

earthUI: An Interactive Interface for the earth (MARS) Package

William Bert Craytor

Vol. 1, Issue 1 · Summer 2026

Abstract. earthUI is a graphical user interface for the R `earth` package, which implements Multivariate Adaptive Regression Splines (MARS). It offers three purpose modes—general predictive modeling, real-estate appraisal, and market-area analysis—and guides the user through data import, model configuration, fitting, and the interpretation of earth’s diagnostics and graphical output. This article documents earthUI’s data-format requirements, modeling workflow, output displays, and complete feature reference.

Keywords: MARS, multivariate adaptive regression splines, earth, R, graphical user interface, real estate valuation

Contents

1	Introduction	3
1.1	What Is earthUI?	3
1.2	Three Purpose Modes	3
1.3	Real Estate–Specific Features	4
1.4	Getting Started	4
2	MLS Input Data Requirements	5
2.1	File Format & Structure	5
2.2	Required Columns for Appraisal Mode	5
2.3	Special Column Naming Conventions	6
2.4	Data Quality & Completeness	6
2.5	Subject Row Placement	6
3	General Purpose Mode	8
3.1	Overview	8
3.2	The Sidebar Workflow	8
3.3	Main Panel Tabs	8
3.4	Settings Persistence	9
3.5	Dark Mode	9
3.6	Locale & Regional Settings	9
	3.6.1 Supported Countries	10
	3.6.2 Override Dropdowns	10
	3.6.3 Saving Defaults	10
4	Appraisal Mode	11
4.1	Subject Row Handling	11
4.2	Effective Date & Sale Age	11
4.3	Special Column Designations	11
4.4	RCA Adjustments Overview	12
5	Market Area Analysis Mode	13
5.1	Differences from Appraisal Mode	13
5.2	When to Use Market Mode	13

6	Variable Selection	14
6.1	Target Variable(s)	14
6.2	The Predictor Table	14
6.3	Data Type Detection & Overrides	14
6.4	Factor and Linear Flags	14
6.5	Special Column Types Reference	15
6.6	Multiple Targets	15
7	Parameter Selection	17
7.1	Basic Parameters	17
7.2	Forward Pass	17
7.3	Allowed Interactions	18
7.4	Pruning	18
7.5	Cross-Validation	18
7.6	Subset Filtering	18
7.7	Recommended Parameter Values	19
7.7.1	Forward Pass Parameters	19
7.7.2	Pruning Parameters	20
7.7.3	Cross-Validation Parameters	20
7.7.4	Variance Model	21
7.7.5	Earth Output File	21
7.8	Settings Defaults	21
8	Fitting the Earth Model	22
8.1	The Fit Button	22
8.2	Fitting Modal & Trace Output	22
8.3	Fit Log	22
9	Downloading Data	23
9.1	Output Columns	23
9.2	Column Ordering & Excel Formatting	23
9.3	CQA Scores	24
9.4	Sorting & Subject Row	24
9.5	mgcvUI Auto-Export	24
9.6	glmnetUI Import	24
9.6.1	Key Benefits	25
9.6.2	Important: Weight Column Must Match	25
9.6.3	Workflow	25
10	RCA Calculations & Downloading	26
10.1	Opening the RCA Dialog	26
10.2	CQA Score Interpolation	26
10.3	Output Columns	26
11	Sales Comparison Grid	28
11.1	Overview & Purpose	28
11.2	Comp Selection Modal	28
11.3	Grid Layout	28
11.4	Sale Price, Concessions & Net SP	29
11.5	Grouped Rows (Location, Site, Age)	29
11.6	Model Variable Rows	29
11.7	CQA Residual & Residual Feature Rows	29

11.8 Adjusted Sale Price Formula	30
11.9 Sheet Protection & Editable Cells	30
11.10 Working with the Grid in Excel	30
12 Downloading Reports	31
12.1 Report Formats	31
12.2 Report Contents	31
13 Demo Dataset: Appraisal_1.csv	32
13.1 Description	32
13.2 Columns	32
13.3 Suggested Quick Start	33

1 Introduction

1.1 What Is earthUI?

earthUI is a graphical user interface for the R `earth` package, which implements Multivariate Adaptive Regression Splines (MARS) (Friedman 1991). The modelling engine is Stephen Milborrow’s `earth` package (Milborrow 2024), documented in detail by Milborrow (n.d.[a]) and Milborrow (n.d.[b]). It runs as a local Shiny application — there is no login, no server, and no accounts. You launch it from R, import a dataset (CSV or Excel), configure your model, and fit it interactively.

