



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Projet de Semestre

# Analyse des déformations des barrages à l'aide des séries temporelles

Nicolas ANTILLE

Responsable :

Reinhard FURRER

Chaire de Statistique  
du Professeur S. MORGENTHALER

Département de Mathématiques

Hiver 2000 – 2001



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Différents modèles pour la prédiction</b>	<b>3</b>
2.1	Le modèle déterministe . . . . .	3
2.2	Le modèle hybride . . . . .	4
2.3	Le modèle statistique . . . . .	4
<b>3</b>	<b>Buts du projet</b>	<b>7</b>
<b>4</b>	<b>Les jeux de données</b>	<b>9</b>
4.1	Le premier jeu de données . . . . .	9
4.2	Le second jeu de données . . . . .	10
<b>5</b>	<b>La régression linéaire</b>	<b>13</b>
<b>6</b>	<b>Le modèle classique</b>	<b>17</b>
6.1	Résultats de la calibration . . . . .	18
6.2	Analyse des résultats . . . . .	18
6.3	Prédiction à l'aide du modèle classique . . . . .	22
6.4	Variantes du modèle classique . . . . .	23
<b>7</b>	<b>Le modèle série temporelle</b>	<b>29</b>
7.1	Résultat de la calibration . . . . .	30
7.2	Analyse des résultats . . . . .	30
7.3	Prédiction à l'aide du modèle ST . . . . .	33
<b>8</b>	<b>Calibrations à l'aide du second jeu</b>	<b>35</b>
8.1	Le modèle classique . . . . .	35
8.2	Le modèle série temporelle . . . . .	35

8.3	Analyse des résultats . . . . .	38
8.4	Prédictions à partir des deux types de modèles . . . . .	40
<b>9</b>	<b>Le modèle mixte</b>	<b>43</b>
<b>10</b>	<b>Quelques justifications</b>	<b>45</b>
10.1	Suppression de données . . . . .	45
10.2	Normalisation des données . . . . .	45
10.3	Choix du pendule . . . . .	47
<b>11</b>	<b>Conclusion</b>	<b>51</b>
	<b>Annexe</b>	<b>53</b>
<b>A</b>	<b>Déviations radiales des pendules A et H</b>	<b>53</b>
<b>B</b>	<b>Quelques codes S-Plus</b>	<b>55</b>
	<b>Bibliographie</b>	<b>57</b>

# Table des figures

6.1	Résidus du modèle classique . . . . .	19
6.2	Densité des résidus du modèle classique . . . . .	20
6.3	Autocorrélations des résidus du modèle classique . . . . .	21
6.4	Valeurs estimées ( $\bullet$ ) et valeurs réelles ( $\diamond$ ) (modèle classique) . . . . .	23
6.5	$R^2$ en fonction du paramètre $\beta$ (jeu de données 1) . . . . .	24
6.6	$R^2$ en fonction du paramètre $\alpha$ (jeu de données 1) . . . . .	25
6.7	$R^2$ en fonction du paramètre $\beta$ (jeu de données 2) . . . . .	26
6.8	$R^2$ en fonction du paramètre $\alpha$ (jeu de données 2) . . . . .	27
7.1	Résidus du modèle ST . . . . .	31
7.2	Densité des résidus du modèle ST . . . . .	32
7.3	Autocorrélations des résidus du modèle ST . . . . .	33
7.4	Valeurs estimées ( $\bullet$ ) et valeurs réelles ( $\diamond$ ) (modèle ST) . . . . .	34
8.1	$R^2$ en fonction de l'exposant de $h$ . . . . .	37
8.2	Structure et densité des résidus . . . . .	38
8.3	Autocorrélations des résidus . . . . .	39
8.4	Modèle classique: valeurs estimées ( $\bullet$ ) et valeurs réelles ( $\diamond$ ) . . . . .	40
8.5	Modèle ST: valeurs estimées ( $\bullet$ ) et valeurs réelles ( $\diamond$ ) . . . . .	41
9.1	Analyse graphique du modèle mixte . . . . .	44
A.1	Déviations radiales du pendule A . . . . .	54
A.2	Déviations radiales du pendule H . . . . .	54



# Liste des tableaux

4.1	Légendes et normalisations des variables . . . . .	10
5.1	Résultat d'une calibration . . . . .	14
6.1	Calibration du modèle classique (jeu de données 1) . . . . .	18
7.1	Calibration du modèle série temporelle (jeu de données 1) . . . . .	30
8.1	Calibration du modèle classique (jeu de données 2) . . . . .	36
8.2	Calibration du modèle série temporelle (jeu de données 2) . . . . .	36
8.3	Intervalles de confiance ponctuels . . . . .	41
9.1	Calibration du modèle mixte . . . . .	43
10.1	Calibration avec la totalité des données . . . . .	46
10.2	Calibration avec des données non normalisées . . . . .	46
10.3	Calibration du modèle classique (pendule M) . . . . .	47
10.4	Calibration du modèle série temporelle (pendule M) . . . . .	48
10.5	Comparaison de modèles ST . . . . .	49





# Chapitre 1

## Introduction

Le 2 décembre 1959, le barrage de Malpasset en France se rompit et l'eau déchaînée inonda la ville de Fréjus. Cette catastrophe, survenue 5 ans après la fin de la construction du barrage, fit 421 victimes. Un tel désastre peut-il se produire en Suisse? Théoriquement, la réponse est oui, car le risque zéro n'existe pas. La Suisse compte plus de 200 barrages sur son territoire et parmi eux la Grande-Dixence, le plus haut barrage en béton du monde. La plupart de ces barrages ont été construits entre 1950 et 1970, c'est-à-dire lorsque les besoins en électricité ont augmenté, mais aujourd'hui encore l'énergie produite par les centrales à accumulation est de première nécessité pour alimenter notre pays en électricité. Sa grande souplesse d'utilisation permet de calquer la courbe de l'offre sur celle de la demande. De plus, à l'heure où le spectre des modifications climatiques hante les esprits et où le recyclage des déchets nucléaires s'apparente à la quadrature du cercle, l'énergie hydraulique renouvelable s'impose comme une énergie respectueuse de l'environnement. C'est pourquoi la Suisse ne peut pas aujourd'hui renoncer à l'exploitation de ses barrages, mais ne doit pas oublier le risque potentiel énorme que représente un barrage et doit par conséquent continuer à faire de la sécurité de ses ouvrages une priorité absolue.

Afin de pouvoir prévenir une catastrophe similaire à celle de Malpasset, tous les barrages de Suisse font l'objet d'un contrôle permanent. Chaque barrage est truffé d'instruments de mesures qui permettent de mesurer la température du béton, les pressions qui s'exercent sur le mur ou encore les mouvements du mur. Ce dernier aspect peut paraître surprenant, mais le mouvement des barrages est un fait avéré et parfaitement normal. Par exemple, le couronnement du barrage de la Grande-Dixence se déplace de

11cm vers l'avant lorsque la cote du lac est maximale, puis reprend progressivement sa position initiale au fur et à mesure que le lac se vide pendant les périodes automnale et hivernale. Le but principal de ces mesures est de pouvoir détecter d'éventuelles anomalies dans le comportement des barrages. Pour ce faire, la méthodologie utilisée est toujours la même: tout d'abord on effectue une prédiction sur la grandeur à mesurer, ensuite on mesure cette grandeur "sur le terrain", puis finalement on compare les deux valeurs obtenues. Toutefois, si la méthodologie en elle-même ne change pas, la méthode mathématique qui permet d'effectuer la prédiction peut prendre elle des formes très variées qui sont brièvement décrites dans le chapitre qui suit.

# Chapitre 2

## Les différents modèles pour la prédiction

Précisons tout d'abord que cette étude s'intéresse uniquement aux modèles de prédiction pour le déplacement d'un point d'un barrage. Ces déplacements sont généralement mesurés grâce à des pendules placés à l'intérieur du barrage lui-même. Un pendule est en fait un fil à plomb parcourant toute la hauteur du mur et tendu en son extrémité inférieure par un poids. Des postes de lecture disposés sur toute la longueur du pendule permettent de connaître la position du mur par rapport à la verticale. Le but des méthodes de prédiction est d'estimer en fonction de différents paramètres (température, cote du lac, . . . ) les valeurs fournies par les pendules. Actuellement, il existe 3 modèles principaux pour la prédiction des déplacements des barrages en béton. Les modèles déterministe et statistique sont antagonistes, tandis que le modèle hybride est quant à lui une combinaison des deux modèles précédents.

### 2.1 Le modèle déterministe

L'approche déterministe consiste fondamentalement en une discrétisation par éléments finis à laquelle on associe une loi de comportement du corps de l'ouvrage et de la fondation. On ne se base aucunement sur les mesures des années antérieures. Cette approche est par conséquent la seule possible lors de la première mise en eau et lors des premières années d'exploitation du barrage. Nous nous dispensons ici de donner les équations génériques de cette approche, ainsi d'ailleurs que leurs justifications physiques, car d'une

part elles appartiennent au domaine de l'analyse numérique et d'autre part elles n'apparaîtront jamais dans la suite de ce document.

## 2.2 Le modèle hybride

Le modèle hybride se base sur le modèle déterministe. Il reprend en fait les mêmes équations, mais introduit des coefficients réels libres qui sont ensuite estimés par une méthode de moindres carrés. Pour des informations détaillées sur ces deux modèles, il est conseillé de se référer à l'ouvrage intitulé *Comparaison "Mesuré-Calculé"* qui a été publié en octobre 2000 par le Comité Suisse des Barrages.

## 2.3 Le modèle statistique

Remarquons pour commencer que le modèle statistique qui est présentée dans cette section est le modèle communément utilisé actuellement. L'approche statistique se base sur l'estimation par une méthode de moindres carrés d'un certain nombre de coefficients. Ces coefficients multiplient des fonctions qui approximent l'effet d'une variable sur le déplacement du mur d'un barrage. Afin de clarifier ces explications, il est nécessaire de décrire le modèle dans le détail.

On admet généralement que seules trois variables affectent le comportement d'un barrage en béton :

- le niveau du plan d'eau
- l'âge, c'est-à dire le temps écoulé depuis la construction
- la température du béton (qui est liée à celle de l'air et de l'eau)

Lors de la modélisation, on substitue généralement la variable *température du béton* par la variable *jours écoulés depuis le 1<sup>er</sup> janvier*. On justifie cette substitution par le fait que la température du béton est fortement influencée par la température de l'air qui est elle-même notoirement influencée par la

période de l'année à laquelle on se trouve. Introduisons alors les variables suivantes :

- $h_0$  = niveau du plan d'eau
- $d_0$  = temps écoulé depuis la construction en années
- $s = \frac{2\pi}{365} \times$  nombre de jours écoulés depuis le 1<sup>er</sup> janvier

Observons maintenant la manière dont ces variables affectent le mouvement d'un barrage, c'est-à-dire décrivons les fonctions communément utilisées pour expliquer l'influence de chacune de ces variables sur les déplacements d'un barrage. Ces fonctions doivent tenter de reproduire le comportement réel et par conséquent elles sont déterminées par la physique du problème, ainsi que par des mesures empiriques.

Le modèle français qu'utilisent la plupart des spécialistes décrit par un polynôme du 4<sup>ème</sup> degré en  $h$  la fonction de prédiction pour les déplacements d'un point d'un barrage en béton. Cette composante du déplacement est appelée composante réversible hydrostatique. On a donc :

$$P_1(h_0) = a_0 + a_1 \cdot h_0 + a_2 \cdot h_0^2 + a_3 \cdot h_0^3 + a_4 \cdot h_0^4 + \epsilon_{h_0}$$

Afin d'exprimer l'effet du vieillissement d'un barrage sur les déplacements de celui-ci, on utilise souvent une combinaison de fonctions exponentielles. Cette composante est appelée la composante irréversible. On obtient alors :

$$P_2(d_0) = a_5 \cdot \exp(-d_0) + a_6 \cdot \exp(d_0) + \epsilon_{d_0}$$

Finalement, l'effet saisonnier est approximé par une combinaison de sinusoides, ce qui paraît adéquat pour traduire l'effet périodique du phénomène. Cette composante est la composante réversible thermique. On pose alors :

$$P_3(s) = a_7 \cdot \cos(s) + a_8 \cdot \sin(s) + a_9 \cdot \sin^2(s) + a_{10} \cdot \sin(s) \cdot \cos(s) + \epsilon_s$$

En additionnant simplement ces trois expressions, on obtient alors la formulation courante pour la prédiction du déplacement d'un point d'un barrage en béton :

$$\begin{aligned}P(h_0, d_0, s) &= P_1(h_0) + P_2(d_0) + P_3(s) + \epsilon \\ &= a_0 + a_1 \cdot h_0 + a_2 \cdot h_0^2 + a_3 \cdot h_0^3 + a_4 \cdot h_0^4 \\ &\quad + a_5 \cdot \exp(-d_0) + a_6 \cdot \exp(d_0) \\ &\quad + a_7 \cdot \cos(s) + a_8 \cdot \sin(s) + a_9 \cdot \sin^2(s) + a_{10} \cdot \sin(s) \cdot \cos(s) + \epsilon\end{aligned}$$

où  $\epsilon = \epsilon_{h_0} + \epsilon_{d_0} + \epsilon_s$ .

Il s'agit ensuite de déterminer les coefficients  $a_i$  par une méthode de moindres carrés en utilisant les mesures réalisées entre la fin de la construction et le moment de la prédiction.

