

# Simulation des incertitudes associées aux calculs des estimateurs «moyenne» et «fréquence de dépassement de seuil» utilisées pour l'évaluation de l'état chimique des masses d'eau souterraine

Rapport final

BRGM/RP-65519-FR

Janvier 2016

1.89 3740.46 -625.5





# **Simulation des incertitudes associées aux calculs des estimateurs «moyenne» et «fréquence de dépassement de seuil» utilisées pour l'évaluation de l'état chimique des masses d'eau souterraine**

Rapport final

**BRGM/RP-65519-FR**  
Janvier 2016

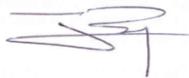
Étude réalisée dans le cadre de la convention ONEMA-BRGM 2013-2015

**A. Henriot, N. Croiset, B. Lopez**

## **Vérificateur :**

Nom : Jean-Philippe Ghestem  
Fonction : Ingénieur chimiste  
Date : 11/02/2016

Signature :



## **Approbateur :**

Nom : Laurence Chery  
Fonction : Responsable d'unité  
Date : 15/02/2016

Signature :



Le système de management de la qualité et de l'environnement  
est certifié par AFNOR selon les normes ISO 9001 et ISO 14001.

## RESUME

L'action réalisée en 2015 s'inscrit dans le cadre de la Directive 2000/60/CE et de la Directive eaux souterraines 2006/118/CE qui imposent aux États-Membres d'évaluer l'état chimique du milieu aquatique à partir des données issues des réseaux de surveillance de la qualité des eaux mis en œuvre en 2007. Les travaux menés en 2015 et présentés dans le rapport sont axés sur l'estimation de l'incertitude associée aux calculs de la moyenne et de la fréquence de dépassement de valeurs seuils, 2 indicateurs largement utilisés pour estimer l'état chimique des masses d'eau.

En 2013 et 2014 des précédents travaux menés sous convention ONEMA-BRGM ont permis de développer et de mettre en œuvre une méthode de simulation de chronique d'évolution des micropolluants dans l'environnement. Ce jeu de « données virtuelles » simulant la réalité environnementale a été utilisé une première fois en 2014 afin de caractériser la capacité des réseaux de surveillance de la qualité des eaux souterraines à fournir des données permettant d'identifier les tendances d'évolution (Lopez et al., rapport BRGM/RP-63845-FR, 2014).

Cette méthode de simulation des chroniques a été reprise pour répondre à d'autres questions qui intéressent l'application de la DCE. Ainsi, en appelant « estimateurs » l'ensemble des indicateurs calculés dans les différents tests caractérisant l'Etat chimique et qualitatif des eaux, l'application de la méthode de simulation des chroniques virtuelles d'évolution des micropolluants a permis :

- d'identifier, pour une série donnée de chroniques affectées par une variabilité environnementale donnée, la gamme de valeur dans laquelle varie les estimateurs environnementaux "vrais" ;
- d'estimer avec quelle performance les estimateurs calculés à partir des mesures ponctuelles s'approchent de la gamme dans laquelle se situent les valeurs vraies, pour diverses incertitudes métrologiques et diverses fréquences de suivi simulées ;
- de développer un outil de génération et d'exploration d'abaques présentant l'évolution des incertitudes simulées des estimateurs en fonction de la longueur, de la fréquence de prélèvement, du type de comportement temporel et de la valeur d'incertitude analytique théorique qui caractérisent les chroniques de suivi de la qualité des eaux.

Ces travaux, bien que développés en premier lieu à partir de données sur les eaux souterraines, concernent aussi le domaine des eaux de surface, les 2 estimateurs « moyenne des concentrations » et « fréquence de dépassement de valeur seuil » étant aussi utilisés pour estimer l'état chimique des cours d'eau et des plans d'eau. L'outil final HYPST-R fournit une aide à tous les gestionnaires en charge de l'évaluation de l'état chimique du milieu aquatique en France en leur permettant d'affiner l'estimation des incertitudes liées aux calculs des indicateurs en prenant en compte les caractéristiques propres à chaque suivi de paramètres chimiques aux points de contrôle. Ces nouvelles informations sont destinées à intégrer la notion de niveau de confiance que les opérateurs doivent associer à l'évaluation de l'état chimique rapporté au niveau européen.

**Mots-clés :** Analyse statistique, niveau de confiance, incertitude, état des lieux, DCE, analyse série chronologique, qualité eau

Couverture géographique : Nationale

Niveau de lecture : Expert

CORRESPONDANTS ONEMA : STAUB Pierre-François

En bibliographie, ce rapport sera cité de la façon suivante :

**Croiset N., Henriot A., Lopez B.** (2016) – Simulation des incertitudes associées aux calculs des estimateurs « moyenne » et « fréquence de dépassement » utilisés pour l'évaluation de l'état chimique des masses d'eau souterraine. Rapport final BRGM/RP-65519-FR, 71 p., 41 ill., 3 ann.



## Synthèse pour l'action opérationnelle

L'estimation du niveau de confiance de l'évaluation de l'état chimique est actuellement réalisée de manière plus ou moins empirique ou à dire d'expert par les personnes en charge du rapportage européen. Il n'existe en effet pas de méthode définie à l'échelle nationale et les paramètres à prendre en compte pour son estimation sont aujourd'hui essentiellement qualitatifs. Si ces paramètres apparaissent tout à fait pertinents pour estimer le niveau de confiance global de l'exercice d'évaluation de l'état, ils sont néanmoins insuffisants et doivent être complétés par des informations sur les incertitudes statistiques associées aux calculs des estimateurs Mma et Fd. Il est en effet important que les gestionnaires puissent disposer d'informations sur les incertitudes relatives aux estimateurs de l'état chimique des masses d'eau souterraines qu'ils calculent.

Le travail réalisé dans le cadre de la convention ONEMA-BRGM 2015 a permis de proposer des pistes concrètes afin d'estimer les incertitudes relatives aux calculs de la moyenne des moyennes annuelles (Mma) et de la fréquence de dépassement de la valeur seuil (Fd) en fonction du comportement de l'élément chimique étudié, de l'incertitude associée aux mesures et du calendrier d'échantillonnage. La méthode suivie repose sur la génération de chroniques virtuelles dont les paramètres sont connus et maîtrisés. De façon analogue aux travaux de (Lopez et al. 2014) les chroniques virtuelles ont été générées selon différentes modalités de programme de surveillance (type de calendrier, fréquence de prélèvement, type d'évolution temporelle, périodicité des cycles variable.). Les estimateurs Mma et Fd ont été calculés sur ces séries temporelles la variabilité de ces résultats en fonction de la variation des paramètres définissant les chroniques a été investiguée.

Ces résultats pourront intéresser les gestionnaires de la ressource en eau à plusieurs titres :

- Tout d'abord, ils permettent de se rendre compte que les incertitudes liées au calcul de la Mma et à la fréquence de dépassement de seuil peuvent être particulièrement élevées, et donc l'évaluation de l'état peut être biaisée ;
- Ensuite, ils peuvent aider le gestionnaire lors de l'établissement d'un programme de surveillance, à optimiser le suivi en fonction des caractéristiques de chaque point d'eau.

Les résultats ont tout d'abord été générés sous la forme d'abaques graphiques présentant l'évolution de l'incertitude associée à l'estimateur Mma ou Fd en fonction des paramètres qui caractérisent l'évolution de l'élément chimique (comportement) et le suivi virutel réalisé (longueur de la chronique, fréquence d'échantillonnage, incertitude analytique). De très nombreuses combinaisons de paramètres (plus de 20 000) ont été testées ; il n'était donc pas possible de représenter dans ce rapport l'ensemble des résultats : une interface graphique nommée Hypst-R a donc été développée afin de permettre à l'opérateur :

- d'afficher et analyser facilement l'ensemble des résultats générés durant l'étude,
- de paramétriser lui-même les chroniques (type de suivi, d'évolution et d'incertitude analytique) pour lesquelles l'incertitude est estimée,
- d'exporter des résultats sélectionnés selon plusieurs formats possibles.

L'interface a été développée sous environnement R, en utilisant le paquet « Shiny ». Ce paquet permet de construire de manière simple et efficace des applications interactives avec R, où les

sorties se mettent à jour au fur et à mesure que les entrées sont modifiées sans nécessiter un rechargeement de l'interface. Pour le moment, Hypst-R n'est disponible qu'en installation locale mais il sera simple de créer également une application web si besoin. L'intérêt d'une version web est que les utilisateurs n'ont alors pas besoin d'installer R et ses extensions sur leur machine et, de plus, ils disposent automatiquement de la version la plus à jour de l'outil. A terme, ce type d'interface permettrait de « piloter » via une application web l'ensemble des outils d'analyse statistique des données sur les eaux souterraines développés sous environnement R.

# Sommaire

<b>1. Introduction .....</b>	<b>11</b>
1.1. CADRE DU PROJET .....	11
1.2. OBJECTIFS .....	12
<b>2. Aspects théoriques et méthodologiques.....</b>	<b>15</b>
2.1. DEFINITIONS .....	15
2.2. METHODE .....	16
<b>3. Génération de chroniques virtuelles et détermination des incertitudes associées au calcul des métriques .....</b>	<b>19</b>
3.1. MODELES D'EVOLUTIONS TEMPORELLES THEORIQUES.....	19
3.1.1. Formulation de Lopez et al. (2014) .....	19
3.1.2. Modèles théoriques d'évolution.....	19
3.1.3. Gamme de paramètres investiguée .....	21
3.1.4. Exemples d'évolutions théoriques.....	23
3.2. GENERATION DES CHRONIQUES VIRTUELLES .....	26
3.2.1. Méthode.....	26
3.2.2. Gamme des paramètres investiguée .....	26
3.3. CALCUL DES METRIQUES THEORIQUES DES CHRONIQUES.....	28
3.4. CALCULS DES METRIQUES SUR LES CHRONIQUES GENEREES .....	31
3.4.1. Nombre de tests, normalité des données.....	31
3.4.2. Calcul de la distribution des moyennes des moyennes annuelles et des fréquences de dépassement de valeur seuil.....	31
3.4.3. Ecart aux valeurs vraies.....	34
<b>4. Résultats .....</b>	<b>37</b>
4.1. TESTS REALISES.....	37
4.2. ABAQUES .....	38
4.2.1. Cas des évolutions monotones .....	38
4.2.2. Cas des évolutions cycliques périodiques.....	40
4.3. INTERFACE « HYPST-R » .....	48
4.3.1. Considérations générales sur l'interface .....	48
4.3.2. Architecture de l'interface « Hypst-R » .....	49
4.3.3. Onglets Mma et Fd .....	50
4.3.4. Génération de chroniques.....	53

<b>5. Conclusions .....</b>	<b>57</b>
<b>6. Bibliographie .....</b>	<b>59</b>

## Liste des illustrations

Illustration 1 : Procédure détaillée du calcul du bon état chimique, Circulaire du 17 décembre 2008, Annexe III, paragraphe 4. ....	12
Illustration 2 : exemple d'évolution théorique de concentration (en haut) et de chronique créée à partir par échantillonnage au sein de cette évolution (en bas). ....	17
Illustration 3 : Exemple de séries temporelles générées suivant une même loi d'évolution avec des paramètres identiques, mettant en évidence le déphasage aléatoire sur la fonction périodique. a) signal périodique annuel + bruit ; b) signal périodique pluriannuel (période 8 ans) + bruit ; c) signal périodique à deux cycles : annuel et pluriannuel (8 ans)+ bruit. ....	21
Illustration 4 : Indice NAO de 1864 à 2000 défini comme la différence de pressions entre Lisbonne et Stykkisholmur pour les mois de décembre à mars. Source : Wanner et al, 2001.22	
Illustration 5 : Exemples d'évolutions de type monotone affectées d'une incertitude analytique de 2.5 %, 10 % et 50 %.....	24
Illustration 6 : Exemples d'évolutions à 1 cycle de longueur 1 an, et d'amplitude égale à 50 % de la valeur moyenne, pour des incertitudes analytiques de 2.5 % et 30 . ....	25
Illustration 7 : Exemples d'évolutions à deux cycles de longueurs 1 an et 6 ans, et d'amplitude égale à 50 %, affectées d'une incertitude analytique de 2.5 %. A gauche, la part du premier cycle (1an) dans le signal représente 50 %, et celle du deuxième cycle (6 ans) représente 50 %. A droite, la part du premier cycle représente 10 %, et la part du second cycle 90 %. ....	25
Illustration 8 : Exemples de répartition dans le temps des dates de prélèvement suivant les différentes modalités de calendriers proposées pour 2 prélèvements par an en moyenne. ....	27
Illustration 9 : Densité de probabilité d'une loi normale de moyenne 1 et d'écart-type 0.15 (c'est-à-dire une incertitude élargie égale à 0.3). La fréquence de dépassement de seuil est calculée pour 3 valeurs seuil : 1.025, 1,1 et 1,25 : le seuil est représenté par la droite verticale de couleur rouge, et la zone colorée en rouge représente la probabilité d'avoir une valeur supérieure au seuil. ....	28
Illustration 10 : Illustration du calcul de la fréquence de dépassement de seuil pour une évolution cyclique. La valeur seuil a été fixée à 1,025. L'incertitude analytique est égale à 0,05. La figure du haut représente l'évolution des concentrations sur un cycle (les lignes pointillées représentent l'enveloppe dans laquelle se situent 95% des valeurs). En bas, évolution de la probabilité de dépassement de la valeur seuil. La probabilité moyenne sur un cycle entier est 0,31. ....	29
Illustration 11 : Illustration du calcul de la fréquence de dépassement de seuil pour une évolution cyclique à double cycle. La valeur seuil et l'incertitude analytique sont identiques à l'Illustration 10. La probabilité de dépassement de la valeur seuil dans ce cas est de 0,32. ....	30
Illustration 12 : Deux exemples de convergence de la Mma moyenne calculée sur les évolutions temporelles simulées en fonction du nombre de tests réalisés A) cas d'une évolution monotone, de longeur 3 ans, 12 prélèvements par an, incertitude analytique = 50 %,	

calendrier régulier; B) cas d'une évolution à 1 cycle (période 1 an, amplitude 0.2), de longueur 18 ans, 1 prélèvement par an, incertitude analytique 50%. ....	31
Illustration 13 : Valeurs de la Mma et de la fréquence de dépassement de la valeur seuil calculées pour 10 000 chroniques et moyennes et écart-type de ces métriques. Dans cet exemple, l'évolution est monotone, l'incertitude analytique est de 50%, et la chronique est composée d'un prélèvement par an pendant 6 ans ; la valeur seuil est fixée à 1,5. ....	32
Illustration 14 : Densité de probabilité de la Mma . Sur un tel graphique, on peut lire la probabilité que la variable soit comprise entre deux valeurs a et b, comme l'aire sous la courbe dans l'intervalle [a,b]. En noir, l'histogramme représente les données expérimentales tandis que la courbe rouge représente la loi normale de moyenne 1 et d'écart-type 0,10 , l'écart-type calculé sur les 10 000 chroniques.....	33
Illustration 15 : Histogramme des fréquences de dépassement de la valeur seuil. Les paramètres des chroniques sont les mêmes que ceux de l'exemple de l'illustration 13. La valeur seuil est considérée égale à 1,125. ....	33
Illustration 16 : Densité de probabilité des Mma calculées sur 10 000 chroniques correspondant à l'exemple précédent (moyenne =1, écart-type= 0.10) et calcul des bornes permettant d'avoir 95% des valeurs de la Mma entre ces bornes.....	34
Illustration 17 : Densité de probabilité des Mma calculées sur 10 000 chroniques correspondant à l'exemple précédent (moyenne =1 , écart-type= 0,104). Est figuré en rouge l'intervalle 0,975-1,025 correspondant à la Mma vraie à 5% près. La probabilité de trouver la Mma vraie est calculée comme l'aire de ce polygone, elle est égale à 19%.....	34
Illustration 18 : Histogramme des fréquence dépassement de seuil et définition de l'intervalle de confiance. Les paramètres sont les mêmes que sur l'illustration 13 et l'illustration 12 (évolution monotone, incertitude analytique = 50%, longueur de la chronique = 6 ans, fréquence =1/an, valeur seuil =1,125). ....	35
Illustration 19 : Gamme des paramètres décrivant les chroniques générées .....	37
Illustration 20 : Incertitude sur la Mma (%) en fonction de la fréquence de prélèvement dans le cas d'une évolution monotone, pour des chroniques de longueur 6 ans. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5, 10, 30 et 50 %. ....	39
Illustration 21 : Largeur de l'intervalle de confiance autour de la fréquence de dépassement de seuil en fonction de la fréquence de prélèvement dans le cas d'une évolution monotone, pour des chroniques de longueur 3 ans et pour une valeur seuil égale à 1.1 fois la moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2.5, 10, 30 et 50 %.....	40
Illustration 23 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique annuelle, pour des chroniques de longueur 3 ans. Les différentes couleurs représentent différents types de calendrier . Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différents cas d'amplitude du signal périodique (en part de la valeur moyenne). ....	41
Illustration 24 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique annuelle, pour des chroniques de longueur 6 ans. Les différentes couleurs représentent différents types de calendrier . Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différents cas d'amplitude du signal périodique (en part de la valeur moyenne). ....	42

- Illustration 25 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique annuelle, pour des chroniques de longueur 18 ans. Les différentes couleurs représentent différents types de calendrier . Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différents cas d'amplitude du signal périodique (en part de la valeur moyenne). ..... 43
- Illustration 26 : Largeur de l'intervalle de confiance autour de la fréquence de dépassement de la valeur seuil en fonction de la fréquence de prélèvement dans le cas d'une évolution à un cycle annuel, pour des chroniques de longueur 12 ans et pour une valeur seuil égale à 1.1 fois la moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 5 % . Les deux colonnes représentent deux valeur d'amplitude du signal périodique, respectivement 10 % et 50 % à gauche et à droite. ..... 44
- Illustration 27 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 6 ans, et pour une amplitude de cycle de 10 % de la valeur moyenne. Les différentes couleurs représentent différents types de calendrier . Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportion du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite)..... 45
- Illustration 28 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 6 ans et pour une amplitude de cycle de 20 % de la valeur moyenne. Les différentes couleurs représentent différents types de calendrier . Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportion du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite)..... 46
- Illustration 29 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 6 ans et pour une amplitude de cycle de 50 % de la valeur moyenne. Les différentes couleurs représentent différents types de calendrier . Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportion du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite)..... 47
- Illustration 30 : Largeur de l'intervalle de confiance autour de la fréquence de dépassement de la valeur seuil en fonction de la fréquence de prélèvement dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 3 ans et pour une amplitude de cycles de 50 % de la valeur moyenne et pour une valeur seuil égale à 1.025 fois la moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportion du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite)..... 48
- Illustration 31 Copie d'écran du bandeau supérieur de l'interface Hypst-R (en haut). Architecture de l'interface Hypst-R, description des fonctions de chacun des onglets (en bas).50
- Illustration 32 Capture d'écran de l'interface "Hypst-R", onglet Mma, décrivant l'organisation de l'onglet en 6 zones dont les fonctions sont distinctes. .... 51
- Illustration 33 Capture d'écran de l'interface Hypst-R. Onglet Mma, détail de l'affichage conditionnel des critères relatifs aux évolutions à un ou deux cycles..... 52
- Illustration 34 : Exemple de graphiques générés grâce à Hypst-R. Exemple du choix de plusieurs types de calendrier distingués par des couleurs différentes (calendrier régulier uniquement (figure de gauche) en rouge, et calendriers bruité, régulier, troué (figure

de droite) respectivement de couleur rouge, verte et bleue). Les graphiques sont organisés en colonne au sein de laquelle chaque ligne correspond à un type d'incertitude analytique (ici, de haut en bas, 2,5 %, 10 %, 30 % et 50 %). ....	52
Illustration 35 : Exemple de graphiques générés grâce à Hypst-R. Exemple du choix de plusieurs types d'évolutions : monotone et à un cycle, pour lesquelles les résultats sont organisés selon deux colonnes (résultats relatifs à l'évolution à un cycle à gauche, résultats relatifs à l'évolution monotone à droite). Chaque ligne correspond à une incertitude analytique différente (de haut en bas, 2,5 %, 10 %, 30 % et 50 %). 53	
Illustration 37 : Exemple de chronique générée avec Hypst-R. Cas d'une chronique de type monotone dont l'incertitude analytique associée est de 30 %. Exemple de calcul et d'affichage des moyennes annuelles (les moyennes annuelles sont représentées sous forme de segment de droite de la même couleur que les points). ....	54
Illustration 38 : Exemple de chronique générée avec Hypst-R. Cas d'une chronique de type monotone dont l'incertitude analytique associée est de 30 %. Les moyennes annuelles (les moyennes annuelles sont représentées sous forme de segment de droite de la même couleur que les points) et la la Mma (segment de couleur orangée, représenté de plus forte épaisseur) sont calculées et affichées. ....	55
Illustration 39 : Capture d'écran de l'interface Hypst-R, onglet « Exemple de génération de » permettant de générer des chroniques selon des critères choisis par l'utilisateur au moyen des différents champs de sélection. Le graphique représenté ici correspond à une évolution de type monotone, selon un calendrier régulier de longueur 12 ans, selon une fréquence de 4 échantillons/an associés à une incertitude analytique de 30 %. ....	56
Illustration 40 : Installation d'un package depuis Internet – Aperçu d'écran	65
Illustration 41 : Copie d'écran du logiciel R lors du choix du site miroir depuis lequel télécharger un paquet	65
Illustration 42 : Choix du répertoire de travail- Aperçu d'écran	66

## Liste des annexes

Annexe 1 : Vérification analytique de quelques résultats calculés	62
Annexe 2 : Installation et lancement de « Hypst'R »	63



# 1. Introduction

## 1.1. CADRE DU PROJET

La directive européenne 2000/60/CE et la directive fille 2006/118/CE introduisent les obligations réglementaires relatives au suivi et à l'évaluation de l'état chimique des masses d'eau souterraine, désignées sous le terme de « programme de surveillance ». L'objectif poursuivi par l'Union Européenne est de disposer, à l'échelle de son territoire, d'indicateurs agrégés permettant de rendre compte, de façon simple, de l'état chimique, quantitatif ou écologique, de l'ensemble des masses d'eau souterraines. Ces indicateurs doivent pouvoir s'appliquer et se comparer entre eux malgré la pluralité et la diversité des contextes géologiques, climatiques, hydrologiques et hydrogéologiques ainsi que la potentielle variabilité spatio-temporelle des pressions anthropiques s'exerçant sur ces masses d'eau.

