



Université du Québec
à Chicoutimi

8INF914 – Résolution de problèmes industriels

Projet de session

Partie I : Modélisation de l'élévation aval et des fonctions de production

Auteurs :

Arthur LEBARON

Maxime CHEVALIER

Encadrant :

Pr. Sara SÉGUIN

Version 1.0 du
14 février 2018

Table des matières

Introduction	1
Revue de littérature sur les méthodes de modélisation de fonctions	3
1.1 MatLab	3
1.2 Moindres carrés	3
1.3 Régression polynomiale	3
1.4 Corrélation des données	4
1.5 Données aberrantes	4
1.6 Bruitage des données	4
1.7 Validation par les moindres carrés	4
1.8 Least Absolute Residuals	4
Méthodologie retenue dans le cadre du projet	7
2.1 Fonction d'élévation	7
2.2 Fonctions de productions	8
Résultats	9
3.1 Fonction d'élévation	9
3.2 Fonctions de productions	10
3.2.1 Turbine 1	10
3.2.2 Turbine 2	11
3.2.3 Turbine 3	12
3.2.4 Turbine 4	14
3.2.5 Turbine 5	15
Discussions	17
Conclusion	19

Table des figures

2.1	Boxplot de l'élévation en fonction de la quantité turbinée	7
3.2	Elevation aval en fonction du débit total à la centrale	9
3.3	Fonction de production de la turbine 1	10
3.4	Graphique des résiduels de la fonction de production 1	10
3.5	Fonction de production de la turbine 2	11
3.6	Graphique des résiduels de la fonction de production 2	12
3.7	Fonction de production de la turbine 3	12
3.8	Graphique des résiduels de la fonction de production 3	13
3.9	Fonction de production de la turbine 4	14
3.10	Graphique des résiduels de la fonction de production 4	14
3.11	Fonction de production de la turbine 5	15
3.12	Graphique des résiduels de la fonction de production 5	15

Introduction

Rio Tinto est l'entreprise minière et métallurgique la plus importante au Canada. Près de 15000 employés sont répartis sur les 35 sites du groupe réparti dans le pays. Cependant, la métallurgie (production d'aluminium et de fer notamment) est une activité très énergivore. Pour pallier à des dépenses énergétiques trop importantes, les compagnies privées et le gouvernement ont très tôt investi dans la production hydroélectrique, notamment au 20^{ème} siècle. Aujourd'hui, cette production représente 97% de l'électricité produite au Québec.

Avec le rachat d'Alcan, le groupe Rio Tinto a acquis différentes centrales hydroélectrique qu'il utilise pour alimenter ses centrales. Cependant, ce type de ressource nécessite une gestion particulière car tout dépend de l'eau disponible. Les ingénieurs font face aux aléas de la nature. En effet, ils sont dépendants des stocks d'eau et des précipitation pour gérer efficacement la production énergétique et donc le fonctionnement des leurs métallurgies. En hiver, les précipitations sont sous forme solide (c'est une réserve non disponible sur le moment) qui permet de remplir les réservoirs lors de la fonte des neiges. Le reste de l'année, ils sont dépendants de leurs réserves et de la pluie. Pour maximiser leurs réserves, le groupe a continué d'investir afin de contrôler toujours plus de vallées et ainsi récupérer l'écoulement d'eau.

Afin de construire leurs modèles de gestion de la production hydroélectrique, les ingénieurs disposent d'énormément de données. Un réseau de capteurs pluviométriques, ainsi que des capteurs de débits dans les rivières sont installés, leurs permettant de prévoir l'arrivée d'eau dans les réservoirs et barrages, de même que la température de l'air (acquise par satellite) et toutes les données des années précédentes. Grâce à cela ils ont pu construire un modèle permettant de gérer efficacement la production hydroélectrique.

Pascal Cote, ingénieur analyste en recherche opérationnelle chez Rio Tinto Alcan, nous a fournis dans le cadre de ce projet les données de la centrale des Chutes du Diables (construit en 1953) datant de 2013. Nous avons donc à disposition, et ce pour chaque heure les données suivantes :

En-tête fichier de données	Description
Elav	Élévation aval
Qtot	Débit total à la centrale
Qvan	Débit déversé à la centrale
Niv Amont	Élévation amont
Q1, Q2, Q3, Q4, Q5	Débit turbiné par chaque turbine
P1, P2, P3, P4, P5	Puissance produite par chaque turbine

La centrale des Chutes du Diable est composée de 5 turbines, ayant chacune une capacité de production différentes. Dans ce rapport, nous allons tenter de déterminer la fonction de production associée à chaque turbine de la centrale afin de créer un modèle, qui nous permettra par la suite de créer un outil de gestion de la centrale qui déterminera quelles turbines activer afin de maximiser la production d'électricité.

Revue de littérature sur les méthodes de modélisation de fonctions

1.1 MatLab

Afin de modéliser et traiter les données procurées par Pascal Cote, nous avons utilisé comme suggéré le logiciel Matlab, et principalement la boîte à outils "Curve Fitting". Cette fonctionnalité permet de facilement déterminer un modèle sur des données tout en ayant des informations sur la précision du modèle généré. Nous avons principalement utilisé la régression polynomiale pour créer nos modèles.

