

# Les déterminants naturels et politiques des AOC viticoles de Côte-d'Or

*Version révisée n° 2*

## Résumé

Cet article présente la construction d'une base de données parcellaire afin de décrire les relations entre les caractéristiques naturelles (topographie, géologie, pédologie) et la hiérarchie des appellations d'origine contrôlée (AOC) viticoles de Côte-d'Or en Bourgogne. Les données constituées comptent 67 variables systématiquement renseignées pour plus de 110000 parcelles situées dans les 31 communes qui forment la côte de Beaune et la côte de Nuits. L'utilisation de ces données est illustrée par une analyse statistique qui met en évidence des relations significatives entre les caractéristiques naturelles des parcelles et leur position dans la hiérarchie des AOC. Ces relations apparaissent toutefois différenciées selon la commune d'appartenance des parcelles. Nous interprétons ce résultat comme la preuve de déterminants politiques aux AOC, en lien avec le lobbying qui a eu lieu lors de leur mise en place. Les données et prédictions obtenues avec ou sans correction des effets politiques sont consultables via une cartographie interactive également présentée.

**Mots clés** : Économie spatiale ; vin ; géohistoire ; système d'information géographique ; économétrie.

## 1. Introduction

Les appellations d'origine contrôlée (AOC) viticoles de Bourgogne sont issues de processus historiques complexes au cours desquels les parcelles ont été classifiées selon leurs caractéristiques biophysiques et selon les rapports économiques, politiques et sociaux en vigueur (Garcia, 2011 ; Wolikow and Jacquet, 2011). La classification actuellement utilisée pour désigner les vins de cette région est issue de plusieurs siècles de culture de la vigne, de production de vin et de négociation sur les dénominations. Ces trois ensembles de pratiques forment les usages loyaux et constants selon la doctrine de l'institut national de l'origine et de la qualité (INAO) pour définir, reconnaître et gérer les AOC (Capus, 1947 ; Humbert, 2011). La complexité des facteurs pris en compte par la référence au lieu de production est à la fois une force et une faiblesse pour le système des AOC. Ces dernières permettent en effet de simplifier les nombreux déterminants de la qualité des vins au prix d'une perte d'information et donc d'une certaine opacité pour les acteurs du marché.

La référence au lieu de production permet de donner une indication composite sur la qualité des vins lors des échanges. La question de la nature de l'information contenue dans les AOC se pose alors, en particulier la distinction de la part relative aux processus naturels de la part relative aux processus humains (Dion, 1952). Cette distinction est déterminante pour identifier les facteurs immobiles et non-reproductibles qui justifient pleinement la référence au lieu de production. La question de l'origine de la qualité des vins – dont la définition est fondamentalement subjective – a fait l'objet d'une littérature abondante avec une diversité remarquable des approches. Sans prétendre à l'exhaustivité, nous pouvons mentionner des publications scientifiques récentes que ce soit dans le domaine des sciences naturelles comme l'agronomie (van Leeuwen, Roby et de Rességuier, 2018 ; Abbal, Sablayrolles, Matzner-Lober et Carbonneau, 2019), la chimie (Polàskova, Herszage et Ebeler, 2008 ; Roullier-Gall, Boutegrabet, Gougeon et Schmitt-Kopplin, 2014), la microbiologie (Bokulich, Thorngate, Richardson et Mills, 2014 ; Knight, Klaere, Fedrizzi et Goddard, 2015) ou dans le domaine des sciences humaines avec l'anthropologie (Barrey et Teil, 2011 ; Demossier, 2011), l'histoire (Jacquet et Laferté, 2006 ; Labbé, 2011), la science politique (Colman, 2008 ; Carter, 2018), la sociologie (Karpik, 2007 ; Beckert, Rössel, et Schenk, 2017), la géographie (Vaudour, 2002 ; Tiefenbacher, 2013) et l'économie (Gergaud et Ginsburgh, 2008 ; Ashenfelter et Storchmann, 2010). Nous pouvons également citer deux des nombreux ouvrages destinés à un public plus large qui ont pour but d'informer les amateurs sur la qualité des vins (Bazin, 2015 ; Parker, 2015).

Les approches quantitatives tiennent une place de choix parmi les travaux scientifiques, comme illustré par Jones et al. (2004) qui effectuent à l'échelle régionale des inventaires de terrains propices au développement ou à l'implantation d'une viticulture de qualité dans l'Oregon en se basent quasi-exclusivement sur les facteurs biophysiques. Ce travail de recherche identifie de manière systématique les aires géographiques offrant une combinaison idéale entre la variété du raisin et le climat, l'altitude, la pente, l'orientation et le type de sol. Nous présentons dans cet article la constitution de données exhaustives similaires et l'estimation d'un modèle statistique qui permet d'analyser les déterminants des AOC viticoles du département de la Côte-d'Or. Dans la continuité des travaux

précédents, nous montrons une forte dépendance des AOC aux caractéristiques biophysiques des parcelles. De manière plus originale, nous montrons la présence d'une hiérarchie implicite entre les communes de la zone, que nous interprétons comme issue des arbitrages politiques qui ont eu lieu lors de la désignation des AOC. En effet, c'est à l'échelle communale que les syndicats de producteurs et les négociants ont fait valoir leurs intérêts dans la mise en place des appellations (Jacquet, 2009 ; Lucand, 2011).

La Section 2 présente la construction de la base de données géographique disponible sous licence GNU GPL V3 sur le serveur <https://data.inrae.fr/>. La parcelle cadastrale est l'unité élémentaire d'observation afin de permettre l'appariement des variables sur les AOC actuelles (produites par l'INAO), sur les AOC de 1936 (produites par la MSH de Dijon), sur les lieux-dits par le Plan Cadastral Informatisé (produit par la DGFIP), sur l'altimétrie par le RGE ALTI à 5 mètres (produit par l'IGN), sur l'occupation du sol (produite par Hilal et al., 2018), sur la géologie par Charm-50 (produit par le BRGM) et sur la pédologie par le Référentiel Pédologique de Bourgogne (produit par le Gis Sol). Les données ainsi constituées concernent l'ensemble des parcelles des 31 communes incluses dans la côte de Beaune et la côte de Nuits, soient l'ensemble des vignobles du département de la Côte-d'Or à l'exception des hautes côtes et du Châtillonnais. Cette base de données permet de relier finement les AOC aux caractéristiques biophysiques des parcelles dont les vins sont issus, et possède ainsi une utilisation plus large que celle illustrée ici.

Nous ne considérons pas les variables climatiques car il n'existe pas à notre connaissance de données disponibles à une résolution suffisamment fine pour être utilisée au niveau des parcelles. Par exemple, les données historiques de Météo-France sont disponibles à la maille SAFRAN pour une résolution de 8 km alors que la zone d'étude fait au plus 5 km de large (Figure 1). Madelin et al. (2010) utilisent les observations des stations climatiques pour étudier la dynamique du climat, mais sans qu'elles fournissent suffisamment de points pour une interpolation systématique au niveau des parcelles. Bois et al. (2020) utilisent les mesures du réseau Climeo du Bureau Interprofessionnel des Vins de Bourgogne mais limitent leur application à quelques communes au nord de Beaune. De plus, la topographie est typiquement utilisée pour interpoler les observations des stations climatiques afin d'augmenter leur résolution (Joly et al., 2010). Sachant que la topographie est déjà incluse dans les données à une échelle très fine, de telles interpolations seraient redondantes. Nous pouvons ainsi considérer que les effets du climat dans la désignation des AOC sont pris en compte par les variables topographiques. Nos données caractérisent les Climats au sens bourguignon du terme, par les caractéristiques géologiques, hydrométriques et d'exposition particulières telles que reconnues par l'UNESCO comme patrimoine mondial en 2015.

