

UNIVERSITÉ CLAUDE BERNARD LYON I

MASTER DATA SCIENCE

MODÈLES DE RÉGRESSION

Régression des données d'extraction de puits de pétrole au Canada

Auteurs :

Bruno DUMAS

Jules SAUVINET

Professeur :

François WAHL

11 janvier 2017



Résumé

L'objectif est d'effectuer des régressions des valeurs d'extractions des puits de pétrole du Canada afin de pouvoir prédire les future données d'extraction. Les courbes des valeurs des données d'extraction donne une classification de la qualité des puits. L'objectif de ce travail est de mettre en place une classification automatique de ces puits : La démarche proposée est décrite ci-dessous. L'idée est de remplacer ces courbes par des fonctions paramétriques. Pour cela, plusieurs r égressions vont être envisagées afin d'avoir la meilleure prédiction/classification possible.

Table des matières

1	Régressions polynomiales	2
2	Régressions exponentielles	3
3	Courbes hautes et basses á 95%	4
4	Reclassement avec régression logistique	5
5	Gestion des spikes et lissage des courbes	6
6	Conclusion	7

1 Régressions polynomiales

Si on se réfère à la figure ci-dessus, on observe que la plupart des simulations présente une remontée au bout de quelques mois, que certaines simulations sont concaves au lieu d'être convexes, voire que certaines d'entre elles pourraient avoir des valeurs négatives (autour des 35 mois d'exploitation).

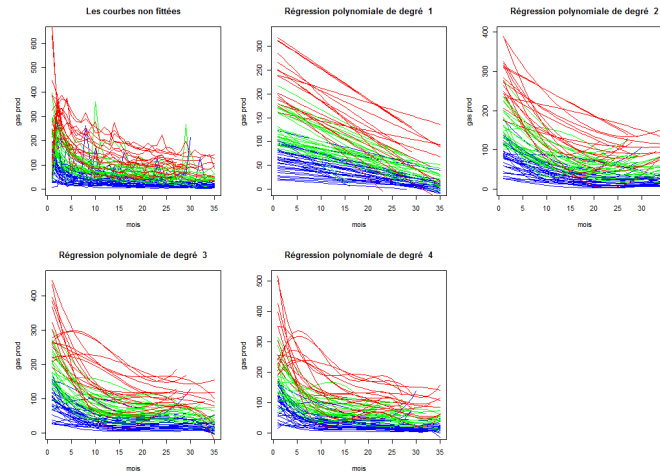


FIGURE 1 – Courbes non régressées, et courbes de régressions polynomiales de degrés 1,2,3 et 4

2 Régressions exponentielles

Une idée simple pour corriger ces défauts est d'utiliser une autre forme paramétrique pour les simulations. Une suggestion immédiate pour qui a un peu l'habitude de ces courbes est une forme exponentielle du style : , où y est la production, t le mois, et k_0 et k_1 deux paramètres à déterminer.

Régression exponentielle avec `lm`

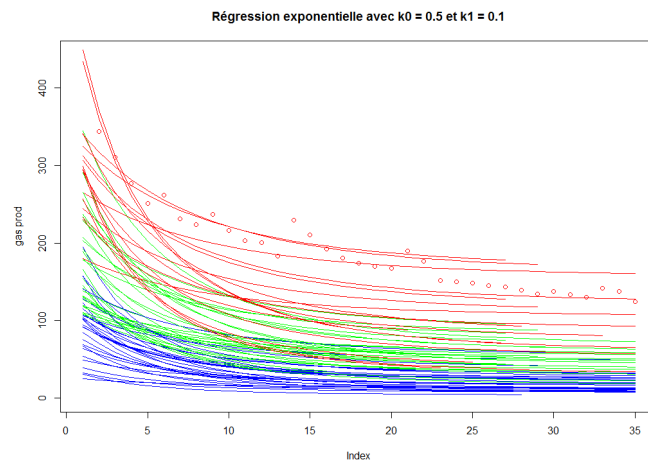


FIGURE 2 – Courbes régressées avec une fonction exponentielle et et la fonction `lm` de R

Régression exponentielle avec `nls`

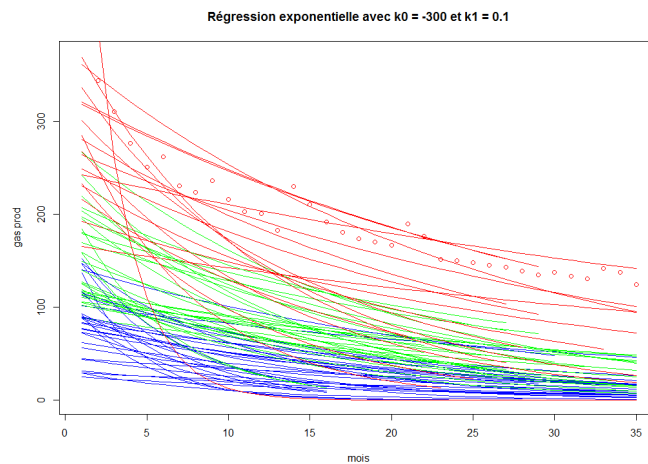


FIGURE 3 – Courbes régressées avec une fonction exponentielle et et la fonction `nls` de R

Est-ce que cela marche ? avez-vous d'autres idées ?

