

mgcvUI : une interface interactive pour le package mgcv (modèles additifs généralisés)

Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.

William Bert Craytor

Volume 1, Numéro 1 · Été 2026

Publié le 1^{er} juillet 2026

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Cet article en libre accès est diffusé selon les termes de la [licence Creative Commons Attribution 4.0](#).

Résumé. mgcvUI est une interface graphique pour le package R **mgcv**, qui ajuste des modèles additifs généralisés (GAM) au moyen de splines de régression pénalisées avec sélection automatique du degré de lissage. Elle propose trois modes d'utilisation — modélisation prédictive générale, évaluation immobilière et analyse de zone de marché — et guide l'utilisateur à travers l'importation des données, la spécification des termes lisses, l'ajustement du modèle, les graphiques de diagnostic et d'effets, ainsi que les rapports téléchargeables. Cet article documente les exigences de format de données de mgcvUI, le flux de travail de modélisation, les affichages de sortie et la référence complète des fonctionnalités.

Mots-clés : modèles additifs généralisés, GAM, splines pénalisées, lissage, mgcv, R, interface graphique, évaluation immobilière

Note de l'éditeur (juillet 2026) : Peu avant la mise sous presse, mgcvUI a été intégré à earthUI (version 0.11.0 et ultérieures), qui fournit désormais le flux de travail GAM décrit ici aux côtés des flux de travail MARS et elastic-net. mgcvUI demeure disponible dans sa dernière version publiée (0.3.0) à l'adresse <https://github.com/wcraytor/mgcvUI> ; les développements futurs se poursuivent dans earthUI. Le présent article s'applique à mgcvUI tel que publié.

Table des matières

1	Introduction	4
1.1	Historique et contexte	4
1.1.1	Modèles additifs généralisés	4
1.1.2	Le package mgcv	4
1.1.3	Comparaison avec MARS (earth) et l'elastic net (glmnet)	5
1.1.4	La chaîne earthUI–mgcvUI	5
1.1.5	Flux de travail recommandé	6
1.2	Qu'est-ce que mgcvUI ?	6
1.3	Trois modes d'utilisation	7
1.4	Fonctionnalités propres à l'immobilier	7
1.5	Premiers pas	7
2	Exigences relatives aux données d'entrée MLS	9
2.1	Format et structure du fichier	9
2.2	Colonnes requises en mode évaluation	9
2.3	Conventions de nommage des colonnes spéciales	10
2.4	Qualité et complétude des données	10
2.5	Emplacement de la ligne objet	10

3	Mode General	11
3.1	Vue d'ensemble	11
3.2	Le flux de travail de la barre latérale	11
3.3	Importation depuis earthUI (facultatif)	12
3.4	Onglets du panneau principal	12
3.5	Persistance des réglages	13
3.6	Mode sombre	13
3.7	Locale et réglages régionaux	13
4	Mode Appraisal	15
4.1	Traitement de la ligne objet	15
4.2	Date d'effet et âge de la vente	15
4.3	Désignations de colonnes spéciales	15
4.4	Aperçu des ajustements RCA	15
5	Mode Market Area Analysis	16
5.1	Différences avec le mode évaluation	16
5.2	Quand utiliser le mode marché	16
6	Sélection des variables	17
6.1	Variable cible	17
6.2	Transformation de la réponse	17
6.3	Le tableau des prédicteurs	17
6.4	Détection et forçage des types de données	18
6.5	Référence des types de colonnes spéciales	18
7	Choix des paramètres	20
7.1	Préréglages de paramètres	20
7.2	Famille	20
7.3	Méthode	20
7.4	Gamma (multiplicateur de lissage)	21
7.5	Validation croisée	21
7.6	Select (rétrécissement)	21
7.7	Base par défaut	21
7.8	k par défaut (dimension de la base)	21
7.9	Type de produit tensoriel	22
7.10	Interactions autorisées	22
7.11	Interactions facteur $\times$ terme lisse	22
7.12	Paramètres avancés	23
8	Ajustement du modèle	24
8.1	Le bouton d'ajustement	24
8.2	Ce qui se passe pendant l'ajustement	24
8.3	Ajustement en arrière-plan et progression	24
9	Onglets de résultats	25
9.1	Data	25
9.2	Equation	25
9.3	Summary	25
9.4	Variable Importance	25
9.5	Graphiques de contribution	26
9.5.1	Termes lisses univariés — $s(x)$	26

9.5.2	Termes lisses par facteur — $s(x, \text{by} = \text{factor})$	26
9.5.3	Termes d'interaction — $ti(x_1, x_2)$ et $te(x_1, x_2)$	26
9.5.4	Termes paramétriques — linéaires et facteurs	26
9.6	Correlation	26
9.7	Diagnostics	27
9.8	RCA Adjustments	27
9.9	Anova	27
9.10	Mgev Output	27
9.11	Sign Check	27
9.12	Concurvity	27
9.13	Basis Check	28
10	Téléchargement des données	29
10.1	Colonnes de sortie	29
10.2	Scores CQA	29
11	Calculs RCA et téléchargement	30
11.1	Ouverture de la boîte de dialogue RCA	30
11.2	Interpolation du score CQA	30
11.3	Colonnes de sortie	30
12	Grille de comparaison des ventes	31
12.1	Boîte de dialogue de sélection des comparables	31
12.2	Structure du classeur	31
12.3	Protection des feuilles et cellules résiduelles	31
13	Téléchargement des rapports	32
13.1	Étape 1 : générer le rapport Quarto	32
13.2	Étape 2 : convertir le rapport Quarto	32
13.3	Contenu du rapport	32
14	Export des fonctions du modèle	34
14.1	Fonctions R	34
14.2	Fonctions Python	34
14.3	En-tête C++	34
14.4	Base de données SQLite	34
15	Comparaison avec earthUI et glmnetUI	36
15.1	Principales différences	36
15.2	Fonctionnalités communes	36
15.3	Flux de travail recommandé	37
16	Jeu de données de démonstration : Appraisal_1.csv	38
16.1	Description	38
16.2	Colonnes	38
16.3	Prise en main rapide suggérée	39
17	Configuration requise et dépannage	40
17.1	Plateformes prises en charge	40
17.2	Version de R	40
17.3	Packages R requis	40
17.4	Dépendances facultatives	41
17.5	Remarques propres à chaque plateforme	41

17.5.1	macOS	41
17.5.2	Windows	41
17.5.3	Linux	41
17.6	Dégradation progressive	42
17.7	Dépannage	42
17.7.1	« PDF option is not available »	42
17.7.2	« Settings will not persist »	42
17.7.3	« Diagnostics plot unavailable »	43
17.7.4	« Report error : number of items to replace... »	43
17.7.5	Le port 7880 est déjà utilisé	43
17.7.6	L'installation des packages échoue sous Linux	43

1 Introduction

1.1 Historique et contexte

1.1.1 Modèles additifs généralisés

Les **modèles additifs généralisés** (GAM, *Generalized Additive Models*) ont été formalisés par Trevor Hastie et Robert Tibshirani dans leur monographie de 1990. Les GAM étendent le modèle linéaire en remplaçant chaque terme linéaire $\beta_j x_j$ par une fonction lisse $f_j(x_j)$, de sorte que le modèle devient :

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

Chaque f_j est estimée à partir des données au moyen d'une spline de régression pénalisée. Les fonctions lisses permettent au modèle de capter des relations non linéaires sans que l'utilisateur ait à spécifier une forme fonctionnelle à l'avance. La pénalité contrôle l'ondulation de chaque terme lisse, évitant le surajustement tout en laissant assez de souplesse pour suivre les structures réelles.

1.1.2 Le package mgcv

Le package R `mgcv` (Mixed GAM Computation Vehicle) a été développé par Simon Wood à l'université de Bath (WOOD 2017). Publié pour la première fois en 2001, il est aujourd'hui l'implémentation R de référence des GAM et est fourni avec toute installation de R. Les méthodes sous-jacentes sont décrites dans une série d'articles portant sur l'estimation stable de paramètres de lissage multiples (WOOD 2004), les splines de régression à plaque mince (WOOD 2003), l'estimation REML rapide et stable (WOOD 2011) et la sélection de modèles lisses généraux (WOOD, PYA et SÄFKEN 2016). Parmi les fonctionnalités clés :

- **Sélection automatique du degré de lissage** : le paramètre de lissage de chaque terme est estimé dans le cadre même de l'ajustement du modèle, par REML (maximum de vraisemblance restreint), GCV (validation croisée généralisée) ou ML (maximum de vraisemblance).
- **Types de bases multiples** : splines de régression à plaque mince (tp), splines de régression cubiques (cr), P-splines (ps) et B-splines (bs), chacune ayant des propriétés différentes.
- **Lissages par produit tensoriel** : `te()` et `ti()` permettent des interactions lisses entre variables sans subir le fléau de la dimension.
- **Sélection de variables** : l'option `select = TRUE` ajoute une pénalité supplémentaire capable de ramener des termes lisses à zéro, ce qui revient à effectuer une sélection de variables.

- **Diagnostics de concurvit ** : mesure l’analogue de la multicol n arit  pour les termes lisses.

1.1.3 Comparaison avec MARS (earth) et l’elastic net (glmnet)

L’application earth et ses deux Post Processing Modules (modules de post-traitement) glmnet et mgcv reposent sur des moteurs de mod lisation diff rents. Le tableau suivant r sume les principales diff rences :

Caract�ristique	mgcv (GAM)	earth (MARS)	glmnet (elastic net)
Type de mod�le	Non lin�aire lisse (splines p�nalis�es)	Lin�aire par morceaux (splines adaptatives)	Lin�aire (avec r�gularisation)
Non-lin�arit�	Automatique, par fonctions de lissage (thin plate, cubique, etc.)	Automatique, par fonctions charni�res � n�uds d�duits des donn�es	Uniquement par interactions ou expansion de base
S�lection de variables	Facultative, par double p�nalit� de r�tr�cissement (<code>select=TRUE</code>)	Int�gr�e, par s�lection pas � pas avant/arri�re avec �lagage GCV	Int�gr�e, via la p�nalit� L_1 (Lasso)
Interactions	Lissages par produit tensoriel (<code>te()</code> , <code>ti()</code>)	Produits de charni�res (interpr�tables, jusqu’au degr� 3)	Produits crois�s deux � deux (difficiles � interpr�ter)
Interpr�tabilit�	Courbes d’effets partiels lisses ; tr�s intuitives	Les g-fonctions montrent des effets partiels lin�aires par morceaux, variable par variable	Les coefficients sont des poids lin�aires, faciles � expliquer un par un
Domaine de pr�dilection	Relations non lin�aires lisses, mod�lisation souple	D�tection automatique des non-lin�arit�s et des interactions	Donn�es de grande dimension, s�lection de variables, petits �chantillons
Ajustements RCA	Ajustements lisses via les effets partiels	Ajustements lin�aires par morceaux via les g-fonctions	Ajustements lin�aires par variable
Contr�le du surajustement	Estimation du param�tre de lissage par REML/GCV	�lagage fond� sur le GCV	Lambda par validation crois�e

MARS (Multivariate Adaptive Regression Splines) a  t  introduit par Jerome Friedman en 1991. La m thode construit des mod les lin aires par morceaux en s lectionnant de mani re adaptative des fonctions charni res et la position de leurs n uds   partir des donn es. L’impl mentation R est le package **earth** de Stephen Milborrow.