# Chapitre 3

## Buts du projet

En Suisse, la surveillance des barrages est placée sous la responsabilité de l'office fédérale des eaux et de la géologie (OFEG) situé à Bienne. Les personnes chargées de cette mission de contrôle de la sécurité des barrages utilisent fréquemment le modèle statistique décrit dans le chapitre précédent. Ce modèle, qui repose uniquement sur des considérations physiques, est d'ailleurs couramment utilisé dans le monde entier pour estimer et prédire les déplacements des murs des barrages. Il est toutefois raisonnable de se poser des questions sur la pertinence de ce modèle. En effet, il s'appuie sur le fait que la saison et la température sont fortement corrélées. Or, d'une année à l'autre, il peut exister à une même période d'importantes différences climatiques. De même, il est avéré que les déplacements des barrages sont des déplacements extrêmement lents. Ainsi, si le niveau d'eau baisse brusquement, le déplacement résultant de cette diminution de la hauteur d'eau s'effectuera avec un certain décalage. Le modèle tient-il compte de tous ces paramètres? La prédiction faite à partir de ce modèle se révèle-t-elle exacte? Les variables qui apparaissent dans la fonction  $P(h_0, d_0, s)$  ont-elles toutes une influence "significative" sur le déplacement du barrage? Afin d'essayer de répondre à ces questions, nous disposons de deux jeux de données provenant de barrages se situant en Suisse. La première étape de ce projet consistera donc à calibrer le modèle classique, c'est-à-dire celui décrit dans le chapitre précédent, à l'aide d'un de ces jeux de données. Cette calibration s'effectuera à l'aide du logiciel S-Plus (Mathsoft Inc.), mais il existe beaucoup d'autres logiciels capables d'effectuer les mêmes calculs. Il s'agira ensuite d'analyser les résultats obtenus et de vérifier la qualité de la prédiction effectuée à partir du modèle calibré. Tous ces éléments seront développés et expliqués au

moment où ils interviendront dans la discussion.

Dans un deuxième temps, ce projet s'attachera à l'étude d'un autre type de modèle, appelé série temporelle. Ce modèle, tout comme le modèle traditionnel, est linéaire, mais calcule l'estimation du déplacement du barrage au temps  $t$  non seulement en fonction des données au temps  $t$ , mais également en fonction des données aux temps  $t-1$ ,  $t-2$ ,  $\dots$ . Contrairement au modèle classique, ce modèle ne repose sur aucun fondement physique, mais plutôt sur une perception intuitive du problème. Comme auparavant, il s'agira de calibrer le modèle à partir du même jeu de données, puis de procéder à une analyse à la fois des résultats et de la qualité de la prédiction. Finalement, lorsque les deux modèles auront été analysés en profondeur, une comparaison des résultats permettra de tirer des conclusions.

Finalement, il s'agira de réaliser le même travail avec le second jeu de données afin de vérifier les conclusions obtenues à partir des calibrations effectuées grâce au premier jeu de données. Avant de présenter les différents résultats, il convient d'une part de décrire avec un peu plus de précision les jeux de données qui vont permettre d'effectuer les calibrations et d'autre part d'expliquer brièvement le principe de la calibration d'un modèle linéaire afin de permettre une meilleure compréhension de la méthodologie utilisée.



# Chapitre 4

## Les jeux de données

### 4.1 Le premier jeu de données

Le premier jeu de données dont nous disposons provient d'un barrage qui se situe dans le Haut-Valais. Il nous a été aimablement fourni par l'OFEG. Ce jeu contient les informations relatives aux mesures effectuées sur ce barrage entre 1954 et 1965, ce qui représente 151 observations. Toutefois, afin de pouvoir calibrer un modèle de type série temporelle, il est préférable de posséder des observations réalisées à intervalles réguliers (cet aspect paraîtra évident lors de la description du modèle). C'est pourquoi, seules 132 observations seront conservées, c'est-à-dire une observation par mois pendant 11 ans. Cette façon de procéder peut sembler étrange, mais ce choix sera justifié dans le chapitre 10.

Pour chacune des observations, nous possédons des informations sur deux pendules que nous appellerons dorénavant pendule A et pendule B. Le pendule A comporte un seul poste de lecture, tandis que le pendule B en possède deux que nous appellerons dorénavant B1 et B2. Pour chacun de ces postes de lecture, nous possédons les valeurs des déplacements radial et tangentiel du barrage mesurées en ces points. Au total, nous avons donc 6 informations relatives au déplacement du barrage. En outre, pour chaque observation, nous connaissons la date de la mesure et le niveau du lac à cette date. Nous disposons ainsi de toute l'information nécessaire pour procéder à la calibration du modèle traditionnel.

Afin d'éviter que les spécificités du barrage ne faussent les résultats, nous utiliserons lors de la calibration des données normalisées au lieu des données

TAB. 4.1 – *Légendes et normalisations des variables*

Variable	Légende	Formule de normalisation
Hauteur du niveau d'eau	$h_0$	
Hauteur du niveau d'eau	$h$	$\frac{h_0 - \min(h_0)}{\max(h_0) - \min(h_0)}$
Année de la mesure	$d_0$	
Année de la mesure	$d$	$\frac{d_0}{10}$
Mois de la mesure	mois	
Jour de la mesure	jour	
Jour de la mesure	$s$	$\frac{30.5 \cdot (\text{mois} - 1) + \text{jour}}{365} \cdot 2 \cdot \pi$
Déplacement radial au poste A	$a_{\text{rad}}$	
Déplacement tangentiel au poste A	$a_{\text{tang}}$	
Déplacement radial au poste B1	$b_{\text{rad}}$	
Déplacement tangentiel au poste B1	$b_{\text{tang}}$	
Déplacement radial au poste B2	$c_{\text{rad}}$	
Déplacement tangentiel au poste B2	$c_{\text{tang}}$	

brutes. Ce choix semble légitime et nous le justifierons d'ailleurs également dans le chapitre 10. Le tableau 4.1 synthétise toutes les légendes qui seront utilisées dans les chapitres suivants ainsi que les formules de normalisation de différentes variables. Toutes les variables sont maintenant définies et les caractéristiques de ce jeu de données sont connues.

## 4.2 Le second jeu de données

Le second jeu de données, également fourni par l'OFEG, provient d'un important barrage-voûte. Il contient des informations relatives aux mesures effectuées sur ce barrage entre 1943 et 1967, ce qui représente 328 observations. Pour chacune de ces observations nous possédons des informations concernant la date de la mesure, la cote du lac et les déviations radiale et tangentielle de 4 pendules. Chacun de ces pendules possède 2 ou 3 postes de lecture. Contrairement au premier jeu, nous ne supprimerons aucune observation, mais nous appliquerons par contre les mêmes conventions pour la

normalisation des données. Maintenant que les deux jeux de données ont été décrits, nous pouvons nous atteler à la calibration des différents modèles sans oublier au préalable de décrire le principe de la régression linéaire.



# Chapitre 5

## La régression linéaire

La régression linéaire, aussi appelée régression par moindres carrés, est une méthode qui permet d'estimer les paramètres d'une fonction linéaire. Ces paramètres sont estimés de telle sorte que l'erreur commise soit minimale au sens  $L^2$ . Ainsi, lorsque l'on souhaite exprimer une variable  $y$  comme combinaison linéaire de variables  $x_1, x_2, \dots, x_p$ , on écrit usuellement :

$$\vec{y} = a_0 \cdot \vec{1} + a_1 \cdot \vec{x}_1 + a_2 \cdot \vec{x}_2 + \dots + a_p \cdot \vec{x}_p + \vec{\epsilon}$$

Dans cette expression,  $\vec{\epsilon}$  est l'erreur aléatoire commise dans l'estimation de  $\vec{y}$  par la fonction linéaire. On fait généralement l'hypothèse que la variable aléatoire  $\vec{\epsilon}$  suit une loi normale centrée et de variance  $\sigma^2$  inconnue. La variable  $y$  est appelée la variable-réponse et  $x_1, x_2, \dots, x_p$  sont appelées les variables explicatives. Les coefficients  $a_0, a_1, a_2, \dots, a_p$  sont les paramètres du problème. Ces paramètres sont estimés de telle sorte que la norme du vecteur  $\vec{\epsilon}$  soit minimale. Le logiciel S-Plus permet de réaliser cette estimation de manière extrêmement rapide. En plus de la valeur des coefficients  $a_i$ , S-Plus fournit pour chaque coefficient toute une série de valeurs qui sont présentées dans un tableau similaire au tableau 5.1

Afin de mieux comprendre les principes de la calibration d'un modèle linéaire, nous allons maintenant décrire une à une toutes les informations que contient ce tableau récapitulatif. La première colonne du tableau contient simplement la liste de toutes les variables explicatives présentes dans le modèle. La variable *Intercept* correspond à la constante de la fonction linéaire et donc à la valeur estimée de la variable-réponse lorsque toutes les autres variables explicatives sont nulles. La deuxième colonne du tableau contient

TAB. 5.1 – Résultat d'une calibration

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
Intercept ( $a_0$ )				
$a_1$				
...				
...				
$a_p$				
Ecart-type des résidus :				
$R^2 =$				

la valeur des coefficients associés aux variables explicatives. La troisième colonne contient l'écart-type de chacun des coefficients. Rappelons que l'écart-type est la racine de la variance. La quatrième colonne donne la  $t$ -valeur de chacun des coefficients. La  $t$ -valeur d'un coefficient est la valeur obtenue en lui appliquant le test de Student. Dans notre cas, c'est simplement le rapport entre la valeur du coefficient et son écart-type. Le test de Student a pour but de tester l'hypothèse nulle  $H_0$  : "le coefficient est égal à zéro" contre l'hypothèse  $H_1$  : "le coefficient n'est pas égal à zéro". La théorie nous apprend que  $t$  est distribué selon une loi de Student avec  $n - 1$  degrés de liberté, où  $n$  est le nombre d'observations sur lequel repose la régression linéaire. Le logiciel S-Plus calcule alors la probabilité que sous  $H_0$  une réalisation d'une loi  $t$  avec  $n - 1$  degrés de liberté soit supérieure à la valeur absolue de la  $t$ -valeur. Cette probabilité s'appelle la  $p$ -valeur et on la retrouve dans la colonne 5 du tableau. Les statisticiens considèrent généralement qu'un coefficient est nul si la  $p$ -valeur qui lui est associée est supérieure à 5%. Dans ce cas, on dit que la variable explicative est non significative et on la retire du modèle. On recalibre ensuite le modèle avec les variables explicatives restantes. On recommence ce processus jusqu'à ce que le modèle ne contienne plus aucune variable non significative. Il est recommandé de ne supprimer qu'une seule variable à la fois dans un modèle, en général celle dont la  $p$ -valeur est la plus grande. Tous les modèles qui seront présentés par la suite, auront été obtenus en suivant scrupuleusement cette méthodologie. Pour des détails plus complets sur les tests statistiques et sur le test de Student en particulier, le lecteur peut se référer à différents ouvrages statistiques (voir par exemple Hogg et Craig, 1970).

La calibration à l'aide de S-Plus fournit encore deux valeurs représentatives

de la qualité du modèle. Il s'agit d'une part de l'écart-type des résidus qui est simplement l'estimation empirique de  $\sigma$  et d'autre part du  $R^2$  du modèle. Le  $R^2$  est une statistique communément utilisée pour juger de la qualité d'ajustement d'un modèle. Plus la valeur du  $R^2$  est proche de 1, plus le modèle est bon. C'est d'ailleurs un des critères importants qui va nous servir pour la comparaison de deux modèles linéaires. Pour de plus amples détails et une formulation plus mathématique du problème de la régression, il existe de nombreux ouvrages de statistique très complets sur le sujet (voir Morgenthaler, 1997, Searle, 1971 et Scheffé, 1959). Ce bref descriptif de la régression linéaire clôt toute la partie introductive de ce projet et permettra certainement une meilleure compréhension des analyses et des résultats qui sont présentés dans les chapitres 6 et 7.





# Chapitre 6

## Le modèle classique

Avant de présenter et de commenter les résultats obtenus en calibrant le modèle classique, il est certainement utile de rappeler la formulation mathématique de ce modèle en substituant le nom des variables par ceux définis dans le tableau 4.1.

$$\begin{aligned}a_{\text{rad}} &= P_1(h) + P_2(d) + P_3(s) + \epsilon \\ &= a_0 + a_1 \cdot h + a_2 \cdot h^2 + a_3 \cdot h^3 + a_4 \cdot h^4 \\ &\quad + a_5 \cdot \exp(-d) + a_6 \cdot \exp(d) \\ &\quad + a_7 \cdot \cos(s) + a_8 \cdot \sin(s) + a_9 \cdot \sin^2(s) + a_{10} \cdot \sin(s) \cdot \cos(s) + \epsilon\end{aligned}$$

Le but de la calibration du modèle est donc d'estimer la valeur des coefficients  $a_0, a_1, \dots, a_{10}$  de telle sorte que le modèle s'approche le plus possible de la vraie valeur de la variable-réponse, c'est-à-dire que le carré moyen de l'erreur (Mean Square Error) soit minimal. Le logiciel S-Plus permet de réaliser très facilement cette calibration grâce à sa grande souplesse d'utilisation. Précisons finalement que toutes les calibrations réalisées dans ce chapitre concernent la déviation radiale du pendule A. Ce choix est personnel et arbitraire et il sera justifié au chapitre 10. La figure A.1 en annexe représente cette déviation pendant toute les années d'observation.

Remarque : certains statisticiens considèrent que si la variable  $x^3$  apparaît dans un modèle linéaire, alors les variables  $x^2$  et  $x$  doivent également impérativement apparaître dans le modèle, même si elles ne sont pas significatives. Dans la suite de ce travail, nous suivrons une voie différente en ne conservant que les variables significatives et en supprimant toutes les variables non significatives. Ainsi, dans certains modèles, apparaît la variable

$h^2$ , alors que la variable  $h$  qui est non significative n'est pas présente dans le modèle.

## 6.1 Résultats de la calibration

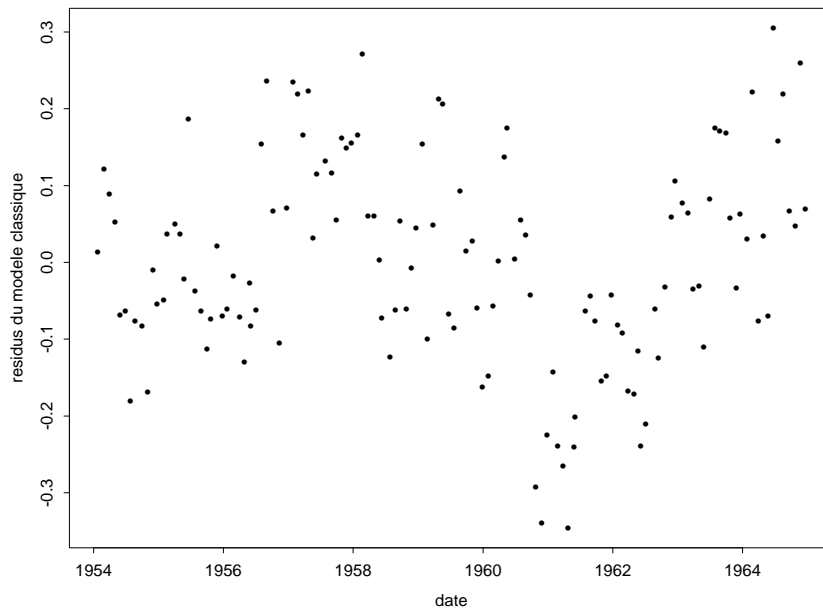
Il est temps de passer à la présentation des résultats, puis à leur analyse. Le tableau 6.1 résume le meilleur modèle obtenu en calibrant le modèle classique à partir des données normalisées provenant du premier barrage.

