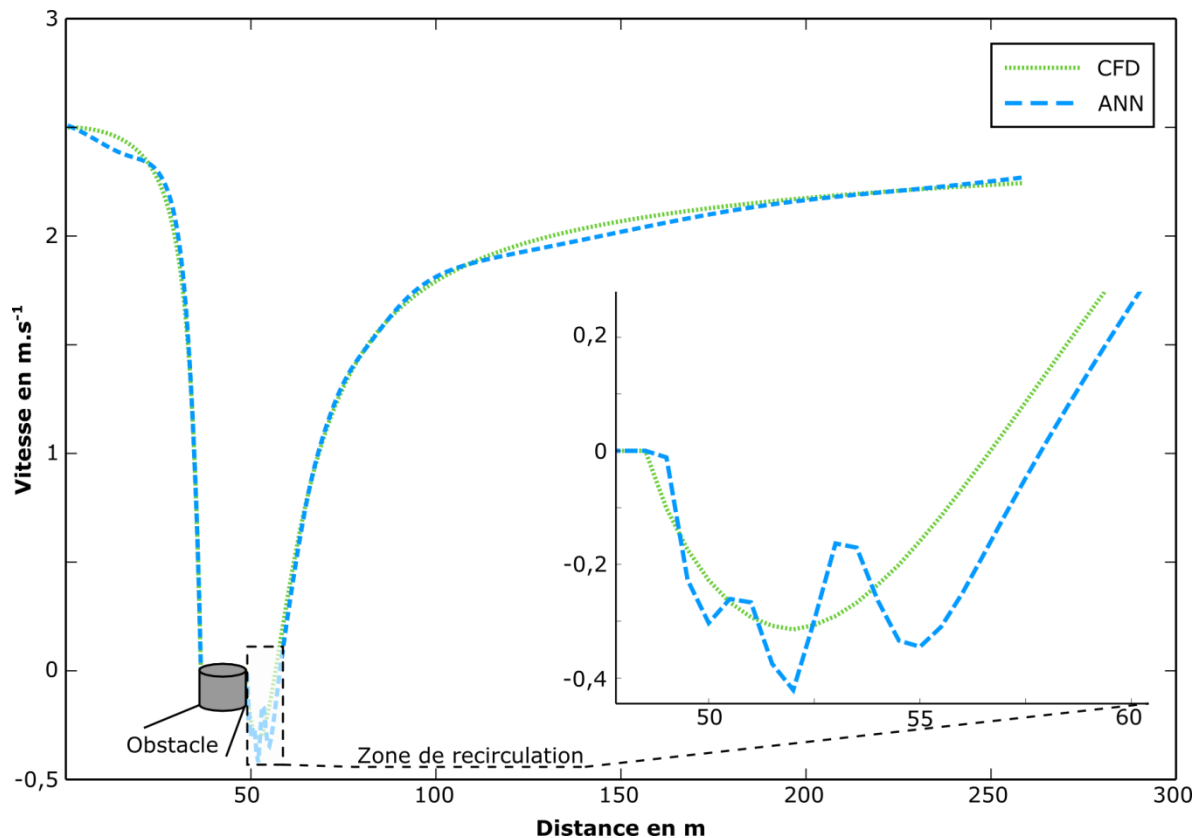


Figure III-69 : Vitesse selon x dans l'axe central du domaine pour la réalité CFD et la modélisation ANN (cas favorable : $D=12\text{ m}$ – $v=2,5\text{ m.s}^{-1}$)

Sur la quasi-totalité du domaine, les courbes de vitesse simulées (CFD) et modélisées (ANN) se confondent. Néanmoins, la zone de recirculation présente quelques écarts. La courbe issue de la réalité CFD diminue, atteint un minimum puis augmente de manière continue. La trajectoire de la courbe issue de la modélisation par réseau de neurones présente des discontinuités. Il sera nécessaire d'évaluer si ces écarts ont un impact sur la modélisation finale dans le paragraphe III-4.6.

III-4.5.2 Vitesse selon y

La deuxième composante de la vitesse, normale à la vitesse d'entrée subit des variations plus faibles. Celles-ci sont principalement localisées près de l'obstacle. L'utilisation du réseau de neurones pour les différents cas tests donne des résultats satisfaisants, répertoriés dans le tableau ci-après :

Tableau III-13 : Prédiction de la vitesse selon y pour des cas non-appris

Cas	Diamètre (en m)	Vitesse (m.s^{-1})	R^2	FAC2
1	12	2,5	0,98	0,46
2	12	5,5	0,98	0,47
3	12	9,5	0,98	0,48
4	26	2,5	0,98	0,45
5	26	5,5	0,98	0,45
6	26	9,5	0,98	0,46

Cas	Diamètre (en m)	Vitesse (m.s^{-1})	R^2	FAC2
7	50	2,5	0,98	0,44
8	50	5,5	0,99	0,45
9	50	9,5	0,98	0,44

Il est difficile d'identifier un cas nettement meilleur qu'un autre à travers l'utilisation du coefficient de détermination. Les valeurs du FAC2 ne sont pas bonnes. Cela est dû aux valeurs proches de zéro. Effectivement, les valeurs sont susceptibles de varier de plusieurs ordres de grandeur et donc d'affecter fortement la valeur de cet indicateur comme observé sur la figure ci-après :

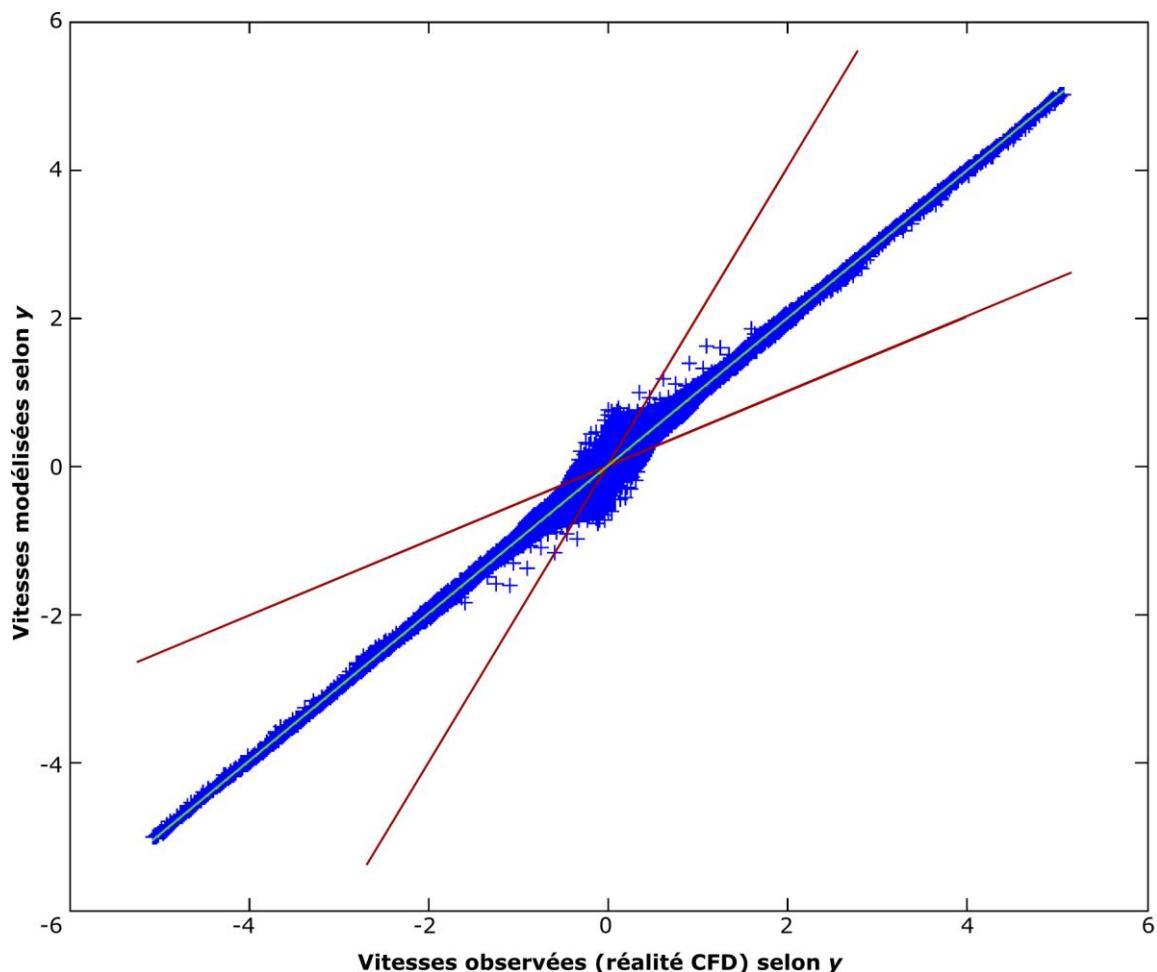


Figure III-70 : Dispersion des valeurs modélisées (ANN) par rapport aux valeurs issues de la réalité CFD (cas favorable $D=50\text{ m}$ – Vitesse de vent : $5,5\text{ m.s}^{-1}$)

Le FB et NMSE ne sont pas pertinents lorsqu'un changement de signe est effectif dans le jeu de données. En visualisant comparativement les champs de vitesse selon y , il est aisé d'identifier les zones où la prédiction est déformée, en amont de l'obstacle :

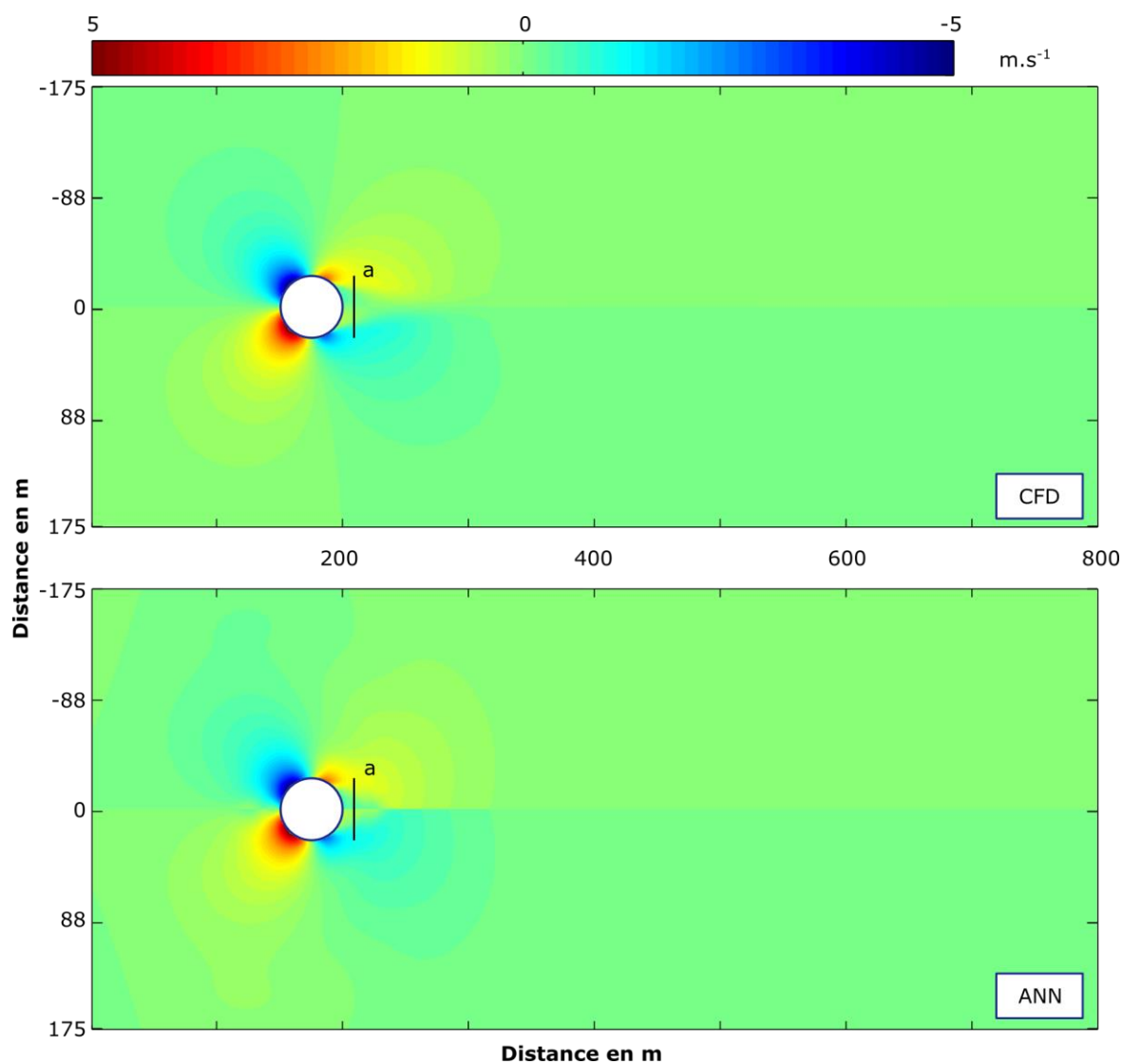


Figure III-71 : Visualisation du champ 2D de vitesse selon y pour la simulation CFD et la modélisation par réseau de neurones pour le cas 8 (Diamètre : 50 m – Vitesse du vent : $5,5 \text{ m.s}^{-1}$)

Le changement de signe sur l'axe central semble être plus prononcé sur la modélisation ANN, notamment au niveau de la zone de recirculation. Pour caractériser pleinement cet écart, il est possible de tracer une coupe selon l'axe y :

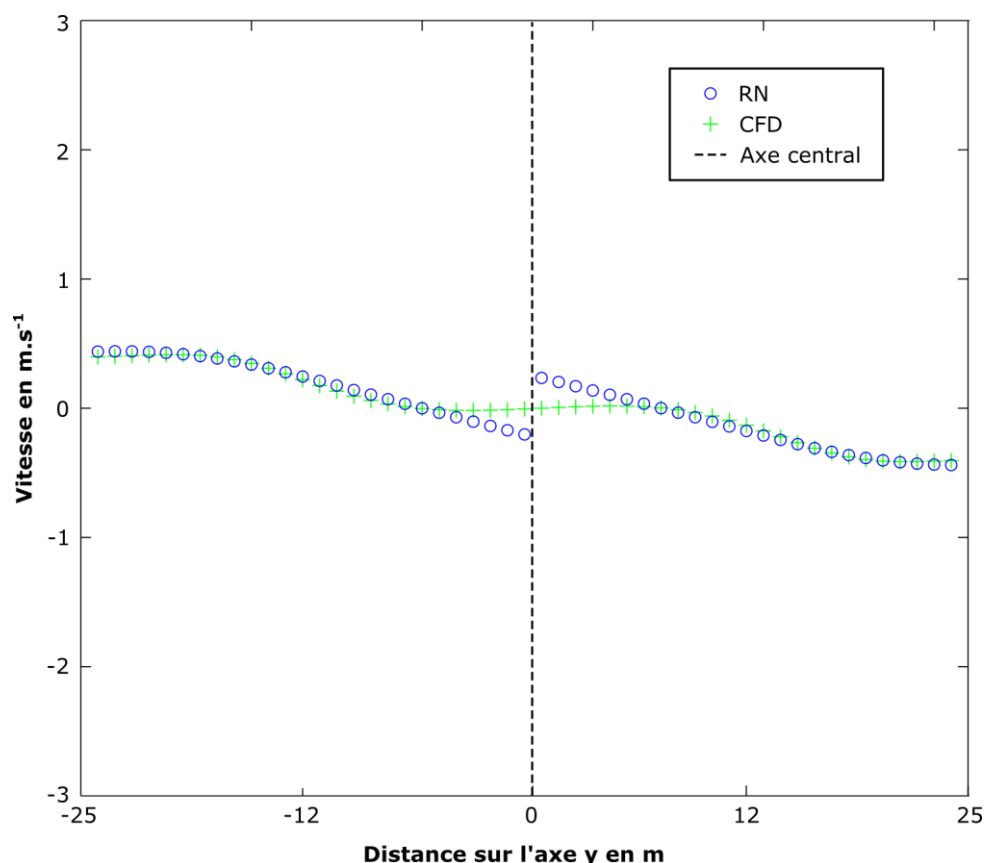


Figure III-72 : Vitesse selon l'axe y : coupe selon l'axe y dans la zone de recirculation (indiquée -a- sur la figure précédente) pour le cas 8 (Diamètre : 50 m – Vitesse du vent : $5,5 \text{ m.s}^{-1}$)

Ce graphique présente un agrandissement de la zone immédiatement en aval de l'obstacle. Dans l'ensemble, les points sont alignés entre la modélisation ANN et la réalité CFD. Cependant, à proximité de l'axe central, le réseau de neurones prédit moins efficacement les valeurs de vitesse proches de zéro. L'angle entre la position du point à modéliser et l'axe central est une des entrées du réseau de neurones. Les valeurs trop faibles peuvent être interprétées comme un manque de représentativité dans la base d'exemples pour cette zone. D'autre part, la discontinuité entre les y positifs et négatifs est due à la reconstruction du champ de vitesse de manière symétrique, le réseau de neurones ne prédisant qu'une moitié du domaine.

III-4.5.3 Coefficient de diffusion turbulente

L'évaluation de la prévision du coefficient de diffusion turbulente est reportée dans le tableau suivant :

Tableau III-14 : Critères de performance pour la prévision du coefficient de diffusion turbulente par le réseau de neurones sur les cas non-appris.

Cas	Diamètre (en m)	Vitesse (m.s^{-1})	R^2	FAC2	FB	NMSE
1	12	2,5	0,98	0,99	$-1,2 \times 10^{-2}$	$1,3 \times 10^{-3}$
2	12	5,5	0,98	0,99	$-1,1 \times 10^{-2}$	$1,2 \times 10^{-3}$
3	12	9,5	0,98	0,99	$-7,5 \times 10^{-3}$	$1,1 \times 10^{-3}$

Cas	Diamètre (en m)	Vitesse (m.s ⁻¹)	R ²	FAC2	FB	NMSE
4	26	2,5	0,98	0,99	-2,3 x 10 ⁻³	4,0 x 10 ⁻³
5	26	5,5	0,98	0,99	-7,1 x 10 ⁻³	3,7 x 10 ⁻³
6	26	9,5	0,98	0,99	-3,4 x 10 ⁻³	3,4 x 10 ⁻³
7	50	2,5	0,93	0,95	-3,2 x 10 ⁻²	3,7 x 10 ⁻²
8	50	5,5	0,92	0,94	-1,8 x 10 ⁻²	3,8 x 10 ⁻²
9	50	9,5	0,91	0,93	2,1 x 10 ⁻²	4,1 x 10 ⁻²

Pour l'ensemble des cas, le coefficient de diffusion turbulente est déterminé avec une valeur de coefficient de détermination et de facteur 2 supérieure à 0,9, démontrant une aptitude satisfaisante à prédire le champ horizontal. Le biais fractionnel comme l'erreur quadratique moyenne normalisée donnent des valeurs faibles. En ce qui concerne l'erreur systématique, la plupart des cas montrent un critère général tendant vers la surestimation (FB négatif). Le NMSE traduit une erreur globale plus importante sur les cas tests à diamètre important, où un domaine plus large est représenté. Ces valeurs restent dans un intervalle correct pour la modélisation. La visualisation comparative sur le cas test le mieux prédit permet d'identifier des zones où le réseau de neurones a des difficultés à prédire le comportement de ce paramètre :

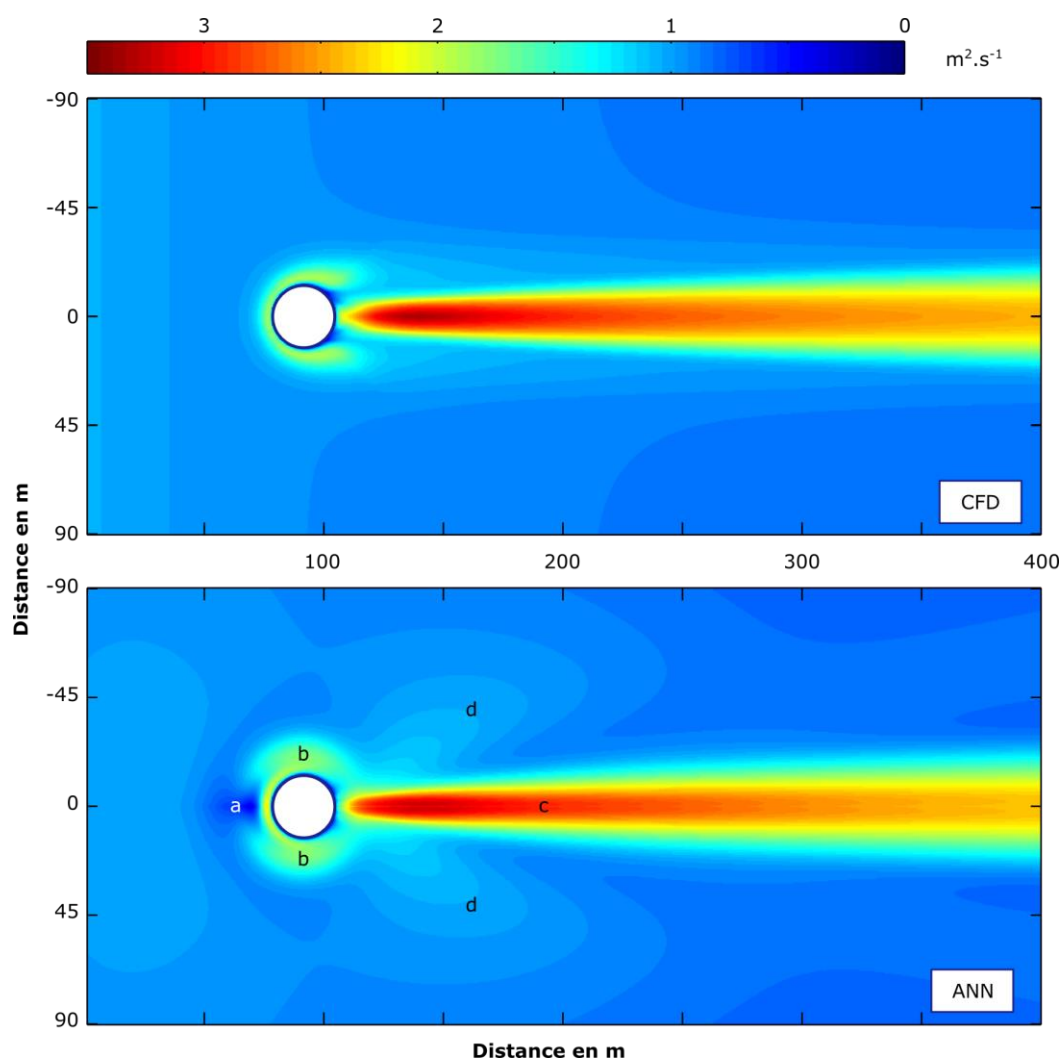


Figure III-73 : Visualisation de la prédiction du coefficient de diffusion turbulente par l'ANN. Cas favorable (Diamètre : 26 m - Vitesse du vent : 9,5 m.s⁻¹)

Une zone en amont de l'obstacle est largement sous-estimée (a), avec des valeurs proches de zéro. Le contournement de l'obstacle (b) présente lui aussi des écarts à la simulation : zone plus large et moins profilée. Le sillage (c) est lui correctement modélisé. Des champs latéraux (d) sont identifiés dans la prévision neuronale, illustrant des difficultés dans cette zone. Des observations similaires peuvent être effectuées sur le cas le moins bien prédit :

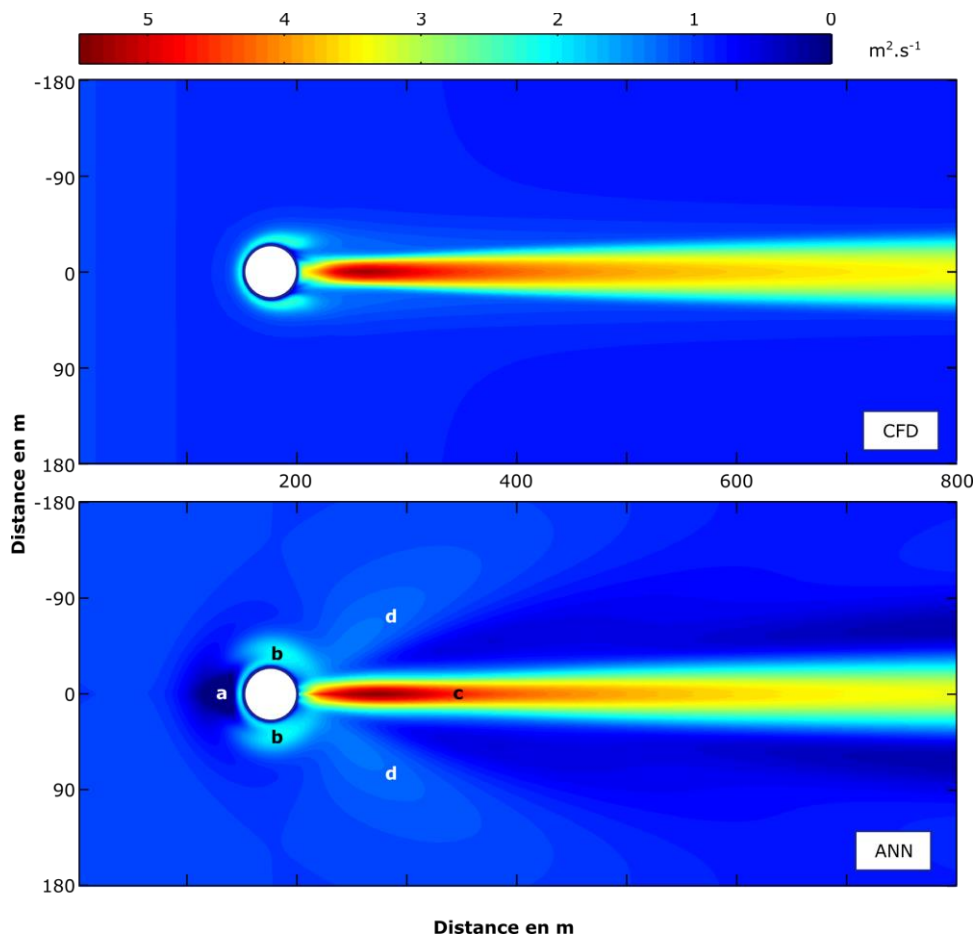


Figure III-74 : Visualisation de la prédiction du coefficient de diffusion turbulente par l'ANN. Cas défavorable (Diamètre : 50 m - Vitesse du vent : $9,5 \text{ m.s}^{-1}$)

Les moins bons résultats obtenus pour ce cas sont identifiables visuellement sur la figure précédente. Plusieurs zones ne sont pas correctement modélisées. Il s'agit notamment de la zone en amont de l'obstacle (a). Certaines valeurs semblent incohérentes, de l'obstacle vers les bords du domaine (b et d). Cela peut s'expliquer par des incohérences également observées dans la base de données, à proximité des bords. Le réseau de neurones tente de les reproduire, puisqu'il ne se rattache pas aux équations physiques mais est issu d'un apprentissage statistique. Le coefficient de diffusion est proche de zéro dans le cas modélisé. Les valeurs du coefficient de diffusion turbulente dans le sillage (c) sont semblables. Par ailleurs, le reste du domaine est relativement correctement modélisé.