La Section 3 présente l'estimation du modèle statistique dont les spécifications sont décrites plus extensivement dans un article associé (Ay, 2020). Le principe est d'utiliser la structure hiérarchique des AOC (Coteaux bourguignons < Bourgogne régional < Villages < Premiers crus < Grands crus) pour la relier aux caractéristiques biophysiques des parcelles par une variable latente théorique de qualité des vignes. Nous montrons que cette variable, non observable par définition, peut être estimée de manière semi-paramétrique avec peu d'*a priori* sur la structure des effets des variables et sur l'observation de la qualité des vins. La mesure de la qualité que nous estimons est simplement révélée à partir de la hiérarchie des AOC actuelles. Nous utilisons pour cela des modèles ordonnés additifs généralisés

(OGAM pour *ordered generalized additive model*, Wood et al., 2016) qui présentent de très bonnes capacités de prédiction (avec seulement 10% d'erreurs). Ils permettent également d'estimer des effets communaux à partir des écarts au modèle où les variables naturelles seraient les seuls déterminants de la hiérarchie des AOC. La méthode proposée permet en outre de corriger ces effets dits « de lobbying » afin de reconstruire des variations de qualité des vignes qui ne proviennent que de leurs seules caractéristiques naturelles.

La Section 4 présente l'utilisation de l'outil de visualisation *Shiny* (Chang et al., 2019) que nous avons développé pour consulter les données produites et les résultats de la modélisation statistique. L'utilisateur peut ainsi saisir les informations typiquement disponibles sur une étiquette de vin de Bourgogne (le niveau de l'AOC dans la hiérarchie, la commune de production et le lieu-dit de la parcelle) pour identifier géographiquement l'ensemble des parcelles dont est issu le vin et le niveau de qualité prédit (avec ou sans correction des effets communaux). Cette information permet une évaluation plus précise de la qualité potentielle du vin sans pour autant introduire de facteurs subjectifs exogènes. Cela permet d'améliorer l'information disponible pour les acteurs du marché à partir d'informations déjà présentes sur les étiquettes des bouteilles. Chaque vin identifié peut être comparé aux autres vins du même niveau hiérarchique ou aux vins d'autres niveaux hiérarchiques afin d'évaluer sa qualité relative.

Les codes R (R Core Team, 2019) utilisés sont fournis dans le matériel additionnel associé pour reproduire l'ensemble des tables et des figures présentées à partir des données disponibles sur le serveur <https://data.inrae.fr/>. L'intégralité du code qui permet de construire l'application *Shiny* est également présenté dans le matériel additionnel, l'application peut ainsi être lancée localement, voire modifiée par les utilisateurs. La dernière version officielle est accessible sur le répertoire <https://github.com/jsay/geoInd>.

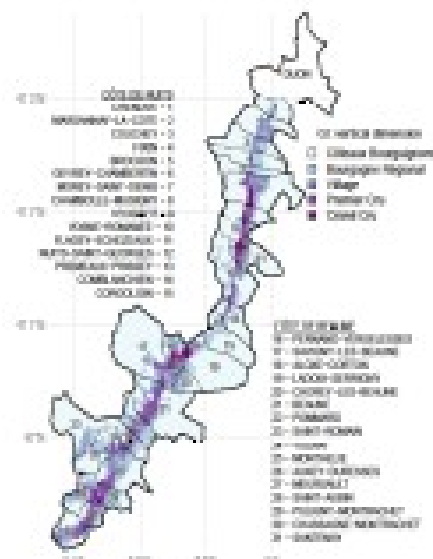
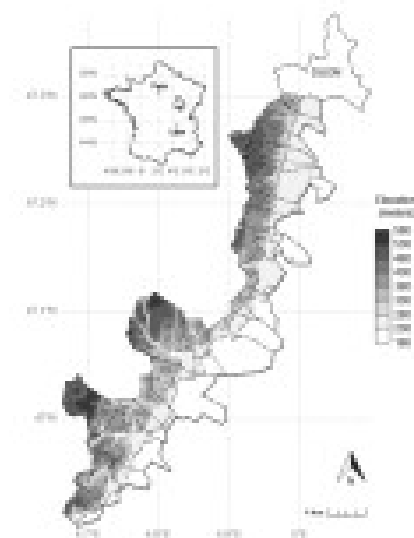


Figure 1 : La zone de la Côte-d'Or étudiée, sa topographie et ses appellations d'origine contrôlées

Sources : INRAE / INAO / MSH / DGFIP / IGN / BRGM / Gis Sol.

Lecture : La carte de gauche représente l'altitude des parcelles, catégorisée en 8 classes de 50 mètres d'amplitude. Située la ville de Dijon au Nord de la zone, elle permet de faire apparaître la topographie de la Côte viticole majoritairement orientée à l'Est mais dont la présence de vallées sèches (appelées combes) produit des variations localisées. La carte de droite présente les 31 communes de la zone, qui constituent la dimension politique des AOC viticoles. La dimension hiérarchique des AOC est représentée par les 5 niveaux reportée sur cette même carte. Ces deux cartes sont reproductibles à partir des données présentées dans cet article, les codes utilisés sont consultables à <https://github.com/jsay/geoInd>.

## 2. Construction des données

L'unité géographique de base est la parcelle cadastrale, au sein des 31 communes du périmètre présenté dans la Figure 1. Leur géométrie est issue de la BD parcellaire de l'IGN dans sa version 2014 (téléchargement le 09/10/2015). Nous l'avons enrichie de variables décrivant la géométrie des parcelles avec l'ajout de la surface, du périmètre et de la distance maximale entre deux sommets pour chaque polygone cadastral (Conrad et al., 2015).

### 2.1. Les AOC actuelles

Les polygones cadastraux ont ensuite été appariés par jointure géographique aux délimitations parcellaires des AOC viticoles produites par l'INAO, disponibles à l'adresse <https://www.data.gouv.fr/fr/datasets/delimitation-parcellaire-des-aoc-viticoles-de-linao> sous licence ouverte (téléchargement le 21/08/18). Selon le principe des replis, les parcelles d'un niveau hiérarchique supérieur peuvent toujours être revendiquées dans un niveau inférieur. Les couches de l'INAO contenant l'ensemble des AOC revendiquable pour une parcelle, leur superposition conduit à la présence de plusieurs niveaux d'AOC sur une même parcelle, ce qui est en contradiction avec l'interdiction de revendiquer des AOC différentes pour un même produit. Dans les faits, les producteurs revendiquent l'AOC maximale à laquelle la parcelle peut prétendre. La variable AOC que nous avons créée représente cette dernière valeur, codée 0 pour les parcelles hors AOC, 1 pour les Coteaux bourguignons, 2 pour les Bourgognes régionaux et jusqu'à 5 pour les Grands crus. Les informations présentes sur les étiquettes peuvent correspondre soit à des AOC soit à des dénominations géographiques complémentaires (le plus souvent sans que cette distinction soit claire pour le consommateur). Cette différence est codée dans la variable AOCTp.

### 2.2. Enrichissement des AOC historiques

Des variables sur les classifications historiques des parcelles, obtenues auprès de la Maison des Sciences de l'Homme de Dijon sous formes de cartes numérisées, sont également appariées au parcellaire. Alors que l'INAO a été créé en 1936, la première délimitation officielle des AOC sur la côte s'est opérée entre 1936 et 1940. Elle fut basée sur deux classements antérieurs : celui de Jules Lavalley de 1855 (Lavalley, 1855) et celui du Comité d'Agriculture et de Viticulture de l'Arrondissement de Beaune de 1860 (Wolikow and Jacquet, 2011). Nous compilons les différentes classifications qui ont eu lieu entre 1936 et 1940 pour obtenir une hiérarchie des parcelles en 3 niveaux : Régional < Village < Grand cru, que nous considérons comme les niveaux d'AOC en 1936. Cette classification historique est moins détaillée que l'actuelle (3 niveaux au lieu de 5) car l'AOC Coteaux bourguignons

n'existait pas encore (les niveaux ordinaires et grands ordinaires qui la précédèrent n'étaient pas délimités) et les Premiers crus ont seulement été instaurés par décret en 1943. Les AOC représentaient 27 % des parcelles de la zone au lieu de 55 % actuellement. Près de 165000 parcelles hors AOC en 1936 le sont actuellement (tous niveaux confondus) alors que seulement 2610 parcelles sont dans le cas inverse de déclassement (ce sont des parcelles construites). La majorité des parcelles classées en niveaux Village et Grand cru l'étaient déjà en 1936, les Premiers Crus actuels étaient principalement en Village, et les Coteaux bourguignons et les Bourgogne niveau régional d'aujourd'hui étaient hors AOC.