1.2 Moindres carrés

Cette méthode permet de faire une approximation d'un nuage de point par une fonction $f(x)$ qui décrit aux mieux les données. Celle-ci est la fonction qui minimise la somme des résidus au carré :

$$\text{Min } F = \sum_{i=1}^n (y_i - f(x_i))^2 = \sum_{i=1}^n r_i^2 \quad (1.1)$$

Avec :

r_i les écart en la fonction et nos mesures, aussi appelé résidu

y_i nos mesures

Lorsque nous exécutons cette méthode dans un logiciel comme matlab, celui-ci nous retourne les coefficients de la fonction (équation linéaire, polynôme ...) permettant d'obtenir la meilleur représentation analytique de nos données.

1.3 Régression polynomiale

Il existe différente méthode d'approximation, l'une d'elles se nomme l'interpolation polynomiale et consiste à déterminer un polynôme de degré n , si possible le plus bas, passant par tous les points. Il faut trouver un polynôme, à coefficients réels, de la forme : $p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x^1 + a_0$ Cette méthode permet de trouver une représentation analytique exacte de nos données. Cependant, ceci n'est pas réellement possible en pratique. Nous utilisons donc la régression polynomiale qui correspond au fait d'utiliser une méthode des moindres carrés pour faire l'interpolation polynomiale.

1.4 Corrélation des données

Afin de déterminer visuellement si les données sont corrélées, il est recommandé de les tracer l'une en fonction de l'autre. Si le graphique obtenu semble suivre un schéma alors il est probable que les données soient corrélées. Si les points du graphique ne suivent pas de schéma alors il est fortement probable que les données ne soient pas corrélées.

1.5 Données aberrantes

Il est très difficile de déterminer si une donnée est aberrante ou non. Lorsqu'une donnée est très éloignée des autres il est tentant de l'éliminer pour « améliorer » le modèle générer. Il faut tout d'abord s'assurer que c'est une donnée aberrante. Cette donnée a pu être générée par un capteur défectueux ou fait partie d'une spécificité du système qu'il est nécessaire de modéliser. On peut tenter de cibler des données dites aberrantes grâce au diagramme à moustache, ou via la méthode `isoutlier()` de Matlab.

1.6 Bruitage des données

Les données reçues peuvent également être bruitées, il est alors nécessaire de les lisser. Cette méthode est très utile lorsque les données sont des séries temporelles.

1.7 Validation par les moindres carrées

Afin de valider les modèles, il faut analyser le graphique des résiduels. Ce dernier permet de voir les erreurs associées. Il faut que les erreurs entre les valeurs prédites et mesurées soient indépendantes, que la moyenne soit égale à 0, que la variance soit constante et que les erreurs soient distribuées normalement. Si ces conditions ne sont pas remplies, c'est que la relation entre le modèle et les données n'est pas la bonne ou qu'une erreur a été faite lors de la modélisation. Sous Matlab et dans l'outil « curve fitting » il est simple d'avoir accès à ces données. Il faut que la répartition soit correcte, que le SSE et le RMSE soient proches de 0, et que le R-square soit proche de 1.

1.8 Least Absolute Residuals

"La méthode LAR trouve une courbe qui minimise la différence absolue des résidus, plutôt que les différences au carré. Par conséquent, les valeurs extrêmes ont une influence

moindre sur l'ajustement." *Documentation Matlab* Grâce à cette méthode, nous avons obtenu de meilleures valeurs pour les indices de confiance (SSE, R^2 , ...). Nous ne l'avons pas choisie seulement pour ça, mais pour le fait qu'elle atténue les valeurs importantes et permet donc de ne pas donner de poids aux valeurs qui pourraient être considérées comme aberrantes.

Méthodologie retenue dans le cadre du projet

2.1 Fonction d'élévation

Pour réaliser la partie I du projet de session, nous avons tout d'abord tracer le graphique de l'élévation aval en fonction du débit total à la centrale afin de pouvoir modéliser la fonction d'élévation. Comme les données sont liées (ce n'est pas un nuage de point aléatoire) nous avons décidé de tenter un pré traitement. Les données en ordonnées pouvant correspondre à une même abscisse nous avons tracé le boxplot Matlab ci-dessous (Via la commande `boxplot(Qtot, Eval)`).