The application provides a complete workflow: data import, variable configuration, model fitting, diagnostic plots, variable importance, model equations, and downloadable reports in HTML, Word, or PDF format.

1.2 Three Purpose Modes

When you launch earthUI, a **Purpose** radio button at the top of the sidebar lets you choose one of three modes:

- **General** — Earth regression for any type of population or dataset. This is the default mode. It provides the full MARS modeling workflow without any domain-specific additions.
- **For Appraisal** — Earth regression tailored for real estate appraisal. Adds features specific to single-property valuation, including subject property handling, special column designations, Reconciliation by Comparable Adjustment (RCA), and Sales Comparison Grid generation.
- **Market Area Analysis** — Earth regression tailored for market area studies. Adds features for analyzing groups of properties in a defined market, with an optional “Skip first row” checkbox for subject property exclusion.

In all three modes, the core modeling engine is identical — you are always fitting an Earth (MARS) model. The purpose setting controls which additional tools and interface elements are available.

Switching purposes clears all state. When you change the purpose radio, earthUI resets to a clean default state: imported data, model results, tabs, variable configuration, and earth parameters are all cleared. You must re-import your file after switching. Previously saved settings for that file and purpose combination will then be automatically restored from your last session.

1.3 Real Estate–Specific Features

When either **For Appraisal** or **Market Area Analysis** is selected, earthUI activates several features designed for real estate analysis:

- **Special column designations** — Each predictor can be tagged with a special role such as `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions`, or `display_only`. These designations control how the column is handled during fitting, output, and Sales Grid generation. See Chapter 6 for the complete reference.
- **Rounding of latitude and longitude** — Columns designated as latitude or longitude are automatically rounded to 3 decimal places to prevent overfitting.
- **Sale Age column** — When a column is designated as `contract_date` and an Effective Date is provided, earthUI computes a `sale_age` column (days between sale date and effective date) and substitutes it as a predictor.
- **RCA computations (appraisal only)** — In appraisal mode, after fitting the model earthUI can compute Reconciliation by Comparable Adjustment (RCA) output, which produces per-comparable adjustments, net/gross adjustment summaries, and an adjusted sale price for the subject property.
- **Sales Comparison Grid (appraisal only)** — Generates a multi-sheet Excel workbook with formulas for comparable adjustments, residual feature breakdowns, and adjusted sale prices. See Chapter 11.

Tip: The General purpose mode works for any dataset — financial, scientific, engineering, or real estate. The appraisal and market modes simply add convenience features for real estate professionals.

1.4 Getting Started

To use earthUI:

1. **Install the package** in R: `install.packages("earthUI")` or install from source.
2. **Launch the application:** run `earthUI::launch()` in R, or from the command line run `Rscript -e 'earthUI::launch()'`. The app opens in your web browser on port 7878. You can also access the app directly by navigating to `http://localhost:7878` in your browser.
3. **Import your data** using the file upload in Section 1 of the sidebar. earthUI accepts CSV and Excel files.
4. **Select your Purpose** (General, For Appraisal, or Market Area Analysis).
5. **Configure variables** — choose your target(s) and predictors, set data types, and assign any special column roles.
6. **Set Earth parameters** — degree, nprune, subset filters, and other `earth()` arguments.
7. **Fit the model** — click “Fit Earth Model” and review the results in the main panel.
8. **Export** — download predictions as Excel, generate reports, or (in appraisal mode) compute RCA adjustments and Sales Comparison Grids.

Settings are automatically persisted in your browser’s local storage and restored when you reload the same input file.

2 MLS Input Data Requirements

For real estate appraisal and market analysis workflows, your input data typically comes from a Multiple Listing Service (MLS) export. This chapter describes the expected file structure and the columns that earthUI can use.

2.1 File Format & Structure

earthUI accepts **CSV** and **Excel** (.xlsx, .xls) files. On import, column names are automatically converted to **snake_case** — for example, “Living SqFt” becomes `living_sqft`, “Contract Date” becomes `contract_date`, and “Sale Price” becomes `sale_price`. This normalization ensures consistent column references throughout the workflow. The CSV separator and decimal mark used during import are determined by the locale settings (see Chapter 3, “Locale & Regional Settings”).

