

La consommation de glyphosate en France

Lancelot Ravier

2024-043-27

1 Introduction

Le glyphosate, découvert en 1970 et commercialisé dès 1974 en France (notamment sous le nom Roundup), est devenu l'herbicide le plus utilisé à travers le monde, tant dans le secteur agricole que chez les particuliers. Sa popularité s'explique par son efficacité remarquable à éradiquer les mauvaises herbes et les plantes indésirables dans une grande variété de milieux, des vastes champs agricoles aux jardins privés. Toutefois, l'utilisation généralisée de cet herbicide n'est pas sans controverses. D'importantes inquiétudes ont été soulevées concernant ses effets potentiellement cancérogènes, des études scientifiques ayant mis en lumière des liens entre l'exposition au glyphosate et certains types de cancer chez l'humain.

Face à ces risques sanitaires, l'Union Européenne a progressivement adopté une position plus restrictive, envisageant même une interdiction totale de l'herbicide malgré les débats scientifiques et politiques toujours en cours. La France, malgré la pression internationale et les recommandations de l'UE, a montré une certaine réticence à bannir complètement l'usage du glyphosate. Cette position s'explique principalement par la forte dépendance des agriculteurs français au glyphosate, dont l'efficacité quasiment inégalée permet de soutenir la productivité agricole tout en limitant les coûts. De plus, le défi de trouver des alternatives aussi efficaces et économiquement viables reste considérable.

Dans ce contexte, notre projet de recherche vise à examiner de manière approfondie l'impact de divers facteurs agricoles sur les quantités de glyphosate achetées en France. Nous analysons les données agrégées par département pour estimer les ventes de glyphosate, en nous appuyant sur des variables telles que la surface agricole utilisée, la production brute standard, et la part des cultures spécifiques comme les vignes et les céréales. Ce cadre économétrique offre une perspective précieuse pour comprendre les dynamiques d'utilisation du glyphosate et pour éclairer les discussions sur les politiques agricoles et environnementales futures.

Quels-sont les déterminants de la consommation de Glyphosate en France métropolitaine ?

2 Données

Nom des bases de données originales :

glyphosate (source : <https://ventes-produits-phytopharmaceutiques.eaufrance.fr/search?filetype=Achats>)



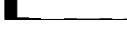



vignes (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-partvigne20/detail/>)

pbs (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-pbs20/detail/>)

sau (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-sau20/detail/>)

cereales (<https://agreste.agriculture.gouv.fr/agreste-web/disaron/Carte-RA-partcereoleopro20/detail/>)

Nom de la base de données agrégée : data

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
annee	1	0	2020.0	0.0	2020.0	2020.0	2020.0	
glyphosate	96	0	88 445.5	68 943.1	749.7	80 946.9	347 598.2	
vignes	41	10	4.7	10.8	0.0	0.4	50.1	
cereales	90	0	38.4	24.2	0.0	38.7	84.9	
pbs	96	0	668 782.6	520 026.7	103.0	534 718.0	2 685 717.0	
sau	96	0	278 602.9	143 130.3	1.0	296 772.5	557 179.0	

2.1 Variables

annee - 2020

code_departement - numéro indicatif pour chaque département

departement - nom du département

pbs - production brut standard, production potentielle totale des exploitations par département, résultent des valeurs moyennes des rendements et des prix observés sur la période 2015 à 2019, exprimé en euros.

sau - superficie agricole utilisée (en ha.), comprenant les céréales, les oléagineux, protéagineux et plantes à fibres, les autres plantes industrielles destinées à la transformation, les cultures fourragères et les surfaces toujours en herbe, les légumes secs et frais, les fraises et les melons, les pommes de terre, les fleurs et plantes ornementales, les vignes, les autres cultures permanentes (vergers, petits fruits, pépinières ligneuses), les jachères, les jardins et vergers familiaux.

cereales - part des cereales et oléagineux dans la sau (en %)

vignes - part de vignes dans la sau (en %)

glyphosate - quantité totale (en kg) de glyphosate achetée

2.2 Import et fusion des différentes bases de données

2.3 Visualisation des données

Certaines observations sont manquantes pour la variable vignes (due au secret statistique). Il est nécessaire de traiter ces valeurs manquantes avant toute analyse de données car ces dernières rendent toute analyse graphique impossible.

2.4 Traitement des valeurs manquantes

Les valeurs manquantes pour la variable vignes sont dues au secret statistique (moins de 3 cultures de vignes). Ainsi, nous pouvons remplacer ces valeurs manquantes par des 0.

2.5 Creation de variables

La base de donnée finale contient maintenant certaines variables ajoutées (log).

3 Analyse des données

3.1 Analyse générale

##	vars	n	mean	sd	median	trimmed	mad
## code_departement*	1	96	48.50	27.86	48.50	48.50	35.58
## departement*	2	96	48.50	27.86	48.50	48.50	35.58

## glyphosate	3	96	88445.47	68943.12	80946.88	82337.76	81775.27
## vignes	4	96	4.19	10.28	0.20	1.38	0.30
## cereales	5	96	38.35	24.17	38.70	38.02	33.73
## pbs	6	96	668782.58	520026.71	534718.00	603190.58	347525.89
## sau	7	96	278602.89	143130.28	296772.50	283144.96	159880.62
## log_pbs	8	96	12.91	1.58	13.19	13.17	0.72
## log_sau	9	96	12.06	1.96	12.60	12.46	0.53
## log_glyphosate	10	96	10.89	1.27	11.30	11.06	0.94
##			min	max	range	skew	kurtosis
## code_departement*	1.00		96.00	95.00	0.00	-1.24	2.84
## departement*	1.00		96.00	95.00	0.00	-1.24	2.84
## glyphosate	749.72		347598.17	346848.46	0.84	0.62	7036.48
## vignes	0.00		50.10	50.10	3.12	8.94	1.05
## cereales	0.00		84.90	84.90	0.05	-1.23	2.47
## pbs	103.00		2685717.00	2685614.00	1.45	2.51	53075.00
## sau	1.00		557179.00	557178.00	-0.22	-1.01	14608.17
## log_pbs	4.63		14.80	10.17	-3.18	12.24	0.16
## log_sau	0.00		13.23	13.23	-4.47	21.85	0.20
## log_glyphosate	6.62		12.76	6.14	-1.27	1.30	0.13