TAB. 6.1 – Calibration du modèle classique

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$h^2$	0.5375	0.0494	10.8775	0.0000
$\cos(s)$	0.0538	0.0182	2.9548	0.0037
$\sin(s)$	-0.0680	0.0175	-3.8941	0.0002
$\sin^2(s)$	0.1165	0.0352	3.3123	0.0012
$\exp(-d)$	0.3165	0.0471	6.7165	0.0000
$\exp(d)$	-0.1003	0.0152	-6.5866	0.0000
Écart-type des résidus : 0.1366				
$R^2 = 0.9448$				

## 6.2 Analyse des résultats

L'analyse d'un modèle statistique linéaire se laisse volontiers découper en deux parties. La première de ces parties consiste essentiellement en une observation des résultats ou plutôt du tableau des résultats. On peut notamment y observer quelles variables sont significatives (évidemment que dans le modèle final toutes les variables encore présentes sont significatives). Cette première phase ne permet pas encore de juger avec certitude la qualité du modèle. Il s'agit dès lors d'effectuer une analyse plus en profondeur des résultats. Pour cela, on procède généralement à une analyse des résidus. On appelle résidus la différence entre la vraie valeur de la variable-réponse à estimer et la valeur approximée par le modèle linéaire. Le résidu est donc l'erreur commise par la modélisation.

FIG. 6.1 – *Résidus du modèle classique*

Si l'on observe le tableau des résultats de la calibration du modèle classique, on se rend compte que seuls 6 coefficients sont significatifs et que la valeur du  $R^2$  est de 0.9448. La calibration a donc permis d'éliminer 5 variables explicatives du modèle initial. On remarque particulièrement qu'il est inutile de considérer les quatre premières puissances de la hauteur d'eau comme variables significatives. En effet, seule la hauteur d'eau au carré intervient dans la formulation finale du modèle. Nous verrons plus tard les implications de ce constat, notamment lors de la calibration du modèle classique à l'aide du deuxième jeu de données dans la chapitre 8. On peut aussi remarquer (heureusement) que la coefficient relatif à la hauteur d'eau au carré est positif ce qui signifie que plus la cote du lac est élevée, plus la déviation radiale du pendule est importante.

La figure 6.1 représente les valeurs des résidus du modèle classique. On peut aisément y déceler une tendance sinusoidale. Or, une des qualités d'un bon modèle réside dans le fait qu'on ne puisse dégager aucune structure dans le graphique des résidus. La présence d'une structure indique généralement une forte corrélation des résidus ce qui ne devrait pas être le cas puisque l'on fait en régression linéaire une hypothèse d'indépendance des résidus.

La figure 6.2 permet de comparer la densité des résidus avec la densité

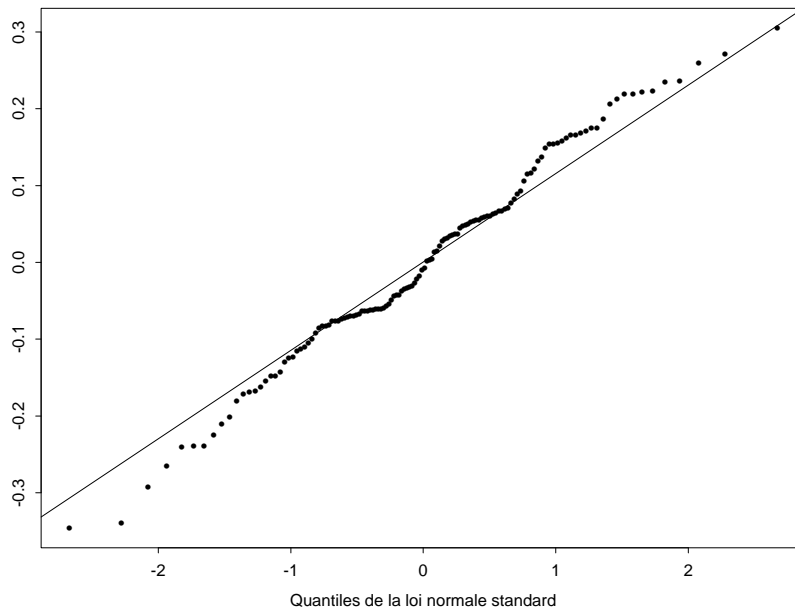


FIG. 6.2 – *Comparaison de la densité des résidus avec une densité normale*

normale centrée et réduite. On peut remarquer que les queues de la distribution des résidus diffèrent assez significativement de celles de la loi normale. Ceci est le cas pour presque tous les modèles, mais la manière dont ces queues diffèrent de celles de la loi normale permet de voir que la densité des résidus possèdent des queues plus lourdes que celles de la loi normale. Cela signifie qu'il y a une probabilité assez forte d'obtenir des résidus élevés, ce que l'on veut éviter au maximum. De plus, on observe que le reste de la distribution ne coïncide que partiellement avec la distribution normale. Une comparaison ultérieure avec la distribution des résidus du modèle de type série temporelle permettra de mettre en évidence la meilleure "normalité" des résidus de ce second modèle.

Un autre aspect intéressant de l'analyse des résidus est d'étudier les autocorrélations entre les résidus. Idéalement, on souhaiterait n'avoir aucune autocorrélation entre les résidus. Dans les faits, cela est quasi impossible et par conséquent on va chercher un modèle pour lequel ces autocorrélations sont faibles. La figure 6.3 représente les autocorrélations entre les résidus du modèle classique. La fonction représentée sur ce graphique peut s'écrire sous la forme mathématique suivante:  $ACF(k) = \rho(k) = Corr(r[t], r[t+k])$  où  $r[t]$  correspond au résidu au temps  $t$ . Il est alors évident que  $\rho(0)=1$  quel que

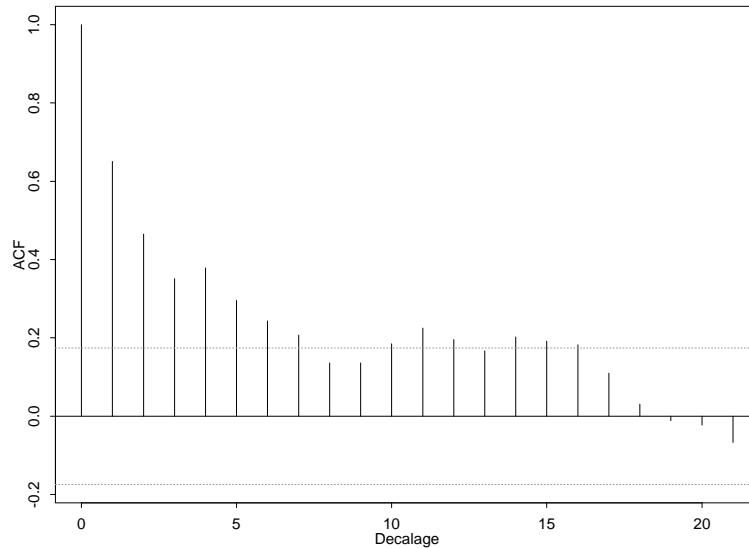


FIG. 6.3 – Autocorrélations des résidus du modèle classique

soit le modèle sous-jacent. Dans le cas qui nous occupe maintenant, nous remarquons que ces autocorrélations sont plutôt grandes. Ce constat sera plus éloquent lors de la comparaison avec les autocorrélations des résidus du modèle série temporelle. Sous l'hypothèse que les résidus sont indépendants et distribués selon une loi normale centrée, la théorie nous apprend que  $\rho(k)$  suit approximativement une loi normale centrée de variance  $\frac{1}{n}$  où  $n$  est le nombre d'observations. On en déduit que  $\left[ \frac{-2}{\sqrt{n}}, \frac{2}{\sqrt{n}} \right]$  est un intervalle de confiance pour  $\rho(k)$  avec un indice de confiance approximativement égal à 95%. Cette intervalle est représenté sur la figure 6.3 par les deux lignes horizontales. On remarque ici que  $\rho(k)$  dépasse fréquemment cette ligne. Cela nous conduit à rejeter l'hypothèse que les résidus sont les réalisations d'un bruit blanc. Or, la régression linéaire se base sur cette hypothèse. Cela représente une nouvelle preuve que le modèle classique n'est pas forcément le modèle adéquat pour résoudre le problème de l'estimation et de la prédiction des déplacements d'un barrage en béton.

### 6.3 Prédiction à l'aide du modèle classique

L'idée de la prédiction est une idée relativement simple : il s'agit de prédire les valeurs futures d'une variable-réponse en se basant sur un modèle calibré à partir de données passées. Dans notre exemple, nous avons calibré le modèle classique à partir des données correspondant à la période 1954–1964. Nous avons obtenu la fonction suivante :

$$a_{\text{rad}} \approx a_2 \cdot h^2 + a_5 \cdot \exp(-d) + a_6 \cdot \exp(d) \\ + a_7 \cdot \cos(s) + a_8 \cdot \sin(s) + a_9 \cdot \sin^2(s)$$

où les  $a_i$  sont les valeurs ajustées des paramètres.

Supposons maintenant que l'on se place en janvier 1966 et que l'on connaisse la valeur de toutes les variables explicatives qui interviennent dans l'approximation ci-dessus. On obtient alors une estimation de la valeur de la variable-réponse  $a_{\text{rad}}$  en remplaçant les valeurs des coefficients  $a_i$  par les valeurs estimées par la calibration du modèle. On peut ensuite comparer la valeur estimée à la vraie valeur mesurée sur le terrain. Cela permet alors, si le modèle est bon, de détecter un éventuel problème sur le barrage. Dans notre cas, cette prédiction va nous servir à juger de la qualité du modèle calibré. Malheureusement, nous ne possédons pas les valeurs nécessaires au-delà de 1964 ; c'est pourquoi nous avons emprunté une démarche légèrement différente. Tout d'abord, nous avons recalibré le modèle classique en nous basant uniquement sur les données relatives à la période 1954–1962. Ensuite, nous avons effectué une prédiction pour les années 1963 et 1964 et enfin nous avons représenté sur la figure 6.4 à la fois les vraies valeurs ( $\diamond$ ) et les valeurs prédites ( $\bullet$ ) pour ces deux années. Nous nous dispensons ici de présenter les résultats de cette recalibration, car ils ressemblent en tout point à ceux de la calibration basée sur les 11 années d'observation.

En observant la figure 6.4, on remarque tout de suite que la prédiction a tendance à sous-estimer la vraie valeur du déplacement radiale du pendule A. Il est très difficile de donner une explication raisonnable à ce phénomène, mais L'OFEG nous a informé que le barrage duquel proviennent les données avait été entièrement vidé en 1963. On ne peut établir aucun lien formel entre cet évènement et la sous-estimation de la réalité par le modèle, mais c'est la seule ébauche d'explication que nous ayons. Avant de développer l'analyse du modèle de type série temporelle, nous allons étudier brièvement deux variantes du modèle classique et discuter de leur intérêt respectif.

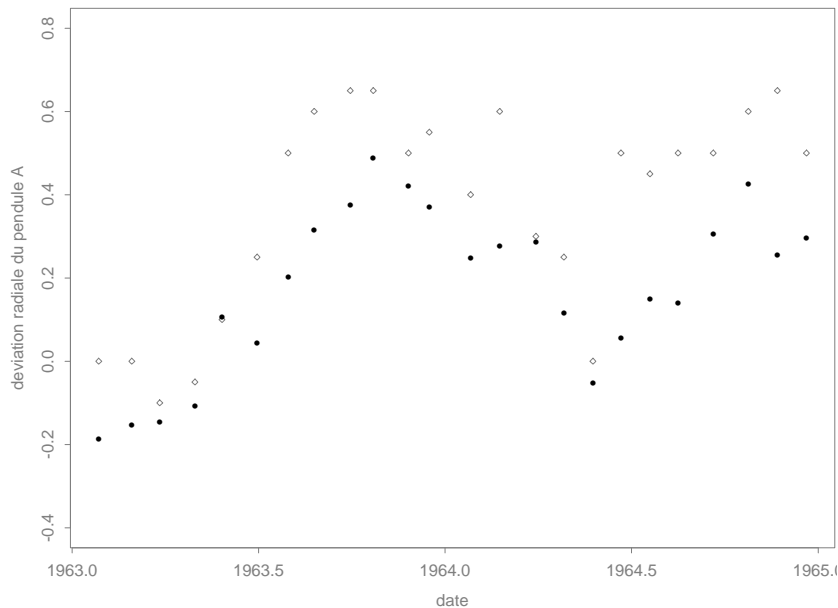


FIG. 6.4 – Valeurs estimées (●) et valeurs réelles (◇)

## 6.4 Variantes du modèle classique

Afin de parfaire l'analyse du modèle classique, nous allons encore étudier deux variantes de ce modèle. La première de ces variantes consiste simplement à introduire un paramètre supplémentaire dans l'exponentielle décroissante de la composante irréversible. Nous proposons donc d'écrire :

$$P_2(d) = a_5 \cdot \exp(-\beta \cdot d) + a_6 \cdot \exp(d) + \epsilon_d$$

En apparence, ce modèle ressemble très fortement au modèle original. Toutefois, l'introduction de ce nouveau coefficient rend le modèle non-linéaire et complique de ce fait énormément le processus d'estimation des paramètres. Cependant, le but de cette étude n'étant pas une analyse théorique détaillée de ce nouveau modèle, nous nous sommes contentés de calibrer le modèle pour certaines valeurs du coefficient et de reporter sur un graphique la valeur du  $R^2$  du modèle en fonction de la valeur du paramètre  $\beta$ . Nous avons ainsi obtenu la figure 6.5. En observant cette figure, on s'aperçoit immédiatement que cette fonction possède un maximum local (et même global) pour  $\beta \approx 1.4$ . En effet, la fonction paraît constante entre 1.2 et 1.6, mais cela est uniquement dû à des faits d'arrondis. Il faudrait connaître les valeurs du  $R^2$  avec une