En s'intéressant au graphe de dispersion des valeurs de la réalité CFD par rapport à la modélisation ANN, ces disparités sont constatées, notamment dans les valeurs faibles (figure ci-après) :

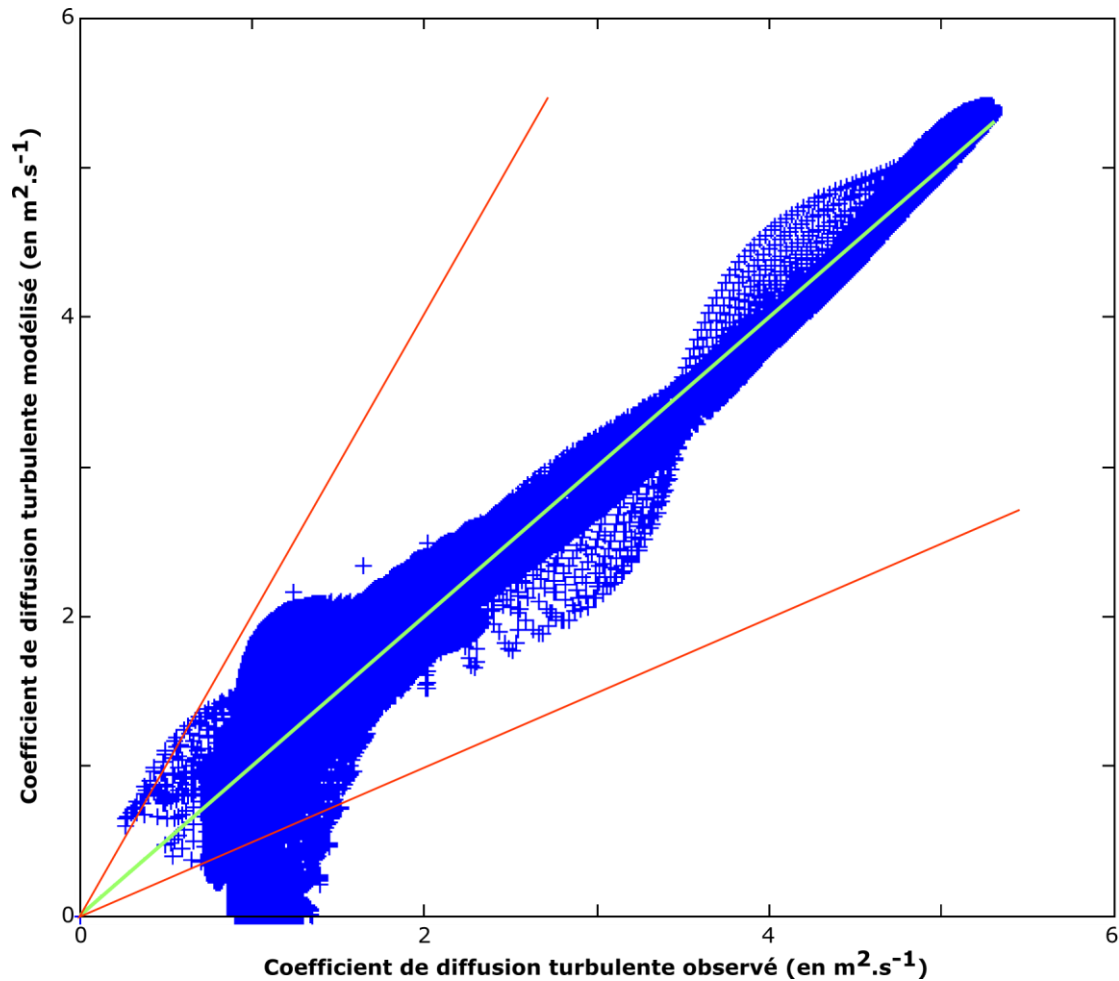


Figure III-75 : Graphe de dispersion des données modélisées et observées pour le cas défavorable (Diamètre : 50 m – vitesse du vent : $9,5 \text{ m.s}^{-1}$)

Le coefficient de diffusion turbulente est correctement modélisé pour une grande partie des valeurs, notamment pour les plus importantes. 93% des valeurs restent cependant dans l'intervalle du facteur deux. Le maximum et minimum modélisés par ANN du coefficient de diffusion turbulente concordent avec les valeurs obtenues par la réalité CFD. Des difficultés de modélisation apparaissent pour les valeurs faibles. Les erreurs les plus importantes sont dans la zone où $D_t < 1$. Cela correspond aux valeurs d'entrée dans le domaine et donc omniprésentes dans la simulation CFD. Comme vu précédemment, la zone en amont de l'obstacle présente des valeurs faibles, pouvant atteindre zéro pour la modélisation ANN. Ces valeurs ne sont pas représentatives de la réalité. Dans une démarche d'amélioration, il est peut être envisagé de contraindre le réseau de neurones à fournir des valeurs à minima égales au coefficient de diffusion turbulente d'une couche limite atmosphérique dans des conditions de stabilité neutre à l'altitude considérée. Le reste du domaine est correctement modélisé.

La caractérisation de cette zone permet d'identifier une faiblesse de la modélisation par réseau de neurones. Néanmoins, cette zone est relativement peu élargie. Aussi, il est intéressant de regarder dans la trajectoire axiale l'erreur commise par le réseau de neurones par rapport à la réalité CFD. La figure ci-après montre que la zone de l'écoulement en contact avec l'obstacle est modélisée de manière correcte. En amont de cette « ceinture », une zone de longueur inférieure à un diamètre présente des valeurs nulles, incohérentes. Par contre, dès l'aval de l'obstacle, la modélisation est de nouveau satisfaisante.

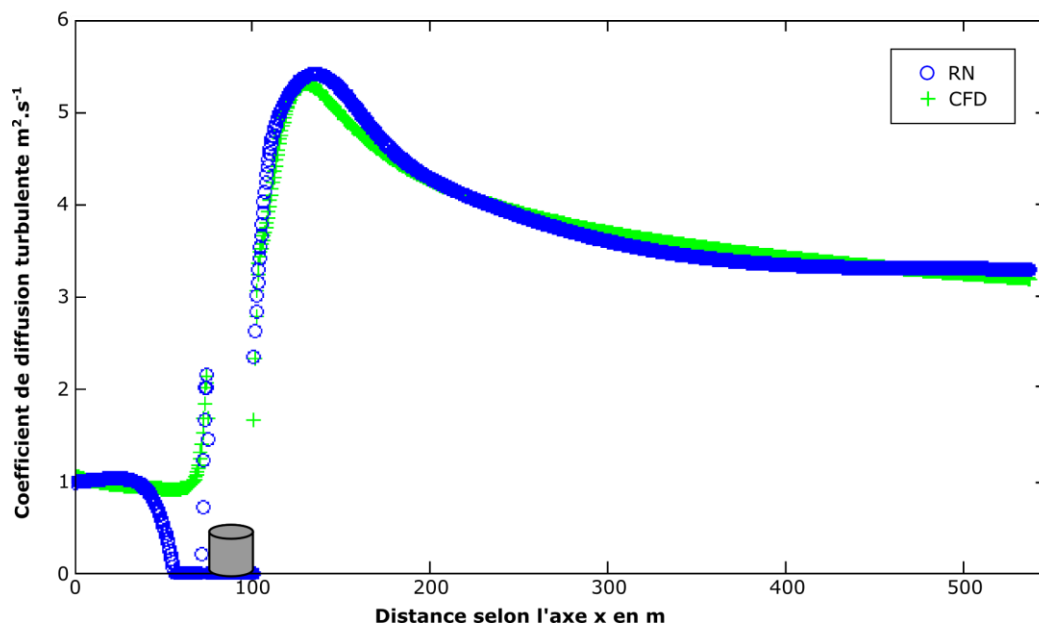


Figure III-76 : Coefficient de diffusion turbulente modélisé et issu de la réalité CFD dans l'axe central pour le cas défavorable (diamètre : 50 m – vitesse de vent : $9,5 \text{ m.s}^{-1}$)

Dans le cas favorable, des valeurs inférieures sont observées en amont de l'obstacle. Cependant, elles n'atteignent pas la valeur nulle. Par ailleurs, le champ lointain est également d'un grand intérêt. C'est là où la perturbation de l'écoulement se fait ressentir. Pour le cas défavorable, il est important que ce champ lointain soit correctement modélisé, les modifications s'étalant sur une zone plus importante que celle en amont de l'obstacle.

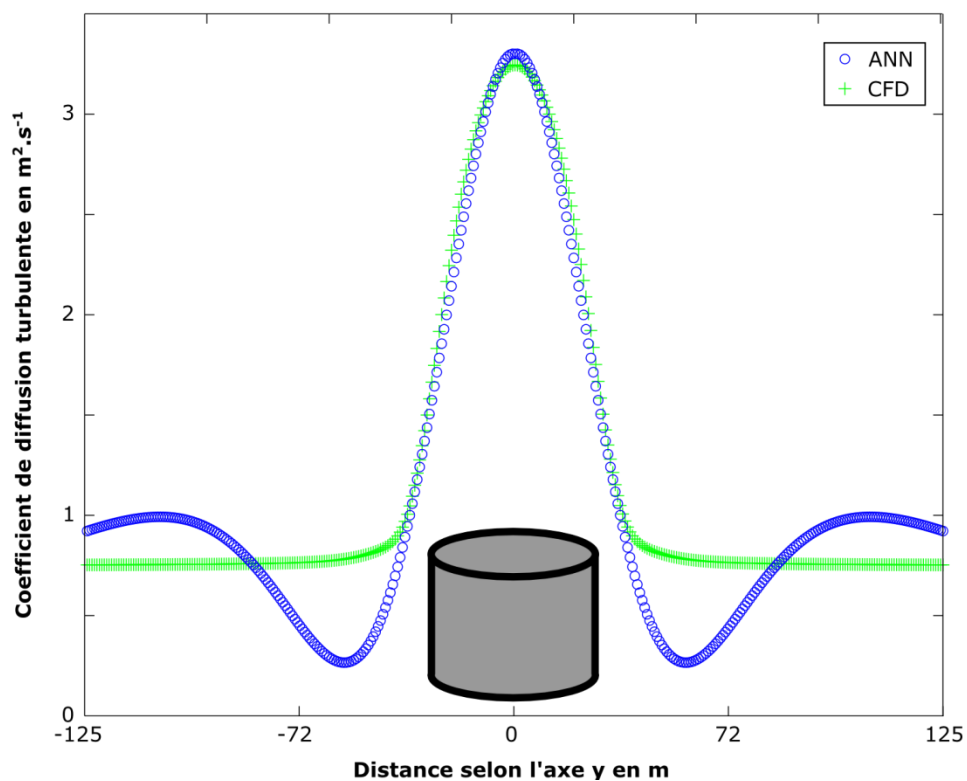


Figure III-77 : Coefficient de diffusion turbulente selon une coupe normale à l'écoulement, en aval de l'obstacle ($y=1000$ cellules)

La zone axiale est correctement modélisée puisque le pic de valeur pour D_t est respecté. Sur les côtés subsistent des valeurs sous-estimées, puis surestimées. Ce sont les singularités observées sur les côtés de la zone de sillage. Ces écarts observés sont susceptibles d'influencer le transport d'une bouffée dans le champ autour de l'obstacle. Cette influence sera évaluée dans le paragraphe III-4.6

III-4.5.4 Temps de calcul

Le temps de calcul pour la prévision des champs de vitesse et du coefficient de diffusion turbulente sont directement reliés au nombre de valeurs exigées. Ainsi, si le pas d'espace diminue ou si les dimensions du domaine augmentent (diamètre de l'obstacle plus élevé) alors le nombre de valeurs à calculer augmente également. Pour les cas précédents, le pas d'espace est de 0,5 m. Afin de simuler les différents domaines, il faut donc calculer de 87 000 valeurs environ à plus de 1 500 000, respectivement pour un domaine de 85 mètres de large pour 260 mètres de long (diamètre 12) et de 350 mètres de large pour 1075 mètres de long (diamètre 50). Avec une station de travail « classique » dont les spécifications sont disponibles en annexe, le temps de calcul varie d'une demi-seconde pour le calcul concernant les faibles diamètres à moins de 1,5 secondes pour le calcul concernant les diamètres importants.

En ce qui concerne les modèles CFD, la prise en compte de la turbulence impose des conditions sur la forme du maillage ainsi que sur le nombre de points à proximité des parois notamment. Conséquemment, le maillage utilisé lors de la simulation CFD est différent de celui interpolé sur le maillage homogène carré utilisé ici. Ainsi, le nombre de nœuds du maillage CFD varie entre 112 000 et 500 000, répartis prioritairement à proximité de l'obstacle. Une interpolation cubique est effectuée sur un maillage carré afin de pouvoir comparer CFD et ANN. Les calculs itératifs effectués grâce au modèle $k - \epsilon$ standard ont une durée variant entre 20 minutes et une heure sur la même station de travail. Le temps de calcul de l'écoulement turbulent autour d'un obstacle cylindrique par la méthode présentée est donc extrêmement rapide et pertinent pour une utilisation en situation de crise ou afin de tester une grande variété de scénarios.

Les résultats présentés dans les trois paragraphes précédents ont montré la possibilité de prévoir les paramètres dirigeant la dispersion autour d'un obstacle cylindrique. En ayant connaissance du champ de vitesse autour de l'obstacle et de la distribution spatiale du coefficient de diffusion turbulente, il est possible de simuler numériquement l'équation d'advection diffusion pour chaque cas test. De multiples solutions existent et le paragraphe suivant utilise une discrétisation par différences finies. Cette méthode est détaillée en annexe 1. Celle-ci sera utilisée avec les paramètres issus du calcul CFD et du calcul ANN afin de pouvoir comparer les deux méthodes.

III-4.6. Application aux cas non-appris

Pour chacun des cas non-appris détaillés au III-4.5, il est possible de simuler la dispersion d'une bouffée de gaz sans vitesse initiale dans un écoulement stationnaire. Ainsi, dans l'évaluation du modèle complet, une bouffée de fraction massique égale à 1 et de forme carrée est relâchée en amont de l'obstacle sur l'axe central. Son évolution est suivie au cours du temps. Les critères de performance développés au I-1 sont utilisés pour évaluer la qualité du modèle.

Le tableau ci-après reporte les valeurs de ces critères évalués en aval de l'obstacle, proche de la fin du domaine, à une distance de 10 diamètres environ.

Cas	Diamètre (en m)	Vitesse (m.s^{-1})	R^2	FAC2	FB	NMSE
1	12	2,5	0,99	0,97	0,09	0,01
2	12	5,5	0,98	0,94	0,09	0,03
3	12	9,5	0,96	0,91	0,09	0,05
4	26	2,5	0,94	0,88	0,04	0,07
5	26	5,5	0,98	0,93	0,04	0,02
6	26	9,5	0,99	0,94	0,04	0,01
7	50	2,5	0,88	0,81	0,09	0,15
8	50	5,5	0,91	0,85	0,08	0,10
9	50	9,5	0,98	0,85	0,05	0,02

L'ensemble des valeurs des indicateurs est comprise dans les intervalles indiqués par Chang et Hanna (2004). Des disparités existent entre les cas testés. Le graphique ci-après permet de situer l'ensemble des cas modélisés en fonction de l'erreur systématique et de l'erreur totale.

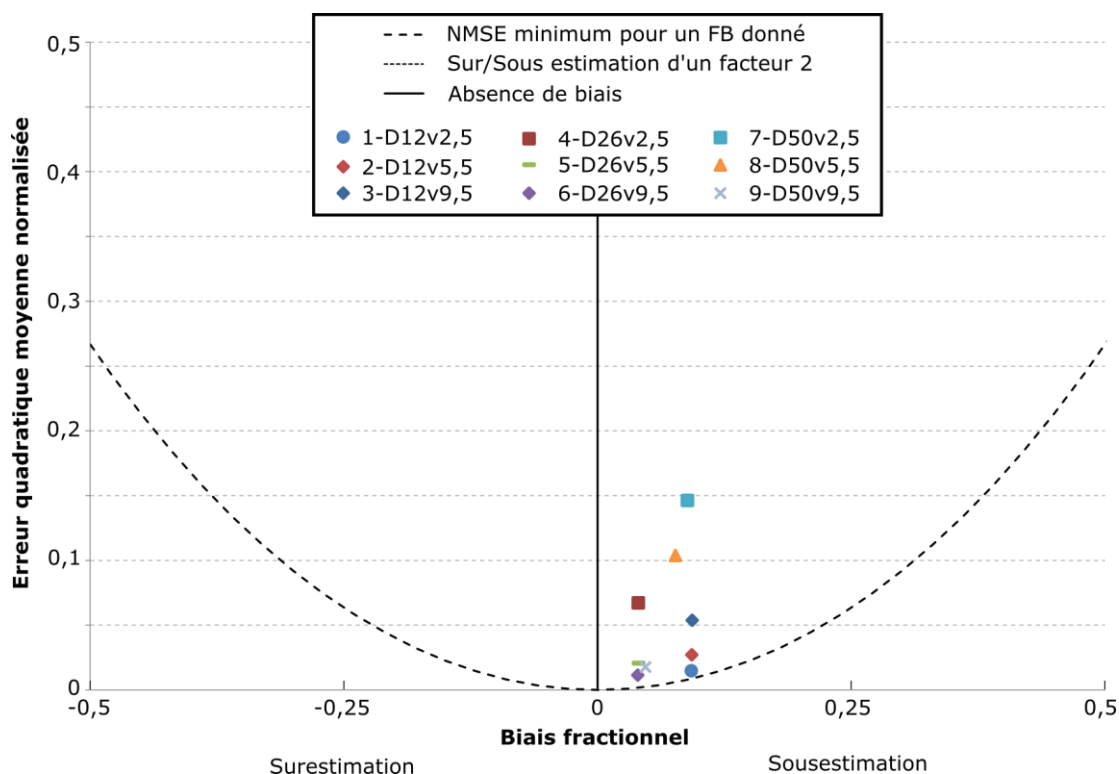


Figure III-78 : Erreur systématique (FB) et erreur totale (NMSE) pour les différents cas non-appris en sortie de domaine (D : Diamètre – v : Vitesse).

Les différents cas non-appris sont proches des axes ce qui est la traduction d'une modélisation précise. L'ensemble des cas est situé dans les valeurs positives du biais fractionnel ce qui traduit une légère sous-estimation des valeurs de fraction massique. Certains cas sont mieux modélisés que d'autres. Le cas 6 est le plus précis alors que le 7 l'est moins. Cependant, les valeurs relevées sont dans des proportions satisfaisantes.

La qualité du modèle peut être appréciée visuellement dans un premier temps, afin d'extraire une tendance générale. Cette analyse est effectuée sur le cas 6, présentant un diamètre intermédiaire de 26 m et une vitesse forte de $9,5 \text{ m.s}^{-1}$.