Pour l'année 2020,

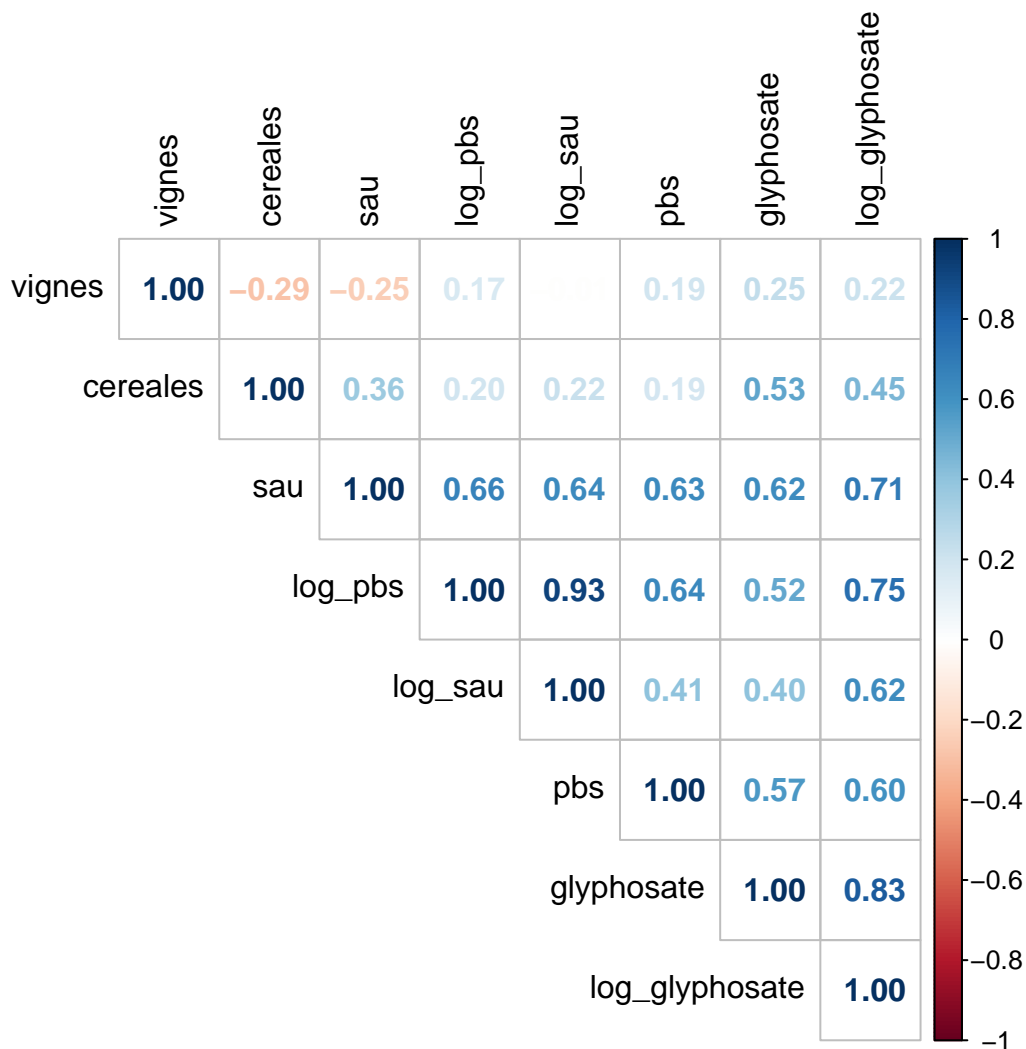
La quantité moyenne de glyphosate acheté par département est de 88445.47 kg. La Charente maritime est le premier acheteur de glyphosate (347598.1719 kg). La Marne arrive en 3ème position.

la PBS moyenne d'un département français est de 668782.6€, variant de 103€ dans les Hauts de Seine (92) à 2685717€ en Marne (51). L'écart-type est de 53075€. La distribution est asymétrique à droite (skew = 1.45) et leptokurtique (kurtosis = 2.51) suggérant des écarts à la moyenne supérieurs à ceux d'une distribution normale.

La SAU moyenne est de 278601.9 Ha, allant de 1 Ha pour Paris (75), à 557179 pour la Marne (51). L'écart-type est de 14608.16 ha. La distribution est légèrement asymétrique à gauche, et platykurtique (kurtosis = -1.01), suggérant un nombre d'écarts à la moyenne inférieurs que dans une distribution normale.

En moyenne, 38.35% de la SAU est destinée à la culture de céréales, et 4.28% de la SAU destinée à la culture viticole. C'est en Eure et Loire que la part de SAU consacrée à la culture de céréales est la plus grande (84.9%) et c'est en Gironde que la part de la SAU consacrée à la culture de vignes est la plus grande (50.1%).

