

**MODÉLISATION DU SURENFONCEMENT
DES NAVIRES MARCHANDS QUI TRANSITENT
PAR LE FLEUVE SAINT-LAURENT**

Rapport de recherche No R-974

Mars 2008

**Modélisation du surenforcement des navires marchands qui transitent
par le fleuve Saint-Laurent**

Par

Claudie Beaulieu

Taha B.M.J. Ouarda

Ousmane Seidou

Chaire en hydrologie statistique Hydro-Québec/CRSNG
Chaire du Canada en estimation des variables hydrologiques
INRS-ETE

490, rue de la Couronne
Québec (Qc), Canada, G1K-9A9

À l'intention de

M. Samir Gharbi, Ph.D.

Génie hydraulique

Services techniques, GCC, Pêches et Océans Canada
101, boul. Champlain
Québec (QC), Canada, G1K 7Y7

Rapport de recherche N° R-974

Mars 2008

ISBN : 978-2-89146-565-6

Table des matières

Table des matières	iii
Liste des figures	v
Liste des tableaux	vii
1. Introduction	1
2. Revue de littérature	3
2.1. Modèle Barrass 1	4
2.2. Modèle Barrass 2	5
2.3. Modèle Eryuzlu 1	5
2.4. Modèle Eryuzlu 2	6
2.5. Modèle Eryuzlu 3	6
2.6. Modèle Eryuzlu 4	6
2.7. Modèle Hooft	7
2.8. Modèle Huuska/Guliev	7
2.9. Modèle Icorels	8
2.10. Modèle Japanese	8
2.11. Modèle Millward 1	8
2.12. Modèle Millward 2	8
2.13. Modèle Norrbin	9
2.14. Modèle Römisch	9
2.15. Modèle Simard	9
2.16. SLS Trial Formula	10
2.17. Modèle Tuck	10
2.18. Modèle VLCC	11
2.19. Modèle VLCC (non restreint)	11
2.20. Comparaison des modèles	11
3. Données	13
4. Analyses préliminaires	17
4.1. Corrélations	17
4.2. Calcul du nombre de Froude	18
4.3. Régression pas-à-pas	18
4.4. Régression ridge	25

5. Nouveaux modèles	31
5.1. Arbres de régression	31
5.2. Modèle Eryuzlu 1 optimisé	38
6. Discussion et conclusion	47
Références	51
Annexe A : Statistiques descriptives des variables susceptibles d'influencer le surenfoncement et du surenfoncement.	57
Annexe B : Étude des corrélations entre les variables susceptibles d'influencer le surenfoncement et du surenfoncement.	59
Annexe C : Arbres de classification et de régression	63

Liste des figures

Figure 1 : Carte des villes de départ et d'arrivée des douze voyages 1) Montréal-Portneuf, 2) Québec-Trois-Rivières, 3) Montréal-Neuveville, 4) Portneuf-Montréal, 5) Neuville-Montréal, 6) Montréal-Neuveville, 7) Sault-au-cochon-Québec, 8) Québec-Contrecoeur, 9) Québec-Montréal, 10) Sault-au-cochon-Québec, 11) Québec-Sault-au-cochon et 12) Sorel-Montréal.....	15
Figure 2 : Résidus en fonction des observations pour le modèle de régression pas-à-pas ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.....	22
Figure 3 : Résidus en fonction des valeurs prédites pour le modèle de régression pas-à-pas ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.....	23
Figure 4 : Droite de Henry des résidus pour le modèle de régression pas-à-pas ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.	24
Figure 5 : Résidus en fonction des observations pour le modèle de régression ridge ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.....	28
Figure 6 : Résidus en fonction des valeurs prédites pour le modèle de régression ridge ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.....	29
Figure 7 : Droite de Henry des résidus pour le modèle de régression ridge ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.	30
Figure 8 : Schéma simplifiée d'un arbre de régression (Breiman et al., 1984).....	31
Figure 9 : Valeurs observées et valeurs prédites obtenues par arbre de régression ajusté sur a) l'ensemble des données b) porte-conteneurs, c) citerne et d) vraquier.....	36
Figure 10 : Valeurs observées et valeurs prédites obtenues par le modèle <i>Eryuzlu 1</i> ajusté sur a) l'ensemble des données b) porte-conteneurs, c) citerne et d) vraquier.....	39
Figure 11 : Résidus en fonction des observations pour le modèle de régression ajusté sur les erreurs du modèle <i>Eryuzlu 1</i> pour a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.....	42
Figure 12 : Résidus en fonction des valeurs prédites pour le modèle de régression ajusté sur les erreurs du modèle <i>Eryuzlu 1</i> pour a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.....	43
Figure 13 : Droite de Henry des résidus pour le modèle de régression ajusté sur les erreurs du modèle <i>Eryuzlu 1</i> pour a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.	44

Figure 14 : Valeurs observées et valeurs prédites obtenues par le modèle *Eryuzlu OPT* ajusté sur a) l'ensemble des données b) porte-conteneurs, c) citerne et d) vraquier.....46

Liste des tableaux

Tableau 1 : Paramètres utilisés pour l'estimation du surenforcement selon les différents modèles.....	4
Tableau 2 : Résumé des douze voyages et caractéristiques des navires utilisés pour la campagne de terrain de 2005 (Gharbi et al., 2007).	14
Tableau 3 : Minimum et maximum des principaux paramètres mesurés durant les douze voyages de la campagne de terrain de 2005 (Gharbi et al., 2007).	14
Tableau 4 : Coefficients des différents modèles de régression pas-à-pas.	20
Tableau 5 : Adéquation des différents modèles de régression pas-à-pas.	20
Tableau 6 : Coefficients des différents modèles de régression ridge.	26
Tableau 7 : Adéquation des différents modèles de régression ridge.....	27
Tableau 8 : Performance des différents modèles ajustés.....	35
Tableau 9 : Performance des modèles empiriques d'estimation du surenforcement et du modèle retenu qui a été développé par arbre de régression.....	37
Tableau 10 : Critères de performance des modèles $RegT$ calculés avec des erreurs de prédiction obtenues par validation croisée.....	37
Tableau 11 : Coefficients des différents modèles de régression pas-à-pas ajustés sur les erreurs du modèle $Eryuzlu 1$	41
Tableau 12 : Performance du modèle $Eryuzlu OPT$, du modèle $RegT$ et des modèles empiriques d'estimation du surenforcement.....	45

1. Introduction

Ce travail s'inscrit dans le cadre du projet d'étude du dégagement sous quille des navires marchands transitant par le fleuve Saint-Laurent, mené par la Garde Côtière Canadienne (GCC) en vue de réévaluer les normes de dégagement sous quille. Le dégagement sous quille représente la distance entre la profondeur nominale du chenal et la partie la plus basse de la quille du navire lorsqu'il est à l'arrêt en eaux calmes (Morse et Simard, 1991). Le surenfoncement, est la réduction du dégagement sous quille résultant du mouvement du navire et de son interaction avec le milieu. Ce phénomène est accentué dans les voies navigables ayant une profondeur d'eau restreinte (Stocks et al., 2002). Le surenfoncement peut faire en sorte qu'un navire s'échoue, même si ce dernier dispose d'un dégagement sous quille suffisant lorsqu'il est immobile. Dans Barrass (2004), des exemples de navires qui ont récemment échoués du à un surenfoncement excessif sont présentés. Par conséquent, une bonne estimation du surenfoncement des navires marchands est nécessaire afin de s'assurer d'un dégagement sous quille sécuritaire. Dans les voies navigables commerciales comme celle du Saint-Laurent, une bonne prédiction du surenfoncement est également importante pour assurer l'efficacité du trafic, permettant ainsi aux navires de transporter des charges maximales en toute sécurité.

Le principal facteur qui influence le surenfoncement est la vitesse du navire. Or, ce phénomène est complexe et plein d'autres facteurs doivent être considérés pour fournir une bonne estimation du surenfoncement. Des expressions analytiques pour le surenfoncement ont été développées et sont présentées dans Gourlay (2000), Gourlay (2006) et Varyani (2006). Ces articles font également référence aux formules théoriques plus anciennes de surenfoncement. En pratique, l'usage de ces modèles n'est pas très commode pour les navigateurs. De ce fait, plusieurs chercheurs ont développés des modèles empiriques pour prédire le surenfoncement des navires. Ces modèles ont été construits par essais et erreurs en se basant sur des résultats de campagne de terrain ou sur des tests en laboratoire. Leurs conditions d'application sont très restrictives et très différentes. Dans des conditions données, les prédictions de surenfoncement obtenues avec les différents modèles empiriques

varient énormément d'un modèle à un autre. Le choix du modèle le plus approprié peut s'avérer ardu (Macpherson, 2002).

Plusieurs modèles de surenfacement ont été développés spécifiquement pour le fleuve Saint-Laurent. Le modèle de Simard (1982) est basé sur des données de terrain. Eryuzlu et al. (1994) propose plutôt quatre équations basées sur des études en laboratoire. Ces deux approches sont considérées les plus appropriées pour le fleuve Saint-Laurent (Morse et Simard, 1991; Morse et al., 2004). La GCC a adopté un modèle proposé dans Eryuzlu et al. (1994), pour prédire le surenfacement des navires transitant par le fleuve Saint-Laurent.

Dans une étude réalisée par Gharbi et Le Page (2007), vingt modèles de surenfacement utilisés un peu partout à travers le monde sont présentés, leur précision est comparée et leur adéquation au fleuve Saint-Laurent est vérifiée. Le modèle présentement utilisé par la GCC pour prédire le surenfacement était parmi les plus performants. Néanmoins, les conclusions de cette étude ont montré que même les meilleurs modèles ne donnent pas une performance satisfaisante. De ce fait, l'optimisation des modèles d'estimation du surenfacement des navires et/ou le développement de nouveaux modèles ont été recommandés. Les objectifs de ce travail sont de :

- 1) Utiliser une nouvelle approche pour développer un modèle de surenfacement.
- 2) Optimiser un modèle de surenfacement prometteur pour la voie navigable du Saint-Laurent.

Dans un premier temps, une brève revue de littérature des modèles empiriques est effectuée (section 2). Dans un deuxième temps, les données utilisées sont présentées (section 3). Dans un troisième temps, des analyses préliminaires des données ont été réalisées afin d'analyser le comportement statistique des différentes variables ainsi que leurs interactions (section 4). Dans un quatrième temps, différentes méthodes de modélisation ont été utilisées pour déterminer l'approche la plus prometteuse. Un nouveau modèle a été développé en utilisant une approche d'arbre de régression et de classification (section 5). Par la suite, le modèle utilisé par la GCC a été optimisé en modélisant les erreurs. Enfin, une discussion et une conclusion sont présentées dans la section 6.

2. Revue de littérature

Sur des voies navigables ayant une profondeur d'eau restreinte, il est nécessaire d'établir un équilibre entre l'efficacité (le besoin des usagers de transporter des charges maximales) et la sécurité du trafic, compromise par le surenfoncement du navire et le risque de toucher le fond. La GCC a fixé des normes afin d'assurer un dégagement sous quille sécuritaire des navires qui empruntent la voie navigable du fleuve Saint-Laurent. Le surenfoncement est la principale composante du dégagement sous quille et une bonne estimation est importante pour s'assurer d'une marge de manœuvre suffisante.

Dans le travail de Gharbi et Le Page (2007), la performance de plusieurs modèles empiriques pour prédire le surenfoncement de trois types de navires (porte-conteneurs, citerne et vraquier) est évaluée. Malgré le fait que les conditions d'application des différents modèles n'étaient pas toujours respectées, tous les modèles ont été utilisés pour prédire le surenfoncement. Les recommandations de cette étude étaient d'améliorer les modèles afin de réduire le biais et les erreurs quadratiques ou encore, de développer de nouvelles équations de surenfoncement. Les modèles de surenfoncement qui ont été comparés dans cette étude sont brièvement présentés dans cette section. Pour une description plus complète, le lecteur peut également se référer aux travaux de Dick et al. (1991), Herreros Sierra et al. (2000), Briggs et al. (2004) et Briggs (2006), qui présentent plusieurs modèles ou encore aux articles souches de chaque modèle tels que présentés ci-bas. Le tableau 1 présente les paramètres utilisés dans les différents modèles d'estimation du surenfoncement.

Tableau 1 : Paramètres utilisés pour l'estimation du surenfoncement selon les différents modèles.

Symbole	Description
∇	Déplacement (m^3)
A_c	Aire de section transversale mouillée du navire (m^2)
A_s	Aire de section transversale du navire (m^2)
A_w	Aire de section transversale du chenal (m^2)
b	Largeur du navire (m)
B	Largeur restreinte du chenal (m)
C_b	Coefficient de blocage
F_{nh}	Nombre de Froude
g	Constante d'accélération gravitationnelle valant 9.80665 m/s^2
h	Hauteur totale d'eau (m)
h_z	Profondeur bathymétrique (m)
K_1	Facteur de correction sur le facteur de blocage du chenal
K_b	Facteur de correction pour la largeur du chenal (fonction de W et b)
K_s	Facteur de correction pour la largeur du chenal (fonction de s_1)
L	Longueur du navire (m)
L_{pp}	Longueur du navire entre les parallèles (m)
PAR_h/T	Réduction de l'assiette due à la propulsion en eaux peu profondes
R_{hTh}	Rapport de la profondeur bathymétrique et de la profondeur d'eau
S	Surenfoncement du navire (m)
S_{obs}	Surenfoncement observé du navire (m)
s	Facteur de blocage du chenal
s_1	Facteur corrigé de blocage du chenal
S_2	Facteur de vitesse de retour
T	Tirant d'eau du navire (m)
V	Vitesse du navire (SOW, <i>speed over water</i>) (m/s)
V_{cr}	Vitesse critique pour un chenal non restreint (m/s)
V_k	Vitesse du navire (noeuds)
W	Largeur du chenal

2.1. Modèle Barrass 1

Ce modèle est présenté dans Barrass (1979) et il est applicable pour des chenaux restreints et non restreints. De plus, l'estimation est satisfaisante lorsque le rapport de la profondeur

d'eau sur le tirant d'eau (h/T) varie entre 1.1 et 1.4. L'équation suivante permet d'estimer le surenfoncement maximum :

$$S = \frac{C_b \cdot S_2^{2/3} \cdot V_k^{2.08}}{30} \quad (1)$$

$$S_2 = \frac{A_s}{A_w} \quad (2)$$

$$A_s = T \cdot b \quad (3)$$

$$A_w = A_c - A_s \quad (4)$$

$$A_c = h \cdot \min\{B, W\} \quad (5)$$

$$B = 10 \cdot b \quad (6)$$

2.2. Modèle Barrass 2

Un autre modèle a été suggéré par Barrass (1981) pour faire la distinction entre les eaux confinées et les eaux libres. Il permet également de prédire le surenfoncement maximum :

$$S = \begin{cases} C_b \cdot V_k^2 / 50, & \text{eaux confinées} \\ C_b \cdot V_k^2 / 100, & \text{eaux libres} \end{cases} \quad (7)$$

2.3. Modèle Eryuzlu 1

Ce modèle est présenté dans Eryuzlu et D'Agnolo (1991). Il a été développé pour les vraquiers ayant un coefficient de blocage de 0.8, un poids qui varie entre 19 000 et 227 000 tonnes et un tirant d'eau statique de 8.13 à 20.32 mètres. Il est applicable seulement aux chenaux non restreints, des vitesses qui varient entre 2 et 9 mètres/seconde, un chenal dont la profondeur ne dépasse pas 34.35 mètres et enfin lorsque $1.1 < h/T < 3.0$. Le modèle s'exprime ainsi :

$$S = 0.181 \cdot \sqrt{T \cdot b} \cdot \left(\frac{V}{\sqrt{g \cdot T}} \right)^{2.269} \cdot \left(\frac{T}{h} \right)^{0.994} \quad (8)$$

Ce modèle est utilisé par la GCC pour prédire le surenfoncement sur le fleuve Saint-Laurent.

