



TITRE :	<p align="center">RAPPORT DE STAGE DE FIN D'ETUDES cycle ingénieur en Mathématiques Appliquées et Calcul Scientifique.</p> <p align="center">Analyse et statistique descriptive des données des débits d'équivalent de dose</p>				
Auteur	BABE Wilfried Stéphane	Ecole :	Université Paris 13 Institut Galilée 	Entreprise de stage :	ELECTRICITE DE FRANCE 
Manager de branche	M. SAINTAMON Fabrice	Poste	Manager première ligne de la branche UNIE-GPEX-IRP		
Maitre de stage	M. ROCHER Alain	Poste	Expert radiochimie		
Encadrant du stage	M. JAUBERT Jerome	Poste	Ingénieur d'études.		
Intervenant Mathématiques dans le stage	M. LACAILLE Jerome	Poste	Ingénieur de recherche à SAFRAN et enseignant à l'école d'ingénieur Sup Galilée.		
Encadrant du stage et Directeur de la formation (MACS)	M. LAFITTE Olivier	Poste	Directeur de la formation MACS de l'Institut Galilée à Université Paris 13.		

REMERCIEMENT

Les travaux présentés dans ce rapport ont été effectués dans la branche Ingénierie de la Radioprotection (IRP) du département Groupe Prévention Environnement Exploitation (GPEX) sous la direction de l'Unité d'Ingénierie d'Exploitation (UNIE).

C'est avec honneur que je remercie M. Fabrice SAINTAMON, le manager de la branche IRP de m'avoir donné le privilège d'intégrer sa branche et de m'avoir appris le professionnalisme dans le monde de l'entreprise.

C'est d'abord à mon maître de stage M. Alain ROCHER, expert radiochimie, que j'exprime ma gratitude pour m'avoir donné la chance afin de peaufiner mes connaissances dans ce stage. Sa pédagogie dans les explications sur la structuration des réacteurs nucléaires et les tâches à faire à chaque niveau de l'évolution de mon travail m'ont permis de mieux comprendre le domaine physique. Je le remercie également pour ses analyses sur les résultats et surtout pour son professionnalisme.

Je tiens tout particulièrement à remercier M. Jérôme JAUBERT, ingénieur d'études qui a participé à mon encadrement, je lui sais gré pour son accueil chaleureux, sa disponibilité, sa pédagogie dans les explications à travers lesquels j'ai beaucoup appris en radioprotection. Ses compétences et ses expériences des sites nucléaires ont été un plus à mon apprentissage.

Je remercie également Mme. Christine PIGNOT, chargée d'affaires qui m'a fourni la base de données sur laquelle j'ai travaillé et a répondu à mes questions concernant la base de données. Merci à elle d'avoir facilité mon intégration dans la branche.

Je souhaite remercier M. Jérôme LACAILLE, ingénieur de recherche à SAFRAN et professeur à Sup Galilée pour son intervention dans le domaine de l'analyse descriptive dans le but de valider mes résultats.

De façon particulière je remercie Thomas CHRISTOPHE, ingénieur d'études, qui, par son accueil chaleureux au sein de son bureau, m'a aidé sur certains détails de la radioprotection. Aussi, il était là pour répondre à mes questions, tout en me donnant des idées pour la rédaction de ce rapport. Son professionnalisme a été un exemple pour moi et il a aussi facilité mon intégration dans la branche.

Du plus profond du cœur je remercie l'ensemble de la branche IRP pour son aimable accueil qui a facilité mon intégration, et pour son accompagnement tout au long de mon stage.

Je remercie M. Olivier LAFFITE et M. Emmanuel AUDUSSE respectivement directeur et responsable d'études de la formation MACS à Sup Galilée pour leur encadrement et leur autorisation pour faire ce stage.

Enfin, je tiens à adresser à mes amis et mes parents un très grand remerciement pour leurs encouragements et leurs conseils durant toute la période du stage.

Résumé

EDF est une entreprise dont l'activité est basée sur la production d'électricité, principalement d'origine nucléaire. Afin d'assurer la radioprotection des travailleurs, EDF met en œuvre un programme de gestion du terme source des tranches ¹. Ce programme repose sur l'état radiologique de la tranche qui est défini sur la base de l'analyse des indices de débits d'équivalent de dose mesurés à proximité des circuits ². A la suite de cette analyse, EDF met en place des méthodes de traitements si besoin met en œuvre les moyens nécessaires au maintien de la l'état radiologique de la tranche.

Mon stage a consisté à mener une étude multi-paramètres en faisant de l'analyse statistique. Mon rôle était de faire une étude statistique, pour comprendre au mieux les paramètres les plus influents dans les tendances des indicateurs radiologiques ² des principaux circuits des centrales (circuit primaire et auxiliaires). Cette étude réalisée sur les sept dernières années (2011-2017) a été instruite dans le but de fiabiliser l'analyse des indicateurs radiologiques des circuits. Dans un second temps, mon travail a consisté à bâtir un modèle statistique prédictif, sur la base des résultats provenant de l'étude multi paramètres, conduisant à anticiper l'évaluation des indicateurs radiologiques ² pour les arrêts ultérieurs. À cet effet, nous avons d'abord étudié les incertitudes sur les données expérimentales provenant des centrales ³ à travers un logiciel R et stabilisé les tendances des indicateurs radiologiques pour construire des modèles prédictifs. Ces modèles prédictifs ont été réalisés avec un logiciel Matlab en utilisant les modèles d'une méthode de machine Learning. La validation des modèles a été effective sur la base de données de l'année 2018 qui ne figuraient pas dans la construction de mon modèle.

Mots-clés :

Energie nucléaire, Radioprotection, Réacteurs, Modèle prédictif, Tendance, Stabilité, Circuit des centrales, indicateurs radiologiques, Machine Learning,

Abstract

EDF is a company whose activity is based on electricity production, mainly of nuclear origin. Therewith To ensure the radiation safety of workers, EDF is implementing a management of the term program source of the unit. This program is based on the radiological State of the unit that is defined on the basis of the analysis of dose equivalent flow indices measured near the circuits. As a result, EDF implements methods of treatment if need implements the means necessary for the maintenance of the radiological State of the unit.

My internship was to carry out a multi-parameter study with statistical analysis. My role was to do a statistical study to better understand the most influential parameters in the trends of the radiological indicators of main circuits of power plants (primary and auxiliary circuit). This study on the past seven years (2011-2017) was heard in order to make reliable analysis of the radiological indicators of circuits. Secondly, my job was to build a predictive statistical model, based on the results from the study multi-parameters, leading to anticipate the evaluation of radiological indicators for subsequent cases. For this purpose, we first studied the uncertainties on the experimental data from power plants through a software R and stabilized trends of radiological indicators to build predictive models. These predictive models have been made with a Matlab software using a method of machine learning models. The validation of the models has been effective on the database of the year 2018 not included in the construction of my model.

Keywords :

Nuclear power, Radiological protection, Reactor, Predictive model, Trend, Stability, power plants circuit, Radiological indicators, Machine Learning

SOMMAIRE

1- PRESENTATION DE L'ENTREPRISE	5
1.1 EDF	5
1.2 Direction, unité et branche	5
2- INTRODUCTION	6
2.1 Contexte	6
2.1.1 Radioprotection et centrales nucléaires	6
2.1.2 Indices radiologique de débits d'équivalent de dose	8
2.2 Objectifs du stage	9
2.2.1 Optimisation du choix des réacteurs à décontaminer	9
2.2.2 Anticipation du comportement des tranches	9
2.3 Plan du travail	10
3- ANALYSES DES DONNEES EXPERIMENTALES	10
3.1 Provenance des données expérimentales	10
3.2 Cohérence des données expérimentales	11
3.3 Conclusion de l'analyse des données	14
4- TRAITEMENT DES DONNEES	15
4.1 Construction des scénarios sur chaque palier	15
4.2 Principe d'ajustement des données aberrantes	16
4.2.1 Raison du traitement des données	16
4.2.2 Méthode de traitement des données	17
4.3 Comparaison des données initiales et finales	22
5- CONSTRUCTION ET EVALUATION DES MODELES PREDICTIFS	24
5.1 Modélisation des scénarios	24
5.2 Machine Learning et Matlab	26
5.2.1 Les types de méthode et modèles de machine Learning	27
5.2.2 Utilisation de machine Learning avec Matlab	29
5.3 Construction et validation des modèles	32
5.3.1 Choix des modèles sur chaque scénario	32
5.3.2 Validation et extrapolation des modèles	35
6- CONCLUSIONS ET PERSPECTIVES	38
7- REFERENCE	39
8- ANNEXE	40

1- PRESENTATION DE L'ENTREPRISE

1.1 EDF

La société anonyme EDF est une entreprise d'envergure et de notoriété dans le domaine de l'énergie nucléaire puisqu'elle est le premier groupe producteur et fournisseur d'électricité en France. En parallèle, le groupe consacre en 2014 37 % de ses investissements bruts de développement aux énergies renouvelables. EDF fait partie d'un groupe avec les filiales de la figure ci-dessous (cf. figure 1)



Figure 1 : les filiales de EDF en 2018

1.2 Direction, unité et branche

C'est à l'intérieur du site CAP AMPERE principalement à la division production nucléaire (DPN) que j'ai effectué mon stage. La DPN est structurée de la manière suivante : (figure 2).

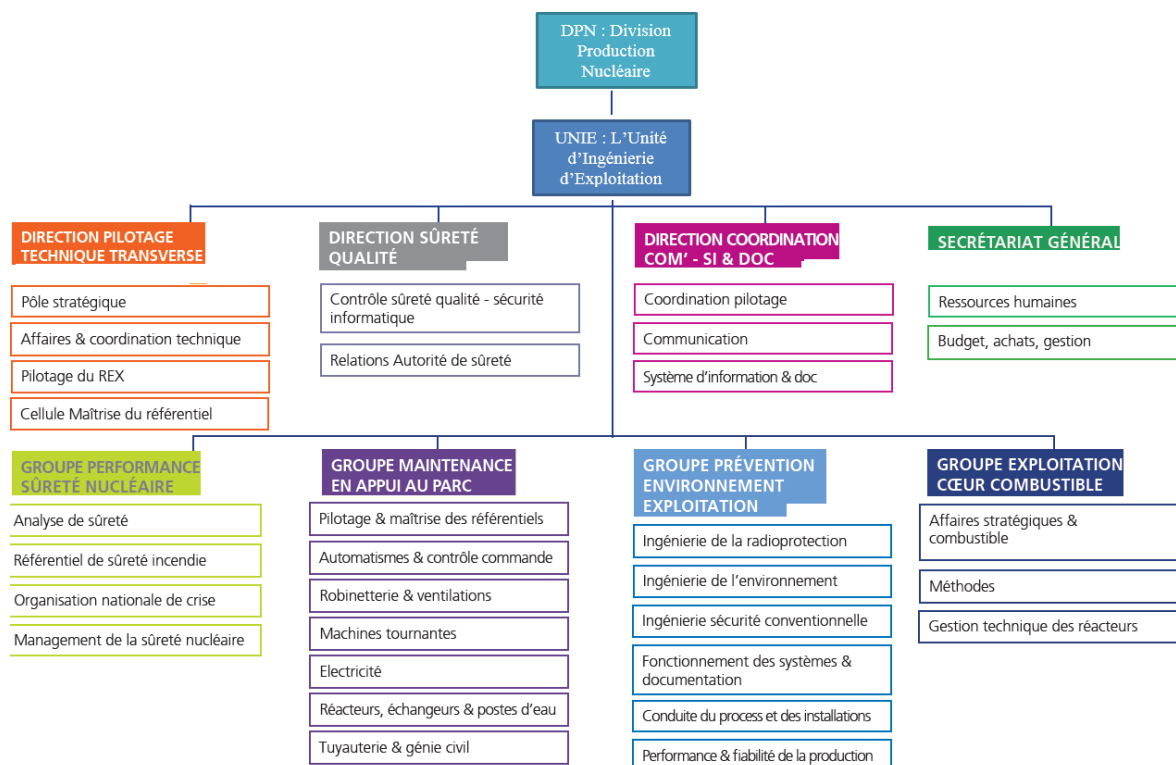


Figure 2 : Organigramme UNIE

C'est au sein du Groupe Prévention environnement Exploitation (GPEX), précisément dans la branche ingénierie de la radioprotection ⁴ (IRP) que j'ai été accueilli pour réaliser mes tâches. Les principales activités de la branche ingénierie radioprotection sont:

- **Piloter** la cohérence d'ensemble du référentiel radioprotection ou traiter des sujets spécifiques du référentiel
- **Contribuer** à la représentation institutionnelle en présentant des éléments dans des congrès, au club RP, à la Société Française de Radioprotection.
- **Rédiger des dossiers** spécifiques concernant la maîtrise des risques : risque xénon, alpha...
- **Associer à des expérimentations** organisationnelles sur les sites.
- **Participer** à l'animation technique des réseaux métiers.
- **Porter des affaires nationales** à enjeux radioprotection.
- **Chercher à optimiser** les conditions d'intervention dans le domaine de la radioprotection dans son domaine d'activité.
- **Réaliser la veille réglementaire** et analyser des nouveaux textes du domaine radioprotection.
- **Piloter des dossiers** concernant son domaine d'activité.

Le stage portait sur l'aspect «Chercher à optimiser les conditions d'intervention dans le domaine de la radioprotection dans son domaine d'activité».

2- INTRODUCTION

Une démarche simple et structurée explique les différents points abordés pour parvenir aux résultats de nos travaux. Cette démarche se présente comme décrite ci-dessous.

2.1 Contexte

2.1.1 Radioprotection et centrales nucléaires

EDF exploite 58 réacteurs nucléaires ³ à eau sous pression (REP). Le choix de l'énergie nucléaire en France par rapport à l'emploi des combustibles fossiles permet d'éviter chaque année le rejet dans l'atmosphère de:

- Poussières
- Dioxyde de soufre
- Dioxyde d'azote
- Gaz carboniques

On retrouve aussi plusieurs réacteurs sur une centrale nucléaire, ces réacteurs REP-EDF sont classés en trois puissances, aussi appelés **paliers**³ qui sont: REP 900 MW, REP 1300 MW, REP 1450 MW. Cependant, tous les réacteurs d'une même centrale sont obligatoirement du même palier, aussi ils présentent un principe de fonctionnement³ identique. (cf. figure 3).

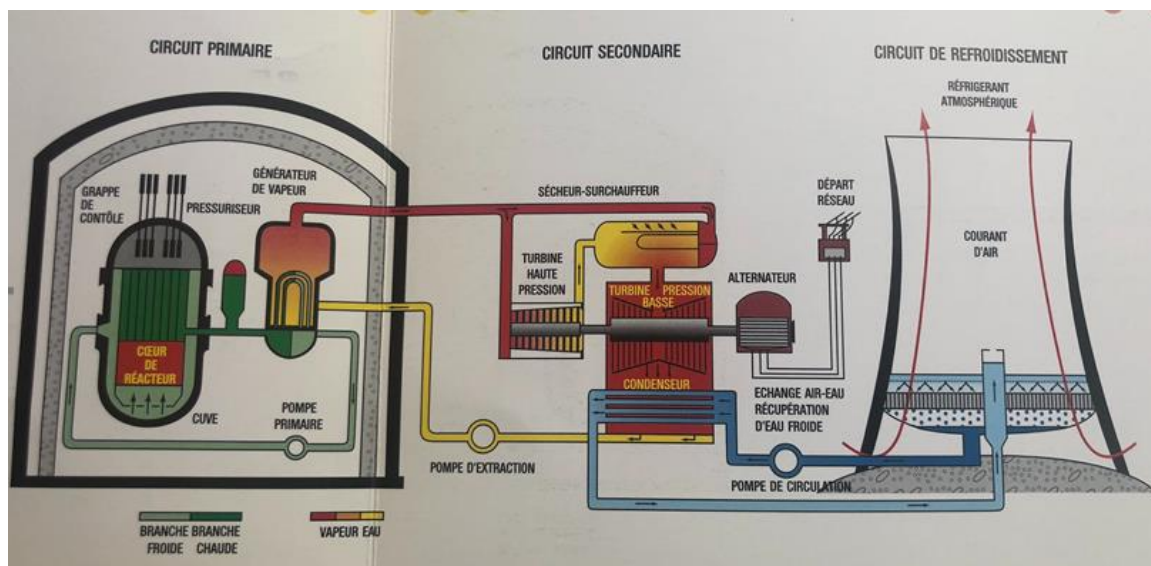


Figure 3 : Exemple de principe de fonctionnement d'une centrale nucléaire REP

En effet le fonctionnement d'une centrale est tel que le combustible nucléaire, sous forme d'assemblages, est logé dans la cuve du réacteur. La fission de noyaux d'uranium s'accompagne d'un très fort dégagement de chaleur. Cette énergie thermique est ensuite convertie en électricité par l'intermédiaire de trois circuits d'eau entièrement indépendants les uns des autres.

- **Le circuit primaire:** de l'eau à la pression de 155 fois la pression atmosphérique extrait la chaleur dégagée par le combustible, la libère dans quatre générateurs de vapeur et retourne dans la cuve du réacteur. Ce circuit renferme le combustible nucléaire, il est de ce fait enfermé dans l'enceinte de confinement.
- **Le circuit secondaire:** En récupérant la chaleur cédée par le circuit primaire dans les générateurs de vapeur, l'eau de ce circuit se vaporise et entraîne les quatre corps de la turbine. Celle-ci entraîne l'alternateur qui produit l'électricité. Après détente de la vapeur, celle-ci se transforme en eau dans le condenseur et retourne vers les générateurs de vapeur.
- **Le circuit de refroidissement :** Après avoir condensé la vapeur du circuit secondaire, l'eau de ce circuit, réchauffée, part vers le réfrigérant atmosphérique où elle se refroidit en cédant sa chaleur à l'air ambiant.

Une centrale nucléaire fonctionne avec des matières nucléaires, donc cela engendre des risques. A ce titre les ingénieurs radioprotection apportent un appui dans des domaines tels que : le zonage radiologique, la maîtrise des chantiers à risque radioprotection, la dosimétrie, la gestion des sources radioactives, etc. Ingénierie de la radioprotection est aussi apte à décrire de manière précise le comportement des systèmes souvent complexes qui se produisent sur les centrales nucléaires. De plus, depuis quelques décennies, les méthodes telles que par exemple, l'assainissement des tranches et l'élimination des points chauds ont été mises en place pour résoudre certains problèmes rencontrés sur les centrales. Ces actions mises en place sont également supervisées par GPEX-IRP.

2.1.2 Indices radiologique de débits d'équivalent de dose

Le débit d'équivalent de dose est un débit d'irradiation mesuré à l'aide d'un radiamètre par une personne physique. Dans le cas des indicateurs radiologiques de tranche, les points de mesure ont été choisis afin d'être représentatifs du débit d'équivalent de dose engendré par un circuit. (cf. figure 4).

