

earthUI : une interface interactive pour le package earth (MARS)

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: [10.67330/7fj41b65](https://doi.org/10.67330/7fj41b65)

© 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. earthUI est une interface graphique pour le package R `earth`, qui implémente les Multivariate Adaptive Regression Splines (MARS). 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 configuration du modèle, l'ajustement, ainsi que l'interprétation des diagnostics et des sorties graphiques d'`earth`. Cet article documente les exigences de format de données d'`earthUI`, le flux de travail de modélisation, les affichages de sortie et la référence complète des fonctionnalités.

Mots-clés : MARS, multivariate adaptive regression splines, `earth`, R, interface graphique, évaluation immobilière

Note de l'éditeur (juillet 2026) : Cet article décrit `earthUI` 0.10.0. Peu avant la mise sous presse, `earthUI` 0.11.0 a absorbé les applications compagnes `glmnetUI` et `mgcvUI` (décrites dans les deux articles suivants), regroupant les flux de travail MARS, `elastic-net` et `GAM` dans une seule application. L'interface et le flux de travail décrits ici demeurent inchangés par ce regroupement.

Table des matières

1	Introduction	4
1.1	Qu'est-ce qu' <code>earthUI</code> ?	4
1.2	Un package, trois routines	4
1.3	Trois modes Purpose	4
1.4	Fonctionnalités spécifiques à l'immobilier	5
1.5	Installation d' <code>earthUI</code> et de <code>vProlog</code>	5
1.6	Premiers pas	6
2	Exigences relatives aux données d'entrée MLS	7
2.1	Format et structure du fichier	7
2.2	Colonnes requises pour le mode Appraisal	7
2.3	Conventions de nommage des colonnes spéciales	9
2.4	Qualité et complétude des données	9
2.5	Placement de la ligne du bien à évaluer	9
3	Mode General Purpose	11
3.1	Vue d'ensemble	11
3.2	Le flux de travail de la barre latérale	11
3.3	Les onglets du panneau principal	12

3.4	Persistence des paramètres	12
3.5	Mode sombre	12
3.6	Locale et paramètres régionaux	13
3.6.1	Pays pris en charge	13
3.6.2	Remplacement du format de papier	13
3.6.3	Enregistrement des valeurs par défaut	14
4	Mode d'évaluation (Appraisal Mode)	15
4.1	Traitement de la ligne du bien à évaluer	15
4.2	Effective Date et ancienneté des ventes (Sale Age)	15
4.3	Désignations de colonnes spéciales (Special)	15
4.4	Residual Contribution (estimation des caractéristiques rares)	16
4.5	Aperçu des ajustements RCA	17
5	Mode d'analyse de zone de marché (Market Area Analysis Mode)	18
5.1	Différences par rapport au mode d'évaluation	18
5.2	Quand utiliser le mode marché	18
6	Sélection des variables	19
6.1	Variable(s) cible(s)	19
6.2	La table des prédicteurs	19
6.3	Détection des types de données et redéfinitions	20
6.4	Indicateurs Factor et Linear	20
6.5	Référence des types de colonnes spéciaux	20
6.6	Cibles multiples	21
7	Sélection des paramètres	22
7.1	Paramètres de base	22
7.2	Passe avant	22
7.3	Interactions autorisées	23
7.4	Élagage	23
7.5	Validation croisée	23
7.6	Filtrage par sous-ensemble	24
7.7	Valeurs de paramètres recommandées	24
7.7.1	Paramètres de la passe avant	24
7.7.2	Paramètres d'élagage	26
7.7.3	Paramètres de validation croisée	26
7.7.4	Modèle de variance	27
7.7.5	Fichier de sortie earth	27
7.8	Réglages par défaut	27
8	Ajustement du modèle Earth	28
8.1	Le bouton Fit	28
8.2	Fenêtre modale d'ajustement et sortie de trace	28
8.3	Journal d'ajustement	28
9	Téléchargement des données	29
9.1	Colonnes de sortie	29
9.2	Ordre des colonnes et mise en forme Excel	29
9.3	Scores CQA	30
9.4	Tri et ligne du bien à évaluer	30
9.5	Export automatique vers mgcvUI	30

9.6	Import dans glmnetUI	31
9.6.1	Principaux avantages	31
9.6.2	Important : la colonne de pondération doit correspondre	31
9.6.3	Flux de travail	31
10	Calculs RCA et téléchargement	32
10.1	Ouverture de la boîte de dialogue RCA	32
10.2	Interpolation du score CQA	32
10.3	Colonnes de sortie	32
10.4	L'onglet RCA Adjustments	33
11	Sales Comparison Grid	34
11.1	Vue d'ensemble et objet	34
11.2	Fenêtre modale de sélection des comparables	34
11.3	Disposition de la grille	34
11.4	Prix de vente, concessions et Net SP	35
11.5	Lignes groupées (localisation, site, âge)	35
11.6	Lignes de variables du modèle	35
11.7	Ligne CQA Residual et lignes de caractéristiques résiduelles	36
11.8	Formule du prix de vente ajusté (Adjusted Sale Price)	36
11.9	Protection des feuilles et cellules modifiables	36
11.10	Utilisation de la grille dans Excel	36
12	Téléchargement des rapports	38
12.1	Generate Quarto Report	38
12.2	Convert Quarto Report	38
12.3	Contenu du rapport	38
13	Les Post Processing Modules : Elastic Net et GAM	40
13.1	Pourquoi des ajustements d'appui ?	40
13.2	Lancement et état partagé	40
13.3	La routine elastic net	40
13.4	La routine GAM	40
13.5	Un flux de travail pratique	41
14	Jeu de données de démonstration : Appraisal_1.csv	42
14.1	Description	42
14.2	Colonnes	43
14.3	Démarrage rapide suggéré	43
15	Fonctionnalité expérimentale : extraction de texte en Prolog et règles de dérivation	45
15.1	Ce qu'elle fait	45
15.2	Prérequis : vProlog (Trealla), et non SWI-Prolog	46
15.3	Activation et exécution	46
15.4	Règles de dérivation	46
15.5	Fichier de données actif versionné	47
15.6	Interface programmatique et fichiers inclus	47
15.7	Expérimentation connexe : a priori par région climatique	47
15.7.1	Au-delà de la Californie : packs juridiques et modules de zone de marché	47

1 Introduction

1.1 Qu'est-ce qu'earthUI ?

earthUI est une interface graphique pour le package R **earth**, qui implémente les Multivariate Adaptive Regression Splines (MARS) (FRIEDMAN 1991). Le moteur de modélisation est le package **earth** de Stephen Milborrow (MILBORROW 2024), documenté en détail par MILBORROW (s. d.[a]) et MILBORROW (s. d.[b]). Il s'exécute sous la forme d'une application Shiny locale — sans identifiant, sans serveur et sans compte. On la lance depuis R, on importe un jeu de données (CSV ou Excel), on configure le modèle et on l'ajuste de manière interactive.

L'application offre un flux de travail complet : importation des données, configuration des variables, ajustement du modèle, graphiques de diagnostic, importance des variables, équations du modèle et rapports téléchargeables aux formats HTML, Word ou PDF.

1.2 Un package, trois routines

Depuis la version actuelle, le package earthUI *contient* également ses deux Post Processing Modules (modules de post-traitement), qui constituaient dans les versions antérieures les packages séparés glmnetUI et mgcvUI. L'installation d'earthUI installe les trois routines ; chacune s'exécute comme une application Shiny locale distincte :

- `earthUI::launch()` — l'application earth (MARS) décrite dans le présent article, sur le port 7878. `earth()` est le moteur central d'évaluation.
- `earthUI::launch_glmnet()` — le Post Processing Module elastic net (port 7879), qui importe les fonctions de base du modèle earth et les réestime sous régularisation lasso/elastic-net (voir « glmnetUI Import » ci-dessous).
- `earthUI::launch_mgcv()` — le Post Processing Module GAM (port 7880), qui utilise les emplacements de nœuds du modèle earth comme points de départ pour les termes de lissage de `mgcv` (voir « mgcvUI Auto-Export » ci-dessous).

Les trois routines partagent une même arborescence de projets, une même base de données de paramètres et une même copie de l'infrastructure de support, de sorte qu'un projet ouvert dans l'une des applications est immédiatement disponible dans les autres. Les routines glmnet et mgcv constituent des éléments de corroboration de la conclusion de valeur du modèle earth, et non des flux de travail concurrents : le schéma prévu consiste à ajuster et à affiner d'abord le modèle earth, puis à exécuter les modules sur le même projet lorsqu'une corroboration est souhaitée.

1.3 Trois modes Purpose

earthUI organise le travail en **projets** — des entités nommées qui regroupent les fichiers d'entrée, les sorties et les paramètres en un emplacement fixe. Chaque projet possède un **Purpose** (finalité), choisi dans l'assistant New Project lors de la création du projet. Il existe trois modes :

- **General** (général) — régression Earth pour tout type de population ou de jeu de données. Il s'agit du mode par défaut. Il offre le flux de travail complet de modélisation MARS, sans aucun ajout propre à un domaine particulier.
- **For Appraisal** (pour l'évaluation immobilière) — régression Earth adaptée à l'évaluation immobilière. Ce mode ajoute des fonctionnalités propres à l'évaluation d'un bien individuel, notamment la prise en charge du bien à évaluer, les désignations spéciales de colonnes, la Residual Constraint Approach (RCA) (fr. : approche par contraintes résiduelles) et la génération de la Sales Comparison Grid (grille de comparaison).

- **Market Area Analysis** (analyse de zone de marché) — régression Earth adaptée aux études de zones de marché. Ce mode ajoute des fonctionnalités destinées à l'analyse de groupes de biens au sein d'un marché défini, plutôt qu'à l'évaluation d'un bien unique.

Dans les trois modes, le moteur de modélisation central est identique — on ajuste toujours un modèle Earth (MARS). Le réglage du Purpose détermine quels outils et éléments d'interface supplémentaires sont disponibles.

Le Purpose est fixé par projet. La finalité d'un projet — de même que son pays, son État, son comté et sa ville — est verrouillée à la création du projet. Le passage à un autre projet réinitialise earthUI à un état par défaut vierge (les données importées, les résultats du modèle, les onglets, la configuration des variables et les paramètres earth sont tous effacés), puis restaure la configuration enregistrée de ce projet pour sa finalité.

1.4 Fonctionnalités spécifiques à l'immobilier

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

- **Désignations spéciales de colonnes** — Chaque prédicteur peut se voir attribuer 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, de la production des sorties et de la génération de la Sales Grid. Voir le chapitre 6 pour la référence complète.
- **Arrondi de la latitude et de la longitude** — Les colonnes désignées comme latitude ou longitude sont automatiquement arrondies à 3 décimales afin de prévenir le surajustement.
- **Colonne Sale Age** — Lorsqu'une colonne est désignée comme `contract_date` et qu'une Effective Date (date d'effet) est fournie, earthUI 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 (évaluation uniquement)** — En mode évaluation, après l'ajustement du modèle, earthUI peut calculer les sorties de la Residual Constraint Approach (RCA), qui produisent des ajustements par bien comparable, des synthèses d'ajustements nets/bruts et un prix de vente ajusté pour le bien à évaluer.
- **Sales Comparison Grid (évaluation uniquement)** — Génère un classeur Excel à plusieurs feuilles comportant des formules pour les ajustements des biens comparables, les décompositions du résidu par caractéristique et les prix de vente ajustés. Voir le chapitre 11.

***Conseil :** le mode General convient à tout jeu de données — financier, scientifique, technique ou immobilier. Les modes évaluation et zone de marché ne font qu'ajouter des fonctionnalités de commodité destinées aux professionnels de l'immobilier.*

1.5 Installation d'earthUI et de vProlog

earthUI est publié sur le CRAN ; l'installation standard tient donc en une seule ligne de R :

```
install.packages("earthUI")
```

Le présent article décrit la version actuelle. Si le CRAN propose encore une version plus ancienne au moment de la lecture, la version de développement s'installe depuis GitHub (noter le `subdir` — le package réside dans le dossier `pkg/` du dépôt) :

```
remotes::install_github("wcraytor/earthUI", subdir = "pkg")
```

vProlog (facultatif). Le dispositif expérimental de traitement des remarques (voir « Prolog Text Extraction » vers la fin du présent article) utilise le package vProlog — un moteur Prolog embarqué autonome. Il ne s'agit délibérément *pas* d'une dépendance obligatoire : earthUI s'installe et s'exécute sans lui, et les fonctionnalités Prolog demeurent simplement désactivées tant qu'il n'est pas présent. vProlog est distribué via le dépôt r-universe de l'auteur, qui fournit des binaires précompilés pour Windows et macOS (aucune chaîne de compilation n'est nécessaire) :

```
install.packages("vProlog",
  repos = c("https://wcraytor.r-universe.dev",
    getOption("repos")))
```

Aucune installation système de Prolog (SWI-Prolog ou autre) n'est requise — le moteur Trealla est compilé dans le package lui-même. Tout le reste du présent article fonctionne sans vProlog.

1.6 Premiers pas

Pour utiliser earthUI :

1. **Installer le package** depuis le CRAN comme décrit ci-dessus.
2. **Lancer l'application** : exécuter `earthUI::launch()` dans R, ou, depuis la ligne de commande, exécuter `Rscript -e 'earthUI::launch()'`. L'application s'ouvre dans le navigateur web sur le port 7878. On peut également y accéder directement en naviguant vers `http://localhost:7878` dans le navigateur. (Les Post Processing Modules se lancent de la même manière : `launch_glmnet()` sur le port 7879, `launch_mgcv()` sur le port 7880.)
3. **Créer ou ouvrir un projet** dans la Section 1 de la barre latérale. L'assistant New Project recueille le nom du projet, le Purpose (General, For Appraisal ou Market Area Analysis), la localisation (pays, État, comté, ville) et, facultativement, un fichier de données initial, qui est copié dans le dossier `in/` du projet.
4. **Importer les données** dans la Section 2, qui répertorie les fichiers CSV et Excel présents dans le dossier `in/` du projet actif. Les fichiers ajoutés manuellement via le Finder ou l'Explorer apparaissent après un clic sur Refresh.
5. **Configurer les variables** — choisir la ou les variables cibles ainsi que les prédicteurs, définir les types de données et attribuer les éventuels rôles spéciaux de colonnes.
6. **Définir les paramètres Earth** — `degree`, `nprune`, filtres de sous-ensembles et autres arguments de `earth()`.
7. **Ajuster le modèle** — cliquer sur « Fit Earth Model » et examiner les résultats dans le panneau principal.
8. **Exporter** — télécharger les prédictions au format Excel, générer des lots de rapports Quarto ou (en mode évaluation) calculer les ajustements RCA et les Sales Comparison Grids.