2.4. Modèle Eryuzlu 2

Ce modèle a les mêmes conditions d'application que le précédent sauf qu'il ne prend pas en compte la largeur du navire (Eryuzlu et al., 1994). Le modèle est tel que :

$$S = 0.298 \cdot T \cdot \left(\frac{V}{\sqrt{g \cdot T}} \right)^{2.289} \cdot \left(\frac{T}{h} \right)^{0.972} \quad (9)$$

2.5. Modèle Eryuzlu 3

Ce modèle est présenté dans Eryuzlu et al. (1994). Il est basé sur le modèle **Eryuzlu 2** avec en plus, un facteur de pondération pour la largeur du chenal. L'étude a été réalisée sur des navires de charge et des vraquiers ayant un coefficient de blocage de 0.8 et plus. Le modèle peut être utilisé à la fois dans les eaux confines et libres. Il est applicable lorsque $1.1 < h/T < 2.5$. L'équation permet de prédire le surenfoncement à la proue.

$$S = 0.298 \cdot T \cdot \left(\frac{V}{\sqrt{g \cdot T}} \right)^{2.289} \cdot \left(\frac{T}{h} \right)^{0.972} \cdot K_b \quad (10)$$

$$K_b = \begin{cases} \frac{3.1}{\sqrt{W/b}}, & \frac{W}{b} < 9.61 \\ 1, & \frac{W}{b} \geq 9.61 \end{cases} \quad (11)$$

2.6. Modèle Eryuzlu 4

Ce modèle est présenté dans Eryuzlu et Hausser (1978). Il permet d'estimer le surenfoncement des citernes avec un coefficient de blocage d'au moins 0.8, dans des eaux non restreintes, mais peu profondes. Il est applicable lorsque $1.08 < h/T < 2.75$. L'équation suivante peut être utilisée pour calculer le surenfoncement à la proue :

$$S = 0.113 \cdot \left(\frac{T}{h} \right)^{0.27} \cdot b \cdot F_{nh}^{1.8} \quad (12)$$

$$F_{nh} = \frac{V}{\sqrt{g \cdot h}} \quad (13)$$

2.7. Modèle Hooft

Ce modèle est présenté en détails dans Hooft (1974) et a été développé à partir du modèle de **Tuck** présenté un peu plus loin dans cette section. Le facteur de 1.96 peut varier de 1.9 à 2.03. La formule permettant de calculer le surenfoncement à la proue est donnée par :

$$S = 1.96 \cdot \frac{\nabla}{L_{pp}^2} \cdot \frac{F_{nh}}{\sqrt{1 - F_{nh}^2}} \quad (14)$$

$$\nabla = C_b \cdot L \cdot T \cdot B \quad (15)$$

2.8. Modèle Huuska/Guliev

Ce modèle est présenté en détails dans Huuska (1976). Il s'agit du modèle de **Hooft** modifié pour permettre son application en eaux restreintes en introduisant un facteur de correction pour la largeur du chenal. L'équation est valide lorsque $F_{nh} < 0.7$ et que $1.1 < h/T < 2$. Le surenfoncement à la proue s'obtient ainsi :

$$S = 2.4 \cdot \frac{\nabla}{L^2} \cdot \frac{F_{nh}}{\sqrt{1 - F_{nh}^2}} \cdot K_s \quad (16)$$

$$K_s = \begin{cases} 7.45 \cdot s_1 + 0.76, & s_1 > 0.03 \\ 1, & s_1 \leq 0.03 \end{cases} \quad (17)$$

$$s_1 = \frac{s}{K_1} \quad (18)$$

$$s = \frac{A_s}{A_c} \quad (19)$$

$$K_1 = 5.2855 \cdot s^{0.7098} + 3.4559 \cdot \frac{s}{R_{hTh}} + \frac{0.2856}{R_{hTh}^{0.2801}} \quad (20)$$

$$R_{hTh} = h_z / h \quad (21)$$

2.9. Modèle Icorels

Ce modèle est présenté dans ICORELS (1980). Il est semblable aux deux précédents (**Hooft** et **Huuska/Guliev**). Il permet de prédire le surenfoncement à la proue des navires :

$$S = 2.4 \cdot \frac{\nabla}{L_{pp}^2} \cdot \frac{F_{nh}^2}{\sqrt{1 - F_{nh}^2}} \quad (22)$$

2.10. Modèle Japanese

Ce modèle a été proposé par OCDI (travail non publié, 2002) pour répondre aux nouveaux standards pour les voies maritimes du Japon. C'est le modèle le plus récent. Il s'exprime ainsi :

$$S = \left[\left(0.7 + 1.5 \cdot \frac{1}{h/T} \right) \left(\frac{C_b}{L_{pp}/B} \right) + 15 \cdot \frac{1}{h/T} \cdot \left(\frac{C_b}{L_{pp}/B} \right)^3 \right] \cdot \frac{V^2}{g} \quad (23)$$

2.11. Modèle Millward 1

Ce modèle, présenté par Millward (1990), a été développé pour différents types de navires qui transitent dans des chenaux non restreints. Il est restreint aux navires ayant une largeur approximative de deux fois L_{pp} et un coefficient de blocage qui varie entre 0.44 et 0.83. Le modèle est donné par :

$$S = 0.01 \cdot L_{pp} \cdot \left(15 \cdot C_b \cdot \frac{1}{L_{pp}/B} - 0.55 \right) \cdot \frac{F_{nh}^2}{1 - 0.9 F_{nh}^2} \quad (24)$$

2.12. Modèle Millward 2

Millward (1992) a apporté quelques modifications au modèle **Millward 1**. Le nouveau modèle s'exprime ainsi :

$$S = 0.01 \cdot L_{pp} \cdot \left(61 \cdot 7 C_b \cdot \frac{1}{L_{pp}/T} - 0.6 \right) \cdot \frac{F_{nh}^2}{\sqrt{1 - F_{nh}^2}} \quad (25)$$

2.13. Modèle Norrbin

Ce modèle est présenté dans Norrbin (1986) et est basé sur le modèle de **Tuck**. Il permet d'estimer le surenfoncement à la proue d'un navire circulant dans des chenaux non restreints. Il s'applique lorsque le nombre de Froude est inférieur à 0.4. Le modèle s'exprime ainsi :

$$S = \frac{C_b}{15} \cdot \left(\frac{1}{L_{pp}/B} \right) \cdot \left(\frac{1}{h/T} \right) \cdot V^2 \quad (26)$$

2.14. Modèle Römisch

Ce modèle a été développé pour estimer le surenfoncement à la proue et à la poupe des navires (Römisch avant et Römisch arrière) pour des chenaux restreints ou non restreints ayant une forme rectangulaire ou trapézoïdale (Römisch, 1989). Le modèle est applicable lorsque $1.19 < h/T < 2.25$. Il est donné par :

$$S = \begin{cases} C_V \cdot C_F \cdot K_{\Delta T} \cdot T, & \text{proue} \\ C_V \cdot K_{\Delta T} \cdot T, & \text{poupe} \end{cases} \quad (27)$$

$$C_V = 8 \cdot \left(\frac{V}{V_{cr}} \right)^2 \cdot \left[\left(\frac{V}{V_{cr}} - 0.5 \right)^4 + 0.0625 \right] \quad (28)$$

$$C_F = \left(\frac{10 \cdot C_b \cdot B}{L_{pp}} \right)^2 \quad (29)$$

$$K_{\Delta T} = 0.155 \sqrt{\frac{h}{T}} \quad (30)$$

$$V_{cr} = K_c \sqrt{gh} \quad (31)$$

$$K_c = 0.2472 \cdot \ln \left(\frac{A_c}{A_s} \right) + 0.0241 \quad (32)$$

2.15. Modèle Simard

Ce modèle a été développé pour estimer le surenfoncement de tous types de navires qui circulent sur la voie maritime du Saint-Laurent (Simard, 1969; Simard, 1982; Morse et

Simard, 1991). Le modèle est applicable lorsque la vitesse par rapport à l'eau varie entre 8 et 15 nœuds, le tirant d'eau du navire est de l'ordre de 10 mètres, la largeur du navire est de l'ordre de 25 mètres et l'aire de section transversale immergée ne dépasse pas 250 m². Il est recommandé d'utiliser le modèle pour les chenaux restreints. L'équation est donnée par :

$$S = \frac{V^2}{2g} \cdot \left[\left(\frac{1.01}{1 - A_s/A_c} \right)^2 - 0.80 \right] \quad (33)$$

2.16. SLS Trial Formula

Ce modèle est présenté dans Stocks et al. (2002). L'information concernant les conditions d'applications du modèle ne sont pas présentées. Le modèle permettant de calculer le surenfacement à la proue s'exprime ainsi :

$$S = L_{pp} \cdot \left[SHF \cdot \left(1 + 0.35 \left(\frac{T}{h} \right)^2 \right) \cdot CH_s + 0.5 \cdot 2.5 \cdot SHF (C_b^2 - 0.3 \cdot CH_s) \cdot PAR_{-h/T} \right] \quad (34)$$

$$SHF = \left[1.5 \frac{\nabla}{L_{pp}^3} + 0.044 \cdot C_b^2 \right] \cdot par_{-F} \quad (35)$$

$$par_{-F} = F_{nh}^{(1.8+0.4F_{nh})} \quad (36)$$

$$CH_s = 1 + 10 \cdot \left(\frac{A_s T}{A_c h} \right) - 1.5 \cdot \left(1 + \frac{A_s T}{A_c h} \right) \cdot \sqrt{\frac{A_s T}{A_c h}} \quad (37)$$

$$PAR_{-h/T} = 1 - e^{\left[\frac{2.5(1-h/T)}{F_{nh}} \right]} \quad (38)$$

2.17. Modèle Tuck

Ce modèle est proposé par Tuck (1966). Il est également présenté dans Stocks et al. (2002), mais le domaine de validité n'est pas spécifié. L'équation est donné par :

$$S = 1.46 \cdot \frac{\nabla}{L^2} \cdot \frac{F_{nh}^2}{\sqrt{1 - F_{nh}^2}} \cdot K_s + 0.5L \cdot \sin \left(\frac{\nabla}{L^3} \cdot \frac{F_{nh}^2}{\sqrt{1 - F_{nh}^2}} \cdot K_s \right) \quad (39)$$

2.18. Modèle VLCC

Ce modèle est présenté dans Stocks et al. (2002). Il permet de prédire le surenfoncement à la proue :

$$S = 1.06 \cdot h \cdot F_{nh}^2 \cdot \frac{\left(\frac{A_c}{A_s}\right) \left(1 - \frac{A_c}{A_s}\right) \left(1 - 0.5 \cdot \frac{A_c}{A_s}\right)}{\left(1 - \frac{A_c}{A_s}\right)^3 - 1.16 \cdot F_{nh}^2} \quad (40)$$

2.19. Modèle VLCC (non restreint)

Ce modèle est présenté dans Stocks et al. (2002). Il permet de prédire le surenfoncement à la proue dans des chenaux non restreints:

$$S = 0.28 \cdot T \cdot \frac{F_{nh}^2}{\sqrt{1 - F_{nh}^2}} \quad (41)$$

2.20. Comparaison des modèles

Dans le travail de Gharbi et Le Page (2007), la performance de ces modèles a été comparée sur des données recueillies lors d'une campagne de terrain en 2005. Ces données sont présentées à la section suivante. Des critères ont été utilisés pour évaluer et comparer la performance des différentes techniques (racine de l'erreur quadratique moyenne, coefficient de détermination et biais moyen relatif). Puisque ces mêmes critères seront utilisés plus loin dans le texte pour comparer les modèles jugés les plus performants avec les modèles développés et optimisés, ils sont présentés ici. L'erreur quadratique moyenne (EQM) permet de mesurer l'écart entre les valeurs prédites et les valeurs observées. Lorsque cette valeur est faible, cela signifie qu'il y a peu d'écart entre les valeurs prédites (\hat{y}_i) et observées (y_i).

$$EQM = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (42)$$

où n représente la taille de l'échantillon. Le coefficient de détermination permet de vérifier l'ajustement du modèle. Lorsqu'il est près de 1, le modèle est vraisemblablement bon tandis que lorsqu'il se rapproche de 0, alors le modèle est probablement mauvais.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (43)$$

Enfin, le dernier critère de performance utilisé est le biais moyen relatif. Ce critère est intéressant puisqu'il permet de vérifier si les estimations tendent à être supérieures ou inférieures aux observations. On calcule le biais moyen relatif pour avoir l'importance du biais (en pourcentage) par rapport aux observations.

$$BMR = \frac{1}{n} \cdot 100 \cdot \sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right) \quad (44)$$

Les dix modèles jugés les plus performants selon ces critères ont été identifiés. Cependant, même avec les meilleurs modèles, la performance peut laisser à désirer et pourrait être améliorée. Les recommandations de cette étude étaient d'optimiser les modèles afin de réduire les erreurs quadratiques et le biais ou encore de développer de nouveaux modèles plus performants.

3. Données

La base de données utilisée a été construite suite à la campagne de terrain menée par la GCC en 2005. Douze voyages ont été effectués sur la voie navigable du fleuve Saint-Laurent avec différents types de navires (porte-conteneurs, citerne et vraquier). Le tableau 2 présente les itinéraires des douze voyages ainsi que les caractéristiques des navires utilisés. La figure 1 présente une carte des villes de départ et d'arrivée des douze voyages. Pendant ces voyages, des mesures du surenfoncement des navires ainsi que les paramètres permettant de prédire le surenfoncement comme la vitesse, la position, les caractéristiques du navire et les caractéristiques du chenal ont été prises. Les valeurs du surenfoncement ont été calculées à partir des relevés GPS-OTF pris à toutes les deux secondes. Cette technologie est décrite dans Pagé et Stocks (2002). Les observations prises par GPS-OTF sont précises et peuvent être utilisées pour valider un modèle de prédiction du squat (Parker et Huff, 1998). Les résultats obtenus ont été analysés et filtrés, pour ensuite avoir une valeur de surenfoncement maximum sur des intervalles de 30 secondes. Ensuite, les données ont été validées en analysant les variations du surenfoncement en fonction des principaux paramètres physiques (nombre de froude, vitesse, etc.). Cette base de données comporte un ensemble de 5785 observations et est présentée en détails dans Gharbi et al. (2007). Le tableau 3 présente un résumé des principaux paramètres physiques. Le nombre de Froude et le rapport de la hauteur d'eau sur le tirant d'eau sont également incorporés dans le tableau 3 puisque ces deux variables sont importantes et définissent le domaine de validité de plusieurs modèles empiriques. Différentes vitesses ont été mesurées durant la campagne de terrain. En pratique, la vitesse utilisée dans les modèles d'estimation du surenfoncement est la vitesse du navire par rapport à l'eau. Dans le reste de ce travail, le terme vitesse désigne celle qui est mesurée par rapport à l'eau.

Tableau 2 : Résumé des douze voyages et caractéristiques des navires utilisés pour la campagne de terrain de 2005 (Gharbi et al., 2007).

Numéro du voyage	Type de navire	Longueur/Largeur (m)	Tirant d'eau (m)	Coefficient de blocage (m)	Itinéraire	Date
1	Porte-conteneur	294/32	10,0	0,70	Montréal - Portneuf	09/07/2005
2	Vraquier	225/32	11,0	0,87	Québec - Trois-Rivières	14/07/2005
3	Porte-conteneur	245/32	9,5	0,70	Montréal - Neuville	18/07/2005
4	Porte-conteneur	294/32	10,0	0,70	Portneuf - Montréal	25/07/2005
5	Porte-conteneur	202/31	10,0	0,69	Neuville - Montréal	29/07/2005
6	Porte-conteneur	173/27	8,9	0,60	Montréal - Neuville	31/07/2005
7	Citerne	274/48	15,5	0,82	Sault-au-cochon - Québec	05/08/2005
8	Vraquier	200/24	10,6	0,90	Québec - Contrecoeur	16/08/2005
9	Citerne	183/32	9,4	0,74	Québec - Montréal	18/08/2005
10	Citerne	183/32	11,4	0,77	Sault-au-cochon - Québec	26/08/2005
11	Vraquier	200/24	10,7	0,90	Québec - Sault-au-cochon	01/09/2005
12	Vraquier	185/24	8,6	0,84	Sorel - Montréal	02/09/2005

Tableau 3 : Minimum et maximum des principaux paramètres mesurés durant les douze voyages de la campagne de terrain de 2005 (Gharbi et al., 2007).

Paramètres mesurés	Minimum	Maximum
Surenfoncement maximum (m)	0,05	1,58
Vitesse (m/s)	1,05	7,83
Largeur du chenal (m)	235	3300
Niveau d'eau (m)	11,56	54,43
Nombre de Froude	0,0794	0,6361
Rapport niveau d'eau/tirant d'eau	1,1232	4,7533

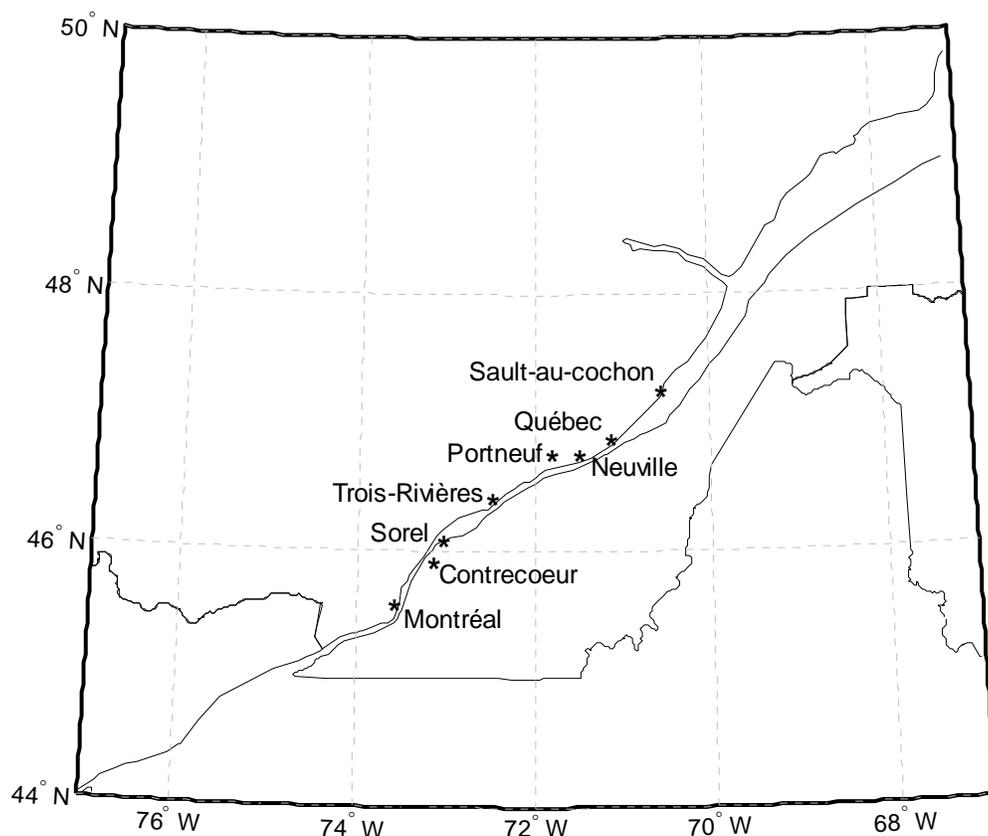


Figure 1 : Carte des villes de départ et d'arrivée des douze voyages 1) Montréal-Portneuf, 2) Québec-Trois-Rivières, 3) Montréal-Neuville, 4) Portneuf-Montréal, 5) Neuville-Montréal, 6) Montréal-Neuville, 7) Sault-au-cochon-Québec, 8) Québec-Contrecoeur, 9) Québec-Montréal, 10) Sault-au-cochon-Québec, 11) Québec- Sault-au-cochon et 12) Sorel-Montréal.

Normalement, il est préférable d'effectuer des tests de valeurs aberrantes dans les données. Il existe plusieurs techniques qui permettent de détecter des valeurs aberrantes dans les données. La plupart d'entre elles sont basées sur la normalité des données. Par contre, la série de données du surenfoncement maximum n'est pas distribuée normalement. La transformation de Box-Cox a été utilisée, mais la transformation suggérée par cette technique ne rend pas les données normales. Un test de détection de valeurs aberrantes non basé sur la normalité comme par exemple les procédures Bayésiennes présentées dans Ondo et al. (2001) pourrait être utilisée. Par contre, étant donné que la détection de valeurs aberrantes n'est pas l'objectif de ce projet et que les données ont déjà été filtrées et validées, la détection de valeurs singulières n'a pas été jugée nécessaire.

