



Fast solutions for predicting air pollution risks Solutions rapides pour la prévision des risques de pollution atmosphérique

Raphaël Périllat
Phimeca Engineering
Paris, France
perillat@phimeca.com

Sylvain Girard
Phimeca Engineering
Paris, France
girard@phimeca.com

Irène Korsakissok
IRSN
Fontenay-aux-Roses, France
irene.korsakissok@irsn.fr

Résumé — En cas de rejet accidentel de radionucléides dans l'atmosphère, l'utilisation d'émulateurs permet d'obtenir rapidement des résultats utiles à la prise de décision. Cependant, ces prédictions sont complexes et l'utilisation de méthodes mathématiques avancées est nécessaire.

Abstract — In the event of an accidental release of radionuclides into the atmosphere, the use of emulators makes it possible to quickly obtain useful results for decision-making. However, these predictions are complex and the use of advanced mathematical methods is necessary.

Mots clés — *émulation, processus gaussien, classification, dispersion atmosphérique*

I. CONTEXTE

En cas de rejet accidentel de radionucléides dans l'atmosphère, les simulations numériques de la dispersion atmosphérique de particules radioactives sont utilisées afin de prédire les zones impactées par des rejets accidentels. À l'IRSN, la plateforme opérationnelle de crise C3X permet de réaliser ces calculs sur la base desquels sont formulées des recommandations d'actions de protection des populations à l'attention des décideurs. Les recommandations émises par les décideurs s'appuient notamment sur des indicateurs opérationnels tels qu'une distance maximum de dépassement de valeurs de référence de dose projetée et une ouverture angulaire correspondant à la largeur des zones considérées. Dans un deuxième temps, ces premières estimations sont affinées à l'aide de simulations plus élaborées.

Les distances maximales de dépassement de valeurs-repères données par simulation ainsi que les ouvertures angulaires du territoire susceptible d'être concerné sont des variables scalaires qui permettent de définir des zones sur lesquelles des actions de protection d'urgence (évacuation, mise à l'abri, prise d'iode) pourraient être recommandées. Malgré leur simplicité et la relative rapidité des calculs qui permettent d'y aboutir, il est indispensable en cas de crise de disposer quasiment instantanément d'ordres de grandeur, notamment pour l'expert lui-même, dans le but de décider du jeu de données le plus approprié, compte-tenu des informations disponibles ou non, et de l'objectif donné à l'évaluation. À cet effet, des calculs sont effectués en amont sur un certain nombre de scénarios accidentels et de situations météorologiques décrites à l'aide de quelques

paramètres (direction et vitesse de vent, stabilité, pluie) supposés homogènes sur le domaine de simulation. Ces données précalculées sont rassemblées dans une Fiche Accident Type (FAT). Cependant, l'ensemble des situations météorologiques possibles ne peut être couvert et les résultats sont donnés pour quelques situations supposées représentatives. En général, la situation météorologique du jour de l'accident correspond rarement exactement à l'une de celles décrites dans la FAT. De même, d'autres paramètres peuvent légèrement varier par rapport aux hypothèses de calcul prises dans la FAT (par exemple la quantité rejetée, la hauteur de rejet ou les écarts-types de dispersion choisis). Il serait impossible d'enrichir les FAT pour couvrir l'ensemble des hypothèses possibles, mais disposer d'un outil fournissant instantanément les ordres de grandeur correspondant aux hypothèses retenues à partir d'une base de données précalculées permettrait de répondre aux contraintes opérationnelles.

Les émulateurs, autrement appelés métamodèles, sont des modèles de substitution au modèle de calcul de départ, utilisés pour approximer certaines de ses réponses scalaires. Ce sont des fonctions statistiques construites à partir d'un grand nombre de simulations, dans le but de prédire de nouvelles réponses avec un temps d'évaluation négligeable par rapport à celui du modèle initial considéré. Dans cette étude ce sont les processus gaussiens [3] qui sont utilisés pour la construction des émulateurs.

Un des premiers objectifs de la construction d'émulateurs est de pouvoir obtenir quasi-instantanément les distances d'atteinte dans les cas où les hypothèses météorologiques ne sont pas exactement les mêmes que celles données par la FAT. Le deuxième objectif est d'obtenir un outil à des fins de sensibilisation et de formation, qui permettrait de faire varier certains paramètres afin d'en observer les conséquences. Avoir un ordre de grandeur de la variation des distances d'atteinte permet de mieux appréhender la sensibilité du modèle aux variables d'entrée. À plus long terme, ces outils peuvent aussi être utilisés pour estimer des incertitudes à l'aide d'un ensemble de simulations. Dès lors qu'un émulateur fiable est construit, il est possible de l'utiliser comme un modèle classique. On peut alors perturber ses variables d'entrées incertaines, et les propager pour évaluer les incertitudes de ses sorties. C'est ce qui a été fait sur le cas de Fukushima dans le cadre de précédentes

études; des émulateurs de concentration ou de dose ont été construits et utilisés pour réaliser une étude de sensibilité aux paramètres d'entrée des simulations de dispersion [1] ainsi que pour quantifier les incertitudes sur les données d'entrée [2].