### 2.3. Enrichissement des lieux-dits

Les lieux-dits (géométrie et toponymie) sont disponibles dans le Plan Cadastral Informatisé (DGFIP) téléchargeable à l'adresse <https://cadastre.data.gouv.fr/datasets/plan-cadastral-informatise>. Le téléchargement pour la Côte-d'Or (21) date du 13/01/2019. Ces données, en licence ouverte Etalab, nous permettent de renseigner les lieux-dits des parcelles viticoles et certaines variables communales agrégées. Une attention particulière est portée sur les lieux-dits dont les intitulés doivent être croisés avec le nom des communes pour être uniques (un même lieu-dit toponymique peut être présent sur plusieurs communes).

Nous obtenons une distance moyenne de 1,2 km avec le chef-lieu de chaque commune, ce qui est cohérent avec la taille des communes du périmètre d'étude (environ 2,5 km<sup>2</sup>). Notons que 6426 parcelles de la BD parcellaire sont absentes du Plan Cadastral Informatisé. Elles correspondent à environ 4 % de la base initiale et n'ont donc pas été appariées (des valeurs omises sont reportées pour les variables associées aux lieux-dits). Ces parcelles sont pour la plupart hors AOC et se concentrent sur les communes les plus urbanisées, telles que Chenôve, Marsannay-la-Côte et Beaune. Ces valeurs manquantes semblent donc correspondre à des espaces bâtis qui ne peuvent pas être revendiqués en AOC. Des vérifications manuelles n'ont pas suffi pour statuer définitivement sur ce point, qui n'est donc pas univoque. Ces valeurs manquantes seront exclues de l'analyse statistique, mais elles ne sont pas décisives pour les résultats issus des estimations.

### 2.4. Enrichissement de la topographie

Les informations de la couche cadastrale sont enrichies de variables topographiques issues d'un modèle numérique de terrain (MNT RGE ALTI 5 m, sous licence IGN "Recherche") et d'une couche d'occupation du sol (MOS) provenant du modèle développé par Hilal et al. (2018) et disponible en téléchargement sur <https://data.inrae.fr/>, voir Hilal (2020). Le MOS est construit à partir de la BD TOPO, du registre parcellaire graphique (RPG, Agence de services et de paiement) et de Corine Land Cover (Agence européenne de l'environnement). Ces deux couches raster sont à une résolution de 5 m. Les informations altimétriques du MNT permettent de produire 3 couches raster supplémentaires, toujours à 5 m de résolution : la pente, l'exposition et les radiations solaires. Ces attributs sont calculés en utilisant le logiciel ArcGis (Rich and Fu, 2000). Les 5 couches raster ainsi constituées (altitude, pente, exposition, radiation solaire et occupation du sol) sont converties en fichiers au format XYZ, avec X et Y les coordonnées Lambert 93 du centre de chaque pixel et Z la variable d'intérêt de chacune des couches. Les fichiers sont regroupés dans une table XYZ contenant un seul couple XY et les 5 attributs Z issus des couches raster. Cette table est



ensuite appariée avec une autre table XYZ produite après rastérisation à 5 m des parcelles cadastrales en vue de récupérer le couple XY du centroïde de chaque pixel et Z l'identifiant PAR2RAS qui reprend l'identifiant cadastral (IDU) de la parcelle dans laquelle se situe le pixel. L'identifiant PAR2RAS servira à l'appariement des parcelles avec les 5 variables topographiques. Le résultat est une base contenant plus de 14 millions de lignes, une pour chaque pixel de 5 m. Les informations ainsi disponibles permettent de calculer à l'échelle des parcelles cadastrales des fréquences décrivant l'occupation non agricole du sol (urbain, forêt, eau) et des valeurs moyennes pour les variables dérivées de l'altimétrie.

## 2.5. Enrichissement de la géologie

Les données géologiques sont extraites de la BD harmonisée Charm-50 produite par le BRGM à l'échelle du 1/50 000. Cette base est disponible sur le site <http://infoterre.brgm.fr> sous licence Ouverte. L'extraction contient les formations géologiques, nommée GE0050K\_HARM\_021\_S\_FGE0L\_CGH\_2154, téléchargées le 25/04/2019 pour le département de la Côte-d'Or. L'appariement est réalisé par intersection des centroïdes des parcelles avec les polygones géologiques. La faible taille des parcelles permet de s'assurer de la validité de cette procédure, vérifiée manuellement par ailleurs. Le dictionnaire associé aux 16 variables sur la géologie est disponible dans la Table 1 en Annexe 3. La description des variables est peu précise actuellement car les données du BRGM sont disponibles depuis peu et ne possèdent pas encore (à notre connaissance) de dictionnaire exploitable. Ce manque de précision n'est pas limitant pour l'analyse statistique qui suit car ces variables géologiques seront utilisées sous forme d'indicatrices. Cela peut néanmoins être différent pour d'autres utilisations de la base de données. Pour diminuer la multicollinéarité lors de l'utilisation statistique de ces notations géologiques et nous assurer d'estimations précises, les catégories géologiques qui comptent moins de 1000 parcelles sont regroupées dans une modalité de référence codée 0AREF. Il reste ainsi les 17 catégories présentées en Annexe qui pourront être utilisées dans la modélisation.

## 2.6. Enrichissement de la pédologie

Les données pédologiques utilisées sont extraites du Référentiel Pédologique de Bourgogne : "Régions naturelles, pédopaysages et sols de Côte-d'Or à l'échelle 1/250 000" (étude 25021 dans le référentiel Gis Sol). Ces données sont compatibles avec la référence nationale DoneSol et correspondent à la meilleure information pédologique actuellement disponible systématiquement sur le périmètre d'étude. La localisation des types de sol et l'appariement avec le parcellaire cadastral s'opèrent par le biais des 194 unités cartographiques de sols (UCS) qui composent le périmètre d'étude. Les UCS sont des polygones construits pour être homogènes en termes de pédopaysages (topographie, climat, géologie). Elles sont typiquement utilisées pour cartographier les caractéristiques des sols, mais peuvent néanmoins contenir différents types de sols. Ces derniers, regroupés en unités typologiques, ne peuvent pas être localisés plus précisément que les unités cartographiques. Cette imprécision dans la localisation des données est une limite importante pour leur usage statistique à l'échelle parcellaire (Ay, 2011). En l'absence de données spatialement plus précises, les données parcellaires du cadastre sont enrichies du libellé de l'UCS et des 11 variables correspondantes à l'unité typologique de sol dominante,

c'est-à-dire celle qui est la plus étendue au sein de chaque UCS. Ce choix à première vue arbitraire ne change pas les résultats obtenus. Comme pour les variables sur la géologie, les variables pédologiques seront intégrées dans les modèles statistiques par des indicatrices, qui correspondent ici aux UCS. Le détail des 11 variables pédologiques est maintenu dans les données constituées pour ne pas limiter les autres usages qui peuvent en être faits. Les libellés des unités cartographiques, reportés dans la variable DESCRp, ont été saisis manuellement à partir du site <https://bourgogne.websol.fr/carto>.

## 2.7. Statistiques descriptives

Nous proposons donc une base de données qui contient 59 113 observations qui seront utilisables pour estimer le modèle statistique de désignation des AOC. Ce nombre provient de plusieurs sélections présentées en matériel additionnel : le principal critère concerne les parcelles ayant au moins une AOC. Cette opération en exclue 49 717 pour n'en conserver que 60 632. Le second critère enlève les observations avec des valeurs manquantes pour au moins une des variables qui ont été construites. Ce critère écarte 1 519 parcelles. Comme indiqué dans la Table 1, les parcelles restantes ont des surfaces faibles (0,2 ha de moyenne), des altitudes comprises entre 200 et 500 m (286 m de moyenne), des pentes entre 0 et 37 degrés (5,75 degrés de moyenne) et des radiations solaires comprises entre 581000 et 1,2 millions de Joules (1 millions de Joules en moyenne). Nous observons également que le niveau village de la hiérarchie des AOC regroupe 42 % des parcelles, les niveaux régionaux et coteaux bourguignons respectivement 23 % et 16,5 %, alors que les niveaux premier et grand cru respectivement 15 % et 3 %. Les vignobles sont globalement orientés à l'Est, avec 55 % des observations qui ont une orientation comprise entre 45 et 135 degrés.