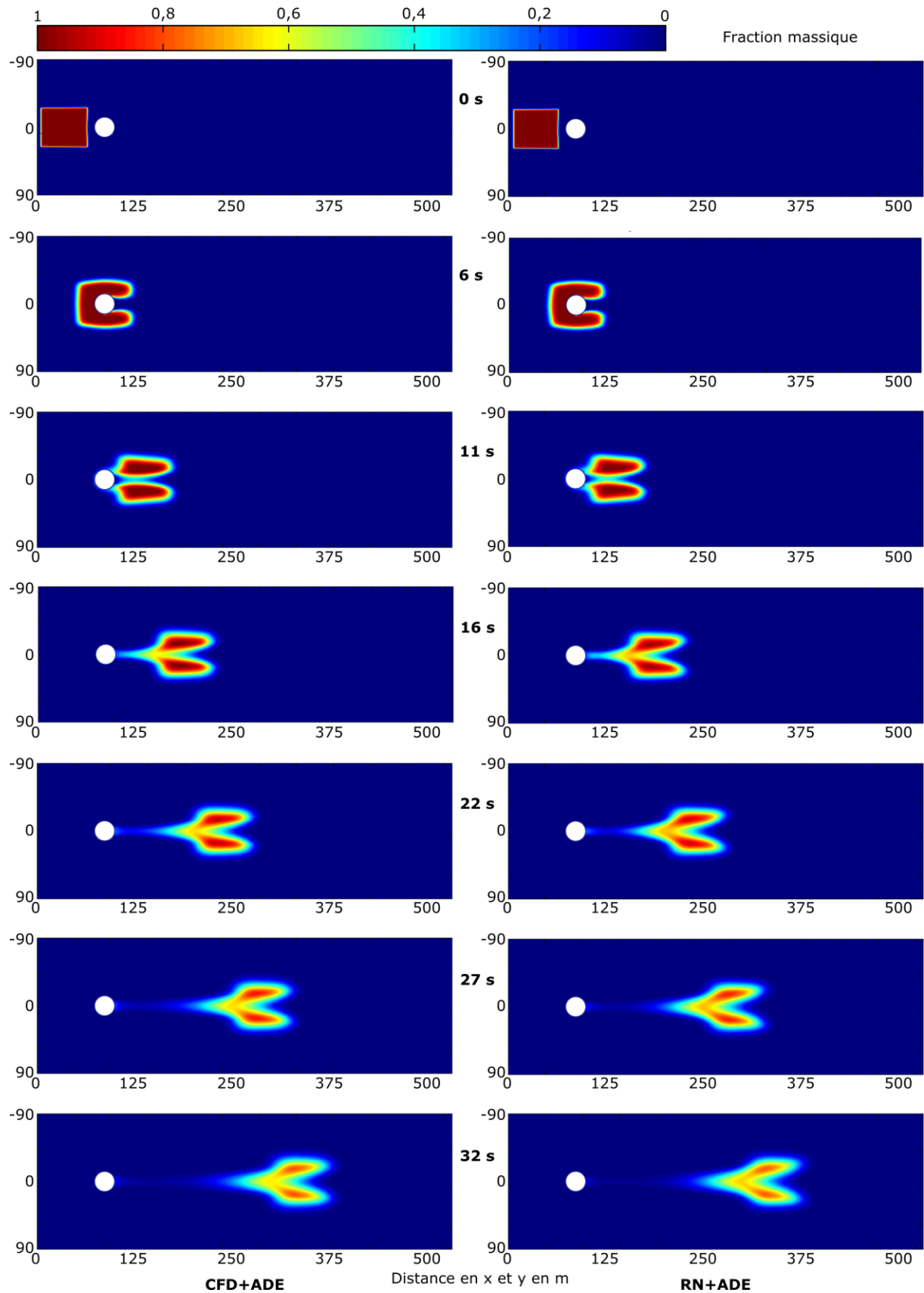


Figure III-79 : Visualisation de la dispersion d'une bouffée autour d'un obstacle. Comparaison entre la réalité CFD et la modélisation ANN pour le cas favorable (Diamètre : 26 m - Vitesse du vent : $9,5 \text{ m.s}^{-1}$)

La figure ci-avant montre que l'application de l'ADE à partir des paramètres modélisés par réseaux de neurones est similaire à la situation référence, représentée par l'utilisation des paramètres issus de simulations CFD. La dilution du nuage semble un peu moins importante dans le cas modélisé mais il est nécessaire de regarder plus en détail. Le graphe de dispersion confirme une bonne prédiction générale : 99% des valeurs estimées se situent dans un facteur deux des valeurs CFD, la plupart des points étant peu éloignés de la première bissectrice :

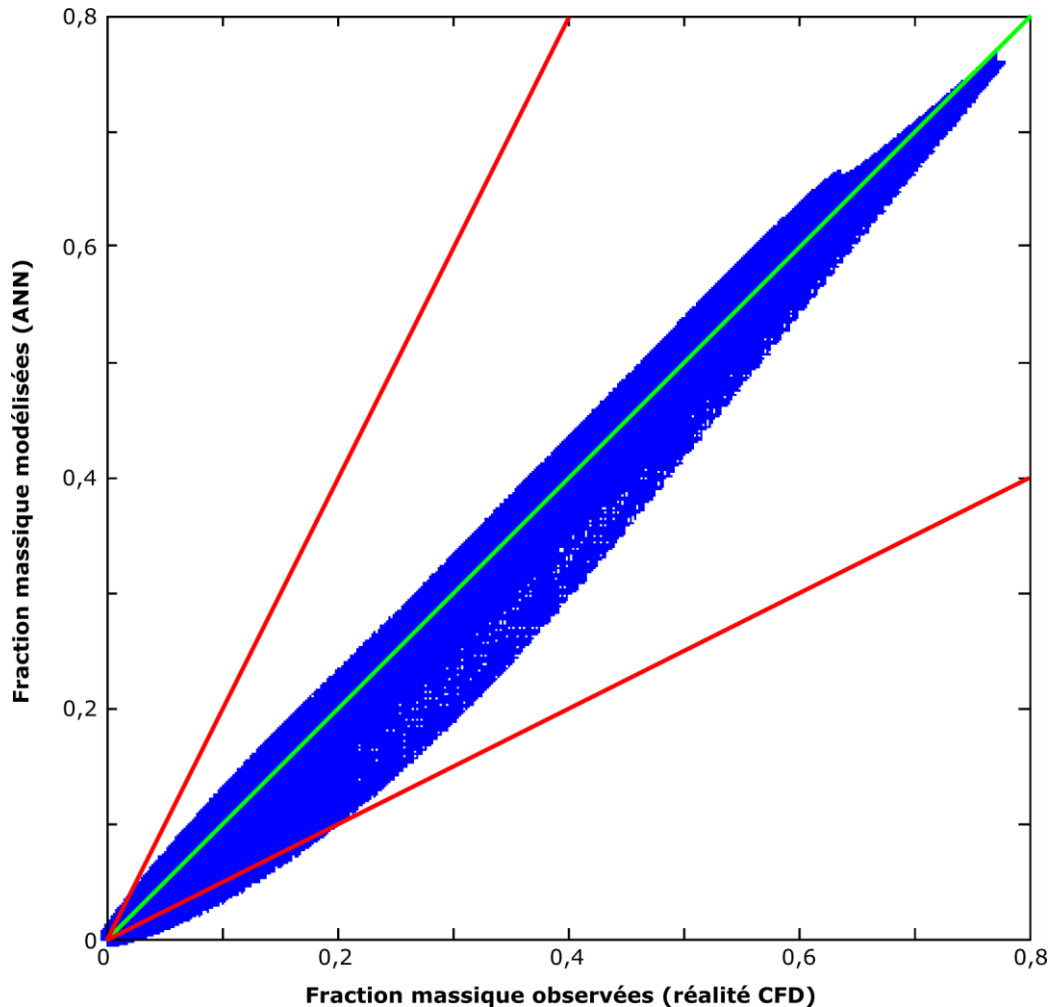


Figure III-80 : Graphe de dispersion entre les fractions massiques observées (réalité CFD) et modélisées (ANN) le cas 6 (diamètre : 26 m – vitesse de vent : $9,5 \text{ m.s}^{-1}$) au temps $t=101 \text{ s}$. La droite verte représente la première bissectrice $y=x$. Les droites rouges représentent les sous-estimations et surestimations d'un facteur deux.

Une légère tendance à la sous-estimation est observée, notamment pour les valeurs moyennes à faibles. Cela vient confirmer la valeur du biais fractionnel, positive. Peu de points se situent en dehors d'un facteur deux de la simulation CFD. Les points figurant sur ce graphique ne permettent pas de localiser les zones où la modélisation est pertinente et là où elle l'est moins.

La figure suivante présente les fractions massiques dans l'axe central pour le même pas de temps :

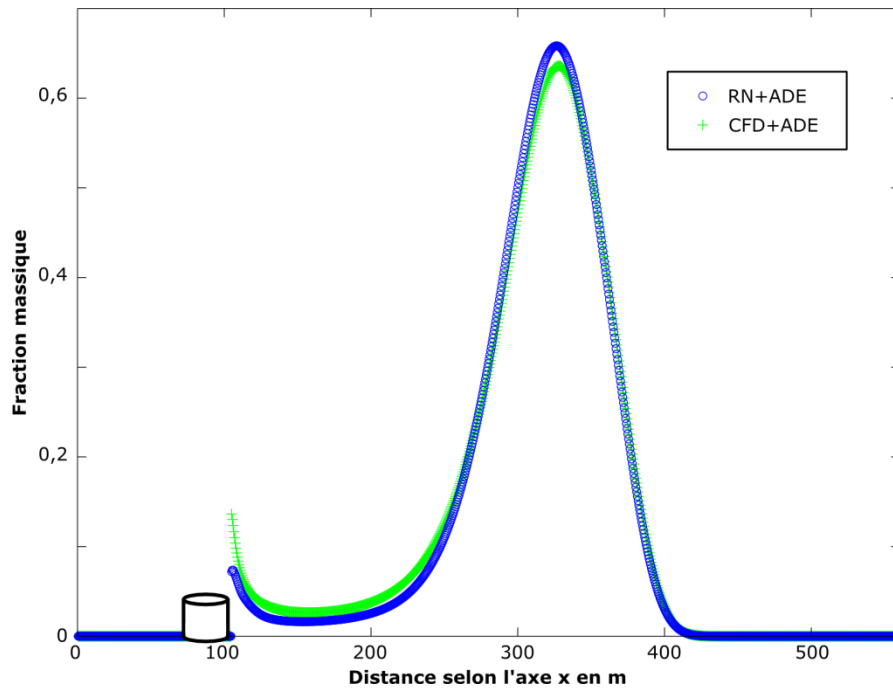


Figure III-81 : Fraction massique dans l'axe central pour simulation CFD et la modélisation ANN au temps $t=101$ s.

Les deux courbes se confondent sur la majorité du domaine. L'allure générale de la fraction massique modélisée est similaire à la référence. La fraction massique maximale est surestimée par le modèle tandis que les valeurs derrière l'obstacle sont sous-estimées. L'ensemble est satisfaisant. En s'intéressant à une coupe selon l'axe y au maximum de concentration, il est possible d'identifier la surestimation au pic de fraction massique. Par ailleurs, la dilution semble moins importante dans le cas de la modélisation, conduisant à une sous-estimation des fractions massiques plus faibles (figure ci-après).

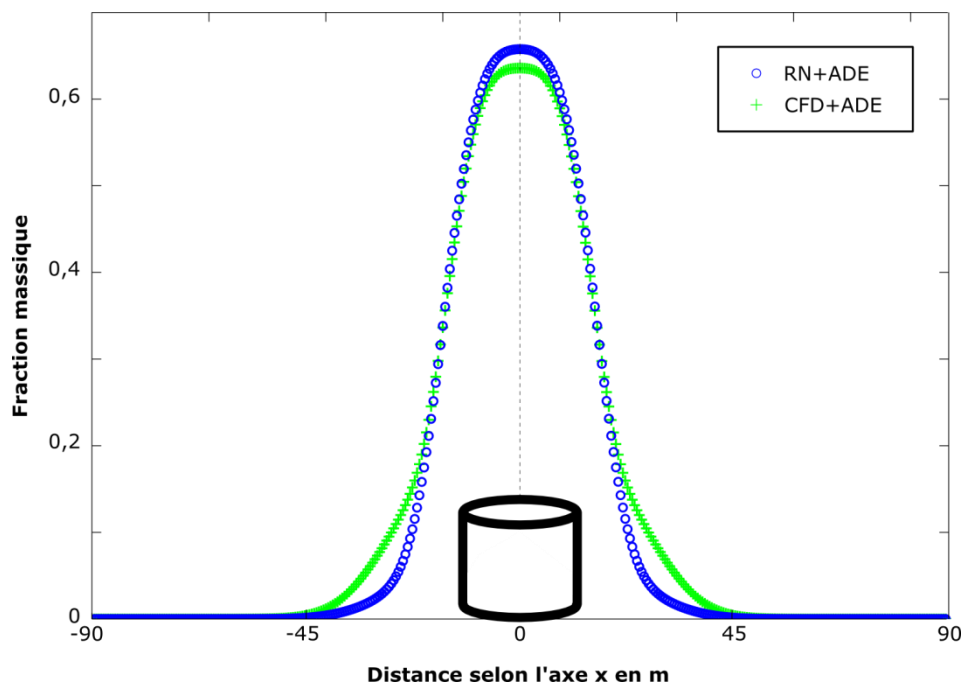


Figure III-82 : Fraction massique pour une coupe selon y au niveau de la concentration maximale ($x = 330$ m) pour la simulation CFD et la modélisation ANN au temps $t = 101$ s.

Ces observations sont effectuées à un instant donné ($t = 101$ secondes). L'évolution temporelle de la bouffée peut être évaluée à travers le suivi des indicateurs au cours du temps. Le coefficient de détermination présente une valeur initiale inférieure à 0,9. Cela peut correspondre à un léger décalage du nuage, impliquant des écarts importants à sa marge. Par la suite, la valeur se stabilise à partir de $t = 6$ secondes. Cela correspond au passage de la bouffée autour de l'obstacle et donc une harmonisation des écarts et une dilution importante de la bouffée. En ce qui concerne le FAC2, la même tendance générale est observée en début de modélisation. La valeur diminue ensuite puis se stabilise.

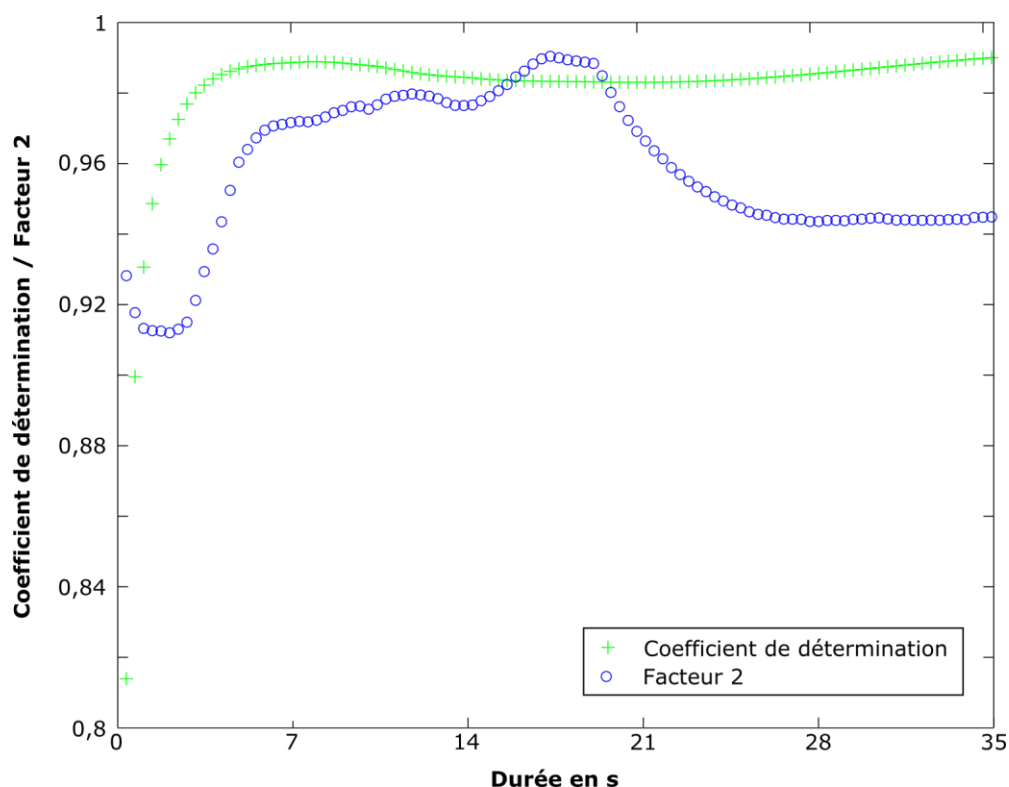


Figure III-83 : Evolution du coefficient de détermination et du facteur 2 au cours de la modélisation pour le cas 6.

L'évolution de l'erreur totale et de l'erreur systématique figure sur le graphique ci-après. Une observation importante est la tendance générale de l'erreur. L'erreur systématique reste globalement constante au cours des pas de temps, illustrant une légère sous-estimation globale (valeur positive). L'erreur quadratique moyenne normalisée suit, elle, la même tendance que le R2 et le FAC2. Une diminution de sa valeur dans un premier temps puis une légère augmentation, se stabilisant avec l'augmentation des pas de temps.

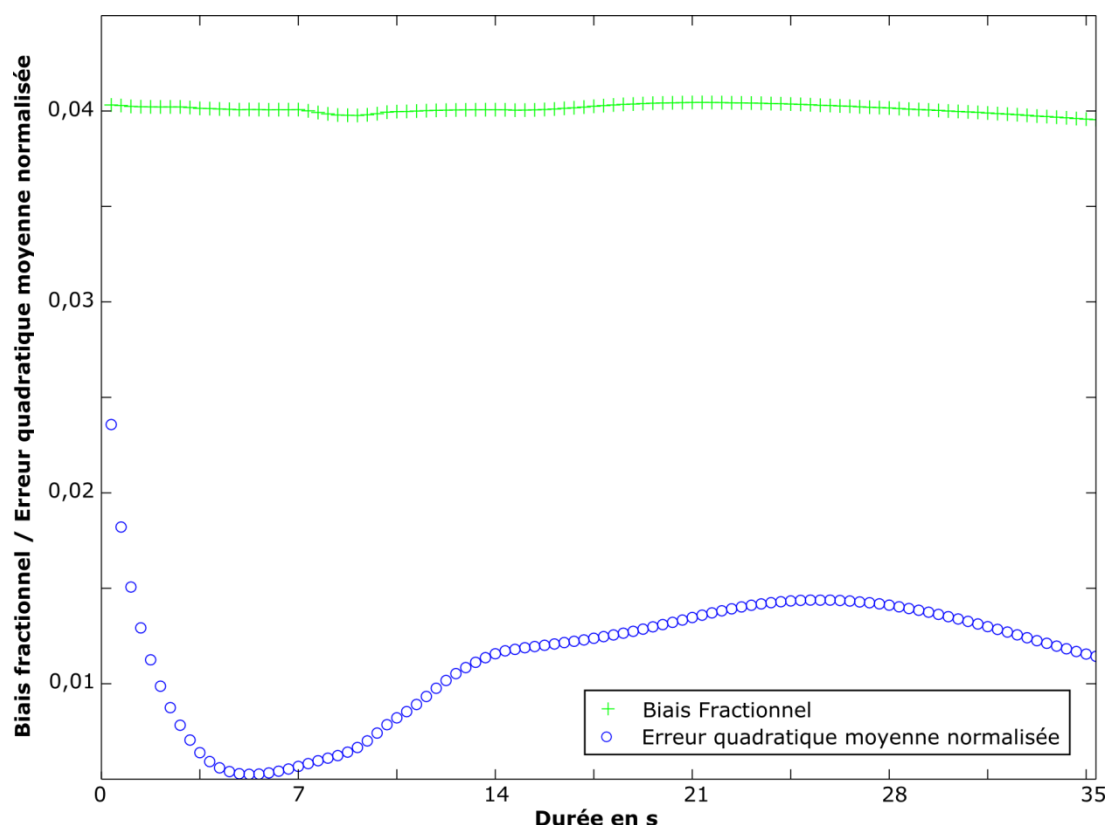


Figure III-84 : Evolution du FB et du NMSE au cours de la modélisation pour le cas 6.

III-4.7. Conclusion

Le modèle développé dans ce chapitre est adapté à la modélisation de la dispersion atmosphérique autour d'un obstacle cylindrique. Une base de données d'écoulements atmosphériques en condition de stabilité neutre autour d'un obstacle a été générée. Celle-ci a permis l'apprentissage de trois réseaux de neurones, chacun permettant de modéliser un des trois paramètres déterminant pour la modélisation bidimensionnelle de la dispersion d'un polluant à partir de l'équation d'advection diffusion : vitesse selon x , vitesse selon y et coefficient de diffusion turbulente. La sélection du meilleur modèle possible pour chacun des trois réseaux de neurones s'est effectuée à partir de la sélection des variables d'entrée en passant par l'échantillonnage de la base de données en base d'exemples et par l'optimisation de l'architecture du réseau. L'évaluation de la qualité des différents apprentissages sur des cas non-présents dans la base de données initiale a démontré les performances de généralisation associées aux réseaux appris. Des disparités apparaissent dans les valeurs des critères de performance pour les différents cas non-appris. L'impact de ces disparités sur le résultat final est évalué lors de la résolution par différences finies de l'équation d'advection diffusion. En simulant le transport turbulent d'une bouffée passive utilisant les paramètres modélisés et issus de simulations CFD, les valeurs des critères de performance obtenus sont satisfaisants. De plus, l'évolution des erreurs commises par le modèle ne conduit pas à des instabilités. Au contraire, ces erreurs se stabilisent. Effectivement, le calcul de la concentration au pas de temps suivant étant réalisé par la résolution de l'équation d'advection diffusion, il n'y a pas de bruit introduit par une estimation. Il existe cependant une approximation, due au schéma numérique de discrétisation. Par ailleurs, la modélisation des paramètres par réseau de neurones indépendamment de la variable temporelle permet d'utiliser des méthodes des schémas numériques connus et éprouvés. La validité de ces schémas n'est pas évaluée ici. On notera simplement que de la diffusion numérique est présente dans la

résolution de l'ADE par utilisation des paramètres issus de la CFD ou de leur modélisation par réseau de neurones. Ainsi, la comparaison est possible en dehors de toutes considérations sur le schéma numérique employé.

Le temps de calcul est un élément important en faveur de la modélisation du champ de vitesse et du coefficient de diffusion turbulente à partir des réseaux de neurones. Effectivement, pour un domaine de plus de 400 000 cellules de 0,5 m de côté (180 m de large pour 550 m de long), le temps de calcul est inférieur à 2 secondes pour chaque paramètre (vitesses selon x et y et coefficient de diffusion turbulente) contre 20 minutes à minima pour un calcul CFD à partir d'un modèle $k - \epsilon$ standard pour le même domaine. Le temps de calcul de l'advection et la diffusion entre les deux modèles n'est pas un élément de différenciation, la même méthode étant appliquée avec les paramètres issus de la CFD et modélisés par les réseaux de neurones. Celle-ci est effectuée en moins de 5 minutes, en fonction du domaine représenté. Cette durée peut être réduite par l'emploi de schémas numériques plus rapides et par une optimisation de codage.

Conclusions

« C'est toujours en bout de piste qu'on a envie de décoller enfin. »

Thomas Boulard

Les conséquences d'accidents majeurs sur des sites industriels peuvent être dévastatrices et sont difficilement maîtrisables. Une dispersion atmosphérique de substances est souvent observée dans la chaîne d'évènements menant à l'accident, que ce soit une étape ou l'état final. Sur un site industriel, les zones de stockage de produits sont les plus susceptibles d'émettre en grande quantité des fuites de substances dangereuses. Elles sont principalement contenues dans des cuves de formes cylindriques. Les écoulements autour de tels obstacles comportent des perturbations spécifiques, notamment dans les conditions turbulentes liées à la couche limite atmosphérique. La dispersion atmosphérique est une combinaison des conditions météorologiques globales et de la configuration de l'environnement proche.

Les réglementations française et européenne imposent des études de dangers aux exploitants afin d'évaluer les conséquences d'un accident sur leur installation. Ces études cartographient notamment les zones de différents niveaux d'aléas. Par ailleurs, en situation de crise lorsque l'accident se produit, il est nécessaire de pouvoir prévoir rapidement les zones à évacuer en priorité. Pour répondre à cette problématique, les industriels peuvent s'appuyer sur les apports de la simulation numérique. Celle-ci permet à la fois d'éviter des essais difficiles à mettre en œuvre et à reproduire mais aussi de multiplier les scénarios pris en compte. Ainsi, l'utilisation d'un modèle doit permettre de répondre à des exigences de précision (prévisions conformes à la réalité), de rapidité de calcul et de facilité de paramétrisation. La première exigence est fonction du niveau de détail requis : la prise en compte des obstacles est indispensable dans la détermination de la dispersion atmosphérique sur site industriel. Les différentes situations à envisager nécessitent de nombreuses simulations, réalisables à la condition que le modèle utilisé soit rapide. Enfin, dans le cas de la situation de crise, les informations peuvent manquer pour décrire parfaitement la situation. Il est alors nécessaire que le modèle puisse prendre en compte cette composante.

Les modèles de CFD permettent une grande précision sur l'évolution de la concentration d'une substance dans un environnement proche et complexe mais les temps de calcul associés sont importants. Par ailleurs, des modèles analytiques permettent d'obtenir des résultats très rapidement au prix d'hypothèses importantes. Des alternatives ont également été proposées, afin d'allier rapidité et précision. Celles-ci sont généralement adaptées à un site unique d'étude.

L'objectif de ce travail était de proposer une méthodologie permettant de répondre aux impératifs de rapidité et de précision de la prévision de la dispersion atmosphérique en situation de crise, sur un site à configuration quelconque et ne nécessitant pas une expertise poussée.

Ce travail de thèse a utilisé les réseaux de neurones, modèles issues de l'apprentissage statistique afin de répondre aux exigences de précision et de rapidité. Leur utilisation a nécessité la constitution et l'apprentissage de bases de données. Devant la complexité de la réalisation de bases de données exhaustives à partir d'essais expérimentaux, le choix a été d'utiliser le modèle de CFD RANS $k - \epsilon$ standard en tant que réalité simulée.

Trois méthodologies ont été développées, prenant en compte les contraintes de manière croissante.

La première méthodologie s'est concentrée sur la dispersion continue d'une source de méthane en champ libre dans un écoulement stationnaire. Un réseau de neurones a été utilisé pour modéliser la concentration pour chaque point du domaine considéré. Les entrées utilisées concernent les conditions atmosphériques, les données relatives à la source du rejet et les coordonnées du point à modéliser. Une base de données de 90 simulations a été générée. Elles ont été réalisées à partir de la compréhension de la physique régissant les écoulements atmosphériques libres à l'intérieur de la couche de surface. L'apprentissage de cette base de données par le réseau de neurones a été effectué en identifiant les paramètres influants : échantillonnage, architecture et initialisation d'un perceptron à deux couches. La vérification du modèle a été réalisée par la comparaison des résultats sur des cas non-appris simulés avec le modèle CFD. La performance du modèle développé montre une modélisation rapide du champ de concentration. La précision de ce modèle s'approche de la précision des résultats obtenus par simulation CFD. Une légère surestimation des cas est observée ainsi que certaines difficultés à modéliser des hauts gradients de concentration. Cette première méthodologie a permis de valider la possibilité d'utilisation d'une réalité simulée pour la création d'un modèle utilisant les réseaux de neurones. D'autre part, la vitesse de calcul est 600 fois plus rapide en moyenne démontrant la pertinence d'un tel modèle pour ce critère par rapport à une utilisation opérationnelle.