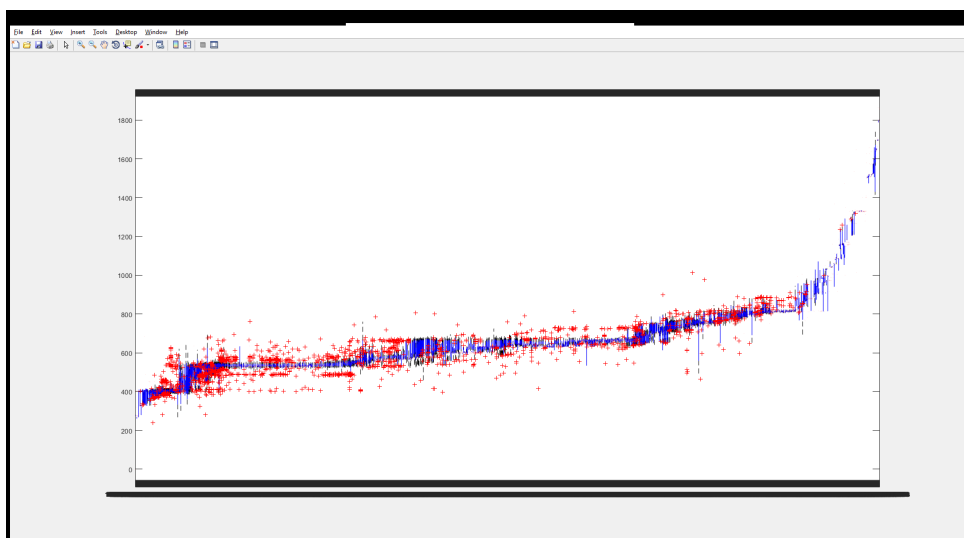


FIGURE 2.1 – Boxplot de l'élévation en fonction de la quantité turbinée

On remarque sur ce graphique que pour chaque valeur de x étant associées à plusieurs ordonnées, un diagramme à moustache a été tracé. Toutes les valeurs en rouge sont considérées par Matlab comme des valeurs aberrantes. Nous n'avons pas réussi à récupérer seulement les valeurs en bleues pour tenter de créer un modèle (même si beaucoup trop de points sont considérés comme aberrants, nous aurions aimé comparer un modèle basé sur cette approche avec le modèle décrit ci-dessous). Ces données étant la base du modèle (car la fonction associée sera utilisée pour le modèle de chaque turbine) nous avons décidé de garder toutes les données. En effet, toutes les valeurs se situent dans la même zone, et comme il peut exister plusieurs y pour un x , nous avons considéré que l'erreur produite

par des données aberrantes potentielles serait amoindrie par la masse des autres points. Ensuite nous avons utilisé l'outil « curve fitting » et nous avons choisi la régression polynomiale pour approximer la fonction. Nous avons testé différents paramètres afin d'obtenir un graphique résiduel convenable et des valeurs de SSE, RMSE proche de 0 ainsi qu'un R^2 et un R ajusté proche de 1. Une fois la fonction d'évaluation trouvée, nous avons mis la formule dans le fichier Excel afin de calculer les hauteurs de chutes de chaque turbine.

2.2 Fonctions de productions

Après avoir calculé la hauteur de chute de chaque turbine, nous avons réutilisé l'outil « curve fitting » pour déterminer la fonction de production de chaque turbine, toujours en choisissant une régression polynomiale puis en adaptant les paramètres pour valider notre modèle avec le graphique des résiduels.

Résultats

3.1 Fonction d'élévation

Après avoir importé les données dans Matlab, nous avons utilisé la boîte à outils "curve fitting" afin de trouver un modèle. Après avoir testé différentes combinaisons, nous avons choisi d'utiliser le modèle suivant :

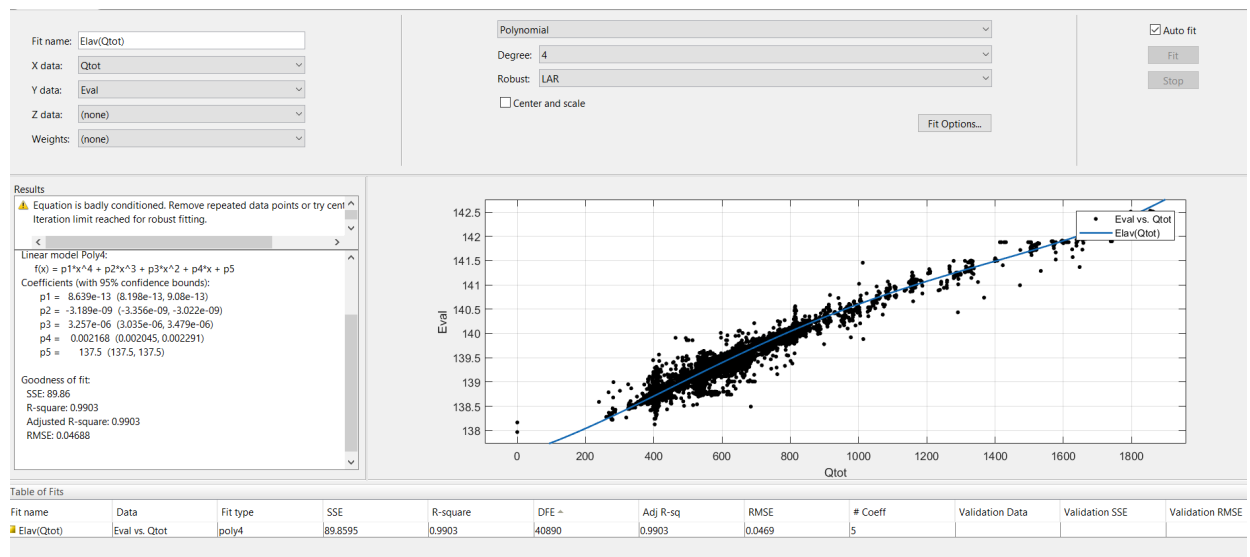


FIGURE 3.2 – Elevation aval en fonction du débit total à la centrale

Nous avons choisi la régression polynomiale, un degré de polynôme égal à 4 et la méthode LAR car pour ces paramètres nous avons le meilleur compromis entre le degré du polynôme et les différents indices de fiabilités (SSE, R^2 , ...). Choisir un degré de polynôme égal à 5 nous permettait d'aller chercher les deux points d'abscisse 0, mais n'améliorait pas de manière significative les indices de fiabilité.