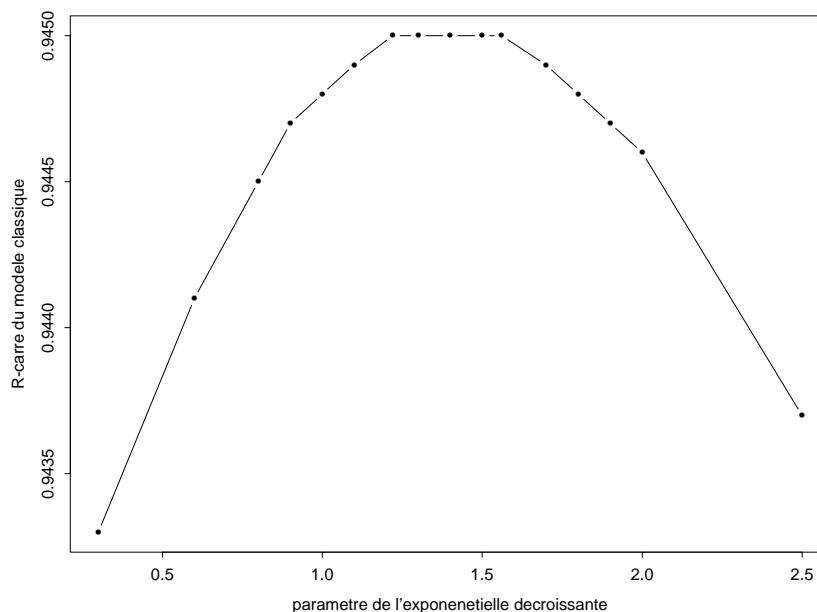
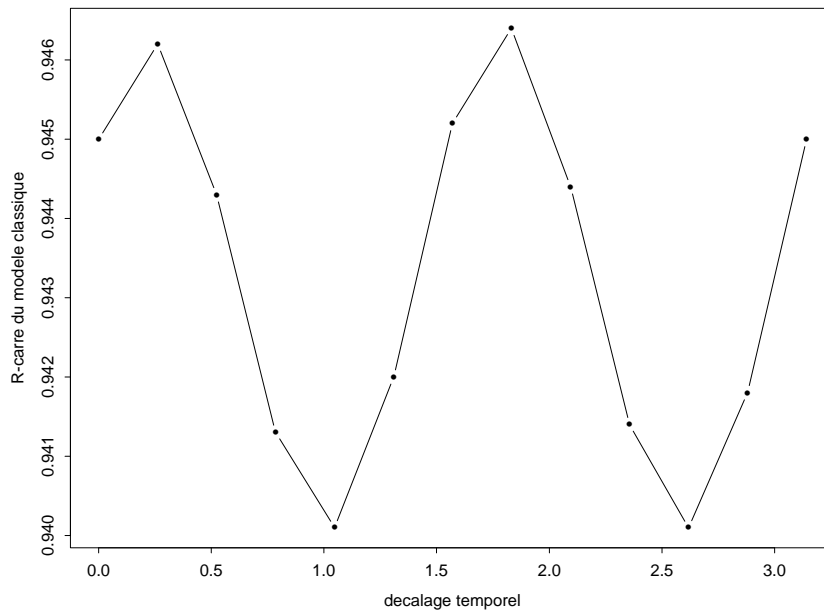


FIG. 6.5 –  $R^2$  en fonction du paramètre  $\beta$

plus grande précision. La valeur de 1.4 pour le maximum de la fonction a été obtenue par extrapolation et, la suite nous le montrera, nous n'avons besoin que d'une valeur approximative de ce maximum. Il convient également d'observer l'échelle de l'axe des ordonnées pour se rendre compte que la différence entre le  $R^2$  du modèle classique ( $\beta=1$ ) et le  $R^2$  maximal est infime (0.02%). Etant donné ces résultats, l'introduction de ce paramètre supplémentaire ne nous paraît pas nécessaire et cette première variante du modèle classique non seulement n'améliore pas la qualité de la calibration, mais induit également des complications calculatoires à cause de sa non-linéarité.

La seconde variante possède beaucoup de similitudes avec la première. Elle consiste aussi à introduire un nouveau coefficient et entraîne également une perte de linéarité. La modification porte cette fois sur la composante réversible thermique. Il s'agit d'introduire un décalage dans toutes les fonctions sinusoïdales. Cette idée repose sur l'observation suivante : la cote du lac atteint généralement son minimum au début du printemps, c'est-à-dire au quart de l'année environ. Or, la fonction  $\cos(s)$  atteint son minimum lorsque  $s=\pi$  (milieu de l'année) et la fonction  $\sin(s)$  atteint son minimum lorsque



FIG. 6.6 –  $R^2$  en fonction du paramètre  $\alpha$ 

$s = \frac{3\pi}{2}$ . Nous proposons donc d'écrire :

$$P_3(s) = a_7 \cdot \cos(s + \alpha) + a_8 \cdot \sin(s + \alpha) + a_9 \cdot \sin^2(s + \alpha) + a_{10} \cdot \sin(s + \alpha) \cdot \cos(s + \alpha) + \epsilon_s$$

Comme avant, nous avons procédé à une étude empirique de cette variante en faisant varier le paramètre de décalage  $\alpha$  entre 0 et  $\pi$  et nous avons alors obtenu la figure 6.6. On remarque tout de suite sur cette figure que la fonction obtenue possède un caractère périodique ce qui confirme que notre intuition était bonne. Toutefois, la variation du  $R^2$  entre les différentes valeurs de  $\alpha$  est beaucoup trop faible (0.1%) pour nous devions privilégier ce modèle au modèle classique. L'infime gain en qualité ne justifie pas la perte de la linéarité du modèle. Par conséquent, il semble préférable de privilégier le modèle classique à cette variante.

Remarquons toutefois qu'il faudrait idéalement calibrer chacune de ces variantes à l'aide de données provenant d'autres barrages. S'il s'avère que les maxima de ces fonctions s'obtiennent presque chaque fois pour les mêmes valeurs de  $\alpha$  ou de  $\beta$ , alors il conviendrait de reconsidérer le jugement effectué précédemment et d'introduire directement les valeurs correspondantes dans le modèle classique. On obtiendrait ainsi un modèle pour ainsi dire optimal

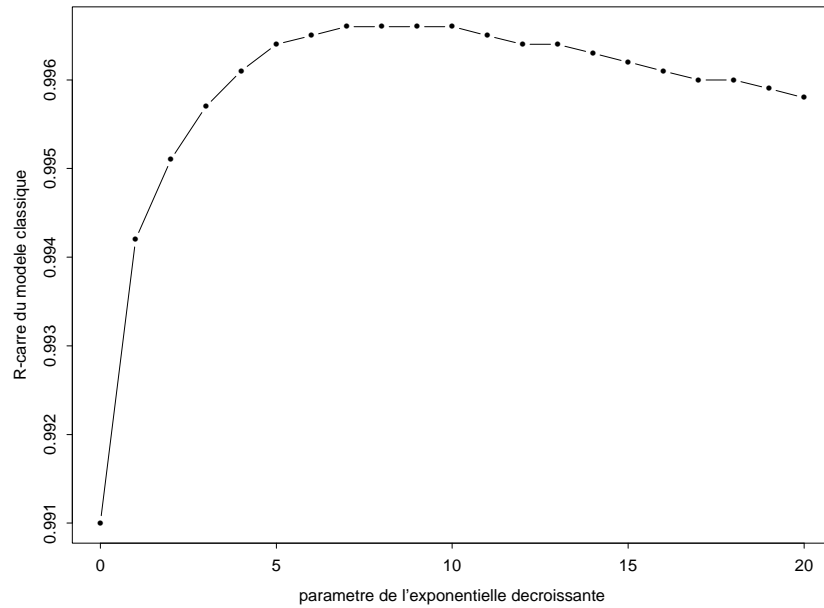
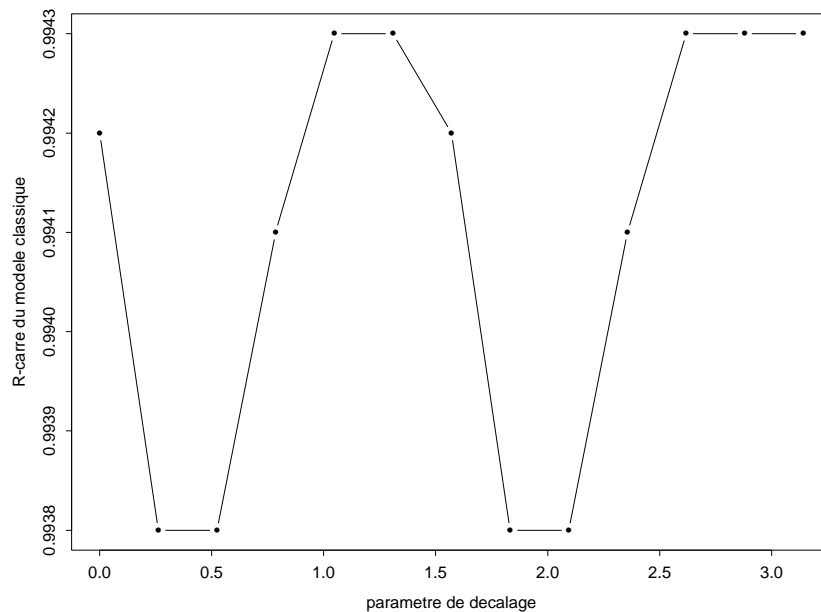


FIG. 6.7 –  $R^2$  en fonction du paramètre  $\beta$

sans avoir à renoncer à la linéarité. Malheureusement, nous ne possédons que deux jeux de données pour effectuer cette étude, ce qui est largement insuffisant. Nous avons donc calibré les deux variantes à l'aide du second jeu de données afin de pouvoir comparer les résultats avec ceux obtenus à partir du premier jeu de données. Les figures 6.7 et 6.8 sont les figures qui correspondent aux figures 6.5 et 6.6 respectivement.

En observant tout d'abord la figure 6.7, on remarque que la valeur de  $\beta$  qui maximise le  $R^2$  du modèle classique est comprise entre 7 et 10, ce qui est relativement différent de la valeur optimale ( $\beta=1.4$ ) trouvée dans le cas du premier jeu de données. Par contre, comme précédemment, la variation du  $R^2$  entre les valeurs  $\beta=1$  et  $\beta=8$  est infime (0.2%) et par conséquent négligeable. Sur la base de ces constatations, nous émettons un préavis négatif concernant cette variante du modèle classique. Il conviendrait toutefois de calibrer sur un nombre plus important de jeux de données. Concernant la seconde variante, la comparaison des graphiques permet de dégager dans les deux cas une structure assez similaire. On y retrouve notamment un caractère périodique. Par contre, les maxima de chacune des fonctions sont atteints pour des valeurs fort différentes les unes des autres. De ce fait, les mêmes conclusions que pour la première variante s'imposent, à savoir conserver le modèle classique comme

FIG. 6.8 –  $R^2$  en fonction du paramètre  $\alpha$ 

modèle de base.

Plus généralement, on peut encore remarquer que dans le cas du premier jeu de données, l'amélioration apportée par la première variante est environ six fois plus faible que celle apportée par la seconde variante, tandis que dans le cas du deuxième jeu de données, l'amélioration apportée par la seconde variante est 25 fois plus faible que celle apportée par la première variante. Cela tend à démontrer que ces variantes ont un effet très variable suivant le jeu de données sous-jacent, mais dans tous les cas elles n'apportent qu'une amélioration minimale au modèle. La conclusion que nous pouvons finalement tirer de l'étude de ces deux variantes est la suivante : les améliorations que nous avons voulu apporter au modèle permettent une dérisoire amélioration de la qualité de l'estimation qui ne justifie pas la complication calculatoire qui en découle.

Remarque : la première variante introduit un coefficient multiplicatif à l'exposant de l'exponentielle décroissante. On peut dès lors se demander pourquoi nous n'avons pas tester une troisième variante qui consisterait à introduire un coefficient multiplicatif à l'exposant de l'exponentielle croissante. En fait, ces deux variantes nous ont été suggérées par l'OFEG et leur étude ne fait que répondre à sa demande précise.



# Chapitre 7

## Le modèle série temporelle

Un modèle série temporelle (ST) est avant tout un modèle linéaire. La seule différence par rapport au modèle classique se situe au niveau des variables explicatives. Dans un modèle ST, les variables explicatives ne sont pas forcément simultanées à la variable que l'on souhaite estimer. On peut ainsi par exemple faire dépendre la valeur de la variable-réponse au temps  $t$  de la valeur de cette même variable aux temps  $t - 1$ ,  $t - 2$ ,  $\dots$ . On remarque aisément qu'il existe une multitude de modèles ST possibles pour modéliser une même variable-réponse. Toutefois, une fois le modèle choisi, le processus d'estimation des paramètres  $a_i$  consiste simplement à effectuer une régression par moindres carrés. Comme avant, nous éliminerons progressivement du modèle les variables non significatives pour aboutir finalement au modèle final. La principale difficulté avec ce type de modèle réside dans le fait de trouver un modèle initial adéquat. Dans le cadre de cette étude, nous avons opté pour un modèle qui s'inspire dans une large mesure de notre intuition du problème et dans une moindre mesure de la physique du problème, c'est-à-dire du modèle classique. Voici donc notre modèle initial :

$$\begin{aligned} a_{\text{rad}}[t] = & a_0 + a_1 \cdot a_{\text{rad}}[t - 1] + a_2 \cdot a_{\text{rad}}[t - 2] \\ & + a_3 \cdot h[t] + a_4 \cdot h[t - 1] + a_5 \cdot h^2[t] + a_6 \cdot h^2[t - 1] \\ & + a_7 \cdot h^3[t] + a_8 \cdot h^3[t - 1] + a_9 \cdot h^4[t] + a_{10} \cdot h^4[t - 1] \\ & + a_{11} \cdot d[t] + a_{12} \cdot d[t - 1] + \epsilon \end{aligned}$$

$a_{\text{rad}}[t]$  représente la mesure de la déviation radiale du pendule A au temps  $t$  et  $a_{\text{rad}}[t - 1]$  la mesure cette déviation radiale au temps  $t - 1$ . Cette notation est bien évidemment la même pour les variables  $d$  et  $h$ . Comme avant, le

but sera d'estimer par moindres carrés les coefficients  $a_i$  et nous utiliserons à nouveau le logiciel S-Plus pour cela. Les résultats de cette estimation sont présentés de la section suivante.