L'état chimique des masses d'eau est évalué en Europe, en référence à la Directive Cadre sur l'Eau (DCE) 2000/60/CE et sa directive fille 2006/118/CE. Les modalités de l'évaluation de l'état chimique des masses d'eau souterraines sont précisées en France dans l'annexe III de la circulaire du 17 décembre 2008. Deux indicateurs doivent être calculés pour évaluer cet état : la moyenne des moyennes annuelles des concentrations (Mma) et la fréquence de dépassement de la valeur seuil (Fd, cf. [Illustration 1](#)) pour un ensemble de paramètres chimiques.

Le programme de surveillance a bénéficié d'améliorations depuis sa première parution (2000/60/CE) à mesure de son déploiement au sein de l'UE. Il est désormais davantage mature et les enjeux s'orientent désormais vers une meilleure prise en compte des incertitudes liées aux calculs des indicateurs et à leur agrégation. Cette étude s'intègre dans ce contexte et vise à proposer un cadre théorique et méthodologique pour l'estimation des incertitudes associées aux indicateurs « moyenne » et « fréquence de dépassement » de la valeur seuil. Elle s'inscrit dans le cadre de la convention Onema BRGM 2013-2015.

#### Etape 1 : Calcul de la moyenne par point

Pour chaque paramètre et chaque point, le calcul de la moyenne des moyennes annuelles (Mma) à comparer aux valeurs seuils ou aux normes de qualité se fera ainsi :

1. Prendre tous les résultats issus des réseaux de surveillance (contrôle de surveillance, contrôle opérationnel et autres) sur la période considérée (cf. remarque ci-dessous pour plus de précisions),
2. Pour chacune des années de contrôle et en chaque point, calculer la moyenne annuelle des concentrations (y compris si le nombre de données disponibles d'une année à l'autre est différent) (cf. remarque ci-après),
3. Faire la moyenne des moyennes annuelles (Mma).
4. Avec les données ainsi collectées, on qualifie individuellement l'état chimique de chaque point d'eau disposant de données pour le paramètre considéré tel que :

Un point d'eau est en bon état chimique si :

- Mma ne dépasse pas la valeur seuil (ou norme) du paramètre étudié ;
- et si, la fréquence de dépassement de la norme ou valeur seuil (Fd) n'excède pas 20% (les chroniques doivent compter au moins 5 valeurs sinon on ne pourra pas appliquer ce critère de 20%).

Si une de ces deux conditions n'est pas respectée, alors le point d'eau est déclaré en mauvais état chimique<sup>1</sup>.

*Illustration 1 : Procédure détaillée du calcul du bon état chimique, Circulaire du 17 décembre 2008, Annexe III, paragraphe 4.*

Le présent travail ne peut se comprendre isolément et représente un des volets de la convention. Il doit donc être considéré en relation avec l'ensemble des études déjà réalisées (Lopez et al. 2011 ; Croiset & Lopez 2013 ; Lopez et al. 2014) et de nombreuses références à ces études sont présentes dans ce travail. Les auteurs encouragent le lecteur à s'y référer.

## 1.2. OBJECTIFS

Si la DCE et la directive fille sur les eaux souterraines demandent que des niveaux de confiance soient respectivement associés aux réseaux de surveillance et à l'évaluation des tendances à la hausse, il n'en est rien en revanche pour l'évaluation du bon état chimique. Toutefois au niveau national, malgré l'absence d'exigence réglementaire, il est proposé dans la circulaire du 23 octobre 2012 sur les modalités de calcul de l'état des lieux des eaux souterraines, d'associer un niveau de confiance à chaque résultat du calcul de l'état des lieux. La circulaire fournit des recommandations pour définir un niveau de confiance de l'évaluation de l'état chimique sans toutefois fournir une méthode bien établie. Il est indiqué dans cette circulaire que « l'estimation du niveau de confiance peut prendre en compte les incertitudes

<sup>1</sup> Pour mémoire, remarques figurant dans l'annexe III

c) L'objectif de l'étape 1 est d'identifier les anomalies dans la qualité des eaux souterraines. Pour cette raison il n'y a pas de restriction minimum quant au nombre de données nécessaires au calcul de la moyenne des moyennes annuelles. Cependant compte tenu du fait qu'une moyenne n'est pas représentative de la même manière lorsqu'elle est calculée à partir d'une ou plusieurs valeurs annuelles, un indice de confiance sera estimé (cf. paragraphe 5) pour permettre de comparer cet estimateur d'un point à un autre, plus particulièrement pour comparer le réseau RCS/RCO respectant une fréquence d'échantillonnage bien défini (arrêté du 25 janvier 2010) et les autres réseaux de surveillance.

d) La période de référence pour l'évaluation de l'état des masses d'eau au titre de la DCE est de six ans (durée d'un programme de surveillance). Pour le premier plan de gestion, les données disponibles ne permettent pas toujours de disposer de ces six années. Les réseaux de suivi de l'état chimique des eaux souterraines (RCS et RCO) ne sont en place que depuis 2007 et 2008 respectivement. De ce fait, pour le prochain état des lieux (2013), compte tenu des délais d'analyse des échantillons, de récupération, de validation et d'exploitation des résultats analytiques, l'évaluation portera sur une période commune à tous : 2007-2010. Pour les réseaux mis en place en 2008, à l'exemple de la Guadeloupe et les réseaux RCO, la période de référence sera 2008-2010. Pour le prochain rapportage communautaire (2016), l'évaluation portera sur l'ensemble des données issues du contrôle de surveillance et du contrôle opérationnel sur la période 2007-2012 (premier cycle du programme de surveillance)

analytiques, les incertitudes liées à la représentativité des réseaux de surveillance ainsi que les incertitudes dues aux variations des concentrations » mais les méthodes proposées pour définir le niveau de confiance des différents tests de l'enquête appropriée sont essentiellement basées sur la quantité d'information existante (nombre de valeurs disponibles pour le test « intrusion salée », représentativité spatiale du réseau pour le test « qualité générale »...). Pour l'indicateur Mma, l'indice de confiance proposé est binaire et basé uniquement sur la comparaison de la fréquence d'échantillonnage avec les prescriptions de l'arrêté du 25 janvier 2010 (respect ou non-respect des prescriptions). Ce niveau de confiance ne prend donc aucunement en compte les incertitudes sur les données de surveillance.

Dans ce travail sont proposées des méthodes pour évaluer les incertitudes relatives aux calculs des Mma et Fd, en prenant en compte les erreurs présentes lors des processus d'échantillonnage et d'analyse et la variabilité temporelle des concentrations. L'évaluation de ces incertitudes est un premier pas indispensable à l'évaluation du niveau de confiance de l'état chimique.

Dans cette étude, le travail est focalisé sur l'estimation de l'incertitude statistique qui affecte le calcul des estimateurs Mma et Fd des paramètres utilisés pour le calcul de l'état chimique des eaux souterraines. Il est proposé d'aller plus loin que les recommandations actuellement faites au niveau national pour estimer les niveaux de confiance de l'évaluation de l'état des lieux. L'objectif est de donner aux gestionnaires les moyens d'estimer la probabilité que les résultats des calculs de la Mma et de la Fd soient corrects.

Les incertitudes sont estimées en fonction du comportement de l'élément chimique étudié et des caractéristiques du programme de surveillance envisagé. Cette contrainte impose de réaliser un grand nombre de tests correspondant à différents type de comportements, différentes incertitudes associées aux mesures et différents calendrier d'échantillonnage.

Pour laisser aux utilisateurs la possibilité de parcourir facilement l'ensemble des résultats produits, une interface interactive, baptisée « Hypst'R », a été créée. Cette interface permet à l'opérateur de simuler les incertitudes attendues sur la Mma et la Fd en fonction des paramètres qui définissent une chronique (fréquence et longueur de suivi, incertitude analytique, présence de variations cycliques...).



## 2. Aspects théoriques et méthodologiques

### 2.1. DEFINITIONS

Cette étude utilise la méthode développée par Lopez et al. (2014). Les définitions proposées dans le travail de 2014, utiles pour la compréhension de cette méthode sont reprises ci-dessous. Les auteurs soulignent que ces définitions ne se veulent pas à portée universelle mais sont propres à l'exercice réalisé.

- **L'évolution** correspond aux variations temporelles continues réelles de la concentration d'un élément chimique dissout dans les eaux souterraines en un point donné. Dans le cas d'un suivi chimique classique, chaque analyse correspond à la valeur de concentration à l'instant  $t$  précis du prélèvement. Une image partielle de l'évolution est donnée par la suite chronologique des analyses issues de prélèvements réalisés en un même point : la chronique. L'évolution est supposée être connue du gestionnaire lorsqu'il cherche à estimer les indicateurs d'état chimique et leurs incertitudes.
- Le **comportement** d'un élément chimique dissout dans un aquifère correspond aux caractéristiques de son évolution temporelle en un point donné de l'aquifère. Le comportement dépend à la fois des propriétés intrinsèques de l'élément chimique qui vont guider sa mobilité et des propriétés hydrogéologiques et hydrodynamiques de l'aquifère dans lequel circulent les eaux contenant l'élément chimique. De grandes classes de comportement peuvent être définies en regroupant des évolutions aux caractéristiques proches.
- La **chronique** est la suite chronologique des résultats des analyses chimiques réalisées sur des échantillons prélevés en un même point. La chronique est constituée de valeurs discrètes de concentration en un élément chimique auxquelles sont associées des dates de prélèvement. La chronique est caractérisée par une **longueur** (date de début et date de fin) et une **fréquence** d'échantillonnage (nombre de prélèvement par unité de temps). Les valeurs de concentrations qui constituent la chronique sont affectées par un bruit formé d'une composante environnementale, d'une composante liée à l'opération d'échantillonnage et d'une composante due à l'incertitude analytique. Ce bruit engendre un écart plus ou moins important entre la valeur vraie de l'évolution au moment du prélèvement et la valeur de concentration reportée dans la chronique.
- L'**incertitude** caractérise la dispersion des résultats qu'on peut attribuer à une caractéristique d'un objet « mesuré ». Dans le cadre de ce rapport, l'incertitude théorique prise en compte correspond à l'incertitude liée aux opérations d'échantillonnage et à analytique. Afin de simplifier le raisonnement et compte tenu du peu d'informations disponibles sur les incertitudes liées au prélèvement, le terme générique utilisé est « incertitude analytique ». Dans ce travail, on considère une incertitude avec un niveau de confiance de 95%. On supposera que les erreurs liées à la mesure obéissent à une loi normale. Dans ce cas, on peut exprimer l'incertitude comme 2 fois l'écart-type.
- Les indicateurs moyenne des moyennes annuelles (Mma) et fréquence de dépassement de la valeur seuil (Fd) sont définis comme suit :

$$\text{Moyenne des moyennes annuelles (Mma) : } M_{\text{ma}} = \frac{x_1}{N_{x_1}}$$

avec :

- $x_1$  moyenne arithmétique du paramètre considéré, pour l'année i  
 $N_{x_1}$  nombre total de moyennes annuelles

$$\text{Fréquence de dépassement (Fd) : } Fd = \frac{n}{N}$$

avec :

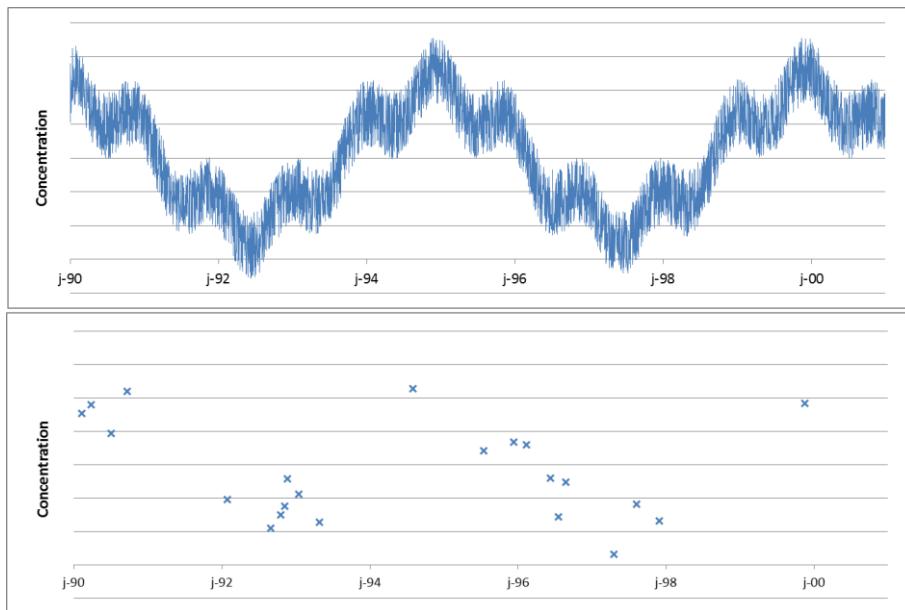
- n nombre de valeurs de concentration strictement supérieures à la valeur seuil  
N nombre total de données

## 2.2. METHODE

L'objectif de ce travail est de simuler l'incertitude associée aux moyennes des moyennes annuelles (Mma) et fréquences de dépassement de la valeur seuil (Fd) sur les chroniques de qualité des eaux souterraines en fonction de l'évolution de la concentration (monotone ou cyclique), de l'incertitude analytique associée aux mesures et du calendrier d'échantillonnage.

Plusieurs approches sont possibles pour traiter cette question. Ghestem et al. (2014) et Yari et al. (2015) ont proposé des méthodes d'estimation de l'incertitude sur les indicateurs réglementaires d'évaluation de l'état chimique des eaux de surface à partir des données élémentaires de concentrations et de leurs incertitudes analytiques. Les incertitudes sur les métriques retenues comme indicateurs réglementaires de l'état (moyennes annuelles et concentrations maximum annuelles dans le cas des eaux de surface) sont calculées puis l'incertitude sur l'état (défini comme la position des métriques par rapport aux valeurs seuil réglementaires) est calculée. Différentes méthodes sont proposées. Deux semblent donner satisfaction et donnent des résultats concordants : une méthode « analytique » et la méthode dite « de Cyrano ». Dans la méthode « analytique », l'incertitude associée à la métrique « moyenne » est estimée en combinant les incertitudes associées à chaque donnée et en tenant compte de la dispersion entre les données. L'incertitude associée à chaque donnée n'étant en général pas disponible, elle est calculée de façon simplifiée en utilisant un modèle de type puissance dans lequel l'incertitude est de l'ordre de 75% quand la concentration est au niveau de la limite de quantification et une incertitude de 30% à partir de 10 fois la limite de quantification. La méthode de Cyrano est une technique de « bootstrap » : la statistique est calculée à partir du rééchantillonnage des données (de très nombreuses simulations sont réalisées engendrant des nouveaux échantillons obtenus par « tirage avec remise » dans une urne contenant l'échantillon initial). Cette technique utilise la seule information des données et se base sur un minimum d'hypothèses. Pour plus de détails, le lecteur est encouragé à se reporter à ces rapports qui fournissent une réflexion complémentaire au travail réalisé dans la présente étude.

La méthode suivie dans le présent travail pour le contexte des eaux souterraines est analogue à celle proposée par Lopez et al. (2014). Elle consiste à créer tout d'abord des évolutions continues théoriques de concentrations suivant des modèles de comportements estimés à partir d'observations. Ces évolutions théoriques sont parfaitement maîtrisées (Mma et Fd sont fixées par construction notamment). Des valeurs sont ensuite échantillonnées dans ces évolutions afin de simuler un suivi de la qualité des eaux. Ces données forment des chroniques virtuelles. Un exemple d'évolution théorique et de chronique virtuelle est présenté sur l'[Illustration 2](#). Pour chacune de ces chroniques, la Mma et la Fd sont calculées.



*Illustration 2 : Exemple d'évolution théorique de concentration (en haut) et de chronique créée à partir par échantillonnage au sein de cette évolution (en bas).*

La procédure adoptée comporte 4 étapes :

- i. Identification des modèles mathématiques pertinents pour décrire les évolutions temporelles les plus généralement constatées dans les eaux souterraines (basé sur Lopez et al., 2014) ;
- ii. Génération d'environ 20 000 évolutions.
- iii. Pour chacune de ces évolutions
  - o Génération de 10 000 chroniques virtuelles :
    - échantillonnage, au sein des évolutions continues, et selon plusieurs modalités (type de calendriers, fréquences d'échantillonnage) de façon à constituer une chronique, analogue du jeu de données dont pourrait disposer le gestionnaire ;
    - calcul des indicateurs Mma et Fd sur cette chronique
  - o Calcul de la dispersion et de l'écart aux valeurs théoriques des 10 000 Mma et Fd calculés.



### **3. Génération de chroniques virtuelles et détermination des incertitudes associées au calcul des métriques**

#### **3.1. MODELES D'EVOLUTIONS TEMPORELLES THEORIQUES**

##### **3.1.1. Formulation de Lopez et al. (2014)**

Les modèles d'évolutions temporelles sont définis en référence aux évolutions observées dans les milieux naturels, notamment des ions nitrates et des produits phytosanitaires, décrites dans les travaux de Baran et al. (2009) et Lopez et al. (2012, 2015). De ces travaux, il ressort que les évolutions temporelles de la qualité des eaux souterraines peuvent être classées en 4 grands types:

- évolution monotone ;
- évolution à cyclicité périodique principalement annuelle ;
- évolution à cyclicité périodique principalement pluriannuelle ;
- évolution à double cyclicité périodique annuelle et pluriannuelle.

Dans le travail de Lopez et al. (2014), 2 modèles mathématiques avaient été retenus pour représenter ces 4 types d'évolution :

- un modèle linéaire avait été utilisé pour l'évolution monotone stationnaire ou avec tendance monotone
- une formulation par variogramme avait été utilisée pour les différentes évolutions cycliques.

##### **3.1.2. Modèles théoriques d'évolution**

Afin d'utiliser une formulation aussi simple que possible, nous avons choisi, dans ce travail, de ne pas utiliser les variogrammes mais de représenter les évolutions cycliques par des fonctions périodiques de type cosinus. Cette formulation est quasiment semblable à la formulation par variogramme. Deux différences sont cependant à noter :

- le terme correspondant à la pépite du variogramme n'est pas pris en compte. Ce terme avait toutefois été considéré comme négligeable par Lopez et al. (2014).
- l'expression de l'évolution temporelle telle que proposée par Lopez et al. (2014) impliquait que la composante aléatoire intervienne comme un terme multiplicatif des valeurs de la série. Cela signifie que le bruit introduit par pour représenter les erreurs liées à la mesure est d'autant plus fort que la valeur de concentration est forte. Ce choix avait été dicté par l'utilisation des variogrammes, mais il n'apparaît pas comme le plus pertinent au vu des sources d'erreur : les erreurs liées au processus d'échantillonnage sont a priori indépendantes des concentrations, et les erreurs analytiques sont maximales lorsque les concentrations mesurées sont proches des limites de quantification. En considérant que les concentrations ne varient que peu dans les évolutions théoriques générées (dans ce travail, aucune évolution à long terme est prise en compte), il est apparu préférable de considérer que l'incertitude analytique était fixée pour chaque évolution.

D'autre part, nous avons choisi dans ce rapport de ne pas considérer les évolutions dans lesquelles des tendances d'évolution monotones étaient présentes.

Finalement, les évolutions générées peuvent s'écrire sous la forme suivante :

$$\mathcal{L}(t) = A_0 \times 1 + \mathcal{F}_{cos}(t) + \mathcal{G}(m, \sigma) \quad \text{Équation 1}$$

avec :

$A_0$	Moyenne des valeurs de la chronique, fixée à 1
$\mathcal{F}_{cos}(t)$	Composante cyclique
$\mathcal{G}(m, \sigma)$	Composante gaussienne
$t$	Temps
$m$	Moyenne de la composante gaussienne, fixée à 0.
$\sigma$	Ecart-type de la composante gaussienne

Pour une évolution périodique à un seul cycle, la composante périodique s'écrit :

$$\mathcal{F}_{cos}(t) = A_{cycle} \frac{1}{2} \cos \frac{2\pi t + \Delta t}{période} \quad \text{Équation 2}$$

Pour une évolution périodique à deux cycles, la composante périodique s'écrit :

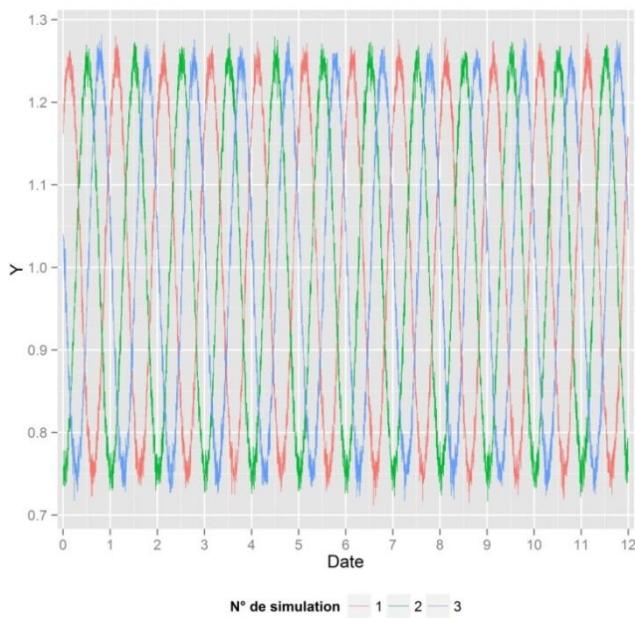
$$\mathcal{F}_{cos}(t) = \alpha A_{cycle} \frac{1}{2} \cos \frac{2\pi t + \Delta t_1}{P_1} + (1 - \alpha) A_{cycle} \frac{1}{2} \cos \frac{2\pi t + \Delta t_2}{P_2} \quad \text{Équation 3}$$

avec :

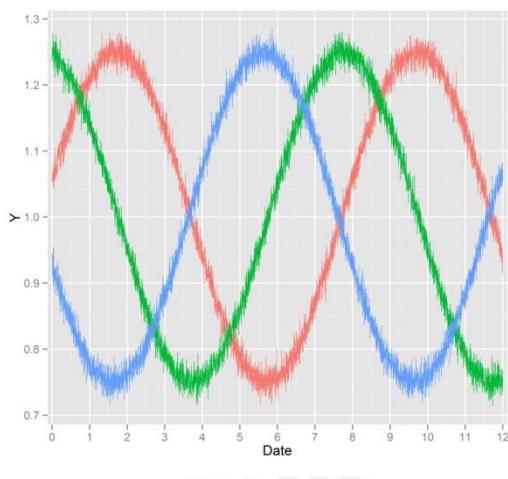
$\alpha$	part relative de chacun des deux signaux périodiques (cycle 1 et cycle 2), lorsque cela est applicable. Ce paramètre varie entre 0 et 1.
$A_{cycle}$	amplitude du cycle (ou somme des amplitudes des deux cycles s'il y a deux cycles)
$P_1$	période du premier cycle (cycle « annuel ») (an)
$P_2$	période du deuxième cycle (cycle « pluri-annuel ») (an)
$\Delta t_1$ et $\Delta t_2$	décalages temporels aléatoires pour les cycles 1 et 2 compris entre 0 et 1

Les décalages temporels sont des termes aléatoires indépendants qui permettent que l'évolution commence à un moment quelconque du cycle. Des exemples de chroniques correspondant à plusieurs simulations selon une même loi d'évolution temporelle sont présentés en [Illustration 3](#) ; les chroniques diffèrent seulement de ce décalage temporel.