$$Eval(Q_{tot}) = 8,639e^{-13} \times Q_{tot}^4 - 3,189e^{-9} \times Q_{tot}^3 + 3,257e^{-6} \times Q_{tot}^2 + 0,002168 \times Q_{tot} + 137.5 \quad (3.2)$$

Nous avons ensuite rentré l'équation (3.2) dans Excel afin de calculer, pour chaque turbine, les fonctions de production grâce à la formule suivante :

$$hChute_i = NivAmont - Elav(Q_{tot}) - (0.5e^{-5} \times Q_i^2) \quad (3.3)$$

3.2 Fonctions de productions

3.2.1 Turbine 1

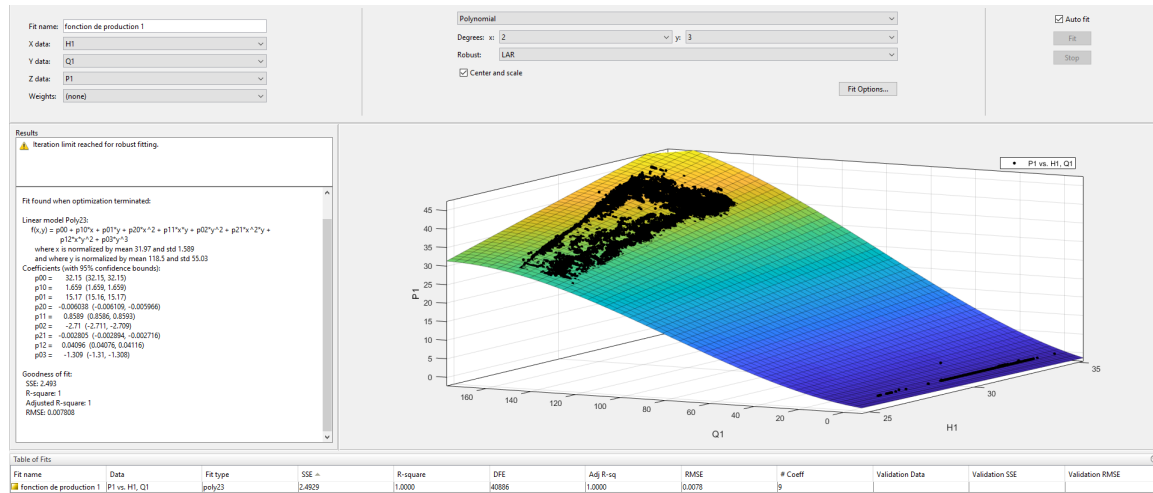


FIGURE 3.3 – Fonction de production de la turbine 1

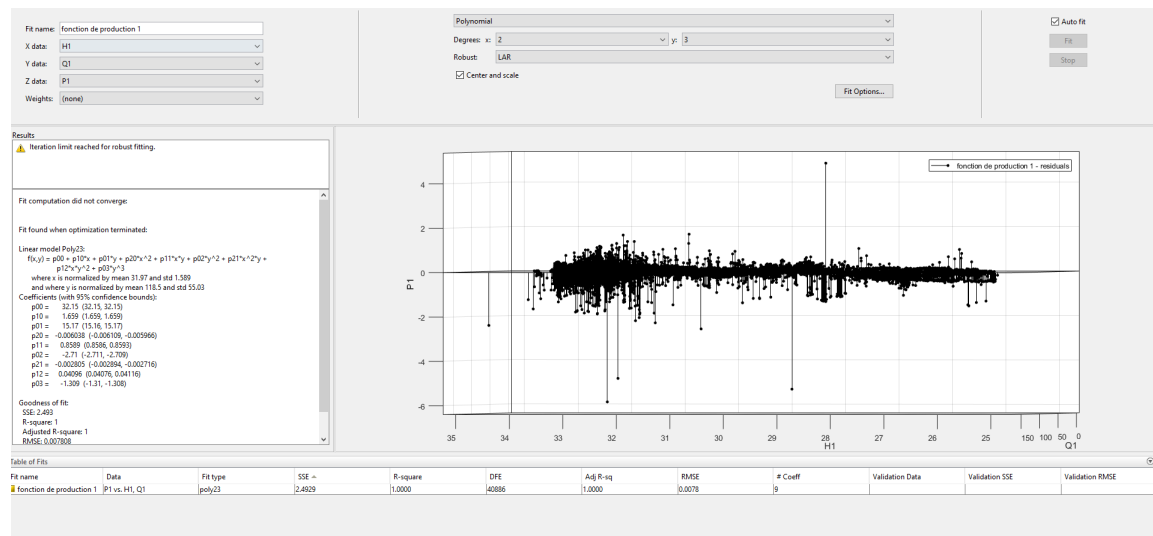


FIGURE 3.4 – Graphique des résiduels de la fonction de production 1

Voici la meilleure représentation de la fonction de production de la turbine 1 et son graphique résiduel. Ce dernier est équitablement distribué et seulement 4 points sont très éloignés. Nous avons testé en les supprimant mais nous n'obtenions pas de gains. C'est

due à la méthode d'approximation LAR. Ici nous avons choisi une approximation avec un degré en X de 2, et en Y de 3.

$$\begin{aligned} Production(H_1, Q_1) = & 32,15 + 1.659H_1 + 15,17Q_1 - 0,006038H_1^2 + \\ & 0.8589H_1Q_1 - 2.71Q_1^2 - 0.002805H_1^2Q_1 + 0.04096H_1Q_1^2 - 1.309Q_1^3 \end{aligned} \quad (3.4)$$