La configuration est enregistrée côté serveur dans une base de données de projet, indexée par projet et par Purpose, via les boutons « Save current as default » des Sections 3 et 4 (voir le chapitre 3, « Settings Persistence »).

2 Exigences relatives aux données d'entrée MLS

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

2.1 Format et structure du fichier

earthUI accepte les fichiers **CSV** et **Excel** (.xlsx, .xls). Lors de l'import, 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 tout au long du flux de travail. Le séparateur CSV et le signe décimal utilisés lors de l'import sont déterminés par les paramètres de locale (voir le chapitre 3, « Locale & Regional Settings » [paramètres régionaux]).

Le fichier de données doit se présenter sous la forme d'une table plate 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 pour le mode Appraisal

Bien que earthUI fonctionne avec n'importe quel ensemble de colonnes, le flux de travail complet d'évaluation (ajustements RCA + Sales Comparison Grid, la grille de comparaison) tire parti de la présence des colonnes suivantes dans l'export MLS :

Colonne	Type spécial	Rôle
Sale Price	(target)	Variable réponse du modèle
Contract Date	<code>contract_date</code>	Sert au calcul de <code>sale_age</code> (nombre de jours par rapport à la date d'effet)
Listing Date	<code>listing_date</code>	Utilisée avec la date de contrat pour calculer le DOM en l'absence de colonne DOM
Days on Market	<code>dom</code>	Nombre de jours sur le marché ; affiché dans la Sales Grid
Concessions	<code>concessions</code>	Concessions de vente ; $\text{Net SP} = \text{Sale Price} - \text{Concessions}$ dans la Sales Grid
Living Area (SF)	<code>living_area</code>	Permet le calcul des résidus par pied carré (<code>residual_sf</code> , <code>cqa_sf</code>)
Lot Size	<code>lot_size</code>	Regroupée dans la ligne « Site Size Dimensions » de la Sales Grid
Site Dimensions	<code>site_dimensions</code>	Regroupée avec la taille du terrain dans la Sales Grid
Latitude	<code>latitude</code>	Arrondie à 3 décimales ; sert au calcul de proximité et au groupe Location
Longitude	<code>longitude</code>	Arrondie à 3 décimales ; sert au calcul de proximité et au groupe Location
Area ID	<code>area</code>	Regroupée dans la ligne « Loc : Long Lat Area » de la Sales Grid
Actual Age	<code>actual_age</code>	Regroupée dans la ligne « Actual Age Effective Age » de la Sales Grid
Effective Age	<code>effective_age</code>	Regroupée avec l'âge réel dans la Sales Grid
Parcel Number	(aucun)	Affiché dans la ligne APN de la Sales Grid
Listing ID	(aucun)	Affiché dans la ligne APN de la Sales Grid
Address	<code>display_only</code>	Affichée dans la ligne Address de la Sales Grid ; exclue du modèle

Les noms de colonnes du tableur peuvent être rédigés dans une langue étrangère — les noms « spéciaux » sont en anglais afin que le programme R puisse leur réserver un traitement particulier. Pour le reste, les noms de colonnes fournis apparaissent tels quels dans les modèles de régression, les graphiques et (dans le cas d'un rapport d'évaluation) l'Intermediate Sales Grid (grille de comparaison intermédiaire).

Toutes les colonnes ne sont pas requises. earthUI s'adapte — si une colonne est absente, la caractéristique correspondante est simplement omise. Par exemple, si aucune colonne **concessions** n'est désignée, la ligne Net SP de la Sales Grid affiche le prix de vente sans déduction des concessions. Toutefois, pour les modèles de prix immobiliers, certaines colonnes sont fortement recommandées afin d'obtenir un ajustement acceptable :

1. **Sale Age** (ancienneté de la vente) — le nombre de jours entre la date de vente du contrat et la date d'effet de l'évaluation ou de l'analyse. Lorsque l'historique des ventes porte sur plusieurs années, en particulier sur des périodes supérieures à 5 ans, `sale_age` joue souvent un rôle central dans l'estimation du prix de vente. De fait, cette variable est souvent si importante que, sans elle, earthUI ne parvient à produire aucun modèle.
2. **Living Area** (surface habitable) — désignée également par des noms tels que « Living Sqft », « GLA » (*gross living area*, surface habitable brute), etc. Il s'agit d'un autre déterminant majeur du prix de vente.
3. **Total Bath Count** (nombre total de salles de bain) — le nombre total de salles de bain complètes, quart, demi et trois-quarts. Par exemple, deux salles de bain complètes et une demi-salle de bain correspondent à une valeur de 2.5.

4. **Garage Bays** ou **Garage Area** (places ou surface de garage) — le nombre de places de garage ou la surface du garage en pieds carrés.
5. **Lot Size** (taille du terrain) — la superficie du terrain, généralement exprimée en pieds carrés ou en acres.
6. **Longitude**, **Latitude** et, si disponible, **Area ID** (identifiant de zone). Les ajustements correspondants seront combinés en un ajustement unique Location dans la Sales Grid.

2.3 Conventions de nommage des colonnes spéciales

earthUI identifie les colonnes par leur **désignation de type spécial**, et non par leur nom de colonne. Les colonnes peuvent porter n'importe quel nom dans l'export MLS — ce qui importe est de leur attribuer le type spécial correct dans la table Variable Configuration (configuration des variables) présentée au chapitre 6.

Par exemple, le MLS peut exporter la surface habitable sous le nom « GLA », « Living SqFt », « liv_area » ou « gross_living_area ». Après l'import (où le nom est converti en snake_case), il suffit de la désigner comme `living_area` dans la liste déroulante Special. earthUI l'utilisera alors pour le calcul des résidus par pied carré et pour le regroupement dans la Sales Grid, quel que soit son nom d'origine.

2.4 Qualité et complétude des données

- **Valeurs manquantes (NA)** : les lignes comportant des valeurs NA dans une colonne prédicteur ou dans la colonne cible sont automatiquement supprimées avant l'ajustement. Le paramètre `na.action` est toujours fixé en interne à `na.fail`, la suppression des NA étant effectuée avant l'appel à `earth()`.
- **Colonnes de dates** : les dates doivent être dans un format que R peut analyser (p. ex. 2025-06-15, 06/15/2025, June 15, 2025). earthUI détecte automatiquement les colonnes de dates lorsque au moins 80% des valeurs sont analysées avec succès. Les années à deux chiffres (p. ex. 06/15/25) sont lues comme l'année correspondante la plus proche, de sorte que les exports MLS récents sont analysés correctement ; les années à quatre chiffres sont toujours prises littéralement.
- **Colonnes numériques** : le prix de vente, la surface habitable, la taille du terrain, les concessions et les champs similaires doivent être numériques. Si le MLS exporte les prix avec des symboles monétaires ou des séparateurs de milliers (p. ex. « \$350,000 » ou « 350.000,00 »), il peut être nécessaire de nettoyer ces valeurs avant l'import.
- **Colonnes facteurs** : les variables catégorielles telles que l'identifiant de zone, le style ou l'état doivent contenir un nombre raisonnable de valeurs uniques. earthUI détecte automatiquement les variables facteurs, mais cette détection peut être outrepassée.

***Conseil** : après l'import, examinez la colonne T/F/NA du tableau des prédicteurs — elle indique, pour chaque colonne, le nombre de valeurs true/false/NA. Les colonnes présentant plus de 30% de valeurs manquantes sont signalées en rouge. Il convient d'envisager d'exclure les colonnes à fort taux de NA ou de nettoyer les données avant l'import.*

2.5 Placement de la ligne du bien à évaluer

En mode **Appraisal** (évaluation), la **ligne 1 doit correspondre au bien à évaluer**. Toutes les lignes restantes sont des ventes comparables. Le prix de vente du bien à évaluer peut être laissé vide (NA) ou fixé à une valeur quelconque — earthUI le traite dans tous les cas comme NA lors de l'ajustement.

En mode **Market Area Analysis** (analyse de zone de marché), placer un bien à évaluer en ligne 1 est facultatif, et aucune ligne n'est exclue automatiquement. Pour tenir la ligne du bien

à évaluer (ou toute autre ligne) hors de l'ajustement, il faut lui attribuer une valeur nulle dans une colonne de pondérations désignée ou utiliser un filtre de sous-ensemble.

En mode **General** (général), aucune ligne ne fait l'objet d'un traitement particulier — toutes les lignes sont traitées de manière identique.

3 Mode General Purpose

3.1 Vue d'ensemble

Le mode General Purpose (mode à usage général) est le mode actif par défaut au lancement d'earthUI. Il offre le flux de travail complet de modélisation MARS pour tout jeu de données — et pas seulement pour l'immobilier. Vous pouvez utiliser earthUI pour des données scientifiques, des analyses financières, des études d'ingénierie ou tout problème de régression dans lequel vous souhaitez explorer des relations non linéaires et des interactions entre variables.

En mode General, l'interface omet les fonctionnalités propres à l'immobilier (colonnes spéciales, ancienneté de la vente, arrondi des coordonnées, RCA). La barre latérale est épurée afin de se concentrer 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 de la sélection du projet jusqu'à l'export :

1. Project (projet) — Sélectionnez un projet existant ou créez-en un nouveau via l'assistant New Project, qui enregistre l'objet, le pays et la localisation administrative (tous verrouillés une fois le projet créé). Le projet actif détermine l'emplacement d'où sont lus les fichiers d'entrée et celui où toutes les sorties sont écrites — il n'existe pas de champ distinct pour le dossier de sortie.

2. Import Data (importation des données) — Choisissez un fichier CSV ou Excel dans le dossier `in/` du projet actif; un lien Refresh réanalyse le dossier à la recherche de fichiers ajoutés manuellement. Pour les fichiers Excel comportant plusieurs feuilles, un sélecteur de feuille apparaît. Les noms de colonnes sont automatiquement convertis en `snake_case`.

3. Variable Configuration (configuration des variables) — Sélecteur de variable cible (plusieurs cibles possibles), Effective Date (date d'effet) et tableau des prédicteurs avec cases à cocher Include, Factor, Linear et Latent. Voir le chapitre 6 pour tous les détails.

4. Earth Call Parameters (paramètres d'appel de earth) — Tous les arguments de la fonction `earth()` : degree, penalty, nk, méthode d'élagage, validation croisée, etc. Voir le chapitre 7 pour la référence complète des paramètres.

Execute Prolog Processing — Section optionnelle et non numérotée, située entre les sections 4 et 5, désactivée tant que la fonctionnalité expérimentale Prolog n'est pas activée dans Settings. Voir le chapitre 14.

5. Fit Earth Model (ajustement du modèle earth) — Un champ de graine aléatoire et le bouton qui exécute le modèle de manière asynchrone. Voir le chapitre 8 pour les détails de l'ajustement.

6. Download Estimated Target Variable(s) & Residuals — (intitulée « Download Estimated Sale Prices & Residuals » dans les modes appraisal et market) Exporte les prédictions, les résidus, les scores CQA et les contributions par fonction g sous forme de fichier Excel. Voir le chapitre 9.

7–8. RCA Adjustments et Sales Grid — Mode appraisal uniquement. Voir les chapitres 10 et 11.

9. Generate Quarto Report (génération du rapport Quarto) — Écrit un ensemble de rapport Quarto autonome (source `.qmd` et ressources associées) dans le dossier de sortie du projet. Numérotée 7 hors mode appraisal. Voir le chapitre 12.

10. Convert Quarto Report (conversion du rapport Quarto) — Convertit tout fichier `.qmd` en HTML, Word ou PDF. Voir le chapitre 12.

3.3 Les onglets du panneau principal

Après l'ajustement, le panneau principal propose dix onglets :

Onglet	Contenu
Data	Aperçu des données importées. En mode appraisal, scindé en deux tables : Subject Property et Comparable Sales.
Equation	L'équation du modèle ajusté, affichée en LaTeX/MathJax, montrant chaque fonction de base et son coefficient.
Summary	Métriques clés (R^2 , GRSq, GCV, RSS, CV R^2) et tableau des coefficients.
Variable Importance	Diagramme en barres et tableau classé des scores d'importance des prédicteurs.
Contribution	Graphiques de contribution des fonctions g. Sélectionnez une fonction g dans le menu déroulant ; les prédicteurs univariés sont représentés par des courbes, les interactions bivariées ($\text{degree} \geq 2$) par des graphiques en perspective 3D et des graphiques de contours.
Correlation	Carte thermique des corrélations entre prédicteurs.
Diagnostics	Trois graphiques : Residuals vs Fitted, Normal Q-Q et Actual vs Predicted.
RCA Adjustments	(mode appraisal uniquement) Histogrammes des valeurs Residual Adj %, Net Adj % et Gross Adj % sur l'ensemble des comparables, renseignés une fois l'étape 7 (Calculate RCA Adjustments) terminée.
Residual Contribution	Estime les contributions de marché des caractéristiques rares ou exclues en régressant conjointement les résidus du modèle sur des colonnes jamais utilisées par le modèle (voir subsection 4.4).
ANOVA	Tableau de décomposition ANOVA montrant la contribution de chaque fonction de base à la RSS.
Earth Output	Sortie texte brute de 'earth : :summary()', y compris les détails de la passe d'élagage.

3.4 Persistance des paramètres

earthUI enregistre votre configuration côté serveur dans une base de données de projets (`projects.sqlite` à la racine de `regProj`), indexée par projet et par objet. L'enregistrement est **déclenché par bouton** : le bouton « Save current as default » de la section 3 enregistre la cible, la date d'effet et la configuration des prédicteurs, tandis que le bouton de la section 4 enregistre les paramètres d'appel de earth et la matrice Allowed Interactions. Chaque bouton ne met à jour que sa propre moitié, de sorte qu'ils ne s'écrasent jamais mutuellement, et **l'ajustement du modèle n'enregistre rien automatiquement** — vous pouvez expérimenter librement sans écraser vos valeurs par défaut enregistrées. Les différents fichiers d'entrée d'un projet (un petit extrait de test, le jeu de données complet) partagent une même configuration et, comme les paramètres résident dans l'arborescence du projet, ils l'accompagnent lorsque vous synchronisez, sauvegardez ou partagez le dossier `regProj`.