Les automates cellulaires en combinaison avec les réseaux de neurones (CA-ANN) ont été introduits dans la deuxième méthodologie pour la modélisation de l'évolution d'une bouffée de méthane dans un écoulement stationnaire. La règle de transition de l'automate cellulaire est réalisée par un réseau de neurones récurrent dont les entrées sont issues de la discrétisation de l'équation d'advection diffusion. La base de données est constituée de 95 simulations. Elle fut générée à partir de simulations bidimensionnelles de l'évolution d'un nuage de méthane dans une atmosphère libre dans des conditions de stabilité neutres. Une prise en compte implicite de la turbulence est réalisée pour garantir une simplicité d'utilisation. Le coefficient de diffusion turbulente ne fait donc pas partie des entrées du réseau de neurones. La résolution de la diffusion turbulence est incluse dans le modèle à travers l'apprentissage de l'évolution de la concentration. Les résultats montrent des simulations correctes de cas tests non-appris (R^2 supérieur à 0,85) pour un nombre d'itérations inférieur à 50. Néanmoins, la comparaison avec les simulations CFD montre une dégradation de la modélisation au cours des pas de temps. Les erreurs se propagent et s'amplifient à chaque passage démontrant une certaine instabilité du modèle. Le CA-ANN permet donc de prévoir correctement l'évolution d'une bouffée d'une substance à la condition de limiter le nombre d'itérations de la règle de transition. Le calcul par le CA-ANN pour une durée de simulation fixée dans le domaine considéré dans ce travail est 1,5 fois plus rapide que le modèle CFD. Le CA-ANN calcule les concentrations dans chaque cellule à chaque pas de temps. Si ces résultats sont demandés par l'utilisateur, alors le modèle est environ 120 fois plus rapide que les simulations CFD. Par conséquent, la modélisation de l'évolution d'une bouffée de substance par cette méthode est possible et peut nécessiter des adaptations sur les dimensions des mailles selon la cinétique du phénomène.

Dans les simulations numériques d'une couche limite atmosphérique en champ libre, la turbulence est généralement considérée comme homogène horizontalement. Cette condition n'est plus vérifiée lorsqu'un obstacle est rencontré dans l'écoulement. En plus de la modification du champ de vitesse, le coefficient de diffusion turbulente varie également spatialement. Les limitations observées lors de l'évaluation de la méthode précédente ont conduit au développement d'une dernière méthode afin de répondre à cette problématique. Ce choix argumenté a été détaillé dans une publication fournie en annexe (Lauret et al. 2014). Cette troisième méthode est basée sur une séparation du problème de la dispersion turbulente en deux étapes : le calcul des caractéristiques de l'écoulement autour d'un cylindre vertical puis le calcul de la dispersion atmosphérique d'un composé. La première étape est réalisée par trois réseaux de neurones distincts, chacun permettant le calcul d'une des trois caractéristiques principales de l'écoulement : vitesse selon x (sens du vent) et selon y (perpendiculaire au vent), coefficient de diffusion turbulente. Les entrées du réseau de neurones sont du même type que pour le premier modèle : la vitesse du vent en entrée qui est supposée homogène, le rayon de l'obstacle ainsi que les coordonnées du point à modéliser. Les trois réseaux de neurones ont été appris par utilisation de la même base de données de 72 simulations différentes. Les résultats des différents apprentissages montrent de bonnes capacités de généralisation de la modélisation. L'application à la prévision d'écoulements autour d'un obstacle pour des cas tests disjoints de la base d'apprentissage permet de prédire correctement les champs horizontaux des différentes grandeurs. La modélisation de l'écoulement par les réseaux de neurones indépendamment de la variable temporelle permet d'utiliser des schémas numériques connus et éprouvés pour résoudre l'équation d'advection diffusion. Les différences finies ont été utilisées dans ce travail. La comparaison par la résolution de l'ADE à partir des prévisions des réseaux de neurones sur un maillage homogène d'une part et de l'interpolation des valeurs obtenues par simulation CFD d'autre part montre la validité de la méthode. L'erreur commise pour la détermination de l'écoulement ne se propage pas et permet d'obtenir des résultats satisfaisant pour des durées de dispersion supérieures à la minute. La prise en compte de la perturbation de l'écoulement par l'obstacle est effective. La précision de ce modèle est donc proche de la précision du modèle CFD utilisé pour la création de la base de données. Par ailleurs, le temps de calcul est grandement amélioré. Effectivement, le calcul de l'écoulement est quasi instantané et permet d'envisager l'utilisation de ce modèle en situation de crise.

En conclusion de ce travail de thèse, plusieurs modèles ont été développés pour répondre à différents cas rencontrés en risque industriel. Les outils employés que sont les réseaux de neurones et les automates cellulaires permettent de répondre aux exigences formulées par l'étude du contexte. Le temps de calcul des différents modèles est plus court que ceux obtenus par simulation CFD et la performance des modèles est du même ordre de grandeur. Cependant, la phase de constitution de la base de données est longue, tout comme la phase d'optimisation du modèle neuronal. La prise en compte de la perturbation de l'écoulement par un obstacle est effective. La zone susceptible d'être simulée correspond au champ proche, avec des échelles temporelles de l'ordre de quelques minutes et des échelles spatiales inférieures au kilomètre.

Perspectives

Les méthodes développées dans ce travail de thèse peuvent être complétées selon plusieurs aspects.

► **Perspectives d'améliorations de l'automate cellulaire à règle de transition réseau de neurones (CA-ANN)**

L'automate cellulaire à règle de transition de type réseau de neurones ou CA-ANN développé au III-3 permet de reproduire l'évolution d'un nuage d'une substance en champ libre avec cependant une augmentation de l'erreur avec le nombre d'itérations. Le réseau de neurones utilisé, de type récurrent, semble devenir instable au cours de la modélisation. En tenant compte de cette tendance, il est possible d'envisager l'utilisation de cette méthode sur des échelles plus vastes, mettant en jeu des cellules plus grandes permettant d'utiliser un pas de temps plus important. Il serait ainsi possible de simuler une même durée avec un nombre de pas de temps inférieur afin d'éviter l'atteinte d'une instabilité.

► **Perspectives d'améliorations du modèle de prévisions de la perturbation d'un écoulement par un obstacle cylindrique (III-4)**

En ce qui concerne cette méthodologie, des améliorations peuvent être également apportées. L'objectif majeur est la prise en compte de plusieurs obstacles dans le champ d'écoulement tout en gardant la contrainte d'une configuration inconnue au préalable et donc non incluse dans la base d'apprentissage du modèle. Dans une telle configuration, une première approche est de considérer chaque obstacle séparément. L'écoulement obtenu est correctement modélisé d'après les résultats décrits dans ce travail. Il est ainsi possible d'effectuer des combinaisons linéaires de ces différents résultats pour obtenir une première approximation. Cette approche n'est pas un résultat correct au niveau physique. Il devient alors nécessaire d'effectuer un bilan de la quantité de mouvement.

Par ailleurs, si les obstacles cylindriques ont été développés dans ce travail puisqu'en nombres plus importants sur les sites industriels, d'autres formes peuvent également être rencontrées. D'autres stockages peuvent exister sous formes de cylindres horizontaux ou de sphères. Les bâtiments administratifs sont généralement de formes rectangulaires. La multitude de rapports de formes possible est un élément qui peut empêcher l'utilisation des réseaux de neurones pour la détermination de l'écoulement autour de tels obstacles à cause de difficultés à aborder chaque situation spécifique. Cependant, dans un objectif de respecter des temps de calcul relativement courts, il pourra être envisagé d'utiliser les transformations conformes pour traiter ce genre de cas. Cette méthode, utilisée au début de l'aérodynamique, permet, par exemple, de transposer un écoulement autour d'un cercle à un écoulement autour d'un profil d'aile. Cette piste est à approfondir.

Les bâtiments « procédés » posent également des défis supplémentaires. Ces zones composées d'un grand nombre de tuyauteries et d'éléments permettant la réalisation d'un procédé ne présentent pas une étanchéité parfaite d'un point de vue global. Effectivement, l'absence d'enveloppe hermétique et la compacité pouvant s'éloigner largement de l'unité impliquent une dispersion à l'intérieur de ces unités. De plus, la

détermination d'un écoulement turbulent se heurte à la complexité de ces géométries particulières. A l'échelle de modélisation considérée, une piste de réflexion est la prise en compte d'une porosité de ces unités pouvant être traversées à la fois par l'écoulement et les substances relâchées. La méthode pour caractériser cette porosité reste à décrire et à implémenter.

Le passage à la modélisation tridimensionnelle doit permettre de déterminer les écoulements autour d'obstacles différents tels que les sphères. Ces éléments de stockage très présents sur les sites industriels présentent un type d'écoulement ne permettant pas de les étudier en deux dimensions uniquement. D'autre part, la topographie des sites industriels peut ne pas être complètement plane et induire des variations de l'écoulement selon l'axe z . La prise en compte de cette composante permet d'envisager notamment l'ajout de la convection thermique, largement tributaire de la nature du sol. Les méthodes jointes à ces modifications devront être développées. Cela pourra s'effectuer à travers l'ajout d'entrée(s) au modèle neuronal.

► **Prise en compte de données réelles pour l'apprentissage des bases de données**

Dans ce travail, trois bases de données différentes ont été créées : rejet stationnaire de méthane dans l'air, évolution temporelle de différentes bouffées dans un champ libre et simulation d'écoulements autour d'un obstacle. Ces bases de données ont été créées à partir d'un modèle de turbulence de mécanique des fluides numériques, le modèle $k - \epsilon$ standard. Comme toute modélisation, ce modèle tente de représenter la réalité. Il n'est cependant pas exempt de défauts. Le travail effectué ici a eu pour but de démontrer les possibilités d'utilisation de modèles issus de l'intelligence artificielle en dispersion atmosphérique. L'intérêt de la base de données se situait dans sa capacité à être utilisée en tant que réalité, simulée dans ce cas. De plus, les difficultés techniques liées à l'obtention de données sans erreurs de mesure sont ainsi évitées. L'amélioration des modèles pourra être effectuée en utilisant cette méthodologie à partir d'une base de données mesurées. Celles-ci représentent la réalité, dans toute sa complexité et son caractère imprévisible. Afin de vérifier que ce travail de thèse est généralisable à une base de données réelle, la deuxième méthodologie a été appliquée à partir d'une base de données CFD bruitée artificiellement. Les détails de ces essais numériques sont disponibles en annexe 2. Les résultats montrent qu'il est tout fait possible d'envisager l'extension de cette méthode à l'apprentissage de données bruitées.

► **Vérification et validation d'un modèle**

Un travail de vérification sur un nombre important de cas tests et des essais de validation devront être effectués afin de qualifier ce modèle. Cette phase devra comprendre notamment une analyse de sensibilité. Enfin, la création de cet outil devra se confronter à la réalité du terrain par l'utilisateur en situation opérationnelle.

► **Améliorations techniques**

Sur le plan technique, une optimisation informatique du code de calcul devra être effectuée. La résolution de l'équation d'advection pourra, de ce fait, être accélérée et permettre l'utilisation du modèle en temps réel. La création de scénarios multiples devra être implémentée, notamment en ce qui concerne le terme source : rejet continu/instantané, quantité de mouvement initiale d'un jet, ... D'autre part, des informations importantes concernant la qualification du risque en fonction du scénario pourront être extraites : dose intégrée en un point, durée de persistance, ... Ces informations devront être complétées en fonction des besoins des utilisateurs.

Bibliographie

- Ak, R., Li, Y., Vitelli, V., Zio, E., 2013. A Genetic Algorithm and Neural Network Technique for Predicting Wind Power under Uncertainty. *Chem. Eng. Trans.* 33, 925–930.
- Allix, A., 1932. A propos des brouillards lyonnais. 4. Le brouillard mortel de Liège et les risques pour Lyon. *Les Etudes Rhodaniennes*. 8, 133–144.
- ANSYS® Academic Research, Release 14.0, Help System, Species Transport, ANSYS, Inc.
- ARIA, 2008. Fuite de gaz toxiques dans une usine agrochimique.
- Asch, G., 2010. Caractéristiques métrologiques, in: Dunod (Ed.), *Les Capteurs En Instrumentation Industrielle*. pp. 17–45.
- Avolio, M.V., Crisci, G.M., Di Gregorio, S., Rongo, R., Spataro, W., D'Ambrosio, D., 2006. Pyroclastic flows modelling using cellular automata. *Comput. Geosci.* 32, 897–911.
- Barron, a. R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* 39, 930–945.
- Bitsuamlak, G.T., Stathopoulos, T., Asce, F., Bédard, C., 2004. Numerical Evaluation of Wind Flow over Complex Terrain : Review. *J. Aerosp. Eng.* 135–145.
- Blocken, B., Gualtieri, C., 2012. Ten iterative steps for model development and evaluation applied to Computational Fluid Dynamics for Environmental Fluid Mechanics. *Environ. Model. Softw.* 33, 1–22.
- Boussinesq, J., 1977. Essai sur la théorie des eaux courantes. Mémoires présentés par Divers savants à l'académie des Sci. l'institut Fr. 23-24, 1–680.
- Boznar, M., Lesjak, M., Mlakar, P., 1993. A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. *Atmos. Environ. Part B. Urban Atmos.* 27, 221–230.
- Brambilla, S., Totaro, R., Manca, D., 2009. Simulation of the LPG release, dispersion, and explosion in the Viareggio railway accident. *Chem. Eng. Trans.* 19, 195–200.
- Brook, D.R., Beck, N. V., Clem, C.M., Strickland, D.C., Griffiths, I.H., Hall, D.J., D., K.R., M., H.J., 2001. Validation of the urban dispersion model (udm), in: 8th Int. Conf. on

- Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes. pp. 8–12.
- Buckingham, E., 1914. On physically similar systems; illustrations of the use of dimensional equations. *Phys. Rev.* 4, 345–376.
- Cao, X., 2007. Modelling the Concentration Distribution of Non-Buoyant Aerosols Released from Transient Point Sources into the Atmosphere. Queen's University, Ontario, Canada.
- Carruthers, D.J., Holroyd, R.J., Hunt, J.C.R., Weng, W.S., Robins, a. G., Apsley, D.D., Thompson, D.J., Smith, F.B., 1994. UK-ADMS: A new approach to modelling dispersion in the earth's atmospheric boundary layer. *J. Wind Eng. Ind. Aerodyn.* 52, 139–153.
- Cavallaro, A., Tebaldi, G., Gualdi, R., 1982. Analysis of transport and ground deposition of the TCDD emitted on 10 July 1976 from the ICMESA factory (Seveso, Italy). *Atmos. Environ.* 16, 731–740.
- Chamayou, R., 1997. Réservoirs métalliques : stockage des liquides . Généralités. *Tech. l'Ingénieur BM 6*, 1–7.
- Chang, J.C., Hanna, S.R., 2004. Air quality model performance evaluation. *Meteorol. Atmos. Phys.* 87, 167–196.
- Charbonnier, J., 2004. Bhopal : La pire catastrophe industrielle de tous les temps, 1st ed. Editions Préventique.
- Chopard, B., Dupuis, A., Masselot, A., Luthi, P., 2002. Cellular automata and lattice boltzmann techniques: an approach to model and simulate complex systems. *Adv. Complex Syst.* 5, 1–144.
- Courant, R., Friedrichs, K., Lewy, H., 1967. On the partial difference equations of mathematical physics. *IBM J. Res. Dev.* 11, 215–234.
- Da Vinci, L., 1508. Codex Leicester.
- Devito, T.J., Andrews, W.S., Roy, G., 2000. Modeling aerosol puff concentration distributions from point sources using artificial neural networks, in: *Proceedings of the 2000 Battlespace Atmospheric and Cloud Impacts on Military Operations Conference*. Cooperative Institute for Research in the Atmosphere and Army Research Laboratory, Fort Collins, Colorado, pp. 24–27.
- Doursat, R., Bienenstock, E., Geman, S., 1992. Neural Networks and the Bias/Variance Dilemma. *Neural Comput.* 4, 1–58.

- Doury, A., 1976. Une méthode de calcul pratique et générale pour la prévision numérique des pollutions véhiculées par l'atmosphère., Commissariat à l'Energie Atomique. Centre d'études nucléaires de Saclay.
- Dreyfus, G., 2004. Neural Networks, Methodology and Applications, Springer. ed, Neural Networks.
- Fletcher, C.A.J., 1991. Computational Techniques for Fluid Dynamics, Volume 1, Fundamental and General Techniques, 2nd editio. ed. Springer-Verlag, New York Berlin Heidelberg.
- Forestier, S., 2011. Etude de l'évaporation d'un liquide réponadu au sol suite à la rupture d'un stockage industriel. Ecole Nationale Supérieure des Mines de Saint Etienne.
- Frisch, U., Hasslacher, B., Pomeau, Y., 1986. Lattice-Gas Automata for the Navier-Stokes Equation. Phys. Rev. Lett. 56, 1505–1508.
- Gailis, R., 2004. Wind Tunnel Simulations of the Mock Urban Setting Test — Experimental Procedures and Data Analysis, Sciences-New York. Victoria, Australia.
- Galle, P., Paulin, R., Coursaget, J., 2003. Données métrologiques et évaluation des risques en France lors de l'accident de Tchernobyl (26 avril 1986). Mise au point historique. C. R. Biol. 326, 699–715.
- Gardner, M., 1970. Mathematical games: the fantastic combinations of John Conway's new solitaire game "life". Sci. Am. 223, 120–123.
- Gardner, M., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. Atmos. Environ. 32, 2627–2636.
- Gousseau, P., Blocken, B., van Heijst, G.J.F., 2011. CFD simulation of pollutant dispersion around isolated buildings: on the role of convective and turbulent mass fluxes in the prediction accuracy. J. Hazard. Mater. 194, 422–434.
- Guariso, G., Maniezzo, V., 1992. Air quality simulation through cellular automata. Environ. Softw. 7, 131–141.
- Hagan, M.T., Menhaj, M.B., 1994. Training Feedforward Networks with the Marquardt Algorithm. IEEE Transations Neural Networks 5, 989–993.
- Hardy, J., Pomeau, Y., De Pazzis, O., 1973. Time evolution of two-dimensional model system. I. Invariant states and time correlation functions. J. Math. Phys. 14, 1746–1759.
- Hargreaves, D.M., Wright, N.G., 2007. On the use of the k– model in commercial CFD software to model the neutral atmospheric boundary layer. J. Wind Eng. Ind. Aerodyn. 95, 355–369.

- Havens, J., Walker, H., Spicer, T., 2012. Bhopal atmospheric dispersion revisited. *J. Hazard. Mater.* 233-234, 33–40.
- Hernández-espinosa, M.F.C., 2001. Weight Initialization Methods for Multilayer Feedforward, in: *ESANN'2011: European Symposium on Artificial Neural Networks*. D-Facto public., Bruges, Belgium, pp. 119–124.
- Heseltine, J.L., 2003. Flow Around a Circular Cylinder with a Free End. University of Saskatchewan, Saskatoon, Canada.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2, 359–366.
- Hosker Jr, R.P., 1985. Flow and Diffusion Near Obstacles, in: *Atmospheric Science and Power Production*. pp. 241–326.
- Hutzler, G., 2006. Cellular automata and agent-based approaches for the modelling and simulation of biological systems : application to the lambda phage. *Model. Complex Biol. Syst. Context Genomics*. 1–17.
- IAEA, 2005. Chernobyl's Legacy: Health, Environmental and Socio-Economic Impacts. Vienne, Austria.
- Itami, M., 1994. Simulating spatial dynamics : cellular automata theory. *Landsc. Urban Plan.* 30, 27–47.
- Kaimal, J.C., Finnigan, J.J., 1994. Atmospheric boundary layer flows. Oxford University Press.
- Kari, J., 2005. Theory of cellular automata: A survey. *Theor. Comput. Sci.* 334, 3–33.
- Koeltzsch, K., 2000. The height dependence of the turbulent Schmidt number within the boundary layer. *Atmos. Environ.* 34, 1147–1151.
- Kolmogorov, A.N., 1941. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk SSSR* 30, 299–303.
- Korsakissok, I., 2009. Changements d'échelle en modélisation de la qualité de l'air et estimation des incertitudes associées. Université Paris-Est.
- Korsakissok, I., Mathieu, a., Didier, D., 2013. Atmospheric dispersion and ground deposition induced by the Fukushima Nuclear Power Plant accident: A local-scale simulation and sensitivity study. *Atmos. Environ.* 70, 267–279.
- Launder, B.E., Spalding, D.B., 1974. The numerical computation of turbulent flows. *Comput. Methods Appl. Mech. Eng.* 3, 269–289.

- Lauret, P., Heymes, F., Aprin, L., Johannet, A., Dusserre, G., Munier, L., Lapébie, E., 2013. Near Field Atmospheric Dispersion Modelling on an Industrial Site Using Neural Networks. *Chem. Eng. Trans.* 31, 151–156.
- Lauret, P., Heymes, F., Aprin, L., Johannet, A., Dusserre, G., Lapébie, E., Osmont, A., 2014. Atmospheric Turbulent Dispersion Modeling Methods using Machine learning Tools. *Chem. Eng. Trans.* 36, 517–522.
- Macdonald, R.W., Hall, D.J., 1998. An improved method for the estimation of surface roughness of obstacle arrays. *Atmos. Environ.* 32, 1857–1864.
- Malek, E., Davis, T., Martin, R.S., Silva, P.J., 2006. Meteorological and environmental aspects of one of the worst national air pollution episodes (January, 2004) in Logan, Cache Valley, Utah, USA. *Atmos. Res.* 79, 108–122.
- Margolus, N., 1984. Physics-like models of computation. *Physica* 10, 81–95.
- Marin, M., Rauch, V., Rojas-Molina, A., 2000. Cellular automata simulation of dispersion of pollutants. *Comput. Mater. Sci.* 18, 132–140.
- Mavroidis, I., Griffiths, R., Hall, D., 2003. Field and wind tunnel investigations of plume dispersion around single surface obstacles. *Atmos. Environ.* 37, 2903–2918.
- Meroney, R.N., 2004. Wind Tunnel and Numerical Simulation of Pollution Dispersion: A Hybrid Approach, in: *Simulation*. Hong Kong, pp. 60.
- Milliez, M., Carissimo, B., 2006. Numerical simulations of pollutant dispersion in an idealized urban area, for different meteorological conditions. *Boundary-Layer Meteorol.* 122, 321–342.
- Minsky, M., Papert, S., 1969. *Perceptrons*. MIT Press, Cambridge MA.
- Oberkampf, W.L., Trucano, T.G., 2002. Verification and Validation in Computational Fluid Dynamics. *Prog. Aerosp. Sci.* 38, 209–272.
- Ollinger, N., 2002. *Automates cellulaires : structures*. Ecole Normale Supérieure de Lyon.
- Oussar, 1998. Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus. *Réseaux*. Université Paris VI.
- Oussar, Y., Dreyfus, G., 2001. How to Be a Gray Box : The Art of Dynamic Semi-Physical Modeling. *Neural Networks* 14, 1161–1172.
- Ozgoren, M., Pinar, E., Sahin, B., Akilli, H., 2011. Comparison of flow structures in the downstream region of a cylinder and sphere. *Int. J. Heat Fluid Flow.* 14 (4), 225–235.