II. ZONES D' ACTIONS APRES UN ACCIDENT NUCLEAIRE

Dans cette étude, nous nous plaçons dans le cas d'une brèche primaire entraînant la fusion totale du cœur en une heure d'un Réacteur à Eau Pressurisée (REP) de 1300 MWe. Ce scénario accidentel est choisi car il fait partie des scénarios précalculés conduisant au dépassement de valeurs-repères sur des distances significatives. En effet, dans le cas d'accidents ne conduisant pas à recommander des actions de protection des populations, ou sur de très faibles distances, même avec des hypothèses pénalisantes, la question de la sensibilité des paramètres d'entrée ne se pose pas dans les mêmes termes.

Le but de ces fiches est de répertorier les conséquences d'un accident type en fonction d'hypothèses météorologiques répertoriées dans la TABLE I. Ces distances sont calculées à partir des doses simulées avec le modèle de dispersion à bouffées gaussiennes pX, inclus dans la plateforme C3X, et à partir de différents seuils de dose de 10 ou 50 millisievert (mSv). Les calculs sont effectués avec les écarts-types gaussiens de Doury [6], pour deux situations de stabilité : diffusion normale (DN) et diffusion faible (DF) et deux valeurs de vitesse de vent (3 et 7 m/s). Une situation météorologique avec une pluie de 2 mm/h est également étudiée. Dans la pratique, il est rare que la situation météorologique du jour corresponde exactement à l'un des cinq scénarios retenus dans la fiche. De plus, la quantité totale rejetée peut varier ; il est courant qu'une analyse des paramètres réacteurs conduise à estimer les rejets comme étant plus faibles que ceux de la FAT à un facteur près; or, si les concentrations dans l'air évoluent de façon linéaire par rapport aux quantités émises, ce n'est pas le cas des distances.

De nouveaux calculs pX doivent donc être lancés pour estimer les nouvelles conséquences à partir des hypothèses qui diffèrent de celles de la fiche. Bien qu'un calcul pX dure environ 45 secondes, le temps d'exécution de toute la chaîne de calcul est de l'ordre de plusieurs minutes (entrée des paramètres de calcul dans l'IHM, pré et post-traitement, calcul des doses, vérification des résultats, etc.). Ces estimations peuvent être répétées plusieurs fois s'il existe une incertitude sur la configuration de calcul à prendre (avec/sans pluie par exemple). La construction d'un émulateur pour ce scénario accidentel permettrait donc de disposer de résultats de façon quasi-instantanée pour toutes les hypothèses à tester dès lors qu'elles se trouvent dans la plage de variation des paramètres prise en compte à la construction de l'émulateur. Il s'agit, en quelque sorte, d'une version améliorée de la FAT permettant d'explorer un plus grand nombre de situations.

L'objet de cette étude est de décrire la construction de l'émulateur pour le scénario accidentel proposé, ainsi que sa validation. En effet, l'émulateur étant une approximation mathématique du modèle physique pX, il est indispensable qu'il donne des résultats similaires à ceux du modèle d'origine.

TABLE I. DISTANCES D' ATTEINTE DES VALEURS REPERES POUR LA FAT N°APRP_1300_7

Valeur-repère	Distances d'atteinte (km)				
	DF 3m/s	DF 7m/s	DN 3m/s	DN 7m/s	DN 7m/s (pluie)
Mise à l'abri (dose efficace totale ≥ 10 mSv)	6,5	9,7	2,7	3,9	4,1
Evacuation (dose efficace totale ≥ 50 mSv)	2,9	4,1	1,2	1,8	1,8
Administration d'iode (dose équiv. thyroïde inhalation ≥ 50 mSv)	10,7	16,3	4,4	6,5	6,1

Avant la construction d'émulateurs pour le calcul des zones d'action, il convient de définir quelles grandeurs scalaires devront être émulées. Dans notre étude, chaque zone d'action sera définie par trois paramètres : sa distance maximale, son ouverture angulaire et son orientation (cf. Fig. 1), qui sont calculables à partir des sorties de dose du modèle pX. Or, puisque l'orientation du panache ne dépend que de la direction du vent, supposé homogène et uniforme dans les cas étudiés, ce paramètre n'a pas besoin d'être émulé, puisqu'il est obtenu directement à partir de la valeur en entrée de modèle. Par ailleurs, 3 zones seront considérées :

- Une zone de mise à l'abri, où la dose efficace totale simulée est supérieure à 10 mSv,
- Une zone d'évacuation, où la dose efficace totale simulée est supérieure à 50 mSv,
- Une zone de prise d'iode, où la prédiction de la dose équivalente à la thyroïde engagée par inhalation est supérieure à 50 mSv.