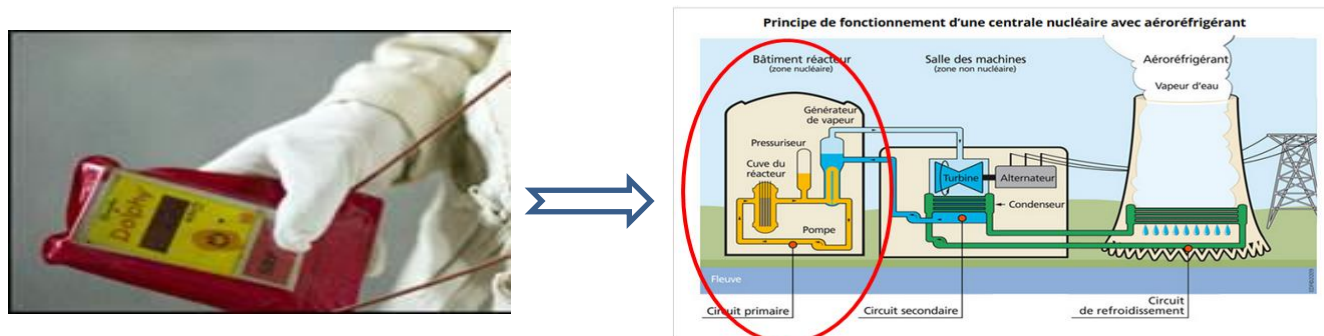


Figure 4 : Radiamètre, bâtiment réacteur et circuit primaire

Ce travail fait intervenir deux grands groupes : les indices de boucle du circuit primaire principal ($I_{Boucle_{RCP}}$)² et les indices du bâtiment réacteur (I_{BR})² qui contient des sous-indices par circuit appelés:

- I_{RRA} : Refroidissement du réacteur à l'arrêt
- I_{RCV} : Contrôle chimique et volumétrique
- I_{RPE} : Purges, événements et exhaures nucléaires
- I_{RCP} : Circuit primaire principal
- I_{PZR} : Pressuriseur
- I_{RIS} : Système d'injection de sécurité
- I_{GV} : Générateur de vapeur
- I_{PTR} : Traitement et réfrigération des piscines

L'indice du bâtiment réacteur est représentatif de l'état radiologique de la tranche. La valeur de l'indice (exprimé en mSv/h) est obtenue en faisant la moyenne de l'ensemble des sous-indices. De la même façon, les sous-indices sont calculés en réalisant la moyenne des points les composants.

EX : Le sous-indice RRA dans le palier 900 MW est la moyenne mesurée sur quatre points dans le circuit refroidissement du réacteur à l'arrêt : $I_{RRA_{900MW}} = \frac{P_1 + P_2 + P_3 + P_4}{4}$

Ces indices des réacteurs (tranches) d'un même palier peuvent être comparés d'un arrêt à l'autre pour évaluer l'état radiologique des différents circuits et détecter des dérives éventuelles afin d'y remédier.

2.2 Objectifs du stage

EDF met en œuvre un programme de gestion du terme source des tranches ¹. Ce programme repose sur l'état radiologique de la tranche qui est défini sur la base de l'analyse des indices de débits d'équivalent de dose mesurés à proximité des circuits. A la suite de cette analyse, EDF met en place des méthodes de traitements si besoin met en œuvre les moyens nécessaires pour maintenir l'état radiologique de la tranche comme tel. Dans le cadre de ce stage j'ai mené une étude multi-paramètres sur les débits d'équivalent de dose afin de consolider l'analyse des indices radiologiques des tranches et appréhender leurs évolutions. Ce travail s'est déroulé en deux étapes:

- ✓ Etude du comportement de chaque point de débits d'équivalent de dose mesuré afin de fiabiliser l'analyse des indicateurs radiologiques des circuits.
- ✓ Construction d'un modèle prédictif dans le but d'anticiper l'évolution des indices de débits d'équivalent de dose des circuits pour les années ultérieures

2.2.1 Optimisation du choix des réacteurs à décontaminer

Afin de mieux gérer de l'état radiologique des tranches, l'un des objectifs d'EDF est de décontaminer les tranches en retrait d'un point de vue radiologique. Cependant la difficulté réside dans la priorisation des tranches à traiter. Cette priorisation étant basée sur l'analyse des indices radiologiques de tranche (eux-mêmes obtenus par des mesures de débits d'équivalent de dose sur les circuits du Centre Nucléaire de Production d'Electricité (CNPE)) ³, il est nécessaire de fiabiliser l'étude de ces indices en utilisant des outils statistiques adaptés. Cette analyse constitue le premier objectif de ce stage. Elle prend en compte l'étude multi-paramètres tout en restituant aux mieux le comportement réel des débits d'équivalents de dose fournis par le radiamètre sur les différents circuits.

2.2.2 Anticipation du comportement des tranches

Après avoir fiabilisé les indices sur la base une étude multi-paramètres, une nouvelle base de données est extraite et elle contient ces indices fiables. Nous utilisons cette base pour construire des scénarios sur les différents circuits (jeux de données) afin de mener une étude statistique plus approfondie. Cette étude est le deuxième objectif de ce stage, qui consiste à bâtir un modèle prédictif à travers les méthodes machine Learning avec les scénarios construits sur les différents circuits afin d'anticiper l'évolution des indices sur années à venir.

2.3 Plan du travail

Pour répondre aux différents objectifs, nous suivrons une méthodologie simple et structurée de la manière suivante :

- En premier nous analyserons les données mesurées sur les centrales
- En second plan nous traiterons ces données en utilisant des méthodes statistiques à travers des logiciels R et Matlab.
- Enfin, nous construirons et évaluerons des modèles prédictifs en utilisant des méthodes de machine Learning sur différents scénarios.

3- ANALYSES DES DONNEES EXPERIMENTALES

Ce chapitre consiste à étudier les données de chaque réacteur sur les différents circuits et déterminer l'incertitude des données expérimentales dans le but de les traiter. L'étude de l'analyse de données est sur l'ensemble des tranches d'un même palier (900 MW, 1300 MW, 1450 MW). Pour comprendre au mieux le comportement des données mesurées à proximité des circuits, une procédure en mise en place de la manière suivante.

3.1 Provenance des données expérimentales

Dans le cadre de ce travail, nous avons utilisé des données expérimentales enregistrées par des radiamètres puis intégrées dans des fichiers Excel qui est la base de données (BDD). Cette BDD qui est rigoureusement contrôlée par une ingénieure de la branche IRP, contient les valeurs des indices (en mSv/h) des débits d'équivalent de dose des différents circuits pour tous les 58 réacteurs de 2011 à 2017. L'indice étant la moyenne des types de point qui le compose, ces types de points varient d'un circuit à un autre en fonction du palier. Pour chacun des paliers, ces fichiers proposent une liste de réacteur³ (BLA1, DAM4, CHB3, ...) et de circuit² (RRA, RCV, RPE, RIS, PTR, PZR, GV, RCP). Dans ce travail, on s'intéresse aux propriétés des indices, dont voici les corrélations:

Indicateur radiologique:
$$indice^{DeD} = \frac{P_1 + P_2 + \dots + P_N}{N}$$

Les tableaux 1,2 et 3 (cf. Annexe) distinguent les réacteurs qui constituent les différents paliers sur lesquels nous ferons ces études. Ces familles sont les suivantes :

- Les REP 900 MW
- Les REP 1300 MW
- Les REP 1450 MW

A la suite de ce travail nous prendrons comme base d'apprentissage l'ensemble de chaque palier. Les résultats seront ensuite analysés par famille.

3.2 Cohérence des données expérimentales

L'étude multi-paramètre implique qu'on utilise des données de plusieurs tranches avec tous les circuits de façon simultanée pour faire une étude statistique. Si ces données ne sont pas cohérentes entre elles, il sera impossible de trouver des modèles fiables qui puissent avoir un sens et permettre une extrapolation vers les années ultérieures (2018-xxxx). Dans l'optique de déterminer l'incertitude des données, nous avons fait un test qui permet d'évaluer la cohérence des indices. Le test est basé sur une procédure d'analyse des tendances des indices. Cette analyse des tendances présente l'évolution temporelle des indices sur différents circuits (cf. figure 5). Les données de ces tendances proviennent des tableaux 1, 2 3 (cf. Annexe).

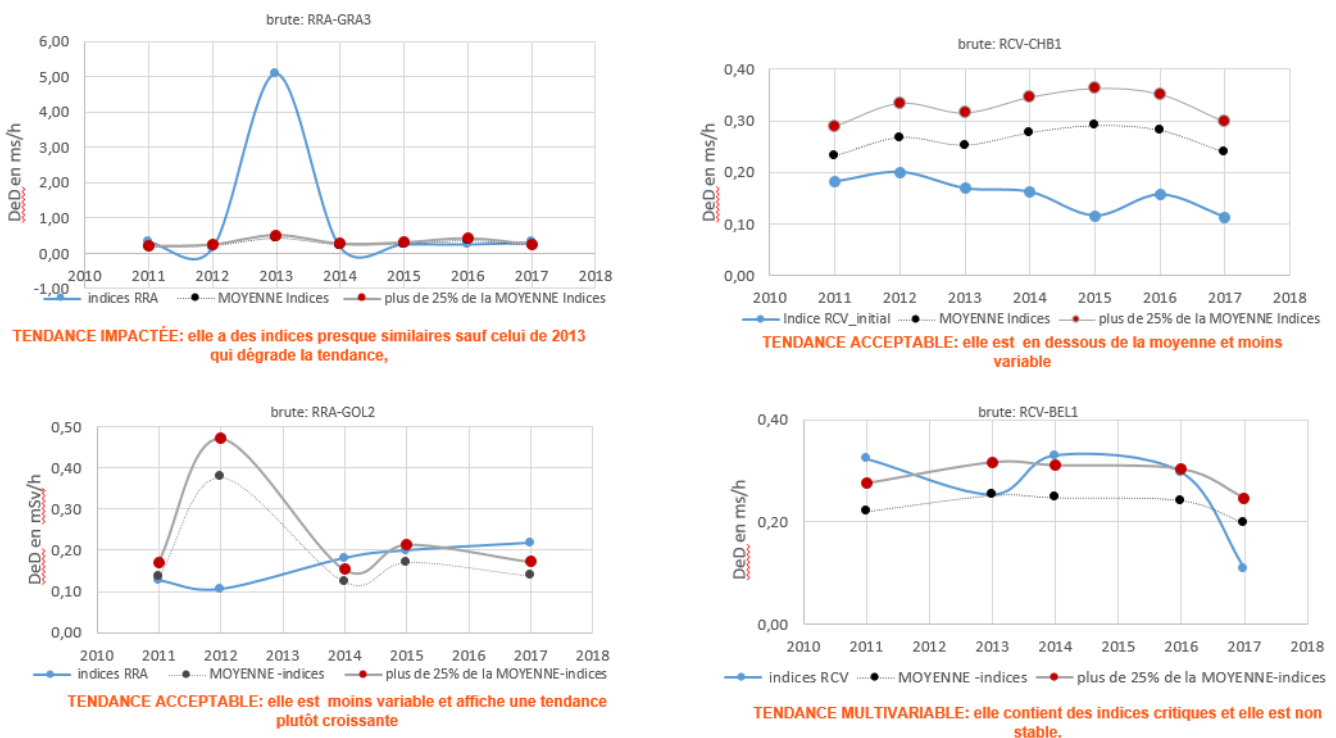


Figure 5 : Présentations et appréciations des tendances des indices sur chaque circuit et réacteur

Après avoir représenté toutes les tendances des données de la BDD on obtient cette première analyse brute en exemple cité ci-dessous



Figure 6 : répartition en % des tendances des indices des circuits en fonction des réacteurs sur chaque palier

Afin de mieux comprendre la cohérence des données cette analyse fait appel à la méthode de boîte à moustache⁵. La boîte à moustaches étant une traduction de Box & Whiskers Plot⁵, est une invention de TUKEY(1977)⁵ pour représenter schématiquement la distribution d'une variable. Cette représentation graphique peut être un moyen pour approcher les concepts abstraits de la statistique, si l'on pratique son usage sur différents jeux de données. Elle utilise 5 valeurs qui résument des données : le minimum, les 3 quartiles Q_1 , Q_2 (médiane), Q_3 , et le maximum.

Procédure:

- 1- Classer les n données par ordre croissant.
- 2- Diviser les données en 2 groupes de tailles égales. On obtient le groupe du bas et le groupe du haut, chacun contenant 50% des observations.
- 3- Calculer à nouveau la médiane du groupe du bas. On obtient le quartile Q_1 , qui correspond à 25 % des observations.
- 4- Calculer à nouveau la médiane du groupe du haut. On obtient le quartile Q_3 , qui correspond à 75 % des observations.

L'écart interquartile est utilisé comme indicateur de dispersion. Il correspond à 50% des effectifs situés dans la partie centrale de la distribution.

Lecture d'une boîte à moustaches (figure 7):

- l'échelle des valeurs de la variable, située sur l'axe vertical.
- la valeur du 1er quartile Q_1 (25% des effectifs), correspondant au trait inférieur de la boîte.
- la valeur du 2ème quartile Q_2 (50% des effectifs), représentée par un trait horizontal à l'intérieur de la boîte.
- la valeur du 3ème quartile Q_3 (75% des effectifs), correspondant au trait supérieur de la boîte,
- les 2 « moustaches » inférieure et supérieure, représentées ici par les petits rectangles verticaux de part et d'autre de la boîte. Ces 2 moustaches, délimitent les valeurs dites adjacentes qui sont déterminées à partir de l'écart interquartile ($Q_3 - Q_1$).
- les valeurs dites extrêmes, atypiques, exceptionnelles, (outliers) situées au-delà des valeurs adjacentes sont individualisées. Elles sont représentées par des marqueurs (carré, ou étoile, etc.).

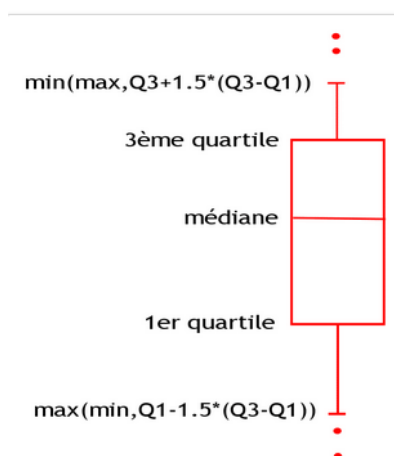


Figure 7 : Présentation d'une boîte à moustache

Pour TUKEY la valeur 1.5 est donc un compromis pour retenir comme atypiques assez d'observations mais pas trop d'observations. Nous utilisons la méthode de boîte à moustache à travers un logiciel R. ce logiciel est conçu pour réaliser au mieux des études statistique. Pour utiliser cette méthode sur R il y a des étapes à suivre qui sont :

- Construire un tableau de données sur Excel par exemple : Classeur2.csv
- Importer le tableau dans R: `Data1 <- read.csv ("chemin ou se trouve le fichier Excel /nom du fichier Excel ", sep= ";")`
- Calculer les 5 valeurs qui résument des données : `summary(Data1)`
- Afficher les outliers et les boîtes à moustache de chaque variable : `boxplot(Data1)$out`

Après avoir expliqué la méthode de boîte à moustache et comment le mettre en pratique sur R, un exemple de d'application est mis en évidence à la suite de ce rapport.

Le tableau 4 suivant représente une partie des données expérimentales de débits d'équivalent de dose (DED) mesurés sur quatre types de points formant l'indice du circuit RRA du palier 900 MW du tableau 1 (cf. Annexe).

réacteurs	années	PTn°4	PTn°11	PTn°12	PTn°13
BLA1	2011	0,28	0,03	0,07	0,26
CRU2	2012	0,24	0,1	0,1	0,06
CRU3	2012	0,15	0,045	0,04	0,13
CRU4	2012	0,25	0,35	1	0,08
GRA1	2013	1	0,45	0,051	0,35
GRA2	2013	0,3	0,08	0,12	0,1
GRA3	2013	20	0,06	0,05	0,2
GRA4	2013	0,3	0,13	0,17	0,3
CRU2	2014	1,3	0,25	0,2	0,05
CRU3	2014	0,15	0,1	0,06	0,08
CRU4	2014	0,25	0,13	0,085	0,045
GRA2	2016	0,5	0,1	0,3	0,3
GRA3	2016	0,56	0,071	0,26	0,14
BLA3	2016	0,09	0,14	0,13	0,07
TRI3	2016	0,42	0,45	0,2	0,2
GRA4	2013	0,3	0,13	0,17	0,3
CRU2	2017	1,3	0,25	0,2	0,05
CRU3	2017	0,15	0,1	0,06	0,08
CRU4	2017	0,25	0,13	0,085	0,045
SLT2	2017	0,6	0,5	0,32	0,52
TRI1	2017	0,4	0,15	0,19	0,35
TRI3	2017	0,8	0,3	0,2	0,11

Tableau 4 : un extrait de DED sur 4 points formant l'indice RRA du palier 900 MW

Remarques : Ce scénario a été fait sur chaque circuit en fonction du type de palier. Chaque colonne représente une variable. A travers la méthode de boîte à moustache nous déterminons les données incohérentes appelées communément outliers ou points aberrants.

Boîte à moustache:

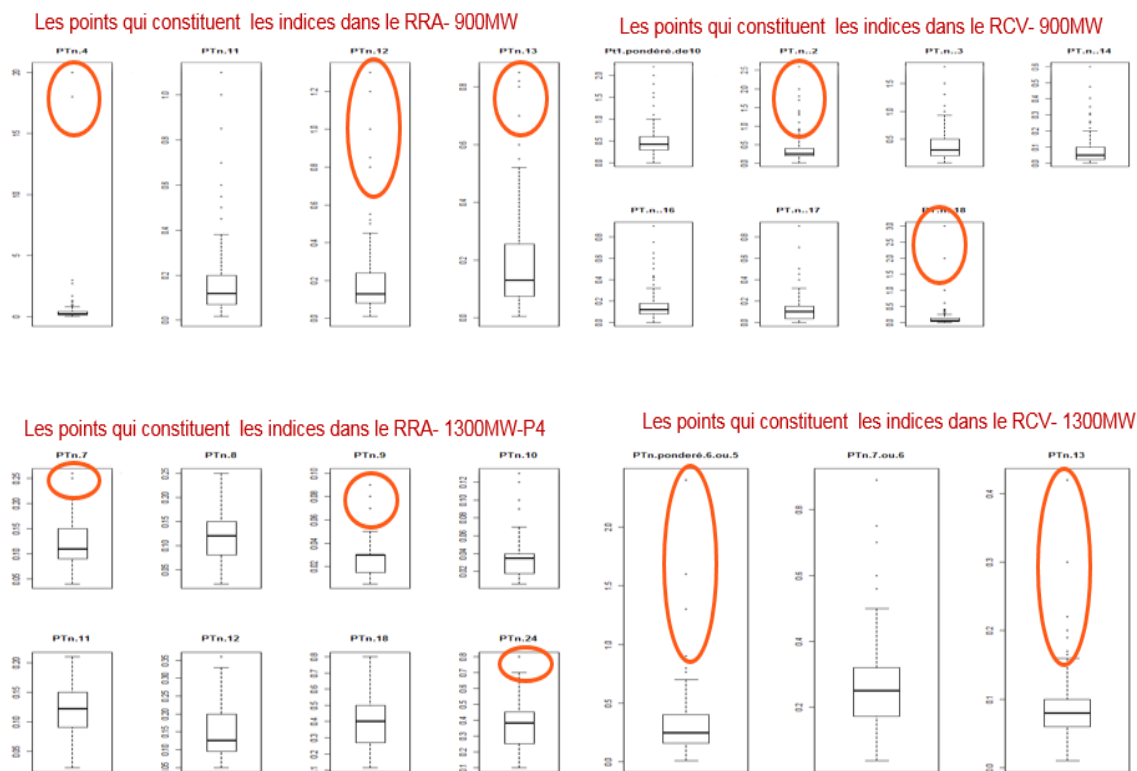


Figure 8: Identifications des points aberrants sur les types de points les indices

Les points entourés en orange représentent les points aberrants sur chaque variable. Ce premier travail détermine les points aberrants du point de vue statistique, cependant tous ne le sont pas. En tenant compte de l'avis de l'expert du domaine de la radioprotection, nous constatons qu'il nous faut valider les résultats fournis par la boîte à moustache en fixant un seuil (valeurs à partir de laquelle on justifie que la donnée est dite aberrante) afin de mieux décrire la cohérence des données car certains points aberrants ont du sens selon le type de circuits.