La r gression **elastic net**, mise en  uvre dans le package **glmnet** de Friedman, Hastie et Tibshirani (2010), combine les p nalit s Lasso (L_1) et ridge (L_2) pour effectuer simultan ment la s lection de variables et le r tr cissement des coefficients.

1.1.4 La cha ne earthUI–mgcvUI

Un flux de travail puissant associe la d couverte automatique des n uds par earthUI   l’estimation lisse de mgcvUI. Lorsque earthUI exporte un fichier de r sultats `.rds` et que mgcvUI l’importe :

1. **Earth découvre les nœuds** — MARS détermine, pour chaque variable, des positions de charnières (points de rupture) déduites des données.
2. **mgcvUI les utilise comme ancrages** — les nœuds d'earth deviennent la base des splines de régression cubiques (base `cr`) du GAM, ce qui donne aux termes lisses une longueur d'avance grâce à des positions de nœuds reflétant la structure réelle des données.
3. **mgcv affine l'ajustement** — l'estimation par splines pénalisées lisse l'ajustement linéaire par morceaux d'earth et produit des courbes d'effets partiels lisses, tout en préservant la forme générale découverte par earth.

Cette chaîne fait le lien entre modélisation exploratoire (earth) et modélisation confirmatoire (GAM), en combinant le meilleur des deux approches.

1.1.5 Flux de travail recommandé

Pour l'évaluation immobilière et les applications similaires exigeant des modèles interprétables et défendables :

1. **Commencez par earthUI** pour découvrir les variables importantes, les relations non linéaires et les interactions présentes dans vos données. La sélection automatique de base d'earth et ses g-fonctions fournissent d'excellents éclairages initiaux.
2. **Affinez avec mgcvUI** si vous souhaitez des effets partiels plus lisses. Importez les nœuds d'earthUI dans mgcvUI pour passer sans heurt des modèles linéaires par morceaux aux modèles lisses.
3. **Utilisez glmnetUI** lorsque vous avez besoin d'une sélection de variables agressive (Lasso), lorsque le nombre de prédicteurs approche ou dépasse le nombre d'observations, ou lorsque vous voulez un modèle purement linéaire avec régularisation, pour des raisons de défendabilité.

1.2 Qu'est-ce que mgcvUI ?

mgcvUI est une interface graphique pour le package R `mgcv`. Elle s'exécute comme une application Shiny locale — sans identification, sans serveur et sans compte utilisateur. Vous la lancez depuis R, vous importez un jeu de données (CSV ou Excel), vous configurez votre modèle et vous ajustez un GAM de manière interactive.

Dans la version actuelle, mgcvUI n'est plus un package installé séparément : il est livré à l'intérieur du package earthUI, dont il constitue le Post Processing Module (module de post-traitement) GAM, démarré par `earthUI::launch_mgcv()`. L'application elle-même — et tout ce qui est décrit dans cet article — demeure inchangée ; seules l'installation et la commande de lancement diffèrent. Les trois routines (earth, glmnet, mgcv) partagent une même arborescence de projet, une même base de données de réglages et un même exemplaire de l'infrastructure de support.

L'application offre un flux de travail complet : importation des données, configuration des variables avec spécification des termes lisses, ajustement du modèle avec traitement en arrière-plan, graphiques de diagnostic, courbes d'effets partiels lisses et rapports téléchargeables aux formats Word, PDF ou HTML.

Les **modèles additifs généralisés** remplacent la relation linéaire βx par une fonction lisse $f(x)$ pour chaque prédicteur. Cela signifie :

- **Aucune forme fonctionnelle à spécifier** — ce sont les données qui déterminent la forme de chaque relation.
- **Des effets partiels lisses** — la contribution de chaque prédicteur est une courbe lisse et non une droite, ce qui permet de voir facilement comment une variable influe sur la réponse selon ses différentes valeurs.

- **Sélection automatique du degré de lissage** — `mgcv` détermine par REML ou GCV le degré d'ondulation de chaque courbe, en équilibrant qualité d'ajustement et surajustement.

1.3 Trois modes d'utilisation

Chaque projet a un **objet** (*purpose*), choisi lors de sa création à la section 1 de la barre latérale (**Project**, projet) puis verrouillé. L'objet du projet actif détermine les outils et éléments d'interface qui apparaissent ; pour travailler dans un autre mode, créez ou ouvrez un projet ayant cet objet. Les trois modes sont :

- **General** (général) — régression GAM pour tout type de population ou de jeu de données. C'est le mode par défaut. Il offre le flux de travail GAM complet, sans ajout propre à un domaine.
- **For Appraisal** (pour l'évaluation) — régression GAM adaptée à l'évaluation immobilière. Ajoute des fonctionnalités propres à l'évaluation d'un bien unique : traitement du bien objet, désignations de colonnes spéciales et approche par contrainte des résidus (RCA, *Residual Constraint Approach*).
- **Market Area Analysis** (analyse de zone de marché) — régression GAM adaptée aux études de zone de marché. Ajoute des fonctionnalités permettant d'analyser des groupes de biens dans un marché défini, avec traitement facultatif d'une ligne objet.

Dans les trois modes, le moteur de modélisation est identique — vous ajustez toujours un GAM au moyen de `mgcv::gam()`. Le réglage de l'objet détermine seulement quels outils et éléments d'interface supplémentaires sont disponibles.

1.4 Fonctionnalités propres à l'immobilier

Lorsque **For Appraisal** ou **Market Area Analysis** est sélectionné, `mgcvUI` active plusieurs fonctionnalités conçues pour l'analyse immobilière :

- **Désignations de colonnes spéciales** — chaque prédicteur peut recevoir un rôle spécial tel que `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Ces désignations déterminent la manière dont la colonne est traitée lors de l'ajustement et dans les sorties.
- **Colonne Sale Age** (âge de la vente) — lorsqu'une colonne est désignée `contract_date` et qu'une date d'effet est fournie, `mgcvUI` calcule une colonne `sale_age` (nombre de jours entre la date de vente et la date d'effet) et la substitue comme prédicteur.
- **Calculs RCA (mode évaluation uniquement)** — en mode évaluation, après l'ajustement du modèle, `mgcvUI` peut produire les sorties de l'approche par contrainte des résidus (RCA) : ajustements par comparable, synthèses des ajustements nets et bruts et prix de vente ajusté du bien objet.

Astuce : Le mode General convient à tout jeu de données — financier, scientifique, technique ou immobilier. Les modes évaluation et marché ne font qu'ajouter des commodités destinées aux professionnels de l'immobilier.

1.5 Premiers pas

Pour utiliser `mgcvUI` :

1. **Installez earthUI** dans R : `install.packages("earthUI")` — la routine `mgcv` y est incluse. (Voir « Installing earthUI and vProlog » dans l'article `earthUI` du présent numéro pour la version de développement et les composants facultatifs.)

2. **Lancez l'application** : exécutez `earthUI::launch_mgcv()` dans R. L'application s'ouvre dans votre navigateur sur le port 7880.
3. **Créez ou ouvrez un projet** à la section 1 de la barre latérale. Un projet fixe l'objet (General, For Appraisal ou Market Area Analysis) et le lieu de travail (pays, état, comté, ville) ainsi que ses dossiers d'entrée et de sortie. Les projets sont stockés sous une racine `regProj` commune et sont partagés avec les applications sœurs earthUI et glmnetUI. Ouvrir un projet ayant un autre objet est la façon de changer de mode.
4. **Importez vos données** à la section 2 en sélectionnant un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajoutez des fichiers à ce dossier, puis cliquez sur **Refresh** (actualiser) pour les lister.
5. **Importez depuis earthUI (facultatif)** — la section 3 permet d'importer un fichier de résultats `.rds` d'earthUI afin d'amorcer le GAM avec les nœuds découverts par earth. Ignorez cette étape si vous n'avez pas de résultat earthUI.
6. **Configurez les variables** — choisissez votre cible et vos prédicteurs, définissez les types de données, attribuez les éventuels rôles de colonne spéciaux et marquez les variables comme facteur ou linéaire.
7. **Réglez les paramètres mgcv** — famille, méthode, gamma, type de base, k, type de produit tensoriel, interactions et autres options.
8. **Ajustez le modèle** — cliquez sur « Fit Mgcv GAM Model » (ajuster le modèle GAM mgcv) et examinez les résultats dans le panneau principal.
9. **Exportez** — téléchargez les prédictions au format Excel, générez et convertissez un rapport Quarto ou, en mode évaluation, calculez les ajustements RCA et une grille de ventes.

Les réglages sont automatiquement conservés dans une base de données SQLite et restaurés lorsque vous rechargez le même fichier d'entrée.

2 Exigences relatives aux données d'entrée MLS

Pour les flux de travail d'évaluation immobilière et d'analyse de marché, vos données d'entrée proviennent généralement d'un export de Multiple Listing Service (MLS). Ce chapitre décrit la structure de fichier attendue et les colonnes que mgcvUI sait exploiter.

2.1 Format et structure du fichier

mgcvUI accepte les fichiers **CSV** et **Excel** (.xlsx, .xls). À l'importation, les noms de colonnes sont automatiquement convertis en **snake_case** — par exemple, « Living SqFt » devient **living_sqft**, « Contract Date » devient **contract_date** et « Sale Price » devient **sale_price**. Cette normalisation garantit des références de colonnes cohérentes dans tout le flux de travail. Le séparateur CSV et le symbole décimal utilisés à l'importation sont déterminés par les réglages de locale (voir le chapitre 3, « Locale & Regional Settings »).

Votre fichier de données doit être un tableau plat comportant une ligne par bien et une colonne par attribut. La première ligne du fichier doit contenir les en-têtes de colonnes.

2.2 Colonnes requises en mode évaluation

Bien que mgcvUI fonctionne avec n'importe quel ensemble de colonnes, le flux de travail complet d'évaluation tire profit de la présence des colonnes suivantes :

Colonne	Type spécial	Rôle
Sale Price	(cible)	Variable réponse du modèle
Contract Date	contract_date	Sert à calculer sale_age (jours écoulés depuis la date d'effet)
Listing Date	listing_date	Utilisée avec la date de contrat pour calculer le DOM en l'absence de colonne DOM
Days on Market	dom	Jours sur le marché ; affichée dans les exports
Concessions	concessions	Concessions de vente ; SP net = prix de vente – concessions
Living Area (SF)	living_area	Active les résidus au pied carré (residual_sf , cqa_sf)
Lot Size	lot_size	Colonne de superficie du terrain
Site Dimensions	site_dimensions	Regroupée avec la superficie du terrain
Latitude	latitude	Arrondie à 3 décimales
Longitude	longitude	Arrondie à 3 décimales
Area ID	area	Identifiant de zone de marché / quartier
Actual Age	actual_age	Âge du bien
Effective Age	effective_age	Âge effectif du bien
Address	display_only	Affichée dans les exports ; exclue du modèle

Les noms de colonnes du tableur peuvent être dans une langue étrangère — les noms « spéciaux » sont en anglais afin que le programme R puisse leur appliquer un traitement particulier. Pour le reste, les noms de colonnes fournis apparaissent tels quels dans les modèles de régression, les graphiques et, si vous réalisez des évaluations, les rapports de sortie.