Avec H_1 normalisé par (31.97,1.589) et Q_1 normalisé par (118.5,55.03)

3.2.2 Turbine 2

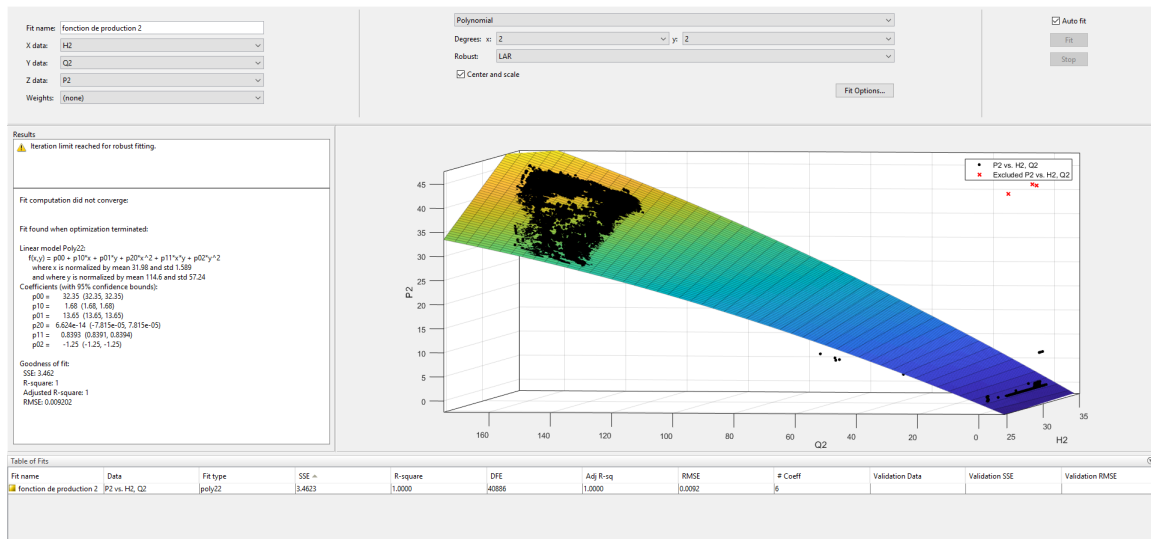


FIGURE 3.5 – Fonction de production de la turbine 2

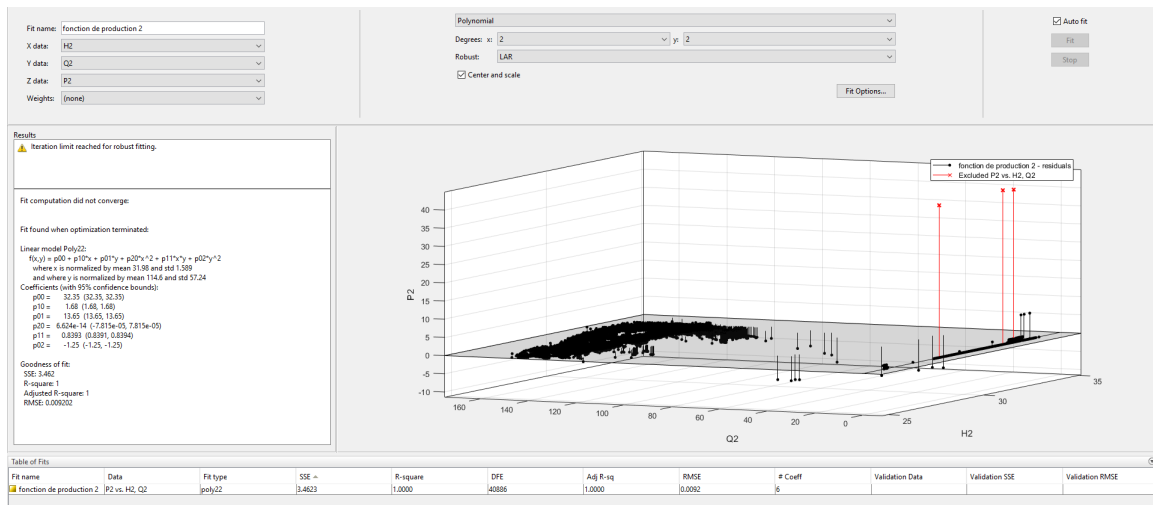


FIGURE 3.6 – Graphique des résiduels de la fonction de production 2

Avec H_2 normalisé par (31.98,1.589) et Q_2 normalisé par (114.6,57.24).

On a ici supprimé 3 points afin d'améliorer les valeurs du SSE (qui passe de 8 à 3).