Dans cette optique nous allons opter pour une étude basée sur les rigueurs de la radioprotection. Elle permet de fixer des seuils de débits d'équivalent de dose sur chaque colonne (cf. Tableau 4). Cette pratique fait appel à l'utilisation de la fonction Hampel ⁶ (voir partie 4).

3.3 Conclusion de l'analyse des données

L'analyse des données expérimentales nous conduit à une approche originale, basée sur une étude statistique de Tukey. Cette étude mise en évidence grâce aux outils informatiques, nous fournit une visualisation graphique de la dispersion des données à l'aide d'un raisonnement basé sur la propagation des incertitudes. La méthode numérique (boîte à moustache) a permis de donner une

conclusion concernant les tranches avec des données irrégulière (moins meilleur) qui méritent d'être traité afin de construire un modèle prédictif sur une bonne base de données.

Après toutes ces analyses nous avons conclu qu'il fallait prioriser le traitement des données des tranches telles que : BUG2 ,3 ,4 ,5 , GRA1,2,3,4,5,6, TRI1,2,3,4, FES1,2 etc. (cf. Annexe).

4- TRAITEMENT DES DONNEES

Dans ce chapitre on se sert des résultats l'analyse des données (voir partie3) pour cibler les données des tranches qui doivent être traitées. De manière générale, il consiste à fiabiliser les données en traitant les points manquants et les points aberrants de la BDD.

4.1 Construction des scénarios sur chaque palier

L'ajustement implique des choix concernant les tranches et les circuits à traiter. Dans notre contexte, on aura trois grands scénarios pour traiter les données manquantes et les données aberrantes des 7 dernières années (2011-2017).

- *Palier 900 MW: RRA, RCV, RCP, RPE, PTR, PZR, GV, RIS, BOUCLE*
- *Palier 1300 MW: RRA, RCV, RCP, RPE, PTR, PZR, GV, RIS, BOUCLE*
- *Palier 1450 MW: RRA, RCV, RCP, RPE, PTR, PZR, GV, RIS, BOUCLE*

Palier 900 MW	Palier 1300 MW	Palier 1450 MW
I_{RRA} : 4 POINTS	I_{RRA} : 7 POINTS	I_{RRA} : 6 POINTS
I_{RCV} : 7 POINTS	I_{RCV} : 3 POINTS	I_{RCV} : 3 POINTS
I_{RCP} : 11 POINTS	I_{RCP} : 18 POINTS	I_{RCP} : 16 POINTS
I_{RPE} : 4 POINTS	I_{RPE} : 3 POINTS	I_{RPE} : 3 POINTS
I_{PTR} : 1 POINT	I_{PTR} : 2 POINTS	I_{PTR} : 2 POINTS
I_{PZR} : 5 POINTS	I_{PZR} : 5 POINTS	I_{PZR} : 5 POINTS
I_{GV} : 12 POINTS	I_{GV} : 4 POINTS	I_{GV} : 4 POINTS
I_{RIS} : 3 POINTS	I_{RIS} : PAS DE POINTS	I_{RIS} : PAS DE POINTS

Tableau 5 : Présentation des scénarios qui ont servi de base d'apprentissage pour traiter les données

Palier 900 MW : ce palier regroupe 34 réacteurs dont la puissance de production électrique est 900 MW.

Palier 1300 MW : ce palier regroupe 20 réacteurs dont la puissance de production électrique est 1300 MW.

Palier 1450 MW : ce palier regroupe 4 réacteurs dont la puissance de production électrique est 1450 MW.

Chaque indice de circuit est traité indépendamment des autres. Pour visualiser les différents réacteurs qui composent les paliers voir les tableaux 1, 2 et 3 (cf. Annexe).

4.2 Principe d'ajustement des données aberrantes

4.2.1 Raison du traitement des données

En général, la tendance que l'on souhaite observer sur les principaux circuits du bâtiment réacteur est que, plus les années avancent plus on observe une tendance monotone sur les indices. Aussi on souhaite voir la tendance des indices d'un circuit suivre la même tendance que celle de la moyenne des indices du palier.

Exemple : la tendance des indices du circuit RPE du réacteur Chinon 3 (palier 900 MW) en bleu et la moyenne du palier 900 MW en pointiez noir, puis celle en rouge représente les 25% de plus de la moyenne.

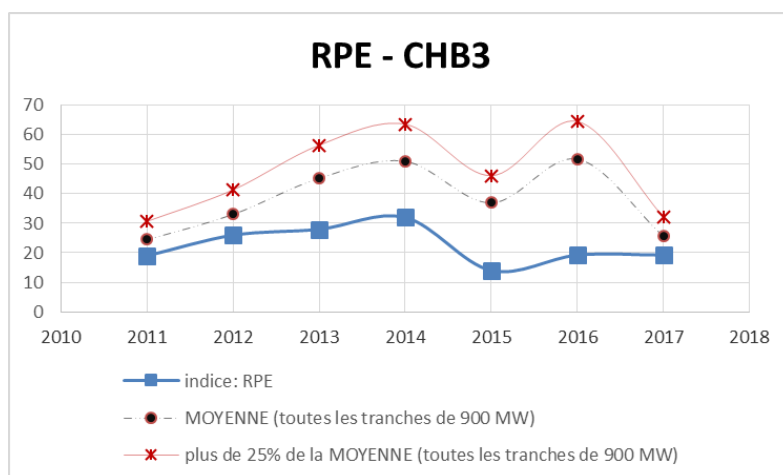


Figure 9: Exemple tendance qui ne nécessite pas de traitement particulière

Sur cette figure 9 nous avons des données d'un circuit qui ne nécessitent pas de traitement particulier puisque la tendance de ces indices suit la même allure que la moyenne du palier. De plus, elle est en dessous de la moyenne du palier tout en étant pratiquement stable car elle ne présente pas de point qui suscite une certaine interprétation.

Malheureusement, après avoir effectué une analyse sur l'ensemble des indices (cf. Tableau 1, 2 et 3) pour mieux observer les cohérences des données nous avons déterminé les circuits qui présentent des points aberrants. Ces points sont dits aberrants puisqu'ils détériorent la monotonie des tendances des indices du circuit.

En outre, certaine tendance des indices de quelques réacteurs dans les différents paliers sont difficilement interprétables, étant donné qu'on ne dispose pas d'assez de données les concernant. Exemple, les réacteurs de la centrale Nogent qui présentent que 3 indices au lieu de 7 indices sur les 7 années.

Le but étant aussi de construire un modèle prédictif, il faut donc s'assurer des cohérences des données, d'où l'intervention de l'ajustement des données qui exigent des traitements.

4.2.2 Méthode de traitement des données

Le traitement de données incertaines est mis en œuvre par le biais d'un logiciel Matlab. Dans le cadre de ce travail, Matlab détecte les points aberrants et manquants des circuits sur l'ensemble des différents paliers, puis les remplacent par la médiane du palier. Pour mieux comprendre le scénario voir tableau 6.

PALIER PUISSANCE	REACTEUR	INDICE BOUCLE RCP FIN CYCLE 1E-2mSv/h	INDICE RRA 1E- 2mSv/h	INDICE RCV 1E- 2mSv/h	INDICE RPE 1E- 2mSv/h	INDICE RIS 1E- 2mSv/h	INDICE PZR 1E- 2mSv/h	INDICE GV 1E- 2mSv/h	INDICE RCP 1E- 2mSv/h	INDICE PTR 1E- 2mSv/h
900	BLA2	22	11	10	10	25	7	17	17	30
900	GRA4	36	12	22	15	18	16	23	29	80
900	TRI1	56	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait
900	BLA2	19	15	10	14	12	6	12	18	30
900	CRU3	19	9	13	19	16	24	14	23	23
900	GRA4	112	20	56	544	50	26	41	53	20
900	TRI1	70	22	36	31	9	25	25	40	3
900	BLA2	51	10	21	14	10	21	32	35	10
900	BUG5	31	10	76	16	3	17	22	32	23
900	CRU3	19	4	12	6	18	16	16	15	12
900	GRA4	83	508	33	569	47	10	33	79	32
900	TRI1	75	27	28	28	11	13	54	58	4
900	BLA2	64	13	26	18	14	12	39	41	20
900	BUG2	84	11	19	19	33	35	21	25	18
900	CRU3	47	13	21	52	3	16	16	30	200
900	GRA4	33	24	30	19	19	20	24	45	20

Tableau 6 : Exemple de jeu de donnée contenant des incertitudes

Cet exemple montre une partie du travail fait sur le palier 900 MW. Les chiffres représentent les données des principaux circuits (Boucles et sous circuit du bâtiment réacteur). Les données encadrées en rouge sont celles qui ont été détectées par le logiciel R (voir chapitre 3) puis validé par l'expert du domaine comme étant un des points aberrants après analyse des données de chaque variable. Ces points dits aberrants sont aussi décelés par Matlab en adaptant une méthode de traitement qui est explicitée ci-après.

Remarque: une colonne de circuit représente une variable et le traitement des données d'une variable à une autre ce fait de façon indépendante.

Pour rappel le logiciel R détecte des points qui sont statistiquement des points aberrants. Toutefois, en prenant en compte l'avis de l'expert domaine de la radioactivité, il y a des points aberrants qui sont justifiables, par conséquent ne nécessitent pas forcément un traitement. Par la suite, les points défectueux qu'affiche le logiciel R pour chaque variable doivent être traités avec une méthode adaptée par rapport à l'interprétation de l'expert. Dans ce traitement il s'agit de fixer un seuil à partir duquel une donnée est considérée comme un point aberrant. Pour cela un programme est mis en place en fixant un seuil significatif aux différentes variables à travers le logiciel Matlab. Le seuil varie d'une variable à une autre, c'est la valeur fixée pour laquelle tout point aberrant détecté par le logiciel R se trouvant supérieur à celle-ci est déclaré comme points aberrants et donc nécessite un traitement dans le cadre de ce travail.

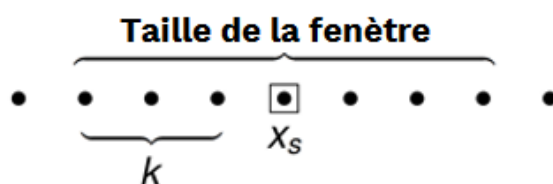
Cette partie du travail se fera avec la fonction de traitement de donnée de Matlab nommée Hampel ⁶. Hampel est une fonction développée par MathWorks qui est incorporée dans Matlab à travers le package Matlab Signal Processing. Pour chaque jeu de données, elle calcule une médiane locale et estime l'écart type d'un ensemble d'échantillon. Si un échantillon diffère de la médiane locale de plus de trois écarts, il est remplacé par la médiane. La fonction Hampel traite chaque colonne des données comme un canal indépendant.

Algorithme général de la fonction Hampel : $\text{hampel}(x, k, n_{\sigma})$:

X = les données

K = le nombre d'échantillon à prendre sur les données

n_{σ} = le seuil



Pour un échantillon de données:

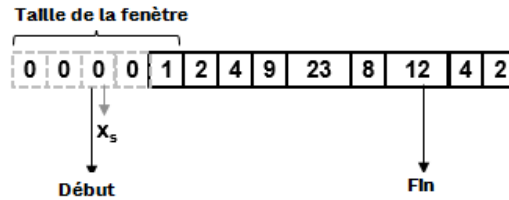
- La fonction Hampel parcourt tous les points avec la même procédure : ici l'exemple est fixé sur x_s puis centre la fenêtre de longueur irrégulière sur l'échantillon actuel.
- On calcule la médiane locale, m_i et l'écart-type σ_i , sur la fenêtre de données actuelle
- Compare l'échantillon courant avec $n_{\sigma} \times \sigma_i$, où n_{σ} est la valeur de seuil.
Si $|x_s - m_i| > n_{\sigma} * \sigma_i$, le filtre identifie l'échantillon courant x_s comme une valeur aberrante et le remplace par la valeur médiane m_i .

Cet algorithme est explicité en prenant un exemple :

La difficulté de cet algorithme est le traitement aux abords des données d'une variable. Considérons un cadre de données qui est transmis au filtre Hampel.

1	2	4	9	23	8	12	4	2
---	---	---	---	----	---	----	---	---

Dans cet exemple, le filtre Hampel fait glisser une fenêtre de longueur 5 (Lon) sur les données. Le filtre a une valeur seuil de 2 (n_σ). Pour avoir une fenêtre complète au début de la trame, l'algorithme de filtrage ajoute le cadre à (Lon - 1) zéros. Pour calculer le premier échantillon de la sortie, la fenêtre est centrée sur le $\left(\frac{Lon-1}{2} + 1\right)^{ième}$ échantillon dans la trame ajoutée. Dans notre cas se sera le troisième zéro puisque Lon=5.



Le filtre calcule la médiane, l'écart absolu médian et l'écart type sur les données de la fenêtre locale. Sur l'échantillon actuel: $x_s=0$, on a :

- La fenêtre de données capter qui est $[0 \ 0 \ 0 \ 0 \ 1]$ avec la médiane locale: $m_i = 0$.
- Déviation absolue médiane: $mad_i = \text{médiane}(|x_{i-k} - m_i|, \dots, |x_{i+k} - m_i|)$. Pour cette fenêtre de données on a, $mad = \text{médiane}(|0 - 0|, \dots, |1 - 0|) = 0$.
- Écart type: $\sigma_i = \delta * mad = 0$, avec $\delta = \frac{1}{\sqrt{2} \operatorname{erfc}^{-1}\left(\frac{1}{2}\right)} \approx 1.4826$

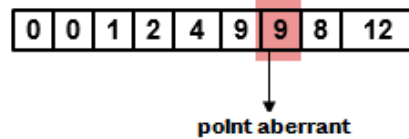
$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$, avec $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\varphi^2} d\varphi$ qui la fonction erreur, permettant d'exprimer la marge d'erreur des évaluations statistiques.

L'échantillon courant, $x_s = 0$ ne respecte pas la relation pour la détection des valeurs aberrantes. $|x_s - m_i| > [(n_\sigma * \sigma_i) = 0]$. Par conséquent, le filtre Hampel génère l'échantillon d'entrée actuel $x_s = 0$.

Répétez cette procédure pour chaque échantillon suivant jusqu'à ce que l'algorithme centre la fenêtre sur le $\left[Fin - \frac{Len-1}{2}\right]^{ième}$ échantillon. Comme la fenêtre centrée sur les derniers échantillons $\frac{Len-1}{2}$ ne peut pas être pleine, ces échantillons sont traités avec la trame suivante des données d'entrée.

DOCUMENT CONFIDENTIEL

Résultat sur le premier cadre de sortie généré par le filtre Hampel:



Le septième échantillon de la trame d'entrée ajoutée, 23, est une valeur aberrante. Le filtre Hampel remplace cet échantillon par la médiane sur la fenêtre locale [4 9 23 8 12].

Maintenant que la fonction Hampel a été décrite à travers l'exemple ci-dessus, un résumé de code sur la fonction Hampel à travers Matlab pour le traitement des indices de débits d'équivalent de dose est présenté de la manière suivante:

- ❖ Construction du scénario sur Excel: c'est à dire le jeu de données à traiter
- ❖ Ouvrir Matlab en double cliquant sur l'icône Matlab. Ensuite, écrivez « ver » dans la console, puis taper sur la touche entrer pour visualiser la liste des packages installés afin de s'assurer que le package «Signal Processing Toolbox » est bien installé. Ce package est l'un des packages qui permet de faire de la Data Analyse. Dans notre contexte il permet l'utilisation de la fonction Hampel.