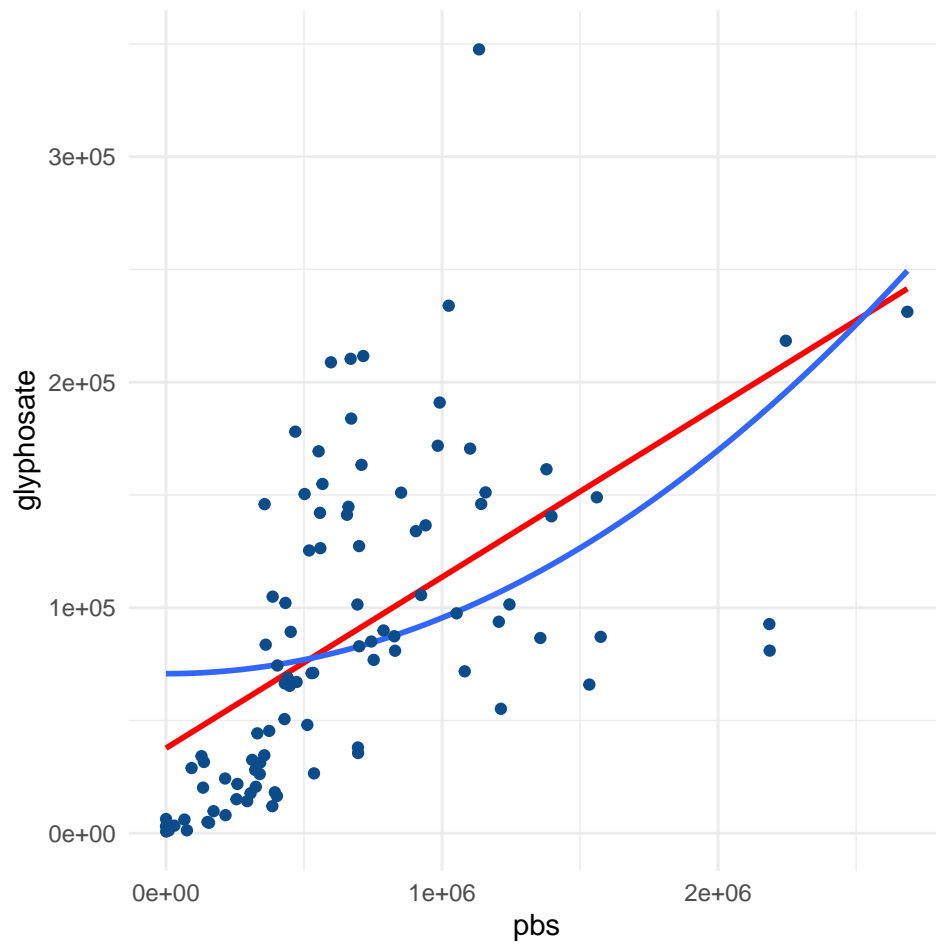
3.2 Matrice de corrélation



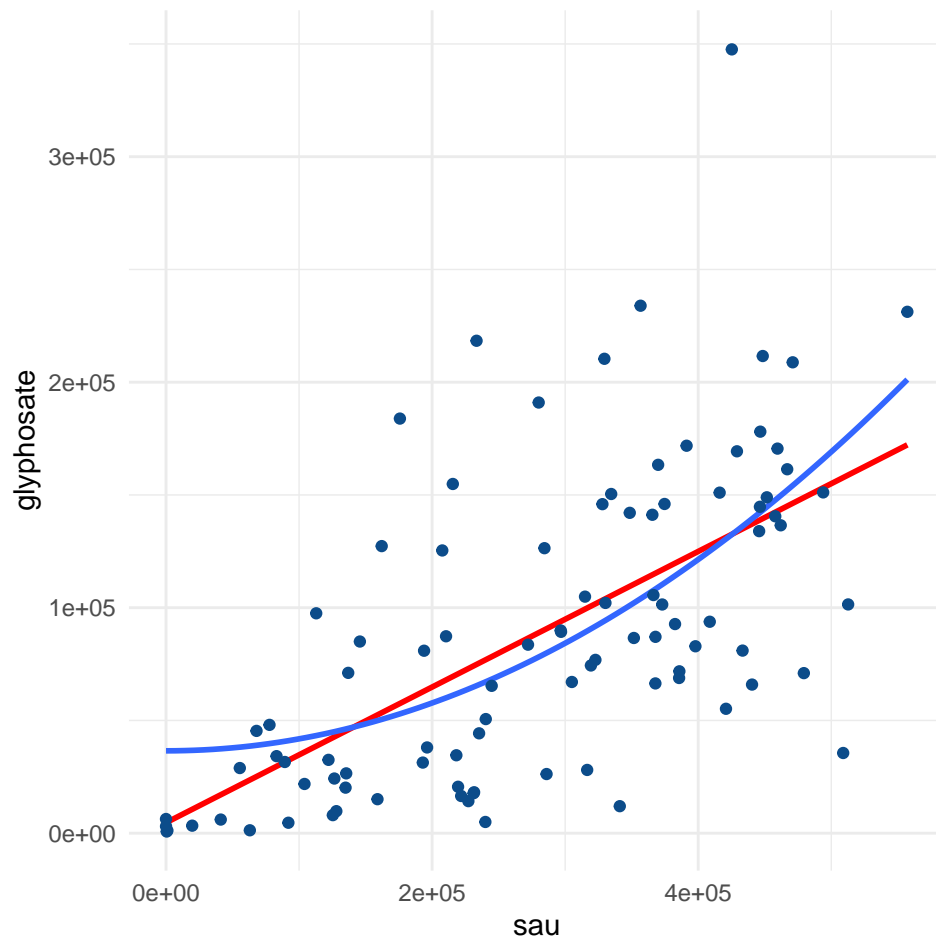
Toutes les variables ont une corrélation positive avec glyphosate. Cette relation est modérée pour les variables cereales, pbs, sau et faible pour la variable vignes.