Your data file should be a flat table with one row per property and one column per attribute. The first row of the file must contain column headers.

2.2 Required Columns for Appraisal Mode

While earthUI works with any set of columns, the full appraisal workflow (RCA adjustments + Sales Comparison Grid) benefits from having the following columns in your MLS export:

Column	Special Type	Purpose
Sale Price	(target)	Response variable for the model
Contract Date	<code>contract_date</code>	Used to compute <code>sale_age</code> (days from effective date)
Listing Date	<code>listing_date</code>	Used with contract date to compute DOM if no DOM column exists
Days on Market	<code>dom</code>	Days on market; displayed in Sales Grid
Concessions	<code>concessions</code>	Sale concessions; Net SP = Sale Price – Concessions in Sales Grid
Living Area (SF)	<code>living_area</code>	Enables per-SF residuals (<code>residual_sf</code> , <code>cqa_sf</code>)
Lot Size	<code>lot_size</code>	Grouped in "Site Size Dimensions" row in Sales Grid
Site Dimensions	<code>site_dimensions</code>	Grouped with lot size in Sales Grid
Latitude	<code>latitude</code>	Rounded to 3 dp; used for proximity calculation and Location group
Longitude	<code>longitude</code>	Rounded to 3 dp; used for proximity calculation and Location group
Area ID	<code>area</code>	Grouped in "Loc: Long Lat Area" row in Sales Grid
Actual Age	<code>actual_age</code>	Grouped in "Actual Age Effective Age" row in Sales Grid
Effective Age	<code>effective_age</code>	Grouped with actual age in Sales Grid
Parcel Number	(none)	Displayed in APN row of Sales Grid
Listing ID	(none)	Displayed in APN row of Sales Grid
Address	<code>display_only</code>	Displayed in Address row of Sales Grid; excluded from model

Spreadsheet column names can be in a foreign language — the “special” names are in English so that the R program can give them special treatment. Otherwise, the given column names show up in the regression models, graphs, and (if doing appraisals) the Intermediate Sales Grid.

Not all columns are required. earthUI adapts — if a column is missing, the corresponding feature is simply omitted. For example, if no `concessions` column is designated, the Net SP

row in the Sales Grid shows Sale Price without a concessions deduction. However, for real estate pricing models certain columns are highly recommended to achieve acceptable fit:

1. **Sale Age** — the number of days between the contract sale date and the effective date of the appraisal or analysis. If multi-year sales history is being used, especially for periods over 5 years, `sale_age` often plays a central role in estimating the sale price. In fact it is often so important that without it, earthUI fails to provide any model at all.
2. **Living Area** — also goes by names such as “Living Sqft,” “GLA” (gross living area) and so on. This is another leading determinant of sale price.
3. **Total Bath Count** — the total number of full, quarter, half, and 3/4 bathrooms. For example, two full baths and one half-bath would be a value of 2.5.
4. **Garage Bays** or **Garage Area** — the number of garage spaces or the garage square footage.
5. **Lot Size** — the land area of the property, typically in square feet or acres.
6. **Longitude, Latitude,** and if available **Area ID**. Adjustments for these will be combined under a single Location adjustment in the Sales Grid.

2.3 Special Column Naming Conventions

earthUI identifies columns by their **special type designation**, not by their column name. You can name your columns anything you like in the MLS export — what matters is that you assign the correct special type in the Variable Configuration table (Chapter 6).

For example, your MLS might export living area as “GLA”, “Living SqFt”, “liv_area”, or “gross_living_area”. After import (where it becomes snake_case), you simply designate it as `living_area` in the Special dropdown. earthUI will then use it for per-SF residual calculations and Sales Grid grouping regardless of its original name.