Toutes les colonnes ne sont pas obligatoires. mgcvUI s'adapte — si une colonne est absente, la fonctionnalité correspondante est simplement omise. Cependant, pour les modèles de prix immobiliers, certaines colonnes sont vivement recommandées :

1. **Sale Age** — le nombre de jours entre la date de vente contractuelle et la date d'effet. Lorsqu'on utilise un historique de ventes pluriannuel, `sale_age` joue souvent un rôle central.
2. **Living Area** — appelée aussi « Living Sqft » ou « GLA » (*gross living area*, surface habitable brute).
3. **Total Bath Count** — le nombre total de salles de bains complètes, aux trois quarts, à moitié et au quart.
4. **Garage Bays** ou **Garage Area** — nombre de places de garage ou surface du garage.
5. **Lot Size** — la superficie du terrain du bien.
6. **Longitude, Latitude** et, si disponible, **Area ID** pour la variation géographique des prix.

2.3 Conventions de nommage des colonnes spéciales

mgcvUI identifie les colonnes par leur **désignation de type spécial**, et non par leur nom. Vous pouvez nommer vos colonnes comme vous le souhaitez dans l'export MLS — ce qui compte, c'est de leur attribuer le bon type spécial dans le tableau de configuration des variables (chapitre 6).

2.4 Qualité et complétude des données

- **Valeurs manquantes (NA)** : les lignes comportant des valeurs NA dans un prédicteur ou dans la colonne cible sont automatiquement supprimées avant l'ajustement. mgcvUI restreint les données aux seules colonnes utilisées par le modèle, de sorte que des NA dans des colonnes inutilisées n'entraînent aucune perte de ligne.
- **Colonnes numériques** : le prix de vente, la surface habitable, la superficie du terrain et les champs analogues doivent être numériques.
- **Colonnes facteur** : les colonnes texte, TRUE/FALSE et facteur R sont automatiquement traitées comme catégorielles — elles ne peuvent entrer dans le modèle qu'en tant que facteurs. Une colonne **numérique** (par exemple un code numérique de zone ou de style) ne devient un facteur que si vous cochez sa case **Factor** ; mgcvUI ne déduit jamais le caractère catégoriel de l'étendue des valeurs d'une variable numérique, si bien que des prédicteurs numériques discrets comme `bath_count` restent continus (lissés). Le rôle spécial est indépendant de la décision relative au facteur.

Astuce : Examinez la colonne NA du tableau des prédicteurs après l'importation. Les colonnes comportant de nombreuses valeurs manquantes peuvent provoquer la suppression de lignes. Les colonnes présentant $\geq 50\%$ de NA sont signalées en rouge et automatiquement exclues.

2.5 Emplacement de la ligne objet

En mode **Appraisal**, la ligne 1 doit être le bien objet. Toutes les autres lignes sont des ventes comparables. La ligne objet est exclue de l'ajustement du modèle. Après l'ajustement, le modèle produit malgré tout des prédictions pour la ligne objet.

En mode **Market Area Analysis**, placer l'objet en ligne 1 est facultatif.

En mode **General**, il n'y a pas de traitement particulier des lignes — toutes sont traitées de la même façon.

3 Mode General

3.1 Vue d'ensemble

Le mode General est le mode par défaut au lancement de mgcvUI. Il offre le flux de travail GAM complet pour n'importe quel jeu de données — pas seulement immobilier. Vous pouvez utiliser mgcvUI pour des données scientifiques, une analyse financière, des études techniques ou tout problème de régression où l'on s'attend à des relations non linéaires lisses.

En mode General, l'interface omet les fonctionnalités propres à l'immobilier (colonnes spéciales, âge de la vente, RCA). La barre latérale est simplifiée et centrée sur la sélection des variables, la configuration des paramètres, l'ajustement du modèle et l'export.

3.2 Le flux de travail de la barre latérale

La barre latérale est organisée en sections numérotées et repliables qui vous guident depuis le choix du projet jusqu'à l'export du rapport. La plupart des sections n'apparaissent qu'une fois un projet actif ouvert, et plusieurs sont propres à l'objet du projet. Le dossier de sortie découle du projet actif — il n'y a pas de champ distinct pour le dossier de sortie.

1. Project (projet) — créez ou ouvrez un projet. Un projet fixe l'objet et le lieu de travail (pays, état, comté, ville) ainsi que ses dossiers d'entrée (`<os>_in/`) et de sortie. Les projets sont stockés sous une racine `regProj` commune (par défaut `~/regProj`, modifiable au moyen de la variable d'environnement `REGPROJ_ROOT` ou du menu Settings) et sont partagés avec les applications sœurs `earthUI` et `glmnetUI`, de sorte que les modèles des trois applications coexistent. Cliquez sur **Close Project** (fermer le projet) pour changer.

2. Import Data (importer les données) — sélectionnez un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajoutez des fichiers au dossier via le Finder ou l'Explorateur, puis cliquez sur **Refresh** pour les lister. Pour les fichiers Excel comportant plusieurs feuilles, un sélecteur de feuille apparaît. Les noms de colonnes sont automatiquement convertis en `snake_case`.

3. Import from earthUI (facultatif) — importez un fichier de résultats `.rds` d'`earthUI` pour amorcer le GAM avec les positions de nœuds découvertes par `earth` (et pour préremplir la sélection des variables). Cette étape est facultative — ignorez-la si vous n'avez pas de résultat `earthUI` à importer.

4. Variable Configuration (configuration des variables) — sélecteur de variable cible, transformation de la réponse (aucune, log, log10), tableau des prédicteurs avec cases à cocher Include, Factor et Linear. La colonne Special n'apparaît qu'en modes Appraisal et Market. Voir le chapitre 6 pour tous les détails.

5. MgcV Call Parameters (paramètres de l'appel mgcv) — toute la configuration du modèle : préréglages de paramètres, famille, méthode, gamma, validation croisée, select, type de base, k, type de produit tensoriel, matrice d'interactions et paramètres avancés. Voir le chapitre 7 pour la référence complète des paramètres.

6. Fit MgcV GAM Model (ajuster le modèle GAM mgcv) — le bouton qui lance le modèle. Les résultats apparaissent dans les onglets de droite.

7. Download Output (télécharger les sorties) — exporte les prédictions, les résidus, les scores CQA et les contributions par variable dans un fichier Excel placé dans le dossier de sortie du projet. Disponible dans tous les modes une fois le modèle ajusté.

8. Calculate RCA Adjustments & Download (calculer et télécharger les ajustements RCA) — calcule les ajustements RCA par variable pour chaque comparable et interpole le résidu de l'objet par CQA. N'apparaît qu'en modes Appraisal et Market, après l'ajustement.

9. Generate Sales Grid & Download (générer et télécharger la grille de ventes) — génère la grille de comparaison des ventes, en recommandant les comparables d'après le pourcentage d'ajustement brut. N'apparaît qu'en mode Appraisal.

10. Generate Quarto Report (générer le rapport Quarto) — écrit un ensemble de rapport Quarto (`.qmd`) autonome — avec tous les graphiques et tableaux — dans le dossier de sortie du projet. Disponible une fois le modèle ajusté.

11. Convert Quarto Report (convertir le rapport Quarto) — rend l'ensemble `.qmd` généré en HTML, Word ou PDF. Le PDF exige une installation LaTeX ; si aucune n'est détectée, l'option PDF est masquée. Voir le chapitre 13.

Les étapes propres à un objet sont omises (et les étapes restantes renumérotées) dans les autres modes : en mode Market, l'étape de la grille de ventes est absente et, en mode General, les étapes RCA et grille de ventes le sont toutes deux.

3.3 Importation depuis earthUI (facultatif)

La section 3 de la barre latérale — **Import from earthUI** — permet d'importer un fichier de résultats `.rds` d'earthUI. Cela active la chaîne `earth-mgcv` : les positions de nœuds déduites des données par earth deviennent des points d'ancrage pour les termes lisses du GAM.

Lorsqu'un résultat earthUI est importé, mgcvUI :

- affiche une synthèse indiquant la variable cible, le R-carré, le nombre de termes et les positions de nœuds par variable ;
- sélectionne automatiquement le pré-réglage de paramètres « Earth Pipeline » (base de splines de régression cubiques, $\gamma = 1,4$) ;
- fixe la variable cible pour qu'elle corresponde à celle d'earth ;
- précoche Include pour les variables prédictrices d'earth ;
- précoche Factor pour les variables catégorielles d'earth et Linear pour ses prédicteurs linéaires ;
- préremplit les interactions détectées par earth (surlignées en jaune dans la matrice d'interactions et verrouillées contre toute modification manuelle) ;
- fournit un bouton **Export Knots CSV** permettant de télécharger les nœuds d'earth au format CSV.

Le tableau de configuration des variables affiche alors une note expliquant que les sélections de variables (Include, Linear, Factor) et les nœuds de splines ont été *amorçés* à partir du modèle earth. Contrairement à glmnetUI, où l'ajustement glmnet est lié à la base d'earth, mgcvUI ajuste le GAM à partir de vos sélections courantes — earth ne fournit qu'une longueur d'avance, et tous les contrôles restent modifiables.

Un bouton **Clear** (effacer) supprime les données earth importées et rétablit le mode autonome.

3.4 Onglets du panneau principal

Après l'importation des données, le panneau principal propose les onglets suivants (les onglets dépendant du modèle se remplissent après l'ajustement) :

Onglet	Contenu
Data	Tableau interactif (DataTable) des données importées. En modes évaluation et marché, scindé en tableaux Bien objet et Ventes comparables.
Equation	Formule GAM complète rendue en MathJax, avec un tableau des définitions des fonctions lisses.
Summary	Métriques clés (R^2 , R^2 de validation croisée, déviance expliquée, AIC, n, nombre de termes lisses) et tableaux des termes lisses et paramétriques.
Variable Importance	Diagramme en barres horizontales des statistiques F (termes lisses) et des valeurs $ t $ (termes paramétriques), avec tableau classé.
Contribution	Courbes interactives (plotly) d'effets partiels lisses, avec bandes de confiance et repères des nœuds earth.
Correlation	Carte de chaleur des corrélations entre prédicteurs numériques (disponible avant l'ajustement).
Diagnostics	Résidus vs valeurs ajustées, diagramme quantile-quantile, histogramme, réponse vs valeurs ajustées (via gratia). Nuage de points observé vs prédit.
RCA Adjustments	Histogrammes de l'analyse RCA (modes évaluation/marché, après calcul).
Anova	Tableau ANOVA combinant les termes paramétriques et lisses.
Mgcv Output	Texte brut : durée d'exécution, formule, synthèse du modèle, famille, concurvit.
Sign Check	Contrôle de cohérence de direction entre earth et le GAM (lorsque earth a été importé).
Concurvity	Matrices de concurvit globale et par paires, avec code couleur.
Basis Check	Tests d'adéquation de la dimension de base fournis par gam.check() .

3.5 Persistance des réglages

mgcvUI enregistre automatiquement votre configuration dans une base de données SQLite, indexée par le nom du fichier d'entrée. Lorsque vous rechargez le même fichier, tous les réglages sont restaurés : choix de la cible, cases à cocher des prédicteurs, types de données, paramètres mgcv et matrice d'interactions. L'objet est une propriété du projet actif plutôt qu'un réglage global, et le dernier fichier d'entrée utilisé est mémorisé par projet puis chargé automatiquement à la réouverture du projet.