D'après la matrice de corrélation, aucune corrélation trop élevée n'est détectée sauf entre pbs et sau, expliquée par le fait que la production d'une exploitation est forcément corrélée à sa taille. Le risque d'overfitting peut donc être écarté.

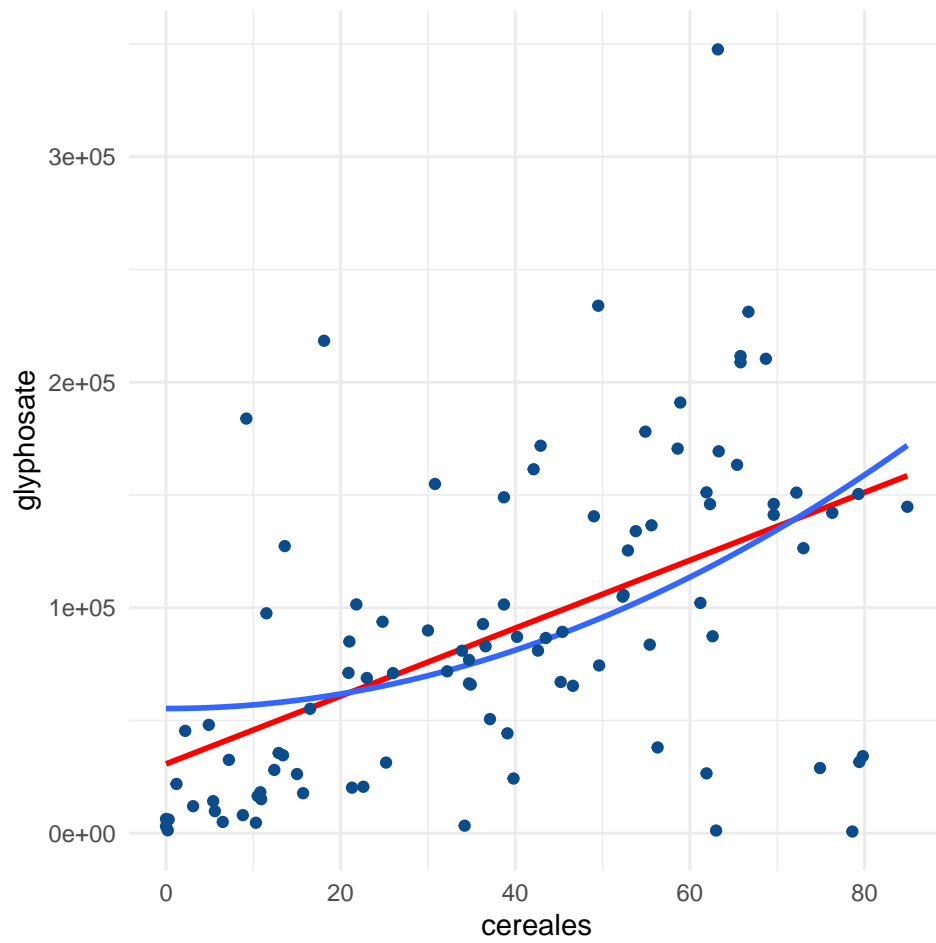
```
## `geom_smooth()` using formula = 'y ~ x'
```



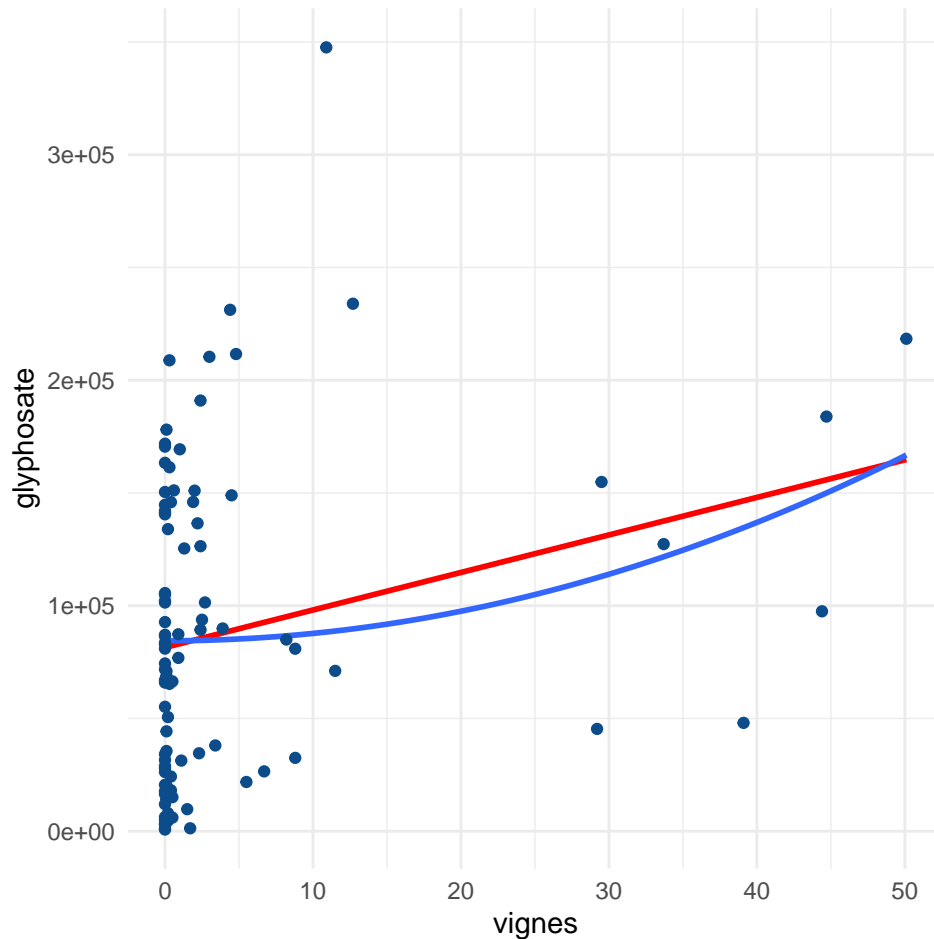
```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'
```