Certaines variables mesurées dans la campagne de terrain ne sont pas utiles pour cette étude. Certains autres paramètres ont du être calculés. La liste des variables qui font partie des différents modèles de surenfacement a été établie. Cette liste est utilisée comme ensemble de variables explicatives pour les sections suivantes :

- 1) le nombre de Froude (F_{nh})
- 2) le coefficient de blocage (C_b)
- 3) le facteur de vitesse de retour (S_2)
- 4) la vitesse du navire (par rapport à l'eau) (V)
- 5) le tirant d'eau (T)
- 6) la largeur du navire (b)
- 7) la hauteur totale d'eau (h)
- 8) la profondeur bathymétrique (h_z)
- 9) la largeur du chenal (W)
- 10) la largeur restreinte du chenal (B)
- 11) la longueur du navire entre les parallèles (L_{pp})
- 12) la longueur du navire (L)
- 13) l'aire de section transversale du navire (A_s)
- 14) l'aire de section transversale mouillée du navire (A_c)
- 15) l'aire nette de section transversale du navire (A_w)
- 16) réduction de l'assiette due à la propulsion en eaux peu profondes (PAR_h/T)
- 17) vitesse critique pour un chenal non restreint (V_{cr})
- 18) un facteur de correction pour la largeur du chenal (K_b)
- 19) le type de navire ($st = 0$ pour tous les types, = 1 pour les porte-conteneurs, = 2 pour les citernes et = 3 pour les vraquiers).

Une autre expression pour le facteur de correction pour la largeur du chenal est utilisée dans certains modèles. Par contre, sur tout l'ensemble de données, ce facteur de correction vaut 1, donc il n'a pas été utilisé dans les analyses.

4. Analyses préliminaires

4.1. Corrélations

Afin d'avoir une première compréhension des interactions entre les paramètres susceptibles d'influencer le surenfoncement, une analyse exploratoire a été réalisée. Dans un premier temps, des statistiques descriptives ont été calculées pour les différentes variables de la base de données afin de voir le comportement statistique des différentes variables (Annexe A). Dans un deuxième temps, les corrélations entre toutes ces variables ont été étudiées afin de choisir des variables non-redondantes pour expliquer le surenfoncement (Annexe B). Soit deux variables quelconques x et y avec des séries de données de taille n , le coefficient de corrélation entre ces deux dernières est donné par :

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (45)$$

Des matrices de corrélations ont été calculées selon cette expression. Cependant, un changement de variable pour le logarithme du surenfoncement s'est imposé. En effet, le coefficient de corrélation implique un lien fonctionnel linéaire entre les variables. La relation entre le surenfoncement et les différentes variables explicatives ne semble pas linéaire. Par ailleurs, l'examen graphique des nuages de points bivariés entre le logarithme du surenfoncement et les autres variables montre que ce lien semble plus linéaire (Annexe B). De ce fait, les corrélations ont été recalculées en utilisant le logarithme du surenfoncement (Annexe B). Il semble que les variables les plus corrélées avec le surenfoncement sont le nombre de Froude et la vitesse du navire par rapport à l'eau. De plus, la corrélation avec le nombre de Froude est augmentée en utilisant cette transformation de données. Des analyses en composantes principales et des arbres de classification ont été réalisés. Comme ces analyses ne donnaient pas de résultats concluants, elles n'ont pas été incluses dans ce travail.

4.2. Calcul du nombre de Froude

Le nombre de Froude a été présenté dans la section 2 du rapport (équation 9). Ce dernier est parmi les composantes les plus importantes pour expliquer le surenfoncement. L'équation du nombre de Froude est répétée ici :

$$F_{nh} = \frac{V}{\sqrt{g \cdot h}} \quad (46)$$

où V est la vitesse de l'écoulement, g l'accélération de la pesanteur et h une dimension linéaire caractéristique du phénomène. Dans le cadre de ce projet, V correspond à la vitesse réelle du navire (la vitesse par rapport à l'eau, en m/s), g vaut 9.80665 m/s^2 et comme dimension linéaire caractéristique du phénomène, la profondeur totale d'eau (h) est utilisée.

Plusieurs paramètres peuvent être utilisés pour représenter la dimension linéaire caractéristique du phénomène. La dimension linéaire a été déterminée de façon à ce que le nombre de Froude soit le plus corrélé possible avec le surenfoncement. Le nombre de Froude a été calculé avec la largeur du navire, le tirant d'eau, la longueur du navire entre les parallèles, la hauteur totale d'eau moins le tirant d'eau et la largeur du chenal moins la largeur du navire. Les matrices de corrélations ont été recalculées, et la première expression du nombre de Froude utilisée a été retenue puisque cette dernière donne la plus forte corrélation avec le logarithme du surenfoncement maximum. Dans le reste du rapport, le nombre de Froude s'exprime en fonction de la hauteur totale d'eau.

4.3. Régression pas-à-pas

Pour sélectionner des variables explicatives dans une grosse base de données, l'utilisation de la régression pas-à-pas est appropriée. Par ailleurs, l'utilisation de la régression pas-à-pas comme outil de prédiction peut être controversée. En effet, la capacité prédictive du modèle peut être affectée par la sélection du modèle (Hurvich et Tsai, 1990). Le test de Fisher et le coefficient de détermination utilisés pour faire la sélection du modèle sont biaisés dans le cas de la régression pas-à-pas (Rencher et Pun, 1980). L'utilisation de la régression pas-à-pas

dans cette section est faite à titre exploratoire, pour déterminer les variables qui semblent les plus importantes pour décrire le surenfoncement.

Cette technique est présentée dans Myers (1986). C'est une procédure automatique, qui étape par étape, se base sur un critère pour ajouter ou retirer une variable explicative du modèle. Le modèle de départ, sans variable explicative, s'exprime ainsi :

$$Y = \beta_0 + \varepsilon \quad (47)$$

où Y représente le vecteur d'observations de la variable dépendante, β_0 l'ordonnée à l'origine et ε les résidus. Pour vérifier si l'ajout ou le retrait d'une variable a un effet sur la contribution des autres variables déjà dans le modèle, le critère le plus communément utilisé est le test de Fisher. Ce dernier consiste à comparer les sommes de carrés résiduelles des différents modèles afin de choisir celui qui ajuste le mieux les données. Le test de Fisher peut s'exprimer ainsi :

$$F = \frac{RSS_2 - RSS_1 / 1}{RSS_2 / l} \quad (48)$$

où RSS_1 représente la somme de carrés des résidus du premier modèle, RSS_2 la somme de carrés des résidus du deuxième modèle et l est le nombre de degrés de liberté du deuxième modèle. La procédure se termine lorsqu'aucune variable explicative ne peut être ajoutée ou retirée du modèle.

Quatre modèles de régression ont été ajustés. Le premier a été ajusté avec l'ensemble des données et les autres modèles avec les données des navires de type porte-conteneurs, citerne et vraquier. Le tableau 4 présente les coefficients de régression estimés pour chaque variable explicative. Une valeur de 0 signifie que cette variable n'ajoute pas un apport significatif au modèle (seuil critique de 5%). Le tableau 4 présente les résultats d'ajustement des différents modèles.

Tableau 4 : Coefficients des différents modèles de régression pas-à-pas.

Variable	Coefficients de régression			
	Ensemble des données	Porte-conteneurs	Citerne	Vraquier
β_0	-2,9490	-25,3632	-21,0766	-5,0917
F_{nh}	-0,9768	7,4335	-2,5653	2,4012
C_b	0,0000	47,3619	32,9460	0,0000
S_2	-1,7714	14,3166	6,8873	5,5464
V	0,4406	-0,2626	0,6209	0,3511
T	0,2759	1,9277	-0,2242	0,2383
b	0,1982	0,0000	0,0000	0,0172
h	0,1507	0,2566	-0,1589	0,0000
W	0,0000	-0,0002	-0,0001	-0,0001
B	0,0000	0,0000	0,0000	0,0000
L_{pp}	-0,0219	0,0006	0,0000	0,0000
L	0,0195	0,0000	0,0000	0,0000
A_s	0,0000	-0,0869	0,0000	0,0000
A_c	-0,0103	0,0000	-0,0001	0,0002
A_w	0,0102	0,0000	0,0000	0,0000
h_z	0,0415	0,0000	0,2419	0,1224
PAR_h/T	0,0000	1,1735	0,0000	0,4820
K_b	-3,2926	-5,0971	-2,9277	-1,5448
V_{cr}	-0,4784	-0,3498	-0,2612	-0,4065
st	0,0645	0,0000	0,0000	0,0000

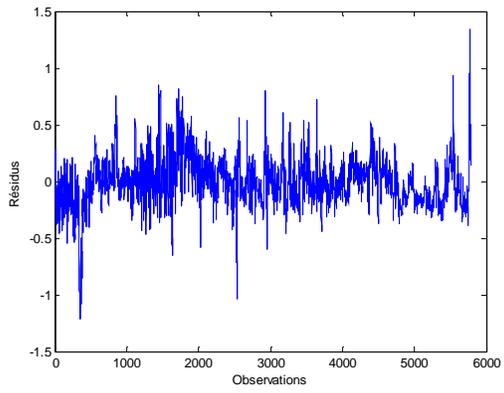
Tableau 5 : Adéquation des différents modèles de régression pas-à-pas.

Statistique	Ensemble des données	Porte-conteneurs	Citerne	Vraquier
R^2	0,8398	0,8028	0,9092	0,8291
EQM	0,0115	0,0092	0,0110	0,0023
BMR	2,1827	1,8237	1,0211	1,2709
Fisher	2160,72	895,17	1299,56	741,93
p-valeur	0,0000	0,0000	0,0000	0,0000

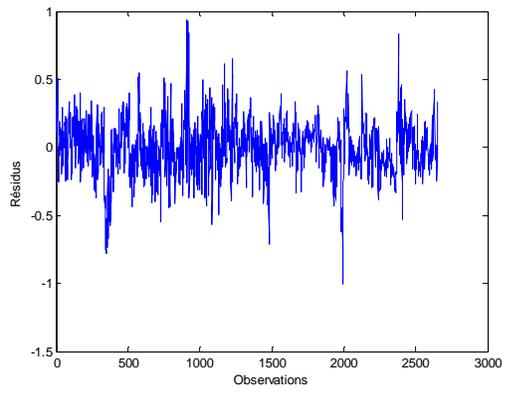
Le tableau 4 montre qu'il est important d'ajuster des modèles séparés pour les différents types de navires puisque les coefficients de régression diffèrent beaucoup d'un modèle à l'autre. Il arrive souvent qu'une variable soit incluse dans le modèle représentant un type de navire et exclue dans le modèle représentant un autre type de navire. Par exemple, le coefficient de blocage est une variable significative dans les modèles ajustés avec les données des navires de types porte-conteneur et citerne, mais ne semble pas significative pour les vraquiers ainsi que pour le modèle ajusté sur l'ensemble des données. De plus, le nombre de Froude a un coefficient négatif pour l'ensemble des données et les navires de type citerne alors que son coefficient est positif pour les modèles des porte-conteneurs et des vraquiers. Cependant, ces différences de coefficients peuvent être liées à un mauvais ajustement du modèle du à la multicollinéarité entre les variables. Il est intéressant de noter que dans tous les cas, la procédure d'inclusion choisit toujours le nombre de Froude comme première variable explicative à entrer dans le modèle. La largeur restreinte du chenal (B) est la seule variable qui n'est pas significative pour tous les modèles à la fois. Cette dernière est directement liée à la largeur du navire ($B = 10 \cdot b$).

La performance des différents modèles semble plutôt satisfaisante en se fiant aux statistiques présentées dans le tableau 5. Par ailleurs, tous ces modèles reposent sur les postulats de normalité, d'homoscédasticité et d'indépendance des résidus. Pour vérifier les hypothèses de base, des graphiques des résidus en fonction des observations (indépendance), des résidus en fonction des valeurs prédites (homoscédasticité) et des droites de Henry (normalité) ont été produits et sont représentés dans les figures 1, 2 et 3. Les résidus des différents modèles semblent indépendants (figure 2). Les variances de ces modèles ne semblent pas homogènes (figure 3). Il semble y avoir une plus grande variabilité dans les résidus pour les valeurs prédites inférieures à -1 que pour les plus grandes valeurs prédites. La normalité semble respectée seulement pour le modèle ajusté sur les données des navires de type citerne (figure 4). Les résidus des autres modèles ne sont pas distribués normalement.

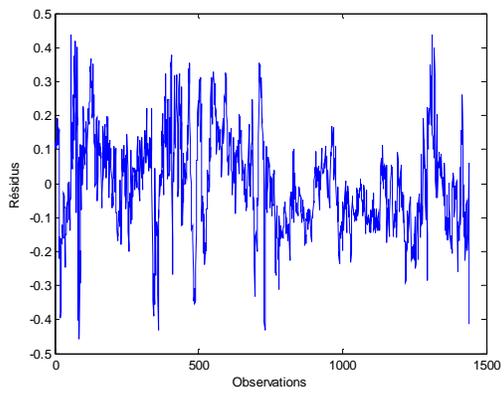
a)



b)



c)



d)

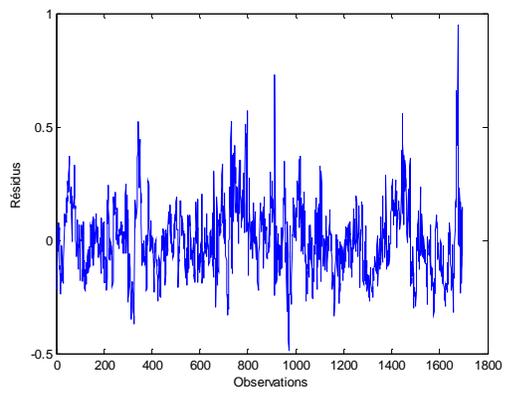
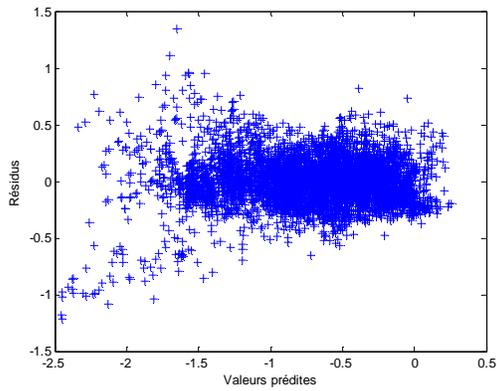
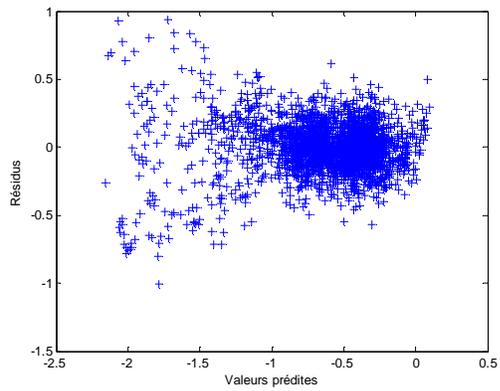


Figure 2 : Résidus en fonction des observations pour le modèle de régression pas-à-pas ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

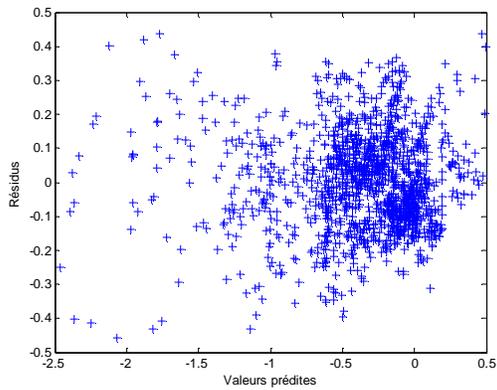
a)



b)



c)



d)

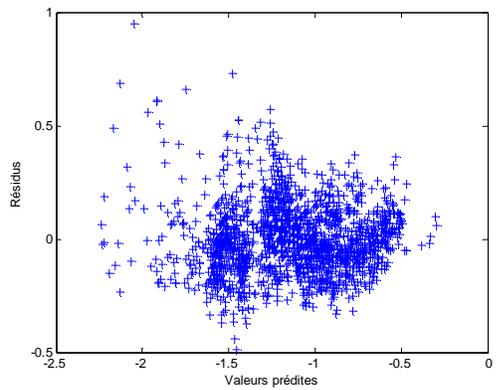


Figure 3 : Résidus en fonction des valeurs prédites pour le modèle de régression pas-à-pas ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

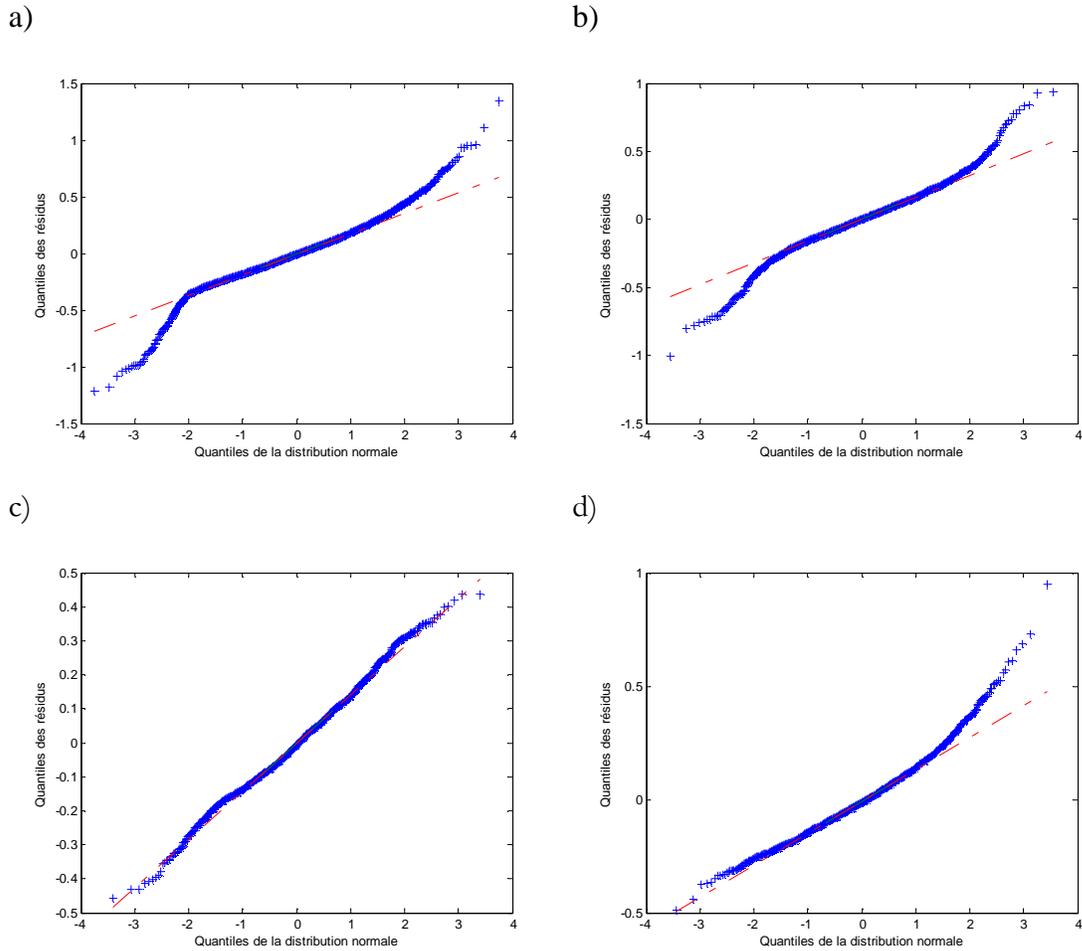


Figure 4 : Droite de Henry des résidus pour le modèle de régression pas-à-pas ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

L'analyse des postulats de ces modèles montre que ces derniers n'ajustent pas très bien les données. De plus, l'hypothèse de linéarité entre le logarithme du surenfoncement et les différentes variables explicatives n'est probablement pas raisonnable. En effet, les différents modèles d'estimation du surenfoncement présentés dans la littérature représentent des relations beaucoup plus complexes.

