

Méthode d'estimation des tendances temporelles de la qualité de l'eau en rivières

Évolution de la qualité de l'eau avec R (EQuER) - Version 1

Dans cette version¹, les tendances temporelles de la qualité de l'eau (période de 10 ans) sont analysées à l'aide de tests non paramétriques (Helsel et Hirsch, 2002) sur les données mensuelles d'une station d'échantillonnage avec le logiciel R (R Core Team, 2023). Si plus d'une mesure est disponible dans le mois, la première valeur du mois où des résultats sont disponibles pour la majorité des paramètres est sélectionnée. Le choix du test non paramétrique de détection de tendance se fait en fonction des caractéristiques de la série temporelle de données. La présence de saisonnalité et d'autocorrélation (dépendance des données), ainsi que la proportion de données censurées (p. ex. : sous la limite de détection de la méthode; LDM), sont prises en compte. Dans certains cas où les données manquantes ou censurées sont plus importantes ou en présence d'autocorrélation, des tests paramétriques additionnels sont effectués pour valider la présence ou non de tendances (tableau 1).

Afin de déterminer la présence de saisonnalité dans les données, le test de Kruskal-Wallis (fonction *kruskal.test* du progiciel « stats »; R Core Team, 2023) est employé. Le seuil de significativité statistique utilisé est de 10 % (*valeur-p* < 0,10). Pour ce test, une valeur équivalant à la moitié de la limite de détection de la méthode (LDM/2) est attribuée aux données censurées qui sont sous la limite de détection de la méthode. Si cette dernière n'est pas fixe dans le temps (p. ex. : lors d'un changement de méthode), la LDM la plus élevée est utilisée.

La présence d'autocorrélation temporelle est évaluée par un test de Rank von Neumann (fonction *serialCorrelationTest* du progiciel « EnvStats »; Millard, 2013) appliqué sur les résidus de la droite constituée d'une pente Theil/Sen (Theil, 1950; Sen, 1968) et d'une ordonnée à l'origine dérivée comme suit :

$$\text{Ordonnée à l'origine} = Y_{0,5} - \text{pente} * X_{0,5} \quad (\text{Équation 1})$$

où $Y_{0,5}$ correspond à la valeur médiane de la variable de qualité de l'eau et $X_{0,5}$ correspond à la valeur médiane de temps. Lorsqu'il y a présence de saisonnalité, les résidus sont calculés par rapport aux droites mensuelles, chacune étant constituée de la pente globale et d'une ordonnée à l'origine spécifique au mois selon l'équation 1. Le seuil de significativité statistique utilisé pour le test de détection d'autocorrélation est de 5 % (*valeur-p* < 0,05).

Tests non paramétriques de tendances monotoniques

En l'absence de saisonnalité, un test de Mann-Kendall (fonction *kendallTrendTest* du progiciel « EnvStats »; Millard, 2013) est utilisé, avec une approche de correction de la variance considérant un coefficient d'autocorrélation lag-1 (fonction *mmky1lag* du progiciel « modifiedmk »; Yue et Wang, 2004; Patakamuri et O'Brien, 2021) s'il y a présence d'autocorrélation. Le test de Mann-Kendall saisonnier est sélectionné lorsque les données comportent de la saisonnalité (fonction *kendallSeasonalTrendTest* du progiciel « EnvStats »; Millard, 2013), avec une approche Hirsch-Slack 1984 en présence d'autocorrélation (Hirsch et Slack, 1984). La pente globale est estimée par la médiane de tous les accroissements des paires de points appartenant à un même mois (Hirsch et coll., 1982), et l'ordonnée à l'origine par l'équation 1. En présence d'autocorrélation, les tests de Mann-Kendall et de Mann-Kendall

¹ Cette version découle d'un algorithme élaboré précédemment dans le logiciel SAS, qui a été utilisé pour estimer les tendances temporelles de la qualité de l'eau dans de précédents rapports du ministère (MDDEP, 2012; Patoine et D'Auteuil-Potvin, 2013). Des versions ultérieures du programme pourraient utiliser des approches et des tests statistiques différents, qui seraient mieux adaptés à la détection de tendances temporelles.