D'après ces graphiques, des relations non-linéaires (carré) peuvent être observées entre les variables explicatives et glyphosate. Il serait donc pertinent de tester la présence de relations non-linéaire et d'inclure certaines variables au carré dans nos futurs modèles, afin de modéliser et prendre en compte ces effets dans nos estimations.

4 Modèles et test

4.1 Equation à estimer

L'équation à estimée est la suivante :

$$\log_glyphosate = \beta_0 + \beta_1 * \log_pbs + \beta_2 * \log_sau + \beta_3 * cereales + \beta_4 * vignes + u$$

Le paramètre lié à pbs est une élasticité (variation en % lorsque le pbs augmente de 1%, ceteris paribus). Le signe du paramètre devrait être positif. En effet, la production est importante, plus le recours au glyphosate serait nécessaire pour protéger et assurer les récoltes aux enjeux davantage important que leur taille serait conséquente.

Le paramètre lié à sau est une élasticité (variation en % lorsque la sau augmente de 1%, ceteris paribus). Le signe du paramètre devrait être positif. En effet, plus la superficie agricole utilisée est importante, plus l'utilisation de glyphosate sera importante, ce dernier étant utilisé sur une superficie plus grande.

Les paramètres liés à cereales et vignes sont des semi-élasticités (variation de $(\beta_i * 100)\%$ lorsque vignes/cereales augmente d'une unité). Le signe du paramètre lié à vignes et cereales devrait aussi être positif car le glyphosate est principalement utilisé dans les vignes et les cultures de céréales.

	(1)
(Intercept)	2.284 (0.648) *** [0.996, 3.571]
log_pbs	0.927 (0.133) *** [0.664, 1.191]
log_sau	-0.349 (0.105) ** [-0.558, -0.140]
cereales	0.020 (0.003) *** [0.014, 0.026]
vignes	0.016 (0.008) * [0.001, 0.032]
Num.Obs.	96
R2	0.739
R2 Adj.	0.728
AIC	200.5
BIC	215.9
Log.Lik.	-94.248
F	64.542
RMSE	0.65

4.2 Modèle MCO

Afin de valider les résultats de ce premier modèle, il est nécessaire de vérifier le respect des hypothèses économétriques.

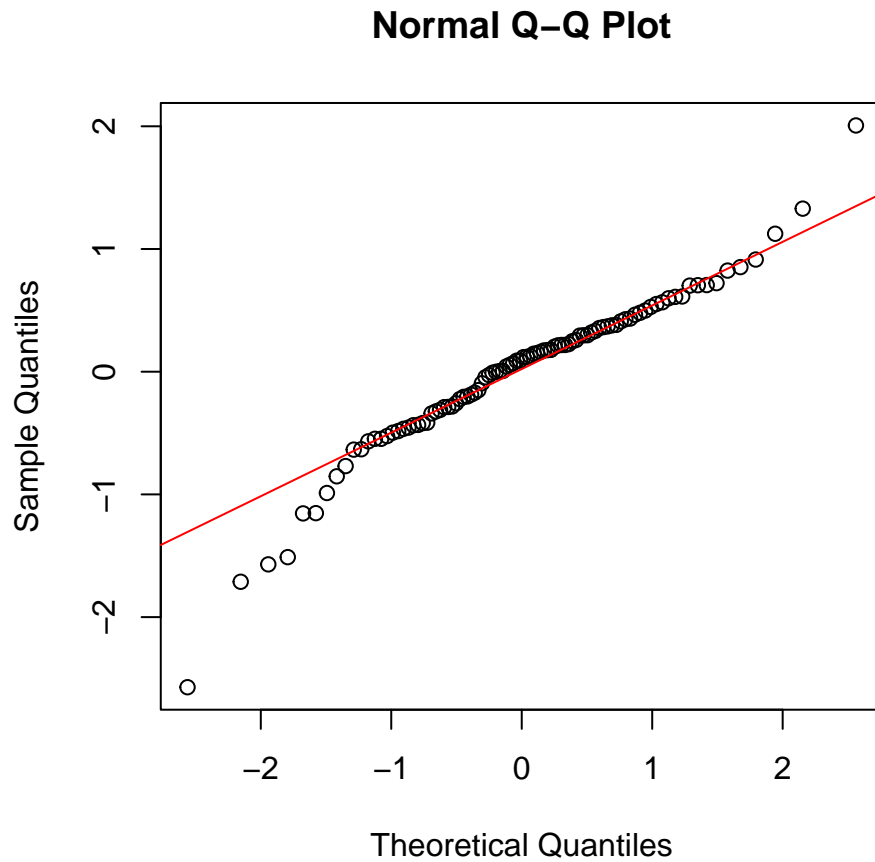
4.3 Hypothèses

4.3.1 H1 : Espérance de l'erreur nulle

```
##
## One Sample t-test
##
## data:  res1
## t = 3.6651e-16, df = 95, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.131548  0.131548
## sample estimates:
## mean of x
## 2.428613e-17
```