## 7.1 Résultat de la calibration

Le tableau 7.1 synthétise les résultats obtenus en calibrant un modèle ST à partir des 132 observations normalisées provenant du barrage du Haut-Valais.

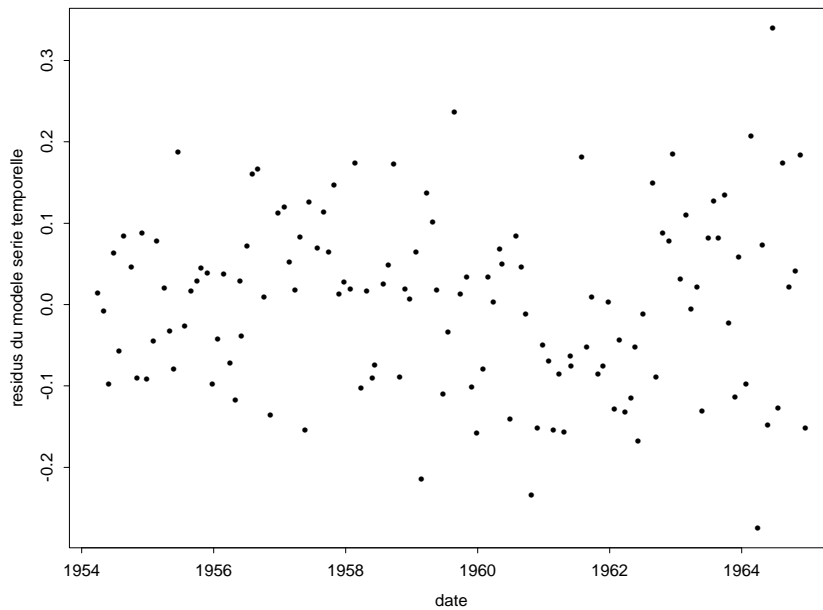
TAB. 7.1 – Calibration du modèle série temporelle

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$a_{\text{rad}}[t - 1]$	0.7847	0.0812	9.6598	0.0000
$a_{\text{rad}}[t - 2]$	-0.1581	0.0548	-2.8825	0.0046
$d[t]$	-0.1045	0.0320	-3.2693	0.0014
$h^2[t]$	0.5726	0.0397	14.4388	0.0000
$h^2[t - 1]$	-0.2706	0.0590	-4.5886	0.0000
Ecart-type des résidus : 0.1094				
$R^2 = 0.9638$				

Remarque : pour calibrer ce modèle, on a bien utilisé les 132 observations, mais certaines d'entre elles de façon partielle seulement. En effet, lorsque  $t = 0$ , nous ne connaissons pas les valeurs de  $a_{\text{rad}}[t - 1]$  et  $a_{\text{rad}}[t - 2]$ . Par conséquent, les valeurs des coefficients qui minimisent le carré moyen de l'erreur ont été calculés sur la base de 130 équations seulement.

## 7.2 Analyse des résultats

Afin de pouvoir effectuer une comparaison objective des deux modèles, nous allons analyser les mêmes points que dans le cas du modèle classique. Remarquons tout d'abord que le modèle obtenu ne comporte que cinq variables explicatives et possède un  $R^2$  de 0.9638 qui est donc encore plus élevé que le  $R^2$  du modèle précédent qui était de 0.9448. On arrive donc à un modèle aussi bon si ce n'est meilleur avec une variable explicative de moins.

FIG. 7.1 – *Résidus du modèle ST*

Il est assez délicat d’essayer d’interpréter le signe des coefficients, mais on peut tout de même remarquer que à la fois  $a_{\text{rad}}[t - 1]$  et  $a_{\text{rad}}[t - 2]$  et  $h^2[t]$  et  $h^2[t - 1]$  sont de signes opposés. Une interprétation possible de ceci serait que  $a_{\text{rad}}[t - 2]$  et  $h^2[t - 1]$  agissent comme des correcteurs des effets exagérés de  $a_{\text{rad}}[t - 1]$  et  $h^2[t]$  respectivement. Toutefois, il est difficile de justifier cette explication en se basant sur la physique du problème.

La figure 7.1 représente les valeurs des résidus pour chaque observation. Contrairement au cas précédent, il est impossible de dégager une structure parmi les résidus. Cette absence de structure est un aspect fort positif. On distingue tout au plus quelques valeurs “extrêmes” ce qui est parfaitement normal dans ce type de régression.

Sur la figure 7.2, on retrouve une comparaison de la densité des résidus avec une densité normale standard. Comme dans le cas du modèle classique, les queues de la distribution ne coïncident que faiblement avec celles de la loi normale. Mais le comportement de ces queues permet de voir que les queues de la densité des résidus sont moins lourdes que celles de la loi normale, ce qui est une excellente propriété. En effet, cela signifie que les résidus possèdent une petite variance. De plus, la densité des résidus se calque partout ailleurs assez bien avec celle de la loi normale. On peut en tous les cas conclure que

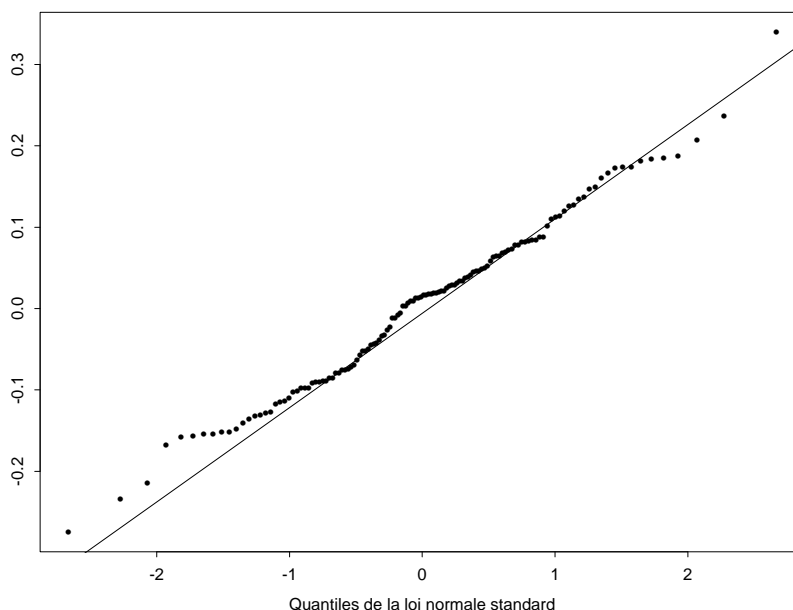


FIG. 7.2 – *Comparaison de la densité des résidus avec une densité normale*

la distribution des résidus se rapproche beaucoup plus d'une distribution normale dans le cas du modèle ST que dans le cas du modèle classique et que la probabilité d'avoir de grands résidus est beaucoup plus faible dans le cas que nous étudions que dans le cas précédent. Cet argument plaide à nouveau en faveur du modèle développé dans ce chapitre.

La figure 7.3 représente les autocorrélations entre les résidus. On remarque immédiatement qu'elles sont beaucoup plus faibles que dans le modèle classique. Cela peut en partie s'expliquer par le fait que dans les modèles de type ST les variables explicatives sont fortement corrélées entre elles ce qui entraîne de faibles corrélations entre les résidus. Toutefois, les immenses différences dans les valeurs des autocorrélations des deux modèles sont trop importantes pour n'être imputées qu'au caractère série temporelle du modèle que nous analysons actuellement. En fait, d'une part les autocorrélations du modèle ST sont très faibles, et d'autre part les autocorrélations du modèle classique sont très fortes. De plus, la fonction  $\rho(k)$  ne dépasse presque jamais les bornes de l'intervalle de confiance, ce qui ne contredit pas l'hypothèse que les résidus sont les réalisations d'un bruit blanc. Signalons toutefois que cela ne prouve pas non plus que les résidus sont les réalisations d'un bruit blanc.

En guise de conclusion, nous pouvons tout d'abord remarquer que tous



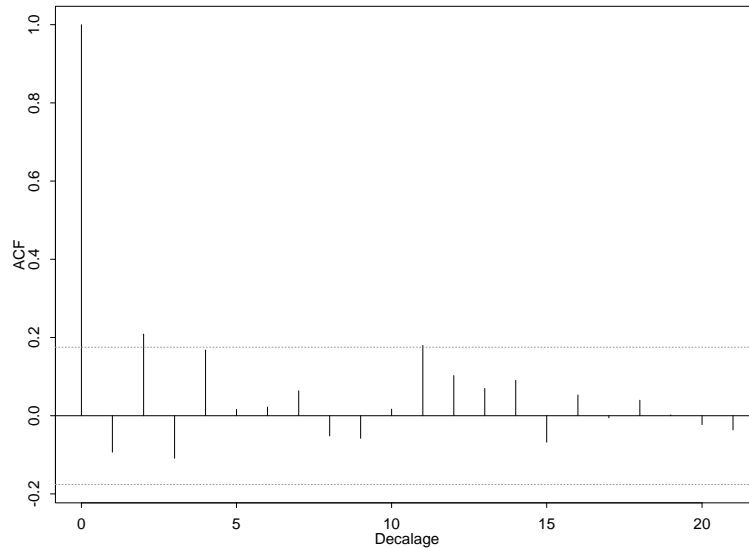


FIG. 7.3 – Autocorrélations des résidus du modèle ST

les aspects abordés au cours de l'analyse des résultats de ce modèle tendent à indiquer qu'il possède de meilleures propriétés que le modèle classique. En effet, il nécessite un nombre inférieur de variables explicatives pour aboutir à un  $R^2$  légèrement supérieur. La faible corrélation de ses résidus, ainsi que leur distribution quasi normale, deux propriétés que ne possèdent pas le modèle classique, sont des signes irréfutables de qualité. Ensuite, il convient de préciser qu'en terme de calibration, les deux modèles se traitent de façon parfaitement analogue et donc que les temps de calibration respectifs des modèles sont relativement semblables. Finalement, afin de vérifier la pertinence de cette analyse, il reste à effectuer la même prédiction que dans le cas du modèle classique, puis de comparer les résultats obtenus.

### 7.3 Prédiction à l'aide du modèle ST

Le processus utilisé pour effectuer la prédiction est en tout point semblable à celui utilisé dans le cas du modèle classique. La figure 7.4 permet de visionner conjointement les vraies valeurs du déplacement radiale du pendule A et les valeurs estimées à l'aide du modèle ST. Une comparaison visuelle entre la figure 6.4 et la figure 7.4 permet immédiatement de se rendre compte

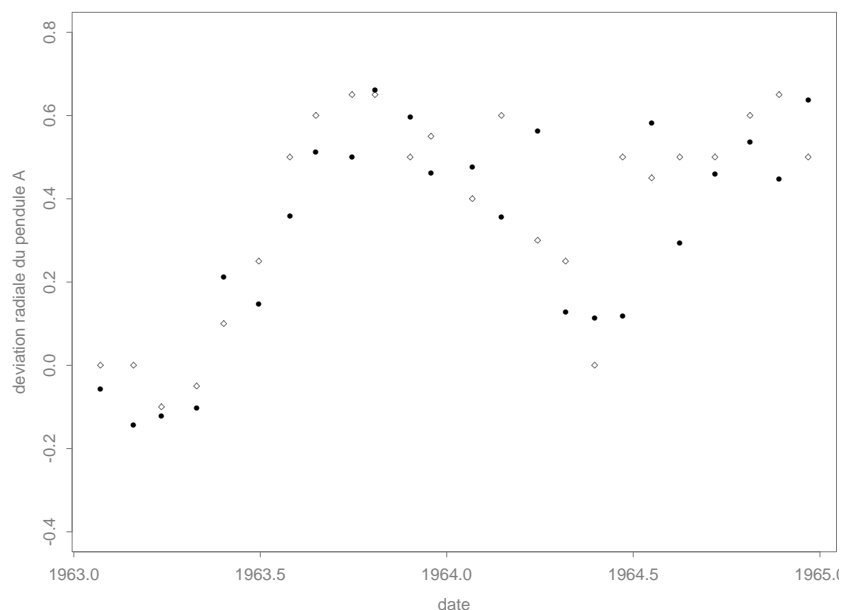


FIG. 7.4 – Valeurs estimées (●) et valeurs réelles (◇)

que la qualité de la prédiction est meilleure dans le cas d'un modèle ST. Ce résultat était quelque peu prévisible si l'on se réfère aux résultats des calibrations et aux analyses des résidus. En effet, ces deux aspects avaient déjà permis de dégager des conclusions favorables au modèle de type série temporelle. Le résultat obtenu grâce à ces prédictions ne fait que confirmer les conclusions tirées précédemment. Toutefois, ces conclusions reposent sur la comparaison de résultats obtenus à partir d'un unique jeu de données et de ce fait, il convient de les traiter avec une certaine prudence. Afin de les confirmer ou de les infirmer, nous avons effectué un travail de calibration et d'analyse tout à fait analogue à partir du second jeu de données et le chapitre qui suit présente de façon assez condensée les résultats obtenus.

# Chapitre 8

## Calibrations à l'aide du second jeu de données

Dans ce chapitre, nous nous contenterons dans un premier temps de présenter les résultats obtenus sans y ajouter beaucoup de commentaires, puis dans un second temps seulement nous analyserons ces résultats avec l'objectif essentiel de les comparer à ceux obtenus à partir du premier jeu de données. Nous nous sommes concentrés sur un seul pendule, que nous appellerons pendule H, pour effectuer toutes les calibrations présentées dans ce chapitre. Comme avant, un bref descriptif des données concernant cette déviation peut être consultée en annexe sur la figure A.2.