Nous valions le le modèle grâce au graphique des résiduels.

$$\begin{aligned} Production(H_2, Q_2) = & 32,35 + 1,68H_2 + 13,65Q_2 \\ & + 6,624e^{-14}H_2^2 + 0,8393H_2Q_2 - 1,25Q_2^2 \end{aligned} \quad (3.5)$$

3.2.3 Turbine 3

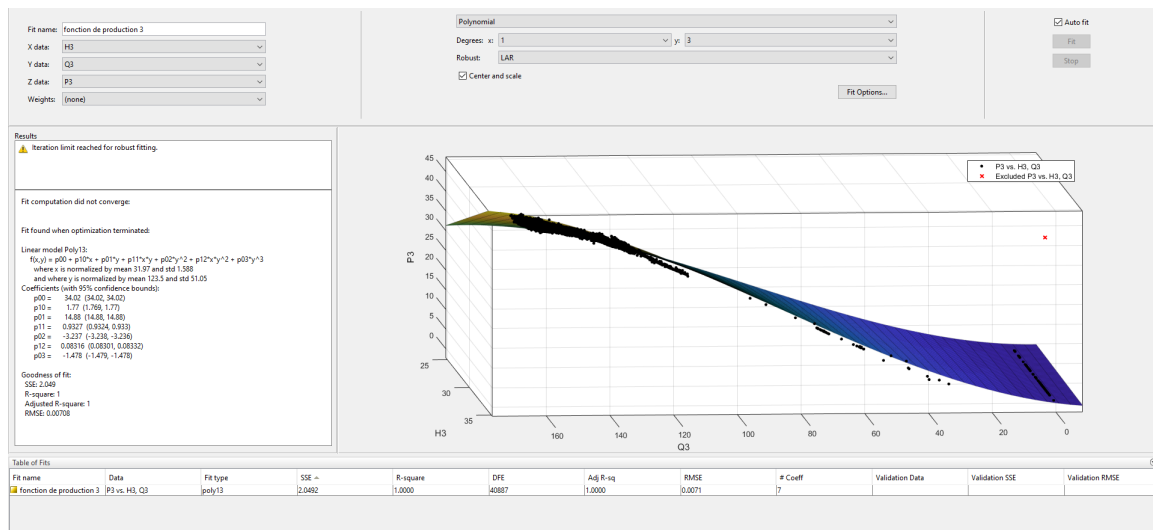


FIGURE 3.7 – Fonction de production de la turbine 3

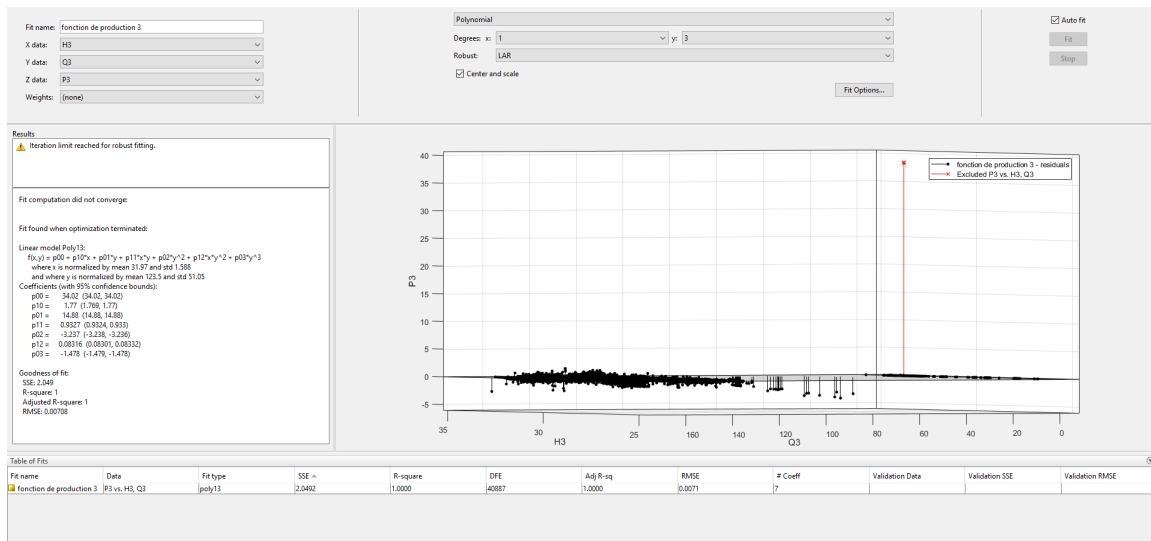


FIGURE 3.8 – Graphique des résiduels de la fonction de production 3

On a supprimé un point qui semble totalement aberrant.

$$\begin{aligned}
 Production(H_3, Q_3) = & 34,02 + 1,77H_3 + 14,88Q_3 \\
 & + 0,9327H_3Q_3 - 3,237Q_3^2 + 0,08316H_3Q_3^2 - \\
 & 1,478Q_3^3
 \end{aligned} \tag{3.6}$$

Avec H_3 normalisé par (31.97,1.588) et Q_3 normalisé par (123.5,51.05).