- Pasquill, F., 1961. The estimation of the dispersion of windborne material. *Meteorol. Mag.* 90, 33–49.
- Pelliccioni, a, Tirabassi, T., 2006. Air dispersion model and neural network: A new perspective for integrated models in the simulation of complex situations. *Environ. Model. Softw.* 21, 539–546.
- Pesatori, A.C., Consonni, D., Bachetti, S., Zocchetti, C., Bonzini, M., Baccarelli, A., Bertazzi, P.A., 2003. Short- and long-term morbidity and mortality in the population exposed to dioxin after the “Seveso accident”. *Ind. Health* 41, 127–38.
- Podnar, D., Koračin, D., Panorska, A., 2002. Application of artificial neural networks to modeling the transport and dispersion of tracers in complex terrain. *Atmos. Environ.* 36, 561–570.
- Pontiggia, M., Busini, V., Gattuso, M., Uguccioni, G., Rota, R., 2012. Consequences Assessment of an Accidental Toxic Gas Release Through a CFD Tool : Effect of the Terrain and Major Obstacles. *Chem. Eng. Trans.* 26, 537–542.
- Pontiggia, M., Derudi, M., Busini, V., Rota, R., 2009. Hazardous gas dispersion: a CFD model accounting for atmospheric stability classes. *J. Hazard. Mater.* 171, 739–47.
- Raga, G.B., Baumgardner, D., Castro, T., Mart, A., 2001. Mexico City air quality : a qualitative review of gas and aerosol measurements (1960 – 2000). *Atmos. Environ.* 35, 4041–4058.
- Reynolds, O., 1883. An Experimental Investigation of the Circumstances Which Determine Whether the Motion of Water Shall Be Direct or Sinuous, and of the Law of Resistance in Parallel Channels. *Philos. Trans. R. Soc. London* 174, 935–982.
- Richards, P.J., Hoxey, R., 1993. Appropriate boundary conditions for computational wind engineering models using the k-epsilon turbulence model. *J. Wind Eng. Ind. Aerodyn.* 47, 145–153.
- Riddle, A., Carruthers, D., Sharpe, A., McHugh, C., Stocker, J., 2004. Comparisons between FLUENT and ADMS for atmospheric dispersion modelling. *Atmos. Environ.* 38, 1029–1038.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Russo, L., Vakalis, D., Siettos, C., 2013. Simulating the Wildfire in Rhodes in 2008 with a Cellular Automata Model. *Chem. Eng. Trans.* 35, 1399–1404.
- Sarkar, C., Abbasi, S. a, 2006. Cellular automata-based forecasting of the impact of accidental fire and toxic dispersion in process industries. *J. Hazard. Mater.* 137, 8–30.

- Schleicher, D., 2013. Interview with John Horton Conway. *Not. AMS* 567–575.
- Schlesinger, S., 1979. Terminology for model credibility. *Simulation* 32, 103–104.
- Schlünzen, K.H., Grawe, D., Bohnenstengel, S.I., Schlüter, I., Koppmann, R., 2011. Joint modelling of obstacle induced and mesoscale changes—Current limits and challenges. *J. Wind Eng. Ind. Aerodyn.* 99, 217–225.
- Sharan, M., Gopalakrishnan, S.G., 1997. Bhopal gas accident: a numerical simulation of the gas dispersion event. *Environ. Model. Softw.* 12, 135–141.
- Singh, M.P., Ghosh, S., 1987. Bhopal gas tragedy: model simulation of the dispersion scenario. *J. Hazard. Mater.* 17, 1–22.
- Siou, L.K. a, Johannet, A., Borrell, V., Pistre, S., 2011. Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). *J. Hydrol.* 403, 367–380.
- Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B., Glorennec, P.Y., 1995. Nonlinear Black-Box Modeling in System Identification: a Unified Overview. *Automatica* 31, 1691–1724.
- Snyder, W.H., 1981. Guideline for fluid modelling of atmospheric dispersion. Office of Air Quality Planning and Standards, Research Triangle Park.
- Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B* 36, 111–147.
- Taneda, S., 1977. Visual study on unsteady separated flows around bodies. *Prog. Aerosp. Sci.* 17, 287–348.
- Taveau, J., 2010. Risk assessment and land-use planning regulations in France following the AZF disaster. *J. Loss Prev. Process Ind.* 23, 813–823.
- Tixier, J., Dusserre, G., Salvi, O., Gaston, D., 2002. Review of 62 risk analysis methodologies of industrial plants. *J. Loss Prev. Process Ind.* 15, 291–303.
- Toffoli, T., 1984. Cellular Automata as an alternative to (rather than an approximation of) differential equations in modeling physics. *Physica* 10D, 117–127.
- Vendel, F., 2011. Modélisation de la dispersion atmosphérique en présence d’obstacles complexes : application à l’étude de sites industriels. Ecole Centrale de Lyon.
- Vick, B., 2007. Multi-physics Modeling Using Cellular Automata. *Complex Syst.* 17, 65–78.
- Von Neumann, J., Burks, A.W., 1966. Theory of Self-Reproducing Automata.

- Wahid, H., Ha, Q.P., Duc, H., Azzi, M., 2013. Neural network-based meta-modelling approach for estimating spatial distribution of air pollutant levels. *Appl. Soft Comput.* 13, 4087–4096.
- Wieringa, J., 1993. Representative roughness parameters for homogeneous terrain. *Boundary-Layer Meteorol.* 63, 323–363.
- Wolfram, S., 1983. Statistical mechanics of cellular automata. *Rev. Mod. Phys.* 55, 601–644.
- Wolfram, S., 2002. *A New Kind of Science, Volume 5.* ed. Wolfram Media.
- Yanenko, N.N., 1971. The method of fractional steps: The solution of problems of mathematical physics in several variables. Springer-Verlag, New York.
- Yi, J., Prybutok, V.R., 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environ. Pollut.* 92, 349–357.

Sites internet :

- Air Liquide, A., 2014. Gas Encyclopedia - Methane [Document internet]. URL : <http://encyclopedia.airliquide.com/encyclopedia.asp>, consultation le 12/04/2014
- Ministère de l'écologie, du développement durable et de l'énergie, 2014. Inspection des installations classées [Document internet] URL : <http://www.installationsclassees.developpement-durable.gouv.fr>, consultation le 12/04/2014
- Sciare, J., 2014. La pollution aux particules fines en Ile-de-France caractérisée en temps réel [Document internet]. URL : <http://www2.cnrs.fr/presse/communiqu/3481.htm>, consultation le 10/04/2014

Liste des figures

Figure I-1 : Illustration de l'effet de smog sur la ville de Paris (copie d'écran du site lemonde.fr du 14.03.2014)	6
Figure I-2 : Cartographie à l'échelle nationale des pics de concentration en PM10 le vendredi 14 mars 2014 (capture d'écran du site http://www.prevoir.org/fr/prevision_pm10.php du 14 mars 2014).....	7
Figure I-3 : Zones de dispersion de la radioactivité entre le 26 avril et le 3 mai 1986 après l'accident de Tchernobyl (document de l'institut royal météorologique de Belgique, issu de Galle et al., 2003).....	8
Figure I-4 : Vue satellite du site pétrochimique de Lavéra, Martigues, France (copie d'écran Google Map ® 2014)	12
Figure I-5 : Stockages et conduites de fluides - Site de Lavéra Martigues - Photo Laurent Valencia ©.....	13
Figure I-6 : Illustration de la chaîne des conséquences de l'impact d'un projectile sur un stockage industriel.....	14
Figure I-7 : Schéma structurel de l'atmosphère terrestre et son profil de température (Korsakissok, 2009).....	18
Figure I-8 : Evolution de la couche limite convective et la couche limite stable en fonction du temps et du transfert de chaleur du sol vers l'atmosphère. (Kaimal and Finnigan, 1994).....	19
Figure I-9 : Illustration des différentes échelles d'observation des mécanismes de l'atmosphère (Schlünzen et al., 2011) et définition de la zone étudiée dans ce travail de recherche.	20
Figure I-10 : Figures de turbulence autour d'un obstacle (Da Vinci, 1508)	22
Figure I-11 : Exemple d'un enregistrement de la vitesse d'un écoulement turbulent représenté par les fluctuations autour de la valeur moyenne.	26
Figure I-12 : Cascade d'énergie des tourbillons et modélisation	28
Figure I-13 : Représentation des écoulements autour d'un cylindre pour différents nombre de Reynolds (A, B et C d'après Taneda, 1977 – D réalisé à l'Institut des Sciences des Risques, 2013).....	30
Figure I-14 : Différentes zones caractéristiques d'un écoulement autour d'un obstacle cylindrique (retranscription d'après Hosker Jr, 1985).....	31

Figure I-15 : Structures tridimensionnelles observées dans un écoulement autour d'un cylindre vertical (Heseltine, 2003).....	32
Figure I-16 : Représentation schématique des liens entre réalité physique, développement théorique et simulations informatiques - adapté de Meroney (2004).....	34
Figure I-17 : Illustration des mécanismes d'advection - diffusion et combinaison des deux phénomènes.	38
Figure I-18 : Schéma de fonctionnement du modèle de Vendel (2011)	43
Figure I-19 : Modélisation de la turbulence pour différents modèles	44
Figure II-1 : Échantillon de configuration de l'automate cellulaire "game of Life" (source : http://commons.wikimedia.org/wiki/File:Puffer_train.svg).....	48
Figure II-2 : Règle de transition de l'automate cellulaire n°90	48
Figure II-3 : Visualisation de l'automate cellulaire de Wolfram - règle 90 - 200 pas de temps	49
Figure II-4 : Différentes formes d'automates cellulaires	50
Figure II-5 : Voisinage de Moore (gauche) et Von Neumann (droite) pour différents rayons ..	51
Figure II-6 : Structure tridimensionnelle du voisinage de Von Neumann	54
Figure II-7 : représentation graphique d'une fonction sigmoïde (source : commons.wikimedia.org/wiki/File:Hyperbolic_Tangent.svg)	56
Figure II-8 : Neurone mathématique - Représentation avec une fonction d'activation de type sigmoïde.....	56
Figure II-9 : Réseau de neurones à n variables (bleu), N_c neurones cachés (rouge), N_s neurones de sortie (vert)	57
Figure II-10 : Illustration d'un réseau de neurones récurrent.....	58
Figure II-11 : Schéma d'un Perceptron Multicouches.....	58
Figure II-12 : Principe de l'apprentissage d'un ANN.....	60
Figure II-13 : illustration du dilemme biais-variance	62
Figure II-14 : Utilisation de la validation croisée.....	64
Figure II-15 : Schéma des différentes étapes de la définition du problème à la sélection du modèle le plus performant.	64
Figure III-1 : Cahier des charges du modèle.....	70
Figure III-2 : Proposition des différentes méthodes de modélisation à utiliser selon les cas à étudier	71
Figure III-3 : Mesure des angles θ_i , θ_{\max} et θ_{\min}	73
Figure III-4 : Schéma du cas CFD simulé et les conditions aux limites associées.....	73
Figure III-5 : Schéma des profils de vitesse, d'énergie cinétique turbulente et de son taux de dissipation utilisés en tant que conditions aux limites en entrée d'un domaine tridimensionnel et plan horizontal associé.	75

Figure III-6 : Représentation des éléments utilisés pour l'apprentissage.....	78
Figure III-7 : Représentation schématique des différents types d'erreurs : a. Modèle fidèle mais non juste (erreurs systématiques) b. Modèle juste mais non fidèle (erreurs aléatoires) c. modèle ni juste ni fidèle (erreurs systématiques et aléatoires) d. modèle juste et fidèle donc précis (peu d'erreurs). Adapté de Asch (2010).	80
Figure III-8 : Définition du maillage et des conditions aux limites de la simulation CFD	82
Figure III-9 : Représentation du contenu de la base de données CFD.....	83
Figure III-10 : Représentation de l'échantillonnage de la base de données CFD pour la création de la base de données ANN.	84
Figure III-11 : Entrées et sortie du réseau de neurones.	84
Figure III-12 : Influence de l'initialisation sur l'Erreur Quadratique Moyenne sur l'ensemble de Test.....	85
Figure III-13 : Influence du nombre de neurones dans la couche cachée sur la qualité de la généralisation (Valeurs médianes pour 20 initialisations – barres d'erreurs : minimum et maximum des 20 initialisations)	86
Figure III-14 : Influence du nombre d'exemples sur la qualité du modèle. (Valeurs médianes pour 20 initialisations – barres d'erreurs : minimum et maximum des 20 initialisations)	87
Figure III-15 : Influence du nombre d'exemples sur la qualité du modèle. Minimum de 20 apprentissages.	87
Figure III-16 : Valeurs du R^2 et du FAC2 pour les neuf cas tests – Comparaison entre les résultats de la simulation CFD et la modélisation ANN.	89
Figure III-17 : Performance du modèle ANN sur les neuf cas tests.....	90
Figure III-18 : Erreur systématique et aléatoire pour l'ensemble des cas tests	91
Figure III-19 : Cas 3 (moins bonne prédiction)- iso-surfaces CFD et iso-contours ANN.	91
Figure III-20 : Cas 3 – Agrandissement de la zone à plus hauts gradients de concentration	92
Figure III-21 : cas test 5 - iso-surfaces CFD et iso-contours ANN.....	92
Figure III-22 : Cas 5 – Agrandissement de la zone à plus haut gradients de concentration ..	93
Figure III-23 : Diagramme de dispersion des valeurs observées et modélisées pour le cas défavorable 3 (appairées dans l'espace sur la figure de gauche et par rang dans la figure de droite).....	94
Figure III-24 : Diagramme de dispersion des valeurs observées et modélisées pour le cas favorable 5 (appairées dans l'espace sur la figure de gauche et par rang dans la figure de droite).....	94
Figure III-25 : Schématisation du fonctionnement d'un réseau de neurones statique tel que celui développé au III-2 avec ajout de la variable temporelle.	96

Figure III-26 : Schématisation du modèle développé : les interactions locales sont prises en compte à travers un maillage régulier. La règle de transition de type réseau de neurones permet d'estimer la concentration dans chaque cellule au pas de temps suivant.	97
Figure III-27 : Représentation schématique du fonctionnement du prédicteur neuronal associé au modèle postulé « bruit d'état ». Evolution sur plusieurs pas de temps dk.	98
Figure III-28 : Schématisation de la mise à jour de l'état de l'automate cellulaire de manière synchrone au cours d'un pas de temps dk.....	99
Figure III-29 : Variables d'entrée et cible du Réseau de Neurones	100
Figure III-30 : Maillage et conditions aux limites du domaine de calcul de l'automate cellulaire	101
Figure III-31 : Procédure de mise à jour de l'état d'une cellule.....	102
Figure III-32 : Schéma du maillage CFD utilisé.....	103
Figure III-33 : Configuration du cas général pour la constitution de la base de données CFD	103
Figure III-34 : Représentation schématique de la base de données CFD.	104
Figure III-35 : Représentation schématique de l'interpolation du maillage CFD (vert) sur le maillage de l'automate cellulaire (bleu).....	104
Figure III-36 : Représentation du seuil définissant le nombre d'éléments extraits de chaque classe de concentration.....	105
Figure III-37 : Représentation schématique de la constitution de la base d'exemples à partir de l'échantillonnage de la base de données CFD.....	105
Figure III-38 : Sélection et apprentissage du meilleur modèle possible.....	106
Figure III-39 : Influence de l'initialisation sur la qualité la généralisation	107
Figure III-40 : Influence du nombre de neurones dans la couche cachée pour des apprentissages sur une base d'exemples de 18 060 exemples (Valeurs médianes pour 10 initialisations – barres d'erreurs : minimum et maximum des 20 initialisations).	108
Figure III-41 : Scores de validation croisée pour des échantillonnages différents de la base de données	109
Figure III-42 : Durée d'apprentissage en fonction de l'échantillonnage	110
Figure III-43 : Valeurs du R^2 et du FAC2 pour les neuf cas tests à 46 pas de temps pour les deux règles de transition ANN testées.....	111
Figure III-44 : Performance des deux modèles selon les critères FB et NMSE.	112
Figure III-45 : Comparaison des deux modèles selon FB et NMSE	113
Figure III-46 : Valeurs du R^2 et du FAC2 pour les 9 cas tests à 90% de la masse restant dans le domaine.....	114
Figure III-47 : Comparaison des isosurfaces de concentrations CFD et isocontours de concentrations CA-ANN ₂ pour le cas 1.....	115

Figure III-48 : Dispersion des points observés (abscisses) et modélisés (ordonnées). Les points sont appairés en espace dans la figure de gauche (diagramme de dispersion). Dans la figure de droite, un tri des valeurs de concentration est effectué au préalable (diagramme quantile-quantile).....	116
Figure III-49 : Erreurs aléatoires et systématiques pour les deux CA-ANN et pour 90 % de la masse restante dans le système	116
Figure III-50 : Différentes erreurs selon les cas testés pour le CA-ANN ₂	117
Figure III-51 : Propagation de l'erreur aléatoire, systématique et totale au cours des pas de temps pour le cas favorable.....	117
Figure III-52 : Evolution de la masse restante dans le domaine pour la simulation CFD et la simulation CA-ANN ₂	118
Figure III-53 : Schématisation du cas Fluent.....	122
Figure III-54 : Maillage du cas CFD	123
Figure III-55 : Lignes de courant obtenues autour d'un obstacle cylindrique de 52 m de diamètre pour une vitesse d'écoulement en entrée de 10 m .s ⁻¹ - Simulation par modèle k – ε standard.	124
Figure III-56 : Contours du coefficient de diffusion turbulente autour d'un obstacle cylindrique de 52 m de diamètre pour une vitesse d'écoulement en entrée de 10 m .s ⁻¹ – Simulation par un modèle k – ε standard.	124
Figure III-57 : Représentation schématique du contenu de la base de données CFD des simulations d'écoulements autour d'obstacles	125
Figure III-58 : Echantillonnage de la base de données CFD des simulations d'écoulements autour d'obstacles pour créer la base d'exemples du réseau de neurones.	126
Figure III-59 : Structure des réseaux de neurones utilisée dans la détermination des champs des vitesses selon x et y et des variations spatiales du coefficient de diffusion turbulente..	127
Figure III-60 : Influence de l'initialisation sur la qualité des modèles réseaux de neurones (D _t , U _x , U _y), rappelons que le vent est orienté selon x.	128
Figure III-61 : Influence du nombre de neurones dans la couche cachée pour les différentes variables de sortie.	129
Figure III-62 : Erreur quadratique moyenne sur l'ensemble de test et nombre de valeurs dans la base d'exemples pour la détermination de la vitesse selon la direction x.	130
Figure III-63 : Qualité de l'apprentissage en fonction du nombre de valeurs dans la base d'exemples pour la détermination de la vitesse selon y.	130
Figure III-64 : Qualité de l'apprentissage du réseau de neurones pour la détermination du coefficient de diffusion turbulente en fonction du nombre total d'exemples dans la base d'apprentissage et du nombre d'intervalles sélectionnés par cas.....	131

Figure III-65 : Visualisation de la réalité CFD et de la modélisation ANN de la vitesse selon la direction x pour le meilleur cas non-appris (Diamètre de l'obstacle : 12 m - Vitesse en entrée : $2,5 \text{ m.s}^{-1}$)	133
Figure III-66 : Visualisation de la réalité CFD et de la modélisation ANN de la vitesse selon la direction x pour le cas non-appris défavorable (Diamètre de l'obstacle : 50 m - Vitesse en entrée : $2,5 \text{ m.s}^{-1}$)	134
Figure III-67 : Diagramme de dispersion entre vitesses observées et vitesses modélisées par ANN. Les droites rouges signalent une sur/sous estimation d'un facteur 2. La droite verte est la fonction $y=x$. Représentation du cas favorable (diamètre : 12m – vitesse du vent : $2,5 \text{ m.s}^{-1}$)	135
Figure III-68 : Diagramme de dispersion entre vitesses observées et vitesses modélisées par l'ANN. Les droites rouges signalent une sur/sous estimation d'un facteur 2. La droite verte est la fonction $y=x$. Représentation du cas favorable (diamètre : 50 m – vitesse de vent : $2,5 \text{ m.s}^{-1}$)	136
Figure III-69 : Vitesse selon x dans l'axe central du domaine pour la réalité CFD et la modélisation ANN (cas favorable : $D=12 \text{ m} - v=2,5 \text{ m.s}^{-1}$).....	137
Figure III-70 : Dispersion des valeurs modélisées (ANN) par rapport aux valeurs issues de la réalité CFD (cas favorable $D=50 \text{ m} - \text{Vitesse de vent} : 5,5 \text{ m.s}^{-1}$).....	138
Figure III-71 : Visualisation du champ 2D de vitesse pour la simulation CFD et la modélisation par réseau de neurones pour le cas 8 (Diamètre : 50 m – Vitesse du vent : $5,5 \text{ m.s}^{-1}$)	139
Figure III-72 : Vitesse selon l'axe y : coupe selon l'axe y dans la zone de recirculation (indiquée –a- sur la figure précédente) pour le cas 8 (Diamètre : 50 m – Vitesse du vent : $5,5 \text{ m.s}^{-1}$)	140
Figure III-73 : Visualisation de la prédiction du coefficient de diffusion turbulente par l'ANN. Cas favorable (Diamètre : 26 m - Vitesse du vent : $9,5 \text{ m.s}^{-1}$)	141
Figure III-74 : Visualisation de la prédiction du coefficient de diffusion turbulente par l'ANN. Cas défavorable (Diamètre : 50 m - Vitesse du vent : $9,5 \text{ m.s}^{-1}$)	142
Figure III-75 : Graphe de dispersion des données modélisées et observées pour le cas défavorable (Diamètre : 50 m – vitesse du vent : $9,5 \text{ m.s}^{-1}$)	143
Figure III-76 : Coefficient de diffusion turbulente modélisé et issu de la réalité CFD dans l'axe central pour le cas défavorable (diamètre : 50 m – vitesse de vent : $9,5 \text{ m.s}^{-1}$)	144
Figure III-77 : Coefficient de diffusion turbulente selon une coupe normale à l'écoulement, en aval de l'obstacle ($y=1000$ cellules)	144
Figure III-78 : Erreur systématique (FB) et erreur totale (NMSE) pour les différents cas non-appris en sortie de domaine (D : Diamètre – v : Vitesse).....	146

Figure III-79 : Visualisation de la dispersion d'une bouffée autour d'un obstacle. Comparaison entre la réalité CFD et la modélisation ANN pour le cas favorable (Diamètre : 26 m - Vitesse du vent : $9,5 \text{ m.s}^{-1}$).....	147
Figure III-80 : Graphe de dispersion entre les fractions massiques observées (réalité CFD) et modélisées (ANN) le cas 6 (diamètre : 26 m – vitesse de vent : $9,5 \text{ m.s}^{-1}$) au temps $t=101 \text{ s}$. La droite verte représente la première bissectrice $y=x$. Les droites rouges représentent les sous-estimations et surestimations d'un facteur deux.....	148
Figure III-81 : Fraction massique dans l'axe central pour simulation CFD et la modélisation ANN au temps $t=101 \text{ s}$	149
Figure III-82 : Fraction massique pour une coupe selon y au niveau de la concentration maximale ($x= 330 \text{ m}$) pour la simulation CFD et la modélisation ANN au temps $t=101 \text{ s}$	150
Figure III-83 : Evolution du coefficient de détermination et du facteur 2 au cours de la modélisation pour le cas 6.	150
Figure III-84 : Evolution du FB et du NMSE au cours de la modélisation pour le cas 6.	151
Figure IV-1 : Score de validation croisée en fonction de la complexité du réseau de neurones pour un apprentissage à partir de données sans bruit et avec un bruit de 30 dB.	181
Figure IV-2 : Score de validation croisée en fonction de la complexité du réseau de neurones pour un apprentissage à partir de données bruitées : 20 et 10 dB.	182

Liste des tableaux

Tableau I-1 : Stabilité de la CLA en fonction du nombre de Richardson	24
Tableau I-2 : Lien entre valeur de la longueur de Monin Obukhov et conditions de stabilité de la Couche Limite Atmosphérique	25
Tableau III-1 : Longueur de rugosité en fonction du terrain.....	75
Tableau III-2 : Propriétés physico-chimiques du milieu ambiant	76
Tableau III-3 : Variables d'entrée du réseau de neurones.....	83
Tableau III-4 : Caractéristiques du modèle sélectionné et de sa conception.....	88
Tableau III-5 : Critères de performance - Comparaison ANN/CFD	89
Tableau III-6 : Base d'apprentissage et score de validation croisée minimum	109
Tableau III-7 : Critères de performance – comparaison CA-ANN/CFD – valeurs pour deux réseaux de neurones différents (46 itérations).....	110
Tableau III-8 : R^2 et FAC2 pour neuf cas modélisés, 90% de la masse restant dans le domaine de calcul	113
Tableau III-9 : Variation des paramètres dans la base de données d'écoulements autour d'un obstacle.....	125
Tableau III-10 : Variables d'entrée et cible des réseaux de neurones.....	126
Tableau III-11 : Caractéristiques de l'apprentissage optimal sélectionné pour chaque paramètre.....	131
Tableau III-12 : Prévion de la vitesse selon x pour des cas non-appris	132
Tableau III-13 : Prévion de la vitesse selon y pour des cas non-appris	137
Tableau III-14 : Critères de performance pour la prévion du coefficient de diffusion turbulente par le réseau de neurones sur les cas non-appris.....	140
Tableau IV-1 : Indicateurs de performance pour la modélisation de cas test non-appris à partir d'une base de données bruitée	182
Tableau IV-2 : Caractéristiques techniques des ordinateurs utilisés	200

Partie IV : Annexes

IV-1.	RESOLUTION NUMERIQUE DE L'EQUATION D'ADVECTION DIFFUSION (ADE)	177
IV-1.1.	Méthode générale	177
IV-1.2.	Résolution de l'advection	177
IV-1.3.	Résolution de la diffusion	178
IV-2.	PRISE EN COMPTE DE DONNEES INITIALES BRUTEES.....	181
IV-3.	ARTICLE PUBLIE DANS CHEMICAL ENGINEERING TRANSACTIONS – VOL. 31, 2013	185
IV-4.	ARTICLE PUBLIE DANS CHEMICAL ENGINEERING TRANSACTIONS – VOL. 36, 2014.....	193
IV-5.	SPECIFICATION DES STATIONS DE TRAVAIL UTILISEES DANS CETTE ETUDE	200

IV-1. Résolution numérique de l'équation d'Advection Diffusion (ADE)

IV-1.1. Méthode générale

On se propose ici d'utiliser la méthode développée par Yanenko, (1971). L'idée importante dégagée est la décomposition des phénomènes et leur résolution de manière indépendante. Ainsi, il est possible de récupérer une concentration intermédiaire à chaque opérateur puis de l'utiliser comme entrée pour un autre opérateur. Dans le cas de la résolution de l'équation d'advection, le schéma peut se résumer ainsi :

- ▶ Résolution de l'advection selon x
- ▶ Résolution de l'advection selon y
- ▶ Résolution de la diffusion selon x
- ▶ Résolution de la diffusion selon y

A chaque fois, le champ de concentration obtenu en sortie est donné comme entrée à l'opérateur suivant.