saisonnier nécessitent des séries de données continues, sans données manquantes. Pour ces tests, les données manquantes sont remplacées par les valeurs prédites par une régression locale non paramétrique (LOESS) utilisant un coefficient d'étendue de lissage (*span*) de 0,75. Pour tous les tests non paramétriques de tendances monotoniques avec moins de 5 % de données censurées, les données sous la limite de détection sont ajustées à une valeur équivalant à la moitié de la limite de détection de la méthode (LDM/2). Lorsque la LDM change au cours de la série de données, la LDM la plus élevée est appliquée à l'ensemble de la série.

Lorsque la proportion de données censurées se situe entre 5 et 50 %, les tests pour données censurées de Mann-Kendall (fonction *cenken* du progiciel « *NADA* »; Lee, 2025) et Mann-Kendall saisonnier (fonction *censeaken* du progiciel « *NADA2* »; Julian et coll., 2025) sont utilisés en l'absence et en présence de saisonnalité, respectivement. Dans ces cas, les valeurs censurées ne sont pas modifiées préalablement au test et les changements de limite de détection de la méthode sont pris en compte. La tendance linéaire non paramétrique est estimée par la méthode *Akritis-Theil-Sen*. Pour la version avec saisonnalité, la tendance globale est obtenue par l'ajout d'une variable bloc et la significativité est testée par permutation (Julian et coll., 2025).

À partir d'une proportion de données censurées de plus de 50 % (Helsel, 1990), une régression logistique sur la fréquence de détection (0 = sous la limite de détection; 1 = au-dessus de la limite de détection) en fonction du temps est effectuée. La probabilité (*valeur-p*) du test de Wald sur la variable temps est utilisée pour déterminer la significativité de la tendance. Contrairement aux autres tests qui prédisent la concentration, la régression logistique prédit la probabilité de détection du paramètre. Celle-ci peut aussi être interprétée comme la fréquence de détection prédite. Lorsque la LDM change au cours de la série de données, la LDM la plus élevée est appliquée à l'ensemble de la série.

Validation par tests paramétriques

La présence de données censurées, de données manquantes ou des changements non monotones dans la série temporelle peuvent influencer le test de détection d'autocorrélation et créer de l'incertitude quant au résultat du test de tendance non paramétrique. Pour cette raison, une validation de la détection ou non d'une tendance est effectuée par un test paramétrique. La détection de la tendance est validée si les deux tests (paramétrique et non paramétrique) sont significatifs, mais elle est incertaine si les deux tests ne concordent pas.

Les tests paramétriques nécessitent le respect de plus d'hypothèses que les tests non paramétriques (normalité, constance de la variance, indépendance des résidus). Une transformation logarithmique ($\log_{10}$) de la variable préalablement au test est faite pour s'approcher de la normalité. En l'absence de saisonnalité, une régression linéaire simple est effectuée en utilisant l'équation suivante :

$$\log(Y) = a + b * \text{temps} \quad (\text{Équation 2})$$

où a = ordonnées à l'origine, b = pente de la tendance. En présence de saisonnalité, une régression linéaire avec une périodicité annuelle (fonction SIN et COS) est effectuée (Hirsch et coll., 1991) en utilisant l'équation suivante :

$$\log(Y) = a + b * \text{temps} + \beta_1 \text{SIN}(2\pi * \text{temps}) + \beta_2 \text{COS}(2\pi * \text{temps}) \quad (\text{Équation 3})$$