Ce sont donc deux émulateurs (distance maximale et ouverture angulaire) qui sont construits pour chacune des 3 zones d'action. Dans ce document, nous nous concentrons principalement sur l'étude des distances maximales, qui sont les plus utilisées en pratique dans la gestion de crise.

Pour cette construction d'émulateurs, nous utilisons le modèle à bouffées gaussiennes pX, avec les écarts-types de dispersion de Doury. Dès lors, il faut également distinguer deux modes de stabilité : le mode de diffusion neutre (DN) et le mode de diffusion faible (DF). Dans la pratique, il serait possible de construire un seul émulateur qui considérerait les deux modes, mais la discontinuité assez forte entre ces classes de stabilités augmenterait considérablement le nombre de simulations nécessaires à la mise en place des émulateurs. C'est pour cela que nous avons dissocié les deux modes pour créer des émulateurs pour chacun d'eux. Ainsi, ce ne sont pas six, mais douze émulateurs qui sont considérés pour la suite de cette étude.

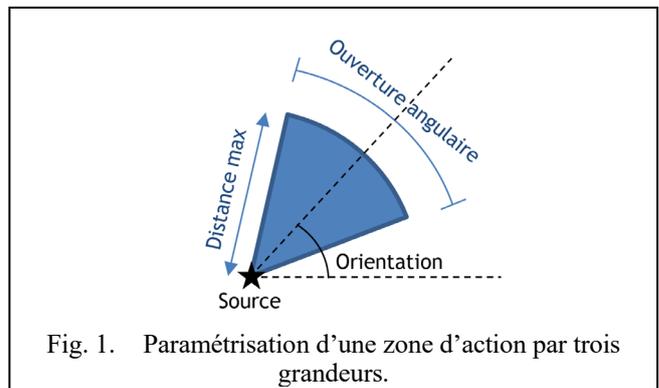


Fig. 1. Paramétrisation d'une zone d'action par trois grandeurs.

TABLE II. PARAMETRES D'ENTREE ET PLAGE DE VARIATION POUR LA CONSTRUCTION DES EMULATEURS

Variable d'entrée	Plage de variation	Unité
Hauteur de rejet	[0, 100]	m
Module du vent	[0, 10]	m/s
Intensité de la pluie	[0, 10]	mm/h
Amplitude du terme source	[10, 100]	%
Facteur de battement de vent	[1, 3]	-

Dans cette étude, nous avons considéré un maillage polaire, pour avoir des mailles plus resserrée dans la zone proche de la source, là où il y a de fortes variations de dose. Les points du maillage sont répartis sur 36 angles sur 360° et sur 61 rayons différents à partir de 500 m et de plus en plus espacés les uns des autres au fur et à mesure que nous nous éloignons de la source, jusqu'à une distance éloignée de 30km pour capturer toutes les zones d'actions de notre ensemble de simulations.

Cinq paramètres d'entrée du modèle pX sont considérés comme des variables incertaines (cf. TABLE II.). Pour chacune de ces variables, un intervalle de variation est défini afin de délimiter l'espace que nous cherchons à explorer. Ces 5 variables sont sélectionnées, car elles font partie des quelques paramètres susceptibles d'être modifiés au cours de la crise, suite à des précisions supplémentaires ou bien pour prendre en compte des incertitudes fortes sur leurs valeurs. Dans le cas où nous chercherions à construire un émulateur adapté à des situations post-accidentelles, nous pourrions rajouter quelques variables supplémentaires, telles que les vitesses de dépôt sec et les différents coefficients de lessivage qui, bien que de second ordre par rapport au terme source et à la météorologie, sont également incertains. Nous aurions alors plutôt 9 variables, mais pour une première construction d'émulateur nous privilégions le cas restreint avec 5 variables afin de faciliter les différentes étapes de l'étude.