3.6 Mode sombre

Cliquez sur l'icône lune/soleil dans le coin supérieur droit pour basculer entre les thèmes Nord Light et Nord Dark. La préférence de thème est enregistrée dans le localStorage et persiste d'une session à l'autre. Tous les éléments de l'interface s'adaptent au thème sélectionné.

3.7 Locale et réglages régionaux

mgcvUI prend en charge les conventions internationales de formatage des nombres et des CSV au moyen d'un système de locale fondé sur le pays. Le menu **Settings** de la barre de titre propose des sélecteurs **Country** (pays) et **Paper** (papier) pour plus de 30 pays. Chaque préréglage configure :

- le **séparateur CSV** — virgule (,) pour les États-Unis, le Royaume-Uni et le Japon, ou point-virgule (;) pour la majeure partie de l'Europe ;
- le **symbole décimal** — point (.) ou virgule (,) ;

- **le séparateur de milliers** — virgule (États-Unis/Royaume-Uni/Japon), point (Allemagne/Italie/Espagne), espace (Finlande/France/Pologne/pays baltes) ou apostrophe (Suisse) ;
- **le format de papier** — Letter (États-Unis/Canada/Mexique) ou A4 (partout ailleurs).

Cliquez sur **Save** (enregistrer) pour conserver globalement vos préférences de locale. La section Import Data comporte également un sélecteur **Locale** propre à chaque fichier, pour les importations ponctuelles suivant d'autres conventions. Le même menu Settings contient le champ **regProj root folder** (dossier racine regProj), qui déplace l'arborescence de projet partagée par mgcvUI, earthUI et glmnetUI.

***Astuce :** Le formatage des nombres sur les axes des graphiques et sur les étiquettes de pente s'adapte automatiquement à votre locale. La locale allemande utilise le point pour les milliers (200.000), la finlandaise l'espace (200 000) et la suisse l'apostrophe (200'000). Aucun symbole monétaire n'est affiché — mgcvUI est indépendant de la devise.*

4 Mode Appraisal

Lorsque l'objet du projet actif est **For Appraisal**, mgcvUI se configure pour l'évaluation d'un bien unique. Toutes les fonctionnalités décrites au chapitre 3 restent disponibles ; le présent chapitre ne traite que des ajouts propres à l'évaluation.

4.1 Traitement de la ligne objet

En mode évaluation, **la ligne 1 de votre jeu de données est le bien objet** et toutes les autres lignes sont des ventes comparables. Votre fichier d'entrée doit être organisé en conséquence (voir le chapitre 2). Le prix de vente de l'objet peut être laissé vide ou fixé à n'importe quelle valeur — mgcvUI le traite automatiquement comme NA lors de l'ajustement.

Après l'ajustement, le modèle produit des prédictions pour la ligne objet, ce qui constitue la base du flux de travail d'ajustement RCA.

4.2 Date d'effet et âge de la vente

En modes évaluation et marché, un champ **Effective Date** (date d'effet) apparaît dans la section Variable Configuration (avec la date du jour par défaut). Si vous désignez une colonne comme `contract_date` dans le menu Special, mgcvUI calcule une colonne `sale_age` — le nombre entier de jours entre la date de contrat de chaque vente et la date d'effet. Cette colonne est ajoutée comme prédicteur.

Lorsque la date d'effet change, `sale_age` est automatiquement recalculée.

4.3 Désignations de colonnes spéciales

En modes évaluation et marché, un menu **Special** apparaît pour chaque prédicteur dans le tableau de configuration des variables. Voir le chapitre 6 pour la référence complète des types spéciaux et de leurs effets.

4.4 Aperçu des ajustements RCA

Le bouton **Calculate RCA Adjustments & Download** (visible en modes évaluation et marché après l'ajustement) calcule, pour chaque comparable, des ajustements dérivés du marché par rapport à l'objet. Le flux de travail RCA complet est décrit au chapitre 11.

***Astuce :** Le score CQA que vous attribuez à l'objet détermine la part de la distribution des résidus qui lui est imputée. Un score de 5,00 place l'objet à la médiane des comparables.*

5 Mode Market Area Analysis

Lorsque l'objet du projet actif est **Market Area Analysis**, mgcvUI offre les mêmes fonctionnalités propres à l'immobilier que le mode évaluation (colonnes spéciales, âge de la vente), mais il est orienté vers l'analyse d'un groupe de biens plutôt que vers l'évaluation d'un objet unique.

5.1 Différences avec le mode évaluation

- **Le traitement de la ligne objet est souple** — en mode marché, la ligne 1 n'est pas automatiquement considérée comme un bien objet.
- **Pas de grille de ventes** — l'étape « Generate Sales Grid & Download » n'est pas disponible.

5.2 Quand utiliser le mode marché

Le mode Market Area Analysis convient lorsque vous :

- construisez un modèle GAM pour un quartier ou une zone de marché afin de comprendre les déterminants non linéaires de la valeur ;
- analysez la manière dont des variables telles que la surface, l'âge, la superficie du terrain et la localisation influent sur les prix de vente au sein d'un groupe de biens ;
- préparez les éléments justificatifs d'une analyse des conditions de marché ou d'une délimitation de quartier ;
- travaillez sur un jeu de données pour lequel vous souhaitez les fonctionnalités de colonnes spéciales (âge de la vente) sans avoir besoin du flux de travail de la grille de ventes.

Astuce : Le mode marché est également utile pour la régression immobilière générale lorsque vous souhaitez les fonctionnalités de colonnes spéciales sans avoir besoin du flux de travail d'évaluation d'un bien unique.

6 Sélection des variables

La section 4 de la barre latérale — **Variable Configuration** — est l'endroit où vous choisissez quelles colonnes participent au modèle et comment elles sont traitées.

6.1 Variable cible

Le menu **Target (response) variable** (variable cible ou réponse) situé en haut de la section 4 liste chaque colonne numérique de votre jeu de données. Sélectionnez une colonne comme variable réponse (par exemple le prix de vente). La colonne cible est automatiquement exclue de la liste des prédicteurs.

6.2 Transformation de la réponse

Sous le sélecteur de cible, un menu **Response Transform** (transformation de la réponse) propose trois options :

- **None (raw)** (aucune) — utilise la réponse telle quelle. Approprié lorsque la relation entre les prédicteurs et la réponse est approximativement additive sur l'échelle d'origine.
- **Log (natural)** (logarithme naturel) — applique une transformation logarithmique népérienne avant l'ajustement. Après transformation inverse, les ajustements deviennent proportionnels (%) plutôt qu'absolus (\$). Particulièrement utile pour les prix de vente, où un ajustement de 10 000 \$ n'a pas du tout la même portée selon le niveau de prix.
- **Log10** — applique une transformation logarithmique en base 10.

***Astuce :** La transformation logarithmique rend les ajustements temporels proportionnels (%) au lieu d'absolus (\$). Pour les modèles immobiliers couvrant une large plage de prix, les transformations logarithmiques produisent souvent des résidus mieux comportés.*

Lorsqu'une transformation logarithmique est sélectionnée, les valeurs ≤ 0 de la réponse sont automatiquement écartées. Toutes les sorties (prédictions, contributions, résidus) subissent automatiquement la transformation inverse pour revenir à l'échelle d'origine.

6.3 Le tableau des prédicteurs

Sous le sélecteur de cible, un tableau défilant liste toutes les colonnes restantes. Les colonnes de cases à cocher utilisent des en-têtes verticaux pivotés, par souci de compacité. L'ordre des colonnes est le suivant :

Colonne	Description
Variable	Nom de la colonne (police à chasse fixe, tronqué par des points de suspension). Cellule encadrée.
Type	Menu du type de données : numeric, integer, character, factor, logical, Date, POSIXct. Menu encadré.
Include	Case à cocher — inclure cette colonne comme prédicteur du modèle. En-tête vertical.
Factor	Case à cocher — traiter comme facteur catégoriel (crée un coefficient distinct par modalité). En-tête vertical.
Linear	Case à cocher — imposer une relation linéaire (pas de terme lisse, simplement βx). En-tête vertical.
Special	Menu (modes évaluation/marché uniquement) — voir la référence des types de colonnes spéciales ci-dessous. Menu encadré.
NAs	Nombre et pourcentage de valeurs manquantes, avec code couleur : rouge ($\geq 50\%$), orange ($\geq 20\%$), gris (> 0). Cellule encadrée.

Facteur, terme lisse ou terme linéaire : par défaut, les variables numériques incluses reçoivent un terme lisse $f(x)$. Cocher **Factor** crée un terme catégoriel (un coefficient par modalité). Cocher **Linear** crée un simple terme linéaire βx au lieu d'un terme lisse. Les variables marquées à la fois Factor et Linear sont traitées comme facteurs.

6.4 Détection et forçage des types de données

mgcvUI détecte automatiquement les types de données à l'importation. Les colonnes numériques, entières, logiques, facteur et date sont reconnues. Vous pouvez forcer n'importe quelle détection en modifiant le menu **Type**. Le changement de type modifie la manière dont la colonne est traitée dans le modèle.

6.5 Référence des types de colonnes spéciales

En modes évaluation et marché, le menu **Special** propose les options suivantes :

Pondération :

- **weight** — colonne de poids des observations (une seule autorisée ; les lignes de poids nul sont exclues de l'ajustement)

Types de date et d'heure :

- **contract_date** — déclenche le calcul automatique de **sale_age** à partir de la date d'effet
- **sale_age** — identifie une colonne d'âge de vente déjà calculée (jours écoulés depuis la date d'effet), utilisée telle quelle
- **listing_date** — utilisée comme solution de repli pour calculer le nombre de jours sur le marché
- **dom** — identifie la colonne des jours sur le marché

Type d'enregistrement :

- **sale_type** — identifie la colonne de type d'enregistrement qui distingue l'objet des ventes comparables

Types monétaires :

- **concessions** — identifie les concessions de vente (crédits du vendeur, incitations à l'acheteur, etc.)

Types de taille et de localisation :

- **latitude** — valeurs automatiquement arrondies à 3 décimales

- `longitude` — même arrondi que la latitude
- `area` — identifiant de zone de marché ou de quartier
- `living_area` — active les calculs de résidus au pied carré (`residual_sf` et `cqa_sf`)
- `lot_size` — colonne de superficie du terrain
- `site_dimensions` — regroupée avec la superficie du terrain

Types d'âge :

- `actual_age` — colonne de l'âge du bien
- `effective_age` — âge effectif du bien

Types d'affichage :

- `display_only` — la colonne est incluse dans les exports Excel mais entièrement exclue de l'ajustement du modèle. À utiliser pour les champs d'adresse, les numéros MLS, les identifiants de parcelle ou d'autres données de référence.

Astuce : Les colonnes désignées *`display_only`* demeurent dans votre jeu de données et apparaissent dans l'export Excel, mais elles sont entièrement exclues du tableau des prédicteurs. À utiliser pour les colonnes d'identifiant, les adresses ou d'autres champs de référence.