2.4 Data Quality & Completeness

- **Missing values (NA):** Rows with NA values in any predictor or target column are automatically removed before fitting. The `na.action` is always set to `na.fail` internally, with NA removal handled before the call to `earth()`.
- **Date columns:** Dates should be in a format R can parse (e.g., 2025-06-15, 06/15/2025, June 15, 2025). earthUI auto-detects date columns when at least 80% of values parse successfully.
- **Numeric columns:** Sale price, living area, lot size, concessions, and similar fields must be numeric. If your MLS exports prices with currency symbols or thousands separators (e.g., “\$350,000” or “350.000,00”), you may need to clean these before import.
- **Factor columns:** Categorical variables like area ID, style, or condition should contain a manageable number of unique values. earthUI auto-detects factors but you can override the detection.

Tip: Review the NA column in the predictor table after import. Columns with more than 30% missing values are flagged in red. Consider excluding high-NA columns or cleaning the data before import.

2.5 Subject Row Placement

In **Appraisal** mode, **row 1 must be the subject property**. All remaining rows are comparable sales. The subject’s sale price can be left blank (NA) or set to any value — earthUI treats it as NA during fitting regardless.

In **Market Area Analysis** mode, placing the subject in row 1 is optional. If present, check “Skip first row (subject property)” to exclude it from fitting.

In **General** mode, there is no special row handling — all rows are treated equally.

3 General Purpose Mode

3.1 Overview

General Purpose mode is the default when you launch earthUI. It provides the complete MARS modeling workflow for any dataset — not just real estate. You can use earthUI for scientific data, financial analysis, engineering studies, or any regression problem where you want to explore non-linear relationships and interactions between variables.

In General mode, the interface omits the real estate-specific features (special columns, sale age, coordinate rounding, RCA). The sidebar is streamlined to focus on variable selection, parameter configuration, model fitting, and export.

3.2 The Sidebar Workflow

The sidebar is organized into numbered, collapsible sections that guide you from data import through export:

1. Import Data — File upload accepting CSV and Excel files. For Excel files with multiple sheets, a sheet selector appears. Column names are automatically converted to snake_case.

2. Project Output Folder — A text field specifying where downloads and fit logs are saved (defaults to ~/Downloads).

3. Variable Configuration — Target variable selector (supports multiple targets), predictor table with checkboxes for Include, Factor, and Linear. See Chapter 6 for full details.

4. Earth Call Parameters — All arguments to the `earth()` function: degree, penalty, nk, pruning method, cross-validation, and more. See Chapter 7 for the complete parameter reference.

5. Fit Earth Model — A single green button that runs the model asynchronously. See Chapter 8 for fitting details.

6. Download Output — Exports predictions, residuals, CQA scores, and per-g-function contributions as an Excel file. See Chapter 9.

7. Download Report — Generates a formatted report (HTML, Word, or PDF) saved to the output folder. See Chapter 12.

3.3 Main Panel Tabs

After fitting, the main panel provides nine tabs:

Tab	Contents
Data	Preview of imported data. In appraisal mode, split into Subject Property and Comparable Sales tables.
Equation	The fitted model equation displayed in LaTeX/MathJax, showing each basis function and its coefficient.
Summary	Key metrics (R^2 , $GRSq$, GCV , RSS , $CV R^2$) and a coefficients table.
Variable Importance	Bar chart and ranked table of predictor importance scores.
Contribution	g-function contribution plots. Select a g-function from the dropdown; univariate predictors show line plots, bivariate interactions ($\text{degree} \geq 2$) show 3D perspective and contour plots.
Correlation	Heatmap of predictor correlations.
Diagnostics	Three plots: Residuals vs Fitted, Normal Q-Q, and Actual vs Predicted.
ANOVA	ANOVA decomposition table showing each basis function's contribution to RSS.
Earth Output	Raw text output from <code>'earth::summary()'</code> , including the pruning pass details.

3.4 Settings Persistence

earthUI automatically saves your configuration to the browser's local storage, keyed by both the input filename and the current purpose mode. When you reload the same file under the same purpose, all settings are restored: target selection, predictor checkboxes, data types, earth parameters, and response weights. This means the same file can have different configurations for General, Appraisal, and Market modes. Settings are also backed up to an SQLite database so they persist across browser sessions. A **Reset to Defaults** button clears all saved settings for the current purpose.

3.5 Dark Mode

Click the moon/sun icon in the upper-right corner to toggle between light and dark themes. The theme preference is saved in local storage and persists across sessions.