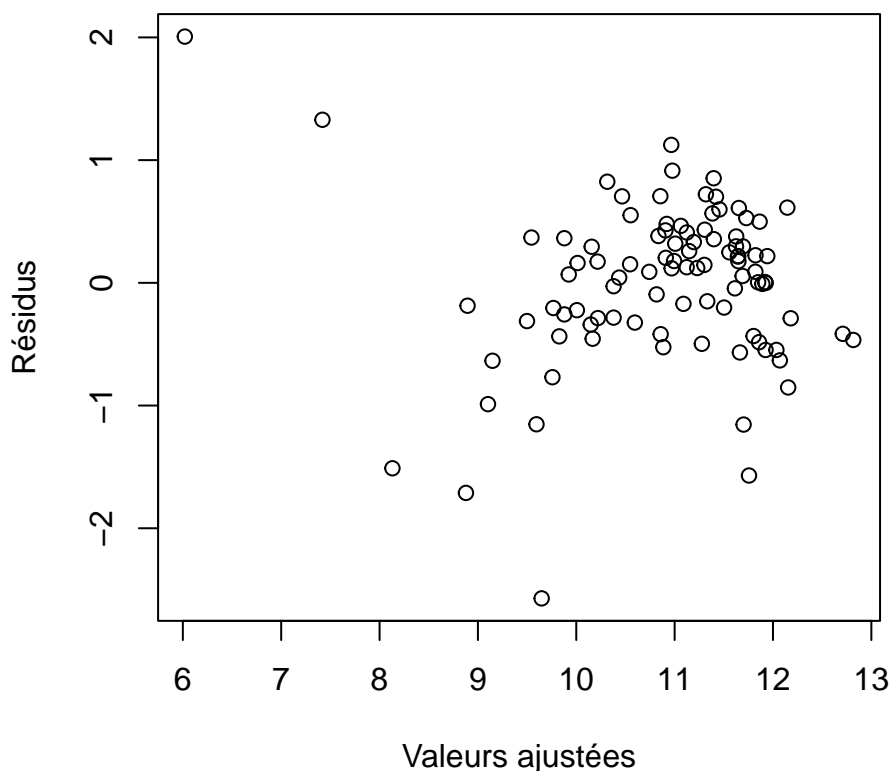
D'après ce test, l'espérance du terme d'erreur n'est pas significativement différente de 0. L'hypothèse H1 est donc respectée.

4.3.2 H2 : Perturbations heteroscedastiques



Les résidus ne suivent pas exactement une loi normale centrée réduite. Afin de s'assurer du calcul non faussé de la matrice var-cov ainsi que le calcul des tests qui en découlent, il est nécessaire de vérifier la présence d'hétéroscédasticité des perturbations.

Résidus vs. Valeurs ajustées



D'après ce graphique, nous pourrions être en présence d'hétéroscédasticité des perturbations (dispersion croissante).

Warning: Heteroscedasticity (non-constant error variance) detected (p < .001).

Le test de Breuch-Pagan vient confirmer la présence d'hétéroscédasticité des perturbations précédemment visualisée. La matrice des variances covariances doit être corrigée afin de réaliser une inférence statistique fiable.

4.3.3 H3 : Variables mesurées avec erreur

Certaines des variables de notre modèle pourraient être mesurées avec erreur.

- La variable pbs pourrait être coréllée avec le terme d'erreur dans le cas où des facteurs non observés influenceraient à la fois la pbs et l'achat de glyphosate, où dans le cas présent car la pbs inclut toutes les exploitations par département, mêmes celles n'utilisant pas de glyphosate.
- La variable sau pourrait être coréllée avec le terme d'erreur dans le cas où es facteurs non mesurés affecteraient à la fois la sau et 'achat de glyphosate (politiques agricoles spécifiques, conditions météorologiques extrêmes)
- Les variables cereales et vignes euvent aussi être sujettes à endogénéité dans le cas où, par exemple, certaines cultures nécessiteraient davantage d'herbicides que d'autres, où si des pratiques agricoles spécifiques sont coréllées à la fois avec la composition de la sau et l'achat de glyphosate.

Dans le cadre de notre modèle mco1, la variable susceptible d'être mesurée avec erreur est la variable sau. En effet, le paramètre lié à `log_sau` est de signe négatif, ce qui va à l'encontre de nos intuitions économique de départ. De plus, ce paramètre est significatif au seuil de 5%, ce qui écarte la non-significativité du paramètre. La base de données analysée ne permet pas d'effectuer des tests quant à l'utilisation de variables instrulmentales (coréllées avec sau, non coréllés avec le terme d'erreur).

4.3.4 H4 : Spécification du modèle

Afin de tester la bonne spécification de notre modèle, certaines variables au carré ont été créés.

```
##
## RESET test
##
## data: mco1
## RESET = 45.031, df1 = 4, df2 = 87, p-value < 2.2e-16
```

D'après le test RESET et les analyses graphiques réalisées en section 3, le modèle n'est pas correctement spécifié car certaines relations non-linéaires (au carré) ne sont pas prises en compte dans l'estimation.

```
##
## RESET test
##
## data: mco_reset_1
## RESET = 2.4876, df1 = 8, df2 = 79, p-value = 0.0184

##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.7836e+00 2.7150e+00  2.8669 0.0051992 **
## log_pbs      2.5955e-01 6.2626e-01  0.4144 0.6795717
## log_sau     -1.1674e+00 2.1968e-01 -5.3144 8.152e-07 ***
## cereales     4.0045e-02 9.1502e-03  4.3764 3.341e-05 ***
## vignes       1.0359e-01 1.5637e-02  6.6244 2.778e-09 ***
## sq_log_pbs  -4.3508e-03 2.3206e-02 -0.1875 0.8517140
## sq_log_sau   8.8507e-02 1.0386e-02  8.5215 4.282e-13 ***
## sq_cereales -2.2373e-04 9.4316e-05 -2.3722 0.0198889 *
## sq_vignes   -1.2518e-03 3.3940e-04 -3.6882 0.0003926 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D'après le coeftest, seul la variable `sq_log_pbs` n'est pas significative. Ainsi, il semble pertinent de la supprimer pour voir l'effet sur les résultats du test RESET.