7 Choix des paramètres

La section 5 de la barre latérale — **Mgcv Call Parameters** — donne accès à toutes les options de configuration du GAM. Chaque paramètre est accompagné d'une icône d'aide bleue (?) affichant une infobulle explicative.

7.1 Préréglages de paramètres

Un menu situé en haut propose deux préréglages qui fixent des valeurs par défaut sensées pour les flux de travail courants :

Préréglage	Base	Réglages
Standalone (découverte)	tp (plaque mince)	k = auto (10), gamma = 1,2, select = TRUE
Earth Pipeline (raffinement)	cr (régression cubique)	k = auto (10), gamma = 1,4, select = TRUE

Le préréglage **Earth Pipeline** est automatiquement sélectionné lorsque des nœuds earthUI sont importés. La base de splines de régression cubiques (**cr**) est indispensable à l'intégration des nœuds d'earth, car elle seule permet de spécifier des positions de nœuds exactes.

7.2 Famille

Choisissez la famille de distribution de votre variable réponse :

Famille	Cas d'usage
gaussian	Réponses continues (par exemple un prix de vente). La plus courante.
Gamma	Réponses continues positives dont la variance est proportionnelle à la moyenne.
poisson	Données de comptage (par exemple un nombre de ventes).
binomial	Résultats binaires.
inverse.gaussian	Données positives fortement asymétriques à droite.

7.3 Méthode

La méthode d'estimation du paramètre de lissage :

Méthode	Description
REML	Maximum de vraisemblance restreint (par défaut, recommandé). La plus robuste face au surajustement.
GCV.Cp	Validation croisée généralisée. Tend à sélectionner des modèles légèrement plus complexes.
ML	Maximum de vraisemblance. Proche de REML, mais peut sous-lisser.
P-REML	Variante REML de Pearson.
P-ML	Variante ML de Pearson.
GACV.Cp	Validation croisée approchée généralisée.

7.4 Gamma (multiplicateur de lissage)

Le paramètre gamma multiplie les degrés de liberté effectifs dans le critère de lissage, ce qui favorise des ajustements plus lisses (moins ondulés). La valeur par défaut est 1,2.

Valeur	Effet
1,0	Lissage standard
1,2–1,4	Légèrement plus lisse que par défaut (recommandé)
2,0 et plus	Beaucoup plus lisse — utile pour des données bruitées ou de petits échantillons

Des valeurs de gamma plus élevées protègent du surajustement en pénalisant davantage la complexité. Le préréglage Earth Pipeline emploie 1,4 pour un lissage supplémentaire lors du raffinement des nœuds d'earth.

7.5 Validation croisée

Lorsque la case est cochée (par défaut), mgcvUI exécute après l'ajustement une validation croisée à 10 blocs afin de calculer un R-carré de validation croisée. Cela fournit une estimation honnête du pouvoir prédictif hors échantillon.

7.6 Select (rétrécissement)

Lorsque la case est cochée (par défaut), une pénalité supplémentaire est ajoutée à chaque terme lisse, capable de le ramener à zéro. Cela permet une sélection automatique de variables — les termes lisses sans importance sont de fait retirés du modèle. Mis en œuvre par `select = TRUE` dans `gam()`.

7.7 Base par défaut

Le type de fonction de base spline des termes lisses :

Base	Nom complet	Description
tp	Spline de régression à plaque mince	Par défaut. Isotrope (aucun placement de nœuds requis). Bon choix polyvalent.
cr	Spline de régression cubique	Permet un placement explicite des nœuds. Indispensable à l'intégration des nœuds d'earth.
ps	P-spline	Base de B-splines à nœuds équidistants avec pénalité de différences. Efficace en calcul.
bs	B-spline	Base de B-splines souple.

7.8 k par défaut (dimension de la base)

La dimension de la base k détermine l'ondulation maximale de chaque terme lisse. La valeur 0 (par défaut) signifie « automatique » — mgcvUI utilise $k = 10$ ou le nombre de nœuds d'earth, selon ce qui convient.

- **k petit (3–5)** : courbes très lisses, presque linéaires.
- **k moyen (8–15)** : souplesse modérée (généralement suffisante).
- **k grand (20 et plus)** : très souple — risque de surajustement en l'absence d'une régularisation forte.

Astuce : Les degrés de liberté effectifs (EDF) affichés dans l'onglet Summary indiquent combien de degrés de liberté chaque terme lisse consomme réellement. Si l'EDF est proche de $k - 1$, la base est peut-être trop petite — augmentez k . Si l'EDF est bien inférieur à k , le terme lisse est correctement pénalisé.

7.9 Type de produit tensoriel

Pour les interactions entre variables continues, deux types de produits tensoriels sont disponibles :

Type	Fonction	Description
ti	<code>ti(x1, x2)</code>	Interaction tensorielle — ne modélise que l'effet d'interaction, les effets principaux étant traités séparément par des termes <code>s()</code> univariés. À préférer pour l'interprétabilité.
te	<code>te(x1, x2)</code>	Produit tensoriel — modélise l'effet conjoint complet, effets principaux compris. Plus difficile à décomposer pour les ajustements RCA.

7.10 Interactions autorisées

Une section repliable **Allowed Interactions** (interactions autorisées) affiche une matrice triangulaire supérieure de cases à cocher pour tous les prédicteurs lisses inclus (ni linéaires, ni facteurs). Chaque case active un lissage par produit tensoriel entre le couple de variables correspondant.

- **Allow All / Clear All** — boutons de bascule globale.
- **Cliquez sur un nom de variable** — bascule toutes les interactions de cette variable.
- **Clic droit sur un nom de variable** — « Block from main effect » (exclure de l'effet principal) — marque la variable pour un usage en interaction seulement (pas de terme lisse univarié).

Lorsque des nœuds earthUI sont importés, les interactions détectées par earth sont précochées et verrouillées (surlignées sur fond jaune). Si le modèle earth utilisait `degree = 1` (sans interactions), un message d'information est affiché.

Avertissement

L'activation d'interactions tensorielles accroît sensiblement la complexité du modèle. N'utilisez des interactions que lorsque vous disposez d'une connaissance du domaine ou d'indices détectés par earth qui les étayent. Vérifiez que les ajustements obtenus sont raisonnables et explicables.

7.11 Interactions facteur $\times$ terme lisse

Une matrice distincte **Factor-by-Smooth Interactions** apparaît lorsque des variables facteur et des variables lisses sont toutes deux incluses. Chaque case crée un terme `s(x, by = factor)` — une courbe lisse distincte pour chaque modalité du facteur.

7.12 Paramètres avancés

Une section repliable « Advanced » (avancé) expose des réglages supplémentaires :

- **Optimizer** (optimiseur) — algorithme d'estimation du paramètre de lissage : `outer/newton` (par défaut), `outer/bfgs` ou `efs` (Fellner-Schall étendu).
- **Scale** (échelle) — paramètre d'échelle. 0 (par défaut) = estimation à partir des données. Fixez une valeur connue pour les modèles à échelle imposée.
- **Discrete covariate method** (méthode à covariables discrètes) — ajustement discrétisé rapide pour les grands jeux de données (`discrete = TRUE`).
- **Threads** (fils d'exécution) — nombre de fils parallèles (1–32). Effectif seulement lorsque `discrete = TRUE`.

8 Ajustement du modèle

8.1 Le bouton d'ajustement

La section 6 de la barre latérale contient le bouton **Fit MgcV GAM Model**. Un clic lance le modèle avec votre configuration courante.

8.2 Ce qui se passe pendant l'ajustement

Lorsque vous cliquez sur Fit, mgcvUI :

1. **Valide la configuration** — vérifie la présence d'au moins un prédicteur et d'au moins 10 lignes complètes. Avertit si moins de 50% des lignes sont complètes (en indiquant les colonnes responsables de la plupart des NA).
2. **Prépare les données** — restreint aux seules colonnes utiles au modèle, supprime les lignes comportant des NA, applique les poids (s'ils sont désignés), exclut la ligne objet (en mode évaluation) et applique les transformations de la réponse.
3. **Construit la formule du GAM** — crée des termes lisses $s(x, bs="cr", k=10)$ pour les prédicteurs numériques, des termes facteur pour les variables catégorielles, des termes linéaires pour les variables marquées Linear et des produits tensoriels pour les interactions sélectionnées. Les nœuds d'earth sont intégrés comme positions de nœuds explicites dans les termes à base `cr`.
4. **Traite les cas limites** — écarte les variables constantes, convertit les prédicteurs binaires en facteurs, plafonne k au nombre de valeurs distinctes moins un, complète les nœuds d'earth par des quantiles des données lorsqu'il en faut davantage.
5. **Ajuste le GAM** — appelle `mgcv::gam()` avec la formule construite, la famille, la méthode, gamma, select et les autres paramètres.
6. **Exécute la validation croisée** (si elle est activée) — validation croisée à 10 blocs calculant le R-carré hors échantillon.
7. **Met à jour tous les onglets de résultats** — Summary, Equation, graphiques de contribution, Diagnostics et tous les autres onglets sont remplis.

8.3 Ajustement en arrière-plan et progression

Lorsque le package `callr` est disponible, l'ajustement du modèle s'exécute dans un processus d'arrière-plan, de sorte que l'application reste réactive. Une surcouche de progression au style terminal sur fond sombre affiche :

- un **chronomètre** indiquant les secondes/minutes écoulées ;
- un **journal de trace en couleurs** : bleu pour l'état, jaune pour la progression des blocs de validation croisée, rouge pour les erreurs, vert pour les résultats ;
- un bouton **Abort** (interrompre) permettant de tuer le processus d'ajustement au besoin ;
- un bouton **Close** (fermer) qui apparaît une fois l'exécution terminée.

Si `callr` n'est pas disponible, l'ajustement s'exécute de manière synchrone (l'application se fige jusqu'à la fin de l'ajustement).

Après un ajustement réussi, une coche blanche apparaît sur le bouton Fit et une ligne d'état affiche : « R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X. »

9 Onglets de résultats

9.1 Data

Affiche les données importées sous forme de tableau interactif (DataTable) avec défilement horizontal et 15 lignes par page. En modes évaluation et marché, l'onglet est scindé en un tableau **Subject Property** (bien objet, ligne 1) et un tableau **Comparable Sales** (ventes comparables, les autres lignes) ; en mode General, un seul tableau est affiché.

Chaque cellule tient sur une seule ligne et est tronquée à la largeur de la colonne, de sorte que les champs de texte libre étendus (tels que les commentaires sur un bien) n'étirent pas une ligne sur toute la hauteur de la page. Un simple clic sélectionne une ligne ; un **double-clic sur n'importe quelle cellule** ouvre une fenêtre affichant son contenu intégral, non tronqué.

9.2 Equation

Affiche le modèle GAM ajusté en deux parties :

Équation du modèle (rendue par MathJax) :

- Le membre de gauche présente la réponse avec la transformation adéquate (log, log10) et la fonction de lien.
- Le membre de droite présente la valeur de l'ordonnée à l'origine, les termes lisses sous la forme $f_i(\text{variable})$, les termes facteur regroupés par variable avec le nombre de modalités, et les termes linéaires avec la valeur de leur coefficient.
- La famille et la méthode sont affichées au-dessus de l'équation.