Par ailleurs, il peut y avoir un problème de multicollinéarité entre les variables explicatives. La relation entre la variable dépendante et les variables explicatives peut être faussée lorsqu'il y a présence de multicollinéarité, ce qui fait en sorte qu'une explication physique de la

relation peut être difficile. Un modèle de régression qui tient compte du problème de multicollinéarité pourrait être utilisé.

4.4. Régression ridge

Comme les modèles de régression pas-à-pas ajustés dans la section précédente ne sont pas satisfaisants, d'autres techniques de modélisation ont été utilisées. Il est possible que les problèmes d'ajustement soient dus en partie à de la multicollinéarité entre les variables explicatives. On peut éviter l'utilisation de variables explicatives trop corrélées entre elles en choisissant seulement les plus pertinentes avec la régression ridge (Marquardt and Snee, 1975). Cette technique, présentée dans (Myers, 1986), permet de tenir compte de la multicollinéarité entre les variables explicatives en modifiant les estimateurs de la régression. Les paramètres sont modifiés en introduisant une constante (k) dans le système d'équations de la méthode des moindres carrés :

$$\beta^R = (r_{XX} + kI_p)^{-1} r_{YX} \quad (49)$$

où $\beta^R = (\beta_1^R, \dots, \beta_p^R)$ est le vecteur des coefficients de régression ridge pour les p variables explicatives, r_{XX} est la matrice des corrélations entre les variables explicatives, r_{YX} est la matrice des corrélations entre la variable dépendante et les variables explicatives et I_p est la matrice identité de taille $p \times p$. Lorsque $k = 0$, les estimateurs sont les mêmes que ceux d'une régression multiple ordinaire. Lorsque $k > 0$, les estimateurs sont biaisés, mais plus stables que ceux d'une régression multiple puisqu'ils sont corrigés pour la multicollinéarité entre les variables.

La constante k se détermine par optimisation. Elle peut être définie par une procédure automatique qui se base sur l'indice ISRM :

$$ISRM = \sum_{i=1}^p \left[\frac{(\lambda_i + k)^2}{\sum_{j=1}^p \lambda_j / (\lambda_j + k)^2} - 1 \right]^2 \quad (50)$$

où $\lambda_i (i=1, \dots, p)$ représente les valeurs propres de la matrice des corrélations entre les variables explicatives. Il est suggéré de choisir la valeur de k qui minimise l'indice *ISRM* (Vinod, 1976).

La régression ridge a été appliquée pour construire quatre modèles : sur l'ensemble des données, les données des navires de type porte-conteneurs, citerne et vraquier en considérant les mêmes variables que dans la section précédente. Le seuil critique utilisé est de 5%. Le tableau 6 présente les coefficients de régression estimés pour chaque variable explicative. Le tableau 7 présente les statistiques des différents modèles.

Tableau 6 : Coefficients des différents modèles de régression ridge.

Variable	Coefficients de régression			
	Ensemble des données	Porte-conteneurs	Citerne	Vraquier
k	0,0400	0,0300	0,0100	0,0200
β_o	-4,2364	-2,3521	-5,1320	-8,1113
F_{nh}	1,2747	1,9569	1,0430	3,1758
C_b	0,4851	3,2990	4,5640	2,3652
S_2	1,7698	4,4781	8,0579	3,0954
V	0,2277	0,1409	0,3457	0,1951
T	0,0390	-0,2527	0,0532	0,0872
b	0,0450	0,0210	-0,0028	0,0099
h	0,0014	0,0003	-0,0039	-0,0194
W	0,0000	-0,0002	-0,0001	-0,0001
B	0,0045	0,0021	-0,0003	0,0010
L_{pp}	-0,0013	-0,0001	-0,0008	0,0040
L	-0,0009	-0,0012	-0,0006	0,0042
A_s	-0,0010	-0,0008	0,0003	0,0010
A_c	-0,0001	0,0000	0,0000	0,0000
A_w	-0,0001	0,0000	0,0000	0,0000
h_z	0,0106	0,0103	0,0168	0,0511
PAR_h/T	-0,2218	-0,1326	0,3929	-0,1018
K_b	-0,6222	-0,7062	-1,7155	-0,7205
V_{cr}	0,0108	-0,0242	-0,0340	-0,0544
st	-0,0291	0,0000	0,0000	0,0000

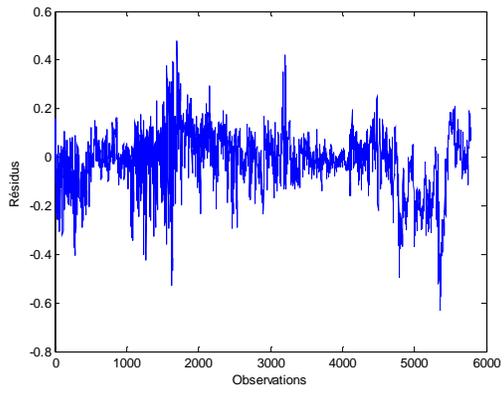
Tableau 7 : Adéquation des différents modèles de régression ridge.

Statistique	Ensemble des données	Porte-conteneurs	Citerne	Vraquier
R²	0,7623	0,7252	0,8828	0,7653
EQM	0,0147	0,0107	0,0116	0,0024
BMR	2,6038	2,3013	1,1395	1,3997
Fisher	1251,7	438,6	689,9182	385,9164
p-valeur	0,0000	0,0000	0,0000	0,0000

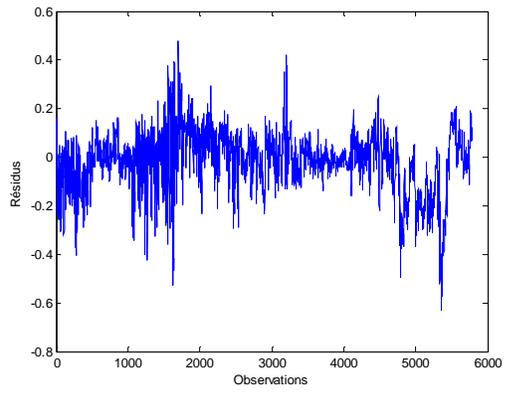
On peut voir que les différences entre les coefficients des paramètres entre les modèles sont moins prononcées avec une approche de régression ridge qu'avec la régression pas-à-pas. Ceci vient probablement du problème de multicollinéarité entre les variables. De plus, la performance des différents modèles est inférieure dans le cas de la régression ridge. En effet, la perte de variance expliquée constitue la contrepartie d'une meilleure stabilité des coefficients. La plupart des variables semblent significatives excepté pour A_c et A_w qui ne sont pas incluses dans trois des quatre modèles ajustés.

Ces modèles sont basés sur la normalité, l'homoscédasticité et l'indépendance des résidus. Pour vérifier les hypothèses de base, des graphiques des résidus en fonction des observations (indépendance), des résidus en fonction des valeurs prédites (homoscédasticité) et des droites de Henry (normalité) ont été produits et sont représentés dans les figures 5, 6 et 7.

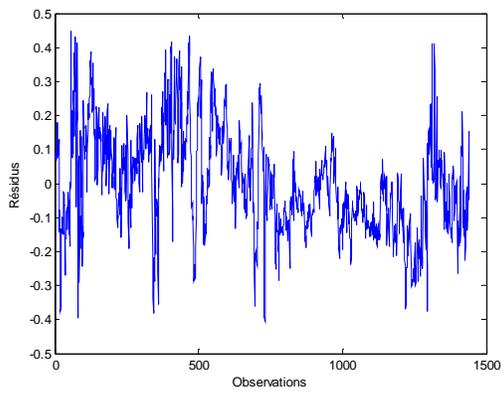
a)



b)



c)



d)

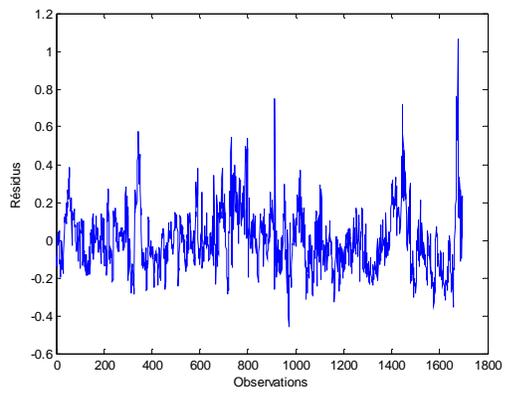
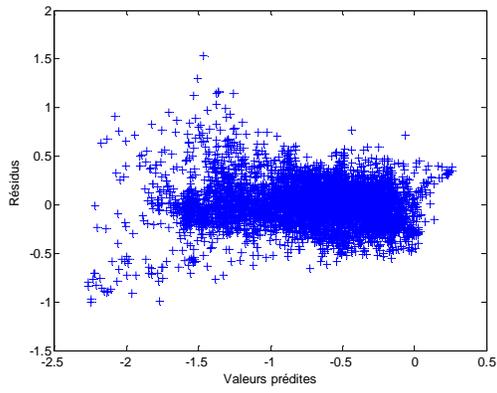
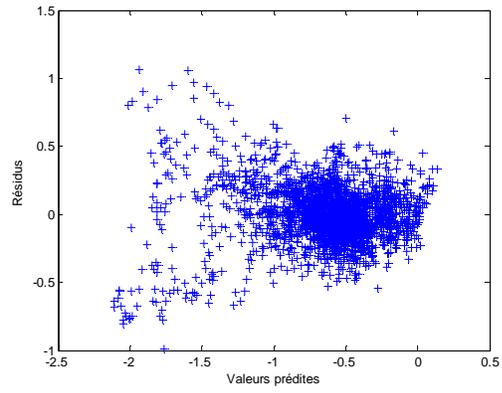


Figure 5 : Résidus en fonction des observations pour le modèle de régression ridge ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

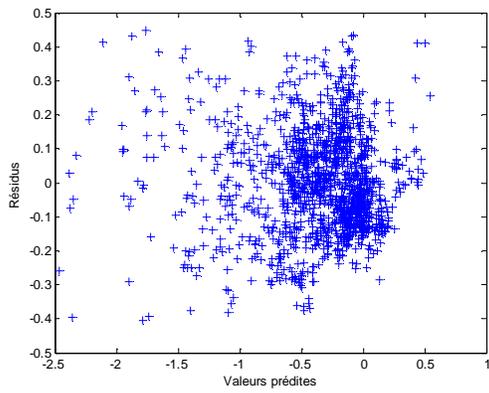
a)



b)



c)



d)

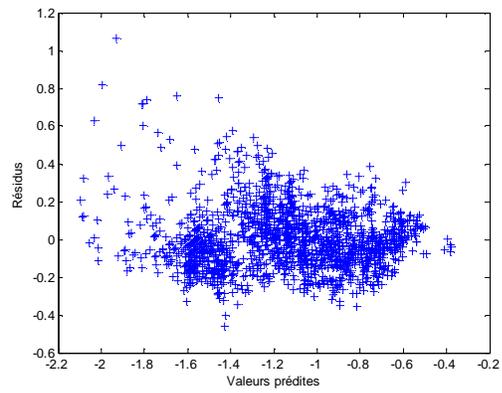
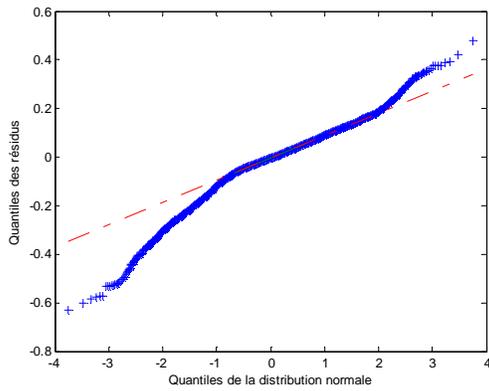
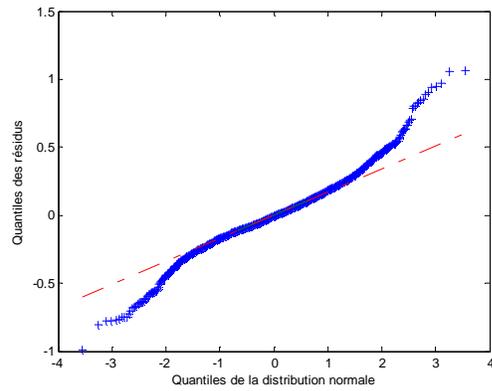


Figure 6 : Résidus en fonction des valeurs prédites pour le modèle de régression ridge ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

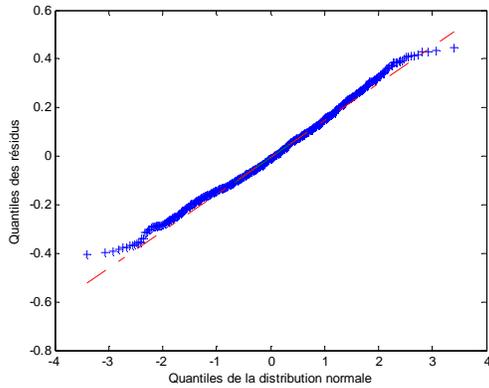
a)



b)



c)



d)

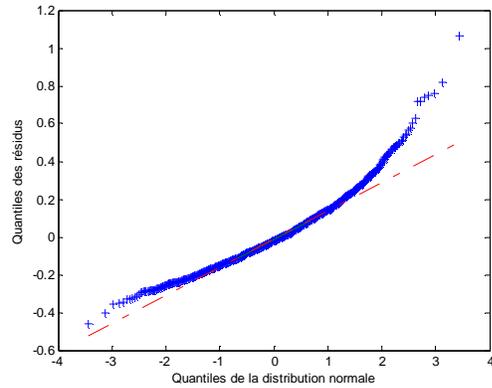


Figure 7: Droite de Henry des résidus pour le modèle de régression ridge ajusté sur a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

L'analyse des postulats de ces modèles montre que ces derniers n'ajustent pas très bien les données. Un modèle non paramétrique qui n'est pas basé sur des hypothèses contraignantes et qui n'assume aucune forme dans la relation entre la variable dépendante et les variables explicatives serait préférable.

5. Nouveaux modèles

5.1. Arbres de régression

Jusqu'à maintenant, les techniques de régression utilisées supposent que l'on connaît la forme de la relation entre le surenfoncement maximum et les variables explicatives. Plus particulièrement, on suppose la linéarité entre le logarithme du surenfoncement et les variables explicatives. Par contre, cette relation ne semble pas linéaire et plutôt complexe. De plus, les hypothèses de base des résidus ne sont pas toujours respectés ce qui fait en sorte que les différents modèles ajustés ne sont pas valides. Dans ce cas, une technique de type non-paramétrique basée sur les moindres carrés comme les arbres de classification et de régression, peut être utilisée (Breiman et al., 1984). C'est un arbre constitué d'une suite de questions qui peuvent être répondues par oui ou non combinée à un ensemble de valeurs prédites. La figure 8 présente un schéma simplifiée de la technique.

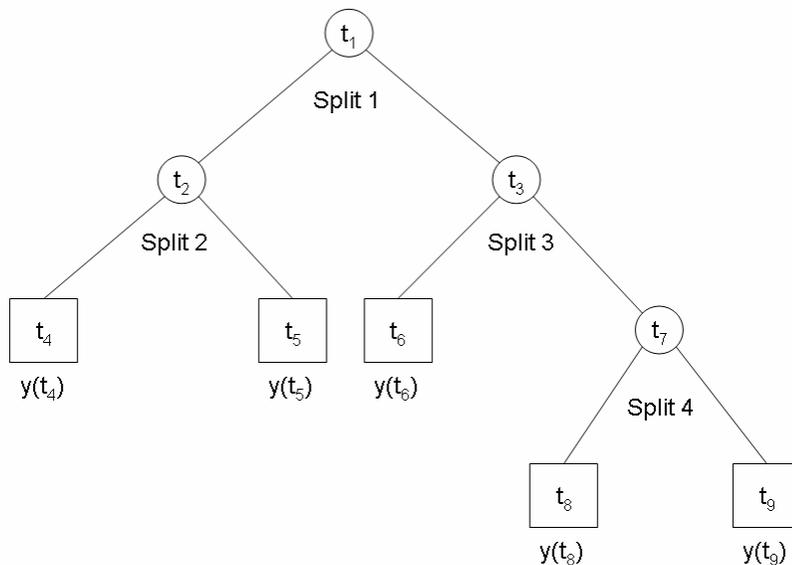


Figure 8 : Schéma simplifiée d'un arbre de régression (Breiman et al., 1984).