III. EMULATION PAR PROCESSUS GAUSSIENS

La construction d'émulateurs passe par la simulation d'un grand nombre de cas dans l'espace des possibles. Un tirage est effectué pour chaque variable d'entrée de manière uniforme sur sa plage de variation, afin de rendre les émulateurs uniformément fiable sur l'espace des possibles. Dans notre cas, cela revient à tirer aléatoirement un point dans un espace de dimension 5. Une simulation est ensuite effectuée avec pX pour chaque point de cet échantillon d'apprentissage, avec les valeurs de ces paramètres tirées au sort. Les distances maximales ainsi que les ouvertures angulaires correspondant aux trois zones d'actions considérées peuvent être extraites des résultats de cette simulation. Ces grandeurs sont stockées pour ensuite être utilisées dans la construction d'émulateurs.

Un processus gaussien est défini comme une collection d'un nombre fini de variables aléatoires ayant une distribution jointe gaussienne. Ainsi, avec un émulateur de ce type, la surface de réponse décrite par la sortie lorsque les entrées varient est modélisée comme la somme d'un terme constant et la réalisation d'un processus gaussien centré. Si la sortie de modèle pX est notée y et que le vecteur de ses entrées est noté x , l'émulateur f s'écrit sous la forme :

$$f(x) = C + \sum_{j=1}^N w_j(x)(g(x^{(j)}) - C) \quad (1)$$

où C est le terme constant, $x^{(j)}$ est un point de l'échantillon d'apprentissage de taille N et $w_j(x)$ sont des coefficients de pondération. La détermination de f par apprentissage est réalisée à l'aide du module python DiceKriging [3]. Dans notre cas, $N = 2048$ simulations sont utilisées pour construire les métamodèles dans le mode de diffusion DN et 2048 simulations supplémentaires sont utilisées pour le mode DF.

Grâce aux émulateurs ainsi construits, nous sommes en mesure d'évaluer très rapidement la valeur de la distance maximale et de l'ouverture des zones d'actions, pour différents seuils de dose et différents modèles de stabilité de Doury. Une interface graphique a été développée afin de représenter géométriquement ces paramètres, pour une meilleure visualisation des résultats.

Théoriquement, avec un nombre infini de simulations utilisées pour leur construction, les émulateurs donnent des résultats identiques à ceux du modèle. Mais dans la pratique, le nombre de simulations est limité et les émulateurs ainsi construits peuvent s'éloigner des résultats du modèle. Pour vérifier s'ils restent fiables, une approche courante consiste à évaluer l'émulateur sur un échantillon de test et d'en comparer les résultats avec le modèle initial. Pour cela, nous tirons M nouveaux points dans notre espace d'entrée à 5 dimensions et nous lançons ensuite M fois le modèle pX avec des paramètres d'entrées correspondants à ces tirages. Nous obtenons alors M résultats de sorties (distances maximales et ouvertures angulaires) que nous pouvons ensuite comparer à ce que prédisent les émulateurs pour ces mêmes hypothèses. Nous évaluons ainsi la qualité de nos émulateurs en plusieurs points répartis aléatoirement dans l'espace, ce qui permet de quantifier l'erreur commise par l'émulateur sur des cas qu'il ne connaît pas. Pour cette étude, nous considérons un échantillon de test de $M = 1000$ points.

Avec les résultats du modèle pX et des émulateurs sur l'échantillon de test, nous pouvons comparer directement les valeurs obtenues, pour observer leurs différences (cf. Fig. 2). Nous pouvons alors remarquer que les résultats sont assez

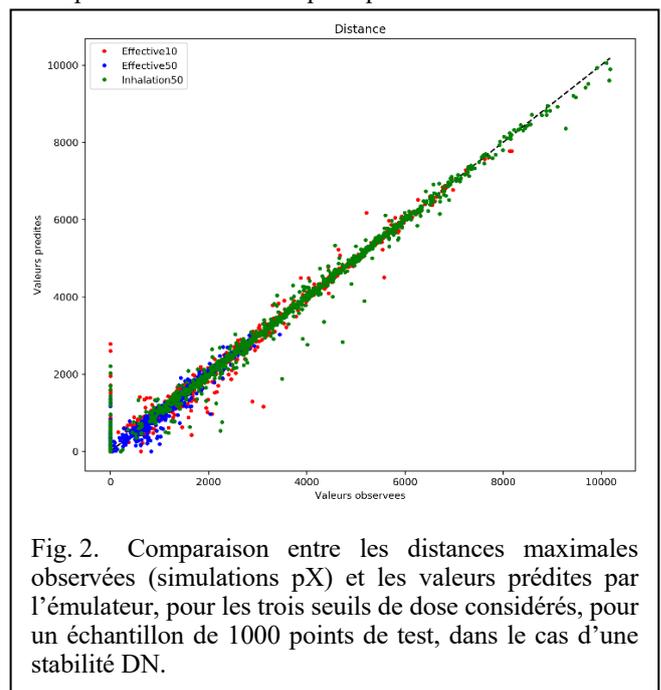


Fig. 2. Comparaison entre les distances maximales observées (simulations pX) et les valeurs prédites par l'émulateur, pour les trois seuils de dose considérés, pour un échantillon de 1000 points de test, dans le cas d'une stabilité DN.