3.5 Mode sombre

Cliquez sur l'icône lune/soleil dans le coin supérieur droit pour basculer entre les thèmes clair et sombre. La préférence de thème est enregistrée dans le stockage local et persiste d'une session à l'autre.

3.6 Locale et paramètres régionaux

earthUI prend en charge les conventions internationales de formatage des nombres, des dates et des fichiers CSV grâce à un système de locale fondé sur le pays. Le menu déroulant **Country** (pays) du menu **Settings** (engrenage) de la barre de menus supérieure sélectionne un préréglage parmi 31 pays pris en charge. Chaque préréglage configure :

- **Séparateur CSV** — virgule (,) pour les États-Unis, le Royaume-Uni et le Japon, ou point-virgule (;) pour la majeure partie de l'Europe, où la virgule sert de marque décimale.
- **Marque décimale** — point (.) ou virgule (,).
- **Séparateur de milliers** — virgule (États-Unis/Royaume-Uni/Japon), point (Allemagne/Italie/Espagne), espace (Finlande/France/Pologne/pays baltes/Ukraine/Russie) ou apostrophe (Suisse).
- **Format de date** — MM/DD/YYYY (États-Unis), DD/MM/YYYY (majeure partie de l'Europe) ou YYYY-MM-DD (Suède/Lituanie/Japon/Canada).
- **Format de papier** — Letter (États-Unis/Canada/Mexique) ou A4 (partout ailleurs).

3.6.1 Pays pris en charge

Pays	Sép. CSV	Décimale	Milliers	Date	Papier	
États-Unis, Canada (csv)	,	.	,	MDY/YMD	Letter	
Royaume-Uni, Irlande, Australie, Nouvelle-Zélande	,	.	,	DMY	A4	
Allemagne, Autriche, Belgique, Pays-Bas, Italie, Espagne, Portugal, Danemark, Brésil	;	,	.	DMY	A4	
France, Pologne, Rép. tchèque, Norvège, Lettonie, Estonie	;	,	(espace)	DMY	A4	
Finlande, Suède	;	,	(espace)	DMY/YMD	A4	
Suisse	;	,	'	DMY	A4	
Ukraine, Russie	;	,	(espace)	DMY	A4	
Turquie	;	,	.	DMY	A4	
Japon, Corée du Sud	,	.	,	YMD	A4	
Mexique	,	.	,	DMY	Letter	
Lituanie	;	,	(espace)	YMD	A4	

3.6.2 Remplacement du format de papier

Sous le sélecteur de pays, un menu déroulant **Paper** (Letter ou A4) vous permet de remplacer le format de page du rapport sans changer de pays. Le séparateur CSV, la marque décimale et l'ordre du format de date suivent le préréglage du pays sélectionné ; lorsque vous changez de pays, le remplacement du format de papier est réinitialisé à la valeur par défaut de ce pays.

3.6.3 Enregistrement des valeurs par défaut

Cliquez sur **Save** dans le menu Settings pour enregistrer globalement vos préférences de locale. Ces valeurs par défaut s'appliquent à toutes les sessions futures, quel que soit le projet ouvert. Les paramètres propres à chaque projet (cible, prédicteurs, paramètres) sont enregistrés séparément dans la base de données du projet ; les valeurs par défaut de locale persistent pour tous les projets via une base de données SQLite au niveau utilisateur. Cette approche à deux niveaux est conçue pour les organisations, telles que les cabinets d'audit, qui travaillent avec des données provenant de plusieurs pays — définissez votre pays le plus fréquent comme valeur par défaut, puis ajustez au besoin.

***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 des points pour les milliers (200.000), la locale finlandaise des espaces (200 000), la locale suisse des apostrophes (200'000). Aucun symbole monétaire n'est affiché — earthUI est indépendant de la devise.*

4 Mode d'évaluation (Appraisal Mode)

Lorsque vous sélectionnez **For Appraisal** (pour évaluation) comme Purpose (objectif), earthUI se configure pour l'évaluation d'un bien unique. Toutes les fonctionnalités décrites au chapitre 3 restent disponibles ; le présent chapitre ne couvre que les ajouts propres à l'évaluation immobilière.

4.1 Traitement de la ligne du bien à évaluer

En mode évaluation, **la ligne 1 de votre jeu de données correspond au bien à évaluer** et toutes les lignes restantes sont des ventes comparables. Votre fichier d'entrée doit être organisé en conséquence (voir chapitre 2). Le prix de vente du bien à évaluer peut être laissé vide ou fixé à une valeur quelconque — earthUI le traite automatiquement comme NA lors de l'ajustement du modèle.

Après l'importation, l'onglet Data se scinde en deux sections : **Subject Property** (bien à évaluer, ligne 1) et **Comparable Sales** (ventes comparables, lignes 2 et suivantes). La ligne 1 est toujours exclue de l'ajustement du modèle — la notification « Skipping row 1 (subject). Fitting on N rows. » (ligne 1 ignorée ; ajustement sur N lignes) le confirme. Après l'ajustement, le modèle génère néanmoins des prédictions pour la ligne du bien à évaluer, affichées sous la forme `est_<target>` dans les résultats.

4.2 Effective Date et ancienneté des ventes (Sale Age)

Un champ **Effective Date** (date d'effet) apparaît dans la section Variable Configuration dans chaque mode d'objectif (avec la date du jour par défaut) — les analyses générales et de marché disposent elles aussi d'une date d'effet, et non les seules évaluations — mais il n'est réellement *utilisé* que dans les modes évaluation et marché. Dans ces modes, si vous désignez une colonne comme `contract_date` dans le menu déroulant Special, earthUI calcule une colonne `sale_age` — le nombre de jours entiers entre la date de contrat de chaque vente et la date d'effet. Cette colonne remplace la colonne de date d'origine comme prédicteur.

La première fois que vous cliquez sur Fit après avoir désigné une date de contrat, earthUI crée la colonne `sale_age` et vous invite à cliquer de nouveau sur Fit pour l'inclure.

4.3 Désignations de colonnes spéciales (Special)

Dans les modes évaluation et marché, un menu déroulant **Special** apparaît pour chaque prédicteur dans le tableau Variable Configuration. Voici la liste complète des types spéciaux et de leurs effets :

Type spécial	Effet
no	Valeur par défaut — aucun traitement particulier
contract_date	Déclenche le calcul automatique de <code>sale_age</code> à partir de l'Effective Date
listing_date	Utilisée avec la date de contrat pour calculer le DOM lorsqu'aucune colonne <code>dom</code> n'existe
dom	Identifie la colonne Days on Market (durée de mise en marché) ; affichée dans la Sales Grid (grille de comparaison)
concessions	Identifie les concessions de vente ; utilisée dans la formule Net SP (Sale Price – Concessions) de la Sales Grid
latitude	Valeurs arrondies à 3 décimales ; utilisées pour la proximité haversine et le groupe Location de la Sales Grid
longitude	Valeurs arrondies à 3 décimales ; utilisées pour la proximité haversine et le groupe Location de la Sales Grid
area	Regroupée dans la ligne « Loc : Long Lat Area » de la Sales Grid
living_area	Active les résidus au pied carré (<code>residual_sf</code> , <code>cqa_sf</code>) dans les fichiers téléchargés
lot_size	Regroupée dans la ligne « Site Size Dimensions » de la Sales Grid
site_dimensions	Regroupée avec la taille du terrain dans la Sales Grid
actual_age	Regroupée dans la ligne « Actual Age Effective Age » de la Sales Grid
effective_age	Regroupée avec l'âge réel dans la Sales Grid
display_only	Colonne incluse dans les exports mais exclue de l'ajustement du modèle (plusieurs colonnes autorisées)
remarks	(Expérimental) Colonne de remarques en texte libre convertie en colonnes de caractéristiques par l'étape Prolog facultative (plusieurs colonnes autorisées) ; voir chapitre 14
list	(Expérimental) Colonne à valeurs séparées par des virgules scindée en colonnes TRUE/FALSE par encodage one-hot lors de l'étape Prolog facultative (plusieurs colonnes autorisées) ; voir chapitre 14

Chaque type spécial ne peut être affecté qu'à une seule colonne (à l'exception de `display_only`, `remarks` et `list`). L'affectation d'un type spécial à une seconde colonne le retire automatiquement de la première. Un petit badge bleu apparaît à côté du nom de la variable pour indiquer le type spécial qui lui est affecté.

4.4 Residual Contribution (estimation des caractéristiques rares)

Les caractéristiques trop rares pour le canal de régression — un atelier présent dans une poignée de ventes — relèvent de la couche résiduelle de la RCA. L'onglet **Residual Contribution** (contribution résiduelle) estime leur contribution de marché en régressant les résidus du modèle ajusté sur des colonnes que le modèle n'a jamais utilisées, par la méthode des moindres carrés ordinaires. La sélection de plusieurs colonnes les estime *conjointement*, de sorte que des aménités corrélées (p. ex. `outbuilding_sqft` et `pr_has_workshop`) se partagent le crédit au lieu d'être comptées deux fois. Les indicateurs binaires produisent des estimations forfaitaires et listent, à titre de preuve, le résidu de chaque vente concernée ; les colonnes numériques produisent des contributions unitaires ; l'erreur type rapportée correspond à un écart type d'estimation ($t \geq 2$ servant de critère usuel de signification). Les termes appuyés par moins de dix observations reçoivent automatiquement la mention CAUTION dans la note de dossier de travail générée, laquelle documente en un seul paragraphe la méthode, les estimations et les réserves pour le rapport. Seuls les prédicteurs effectivement *utilisés* par le modèle sont exclus de la liste des candidats — un indicateur rare proposé à l'ajustement mais jamais retenu par la passe avant (forward pass) demeure éligible, puisque le modèle n'a absorbé aucune part de sa valeur.

4.5 Aperçu des ajustements RCA

Le bouton **Calculate RCA Adjustments & Download** (calcul et téléchargement des ajustements RCA ; section 7 de la barre latérale, visible uniquement en mode évaluation après l'ajustement du modèle) calcule des ajustements dérivés du marché pour chaque bien comparable par rapport au bien à évaluer. Le flux de travail RCA complet est décrit au chapitre 10.

Après le calcul des ajustements RCA, le bouton **Generate Sales Grid & Download** (section 8 de la barre latérale) devient disponible. Le flux de travail de la Sales Grid est décrit au chapitre 11.

***Conseil :** le score CQA (état, qualité, attrait) que vous attribuez au bien à évaluer détermine la part de la distribution des résidus attribuée à celui-ci. Un score de 5.00 place le bien à évaluer à la médiane des biens comparables.*

5 Mode d'analyse de zone de marché (Market Area Analysis Mode)

Lorsque vous sélectionnez **Market Area Analysis** (analyse de zone de marché) comme Purpose, earthUI offre les mêmes fonctionnalités propres à l'immobilier que le mode évaluation (colonnes spéciales, ancienneté des ventes, arrondi des coordonnées), mais est orienté vers l'analyse d'un groupe de biens plutôt que vers l'évaluation d'un unique bien à évaluer.

5.1 Différences par rapport au mode d'évaluation

- **Pas d'exclusion automatique du bien à évaluer** — contrairement au mode évaluation, le mode marché n'exclut pas automatiquement la ligne 1 de l'ajustement. Toutes les lignes sont incluses ; pour tenir une ligne de bien à évaluer hors de l'ajustement, attribuez-lui une valeur nulle dans une colonne de pondérations désignée (comme dans le jeu de données de démonstration) ou utilisez un filtre de sous-ensemble.
- **Pas de sections RCA ni Sales Grid** — les étapes « Calculate RCA Adjustments & Download » et « Generate Sales Grid & Download » ne sont pas disponibles. Le mode marché est centré sur l'ajustement du modèle et ses sorties, et non sur les ajustements par bien comparable.
- **Numérotation de la barre latérale** — en l'absence des étapes RCA et Sales Grid, la section Generate Quarto Report porte le numéro 7 au lieu de 9 (la section Convert Quarto Report conserve le numéro 10).

5.2 Quand utiliser le mode marché

Le mode Market Area Analysis convient lorsque vous :

- construisez un modèle de régression pour un quartier ou une zone de marché afin d'en comprendre les déterminants de la valeur ;
- analysez la manière dont des variables telles que la surface, l'âge, la taille du terrain et la localisation influent sur les prix de vente d'un groupe de biens ;
- préparez des éléments à l'appui d'une analyse des conditions de marché ou d'une délimitation de quartier ;
- travaillez avec un jeu de données dont la ligne 1 contient un bien à évaluer, tout en souhaitant pouvoir l'inclure dans l'ajustement ou l'en exclure.

***Conseil :** le mode marché est également utile pour la régression immobilière générale lorsque vous souhaitez bénéficier des fonctionnalités de colonnes spéciales (ancienneté des ventes, arrondi des coordonnées) sans avoir besoin du flux de travail d'ajustements RCA.*

6 Sélection des variables

La section 3 de la barre latérale — **Variable Configuration** (configuration des variables) — est l'endroit où l'on choisit les colonnes qui participent au modèle et la manière dont elles sont traitées.

6.1 Variable(s) cible(s)

La liste déroulante **Target (response) variable(s)** (variable(s) cible(s), ou de réponse), en haut de la section 3, énumère toutes les colonnes du jeu de données. Sélectionnez une seule colonne pour un modèle standard à réponse unique, ou plusieurs colonnes pour un modèle à réponses multiples. Lorsque plusieurs cibles sont sélectionnées, le modèle ajuste toutes les réponses simultanément au moyen de `earth(cbind(y1, y2, ...) ~ .)`.

Les colonnes sélectionnées comme cibles sont automatiquement exclues de la liste des prédicteurs.