Table 1 : Statistiques descriptives des variables utilisées.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Surface [1000 m2]	59 113	1.908	3.399	0.000	0.517	2.178	177.200
Altitude [1000 m]	59 113	0.286	0.056	0.210	0.241	0.319	0.505
Pente [degrés]	59 113	5.772	5.478	0.000	1.556	8.747	36.970
Radiation solaire [millions J]	59 113	1.060	0.049	0.581	1.048	1.076	1.230
Longitude [degrés]	59 113	4.837	0.104	4.665	4.740	4.955	5.003
Latitude [degrés]	59 113	47.060	0.110	46.900	46.980	47.170	47.300
Niveau AOC Coteaux	59 113	0.164	0.370	0	0	0	1
Niveau AOC Régional	59 113	0.229	0.420	0	0	0	1
Niveau AOC Village	59 113	0.428	0.495	0	0	1	1
Niveau AOC Premier Cru	59 113	0.147	0.354	0	0	0	1
Niveau AOC Grand Cru	59 113	0.032	0.177	0	0	0	1
Exposition [0-45]	59 113	0.046	0.210	0	0	0	1
Exposition [45-90]	59 113	0.186	0.389	0	0	0	1
Exposition [90-135]	59 113	0.362	0.481	0	0	1	1
Exposition [135-180]	59 113	0.212	0.409	0	0	0	1
Exposition [180-225]	59 113	0.100	0.300	0	0	0	1
Exposition [225-270]	59 113	0.044	0.206	0	0	0	1
Exposition [270-315]	59 113	0.030	0.170	0	0	0	1
Exposition [315-360]	59 113	0.021	0.142	0	0	0	1

Sources : INRAE / INAO / MSH / DGFIP / IGN / BRGM / Gis Sol.

Notes : L'échantillon est composé de 59 113 parcelles (pas nécessairement en vignes) sous AOC dans la zone des 31 communes considérées. Le tableau reporte l'effectif, la moyenne, l'écart-type, le minimum, les quartiles et le maximum pour les variables biophysiques issues du processus d'enrichissement présenté dans le texte.



### 3. Modèle statistique

Le modèle statistique a pour but d'étudier la relation entre la hiérarchie AOC des parcelles, leurs caractéristiques biophysiques (topographie, géologie, pédologie) et leur commune d'appartenance. Cette modélisation fournit des prédictions qui permettent de préciser la classification sous-jacente aux AOC en positionnant chaque parcelle sur une échelle de qualité continue entre 0 et 100 (après normalisation de la variable latente prédite).

#### 3.1. Estimation du modèle

Le modèle utilisé et le processus de spécification sont tirés d'un article associé (Ay, 2020). Il s'agit d'estimer un modèle ordonné additivement semi-paramétrique (OGAM) qui prend en compte la structure hiérarchique des AOC de la zone, notée  $y \in \{1,2,3,4,5\}$  par ordre croissant des niveaux d'AOC (Coteaux bourguignons < Bourgogne régional < Villages < Premiers crus < Grands crus). Les désignations AOC sont supposées suivre une règle de décision basée sur une variable latente de qualité non observable qui franchit des seuils différents selon la commune d'appartenance. Notons  $X_i$  le vecteur des caractéristiques biophysiques de la parcelle de vigne  $i$  (avec  $i=1, \dots, N$ ) et  $C_i$  le vecteur de dimension 31 qui a pour élément générique  $c_{ih}$  égal à 1 si la parcelle  $i$  se situe dans la *commune*  $h$  et 0 sinon. L'hypothèse d'une distribution logistique de la partie aléatoire de la variable latente produit un modèle de logit ordonné classique (Agresti and Kateri, 2017) :

$$\text{Prob}(y_i > j | X_i, C_i) = \Lambda \left[ B(X_i)^\top \beta + C_i^\top \mu - \alpha_j \right],$$

où  $\Lambda$  est la fonction cumulative de la loi logistique. Les déterminants politiques qui ont impactés la classification AOC au cours de l'histoire sont représentés par le vecteur des effets fixes communaux noté  $\mu$ . En l'absence d'*a priori* théorique sur l'effet de chaque variable biophysique  $X_i$ , nous les spécifions au travers d'une série de transformations additives *B-splines* que nous notons  $B(\cdot)$  avec  $\beta$  le vecteur des coefficients associés. Ce modèle de désignation peut alors être estimé avec la fonction `gam` du package `mgcv` comme décrit dans Wood et al. (2016). Le manuel d'utilisation de l'auteur du package (Wood, 2017) contient de nombreux détails méthodologiques sur le processus de pénalisation semi-paramétrique des variables biophysiques continues. Les codes R utilisés pour estimer différentes versions de ce modèle sont reportées dans le matériel additionnel.

Un point important de la spécification du modèle est l'utilisation des coordonnées géographiques des centroïdes des parcelles de vigne (conjointement aux variables biophysiques) pour contrôler les effets de terroir qui ne seraient pas pris en compte par les variables. Ces effets peuvent apparaître à cause de certaines données non disponibles (sur la micro-biologie des sols ou les variations locales du climat par exemple) ou à cause d'erreurs de mesure dans les variables disponibles (par la présence de variations topographiques très fines ou d'effets d'ombrage). L'inclusion des coordonnées géographiques comme variables explicatives avec un lissage en deux dimensions permet d'améliorer sensiblement les capacités prédictives du modèle et de proposer une estimation non biaisée des effets communaux. Nous estimons des modèles OGAM à des degrés divers d'ajustement des fonctions splines, bien que nous ne reportons dans cet

article uniquement les résultats du modèle présentant les meilleures prédictions. Sur la base des données que nous avons constituées, le lecteur est libre d'estimer une version différente du modèle avec, par exemple, des degrés plus forts de lissage. Notons que les effets politiques estimés par le modèle (via les coefficients  $\mu$ ) sont déduites des désignations AOC et ne nécessitent aucun *a priori* sur le prestige relatif des communes de la zone. La principale considération statistique qui assure la validité des résultats est l'absence de corrélation entre la partie inobservée du terroir biophysique et les délimitations communales, un élément que nous contrôlons par l'inclusion des coordonnées géographiques des centroïdes des parcelles dans la modélisation.

Notons que toutes les variables introduites dans le modèle sont significatives au seuil de 99 %, ce qui conforte notre hypothèse d'un modèle de désignation des AOC basé sur une variable latente de qualité des vignes. Dans ce modèle, qui s'ajuste très précisément aux données, les coordonnées géographiques apparaissent comme les variables explicatives les plus importantes au sens des tests du  $\chi^2$ , suivies des indicatrices communales, de l'altitude, du rayonnement solaire, de la géologie, de la pédologie, de la pente et enfin de l'exposition. Ce modèle avec un lissage spatial fort produit près de 90 % de bonnes prédictions des niveaux d'AOC pour un pseudo- $R^2$  (au sens de McFadden) égal à 0,76.

### 3.2. Effets des variables biophysiques

La Figure 2 présente les effets marginaux de chaque variable biophysique sur la variable latente de qualité des vignes, soient les fonctions  $B(\cdot)$  de l'équation (1). Des effets en U inversés sont obtenus pour l'altitude et la pente, avec les vignes les mieux classées en termes d'AOC qui sont situées à environ 300 mètres d'altitude et 10 degrés de pente. L'effet du rayonnement solaire est plus linéaire, contrairement à ce que le troisième quadrant de la Figure peut laisser apparaître (en bas à gauche). En effet, la plupart des parcelles sous AOC présentent un rayonnement solaire centré-réduit compris entre  $-2$  et  $2$ , soit la partie linéaire de la courbe représentée. Enfin, les effets spatiaux en bas à droite semblent se structurer dans une relation de centre/ périphérie par rapport aux altitudes intermédiaires de la zone. Des figures plus détaillées, qui contiennent en particulier les effets associés aux autres modèles plus lissés spatialement, sont reportées dans l'article associé (Ay, 2020). Le lecteur peut aussi reproduire ces effets pour d'autres modèles avec la fonction `plot` du package `mgcv`. La structure des effets reste cependant robuste à l'ajustement des effets spatiaux, elle reste proche de ce qui est obtenu ici pour le modèle avec  $k=900$ .