Tableau **Smooth Function Definitions** (définitions des fonctions lisses) :

- chaîne de formule du terme (par exemple `s(living_sqft, bs="cr", k=8)`)
- nom de la variable
- type (s, te, ti ou linéaire)
- type de fonction de base
- valeur de k
- positions des nœuds d'earth (en cas d'importation)

9.3 Summary

Six cartes de métriques affichées en haut :

- **R-squared** — proportion de déviance expliquée
- **CV R-squared** — R-carré de validation croisée (si la validation croisée est activée)
- **Dev. Explained %** — pourcentage de déviance expliquée
- **AIC** — critère d'information d'Akaike
- **n** — nombre d'observations utilisées
- **Smooths** — nombre de termes lisses

Sous les cartes :

- **Tableau Smooth Terms** (termes lisses) — terme, EDF (degrés de liberté effectifs), Ref.df, statistique F, valeur p
- **Tableau Parametric Terms** (termes paramétriques) — terme, estimation, erreur type, valeur t, valeur p

9.4 Variable Importance

Un diagramme en barres horizontales montrant l'importance relative de chaque terme du modèle :

- **Termes lisses** : classés par statistique F (barres bleues)
- **Termes paramétriques** : classés par valeur $|t|$ (barres vertes)

Sous le graphique, un DataTable indique le terme, le type, l'EDF, la statistique et la valeur p, triés par statistique décroissante.

9.5 Graphiques de contribution

Chaque terme du modèle qui contribue à la valeur prédite dispose d'un graphique de contribution interactif **plotly**, présenté dans une carte encadrée. L'onglet affiche des graphiques pour tous les types de termes :

9.5.1 Termes lisses univariés — $s(x)$

- **Courbe bleue** avec bande de confiance à 95%
- **Nuage de points gris** représentant les résidus partiels (pour les modèles sans transformation logarithmique)
- **Lignes verticales rouges en tirets** aux positions des nœuds d'earth (si earth a été importé)
- **Infobulle au survol** : valeur de la variable, contribution (\$), IC à 95%, taux (pente par unité)

Pour les **modèles à transformation logarithmique**, l'axe des ordonnées subit la transformation inverse pour exprimer des contributions en dollars, selon la formule $\bar{y} \times (e^{f(x)} - 1)$, ce qui rend les courbes directement interprétables en dollars.

Pour les variables de **latitude et longitude**, la pente est affichée par tranche de 0,001 degré plutôt que par degré, cet incrément géographique étant plus parlant.

9.5.2 Termes lisses par facteur — $s(x, \text{by} = \text{factor})$

Une courbe colorée par modalité du facteur, superposées sur un même graphique. Utile pour voir comment l'effet d'un prédicteur varie d'un groupe à l'autre (par exemple selon les quartiers).

9.5.3 Termes d'interaction — $ti(x_1, x_2)$ et $te(x_1, x_2)$

Carte de chaleur représentant la surface de contribution en 2D. Échelle de couleurs : rouge (négatif) → blanc (nul) → bleu (positif). Le survol affiche les valeurs des deux variables ainsi que le montant de la contribution.

9.5.4 Termes paramétriques — linéaires et facteurs

- **Termes linéaires numériques** : nuage de points des données avec la droite ajustée superposée
- **Termes facteur** : diagramme en barres de la contribution moyenne par modalité

9.6 Correlation

Une matrice en carte de chaleur des corrélations de Pearson entre tous les prédicteurs numériques. Disponible dès l'importation des données — aucun ajustement de modèle n'est requis.

- Dégradé de couleurs : bleu (−1) → blanc (0) → rouge (+1)
- Valeurs de corrélation imprimées dans chaque cellule
- Utile pour repérer la concurvité (l'analogue lisse de la multicolinéarité) avant l'ajustement

9.7 Diagnostics

Deux ensembles de visualisations de diagnostic :

Panneau de diagnostic (via `gratia::appraise`) : quatre graphiques — résidus vs valeurs ajustées, diagramme quantile-quantile, histogramme des résidus et réponse vs valeurs ajustées.

Nuage de points **Actual vs Predicted** (observé vs prédit) : valeurs observées comparées aux valeurs prédites, avec une droite de référence en tirets à 45 degrés. Pour les modèles à transformation logarithmique, les valeurs subissent la transformation inverse vers l'échelle d'origine.

9.8 RCA Adjustments

Trois histogrammes affichés après le calcul RCA (modes évaluation/marché uniquement) :

- **Residual Adjustment %** — distribution des ajustements de résidu en pourcentage du prix de vente
- **Net Adjustment %** — distribution des ajustements nets
- **Gross Adjustment %** — distribution des ajustements bruts

Chaque histogramme indique la moyenne, la médiane et l'écart type dans son sous-titre, avec des droites de référence en tirets pour la moyenne et la médiane.

9.9 Anova

Un tableau ANOVA combinant les termes paramétriques et lisses issus de la synthèse du GAM. Il indique le terme, le type (paramétrique ou lisse) et toutes les statistiques associées.

9.10 Mgcvc Output

Sortie en texte brut affichant :

- la durée de l'ajustement
- la formule du GAM
- la sortie d'impression du modèle
- la sortie complète de `summary(model)`
- la famille et la fonction de lien
- les mesures globales de concurvité

9.11 Sign Check

Lorsque des nœuds earthUI ont été importés, cet onglet compare la direction de chaque variable dans le modèle earth avec sa direction dans le GAM :

- **earth_direction** : croissante, décroissante ou mixte (d'après le signe des fonctions charnières)
- **gam_direction** : croissante, décroissante ou mixte (estimée à partir de la dérivée du terme lisse sur une grille de 100 points)
- **consistent** : TRUE/FALSE
- **warning** : description de toute discordance

Une ligne d'état affiche « All directions consistent » (toutes les directions sont cohérentes) ou le nombre de variables présentant des incohérences.

9.12 Concurvity

La concurvité est l'analogue lisse de la multicolinéarité. Cet onglet affiche :

- **Tableau Overall Concurvity** (concurvité globale) : lignes pour les mesures worst/observed/estimate, colonnes pour chaque terme lisse. Code couleur : blanc ($<0,5$), jaune ($0,5-0,8$), rouge ($>0,8$).
- **Tableau Pairwise Concurvity** (concurvité par paires) : matrice des pires cas par paire, avec le même code couleur.

Des valeurs supérieures à $\sim 0,8$ dans la ligne « worst » indiquent qu'un terme lisse peut être bien approché par les autres termes lisses, ce qui suggère une redondance.

9.13 Basis Check

Sortie de `mgcv::gam.check()` : tests d'adéquation de la dimension de base pour chaque terme lisse. Si le k d'un terme lisse est trop petit, ce test le signale par une valeur p faible, ce qui suggère d'augmenter k .

10 Téléchargement des données

Après l'ajustement, le bouton **Download Output (Excel)** (section 7 de la barre latérale) exporte un fichier Excel contenant les prédictions et les diagnostics dans le dossier de sortie mgcv du projet actif.

10.1 Colonnes de sortie

Colonne	Description
basis	Valeur de l'ordonnée à l'origine après transformation inverse (identique pour toutes les lignes)
<var>_contribution	Contribution de chaque terme lisse à la prédiction. Pour les modèles log, répartie proportionnellement en valeurs monétaires.
est_<target>	Prédiction du modèle (par exemple est_sale_price), après transformation inverse depuis l'échelle logarithmique le cas échéant
residual	Observé – prédit
calc_residual	Vérification : observé – somme des contributions (après transformation inverse)
cqa	Score CQA (Condition-Quality-Appeal : état, qualité, attrait), sur une échelle de 0 à 10
residual_sf	Résidu / surface habitable (si living_area est désignée)
cqa_sf	CQA calculé à partir du classement selon residual_sf

En mode évaluation, la ligne 1 (l'objet) a **residual**, **cqa** et **cqa_sf** fixés à NA. Les lignes sont triées par **residual_sf** décroissant lorsque une colonne **living_area** est désignée.

Une coche blanche apparaît sur le bouton de téléchargement une fois l'opération réussie.

10.2 Scores CQA

Le CQA classe le résidu de chaque ligne par rapport à tous les autres sur une échelle de 0 à 10 :

- **CQA élevé (~9–10)** : vendu bien au-dessus du prix prédit — état, qualité ou attrait probablement supérieurs
- **CQA faible (~0–1)** : vendu bien au-dessous du prix prédit — état probablement inférieur ou vente en difficulté
- **CQA ~5** : proche de la médiane

11 Calculs RCA et téléchargement

Le flux de travail RCA (*Residual Constraint Approach*, approche par contrainte des résidus) est disponible en modes **Appraisal** et **Market Area Analysis**, après l'ajustement du modèle. Le bien objet doit se trouver en ligne 1.

11.1 Ouverture de la boîte de dialogue RCA

Cliquez sur le bouton **Calculate RCA Adjustments & Download** à la section 8 de la barre latérale. Une boîte de dialogue modale apparaît, comportant :

- **Score type** (type de score) : CQA ou CQA par pied carré (si **living_area** est désignée). L'option par pied carré met à l'échelle le résidu de l'objet selon la surface habitable.
- **Subject CQA Score** (score CQA de l'objet) : accepte des valeurs de 0,00 à 10,00. Le champ est initialement vide ; lorsqu'un modèle earthUI a été importé, le type de score et la valeur utilisés dans earthUI sont repris depuis les réglages du projet partagé.

Cliquez sur **Generate** (générer) pour calculer et télécharger. Le fichier Excel est écrit dans le dossier de sortie mgcv du projet actif (<os>_out_mgcv).

11.2 Interpolation du score CQA

1. Les scores CQA et les résidus des comparables sont triés
2. Une interpolation linéaire (**approx()**) fait correspondre votre valeur de CQA à un résidu
3. Valeur de l'objet = prédiction du modèle + résidu interpolé
4. Les lignes de poids nul sont traitées comme si elles avaient le même résidu que l'objet

11.3 Colonnes de sortie

Colonne	Description
subject_value	Prédiction du modèle + résidu interpolé
subject_cqa	Score CQA saisi par l'utilisateur
<var>_adjustment	Contribution de l'objet – contribution du comparable (en dollars)
residual_adjustment	Résidu de l'objet – résidu du comparable
net_adjustments	Somme de tous les ajustements
gross_adjustments	Somme des valeurs absolues des ajustements
residual_adj_pct	Ajustement de résidu en % du prix de vente
net_adj_pct	Ajustements nets en % du prix de vente
gross_adj_pct	Ajustements bruts en % du prix de vente
adjusted_sale_price	Prix de vente observé + ajustements nets

Une coche blanche apparaît sur le bouton une fois le calcul réussi.

12 Grille de comparaison des ventes

La grille de comparaison des ventes n'est disponible qu'en mode **Appraisal**, après le calcul des ajustements RCA. Elle génère un classeur Excel mis en forme, adapté à l'inclusion dans un rapport d'évaluation.

12.1 Boîte de dialogue de sélection des comparables

Un clic sur le bouton **Generate Sales Grid & Download** ouvre une boîte de dialogue modale listant toutes les ventes comparables. Les comparables sont répartis en deux groupes :

- **Recommended comps** (comparables recommandés, précochés) : ajustement brut < 25% du prix de vente, triés par âge de la vente (les plus récents d'abord). Maximum 30 comparables.
- **Other comps** (autres comparables, décochés par défaut) : les comparables restants, triés par pourcentage d'ajustement brut.