6.2 La table des prédicteurs

Sous le sélecteur de cible, une table énumère toutes les colonnes restantes avec les champs suivants :

Colonne	Description
Variable	Nom de la colonne (le nom complet apparaît dans une infobulle s'il est tronqué), précédé d'un bouton à bascule $\times/+$ (voir ci-dessous). Dans les modes évaluation immobilière et marché, un badge bleu indique le type spécial assigné.
Type	Liste déroulante du type de données : <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>logical</code> , <code>factor</code> , <code>Date</code> , <code>POSIXct</code>
Include	Case à cocher — inclure cette colonne comme prédicteur dans le modèle
Factor	Case à cocher — traiter cette colonne comme variable catégorielle
Linear	Case à cocher — imposer une entrée linéaire uniquement (aucune fonction charnière (hinge))
Latent	Case à cocher — écarter entièrement cette colonne du modèle (exclue de <code>earth</code> ainsi que des données exportées vers <code>glmnetUI</code> / <code>mgcvUI</code>), en la réservant au travail sur les variables latentes
Special	Liste déroulante (modes évaluation/marché uniquement) — voir la référence des types de colonnes spéciaux ci-dessous
T/F/NA	Trois comptages par colonne : pour les colonnes logiques, <code>TRUE/FALSE/NA</code> ; pour les colonnes numériques, $\neq 0/=0/NA$. Affichés en rouge lorsque plus de 30% des valeurs sont manquantes

Une ligne d'aide au-dessus de la table explique les abréviations : « Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`). » (soit : Type = type de données de la colonne, Inc

= inclure comme prédicteur, Factor = traiter comme catégorielle, Linear = linéaire uniquement (sans charnières), Special = rôle de la colonne, p. ex. `contract_date`).

Désactivation douce plutôt que suppression. Le bouton à bascule $\times$ placé à côté de chaque nom de variable grise la ligne et écarte la colonne du modèle sans la retirer des données ; il se transforme en un $+$ vert qui réactive la colonne. Les colonnes ne sont jamais détruites.

6.3 Détection des types de données et redéfinitions

earthUI détecte automatiquement les types de données à l'importation. Les colonnes numériques, entières, logiques, facteurs et dates sont reconnues. Les colonnes de type caractère qui ressemblent à des dates (au moins 80% des valeurs s'analysent selon des formats de date courants) sont classées comme **Date**.

Toute détection peut être redéfinie en modifiant la liste déroulante **Type**. Lorsque l'on change une colonne en **character** ou **factor**, la case à cocher Factor est automatiquement cochée. Le changement de type influe sur la manière dont la colonne est transmise à la fonction `earth()`.

6.4 Indicateurs Factor et Linear

- **Factor** — Lorsque la case est cochée, la colonne est traitée comme une variable catégorielle. Cela convient aux colonnes telles que **style**, **area_id** ou **grade**, qui représentent des groupes discrets plutôt que des mesures continues. Les colonnes facteurs entrent dans le modèle sous forme de variables indicatrices (dummy).
- **Linear** — Lorsque la case est cochée, la colonne est contrainte d'entrer linéairement dans le modèle — aucune fonction charnière (linéaire par morceaux) n'est créée pour elle. Cela est utile lorsque l'on sait qu'une variable entretient une relation strictement linéaire avec la cible.

6.5 Référence des types de colonnes spéciaux

Dans les modes évaluation immobilière et marché, la liste déroulante **Special** propose les options suivantes. Chaque type ne peut être assigné qu'à une seule colonne au plus (à l'exception de **display_only**, **remarks** et **list**, qui en admettent plusieurs) :

Types de date et d'heure :

- **contract_date** — Déclenche le calcul automatique de **sale_age**. La colonne de date d'origine est remplacée par une colonne d'entiers mesurant le nombre de jours entre la date de vente et l'Effective Date (date d'effet de l'évaluation).
- **listing_date** — Utilisée comme solution de repli pour calculer les Days on Market (DOM = date de contrat – date de mise en vente ; nombre de jours de mise en marché) lorsqu'aucune colonne **dom** explicite n'est désignée.
- **dom** — Identifie la colonne des Days on Market. Affichée dans la ligne APN et la ligne Date of Sale du Sales Grid (grille de comparaison).

Types monétaires :

- **concessions** — Identifie les concessions de vente (crédits du vendeur, incitations à l'acheteur, etc.). Utilisée dans le Sales Grid pour calculer le prix de vente net : Net SP = Sale Price – Concessions.

Types de taille et de localisation :

- **latitude** — Les valeurs sont automatiquement arrondies à 3 décimales afin de prévenir le surajustement. Utilisée pour les calculs de proximité par la formule de Haversine (distance entre le bien à évaluer et chaque bien comparable) et regroupée dans la ligne Location du Sales Grid.

- **longitude** — Même traitement d'arrondi que la latitude. Utilisée pour la proximité et le groupe Location.
- **area** — Typiquement un identifiant de quartier ou de zone. Regroupée avec la latitude et la longitude dans la ligne « Loc : Long | Lat | Area » du Sales Grid.
- **living_area** — Active les calculs de résidus par pied carré (**residual_sf** et **cqa_sf**) dans le fichier téléchargé en sortie.
- **lot_size** — Regroupée dans la ligne « Site Size | Dimensions » du Sales Grid.
- **site_dimensions** — Regroupée avec la taille du terrain dans le Sales Grid (p. ex. « 75x120 »).

Types d'âge :

- **actual_age** — Regroupée dans la ligne « Actual Age | Effective Age » du Sales Grid.
- **effective_age** — Regroupée avec l'âge réel dans le Sales Grid.

Types d'affichage :

- **display_only** — La colonne est incluse dans les exportations 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 qui ne doivent pas servir de prédicteur. Plusieurs colonnes peuvent recevoir cette désignation.

Types d'extraction de texte (expérimental) :

- **remarks** — Une colonne de texte libre (p. ex. **public_remarks**) destinée à être analysée en colonnes de caractéristiques par l'étape Prolog optionnelle. Sans effet tant que cette fonctionnalité n'est pas activée dans Settings. Plusieurs colonnes admises. Voir le chapitre 14.
- **list** — Une colonne délimitée par des virgules (p. ex. **existing_structures**) destinée à être scindée en colonnes TRUE/FALSE **<col>_<value>** en encodage one-hot (encodage disjonctif complet) par la même étape. Plusieurs colonnes admises. Voir le chapitre 14.

6.6 Cibles multiples

La sélection de plus d'une variable cible ajuste un modèle Earth à réponses multiples. Lorsque plusieurs cibles sont sélectionnées :

- Le modèle prédit toutes les réponses simultanément, en partageant les fonctions de base entre les cibles
- Le bouton **wp (response weights)** (pondérations des réponses) devient actif et permet d'assigner un poids numérique à chaque cible
- Les modèles de variance (**varmod.method**) sont désactivés (non pris en charge en réponses multiples)
- Les onglets de résultats affichent, pour chaque réponse, les métriques, les équations et les graphiques de diagnostic

Conseil : les colonnes désignées comme **display_only** demeurent dans le jeu de données et figurent dans l'exportation Excel, mais sont entièrement exclues de la table des prédicteurs. À utiliser pour les colonnes d'identifiants, les adresses ou d'autres champs de référence.

7 Sélection des paramètres

La section 4 de la barre latérale — **Earth Call Parameters** (paramètres d'appel de earth) — donne accès à l'ensemble des arguments acceptés par la fonction `earth()`. Chaque paramètre est accompagné d'une icône d'aide bleue (?) dotée d'une infobulle explicative. Les paramètres sont organisés en sous-sections repliables.

7.1 Paramètres de base

Paramètre	Valeur par défaut	Description
subset	(vide)	Expression de filtrage des lignes. Voir « Filtrage par sous-ensemble » ci-dessous.
weights	NULL	Sélecteur de colonne pour les poids d'observation (par ligne). Seules les colonnes numériques sont proposées.
wp	NULL	Poids des réponses pour les modèles à cibles multiples. Un bouton ouvre une boîte de dialogue comportant un champ numérique par cible (1.0 par défaut pour chacune).
keepxy	off	Conserve x, y, subset et weights dans l'objet modèle.
trace	0	Niveau de trace pour la sortie de l'ajustement (de 0 à 5).
glm	none	Famille GLM optionnelle : gaussian , binomial ou poisson .
degree	1	Ordre maximal d'interaction. Fixer $\text{degree} \geq 2$ active automatiquement la validation croisée et fait apparaître la matrice Allowed Interactions (interactions autorisées).
penalty	2	Pénalité GCV par noeud. Des valeurs plus élevées produisent des modèles plus simples.

7.2 Passe avant

Paramètre	Valeur par défaut	Description
nk	auto	Nombre maximal de termes avant élagage. Prérempli avec la valeur recommandée pour les données chargées (voir « Valeurs de paramètres recommandées » ci-dessous).

Paramètre	Valeur par défaut	Description
thresh	0.001	Seuil de la passe avant. Des valeurs plus petites autorisent davantage de termes.
minspan	auto	Espacement minimal entre les nœuds. Prérempli avec la valeur recommandée. Les valeurs négatives fixent le nombre maximal de nœuds par prédicteur.
endspan	auto	Espacement aux extrémités — distance minimale entre un nœud et le bord des données. Prérempli avec la valeur recommandée.
newvar.penalty	0	Pénalité pour l'introduction d'une nouvelle variable (favorise la réutilisation des prédicteurs déjà présents).
fast.k	20	Nombre de termes parents considérés dans l'algorithme fast MARS.
fast.beta	1	Contrôle le facteur de vieillissement de fast MARS.

7.3 Interactions autorisées

Lorsque **degree** est fixé à 2 ou plus, une matrice d'interactions apparaît sous les paramètres de base. Il s'agit d'une grille triangulaire supérieure de cases à cocher, une pour chaque paire de prédicteurs. Une case cochée signifie que les deux prédicteurs sont autorisés à interagir ; la décocher interdit cette interaction spécifique.

Un clic sur le nom d'un prédicteur (étiquette de ligne ou de colonne) bascule toutes les interactions de cette variable. Les cases **Allow All** (tout autoriser) et **Clear All** (tout effacer), en haut, permettent un contrôle global. La matrice utilise des en-têtes fixes (« sticky ») afin que les étiquettes de lignes et de colonnes restent visibles lors du défilement.

Un message d'information le rappelle : « Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold). » (les termes d'interaction accroissent le risque de surajustement ; la validation croisée a été activée, en 10 plis).

7.4 Élagage

Paramètre	Valeur par défaut	Description
pmethod	backward	Méthode d'élagage : backward , none , exhaustive , forward , seqrep ou cv .
nprune	NULL	Nombre maximal de termes après élagage. Laisser vide pour une sélection automatique.

7.5 Validation croisée

Paramètre	Valeur par défaut	Description
nfold	10	Nombre de plis de validation croisée. Fixer à 0 pour désactiver la CV.
ncross	20	Nombre de répétitions de la validation croisée.
stratify	on	Stratifie les échantillons de CV de sorte que chaque pli présente une distribution de la réponse similaire.

Lorsque $\text{degree} \geq 2$, la validation croisée est activée automatiquement (nfold fixé à 10) afin de mieux se prémunir contre le surajustement induit par les interactions.

7.6 Filtrage par sous-ensemble

Le champ de saisie **subset** accepte une expression R qui filtre les lignes utilisées pour l'ajustement du modèle. On peut saisir une expression directement (p. ex. `sale_age < 365 & area_id == 460`) ou utiliser le bouton **Build filter...** (construire un filtre) pour la composer visuellement.

Boîte de dialogue Build Filter — Cliquer sur « Build filter... » ouvre une boîte de dialogue guidée. Chaque ligne de condition comporte une liste déroulante de colonnes, un opérateur (<, >, <=, >=, ==, !=) et un champ de valeur qui s'adapte au type de la colonne : saisie numérique pour les nombres, sélecteur de date pour les dates et liste déroulante des valeurs uniques pour les colonnes de type caractère ou factor. Les conditions sont reliées par des connecteurs AND (&) ou OR (|). Un aperçu en bas de la boîte de dialogue affiche l'expression et le nombre de lignes correspondantes. Cliquer sur **Apply** (appliquer) insère l'expression dans le champ de saisie.

Les colonnes de dates doivent utiliser les enveloppes `as.Date("...")` ou `as.POSIXct("...")` dans les expressions saisies manuellement. La boîte de dialogue Build Filter s'en charge automatiquement.

Le filtrage par sous-ensemble est non destructif — les lignes exclues restent dans le jeu de données et reçoivent des prédictions dans l'export Excel.

***Conseil :** utilisez des parenthèses pour grouper les conditions OR lorsqu'elles sont combinées avec AND. Sans parenthèses, $A \ \& \ B \ | \ C$ est évalué comme $(A \ \& \ B) \ | \ C$, ce qui peut ne pas correspondre à votre intention.*

7.7 Valeurs de paramètres recommandées

earthUI affiche une valeur recommandée sous les paramètres clés de la section 4. Ces recommandations se mettent à jour de manière réactive en fonction du nombre de lignes d'ajustement (n) et des prédicteurs sélectionnés (p). Les formules sont dérivées de l'article MARS de Friedman, des algorithmes internes de earth et de tests empiriques. Les champs **nk**, **minspan** et **endspan** sont préremplis avec leurs valeurs recommandées au chargement des données ; une modification manuelle est conservée.