### 8.1 Le modèle classique

Le tableau 8.1 résume les résultats de la calibration du modèle classique à partir du second jeu de données. La méthodologie utilisée est strictement identique à celle utilisée dans les calibrations précédentes.

Le modèle obtenu comporte 7 variables significatives, ce qui représente une variable de plus que lors de la calibration à partir du premier jeu de données. La valeur du  $R^2$  est excellente.

### 8.2 Le modèle série temporelle

Le meilleur modèle ST obtenu à partir du second jeu de données est synthétisé dans le tableau 8.2.

TAB. 8.1 – Calibration du modèle classique

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$h$	36.7769	4.2899	8,5730	0.0000
$h^2$	-28.2451	7.8501	-3.5981	0.0004
$h^4$	50.7733	3.6268	13.9993	0.0000
$\cos(s)$	2.9448	0.2747	10.7188	0.0000
$\sin(s)$	8.7445	0.5015	17.4352	0.0000
$\sin(s) \cdot \cos(s)$	-2.1028	0.4190	-5.0185	0.0000
$\exp(-d)$	-8.1120	0.6088	-13.3251	0.0000
Ecart-type des résidus : 2.116				
$R^2 = 0.9942$				

TAB. 8.2 – Calibration du modèle série temporelle

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$h_{\text{rad}}[t-1]$	0.8061	0.0117	68.6526	0.0000
$h^4[t]$	33.2406	0.6133	54.1985	0.0000
$h^4[t-1]$	-20.7692	0.9618	-21.5934	0.0000
Ecart-type des résidus : 1.8070				
$R^2 = 0.9957$				

Le modèle obtenu comporte seulement 3 variables explicatives, c'est-à-dire deux de moins que précédemment. Ici encore, la valeur du  $R^2$ , proche de 1, est excellente. Avant d'étudier les caractéristiques des résidus de chacun des modèles, il convient de faire une remarque. En effet, si l'on compare les deux modèles ST calibrés jusqu'à présent, on s'aperçoit que la variable explicative  $h$  apparaît une fois avec l'exposant 2 et une fois avec l'exposant 4. Suite à cette observation, nous avons analysé la variation du  $R^2$  en fonction de l'exposant de la variable  $h$ . Les résultats sont synthétisés sur la figure 8.1.

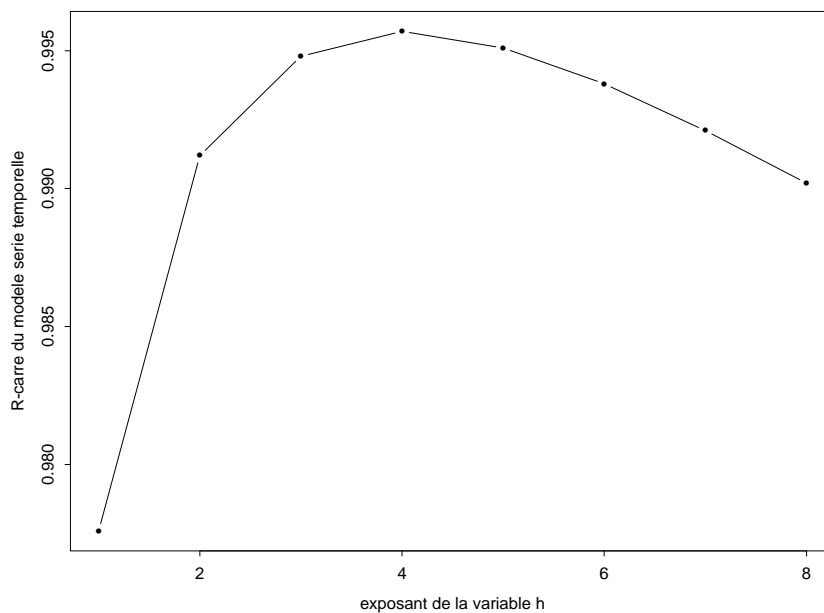


FIG. 8.1 –  $R^2$  en fonction de l'exposant de  $h$

Cette figure permet de constater que la variation du  $R^2$  entre 2 et 4 est très faible. On pourrait donc remplacer sans perte de qualité  $h^4$  par  $h^2$ , le but de cette substitution étant d'obtenir un modèle ST le plus général possible. Idéalement, on souhaiterait arriver à un modèle commun à un maximum de barrages où le travail de calibration se résumerait à la seule estimation des coefficients. Toutes les étapes successives qui permettent de détecter les variables non significatives seraient ainsi supprimées, ce qui engendrerait un gain de temps substantiel. Mais avant d'aboutir un jour peut-être à ce résultat, il conviendra d'effectuer un certain nombre d'analyses de modèles ST, bien plus poussées que celle qui est présentée dans ce document.

### 8.3 Analyse des résultats

Intéressons-nous pour commencer à la structure des résidus. Une rapide observation de la partie gauche de la figure 8.2 permet immédiatement de remarquer que les résidus du modèle classique (en haut) possèdent une structure, tandis qu'il est impossible de déceler la moindre structure dans les résidus du modèle ST (en bas). On retrouve donc exactement la même situation que lors des calibrations sur les données provenant du premier barrage.

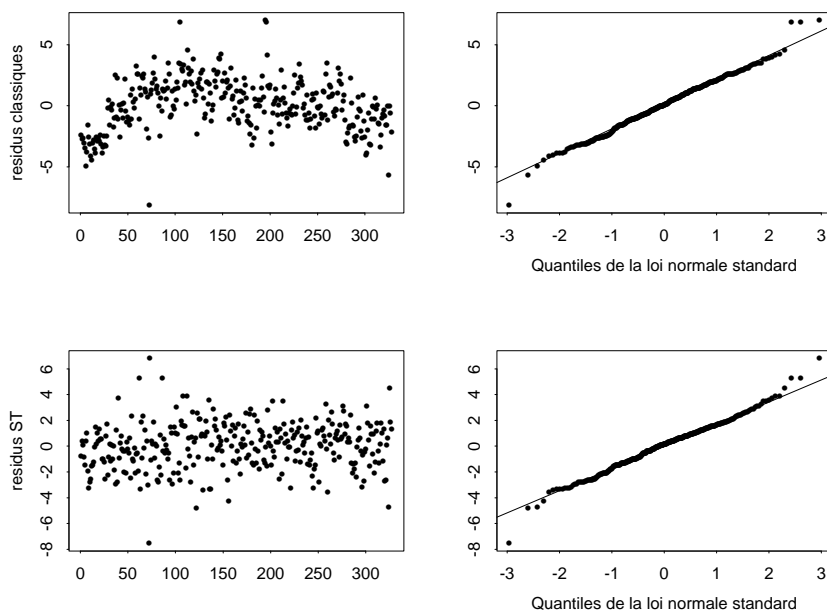


FIG. 8.2 – *Structure et densité des résidus*

Si l'on observe maintenant sur la partie droite de la figure 8.2 les densités de ces mêmes résidus, on s'aperçoit que les deux densités approximent extrêmement bien la densité normale standard. Dans le cas des résidus du modèle classique, cela est plutôt surprenant, car le fait pour une variable de posséder une structure est généralement incompatible avec le fait de suivre une loi normale. On peut également remarquer que les deux densités possèdent des queues plus lourdes que celles de la loi normale, mais si l'on supprime les huit résidus les plus extrêmes parmi les 328 observations, on obtient une densité qui se calque très bien sur la loi normale centrée réduite. En fait, l'analyse de ces densités ne permet pas ici de tirer des enseignements importants. De ce fait, nous avons appliqué le test de Kolmogorov-Smirnov

aux résidus de chacun des modèles. Le test de Kolmogorov-Smirnov (voir Kendall et Stuart, 1979) permet de tester si une suite de variables aléatoires possèdent une distribution normale. Dans notre exemple, l'hypothèse d'une distribution normale n'a pu être rejetée dans aucun des cas, ce qui témoigne tout de même de la qualité de chacun des modèles.

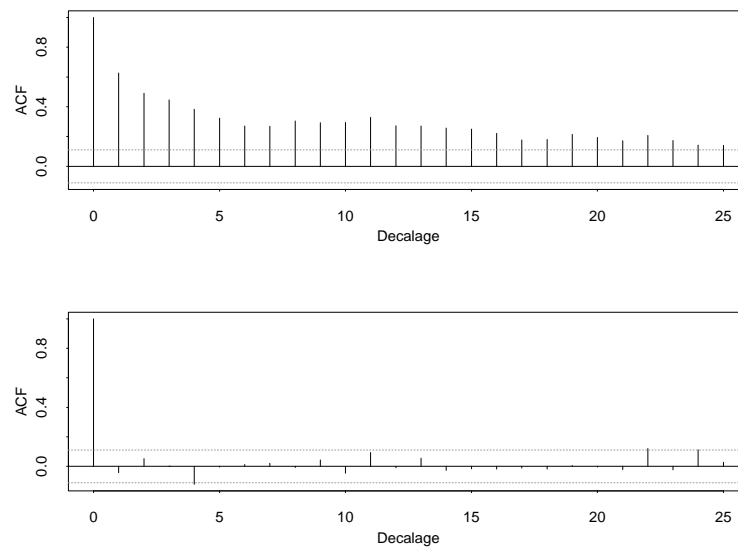


FIG. 8.3 – *Autocorrélations des résidus*

Etudions finalement les autocorrélations des résidus pour chacun des modèles. La figure 8.3 se passe presque de commentaire. Les résidus du modèle de ST sont très faiblement corrélés, tandis que ceux issus du modèle classique sont très fortement corrélés. Cette forte corrélation n'est pas surprenante à la vue de la structure très marquée des résidus. En effet, il est assez intuitif que la présence d'une structure implique une corrélation des données. Les résidus du modèle classique ne font ici que confirmer cette intuition.

Cette brève analyse des résultats permet simplement de confirmer les conclusions obtenues suite à l'analyse des calibrations précédentes. Elle met en exergue à la fois les qualités du modèle ST et les imperfections du modèle classique. Afin de vérifier la justesse de cette analyse, il reste à effectuer des prédictions à partir de chacun des modèles sur le même schéma qu'au paravant. Ces prédictions, ainsi que leur analyse, sont l'objet de la section suivante.

## 8.4 Prédications à partir des deux types de modèles

Le jeu de données comprend 24 années d'observations. Nous avons donc recalibré chacun des modèles en éliminant les trois dernières années d'observations. La prédiction porte donc sur ces 3 années. La figure 8.4 représente la prédiction relative au modèle classique, tandis que la figure 8.5 concerne la prédiction à partir du modèle ST.

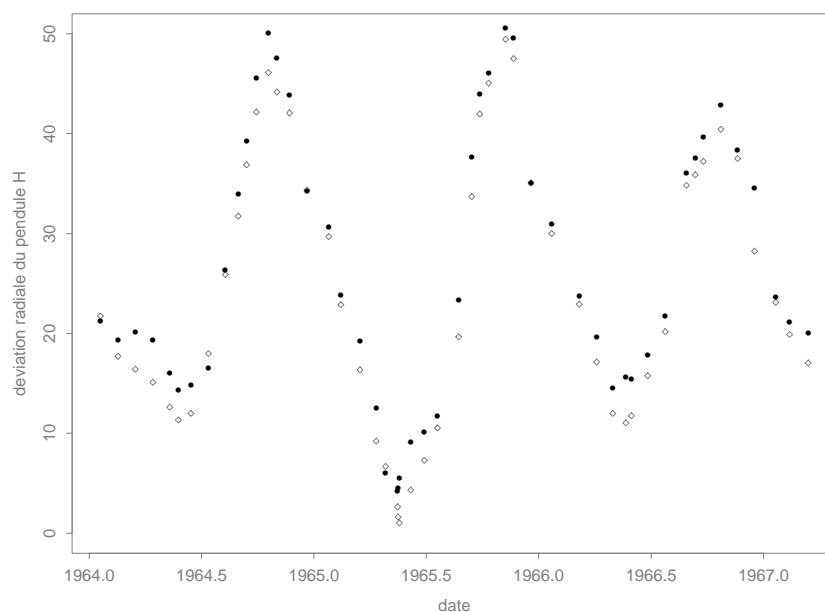
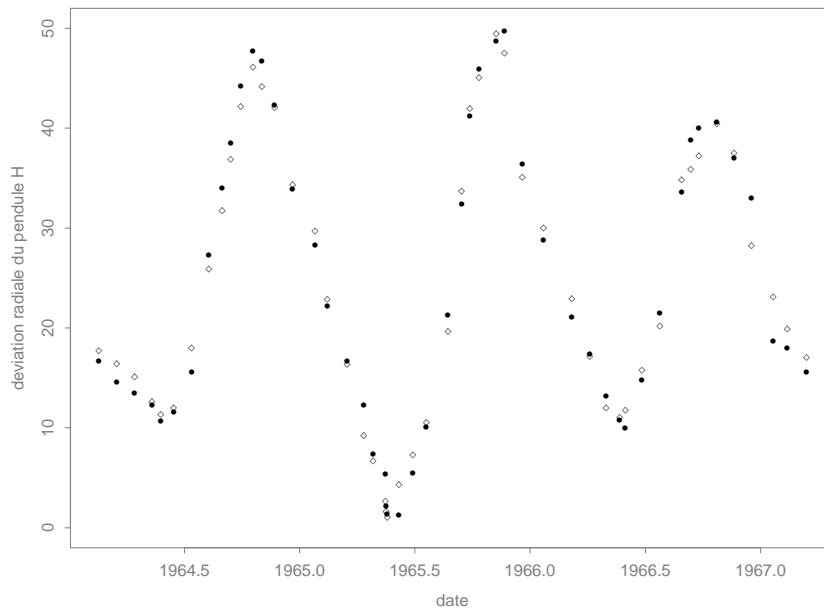


FIG. 8.4 – *Modèle classique : valeurs estimées (●) et valeurs réelles (◇)*

Comme précédemment, en observant attentivement les figures 8.4 et 8.5, on se rend compte que la qualité de la prédiction est meilleure dans le cas du modèle ST. Cela se remarque essentiellement aux extrema de la fonction. On remarque également que la prédiction réalisée à partir du modèle classique a tendance à sur-estimer les valeurs réelles. Rappelons que dans le cas précédent, cette prédiction avait tendance à sous-estimer les valeurs réelles. Cela ne fait que confirmer l'idée que le modèle classique n'est peut-être pas le modèle adéquat pour prédire les déplacements des murs des barrages.