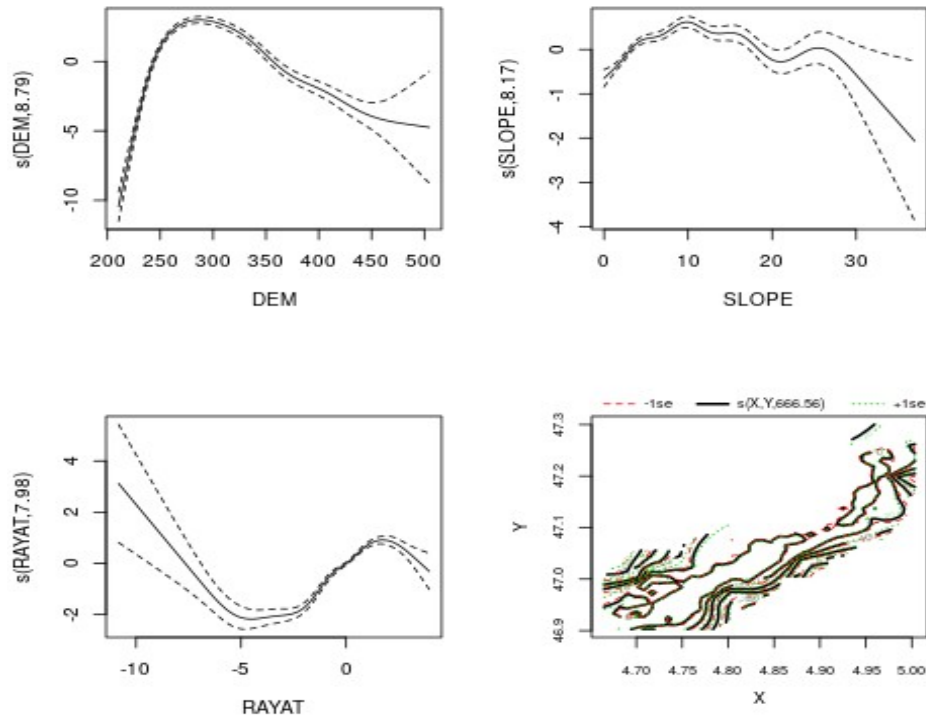


Figure 2 : Effets de la topographie et de la localisation sur la hiérarchie des vignes.

Sources : INRAE / INAO / MSH / DGFIP / IGN / BRGM / GisSol.

Lecture : Les trois premiers quadrants représentent les effets de l'altitude, de la pente et du rayonnement solaire en fixant toutes les autres variables explicatives du modèle à leurs moyennes de l'échantillon. Les effets ont une moyenne normalisée à 0 car leurs niveaux moyens ne sont pas identifiables semi-paramétriquement (Wood, 2017). La dernière fenêtre en bas à droite représente l'effet joint de la longitude et de la latitude par le biais de lignes de niveau et de leurs intervalles de confiance.

### 3.3. Effets communaux

Les coefficients associés aux effets communaux sont d'un intérêt particulier car ils correspondent à la partie des AOC expliquée par une délimitation administrative et non par des caractéristiques biophysiques, soit la partie politiquement déterminée des AOC. Cette interprétation des effets fixes communaux fait écho à certains travaux d'historiens pour lesquels nos résultats offrent une confirmation statistique. En effet, Lucand dans Wolikow and Jacquet (2011) évoque l'existence d'une hiérarchie implicite des communes comme des "identifications commerciales communes, investies d'un plus ou moins grand capital symbolique hérité. Ce capital symbolique hérité attribut un prestige plus ou moins grand à certaines communes ou propriétaires particuliers" (p. 68). Les effets fixes que nous estimons peuvent alors être vus comme des mesures de ce capital symbolique. De manière complémentaire, Jacquet (2009) étudie la structuration des syndicats de viticulteurs aux XIXe et XXe siècles, qui s'opère quasi-exclusivement à l'échelle communale et mentionne le fait que (p. 193) "plus l'appellation requise se calque sur le syndicat qui la défend, plus elle a de chance d'émerger et d'être délimitée strictement". Les effets fixes communaux peuvent donc également mesurer l'action des syndicats, qui apparaît ainsi avoir de l'inertie.

Pour faciliter l'interprétation des effets fixes communaux, nous traduisons les coefficients estimés en mesures de supériorités ordinales  $\gamma_A$  pour la commune  $A$  par rapport à la commune moyenne de la zone (Agresti and Kateri, 2017). Par définition,

$$\gamma_A = \Lambda \left[ \left( \mu_A - \bar{\mu} \right) / \sqrt{2} \right],$$

où  $\mu_A$  représente l'effet fixe de la commune  $A$  et  $\bar{\mu}$  la moyenne des effets fixes sur la zone d'intérêt. Ces mesures de supériorité ordinale, comprises entre  $-1$  et  $1$  pour chacune des communes de la zone, représentent l'écart de probabilité qu'une parcelle de la commune  $A$  soit mieux classée qu'une parcelle aux caractéristiques biophysiques identiques mais localisée dans une commune au hasard. Des valeurs positives indiquent des communes avantagées et des valeurs négatives des communes désavantagées par les désignations AOC.

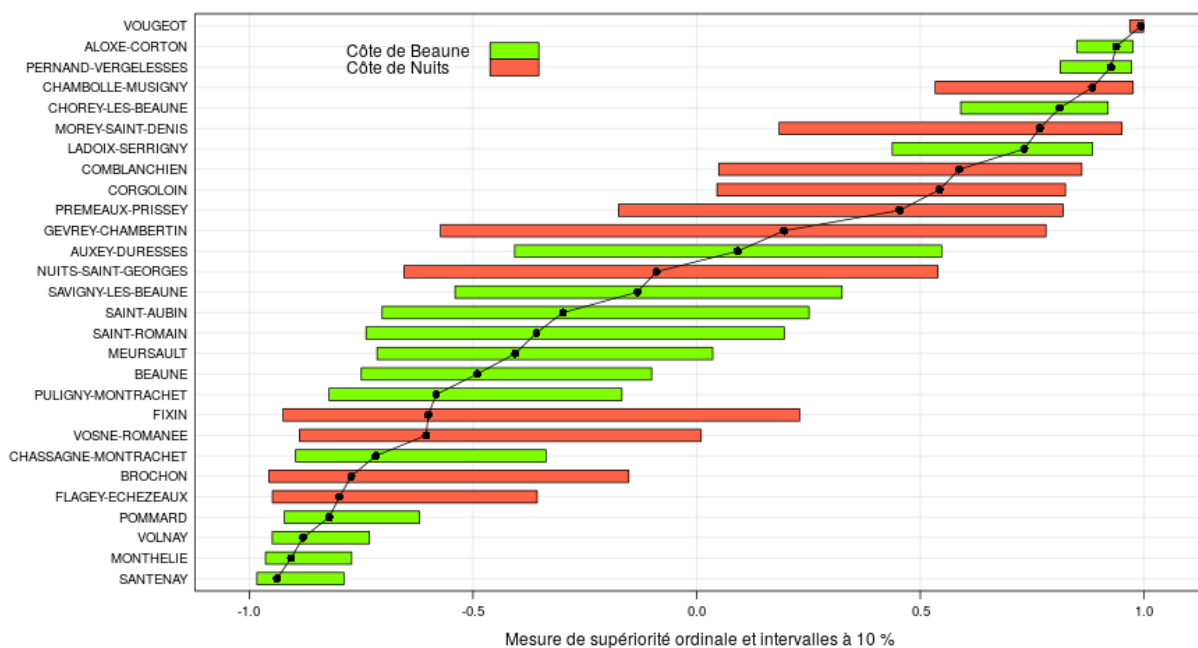


Figure 3 : Classification des communes selon les mesures de supériorité ordinale

Sources : INRA / INAO / MSH / DGFIP / IGN / BRGM / Gis Sol.

Lecture : Pour chacune des communes de la zone (en ordonnées), la figure reporte la mesure de supériorité ordinale de la désignation des AOC par rapport à la commune moyenne, c'est-à-dire l'écart de probabilité qu'une même parcelle soit mieux classée dans la commune considérée que dans une commune de la zone prise au hasard. Les intervalles de confiance représentent l'incertitude associée à l'estimation des effets fixes communaux dans la modélisation statistique (pour un intervalle de confiance fixé à 90 %).