3.2.4 Turbine 4

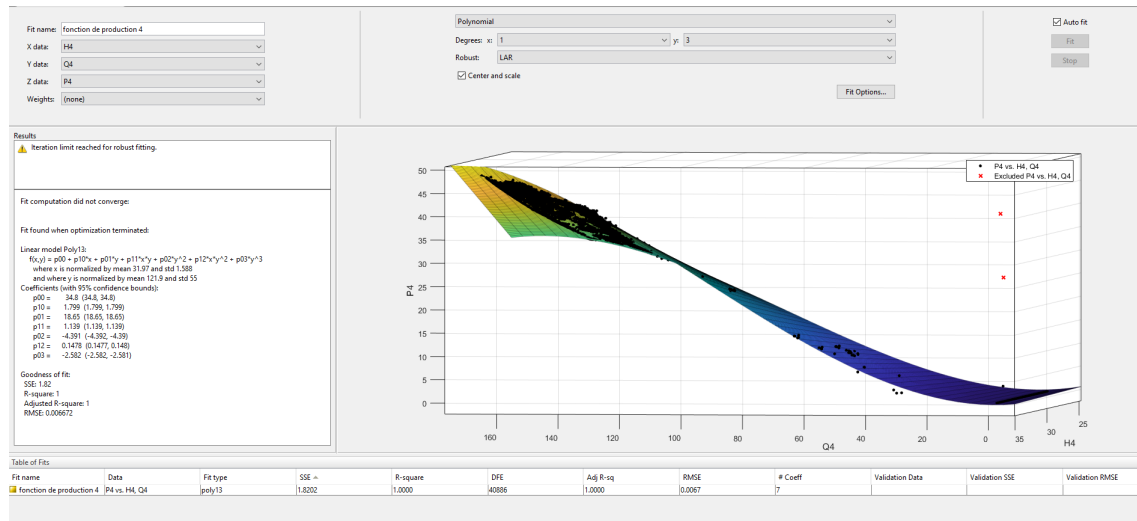


FIGURE 3.9 – Fonction de production de la turbine 4

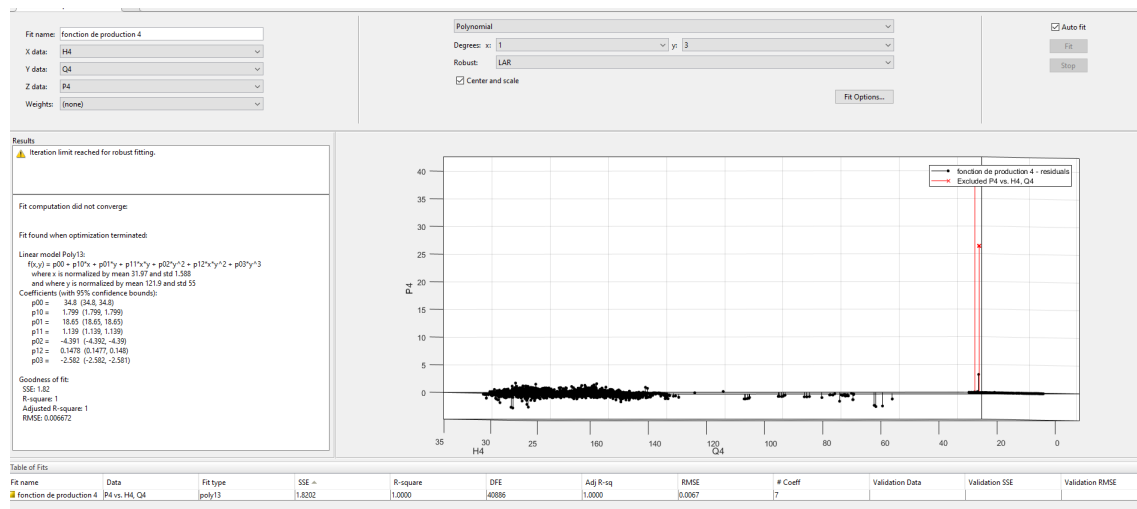


FIGURE 3.10 – Graphique des résiduels de la fonction de production 4

$$\begin{aligned}
 Production(H_4, Q_4) = & 34,8 + 1.799H_4 + 18,65Q_4 \\
 & + 1,139H_4Q_4 - 4,391Q_4^2 + 0,1478H_4Q_4^2 - \\
 & 2,582Q_4^3
 \end{aligned} \tag{3.7}$$

Avec H_4 normalisé par (31.97,1.588) et Q_4 normalisé par (121.9,55).