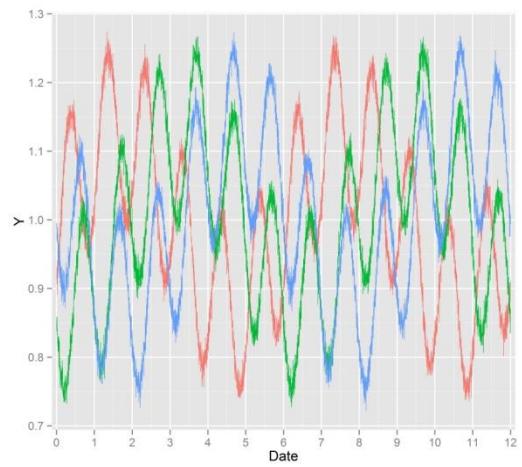
a)



b)



c)



**Illustration 3 : Exemple de séries temporelles générées suivant une même loi d'évolution avec des paramètres identiques, mettant en évidence le déphasage aléatoire sur la fonction périodique.** a) signal périodique annuel + bruit ; b) signal périodique pluriannuel (période 8 ans) + bruit ; c) signal périodique à deux cycles : annuel et pluriannuel (8 ans)+ bruit.

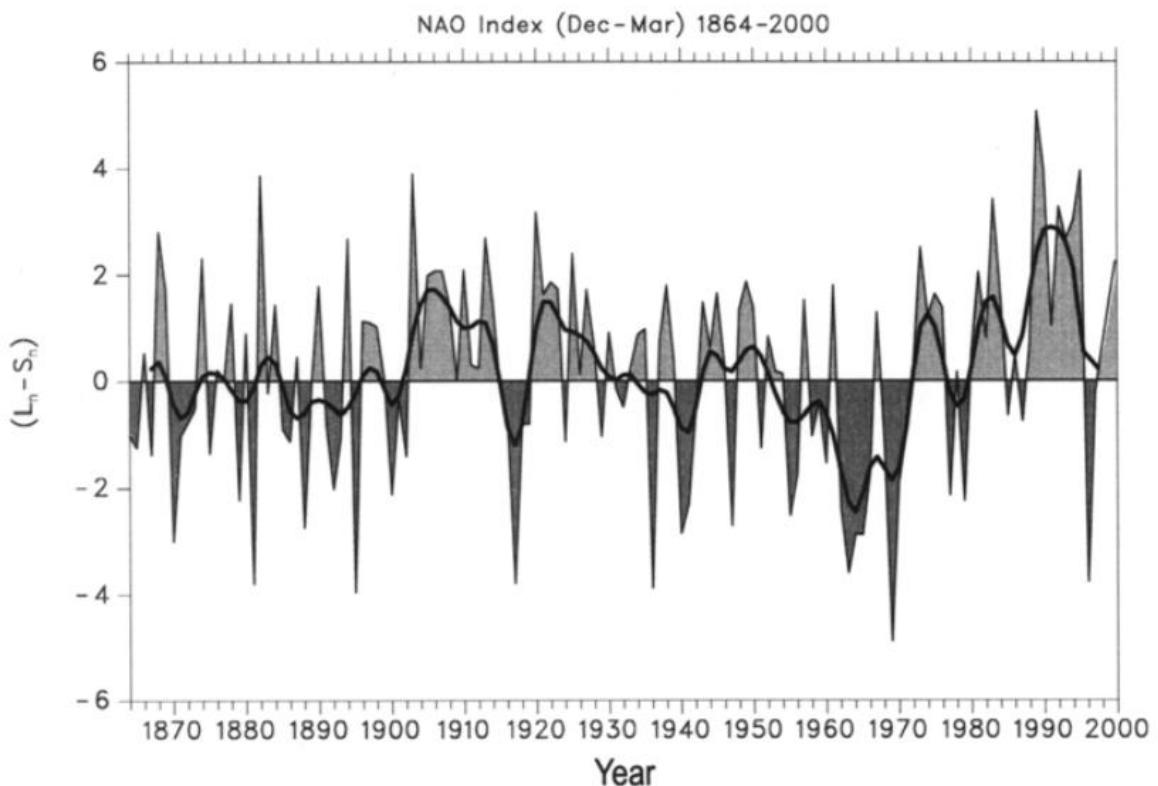
### 3.1.3. Gamme de paramètres investiguée

- **Période des variations pluri-annuelles**

Pour ce qui est de l'estimation de la période des cycles pluriannuels à considérer, l'oscillation Nord-Atlantique (NAO) est très largement documentée comme étant responsable de variations

climatiques d'échelle saisonnière à décennale (Walker & Bliss 1932; Hurrell 1995; Wanner et al. 2001). De manière simplifiée (Wanner et al. 2001), la NAO décrit l'oscillation, à large échelle, de l'atmosphère dans les régions nord atlantique (entre l'anticyclone subtropical de la région des Açores et les basses pressions subpolaires près de l'Islande). Cette variation est associée à des changements dans le régime des vents d'Ouest à la surface de l'Atlantique Nord (Hurrell 1995).

Ces variations ont pour conséquence de moduler les conditions climatiques en Europe et de conduire à des périodes plus ou moins chaudes en comparaison de la température moyenne, ou à des périodes plus ou moins sèches. Pour rendre compte des variations des descripteurs climatiques, un indice NAO peut-être défini. Plusieurs méthodologies de calcul de cet indice existent, assez peu différentes entre elles (un exemple est présenté sur l'[Illustration 4](#)).



*Illustration 4 : Indice NAO de 1864 à 2000 défini comme la différence de pressions entre Lisbonne et Stykkisholmur pour les mois de décembre à mars. Source : Wanner et al, 2001.*

Les périodes de temps de ces changements sont généralement de l'ordre de 2.5, 5-6 et 8 ans (Hurrell, 1995). Or les changements hydroclimatiques peuvent impacter l'évolution de la qualité des eaux souterraines par l'intermédiaire :

- de l'arrivée cyclique dans l'aquifère d'un volume plus ou moins important d'eaux de recharge à la chimie différente,
- du battement cyclique de la nappe à un niveau plus ou moins élevé dans la zone non saturée à la qualité d'eau stockée et aux conditions physicochimiques différentes.

Il apparaît donc pertinent de simuler des évolutions cycliques périodiques de la qualité des eaux souterraines en reprenant les périodes de temps des changements hydroclimatiques globaux. Ainsi, pour ce travail, les périodes retenues pour simuler les évolutions à cycles pluriannuels sont 3, 6 et 8 ans.

- **Amplitudes des cycles**

Les amplitudes de cycles (ou la somme des amplitudes des cycles dans le cas où l'évolution comporte deux cycles) testées sont les suivantes : 10, 20 et 50% de la valeur moyenne de la chronique

- **Poids des cycles**

Lorsque l'évolution comporte deux cycles, la part relative de l'influence d'un cycle par rapport à l'autre est variable. Trois cas de figure ont été testés dans lesquels le cycle 1 représente 10, 50 et 90% de l'amplitude du signal (et le cycle 2 à l'inverse 90, 50 et 10%).

- **Incertitude analytique**

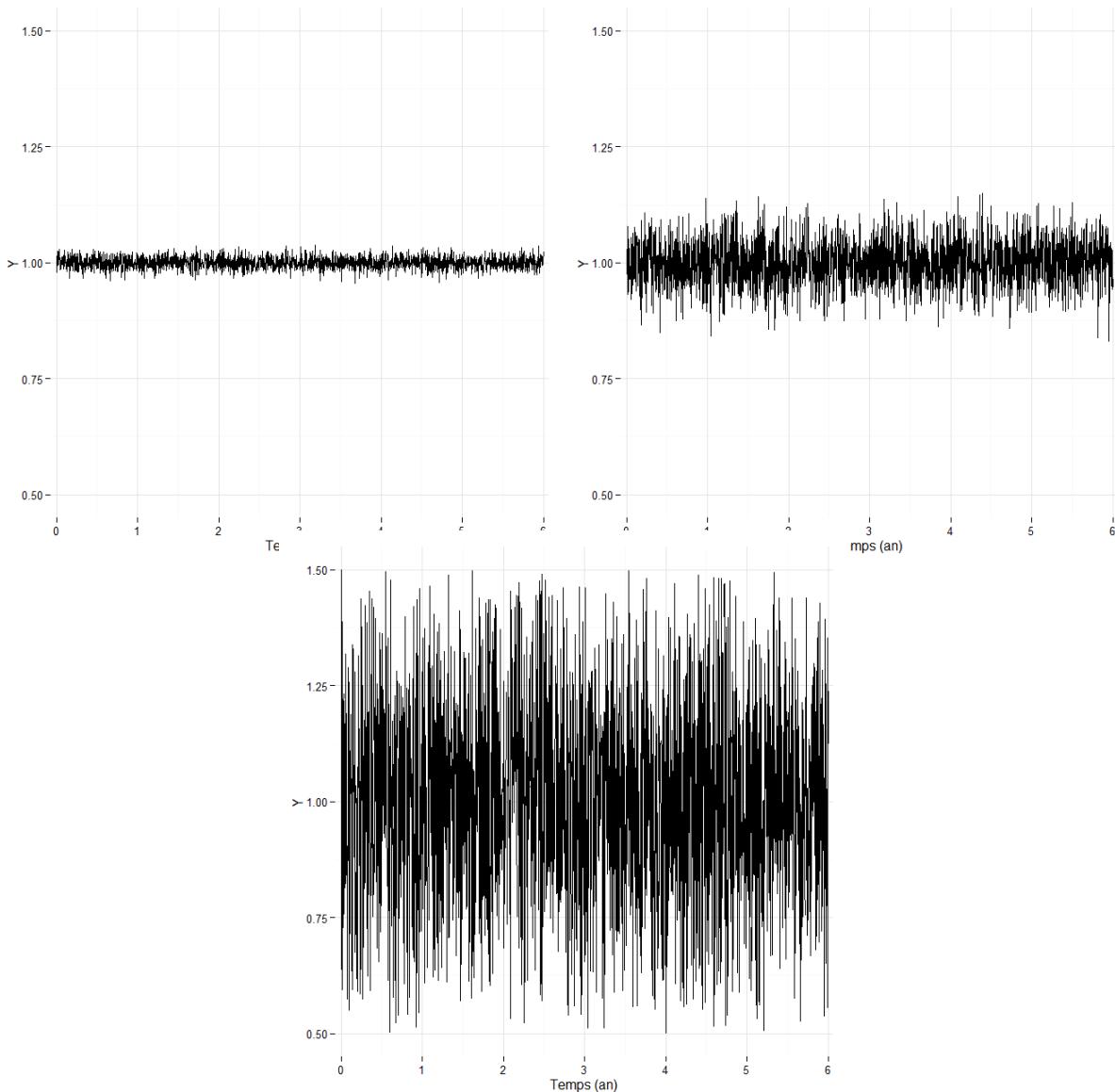
Différentes valeurs d'incertitudes analytiques ont été testées. Dans la suite du document, la notion d'incertitude renvoie à des incertitudes élargies ( $k=2$ ). Ces valeurs peuvent correspondre de façon simplifiée à différents types de paramètres et différentes concentrations (les valeurs d'incertitude mentionnées doivent être considérées comme des estimations minimales pour les paramètres citées) :

- +/- 2,5 % : représente, par exemple, le cas des nitrates à des concentrations de 25-50 mg/l
- +/- 10 % : représente, par exemple, le cas des métaux traces aux alentours de 1-10 µg/l
- +/- 30 % : représente, par exemple, le cas des pesticides classiques à des concentrations de l'ordre de 0,1 µg/l
- +/- 50 % (de 40 à plus de 80 %) : représente, par exemple, le cas des émergents ou des substances plus classiques mais à des concentrations proches des LQ.

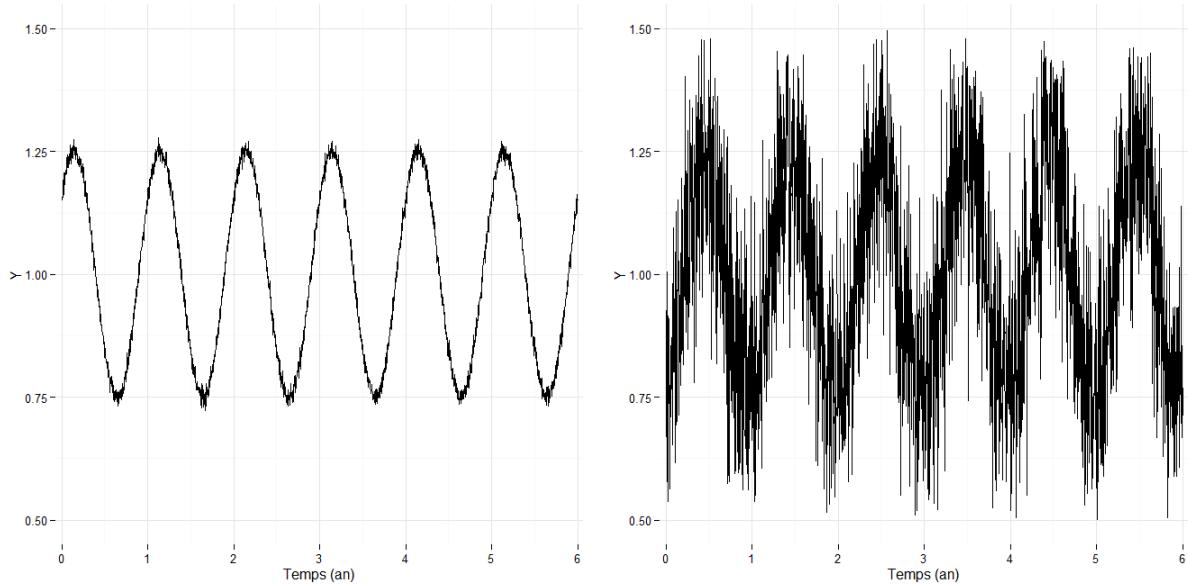
### 3.1.4. Exemples d'évolutions théoriques

Les [Illustration 5](#), [l'Illustration 6](#) et [l'Illustration 7](#) présentent quelques exemples d'évolutions générées selon la méthode exposée au paragraphe 3.1.2 et les paramètres détaillés au paragraphe 3.1.3.

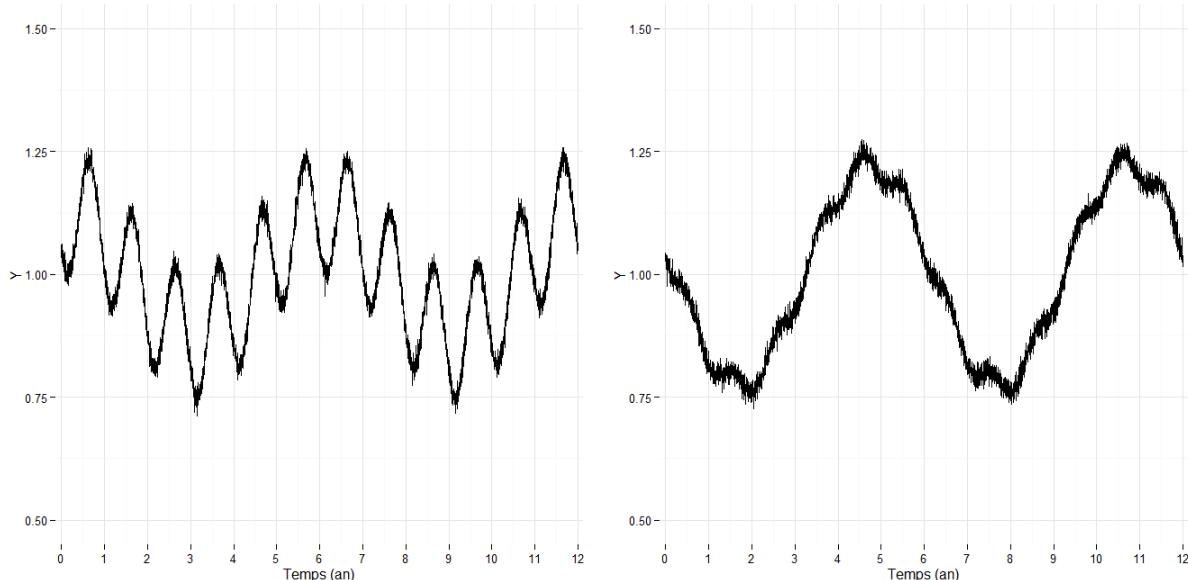
## Simulation des incertitudes associées aux calculs des Mma et Fd



*Illustration 5 : Exemples d'évolutions de type monotone affectées d'une incertitude analytique de 2.5 %, 10 % et 50 %.*



*Illustration 6 : Exemples d'évolutions à 1 cycle de longueur 1 an, et d'amplitude égale à 50 % de la valeur moyenne, pour des incertitudes analytiques de 2.5 % et 30 .*



*Illustration 7 : Exemples d'évolutions à deux cycles de longueurs 1 an et 6 ans, et d'amplitude égale à 50 %, affectées d'une incertitude analytique de 2.5 %. A gauche, la part du premier cycle (1an) dans le signal représente 50 %, et celle du deuxième cycle (6 ans) représente 50 %. A droite, la part du premier cycle représente 10 %, et la part du second cycle 90 %.*

## 3.2. GENERATION DES CHRONIQUES VIRTUELLES

### 3.2.1. Méthode

Des chroniques virtuelles de suivi sont générées à partir des évolutions théoriques de manière à créer un ensemble de séries temporelles similaires aux données dont disposent les gestionnaires.

De façon à reproduire les cas les plus généralement observés d'après les données de suivi existantes (des exemples de chroniques hydrochimiques tirées de la base ADES peuvent être trouvés dans les travaux de (Lopez et al. 2011)), une typologie de prélèvement a été établie suivant trois paramètres : la fréquence de prélèvement, la longueur du suivi et la régularité des prélèvements. Ces trois paramètres sont des paramètres sur lesquels le gestionnaire est en capacité d'agir.

### 3.2.2. Gamme des paramètres investiguée

- **Fréquence de prélèvement**

Les fréquences de prélèvement retenues sont reprises des travaux de (Lopez et al. 2014) et sont celles préconisées dans la directive 2000/60/CE pour les suivis de la chimie des eaux souterraines (12 fois par an pour un aquifère karstique, 1 fois par an seulement sinon) et pour la surveillance des points de captage d'eau potable – 4 fois par an pour une population desservie < à 10 000 habitants et 12 fois par an pour plus de 30 000 usagers (art. 1.3.5).

Cinq valeurs sont retenues dans ce travail : 1 tous les 2 ans, et 1, 2, 4 et 12 par an.

- **Longueur du suivi**

Les cycles de rapportage tels que défini dans la DCE sont d'une durée de six ans. Un rapportage au milieu du cycle peut être demandé.

Les périodes relative aux calculs des moyennes de moyennes ont donc été fixées dans ce travail à 3, 6, 12 et 18 ans, c'est-à-dire de la moitié d'un cycle jusqu'à 3 cycles de rapportage.

- **Calendriers**

Le calendrier de prélèvement décrit l'agencement dans le temps des différentes dates de prélèvements. Différentes modalités ont été envisagées et détaillées dans les travaux de (Lopez et al. 2011) sous la dénomination « calendrier ». Elles sont reprises et détaillées sous les 4 dénominations suivantes ([Illustration 8](#)):

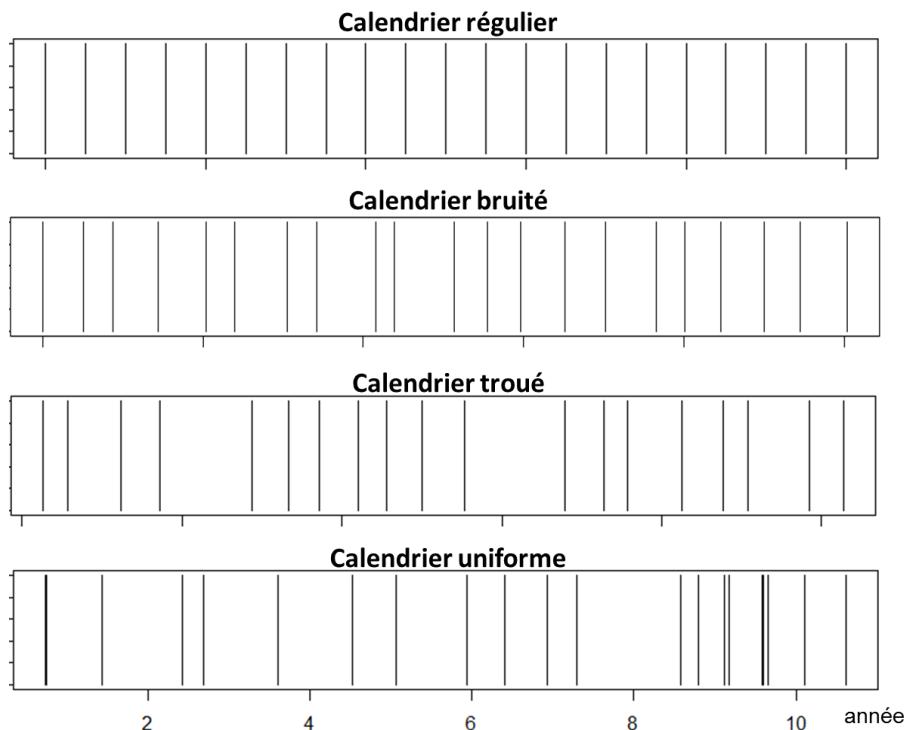
- i. calendrier régulier : prélèvements avec un intervalle de temps strictement constant au cours du temps ;
- ii. calendrier bruité : prélèvements avec un intervalle plus ou moins constant, c'est-à-dire qu'un certain degré de liberté existe sur l'intervalle de temps (quelques jours à quelques mois) ;
- iii. calendrier troué : prélèvements avec un intervalle de temps constant ou quasi constant (cas i. ou ii.) avec des périodes lacunaires ;
- iv. calendrier uniforme : prélèvements avec un intervalle complètement irrégulier.