bons si nous nous plaçons dans le cadre classique de comparaison modèle-mesure. Malgré tout, il existe une vingtaine de points où la valeur prédite est bien plus faible que la valeur observée, ce qui peut poser problème en situation de crise où l'objectif est de maximiser les résultats pour considérer le pire cas possible. Sur cette même figure, nous observons également qu'il y a beaucoup de points où la distance prédite est forte, alors que la distance observée est nulle. Ce problème apparaît principalement pour deux raisons :

- La première cause de cette imprécision est qu'il est impossible de calculer une distance maximale de dépassement inférieure à 500 m qui soit non-nulle. C'est en effet la distance minimale à partir de laquelle sont calculées des concentrations dans notre modélisation. Dès lors qu'un dépassement de seuil se fait sur une petite zone, à moins 500 m de la source, la distance maximale calculée sera forcément estimée à 0 m alors que sa valeur réelle est non-nulle. Cette discontinuité induite par la discrétisation de l'espace peut créer des problèmes lors de la création d'émulateur par processus gaussiens, qui sont des fonctions « lisses ».
- La deuxième cause est qu'il existe dans la pratique un très grand nombre de cas (entre 63 et 97%, en fonction du cas) où la distance maximale de dépassement de seuil est nulle (quand la vitesse du vent est très élevée et que la hauteur de rejet est très haute, par exemple). La variable de sortie possède alors deux types d'évolutions sur deux régions de l'espace différentes. Cet effet de seuil dans le comportement de la variable de sortie perturbe la construction de l'émulateur et le rend moins fiable dans les zones de l'espace où les distances sont supposées être nulles.

Dans la TABLE III, nous pouvons observer que lorsqu'il s'agit de prédire si un résultat sera nul ou non, les émulateurs sont particulièrement mal adaptés. Les nombres de faux-positifs et de faux-négatifs peuvent être particulièrement élevés avec des émulateurs construits de la sorte (jusqu'à 46% d'erreur), c'est pourquoi il est important de corriger ces écarts.

TABLE III. NOMBRE DE CAS OU L'EMULATEUR DONNE UNE REPONSE NON NULLE ALORS QU'ELLE DEVRAIT ETRE A ZERO (FAUX POSITIF) ET OU IL DONNE UNE REPONSE NULLE ALORS QU'ELLE DEVRAIT ETRE POSITIVE (FAUX NEGATIF), SUR UN ECHANTILLON DE TEST DE 1000 SIMULATIONS.

Critère		Faux positif	Faux négatif	Taux d'erreur
DN	Dose efficace totale ≥ 10 mSv	38	1	3,9 %
	Dose efficace totale ≥ 50 mSv	233	3	23,6 %
	Dose équiv. Thyroïde inhalation ≥ 50 mSv	36	1	3,7 %
DF	Dose efficace totale ≥ 10 mSv	368	0	36,8 %
	Dose efficace totale ≥ 50 mSv	460	0	46,0 %
	Dose équiv. Thyroïde inhalation ≥ 50 mSv	399	1	40,0 %

Pour pallier ces problèmes, une classification préalable des résultats peut être envisagée. Si nous sommes en mesure de connaître les zones de l'espace des entrées où les distances maximales de dépassement de seuils sont nulles,

alors nous pouvons nous contenter de construire des émulateurs uniquement en dehors de ces zones et obtenir des estimations plus cohérentes puisque les zones à problèmes seront évitées. Les avantages de cette méthode sont présentés en section 4.

TABLE IV. PARAMETRES D'ENTREE ET PLAGE DE VARIATION POUR LA CONSTRUCTION DES EMULATEURS

Stabilité	Score	Dose efficace ≥ 10 mSv	Dose efficace ≥ 50 mSv	Dose inhalation ≥ 50 mSv
DN	Erreur moyenne	109 m	97 m	116 m
	Q95% de l'erreur	189 m	66 m	293 m
DF	Erreur moyenne	419 m	390 m	462 m
	Q95% de l'erreur	874 m	361 m	1391 m

Avec ces résultats, sans utiliser de classification, l'erreur moyenne des émulateurs a été calculée sur l'échantillon de test, ainsi que le percentile 95% sur ces erreurs, qui permet d'estimer une erreur maximale que l'on peut obtenir. Ainsi, dans le cas d'une stabilité DN, pour un seuil de 10 mSv sur la dose efficace, l'erreur moyenne est de 109 m et elle est inférieure à 419 m dans 95% des cas. (cf. TABLE III). Les deux scores étudiés sont plus faibles dans le cas DN que dans le cas DF (jusqu'à 293 m pour le seuil de dose inhalation > 50 mSv), à cause des problèmes cités précédemment et du fait que, dans ce mode de stabilité, le panache est plus étroit, les gradients de valeurs sont alors très élevés, ce qui rend la prédiction plus difficile à réaliser.