3.2.5 Turbine 5

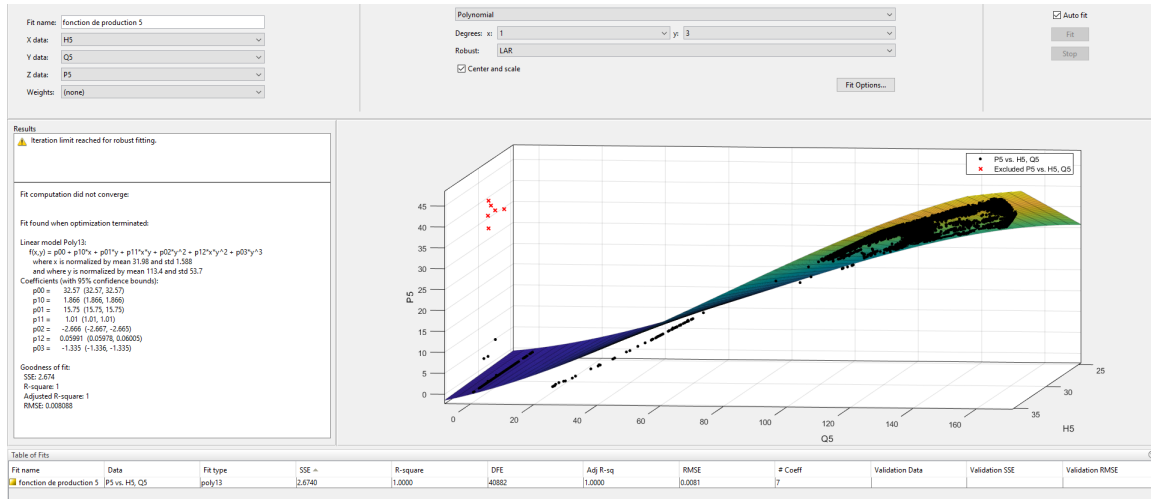


FIGURE 3.11 – Fonction de production de la turbine 5

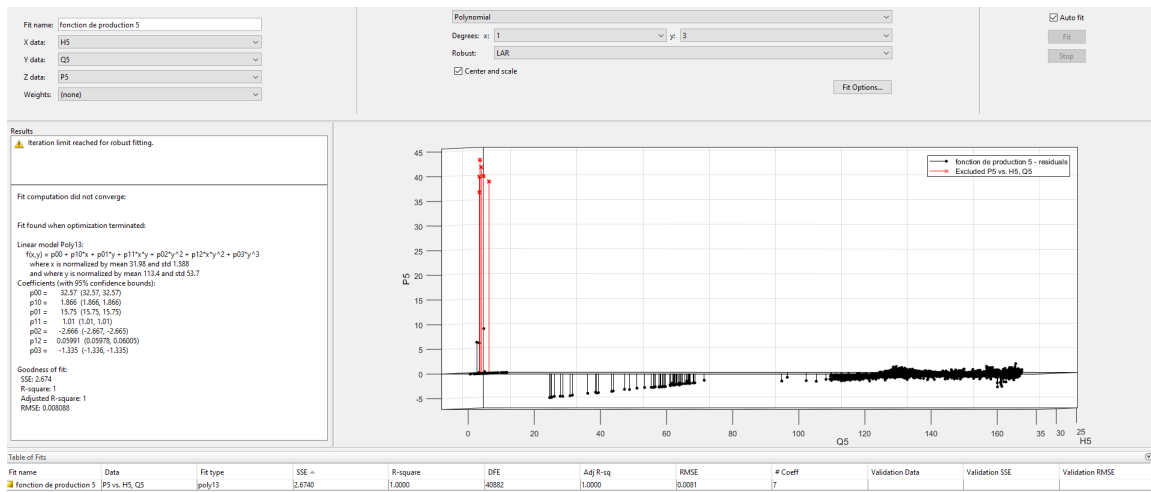


FIGURE 3.12 – Graphique des résiduels de la fonction de production 5

On voit sur cette dernière figure que 4 points sont situés dans la même zone. Cela peut être due à une spécificité de la turbine (qui est plus récente et donc différente des autres) ou à un capteur défectueux. On voit également que le diagramme des résiduel n'est pas équitable réparti. Le modèle peut donc être amélioré mais nous n'avons pas réussi à trouver

une meilleure régression.

$$\begin{aligned}
 Production(H_5, Q_5) = & 32,57 + 1,866H_5 + 15,75Q_5 \\
 & + 1,01H_5Q_5 - 2,666Q_5^2 + 0,05991H_5Q_5^2 - \\
 & 1,335Q_5^3
 \end{aligned} \tag{3.8}$$

Avec H_5 normalisé par (31.98,1.588) et Q_5 normalisé par (113.4,53.7).

Discussions

Nous pensons que nos modèles pour l'élévation aval et les turbines 1, 2, 3 et 4 sont corrects. En effet, les graphiques résiduels montrent des erreurs distribuées normalement et indépendantes. Leurs moyennes sont proches de 0. De plus, les coefficients SSR et RMSE sont proches de 0 et le coefficient de détermination R^2 est proche de 1. Nous avons eu plus de difficultés pour la turbine 5, dont le modèle donne de bons coefficients, mais dont le graphique résiduel n'est pas excellent. En effet, entre les valeurs 20 et 70 (Q5), il y a un problème d'approximation. De plus, on remarque un pull de 6 points détaché du reste qui pourrait peut-être être une spécificité du système.

De même, nous n'avons jamais fait de tris ou lissage de données via des algorithmes ou en utilisant des fonctions. Nous avons seulement utilisé l'outil graphique "curve fitting" afin de supprimer les valeurs aberrantes.

Nous avons également tenter de limiter la complexité de nos polynômes afin de simplifier les traitements futurs. Il était parfois possible d'améliorer nos modèles de quelques centièmes de pourcent au prix d'un ou deux degrés supplémentaires.

Nous prenons donc en hypothèse de départs pour les prochains travaux, que nos équations approchent bien la réalité.

Conclusion

Ce document contient la base de travail pour les jalons futurs. Nous avons réussi à modéliser les différentes fonctions de production grâce à Matlab et à sa boîte à outils "Curve Fitting". Les turbines 1, 2, 3 et 4 possèdent des modèles similaires car se sont les quatre turbines d'origine. Les coefficients des équations modélisant leur production sont cependant différents car chaque turbine possède sa propre fonction de production (se ne sont pas parfaitement les mêmes). La turbine 5, installée plus tard, possède elle une fonction de production différente des précédentes. Cette turbine est plus puissante. Dans notre future travail, nous implémenterons un algorithme de programmation dynamique permettant le chargement optimal des groupes, ainsi qu'une interface utilisateur permettant aux opérateurs de centrales hydroélectrique de les contrôler. Nous prendrons soins de réaliser un logiciel modulaire, permettant de modifier simplement les équations de production. Ainsi, de nouvelles équations pourront êtres entrées lors d'un changement de turbine, ou lorsque, avec l'âge, la turbine devient moins efficace.