Les données sont partitionnées dans une séquence de partages binaires jusqu'à des nœuds terminaux qui donnent tous une valeur prédite. Le partage se fait sur les variables explicatives qui peuvent être quantitatives ou catégoriques. Par exemple, les conditions de partage pourraient être de la forme : $\{F_{mh} \leq 0.3\}$ ou $\{st = 1\}$. Il y a trois éléments nécessaires à la construction d'un arbre de régression : 1) une façon de partager les données à chaque nœud intermédiaire, 2) une règle pour déterminer les nœuds terminaux et 3) une règle permettant d'assigner une valeur prédite à un nœud terminal (Breiman et al., 1984). Premièrement, à chaque nœud intermédiaire, la détermination du partage des données vient minimiser la somme de carrés résiduelle (c'est la meilleure façon de séparer les fortes valeurs des faibles valeurs de la variable dépendante). Deuxièmement, un nœud est considéré terminal lorsqu'il est pur (toutes les valeurs de la variable dépendante à l'intérieur du nœud sont égales) ou que le nombre de cas à l'intérieur du nœud est inférieur ou égal à la taille minimale fixée (cette valeur est arbitraire et dépend du nombre d'observations). Pour ce travail, la taille minimale a été fixée à 10. Troisièmement, les valeurs prédites assignées à chaque nœud sont en fait les moyennes respectives des valeurs de la variable réponse dans chaque nœud.

En construisant l'arbre de départ (T_{\max}), une séquence d'arbre est obtenue $\{T_1, T_2, \dots, T_{\max}\}$. Le problème avec l'arbre de départ est que sa taille n'est pas nécessairement optimale et ses valeurs prédites peuvent être trop optimistes. Un processus d'élagage peut être entrepris pour déterminer l'arbre ayant la taille optimale qui donne des estimations précises, tout en respectant le principe de parcimonie. L'élagage consiste à calculer l'erreur quadratique relative moyenne pour chaque arbre de la séquence. L'erreur quadratique relative moyenne est calculée à partir des erreurs obtenues par validation croisée pour avoir une estimation plus honnête. Pour ce travail, une validation croisée en 10 blocs a été employée. Cela consiste à diviser les données en 10 groupes aléatoires de même taille. Les valeurs de la variable réponse dans chaque groupe sont prédites avec les observations restantes. Dans Breiman et al. (1984), il est suggéré de choisir l'arbre de plus petite taille pour lequel l'erreur quadratique relative moyenne est de plus de un écart-type que l'erreur quadratique relative minimale. L'arbre sélectionné peut être utilisé comme prédicteur. Dans le reste du rapport, les modèles ajustés par arbre de régression seront notés ***RegT***.

Cette technique a été appliquée pour estimer le logarithme du surenfacement avec les mêmes variables explicatives qui ont servi à ajuster les modèles de régression pour l'ensemble des données et par type de navire également. On modélise le logarithme du surenfacement et non le surenfacement puisque la performance s'avère supérieure dans ce cas. Dans la section 2, une liste de toutes les variables susceptibles d'influencer le surenfacement a été dressée. À partir de cette liste, des variables jugées peu pertinentes selon l'expert Dr. Samir Gharbi (communication personnelle, 2007) ont été retirées du jeu de données. De plus, une variable a été ajoutée au modèle, toujours selon les conseils de l'expert Dr. Samir Gharbi (communication personnelle, 2007). Le but était de fournir un modèle contenant des variables explicatives pertinentes au sens physique également. Les variables suivantes ont été sélectionnées :

- 1) le nombre de Froude (F_{nh})
- 2) le coefficient de blocage (C_b)
- 3) le facteur de vitesse de retour (S_2)
- 4) la vitesse du navire (par rapport à l'eau) (V)
- 5) le tirant d'eau (T)
- 6) la largeur du navire (b)
- 7) la hauteur totale d'eau (h)
- 8) la largeur restreinte du chenal (B)
- 9) la longueur du navire entre les parallèles (L_{pp})
- 10) la longueur du navire (L)
- 11) l'aire de section transversale du navire (A_s)
- 12) l'aire de section transversale mouillée du navire (A_c)
- 13) l'aire nette de section transversale du navire (A_w)
- 14) réduction de l'assiette due à la propulsion en eaux peu profondes (PAR_h/T)
- 15) vitesse critique pour un chenal non restreint (V_{cr})
- 16) le rapport de la hauteur totale d'eau sur le tirant d'eau (h/T)
- 17) le type de navire ($st = 0$ pour tous les types, $= 1$ pour les porte-conteneurs, $= 2$ pour les citernes et $= 3$ pour les vraquiers).

La largeur du chenal n'a pas été incluse puisque cette variable est différente de la largeur de la voie navigable. En fait, la largeur restreinte du chenal est plus significative au sens physique. De plus, la largeur du chenal ne semblait pas très significative si on se base sur les résultats des procédures de régression pas-à-pas et de régression ridge utilisées dans la section précédente. Par le fait même, les facteurs de correction pour la largeur du chenal n'ont pas été inclus dans le modèle. Enfin, la profondeur bathymétrique a été retirée du jeu de données parce que cette profondeur est calculée par rapport au zéro des cartes et a été jugée moins significative que la hauteur totale d'eau pour le surenfoncement. Ces quatre variables ont été retirées systématiquement du jeu de données. Par ailleurs, la variable h/T , le rapport entre la hauteur totale d'eau et le tirant d'eau, a été ajoutée à la liste des variables explicatives selon les conseils de l'expert Samir Gharbi.

Le modèle complet comprenant cette liste de variables explicatives a d'abord été ajusté. Par la suite, d'autres variables ont été retirées du jeu de données en comparant à chaque fois les critères de performance des modèles avec et sans ces variables. L'objectif était de réduire le nombre de variables explicatives sans trop diminuer la capacité prédictive du modèle. On a choisi entre L_{pp} et L , la variable qui affecte le moins la performance des modèles puisque ces deux dernières sont un peu redondantes. De plus, dans le but d'avoir un modèle simple, on a vérifié si les variables $PAR_{h/T}$, A_s et A_c donnaient un apport significatif aux modèles. Pour évaluer la performance des modèles, les critères retenus par Gharbi et Le Page (2007) soit : le coefficient de détermination (R^2), l'erreur quadratique moyenne (EQM) et le biais moyen relatif (BMR) ont été calculés. Le tableau 8 présente les différents modèles ajustés.

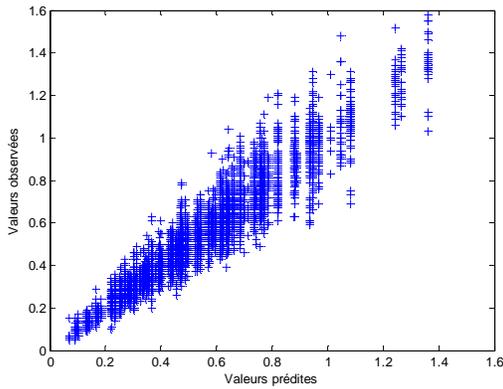
Tableau 8 : Performance des différents modèles ajustés.

Variables retirées du modèle <i>RegT</i>	Critères de performance											
	Tous			Porte-conteneurs			Citerne			Vraquier		
	R ²	EQM	BMR	R ²	EQM	BMR	R ²	EQM	BMR	R ²	EQM	BMR
Aucune	0,91	0,01	1,0	0,83	0,01	1,1	0,88	0,01	0,9	0,88	0,00	0,7
L	0,92	0,01	0,9	0,82	0,01	1,1	0,88	0,01	0,9	0,89	0,00	0,7
Lpp	0,92	0,01	0,9	0,83	0,01	1,1	0,88	0,01	0,9	0,89	0,00	0,7
par_ht	0,92	0,01	0,9	0,84	0,01	1,0	0,89	0,01	0,8	0,89	0,00	0,7
As	0,91	0,01	0,9	0,83	0,01	1,1	0,88	0,01	0,9	0,88	0,00	0,7
Ac	0,92	0,01	0,9	0,85	0,01	1,0	0,88	0,01	0,9	0,88	0,00	0,7
Ac, As	0,91	0,01	1,0	0,83	0,01	1,1	0,88	0,01	0,9	0,87	0,00	0,9
Ac, As, par_ht	0,91	0,01	0,9	0,84	0,01	1,0	0,88	0,01	0,9	0,89	0,00	0,7
Ac, As, par_ht, Lpp	0,92	0,01	0,9	0,84	0,01	1,0	0,88	0,01	0,9	0,88	0,00	0,8
Ac, As, par_ht, L *	0,91	0,01	0,9	0,84	0,01	1,0	0,89	0,01	0,8	0,89	0,00	0,7

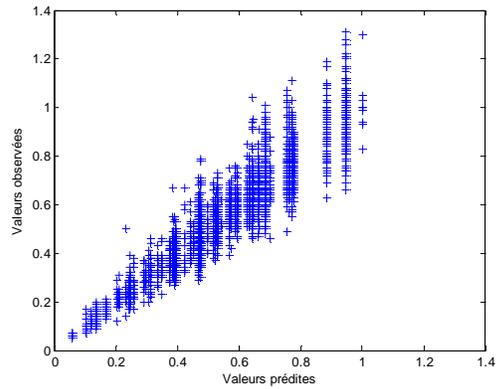
*=Modèle proposé

Les critères de performance entre les différents modèles varient très légèrement entre les différents modèles. Le dernier modèle ajusté (tableau 8) a été retenu puisque c'est un des modèles les plus simples avec une bonne performance. Le choix est un peu arbitraire étant donné que les critères de performance varient très peu selon les variables retirées du modèle. Les quatre arbres de régression correspondants sont placés en annexe C, ils sont difficilement interprétables visuellement étant donné leur taille. La figure 8 présente les valeurs observées et prédites par les quatre modèles retenues. Les prédictions semblent un peu plus problématiques dans le cas des grandes valeurs observées. Néanmoins, les valeurs prédites sont près des valeurs observées en général.

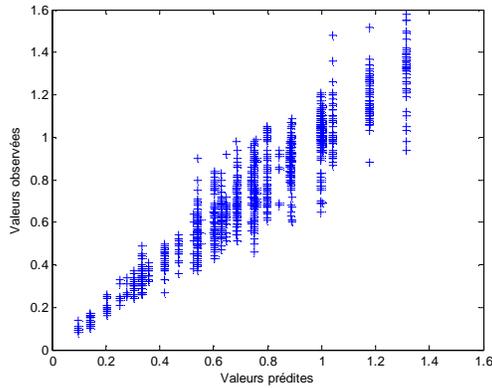
a)



b)



c)



d)

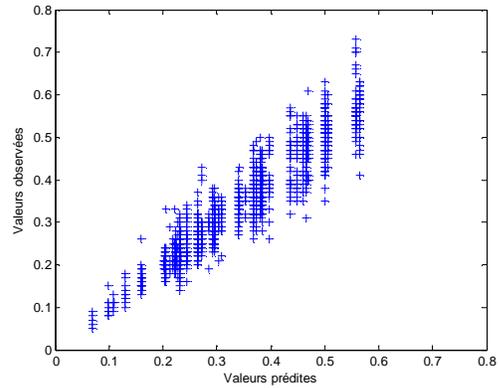


Figure 9 : Valeurs observées et valeurs prédites obtenues par arbre de régression ajusté sur a) l'ensemble des données b) porte-conteneurs, c) citerne et d) vraquier.

Dans le tableau 9, les résultats des dix modèles jugés les plus performants dans l'étude de Gharbi et Le Page (2007) sont présentés pour comparer avec le modèle retenu. En comparant ces résultats avec ceux obtenus des dix modèles jugés les plus performants, on peut voir que les coefficients de détermination obtenus par arbre de régression sont supérieurs à tous ceux des autres modèles et les erreurs quadratiques moyennes sont inférieures dans tous les cas. Par contre, le biais moyen relatif n'est pas systématiquement meilleur dans tous les cas. Les modèles *Eryuzlu 3* et *Simard* ont un meilleur biais moyen relatif dans le cas de l'ensemble des données, *Eryuzlu 2* et *Japanese* pour le type porte-conteneurs et *Romisch arrière* pour le type vraquier.

Tableau 9 : Performance des modèles empiriques d'estimation du surenfoncement et du modèle retenu qui a été développé par arbre de régression.

Modèle	Critères de performance											
	Tous			Porte-conteneurs			Citerne			Vraquier		
	R ²	EQM	BMR	R ²	EQM	BMR	R ²	EQM	BMR	R ²	EQM	BMR
<i>RegT</i>	0,91	0,01	0,9	0,84	0,01	1,0	0,89	0,01	0,8	0,89	0,00	0,7
<i>Barrass1*</i>	0,54	0,03	18,0	0,66	0,02	6,0	0,80	0,03	-12,7	0,51	0,06	63,0
<i>Eryuzlu1*</i>	0,62	0,03	3,1	0,65	0,02	9,0	0,77	0,05	-20,6	0,51	0,02	13,8
<i>Eryuzlu2*</i>	0,56	0,03	-1,7	0,62	0,02	0,3	0,80	0,07	-28,0	0,45	0,02	17,5
<i>Eryuzlu3*</i>	0,59	0,03	0,2	0,64	0,02	2,9	0,81	0,06	-25,8	0,48	0,02	18,2
<i>Japanese*</i>	0,66	0,03	19,3	0,56	0,02	-0,4	0,80	0,02	14,0	0,59	0,04	54,6
<i>Millward2*</i>	0,45	0,09	51,4	0,63	0,04	23,4	0,82	0,02	13,1	0,46	0,21	127,8
<i>Norrbir*</i>	0,67	0,02	6,7	0,57	0,02	-8,7	0,81	0,02	-3,1	0,61	0,02	39,2
<i>Romisch avant*</i>	0,61	0,04	7,5	0,51	0,04	-18,7	0,66	0,07	22,0	0,72	0,02	36,1
<i>Romisch arrière*</i>	0,53	0,04	-13,9	0,57	0,03	-10,7	0,63	0,11	-36,1	0,52	0,01	0,0
<i>Simard*</i>	0,59	0,03	0,7	0,64	0,02	1,0	0,79	0,07	-24,0	0,48	0,02	21,3

*=Résultats tirés de Gharbi et Le Page (2007)

Il est à noter que le modèle par arbre de régression a été construit pour prédire le logarithme du surenfoncement, mais pour calculer les critères de performance, on fait la transformation inverse (exponentielle) pour revenir au surenfoncement prédit. Les modèles ajustés par arbre de régression se montrent performants puisqu'ils décrivent bien l'ensemble des données utilisées pour les ajuster. Cependant, il est nécessaire de vérifier la qualité des prédictions et d'obtenir des critères de performance plus 'honnêtes'. Pour ce faire, la validation croisée de type 'leave-one-out' a été utilisée. Cela consiste à prédire tour à tour toutes les observations de la variable dépendante, en ajustant l'arbre avec toutes les observations restantes. Par la suite, les erreurs sont calculées. Le tableau 10 présente les critères de performance du modèle retenu obtenus par validation croisée.

Tableau 10 : Critères de performance des modèles *RegT* calculés avec des erreurs de prédiction obtenues par validation croisée.

Critère	Ensemble des données	Porte-conteneurs	Citerne	Vraquier
BMR	1,3	1,1	1,2	0,9
EQM	0,01	0,01	0,01	0,00
R²	0,89	0,80	0,87	0,85

Dans tous les cas, le biais moyen relatif et l'erreur quadratique moyenne sont faibles. Donc, il semble que les arbres de régression ajustés peuvent fournir des valeurs prédites de qualité.

5.2. Modèle *Eryuzlu 1* optimisé

Le modèle d'Eryuzlu 1 a été adopté par la Garde Côtière Canadienne pour estimer le surenfoncement (MacPherson, 2002). Il a été présenté dans la section 2.3 du document (équation 8). Ce modèle a été jugé un des plus performants dans l'étude de Gharbi et Le Page (2007). Cependant, selon les recommandations de cette même étude, ce modèle pourrait être optimisé afin de réduire les erreurs et mieux estimer le surenfoncement à basse et haute vitesse. Pour optimiser le modèle, ses erreurs ont été modélisées par régression pas-à-pas. Dans un premier temps, le modèle d'Eryuzlu a été utilisé pour calculer les valeurs prédites de surenfoncement et les erreurs du modèle ont été obtenues :

$$E_{E1} = S_{obs} - 0.181 \cdot \sqrt{T} \cdot b \cdot \left(\frac{V}{\sqrt{g \cdot T}} \right)^{2.269} \cdot \left(\frac{T}{h} \right)^{0.994} \quad (51)$$

où E_{E1} représente les erreurs du modèle *Eryuzlu 1*, S_{obs} est le surenfoncement observé, T est le tirant d'eau, b est la largeur du navire, F_{nh} est le nombre de Froude et h est la profondeur totale de l'eau. La figure 9 présente les valeurs observées et prédites obtenues par le modèle *Eryuzlu 1*. Dans la plupart des cas, les grandes valeurs de surenfoncement sont sous-estimées par le modèle (le BMR est positif). Pour les navires de type vraquier, le modèle semble surestimer le surenfoncement.