7.7.1 Paramètres de la passe avant

nk (nombre maximal de termes avant élagage) :

$$nk = \min(100, \max(21, 2p + 1, \lfloor n/10 \rfloor))$$

n	$p=5$	$p=10$
30	21	21
50	21	21
100	21	21
200	21	21
500	50	50
1000	100	100
1500	100	100

La valeur par défaut de `earth`, $\min(200, \max(20, 2p)) + 1$, ne tient pas compte de la taille du jeu de données et peut se révéler trop faible pour de grands jeux de données, ce qui contraint la passe avant avant qu'elle n'ait exploré adéquatement tous les prédicteurs. Le terme $\lfloor n/10 \rfloor$ autorise environ un terme pour 10 observations — une règle empirique standard pour éviter la surparamétrisation. Le plafond de 100 prévient les rendements décroissants.

minspan (nombre minimal d'observations entre les nœuds) :

$$\text{minspan} = \min(16, \lfloor 5 + n/50 \rfloor)$$

endspan (nombre minimal d'observations depuis les bords des données) :

$$\text{endspan} = \min(16, \lfloor 5 + n/28 \rfloor)$$

n	minspan	endspan
30	5	6
50	6	6
100	7	8
200	9	12
300	11	15
500	15	16
1500	16	16

Le `minspan` calculé automatiquement par `earth` utilise l'équation 43 de Friedman : $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, qui croît en $\ln(np)$ et augmente trop lentement pour les grands jeux de données. À $n=1500$, elle ne donne que ≈ 7 , soit environ 245 emplacements de nœuds candidats par prédicteur continu — un nombre trop élevé, qui permet à la passe avant d'ajuster le bruit. La formule recommandée vise environ 100 candidats par prédicteur pour les grands n , tout en s'en remettant au calcul automatique de `earth` pour les petits n (où celui-ci est bien éprouvé). `Endspan` est fixé légèrement au-dessus de `minspan` afin d'apporter une protection supplémentaire aux bords, ce qui est utile lorsque les données présentent des queues de distribution peu fournies (situation courante en immobilier).

penalty (pénalité GCV par nœud) :

$$\text{penalty} = \begin{cases} 3 & \text{if degree} > 1 \\ 2 & \text{if degree} = 1 \end{cases}$$

Il s'agit de la valeur par défaut de `earth` — 2 pour les modèles additifs, 3 pour les modèles avec interactions (la pénalité plus élevée compense l'espace de recherche plus vaste).

newvar.penalty (pénalité pour l'introduction d'un nouveau prédicteur) :

Recommandation : **0.1** lorsque les prédicteurs sont corrélés (p. ex. la surface habitable avec le nombre de chambres, le nombre de salles de bains, la superficie du terrain). Cette pénalité

oriente la passe avant vers la réutilisation des prédicteurs déjà présents dans le modèle plutôt que vers l'introduction d'alternatives corrélées. Le mécanisme : l'amélioration de la RSS apportée par chaque nouveau prédicteur est multipliée par $1/(1 + \text{penalty})$. Avec `newvar.penalty = 0.1`, une nouvelle variable doit améliorer la RSS d'au moins 10% de plus qu'un nœud supplémentaire sur une variable existante. On obtient ainsi des modèles plus simples, comportant moins de prédicteurs, sans affecter les estimations finales des coefficients (la pénalité est retirée après la sélection). Laisser à 0 si les prédicteurs ne sont pas corrélés. Noter que `earth` ne prend pas en charge un `newvar.penalty` non nul conjointement avec des poids d'observation — `earthUI` désactive le champ (en forçant 0) dès qu'une colonne de poids est désignée.

7.7.2 Paramètres d'élagage

pmethod : recommandation : **backward**. L'élagage arrière fondé sur le GCV, méthode par défaut, est déterministe — les mêmes données produisent toujours le même modèle, sans dépendance à des graines aléatoires. Ce point est important pour la reproductibilité, en particulier dans le travail d'évaluation immobilière. La méthode `cv` utilise la validation croisée pour l'élagage, ce qui introduit une dépendance à la graine aléatoire.

nprune : recommandation : **laisser vide** (NULL). Laisser le GCV sélectionner le nombre optimal de termes. Fixer `nprune` impose un plafond strict qui prime sur le jugement du GCV.

7.7.3 Paramètres de validation croisée

nfold (plis de CV) : recommandation par tranche de taille d'échantillon :

n (lignes d'ajustement)	nfold
0–299	5
300–399	6
400–499	7
500–599	8
600–699	9
700+	10

Un plus grand nombre de plis entraîne chaque pli sur davantage de données (une estimation moins biaisée) mais réduit le pli de test mis de côté. Le tableau progresse jusqu'au standard de 10 plis à mesure que l'échantillon croît, tout en maintenant chaque pli de test suffisamment grand (environ 50 lignes ou plus) pour une métrique par pli stable, et retombe à 5 plis pour les petits échantillons — où l'estimation du CVR^2 demeure bruitée ; il convient alors d'augmenter `ncross` (p. ex. 15–20) pour moyenner la variance. Le plafond de 10 traduit le fait que chaque pli supplémentaire représente un ajustement `earth` complet pour un gain diagnostique négligeable. Avec `pmethod = "backward"`, la validation croisée n'affecte pas le modèle — elle ne fait que calculer le CVR^2 diagnostique et fournir des résidus pour le modèle de variance.

ncross (répétitions de CV) : recommandation par tranche de taille d'échantillon :

n (lignes d'ajustement)	ncross
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

Les répétitions moyennent les aléas du partitionnement en plis afin de stabiliser le CVR^2 ; les grands échantillons nécessitent moins de répétitions, car une seule validation croisée y est déjà plus stable. Les résidus de la CV répétée alimentent également le modèle de variance.

7.7.4 Modèle de variance

varmod.method : recommandation : **lm** — la valeur par défaut de Milborrow : un modèle de variance linéaire robuste pour les intervalles de prédiction (la dispersion étant estimée à partir de résidus bruités, un modèle simple généralise mieux). Si **lm** ne converge pas sur vos données, essayez **rlm** ou **const** (homoscédastique, converge toujours), ou repliez-vous sur **earth** (qui gère aussi une dispersion non linéaire). Cela exige **nfold** > 1 et **ncross** > 1.

7.7.5 Fichier de sortie earth

Après chaque ajustement réussi, earthUI écrit `<filename>_earth_output_<timestamp>.txt` dans le dossier de sortie. Ce fichier contient les termes du modèle, les statistiques récapitulatives (R^2 , $GRSq$, $CVRSq$), les détails du modèle de variance et le journal de trace. Un fichier est créé par ajustement, constituant un registre cumulatif permettant de comparer les configurations de paramètres.

7.8 Réglages par défaut

Un bouton radio en haut de la section 4 détermine quelles valeurs par défaut sont chargées :

- **Use last settings for input file** (utiliser les derniers réglages pour le fichier d'entrée) — restaure les réglages utilisés la dernière fois avec ce fichier particulier
- **Use default settings** (utiliser les réglages par défaut) — applique vos valeurs par défaut personnalisées enregistrées
- **Earth defaults** (valeurs par défaut de earth) — réinitialise tous les paramètres aux valeurs d'usine

Cliquer sur **Save current as default** (enregistrer comme valeurs par défaut), en bas de la section 4, mémorise les paramètres earth courants et la matrice Allowed Interactions ; le bouton correspondant de la section 3 mémorise la configuration de la cible, de la date d'effet et des prédicteurs. Les deux sont enregistrés par projet et par mission dans la base de données du projet.

8 Ajustement du modèle Earth

8.1 Le bouton Fit

La section 5 de la barre latérale contient le bouton **Fit Earth Model** (ajuster le modèle Earth) accompagné d'un champ **Random seed** (graine aléatoire). La graine rend reproductible l'affectation des plis de validation croisée — une liste déroulante rappelle les 10 dernières graines utilisées, et la graine employée est consignée dans l'onglet Earth Output après l'ajustement. Avant l'ajustement, earthUI valide votre configuration — une variable cible doit être sélectionnée et au moins un prédicteur doit être inclus. Dans les modes évaluation immobilière et marché, les colonnes de latitude et de longitude sont arrondies, et l'ancienneté de la vente est calculée à partir de la date d'effet et de la date de contrat (si elles sont désignées).

8.2 Fenêtre modale d'ajustement et sortie de trace

Lorsque vous cliquez sur Fit, une superposition modale sombre apparaît, comportant :

- **En-tête** — « Fitting Earth Model » avec un chronomètre affichant les secondes écoulées et un bouton **Abort** (interrompre) qui annule l'ajustement en arrière-plan
- **Journal de trace** — une zone de texte défilante à chasse fixe affichant en temps réel la sortie de la fonction `earth()` : informations sur le jeu de données, progression de la passe avant, plis de validation croisée et temps d'achèvement
- **Code couleur** — lignes ordinaires en vert, lignes de validation croisée en jaune, erreurs en rouge

earthUI ajuste les modèles de manière asynchrone au moyen de `callr::r_bg()`, qui exécute le calcul dans un processus R distinct. L'application reste ainsi réactive pendant les ajustements de longue durée. Un observateur d'interrogation cadencé à 300 ms lit la sortie du processus d'arrière-plan et la transmet en continu au journal de trace.

Lorsque l'ajustement se termine, un message « Done in X.Xs » apparaît, un bouton de fermeture (X) est ajouté à la fenêtre modale, et l'arrière-plan est retiré afin que les onglets de résultats puissent s'afficher à la demande. La fenêtre modale se ferme ensuite automatiquement une fois que les sorties visibles ont fini de s'afficher — ou vous pouvez la fermer plus tôt.

En cas de succès, une coche verte est apposée au bouton Fit. Si une erreur survient, le message d'erreur est affiché dans le journal de trace et une notification apparaît.

8.3 Journal d'ajustement

Chaque ajustement (réussi ou non) écrit un fichier journal dans votre dossier de sortie :

- **Nom de fichier** : `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Contenu** : horodatage, l'ensemble de la sortie de trace du processus d'ajustement, ainsi que tout message d'erreur
- **Emplacement** : le dossier de sortie du projet actif

Astuce : si `callr` n'est pas installé, earthUI revient à un ajustement synchrone avec une simple barre de progression. Installez `callr` pour bénéficier de l'expérience complète de sortie de trace.

9 Téléchargement des données

Après l’ajustement, la section 6 de la barre latérale — intitulée « Download Estimated Target Variable(s) & Residuals » en mode général et « Download Estimated Sale Prices & Residuals » dans les modes évaluation immobilière et marché — télécharge un fichier Excel contenant les prédictions et les diagnostics. Cette sortie est utilisée à l’étape 7 (RCA) pour attribuer une note CQA (Condition-Quality-Appeal, soit état, qualité, attrait) au bien à évaluer. La sortie est triée selon `residual_sf` et `cqa_sf` afin de vous aider à apprécier la position du bien à évaluer dans le classement. Si le modèle est de bonne qualité, les biens devraient être classés du moins attrayant au plus attrayant sur la base des caractéristiques résiduelles qui ne sont pas entrées dans la régression. La valeur médiane devrait être approximativement 0, la moitié inférieure présentant des valeurs négatives et la moitié supérieure des valeurs positives. Vous devriez trouver les logements de la plus mauvaise qualité, ou « fixers » (biens à rénover), vers le bas du classement et les plus beaux logements en haut. Il existe généralement des exceptions correspondant à des anomalies telles que les saisies immobilières, les ventes à découvert (short sales), les ventes successorales (liées à un héritage) et les ventes rapides motivées par un changement d’emploi ou d’autres raisons. L’examen des anomalies révèle habituellement une raison pertinente expliquant l’anomalie de prix.

Le libellé du bouton s’adapte au mode d’utilisation :

- Général/Marché : **Download Output (Excel)**
- Évaluation immobilière : **Download Intermediate Output (Excel)**

Le format du nom de fichier est `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`.

9.1 Colonnes de sortie

Pour chaque variable cible, les colonnes suivantes sont ajoutées :

Colonne	Description
<code>est_<target></code>	Prédiction du modèle, p. ex. <code>est_sale_price</code> (1 décimale)
<code>residual</code>	Valeur observée moins valeur prédite (1 décimale)
<code>cqa</code>	Score Condition-Quality-Appeal (2 décimales, plage 0–10)
<code>residual_sf</code>	Résidu divisé par la surface habitable (si désignée, 1 décimale)
<code>cqa_sf</code>	CQA calculé à partir du classement via <code>residual_sf</code> (2 décimales)
<code><variable>_contribution</code>	Valeur de contribution par fonction g (1 décimale, une colonne par fonction g)
<code>basis</code>	Contribution à la valeur de l’ordonnée à l’origine, identique pour tous les biens (1 décimale)
<code>calc_residual</code>	Valeur observée moins (basis + toutes les contributions) — colonne de vérification (1 décimale)

Pour les modèles à cibles multiples, un suffixe `<i>` est ajouté afin de distinguer les colonnes de chaque cible.

9.2 Ordre des colonnes et mise en forme Excel

Les colonnes de classement sont placées **le plus à gauche** dans le fichier de sortie, dans cet ordre : `residual_sf`, `cqa_sf`, `residual`, `cqa`. Cela facilite le parcours de la liste triée et l’appréciation de la position du bien à évaluer dans la distribution des CQA.

Ces colonnes reçoivent une mise en forme numérique Excel :

Colonne	Format	Exemple
<code>residual_sf</code>	Numérique, 2 décimales	12.34
<code>cqa_sf</code>	Numérique, 2 décimales	7.25
<code>residual</code>	Numérique, 0 décimale	15,200
<code>cqa</code>	Numérique, 2 décimales	6.80

9.3 Scores CQA

Le score CQA (Condition-Quality-Appeal) classe le résidu de chaque ligne par rapport à l'ensemble des autres résidus. Pour une ligne donnée, le CQA est le pourcentage de lignes présentant un résidu signé plus petit, multiplié par 10. On obtient ainsi une échelle de 0 à 10 où :

- **CQA élevé** ($\approx 9-10$) — le bien s'est vendu beaucoup plus cher que la prédiction du modèle (résidu positif important)
- **CQA faible** ($\approx 0-1$) — le bien s'est vendu beaucoup moins cher que la prédiction (résidu négatif important)
- **CQA ≈ 5** — le résidu est proche de la médiane de l'ensemble des résidus

Lorsqu'une colonne `living_area` est désignée, `cqa_sf` fournit le même classement sur la base des résidus par pied carré.

9.4 Tri et ligne du bien à évaluer

Dans les modes évaluation immobilière et marché, les lignes des biens comparables sont triées par `residual_sf` décroissant (ou `residual` si aucune surface habitable n'est désignée). La ligne du bien à évaluer (ligne 1) demeure en position 1 et n'est pas triée. En mode général, les lignes sont exportées dans leur ordre d'origine.

Ce tri, combiné aux colonnes de classement placées le plus à gauche, permet à l'évaluateur de parcourir rapidement les comparables, du plus surestimé au plus sous-estimé par le modèle, et d'apprécier la position du score CQA attribué au bien à évaluer dans cette distribution.

Les niveaux des variables facteurs dans les données de prédiction sont alignés sur les données d'entraînement avant l'appel à `predict()`. Les lignes comportant des niveaux de facteur non rencontrés à l'entraînement produiront des prédictions NA.