Il est important d'identifier les erreurs acceptables ou non dans ce contexte. Une erreur de quelques centaines de mètres une distance maximale de 1 km peut sembler moins acceptable qu'une erreur identique sur une distance maximale de 15 km, mais paradoxalement si nous raisonnons en termes de zone d'action, la surface de la zone où l'émulateur donne une mauvaise prédiction sera bien moins large dans le cas d'une distance faible que dans le cas d'une distance élevée. Dans la pratique, ces valeurs ne sont pas utilisées en tant que valeurs exactes, mais c'est plutôt leur ordre de grandeur qui permet de déterminer les zones réelles sur lesquelles des actions seront menées (souvent sur des entités administratives telles que les communes). Il faut donc que l'erreur commise sur l'estimation des distances soit plus faible que les marges qui sont prises par les décideurs par sécurité.

Pour ce qui est de l'ouverture angulaire de ces zones, les émulateurs fonctionnent de la même manière, mais ce sont des paramètres plus difficiles à prédire, notamment car il existe une discontinuité supplémentaire : certaines simulations où les vents sont très faibles obtiennent des zones de dépassement de seuils qui ne sont pas orientés dans une direction particulière, mais où la zone d'action est à 360° autour de la source. Ces cas, qui sortent de l'ordinaire des autres simulations, perturbent également la construction des émulateurs, car il n'existe aucun cas intermédiaire où l'ouverture se situe entre 180° et 360°. Tout comme pour les distances maximales, une classification pourrait permettre de prédire les cas de panache à 360°, afin de construire un émulateur uniquement sur les cas classiques où l'ouverture est inférieure à 180°.

IV. METHODES DE CLASSIFICATION POUR L'AMELIORATION DES PREDICTIONS

Comme mentionné précédemment, les distances et ouvertures angulaires que nous essayons d'émuler possèdent des comportements différents sur certaines zones de l'espace. Le fait que les distances soient toutes nulles pour certaines valeurs des variables d'entrée crée une discontinuité difficile à reconstruire par les émulateurs. Pour éviter ce problème, nous utilisons une méthode de classification pour déterminer, en amont de l'émulation, si les distances d'atteintes vont être nulles ou non, avec une méthode des k plus proches voisins [5].

Pour un seuil de dose donné, l'espace des entrées du modèle peut être divisé en deux sous-espaces : un premier dans lequel la distance d'atteinte est nulle et un deuxième dans lequel elle est strictement positive. Le but de la classification est d'estimer si un point P de l'espace se situe dans l'un ou l'autre de ces sous-espaces.

Pour réaliser cette classification, nous avons utilisé les mêmes N points qui ont servi à construire l'émulateur. En construisant alors deux classes, $\{d=0\}$ et $\{d>0\}$, il est possible d'attribuer à tous les points de l'espace une de ces classes, en fonction de l'état de ses k plus proches voisins. Dans notre étude, nous avons choisi $k=5$ et nous avons pondéré l'influence de chacun de ses 5 points par l'inverse de la distance qui le sépare du point à évaluer. Cette distance est définie comme la distance euclidienne entre les points dans un espace à 5 dimensions normalisé de sorte que la plage de variation de chaque variable soit entre 0 et 1.

Une fois cette classification construite, l'évaluation pour savoir si un point de l'espace va donner une distance d'atteinte nulle est quasi-instantanée avec cette méthode. Elle peut donc être utilisée en amont de l'émulateur pour prédire si un résultat sera nul ou bien s'il a besoin d'être évalué par un émulateur.