Les communes relativement favorisées par la classification des AOC apparaissent en haut de la figure 3 et les communes relativement défavorisées en bas. Les intervalles de confiance qui encadrent les valeurs moyennes sont différents de ceux reportés dans Ay (2020), ils représentent ici l'incertitude associée à l'estimation des effets fixes communaux plutôt que l'incertitude associée à la spécification du lissage spatial. Les ordres de grandeur obtenus pour ces deux sources d'incertitude sont toutefois similaires. Nous observons que certaines mesures de supériorité ordinale suivent la hiérarchie des dotations brutes en AOC, où les communes privilégiées sont celles qui possèdent les plus grosses proportions

d'AOC en haut de la hiérarchie (Premiers crus et Grands crus). Mais cette relation n'est pas systématique, certaines communes peu dotées en hauts niveaux d'AOC apparaissent également privilégiées. Parmi les 5 communes les plus privilégiées par la classification AOC, les communes de Vougeot et d'Aloxe-Corton sont relativement bien dotées en Premiers et Grands Crus, alors que ce n'est pas le cas pour les communes de Pernand-Vergelesses et de Chorey-les-Beaunes. À l'inverse, Chassagne-Montrachet ou Vosne-Romanée possèdent de fortes proportions de Premiers et Grands Crus, sans que cela semble venir, au regard de nos résultats, d'un traitement politique préférentiel dans la classification.

### 3.4. Prédiction de la qualité continue

Les prédictions de la variable latente à partir des variables biophysiques et des coordonnées géographiques, soient  $\hat{y}_i = B(X_i)^T \hat{\beta}$  pour toute parcelle  $i$ , vont représenter les valeurs estimées de la qualité des vignes telle que déduite de la hiérarchie des AOC actuelles. Nous estimons un score continue pour chaque parcelle uniquement selon ses caractéristiques biophysiques et, selon que l'on prenne ou pas en compte sa commune d'appartenance, le traitement préférentiel dont elle a fait l'objet au cours de l'histoire. Notons que cette classification statistique des parcelles est directement issue des AOC qui existent aujourd'hui et ne se base pas sur des appréciations subjectives sur ce qui fait la qualité d'une vigne ou d'un vin, ou la réputation de telle ou telle commune. Nous avons normalisé cette variable pour produire un score de classification avec des valeurs comprises entre 0 et 100 avec une distribution aplatie sur la gauche.

Pour faciliter la consultation de ces résultats, nous agrégeons dans l'application *Shiny* les scores prédits au niveau des parcelles sur la base d'un recodage des dénominations et des lieux-dits (voir le matériel additionnel). Nous utilisons les lieux-dits administratifs qui permettent de localiser plus précisément les parcelles en niveaux Coteaux bourguignons, Bourgogne régional et Village pour lesquels la mention du lieu-dit n'est pas reportée sur l'étiquette des bouteilles (cette pratique est néanmoins de plus en plus fréquente en Côte-d'Or). Il s'agit également de renommer les dénominations géographiques complémentaires associées aux Premiers crus pour qu'ils soient plus lisibles dans l'application.

Comme on pouvait s'y attendre, les Grands Crus arrivent en haut de la classification statistique, qu'ils appartiennent à la côte de Nuits (comme "Chambertin" et "Grands-Echezeaux") ou à la côte de Beaune (comme "Montrachet" et "Bâtard-Montrachet"). Notons tout de même qu'un Premier cru arrive en dixième position, ce qui signifie qu'il dépasse les 2/3 des Grands crus de la zone. Ce Premier cru "La Combe d'Orveau" se trouve sur la commune de Chambolle-Musigny, qui n'apparaît pas relativement désavantagée selon la figure 3. Cela indique que la haute classification de ce Premier cru (au-dessus du Grand cru "Musigny" situé sur la même commune) provient des caractéristiques biophysiques et non de la correction communale. Également étonnant, le Grand cru "Romanée-Conti" qui apparaît souvent parmi les vins les plus chers du monde (<https://www.wine-searcher.com/most-expensive-wines>) n'apparaît qu'en 26e position. Le lieu-dit est tout de même dans les 2 % meilleures localités de la zone. Ce résultat amène à penser que la situation de monopole du domaine de la Romanée-Conti qui exploite ce climat peut

expliquer le fort prix observé des bouteilles, indépendamment des caractéristiques biophysiques.

## 4. Application *Shiny*

Nous proposons deux manières d'utiliser l'application *Shiny* (Chang et al., 2019) pour consulter la classification statistique issue de la modélisation, soit en copiant le code du matériel additionnel pour un déploiement local, soit en consultant le site internet <https://cesaer-datas.inra.fr/geoind> pour un déploiement à partir d'un serveur.

L'application est structurée en trois parties. En haut à gauche, trois zones de saisie permettent de renseigner les caractéristiques d'un vin par menus déroulants (à partir des informations typiquement disponibles sur l'étiquette d'une bouteille ou à partir des connaissances personnelles de l'utilisateur). En bas à gauche, un graphique positionne la qualité de la localité sélectionnée dans la distribution générale des qualités selon le niveaux d'AOC des vignes de la région étudiée. À droite, la carte dynamique permet de faire apparaître la localité sélectionnée dont est issu le vin. L'utilisateur peut sélectionner un vin par ses références mais peut aussi se déplacer librement sur le périmètre d'étude, dans les parcelles voisines d'un lieu-dit par exemple. Ainsi, un clic sur une parcelle permet de faire apparaître ses caractéristiques, les prévisions corrigées et non corrigées en particulier.

Prenons l'exemple d'un vin de niveau Premier cru, sur la commune de Flagey-Echezeaux, qui a pour nom de lieu-dit (dénomination géographique complémentaire en l'occurrence) "Les Rouges". Suite à la saisie de ces caractéristiques dans les zones dédiées en haut à gauche, le score corrigé prédit pour ce vin égal à 83,82 apparaît dans le graphique en bas à gauche de l'application. Ce score est supérieur à l'ensemble des Coteaux bourguignons, des Bourgognes régionaux et des villages du périmètre de l'étude. Ce vin de niveau Premier cru est parmi les 10 % des Premiers crus qui ont les plus hauts scores et il dépasse même 30 % des Grands crus du périmètre. La partie à droite de l'application a zoomé sur cette zone, un clic sur le lieu-dit ciblé permet de faire apparaître la différence entre les prédictions brutes et corrigées. Ainsi, le score non corrigé de la localité est sensiblement plus bas (80,92), ce qui implique que suite à la correction des effets communaux, le vin passe du top 7 % au top 3 % sur l'ensemble de la zone étudiée. Ce résultat est consistant avec les effets communaux reportés dans la Figure 3, où la commune de Flagey-Echezeaux apparaît comme défavorisée dans la hiérarchie. D'après la carte reportée dans l'application, on peut constater que ce Premier cru est mitoyen du Grand cru Echezeaux qui se trouve tout juste à l'est.

## 5. Conclusion

Ce travail détaille de façon reproductible toutes les étapes qui permettent de construire une base de données des parcelles cadastrales pour les 31 communes viticoles de la Côte-d'Or. La base contient les caractéristiques géométriques des parcelles ainsi que leur classement AOC et leurs informations biophysiques (topographie, géologie, pédologie).



L'utilisation des données issues de cette recherche est ensuite illustrée par une modélisation statistique des classifications AOC en 5 niveaux hiérarchiques. Les résultats sont consultables via une application de cartographie interactive de type *Shiny*. Toutes les étapes de la mise en œuvre et les modalités de consultation sont décrites dans l'article, ils peuvent ouvrir la voie à d'autres applications comme l'utilisation de méthodes statistiques alternatives pour démêler les déterminants naturels et humains des AOC, la recherche d'explications alternatives aux effets communaux ou le traitement d'autres problématiques. Sur la base des données que nous avons constituées, le lecteur est libre d'estimer une version différente du modèle avec, par exemple, des degrés variables de lissage spatial.

