# Rappel sur le principe de la méthode et influences prises en compte

Pour cette étude, nous utiliserons la méthode d’analyse statistique dite HST développé par EDF (Hydrostatique – Saison – Temps) pour analyser les différentes contributions des variables les plus représentatives. Le modèle est un modèle régression linéaire multiple basé sur le principe de superposition, c’est-à-dire que la mesure est supposée être la somme de plusieurs effets indépendants les uns des autres. La variable à expliquer (la mesure) s’écrit donc sous la forme d’une combinaison linéaire de plusieurs variables explicatives  :

Les coefficients de linéarité à de chacune des variables sont ajustés statistiquement de manière à minimiser la dispersion du résidu , c’est-à-dire l’écart entre le modèle et la mesure. Dans la pratique, c’est la somme des carrés des résidus qui est minimisée (méthode des moindres carrés).

Le modèle HST, adapté en H3SJT (développé par Cementys) pour l’analyse des données bassin 8, est basé sur les influences suivantes:

* Les influences hydrostatiques (H3) :
  + La cote du bassin 9
  + La marée
  + La cote du bassin 8
* Les influences thermiques (SJ) :
  + La température saisonnière moyenne
  + La température journalière moyenne
* L’influence temporelle (T)

Les influences hydrostatiques et thermiques sont postulées réversibles et représentent le comportement normal de l’ouvrage dans son environnement. **L’influence temporelle représente quant à elle les évolutions irréversibles de la mesure liées principalement au vieillissement de l’ouvrage**.

Dans notre analyse, seuls trois effets sont pris en compte:

* L’effet de la marée (effet hydrostatique), noté H. Cet effet est modélisé par un polynôme de degré 4 du niveau où représente le niveau de la mer, sa moyenne et son écart-type sur la période d’analyse.
* L’effet saisonnier (lié aux variations saisonnières de la température), noté S. Cet effet est modélisé par une fonction périodique du temps de période. Les deux premiers termes de la décomposition en série de Fourier sont considérés pour représenter cette fonction périodique :
* Effet journalier J (lié aux variations journalières de la température). Cet effet est modélisé par une fonction périodique du temps de période. Les deux premiers termes de la décomposition en série de Fourier sont considérés pour représenter cette fonction périodique :
* L’effet de dérive temporelle est ajouté car la plage de mesure est suffisante pour mettre en perspective une dérive temporelle. Cet effet est modélisé par une fonction périodique du temps t de période 〖∆T〗\_J=1 jour.

Dans les modèles présentés, s’agissant d’une réactualisation, l’influence du niveau d’eau dans les bassins B8 et B9 n’a été étudiée. Cela pourra faire l’objet d’un second rapport d’analyse.

Ainsi le premier modèle s’écrit sous la forme suivante :

Où Y est la variable à expliquer (mesure brute) et le résidu du modèle. Les coefficients à sont ajustés statistiquement par le modèle. Ce modèle sera appelé **HSJ** dans la suite du rapport.

Un second modèle est proposé prenant en compte l’influence du temps par une variable linéaire T. Le second modèle s’écrit alors :

Etant donné la forte influence de la température sur les mesures (voir le rapport annuel de surveillance ISI.RP.SID.dossier\_de\_surveillance\_annuelle.fev15), la température mesurée au niveau de chaque capteur sera également intégrée comme variable explicative dans les analyses. Ce nouveau modèle, appelé **HSJTTh** dans la suite du rapport, s’écrit alors :

Avec :

L’effet thermique identifié ne pourra pas distinguer la part due à la dilatation du capteur et celle due à la dilatation de la structure mais les données corrigées des influences réversibles seront plus précises. Par ailleurs, les fortes corrélations entre les effets Th et S pourront gêner la bonne séparation de ces deux influences mais l’influence globale (S+Th) ne sera pas perturbée.

Concernant la qualité de l’analyse, le principal indicateur est le coefficient de corrélation (R²). Il permet d’apprécier la capacité du modèle à expliquer les variations de ces mesures. On considèrera une analyse comme statistiquement significative dès lors que **R² > 0.45** et comme excellente pour **R² > 0.80**. L’analyse de chaque effet reste néanmoins primordiale pour juger de la qualité et de la signification physique de l’ajustement.

L’interprétation est ensuite effectuée à l’aide de représentations graphiques, dites de « mesures corrigées des influences réversibles » : on soustrait à la mesure brute les différents effets réversibles identifiés par l’analyse. Les mesures corrigées permettent alors de faire ressortir le comportement irréversible de l’ouvrage (vieillissement) où d’éventuelles anomalies (capteurs défectueux, erreur de mesure, comportement anormal de l’ouvrage, etc.).

Ces modèles statistiques sont établis pour une période de référence, lorsqu’elle existe (cette période de référence doit être suffisamment longue pour obtenir un bon calage des influences). On y confronte ensuite les observations de la période d’auscultation pour détecter les écarts sortant de la plage de dispersion habituelle. Dans ce rapport, l’analyse concernera uniquement la période de référence (calage du modèle).

Les critères suivants sont étudiés pour analyser les résultats du modèle selon chaque paramètre d’influence :

* Le coefficient B mesurant les amplitudes des variations saisonnières
* La vitesse d’évolution du déplacement en mm/an v telle que :