3.6 Locale & Regional Settings

earthUI supports international number, date, and CSV formatting conventions through a country-based locale system. The **Country** dropdown in Section 1 of the sidebar (below the file upload) selects a preset for 31 supported countries. Each preset configures:

- **CSV separator** — comma (,) for US/UK/Japan or semicolon (;) for most of Europe, where the comma is used as a decimal mark.
- **Decimal mark** — period (.) or comma (,).
- **Thousands separator** — comma (US/UK/Japan), period (Germany/Italy/Spain), space (Finland/France/Poland/Baltics/Ukraine/Russia), or apostrophe (Switzerland).
- **Date format** — MM/DD/YYYY (US), DD/MM/YYYY (most of Europe), or YYYY-MM-DD (Sweden/Lithuania/Japan/Canada).
- **Paper size** — Letter (US/Canada/Mexico) or A4 (everywhere else).

3.6.1 Supported Countries

Country	CSV Sep	Decimal	Thousands	Date	Paper
US, Canada (csv)	,	.	,	MDY/YMD	Letter
UK, Ireland, AU, NZ	,	.	,	DMY	A4
Germany, Austria, Belgium, Netherlands, Italy, Spain, Portugal, Denmark, Brazil	;	,	.	DMY	A4
France, Poland, Czech Rep., Norway, Latvia, Estonia	;	,	(space)	DMY	A4
Finland, Sweden	;	,	(space)	DMY/YMD	A4
Switzerland	;	,	'	DMY	A4
Ukraine, Russia	;	,	(space)	DMY	A4
Turkey	;	,	.	DMY	A4
Japan, South Korea	,	.	,	YMD	A4
Mexico	,	.	,	DMY	Letter
Lithuania	;	,	(space)	YMD	A4

3.6.2 Override Dropdowns

Below the country selector, four override dropdowns let you change individual settings without switching countries:

- **Paper** — Letter or A4 (affects PDF report page size)
- **CSV sep** — comma or semicolon (used when importing CSV files)
- **Decimal** — period or comma (used in number display on plots and axes)
- **Date** — MDY, DMY, or YMD (controls the order in which date formats are tried when parsing date columns)

When you change the country, all overrides reset to that country's defaults. Changing an override only affects that one setting.

3.6.3 Saving Defaults

Click **Save as my default** to store your locale preferences globally. These defaults apply to all future sessions regardless of which data file you load. Per-file settings (target, predictors, parameters) are saved separately in the browser's local storage, but locale defaults persist across all files via an SQLite database. This two-level approach is designed for organizations like audit firms that work with data from multiple countries — set your most common country as the default, then override per-file when needed.

***Tip:** Number formatting on plot axes and slope labels automatically adapts to your locale. German locale uses periods for thousands (200.000), Finnish uses spaces (200 000), Swiss uses apostrophes (200'000). No currency symbols are displayed — earthUI is currency-agnostic.*

4 Appraisal Mode

When you select **For Appraisal** as the Purpose, earthUI configures itself for single-property valuation. All features described in Chapter 3 remain available; this chapter covers only the appraisal-specific additions.

4.1 Subject Row Handling

In appraisal mode, **row 1 of your dataset is the subject property** and all remaining rows are comparable sales. Your input file must be organized accordingly (see Chapter 2). The subject's sale price can be left blank or set to any value — earthUI automatically treats it as NA during fitting.

After importing, the Data tab splits into two sections: **Subject Property** (row 1) and **Comparable Sales** (rows 2+). Row 1 is always excluded from model fitting — the notification “Skipping row 1 (subject). Fitting on N rows.” confirms this. After fitting, the model still generates predictions for the subject row, shown as `est_<target>` in the output.

4.2 Effective Date & Sale Age

In appraisal and market modes, an **Effective Date** field appears in the Variable Configuration section (defaulting to today's date). If you designate a column as `contract_date` in the Special column dropdown, earthUI computes a `sale_age` column — the number of integer days between each sale's contract date and the effective date. This column replaces the original date column as a predictor.

The first time you click Fit after designating a contract date, earthUI creates the `sale_age` column and notifies you to click Fit again to include it.