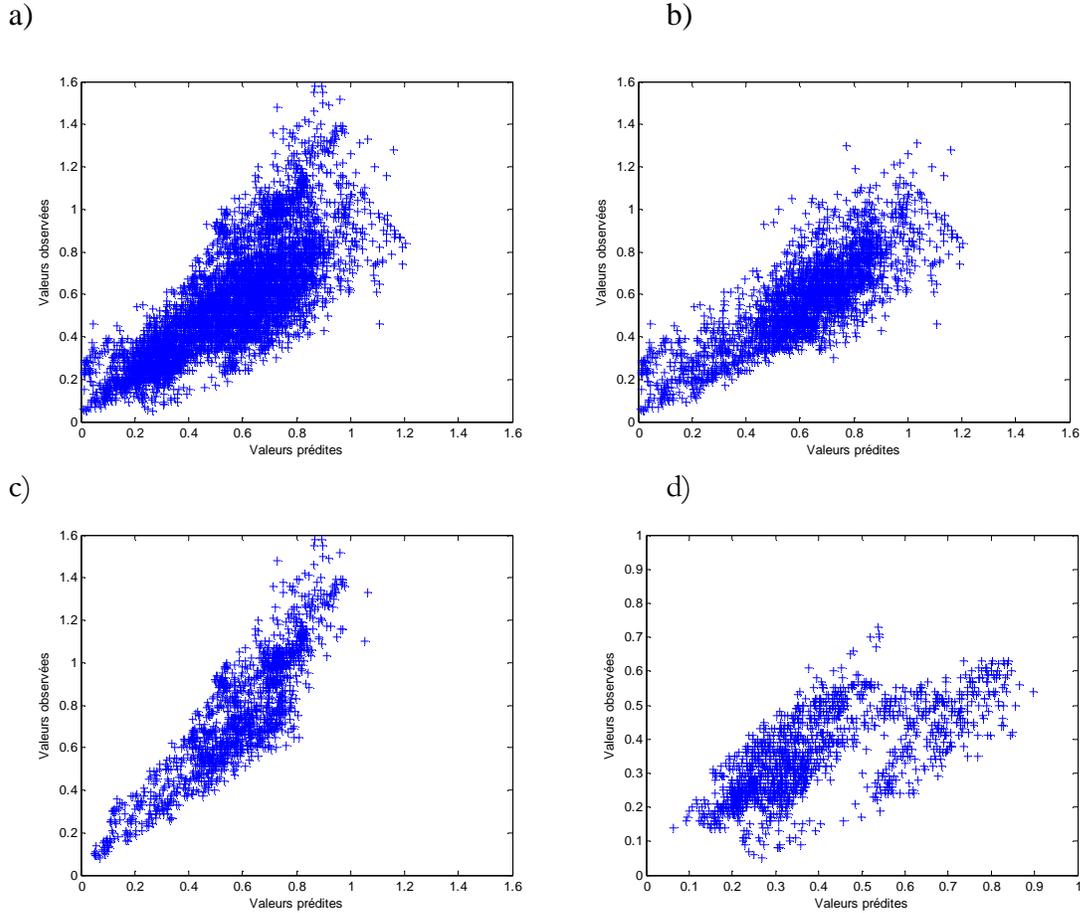


Figure 10 : Valeurs observées et valeurs prédites obtenues par le modèle *Eryuzlu 1* ajusté sur a) l'ensemble des données b) porte-conteneurs, c) citerne et d) vraquier.

Afin d'optimiser le modèle, les erreurs ont été modélisées par une approche de régression pas-à-pas tel que présenté dans la section 4.1. Les variables explicatives utilisées sont les mêmes que dans la section précédente. Le modèle de départ s'exprime ainsi :

$$E_{E1} = \beta_0 + \varepsilon \quad (52)$$

où β_0 est l'ordonnée à l'origine et ε représente les résidus qui sont normalement distribués avec une moyenne de 0 et une variance de σ^2 . Selon l'ensemble de données, les 17 variables explicatives ont été ajoutées ou retirées du modèle. Le modèle complet a la forme suivante:

$$E_{E1} = \beta_0 + \beta_1 F_{nh} + \dots + \beta_{17} st + \varepsilon \quad (53)$$

où $\beta_i (i=1, \dots, 17)$ représente les coefficients de chaque variables explicative. Pour obtenir une estimation du surenfoncement, le modèle d'erreurs et le modèle ***Eryuzlu 1*** sont joints :

$$S = E_{E1} + 0.181 \cdot \sqrt{T} \cdot b \cdot \left(\frac{V}{\sqrt{g \cdot T}} \right)^{2.269} \cdot \left(\frac{T}{h} \right)^{0.994} \quad (54)$$

Dans le reste du document, ce modèle sera noté ***Eryuzlu OPT***.

Le tableau 11 présente les coefficients des différents modèles de régression pas-à-pas ajustés sur les erreurs du modèle ***Eryuzlu 1***. Plusieurs variables explicatives ont été introduites dans le modèle. Il semble que le modèle ***Eryuzlu 1*** est incomplet puisque la plupart des variables explicatives qui sont déjà intégrées dans le modèle, ont été jugées significatives. Encore une fois, on peut remarquer que les variables introduites ne sont pas les mêmes selon que le modèle est ajusté sur l'ensemble des données ou sur les données de type porte-conteneurs, citerne ou vraquier.

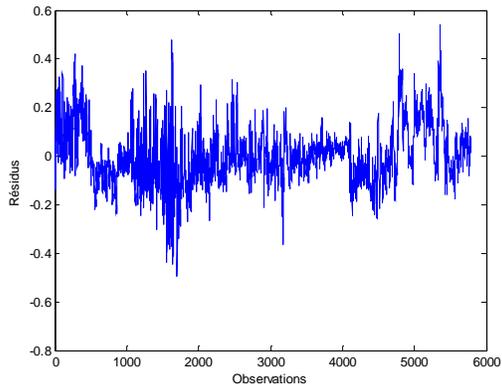
Tableau 11: Coefficients des différents modèles de régression pas-à-pas ajustés sur les erreurs du modèle *Eryuzlu 1*.

Variable	Coefficients de régression			
	Ensemble des données	Porte-conteneurs	Citerne	Vraquier
β_o	2.7907	1.5143	7.7430	-10.5487
F_{nh}	0.9934	-3.5966	-0.4332	0.4941
C_b	-0.9074	0.0000	-10.0913	0.0000
S_2	-1.0178	0.0000	-2.7606	-0.2990
V	-0.0544	0.3238	-0.0282	0.0000
T	-0.1451	-0.1121	0.0000	-1.3506
b	-0.0809	-0.0170	0.0000	-0.4912
h	0.0000	0.2186	0.0000	0.0000
B	0.0000	0.0000	0.0000	0.0000
L_{pp}	0.0248	-0.0399	0.0000	0.0000
L	-0.0223	0.0401	0.0000	0.1824
A_s	0.0043	0.0000	0.0000	0.0000
A_c	0.0000	0.0000	0.0015	0.0000
A_w	0.0000	-0.0003	-0.0015	0.0000
PAR_h/T	-0.0790	-0.9393	-0.3530	0.0613
V_{cr}	0.0000	0.5155	0.0000	0.0000
h/T	0.0000	-3.7300	0.0000	0.0000
st	0.0359	0.0000	0.0000	0.0000

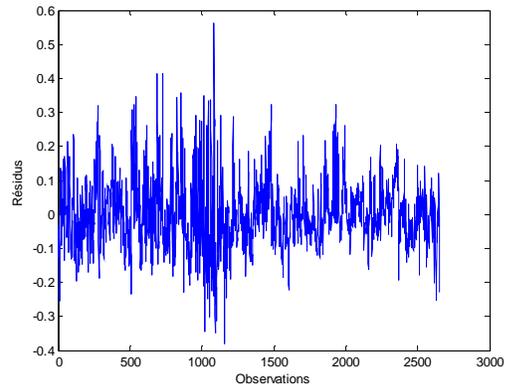
Même si ce sont les erreurs qui sont modélisées, les hypothèses de base classiques (normalité, homoscedasticité et indépendance) doivent être respectées. Pour vérifier les hypothèses de base, des graphiques des résidus en fonction des observations (indépendance), des résidus en fonction des valeurs prédites (homoscedasticité) et des droites de Henry (normalité) ont été produits et sont représentés dans les figures 10, 11 et 12. Les hypothèses de base ne sont pas toujours respectées.

Le tableau 12 présente les différents critères de performance du modèle *Eryuzlu OPT*. Les résultats obtenus par arbre de régression et les résultats des différents modèles empiriques ont également été ajoutés au tableau à titre comparatif.

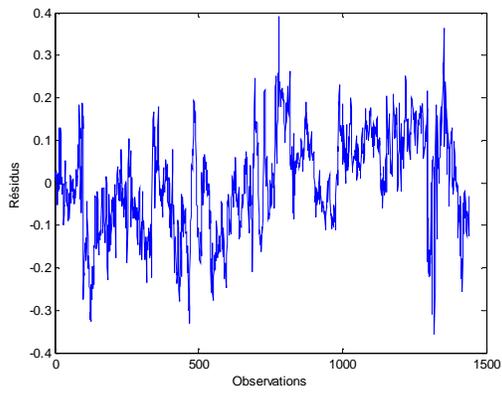
a)



b)



c)



d)

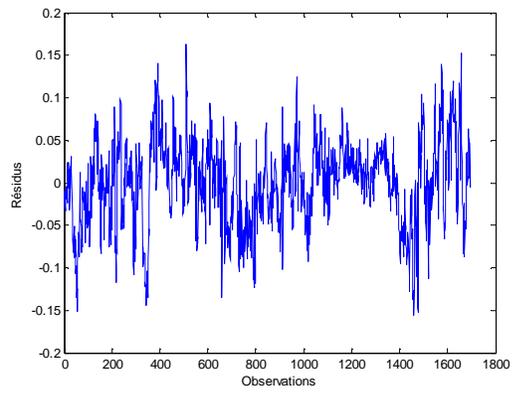
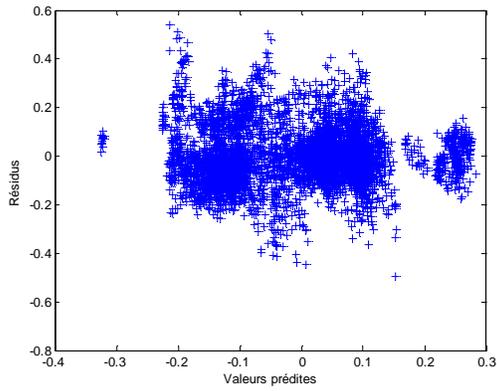
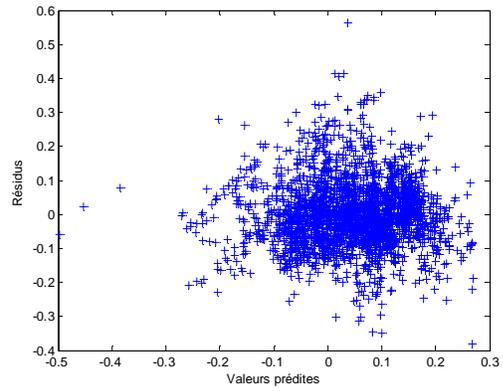


Figure 11 : Résidus en fonction des observations pour le modèle de régression ajusté sur les erreurs du modèle *Eryuzlu 1* pour a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

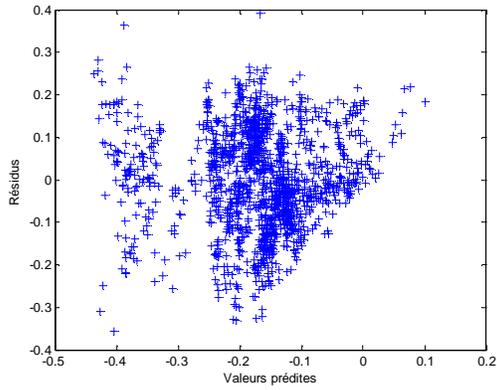
a)



b)



c)



d)

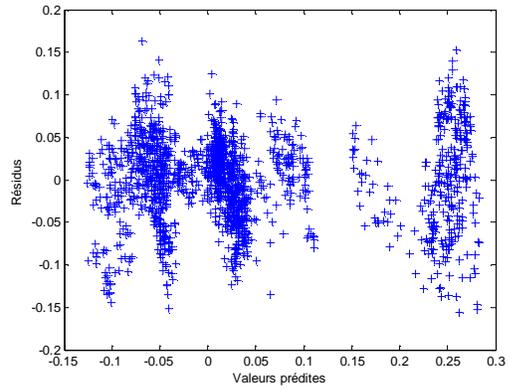
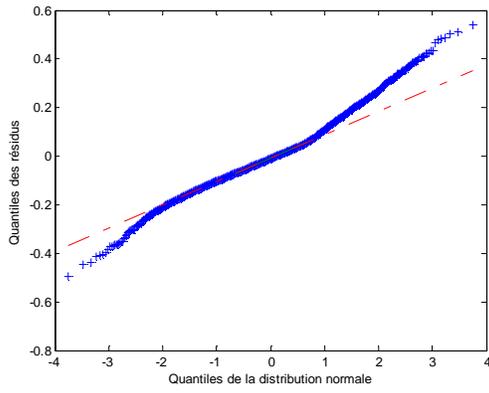
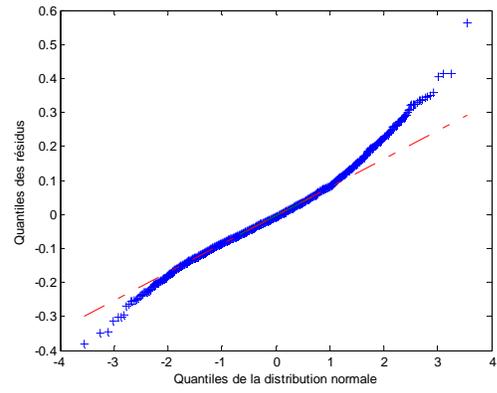


Figure 12 : Résidus en fonction des valeurs prédites pour le modèle de régression ajusté sur les erreurs du modèle *Eryuzlu 1* pour a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

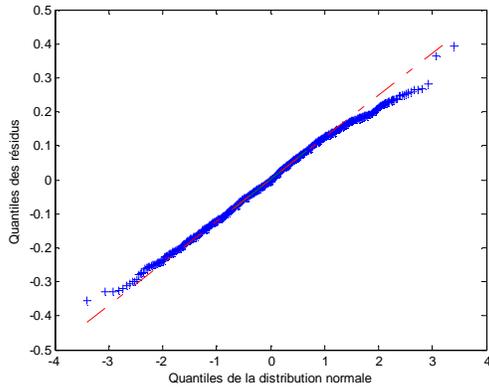
a)



b)



c)



d)

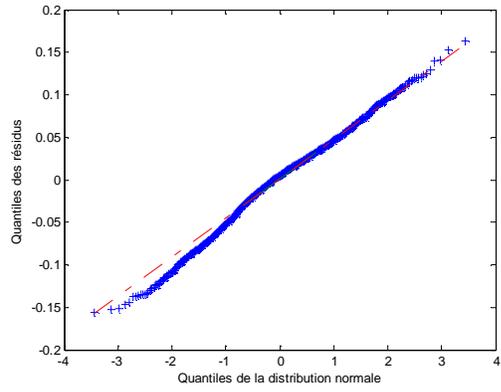


Figure 13 : Droite de Henry des résidus pour le modèle de régression ajusté sur les erreurs du modèle *Eryuzlu 1* pour a) l'ensemble des données, b) porte-conteneurs, c) citerne et d) vraquier.

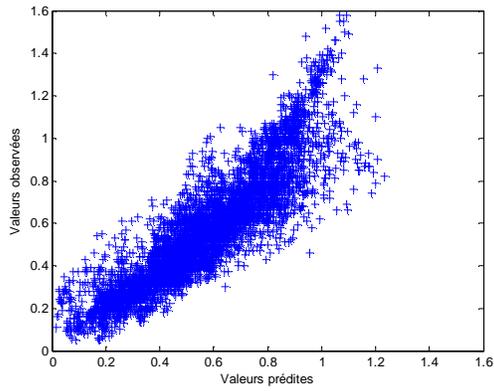
Tableau 12 : Performance du modèle *Eryuzlu OPT*, du modèle *RegT* et des modèles empiriques d'estimation du surenforcement.

Modèle	Critères de performance											
	Tous			Porte-conteneurs			Citerne			Vraquier		
	R ²	EQM	BMR	R ²	EQM	BMR	R ²	EQM	BMR	R ²	EQM	BMR
<i>RegT</i>	0.91	0.01	0.9	0.84	0.01	1.0	0.89	0.01	0.8	0.89	0.00	0.7
<i>Eryuzlu OPT</i>	0.80	0.01	3.9	0.78	0.01	3.7	0.83	0.01	1.8	0.83	0.00	2.4
<i>Eryuzlu1*</i>	0.62	0.03	3.1	0.65	0.02	9.0	0.77	0.05	-20.6	0.51	0.02	13.8
<i>Barrass1*</i>	0.54	0.03	18.0	0.66	0.02	6.0	0.80	0.03	-12.7	0.51	0.06	63.0
<i>Eryuzlu2*</i>	0.56	0.03	-1.7	0.62	0.02	0.3	0.80	0.07	-28.0	0.45	0.02	17.5
<i>Eryuzlu3*</i>	0.59	0.03	0.2	0.64	0.02	2.9	0.81	0.06	-25.8	0.48	0.02	18.2
<i>Japanese*</i>	0.66	0.03	19.3	0.56	0.02	-0.4	0.80	0.02	14.0	0.59	0.04	54.6
<i>Millward2*</i>	0.45	0.09	51.4	0.63	0.04	23.4	0.82	0.02	13.1	0.46	0.21	127.8
<i>Norrbin*</i>	0.67	0.02	6.7	0.57	0.02	-8.7	0.81	0.02	-3.1	0.61	0.02	39.2
<i>Romisch avant*</i>	0.61	0.04	7.5	0.51	0.04	-18.7	0.66	0.07	22.0	0.72	0.02	36.1
<i>Romisch arrière*</i>	0.53	0.04	-13.9	0.57	0.03	-10.7	0.63	0.11	-36.1	0.52	0.01	0.0
<i>Simard*</i>	0.59	0.03	0.7	0.64	0.02	1.0	0.79	0.07	-24.0	0.48	0.02	21.3

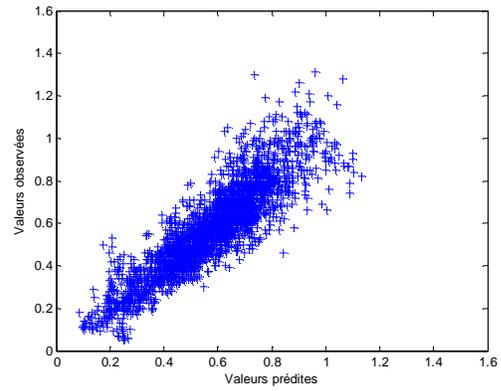
*=Résultats tirés de Gharbi et Le Page (2007)

Ces résultats montrent que la modélisation des erreurs est une autre approche prometteuse pour optimiser les modèles d'estimation du surenforcement. En effet, le coefficient de détermination a augmenté tandis que l'erreur quadratique moyenne et le biais moyen relatif ont diminué par rapport au modèle *Eryuzlu 1*. La figure 14 présente les valeurs observées et les valeurs prédites par le modèle *Eryuzlu OPT*. Les grandes valeurs de surenforcement semblent encore problématiques. Néanmoins, on remarque une amélioration par rapport au modèle *Eryuzlu 1* puisque les grandes valeurs sont mieux représentées. Par ailleurs, les modèles *Eryuzlu OPT* ne respectent pas les hypothèses sur lesquelles ils sont basés. Leur utilisation comme prédicteurs pourrait être controversée.