Le calendrier régulier est le cas idéal, dans lequel les données sont équitablement réparties dans l'année. Ce cas est abordé dans cette étude dans la mesure où il représente, dans la plupart des cas, la modalité la meilleure pour espérer qualifier les données hydrochimiques avec précision.

Un cas plus réaliste est décrit par le calendrier de type bruité. Dans ce cas, les données sont prélevées, en principe, à intervalle régulier, mais un décalage de quelques jours à quelques mois est possible. Ce phénomène est décrit par un bruit de type gaussien sur la date de prélèvement tel que l'incertitude élargie sur la date de prélèvement soit égale à 1/3 de l'écart entre 2 dates (par exemple pour un prélèvement ayant lieu tous les 6 mois, l'incertitude élargie est de 2 mois).

Si quelques valeurs sont manquantes dans la chronique le calendrier est décrit comme « troué ». C'est le cas d'un prélèvement qui, en principe, est réalisé à intervalle presque fixe, mais qui peut, parfois, ne pas être réalisé. C'est le cas qui représente sans doute le mieux ce qui est effectivement réalisé dans les plan de surveillance. Pour ne pas multiplier le nombre de cas possibles, la proportion de données lacunaires dans la série temporelle a été fixée à 10%, ce qui correspond, par exemple à environ un mois manquant pour un suivi mensuel sur une année.

Enfin, une distribution aléatoire des données dans le temps peut être décrite par le calendrier uniforme qui représente, ici, l'analogie avec les points de surveillance ayant appartenu historiquement, à un ou plusieurs réseaux de surveillance dont les fréquences de prélèvement étaient différentes expliquant des données irrégulièrement espacées dans le temps, parfois séparées par des périodes d'absence de prélèvements.



*Illustration 8 : Exemples de répartition dans le temps des dates de prélèvement suivant les différentes modalités de calendriers proposées pour 2 prélèvements par an en moyenne.*

### 3.3. CALCUL DES METRIQUES THEORIQUES DES CHRONIQUES

- **Mma**

Par construction, la moyenne des évolutions théoriques simulées a été fixée à 1. En effet :

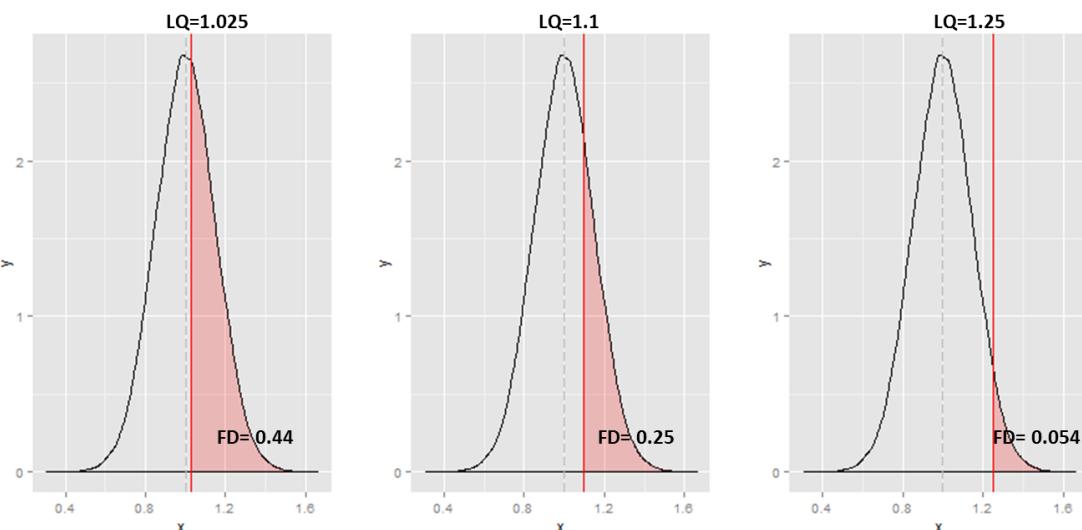
- pour un signal monotone, la composante cyclique est nulle, et l'évolution théorique n'est plus fonction que de deux termes : la valeur initiale et la composante gaussienne. La valeur initiale est fixée à 1 pour l'ensemble des simulations, et la composante gaussienne,  $\mathcal{G} m, \sigma$ , varie autour de la valeur moyenne (m) fixée à 0,
- pour un signal périodique, la composante  $\mathcal{F}_{cos}$  est de moyenne nulle.
- la moyenne des deux termes (gaussien et périodique) est donc de 1.

- **Fréquence de dépassement de la valeur seuil**

La valeur seuil a été fixée dans ce travail à 1, 1.025, 1.1, 1.25 et 1.5.

- *Cas d'une évolution monotone*

Dans le cas d'une évolution monotone, décrite par une distribution purement gaussienne des données, le calcul de la fréquence de dépassement de la valeur seuil revient à calculer la probabilité de trouver une valeur supérieure à une valeur cible (voir [Illustration 9](#)), connaissant la moyenne et l'écart type des données.



*Illustration 9 : Densité de probabilité d'une loi normale de moyenne 1 et d'écart-type 0.15 (c'est-à-dire une incertitude élargie égale à 0.3). La fréquence de dépassement de seuil est calculée pour 3 valeurs seuil : 1.025, 1,1 et 1,25 : le seuil est représenté par la droite verticale de couleur rouge, et la zone colorée en rouge représente la probabilité d'avoir une valeur supérieure au seuil.*

- *Cas d'une évolution cyclique*

Dans le cas des évolutions périodiques à un ou deux cycles, on peut calculer à chaque pas de temps la probabilité que la concentration échantillonnée dépasse la valeur seuil :

$$p(t) = \Phi(C(t), m = VS, sd = ia/2)$$

avec

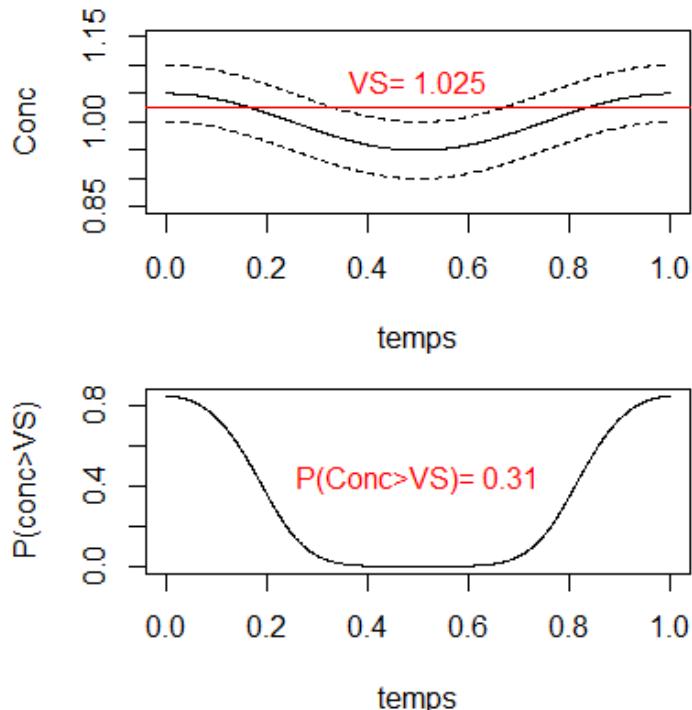
$\Phi(x, m, sd)$  fonction de répartition de la loi normale de moyenne  $m$  et d'écart-type  $sd$

$C(t)$  concentration de l'élément chimique égale à  $A_0 \times 1 + \mathcal{F}_{cos}(t)$

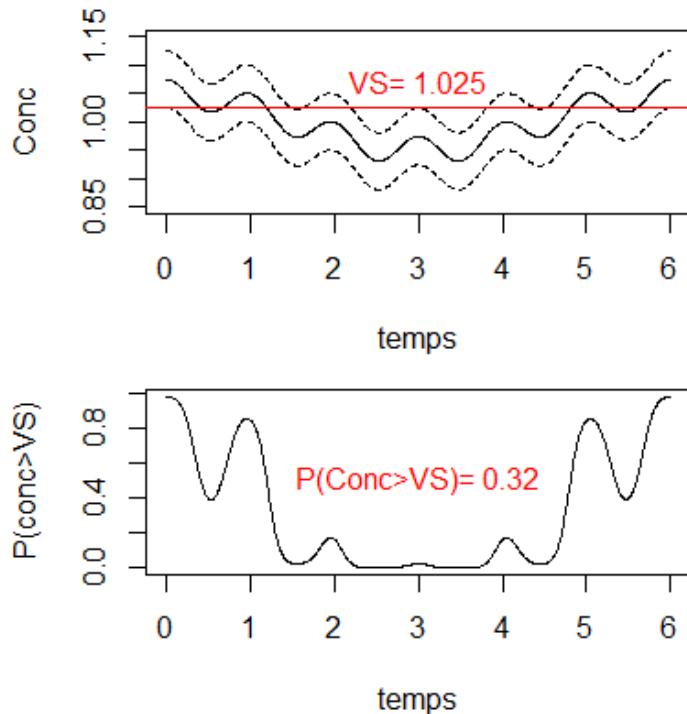
$VS$  valeur seuil

$ia$  incertitude analytique élargie

L'[Illustration 10](#) et l'[Illustration 11](#) présentent deux exemples de calcul de la probabilité de dépassement de seuil pour des évolutions à 1 et 2 cycles.



*Illustration 10 : Illustration du calcul de la fréquence de dépassement de seuil pour une évolution cyclique. La valeur seuil a été fixée à 1,025. L'incertitude analytique est égale à 0,05. La figure du haut représente l'évolution des concentrations sur un cycle (les lignes pointillées représentent l'enveloppe dans laquelle se situent 95% des valeurs). En bas, évolution de la probabilité de dépassement de la valeur seuil. La probabilité moyenne sur un cycle entier est 0,31.*

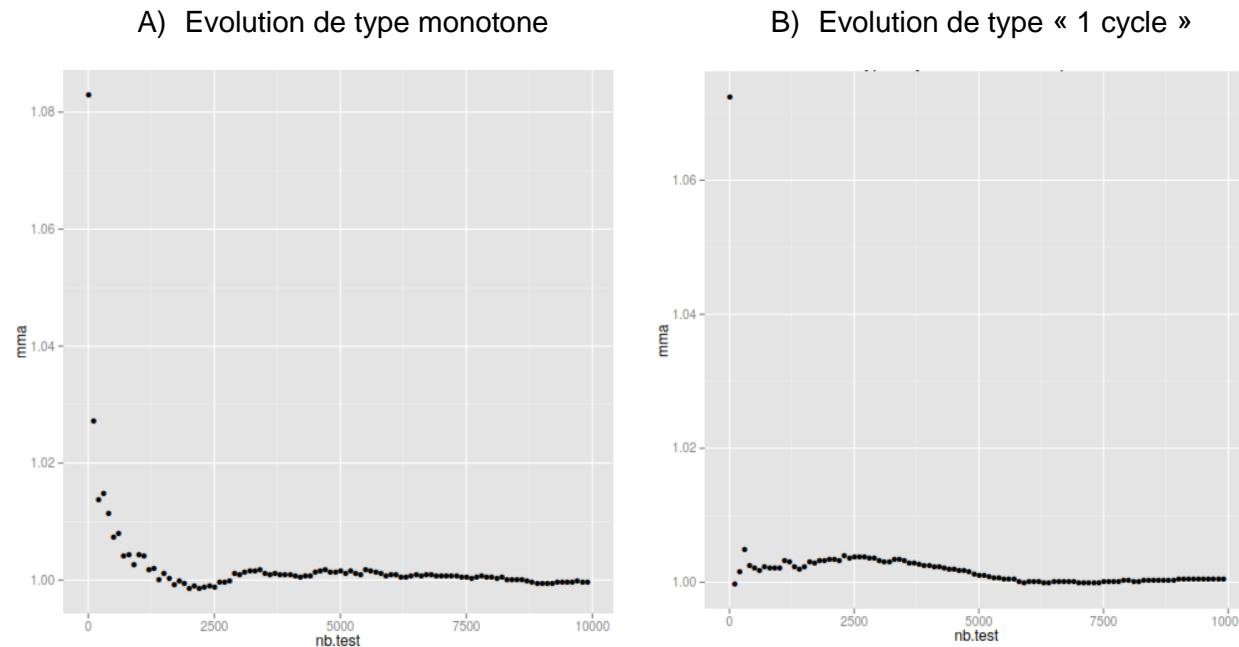


*Illustration 11 : Illustration du calcul de la fréquence de dépassement de seuil pour une évolution cyclique à double cycle. La valeur seuil et l'incertitude analytique sont identiques à l'[Illustration 10](#). La probabilité de dépassement de la valeur seuil dans ce cas est de 0,32.*

### 3.4. CALCULS DES METRIQUES SUR LES CHRONIQUES GENEREES

#### 3.4.1. Nombre de tests, normalité des données

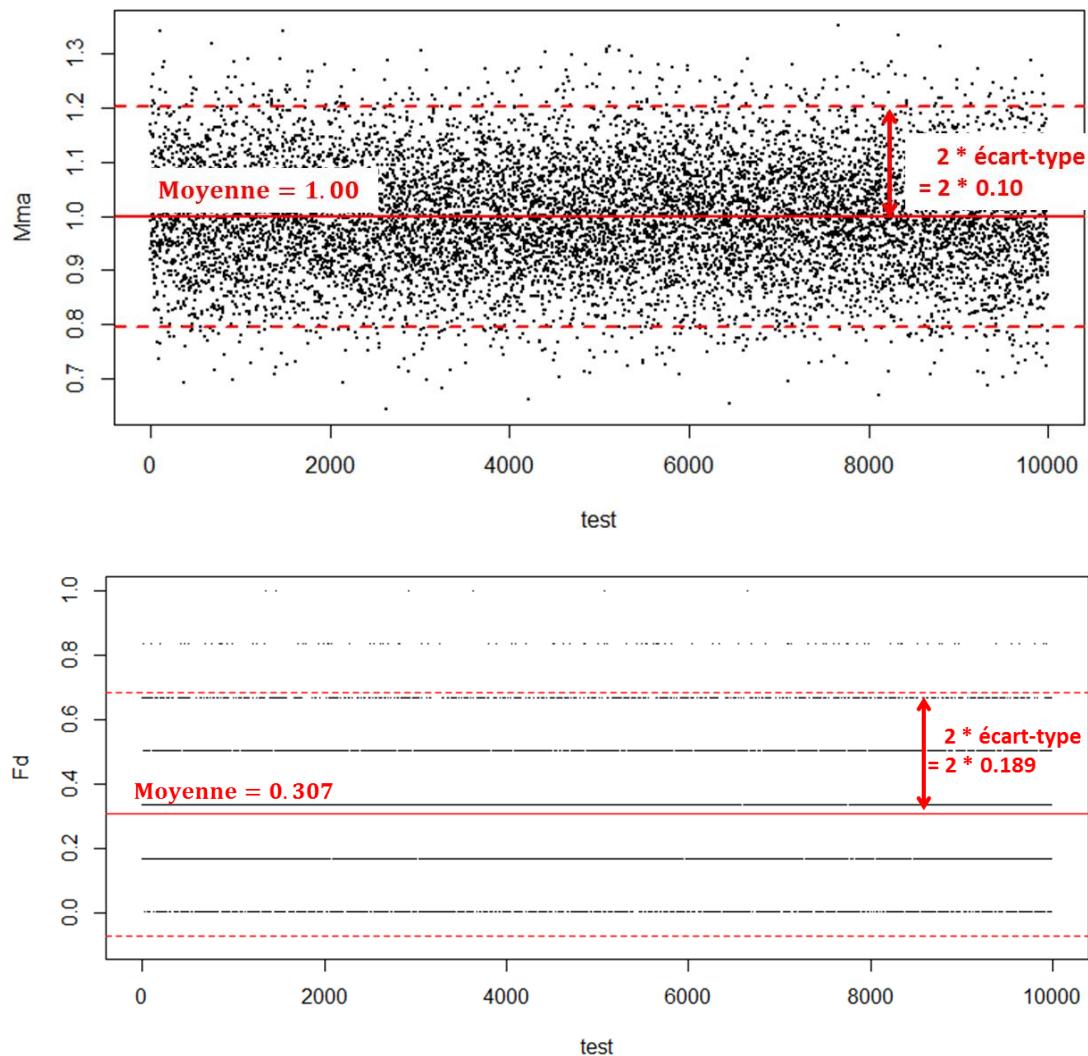
De façon analogue à ce qui avait été réalisé par Lopez et al. 2014, 10 000 tests ont été réalisés pour chaque tirage de paramètres. Ce nombre de tests permet de s'assurer que les valeurs des indicateurs retenues pour décrire les séries (Mma et Fd) convergent ([Illustration 12](#)).



*Illustration 12 : Deux exemples de convergence de la Mma moyenne calculée sur les évolutions temporelles simulées en fonction du nombre de tests réalisés A) cas d'une évolution monotone, de longueur 3 ans, 12 prélèvements par an, incertitude analytique = 50 %, calendrier régulier; B) cas d'une évolution à 1 cycle (période 1 an, amplitude 0.2), de longueur 18 ans, 1 prélèvement par an, incertitude analytique 50%.*

#### 3.4.2. Calcul de la distribution des moyennes des moyennes annuelles et des fréquences de dépassement de valeur seuil

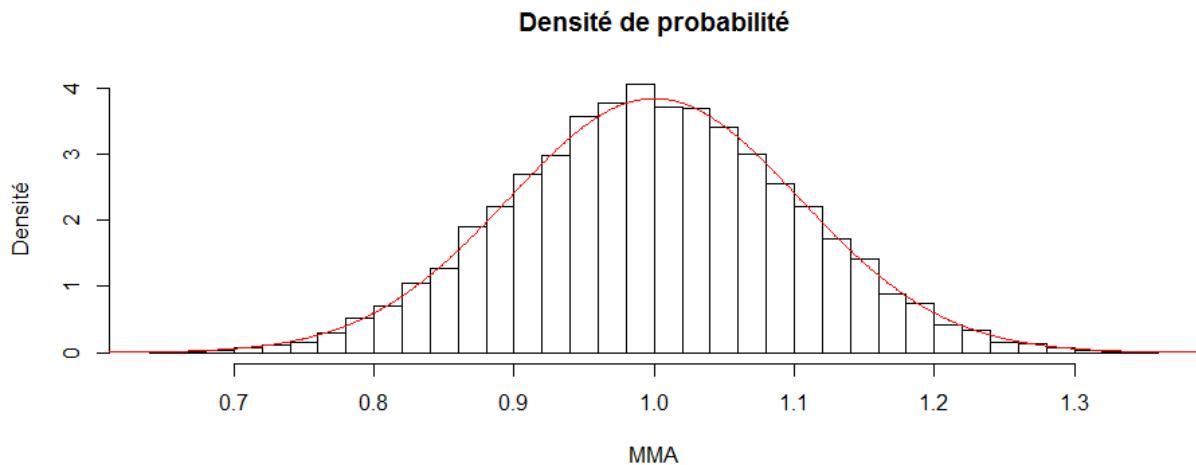
Pour une évolution, un calendrier et une incertitude analytique donnés, pour les 10 000 chroniques créées, on calcule les valeurs de Mma et de Fd. Un exemple de résultat est proposé sur l'[Illustration 13](#).



*Illustration 13 : Valeurs de la  $M_{ma}$  et de la fréquence de dépassement de la valeur seuil calculées pour 10 000 chroniques et moyennes et écart-type de ces métriques. Dans cet exemple, l'évolution est monotone, l'incertitude analytique est de 50%, et la chronique est composée d'un prélèvement par an pendant 6 ans ; la valeur seuil est fixée à 1,5.*

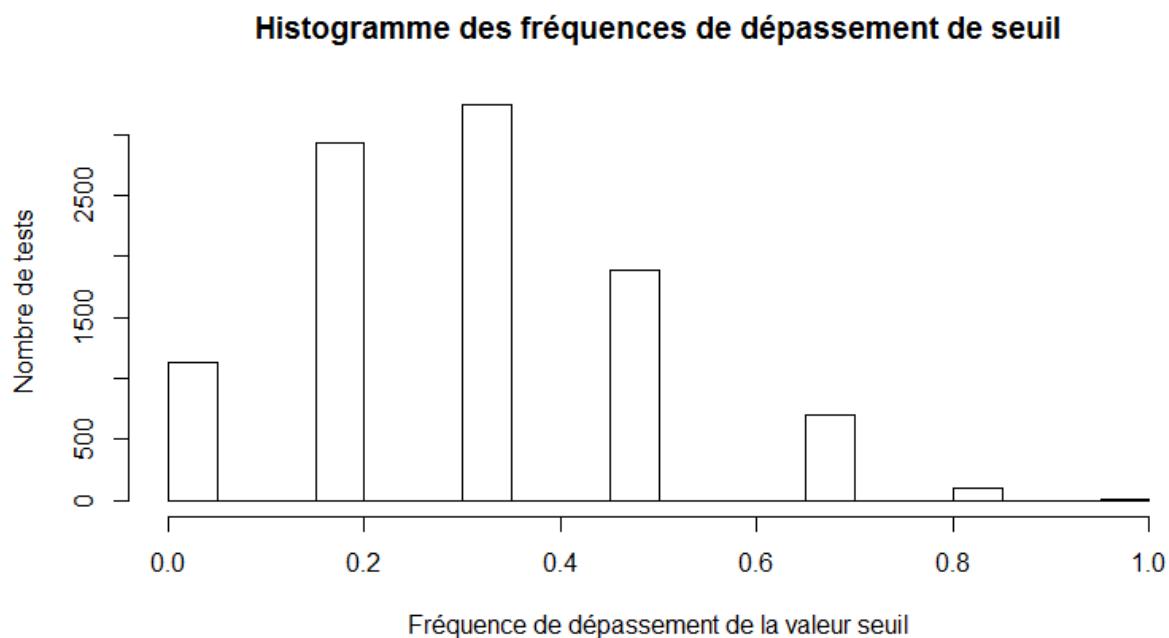
Les valeurs de  $M_{ma}$  sont aléatoires, centrées autour de 1 et l'étalement des valeurs est fonction de l'incertitude analytique.