3 Courbes hautes et basses à 95%

Quelles sont les incertitudes sur les régressions des points 2 ? Plus concrètement, on vous demande de tracer pour un exemple de chaque type de courbe, la courbe haute (à 95%) et la courbe basse (toujours à 95%)

Avec `predict nls`

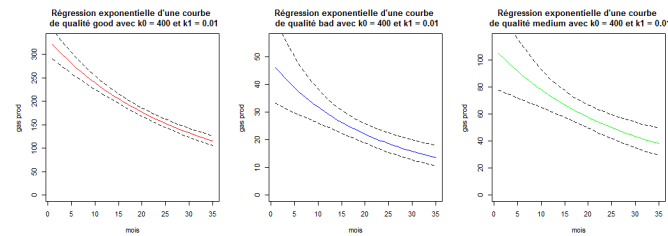


FIGURE 4 – Courbes hautes et basses pour un exemple de 3 classes de courbes

Avec `bootstrapping`

4 Reclassement avec régression logistique

k_1 en fonction de k_0

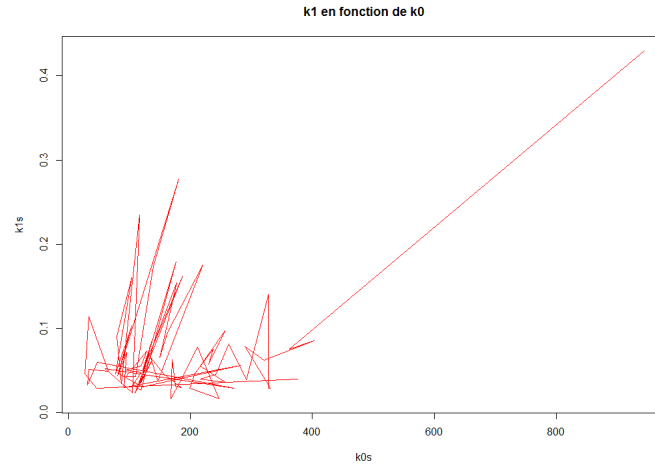


FIGURE 5 – Coefficients k_1 en fonction de k_0

En examinant le graphe k_1 fonction de k_0 , on se rend compte que certaines courbes classées 'Good' par les experts donnent l'impression d'être plutôt 'medium', tandis que certaines 'bad' pourraient être aussi 'medium'. // Avez-vous des suggestions sur 5 courbes au plus qui pourraient être mal classées? Justifiez vos choix (i.e. une façon de faire est d'effectuer une régression logistique dont le y est la classe prédite par l'expert et les x sont les coefficients k_0 et k_1 , et d'examiner comment la régression est améliorée en changeant la classe d'un point; une autre façon plus empirique est de déterminer deux droites $x=k_{11}$ et $x=k_{12}$ qui partitionnent au mieux les classes et d'examiner comme précédemment comment la classification est changée en basculant certains points d'une classe à l'autre).

5 Gestion des spikes et lissage des courbes

Régression polynomiale de degré 3 avec lissage loess

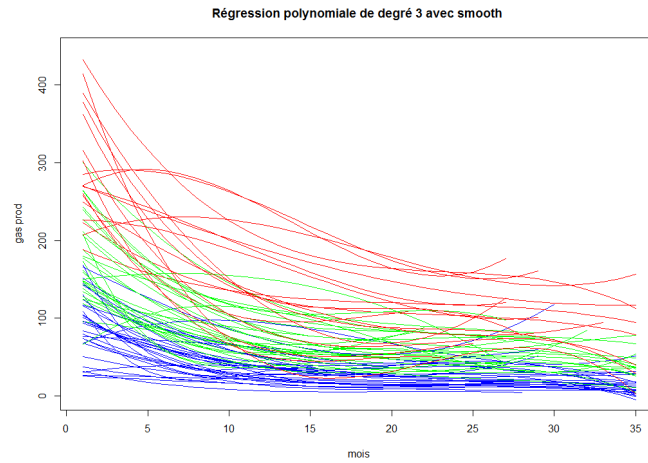


FIGURE 6 – Régression polynomiale de degré 3 avec courbes lissée au préalable avec loess

Régression exponentielle avec lissage loess

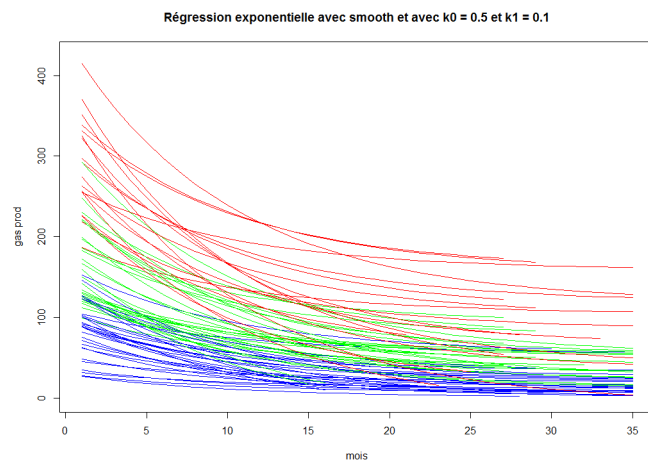


FIGURE 7 – Régression exponentielle lm avec courbes lissée au préalable avec loess

6 Conclusion

TODO TOO

Acknowledgments

Merci á François Wahl.