Le principal avantage de cette méthode est de pouvoir construire un émulateur uniquement sur l'espace restreint des

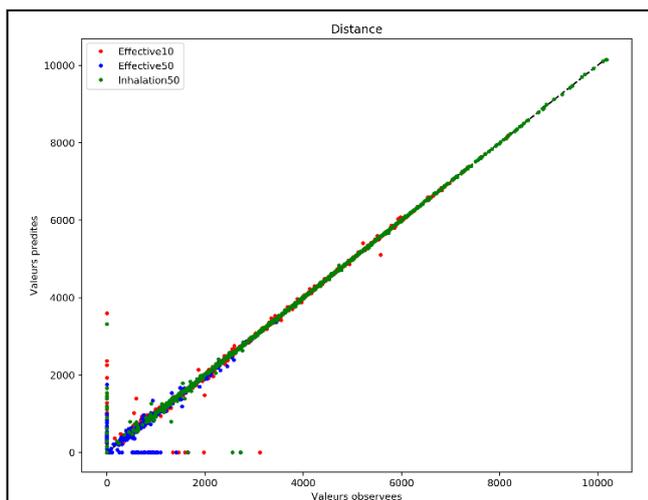


Fig. 2. Comparaison entre les distances maximales observées (simulations pX) et les valeurs prédites par l'émulateur associé à une classification par k plus proches voisins, pour les trois seuils de dose considérés, pour un échantillon de 1000 points de test, dans le cas d'une stabilité DN.

entrées où les distances sont non-nulles. Ce faisant, les émulateurs ainsi construits sont plus précis, car les points problématiques, où les distances se comportent différemment, sont retirés pour sa construction. Lorsque nous voulons alors évaluer un point de l'espace, nous déterminons en premier si la classification prédit qu'il aura une distance nulle. Si c'est le cas, l'algorithme s'arrête, sinon, l'émulateur permet d'évaluer la distance.

TABLE V. NOMBRE DE CAS OU L'EMULATEUR, ASSOCIE A UNE METHODE DE CLASSIFICATION PAR MOINDRE CARRE, DONNE UNE REPONSE NON NULLE ALORS QU'ELLE DEVRAIT ETRE A ZERO (FAUX POSITIF) ET OU IL DONNE UNE REPONSE NULLE ALORS QU'ELLE DEVRAIT ETRE POSITIVE (FAUX NEGATIF), SUR UN ECHANTILLON DE TEST DE 1000 SIMULATIONS.

Critère		Faux positif	Faux négatif	Taux d'erreur
DN	Dose efficace totale ≥ 10 mSv	11	9	2,0 %
	Dose efficace totale ≥ 50 mSv	45	35	8,0 %
	Dose équiv. Thyroïde inhalation ≥ 50 mSv	14	4	1,8 %
DF	Dose efficace totale ≥ 10 mSv	18	26	4,4 %
	Dose efficace totale ≥ 50 mSv	20	27	4,7 %
	Dose équiv. Thyroïde inhalation ≥ 50 mSv	18	25	4,3 %

TABLE VI. PARAMETRES D'ENTREE ET PLAGE DE VARIATION POUR LA CONSTRUCTION DES EMULATEURS

Stabilité	Score	Dose efficace ≥ 10 mSv	Dose efficace ≥ 50 mSv	Dose inhalation ≥ 50 mSv
DN	Erreur moyenne	53 m	60 m	42 m
	Q95% de l'erreur	142 m	78 m	230 m
DF	Erreur moyenne	101 m	414 m	67 m
	Q95% de l'erreur	646 m	222 m	492 m

Cependant, la classification mise en place n'est pas parfaite et dépend des points qui ont été utilisés pour sa construction. Il est possible que la classification donne des résultats erronés, notamment dans la zone proche de l'interface entre les deux sous-espaces que nous cherchons à définir. Pour évaluer les résultats de cette classification, nous utilisons à nouveau un échantillon de test de 1000 points afin d'estimer le nombre de faux-positifs et faux-négatifs que cette méthode nous donne. Avec la combinaison de la classification suivie par l'émulation, nous nous retrouvons avec un certain nombre de points (entre 1,8 et 8,0 %) erronés (cf. TABLE V.), mais ces nombres sont bien inférieurs à ceux des cas problématiques observés dans le TABLE III où aucune classification n'était faite. De plus, l'évaluation de l'émulateur sur les zones où les distances sont non-nulles est dans la plupart des cas bien meilleure, comme le montrent les erreurs d'estimation en TABLE VI.

Pour aller encore plus loin, il conviendrait de chercher à réduire le nombre de faux positifs et de faux négatifs de la classification. Il est possible d'aller dans cette direction en utilisant un autre type de classification par exemple. En effet, si au lieu de procéder par plus proches voisins, nous utilisons une classification par processus gaussien, il s'avère que nous pouvons réduire encore le nombre de points erronés (entre 0.3 et 4.3%), mais en contrepartie l'appel à la classification est bien plus lente, ce qui rend l'utilisation de l'émulateur bien moins réactif : de l'ordre de 2 secondes pour une évaluation, contre 0,2 seconde seulement avec une

classification par plus proches voisins. Il peut donc être intéressant d'utiliser cette autre méthode, mais la première classification sera plus intéressante dans les cas où un très grand nombre d'évaluations est à prévoir.