Pour étudier la distribution de cette métrique, il est possible de représenter la densité de probabilité ([Illustration 14](#)). Les données expérimentales (représentées par l'histogramme de couleur noire) décrivent une loi normale (représentée en rouge sur la figure)



*Illustration 14 : Densité de probabilité de la MMA . Sur un tel graphique, on peut lire la probabilité que la variable soit comprise entre deux valeurs  $a$  et  $b$ , comme l'aire sous la courbe dans l'intervalle  $[a,b]$ . En noir, l'histogramme représente les données expérimentales tandis que la courbe rouge représente la loi normale de moyenne 1 et d'écart-type 0,10 , l'écart-type calculé sur les 10 000 chroniques.*

La fréquence de dépassement de la valeur seuil prend des valeurs discrètes qui dépendent du nombre de mesures dans la chronique. Dans notre exemple où la série comporte uniquement 6 mesures, cette fréquence ne peut prendre que 7 valeurs :  $k/6$  où  $k$  varie entre 0 et 6. L'[Illustration 15](#) représente l'histogramme des fréquences de dépassement de la valeur seuil. |



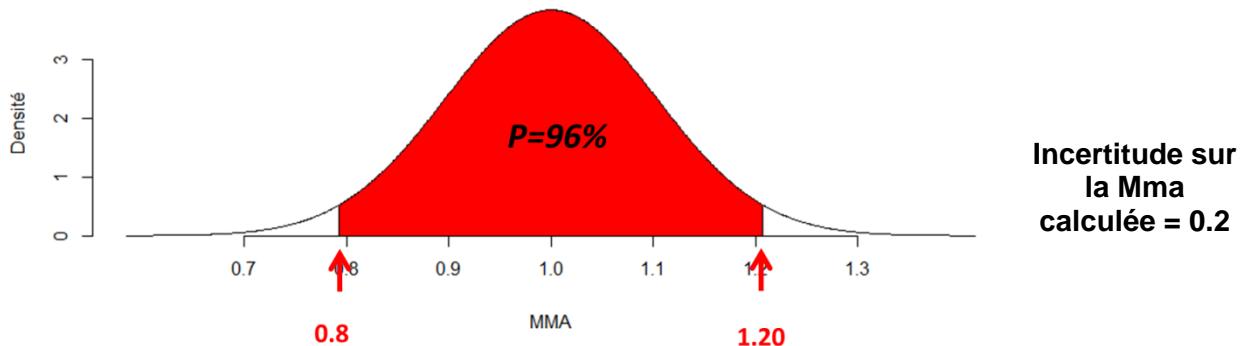
*Illustration 15 : Histogramme des fréquences de dépassement de la valeur seuil. Les paramètres des chroniques sont les mêmes que ceux de l'exemple de l'[Illustration 13](#). La valeur seuil est considérée égale à 1,125.* |

### 3.4.3. Ecart aux valeurs vraies

- **Mma**

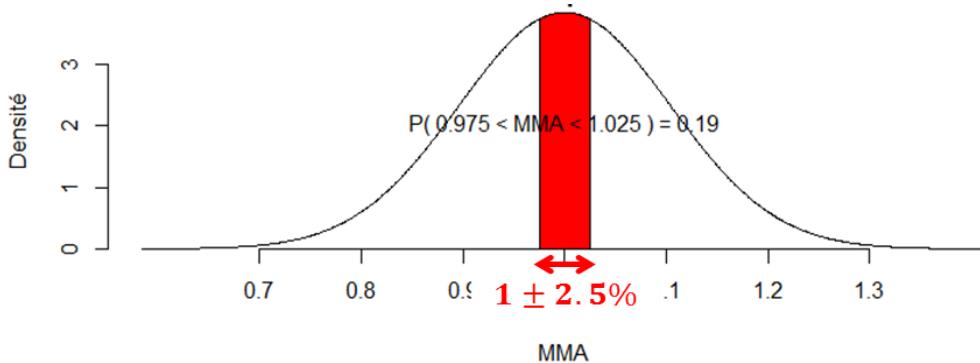
Pour appréhender les écarts des Mma des chroniques à la valeur théorique, deux valeurs sont calculées :

- l'incertitude élargie sur la Mma, c'est-à-dire l'intervalle dans lequel 95% des valeurs de Mma se trouvent, comme présenté sur l'[Illustration 16](#)



*Illustration 16 : Densité de probabilité des Mma calculées sur 10 000 chroniques correspondant à l'exemple précédent (moyenne =1, écart-type= 0.10) et calcul des bornes permettant d'avoir 95% des valeurs de la Mma entre ces bornes*

- la probabilité de trouver la Mma à 5% près, comme figuré sur l'[Illustration 17](#).



*Illustration 17 : Densité de probabilité des Mma calculées sur 10 000 chroniques correspondant à l'exemple précédent (moyenne =1 , écart-type= 0,104). Est figuré en rouge l'intervalle 0,975-1,025 correspondant à la Mma vraie à 5% près. La probabilité de trouver la Mma vraie est calculée comme l'aire de ce polygone, elle est égale à 19%.*

Par la suite dans ce travail, les résultats seront présentés sous la forme de l'incertitude élargie sur la Mma. Cependant, dans l'interface qui permet d'avoir accès à l'ensemble des résultats (cf paragraphe 4.3), les deux valeurs (incertitude et probabilité de trouver la Mma vraie) peuvent être affichées.

- **Fréquence de dépassement de la valeur seuil**

L'incertitude sur la fréquence de dépassement de la valeur seuil sera donnée comme la largeur de l'intervalle de confiance autour de la fréquence de dépassement moyenne calculée, c'est-à-dire la quantité :  $Q_{97.5}(Fd) - Q_{2.5} Fd$  , où  $Q_x Fd$  est le  $x$ ème percentile de la fréquence de

dépassement de seuil (95% des 10.000 valeurs de Fd se trouvent dans cet intervalle). La largeur de l'intervalle de confiance représente donc la dispersion des fréquences de dépassement de la valeur seuil calculées pour les chroniques autour de la valeur vraie de l'évolution. Ainsi, plus la largeur de l'intervalle de confiance est grande, plus l'incertitude associée au calcul de la fréquence de dépassement est importante. A l'inverse, plus la largeur de l'intervalle de confiance est étroite, moins l'incertitude associée au calcul de la fréquence de dépassement de la valeur seuil est grande. On cherchera donc à minimiser autant que possible la largeur de l'intervalle de confiance, une incertitude nulle correspondant à une largeur d'intervalle de confiance nulle.

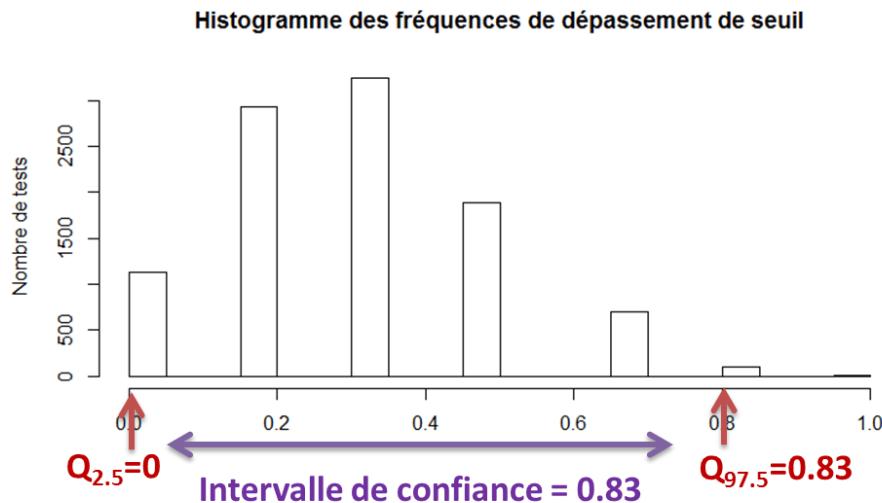


Illustration 18 : Histogramme des fréquences de dépassement de seuil et définition de l'intervalle de confiance. Les paramètres sont les mêmes que sur l'[Illustration 13](#) et l'[Illustration 12](#) (évolution monotone, incertitude analytique = 50%, longueur de la chronique = 6 ans, fréquence = 1/an, valeur seuil = 1,125).

Dans les cas les plus simples, il est possible de connaître l'incertitude associée au calcul de Mma et de la fréquence de dépassement de seuil de façon exacte en utilisant des formules analytiques. Une validation de quelques résultats avec ces formules analytiques est présentée en [Annexe 1](#).



## 4. Résultats

### 4.1. TESTS REALISES

L'ensemble des cas possibles investigués représente un jeu de données particulièrement fourni. Au total, 17 088 combinaisons des paramètres décrivant l'évolution et le calendrier d'échantillonnage ont été étudiées. L'ensemble des valeurs que peuvent prendre les différents paramètres sont rappelées dans l'[Illustration 19](#).

Nom Variable	Type de variable	Description	Valeurs possibles	Unités
type	Texte	Description du type d'évolution de la chronique	monotone, 1cycle, 2cycles	-
$A_0$	entier	moyenne de la chronique	fixée à 1	[c]
freq_ts	entier	Fréquence d'échantillonnage	0.5, 1, 2, 4, 12	an-1
long_ts	entier	Longueur de la série temporelle générée	3, 6, 12, 18	an
ia=0.1	entier	Ensemble des incertitudes (environnementales et analytiques)	2.5, 10, 30, 50	% de [c]
$A_{cycle}$	entier	Amplitude du cycle (ou somme des amplitudes des cycles)	10, 20, 50	%
$\alpha$	entier	Part relative du cycle annuel par rapport au cycle pluriannuel	0.1, 0.5, 0.9	
période	entier	Longueur de la période du cycle annuel	1, 10	an
période2	entier	Longueur de la période du cycle pluriannuel	3, 6, 8, 12	an

Illustration 19 : Gamme des paramètres décrivant les chroniques générées

La lecture des résultats peut être faite au regard d'intérêts différents suivant le public à qui ils s'adressent. Certains seront, par exemple, intéressé par le fait d'optimiser la fréquence d'échantillonnage dans un contexte d'évolution déjà bien connu quand d'autres voudront vérifier, à l'inverse, l'incertitude sur la Mma et la Fd, à calendrier et fréquence de prélèvement fixés. Les clés de lecture sont en réalité assez nombreuses, et il serait difficile de présenter ici l'ensemble des possibles.

Deux approches ont été retenues. La première (au paragraphe 4.2) est la présentation des principaux résultats sous forme de graphiques permettant de représenter l'évolution des incertitudes sur les métriques Mma et Fd selon plusieurs critères (démarche analogue à celle proposée par Lopez et al. (2014)).

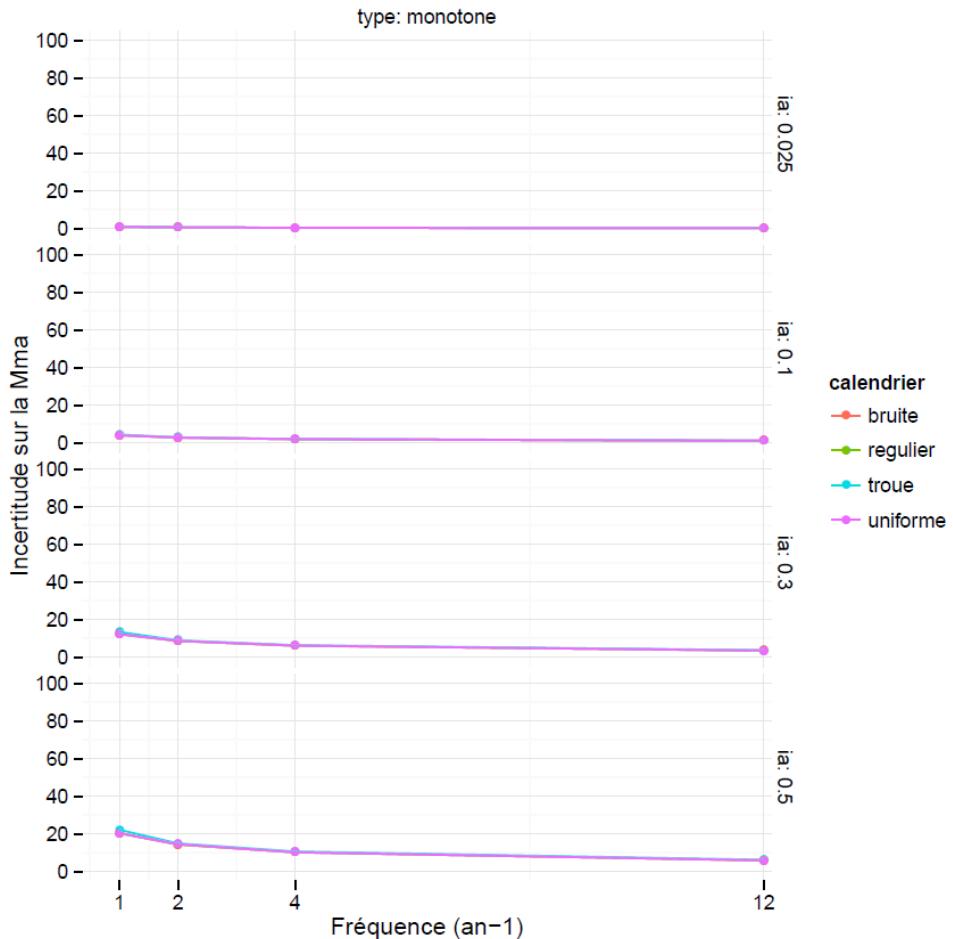
La seconde approche permet la consultation de la totalité des résultats ; Il s'agit d'une interface interactive, baptisée « Hypst-R ». Elle est présentée au §4.3.

## 4.2. ABAQUES

### 4.2.1. Cas des évolutions monotones

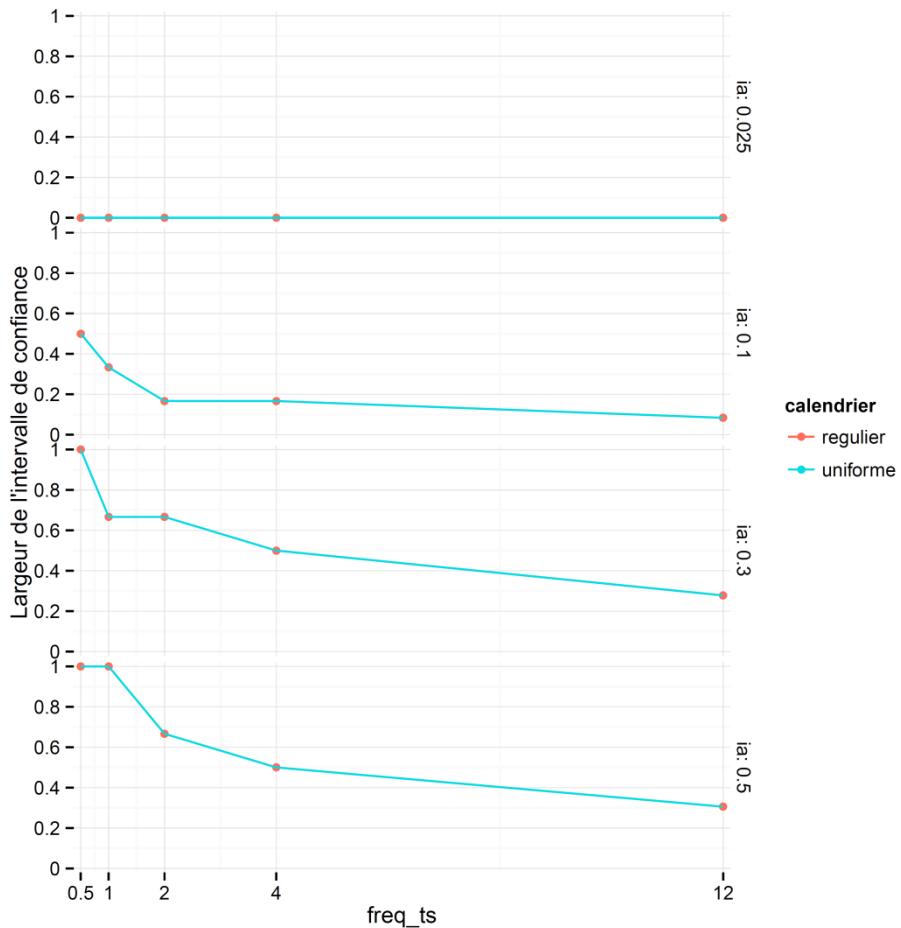
L'analyse des résultats obtenus pour les évolutions monotones ([Illustration 20](#)) montrent que l'incertitude sur la Mma est relativement faible par rapport à l'incertitude analytique. Par exemple, dans le cas, d'une incertitude analytique de 50%, l'incertitude sur la MMA dans une chronique ne disposant que de 6 valeurs est de 20% environ.

Par ailleurs, l'incertitude diminue fortement avec le nombre de valeurs (l'approche théorique montre que cette incertitude diminue en  $\frac{1}{\sqrt{n}}$ , voir en [Annexe 1](#)), par exemple lorsque la fréquence augmente, ou la longueur du suivi est la plus grande. Comme attendu, le type de calendrier n'a quasiment pas d'influence sur les résultats dans le cas d'évolutions monotones ; le calendrier trouvé montre des résultats légèrement différents, ce qui est cohérent avec le fait que dans l'échantillonnage trouvé, les données sont 10% moins nombreuses (voir définition des calendriers au §3.2.2).



*Illustration 20 : Incertitude sur la Mma (%) en fonction de la fréquence de prélèvement dans le cas d'une évolution monotone, pour des chroniques de longueur 6 ans. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5, 10, 30 et 50 %.*

Sur l'[Illustration 21](#) est représenté l'intervalle de confiance sur la fréquence de dépassement de seuil pour des chroniques non cycliques de longueur 3 ans et pour lesquelles la valeur seuil est égale 1.1 fois la moyenne. La largeur des intervalles de confiance autour de la Fd est très importante surtout quand le nombre de valeurs est faible et que l'incertitude analytique est forte.



*Illustration 21 : Largeur de l'intervalle de confiance autour de la fréquence de dépassement de seuil en fonction de la fréquence de prélèvement dans le cas d'une évolution monotone, pour des chroniques de longueur 3 ans et pour une valeur seuil égale à 1.1 fois la moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2.5, 10, 30 et 50 %..*

#### 4.2.2. Cas des évolutions cycliques périodiques

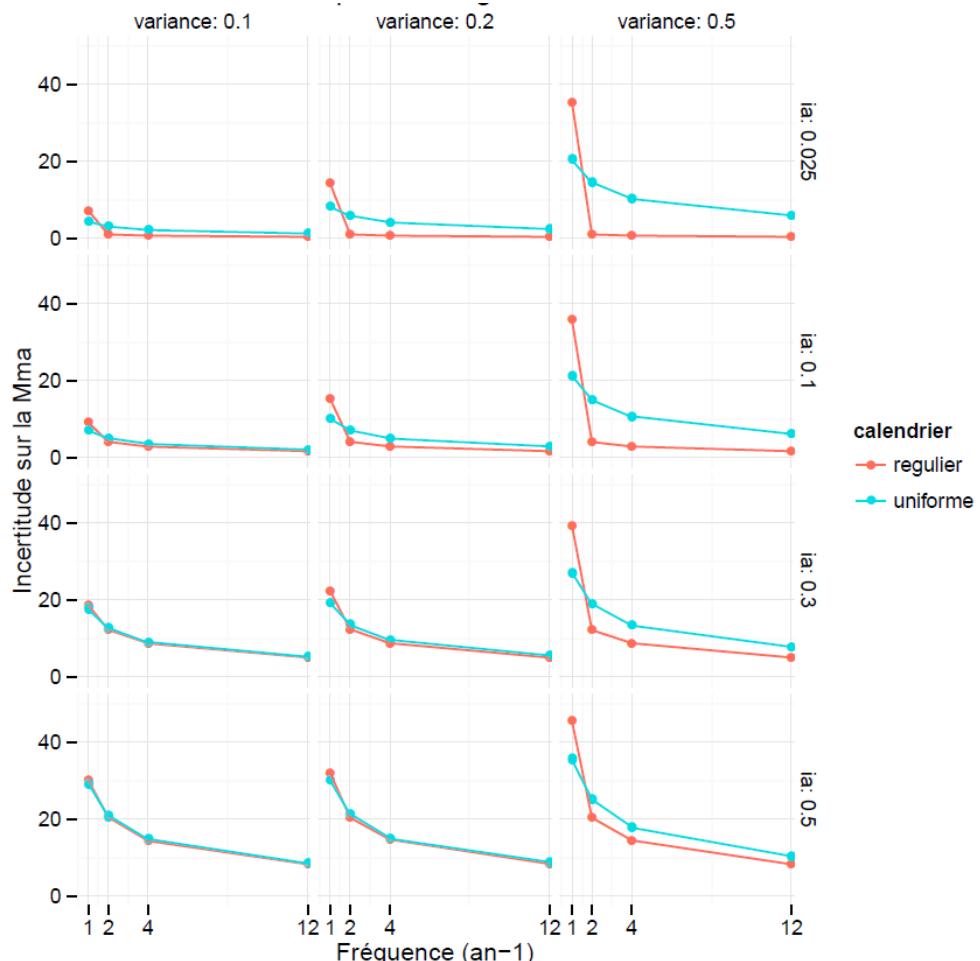
- **Cas des évolutions à un cycle**

Dans le cas d'une évolution cyclique à un seul cycle, l'incertitude sur la Mma ([illustration 22](#)) diminue de façon logique selon que l'incertitude analytique diminue et que l'amplitude des cycles diminue.

L'effet de la répartition des prélèvements dans le temps (le type de calendrier) est, dans le cas des évolutions cycliques, bien visible. Dans le cas où la fréquence de prélèvement est de 1/an, les prélèvements réalisés selon un calendrier régulier conduisent à des incertitudes sur les Mma toujours supérieures aux autres calendriers. En effet, le fait de prélever systématiquement à la même période dans le cycle, conduit à biaiser le calcul de la moyenne. Inversement si les prélèvements ont lieu un nombre pair de fois dans l'année, le cas du calendrier régulier est très favorisé.

Le nombre de données total dans la chronique est, de façon logique, un facteur déterminant pour augmenter les chances de trouver la Mma vraie. Pour une longueur de chronique donnée, les incertitudes diminuent en effet logiquement à mesure que la fréquence de mesure

augmente. De même pour des fréquences de prélèvement de 4 échantillons par an, l'incertitude sur la Mma est toujours inférieure à 10 % lorsque l'incertitude analytique est inférieure à 10%, y compris lorsque l'amplitude du signal périodique est forte. Les conclusions sont les mêmes si la longueur de la chronique augmente ([Illustration 23](#) : longueur de chronique de 6 ans, [Illustration 24](#) : longueur de chronique de 18 ans).