4.3 Special Column Designations

In appraisal and market modes, a **Special** dropdown appears for each predictor in the Variable Configuration table. The complete list of special types and their effects:

Special Type	Effect
no	Default — no special handling
contract_date	Triggers automatic <code>sale_age</code> computation from the Effective Date
listing_date	Used with contract date to compute DOM when no <code>dom</code> column exists
dom	Identifies the Days on Market column; displayed in Sales Grid
concessions	Identifies sale concessions; used in Net SP formula (Sale Price – Concessions) in Sales Grid
latitude	Values rounded to 3 dp; used for Haversine proximity and Location group in Sales Grid
longitude	Values rounded to 3 dp; used for Haversine proximity and Location group in Sales Grid
area	Grouped in "Loc: Long Lat Area" row in Sales Grid
living_area	Enables per-SF residuals (<code>residual_sf</code> , <code>cqa_sf</code>) in download output
lot_size	Grouped in "Site Size Dimensions" row in Sales Grid
site_dimensions	Grouped with lot size in Sales Grid
actual_age	Grouped in "Actual Age Effective Age" row in Sales Grid
effective_age	Grouped with actual age in Sales Grid
display_only	Column included in exports but excluded from model fitting (multiple allowed)

Only one column per special type is allowed (except `display_only`). Assigning a special to a second column automatically clears it from the first. A small blue badge appears next to the variable name showing its assigned special type.

4.4 RCA Adjustments Overview

The **Calculate RCA Adjustments & Download** button (sidebar section 7, visible only in appraisal mode after fitting) computes market-derived adjustments for each comparable relative to the subject. The full RCA workflow is described in Chapter 10.

After computing RCA adjustments, the **Generate Sales Grid & Download** button (sidebar section 8) becomes available. The Sales Grid workflow is described in Chapter 11.

***Tip:** The CQA score you assign to the subject controls how much of the residual distribution is attributed to the subject. A score of 5.00 places the subject at the median of the comparables.*

5 Market Area Analysis Mode

When you select **Market Area Analysis** as the Purpose, earthUI provides the same real estate-specific features as appraisal mode (special columns, sale age, coordinate rounding) but is oriented toward analyzing a group of properties rather than valuing a single subject.

5.1 Differences from Appraisal Mode

- **Skip first row is optional** — a checkbox labeled “Skip first row (subject property)” appears below the Purpose selector. When checked, row 1 is excluded from fitting (like appraisal mode). When unchecked, all rows are included.
- **No RCA or Sales Grid sections** — the “Calculate RCA Adjustments & Download” and “Generate Sales Grid & Download” steps are not available. Market mode focuses on model fitting and output, not per-comparable adjustments.
- **Sidebar numbering** — without the RCA and Sales Grid steps, the Download Report section is numbered 7 instead of 9.

5.2 When to Use Market Mode

Market Area Analysis mode is appropriate when you are:

- Building a regression model for a neighborhood or market area to understand value drivers
- Analyzing how variables like square footage, age, lot size, and location affect sale prices across a group of properties
- Preparing support for a market conditions analysis or neighborhood delineation
- Working with a dataset that includes a subject property in row 1 but you want the option of including or excluding it from the fit

***Tip:** Market mode is also useful for general real estate regression where you want special column features (sale age, coordinate rounding) but do not need the RCA adjustment workflow.*

6 Variable Selection

Section 3 of the sidebar — **Variable Configuration** — is where you choose which columns participate in the model and how they are treated.

6.1 Target Variable(s)

The **Target (response) variable(s)** dropdown at the top of Section 3 lists every column in your dataset. Select one column for a standard single-response model, or multiple columns for a multi-response model. When multiple targets are selected, the model fits all responses simultaneously using `earth(cbind(y1, y2, ...) ~ .)`.

Columns selected as targets are automatically excluded from the predictor list.

6.2 The Predictor Table

Below the target selector, a table lists every remaining column with the following fields:

Column	Description
Variable	Column name (full name shown in tooltip if truncated). In appraisal/market modes, a blue badge shows the assigned special type.
Type	Data type dropdown: <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>logical</code> , <code>factor</code> , <code>Date</code> , <code>POSIXct</code>
Inc?	Checkbox — include this column as a predictor in the model
Special	Dropdown (appraisal/market only) — see Special Column Types Reference below
Factor	Checkbox — treat this column as a categorical variable
Linear	Checkbox — force linear entry only (no hinge functions)
NAs	Count of missing values. Shown in red when more than 30% of values are missing

A hint line above the table explains the abbreviations: “Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges).”