V. OUVERTURE SUR LES MODELES AUTO-ASSOCIATIFS

Dans cette étude, nous avons émulé les distances maximales de dépassement de seuils, qui peuvent avoir des évolutions bien différentes dans certaines plages de variations des paramètres d'entrée. Le fait que ces distances puissent être à zéro dans de nombreux cas, rend la prédiction difficile pour les émulateurs, comme nous l'avons montré dans les sections 3 et 4.

Pour contourner cette difficulté, il est possible de se concentrer sur les valeurs de doses, plutôt que sur un dépassement de seuil. Cependant, les émulateurs par processus gaussien que nous utilisons sont applicables uniquement sur des variables scalaires, tandis que les sorties de dose du modèle sont des champs 2D. Il convient alors de trouver un bon moyen de réduire la dimension de ces données, afin de se ramener à un faible nombre de variables, qu'il sera possible d'émuler indépendamment les unes des autres, par des processus gaussiens. L'analyse en composantes principales (ACP) est une méthode très utilisée pour ce genre de réduction de dimension, mais dans notre cas, les forts changements de forme des panaches sont mal adaptés pour des décompositions linéaires. Nous nous sommes intéressés à une méthode non-linéaire par modèles auto-associatifs (MAA) [4], qui peuvent être vus comme une généralisation non-linéaire de l'ACP.

Grace à une analyse avec les modèles auto-associatifs, il est possible de paramétrer chaque résultat de dose par un nombre restreint de variables scalaires (proche de 5), avec une perte d'information qui tend à être négligeable. Par la suite, ces variables scalaires peuvent être émulées et ainsi permettre la prédiction de champs 2D, une fois les champs reconstruits.

La mise en place de cette méthode est en cours et seuls des résultats préliminaires sont disponibles.

VI. CONCLUSION

Créer des émulateurs de grandeurs scalaires de sortie de modèle permet d'obtenir une estimation quasi-instantanée des zones de dépassement des valeurs repères utilisées en réponse à un accident nucléaire. Dès lors qu'un catalogue

d'émulateurs est construit à partir de scénarios accidentels « type », la production de l'expertise, et donc la mise à disposition des avis techniques pour les décideurs peut être accélérée, en particulier dans les premières phases de la crise. Ces émulateurs doivent cependant être évalués face à des échantillons de test pour vérifier leur fiabilité et parfois complétés de méthodes de classification pour augmenter leurs performances.

La première phase d'émulation a mis en évidence les principales sources d'imprécision. Des outils de classification en amont, pour arriver à prévoir un résultat de sortie nul ou une zone à 360°, ont grandement amélioré les résultats des émulateurs, mais ralentissent aussi parfois l'émulateur ainsi créé.

La discontinuité créée par le dépassement d'un seuil fixé vient complexifier l'émulation d'un tel problème. Pour améliorer les résultats, l'une des perspectives est l'étude d'une paramétrisation différente des données de sortie, qui permette d'éviter l'utilisation de seuil et donc d'éviter les discontinuités liées aux effets de seuil.

REFERENCES

- [1] Girard, S., Mallet, V., Korsakissok, I., Mathieu, A., 2016. Emulation and Sobol' sensitivity analysis of an atmospheric dispersion model applied to the Fukushima nuclear accident. *J. Geophys. Res. Atmospheres* 121, 3484–3496. <https://doi.org/10.1002/2015jd023993>
- [2] Le, N.B.T., Mallet, V., Korsakissok, I., Mathieu, A., Périllat, R., 2019. Meta-modeling and statistical calibration of the input parameters for atmospheric dispersion model applied to the Fukushima nuclear disaster., in: UNCECOMP 2019 Proceedings. Presented at the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, Institute of Structural Analysis and Antiseismic Research School of Civil Engineering National Technical University of Athens, Crete, Greece, pp. 215–228.
- [3] Roustant, O., Ginsbourger, D., Deville, Y., 2012. DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *J. Stat. Softw.* Vol 1 Issue 1 2012.
- [4] Girard, S., Iovleff, S., 2008. Auto-Associative Models, Nonlinear Principal Component Analysis, Manifolds and Projection Pursuit. In : *Lecture Notes in Computational Science and Engineering*. Springer Berlin Heidelberg, p. 202–218. doi : 10.1007/978-3-540-73750-6_8. url : <https://arxiv.org/pdf/1103.6119>.
- [5] Altman, N.S., 1991, An Introduction to Kernel and Nearest Neighbor. *The American Statistician*. 46 (3): 175–185.
- [6] Doury, A., 1981. Le vademecum des transferts atmosphériques. Rapport n° CEA-DSN-440, CEA. No CEA-DSN-440.