*Illustration 22 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique annuelle, pour des chroniques de longueur 3 ans. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différents cas d'amplitude du signal périodique (en part de la valeur moyenne).*

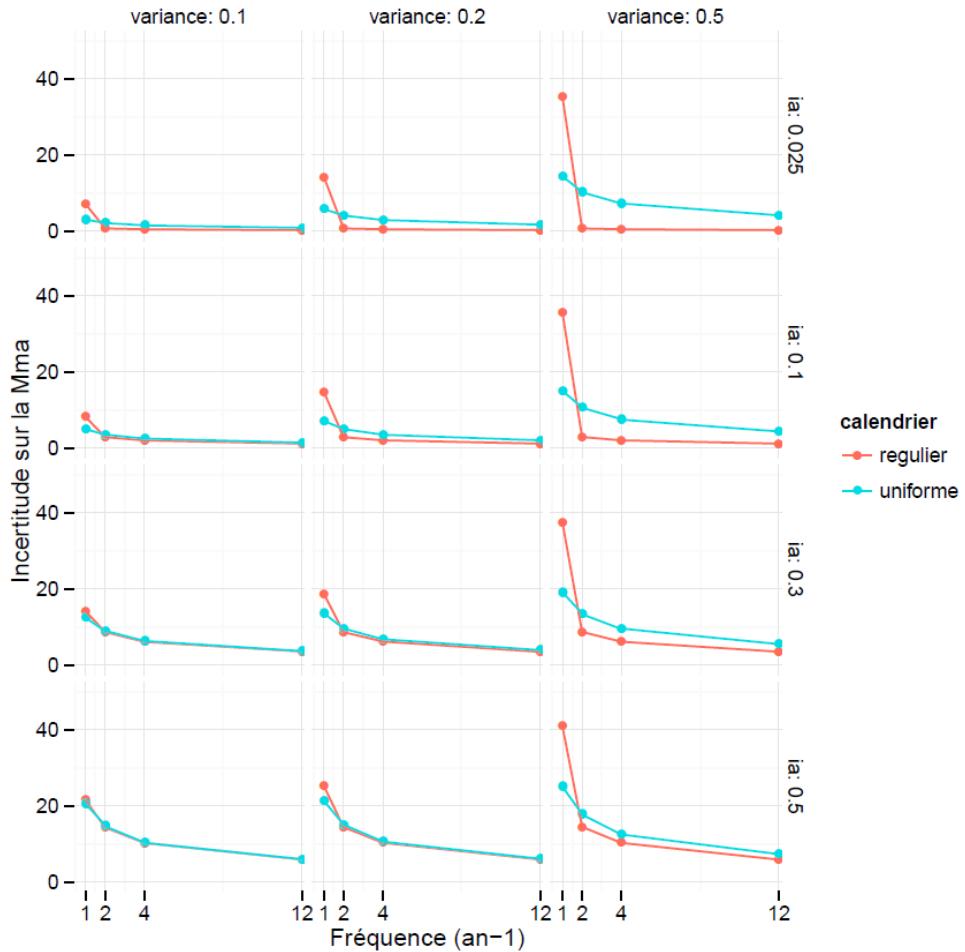


Illustration 23 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique annuelle, pour des chroniques de longueur 6 ans. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différents cas d'amplitude du signal périodique (en part de la valeur moyenne).

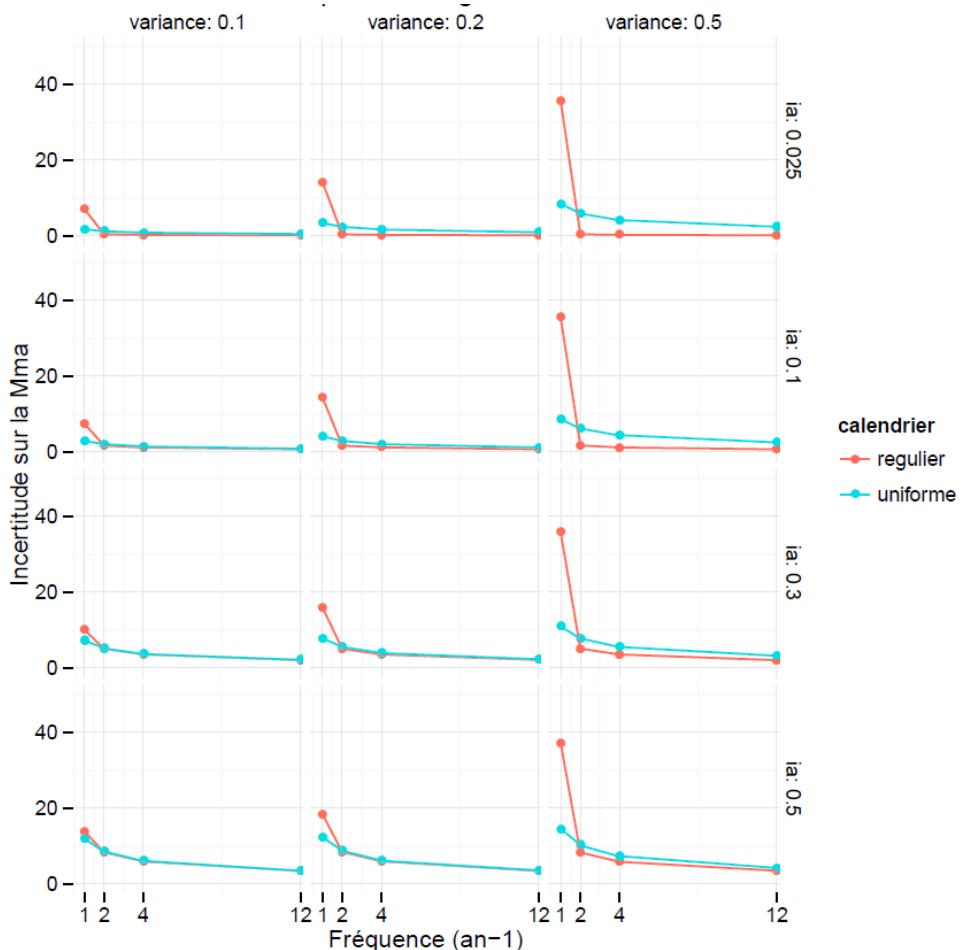
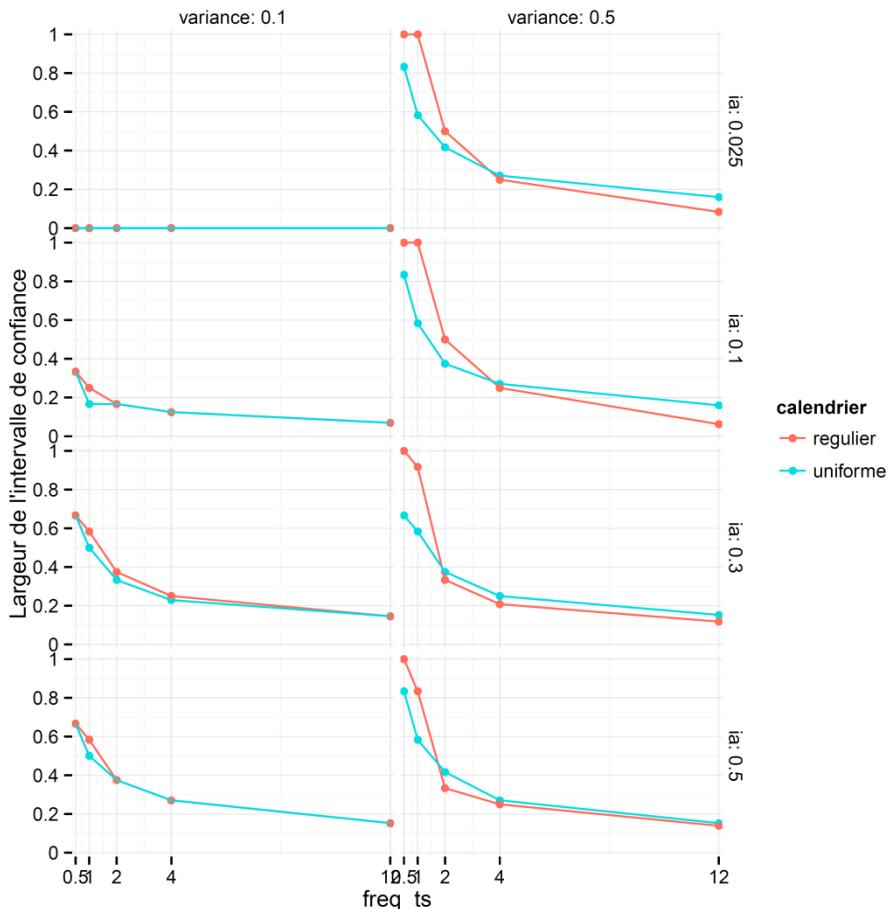


Illustration 24 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique annuelle, pour des chroniques de longueur 18 ans. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différents cas d'amplitude du signal périodique (en part de la valeur moyenne).

L'[Illustration 25](#) montre l'évolution de la largeur de l'intervalle de confiance sur la Fd en fonction de l'incertitude analytique, du type de calendrier et de l'amplitude du cycle. Comme dans le cas des évolutions monotones, l'incertitude sur la fréquence de dépassement de seuil est très importante.

De la même manière que pour la Mma, le calendrier régulier est très défavorisé par rapport à un échantillonnage aléatoire (calendrier uniforme) lorsqu'il n'y a qu'un prélèvement par an car on est amené à échantillonner toujours au même endroit du cycle et la fréquence de dépassement de seuil peut être très surestimée ou très sous-estimée. Plus l'incertitude analytique est petite par rapport au cycle, plus cet effet est visible.

De manière générale, à incertitude analytique et valeur seuil fixes, lorsque l'amplitude du cycle augmente, l'intervalle de confiance s'élargit mais cette influence est peu visible, notamment lorsque le nombre de valeurs est important. De la même manière, lorsque l'incertitude analytique et la valeur seuil augmentent par rapport au cycle, la largeur de l'intervalle de confiance diminue.



*Illustration 25 : Largeur de l'intervalle de confiance autour de la fréquence de dépassement de la valeur seuil en fonction de la fréquence de prélèvement dans le cas d'une évolution à un cycle annuel, pour des chroniques de longueur 12 ans et pour une valeur seuil égale à 1.1 fois la moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 5 %. Les deux colonnes représentent deux valeur d'amplitude du signal périodique, respectivement 10 % et 50 % à gauche et à droite.*

- **Cas des évolutions à deux cycles**

Les résultats associés à des évolutions à deux cycles ([Illustration 26](#), séries temporelles de longueur 3 ans, amplitude des cycles 10 % de la valeur moyenne) montrent une évolution des probabilités semblable à ce qui a été observé pour les évolutions à un seul cycle. Dans les exemples présentés, lorsque la part du cycle annuel augmente (10 %, 20 %, 90 %), l'incertitude associée à la Mma augmente, ce qui est cohérent avec ce qui est observé dans le cas d'évolutions à un cycle (exemples précédents).

Lorsque l'incertitude analytique est grande, l'effet de la périodicité est très faible, voire nul devant l'effet de l'incertitude analytique ([Illustration 26](#), ligne du bas, incertitude analytique de 50 %). L'augmentation de la fréquence d'échantillonnage s'avère être à nouveau un facteur qui permet de diminuer l'incertitude sur la Mma.

L'effet de l'augmentation de l'amplitude des cycles est surtout visible sur les calendriers qui ne sont pas réguliers ([Illustration 27](#) : amplitude des cycles de 20 %, [Illustration 28](#) : amplitude des cycles de 50 %, tous les autres paramètres restant identiques).

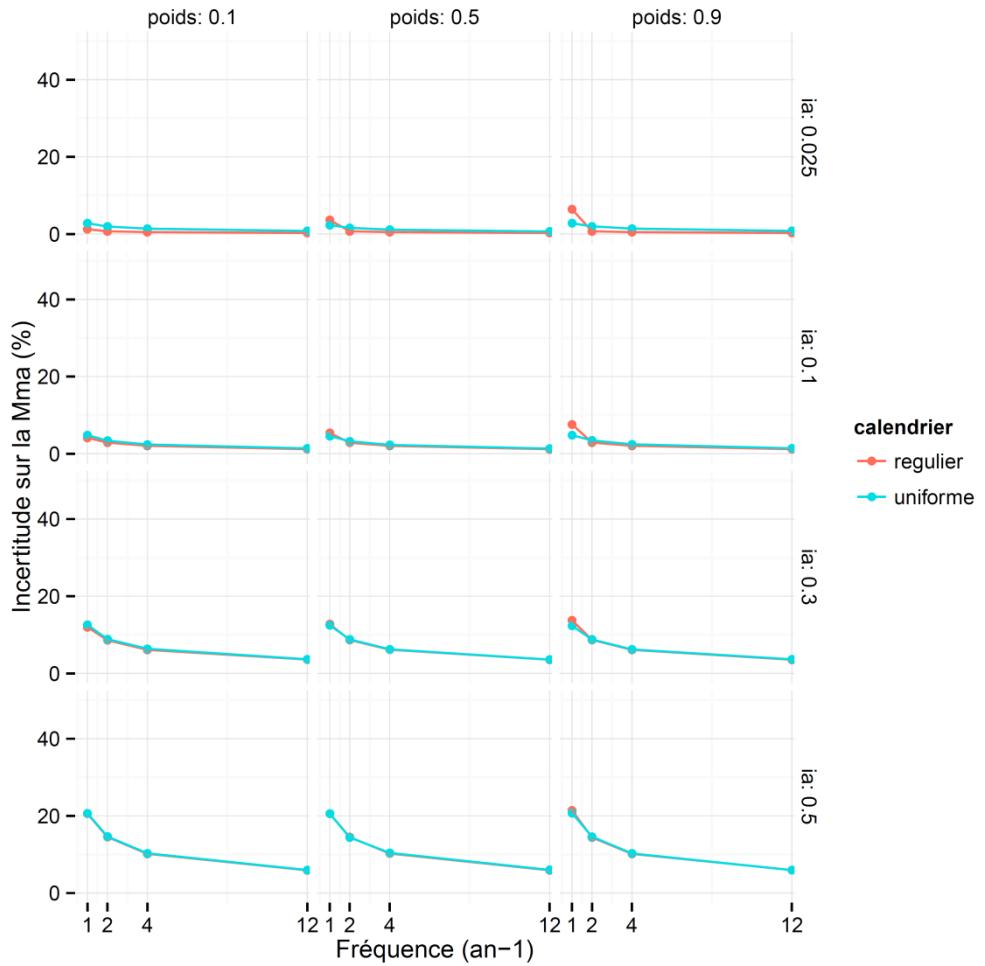
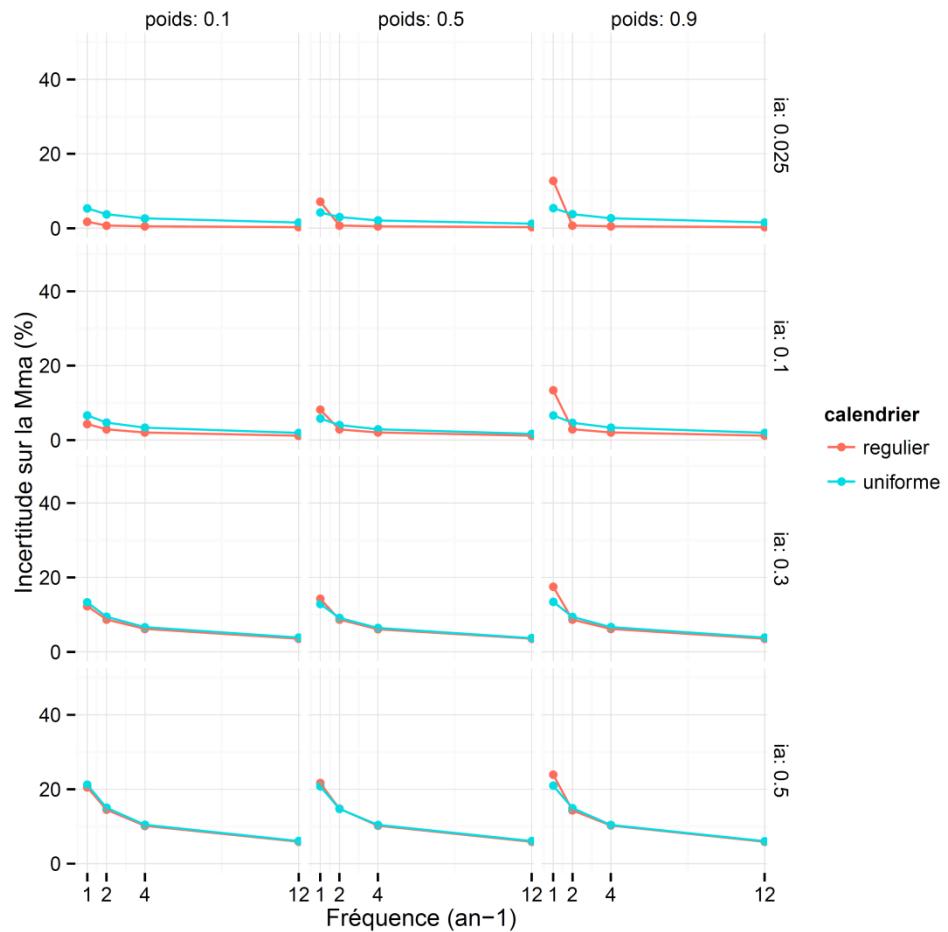
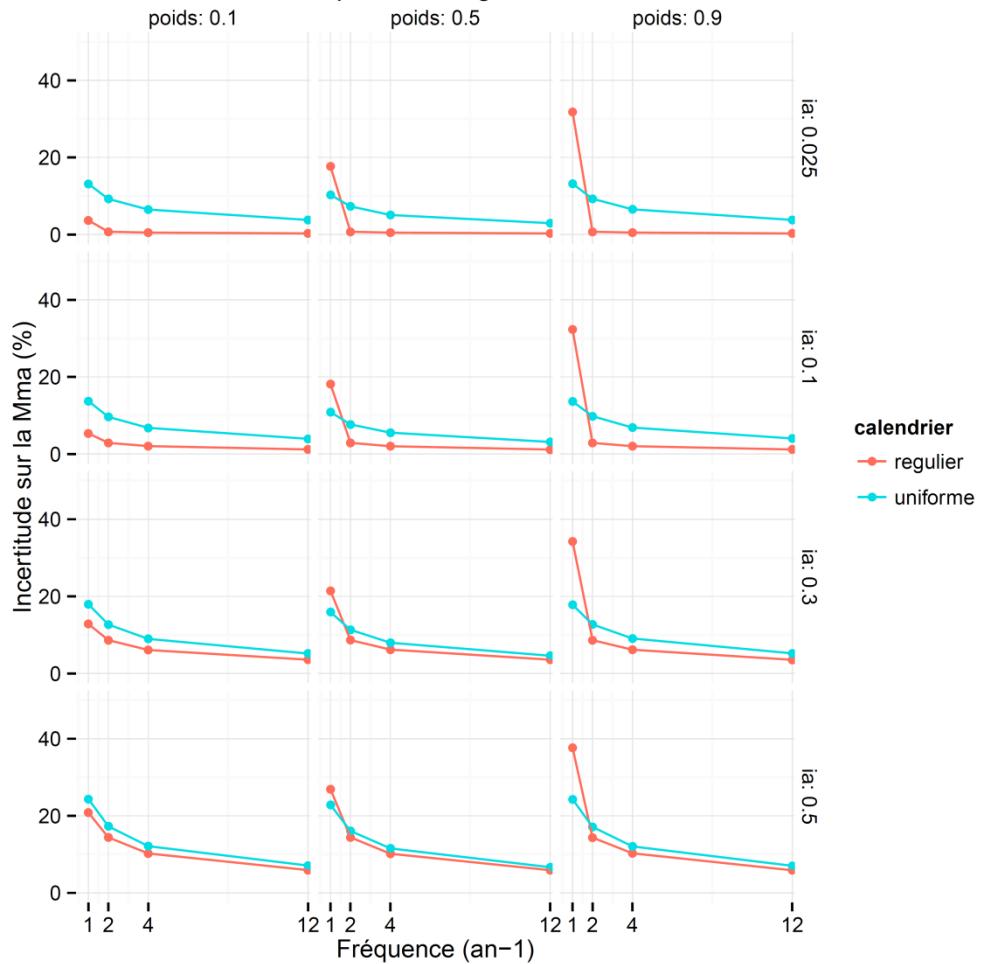


Illustration 26 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 6 ans, et pour une amplitude de cycle de 10 % de la valeur moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportions du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite).



*Illustration 27 : Incertitude sur la Mma (%) dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 6 ans et pour une amplitude de cycle de 20 % de la valeur moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportion du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite).*



*Illustration 28 : Incertitude sur la MMA (%) dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 6 ans et pour une amplitude de cycle de 50 % de la valeur moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportion du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite).*

Les résultats de la largeur de l'intervalle de confiance autour de la fréquence de dépassement de la valeur seuil montrent des évolutions semblables à ce qui est observé pour la MMA ([Illustration 29](#)). L'augmentation de la fréquence d'échantillonnage s'avère être à nouveau un facteur qui permet de diminuer la largeur de l'intervalle de confiance autour de la Fd.

Plus le cycle annuel est important (c'est-à-dire plus le poids du cycle annuel est élevé), plus l'influence du calendrier est forte et plus les fréquences de prélèvement de 1 par an entraînent des incertitudes importantes dans le cas d'un calendrier régulier. Lorsque le poids du cycle annuel est faible, le calendrier régulier conduit à des incertitudes nettement plus faibles et la différence avec le calendrier uniforme est d'autant plus importante que l'incertitude analytique est faible (et donc la part relative du cycle importante). En effet, le cycle pluri-annuel a, dans l'exemple, une longueur de 6 ans, et l'échantillonnage 1, 2, 4,.. fois par an conduit à prélever 6, 12, 24 ... fois dans le cycle pluri-annuel, donc toujours un nombre pair de fois, ce qui donne un net avantage au calendrier régulier.