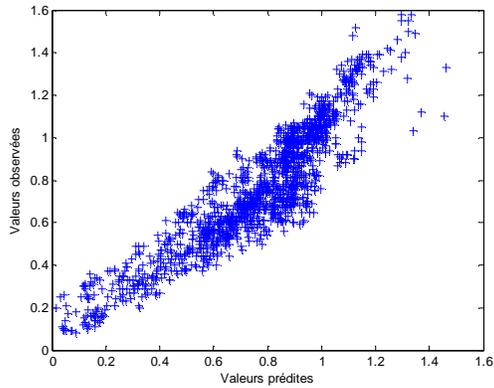
a)



b)



c)



d)

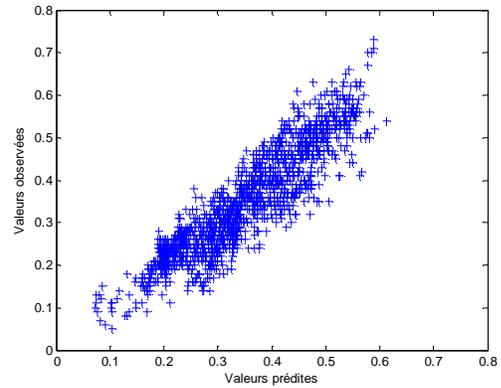


Figure 14 : Valeurs observées et valeurs prédites obtenues par le modèle *Eryuzlu OPT* ajusté sur a) l'ensemble des données b) porte-conteneurs, c) citerne et d) vraquier.

6. Discussion et conclusion

Ce travail a été effectué en suivant les recommandations de Gharbi et Le Page (2007) qui étaient soit d'optimiser les modèles de surenfacement existants ou de développer de nouveaux modèles plus performants. Des analyses préliminaires ont d'abord été réalisées afin de s'appropriier le signal dans les données (section 4). De plus, des nombres de Froude selon différents paramètres pour représenter la composante linéaire caractéristique ont été calculés dans le but de choisir celui pour lequel la corrélation avec le logarithme du surenfacement est la plus élevée. Par la suite, des approches régressives classiques ont été appliquées comme première étape vers la modélisation du surenfacement. À chaque fois, quatre modèles ont été ajustés. Le premier a été ajusté avec toute la base de données avec une variable indicatrice pour le type de navire. Les autres modèles ont été ajustés avec les données recueillies avec des navires de type porte-conteneur, citerne et vraquier respectivement. Deux approches de régression ont été utilisées dans le but de sélectionner les variables explicatives du surenfacement : la régression pas-à-pas et la régression ridge. Les approches pas-à-pas et régression ridge ne se sont pas montrées satisfaisantes puisque les hypothèses de base sur lesquelles reposent les modèles ajustés ne sont pas respectées. Les résultats ont été présentés à titre exploratoire uniquement.

Pour modéliser le surenfacement, l'approche par arbre de régression a été adoptée. L'avantage de cette technique est qu'elle est non paramétrique, donc les données n'ont pas à suivre de distribution particulière. De plus, elle n'assume aucune forme dans le lien entre la variable dépendante et les variables explicatives. Enfin, il n'y a pas d'hypothèse contraignante à faire comme avec les approches régressives classiques. Les variables explicatives à inclure dans le modèle ont été sélectionnées selon le jugement de l'expert Samir Gharbi. L'ajustement du modèle a été vérifié avec trois critères : le coefficient de détermination, l'erreur quadratique moyenne et le biais moyen relatif. Ces trois critères ont été choisis pour pouvoir comparer les résultats avec ceux obtenus de l'ajustement de quelques modèles empiriques jugés performants dans le travail de Gharbi et Le Page (2007). En général, la performance des modèles construits par arbre de régression a surpassé celle des différents

modèles empiriques. Enfin, des estimations plus honnêtes des critères de performance ont été obtenues par une validation croisée de type *'leave-one-out'*.

La modélisation des erreurs du modèle ***Eryuzlu 1*** est une approche prometteuse pour améliorer les prédictions du surenfoncement des navires. Le modèle ***Eryuzlu 1*** est parmi les plus performants de tous les modèles empiriques comparés dans Gharbi et Le Page (2007) et il a été adopté par la GCC comme outil d'estimation du surenfoncement des navires. Pour améliorer sa performance, les erreurs du modèle ont été modélisées par une régression pas-à-pas. En résultat, les critères de performance ont tous été améliorés. Cependant, les hypothèses de base n'étaient pas toujours respectées. La modélisation pourrait être améliorée encore. Le modèle obtenu par arbre de régression demeure le plus performant et le plus recommandable.

Les modèles ***RegT*** sont valides pour prédire le surenfoncement des navires marchands qui transitent sur le fleuve St-Laurent ou sur des voies navigables similaires. Le domaine de validité est moins restrictif que la plupart des modèles empiriques qui sont également basés sur des données recueillies lors de campagnes de terrain ou encore sur des tests en laboratoire. Pour étendre le domaine de validité à d'autres types de navires ou d'autres voies navigables, d'autres études sont nécessaires.

En pratique, les modèles ***RegT*** sont très faciles à utiliser. Une fois programmé, l'utilisateur n'a qu'à entrer les valeurs des variables explicatives et le surenfoncement prédit s'obtient instantanément. Ces modèles ont été construits avec les mêmes variables explicatives que dans les différents modèles empiriques, donc ce n'est pas plus compliqué à utiliser. Ces modèles, ayant été validés avec des données prises par GPS-OTF, pourraient être intégrés dans un système de prédiction de réponse des navires. Le nombre de ports qui implémente des systèmes de prédiction, ainsi que le nombre de navires qui sont dotés de système de prédiction de réponse du navire tendent à augmenter (Parker et Huff, 1998). Ces systèmes, combinés à des cartes marines et des niveaux d'eau précis, et à des formules de prédiction du tirant d'eau dynamique spécifiques au type de navire et de chenal, fournissent au navigateur toute l'information nécessaire pour gérer le dégagement sous quille de manière efficace et sécuritaire. Avec ces systèmes, le navigateur peut planifier son voyage tout en s'assurant qu'il

y a toujours une marge de sécurité/manoeuvrabilité suffisante dans des sections de chenal peu profondes. Il pourra, par exemple, réduire la vitesse du navire dans une section critique pour réduire le surenfoncement et éviter de toucher le fond du chenal.

Dans les travaux futurs, la GCC planifie le développement d'un outil permettant de gérer le dégagement sous quille du trafic maritime sur le fleuve Saint-Laurent. Les modèles de prédiction du surenfoncement par arbre de régression pourraient être implantés dans ce type de système.

Références

Barrass C.B. (1979) A unified approach to squat calculations for ships. *PLANC BULLETIN*, 32, 3-10.

Barrass C.B. (1981) Ship squat – A reply. *The Naval Architect*, Vol. November, 24.

Barrass C.B. (2004) Thirty-two years of research into ship squat. Dans: Proceedings of the Squat Workshop 2004, 3-4 mars, Oldenbourg, Allemagne.

Breiman L., Friedman J.H., Olshen R., Stone C.J. (1984) Classification and Regression Trees, Wadsworth, Belmont, CA, 358 p.

Briggs M.J. (2006) Ship squat predictions for ship/tow simulator. Coastal and Hydraulics Engineering Technical Note CHETN-I-72, U.S. Army Engineer Research and Development Center, Vicksburg, MS. <http://chl.erdc.usace.army.mil.chetn/>

Briggs M.J., Maynard S.T., Nickles C.R., Walker T.N. (2004) Charleston harbour ship motion data collection and squat analysis. Coastal and Hydraulics Engineering Technical Note CHETN-IX-14, U.S. Army Engineer Research and Development Center, Vicksburg, MS. <http://chl.erdc.usace.army.mil.chetn/>

Dick R.A., Prior A., Lantos S. (1991). Criteria for commercial navigation in inland waterways. Transport Canada Publication No TP 10911, Melville Shipping Ltd., Ottawa, ON, 118 p.

Eryuzlu N., Hausser R. (1978) Experimental investigation into some aspects of large vessel navigation in restricted waterways. Dans: Proceedings of Symposium of Aspects of Navigability of Constraint Waterways Including Harbour Entrances, Delft, Hollande, Vol. 2, 1-15.

Eryuzlu N., D'Agnolo F. (1991) Calculation of squat of large vessels in shallow waterways. Rapport interne TP 11085E, Développement des voies navigables, Services d'aide à la navigation maritime, Garde Côtière Canadienne, Pêches et Océans Canada, 8 p.

Eryuzlu N., Cao Y., D'Agnolo F. (1994) Underkeel requirements for large vessels in shallow waterways, Dans: 28th International Navigation Congress, PIANC, Séville, Espagne, S II-2, 17-25.

Gharbi S., Joseph G., Le Page, A. (2007). Étude du squat des navires marchands transitant par la voie navigable du Saint-Laurent. Volume I : Analyse des résultats de la campagne de mesure du squat. Rapport de recherche, Garde Côtière Canadienne, Pêches et Océans Canada, 185 p.

Gharbi, S., and Le Page, A. (2007). Étude du squat des navires marchands. Volume II : Analyse des modèles de prédiction du squat. Rapport de recherche, Garde Côtière Canadienne, Pêches et Océans Canada, 50 p.

Gharbi, S. (2007). Etude du squat des navires marchands transitant par la voie navigable du fleuve Saint-Laurent. Résumé du rapport principal, Garde Côtière Canadienne, Pêches et Océans Canada, 25 p.

Gourlay T. (2000) Mathematical and computational techniques for predicting the squat of ships. Thèse de doctorat, Université d'Adelaïde, Adelaïde, Australie, 139 p.

Gourlay T. (2006) A simple method for predicting the maximum squat of a high-speed displacement ship. *Marine technology*, 43 (3), 146-151.

Herreros Sierra M. H., Zamora Rodriguez R., Pérez Rojas L. (2000). El fenómeno squat en áreas de profundidad variable y limitada. Proc., Second International Conference on Oceanic Engineering. Ocean 2000. Universidad Austral de Chile, Valdivia, Chili.

Hooft J.P. (1974) The behaviour of a ship in head waves as restricted water depth. *International Shipbuilding Progress*, 244 (21), 367.

Hurvich C.M., Tsai C.-L. (1990). The impact of model selection on inference in linear regression. *The American Statistician*, 44 (3), 214-217.

Huuska O. (1976) On the evaluation of underkeel clearances in Finnish waterways. Report No. 9. Otaniemi: University of technology, Ship Hydrodynamics Laboratory.

ICORELS (International Commission for the Reception of Large Ships). (1980) Report of Working group IV, PIANC Bulletin No. 35, Supplement.

MacPherson D.M. (2002) Squat effects: A practical guide to its nature, measurement and prediction. HydroComp Inc. Presented to the Society of Naval Architects and Naval Engineer (SNAME), New England Section.

Marquardt D.W., Snee R.D. (1975). Ridge regression in practice. *The American Statistician*, 29 (1), 3-20.

Millward A. (1990) A preliminary design method for the prediction of squat in shallow water. *Marine Technology*, 27(1), 10-19.

Millward A. (1992) A comparison of the theoretical and empirical prediction of squat in shallow water. *International Shipbuilding Progress*, 39(417), 69-78.

Morse B., Simard L. (1991) La détermination du dégagement sous quille. Rapport de recherche, Développement des voies navigables, Services d'aide à la navigation maritime, Garde Côtière Canadienne, Pêches et Océans Canada, 10 p.

Morse B., Michaud S., Taschereau A. Santerre R. (2004). Ship squat in shallow and confined channels – The Canadian experience. Dans: Proceedings of the Squat-Workshop 2004, 3-4 mars, Oldenbourg, Allemagne.

Myers R.H. (1986) Classical and modern regression with applications, Duxbury Press, Belmont CA, 488 p.

Norrbin N.H. (1986) Fairway design with respect to ship dynamics and operational requirements. SSPA Research Report No. 102. SSPA Maritime Consulting, Goteborg, Suède.

Ondo et al. (2001). Procédures Bayésiennes pour la détection d'observations singulières : synthèse bibliographique, *Journal de la Société Française de Statistique*, 142 (2), 41-64.

Pagé Y., Stocks D.T. (2002) Feasibility study. Squat measurement leading to examination of under keel clearance in the St. Lawrence River. Interim Report 1. Géolocation Inc, Québec, Canada and Fleet technology Limited, Kanata, Canada, 16 p.

Parker B.B., Huff L.C. (1998) Modern under-keel clearance management. *International Hydrographic Review*, 75 (2), 143-165.

Rencher A.C., Pun F.C. (1980). Inflation of R^2 in best subset regression. *Technometrics*, 22 (1), 49-53.

Romisch K. (1989) Empfehlungen zur Bemessung Von Hafeneinfahrten. Wasserbauliche Mitteilungen der Technischen, Universitat Dresden, Helft, Allemagne, 1, 39-63.

Simard L. (1969) Engineering Field Investigations: Squat Study. Rapport interne, Division du chenal maritime du Saint-Laurent, Ministère des Transports, 11 p.

Simard L. (1982) Squat of Ships in Navigation Channels. Thèse de doctorat, Columbia Pacific University, Mill Valley, 268 p.

Stocks D.T., Pagé Y. (2002) Feasibility study squat measurement leading to examination of under keel clearance in the St-Lawrence River. Pêches et Océans Canada, Garde Côtière Canadienne, Direction des Services Techniques Intégrés, Service du génie hydraulique, Région de Québec.

Stocks D.T., Dagget L.L., Pagé Y. (2002) Maximization of ship draft in the St Lawrence seaway volume 1: squat study. Rapport no TP 13888, Centre de Développement des Transports, Transports Canada, 79 p.

Tuck E.O. (1966) Shallow water flow past slender bodies. *Journal of fluid mechanics*, 2 (Part 1): 81-95.

Varyani K.S. (2006) Squat effects on high speed craft in restricted waterways. *Ocean Engineering*, 33, 365-381.

Vinod H.D. (1976) Application of new ridge regression methods to a study of bell system scale economies. *Journal of the American Statistical Association*, 71, 835-841.

Annexe A : Statistiques descriptives des variables susceptibles d'influencer le surenfoncement et du surenfoncement.

Tableau A1 : Statistiques descriptives des différentes variables explicatives et du surenfoncement maximum.

Variable	Statistiques								
	Moyenne	Écart-type	Médiane	Mode	Minimum	Maximum	Coef. Variation	Coef. Asymétrie	Coef. Aplatissement
Fnh	0,44	0,10	0,46	0,49	0,08	0,64	0,22	-0,68	3,22
Cb	0,75	0,09	0,74	0,74	0,60	0,90	0,12	0,29	2,05
V	5,30	1,09	5,50	3,11	1,05	7,83	0,21	-0,60	3,38
T	9,96	0,98	9,60	9,60	8,60	15,55	0,10	2,71	15,92
b	30,08	4,05	32,20	32,20	23,70	47,85	0,13	0,52	6,71
h	15,24	3,91	13,95	13,54	11,56	54,43	0,26	3,67	22,05
hz	14,49	3,37	13,49	14,29	11,00	50,00	0,23	4,00	27,10
W	470,06	326,81	355,00	245,00	235,00	3300,00	0,70	3,65	22,32
B	300,76	40,52	322,00	322,00	237,00	478,50	0,13	0,52	6,71
Lpp	205,05	39,01	191,00	173,90	164,70	282,00	0,19	0,96	2,61
L	215,09	39,86	200,00	183,00	177,00	294,00	0,19	1,00	2,61
As	301,12	67,40	308,10	309,10	203,80	744,10	0,22	3,98	27,83
Ac	4411,83	1774,68	4013,70	4034,90	2804,80	26042,40	0,40	4,71	36,82
Aw	4110,70	1726,53	3713,00	2845,30	2499,30	25298,30	0,42	4,69	36,68
par_ht	0,90	0,09	0,91	1,00	0,56	1,00	0,10	-0,84	3,11
Vcr_cr	8,33	1,61	7,88	6,21	6,06	20,86	0,19	2,71	14,15
S	0,55	0,26	0,53	0,60	0,05	1,58	0,47	0,65	3,17

Annexe B : Étude des corrélations entre les variables susceptibles d'influencer le surenfacement et du surenfacement.

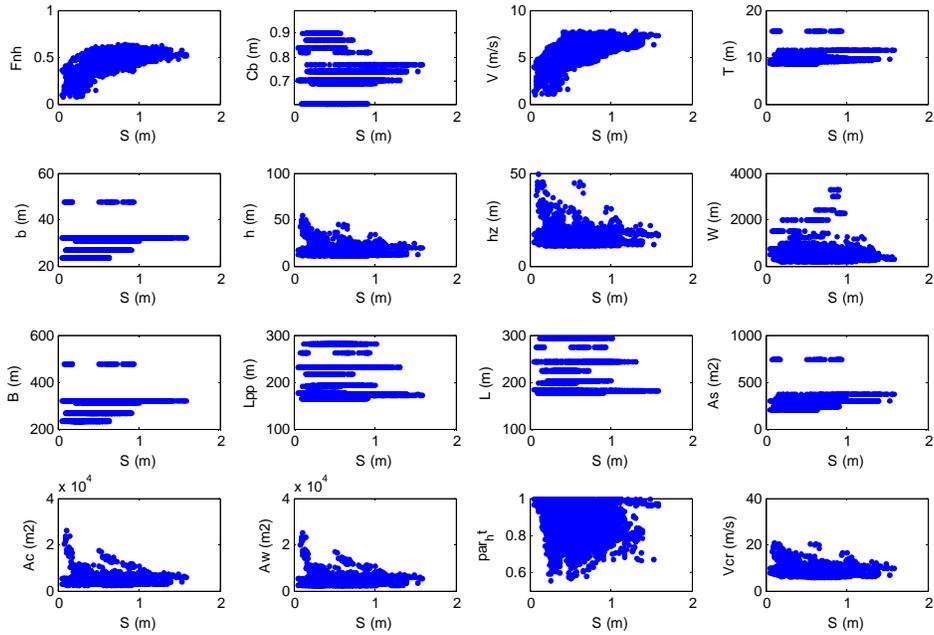


Figure B1: Nuages de points bivariés des différentes variables avec le surenfacement maximum.