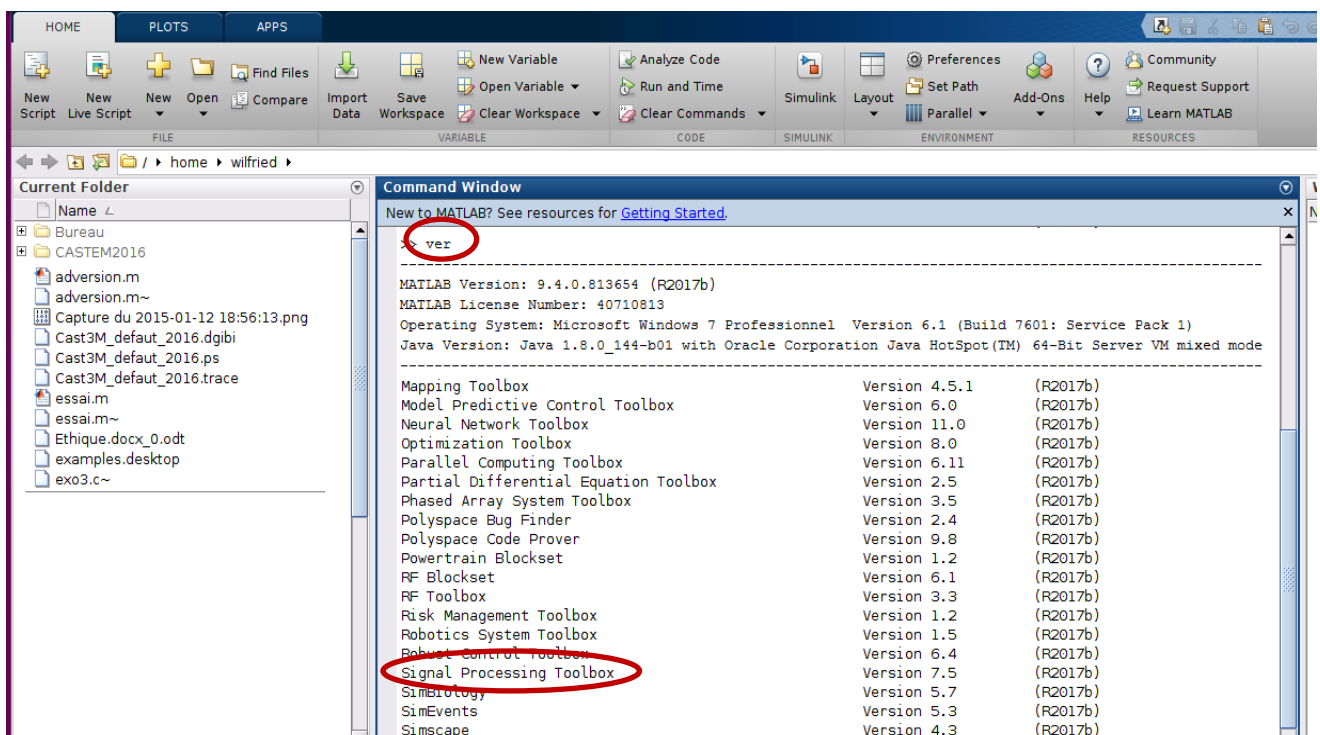
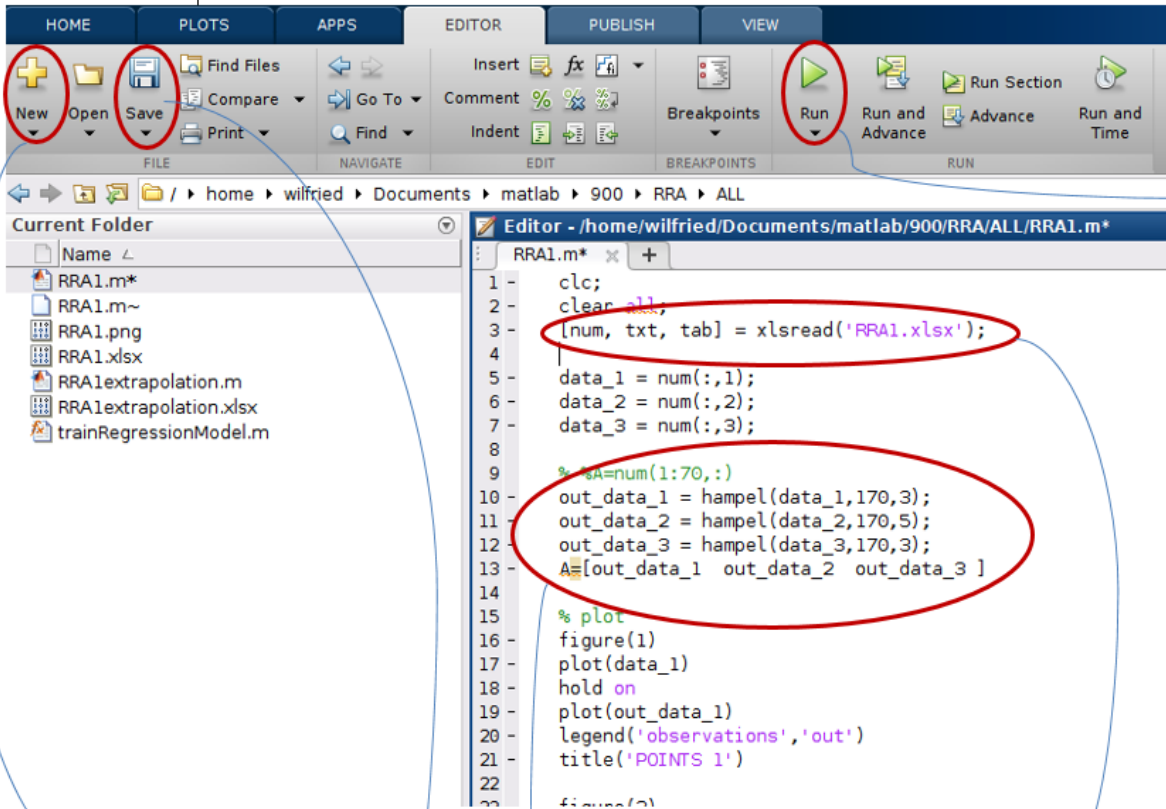


Figure 10: liste de package qu'on peut trouver dans un logiciel Matlab

- ❖ Créer un fichier .m sur Matlab et enregistrer le dans le même dossier que le fichier Excel, puisque le fichier Excel et l'interface du logiciel Matlab doivent être liés pour interagir.
- ❖ Importer le fichier Excel contenant le jeu de données à l'aide d'une ligne de code indiquée dans le programme ci-dessous.
- ❖ Faire appel à la fonction Hampel pour traiter les données de chaque variable.



Créer un nouveau fichier .m dans lequel on écrit le code source de notre programme

Sauvegarder le fichier créé dans le même dossier que le fichier Excel

La ligne de code qui permet à l'interface de Matlab de récupérer les données du fichier Excel. Selon la ligne de code le nom du fichier est RRA1.xlsx

Ligne de code faisant appel à la fonction Hampel qui permet de traiter chaque colonne de notre fichier Excel

Compiler et exécuter le programme

```

1 - clc;
2 - clear all;
3 - [num, txt, tab] = xlsread('RRA1.xlsx');
4 -
5 - data_1 = num(:,1);
6 - data_2 = num(:,2);
7 - data_3 = num(:,3);
8 -
9 - % A=num(1:70,:);
10 - out_data_1 = hampel(data_1,170,3);
11 - out_data_2 = hampel(data_2,170,5);
12 - out_data_3 = hampel(data_3,170,3);
13 - A=[out_data_1 out_data_2 out_data_3 ]
14 -
15 - % plot
16 - figure(1)
17 - plot(data_1)
18 - hold on
19 - plot(out_data_1)
20 - legend('observations','out')
21 - title('POINTS 1')
22 -
23 - figure(2)
  
```

Figure 11: détail de l'utilisation de la fonction Hampel à travers le logiciel Matlab

Pour résumer cette partie (partie 4) du travail consistait à montrer pourquoi traiter les données et comment les traiter.

- Pourquoi traiter les données ? : pour ajuster les points aberrants des débits d'équivalent de dose du BDD afin de fiabiliser les indices et construire un modèle prédictif (voir chapitre 5).

- Comment les traiter ? : en prenant en compte l'avis de l'expert du domaine radiochimie pour fixer un seuil à partir duquel un point est dit aberrant, puis utiliser la fonction Hampel à travers le logiciel Matlab pour mener à bien le traitement de ces données.

4.3 Comparaison des données initiales et finales

Lorsqu'on traite les données expérimentales dans le but d'ajuster de façon correcte sans biaiser les résultats, il est faux de vouloir traiter tous les points qui présentent un pique élevé. Bien que l'opération ait été faite sur la rigueur des statistiques, l'incertitude sur les données existe toujours. Cette incertitude est due à la manière dont sont mesurés les Débits d'équivalent de doses depuis les centrales. Cependant, le traitement de ces données a permis de reconstruire une base de données fiable à l'attendu (relative au secteur d'activité). Ces nouvelles données sont extraites du logiciel Matlab après traitement, puis elles sont comparées aux données brutes. (Figure 12).

- La figure 12 présente deux groupes de données sur l'ensemble des indices de débits d'équivalents de doses qui sont les indices du circuit de boucle RCP et du circuit RRA sur l'ensemble des 34 réacteurs du palier 900 MW de 2011-2017. Chaque groupe contient les données initiales en bleu représentant les indices qui proviennent des centrales et les données finales en orange qui sont celles obtenues après traitement des données initiales. Sur la figure de gauche on constate qu'il n'y a pas d'apparition de points aberrants et sur la figure de droite on a six points aberrants traités par la fonction Hampel.

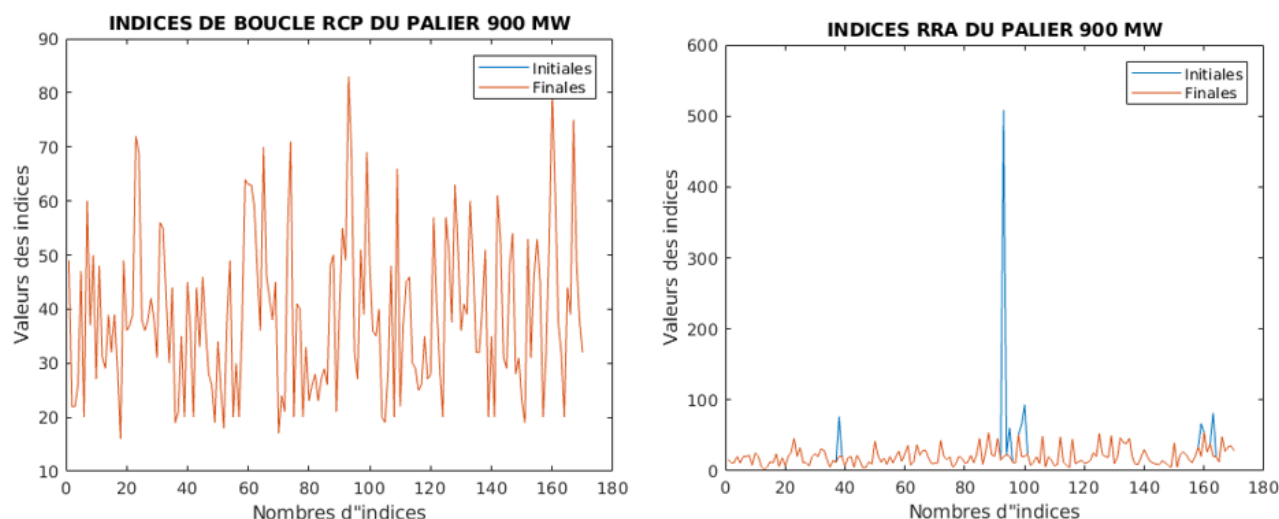


Figure 12: résultats des traitements des points aberrants sur les indices boucles-RCP et les indices RRA

Remarque : le traitement s'est fait aussi sur les autres circuits de tous différents paliers (cf. tableau 1, 2 et 3).

Après avoir comparé l'ensemble des données sur tout le palier 900 MW, une comparaison sur chaque réacteur est nécessaire afin de déterminer les circuits des réacteurs qui présentent un ou des point(s) aberrant(s) sur les sept dernières années (2011-2017).

- La figure 13 présente les données initiales et finales de l'indice de débits d'équivalents de dose du circuit RRA du réacteur Gravelines3 de 2011-2017. Ces données sont extraites du palier 900 MW de la figure ci-dessus (figure 12).

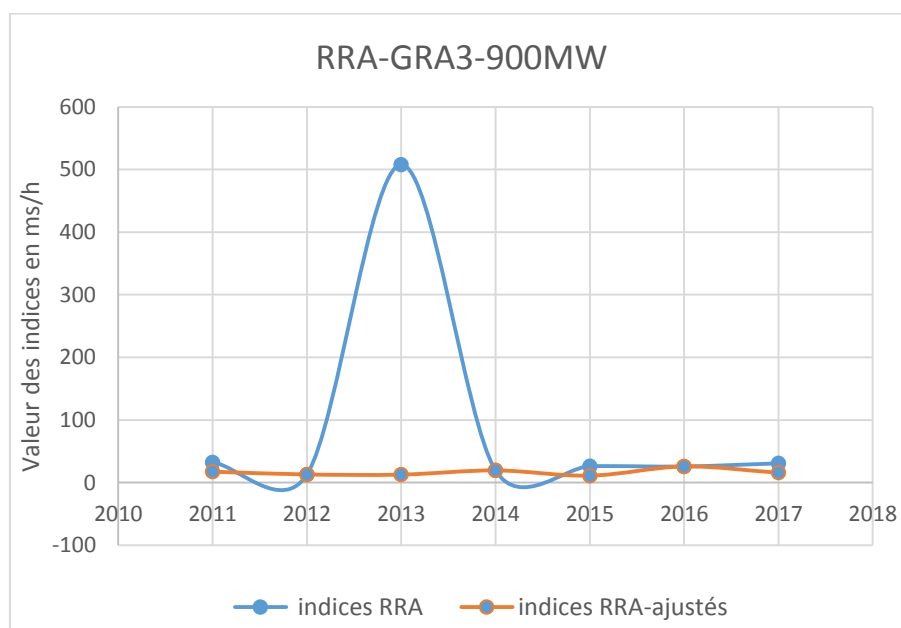


Figure 13: indices du circuit RRA du réacteur Gravelines 3 extraites de la figure 12 (figure droite)

Cette extraction sur chaque réacteur permet de faire une analyse approfondie sur les résultats obtenus après traitement puisqu'elle permet de zoomer sur l'évolution des indices du circuit RRA sur les réacteurs du palier 900 MW. Ce processus montre bien la présence d'un point aberrant en 2013 qui cassait la tendance des indices du circuit RRA de Gravelines3.

En résumé, les figures 12 et 13 justifient assez bien l'intervention de la fonction Hampel pour ce travail puisqu'il y avait des points aberrants dans la BDD initiales. Grâce à la fonction Hampel ces points aberrants ont été traités dans le but de nous fournir des données fiables. Sachant que la construire d'un modèle prédictif requière une base de données fiable avec moins de dispersion non justifié, nous pouvons maintenant envisager de construire un modèle prédictif. Les présentations ci-dessous présentent le travail complet de la fonction Hampel sur tous les paliers 900 MW et 1300 MW.

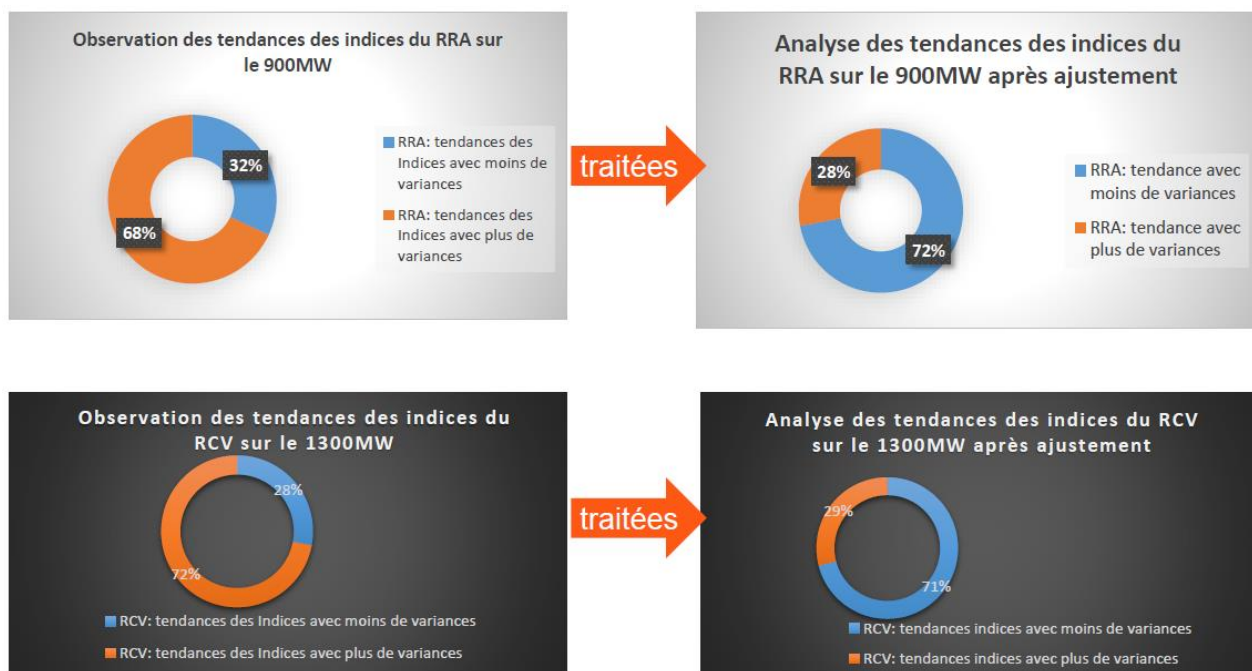


Figure 14: répartitions % des tendances des indices variantes et moins variantes après et avant traitements

5- CONSTRUCTION ET EVALUATION DES MODELES PREDICTIFS

Dans ce chapitre on étudiera l'influence des remplacements de générateur de vapeur (RGV) sur l'évolution des indices. D'abord, on construit deux modèles un avec les données des réacteurs ayant subi un RGV et l'autre avec les données de tout l'ensemble du palier 900 MW (RGV ou non).

5.1 Modélisation des scénarios

Cette partie du travail consiste à modéliser les différents scénarios sur lesquels sont bâtis les modèles prédictifs. Il y a principalement trois types scénarios:

- 1^{er} scénario: la construction se fait sur les types de points qui constituent un indice radiologique en fonction du circuit et du palier.

$$\text{Exemple : } \text{indice}^{RRA_{900}} = \frac{P_4 + P_{11} + P_{12} + P_{13}}{4}$$

Dans cet exemple, les indices RRA des réacteurs du 900 MW sont formés à partir de la moyenne arithmétique de 4 types de point (P_4 , P_{11} , P_{12} , P_{13}). Par la suite le scénario est constitué de ces 4 types de points (voir tableau 4) de sorte que le modèle utilise les trois premières colonnes (P_4 , P_{11} , P_{12} ,) pour prédire la 4^{ème} colonne (P_{13}).

DOCUMENT CONFIDENTIEL

PTn°4	PTn°11	PTn°12	PTn°13
28	3	7	26
16	9	13	5
15	11	14	9
28	20	20	10
11	8	18	6
14	18	20	36
15	3	6	9
46	20	30	4
20	25	15	22
10	8	1	5
2,7	1,7	3,2	2
12	3	3	3,5
15	15	15	6

Tableau 7 : Exemple de données des types de points qui de l'indice RRA décrivant le scénario 1

- 2^{ème} scénario : ce scénario prend tous les indices de chaque circuit sur l'ensemble des réacteurs du palier 900 MW.

○ **Palier 900 MW: RRA, RCV, RCP, RPE, PTR, PZR, GV, RIS, BOUCLE**

Ce scénario est construit directement à partir des indices de différent circuit (cf. Tableau 1). Le modèle réorganise toutes les données de 2011 à 2017 des indices de circuits et le range dans un autre tableau en 3 colonnes (voir tableau 8). La première colonne correspond aux données des indices de 2011 à 2015, la 2^{ème} de 2012 à 2016 et la 3^{ème} de 2013 à 2017. Le but de ce modèle est de prédire les indices des années ultérieures. C'est-à-dire, ayant les données de 2011 et 2012 on prédire celles de 2013, ensuite ayant ceux 2012 et 2013 on prédire 2014 ainsi de suite jusqu'au 2016 et 2017 on prédire 2018.