***Astuce :** la colonne `calc_residual` doit toujours être égale à `residual` (aux arrondis près). Un écart significatif entre les deux signale un problème de calcul méritant investigation.*

9.5 Export automatique vers mgcvUI

À chaque ajustement réussi avec `degree` ≤ 2 , `earthUI` enregistre automatiquement l'objet résultat complet sous forme de fichier `.rds` dans le dossier de sortie du projet. Le nom de fichier suit le modèle `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

Ce fichier peut être chargé par le module de post-traitement `mgcv` — inclus dans le package `earthUI` et lancé au moyen de `earthUI::launch_mgcv()` — via `readRDS()`. La routine GAM utilise les emplacements de nœuds et les fonctions de base du modèle `earth` comme points de départ pour les termes de lissage GAM, permettant une transition sans rupture de la modélisation MARS vers la modélisation GAM.

Les modèles avec `degree` > 2 sont ignorés, car `mgcvUI` ne prend en charge que les interactions par paires. Lorsque `earthUI` est lancé en mode Trilogy (depuis l'interface Trilogy UI), une section supplémentaire **Lock Model Output** (verrouillage de la sortie du modèle) permet de verrouiller

le fichier `.rds` d'un ajustement donné — avec sa base de prédicteurs et ses paramètres CQA partagés — comme modèle earth canonique importé par `glmnetUI` et `mgcvUI`.

9.6 Import dans `glmnetUI`

Le même fichier `.rds` enregistré pour la routine GAM peut également être importé dans le **module de post-traitement elastic net** — lui aussi inclus dans le package `earthUI` et lancé au moyen de `earthUI::launch_glmnet()`. Dans l'application `glmnet`, rendez-vous dans **Section 2 : Import from earthUI** et utilisez le bouton Browse (parcourir) pour sélectionner le fichier `.rds`.

`glmnetUI` recourt à l'approche standard `earth::model.matrix()` : les fonctions de base du modèle earth (fonctions charnières, interactions et termes linéaires) deviennent les colonnes de la matrice de conception de `glmnet`. `glmnet` effectue ensuite une régression régularisée sur ces colonnes de base, en les sélectionnant et en les contractant via le lasso ou l'elastic net. Cette démarche combine la construction adaptative de la base par earth et la régularisation de `glmnet`.

9.6.1 Principaux avantages

- **Interactions interprétables** : les interactions d'earth fondées sur les fonctions charnières (y compris les termes à trois facteurs) ont une signification géométrique claire, contrairement aux interactions natives de `glmnet` par produits croisés, difficiles à expliquer devant un tribunal ou dans un contexte d'audit.
- **Non-linéarité automatique** : les fonctions charnières d'earth capturent des relations non linéaires qu'un modèle `glmnet` standard fondé sur les prédicteurs bruts ne détecterait pas.
- **Sélection de variables** : `glmnet` peut élaguer davantage la base earth, en éliminant les termes charnières qui ne contribuent pas.

9.6.2 Important : la colonne de pondération doit correspondre

Avertissement

Si `earthUI` utilise une colonne de pondération (p. ex. pour exclure la ligne du bien à évaluer avec un poids de 0), la même colonne de pondération doit être désignée dans la configuration des variables de `glmnetUI`. Si les paramètres de pondération diffèrent, les nombres de lignes ne concorderont pas et la base earth ne pourra pas être adjointe à la matrice de modèle de `glmnet`.

9.6.3 Flux de travail

1. Ajuster un modèle earth dans `earthUI` (degré quelconque).
2. Le fichier résultat `.rds` est enregistré automatiquement dans le dossier de sortie du projet.
3. Dans `glmnetUI`, ouvrir **Section 2 : Import from earthUI** et parcourir jusqu'au fichier `.rds`.
4. Vérifier que le **même fichier de données** et la **même colonne de pondération** sont utilisés dans les deux applications.
5. Configurer les paramètres de `glmnet` dans la section 5 (la matrice d'interaction est automatiquement désactivée lorsqu'un import `earthUI` est actif).
6. Cliquer sur **Fit Model**. Les colonnes de la base earth apparaissent dans les onglets Equation, Coefficients et Contributions.

10 Calculs RCA et téléchargement

Le bouton **Calculate RCA Adjustments & Download** (calcul des ajustements RCA et téléchargement) apparaît dans la section 7 de la barre latérale, visible uniquement en mode évaluation après l’ajustement d’un modèle. La RCA (Residual Constraint Approach) produit des ajustements dérivés du marché, par bien comparable, relatifs au bien à évaluer.

10.1 Ouverture de la boîte de dialogue RCA

Un clic sur le bouton ouvre une petite boîte de dialogue modale comportant :

- **Score type** (type de score) — boutons radio permettant de choisir **CQA** ou **CQA per SF** (l’option SF n’est disponible que lorsqu’une colonne `living_area` a été désignée). Si l’on choisit « CQA per SF », alors, sur la base du score CQA attribué au bien à évaluer, son score résiduel sera égal à sa surface habitable multipliée par le `residual_sf` correspondant au score CQA_SF donné.
- **CQA value** (valeur CQA) — champ de saisie du score CQA du bien à évaluer (0.00–10.00). Ce score représente la position que vous attribuez au bien à évaluer dans la distribution de qualité des biens comparables.
- Bouton **Generate** (générer) — calcule la sortie RCA et la télécharge au format Excel. Le bouton reste désactivé tant qu’un score valide n’a pas été saisi.

10.2 Interpolation du score CQA

Lorsque vous cliquez sur Generate, earthUI interpole le résidu du bien à évaluer à partir des couples CQA/résidu des biens comparables :

1. Les scores CQA et les résidus des biens comparables sont triés par CQA croissant
2. Une interpolation linéaire (`stats::approx()`) fait correspondre la valeur CQA saisie à un résidu
3. En cas d’utilisation de CQA per SF, le résidu par SF est reconverti en résidu total par multiplication par la surface habitable du bien à évaluer
4. La valeur estimée du bien à évaluer est calculée comme suit : prédiction du modèle + résidu interpolé

10.3 Colonnes de sortie

Le fichier Excel RCA (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) comprend toutes les colonnes de sortie intermédiaires, auxquelles s’ajoutent :

Colonne	Description
<code>subject_value</code>	Prédiction du modèle + résidu interpolé (ligne 1 uniquement)
<code>subject_cqa</code>	Le score CQA saisi (ligne 1 uniquement)
<code><variable>_adjustment</code>	Contribution du bien à évaluer moins contribution du comparable (par fonction g)
<code>residual_adjustment</code>	Résidu du bien à évaluer moins résidu du comparable
<code>net_adjustments</code>	Somme de tous les ajustements (contribution + résidu)
<code>gross_adjustments</code>	Somme des valeurs absolues de tous les ajustements
<code>residual_pct</code>	Ajustement résiduel en fraction du prix de vente du comparable
<code>net_adj_pct</code>	Ajustements nets en fraction du prix de vente du comparable

Colonne	Description
gross_adj_pct	Ajustements bruts en fraction du prix de vente du comparable
adjusted_sale_price	Prix de vente du comparable + ajustements nets

Les colonnes d'ajustement indiquent, pour chaque bien comparable, la part de la différence que le modèle attribue à chaque variable. **adjusted_sale_price** est le prix de vente du comparable après application de tous les ajustements dérivés du modèle — un ensemble de prix de vente ajustés qui devraient se regrouper autour de la valeur estimée du bien à évaluer.

10.4 L'onglet RCA Adjustments

Une fois la sortie RCA calculée, l'onglet **RCA Adjustments** (ajustements RCA) du panneau principal affiche des histogrammes des distributions Residual Adj %, Net Adj % et Gross Adj % sur l'ensemble des biens comparables (un jeu par cible pour les modèles multi-cibles), offrant un contrôle visuel rapide du degré de regroupement des ajustements.

***Conseil :** un score CQA de 5.00 place le bien à évaluer à la médiane des biens comparables. Un score supérieur à 5 indique un bien de qualité supérieure à la moyenne ; un score inférieur à 5, une qualité inférieure à la moyenne.*

11 Sales Comparison Grid

11.1 Vue d'ensemble et objet

La **Sales Comparison Grid** (grille de comparaison) est un classeur Excel généré en mode évaluation immobilière (section 8 de la barre latérale) après le calcul des ajustements RCA. Elle présente le bien à évaluer aux côtés des biens comparables sélectionnés dans un format de grille structuré, avec des formules Excel qui calculent automatiquement les ajustements ainsi qu'un prix de vente ajusté pour chaque comparable.

La grille est conçue pour le dossier de travail de l'évaluateur — elle combine les ajustements issus de la régression du modèle Earth avec des cellules modifiables dans lesquelles l'évaluateur peut ventiler le résidu CQA entre des caractéristiques spécifiques du bien. Le nom du fichier de sortie est **SalesGrid_<YYYYMMDD_HHMMSS>.xlsx**.

11.2 Fenêtre modale de sélection des comparables

Un clic sur « Generate Sales Grid & Download » ouvre une fenêtre modale dans laquelle on sélectionne les biens comparables à inclure :

- Les **Recommended comps** (comparables recommandés) sont cochés par défaut. Il s'agit de biens comparables présentant un pourcentage d'ajustement brut inférieur à 25% du prix de vente, triés par ancienneté de vente croissante (ventes les plus récentes en premier) et plafonnés à 30.
- Les **Additional comps** (comparables supplémentaires) sont listés en dessous, non cochés par défaut. Leurs ajustements bruts atteignent 25% ou plus, mais ils restent disponibles à la sélection (jusqu'à 50 affichés, triés par pourcentage d'ajustement brut croissant).
- Un maximum de **30 comparables** peut être sélectionné, produisant jusqu'à **10 feuilles** (3 comparables par feuille).

Après confirmation de la sélection, les comparables retenus sont triés par pourcentage d'ajustement brut croissant, de sorte que les biens comparables les plus similaires apparaissent sur la feuille 1.

11.3 Disposition de la grille

Chaque feuille comporte une disposition sur 20 colonnes accueillant le bien à évaluer plus 3 biens comparables. Les colonnes de chaque entité sont les suivantes :

Entité	Colonnes (5 par entité)
Bien à évaluer	Label, Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC)
Chaque comparable	Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC), Adjustment

Les lignes, de haut en bas, sont les suivantes :

1. **Titre** — nom de la feuille (p. ex. « Intermediate Sales Comparable Grid — Sheet 1 of 3 »)
2. **En-têtes** — en-têtes de colonnes « Subject » et comparables, avec l'adresse
3. **Adresse** — adresse complète du bien
4. **APN | MLS# | DOM | Subj.Prox** — numéro de parcelle, identifiant du listing, nombre de jours sur le marché et distance de Haversine (en miles) par rapport au bien à évaluer
5. **Sales Price | Concess. | Net SP** — prix de vente, concessions et formule du prix de vente net

6. Ligne d'en-tête **Regression Features**
7. **BASE VALUE** — l'ordonnée à l'origine du modèle
8. **Date of Sale | OffMkt | OnMkt** — date du contrat, ancienneté de la vente et DOM
9. **Lignes groupées** (conditionnelles) — groupes Location, Site Size et/ou Age
10. **Lignes de variables du modèle** — une ligne par prédictor (hors variables groupées)
11. **Ligne de séparation vide**
12. Ligne d'en-tête **Residual Features**
13. **CQA|Residual** — score CQA + formule du résidu restant
14. **Lignes de caractéristiques résiduelles** — lignes nommées + lignes vides pour la saisie de l'évaluateur
15. **Total VC / Net Adjustment**
16. **Net Adjustment %**
17. **Gross Adjustment %**
18. **Adjusted Sale Price** — ligne de formules
19. Pied de page **Copyright**

11.4 Prix de vente, concessions et Net SP

La ligne 5 affiche trois valeurs pour chaque bien comparable :

- **Sale Price** — le prix de vente du comparable issu des données
- **Concessions** — la valeur de la colonne désignée comme **concessions** (0 si aucune n'est désignée)
- **Net SP** — une formule Excel : Sale Price – Concessions

La colonne du bien à évaluer affiche « N/A » pour le prix de vente (puisqu'il est inconnu) et, le cas échéant, la valeur des concessions si elle est disponible.

11.5 Lignes groupées (localisation, site, âge)

Lorsque certains types spéciaux sont désignés et que ces variables figurent dans le modèle, earthUI crée des lignes groupées qui combinent les variables apparentées :

Loc : Long | Lat | Area — Apparaît lorsque l'une des variables **longitude**, **latitude** ou **area** figure dans le modèle. Affiche les valeurs factuelles de chaque variable constitutive, une contribution à la valeur combinée (somme des VC individuelles) et, pour les comparables, un ajustement (VC combinée du bien à évaluer – VC combinée du comparable). Les cellules VC sont stylisées avec un fond bleu clair.

Site Size | Dimensions — Apparaît lorsque **lot_size** ou **site_dimensions** figure dans le modèle. Même structure que le groupe Location.

Actual Age | Effective Age — Apparaît lorsque **actual_age** ou **effective_age** figure dans le modèle. Même structure.

Les variables consommées par les lignes groupées sont **exclues des lignes de variables individuelles du modèle** situées en dessous, ce qui évite tout double comptage dans les totaux d'ajustement.

11.6 Lignes de variables du modèle

Sous les lignes groupées (ou directement sous BASE VALUE en l'absence de lignes groupées), une ligne par prédictor restant du modèle affiche :

- Les **valeurs factuelles** pour le bien à évaluer et chaque comparable (les valeurs réelles des données)
- La **contribution à la valeur (VC)** — la contribution de la fonction g à la valeur prédite pour cette ligne

- L'**ajustement** (comparables uniquement) — VC du bien à évaluer moins VC du comparable

11.7 Ligne CQA|Residual et lignes de caractéristiques résiduelles

La ligne **CQA|Residual** affiche le score CQA de chaque bien et contient une formule pour le **résidu restant** — la part du résidu non encore ventilée entre des caractéristiques spécifiques. La formule est :

$$\text{Remaining Residual} = \text{Total Residual} - \sum (\text{Residual Feature VCs below})$$

En dessous se trouvent les **lignes de caractéristiques résiduelles** — 5 lignes nommées (Location, View/Appeal, Condition, Quality, Other, soit localisation, vue/attractif, état, qualité, autre) plus 6 lignes vides. Ce sont des cellules de saisie dans lesquelles l'évaluateur inscrit les contributions à la valeur des caractéristiques non capturées par le modèle. À mesure que l'évaluateur saisit des valeurs, la formule du résidu restant (Remaining Residual) diminue automatiquement.

Pour chaque comparable, la colonne d'ajustement contient une formule : VC de la caractéristique du bien à évaluer — VC de la caractéristique du comparable.