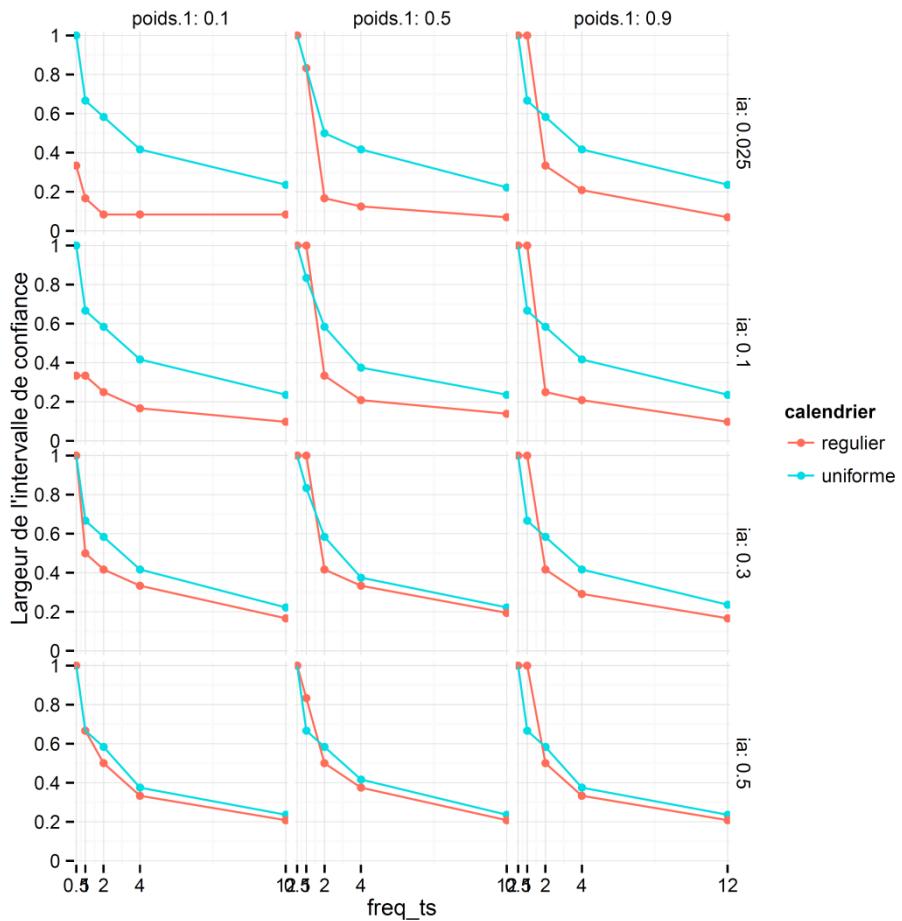


Illustration 29 : Largeur de l'intervalle de confiance autour de la fréquence de dépassement de la valeur seuil en fonction de la fréquence de prélèvement dans le cas d'une évolution cyclique à deux cycles : cycle annuel et cycle pluriannuel de longueur 6 ans, pour des chroniques de longueur 3 ans et pour une amplitude de cycles de 50 % de la valeur moyenne et pour une valeur seuil égale à 1.025 fois la moyenne. Les différentes couleurs représentent différents types de calendrier. Les différentes lignes représentent différentes valeurs d'incertitude analytique, de haut en bas : 2,5 ; 10, 30 et 50 %. Les différentes colonnes représentent différentes proportions du cycle annuel (10 %, 50 % et 90 % respectivement de gauche à droite)

## 4.3. INTERFACE « HYPST-R »

### 4.3.1. Considérations générales sur l'interface

L'interface interactive « Hypst-R » permet de visualiser l'ensemble des résultats associés à l'étude sous forme de graphiques et de tableaux. La plupart des paramètres des graphiques sont paramétrables par l'utilisateur en fonction de ses besoins.

L'interface a été développée sous environnement R, en utilisant le paquet « Shiny ». Ce paquet permet de construire de manière simple et efficace des applications interactives avec R, où les sorties se mettent à jour au fur et à mesure que les entrées sont modifiées sans nécessiter un recharge de l'interface. Il permet de laisser aux utilisateurs le choix des paramètres d'entrée à l'aide des différentes commandes (curseurs, menus déroulants, boutons radio...). Il

est également possible d'intégrer facilement toutes sortes de sorties comme des tableaux, des graphiques. Une application développée avec Shiny peut être accessible sur le web ou en local.

Pour le moment, Hypst-R n'est disponible qu'en installation locale mais il sera simple de créer également une application web si besoin. L'intérêt d'une version web est que les utilisateurs n'ont alors pas besoin d'installer R et ses extensions sur leur machine et de plus, ils disposent automatiquement de la version la plus à jour de l'outil.

Pour l'installation locale, le lancement se fait via R, en exécutant un script de quelques lignes, ce qui ne nécessite aucune compétence en programmation ni maîtrise du logiciel R. L'interface s'ouvre ensuite, au choix dans une nouvelle fenêtre ou dans un navigateur web (la première solution étant plus rapide). L'utilisateur a alors à sa disposition une interface interactive très facilement utilisable.

La procédure d'installation et de lancement de l'outil est présentée en détail en [Annexe 2](#).

#### 4.3.2. Architecture de l'interface « Hypst-R »

L'interface « Hypst-R » est organisée en onglets. L'interface s'ouvre systématiquement sur l'onglet menu (voir [Illustration 30](#)). Trois autres onglets sont disponibles, permettant de visualiser les incertitudes liées au calcul de la Mma, de la Fd et de générer des chroniques à la demande de façon à pouvoir, graphiquement, aider l'utilisateur à choisir le type d'évolution correspondant le mieux aux données dont il dispose. La navigation entre les onglets se fait à l'aide de la souris, en cliquant sur l'un des items présent dans la bande du menu.

**Hypst-R**

Outil de visualisation des incertitudes associées aux calculs de la Mma et de la Fd

brgm

Menu   Mma   Fréquence dépassement   Génération de chroniques virtuelles

Menu	Mma	Fréquence dépassement	Génération de chroniques virtuelles
• Accueil, informations générales	• Résultats relatifs au calcul des MMA • Sélection du type d'évolution, de la longueur de la chronique, de l'incertitude analytique, etc. par l'utilisateur • Représentation graphique • Export des résultats (graphiques)	• Résultats relatifs au calcul de la FD • Sélection du type d'évolution, de la longueur de la chronique, de l'incertitude analytique, etc. par l'utilisateur • Représentation graphique • Export des résultats (graphiques)	• Exemples d'évolutions temporelles générées au cours de cette étude • Choix des caractéristiques de la chronique réalisé par l'utilisateur • Export du graphique

Illustration 30 : Copie d'écran du bandeau supérieur de l'interface Hypst-R (en haut). Architecture de l'interface Hypst-R, description des fonctions de chacun des onglets (en bas).

#### 4.3.3. Onglets Mma et Fd

Les onglets « Mma » et « Fréquence dépassement » sont organisés de manière très similaire. La description détaillée de l'onglet Mma réalisée ci-dessous, s'applique donc aussi pour l'onglet « Fréquence dépassement ».



Illustration 31 : Capture d'écran de l'interface "Hypst-R", onglet Mma, décrivant l'organisation de l'onglet en 6 zones dont les fonctions sont distinctes.

L'onglet Mma de l'interface « Hypst-R » se compose de six zones (Illustration 31) dont les fonctions sont distinctes :

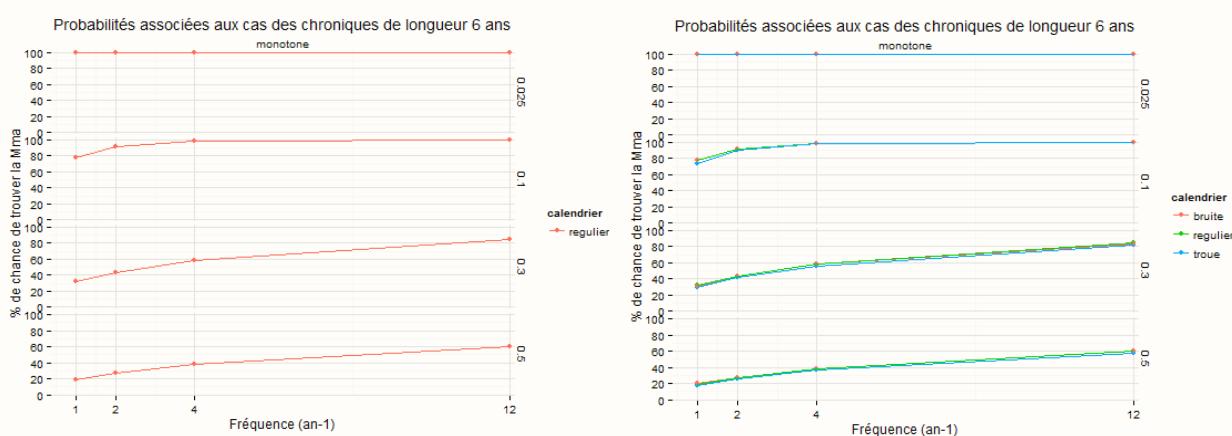
- la zone 1 reprend le titre de l'onglet et propose une courte description des fonctions associées à l'onglet ;
- la zone 2 permet à l'utilisateur de procéder à une sélection dans les résultats de cette étude en utilisant des critères distingués selon i) les variations environnementales (le type d'évolution : monotone, 1 cycle ou 2 cycles) sur un fond de couleur verte ; ii) le type de suivi (type de calendrier, longueur de la série temporelle, fréquence de prélèvement) sur un fond de couleur orangée ; et iii) les conditions d'analyses (incertitude analytique ou variabilité environnementale du paramètre) sur un fond de couleur bleue ;

- la zone 3 est une zone d'affichage conditionnel et ne s'affiche que lorsque des évolutions à un ou deux cycles ont été sélectionnées. Cette zone permet de spécifier l'amplitude et la longueur des cycles, et la part relative des cycles dans le cas d'une évolution à deux cycles ([Illustration 32](#)) ;



*Illustration 32 : Capture d'écran de l'interface Hypst-R. Onglet Mma, détail de l'affichage conditionnel des critères relatifs aux évolutions à un ou deux cycles.*

- la zone 4 est une zone de graphique se mettant à jour automatiquement à mesure que l'utilisateur spécifie les conditions qui l'intéressent (zones 2 et 3) et les paramètres des graphiques (zone 5). Par défaut, les graphiques ([Illustration 32](#), [Illustration 33](#)) représentent la probabilité d'obtenir la Mma à 5% près (axe des ordonnées) en fonction de la fréquence d'échantillonnage (axe des abscisses). Une couleur distincte est donnée à chaque type de calendrier ([Illustration 33](#)), et le cas des différentes valeurs d'incertitudes analytiques sont séparées en différentes lignes, légendées à droite.



*Illustration 33 : Exemple de graphiques générés grâce à Hypst-R. Exemple du choix de plusieurs types de calendrier distingués par des couleurs différentes (calendrier régulier uniquement (figure de gauche) en rouge, et calendriers bruité, régulier, troué (figure de droite) respectivement de couleur rouge, verte et bleue). Les graphiques sont organisés en colonne au sein de laquelle chaque ligne correspond à un type d'incertitude analytique (ici, de haut en bas, 2,5 %, 10 %, 30 % et 50 %).*

Si l'utilisateur choisit de comparer plusieurs types d'évolutions ([Illustration 34](#)), les résultats associés seront ajoutés comme une nouvelle colonne.

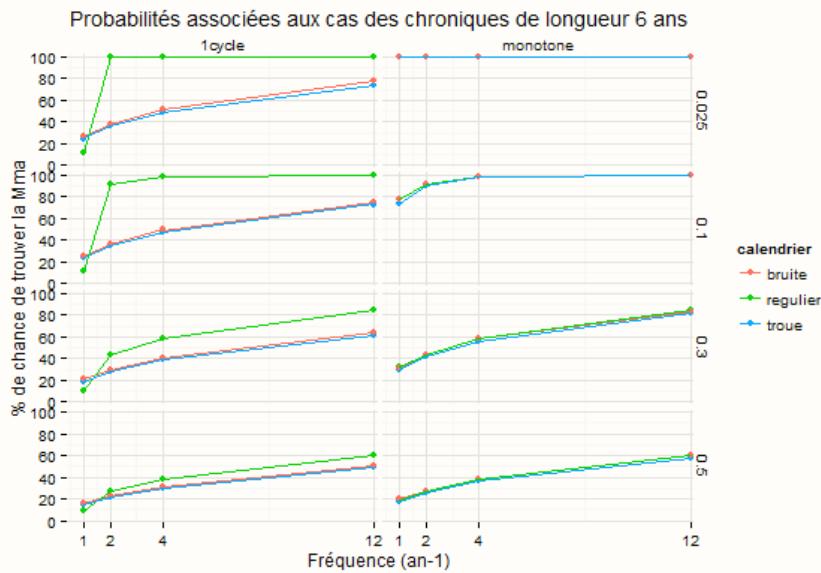


Illustration 34 : Exemple de graphiques générés grâce à Hypst-R. Exemple du choix de plusieurs types d'évolutions : monotone et à un cycle, pour lesquelles les résultats sont organisés selon deux colonnes (résultats relatifs à l'évolution à un cycle à gauche, résultats relatifs à l'évolution monotone à droite). Chaque ligne correspond à une incertitude analytique différente (de haut en bas, 2,5 %, 10 %, 30 % et 50 %).

- la zone 5 (Illustration 31) permet à l'utilisateur de spécifier une partie des paramètres des graphiques. Il choisit tout d'abord sous quelle forme il souhaite afficher les résultats : en incertitude sur la Mma ou en probabilité de trouver la Mma vraie (dans l'onglet « Fd », l'utilisateur n'a pas le choix, seul la largeur de l'intervalle de confiance est proposée). Il choisit ensuite les paramètres qui permettent de détailler les graphiques en lignes ou en colonnes.  
Dans cette zone de l'interface, il est également possible de modifier des paramètres d'apparence des graphiques et de réaliser des exports. Deux types d'export peuvent être réalisés :
  - un export du graphique réalisé sous format pdf : lorsqu'un export est demandé, une fenêtre s'ouvre dans l'éditeur PDF par défaut pour le système d'exploitation sous lequel le logiciel R fonctionne (Acrobat Reader le plus souvent) ;
  - un export des données utilisées pour la réalisation du graphique : lorsqu'un export est demandé un fichier au format csv est créé que l'utilisateur peut télécharger.

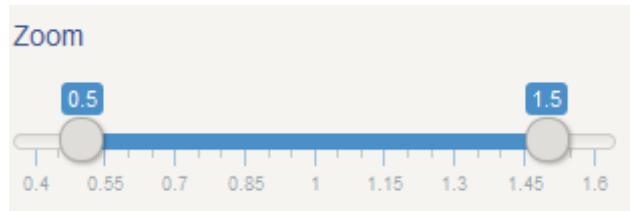
#### 4.3.4. Génération de chroniques

L'onglet « Génération de chroniques virtuelles » (Illustration 37) a été conçu afin de permettre à l'utilisateur i) d'une part de se familiariser avec les termes et les concepts abordés dans cette étude et de faire varier les paramètres des séries temporelles et ii) d'autre part, s'il en ressentait le besoin, de comparer visuellement les données dont il dispose avec des exemples de chroniques virtuelles afin de plus facilement se figurer les conditions dans lesquelles il se trouve pour estimer l'incertitude liée au calcul de la Mma et de la Fd sur ces données.

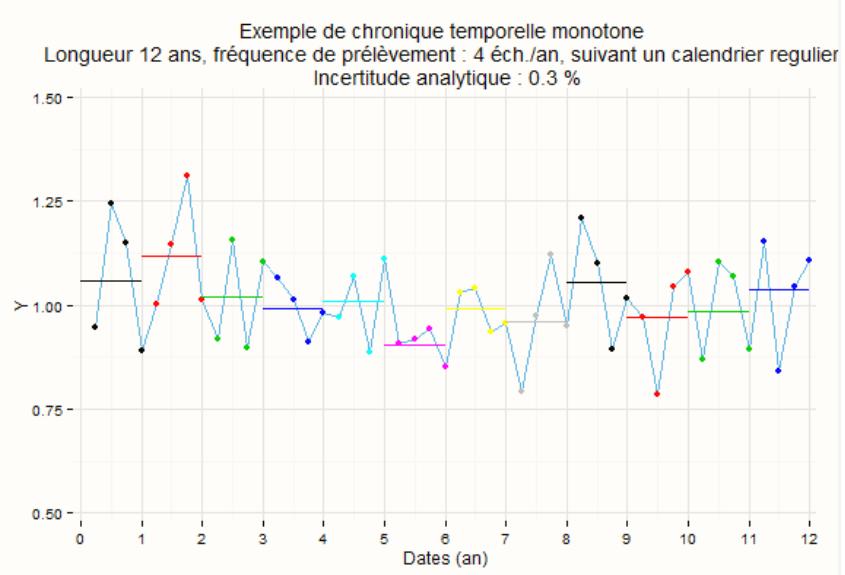
Cet onglet est organisé de façon très similaire aux onglets « Mma » et « Fréquence de dépassement ». Des zones de sélection des paramètres constitutifs de la chronique sont à gauche de la fenêtre et en haut du graphique (variation environnementale sur fond vert, paramètres de suivi sur fond orange et paramètres liés aux analyses sur fond bleu).

La partie relative au graphique représente la série temporelle simulée d'après les caractéristiques sélectionnées. Une nouvelle série temporelle est simulée à chaque fois que l'utilisateur modifie un paramètre, y compris les bornes minimum et maximum de l'axe des ordonnées. Etant donné que les séries temporelles comportent une composante gaussienne aléatoire, la nouvelle chronique simulée sera unique et différente de la précédente. Chaque point (noir ou de couleur) représente une valeur calculée, les valeurs calculées étant reliées par un trait de couleur bleue. Sous le graphique, différentes options de représentation sont présentes :

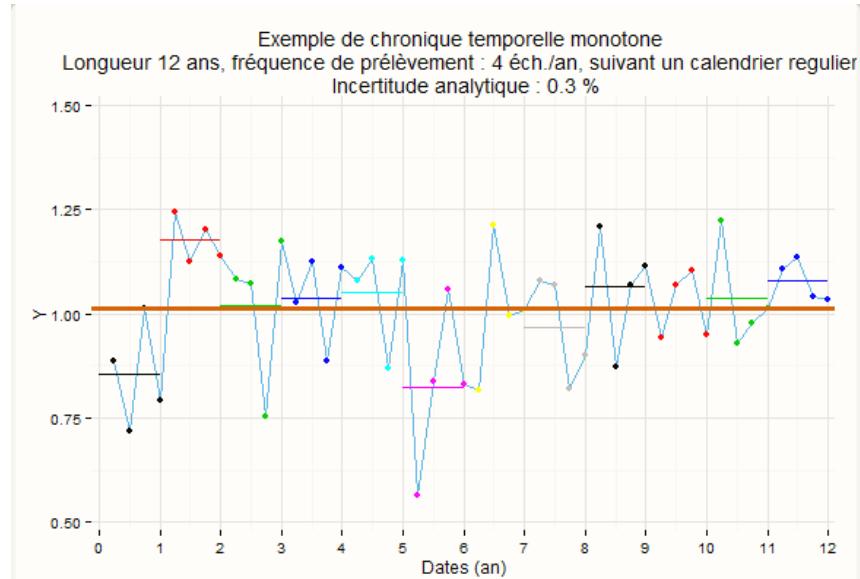
- outil de zoom permettant de fixer les bornes minimum et maximum de l'axe des ordonnées :



- plusieurs options permettant de i) colorer les points selon l'année (permet de mettre en évidence sur quel ensemble est calculée la moyenne annuelle), ii) calculer et tracer sur le graphique les moyennes annuelles ([Illustration 35](#)) et iii) calculer et afficher la moyenne des moyennes annuelles (Mma) sur le graphique ([Illustration 36](#));

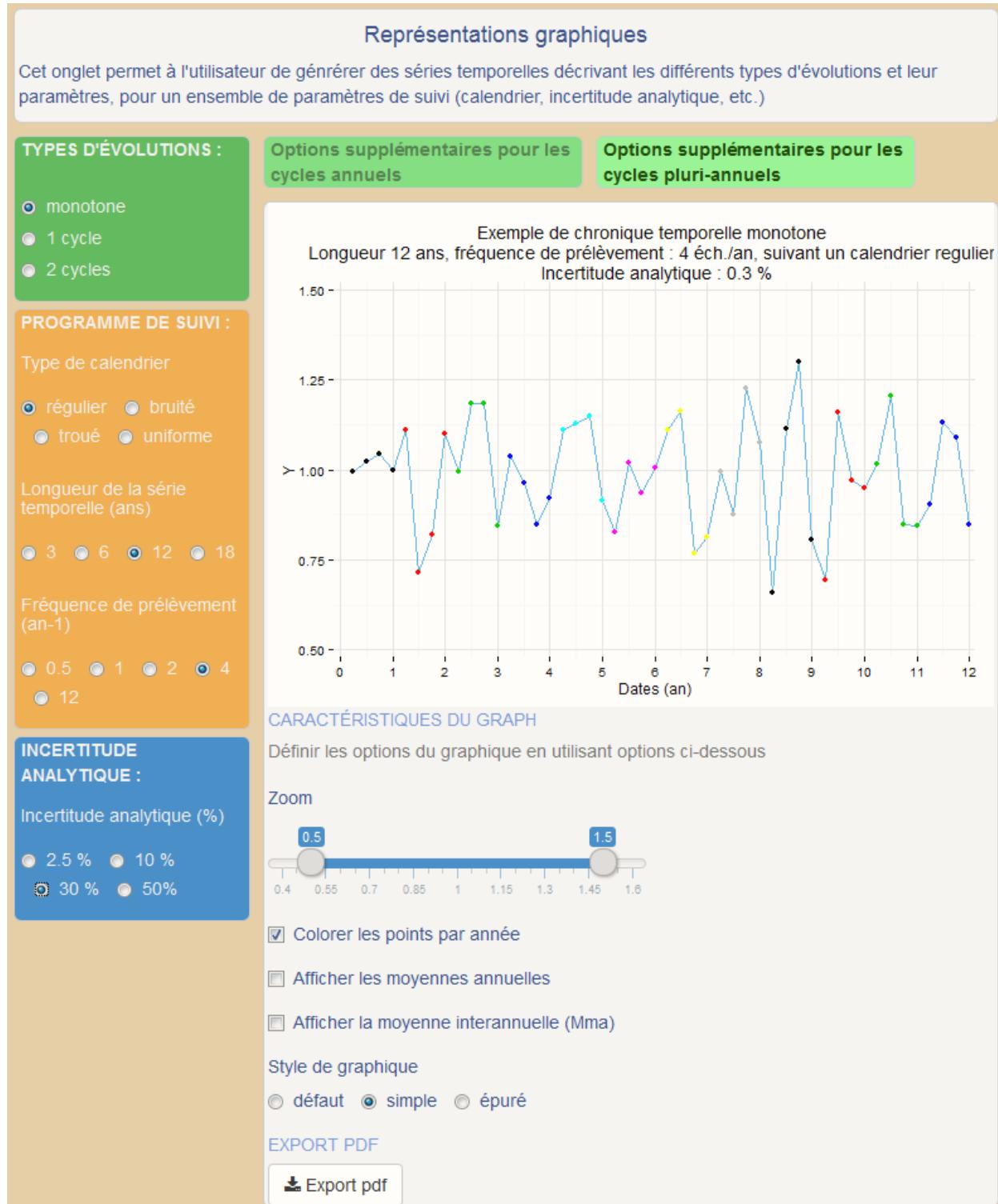


*Illustration 35 : Exemple de chronique générée avec Hypst-R. Cas d'une chronique de type monotone dont l'incertitude analytique associée est de 30 %. Exemple de calcul et d'affichage des moyennes annuelles (les moyennes annuelles sont représentées sous forme de segment de droite de la même couleur que les points).*



*Illustration 36 : Exemple de chronique générée avec Hypst-R. Cas d'une chronique de type monotone dont l'incertitude analytique associée est de 30 %. Les moyennes annuelles (les moyennes annuelles sont représentées sous forme de segment de droite de la même couleur que les points) et la la Mma (segment de couleur orangée, représenté de plus forte épaisseur) sont calculées et affichées.*

- une option permettant de choisir parmi trois types de mise en forme pour le graphique ;
- un bouton réactif permettant d'exporter au format PDF le graphique affiché à l'écran.