où a = ordonnées à l'origine, b = pente de la tendance, β_1 et β_2 sont des coefficients d'ajustement des fonctions sinus et cosinus décrivant la périodicité. La significativité de la tendance est déterminée par le niveau de probabilité (*valeur-p*) du paramètre b (pentes dans les équations 2 et 3). Pour les stations de moins de 12 mois (c'est-à-dire, sans échantillons pour les mois d'hiver), ou lorsqu'il y a présence d'autocorrélation (dépendance des résidus), la régression linéaire (avec ou sans saisonnalité) est effectuée selon un modèle mixte considérant la variable « année » comme aléatoire. Cette fonction permet de s'assurer que les échantillons d'une année à l'autre appartiennent à des groupes différents (p. ex. : une valeur prise en octobre et la suivante, prise en mai, n'appartiennent pas au même groupe d'échantillons). Il est ainsi présumé que les résidus sont indépendants d'une année à l'autre. Pour ces tests paramétriques, les données sous la limite de détection sont ajustées à une valeur

équivalant à la moitié de la limite de détection de la méthode (LDM/2). Lorsque la LDM change au cours de la série de données, la LDM la plus élevée est appliquée à l'ensemble de la série. Pour les séries temporelles comportant entre 5 et 50 % de données censurées, les régressions linéaires avec ou sans périodicité sont effectuées par régressions tobit (fonction *tobit* du progiciel « AER »; Kleiber et Zeileis, 2008). Ce test prend en compte les changements de la LDM.

Barèmes de significativité des tendances

Les tests pour la détection de tendances sont considérés comme significatifs à différents niveaux de probabilité (*valeur-p*) :

- *valeur-p* < 0,05 (significatif *p* < 0,05)
- $0,05 \leq \textit{valeur-p} < 0,10$ (significatif *p* < 0,1)
- $0,10 \leq \textit{valeur-p} < 0,20$ (significatif *p* < 0,2)
- *valeur-p* ≥ 0,20 (non significatif)

Tableau 1 : Choix des tests principaux et de validation selon le % de données censurées, la présence de saisonnalité et d'autocorrélation et la fréquence d'échantillonnage

% de données censurées		Test principal		Test de validation	
		<i>Sans saisonnalité</i>	<i>Avec saisonnalité</i>	<i>Sans saisonnalité</i>	<i>Avec saisonnalité</i>
< 5 % ^a	<i>Sans autocorrélation</i>	Mann-Kendall	Mann-Kendall saisonnier	Aucun (sauf si < 12 mois) ^b	Aucun (sauf si < 12 mois) ^c
	<i>Avec autocorrélation</i>	Mann-Kendall modifié pour autocorrélation	Mann-Kendall saisonnier modifié pour autocorrélation	Modèle linéaire mixte	Modèle linéaire mixte avec périodicité
5 – 50 % ^d		Mann-Kendall pour données censurées	Mann-Kendall saisonnier pour données censurées	Régression linéaire avec données censurées (tobit)	Régression linéaire avec périodicité et données censurées (tobit)
> 50 % ^e		Régression logistique (logit)		Aucun	

^a Les données sous la LDM sont ajustées à LDM/2 pour les tests dans cette catégorie. Si la LDM change, la LDM la plus élevée est utilisée.

^b Modèle linéaire avec effet aléatoire (modèle mixte) si l'échantillonnage s'effectue sur moins de 12 mois (absence de données en période hivernale).

^c Modèle linéaire avec périodicité et effet aléatoire (modèle mixte) si l'échantillonnage s'effectue sur moins de 12 mois (absence de données en période hivernale).

^d Les tests dans cette catégorie prennent en compte les changements de LDM.

^e Lorsqu'il y a un changement de la LDM, la LDM la plus élevée est utilisée pour ce test.

Références bibliographiques

HELSEL, D. R. et R. M. HIRSCH (2002). « Statistical Methods in Water Resources », *Techniques of Water Resources Investigations*, Book 4, chapter A3, U.S. Geological Survey.

HELSEL, Dennis R. (1990). « Less than obvious - statistical treatment of data below the detection limit », *Environ. Sci. Technol.* vol. 24, n° 12, p. 1766–1774.

HIRSCH, R. M. et J. R. SLACK (1984). « A Nonparametric Trend Test for Seasonal Data with Serial Dependence », *Water Resources Research*. vol. 20, n° 6, p. 727-732.