Remarque: afin de pouvoir juger de manière plus objective la qualité d'une prédiction, nous avons calculé des intervalles de confiance pour les va-



FIG. 8.5 – *Modèle ST: valeurs estimées (●) et valeurs réelles (◇)*

leurs prédites. Il s'agit en fait d'intervalles de confiance ponctuels calculés individuellement pour chacune des valeurs. Ces intervalles sont centrés en la valeur prédite et leur demi-longueur est le produit du quantile à 97.5% de la loi normale centrée réduite et de l'écart-type de la valeur prédite. Le code S-Plus pour construire de telles intervalles de confiance est disponible en annexe. A titre d'exemple, nous avons calculé les intervalles de confiance dans le cas de la prédiction par le modèle ST pour les trois dernières valeurs de la prédiction, c'est-à-dire les observations 130, 131 et 132. Nous avons répercuté les résultats dans le tableau 8.3. Les valeurs  $y_{inf}$  et  $y_{sup}$  représentent les bornes

TAB. 8.3 – *Intervalles de confiance ponctuels*

observation	valeur réelle	prédiction	$y_{inf}$	$y_{sup}$
130	0.6	0.5371	0.4775	0.5966
131	0.65	0.4455	0.3887	0.5024
132	0.5	0.6365	0.5616	0.7115

inférieures et supérieures de ces intervalles de confiance. Le premier constat qui s'impose en consultant le tableau semble inquiétant. En effet, aucune des

vraies valeurs n'appartient à l'intervalle de confiance. On pourrait facilement en conclure que la prédiction est mauvaise. Toutefois, il ne s'agit que d'intervalles ponctuels et en fait, ils ne permettent pas de juger la prédiction dans sa globalité. Cet exemple sert simplement à illustrer comment calculer de tels intervalles, mais a aussi pour but de mettre en garde contre les dangers d'une mauvaise interprétation des résultats. En fait, dans le cas d'une régression linéaire multiple, il convient de calculer ce que l'on appelle des bandes de confiance. Malheureusement, le calcul de ces bandes est extrêmement difficile et c'est la raison pour laquelle nous n'avons pas traité cet aspect lors de la prédiction. Pour obtenir néanmoins certains détails sur la manière de calculer des bandes de confiance, nous conseillons le lecteur de se référer à des ouvrages statistiques spécifiques à ce domaine (voir par exemple Sun, J. , Loader, C. and McCormick, W. P. , 2000).

# Chapitre 9

## Le modèle mixte

En marge de cette comparaison entre les modèles classiques et ST, nous avons développé un troisième type de modèle. Ce modèle est en fait un mixte des deux modèles précédents puisqu'il combine des variables explicatives qui apparaissent dans l'un ou l'autre des modèles. Ce modèle est très personnel et nous nous sommes contentés de le calibrer sur le premier jeu de données. Les résultats obtenus sont présentés dans le tableau 9.1.

TAB. 9.1 – *Calibration du modèle mixte*

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$a_{\text{rad}}[t - 1]$	0.6030	0.0632	9.5444	0.0000
$h^2[t]$	0.4968	0.0433	11.4689	0.0000
$h^2[t - 1]$	-0.2155	0.0583	-3.6938	0.0003
$\sin(s)$	-0.0559	0.0140	-3.9940	0.0001
$\exp(-d)$	0.0866	0.0411	2.1061	0.0016
$\exp(d)$	-0.0437	0.0136	-3.2203	0.0372
Ecart-type des résidus : 0.1070				
$R^2 = 0.9657$				

Commençons par remarquer que le  $R^2$  de ce modèle (0.9657) est supérieure aux  $R^2$  des deux modèles précédemment calibrés à partir de ce même jeu de données (0.9448 et 0.9638), alors que le nombre de variables explicatives est sensiblement le même que dans les modèles précédents. Ces premières observations laissent donc supposer que ce type de modèle est aussi performant, si ce n'est plus performant que les modèles classique et ST. Pour confirmer

cette idée, il conviendrait d'analyser plus en détail les caractéristiques de ce modèle et notamment et d'étudier les propriétés des résidus. Nous nous contenterons ici de présenter les principaux graphiques sans entrer dans les détails de l'analyse.

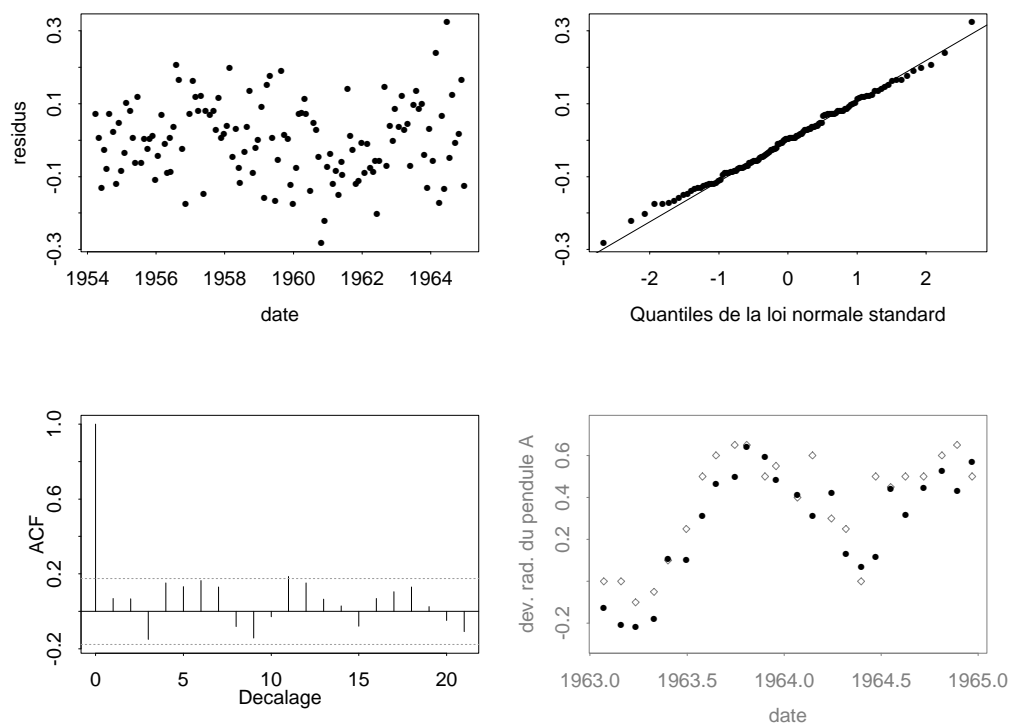


FIG. 9.1 – Analyse graphique du modèle mixte

Une simple observation de la figure 9.1 permet tout de suite d'affirmer que ce modèle n'est pas mauvais et donc qu'il mérite une certaine attention. Toutefois, le modèle mixte présenté n'est qu'un modèle parmi beaucoup d'autres et pour juger de ses qualités réelles, il faudrait analyser de façon plus poussée la grande famille des modèles mixtes. Ce travail n'entre pas dans le cadre de ce projet et nous laissons donc la porte ouverte à de futures recherches dans le domaine de la prédiction des déplacements des murs des barrages basées sur les modèles mixtes. Avant de présenter les conclusions de cette étude, il convient de justifier certains choix effectués avant et pendant les calibrations.

# Chapitre 10

## Quelques justifications

### 10.1 Suppression de données

Toutes les calibrations effectuées jusqu'à présent et relatives au premier jeu de données ne reposent que sur 132 observations du jeu de données. 19 observations ont donc été purement et simplement éliminées. Les raisons de cette suppression de données ont déjà été expliquées auparavant. Il est toutefois légitime de se demander si cette perte d'information conduit à un modèle fort différent de celui que l'on obtient en utilisant les 151 observations disponibles. Le meilleur moyen de répondre à cette question consiste simplement à calibrer le modèle classique à l'aide de l'ensemble des observations. En effectuant ceci on obtient le modèle présenté dans le tableau 10.1. En comparant les valeurs des coefficients obtenus dans la calibration présentée sur le tableau 6.1 avec ceux ci-dessus, il apparaît évident que les deux modèles ne sont pas significativement différents l'un de l'autre et que par conséquent, la suppression de ces 19 observations n'a à aucun moment faussé les résultats.

### 10.2 Normalisation des données

Toutes les calibrations effectuées jusqu'à présent se sont basées sur des données normalisées. Le processus de normalisation est d'ailleurs décrit dans le tableau 4.1. Cette normalisation n'est en fait qu'une transformation linéaire des données et a pour but d'une part d'éliminer l'effet de caractéristiques spécifiques à un barrage et d'autre part de pouvoir travailler avec des nombres raisonnables. Par exemple, il existe des barrages dont la hauteur du plan

TAB. 10.1 – Calibration avec la totalité des données

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$h^2$	0.5197	0.0452	11.4934	0.0000
$\cos(s)$	0.0597	0.0173	2.4510	0.0007
$\sin(s)$	-0.0576	0.0169	-3.4171	0.0008
$\sin^2(s)$	0.1223	0.0333	3.6713	0.0003
$\exp(-d)$	0.3778	0.0455	8.3117	0.0000
$\exp(d)$	-0.1154	0.0155	-7.4455	0.0000
Ecart-type des résidus : 0.1376				
$R^2 = 0.9410$				

d'eau ne dépasse pas 30 mètres alors que la hauteur du plan d'eau d'autres barrages dépasse largement 200 mètres. Afin d'éviter que cette différence n'induisse un modèle trop spécifique au barrage, il est utile de normaliser la variable  $h_0$ . De même, si pour représenter l'âge du barrage, on utilise le nombre d'années écoulées depuis sa construction, on aboutit après 20 ans à des termes de l'ordre de  $e^{20}$  qui induisent des problèmes numériques avec les ordinateurs. Afin de vérifier la pertinence de cette normalisation, nous avons calibré le modèle classique sans utiliser aucune normalisation. Nous avons obtenu le modèle résumé dans le tableau 10.2.

TAB. 10.2 – Calibration avec des données non normalisées

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$h_0^2$	0.0000	0.0000	14.5304	0.0000
$\cos(s)$	0.1013	0.0231	4.3924	0.0007
$\sin(s)$	-0.0981	0.0228	-4.2921	0.0000
$\sin^2(s)$	0.1003	0.0478	2.0999	0.0377
$\exp(-d_0)$	0.1868	0.0573	3.2614	0.0014
$\exp(d_0)$	-0.0000	0.0000	-2.7737	0.0064
Ecart-type des résidus : 0.1859				
$R^2 = 0.8978$				

En comparant le  $R^2$  du modèle calibré avec des données normalisées (0.9448) et celui du modèle ci-dessus (0.8978), on s'aperçoit que le fait de normaliser les données permet d'obtenir un meilleur modèle. De plus cela

permet d'éviter que les coefficients  $h_0^2$  et  $\exp(d_0)$  prennent des valeurs trop petites (dans l'exemple ci-dessus inférieures à  $10^{-4}$ ). L'idée de la normalisation s'avère donc être une excellente idée qui de plus conduit au modèle le plus général possible.

### 10.3 Choix du pendule

Afin de pouvoir comparer objectivement les différents modèles calibrés, nous nous sommes concentrés pour chacun des jeux de données sur un unique pendule. Nous avons évidemment au préalable étudié brièvement le comportement des différents pendules et alors remarqué qu'ils se comportaient tous de façon plus ou moins similaire. Il existe bien sûr des différences entre les déviations radiale et tangentielle, mais en considérant de ces premières observations, il ne nous est pas apparu nécessaire de traiter chaque pendule individuellement. On peut tout de même légitimement se demander si les résultats auraient été différents si l'on avait effectué les calibrations pour d'autres pendules que les pendules A et H. Pour répondre à cette question, nous avons utilisé les deux modèles calibrés au chapitre 8 et dont les résultats sont présentés dans les tableaux 8.1 et 8.2 respectivement. Nous avons alors simplement recalibré ces modèles à partir d'un autre pendule (pendule M) pour obtenir les deux modèles présentés dans les tableaux 10.3 et 10.4.

TAB. 10.3 – Calibration du modèle classique (pendule M)

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$h$	-2.9905	1.5618	-1.9148	0.0564
$h^2$	35.8803	2.8579	12.5547	0.0000
$h^4$	4.6056	1.3204	3.4880	0.0006
$\cos(s)$	1.0244	0.1000	10.2424	0.0000
$\sin(s)$	3.1925	0.1826	17.4839	0.0000
$\sin(s) \cdot \cos(s)$	-0.3310	0.1525	-2.1699	0.0308
$\exp(-d)$	-1.9938	0.2216	-8.9959	0.0000
Ecart-type des résidus : 0.7703				
$R^2 = 0.9986$				

En observant notamment les valeurs des  $R^2$ , mais aussi le fait que presque tous les coefficients sont significatifs, on constate immédiatement que ces

TAB. 10.4 – Calibration du modèle série temporelle (pendule M)

coefficient	valeur	écart-type	$t$ -valeur	$p$ -valeur
$m_{\text{rad}}[t - 1]$	0.8202	0.0180	45.5953	0.0000
$h^4[t]$	27.4431	0.5275	52.0278	0.0000
$h^4[t - 1]$	-19.5508	0.9050	-21.6034	0.0000
Ecart-type des résidus : 1.6060				
$R^2 = 0.9937$				

deux modèles sont bons. Cela tend à démontrer une certaine symétrie entre les pendules que l'on peut déjà comprendre intuitivement. En effet, si le mur d'un barrage avance en un point de contrôle, on imagine aisément qu'il avance également à tous les autres points de contrôle et donc que les différents pendules réagissent de manière assez uniforme. Le fait de se restreindre à l'analyse d'un seul pendule ne réduit donc pas sensiblement la généralité de cette étude.