ANNEE1	ANNEE2	ANNEE3
16	5	10
11	15	11
12	12	11
20	16,375	43
11	16,375	20
15	7	16
15	18	18,375
8	5	10
25	22	18,375
2	5	12
11	42	23
6	12	9
7	9	18,375

Tableau 8 : Exemple de jeu de données de l'indice RRA décrivant le scénario 2

- 3^{ème} scénario : cette partie tient compte de l'apparition d'un paramètre qui est le remplacement du générateur de vapeur (RGV). Selon l'expertise radiologique, un réacteur ayant subi un RGV pourrait avoir de l'impact sur l'évolution des indices à travers les radioéléments dans le réacteur.

Pour réponse à cette incertitude, un scénario est mis en place pour comprendre l'impact des RGV sur l'évolution des indices d'un circuit du réacteur. Ce scénario reprend les mêmes principes du scénario 2 à la différence qu'il contient uniquement les réacteurs ayant subi un RGV (20 réacteurs sur le palier 900 MW).

BLAYAIS	BUGEY	CHINON	CRUAS	DAMPIERE	FESSENAIM	GRAVELINE	TRICASTIN	SAINT LAURENT
1, 2, 3, 4	2, 3, 4	1, 2	1, 4	2, 4	1,2	4,3	3, 4	2

Tableau 9: liste des réacteurs ayant subi un RGV

Tous ces différents scénarios ont permis de construire différents modèles prédictifs en utilisant des méthodes statistiques et numériques.

5.2 Machine Learning et Matlab

Pour construire des modèles prédictifs, des outils informatiques sont utilisés afin d'éditer manuellement un programme contenant les équations et les lois statistiques. Cette manière de procéder est appelée ancienne approche. La nouvelle approche qui est la méthode de machine learning⁷, elle est grandement utilisée dans le domaine des statistiques. Cette méthode est fortement mise en évidence sur la construction des modèles prédictifs. Aussi appelée apprentissage machine ou apprentissage statistique, c'est un champ d'étude de l'intelligence artificielle puisqu'elle utilise des données afin de produire un programme pour performer une tâche. Le machine Learning s'applique sur des logiciels ou langages tels que : R, Matlab, python ou java. Cependant, tous les différents travaux suivants (construction de modèle) ont été conçus par le biais de Matlab qui est un logiciel renfermant les packages adaptés au machine Learning.

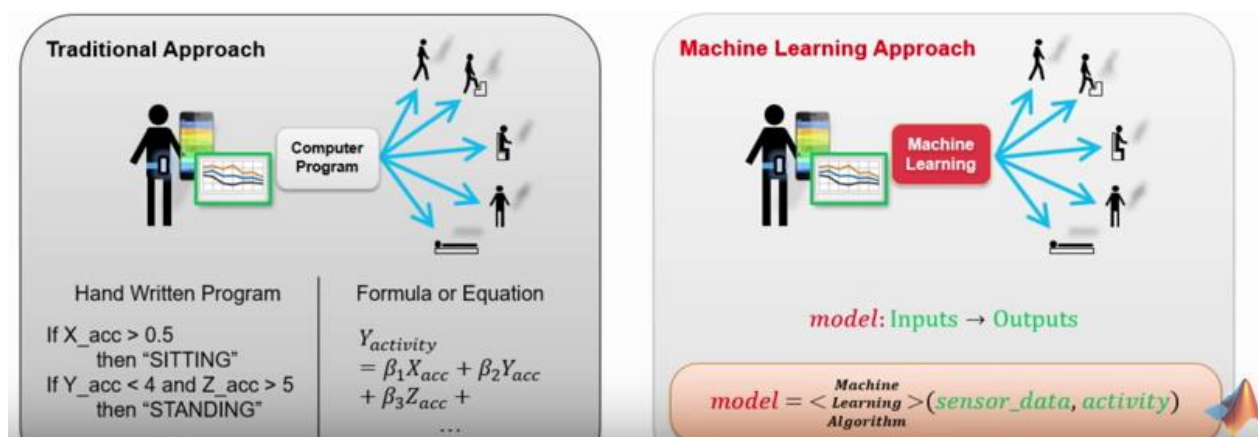


Figure 14: présentation de l'ancienne approche (manuelle) et de la nouvelle approche (machine learning)

5.2.1 Les types de méthode et modèles de machine Learning

Machine Learning⁷ contient deux types d'apprentissages⁷ qui présentent des algorithmes contenant plusieurs modèles pour faire de la prédiction. Ces apprentissages sont :

- L'apprentissage supervisé (supervised Learning):

L'apprentissage supervisé⁷ est une tâche d'apprentissage automatique consistant à faire tourner un algorithme sur des données déjà calé afin que celui-ci s'adapte à prédire des données ultérieure.

Plus formellement, étant donné un ensemble de données D , décrit par un ensemble de caractéristique x , un algorithme d'apprentissage supervisé va faire trouver une fonction liant les variables prédictives en entrée x et la variable à prédire y . Cette fonction décrivant la relation entre x et y s'appelle un modèle de prédiction. Les caractéristiques x peuvent être des valeurs numériques, alphanumériques, des images. Quant à la variable prédite y , elle peut être de deux catégories :

Variable discrète : La variable à prédire peut prendre une valeur d'un ensemble fini de valeurs.

Variable continue : La variable y peut prendre n'importe quelle valeur.

On distingue deux catégories qui peuvent est être attribué à la variable à prédire y qui sont les problèmes de régression et de classification. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de régression tandis que les problèmes de prédiction d'une variable qualitative sont des problèmes de classification. Ces deux algorithmes (régressions et classifications) contiennent plusieurs modèles prédictifs (voir figure 15).

- L'apprentissage non supervisé (unsupervised Learning):

A l'inverse de l'apprentissage supervisé qui tente de trouver un modèle depuis des données rangées $f(x) \rightarrow y$, l'apprentissage non supervisé⁷ prend uniquement des données non classées. Un algorithme de L'apprentissage non supervisé va trouver la structure des données, puis le classer à travers les modèles dont il renferme (voir figure 15).

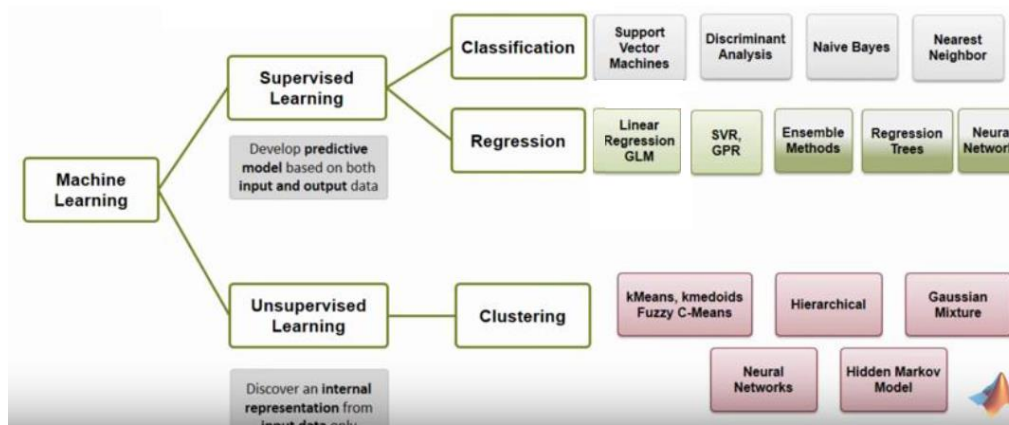


Figure 15: les différents algorithmes et modèles de machine learning

Les données sur lesquelles le travail est effectué sont des données quantitatives et déjà structurées, par conséquent, dans notre contexte l'apprentissage supervisé sera privilégié et l'algorithme utilisé est celui de la régression.

Les modèles de régressions sont sous la forme suivante :

$$y = f(x) + \varepsilon, \text{ avec } \varepsilon = N(0, \sigma^2)$$

L'algorithme de régression contient plusieurs modèles prédictifs comme le montre la figure 15.

- Modèle Régression Linéaire GLM : ce modèle possède trois propriétés :
 - La structure d'erreur est connue à priori et appartient à une famille connue (e.g., gaussien, poisson, binomial)
 - La droite de l'équation du modèle est constitué d'un prédicteur linéaire ($\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$)
 - La gauche de l'équation du modèle est constitué d'une fonction g de lien (e.g., identity, log, ou logit) telle que $g(y) = \eta$
- Modèle SVR ⁸ (support vector networks): le modèles de SVR appelé aussi machine à vecteur de support contient deux sous modèles ϵ -SVR and ν -SVR dont le but est de trouver une estimation de la vraie fonction : $f(x) = \langle w^T, x \rangle + b$ avec $w \in R^N, n \in R$ pour cela il traite des problèmes de minimisations.
- Modèle GPR ⁸ (Gaussian Processes Regression) : Une alternative puissante pour l'approximation précise des fonctions dans un espace de grande dimension est le processus gaussien régression (GPR). Plutôt que de modéliser une distribution sur les points de données le GPR modélise une distribution sur des fonctions. Étant donné un ensemble n de points de données de formation $y_i = f(x_i) + \varepsilon_i$, avec $\varepsilon_i = N(0, \sigma^2)$. En conséquence, les cibles observées peuvent également être décrit par une distribution gaussienne.

SVR et GPR sont tous deux des méthodes basées sur la mémoire en stockant une partie ou la totalité des données d'entraînement pour les tests. SVR offre une excellente capacité de généralisation et de robustesse aux valeurs aberrantes. En revanche, le GPR aborde la régression non linéaire dans une perspective bayésienne, ce qui produit à la fois la fonction estimée optimale et la covariance associée à l'estimation.

- Modèle Réseau de Neurone ⁹ : c'est une unité de traitement qui reçoit des données en entrée, sous la forme d'un vecteur, et produit une sortie réelle. Cette sortie est une fonction des entrées et des poids des connexions.

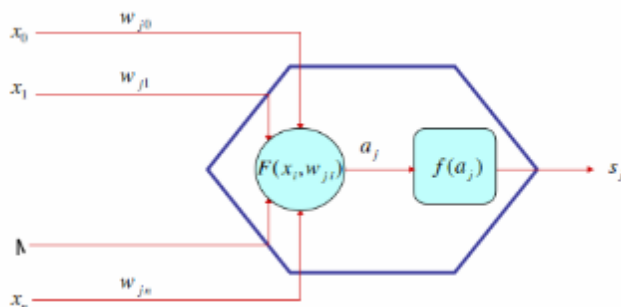


Figure 16: modèle de réseau de neurone

Le réseau est utilisé pour traiter des données

- Les poids et l'architecture sont fixés
- On lui présente un exemple x en entrée et le réseau calcule une sortie $y = F(x, w)$ qui dépend de l'architecture A et des poids w .

5.2.2 Utilisation de machine Learning avec Matlab

Les méthodes de machine Learning s'appliquent parfaitement sur le logiciel Matlab.

Comment utiliser les méthodes de machine Learning sur Matlab ?

Rappelons que le but est de construire un modèle prédictif en utilisant du machine Learning à travers le logiciel Matlab. Avant de commencer la construction du modèle assurez-vous que les packages de la data analysis sont bien installés dans le logiciel Matlab. L'outil principal qui permet d'accéder au package est ML_toolbox qui contient un ensemble de méthodes Matlab et des exemples de machine Learning. Après que l'outil soit installé afin de construire un modèle prédictif, appliquez les procédures suivantes:

- Analyser les données brutes (voir chapitre 3): cette partie consiste à comprendre les imperfections des données brutes.
- Traiter les données brutes (voir chapitre 4): après avoir énuméré des anomalies dans les données brutes, le traitement de ces données s'impose dans le but de bâtir un modèle prédictif sur des données fiables.
- Modéliser les jeux de données: c'est-à-dire ranger de manière claire les données qui permettront de construire et de valider le modèle.
- Construire le modèle : utilisation des jeux de données en comparant tous les modèles prédictifs et choisir celui qui prédire des données proches à celles des données brutes.
- Valider et extrapoler le modèle : utiliser des jeux de données qui n'ont pas participé à la construction du modèle afin de le valider, ensuite faire tourner le modèle pour anticiper l'évolution les données sur les périodes à venir.

Après avoir analysé et traité les données initiales, une méthodologie pour pourvoir construire un modèle prédictif est présentée dans les instructions suivantes :

Cliquer sur l'onglet **Apps** et il affichera une liste d'architecture qui permet de construire un modèle prédictif. Dans notre conteste le choix se fait en sélectionnant **regression learner** pour faire appel au modèle utilisant les algorithmes de la régression.

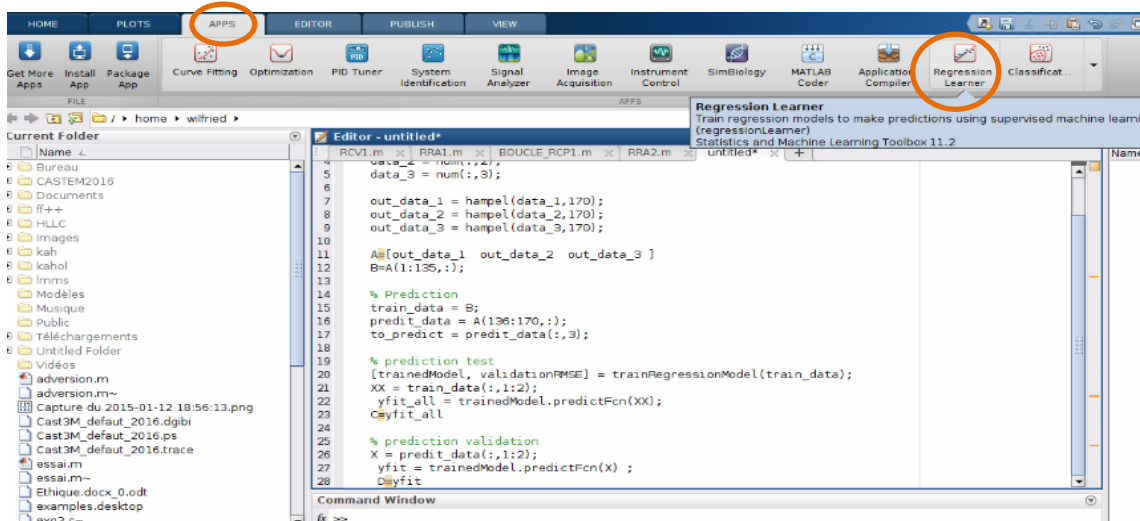


Figure 17: instruction pour choisir l'algorithme sur lesquelles les données seront lancées

Après avoir sélectionné **regression learner** une nouvelle fenêtre s'affiche comme suite (figure 18). Cette fenêtre permet de charger les données sur lesquelles l'on souhaite construire le modèle. Pour y parvenir vous devez cliquer sur l'onglet **new session** ensuite sur **from workspace**. Par la suite on charge les données fiables à travers **workspace variable** et enfin on termine ces tâches avec **start session**

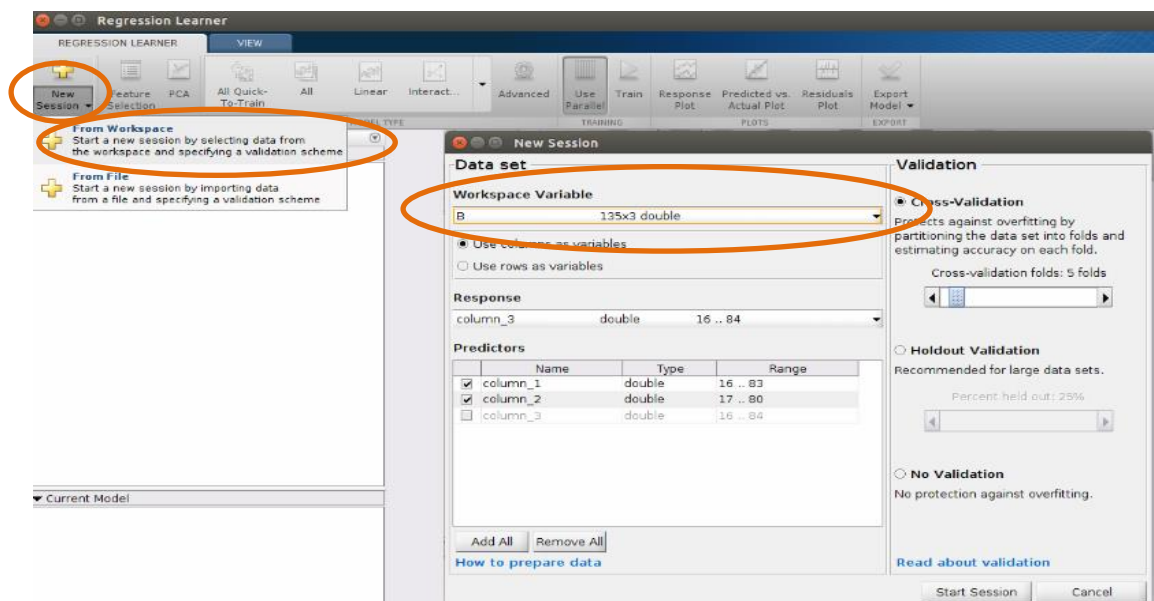


Figure 18: instruction pour charger les données fiables à Matlab

Cette partie consiste à lancer tous les modèles intégrés dans l'algorithme de régression en cliquant sur **Train**. Ensuite on charge et exporte le modèle choisi par l'algorithme à travers l'onglet **expert Model**.

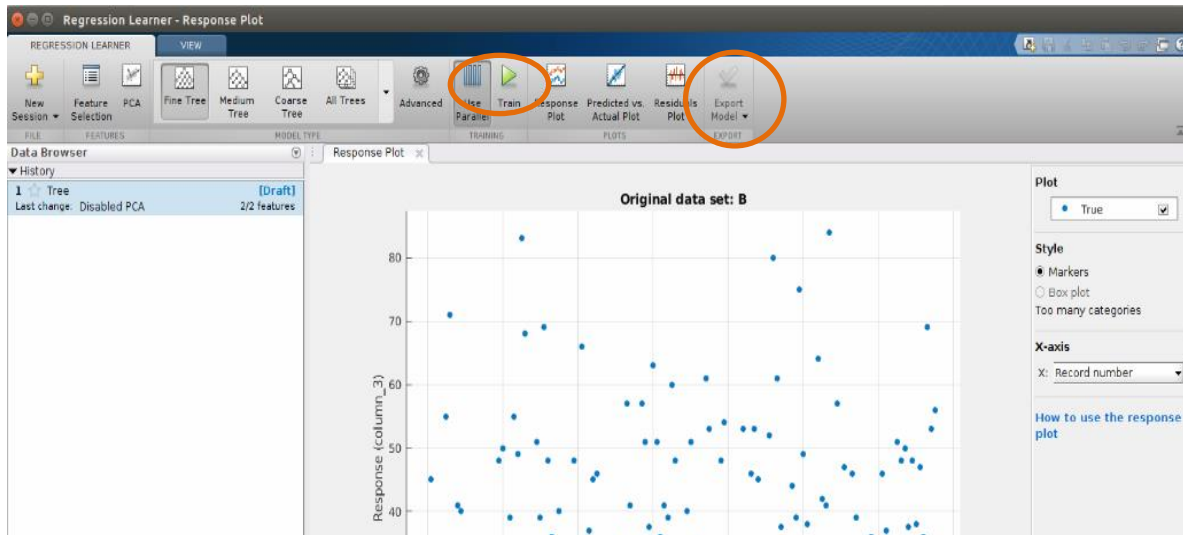


Figure 19: instruction pour lancer tous les modèles et récupérer le meilleur

La figure 20 présente une partie du programme qui a servie à construire le modèle. Sur ce programme on a deux parties :

- La 1^{ère} partie représente les lignes de codes qui ont permis de traiter les données initiales (voir chapitre 4).
- La deuxième partie présente les lignes de codes qui ont permis de construire les données de prédictions, puis valider le modèle sur des nouvelles données.