```
##
## RESET test
##
## data: mco_reset_2
## RESET = 0.6929, df1 = 7, df2 = 81, p-value = 0.6778

##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.1874e+00 7.9495e-01 10.2992 < 2.2e-16 ***
## log_pbs      1.5295e-01 1.1792e-01  1.2970 0.1980074
## log_sau     -1.1290e+00 1.0680e-01 -10.5715 < 2.2e-16 ***
## cereales     3.9648e-02 9.0692e-03  4.3717 3.366e-05 ***
## vignes       1.0379e-01 1.5368e-02  6.7538 1.489e-09 ***
## sq_log_sau   8.7041e-02 8.6616e-03 10.0491 2.850e-16 ***
## sq_cereales -2.2012e-04 9.3634e-05 -2.3509 0.0209651 *
## sq_vignes   -1.2601e-03 3.3429e-04 -3.7694 0.0002956 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Avant d'interpréter les résultats du modèle, il est nécessaire de vérifier la colinéarité entre les variables ajoutées.

```
vif(mco_reset_2)
```

```
##      log_pbs      log_sau      cereales      vignes      sq_log_sau      sq_cereales
##      15.56609      19.04092      21.15698      17.69592      26.42092      19.64957
##      sq_vignes
##      15.37564
```

La colinéarité trop haute entre les variables augmente considérablement le risque d'overfitting. Ainsi, nous faisons le choix de ne pas inclure de variables au carré dans notre modèle, malgré la présence de relations non-linéaires. Ainsi, le modèle estimé sera davantage adaptable à d'autres bases de données car moins afin de présenter un risque de surajustement.

4.3.5 H5 : Variable X non constante

```
apply(data_numeric, 2, var)
```

```
##      glyphosate      vignes      cereales      pbs      sau
##      4.753154e+09      1.057573e+02      5.840983e+02      2.704278e+11      2.048628e+10
##      log_pbs      log_sau      log_glyphosate
##      2.511828e+00      3.829113e+00      1.617348e+00
```

D'après ces résultats, aucune des variables n'est constante (variances différentes de 0).

5 Modèle final : interprétations et performance

```
##
## =====
##                      Dependent variable:
##                      -----
##                      log_glyphosate
##                      -----
## log_pbs              0.927***
##                      (0.184)
##
## log_sau              -0.349**
##                      (0.168)
##
## cereales             0.020***
##                      (0.004)
##
## vignes              0.016**
##                      (0.008)
##
## Constant            2.284*
##                      (1.261)
##
## -----
## Observations          96
## R2                    0.739
## Adjusted R2           0.728
## Residual Std. Error   0.663 (df = 91)
## F Statistic           64.542*** (df = 4; 91)
```

```
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

L'équation estimée est la suivante : $\widehat{\log_glyphosate} = 2.284 - 0.349 * \log_sau + 0.927 * \log_pbs + 0.020 * cereales + 0.016 * vignes$

5.1 Interprétation des paramètres

Lorsque tout les paramètres sont égaux à 0, le $\log_glyphosate$ est de 2.284.

Lorsque la pbs augmente de 1%, la quantité de glyphosate achetée augmente de 0.927%.

Lorsque la sau augmente de 1%, la quantité de glyphosate achetée baisse de 0.349%.

Lorsque la part de céréales dans la sau augmente de 1%, la quantité de glyphosate achetée augmente de 0.02%.

Lorsque la part de vignes dans la sau augmente de 1%, la quantité de glyphosate achetée 0.016%.

Tout ces paramètres sont significatifs au seuil de 5%, sauf la constante qui est significative au seuil de 10%.

La part de la variance expliquée par notre modèle est de 73.9%, et celui-ci est globalement significatif.

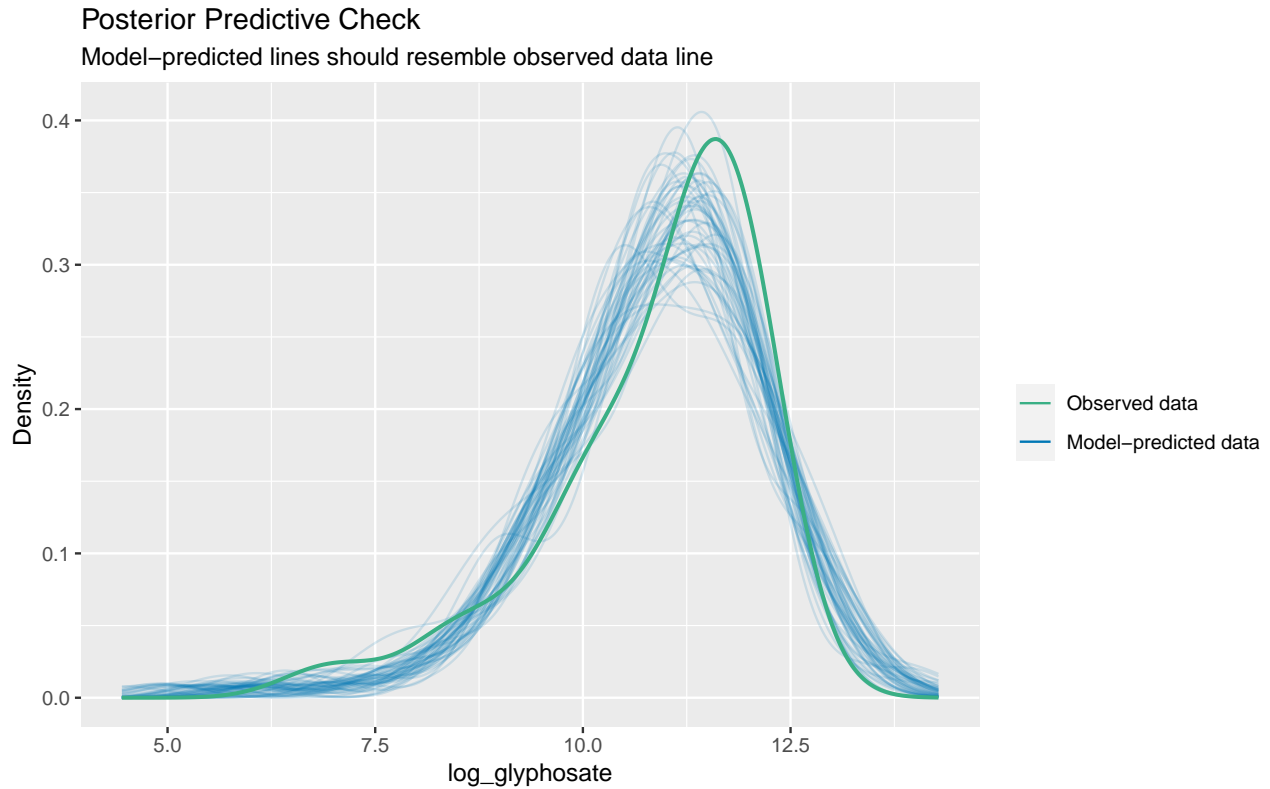
5.2 Performance du modèle