11.8 Formule du prix de vente ajusté (Adjusted Sale Price)

La ligne **Adjusted Sale Price** contient des formules Excel :

- **Bien à évaluer** : somme de toutes les cellules de contribution à la valeur, depuis BASE VALUE jusqu'à la dernière ligne de caractéristique résiduelle. Elle représente la prédiction totale du modèle pour le bien à évaluer, augmentée des caractéristiques résiduelles ventilées par l'évaluateur.
- **Comparables** : Net SP + somme de toutes les cellules d'ajustement, depuis la première ligne groupée ou de variable jusqu'à la dernière ligne de caractéristique résiduelle. On obtient ainsi le prix de vente du comparable après tous les ajustements issus de la régression et saisis par l'évaluateur.

11.9 Protection des feuilles et cellules modifiables

Chaque feuille est **protégée** afin d'empêcher toute modification accidentelle des formules et des données. La protection verrouille :

- Toutes les cellules de formules (Net SP, résidu restant, ajustements, prix de vente ajusté, etc.)
- Toutes les cellules de données (adresses, prix de vente, valeurs factuelles, contributions à la valeur issues du modèle)
- Les libellés et les en-têtes

Les seules cellules **déverrouillées** (modifiables) sont les cellules de saisie des contributions à la valeur des caractéristiques résiduelles — les cellules situées sous la ligne CQA|Residual dans lesquelles l'évaluateur saisit ses ventilations. Elles sont stylisées avec un fond jaune clair pour indiquer qu'elles sont modifiables.

11.10 Utilisation de la grille dans Excel

1. **Ouvrir le fichier** dans Excel ou dans un tableur compatible.
2. **Examiner les ajustements de régression** — les ajustements dérivés du modèle sont pré-remplis et verrouillés.

3. **Ventiler le résidu** — dans les lignes de caractéristiques résiduelles (cellules jaunes), saisir votre appréciation de la part du résidu restant attribuable à chaque caractéristique (Location, View, Condition, Quality, etc.). La formule du résidu restant diminue au fur et à mesure de la ventilation.
4. **Vérifier l'Adjusted Sale Price** — cette formule se met à jour automatiquement à mesure que les valeurs des caractéristiques résiduelles sont saisies.
5. **Feuilles multiples** — si plus de 3 comparables ont été sélectionnés, naviguer entre les feuilles. Chaque feuille présente le même bien à évaluer dans la colonne de gauche, avec 3 comparables différents.

***Conseil :** l'objectif est de ventiler le résidu CQA entre les lignes de caractéristiques jusqu'à ce que le résidu restant soit proche de zéro. Cela démontre que l'on peut rendre compte de l'écart entre l'estimation par régression de chaque comparable et son prix de vente effectif.*

12 Téléchargement des rapports

Après l'ajustement d'un modèle, la génération de rapports suit un flux de travail Quarto en deux étapes : **Generate Quarto Report** (section 9 de la barre latérale en mode évaluation immobilière, 7 sinon) écrit une archive de rapport autonome, et **Convert Quarto Report** (section 10) en effectue le rendu — ou celui de tout autre fichier Quarto — vers les formats finaux.

12.1 Generate Quarto Report

Un clic sur **Generate Quarto Report** écrit une archive Quarto autonome dans le dossier de sortie du projet, à l'emplacement `<base>_qmd/` ; celle-ci contient la source `.qmd` renseignée, tous les graphiques pré-générés (PNG et PDF), les données du rapport et un modèle de référence Word. Aucun rendu n'a lieu à cette étape — l'archive peut être modifiée à la main, combinée à d'autres sources Quarto au moyen des inclusions Quarto, ou prévisualisée de manière autonome avec `quarto preview`.

12.2 Convert Quarto Report

La section 10 propose un sélecteur de fichiers (pointant par défaut vers l'archive la plus récemment générée du projet courant — mais on peut naviguer vers n'importe quel fichier `.qmd`) ainsi que des cases à cocher de format :

- **HTML** — un fichier `.html` avec rendu mathématique KaTeX, images intégrées et table des matières. Utilise le thème Bootstrap Flatly et la police Roboto Condensed.
- **Word** — un fichier `.docx` adapté à l'édition et à la diffusion. Comprend une table des matières, la numérotation des pages, et s'appuie sur un modèle de référence personnalisé assurant une mise en forme homogène.
- **PDF** — composé avec LuaLaTeX pour une sortie de qualité professionnelle. Le format du papier suit le paramètre régional (Letter ou A4). Comprend des pages en orientation paysage pour les grandes matrices d'interactions et utilise Roboto Condensed avec Latin Modern Math pour les équations. L'option PDF n'est proposée que lorsqu'une distribution LaTeX est détectée.

Cliquez sur **Convert** pour effectuer le rendu dans les formats sélectionnés. Le rendu s'exécute en arrière-plan — une boîte de dialogue modale affiche le temps écoulé et la progression de Quarto tandis que l'application reste réactive. Une notification confirme l'emplacement du fichier à l'achèvement.

Toutes les opérations (ajustement de modèles, téléchargements de données, calculs RCA, génération de Sales Grids et rendu de rapports) sont consignées, avec horodatages de début et de fin et durées écoulées, dans `<datafile>_earthui_log.txt` au sein du dossier de sortie, à des fins de débogage et de suivi des performances.

12.3 Contenu du rapport

Chaque rapport généré comprend les sections suivantes :

1. **Description du jeu de données** — nombre d'observations, variable(s) cible(s), prédicteurs utilisés, prédicteurs catégoriels
2. **Spécification du modèle** — degré, statut de la validation croisée, nombre de termes, prédicteurs présents dans le modèle
3. **Interactions autorisées** (degré ≥ 2 uniquement) — la matrice d'interactions avec coches, mise en forme selon le type de sortie

4. **Synthèse des résultats** — R^2 , GRSq, GCV, RSS et CV R^2 (si la validation croisée était activée)
5. **Équation du modèle** — l'équation complète du modèle Earth en notation LaTeX
6. **Coefficients et fonctions de base** — tableau de tous les termes, de leurs coefficients et des fonctions de base qui les définissent
7. **Importance des variables** — diagramme en barres et tableau ordonné des scores d'importance des prédicteurs
8. **Contributions des fonctions g** — graphiques pour chaque fonction g : courbes pour les termes univariés, graphiques en perspective 3D et graphiques de contours pour les interactions bivariées
9. **Matrice de corrélation** — carte thermique des corrélations entre prédicteurs
10. **Diagnostics** — graphiques « Residuals vs Fitted », « Normal Q-Q » et « Actual vs Predicted »
11. **Décomposition ANOVA** — tableau présentant la contribution de chaque fonction de base à la RSS
12. **Sortie Earth** — sortie texte brute de `earth::summary()`, y compris le détail de la passe d'élagage

Pour les modèles à réponses multiples, les sections 4–8 et 10 sont répétées pour chaque variable cible.

***Conseil** : les rapports nécessitent le package R `quarto` et une installation de Quarto. Pour la sortie PDF, une distribution LaTeX doit également être installée (p. ex. via `tinytex::install_tinytex()`) — à défaut, la case à cocher PDF est masquée.*

13 Les Post Processing Modules : Elastic Net et GAM

Le package earthUI est livré avec deux Post Processing Modules (modules de post-traitement) — une routine elastic net construite sur `glmnet` et une routine GAM construite sur `mgcv`. Chacune fait l’objet d’un traitement approfondi dans son propre article de ce numéro ; la présente section explique à quoi elles *servent* dans le flux de travail earthUI et comment les trois routines coopèrent.

13.1 Pourquoi des ajustements d’appui ?

Le modèle earth produit la conclusion de valeur. Les Post Processing Modules existent pour la **corroborer** : deux estimations supplémentaires du même bien à évaluer, produites par des moteurs méthodologiquement différents, ajustées sur les mêmes données de projet. Lorsqu’une ré-estimation elastic net de la base earth et un lissage GAM indépendant aboutissent à une valeur matériellement identique, la conclusion de l’évaluateur ne repose plus sur les hypothèses d’un seul modèle — un avantage significatif dans les contextes de revue d’évaluation et de contentieux. Lorsqu’elles sont en *désaccord*, ce désaccord est en lui-même un diagnostic, signalant le plus souvent des données clairsemées, un terme trop flexible ou un comparable aberrant méritant d’être réexaminé. Aucun des deux modules ne prime sur le modèle earth ; ils constituent des éléments de preuve à l’appui, délibérément présentés sans pondération.

13.2 Lancement et état partagé

Chaque routine est une application Shiny locale distincte :

- `earthUI::launch_glmnet()` — l’application elastic net, sur le port 7879.
- `earthUI::launch_mgcv()` — l’application GAM, sur le port 7880.

Les trois applications lisent la même arborescence de projet et la même base de données de paramètres, de sorte que le projet ouvert dans earthUI est immédiatement disponible dans les modules — de même que les champs à report automatique qu’earthUI enregistre pour lui : la Response (cible), l’Effective Date (date d’effet), le type et la valeur de CQA du bien à évaluer pour la RCA, et la désignation de la surface habitable. Vous ne les ressaisissez pas ; les modules les lisent, si bien que les mêmes valeurs circulent à travers les trois ajustements.

13.3 La routine elastic net

La routine `glmnet` importe les fonctions de base du modèle earth — ses charnières, ses interactions et ses termes linéaires — comme colonnes d’une matrice de conception et les ré-estime sous régularisation lasso/elastic-net (voir « `glmnetUI` Import » ci-dessus pour les mécanismes et la mise en garde relative à la colonne de pondération). Le résultat conserve les termes interprétables et géométriquement signifiants d’earth, tout en laissant le régularisateur rétrécir ou éliminer ceux qui ne justifient pas leur maintien. Voir l’article `glmnetUI` de ce numéro pour le traitement complet.

13.4 La routine GAM

La routine `mgcv` procède en sens inverse : plutôt que de réutiliser la base d’earth, elle ajuste de nouveaux termes de lissage pénalisés, en prenant les emplacements de nœuds du modèle earth comme points de départ (voir « `mgcvUI` Auto-Export » ci-dessus). Comme le GAM re-dérive les formes fonctionnelles sous un paradigme de lissage différent, sa concordance avec le modèle earth constitue une corroboration véritablement indépendante et non un simple écho. Les modèles earth avec $\text{degree} \leq 2$ sont pris en charge. Voir l’article `mgcvUI` de ce numéro pour le traitement complet.

13.5 Un flux de travail pratique

1. Ajustez et affinez d'abord le **modèle earth**, dans earthUI, jusqu'à ce qu'il vous satisfasse. Chaque ajustement réussi ($\text{degré} \leq 2$) exporte automatiquement le fichier **.rds** que consomment les modules.
2. Ouvrez les routines **glmnet** et **mgcv** sur le même projet, importez le résultat earth et ajustez. Travaillez dans une routine à la fois — les ajustements sont de longue durée et les routines se coordonnent via le projet partagé, si bien que le rythme raisonnable est séquentiel, et non parallèle.
3. Itérez librement. Chaque application conserve ses résultats pendant que vous passez de l'une à l'autre, et rien ne vous oblige à générer des rapports en cours de route.
4. Ce n'est que lorsque les trois résultats vous satisfont que vous revenez dans chaque routine pour générer son rapport Quarto. Chaque méthode enregistre son ajustement et sa conclusion de valeur dans le registre de trilogie du projet, ce qui maintient la traçabilité des trois rapports vers une même exécution comparative.

***Conseil :** les modules sont facultatifs à tous les niveaux : le flux de travail propre d'earthUI — ajustement, RCA, Sales Grids, rapports — est complet sans eux. Recourez-y lorsqu'une conclusion mérite corroboration : un bien à évaluer difficile, un segment de marché étroit, ou une mission susceptible de faire l'objet d'une revue.*

14 Jeu de données de démonstration : Appraisal_1.csv

earthUI inclut un jeu de données MLS de démonstration permettant d'explorer le flux de travail d'évaluation immobilière. On le charge par programmation avec :

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

Ou bien, dans l'application Shiny, copiez-le dans le dossier `in/` d'un projet (ou joignez-le comme fichier de données initial lors de la création du projet) et sélectionnez-le dans la Section 2 (Import Data — importation des données).

14.1 Description

Le fichier contient 1,501 ventes résidentielles plus 1 bien à évaluer en ligne 1 (soit 1,502 lignes au total), issues d'un export MLS simulé. Les données représentent des ventes de maisons individuelles dans un marché à plusieurs zones, avec une gamme variée de surfaces, d'âges et de localisations.

Il ne s'agit pas de données réelles, mais elles s'appuient sur un quartier réaliste de Californie du Nord. Toutes les informations d'identification ont été modifiées ou supprimées.

14.2 Colonnes

Colonne	Type	Type spécial	Description
weight	numeric	—	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 de listing MLS
parcel_number	character	display_only	Numéro de parcelle du cadastre du comté (APN)
street_address	character	display_only	Adresse du bien
city_name	character	display_only	Ville
postal_code	character	display_only	Code postal (ZIP)
county_name	character	display_only	Comté
contract_date	Date	contract_date	Date du contrat de vente (sert au calcul de sale_age)
sale_age	numeric	—	Nombre de jours entre la date du contrat et la date d'effet (précalculé)
coe_date	Date	display_only	Date de clôture de l'entiercement (close of escrow)
listing_status	character	display_only	Statut du listing (p. ex. "Sold")
sale_price	numeric	(target)	Prix de vente — variable réponse
rent	numeric	—	Loyer mensuel (pour les modèles à cibles multiples)
list_price	numeric	display_only	Prix affiché du listing
original_list_price	numeric	display_only	Prix affiché initial
living_sqft	numeric	living_area	Surface habitable brute en pieds carrés
beds_total	integer	—	Nombre de chambres
baths_total	numeric	—	Nombre total de salles de bains (p. ex. 2.5 = 2 complètes + 1 demi)
lot_size	numeric	lot_size	Surface 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 pleine précision
longitude6	numeric	display_only	Longitude en pleine précision
garage_spaces	integer	—	Nombre de places de garage
fp_count	integer	—	Nombre de cheminées
no_of_stories	numeric	—	Nombre de niveaux
style	character	—	Style architectural
view	character	—	Type de vue (p. ex. "Neighborhood", "Hills")
days_on_market	integer	dom	Nombre de jours sur le marché
listing_date	Date	listing_date	Date de mise en vente
sale_concessions	numeric	concessions	Concessions du vendeur

14.3 Démarrage rapide suggéré

1. Lancer earthUI : `earthUI::launch()`
2. Créer un projet dont le champ Purpose (finalité) est réglé sur **For Appraisal** dans l'assistant New Project, en joignant `Appraisal_1.csv` comme fichier de données initial
3. Sélectionner `Appraisal_1.csv` dans la Section 2 (Import Data)
4. Sélectionner `sale_price` comme cible
5. Attribuer les types spéciaux indiqués dans le tableau ci-dessus
6. Inclure les prédicteurs : `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (comme facteur), `age`, `latitude`, `longitude`, `garage_spaces`
7. Fixer le degré à 1, puis cliquer sur **Fit Earth Model**
8. Télécharger la sortie intermédiaire (étape 6) et examiner le classement CQA
9. Calculer les ajustements RCA (étape 7) avec un score CQA de ~5.00
10. Générer la Sales Comparison Grid (grille de comparaison des ventes) (étape 8)

Conseil : le jeu de données de démonstration constitue un bon moyen d'explorer toutes les fonctionnalités d'earthUI avant d'importer vos propres données MLS. Les prédicteurs suggérés ci-dessus produisent généralement un modèle dont le R^2 dépasse 0.85.