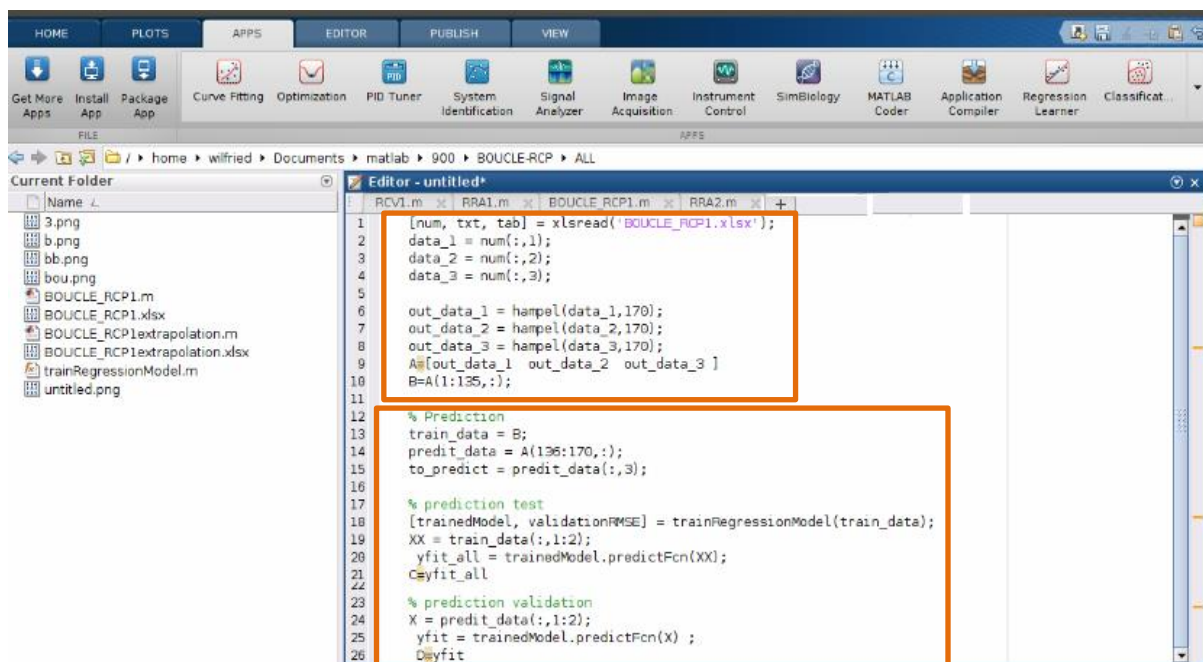


Figure 20: lignes de codes pour faire la prédiction en utilisant le modèle choisi

5.3 Construction et validation des modèles

Cette fin de travail présente les modèles construites sur les différents circuits des réacteurs du palier 900 MW.

- Palier 900 MW: RRA, RCV, RCP, RPE, PTR, PZR, GV, RIS, BOUCLE

5.3.1 Choix des modèles sur chaque scenario

Pendant tout le travail effectué, seul l'algorithme de régression contenant tous ces modèles a été utilisé.



Lorsque des données sont sectionnées à travers l'algorithme de régression il fait un calcul sur chaque modèle afin de déterminer le meilleur d'entre eux. L'algorithme utilise une mesure pendant cette pratique pour comparer les performances de chaque régresseur, cette mesure est l'erreur quadratique normalisée(NMSE). C'est à dire, si on a un vecteur \hat{Y} constituant n prédictions et un vecteur Y des valeurs observées correspondant à ces prédictions, un calcul concernant l'erreur quadratique moyenne (MSE) du prédicteur se fait pour déterminer le mieux regresseur :

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad \text{et} \quad NMSE = \frac{MSE}{VAR(Y)} = \frac{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu)^2} \quad \text{avec} \quad \mu = \frac{1}{n} \sum_{i=1}^n Y_i$$

Une fois le MSE et le NMSE sont calculés, l'algorithme faire son choix en présentant de façon visuel le régresseur qui a le plus petit RMSE (Root Mean Square Error). RMSE mesure l'écart des résidus. En d'autres termes, il vous indique la concentration des données autour du meilleur ajustement.

$$RMSE = \sqrt{MSE} \quad \text{La racine donne un poids relatif plus élevé aux erreurs importantes.}$$

Rappelons que le but est de construire un modèle prédictif à travers les données dont nous disposons. Ces données sont les indices de débits d'équivalent de dose reparti en plusieurs scénarios. D'après la partie 5.1 de ce rapport il y a trois grands scénarios :

- 1^{er} scénario sur les points de de débits d'équivalent de dose qui constitue un indice
- 2^{ème} scénario sur l'ensemble des indices
- 3^{ème} scénario sur l'ensemble des indices faisant intervenir le paramètre RGV

Chaque circuit (RRA, RCV, RCP, RPE, PTR, PZR, GV, RIS, BOUCLE) dispose d'un modèle qui a été bâti sur ces différents scénarios.

Pour le 1^{er} scénario un modèle prédictif est bâti, exemple sur les indices du circuit RRA 900 MW : l'indice RRA est la moyenne de 4 types de points $I_{RRA} = \frac{P_4 + P_{11} + P_{12} + P_{13}}{4}$. A travers ce scénario l'algorithme de régression a choisi comme meilleur modèle prédictif **la régression linéaire**.

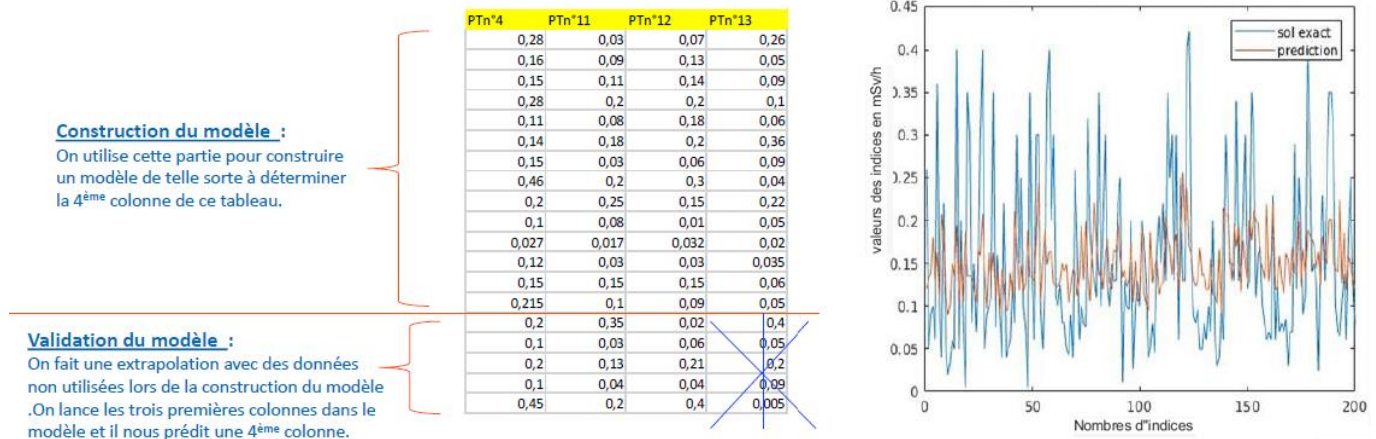


Figure 21: exemple de modèle de prédiction construit sur les indices du circuit RRA 900 MW

Pour le 2^{ème} scénario un modèle prédictif est bâti, exemple sur les indices du circuit BOUCLE-RCP : le modèle se fait directement sur les indices c'est à dire en connaissant l'indice de 2011 et 2012 d'un recteur on prédire celui de 2013, ensuite si on a celui de 2012 et 2013 on prédire 2014 ainsi de suite. Quand on lance les données de BOUCLES-RCP on obtient le modèle prédictif **Stepwise Linear Regression** tel que la figure ci-dessous la montre. En jaune sont les données prédites et en bleu les données observées.

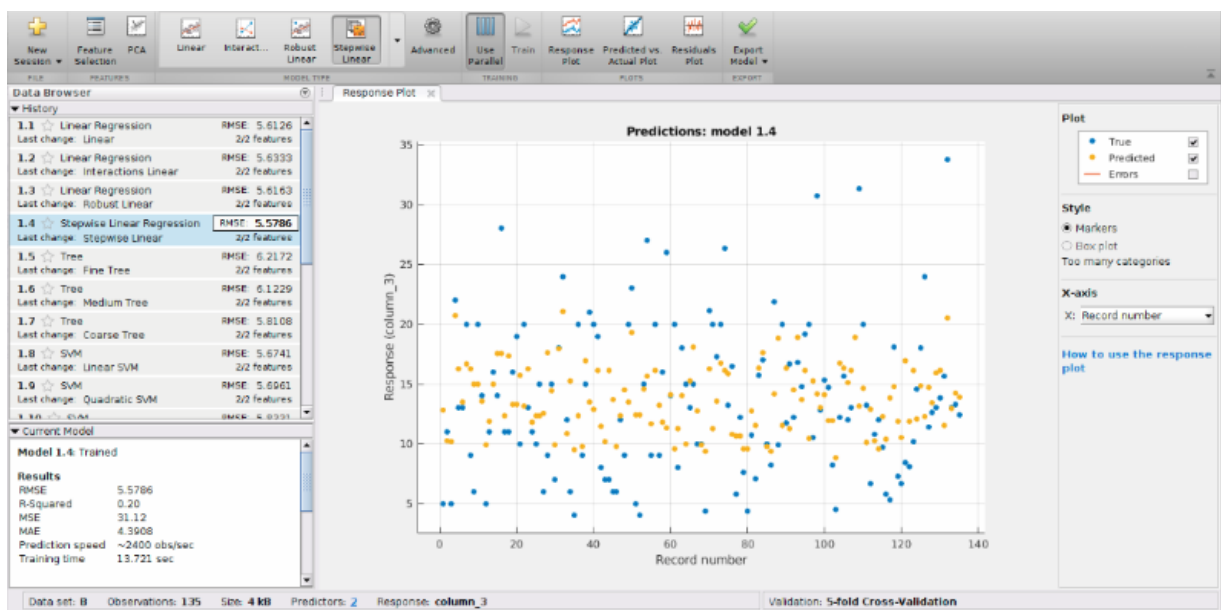
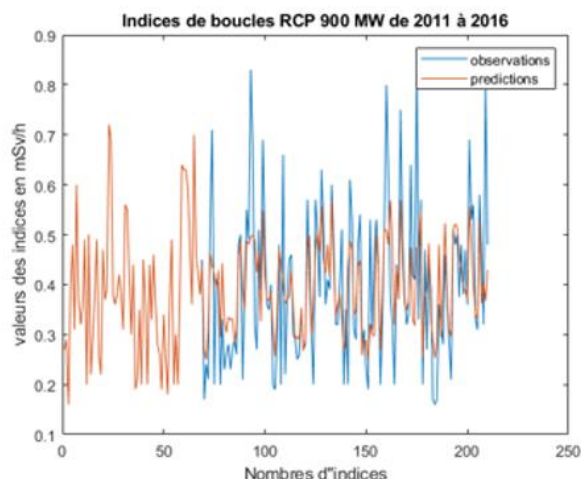


Figure 22: exemple de modèle de prédiction construit sur les indices du circuit boucle-RCP 900 MW

Le modèle prédictif **Stepwise Linear Regression** a été construit en utilisant les indices de 2011-2016 de la boucle RCP 900 MW et il présente les données prédites et les données initiales (observées) comme suite.



Sur le 3^{ème} scénario un modèle prédictif est bâti sur les indices du circuit BOUCLE-RCP des recteurs ayant subi un RGV : ce scénario est identique au 2^{ème} scénario à la différence qu'il utilise uniquement les recteurs avec RGV. En lançant les données on obtient le modèle prédictif **de méthode d'ensemble (bagged tree)**¹⁰. Les traits représentent l'écart entre les données prédites et les données observées en bleu.

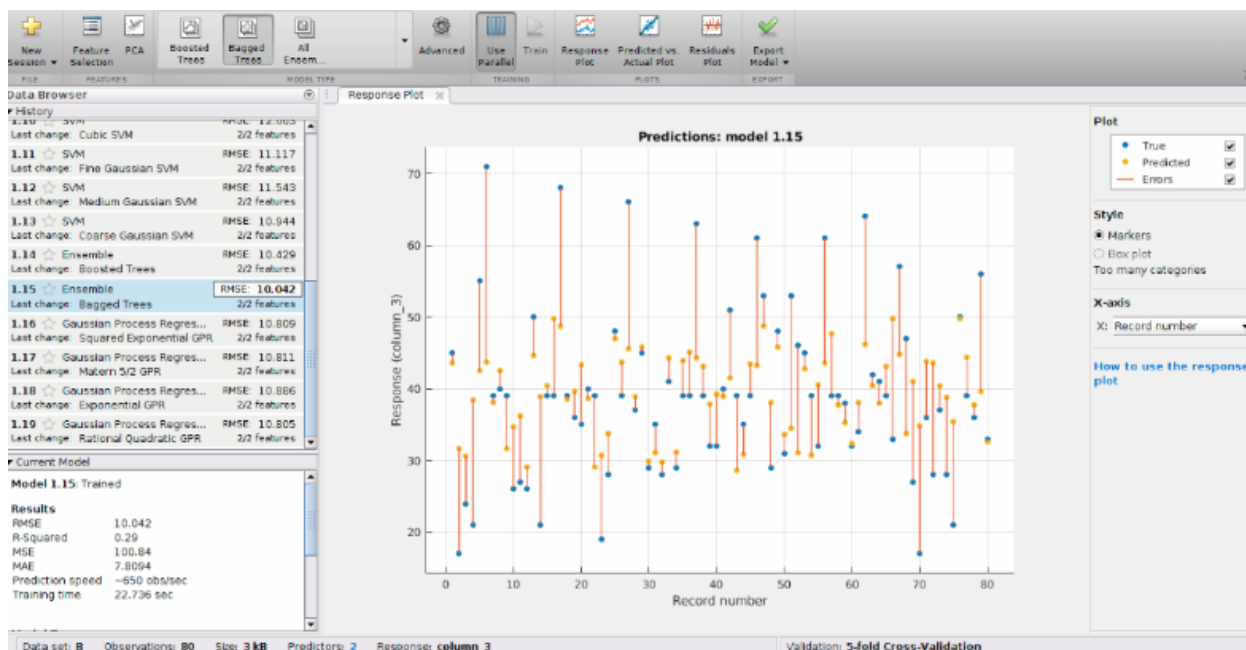
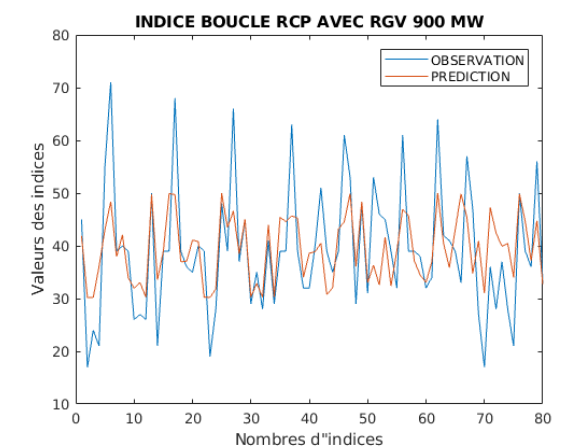


Figure 23: modèle de prédiction sur les indices du circuit boucle-RCP 900 MW sur les réacteurs avec RVG

La modèle prédictive **méthode d'ensemble (bagged tree)** a été construit en utilisant les indices de 2011-2016 de la boucle RCP 900 MW et il présente les données prédites et les données initiales (observées) comme suite.

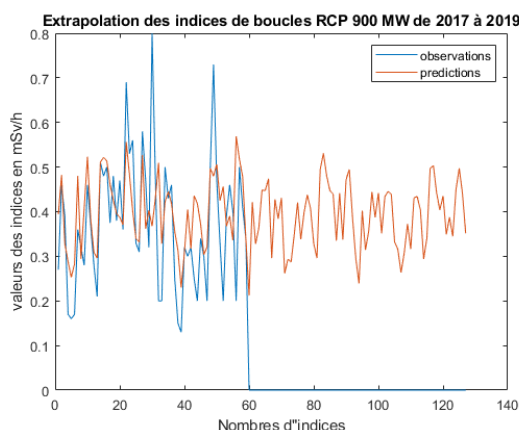


Pour conclure sur ces différents scénarios nous pouvons donc dire, lorsqu'on construit un modèle prédictif, il est illusoire d'espérer atteindre une meilleure précision puisque il y a une hypothèse que ces données ont au préalable certaines incertitudes par le fait qu'elles sont mesurées par des personnes physiques. Toutefois elles suivent les tendances des indices. De plus, la plus part de nos modèles ont été validés avec des données de 2017 et 2018.

5.3.2 Validation et extrapolation des modèles

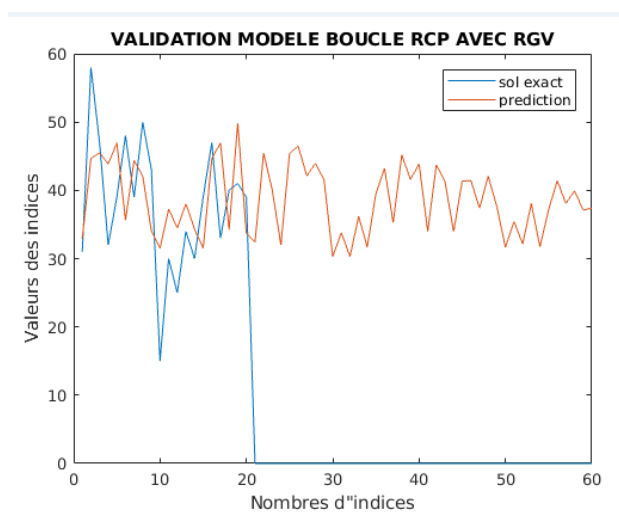
Le modèle obtenu sur le scénario 1 nous permettra de pouvoir calculer la valeur d'un indice pour les années ultérieures même si celui-ci manque un dernier type de points. L'objectif étant d'obtenir un modèle prédictif sur les indices sans avoir nécessairement à passer par ces types de points qui le compose, on souhaite aussi vérifier si la présence d'un paramètre tel que le RGV a une influence significative sur l'évolution des indices de débits d'équivalent de doses. Pour ce faire nous allons valider et extrapoler les modèles obtenus sur les scénarios 2 et 3 afin de conclure sur l'apport du RGV sur l'évolution des indices.