*Illustration 37 : Capture d'écran de l'interface Hypst-R, onglet « Exemple de génération de » permettant de générer des chroniques selon des critères choisis par l'utilisateur au moyen des différents champs de sélection. Le graphique représenté ici correspond à une évolution de type monotone, selon un calendrier régulier de longueur 12 ans, selon une fréquence de 4 échantillons/an associés à une incertitude analytique de 30 %.*

## 5. Conclusions

Le travail réalisé dans le cadre de la convention ONEMA-BRGM 2015 a permis de proposer des pistes concrètes afin d'estimer les incertitudes liées au calcul de la moyenne des moyennes annuelles et de la fréquence de dépassement de la valeur seuil. Ces incertitudes sont calculées en fonction du comportement de l'élément chimique étudié, de l'incertitude associée aux mesures et du calendrier d'échantillonnage. La méthode suivie repose sur la génération des chroniques virtuelles dont les paramètres sont connus et maîtrisés.

Les résultats montrent que, de manière générale, le nombre d'échantillons total dans les chroniques est déterminant pour minimiser l'incertitude sur le calcul des indicateurs qui servent à l'évaluation de l'état chimique des masses d'eau : les chroniques les plus longues et les plus fréquemment échantillonnées sont en effet celles qui permettent d'améliorer sensiblement les taux de détection des métriques vraies. De plus, les tests réalisés durant l'étude montrent que pour des incertitudes analytiques assez faibles (2,5 % et 10 %), la Mma peut être estimée assez précisément. Par contre, dans le cas d'incertitudes analytiques importantes (produits phytosanitaires en de faibles concentrations ou polluants émergents par exemple), il apparaît nécessaire d'augmenter sensiblement les fréquences d'échantillonnage dans le but de diminuer l'incertitude sur l'évaluation de la Mma.

L'incertitude sur la fréquence de dépassement de seuil est quant à elle fortement dépendante de la position de la valeur seuil par rapport à la série (moyenne, incertitude analytique et amplitude des cycles s'il y a lieu). Elle est également très sensible au nombre d'échantillons. Le calcul de la Fd est particulièrement critique pour les courtes séries.

Dans le cas d'évolutions cycliques périodiques, le type de calendrier a une influence importante sur les résultats : un échantillonnage selon un calendrier régulier est particulièrement pénalisant si la fréquence de prélèvement est d'une fois par période de cycle, par contre, cet échantillonnage sera très intéressant si la fréquence est un nombre pair de fois dans une période du cycle. Dans le cas où un seul prélèvement par an est envisagé, il est donc recommandé de l'effectuer de façon irrégulière.

Une interface informatique nommée Hypst-R a été développée afin de permettre à l'opérateur :

- de lire et étudier facilement l'ensemble des résultats générés durant l'étude,
- de paramétriser lui-même les chroniques (type de suivi, d'évolution et d'incertitude analytique) pour lesquelles l'incertitude est estimée,
- d'exporter des résultats sélectionnés selon plusieurs formats possibles.

Cette interface qui contient actuellement les résultats sur les estimateurs Mma et Fd pourrait facilement évoluer afin d'intégrer ceux générés précédemment dans le cadre de la convention ONEMA-BRGM 2014 sur l'incertitude associée au calcul des tendances. Plus généralement, Hypst-R pourrait faciliter l'utilisation et la diffusion d'outils d'analyse statistique des données sur les eaux souterraines actuellement développés sous environnement R et qui ne possèdent pas encore d'interface utilisateur (HYPE, ESTR etc...).



## 6. Bibliographie

- Baran, N., Gourcy, L., Lopez, B., Bourgine, B., & Mardhel, V. (2009). Transfert des nitrates vers les eaux souterraines à l'échelle du bassin Loire-Bretagne. Phase 1 : temps de transfert, tendances et typologie des aquifères. Rapport BRGM RP-56884-FR.
- Croiset N., Lopez B. (2013) – HYPE : Outil d'analyse statistique des séries temporelles d'évolution de la qualité des eaux souterraines – Manuel d'utilisation. BRGM/RP-63066-FR. 64 p., 33 fig.
- Circulaire du 23 octobre 2012 relative à l'application de l'arrêté du 17 décembre 2008 établissant les critères d'évaluation et les modalités de détermination de l'état des eaux souterraines et des tendances significatives et durables de dégradation de l'état chimique des eaux souterraines NOR : DEVL1227826C.
- Ghestem JP, Leynet A., Demeyer S, Fischer N., Lalere B. (2014). Niveau de confiance relatif à l'indicateur réglementaire de l'état chimique d'une masse d'eau. Rapport BRGM RP-63228-FR.
- Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science*, 269(5224), 676–679.
- Lopez, B., Blum, A., & Baran, N. (2011). Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines. Guide méthodologique. Rapport BRGM-59931-FR, 29p.
- Lopez B. et Leynet A. (2011). Evaluation des tendances d'évolution des concentrations en polluants dans les eaux souterraines. Revue des méthodes statistiques existantes et recommandations pour la mise en œuvre de la DCE. Rapport BRGM/RP-59515-FR
- Lopez B., Baran N., Bourgine B., Brugeron A., Gourcy L. (2012) - Pollution diffuse des aquifères du bassin Seine-Normandie par les nitrates et les produits phytosanitaires : temps de transfert et tendances. Rapport final BRGM/RP-60402-FR ; 326p
- Lopez B., Croiset N., Surdyk N., Brugeron A. (2013) – Développement d'outils d'aide à l'évaluation des tendances dans les eaux souterraines au titre de la DCE. Rapport final. BRGM/RP-61855-FR, 98 p., 45 ill., 1 ann.
- Lopez B., Leynet A., Croiset N., Ghestem J.P. (2014) – Surveillance chimique : optimisation des fréquences de mesure pour la surveillance des ESO en vue de répondre à la DCE s'agissant de la prise en compte des tendances d'évolution des contaminations. Rapport final BRGM/RP-63845-FR. 102 p., 7 ill., 16 ann.
- Lopez B., Baran N., Bourgine B., (2015) - An innovative procedure to assess multi-scale temporal trends in groundwater quality: Example of the nitrate in the Seine–Normandy basin, France. *Journal of Hydrology* 522, 1–10.
- Yari A., Ghestem JP (2015). Niveau de confiance relatif à l'indicateur réglementaire de l'état chimique d'une masse d'eau de surface : application à des jeux de données réelles. Rapport BRGM RP-65033-FR.

Walker, G. T., & Bliss, E. W. (1932). World Weather V. *Memoirs of the Royal Meteorological Society*, IV(36), 53–84. <http://doi.org/10.1002/qj.49705422601>

Wanner, H., Brönnimann, S., Casty, C., Gyalistras, D., Luterbacher, J., Schmutz, C., ... Xoplaki, E. (2001). North Atlantic oscillation - Concepts and studies. *Surveys in Geophysics*, 22(1984), 321–382. <http://doi.org/10.1023/A:1014217317898>



## Annexe 1 : Vérification analytique de quelques résultats calculés

Dans les cas les plus simples, il est possible de connaître l'incertitude associée au calcul de Mma et de la fréquence de dépassement de seuil. Il est intéressant de valider les résultats issus de la génération de 10.000 chroniques avec ces formules analytiques.

- **Mma**

Dans le cas d'une évolution monotone, l'incertitude sur la moyenne des moyennes annuelles peut être calculées à partir de l'incertitude sur chaque mesure :  $\sigma_{MMA} = \sigma / \sqrt{n}$  où  $\sigma$  est l'écart-type de la composante gaussienne et  $n$  est le nombre de mesures entrant dans le calcul de la Mma.

Si on reprend les valeurs de l'exemple présenté dans le paragraphe 3.4.2, l'incertitude analytique était de 50% (l'écart-type est donc égal à 0.25) et le nombre de valeur dans la chronique était égal à 6. On a donc une incertitude sur la Mma égale à  $\frac{0.25}{\sqrt{6}} = 0.102$ , très proche de la valeur de 0.104 calculé sur les 10.000 chroniques.

Le calcul théorique devient nettement plus compliqué lorsque des variations cycliques affectent le signal lorsque le calendrier n'est pas régulier.

- **Fréquence de dépassement de la valeur seuil**

Il est également possible de calculer de manière théorique l'incertitude sur la fréquence de dépassement d'une valeur seuil dans le cas où l'évolution est monotone.

Dans ce cas monotone, on peut modéliser cette probabilité par une loi binomiale. La loi binomiale permet de modéliser le nombre de succès obtenus lors de la répétition indépendante de plusieurs expériences aléatoires identiques. Lorsque l'on prélève et on analyse un échantillon d'eau, on a à chaque prélèvement une probabilité  $p$  d'obtenir un résultat supérieur à la valeur seuil.

## Annexe 2 : Installation et lancement de « Hypst'R »

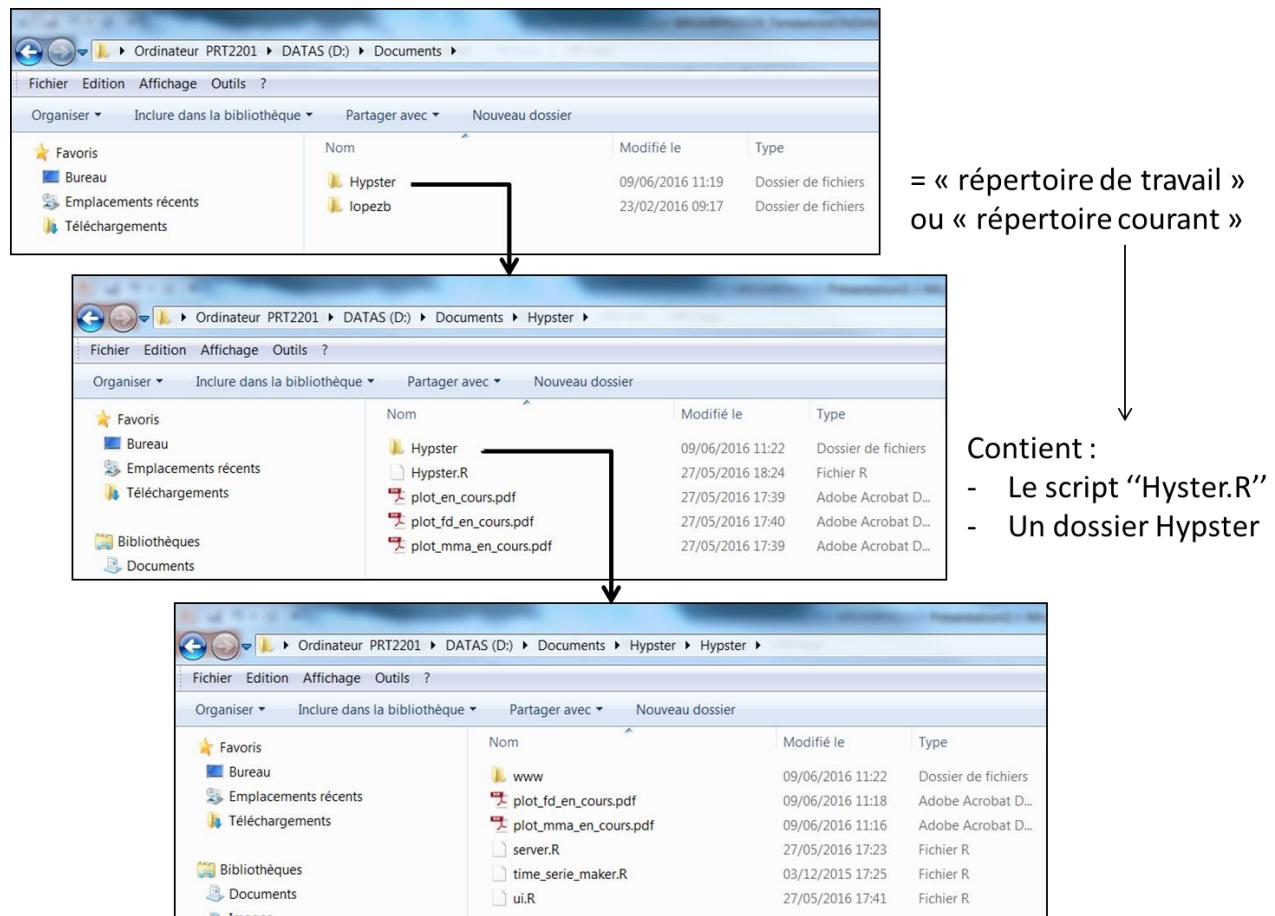
- Diffusion des scripts R de l'outil Hypst'R

L'ensemble des scripts nécessaires au fonctionnement de l'outil Hypst'R est diffusé via le site public du BRGM à l'adresse internet suivante :

<http://www.brgm.fr/production-scientifique/logiciels-scientifiques/logiciels-scientifiques>

Les scripts sont à télécharger à partir du site internet et doivent être sauvegardés dans un répertoire de travail qui sera appelé pour le lancement de l'outil.

Attention, l'architecture et l'arborescence du répertoire de travail est importante et doit être respectée ([Illustration 38](#)).



*Illustration 38 : Architecture du répertoire de travail à respecter pour le fonctionnement de l'outil Hypst'R.*

- Installation du logiciel R

Le logiciel R a été créé en 1993 par Robert Gentleman et Ross Ihaka. Ce logiciel est libre et gratuit. Il peut fonctionner sur différents systèmes d'exploitation (Linux, Windows, MacOS).

R dispose d'une version de base comprenant la plupart des fonctionnalités utiles pour la statistique de base et de nombreux « packages » (ou « extensions »), mis librement à

disposition. Quelques-uns de ces « packages » sont nécessaires au fonctionnement de l'outil HYPE.

### **Une version postérieure à la version 2.1.0 est nécessaire au fonctionnement de l'outil.**

Il est donc nécessaire de télécharger d'abord la version de base de R comme présenté ci-dessous, puis les packages nécessaires.

Pour installer le logiciel, il faut cliquer sur le lien suivant : <http://cran.r-project.org>

Il faut ensuite choisir votre système d'exploitation (Linux, Windows ou MacOS), puis cliquer sur « base » puis télécharger la dernière version.

D'abord, enregistrez le fichier nommé « R-2.XX.X-win32.exe » sur votre disque de façon à le retrouver (par défaut, le fichier s'enregistre le plus souvent dans le dossier « téléchargement » du répertoire « mes documents »).

Ensuite, double cliquez sur ce fichier et suivez les consignes d'installation en laissant les valeurs par défaut. Une fois l'installation terminée, vous aurez la possibilité de lancer le programme depuis le menu « démarrer » de Windows ou en cliquant sur l'icône  sur le bureau.

- Installation des paquets nécessaires au fonctionnement de l'outil**

R est composé d'un socle commun et d'une bibliothèque de fonctions implémentées par les utilisateurs, appelées *paquets*, mises à disposition de tous. Plusieurs de ces *paquets* sont nécessaires au fonctionnement de l'outil

Il faut les **installer** (c'est-à-dire les télécharger sur Internet ou à partir d'un fichier zip) puis les **charger** (c'est-à-dire les rendre accessibles).

La première chose à faire est de télécharger les *paquets* qui seront nécessaires à l'utilisation de l'outil. L'installation est pérenne, il n'y a pas besoin de réinstaller les *paquets* à chaque démarrage de session.

Les *paquets* à installer sont les 3 *paquets* suivants : ggplot2, shiny, data.table

Pour télécharger les *paquets* depuis Internet, il y a deux solutions :

- Par la commande `install.packages(nomdupackage)`
- En cliquant sur « Installer le package » dans le menu « Package »

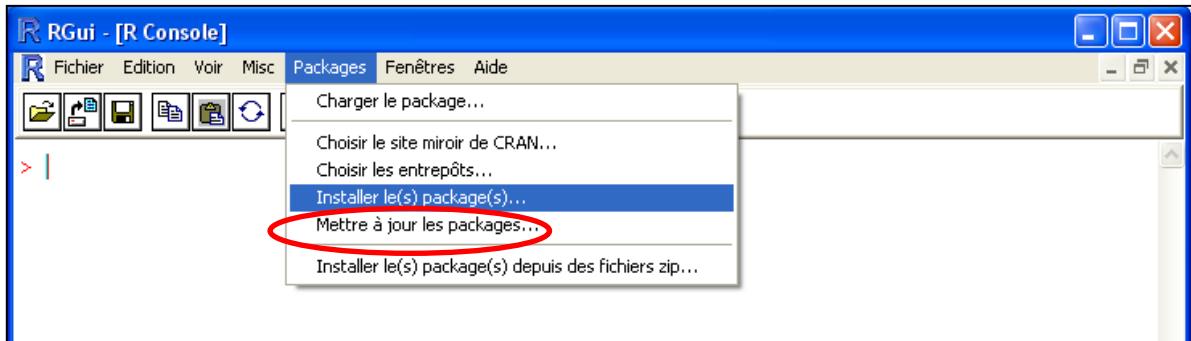


Illustration 39 : Installation d'un package depuis Internet – Aperçu d'écran

Il faut alors sélectionner un site miroir sur lequel le paquet sera téléchargé (voir [Illustration 40](#)). Choisissez le site le plus près de chez vous.



Illustration 40 : Copie d'écran du logiciel R lors du choix du site miroir depuis lequel télécharger un paquet

Choisissez ensuite, dans la liste qui s'affiche, le paquet que vous voulez installer.

- **Lancement de l'outil**

Avant l'exécution d'un module, l'utilisateur doit définir le répertoire de travail. Le répertoire de travail doit contenir le script « Hypster.r ».

Il y a deux possibilités pour définir le répertoire de travail :

- A l'aide de la commande `setwd("chemin")`. Le chemin doit être renseigné entre guillemets, la séparation des répertoires s'indique par une barre oblique (slash) (**attention si vous effectuez un copier-coller depuis l'explorateur de fichier de Windows qui utilise l'antislash pour séparer les répertoires !**) comme montré dans l'exemple ci-dessous. La barre oblique en fin de chemin est optionnelle.

Remarque : pour éviter les erreurs de frappe, vous pouvez ne taper que les premières lettres du nom des répertoires et utiliser la touche de tabulation, qui complète automatiquement le nom des répertoires.

- En sélectionnant « Changer le répertoire courant... » dans le Menu « Fichier » comme présenté sur l'illustration ci-dessous.

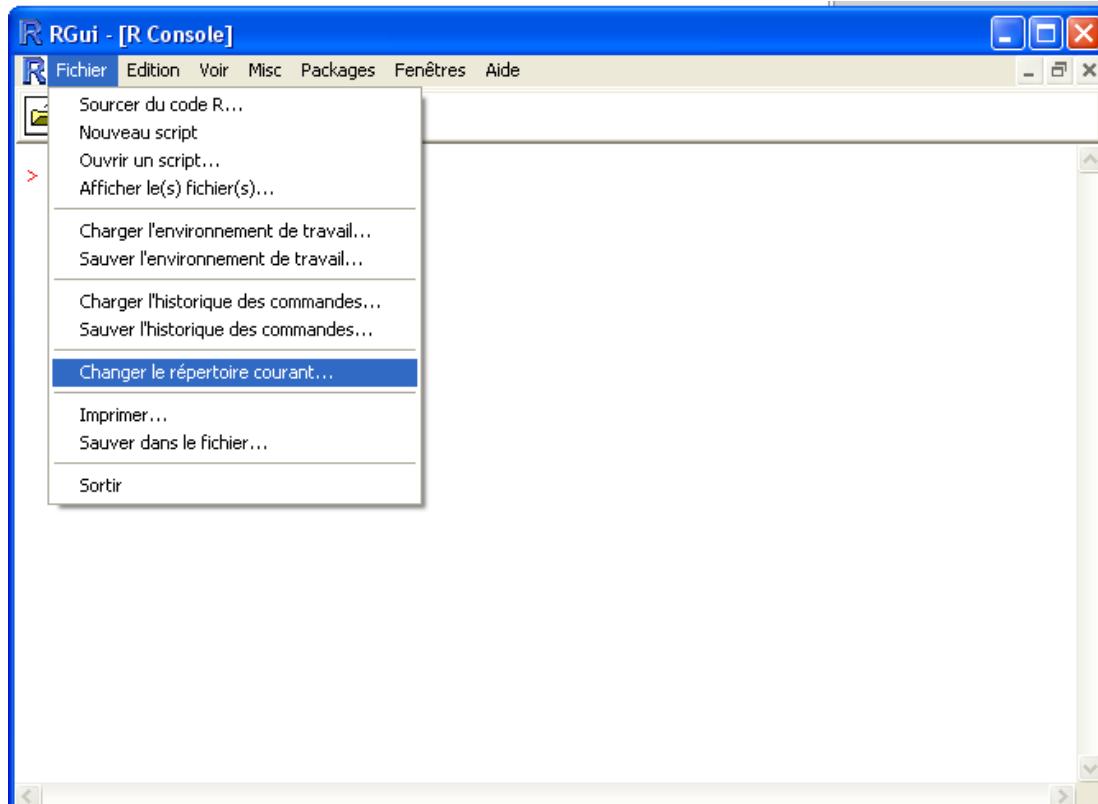


Illustration 41 : Choix du répertoire de travail- Aperçu d'écran

Le lancement de l'outil s'effectue ensuite en tapant la commande `source("Hypster.r")`

La clôture de l'outil s'effectue en fermant la fenêtre de l'explorateur internet à partir de laquelle Hypst'R est commandé. Il est aussi nécessaire de forcer l'arrêt du logiciel R en tapant la touche « Echap ».



Centre scientifique et technique  
**Direction Eau, Environnement et Ecotechnologies (D3E)**  
3, avenue Claude-Guillemain  
BP 36009 – 45060 Orléans Cedex 2 – France – Tél. : 02 38 64 34 34  
[www.brgm.fr](http://www.brgm.fr)