L'équation générale est :

$$\frac{\partial C}{\partial t} + U_{x_i} \frac{\partial C}{\partial x_i} = \frac{\partial}{\partial x_i} \left(D_t(x) \cdot \frac{\partial C}{\partial x_i} \right) \quad (IV-1)$$

IV-1.2. Résolution de l'advection

L'équation à résoudre est la suivante :

$$\frac{\partial C}{\partial t} + U_x \frac{\partial C}{\partial x} = 0 \quad (IV-2)$$

C	Concentration	ppm
U_x	Vitesse selon l'axe x	$m.s^{-1}$
x	Distance	m
t	Temps	s

La discrétisation est effectuée en utilisant les différences finies. On utilisera classiquement un schéma de discrétisation explicite décentré amont pour le terme d'advection. L'équation devient :

$$\frac{C_{i,j}^{k+1} - C_{i,j}^k}{\Delta k} + U_x \frac{C_{i,j}^k - C_{i-1,j}^k}{\Delta x} = 0 \quad (IV-3)$$

Avec $C_{i,j}^k$, la concentration au pas de temps k dans la cellule i dans la direction x et j dans la direction y .

Dans le cas où la vitesse selon x est négative, l'équation devient :

$$\frac{C_{i,j}^{k+1} - C_{i,j}^k}{\Delta k} + U_x \frac{C_{i,j}^k - C_{i+1,j}^k}{\Delta x} = 0 \quad (IV-4)$$

En réarrangeant les termes, l'équation devient :

$$C_{i,j}^{k+1} = C_{i,j}^k \left(1 - \frac{U_x \Delta k}{\Delta x}\right) - C_{i-1,j}^k \left(\frac{U_x \Delta k}{\Delta x}\right) \quad (IV-5)$$

Le critère de stabilité associé à cette équation est le nombre de Courant-Friedrich-Lewy :

$$\Delta k \leq \frac{\Delta x}{\max(U_x)} \quad (IV-6)$$

On récupère en sortie de l'opérateur d'advection un champ de concentration intermédiaire. Le passage selon y ne présente pas de spécificité particulière. La concentration intermédiaire est utilisée comme entrée pour l'opérateur diffusion.

IV-1.3. Résolution de la diffusion

Deux méthodes sont détaillées ici. La première méthode est une résolution explicite de la diffusion. L'équation initiale est de la forme :

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x_i} \left(D_t(x) \cdot \frac{\partial C}{\partial x_i} \right) \quad (IV-7)$$

Avec D_t , le coefficient de diffusion turbulente. Celui-ci n'est pas uniforme dans l'espace. Il caractérise l'écoulement et par conséquent est stationnaire si l'écoulement l'est également.

Afin de résoudre cette équation on considère un maillage uniforme et deux cellules i et $i+1$. Le flux à l'interface de ces cellules est continu. On a donc :

$$q_{i+1/2}^- = q_{i+1/2}^+$$

C'est-à-dire :

$$D_{t_i} \frac{C_{i+1/2} - C_i}{\frac{\Delta x}{2}} = D_{t_{i+1}} \frac{C_{i+1} - C_{i+1/2}}{\frac{\Delta x}{2}}$$

Il vient :

$$C_{i+1/2} = \frac{D_{t_i} C_i + D_{t_{i+1}} C_{i+1}}{D_{t_i} + D_{t_{i+1}}}$$

Le flux devient :

$$q_{i+1/2} = a_{i+1/2} \frac{C_{i+1} - C_i}{\Delta x}$$

$$\text{Avec } a_{i+1/2} = 2 \frac{D_{t_i} D_{t_{i+1}}}{D_{t_i} + D_{t_{i+1}}}.$$

On peut alors discrétiser l'équation initiale :

$$\frac{C_{i,j}^{k+1} - C_{i,j}^k}{\Delta k} = \frac{q_{i+1/2} - q_{i-1/2}}{\Delta x}$$

En exprimant les flux en fonction de la concentration aux temps k , on obtient un schéma de discrétisation explicite :

$$C_i^{k+1} = C_i^k \left(1 - \frac{\Delta k}{\Delta x^2} (a_{i+1/2} - a_{i-1/2})\right) + C_{i+1,j}^k \left(a_{i+1/2} \frac{\Delta k}{\Delta x^2}\right) + C_{i-1,j}^k \left(a_{i-1/2} \frac{\Delta k}{\Delta x^2}\right) \quad (IV-8)$$

Cette équation est stable sous la condition :

$$\Delta k \leq \frac{\Delta x^2}{(a_{i+1/2} - a_{i-1/2})}$$

Dans le cas où D_{t_i} varie peu de cellule en cellule, on retrouve la condition de stabilité associée à la diffusion :

$$\Delta k \leq \frac{\Delta x^2}{2D_t} \quad (IV-9)$$

On remarque donc que ce critère de stabilité sur la diffusion est plus contraignant que le critère de stabilité relié à l'advection, le terme relié au pas d'espace étant élevé au carré. Une méthode implicite, inconditionnellement stable est développée en utilisant un développement de Taylor pour le flux :

$$q_{i+1/2}^{k+1} = q_{i+1/2}^k + \frac{\partial q_{i+1/2}}{\partial C_i} (C_i^{k+1} - C_i^k) + \frac{\partial q_{i+1/2}}{\partial C_{i+1}} (C_{i+1}^{k+1} - C_{i+1}^k) \quad (IV-10)$$

En réarrangeant, le développement devient :

$$q_{i+1/2}^{k+1} = q_{i+1/2}^k - \frac{a_{i+1/2}}{\Delta x} \Delta C_i + \frac{a_{i+1/2}}{\Delta x} \Delta C_{i+1} \quad (IV-11)$$

Avec $\Delta C_i = C_i^{k+1} - C_i^k$.

En reprenant l'équation x, on obtient :

$$\frac{C_{i,j}^{k+1} - C_{i,j}^k}{\Delta k} = \frac{q_{i+1/2}^{n+1} - q_{i-1/2}^{n+1}}{\Delta x} \quad (IV-12)$$

Le schéma est implicite. On remplace le flux par son expression en fonction de la variation de concentration :

$$\frac{\Delta C_i}{\Delta k} = \frac{(q_{i+1/2}^n - \frac{a_{i+1/2}}{\Delta x} \Delta C_i + \frac{a_{i+1/2}}{\Delta x} \Delta C_{i+1}) - (q_{i-1/2}^n - \frac{a_{i-1/2}}{\Delta x} \Delta C_{i-1} + \frac{a_{i-1/2}}{\Delta x} \Delta C_i)}{\Delta x} \quad (IV-13)$$

En réarrangeant les différents termes, on trouve l'expression :

$$\Delta C_{i+1} \left(-a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) + \Delta C_i \left(1 + \frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2}) \right) + \Delta C_{i-1} \left(-a_{i-1/2} \frac{\Delta k}{\Delta x^2} \right) = \frac{\Delta k}{\Delta x} (q_{i+1/2}^k - q_{i-1/2}^k) \quad (IV-14)$$

$$\Delta C_{i+1} \left(-a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) + \Delta C_i \left(1 + \frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2}) \right) + \Delta C_{i-1} \left(-a_{i-1/2} \frac{\Delta k}{\Delta x^2} \right) = \frac{\Delta k}{\Delta x} \left(a_{i+1/2} \frac{C_{i+1} - C_i}{\Delta x} - a_{i-1/2} \frac{C_i - C_{i-1}}{\Delta x} \right) \quad (IV-15)$$

$$\Delta C_{i+1} \left(-a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) + \Delta C_i \left(1 + \frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2}) \right) + \Delta C_{i-1} \left(-a_{i-1/2} \frac{\Delta k}{\Delta x^2} \right) = \frac{\Delta k}{\Delta x^2} (a_{i+1/2} (C_{i+1} - C_i) - a_{i-1/2} (C_i - C_{i-1})) \quad (IV-16)$$

Or : $\Delta C_i = C_i^{k+1} - C_i^k$ implique :

$$\begin{aligned}
& C_{i+1}^{k+1} \left(-a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) - C_{i+1}^k \left(-a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) + C_i^{k+1} \left(1 + \frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2}) \right) - \\
& C_i^k \left(1 + \frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2}) \right) + C_{i-1}^{k+1} \left(-a_{i-1/2} \frac{\Delta k}{\Delta x^2} \right) - C_{i-1}^k \left(-a_{i-1/2} \frac{\Delta k}{\Delta x^2} \right) = \\
& C_{i+1}^k \left(a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) + C_i^k \left(-\frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2}) \right) + C_{i-1}^k \left(a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) \quad (IV-17)
\end{aligned}$$

En réarrangeant :

$$C_i^{k+1} \left(1 + \frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2}) \right) = C_{i+1}^{k+1} \left(a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) + C_i^k + C_{i-1}^{k+1} \left(a_{i-1/2} \frac{\Delta k}{\Delta x^2} \right) \quad (IV-18)$$

$$C_i^{k+1} = \frac{C_{i+1}^{k+1} \left(a_{i+1/2} \frac{\Delta k}{\Delta x^2} \right) + C_i^k + C_{i-1}^{k+1} \left(a_{i-1/2} \frac{\Delta k}{\Delta x^2} \right)}{1 + \frac{\Delta k}{\Delta x^2} (a_{i+1/2} + a_{i-1/2})} \quad (IV-19)$$

La résolution de ce système peut être effectuée par la méthode de Gauss, la matrice à inverser étant tridiagonale.

IV-2. Prise en compte de données initiales bruitées

Comme abordé au III-3, il est possible d'utiliser un automate cellulaire à règle de transition de type réseau de neurones pour modéliser la dispersion atmosphérique en champ libre. L'apprentissage a été effectué à partir d'une base de données de modélisation CFD. Le très faible niveau de bruit contenu dans ces données issues d'un modèle a été démontré par la qualité de généralisation des apprentissages. Effectivement, l'augmentation du nombre de neurones dans la couche cachée a conduit à une diminution continue de la valeur de l'erreur quadratique moyenne. Ce bruit a été considéré comme négligeable dans cette approche. L'extension de la méthodologie développée ici à des données issues d'essais terrain se confronte à la question de l'apprentissage par le réseau de neurones d'une base d'exemples bruitée. Dans cette étude, aucun essai de dispersion atmosphérique n'a été effectué. Afin de vérifier la possibilité d'extension de la méthode à des données issues d'essais réels, la base de données CFD a été modifiée en ajoutant un bruit blanc, avec un rapport signal sur bruit (Signal on Noise Ratio S_{NR}) diminuant de 30 à 10 dB par pas de 10 :

$$S_{NR_{dB}} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (IV-20)$$

La figure ci-après représente les différentes valeurs du score de validation croisée en fonction de la complexité du réseau de neurones et du niveau de bruit introduit dans la base de données :

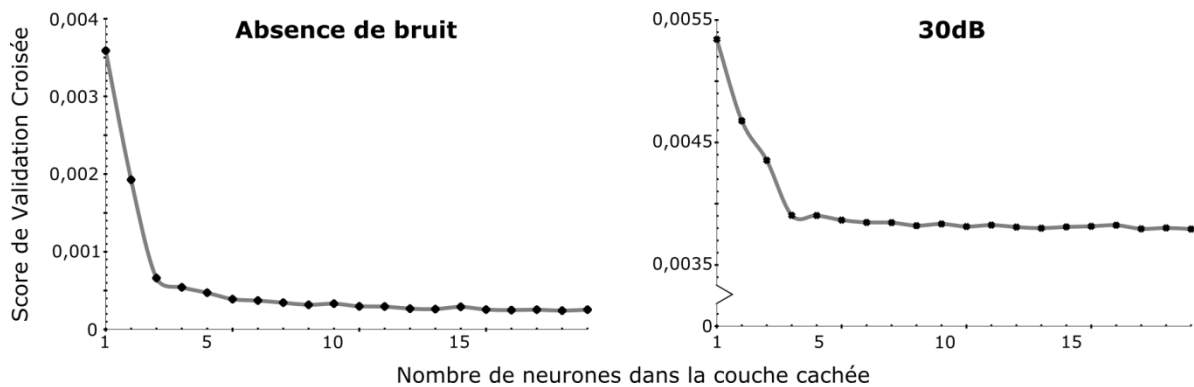


Figure IV-1 : Score de validation croisée en fonction de la complexité du réseau de neurones pour un apprentissage à partir de données sans bruit et avec un bruit de 30 dB.

Dans le cas de la base de données initiale, le seul bruit identifié est le bruit numérique créé par l'interpolation des valeurs sur le maillage de l'automate cellulaire. Dans ce cas là, la diminution du score de validation croisée est continue jusqu'à atteindre une valeur et une pente faible. Lorsqu'un bruit un peu plus important est imposé (30dB), la courbe prend la même allure que précédemment, avec des valeurs de score de validation croisée plus importantes et une asymptote horizontale. L'apprentissage reste satisfaisant. En s'intéressant à des valeurs plus importantes de bruit, l'apprentissage devient de plus en plus difficile :

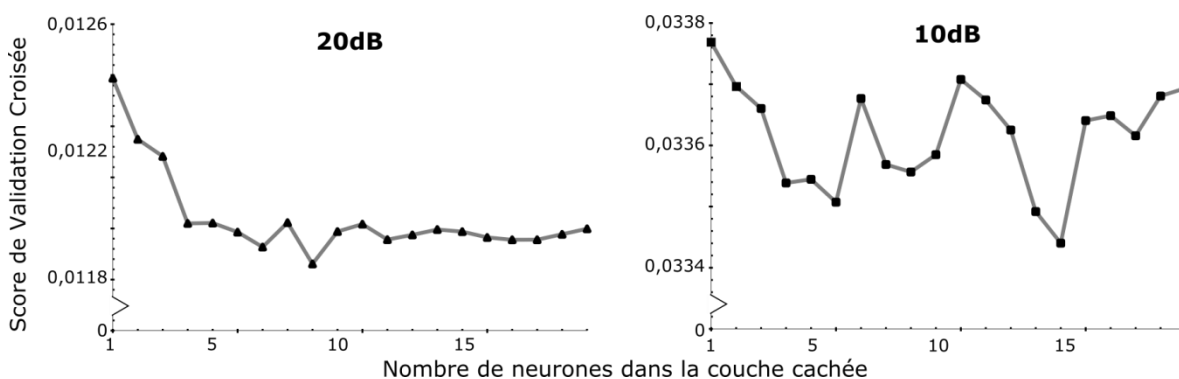


Figure IV-2 : Score de validation croisée en fonction de la complexité du réseau de neurones pour un apprentissage à partir de données bruitées : 20 et 10 dB.

L'augmentation du nombre de neurones dans la couche cachée dans l'apprentissage de la base d'exemples bruitée à 20 dB montre une tendance générale similaire aux apprentissages précédents. L'augmentation du score de validation croisée au-delà de 17 neurones dans la couche cachée montre que l'ajout de complexité nuit à la capacité de généralisation à des données non-apprises. Dans le cas de la base d'exemples fortement bruitée (10 dB), le réseau de neurones a de grosses difficultés à extraire un comportement général de la dispersion. L'augmentation du nombre de neurones n'améliore pas les performances de généralisation. Il semble compliqué d'utiliser une telle base d'exemples pour utilisation en combinaison avec un automate cellulaire.

Les résultats de l'utilisation du meilleur réseau de neurones pour la base contenant un bruit de 30 dB pour les cas tests présentés au III-3.5 sont répertoriés ci-après :

Tableau IV-1 : Indicateurs de performance pour la modélisation de cas test non-appris à partir d'une base de données bruitée

Vitesse du vent (m.s ⁻¹)	Fraction massique	Pas de temps	R2	FAC 2	FB	NMSE
18,8	0,89	46	0,97	0,93	0,02	2,43 x 10 ⁻¹
18,8	0,5	46	0,95	0,84	0,06	2,04 x 10 ⁻¹
10,2	0,89	85	0,9	0,91	-0,01	3,48 x 10 ⁻¹
10,2	0,5	85	0,9	0,88	0,02	2,47 x 10 ⁻¹
10,2	0,26	85	0,83	0,79	0,08	1,89 x 10 ⁻⁴
18,8	0,26	46	0,81	0,71	0,12	2,50 x 10 ⁻¹
3,2	0,26	273	0,03	0,55	0,03	4,43 x 10 ⁻¹
3,2	0,5	273	-0,21	0,58	-0,04	8,25 x 10 ⁻¹
3,2	0,89	273	-0,48	0,58	-0,09	1,33

Ces résultats montrent une bonne adéquation entre le modèle et la réalité CFD pour un faible nombre de pas de temps (vitesses supérieures à $3,2 \text{ m.s}^{-1}$). Au-delà d'un certain nombre, des instabilités apparaissent et le modèle diverge ce qui peut être interprété comme une amplification de l'instabilité constatée préalablement. Ces résultats, bien que moins bons que ceux obtenus à partir de la base d'exemples initiale, montrent qu'il est tout à fait possible d'utiliser des données bruitées et donc des données issues d'essais réels pour appliquer les méthodologies développées dans ce travail.

Une transposition de ce travail à une base de données bruitée a été démontrée pour l'approche CA-ANN. Dans la poursuite de ce travail, il pourra être envisagé de créer une base de données d'écoulements autour d'un obstacle en soufflerie. Ainsi, un apprentissage pour être effectué et la méthode validée à partir de données expérimentale.

IV-3. Article publié dans Chemical Engineering Transactions – Vol. 31, 2013

L'article suivant a été publié dans le volume 31 de Chemical Engineering Transactions, en 2013. Il a de plus été le support d'une présentation orale à la 14th International Symposium on Loss Prevention and Safety Promotion in the Process Industries le 14 mai 2013, à Florence.

Near Field Atmospheric Dispersion Modeling on an Industrial Site Using Neural Networks

Pierre Lauret^{*a}, Frédéric Heymes^a, Laurent Aprin^a, Anne Johannet^b, Gilles Dusserre^a, Laurent Munier^c, Emmanuel Lapébie^c

^aInstitute of Risk Science (ISR), Ecole des Mines d'Alès, Alès, France

^bCentre des Matériaux de Grande Diffusion (CMGD), Ecole des Mines d'Alès, Alès, France

^cCEA, DAM, GRAMAT, F-46500 Gramat, France

*pierre.lauret@mines-ales.fr

Assessment of likely consequences of a potential accident is a major concern of loss prevention and safety promotion in process industry. Loss of confinement on a storage tank, vessel or piping on industrial sites may imply atmospheric dispersion of toxic or flammable gases. Gas dispersion forecasting is a difficult task since turbulence modeling at large scale involves expensive calculations. Therefore simpler models are used but remain inaccurate especially in near field of the gas source. The present work aims to study if Neural Networks and Cellular Automata could be relevant to overcome these gaps. These tools were investigated on steady state and dynamic state. A database was designed from RANS $k-\epsilon$ CFD and Gaussian plume models. Both methods were then applied. Their efficiencies are compared and discussed in terms of quality, real-time applicability and real-life plausibility.

1. Introduction

Major industrial accidents entailing atmospheric dispersion are leading concerns for risk prevention research because of the huge consequences involved. Bhopal disaster sadly illustrates the impact of dispersion of a toxic gas such as methyl isocyanate (MIC) inside and beyond an industrial site. Sharan (1997) showed the importance of both atmospheric conditions and topography. Indeed, local characteristics such as presence of lakes near Bhopal had a major contribution in transporting the MIC into the city area. Flammable gases emissions are also in the studied area because of the UVCE possibility. The Viareggio is a tragic example of LPG dispersion. Brambilla (2009) reported the requirement for taking account of congested environments in such cases.

Industrial sites are defined by various configuration areas. Process buildings, storages or free spaces influence in different ways the atmospheric dispersion because of wind field modifications. Important losses of confinement mainly occur on storages so this study is focused on this type of obstacles.

Models allowing predictions of gas dispersion are abundant. The main strategies are Computational Fluid Dynamics (CFD), semi-empirical and empirical models. CFD models attempt to solve the

Navier-Stokes equations using different kinds of refinement depending of the goal to achieve. Duration of computing, numerical convergence and required expertise (especially for the choice of appropriate turbulence modeling) are the drawbacks of such fineness. Moreover, in several cases, experiments are needed in order to calibrate coefficients. Semi-empirical models such as integral approaches are designed from theory with some assumptions in order to simplify the use of such models while keeping physical soundness. It results in faster modeling and relatively accurate results, especially on simple cases. This explains why their use is widely spread in the industry. Empirical models are based on experiments. From those, correlations are made and can provide accurate results in the range of the experiments database. Beyond this range, the results are not ensured. Gaussian models are one of those. While empirical models can provide inaccurate concentrations, semi-empirical ones preserve consistent results even if the error margin could be important. In any cases, it is crucial to confront models and experimentations (Meroney, 2004).

Starting from this description of the atmospheric dispersion modeling landscape, the proposed approach includes data obtained either by simulation or experiments. It consists in using Machine Learning tools such as Artificial Neural Networks (ANNs) and Cellular Automata (CA). The training database consists in simulation results instead of *in situ* experiments to first design the methodology. Experimental database could be used in second place, to ensure the model to better fit reality.

1.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) came from the original idea to emulate the structure and behavior of the brain (Minsky, 1969). It consists of several mathematical functions, termed as neurons, which are linked in a network. ANN are powerful non-linear statistical data modeling tools. They are generally used when the process to model is not fully known thanks to two essential properties: first the universal approximation (Hornik, 1989), and second the parsimony (Barron, 1993). Thanks to these properties ANNs are able to predict efficiently future behaviors on never encountered situations. ANNs can be used in classification, in text recognition for example (Dreyfus, 2004). They can also be used to forecast physical phenomenon, presenting powerful models (Siou, 2011, Toukourou, 2011). The information about the non-linear phenomenon to simulate or forecast must be provided using a database. As previously presented, ANNs act generally like a black-box: the physics cannot be extracted from the results.

A neuron is a nonlinear, parameterized, bounded function. Variables are assigned to the inputs of the neuron. Output of a neuron is the result of nonlinear combination of the inputs, weighted by the parameters and using an s-shaped function like a sigmoid. A neural network is the composition of several neurons. Parameters calibration is done through application of an algorithm using the training database and designed to decrease the model error (in this work the Levenberg-Marquardt method (Hagan, 1994)). The function realized by the ANN is continuously tested on a disjointed set of examples, namely the validation set. This last set is employed to avoid overtraining using early stopping (Sjöberg, 1995). Lastly, performances of the model must be measured on another set, never used during training or stopping: the test, or validation set.