Sélectionnez vos comparables, puis cliquez sur **Generate Sales Grid** pour créer le classeur.

Types spéciaux requis : `contract_date` et `living_area` doivent être désignés. **Types spéciaux recommandés** : `latitude`, `longitude` et `lot_size` — la grille fonctionne sans eux, mais certains champs resteront vides.

12.2 Structure du classeur

Le classeur généré contient plusieurs feuilles, chacune présentant 3 comparables côte à côte :

- **Lignes d'en-tête** : adresses de l'objet et des comparables, prix de vente, dates de vente et jours sur le marché
- **Lignes groupées** : localisation (latitude, longitude, zone), taille du terrain (superficie, dimensions du site) et âge (âge réel, âge effectif), avec en-têtes de groupe et sous-éléments
- **Lignes des variables du modèle** : une ligne par terme lisse indiquant la contribution de l'objet, celles des comparables et les montants d'ajustement
- **Lignes des caractéristiques résiduelles** : cellules vides et déverrouillées permettant à l'évaluateur de saisir les caractéristiques non captées par le modèle (état, qualité, vue, etc.)
- **Lignes de synthèse** : ajustement net en %, ajustement brut en % et prix de vente ajusté, avec des formules Excel actives

12.3 Protection des feuilles et cellules résiduelles

Chaque feuille est protégée afin d'éviter les modifications accidentelles, mais les cellules de valeur des caractéristiques résiduelles sont explicitement déverrouillées. Les évaluateurs peuvent ainsi saisir des valeurs pour les caractéristiques absentes du modèle tout en préservant les calculs d'ajustement pilotés par formules.

13 Téléchargement des rapports

La génération de rapports suit un flux Quarto en deux étapes, qui remplace le rapport en une seule étape des versions antérieures.

13.1 Étape 1 : générer le rapport Quarto

Le bouton **Generate Quarto Report** (disponible une fois le modèle ajusté) écrit un ensemble `.qmd` autonome dans le dossier de sortie `mgcv` du projet actif. L'ensemble contient la source Quarto, les fichiers de graphiques prérendus (PNG et PDF), une charge utile `report_data.rds` regroupant tous les tableaux et les métadonnées du modèle, ainsi qu'un modèle de style `reference.docx` pour la sortie Word. Parce que l'ensemble est autonome, vous pouvez modifier la source `.qmd` ou la combiner avec d'autres documents Quarto avant le rendu.

Les graphiques sont prérendus lors de la génération de l'ensemble ; si un graphique échoue, une image de remplacement est écrite et le reste de l'ensemble est généré normalement.

13.2 Étape 2 : convertir le rapport Quarto

La section **Convert Quarto Report** rend un fichier `.qmd` vers un ou plusieurs formats de sortie au moyen de `quarto::quarto_render()`. Elle sélectionne par défaut le dernier ensemble généré, mais le bouton **Browse** (parcourir) permet de choisir n'importe quel fichier `.qmd`. Trois formats sont disponibles :

- **HTML** — document HTML autonome. Fonctionne sur toutes les plateformes sans logiciel supplémentaire. Choix recommandé par défaut.
- **Word (.docx)** — rendu avec le modèle de style `reference.docx` fourni. Adapté à l'édition et à la diffusion.
- **PDF** — **exige une installation LaTeX** (voir la configuration requise ci-dessous). Si aucune installation LaTeX n'est détectée, l'option PDF est automatiquement masquée dans les cases de format.

Les générateurs historiques en une seule étape, `render_gam_report()` et `export_gam_docx()`, demeurent disponibles pour un usage programmatique.

13.3 Contenu du rapport

Le rapport généré comprend :

- **Dataset & Model** : fichier de données, taille de l'échantillon, famille, méthode et informations d'ajustement
- **Results Summary** : R^2 , R^2 de validation croisée, déviance expliquée, AIC et métriques d'ajustement associées
- **Model Equation** : équation du modèle, avec la notation f_i pour les termes lisses et des coefficients explicites pour les termes linéaires, ainsi que les définitions des fonctions lisses. Les équations longues se répartissent sur plusieurs lignes dans le PDF.
- **Tableaux de synthèse** : termes lisses (EDF, F, valeur p) et termes paramétriques (estimation, erreur type, t, valeur p)
- **Interactions activées** : une matrice des termes d'interaction tensoriels et facteur $\times$ terme lisse inclus dans le modèle (affichée uniquement en présence d'interactions)
- **Variable Importance** : diagramme en barres classé par statistique F (termes lisses) ou par $|t|$ (termes paramétriques)
- **Smooth Effects** (tous les types de termes) :
 - graphiques de termes lisses univariés (versions statiques ggplot2 des graphiques interactifs)

- graphiques multicourbes des termes lisses par facteur
- cartes de chaleur des interactions (termes ti et te)
- graphiques des termes paramétriques (nuage de points pour les termes numériques, diagramme en barres pour les facteurs)
- carte de chaleur **Predictor Correlations** (corrélations entre prédicteurs)
- **Diagnostics** : diagnostics de résidus et nuage de points observé vs prédit
- **mgcv Summary** : la sortie texte complète du modèle ajusté
- **Annexe de notation** : une section de clôture non numérotée expliquant la notation des fonctions de contribution à la valeur employée dans le rapport

Une coche blanche apparaît sur le bouton Generate Quarto Report une fois la génération réussie.

14 Export des fonctions du modèle

mgcvUI peut exporter les fonctions lisses du GAM ajusté sous forme de code autonome dans plusieurs langages. Chaque terme lisse est évalué sur une grille de 200 points et exporté sous forme de table de correspondance avec interpolation linéaire, ce qui permet d'utiliser le modèle en dehors de R sans aucune dépendance à `mgcv`.

L'export de fonctions est programmatique dans la version actuelle. Les fonctions utilitaires exportées prennent un objet de résultat ajusté (produit par `fit_gam()`) et écrivent les fichiers décrits ci-dessous :

- `generate_r_code()`, `generate_python_code()` et `generate_cpp_code()` — fichiers source par langage
- `export_functions_sqlite()` — base de données SQLite de tables de correspondance
- `export_functions_zip()` — regroupe n'importe quelle combinaison des quatre formats dans une archive zip

14.1 Fonctions R

Fichier : `gam_functions.R`

- Commentaires d'en-tête indiquant la variable réponse, le R^2 , le R^2 de validation croisée et la formule.
- Une constante `INTERCEPT`.
- Un `approxfun()` par variable lisse, avec des vecteurs `x/y` précalculés.
- Les coefficients des termes linéaires sous forme de constantes nommées.
- Utilisation : additionner l'ordonnée à l'origine, chaque `g_var(value)` et chaque `coef_var * value`.

14.2 Fonctions Python

Fichier : `gam_functions.py`

- Import de `numpy`.
- Constante `INTERCEPT`.
- Par variable : tableaux `_x_var` et `_fx_var`, fonction `g_var(x)` utilisant `np.interp()`.
- Coefficients des termes linéaires sous forme de constantes `COEF_*`.

14.3 En-tête C++

Fichier : `gam_functions.hpp`

- En-tête seul, avec `#pragma once`.
- Espace de noms `gam_functions`.
- Fonction utilitaire `interp()` reposant sur `std::lower_bound`.
- Par variable : fonction `inline` avec des tableaux `static const`.
- Coefficients des termes linéaires en `constexpr double`.

14.4 Base de données SQLite

Fichier : `gam_functions.sqlite`

- Table `smooth_grids` : colonnes (variable, `x`, `fx`) — les tables de correspondance à 200 points.
- Table `model_info` : `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.
- Table `linear_terms` : variable, coefficient.

— Table **intercept** : valeur de l'ordonnée à l'origine.

15 Comparaison avec earthUI et glmnetUI

L'application earth et ses Post Processing Modules forment un unique ensemble d'outils de modélisation par régression. Ils partagent le même format de données, les mêmes types de colonnes spéciales, le même flux de travail RCA et les mêmes jeux de données de démonstration, mais reposent sur des moteurs de modélisation différents.

15.1 Principales différences

Caractéristique	mgcvUI	earthUI	glmnetUI
Méthode	GAM (mgcv)	MARS/Earth (earth)	Elastic net (glmnet)
Relations	Non linéaires lisses	Linéaires par morceaux	Linéaires
Sélection de variables	Rétrécissement <code>select=TRUE</code>	Élagage pas à pas	Pénalité Lasso
Interactions	Produits tensoriels (te/ti)	Produits de charnières (degré 2–3)	Produits deux à deux
Effets partiels	Courbes lisses avec IC	g-fonctions (par morceaux)	Pentes linéaires
Chaîne earth	Importe les nœuds d'earth	Source des nœuds	Importe la base de charnières
Export de code	R, Python, C++, SQLite	R, Python, C++, SQLite	—
Formats de rapport	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

Un élément différenciant propre à earthUI mérite d'être signalé : earthUI comprend une fonctionnalité expérimentale Prolog/DCG qui extrait des attributs structurés de biens à partir des commentaires MLS en texte libre.

15.2 Fonctionnalités communes

Les trois outils offrent :

- une arborescence de projet `regProj` partagée (par défaut `~/regProj`) — projets, fichiers d'entrée et sorties propres à chaque méthode sont partagés entre les applications sœurs
- l'importation CSV et Excel avec conversion des colonnes en `snake_case`
- la prise en charge des locales de plus de 30 pays, avec séparateur CSV, symbole décimal et format de papier
- les désignations de colonnes spéciales (`contract_date`, `living_area`, `area`, `latitude`, `longitude`, etc.)
- le calcul automatique de `sale_age` à partir de la date d'effet
- la bascule mode sombre/clair (thème Nord)
- la persistance des réglages (SQLite et/ou `localStorage`)
- la carte de chaleur des corrélations, l'importance des variables et les diagnostics
- le score CQA et le flux de travail d'ajustement RCA
- la grille de comparaison des ventes avec sélection des comparables et sortie Excel mise en forme
- la compatibilité avec les mêmes jeux de données de démonstration (`Appraisal_1.csv`)

15.3 Flux de travail recommandé

- **Commencez par earthUI** pour découvrir les variables importantes, les relations non linéaires et les interactions. La construction automatique de modèle d'earth offre le meilleur point de départ.
- **Affinez avec mgcvUI** si vous souhaitez des effets partiels plus lisses. Importez les nœuds d'earthUI pour une transition sans heurt.
- **Utilisez glmnetUI** lorsque vous avez besoin d'un modèle purement linéaire avec une sélection de variables agressive ou des contraintes de signe sur les coefficients.
- **Comparez les trois** sur les mêmes données afin d'évaluer si la complexité supplémentaire des modèles non linéaires est justifiée.