Remarque : au cours de cette étude, nous avons calibré 3 modèles ST pour les pendules A, H et M respectivement. En tenant compte de la remarque qui concerne l'exposant de  $h$  au chapitre 8, on peut affirmer que plusieurs variables explicatives apparaissent simultanément dans chacun de ces modèles. Nous avons alors recalibré le modèle ST sur chacun de ces pendules en utilisant pour seules variables explicatives la déviation au temps  $t - 1$ , la cote du lac au carré au temps  $t$  et la cote du lac au carré au temps  $t - 1$ . Nous avons ensuite répercuté dans le tableau 10.5 un certain nombre de résultats que nous allons maintenant détailler. Pour chacun des modèles, on retrouve d'une part la valeur des trois coefficients estimés et d'autre part le  $R^2$  du modèle. La dernière ligne du tableau indique ce que nous avons appelé le  $\Delta R^2$ , c'est-à-dire le rapport entre le  $R^2$  du modèle précédemment calibré et le  $R^2$  du modèle que nous avons calibré pour effectuer cette comparaison. Commençons par observer que dans chacun des cas le  $\Delta R^2$  est très proche de 1, ce qui signifie que les nouveaux modèles ne sont pas significativement différents des précédents. Le modèle calibré pour le pendule M est même meilleur. Cela tient au fait que le modèle précédemment calibré pour ce pendule avait simplement été calibré de façon à permettre une comparaison avec le modèle calibré pour le pendule H et justifier ainsi le choix arbitraire du pendule. On peut déjà affirmer que les trois modèles que nous



TAB. 10.5 – Comparaison de modèles ST

coefficient	pendule A	pendule H	pendule M
$dev[t - 1]$	0.7657	0.5875	0.6222
$h^2[t]$	0.5187	28.5198	29.4162
$h^2[t - 1]$	-0.3637	-9.1151	-16.5236
$R^2$	0.9580	0.9912	0.9983
$\Delta R^2$	0.994	0.9954	1.0046

venons de calibrer sont d'excellents modèles. Si l'on observe plus en détail la valeur des différents coefficients, nous pouvons tout d'abord remarquer que l'ordre de grandeur du coefficient relatif à la déviation au temps  $t - 1$  est toujours le même, alors que tous les pendules ne se trouvent pas sur le même barrage! Nous pouvons ensuite observer que l'ordre de grandeur du rapport des deux coefficients qui se rapportent à la cote du lac est également dans les trois modèles sensiblement le même. Signalons finalement que le nombre de variables explicatives présentes dans ces modèles est très faible et que ce sont malgré tout des modèles de grande qualité. Le but de cette comparaison de modèles ST est simplement de montrer qu'il existe certainement des possibilités de construire un modèle de ce type qui permette d'effectuer des prédictions non pas sur un unique barrage, mais sur tout un ensemble de barrages. Le modèle présenté ici ne comporte que trois variables explicatives, mais il parvient à modéliser de manière très satisfaisante les déplacements de pendules provenant de deux barrages différents. Une étude plus approfondie des modèles ST permettra peut-être de construire une gamme restreinte de modèles généraux susceptibles de modéliser les déplacements de n'importe quel barrage.



# Chapitre 11

## Conclusion

Modèle classique ou modèle ST? Ce projet a permis d'ébaucher une réponse à cette question. En effet, toutes les analyses effectuées semblent indiquer que le modèle ST est plus adapté pour modéliser les déplacements des murs des barrages en béton. La structure des résidus et la qualité de prédiction de chacun des modèles permettent sans équivoque de dégager une préférence pour le modèle ST. Toutefois, au moment de tirer les conclusions d'une telle étude, il convient de faire preuve d'une certaine prudence. En effet, toutes les comparaisons entre les deux modèles, aussi explicites soient-elles, reposent sur les jeux de données provenant de deux barrages uniquement. Or, la Suisse compte plus de 200 barrages et le monde plusieurs milliers. On ignore de plus si ces deux modèles sont capables de détecter une anomalie sur un barrage. Il faudrait pour cela les calibrer grâce à un jeu de données provenant d'un barrage sur lequel est survenu par le passé un problème.

Intéressons-nous tout d'abord au modèle classique. Il s'agit sans aucun doute d'un bon modèle, mais il semble montrer certaines limites, notamment au niveau de la qualité de prédiction. De plus, les variables explicatives qui interviennent dans les différentes calibrations ne sont pas toutes les mêmes, ce qui nous conduit à penser que l'existence d'un modèle unique, applicable à presque tous les barrages, relève de l'utopie dans le cadre du modèle classique.

Regardons maintenant de plus près le modèle ST. Ce modèle possède d'indéniables qualités. Dans le cas des deux barrages que nous avons analysés, ce modèle a permis de réaliser d'excellentes prédictions. De plus, le nombre plus restreint de variables explicatives significatives pourraient permettre de proposer un modèle plus général ou en tout cas de classifier les barrages en deux ou trois groupes seulement. Pour vérifier cette hypothèse, il faudrait

bien évidemment calibrer ce modèle sur d'autres barrages.

De la sécurité des barrages dépend la vie de milliers de personnes en Suisse et de ce fait il faut se garder d'agir avec précipitation dans ce domaine. Dès lors, il nous paraît préférable de modéliser individuellement chaque barrage et de trouver dans chaque cas le modèle adéquat. Pour certains ouvrages, le modèle classique sera certainement plus performant, tandis que pour d'autres, le modèle ST permettra d'aboutir à de meilleurs résultats. L'introduction du modèle ST permet donc d'élargir l'éventail des modèles disponibles pour effectuer le travail de contrôle et de prédiction nécessaire sur les barrages. Ce modèle offre aussi une alternative au modèle classique lorsque ce dernier révèle quelques faiblesses et permet d'utiliser deux modèles différents de façon complémentaire.

La recherche de modèles statistiques qui permettent d'estimer les déplacements des barrages est certainement loin d'être terminée. Elle n'en est même qu'à ses balbutiements. Nous avons simplement ouvert une voie quelque peu novatrice qui ne demande qu'à être explorée plus en profondeur. Pour notre sécurité et celle des générations futures.

# Annexe A

## Déviations radiales des pendules A et H

Les figures A.1 et A.2 décrivent les valeurs de la déviation radiale des pendules A et H au cours des années d'observation. En comparant les deux figures, on comprend mieux pourquoi les modèles calibrés à l'aide du second jeu de données étaient meilleurs. En effet, bien que l'on devine une certaine périodicité, la déviation radiale du pendule A est très irrégulière ce qui rend son estimation par un modèle linéaire beaucoup plus délicate. La déviation radiale du pendule H est, à quelques exceptions près, très régulière et l'on repère au premier coup d'oeil le caractère périodique de cette déviation. Au début de cette étude, nous avons fait une ANOVA (analysis of variance) à partir de la déviation radiale du pendule A. Nous avons notamment tenté de déceler une tendance dans cette déviation et malgré le caractère périodique du phénomène, cette analyse s'est avérée totalement improductive, ce qui montre encore une fois le caractère très irrégulier de ce jeu de données.

Signalons encore qu'il existe un autre facteur susceptible d'expliquer la meilleure qualité de l'estimation par les modèles calibrés à partir du second jeu de données. Il s'agit du nombre d'années d'observation. En effet, plus nous possédons d'informations sur une variable, moins les valeurs "aberrantes" de cette valeur influencent l'estimation des paramètres. Dans notre cas, nous possédons pour le premier jeu de données 11 années d'observations et pour le second jeu de données 23 années d'observations. Cette différence, ajoutée à la meilleure régularité de la déviation du pendule H peut en partie expliquer les différences dans la qualité des estimations et surtout des prédictions.

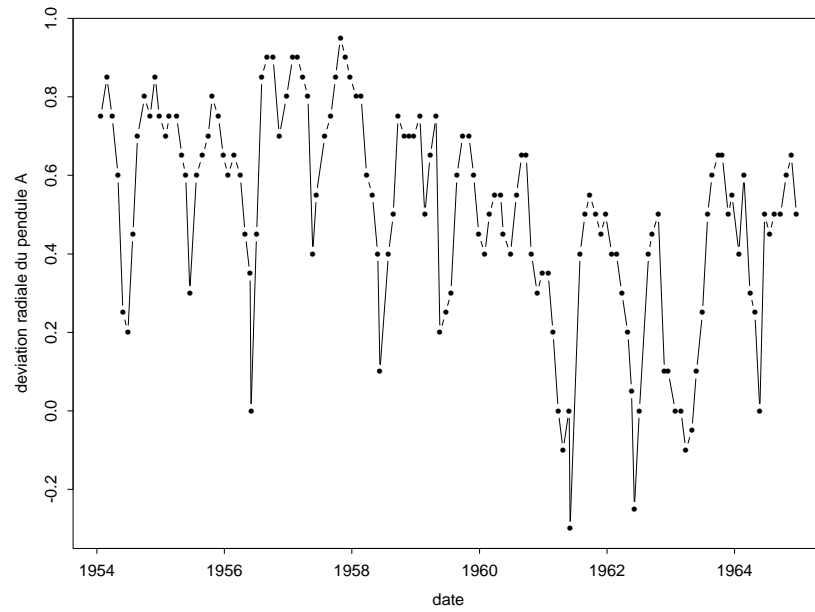


FIG. A.1 – *Déviatiion radiale du pendule A*

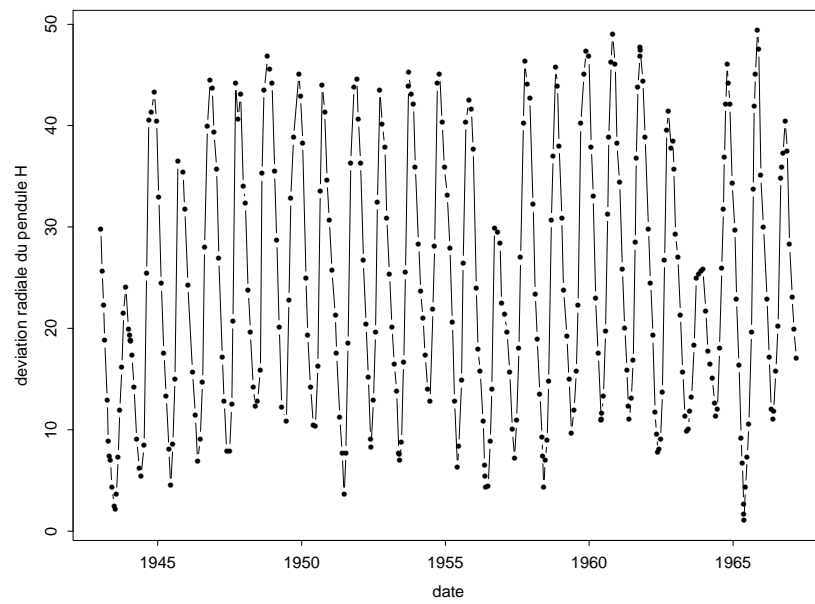


FIG. A.2 – *Déviatiion radiale du pendule H*

# Annexe B

## Quelques codes S-Plus

Voici la marche à suivre pour calibrer un modèle linéaire avec le logiciel S-Plus et les commandes à utiliser pour obtenir les différents graphiques rencontrés dans les pages précédentes. Nous partons du principe que nous possédons un jeu de données (*jeudedonnees.data*) de la forme adéquate. La variable-réponse est  $y$  et les variables explicatives sont  $X_1, \dots, X_p$ .

```
barrage <- read.table('jeudedonnees.data', header=T)
attach(barrage)
```

```
trellis.device()
plot(y)
```

```
modele1 <- lm(y ~ X1 + ... + Xp)
summary(modele1)
```

```
plot(modele1$resid)
qqnorm(modele1$resid)
qqline(modele1$resid)
acf(modele1$resid)
```

Voici maintenant la fonction qui permet de représenter sur le même graphique les vraies valeurs de  $y$  et les valeurs prédites par le modèle. On suppose que le jeu de données se compose de  $n$  observations et que pour effectuer la prédiction on se base sur les  $m$  premières observations. Il est évident que pour pouvoir effectuer cette prédiction il faut au préalable avoir calibré le

modèle 1 comme expliqué précédemment, mais en utilisant uniquement les  $m$  premières observations. Nous appellerons ce modèle *modele2*.

```
prediction <- fonction(X1, X2, ... , Xp)
{
trellis.device()
date <- (annee + (mois - 1)/12 + jour/365)[m+1,n]
Z <- summary.lm(modele2)$cov
X <- cbind(X1[1:m], X2[1:m], ... , Xp[1:m])
T <- cbind(X1[m+1:n], X2[m+1:n], ... , Xp[m+1,n])
yhat <- T %*% Z %*% t(X) %*% y[1:m]
plot(date, y[m+1:n], pch = 5)
points(date, yhat)
}
```

Voici finalement le code pour obtenir une prédiction avec intervalles de confiance ponctuels. Il ressemble fortement au précédent.

```
prediction <- fonction(X1, X2, ... , Xp)
{
trellis.device()
date <- (annee + (mois - 1)/12 + jour/365)[m+1,n]
Z <- summary.lm(modele2)$cov
X <- cbind(X1[1:m], X2[1:m], ... , Xp[1:m])
T <- cbind(X1[m+1:n], X2[m+1:n], ... , Xp[m+1,n])
yhat <- T %*% Z %*% t(X) %*% y[1:m]
sig <- summary.lm(modele2)$sigma
v <- sig^2 * T %*% Z %*% t(T)
yinf <- yhat - 1.96 * diag(v)^(0.5)
ysup <- yhat + 1.96 * diag(v)^(0.5)
plot(date, y[m+1:n], pch = 5)
points(date, yhat, pch = 3)
points(date, yinf, col = 7)
points(date, ysup, col = 7)
}
```



# Bibliographie

- Scheffé, H. (1959). *Analysis of Variance*. Wiley and Sons, New-York.
- Searle, S. R. (1971). *Linear Models*. Wiley and Sons, New-York.
- Morgenthaler, S. (1997). *Introduction à la Statistique*. Presses Polytechniques et Universitaires Romandes.
- Kendall, M. G. and Stuart, A. (1979). *The Advanced Theory of Statistics*, Vol. 2: Inference and Relationship,(4th edition). New York: Oxford University Press.
- Hogg, R. V. and Craig, A. T. (1970). *Introduction to Mathematical Statistics*, 3rd edition Toronto, Canada: Macmillan.
- Comité Suisse des Barrages (2000). *Comparaison Mesuré et Calculé*.
- Sun, J. , Loader, C. and McCormick, W. P. (2000). *Confidence Bands in Generalized Linear Models*. The Annals of Statistic, Vol. 28, No. 2, 429-460.