ANN have already been used in atmospheric dispersion. Prevision of concentrations of tracers in complex terrain have been made by training an ANN using databases of values coming from various sensors spatially distributed (Podnar, 2002). In this case, the output variable is the predicted concentration at a specific point. Moreover, these studies are based on chronic pollution instead of accidental pollution, which is the topic of the present work. Cao (2007) gives an example of the use of neural networks to model the evolution of an aerosol plume. The database has been made especially for the study. It results that the neural networks forecasted more accurate estimations than a Gaussian model. Despite these good results, it appears the difficulty to apply this model in real life situations because of the difficulty to collect inputs information for the neural network.

1.2 Cellular automata

Cellular automata (CA) are tools used for modeling physical phenomena in discrete space-time coordinates. Initially, Von Neumann was interested in seeking logical abstraction of the self-reproduction mechanisms seen in many life-based processes. He adopted a discrete approach for the time, the space and others dynamical variables. He made up a self-replicating theoretical machine, a two dimensional cellular automaton able to reach 29 different states. Since then, CA have been used in simulating various physical phenomena. For instance, Avolio (2006) models

volcanic eruption pyroclastic flows using CA whereas differential equations systems and shows good agreements with observations of the real event. The impact of local interactions on the evolution of the phenomenon is an important feature that promotes the use of CA. From this statement, Wolfram (1983) designed systematic local rules to study different influences from direct neighborhood. Conclusions of his work were that from simple local rules, it is possible to observe very complex phenomena such as biological systems evolution or structure and patterns development in the growth of organisms. He demonstrated that CA can emulate specific behavior of biological or physical phenomenon observed in real life. Itami (1994) formalized cellular automata, defining Q , as the global state of the system:

$$Q = \langle S, N, T \rangle \quad (1)$$

S represents the discrete states accepted for the cellular automaton. N represents the neighborhood of cells providing input values for the transition rules. Depending on the aim of the study, different neighborhoods can be set such as Moore or Von Neumann type. The transition rule T defines how a cell changes his state from the current time step to the next. A large number of quantitative mathematical techniques can be used.

As in classical numerical simulations, the system is defined on a domain. Dimensions and shapes of the cells are determined depending on the phenomenon to model and the acuteness requested.

The transition rule updates synchronously the states of each cell between each time step. According to classical methods, some stability criteria have to be described before using the CA to ensure limitations of numerical errors.

Marin (2000), based on the work of Guariso (1992), identified the different phenomena involved in atmospheric dispersion in order to determine transition rules. Calibration parameters were determined by using measurements from three petrochemical complexes. Comparisons with field measurements gave good agreements and the importance of parameters calibration was strongly underlined. Sarkar (2006) specified a similar method based on simplification of the advection-diffusion equation and developed a model to assess consequences of a loss of confinement on an industrial site, taking account the particular configuration (storage, buildings, type of the area). Vick (2007) gave guidelines to model multi-physics complex processes through CA. These studies pointed out the importance of the determination of the calibration parameters to better fit measurements and the accuracy increase with time step shortening.

In this work, the aim is to check relevance of coupled ANN and CA to predict concentrations of several case studies. The ANN ability to take over large amount of data and to extract a specific behavior is engaged through static and dynamic modeling of atmospheric dispersion.

2. Methodology

2.1 Artificial Neural Networks for stationary modeling

Industrial site observation allows defining various areas corresponding to various pollutant behaviors. Indeed, process buildings, storages or free spaces interact in different ways with atmospheric dispersion because of wind field modification. To gradually increase the difficulty, we initially consider a stationary source term. A database is built from CFD, using resolution of the Reynolds Averaged Navier-Stokes equations by adding equations of turbulent kinetic energy k and its dissipation rate ϵ . The accuracy of such a model and development of all the parameters used are not discussed here, since the objective of this study is to establish a consistent method. We will consider in the future using ANNs with an experimental database. A loss of confinement of gaseous methane into air is modeled. The simulations were set in steady state. The dispersion is modeled in two dimensions. The mesh is composed of 401 301 nodes, with greater density near the methane source. The wind direction is parallel to the ejection direction of the gas. The mix of air and methane is done on a free field of 40 m long and 20 m large. Symmetry conditions are applied at the half of the two meters diameter source. The input of air is set as velocity inlet. Mass flow inlet represents the methane input with a volume fraction of 1. The domain output is set as pressure outlet. In order to simulate correctly the different behavior of the plume, a range from 0.5 to 5 kg.s^{-1} is considered for the methane. The velocity range for the wind input is included between 2 and 10 m.s^{-1} .

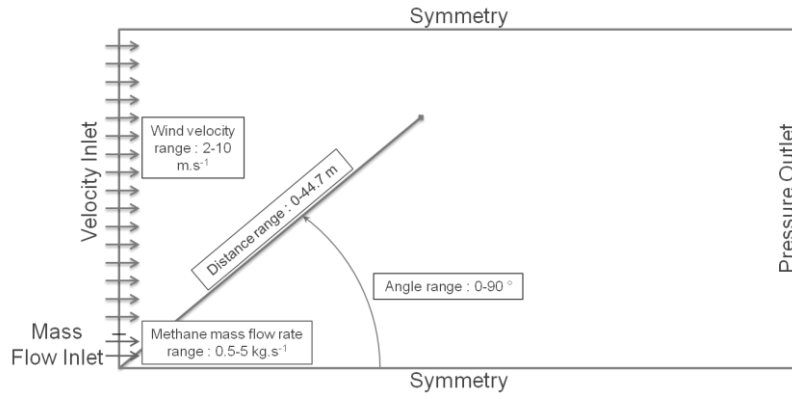


Fig. 1: Configuration used for the simulation.

The database was built from 46 experiments with variations within these ranges. It contains the inputs used by the neural network: the distance (m) and the angle ($^{\circ}$) from source to the point to be evaluated, the wind velocity (m.s^{-1}) and the methane mass flow rate (kg.s^{-1}). The output, or target in training phase, is the methane concentration (mol.m^{-3}).

In order to optimize the computational time, a selection of several data is done instead of taking the wholeness of the >18 000 000 values for each variable. For each case, 20 concentration intervals were created. 50 points per interval including coordinates and concentrations were saved. 36,800 points formed the training set. 4,599 points were used as validation set.

2.2 A Cellular automaton with neural network transition rule

The goal was to model the evolution of the concentration of a pollutant from a source term (point or surface) over time. CA is thus used with ANN transition rule (CA-ANN). As seen above, CA are mostly used with discretization and simplification of physics laws. ANN are used to forecast variables evolution of a phenomenon based on interpolation of a representative database describing it.

First, it is important to define the time step and the dimensions of the cells. According to classical numerical methods, Courant–Friedrichs–Lewy condition (CFL condition) has to be fulfilled. This number corresponds to a threshold over which instability of calculation is observed, growing quickly at each time step. It appears if cells dimensions are less than the distance travelled by the faster wave of the phenomenon during one time step. The minimum time steps and cell dimensions are thus linked by the equation:

$$\text{CFL} = \frac{u \cdot \Delta t}{\Delta x} \leq \text{CFL}_{\max} \quad (2)$$

u is the faster wave velocity, Δt and Δx are the discrete time and distance steps.

To elaborate and evaluate the method, Gaussian puff model is used to create a database. In a first time, we only consider the wind direction dimension. These models are based on the resolution of the advection-diffusion equation (ADE) with several assumptions not described here:

$$\frac{\partial c}{\partial t} + \frac{\partial U_i c}{\partial x_i} = D \frac{\partial^2 c}{\partial x_i^2} \quad (3)$$

As neural networks need relevant variables to work efficiently, the discretization of this equation is done using an explicit scheme. The CFL number is set to one. Consequently, setting the cell dimension Δx and knowing the maximum velocity u gives the time step Δt to be used.

The neural network inputs correspond to cell variables: concentration parameters (C_i^t , C_{i-1}^t , C_{i+1}^t) are collected from gaussian model database. First and second discretized derivative are calculated from those data for each cell i considered. The output, or target in training phase, is the concentration (g.m^{-3}) at time $t+1$ in the cell i considered. Before training, the database is normalized. The learning algorithm is then applied (Levenberg Marquardt). Validation sequence is applied on non-trained data. Fig. 2 sums up the sequence.

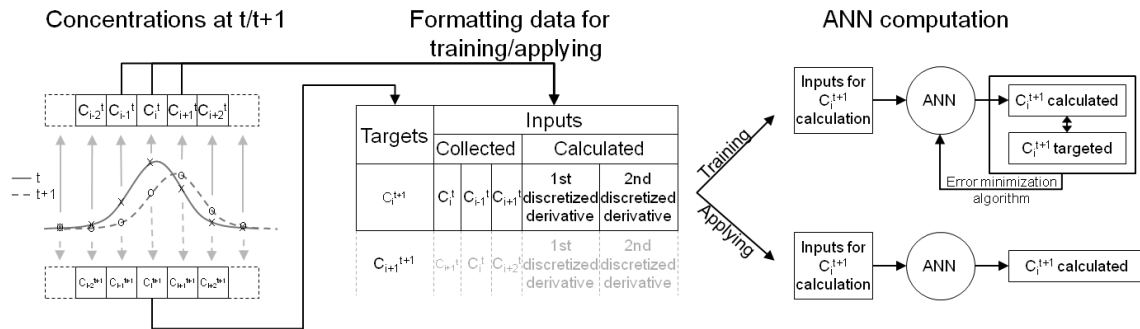


Fig. 2: Logical steps to compute cell concentration à time $t+1$.

In order to evaluate the relevance of such a method, a non-learned example is presented to the cellular automaton and compared to the solution of the Gaussian model.

3. Results

3.1 Artificial Neural Networks

In order to prevent overfitting, early stopping is applied. Results are shown on figure 3 and 4. It provides good performance in generalization. The error graph points out a correlation coefficient R^2 of 0.99. Figure 3 shows that an important number of points are well predicted. Few are less correctly predicted and seem to be isolated cases. Figure 4 compares results from the CFD software and neural network method. Good global agreements are shown. However, several types of error can be identified, such as overestimations for high gradients or negative concentration. The latter requires a correction step to insure physical soundness, but this problem is inherent to empirical modeling.

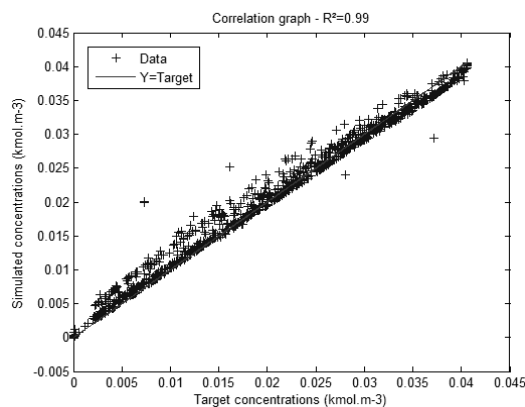


Fig.3: Correlation graph between CFD data and ANN simulation.

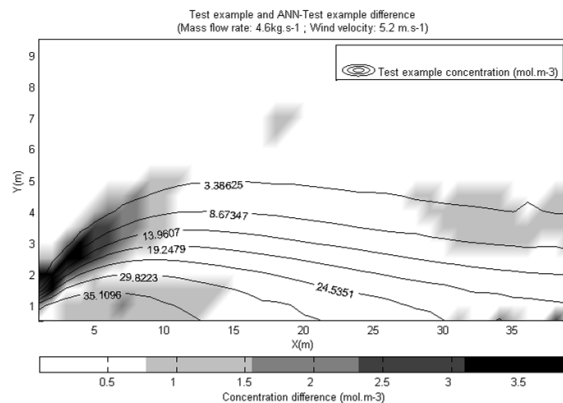


Fig. 4: ANN simulation relative error.

3.2 Cellular automata with neural networks transition rules

The training phase gives satisfying results with a correlation coefficient R^2 near to one. This outcome was expected due to non-linear fitting skills of ANN on equation based data. A validation example is tested with present method. CA is initialized with $t_0=40$ s concentration distribution from a Gaussian puff model. Rating the CA is done by confronting results from both models. Good agreements are obtained for the first time steps of the modeling. As the iteration number increases, the correlation between CA-ANN model and Gaussian plume model data decreases as seen on Figure 5. It can be explained because of the error propagation with iterations. The first steps are well modeled but few differences appear and grow. A divergent behavior is observed after 32th iteration. The error made during each time step is expanded.

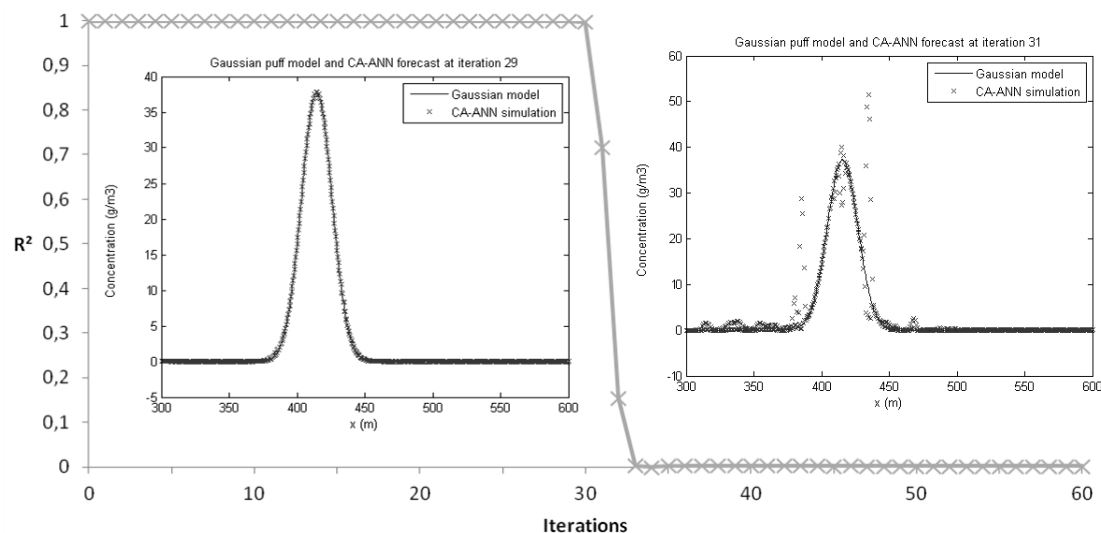


Fig.5.: Decrease of correlation coefficient (between gaussian model and CA-ANN) with iteration number.

4. Improvements

Several improvements can be made on the ANN stationary part. In addition to some errors fixing, special care will be taken in submitting single obstacle to ANN atmospheric modeling. Several different configurations will have to be considered like the position, the shape, and the dimensions of the obstacles. The next step will be to aggregate such different cases according to geometry observed on industrial sites.

To ensure realistic results on the CA with ANN transition rules, a two-dimension CFD database will be created using adequate CFL conditions. Perspectives of this work will be obstacle consideration by adding cell environmental characteristics.

5. Conclusion

Existing forecasting models in atmospheric dispersion have the drawback of being slow but accurate or fast but not enough appropriate. Two methods both using machine learning have been developed and compared to existing models to forecast atmospheric dispersion in open terrain. The stationary ANN model gave good agreement with CFD software with the advantage of faster processing. The 1-D dynamic CA-NN model shows hopeful trends in modeling dynamic atmospheric dispersion and has to be implemented with a 2D CFD learning database to better assess its potential.

References

- Avolio M.V., Crisci G. M., Di Gregorio S., Rongo R., Spataro W., D'Ambrosio D., 2006, Pyroclastic Flows Modelling Using Cellular Automata, *Computers & Geosciences*, 32(7), 897-911.
- Barron A.R., 1963, Universal approximation bounds for superpositions of a sigmoidal function, *IEEE Transactions on Information Theory*, 39(3), 930-945.
- Brambilla S., Totaro R., Manca D., 2009, Simulation of the LPG Release, Dispersion, and Explosion in the Viareggio Railway Accident, *Chemical Engineering Transactions*, 19, 195-200.
- Cao X., 2007, Modelling the Concentration Distribution of Non-Buoyant Aerosols Released from Transient Point Sources into the Atmosphere, Thesis report, Queen's University, Kingston, Ontario, Canada.
- Dreyfus G., 2004, *Neural Networks, Methodology and Applications*, Springer-Verlag, Berlin, Heidelberg.
- Guariso G., Maniezzo V., 1992, Air Quality Simulation through Cellular Automata, *Environmental Software*, 7, 131-141.
- Hagan M.T., Menhaj M.B., 1994, Training feedforward networks with the Marquardt Algorithm, *IEEE Transaction on Neural Networks*, 5(6), 989-993.

- Hornik K., Stinchcombe M., White H., 1989, Multilayer Feedforward networks are universal approximator, *Neural Networks*, 2, 359-366.
- Itami M., 1994, Simulating Spatial Dynamics: Cellular Automata Theory, *Landscape and Urban Planning*, 30, 24-47.
- Marin M., Rauch V., Rojas-Molina A., Lopez-Cajun C.S., Herrera A., Castano V.M., 2000, Cellular Automata Simulation of Dispersion of Pollutants, *Computational Materials Science*, 18, 132-140.
- Meroney R.N., 2004, Wind Tunnel and Numerical Simulation of Pollution Dispersion: a Hybrid Approach, Working paper, Croucher Advanced Study Institute on Wind Tunnel Modeling, Hong Kong University of Science and Technology, 60pp.
- Minsky M., Papert S., 1969, *Perceptrons*, MIT Press, Cambridge MA.
- Podnar D., Koracin D., Panorska, A., 2002, Application of Artificial Neural Networks to Modeling the Transport and Dispersion of Tracers in Complex Terrain, *Atmospheric Environment*, 36, 561-570.
- Sarkar C., Abbasi S.A., 2006, Cellular Automata-Based Forecasting of the Impact of Accidental Fire and Toxic Dispersion in Process Industries, *Journal of hazardous Materials*, A137, 8-30.
- Sharan M., Gopalakrishnan S.G., 1997, Bhopal Gas Accident: a Numerical Simulation of the Gas Dispersion Event, *Environmental Modelling & Software*, 12(2-3), 135-141.
- Siou L.K.A., Johannet A., Borrell V., Pistre S., 2011, Complexity Selection of a Neural Network Model for Karst Flood Forecasting: the Case of the Lez Basin (Southern France), *Journal of Hydrology*, 403, 367-380.
- Sjöberg J., Zhang Q., Ljung L., Benveniste A., Deylon B, Glorennec P.Y., 1995, Nonlinear Black-Box Modeling in System Identification: a Unified Overview, *Automatica*, 31(12), 1691-1724.
- Toukourou M.S., Johannet A., Ayrat P. A., Dreyfus G., 2011, Rainfall-runoff modeling of flash floods in the absence of rainfall forecasts: the case of "Cévenol flash floods", *Journal of Applied Intelligence* 35 (2), 1078-189.
- Vick B., 2007, Multi-Physics Modeling Using Cellular Automata, *Complex Systems*, 17, 65-78.
- Wolfram S., 1983, Statistical Mechanics of Cellular Automata, *Review of Modern Physics*, 55, 601-644.

IV-4. Article publié dans Chemical Engineering Transactions – Vol. 36, 2014

L'article suivant a été publié dans le volume 36 de Chemical Engineering Transactions, en 2014. Il a de plus été le support d'une présentation orale à la 6th International Conference on Safety and Environment in Process and Power Industry le 15 avril 2014, à Bologne.

Atmospheric Turbulent Dispersion Modeling Methods using Machine learning Tools

Pierre Lauret^a, Frédéric Heymes^a, Laurent Aprin^a, Anne Johannet^a, Gilles Dusserre^a, Emmanuel Lapébie^b, Antoine Osmont^b

^aLaboratoire de Génie de l'Environnement Industriel (LGEI), Ecole des Mines d'Alès, Alès, France

^bCEA, DAM, GRAMAT, F-46500 Gramat, France
pierre.lauret@mines-ales.fr

Assessment of likely consequences of a potential accident is a major concern for loss prevention and safety promotion in process industry. Loss of confinement on a storage tank, vessel or piping on industrial sites could imply atmospheric dispersion of toxic or flammable gases. Gas dispersion forecasting is a difficult task since turbulence modeling at large scale involves expensive calculations. Therefore simpler models are used but remain inaccurate especially when turbulence is heterogeneous. The present work aims to study if Artificial Neural Networks coupled with Cellular Automata could be relevant to overcome these gaps. Two methods are reviewed and compared. An example database was designed from RANS $k-\varepsilon$ CFD model. Both methods were then applied. Their efficiencies are compared and discussed in terms of quality, real-time applicability and real-life plausibility.

1. Introduction

Major industrial accidents are generally characterized by important mass release and huge consequences.

Bhopal disaster sadly illustrates the impact of dispersion of a toxic gas such as methyl isocyanate (MIC) inside and beyond an industrial site. Sharan et al. (1997) showed the importance of both atmospheric conditions and topography. Indeed, local characteristics such as presence of lakes near Bhopal had a major contribution in transporting the MIC into the city area. Flammable gases emissions are also of interest in the studied topic because of the UVCE possibility. The Viareggio accident is a tragic example of LPG dispersion. Brambilla et al. (2009) reported the requirement for taking account of congested environments in such cases. Indeed, effects of the terrain and obstacles play a major role in gas dispersion, due to the eddies, wakes, stagnation and recirculation points they can introduce (Pontiggia et al., 2012).

In this context, it is important for decision maker to anticipate crisis situations by modeling several realistic scenarios. Need for operational tools gives important guidelines to create such a model: fast, accurate, reliable, easy-to-use and accounting well for obstacles in the simulation. Models presented in this paper have been designed to answer these criteria.

2. Atmospheric dispersion modeling

Models predicting gas dispersion are abundant. Computation of velocity field and gas dispersion of the commodity released is generally done separately. Atmospheric dispersion models use velocity field as an input. The main strategies are described hereafter.

2.1 Flow velocity field determination

Computational Fluid Dynamics (CFD) first aim is to describe a specific configuration of the flow. It solves the Navier-Stokes equations using different kinds of refinement depending of the goal to achieve. Reynolds decomposition is usually used. It divides the velocity in mean and fluctuation components. Turbulence models are used in order to close the equations system. Computation time, numerical convergence and required expertise are the drawbacks of such fineness. Experiments are often needed in order to calibrate coefficients. CFD models require solving continuity, momentum and energy equations which is time consuming.

Semi-empirical models use Monin-Obukhov theory to determine vertical velocity profile. These profiles are dependent of the atmospheric stability state. Additionally, Hosker (1984) gives relations derived from experiments and theory to model flow behind obstacles. These methods are used to avoid long computing time and can be useful in first approach.

While semi-empirical models can provide quick results, CFD ones preserve consistent results with little error margin. In any cases, it is crucial to confront models and experimentations (Meroney, 2004).

2.2 Turbulent dispersion

Turbulent diffusion is involved through every turbulent dispersion model with a specific form. Gaussian models result from a simplified analytical solution of the Advection Diffusion Equation (ADE) (1). Indeed, some assumptions are made on the wind field: it is considered as constant in space and time, turbulence is considered isotropic and stationary. Standard deviation coefficients are determined from weather stability, and calibrated from experimental data.

$$\frac{\partial C}{\partial t} + u_i \frac{\partial C}{\partial x_i} = D_t \frac{\partial^2 C}{\partial x_i^2} + S_C \quad (1)$$

With C the concentration of a pollutant,

U_i the wind velocity in direction x_i ,

S_C an eventual pollutant source term,

D_t the turbulent diffusion coefficient,

$u_i \frac{\partial C}{\partial x_i}$ is the advection term, $D_t \frac{\partial^2 C}{\partial x_i^2}$ is the diffusion term.

It is also possible to solve the ADE in space and time using discrete approach. This method is used by CFD software, calculating D_t from Navier-Stokes system closure equations. This is an iterative computing, using discretization in space and time.