HIRSCH, R. M., R. B. ALEXANDER et R. A. SMITH (1991). « Selection of methods for the detection and estimation of trends in water quality », *Water Resources Research*. vol. 27, n° 5, p. 803-813.

HIRSCH, R. M., J. R. SLACK et R. A. SMITH (1982). « Techniques of Trend Analysis for Monthly Water Quality Data », *Water Resources Research*. vol. 18, n° 1, p. 107-121.

JULIAN, P., D. HELSEL et L. LEE. *NADA2: Data Analysis for Censored Environmental Data*, [R package version 2.0.1], 2025, [<https://CRAN.R-project.org/package=NADA2>].

KLEIBER, C. et A. ZEILEIS. « Applied Econometrics with R », Springer-Verlag, New York, 2008, doi:10.1007/978-0-387-77318-6.

LEE, L. *NADA: Nondetects and Data Analysis for Environmental Data*, [R package version 1.6-1.2], 2025, [<https://CRAN.R-project.org/package=NADA>].

MILLARD, Steven. P. « EnvStats: An R Package for Environmental Statistics », Springer, New York, 2013, ISBN 978-1-4614-8455-4.

MINISTÈRE DU DÉVELOPPEMENT DURABLE, DE L'ENVIRONNEMENT ET DES PARCS (MDDEP). *Portrait de la qualité des eaux de surface au Québec 1999 – 2008*, Direction du suivi de l'état de l'environnement, Québec, 2012, 97 p. ISBN 978-2-550-63649-6 (PDF).

PATAKAMURI, S. et N. O'BRIEN. *modifiedmk: Modified Versions of Mann Kendall and Spearman's Rho Trend Tests*, [R package version 1.6.], 2021, [<https://CRAN.R-project.org/package=modifiedmk>].

PATOINE, M. et F. D'AUTEUIL-POTVIN. *Tendances de la qualité de l'eau de 1999 à 2008 dans dix bassins versants agricoles au Québec*, Québec, ministère du Développement durable, de l'Environnement, de la Faune et des Parcs, Direction du suivi de l'état de l'environnement, 2013, 22 p. + 7 ann. ISBN 978-2-550-68544-9 (PDF).

R CORE TEAM. *R. A Language and Environment for Statistical Computing*, [Logiciel version 4.4.2], Vienna, Austria. R Foundation for Statistical Computing, 2024, [<https://www.R-project.org/>].

SEN, Pranab Kumar. (1968). « Estimates of the Regression Coefficient Based on Kendall's Tau », *Journal of the American Statistical Association*, vol. 63, n° 324, p. 1379-1389. doi:10.1080/01621459.1968.10480934

THEIL, Henri (1950). « A Rank-Invariant Method of Linear and Polynomial Regression Analysis », *Proc. Kon. Ned. Akad. v. Wetensch. A*. vol. 53, p. 386-392 (I), p. 521-525 (II), p. 1397-1412 (III).

YUE, S. et C. Y. WANG (2004). « The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series », *Water Resources Management*, vol. 3, p. 18.

Auteur

Dominic Vachon, Ph. D., biologiste
Direction du suivi et de l'évaluation de l'état des milieux aquatiques

Réviseurs et collaborateurs scientifiques

Michel Patoine, ing., M. Sc.
Direction du suivi et de l'évaluation de l'état des milieux aquatiques

François D'Auteuil-Potvin, M. Sc., statisticien
Direction principale de la qualité de l'air et du climat

Pour citer ce document :

Vachon, Dominic. *Méthode d'estimation des tendances temporelles de la qualité de l'eau en rivières : Évolution de la qualité de l'eau avec R (EQuER) – Version 1*. Direction du suivi et de l'évaluation de l'état des milieux aquatiques, ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs. 4 p.

© Gouvernement du Québec

Ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs
Dépôt légal – Bibliothèque et Archives nationales du Québec, [2026]
ISBN (PDF) : 978-2-555-03735-9