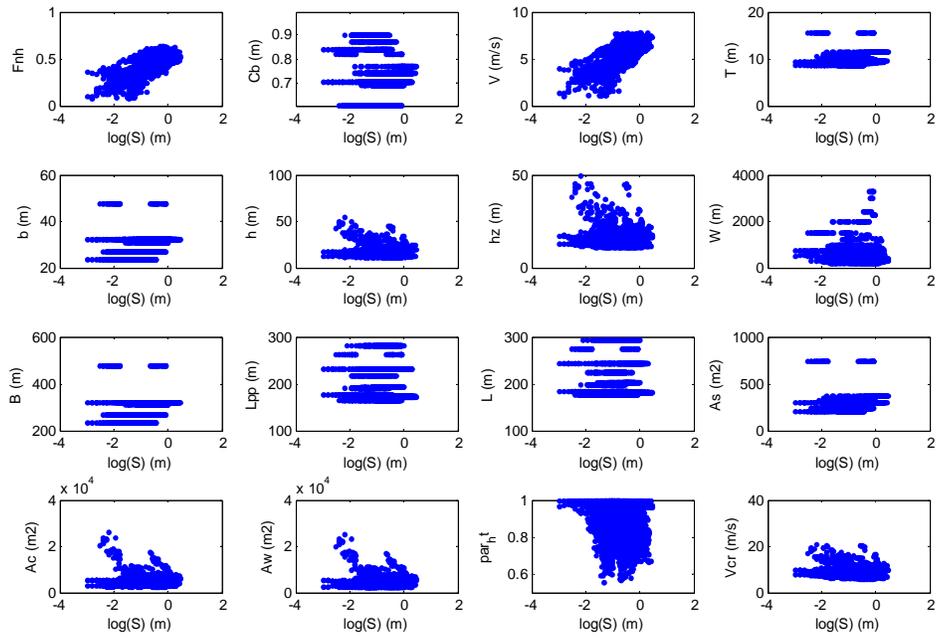


Figure B2: Nuages de points bivariés des différentes variables avec le logarithme du surenfoncement maximum.

Tableau B1 : Corrélations du surenforcement et des différentes variables susceptibles d'influencer le surenforcement pour l'ensemble des données.

	Fnh	Cb	V	T	b	h	hz	W	B	Lpp	L	As	Ac	Aw	par_ht	Vcr_cr	S
Fnh	1,00	-0,44	0,91	-0,48	-0,02	-0,36	-0,35	-0,27	-0,02	-0,12	-0,11	-0,25	-0,33	-0,33	-0,32	-0,32	0,70
Cb		1,00	-0,39	0,53	-0,30	0,16	0,09	0,11	-0,30	-0,04	-0,07	0,06	0,05	0,05	-0,05	0,10	-0,37
V			1,00	-0,31	0,07	0,02	0,02	-0,04	0,07	-0,11	-0,10	-0,09	-0,01	-0,01	-0,04	0,07	0,70
T				1,00	0,39	0,53	0,40	0,45	0,39	0,21	0,19	0,78	0,60	0,58	0,01	0,35	-0,15
b					1,00	0,31	0,25	0,31	1,00	0,44	0,43	0,87	0,57	0,55	0,05	0,18	0,39
h						1,00	0,97	0,60	0,31	0,06	0,06	0,50	0,90	0,91	0,56	0,96	-0,12
hz							1,00	0,55	0,25	0,01	0,01	0,40	0,86	0,87	0,57	0,96	-0,12
W								1,00	0,31	0,03	0,03	0,46	0,67	0,67	0,28	0,60	-0,09
B									1,00	0,44	0,43	0,87	0,57	0,55	0,05	0,18	0,39
Lpp										1,00	1,00	0,38	0,18	0,17	0,04	0,00	-0,08
L											1,00	0,37	0,18	0,17	0,05	0,00	-0,07
As												1,00	0,72	0,71	0,07	0,33	0,17
Ac													1,00	1,00	0,44	0,86	-0,03
Aw														1,00	0,45	0,87	-0,04
par_ht															1,00	0,64	-0,22
Vcr_cr																1,00	-0,15
S																	1,00

Tableau B2 : Corrélations du logarithme du surenforcement et des différentes variables susceptibles d'influencer le surenforcement pour l'ensemble des données.

	Fnh	Cb	V	T	b	h	hz	W	B	Lpp	L	As	Ac	Aw	par_ht	Vcr_cr	ln(S)
Fnh	1,00	-0,44	0,91	-0,48	-0,02	-0,36	-0,35	-0,27	-0,02	-0,12	-0,11	-0,25	-0,33	-0,33	-0,32	-0,32	0,76
Cb		1,00	-0,39	0,53	-0,30	0,16	0,09	0,11	-0,30	-0,04	-0,07	0,06	0,05	0,05	-0,05	0,10	-0,41
V			1,00	-0,31	0,07	0,02	0,02	-0,04	0,07	-0,11	-0,10	-0,09	-0,01	-0,01	-0,04	0,07	0,75
T				1,00	0,39	0,53	0,40	0,45	0,39	0,21	0,19	0,78	0,60	0,58	0,01	0,35	-0,19
b					1,00	0,31	0,25	0,31	1,00	0,44	0,43	0,87	0,57	0,55	0,05	0,18	0,39
h						1,00	0,97	0,60	0,31	0,06	0,06	0,50	0,90	0,91	0,56	0,96	-0,17
hz							1,00	0,55	0,25	0,01	0,01	0,40	0,86	0,87	0,57	0,96	-0,18
W								1,00	0,31	0,03	0,03	0,46	0,67	0,67	0,28	0,60	-0,11
B									1,00	0,44	0,43	0,87	0,57	0,55	0,05	0,18	0,39
Lpp										1,00	1,00	0,38	0,18	0,17	0,04	0,00	-0,03
L											1,00	0,37	0,18	0,17	0,05	0,00	-0,03
As												1,00	0,72	0,71	0,07	0,33	0,14
Ac													1,00	1,00	0,44	0,86	-0,08
Aw														1,00	0,45	0,87	-0,09
par_ht															1,00	0,64	-0,24
Vcr_cr																1,00	-0,19
ln(S)																	1,00

Tableau B3 : Corrélations du logarithme du surenforcement et du nombre de Froude en faisant varier la dimension linéaire caractéristique du nombre de Froude.

Dimension linéaire caractéristique	h	b	T	Lpp	h-T	W-b
Corrélation du Fnh avec log(S)	0,76	0,59	0,72	0,68	0,61	0,60

Annexe C : Arbres de classification et de régression

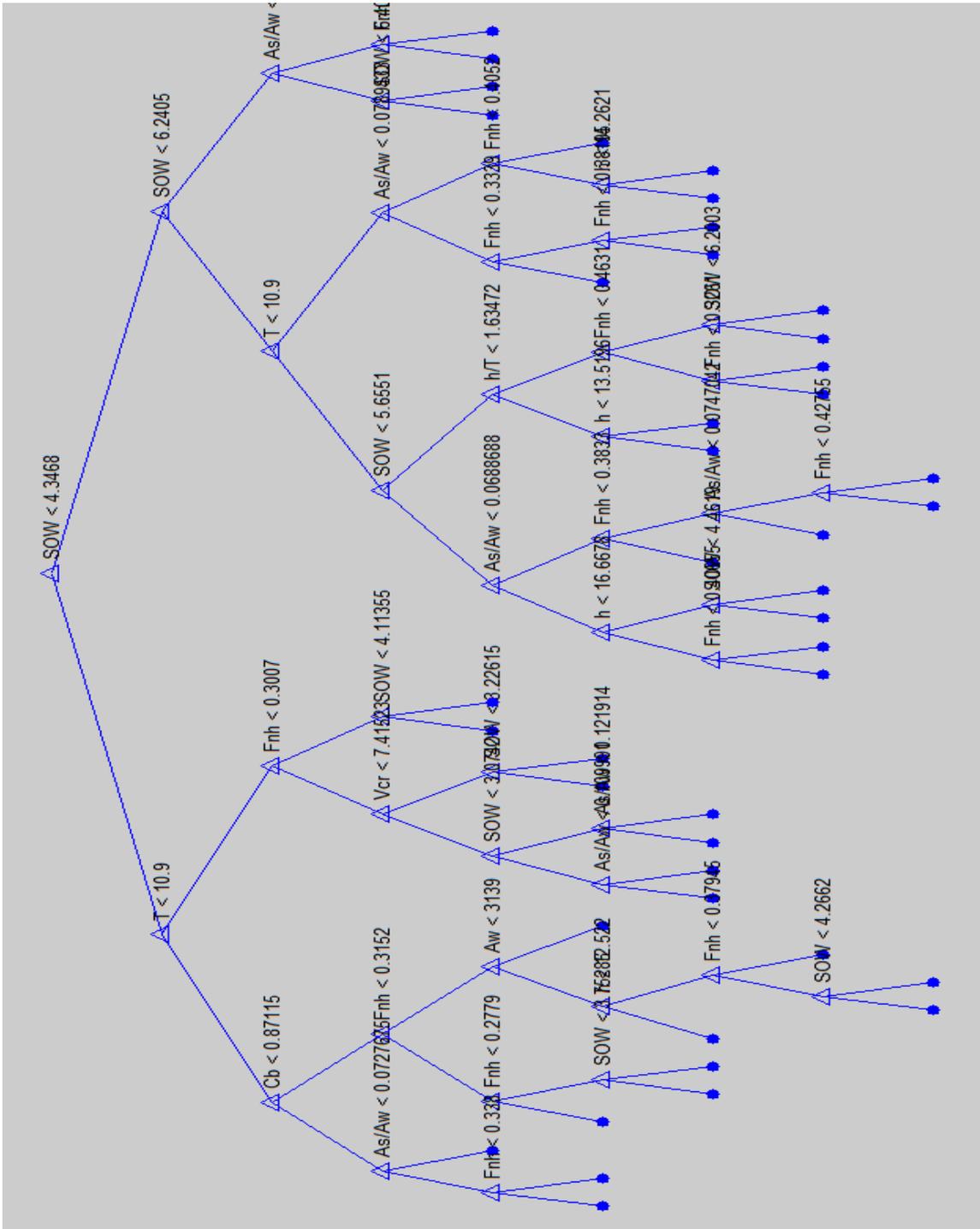


Figure C1 : Arbre de régression ajusté pour le modèle retenu sur toutes les données.

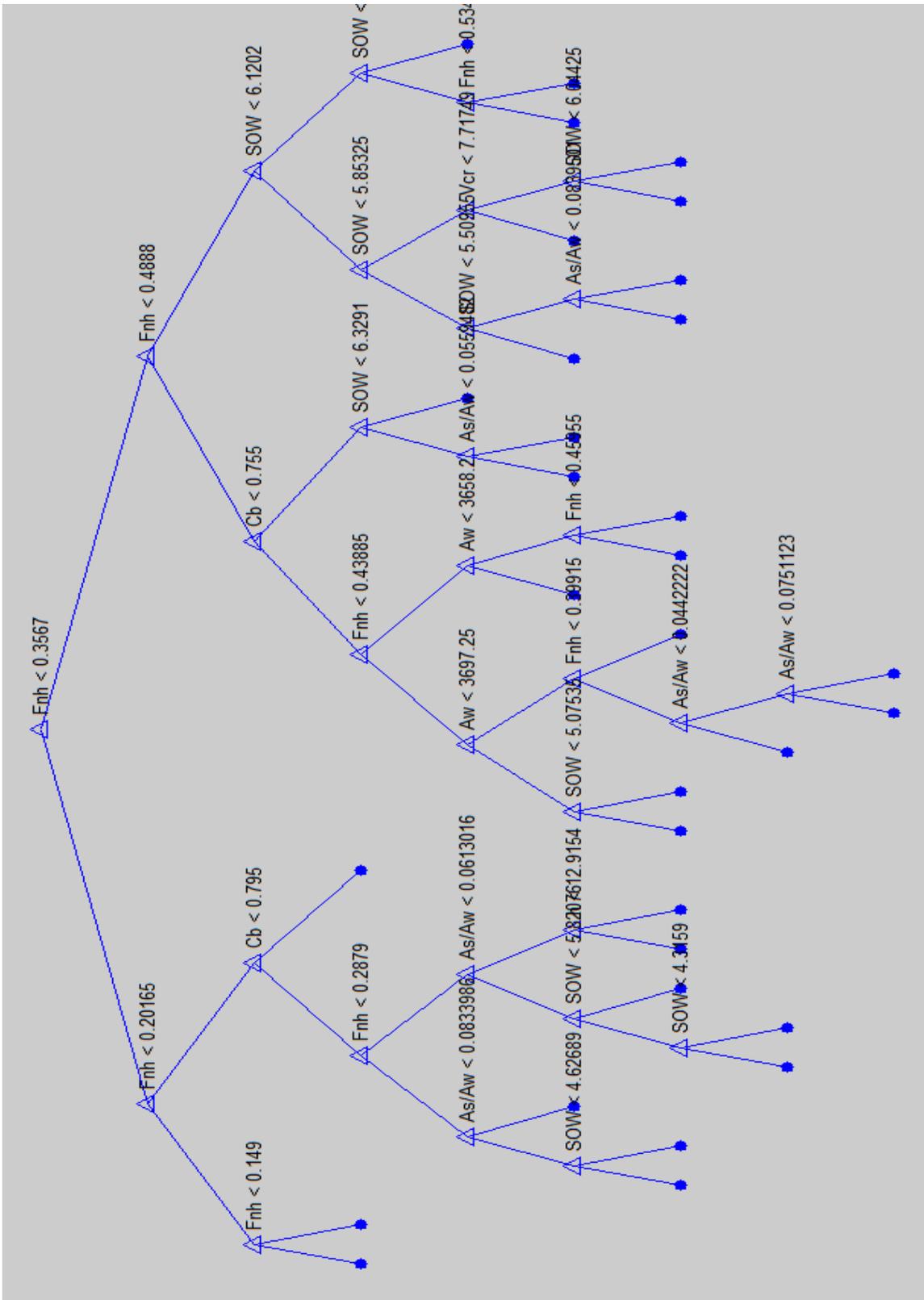


Figure C3 : Arbre de régression ajusté pour le modèle retenu sur les données des navires de type citerne.

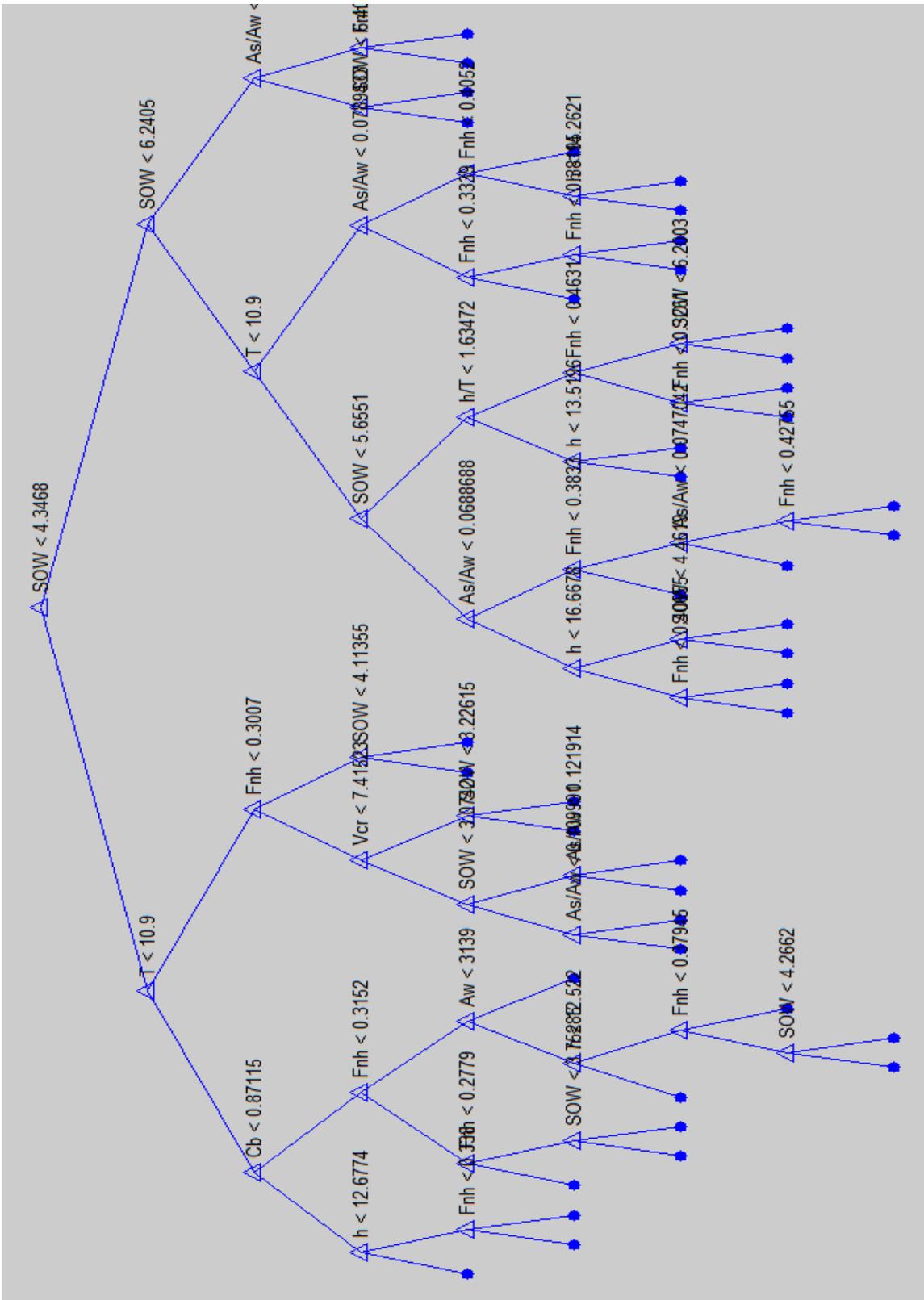


Figure C4 : Arbre de régression ajusté pour le modèle retenu sur les données des navires de type vraquier.