15 Fonctionnalité expérimentale : extraction de texte en Prolog et règles de dérivation

La version 0.10.0 introduit un dispositif **optionnel et authentiquement expérimental** d'extraction de caractéristiques prêtes pour le modèle à partir des colonnes de texte descriptif d'un export MLS — la colonne « Public Remarks » avant tout, mais aussi les remarques d'agent et les listes de caractéristiques délimitées par des virgules, telles que les équipements des pièces ou les constructions existantes. Le moteur d'extraction est Prolog, et plus précisément les **grammaires à clauses définies (DCG, Definite Clause Grammars)** : les règles qui reconnaissent des expressions comme « move in ready », « 3 car garage » ou « owned solar » dans le texte des remarques sont écrites sous forme de règles de grammaire Prolog.

Cette fonctionnalité est encore en développement. Elle peut être utilisée *partiellement* pour l'extraction dès aujourd'hui — la grammaire fournie reconnaît déjà un ensemble utile de caractéristiques — mais son adaptation au vocabulaire de votre propre MLS exige de savoir lire et écrire du Prolog, et des DCG en particulier. Si ce n'est pas votre cas, ignorez entièrement ce chapitre : la fonctionnalité est désactivée par défaut, et le reste d'earthUI fonctionne exactement de la même manière sans elle.

Avertissement

Il s'agit d'une fonctionnalité expérimentale en cours de développement actif. Sa grammaire d'extraction n'est ni complète ni validée pour un travail d'évaluation immobilière en production — les colonnes extraites doivent être vérifiées par rapport au texte source avant d'être utilisées dans un modèle. Un usage efficace de cette fonctionnalité requiert une connaissance pratique de Prolog et des DCG. Tout ce qu'elle fait est additif et réversible ; en cas de comportement anormal, désactivez l'option dans Settings et poursuivez normalement.

15.1 Ce qu'elle fait

Les listings MLS véhiculent une grande quantité d'informations pertinentes pour la valeur dans du texte libre et dans des champs de listes délimitées. Cette étape optionnelle transforme ce texte en colonnes ordinaires exploitables par votre régression, via deux rôles de colonne spéciaux (chapitre 6) :

- les **colonnes remarks** sont analysées par la grammaire DCG Prolog, qui produit une colonne de caractéristique par attribut extrait. Le préfixe de colonne est construit à partir des premières lettres du nom de la colonne source : **public_remarks** donne des colonnes **pr_***, **agent_remark** donne **ar_***, et ainsi de suite.
- les **colonnes list** (p. ex. **existing_structures = "Shed, Workshop"**) sont scindées en colonnes TRUE/FALSE à encodage one-hot **<col>_<value>**. Cette partie est du R pur et ne requiert aucun Prolog.

La grammaire fournie (**inst/prolog/mls_remarks.pl**) extrait actuellement, entre autres : un niveau d'état gradué avec un **condition_score** de 1–5 ; des indicateurs de vue et des types de vue ; des indicateurs de rénovation de cuisine et de salle de bains ; le parquet et d'autres revêtements de sol ; une liste et un décompte des améliorations ; les places et la surface de garage ; piscine, spa et autres aménagements extérieurs ; le nombre de cheminées ; les dépendances (shed, barn, workshop, casita, guest house, kennel, etc.) avec leurs tailles analysées à partir de dimensions telles que « 40x50 » ; le stationnement et le remisage de camping-cars (RV) ; la propriété ou la location des panneaux solaires, le stockage par batterie ; la surface du terrain (y compris la conversion acres-pieds carrés) ; des montants de prix liés au contexte de vente rattachés à

un sujet (réduction de prix, cotisations HOA, location ou rachat du solaire, Mello-Roos) ; des indices de localisation ; et un score de sentiment points-forts/points-faibles dans $[-1, 1]$.

15.2 Prérequis : vProlog (Trealla), et non SWI-Prolog

L'analyse des remarques s'exécute sur le package R **vProlog** — un moteur Prolog Trealla autonome. Il n'est pas disponible sur CRAN ; installez-le depuis le dépôt r-universe de l'auteur comme indiqué dans « Installation d'earthUI et de vProlog » au début de cet article (binaires précompilés pour Windows et macOS, sources à github.com/wcraytor/vProlog). Aucune installation système de Prolog (SWI-Prolog ou autre) n'est requise ni utilisée. Si vProlog est absent, l'analyse des remarques est ignorée avec un message, tandis que la scission des colonnes `list` s'exécute quand même. La fonction utilitaire `vProlog_available()` indique si le moteur et la grammaire fournie sont tous deux présents.

15.3 Activation et exécution

1. Ouvrez **Settings** (engrenage) et cochez **Enable Prolog / list column processing** (activer le traitement Prolog / des colonnes de liste). L'option est désactivée par défaut, porte la mention « advanced feature — for users with Prolog expertise » (fonctionnalité avancée — pour les utilisateurs maîtrisant Prolog), est mémorisée par projet et apparaît dans la boîte de dialogue Project Information.
2. Dans le tableau des prédicteurs, attribuez le rôle **Special remarks** ou `list` aux colonnes concernées (plusieurs colonnes peuvent porter chacun de ces rôles).
3. Utilisez la section **Execute Prolog Processing** (exécuter le traitement Prolog), située entre les Sections 4 et 5 de la barre latérale. Elle propose deux étapes :

1. Expand lists & remarks — exécute la grammaire DCG sur chaque colonne `remarks` et scinde en one-hot chaque colonne `list`. L'expansion est *incrémentale* : seules les colonnes désignées depuis la dernière exécution sont traitées, et le bouton est désactivé dès que tout ce qui a été désigné a déjà été développé. Une barre de progression suit l'analyse.

2. Execute derivation rules — applique vos propres règles Prolog `derive/2` aux colonnes développées (voir ci-dessous). Les règles sont vérifiées syntaxiquement et validées par une exécution à blanc avant leur application.

15.4 Règles de dérivation

Pour chaque ligne de données, earthUI affirme les colonnes de la ligne sous forme de faits `cell(Name, Value)` et collecte toutes les solutions de vos règles `derive(Col, Val)` — chaque `Col` distinct devient une nouvelle colonne, et la *première* valeur trouvée par colonne l'emporte, ce qui autorise une clause de repli à suivre. Des prédicats utilitaires (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) facilitent l'agrégation et le traitement des NA, et des faits `drop/1` suppriment des colonnes après l'exécution des règles. Par exemple :

```
% Merge a collinear cluster of solar flags into one boolean.
derive(solar_any, B) :-
  ( cell(pr_has_solar, true)
  ; cell(pr_solar_owned, true)
  ; cell(pr_has_battery, true) ) -> B = true ; B = false.

% Count amenities instead of OR-ing them (numeric column).
derive(amenity_count, N) :- count_true_prefix(pr_has_, N).

% Threshold -> boolean, with a fallback clause.
derive(big_outbuilding, true) :- cell(pr_outbuilding_sqft, S), S >= 1000.
```

```
derive(big_outbuilding, false).

% Discard a raw granular column once consolidated.
drop(pr_price_solar_lease).
```

Les règles résident dans un fichier `<project>_rules.pl` propre à chaque projet, initialisé à partir d'un gabarit entièrement commenté, et s'éditent via **Edit derivation rules...** (éditer les règles de dérivation) dans un éditeur de code flottant et déplaçable, qui peut aussi être ancré dans la barre latérale. Un bouton **Check** (vérifier) valide les règles (analyse syntaxique, vérification du chargement et exécution à blanc sur la ligne 1 de vos données) sans les exécuter.

15.5 Fichier de données actif versionné

Après l'une ou l'autre étape, le data frame augmenté est écrit dans le dossier d'entrée du projet sous le nom `<original_name>_<YYMMDD_HHMMSS>.xlsx` et devient le fichier de données **actif** du projet — la réouverture du projet charge les données augmentées, de sorte que les expansions ne sont jamais recalculées, et c'est ce même fichier que consomment glmnetUI et mgcvUI. Les colonnes sources de remarques brutes sont automatiquement écartées du modèle ; les colonnes générées ne sont *pas* incluses automatiquement comme prédicteurs. L'analyse de texte peut produire de nombreuses colonnes ; marquez donc celles que vous ne voulez pas (encore) dans la régression comme **Latent**, ou désactivez-les via $\times$ (chapitre 6).

15.6 Interface programmatique et fichiers inclus

L'ensemble du pipeline est exporté pour un usage en traitement par lots : `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` orchestre les passes, en s'appuyant sur `remarks_to_columns()`, `split_list_column()`, `apply_user_rules()` et `validate_rules()`. Les sources Prolog fournies sous `inst/prolog/` sont `mls_remarks.pl` (la grammaire DCG et l'agrégation des caractéristiques), `mls_analyze.pl` (un pilote d'entrées-sorties de fichiers pour l'analyse par lots), `mls_rules.pl` (les prédicats utilitaires accessibles aux règles de dérivation) et `climate_regions_ca.pl` (ci-dessous).

15.7 Expérimentation connexe : a priori par région climatique

Un dispositif expérimental complémentaire traite les caractéristiques à *données clairsemées* — celles que la régression ne peut pas estimer parce qu'elles n'apparaissent que dans une poignée de biens comparables (un refroidisseur évaporatif, par exemple). Trois fonctions utilitaires R — `climate_region_for()`, `market_area_profile()` et `climate_feature_priors()` — classent la localisation d'un projet (comté + ville) dans l'une des sept régions climatiques californiennes reconnues par les acheteurs (une réduction pratique des 16 zones climatiques du bâtiment de la California Energy Commission) et renvoient des a priori ordinaux de valeur contributive pour les équipements de chauffage, ventilation et climatisation (chauffage, climatisation, refroidisseur évaporatif, ventilateur de maison entière, filtration) par région. Le module `Prolog climate_regions_ca.pl` reproduit la même classification pour un usage au sein des règles de dérivation ; les deux versants sont générés à partir d'une source R canonique unique, de sorte qu'ils ne peuvent pas diverger. Ces a priori sont destinés à *éclairer* l'évaluation des caractéristiques et à servir de contrôle de plausibilité des coefficients estimés — les affectations de région (en particulier pour les comtés s'étendant sur plusieurs régions) doivent être revues par l'évaluateur.

15.7.1 Au-delà de la Californie : packs juridictionnels et modules de zone de marché

La Californie est la première juridiction, non la frontière. earthUI est utilisé à l'international, et la connaissance des a priori climatiques est en cours de restructuration en **packs juridictionnels**

(« jurisdiction packs ») : des tables de faits autonomes, de forme purement déclarative (définitions de zones, affectations d'unités administratives, a priori de caractéristiques), indexées sur les codes de pays et de niveaux administratifs du système de projet, de sorte qu'un pack peut décrire un État américain, un Land allemand ou un petit pays tout entier. Une couche complémentaire de **modules de zone de marché** (« market-area modules ») capture le vocabulaire local et des a priori affinés pour une zone de marché nommée.

Les deux sont conçus pour être rédigés *sur le terrain* par des Valuation Engineers et des évaluateurs, dans leur propre langue — et, de façon réaliste, avec l'assistance d'un LLM : le monde ne compte pas beaucoup de programmeurs Prolog, aussi les artefacts de terrain sont-ils délibérément des tables plutôt que des programmes. Le format de pack est un schéma documenté accompagné d'un exemple complet (la Californie elle-même), expressément pour qu'un assistant tel que Claude puisse rédiger un pack valide à partir d'une description en langage courant d'un marché ; earthUI génère le Prolog sous-jacent à partir du pack, et un validateur bloquant (contrôles de schéma, de couverture, d'intégrité référentielle et d'exécution à blanc) doit être satisfait avant qu'un pack puisse être activé. Chaque pack rédigé sur le terrain porte une provenance obligatoire — l'auteur en titre, la date et l'autorité citée — de sorte que c'est le professionnel responsable, et non l'outil de rédaction, qui répond de la connaissance, et que les rapports peuvent citer le pack exact et la version utilisée.

Conseil : un schéma pratique : développez les remarques, consolidez les indicateurs granulaires *pr_** en une poignée de décomptes ou de booléens dérivés au moyen de règles *derive/2*, supprimez avec *drop/1* les indicateurs bruts ainsi consolidés, et ne conservez comme prédicteurs candidats que les colonnes dérivées. La liste des prédicteurs reste ainsi courte et la colinéarité maîtrisable.

Références

- FRIEDMAN, Jerome H. (1991). « Multivariate Adaptive Regression Splines ». In : *The Annals of Statistics* 19.1, p. 1-141. DOI : [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- MILBORROW, Stephen (1^{er} oct. 2024). *earth: Multivariate Adaptive Regression Splines*. R package. URL : <https://CRAN.R-project.org/package=earth>.
- (s. d.[a]). *Notes on the earth Package*. URL : <http://www.milbo.org/doc/earth-notes.pdf>.
- (s. d.[b]). *Variance Models in earth*. URL : <http://www.milbo.org/doc/earth-varmod.pdf>.