```
## # Check for Multicollinearity
##
## Low Correlation
##
##      Term  VIF      VIF 95% CI Increased SE Tolerance Tolerance 95% CI
## cereales 1.17 [1.04,  1.73]          1.08      0.86      [0.58, 0.96]
## vignes   1.42 [1.19,  1.94]          1.19      0.70      [0.51, 0.84]
##
## Moderate Correlation
##
##      Term  VIF      VIF 95% CI Increased SE Tolerance Tolerance 95% CI
## log_pbs  9.54 [6.78, 13.63]          3.09      0.10      [0.07, 0.15]
## log_sau  9.16 [6.51, 13.07]          3.03      0.11      [0.08, 0.15]
```

Aucune colinéarité extrême n'est détectée dans notre modèle, ce qui permet d'écarter le risque de surajustement.

```
## 2 outliers detected: cases 76, 93.
## - Based on the following method and threshold: cook (0.9).
## - For variable: (Whole model).
```

Les deux outliers détectés sont Paris et les hauts-de-Seine. Paris et son agglomération ne possèdent pas de cultures significatives. Ainsi, cette justification explique ces outliers. De ce fait, nous avons décidé de les garder dans notre base de données.



D'après ce graphique de prédictions, les estimations de notre modèle se rapprochent des valeurs observées. Notre modèle est donc performant.

```
## [1] FALSE
```

D'après le test ci-dessus, notre modèle n'est pas singulier (aucune des dimensions de la matrice des variance-covariance n'est nulle). Ainsi, le risque de surajustement et le risque de problèmes numériques et de mauvaises convergences sont moins élevés pour notre modèle.

6 Conclusion

Conclusion

Notre étude a permis de développer un modèle économétrique robuste visant à estimer les quantités de glyphosate achetées en France, en fonction de divers facteurs agricoles significatifs: la surface agricole utilisée, la production brute standardisée, ainsi que les parts respectives de cultures de vignes et de céréales dans la surface agricole utilisable, ces données étant agrégées par département.

Le modèle a démontré une capacité à générer des estimations qui se rapprochent significativement des valeurs observées, comme le montre notre test de prédictions. Cela indique une bonne adéquation du modèle avec les données réelles, renforçant ainsi sa validité pour des applications futures dans des études similaires ou des politiques agricoles.

L'approche retenue a révélé quelques défis, notamment la présence d'hétéroscédasticité, que nous avons réussi à corriger, et la mesure imparfaite de certaines variables, comme la surface agricole utilisable (SAU) dont le signe du paramètre qui n'était pas conforme à nos intuitions de départ malgré sa significativité. Néanmoins, le modèle reste exempt de colinéarité extrême et est statistiquement non singulier, ce qui renforce la fiabilité des coefficients estimés.

Les restrictions du modèle, telles que l'impossibilité d'ajouter des variables au carré en raison de la multicollinéarité élevée (VIF), ont été abordées par des méthodes adaptatives et prudentes, garantissant ainsi que

le modèle reste robuste et pertinent malgré ces contraintes.

En conclusion, ce projet a non seulement permis de comprendre mieux les dynamiques de vente du glyphosate en relation avec l'utilisation des terres agricoles en France, mais il offre également une base solide pour des recherches futures. Ces recherches pourraient explorer des variables additionnelles ou affiner davantage les méthodes de correction des erreurs de mesure pour améliorer la précision des prévisions.