La classification statistique obtenue permet de préciser la hiérarchie des AOC actuelles, ce qui peut se révéler intéressant pour l'amateur de vin. Il est important de mentionner que cette classification se base exclusivement sur les caractéristiques biophysiques des parcelles et ne contient aucune appréciation subjective exogène sur leur pondération ni sur l'importance d'autres facteurs (olfactifs, de réputation, de prix) en lien avec la qualité d'un vin. L'approche statistique permet en outre de corriger les effets communaux issus des arbitrages politiques qui ont eu lieu au cours de l'histoire, et qui apparaissent avoir une influence sur le classement AOC sans que ce soit justifié du point de vue des caractéristiques biophysiques des parcelles. La classification obtenue est donc directement déduite de la hiérarchie actuelle des AOC qui, au regard de la hiérarchie de prix qu'elle produit, est une information jugée crédible pour les acteurs des marchés du vin. En termes de perspectives, il peut être intéressant de croiser cette qualité des vignes prédites avec d'autres référentiels qualitatifs qui peuvent être issus de l'analyse sensorielle, des notes données par les experts ou des prix des vins dans différents circuits de commercialisation. Cela permettrait de gagner en précision sur les liens entre les facteurs naturels et humains à l'origine de la qualité des vins et le rôle des producteurs dans l'obtention de cette qualité.

En revanche, le modèle proposé n'est pas déterministe et il persiste de l'incertitude dans la classification obtenue. La hiérarchie des parcelles dépend de la spécification du modèle statistique et le fait d'avoir favorisé des variables biophysiques pour décrire la relation entre les lieux de production et la qualité des vins peut faire débat, par opposition aux facteurs humains qui ont un effet indéniable sur la qualité des vins. Le débat sur l'articulation des facteurs humains et des facteurs naturels de la qualité du vin existe depuis plus d'un demi-siècle et produit des discussions toujours d'actualité (Delay and Chevalier, 2015 ; Lammoglia et al. 2019). C'est pour alimenter ce débat que les analyses présentées dans cet article sont totalement reproductibles à partir de la base de données produite. Les codes pour l'estimation du modèle sont reportés dans le Matériel Additionnel associé, afin de permettre l'estimation de modèles alternatifs. L'appariement de données supplémentaires pour relier les AOC aux caractéristiques des lieux est également possible par la géolocalisation des parcelles cadastrales. La transparence des analyses permet aux résultats d'être discutés et contestés afin de faire l'objet d'une appropriation par les chercheurs, les décideurs, les professionnels du secteur, ou les amateurs de vin.

## 6. Remerciements

Nous tenons à remercier Pauline Mialhe (CRC, Centre de Recherches de Climatologie, Laboratoire Biogéosciences - Université de Bourgogne), Vincent Larmet et Guillaume Royer (INRA CESAER, Centre d'Economie et de Sociologie appliqués à l'Agriculture et aux Espaces Ruraux) pour leur aide sur le développement de l'application issue de cette recherche.

## 7. Références

Abbal, P., Sablayrolles, J. M., Matzner-Lober, E., et Carbonneau, A. (2019). A Model for Predicting Wine Quality in a Rhône Valley Vineyard. *Agronomy Journal* 111:2 545-554.

Agresti, A. et Kateri, M. (2017). Ordinal probability effect measures for group comparisons in multinomial cumulative link models. *Biometrics* 73: 214–219.

Appelhans, T., Detsch, F., Reudenbach, C. et Woellauer, S. (2018). mapview: Interactive Viewing of Spatial Data in R. R package version 2.6.3.

Ashenfelter, O. et Storchmann, K. (2010). Using hedonic models of solar radiation and weather to assess the economic effect of climate change: the case of Mosel valley vineyards. *The Review of Economics and Statistics* 92:2 333–349.

Ay, J.-S. (2011). Hétérogénéité de la terre et rareté économique. Theses, Université de Bourgogne.

Ay, J.-S. (2020). The informational content of geographical indications. *American Journal of Agricultural Economics*. Sous presse.

Barrey, S. et , Teil, G. (2011). Faire la preuve de l'« authenticité » du patrimoine alimentaire. *Anthropology of Food* 8: 1-18.

Bazin, J.-F. (2015). *Le vin de Bourgogne*. Paris, Dunod. 264 p. ISBN: 978-2100585182.

Beckert, J., Rössel, J. et Schenk, P. (2017). Wine as a cultural product: Symbolic capital and price formation in the wine field. *Sociological Perspectives* 60:1 206-22.

Bivand, R. S., Pebesma, E. and Gomez-Rubio, V. (2013). Applied spatial data analysis with R, Second edition. Springer, NY.

Bois B., Pauthier B., Brillante L., Mathieu O., Leveque J., Van Leeuwen C., Castel T. and Richard Y. (2020) Sensitivity of Grapevine Soil–Water Balance to Rainfall Spatial Variability at Local Scale Level. *Front. Environ. Sci.* 8:110.

Bokulich, N. A., Thorngate, J. H., Richardson, P. M. et Mills, D. A. (2014). Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proceedings of the National Academy of Sciences* 111:1 139-148.

Capus, J. (1947). L'évolution de la législation sur les appellations d'origine. Genèse des appellations contrôlées. L. Larmat.

Carter, E. (2018). For what it's worth: The political construction of quality in French and Italian wine markets. *Socio-Economic Review* 16:3 479-498.

Chang, W., Cheng, J., Allaire, J., Xie, Y. and McPherson, J. (2019). shiny: Web Application Framework for R. R package version 1.4.0.

Colman, T. (2008). *Wine Politics: How Governments, Environmentalists, Mobsters, and Critics Influence the Wines We Drink*. Berkeley, University of California Press, 208 p., ISBN 978-0520267886.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J. (2015). System for automated geoscientific analyses (saga) v. 2.1. 4. *Geoscientific Model Development* 8: 1991–2007.

Delay, E. and Chevallier, M. (2015). Roger Dion, toujours vivant! *Cybergeo: European Journal of Geography* GeOpenMod, document 721.

Demossier, M. (2011). Beyond *terroir*: Territorial construction, hegemonic discourses, and French wine culture. *Journal of the Royal Anthropological Institute*. 17:685-705.

Dion, R. (1952). Querelle des anciens et des modernes sur les facteurs de la qualité du vin. *Annales de géographie* 61: 417–431.

Garcia, J.-P. (2011). Les climats du vignoble de Bourgogne comme patrimoine mondial de l'humanité. Editions Universitaires de Dijon.

Gergaud, O. et Ginsburgh, V. (2008). Natural Endowments, Production Technologies and the Quality of Wines in Bordeaux. Does Terroir Matter? *Economic Journal* 118: 529 142-157.

Hilal, M. (2020). Land use land cover high resolution map (5-m) for Côte-d'Or (21). Portail Data INRAE <https://doi.org/10.15454/ECLYGT>.

Hilal, M., Joly, D., Roy, D. and Vuidel, G. (2018). Visual structure of landscapes seen from built environment. *Urban Forestry & Urban Greening* 32: 71–80.

Humbert, F. (2011). L'INAO, de ses origines à la fin des années 1960: genèse et évolutions du système des vins d'AOC. Ph.D. thesis, Université de Bourgogne.

Jacquet, O. et Laferté, G. (2006). Le contrôle républicain du marché: Vignerons et négociants sous la Troisième République. *Annales. Histoire, Sciences Sociales* 61:5 1147-1170.

Jacquet, O. (2009). Un siècle de construction du vignoble bourguignon. Les organisations vitivinicoles de 1884 aux AOC. Editions Universitaires de Dijon.

Joly, D., Brossard, T., Cardot, H., Cavailles, J., Hilal, M. et Wavresky, P. (2010) Les types de climats en France, une construction spatiale. *Cybergeo : European Journal of Geography* 23155.

Jones G V., Snead N, Nelson P. Geology and Wine 8. Modeling Viticultural Landscapes: A GIS Analysis of the Terroir Potential in the Umpqua Valley of Oregon. *Geoscience Canada*. 2004;31(4):167–178. Karpik, L. (2007). *L'économie des singularités*. Paris, NRF Éditions Gallimard, 384 p. ISBN 978-2070775323.