❖ Validation et extrapolation du modèle sur le 2^{ème} scénario sur le circuit BOUCLE-RCP



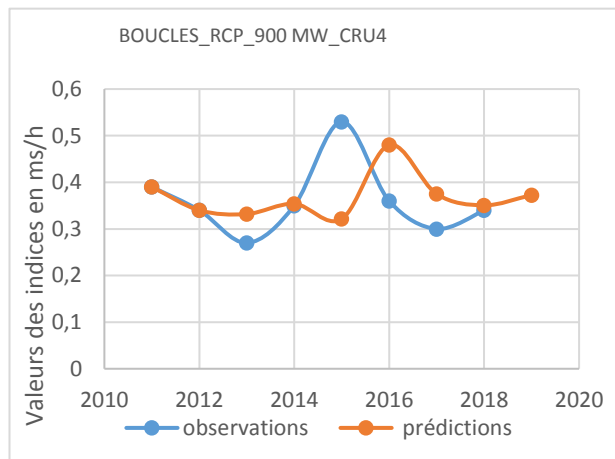
La figure précédente présente la validation du modèle de 0-59 qui est le nombre d'indice de 2017 (34 réacteurs) incluant une partie des indices de 2016 (25 réacteurs). Ensuite, de 60-127 est le nombre d'indice représentant l'extrapolation du modèle, donc une prédiction pour l'ensemble des 34 réacteurs sur 2018 et 2019.

❖ Validation et extrapolation du modèle sur le 3^{ème} scénario sur le circuit BOUCLE-RCP avec RGV

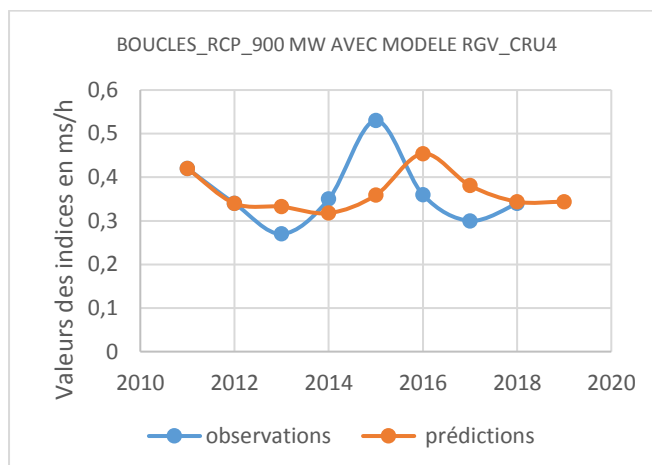


La figure ci-dessus, présente la validation du modèle de 0-20 (nombres d'indices) de 2017 (20 réacteurs). Ensuite, de 21-60 (nombres d'indices) qui représente l'extrapolation du modèle, donc une prédiction pour l'ensemble des 20 réacteurs ayant eu un RGV sur le palier 900 MW sur 2018 et 2019.

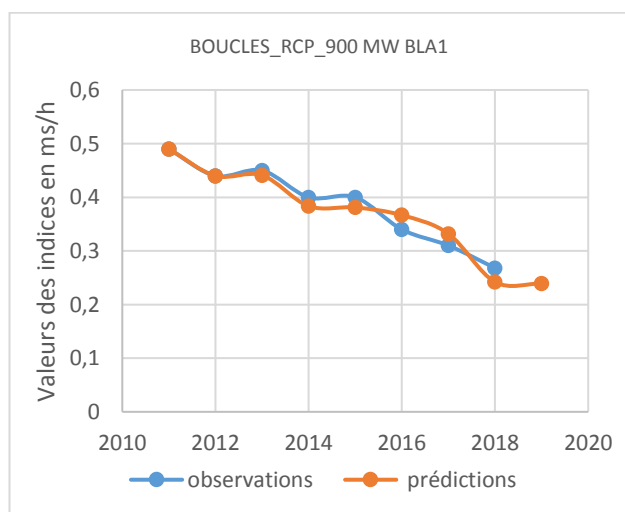
En résumé ces deux modèles présentent certains écarts entre les données prédites et les données observées. Par ailleurs, ils suivent les tendances des indices observés, ce qui est d'ailleurs notre premier objectif recherché à travers ces modèles puisqu'on souhaite avoir une idée sur de l'évolution des indices et non obligatoirement les valeurs exactes. Donc on conclure que ces deux modèles sont acceptables. Cependant, pour certifier le scénario à choisir et pour répondre au mieux concernant l'influence du RGV sur l'évolution des indices, nous zoomons sur les données en répartissant les données des différents circuits sur chaque réacteur afin d'interpréter les tendances qu'elles présentent de 2011-2019 (voir les figures ci-dessous).



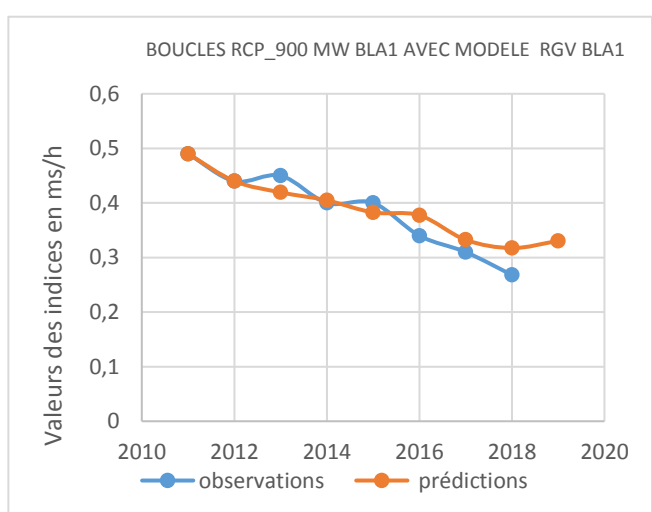
L'écart entre la donnée prédite de 2018 et la donnée en 2018 est négligeable



L'écart entre la donnée prédite de 2018 et la donnée en 2018 négligeable



Tendance très satisfaisante puisque sur la prédiction en 2018 on a une prédiction décroissance comme celles des données observées avec 8% d'écart



Tendance très satisfaisante puisque sur la prédiction en 2018 on a une prédiction décroissance comme celles des données observées avec 18% d'écart

Toutes les figures à gauche: Ces données sont extraites du modèle dont la construction est basée sur l'ensemble des réacteurs du palier 900 MW.

Toutes les figures à droite: Ces données sont extraites du modèle dont la construction est basée uniquement sur les 20 réacteurs du palier 900 MW avec RGV.

Après avoir zoomé les tendances des indices sur quelques réacteurs et tester les en utilisant les données de 2018, on constate que ces modèles sont de très bon modèles prédictifs. En outre, le remplacement de générateur de valeur (RGV) a un impact sur l'évolution des indices sur palier 900 MW puisque le résultat est conforme à l'attendu. Par conséquent, le scénario à retenir pour continuer la validation du modèles sur les années à venir est celui qui est construire sur l'ensemble des réacteurs du palier 900 MW.

6- CONCLUSIONS ET PERSPECTIVES

Ce travail présent des résultats des travaux effectués sur l'ensemble des données de 2011-2017 des indicateurs radiologiques (indices) des 58 réacteurs dont dispose EDF. Ces réacteurs sont repartie en trois grands groupes : REP 900 MW, REP 1300 MW, REP 1450MW. L'objectif de ce travail étant de mener une étude multi-paramètres sur les débits d'équivalent de dose afin de consolider l'analyse des indices et d'appréhender leurs évolutions, ce travail présent donc l'efficacité des outils informatiques tels que le logiciel R et le logiciel Matlab pour la construction des modèles prédictifs.

Par le biais des outils informatiques, l'objectif a été scindé en deux parties. La première consistait à étudier le comportement de chaque point de débits d'équivalent de dose mesuré pour fiabiliser l'analyse des indicateurs radiologiques des circuits et la seconde était de bâtir un modèle prédictif afin d'anticiper l'évolution des indices radiologiques des circuits pour les années ultérieures.

D'abord, pour mener à bien cette étude nous avons commencé par analyser les données provenant des centrales. Les résultats suites suite de cette analyse nous on permit d'identifier les circuits de certains réacteurs qui avaient des tendances des indices très variables avec des indices élevés tel que BUG2 ,3 ,4 ,5 , GRA1,2,3,4,5,6, TRI1,2,3,4, FES1,2... (cf. Annexe).

Ensuite, un traitement de ces indices a été mis en place afin de les fiabiliser puisqu'ils seront utilisés pour bâtir un modèle prédictif. Le traitement a conduit à une nouvelle base de données sur laquelle la méthode de machine Learning modélisée à travers le logiciel Matlab afin de construire les différents modèles prédictifs. Il avait des modèles prédictifs qui était bâti sur l'ensemble des recteurs du palier 900 MW et les autres uniquement sur les réacteurs ayant subi un remplacement de générateur de vapeur (RGV) sur les données de 2011-2016. Les modèles construites sur les RGV avaient pour but de montrer si un remplacement de générateur de vapeur sur un réacteur influencera l'évolution d'indicateurs radiologiques.

Enfin, ces deux scénarios ont été validés sur les données de 2017 et 2018. A la suite de cette validation, une conclusion qui ressort est que le remplacement de générateur de valeur (RGV) a un impact sur l'évolution des indices ce qui est conforme avec le modèle construit sur l'ensemble des réacteurs du palier 900 MW.

La motivation pour laquelle ce travail a été réalisé a en effet pour objectif de proposer une approche statistique aux applications de la radioprotection afin de consolider l'analyse des indicateurs radiologiques et de connaître l'évolution ces indicateurs. Donc d'après le travail effectué les modèles prédictifs construits pendant le stage vont permettre aux experts d'anticiper l'évolution des indices radiologiques des tranches pour les années à venir.

7- REFERENCE

- (1) Expert radioprotection EDF mémento de la radioprotection en exploitation. 2014, 253-261.
- (2) P. REMY ; P.A.MUNOZ Argumentaire pour la mise en place de l'indice BR sur l'ensemble des CNPE, 2010, 4-24.
- (3) Les centrales nucléaires d'EDF, 2010, chapitre 12.
- (4) F. RENARD Présentation UNIE, Qui Fait Quoi UNIE?, 2017, 50-89
- (5) Monique Le Guen. La boîte à moustaches de TUKEY, un outil pour initier à la Statistique. Statistiquement Votre - SFDS, 2001, pp.1-3.
- (6) R. CLEROUX, J.-M. HELBLING, N. RANGER, Détection d'ensembles de données aberrantes en analyse des données multi variées, Revue de statistique appliquée, tome 38, n°1 (1990), p. 5-21.
- (7) Fabien Moutarde, Centre de Robotique (CAOR), MINES ParisTech (Ecole des Mines de Paris).
- (8) Aude Billard, Guillaume de Chambrier, Advanced Machine Learning, Practical 4: Regression (SVR, RVR, GPR), 2016, 2-20.
- (9) Younes BENNAN, Apprentissage par réseaux de neurones artificiels, EPAT'14 : École de Printemps sur l'Apprentissage artificiel 2014 7-12 juin 2014, 10-37.
- (10) Kévin Bailly, Méthodes d'ensembles, Boosting, bagging et Random forests

8- ANNEXE

REACTEUR	CAMP AGNE D'ARR ET année	INDICE BOUCLE RCP FIN CYCLE 1E-2mSv/h	INDICE BR 1E- 2mSv/h	INDICE BRA 1E- 2mSv/h	INDICE RCV 1E- 2mSv/h	INDICE RPE 1E- 2mSv/h	INDICE RIS 1E- 2mSv/h	INDICE PZR 1E- 2mSv/h	INDICE GV 1E- 2mSv/h	INDICE RCP 1E- 2mSv/h	INDICE PTR 1E- 2mSv/h
BLA1	2011	49	19	16	20	14	2	7	22	28	3
BLA2	2011	22	15	11	10	10	25	7	17	17	30
BLA3	2011	22	15	12	15	16	5	6	20	17	9
BLA4	2011	26	30	20	25	77	9	25	29	26	47
BUG2	2011	47	32	11	14	13	15	33	24	68	30
BUG3	2011										
BUG4	2011	60	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait
BUG5	2011	37	20	22	23	33	3	5	16	28	27
CHB1	2011	50	25	8	18	10	4	19	30	44	8
CHB2	2011	27	30	25	29	55	20	6	26	42	25
CHB3	2011	48	23	21	17	19	13	7	22	42	8
CHB4	2011	31	19	6	16	4	4	4	30	31	7
CRU1	2011	29	17	2	13	6	2	12	33	15	30
CRU2	2011	39	15	5	12	14	22	10	17	17	30
CRU3	2011	32	17	13	13	14	16	14	20	20	15
CRU4	2011	39	42	11	19	57	152	20	46	38	20
DAM1	2011	29	27	24	17	32	4	12	20	46	100
DAM2	2011	16	12	6	8	7	3	14	21	12	4
DAM3	2011	49	45	19	32	11	9	33	26	102	100
DAM4	2011	36	22	7	13	12	12	15	22	43	3
FES1	2011	37	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait
FES2	2011	39	55	26	55	14	9	25	21	128	230
GRA1	2011	72	35	46	47	39	4	13	28	43	70
GRA2	2011	69	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait
GRA3	2011	38	43	33	19	50	293	23	17	28	80
GRA4	2011	36	23	12	22	15	18	16	23	29	80
GRA5	2011	38	26	12	32	16	4	9	28	44	13
GRA6	2011	42	25	7	42	14	8	36	29	19	20
SLB1	2011	38	24	20	42	18	6	15	17	33	14
SLB2	2011	31	30	24	24	71	19	17	30	27	60
TRI1	2011	56	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait	pas fait
TRI2	2011	55	37	31	35	47	24	34	42	37	40
TRI3	2011	42	19	29	74	0	0	0	7	42	0
TRI4	2011	30	18	15	16	24	20	6	20	23	1
BLA1	2012	44	20	5	25	6	1	12	24	35	3
BLA2	2012	19	13	15	10	14	12	6	12	18	30
BLA3	2012	21	15	12	17	21	5	6	18	15	7

DOCUMENT CONFIDENTIEL

BLA4	2012	35	46	76	28	78	23	31	57	36	60
BUG2	2012										
BUG3	2012	45	52	7	20	80	7	39	18	125	60
BUG4	2012	36	22	18	32	40	Pas fait	22	14	24	15
BUG5	2012										
CHB1	2012	44	19	5	20	9	3	17	27	26	9
CHB2	2012	33	31	22	25	68	26	17	26	39	32
CHB3	2012	46	20	14	17	26	8	14	20	30	3
CHB4	2012	36	14	4	25	8	5	6	18	15	6
CRU1	2012	28	18	5	15	10	3	10	34	21	6
CRU2	2012	26	29	13	22	15	26	17	43	37	36
CRU3	2012	19	17	9	13	19	16	24	14	23	23
CRU4	2012	34	41	42	23	48	143	23	41	32	15
DAM1	2012	25	45	22	47	17	3	11	29	108	8
DAM2	2012	18	18	12	15	14	7	23	15	24	60
DAM3	2012	87	26	18	45	15	6	13	19	44	2
DAM4	2012	49	77	9	37	33	7	182	52	148	6
FES1	2012										
FES1	2012	30	19	11	19	14	3	13	14	36	23
FES2	2012										
GRA1	2012										
GRA2	2012	94	31	28	32	19	8	11	34	47	50
GRA3	2012	64	27	13	14	41	54	11	29	32	40
GRA4	2012	63	24	24	18	39	22	11	25	27	35
GRA5	2012	63	56	36	53	95	18	12	79	41	250
GRA6	2012	59	34	8	24	41	22	23	46	36	87
SLB1	2012	47	26	13	21	52	3	16	16	30	200
SLB2	2012	36	28	37	36	59	27	6	22	25	30
TRI1	2012	70	29	22	36	31	9	25	25	40	3
TRI2	2012	46	26	29	28	28	4	32	19	35	12
TRI3	2012	42	34	29	15	24	11	9	72	29	30
TRI4	2012	38	23	17	16	29	12	18	26	27	25
BLA1	2013	45	19	10	20	8	2	5	26	31	4
BLA2	2013	17	14	11	10	19	23	11	15	16	20
BLA3	2013	24	15	11	17	18	7	5	22	14	7
BLA4	2013	21	34	43	16	75	55	22	29	32	60
BUG2	2013	55	35	20	29	25	13	48	33	50	31
BUG3	2013	71	38	16	50	31	33	44	20	62	30
BUG4	2013										
BUG5	2013	41	15	5	21	8	2	9	15	22	20
CHB1	2013	40	22	10	17	9	24	6	26	32	60
CHB2	2013										
CHB3	2013	33	41	18	16	28	36	14	32	94	40
CHB4	2013	23	15	11	10	9	31	5	26	13	6

DOCUMENT CONFIDENTIEL

CRU1	2013	26	18	12	14	20	10	11	26	17	45
CRU2	2013	28	28	22	25	11	25	16	42	28	30
CRU3	2013	23	22	11	15	16	30	14	28	23	70
CRU4	2013	27	37	23	22	39	110	28	38	37	1
DAM1	2013	29	21	46	25	11	4	11	16	26	8
DAM2	2013	26	23	9	21	9	7	11	20	49	5
DAM3	2013	48	31	27	37	16	7	16	25	58	6
DAM4	2013	50	43	54	40	38	140	19	26	47	13
FES1	2013	21	16	23	20	19	9	10	7	23	30
FES2	2013	39	pas mesuré	pas mesuré	pas mesuré	pas mesuré	pas mesuré	pas mesuré	pas mesuré	pas mesuré	pas mesuré
GRA1	2013	55	35	46	47	39	4	13	28	43	70
GRA2	2013	49	38	15	35	26	14	11	38	72	25
GRA3	2013	83	128	508	33	569	47	10	33	79	32
GRA4	2013	68	28	23	48	30	24	15	25	26	50
GRA5	2013	32	36	61	23	83	5	6	30	48	40
GRA6	2013	27	20	12	17	10	11	9	31	27	15
SLB1	2013	51	20	12	24	22	6	15	19	23	60
SLB2	2013	39	36	51	29	45	92	7	30	37	25
TRI1	2013	69	33	65	43	43	27	18	27	31	0
TRI2	2013	48	33	93	13	81	23	24	40	9	0
TRI3	2013	36	21	24	26	23	3	12	22	26	2
TRI4	2013	35	19	8	21	22	35	6	16	24	16
BLA1	2014	40	19	12	24	7	2	4	18	35	3
BLA2	2014										
BLA3	2014	19	15	10	16	8	11	9	12	24	30
BLA4	2014	28	38	49	22	92	97	15	26	31	40
BUG2	2014	48	27	5	22	18	9	21	21	55	17
BUG3	2014										
BUG4	2014	66	34	15	58	18	3	19	22	64	20
BUG5	2014	22	21	7	44	4	5	8	16	37	15
CHB1	2014	37	16	9	16	9	4	7	26	17	15
CHB2	2014	45	45	48	23	127	63	7	34	55	15
CHB3	2014	46	17	12	16	32	5	6	22	17	6
CHB4	2014	30	17	8	19	8	4	6	18	30	10
CRU1	2014	29	14	4	14	3	3	12	17	21	40
CRU2	2014	25	33	45	27	47	43	9	28	40	26
CRU3	2014	26	18	10	16	10	12	20	24	17	50
CRU4	2014	35	39	13	21	102	88	23	42	30	25
DAM1	2014	27	18	15	23	21	4	5	19	23	6
DAM2	2014	28	12	11	18	10	4	4	12	14	5
DAM3	2014	57	22	12	22	13	5	15	31	26	3
DAM4	2014	41	41	15	26	23	27	27	13	110	6
FES1	2014	29	24	26	24	30	22	9	15	36	pas mesuré