The last method to mention here is Lagrangian particle tracking. A wide number of particles are introduced in the numerical domain. Using advection from the previous calculated velocity field, particles are followed through the iterations. To take turbulence into account, a random controlled distribution is added to the mean velocity value. At the end of the simulations, the particle density is evaluated and linked to a pollutant concentration. Lagrangian models need expert control to adjust the correct velocity fluctuation. Summary of method cited here is schematize on Figure 1:

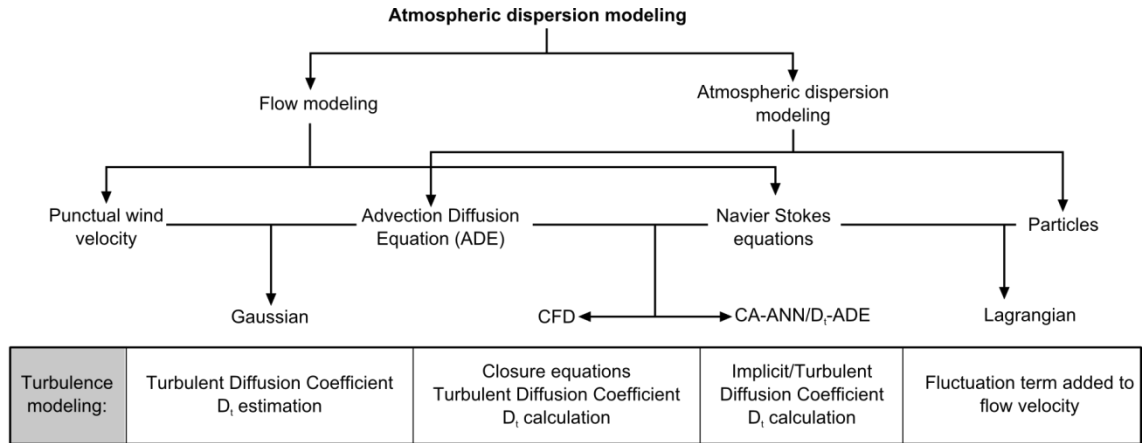


Figure 1: Flow and atmospheric dispersion modelling.

This study considers ADE resolution because no important expertise is needed once the setup is made. Two methods both using artificial neural networks are used here to take turbulence into account.

3. Numerical scheme for Advection Diffusion Equation (ADE) resolution

Solving ADE on a grid can be built using several discretization techniques such as finite differences, finite volumes, finite elements or spectral method. Finite differences are used in this work because of the setup simplicity.

Many different methods exist to discretize partial differential equations (Fletcher, 1991). It is commonly assumed to use respectively classic upwind scheme and centered scheme for advection and diffusion term:

$$\frac{c_{i,j}^{t+1} - c_{i,j}^t}{\Delta t} + U_{x,i,j}^t \frac{c_{i,j}^t - c_{i-1,j}^t}{\Delta x} + U_{y,i,j}^t \frac{c_{i,j}^t - c_{i,j-1}^t}{2\Delta y} = D_t \cdot \frac{c_{i+1,j}^t - 2c_{i,j}^t + c_{i-1,j}^t}{\Delta x^2} + D_t \cdot \frac{c_{i,j+1}^t - 2c_{i,j}^t + c_{i,j-1}^t}{\Delta y^2} \quad (2)$$

With $c_{i,j}^t$ the concentration in the cell i (x direction), j (y direction) at time step t .

In this case, discretization of the time dependent equation is explicit. The concentration at time step $t+1$ depends only of the concentration at time step t .

Each numerical scheme has to be consistent and stable i.e. convergent. Stability condition of this scheme had been largely discussed (Fletcher, 1991):

$$\Delta t \leq \frac{\Delta x^2}{U_x \Delta x + 4D_t + U_y \Delta x} \quad (3)$$

This numerical scheme and associated limitations are used in the following.

4. Using Artificial Neural Networks to solve ADE

Artificial Neural Networks (ANNs) consist of several mathematical functions, termed as neurons, which are linked in a network. ANN are powerful as non-linear statistical data modeling tools. They are generally used when the process to model is not fully known thanks to two essential properties: first the universal approximation (Hornik et al., 1989), and second the parsimony (Barron, 1993). Thanks to these properties ANNs are able to predict efficiently future behaviors on never encountered situations. ANN can be used to forecast physical phenomenon, presenting powerful models (Ak et al., 2013). The information about the non-linear phenomenon to simulate or forecast must be provided using a database. As previously presented, ANNs act generally like a black-box: the physics cannot be extracted from the results. A neuron is a nonlinear, parameterized, bounded function. Variables are assigned to the inputs of the neuron. Output of a neuron is the result of nonlinear combination of the inputs, weighted by the parameters and using an activation function. Sigmoid s-shaped functions are generally used. A neural network is the composition of several neurons. Parameters calibration is done through application of an algorithm using the training database and designed to decrease the model error, in this work the Levenberg-Marquardt method is adopted (Hagan et al., 1994). The function realized by the ANN is

continuously tested on a disjointed set of examples, namely the stop set. This last set is employed to avoid overtraining using early stopping (Sjöberg et al., 1995). Lastly, performances of the model must be measured on another set, never used during training or stopping: test or validation set.

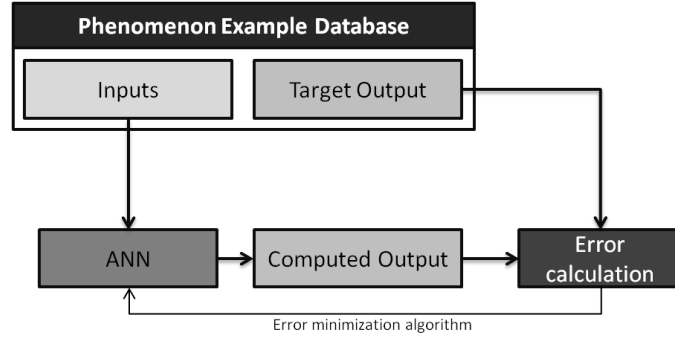


Figure 2: ANN Training method

4.1 Artificial Neural Networks ruled (CA-ANN) Cellular Automata

This method uses both Cellular Automata (CA) and Artificial Neural Networks. Cellular automata are tools used for modeling physical phenomena in discrete space-time coordinates. The impact of local interactions on the evolution of the phenomenon is an important feature that promotes the use of CA. Itami (1994) formalized cellular automata, defining Q , as the global state of the system:

$$Q = \langle S, N, T \rangle \quad (4)$$

S represents the discrete states accepted for the cellular automaton. N represents the neighborhood of cells providing input values for the transition rules, depending on the aim of the study. The transition rule T defines how a cell changes his state from the current time step to the next. A large number of quantitative mathematical techniques can be used such as Artificial Neural Networks, genetic algorithms, Markov Chain, Monte Carlo simulations, fuzzy logic. Cellular automata can be used in real-time to forecast phenomena (Russo et al., 2013). In the present work, ANN are used as transition rules. The inputs are directly selected from the ADE as described in following Figure 3:

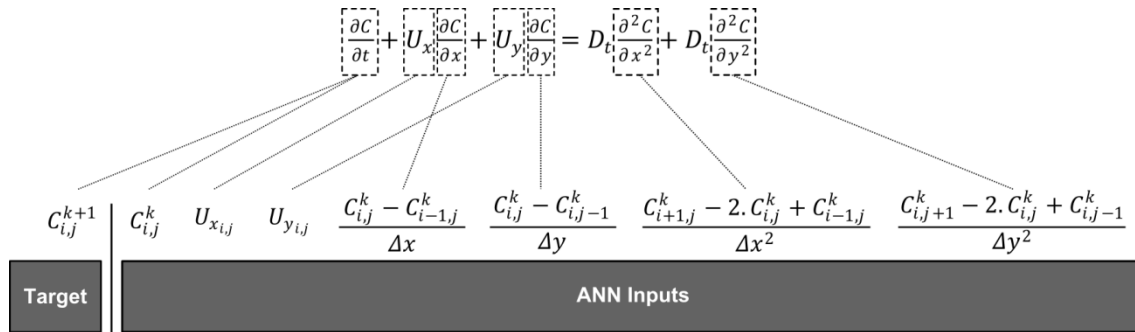


Figure 3: ANN inputs creation

An example database has to be created for the training phase of the ANN. It may be constituted of real life or small scale experiments or from CFD simulation representing reality. Database has to be representative of the atmospheric dispersion to be forecast. Data are recovered from a grid corresponding to the one used with the CA. Each input of the ANN is built from the neighborhood of the considered cell. According to 3.2, spatial and temporal steps are linked together. Data are then normalized and can be used to calibrate ANN parameters.

Once the training algorithm is achieved, the process is similar to CFD techniques. Boundary conditions of the domain to simulate are set. Cells are initialized with concentration and wind field values corresponding to test scenario. The iterative process is then applied: cells concentrations and velocities are recovered. They are combined to compute the ANN inputs, using scheme of

Figure 3. Inputs data feed the ANN synchronously through the entire domain and an output concentration is provided for each cell i,j . Hence, these concentrations are used to compute the next time step. The Figure 4 sums up the iterative process.

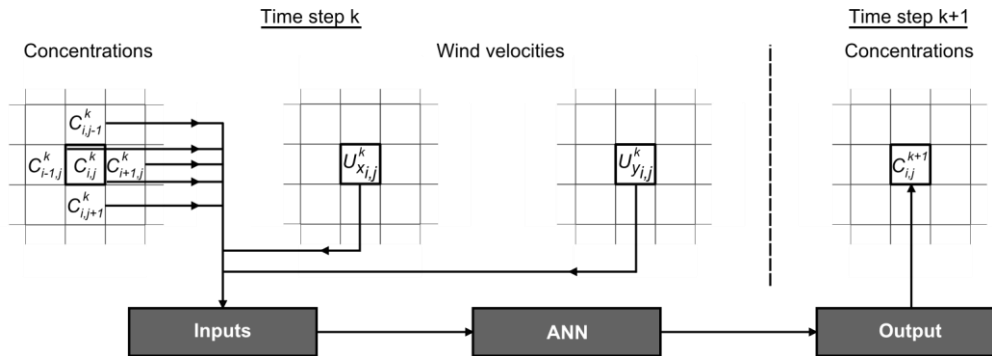


Figure 4: CA transition rule algorithm

Advantages and drawbacks of this method are discussed further.

4.2 ANN determination of the turbulent diffusion coefficient

In this work, the stationary velocity field is considered as known. ADE represents dispersion through time of an initial concentration distribution. Thereby, considering terms of equation 1, the only missing parameter to solve the ADE is the turbulent diffusion coefficient. Method presented here is to forecast this coefficient using ANN. As explained previously, turbulence over horizontal plane is not homogeneous. Especially, obstacles within the flow can critically affect the values of the turbulent diffusion coefficient. Moreover, meteorological stability conditions also affect this coefficient. Finally, altitude of the horizontal plane considered is a concern, related to the height of the atmospheric boundary layer. All these criteria have to be studied to determine the best ANN inputs. Because the ANN output is a single value, inputs have to include geometric data (Lauret et al., 2013). As in 4.1, an example database is built. Because initial hypothesis deals with passive gas, there is no need of information about atmospheric dispersion in the database. Instead, the different examples represent different flows with turbulence variations.

Once the training algorithm is done, the turbulent diffusion coefficient is computed on each cell of the grid. Then, a classical iterative process of the numerical scheme of the ADE described at 3 is applied as described in Figure 5.

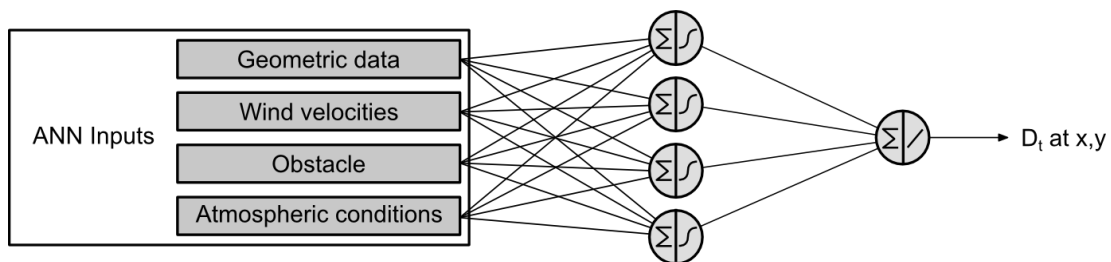


Figure 5: ANN inputs for D_t determination

5. Pros and cons of both methods

Both methods are fast because no direct/indirect solution of a linear system is required compared to CFD models. Moreover, modeling atmospheric dispersion over complex terrain can be considered: spatial heterogeneity of turbulence and obstacles can be implemented. Both methods use ANN. The CA-ANN uses it as a transition rule for the cellular automata whereas in the other hand, ANN are used to determine turbulent diffusion coefficient on a geometrical domain, once (Figure 6).

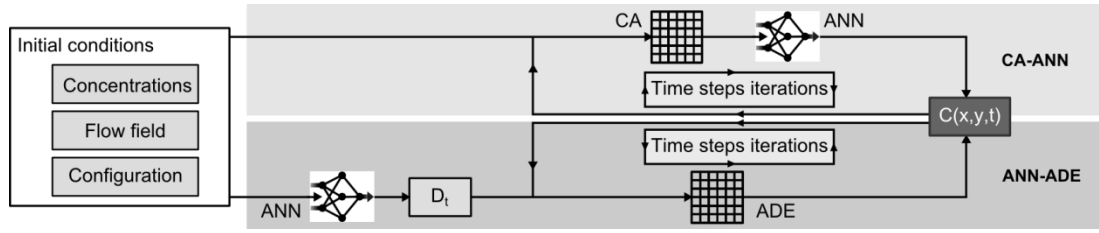


Figure 6: Sum up of both methods

CA-ANN is strictly a black box model: concentration is directly forecast. ANN-ADE can be considered as a grey box: D_t is computed from statistical training method and the phenomenon is modeled using physical ADE.

Even if the training phase is very accurate, it implies some error on the prediction. This error is made at each time step for the CA-ANN. Therefore, global error potentially increases. This configuration can move towards solution instability as can be seen on Figure 7:

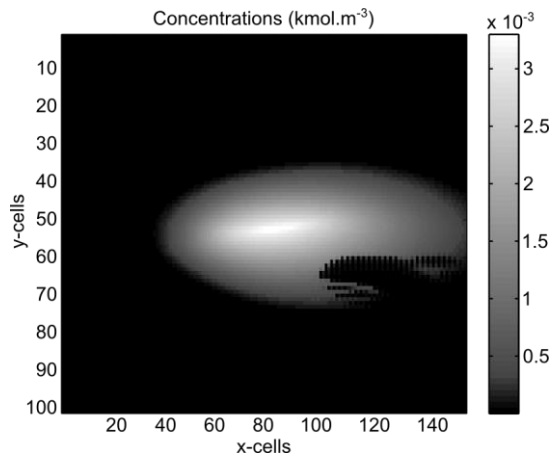


Figure 7: CA-ANN error propagation calculation error

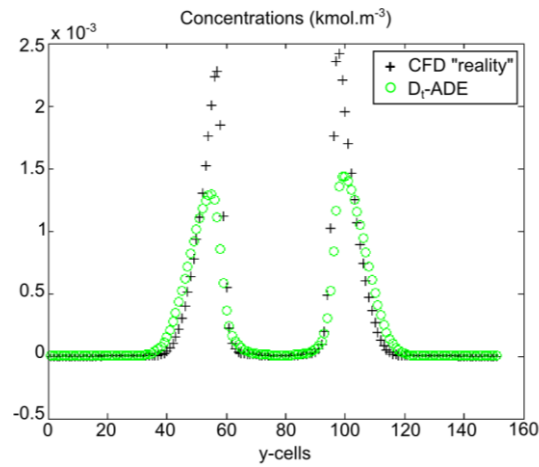


Figure 8: Diffusion error due to D_t

Dealing with CPU-time savings, using ANN rule is nearly instantaneous, because of the application of a simple explicit equation. Nevertheless, the rule is applied for each cell. As the number of cells increases, the CPU-time increases too. For the same domain, CA-ANN appears to be approximately 1.5 times faster than CFD calculation to get final state. If each time step is required, then, the CA-ANN is 20 times faster.

In the case of D_t calculation, error is made once. It implies increase or decrease of diffusion process over advection. This kind of error is illustrated on Figure 8. Peaks of concentration are synchronized in space but not in value. In this case, the coefficient of turbulent diffusion is overestimated compared to CFD transport process, resulting in ANN-ADE peak decrease. In addition, mass balance is difficult to do for the ANN because the lack of information on where the maximum error is located. Nevertheless, a global mass balance can be done and used for basic error correction. As the objective to take into account for 3D atmospheric dispersion, improving the calculation of D_t seems easier. Modifications in the ANN inputs, especially for geometric data, are required. Then, application of finite differences techniques seen at §3 poses no major difficulties. In case of CA-ANN modification, database creation and ANN training phase mobilize considerable computer capacities.

6. Conclusions

Both methods presented here use machine training tools. CA-ANN method is a black box model, focused on forecasting directly concentration of atmospheric gas. Cellular automata are used to emulate ADE through specific domain. Transition rule is implemented by an Artificial Neural Networks. Inputs are created based on ADE parameters. General behavior is satisfactory. Main

difficulty is to avoid error amplification due to small ANN training error inevitably increasing inside the domain. ANN-ADE method is a grey box model. It combines the use of ANN to determine heterogeneous Turbulent Diffusion Coefficient throughout the domain. ADE finite differences resolution is then applied. Major concerns are on the error made during the D_t ANN training phase. It can lead to change the ratio between advection and diffusion in the considered phenomenon. Nevertheless, physics of atmospheric dispersion is kept.

Future work will be focused on tridimensional atmospheric dispersion over multiple obstacles and error correction.

References

- Ak, R., Li, Y., Vitelli, V., Zio, E., 2013. A Genetic Algorithm and Neural Network Technique for Predicting Wind Power under Uncertainty. *Chem. Eng. Trans.* 33, 925–930.
- Barron, a. R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* 39, 930–945.
- Brambilla, S., Totaro, R., Manca, D., 2009. Simulation of the LPG release, dispersion, and explosion in the Viareggio railway accident. *Chem. Eng. Trans.* 19, 195–200.
- Fletcher, C.A.J., 1991. *Computational Techniques for Fluid Dynamics, Volume 1, Fundamental and General Techniques*, 2nd editio. ed. Springer Verlag, New York.
- Hagan, M.T., Menhaj, M.B., 1994. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions Neural Networks* 5, 989–993.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2, 359–366.
- Hosker Jr, R.P., 1985. Flow and Diffusion Near Obstacles. In: *Atmospheric Science and Power Production*. pp. 241–326.
- Itami, M., 1994. Simulating spatial dynamics?: cellular automata theory. *Landsc. Urban Plan.* 30, 27–47.
- Lauret, P., Heymes, F., Aprin, L., Johannet, A., Munier, L., Lapébie, E., 2013. Near Field Atmospheric Dispersion Modelling on an Industrial Site Using Neural Networks. *Chem. Eng. Trans.* 31, 151–156.
- Meroney, R.N., 2004. Wind Tunnel and Numerical Simulation of Pollution Dispersion: A Hybrid Approach. Working paper, Croucher Advanced Study Institute on Wind Tunnel Modeling, Hong Kong University of Science and Technology, -610 December 2004, 60pp.
- Pontiggia, M., Busini, V., Gattuso, M., Uguccioni, G., Rota, R., 2012. Consequences Assessment of an Accidental Toxic Gas Release Through a CFD Tool?: Effect of the Terrain and Major Obstacles. *Chem. Eng. Trans.* 26, 537–542.
- Russo, L., Vakalis, D., Siettos, C., 2013. Simulating the Wildfire in Rhodes in 2008 with a Cellular Automata Model. *Chem. Eng. Trans.* 35, 1399–1404.
- Sharan, M., Gopalakrishnan, S.G., 1997. Bhopal gas accident: a numerical simulation of the gas dispersion event. *Environ. Model. Softw.* 12, 135–141.
- Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B., Glorennec, P.Y., 1995. Nonlinear Black-Box Modeling in System Identification: a Unified Overview. *Automatica* 31, 1691–1724.

IV-5. Spécification des stations de travail utilisées dans cette étude

Les caractéristiques techniques des ordinateurs utilisés pour les simulations numériques à l'aide du code CFD FLUENT 14 sont les suivantes :

Tableau IV-2 : Caractéristiques techniques des ordinateurs utilisés

Types de machine	Fujitsu ESPRIMO P3521 E star5	Dell PRECISION T7400
Processeurs	1	2
Type de processeurs	Intel® Core™ 2 duo	Intel® Xeon® E5440
Fréquence	2,93 GHz	2,83 GHz
Nombre de Coeur /processeur	2	4
RAM	4 Go	12 Go

NNT : 2014 EMSE 0745

Pierre LAURET

ATMOSPHERIC DISPERSION MODELING ON INDUSTRIAL SITE BY COMBINATION OF CELLULAR AUTOMATA AND NEURAL NETWORKS

Speciality: Environmental Engineering

Keywords: Atmospheric dispersion, artificial neural networks, cellular automata, turbulent diffusion coefficient, fluid mechanics, flow around a cylinder

Abstract:

Atmospheric dispersion of toxic or flammable compounds is an event that could lead to serious consequences for man and the environment. Atmospheric dispersion is studied in particular in this work. Modeling of atmospheric dispersion is a tool to assess the geographical areas of danger in the event of a crisis or in risk assessment. The objective of this work was to develop a model that is both fast and accurate, considering the dispersion in the near field (with obstacles) on an industrial site. The approach developed is based on models from artificial intelligence: neural networks and cellular automata. Using neural networks requires training a database typical of the phenomenon. This database is created from atmospheric dispersion CFD simulations ($k - \epsilon$ standard model) in this work. Training the neural network is carried out by identifying the important parameters: database sampling and network architecture. Three methodologies allow responding to different situations: continuous dispersion in open field, puff evolution in the free field and evolution of puff or a plume around a cylindrical obstacle. The first method for estimating the concentrations only uses neural networks. The second method uses the neural network as transition rule of a cellular automaton. The duration of simulation is controlled by the number of iterations of the cellular automaton. The third method divides the problem: the flow calculation is performed by the neural network and the calculation of the dispersion is realized by solving the advection diffusion equation by finite differences method. For the three methods, assessment of the generalization capabilities of the neural network has been validated on a test database and on unlearned cases. The second method uses a recurrent neural network as model, and so a gradual increase of the error is denoted. The third method (calculation of the advection diffusion equation from the neural network modeling of the flow) does not increase the error. A comparison between developed method and CFD simulations is done on unlearned cases in order to validate them. Simulations computing time are low according to crisis duration. Application to real data should be developed to make these models operational.

NNT : 2014 EMSE 0745

Pierre LAURET

MODELISATION DE LA DISPERSION ATMOSPHERIQUE SUR UN SITE INDUSTRIEL PAR COMBINAISON D'AUTOMATES CELLULAIRES ET DE RESEAUX DE NEURONES

Spécialité : Génie de l'Environnement

Mots clefs : Dispersion atmosphérique, réseau de neurones, automates cellulaires, coefficient de diffusion turbulente, écoulement autour d'un cylindre, champ obstrué

Résumé : La dispersion atmosphérique de composés toxiques ou inflammables est un évènement susceptible d'entraîner de graves conséquences sur l'homme et l'environnement. Sa modélisation permet la cartographie des concentrations autour du site en situation de crise ou dans l'évaluation préventive de scénarios potentiels d'accident. L'objectif de ce travail fut de développer un modèle opérationnel, à la fois rapide et précis, prenant en compte la dispersion en champ proche (prise en compte des obstacles) sur un site industriel. L'approche développée s'appuie sur des modèles issus de l'intelligence artificielle : les réseaux de neurones et les automates cellulaires. L'utilisation des réseaux de neurones requiert l'apprentissage d'une base de données caractéristique du phénomène à modéliser. Celle-ci est créée à partir de simulations CFD (modèle $k - \epsilon$ standard) dans le cadre de ce travail. L'apprentissage de la base de données par le réseau de neurones est effectué en identifiant les paramètres importants : échantillonnage et architecture du réseau. Trois méthodologies permettent de répondre à différentes situations : dispersion continue en champ libre, suivi de l'évolution d'une bouffée en champ libre et suivi de l'évolution d'un nuage autour d'un obstacle cylindrique. La première méthode permet d'estimer la concentration dans l'espace, en utilisant les réseaux de neurones seuls. La deuxième méthode utilise le réseau de neurones en tant que règle de transition de l'automate cellulaire. La durée de simulation est pilotée par le nombre d'itérations de l'automate cellulaire. La troisième méthode sépare la problématique : le calcul de l'écoulement est effectué par les réseaux de neurones alors que le calcul de la dispersion est réalisé par la résolution de l'équation d'advection diffusion. Les qualités de généralisation des apprentissages du réseau de neurones ont été validées sur un ensemble de test et en comparaison avec des cas tests non-appris. La deuxième méthode utilise un réseau de neurones récurrent en tant que simulateur, induisant une augmentation progressive de l'erreur alors que la troisième méthode (résolution de l'équation d'advection diffusion à partir de la modélisation de l'écoulement par réseau de neurones) ne fait pas apparaître d'augmentation de l'erreur. La simulation de cas tests non-appris avec des simulations CFD permet de valider les méthodes développées. Les temps de calcul mis en œuvre pour réaliser la dispersion sont en accord avec la cinétique d'une situation de crise. L'application à des données réelles doit être développée dans la perspective de rendre les modèles opérationnels.