- Knight, S., Klaere, S., Fedrizzi, B. et Goddard, M. R. (2015). Regional microbial signatures positively correlate with differential wine phenotypes: Evidence for a microbial aspect to *terroir*. *Scientific reports*. 5.
- Labbé, T. (2011). La revedication d'un terroir viticole : La Côte-de-Beaune à la fin du XVIIIe siècle. *Histoire et Sociétés Rurales* 35: 99-126.
- Lammoglia, A., Leturcq, S. and Delay, E. (2019). Le modèle VitiTerroir pour simuler la dynamique spatiale des vignobles sur le temps long (1836-2014). *Cybergeo: European Journal of Geography GeOpenMod*, document 820.
- Lavalle, J. (1855). Histoire et statistique de la vigne et des grands vins de la Côte-d'Or. Daumier, Dijon. 23.
- Leeper, T. J. (2017). dataverse: R Client for Dataverse 4. R package version 0.2.0.
- Lucand, C. (2011). Les négociants en vins de Bourgogne. Armand Colin.
- Madelin, M., Bois, B., Chabin, J.-P. (2010). Modification des conditions de maturation du raisin en Bourgogne viticole liée au réchauffement climatique. L'exemple des vignobles de la Côte et des Hautes-Côtes de Beaune. *EchoGéo*. 12176.
- Parker, T. (2015) *Tasting French Terroir. The History of an Idea*, Berkeley, University of California Press, 229 p., ISBN 978-0-520-27750-2.
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10:439–446, doi:10.32614/RJ-2018-009.
- Polàskova, P., Herszage, J. et Ebeler, S. (2008). Wine flavor: Chemistry in a glass. *Chemical Society Reviews* 37:11 2478-2489.
- R Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rich, P. M. and Fu, P. (2000). Topoclimatic habitat models. *Proceedings of the Fourth International Conference on Integrating GIS and Environmental Modeling*.
- Roullier-Gall, C., Boutegrabet, L., Gougeon, R. D. et Schmitt-Kopplin, P. (2014). A grape and wine chemodiversity comparison of different appellations in Burgundy: Vintage vs *terroir* effects. *Food chemistry* 152: 100-107.

Tiefenbacher, J. P. (2013). Themes of U.S. Wine Advertising and the Use of Geography and Place to Market Wine. *EchoGéo* 23:1-21.

van Leeuwen, C., Roby, J.-P., et de Rességuier, L. (2018). Soil-related terroir factors: A review. *OENO one* 52:2 173-188.

Vaudour, E. (2002). The Quality of Grapes and Wine in Relation to Geography: Notions of Terroir at Various Scales. *Journal of Wine Research*, 13:2, 117-141,

Wolikow, S. and Jacquet, O. (2011). Territoires et terroirs du vin du XVIIIe au XXIe siècles. Éditions Universitaires de Dijon.

Wood, S. N. (2017). Generalized additive models: An introduction with R. Chapman and Hall/CRC, second edition.

Wood, S. N., Pya, N. and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* 111: 1548–1563.

## Annexes

### Annexe 1 : Dictionnaire des variables

Table 2: Dictionnaire des variables disponibles dans la base de données

NOM	TYPE	DESCRIPTION
IDU	Caractère	Identifiant de la parcelle cadastrale (14 caractères)
CODECOM	Caractère	Code INSEE de la commune d'appartenance (5 caractères)
AREA	Numérique	Surface calculée de la parcelle cadastrale (en mètres carrés)
PERIM	Numérique	Périmètre calculé de la parcelle cadastrale (en mètres)
MAXDIST	Numérique	Distance maximale calculée entre deux sommets (en mètres)
PAOC	Indicatrice	1 si la parcelle est dans au moins une AOC
BGOR	Indicatrice	1 si la parcelle est dans le niveau Coteaux bourguignon
BOUR	Indicatrice	1 si la parcelle est dans le niveau Bourgogne régional
VILL	Indicatrice	1 si la parcelle est dans le niveau Bourgogne village
COMM	Indicatrice	1 si la parcelle est dans le niveau Bourgogne communal
PCRU	Indicatrice	1 si la parcelle est dans le niveau Premier cru
GCRU	Indicatrice	1 si la parcelle est dans le niveau Grand cru
AOC	Numérique	Rang de la parcelle dans la hiérarchie des AOC (entre 0 et 5)
AOCTp	Caractère	Modalité Appel pour une appellation ou Denom pour une dénomination
AOC1b	Caractère	Libellé de l'appellation ou de la dénomination selon la variable AOCTp
AOC36lab	Caractère	Libellé de l'appellation en 1936 (56 modalités)
AOC36lv1	Caractère	Rang de la parcelle dans la hiérarchie des AOC de 1936 (0, 3 ou 5)
LIEUDIT	Caractère	Libellé du lieu-dit de la parcelle (2691 modalités)
CLDVIN	Caractère	Identifiant du lieu-dit de la parcelle (2691 modalités)
LIBCOM	Caractère	Libellé de la commune de la parcelle (31 modalités)
XCHF	Numérique	Latitude du chef-lieu de la commune (système Lambert 93)
YCHF	Numérique	Longitude du chef-lieu de la commune (système Lambert 93)
ALTCOM	Numérique	Altitude du point culminant de la commune (mètre)
SUPCOM	Caractère	Superficie de la commune de la parcelle (hectare)
POPCOM	Numérique	Population de la commune de la parcelle en 2015 (millier d'hab)
CODECANT	Caractère	Identifiant du canton d'appartenance (2 caractères)
REGION	Caractère	Région viticole (CDB pour côte de Beaune, CDN pour côte de Nuits)
PAR2RAS	Numérique	Identifiant pour appariement entre vecteurs et raster
XL93	Numérique	Latitude du centroïde de la parcelle (système Lambert 93)



YL93	Numérique	Longitude du centroïde de la parcelle (système Lambert 93)
NOMOS	Numérique	Part de la parcelle hors du mode d'occupation des sol (entre 0 et 1)
URBAN	Numérique	Part de la parcelle en usage urbain selon le MOS (entre 0 et 1)
FOREST	Numérique	Part de la parcelle en usage forestier selon le MOS (entre 0 et 1)
WATER	Numérique	Part de la parcelle en eau selon le MOS (entre 0 et 1)
DEM	Numérique	Altitude moyenne de la parcelle selon le MNT (en mètres)
SLOPE	Numérique	Pente moyenne de la parcelle selon le MNT (en degrés)
ASPECT	Numérique	Exposition moyenne de la parcelle selon le MNT (en degrés)
SOLAR	Numérique	Radiation solaire moyenne sur la parcelle (en Joules)
CODE	Caractère	Code de la géologie (31 modalités)
NOTATION	Caractère	Notation stratigraphique harmonisée (31 modalités)
DESCR	Caractère	Description géologie (31 modalités)
TYPEGEOL	Caractère	Type superficiel (4 modalités)
APLOCALE	Caractère	Colluvions, Eboulis, etc. (28 modalités)
TYPEAP	Caractère	Type de formation (7 modalités)
GEOLNAT	Caractère	Nature Géologique (3 modalités)
ISOPIQUE	Caractère	Faciès des couches (4 modalités)
AGEDEB	Caractère	Age de la couche (24 modalités)
ERADEB	Caractère	Céno ou Méso (2 modalités)
SYSDEB	Caractère	Age autre (5 modalités)
LITHOLOGIE	Caractère	Litho (16 modalités)
DURETE	Caractère	Dureté (3 modalités)
ENVIRONMT	Caractère	Environnement (9 modalités)
GEOCHIMIE	Caractère	Géochimie (5 modalités)
LITHOCOM	Caractère	Litho détaillée (30 modalités)
NOUC	Caractère	Numéro de l'unité cartographique (2 caractères)
SURFUC	Numérique	Surface de l'unité cartographique (en hectares)
TARG	Numérique	Taux d'argile de l'unité typologique dominante (pourcentage)
TSAB	Numérique	Taux de sable de l'unité typologique dominante (pourcentage)
TLIM	Numérique	Taux de limons de l'unité typologique dominante (pourcentage)
TEXTAG	Caractère	Classes de textures agrégées en 9 modalités (voir Ay, 2011)
EPAIS	Numérique	Épaisseur des sols de l'unité typologique dominante (centimètre)
TEG	Numérique	Taux d'éléments grossiers de l'unité typologique dominante (pour mille)
TMO	Numérique	Taux de Matière organique de l'unité typologique dominante (pourcentage)
RUE	Numérique	Réserve Utile par excès de l'unité typologique dominante (millimètre)

---

RUD	Numérique	Réserve Utile par défaut de l'unité typologique dominante (millimètre)
OCCUP	Numérique	Part de l'unité typologique dominante dans l'unité cartographique (entre 0 et 1)
DESCRp	Caractère	Libellé de la classe pédologique en 33 modalités

---