DOCUMENT CONFIDENTIEL

FES2	2014										
GRA1	2014	57	37	53	40	52	8	9	38	42	37
GRA2	2014	51	25	23	24	19	29	16	28	31	13
GRA3	2014	112	86	20	56	544	50	26	41	53	20
GRA4	2014	63	28	19	39	32	10	14	27	37	30
GRA5	2014	51	66	50	43	116	47	20	69	91	50
GRA6	2014	36	24	10	47	28	9	8	24	21	68
SLB1	2014	41	24	19	21	27	9	18	20	39	16
SLB2	2014	39	34	47	26	58	24	15	37	31	52
TRI1	2014	60	36	40	60	36	26	13	34	38	5
TRI2	2014	48	30	38	23	30	6	15	33	43	25
TRI3	2014	32	21	46	17	32	5	10	19	24	12
TRI4	2014	32	18	19	17	22	23	10	16	24	10
BLA1	2015	40	18	10	8	8	2	4	17	35	3
BLA2	2015	51	25	10	21	14	10	21	32	35	10
BLA3	2015										
BLA4	2015	35	32	30	28	55	56	17	23	37	60
BUG2	2015										
BUG3	2015	61	25	13	19	28	12	26	16	42	50
BUG4	2015	53	24	11	26	6	5	13	19	50	30
BUG5	2015	31	27	10	76	16	3	17	22	32	23
CHB1	2015	29	13	9	10	10	4	6	19	16	15
CHB2	2015	48	32	14	34	110	42	12	21	28	13
CHB3	2015	54	21	12	13	14	6	8	18	48	7
CHB4	2015	28	16	7	15	7	6	4	19	27	40
CRU1	2015	31	27	5	4	4	23	11	48	38	40
CRU2	2015	23	19	40	15	13	33	7	19	19	18
CRU3	2015	19	13	4	12	6	18	16	16	15	12
CRU4	2015	53	25	23	29	14	84	17	20	23	25
DAM1	2015	31	24	27	22	22	5	10	44	19	2
DAM2	2015	46	16	23	10	19	7	8	18	23	5
DAM3	2015	53	21	15	19	12	4	22	24	29	5
DAM4	2015	45	19	12	22	27	27	10	11	30	5
FES1	2015										
FES2	2015	32	49	34	97	19	5	12	19	93	pas mesuré
GRA1	2015	52	44	67	51	63	8	17	36	55	30
GRA2	2015	80	36	54	27	45	19	12	33	51	20
GRA3	2015	61	50	26	40	276	35	17	18	42	20
GRA4	2015	89	27	38	31	31	12	15	27	30	27
GRA5	2015	32	34	82	15	114	12	19	25	26	15
GRA6	2015										
SLB1	2015	44	22	12	23	17	5	11	27	28	60
SLB2	2015	39	35	49	42	53	73	31	23	23	35
TRI1	2015	75	43	27	28	28	11	13	54	58	4

DOCUMENT CONFIDENTIEL

TRI2	2015	49	27	34	28	12	5	15	28	37	50
TRI3	2015	38	27	35	30	39	22	15	31	24	5
TRI4	2015	32	21	29	18	30	31	8	16	28	12
BLA1	2016	34	21	16	25	10	3	4	25	33	2
BLA2	2016	64	28	13	26	18	14	12	39	41	20
BLA3	2016	42	26	11	33	15	8	16	29	36	9
BLA4	2016	41	35	36	23	62	53	12	23	51	25
BUG2	2016	84	25	11	19	19	33	35	21	25	18
BUG3	2016	33	29	16	16	28	11	19	15	61	40
BUG4	2016	57	65	33	90	23	3	31	20	157	20
BUG5	2016	Prolongation AT									
CHB1	2016	47	18	12	16	12	4	13	22	27	7
CHB2	2016	27	23	15	13	51	36	7	26	22	13
CHB3	2016	46	20	12	17	19	6	11	29	22	16
CHB4	2016	39	10	8	12	7	3	0	16	13	1
CRU1	2016	17	17	16	14	22	16	10	23	14	20
CRU2	2016	16	18	15	13	27	86	6	11	14	1
CRU3	2016	17	15	9	10	9	14	5	18	21	50
CRU4	2016	36	34	17	19	35	72	18	46	31	50
DAM1	2016	32	18	18	23	17	4	7	17	25	4
DAM2	2016	28	17	13	14	13	4	7	26	21	28
DAM3	2016	46	20	23	18	12	3	8	25	29	6
DAM4	2016	37	23	11	23	28	30	8	28	29	5
FES1	2016	28	16	19	18	16	6	10	8	29	0
FES2	2016	21	32	20	20	54	18	15	6	79	0
GRA1	2016	51	34	29	25	22	14	18	27	67	15
GRA2	2016	48	35	30	28	28	25	24	39	44	50
GRA3	2016	50	65	26	45	297	22	11	37	74	50
GRA4	2016	110	28	19	31	20	14	13	27	42	36
GRA5	2016	48	153	505	73	581	16	63	70	67	420
GRA6	2016	38	25	18	37	36	5	14	32	20	10
SLB1	2016	47	19	13	17	28	6	16	14	28	50
SLB2	2016	36	29	28	28	52	16	34	22	32	26
TRI1	2016	69	42	29	52	52	22	36	45	47	13
TRI2	2016	53	36	47	21	46	8	13	29	59	80
TRI3	2016	56	21	36	20	25	29	12	19	19	20
TRI4	2016	33	21	36	20	25	29	12	19	19	20
BLA1	2017	31	18	18	16	9	3	5	30	23	2
BLA2	2017	58	27	13	26	15	14	9	34	40	25
BLA3	2017	47	31	11	26	14	9	9	48	48	9
BLA4	2017	32	30	23	20	44	92	15	24	31	25
BUG2	2017	80	19	10	6	6	11	30	14	27	45
BUG3	2017	48	25	24	21	53	7	15	16	33	35
BUG4	2017	Prolongation AT									

DOCUMENT CONFIDENTIEL

BUG5	2017	Prolongation AT									
CHB1	2017	50	15	22	11	19	4	6	16	21	2
CHB2	2017	43	29	17	23	62	13	11	39	30	9
CHB3	2017	46	18	13	16	19	5	6	19	28	20
CHB4	2017	25	14	9	7	7	4	8	28	15	3
CRU1	2017	15	13	8	10	15	8	7	16	16	15
CRU2	2017	13	14	19	12	22	19	8	11	16	14
CRU3	2017	32	12	8	14	13	15	12	10	15	9
CRU4	2017	30	21	16	16	15	21	9	21	32	50
DAM1	2017	32	19	26	18	15	4	8	25	24	7
DAM2	2017	25	16	22	14	12	12	7	17	20	12
DAM3	2017	pas fait	13	12	15	11	4	8	10	23	7
DAM4	2017	34	19	18	26	19	35	10	12	25	5
FES1	2017	30	23	19	18	24	27	0	9	49	0
FES2	2017	Prolongation AT									
GRA1	2017	50	49	40	49	48	23	21	65	54	90
GRA2	2017	73	26	31	30	26	14	12	28	30	25
GRA3	2017	47	44	31	25	124	8	13	53	44	45
GRA4	2017	33	29	24	30	19	19	20	24	45	20
GRA5	2017	Prolongation AT									
GRA6	2017	39	22	10	20	21	7	14	17	42	20
SLB1	2017	46	29	17	18	16	6	9	48	40	8
SLB2	2017	40	31	30	30	26	42	39	22	39	51
TRI1	2017										
TRI2	2017	50	35	62	27	23	13	11	34	49	80
TRI3	2017	41	26	32	26	28	9	14	30	31	25
TRI4	2017										

Tableau 1 : données de tous les réacteurs du palier 900 MW sur les différents circuits de 2011-2017

DOCUMENT CONFIDENTIEL

REACTEUR	CAMPA GNE D'ARRE T année	INDICE BOUCLE RCP FIN CYCLE 1E-2mSv/h	INDICE BR 1E- 2mSv/h	INDICE RRA 1E- 2mSv/h	INDICE RCV 1E- 2mSv/h	INDICE RPE 1E- 2mSv/h	INDICE RIS 1E- 2mSv/h	INDICE PZR 1E- 2mSv/h	INDICE GV 1E- 2mSv/h	INDICE RCP 1E- 2mSv/h	INDICE PTR 1E- 2mSv/h
BEL1	2011	30	15	12	32	12		11	6	17	9
BEL2	2011										
CAT1	2011										
CAT2	2011	12	11	8	21	1		6	5	15	13
CAT3	2011	14	19	17	18	11		14	7	24	31
CAT4	2011	13	11	15	23	2		12	4	12	6
FLA1	2011	41	61	29	19	30		33	46	102	39
FLA2	2011	26	50	18	20	17		22	13	94	31
GOL1	2011										
GOL2	2011	27	13	13	21	4		2	4	17	2
NOG1	2011	20	14	15	23	4		12	12	16	1
NOG2	2011	26	18	8	11	5		9	10	30	1
PAL1	2011										
PAL2	2011	23	29	14	23	8		32	6	45	12
PAL3	2011	19	18	11	23	6		19	7	24	15
PAL4	2011	22	37	23	26	30		22	15	48	113
PEN1	2011	15	15	10	20	5		9	5	22	5
PEN2	2011	20	17	17	23	9		21	6	18	36
SAL1	2011	17	26	17	11	5		12	12	44	6
SAL2	2011										
BEL1	2012										
BEL2	2012	17	13	11	18	4		10	6	15	29
CAT1	2012	19	34	34	22	25		22	7	26	225
CAT2	2012	13	12	6	16	3		24	5	13	16
CAT3	2012	13	15	14	20	13		6	7	20	13
CAT1	2012										
FLA1	2012	38	47	28	17	22		56	28	60	98
FLA2	2012										
GOL1	2012	27	46	222	8	5		15	5	14	8
GOL2	2012	21	13	11	17	5		14	3	17	14
NOG1	2012	18	10	6	6	3		5	7	16	1
NOG2	2012	20	15	11	10	7		17	6	20	13
PAL1	2012	38	39	39	27	8		43	15	52	32
PAL2	2012	29	46	23	35	7		71	9	53	167
PAL3	2012	30	30	18	21	8		23	10	45	44
PAL4	2012										
PEN1	2012										
PEN2	2012	20	18	23	20	9		12	11	21	16
SAL1	2012	13	13	8	9	7		10	4	20	2
SAL2	2012	17	23	14	16	9		24	16	33	19

DOCUMENT CONFIDENTIEL

BEL1	2013	32	17	12	25	12		13	6	19	44
BEL2	2013	16	14	14	9	4		14	6	13	64
CAT1	2013	18	28	21	23	44		11	6	21	180
CAT2	2013										
CAT3	2013	10	19	12	17	8		16	5	12	59
CAT4	2013	15	16	16	19	2		20	6	12	35
FLA1	2013										
FLA2	2013	29	33	15	33	14		36	16	51	4
GOL1	2013	15	10	7	12	8		16	4	12	3
GOL2	2013										
NOG1	2013										
NOG2	2013										
PAL1	2013	31	52	27	40	9		37	13	68	200
PAL2	2013										
PAL3	2013	24	34	14	16	9		21	9	59	38
PAL4	2013	24	67	38	100	20		35	14	91	170
PEN1	2013	15	15	11	18	8		19	9	19	9
PEN2	2013										
SAL1	2013										
SAL2	2013	22	28	16	16	22		39	13	39	2
BEL1	2014	30	15	13	33	11		5	6	20	3
BEL2	2014										
CAT1	2014										
CAT2	2014	12	12	13	18	3		14	6	14	5
CAT3	2014										
CAT4	2014	14	12	17	19	2		12	5	12	16
FLA1	2014	30	31	14	20	13		31	14	50	5
FLA2	2014	28	30	20	33	11		29	10	45	7
GOL1	2014	16	10	7	13	7		12	5	12	4
GOL2	2014	21	18	18	37	6		12	12	22	13
NOG1	2014	17	9	7	14	4		7	4	13	1
NOG2	2014	22	13	10	13	9		21	5	15	11
PAL1	2014	38	71	24	57	18		42	31	125	32
PAL2	2014	32	83	46	39	9		38	16	152	23
PAL3	2014										
PAL4	2014	28	32	18	32	25		25	16	44	40
PEN1	2014	16	13	10	14	4		24	6	15	4
PEN2	2014	25	19	12	34	8		23	8	22	8
SAL1	2014	13	19	6	8	10		13	5	30	36
SAL2	2014										
BEL1	2015										
BEL2	2015	16	12	13	25	4		14	4	13	1
CAT1	2015	15	26	21	7	7		24	6	12	215

DOCUMENT CONFIDENTIEL

CAT2	2015	14	15	13	15	4		9	5	17	65
CAT3	2015	11	13	13	20	5		10	6	14	30
CAT4	2015										
FLA1	2015	36	39	20	16	15		32	23	65	8
FLA2	2015	30	30	16	27	8		19	19	49	3
GOL1	2015										
GOL2	2015	34	20	20	43	11		16	9	22	11
NOG1	2015	20	12	9	16	12		8	14	12	12
NOG2	2015	17	9	8	12	7		5	4	12	0
PAL1	2015										
PAL2	2015	24	33	22	33	6		37	10	46	33
PAL3	2015	32	54	21	39	8		53	12	90	40
PAL4	2015										
PEN1	2015										
PEN2	2015	21	27	39	38	6		35	10	27	17
SAL1	2015	12	17	11	6	4		21	17	24	11
SAL2	2015	20	37	13	14	6		25	9	69	22
BEL1	2016	30	19	17	30	10		9	10	25	23
BEL2	2016	15	11	17	17	3		12	3	11	33
CAT1	2016	16	17	24	22	10		22	5	13	52
CAT2	2016										
CAT3	2016	19	29	34	30	6		17	24	31	58
CAT4	2016	16	19	15	21	2		11	16	17	105
FLA1	2016	34	37	22	19	21		33	24	58	4
FLA2	2016										
GOL1	2016	15	13	8	13	5		27	5	14	14
GOL2	2016										
NOG1	2016	32	13	11	12	3		9	6	18	26
NOG2	2016										
PAL1	2016	42	51	36	72	14		52	14	71	31
PAL2	2016	Prolongatio n AT									
PAL3	2016										
PAL4	2016	22	30	14	17	47		30	13	36	60
PEN1	2016	27	15	11	20	15		33	4	12	19
PEN2	2016	20	19	27	27	8		21	12	19	27
SAL1	2016										
SAL2	2016	Pas fait	23	16	16	3		24	8	31	71
BEL1	2017	26	17	12	11	9		25	5	22	24
BEL2	2017	15	14	16	16	3		22	5	13	62
CAT1	2017										
CAT2	2017	15	19	11	20	3		20	5	24	43
CAT3	2017										
CAT4	2017	11	12	13	16	2		13	3	13	32

FLA1	2017										
FLA2	2017	24	29	27	27	17		22	36	41	2
GOL1	2017	17	14	11	13	9		13	6	17	24
GOL2	2017	23	18	41	41	8		14	7	19	7
NOG1	2017	20	20	20	17	6		17	31	21	17
NOG2	2017	28	17	22	22	6		9	9	24	24
PAL1	2017										
PAL2	2017	Prolongatio n AT									
PAL3	2017	23	21	13	21	5		29	7	31	2
PAL4	2017	24	28	22	20	24		32	21	30	60
PEN1	2017	12	11	12	14	2		15	4	13	8
PEN2	2017										
SAL1	2017	15	20	11	7	7		15	6	33	11
SAL2	2017										

Tableau 2 : données de tous les réacteurs du palier 1300 MW sur les différents circuits de 2011-2017

DOCUMENT CONFIDENTIEL

REACTEUR	CAMPAGNE D'ARRET année	INDICE BOUCLE RCP FIN CYCLE 1E- 2mSv/h	INDICE BR 1E- 2mSv/h	INDICE RRA 1E- 2mSv/h	INDICE RCV 1E- 2mSv/h	INDICE RPE 1E- 2mSv/h	INDICE RIS 1E- 2mSv/h	INDICE PZR 1E- 2mSv/h	INDICE GV 1E- 2mSv/h	INDICE RCP 1E- 2mSv/h	INDICE PTR 1E- 2mSv/h
CHO1	2011	35	8	6	13	3		11	3	10	0
CHO2	2011										
CIV1	2011	16	11	6	23	7		11	26	8	2
CIV2	2011										
CHO1	2012										
CHO2	2012	27	7	8	7	7		13	3	8	0
CIV1	2012										
CIV2	2012	25	14	20	24	30		26	3	8	5
CHO1	2013	19	7	2	18	7		10	5	6	0
CHO2	2013	28	19	13	13	26		65	4	14	9
CIV1	2013	20	9	8	23	8		18	2	8	1
CIV2	2013	24	18	43	53	17		13	2	12	3
CHO1	2014	19	5	7	7	3		7	3	6	2
CHO2	2014										
CIV1	2014	19	13	13	43	8		11	17	9	13
CIV2	2014										
CHO1	2015	21	13	16	9	9		23	2	15	4
CHO2	2015	42	20	17	25	31		65	5	11	9
CIV1	2015										
CIV2	2015	28	27	34	26	25		70	37	13	3
CHO1	2016										
CHO2	2016	39	13	38	38	23		31	2	5	2
CIV1	2016	20	8	8	17	8		19	3	5	2
CIV2	2016	23	22	31	31	69		22	30	13	10
CHO1	2017	33	12	13	6	11		21	4	14	3
CHO2	2017	50	21	25	25	19		75	3	12	15
CIV1	2017	20	13	10	50	19		13	3	9	7
CIV2	2017										

Tableau 3 : données de tous les réacteurs du palier 1450 MW sur les différents circuits de 2011-2017