16 Jeu de données de démonstration : Appraisal_1.csv

16.1 Description

Le jeu de données de démonstration est partagé avec earthUI et glmnetUI. Il se trouve dans le dossier `demo_mls/` des sources de mgcvUI ou, si earthUI est installé :

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
```

Le fichier contient environ 1 500 enregistrements de ventes résidentielles issus d'un export MLS simulé, avec le bien objet en ligne 1 et les ventes comparables sur les lignes suivantes. Les données représentent des ventes de maisons individuelles dans un marché à plusieurs zones, avec une gamme variée de tailles, d'âges et de localisations.

Il ne s'agit pas de données réelles, mais elles s'inspirent d'un quartier réaliste du nord de la Californie. Toute information d'identification a été modifiée ou supprimée.

16.2 Colonnes

Colonne	Type	Type spécial	Description
sale_type	character	sale_type	Code de type d'enregistrement distinguant l'objet des ventes comparables
weight	numeric	weight	Poids de l'observation (0 = exclue de l'ajustement)
id	numeric	display_only	Identifiant interne de l'enregistrement
property_id	numeric	display_only	Identifiant MLS du bien
listing_id	character	display_only	Numéro d'annonce MLS
parcel_number	character	display_only	Numéro de parcelle cadastrale (APN)
street_address	character	display_only	Adresse du bien
city_name	character	display_only	Ville
postal_code	character	display_only	Code postal
county_name	character	display_only	Comté
contract_date	Date	contract_date	Date du contrat de vente (permet de calculer sale_age)
sale_age	numeric	—	Jours entre la date de contrat et la date d'effet (précalculé)
coe_date	Date	display_only	Date de clôture de l'entiercement
listing_status	character	display_only	Statut de l'annonce (par exemple « Sold »)
sale_price	numeric	(cible)	Prix de vente — variable réponse
rent	numeric	—	Loyer mensuel (pour les modèles à cibles multiples)
list_price	numeric	display_only	Prix affiché
original_list_price	numeric	display_only	Prix affiché d'origine
living_sqft	numeric	living_area	Surface habitable brute en pieds carrés
ppsf	numeric	display_only	Prix au pied carré
beds_total	integer	—	Nombre de chambres
baths_total	numeric	—	Nombre total de salles de bains (par exemple 2,5 = 2 complètes + 1 demi)
lot_size	numeric	lot_size	Superficie du terrain en pieds carrés
area_id	integer	area	Identifiant de zone MLS
area_text	character	display_only	Nom de la zone
age	numeric	actual_age	Âge du bien en années
year_built	integer	display_only	Année de construction
latitude	numeric	latitude	Latitude (arrondie à 3 décimales pour le modèle)
longitude	numeric	longitude	Longitude (arrondie à 3 décimales pour le modèle)
latitude6	numeric	display_only	Latitude en précision complète
longitude6	numeric	display_only	Longitude en précision complète
garage_spaces	integer	—	Nombre de places de garage
fp_count	integer	—	Nombre de cheminées

Colonne	Type	Type spécial	Description
no_of_stories	numeric	—	Nombre d'étages
style	character	—	Style architectural
view	character	—	Type de vue (par exemple « Neighborhood », « Hills »)
days_on_market	integer	dom	Jours sur le marché
listing_date	Date	listing_date	Date de mise en vente
sale_concessions	numeric	concessions	Concessions du vendeur

16.3 Prise en main rapide suggérée

1. Lancez la routine mgcv : `earthUI::launch_mgcv()`
2. Créez un nouveau projet (section 1) ayant pour objet **For Appraisal**
3. Placez `Appraisal_1.csv` dans le dossier d'entrée du projet et sélectionnez-le à la section 2 (cliquez sur **Refresh** si nécessaire)
4. Sélectionnez `sale_price` comme cible
5. Éventuellement, réglez Response Transform sur **Log (natural)** pour obtenir des ajustements proportionnels
6. Attribuez les types spéciaux comme indiqué dans le tableau ci-dessus
7. Incluez les prédicteurs : `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (comme facteur), `age`, `latitude`, `longitude`, `garage_spaces`
8. Conservez les valeurs par défaut : REML, gamma = 1,2, base tp, k = auto, select = TRUE
9. Cliquez sur **Fit Mgcvm GAM Model**
10. Examinez les onglets Summary, Contribution (graphiques de termes lisses) et Diagnostics
11. Téléchargez les sorties (étape 7) et examinez le classement CQA
12. Calculez les ajustements RCA (étape 8) avec un score CQA d'environ 5,00
13. Générez une grille de comparaison des ventes (étape 9) avec les comparables recommandés

17 Configuration requise et dépannage

17.1 Plateformes prises en charge

mgcvUI fonctionne sous **macOS**, **Windows** et **Linux**. L'application est développée et principalement testée sous macOS avec RStudio. Les remarques propres à chaque plateforme figurent ci-dessous.

17.2 Version de R

$R \geq 4.1.0$ est requis. RStudio Desktop (2023.06 ou ultérieur) est vivement recommandé — il embarque **Quarto** et **pandoc** (nécessaires au rendu des rapports) et fournit un environnement pratique pour lancer l'application.

17.3 Packages R requis

Les packages suivants sont installés automatiquement en tant que dépendances :

Package	Rôle
shiny	Cadre d'application web
mgcv	Ajustement des modèles GAM
valengrCore	Noyau d'évaluation partagé (rôles spéciaux, report des réglages earthUI)
ggplot2, plotly	Graphiques de contribution statiques et interactifs
gratia	Extraction des termes lisses et diagnostics
DT	Tableaux de données interactifs
quarto, rmarkdown	Génération et rendu des rapports Quarto
readxl	Importation de fichiers Excel
shinyFiles	Explorateurs de fichiers et de dossiers
RSQLite, DBI	Bases de données des réglages et des projets (SQLite)
jsonlite	Sérialisation de l'état des variables

17.4 Dépendances facultatives

Composant	Quand il est nécessaire
LaTeX (TinyTeX, MiKTeX ou MacTeX)	Génération de rapports PDF uniquement. Si aucune installation LaTeX n'est détectée, l'option PDF est automatiquement masquée dans le menu des formats. Installation : <code>tinytex::install_tinytex()</code>
sysfonts + showtext	Police Roboto Condensed pour les graphiques ggplot2. En cas d'indisponibilité ou de fonctionnement hors ligne, l'application bascule automatiquement sur la police sans empattement du système.
earth ($\geq 5.3.0$)	Nécessaire uniquement à la chaîne d'importation earthUI. Inutile pour un usage autonome.
callr	Ajustement du modèle en arrière-plan (l'application reste réactive). Sans lui, l'ajustement s'exécute de manière synchrone.
writexl, openxlsx	Téléchargements Excel (sorties du modèle, RCA, grille de ventes). L'application invite à les installer s'ils manquent.
bslib	Thématisation Bootstrap 5 (Nord Light/Dark).
officer	Export Word programmatique historique (<code>export_gam_docx()</code>).

17.5 Remarques propres à chaque plateforme

17.5.1 macOS

C'est la plateforme qui pose le moins de problèmes. Homebrew est recommandé pour les bibliothèques système. Si vous avez besoin de rapports PDF :

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

Fonctionne bien avec RStudio Desktop. Points essentiels :

- **Machines d'entreprise verrouillées** : si le répertoire temporaire ou le répertoire de données utilisateur est restreint, la persistance des réglages et la génération des rapports peuvent échouer. L'application affiche un avertissement dans la console R mais continue de fonctionner sans persistance.
- **LaTeX** : installez TinyTeX depuis R (même commande que sous macOS ci-dessus) ou installez MiKTeX depuis miktex.org.
- **Chemins de fichiers longs** : Windows impose une limite de 260 caractères pour les chemins. Gardez un chemin de répertoire de travail court.

17.5.3 Linux

C'est la plateforme la plus variable d'une distribution à l'autre. Les utilisateurs d'Ubuntu et de Debian peuvent avoir besoin de bibliothèques système avant que les packages R ne compilent :

```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

Pour les rapports PDF : `tinytex::install_tinytex()` ou installation de `texlive-full` depuis le gestionnaire de packages du système.

Sur les serveurs sans affichage, les graphiques interactifs plotly sont rendus sous forme d'images statiques de repli dans les rapports. L'application Shiny elle-même nécessite une connexion depuis un navigateur web.

17.6 Dégradation progressive

mgcvUI est conçu pour fonctionner même lorsque des composants facultatifs sont absents :

Composant absent	Comportement
Pas de LaTeX	L'option PDF est masquée dans le menu des formats de rapport. HTML et Word restent disponibles.
Pas d'accès Internet / échec des polices	La police Roboto Condensed est remplacée par la police sans empattement du système. Un message est consigné dans la console.
Pas de RSQLite / système de fichiers en lecture seule	Les réglages ne persistent pas d'une session à l'autre. Un avertissement s'affiche dans la console. L'application fonctionne normalement par ailleurs.
Échec de <code>gratia : :appraise</code>	L'onglet Diagnostics et les rapports affichent un message de remplacement. Utilisez <code>mgcv::gam.check()</code> dans la console comme solution de rechange.
Répertoire temporaire non accessible en écriture	La génération de rapports échoue avec un message d'erreur explicite. Corrigez en pointant la variable d'environnement <code>TMPDIR</code> vers un emplacement accessible en écriture.

17.7 Dépannage

17.7.1 « PDF option is not available »

Aucune installation LaTeX n'a été détectée. Exécutez dans R :

```
install.packages("tinytex")
tinytex::install_tinytex()
```

Redémarrez ensuite l'application. L'option PDF apparaîtra.

17.7.2 « Settings will not persist »

Le répertoire de la base de données SQLite ne peut être ni créé ni écrit. Vérifiez les permissions sur :

- macOS/Linux : `~/Library/Application Support/R/mgcvUI/` ou `~/.local/share/R/mgcvUI/`
- Windows : `%APPDATA%/R/data/mgcvUI/`

17.7.3 « Diagnostics plot unavailable »

La fonction `gratia::appraise()` échoue occasionnellement sur des modèles complexes (nombreuses interactions tensorielles, familles non standard). Il s'agit d'un problème connu en amont. Contournement : exécutez `mgcv::gam.check(model)` dans la console R pour afficher directement les graphiques de diagnostic.

17.7.4 « Report error : number of items to replace... »

Il s'agissait d'un problème connu des versions antérieures, lié au package `gratia`. Assurez-vous d'avoir installé la dernière version de `mgcvUI`. La version actuelle intercepte cette erreur et produit le rapport avec un contenu de remplacement pour la section des diagnostics.

17.7.5 Le port 7880 est déjà utilisé

Si l'application ne démarre pas parce que le port 7880 est occupé (par une session précédente), exécutez :

```
# macOS/Linux:
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 L'installation des packages échoue sous Linux

Bibliothèques système manquantes. Installez les en-têtes de développement listés dans la section Linux ci-dessus, puis relancez `install.packages()`.

Références

- WOOD, S. N. (2003). « Thin-plate regression splines ». In : *Journal of the Royal Statistical Society (B)* 65.1, p. 95-114. DOI : [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). « Stable and efficient multiple smoothing parameter estimation for generalized additive models ». In : *Journal of the American Statistical Association* 99.467, p. 673-686. DOI : [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). « Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models ». In : *Journal of the Royal Statistical Society (B)* 73.1, p. 3-36. DOI : [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- (2017). *Generalized Additive Models: An Introduction with R*. 2^e éd. Chapman et Hall/CRC.
- WOOD, S. N., N. PYA et B. SÄFKEN (2016). « Smoothing parameter and model selection for general smooth models (with discussion) ». In : *Journal of the American Statistical Association* 111, p. 1548-1575. DOI : [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).