6.3 Data Type Detection & Overrides

earthUI automatically detects data types on import. Numeric, integer, logical, factor, and date columns are recognized. Character columns that look like dates (at least 80% of values parse against common date formats) are classified as **Date**.

You can override any detection by changing the **Type** dropdown. When you change a column to `character` or `factor`, the Factor checkbox is automatically checked. Changing types affects how the column is passed to the `earth()` function.

6.4 Factor and Linear Flags

- **Factor** — When checked, the column is treated as a categorical variable. This is appropriate for columns like `style`, `area_id`, or `grade` that represent discrete groups rather than continuous measurements. Factor columns enter the model as indicator (dummy) variables.

- **Linear** — When checked, the column is forced to enter the model linearly — no hinge (piecewise-linear) functions are created for it. This is useful when you know a variable has a strictly linear relationship with the target.

6.5 Special Column Types Reference

In appraisal and market modes, the **Special** dropdown provides the following options. Each type can be assigned to at most one column (except `display_only`, which allows multiple):

Date & Time Types:

- `contract_date` — Triggers automatic `sale_age` computation. The original date column is replaced by an integer column measuring days between the sale date and the Effective Date.
- `listing_date` — Used as a fallback for computing Days on Market (DOM = contract date – listing date) when no explicit `dom` column is designated.
- `dom` — Identifies the Days on Market column. Displayed in the Sales Grid’s APN row and Date of Sale row.

Monetary Types:

- `concessions` — Identifies sale concessions (seller credits, buyer incentives, etc.). Used in the Sales Grid to compute Net Sale Price: $\text{Net SP} = \text{Sale Price} - \text{Concessions}$.

Size & Location Types:

- `latitude` — Values are automatically rounded to 3 decimal places to prevent overfitting. Used for Haversine proximity calculations (distance from subject to each comp) and grouped in the Location row of the Sales Grid.
- `longitude` — Same rounding treatment as latitude. Used for proximity and the Location group.
- `area` — Typically a neighborhood or area identifier. Grouped with latitude and longitude in the “Loc: Long | Lat | Area” row of the Sales Grid.
- `living_area` — Enables per-square-foot residual calculations (`residual_sf` and `cqa_sf`) in the download output.
- `lot_size` — Grouped in the “Site Size | Dimensions” row of the Sales Grid.
- `site_dimensions` — Grouped with lot size in the Sales Grid (e.g., “75x120”).

Age Types:

- `actual_age` — Grouped in the “Actual Age | Effective Age” row of the Sales Grid.
- `effective_age` — Grouped with actual age in the Sales Grid.

Display Types:

- `display_only` — The column is included in Excel exports but excluded from model fitting entirely. Use this for address fields, MLS numbers, parcel IDs, or other reference data that should not be a predictor. Multiple columns can have this designation.

6.6 Multiple Targets

Selecting more than one target variable fits a multi-response Earth model. When multiple targets are selected:

- The model predicts all responses simultaneously, sharing basis functions across targets
- The **wp (response weights)** button becomes active, allowing you to assign a numeric weight to each target
- Variance models (`varmod.method`) are disabled (not supported for multi-response)
- Results tabs display per-response metrics, equations, and diagnostic plots

Tip: Columns designated as *display_only* remain in your dataset and appear in the Excel export, but are excluded from the predictor table entirely. Use this for ID columns, addresses, or other reference fields.

7 Parameter Selection

Section 4 of the sidebar — **Earth Call Parameters** — provides access to all arguments accepted by the `earth()` function. Each parameter has a blue help icon (?) with a tooltip explanation. Parameters are organized into collapsible subsections.

7.1 Basic Parameters

Parameter	Default	Description
subset	(empty)	Row filter expression. See “Subset Filtering” below.
weights	NULL	Column selector for case (row) weights. Only numeric columns are listed.
wp	NULL	Response weights for multi-target models. Button opens a dialog with one numeric input per target (default 1.0 each).
keepxy	off	Retain x, y, subset, and weights in the model object.
trace	0	Trace level for fitting output (0 through 5).
glm	none	Optional GLM family: <code>gaussian</code> , <code>binomial</code> , or <code>poisson</code> .
degree	1	Maximum interaction order. Setting $\text{degree} \geq 2$ auto-enables cross-validation and reveals the Allowed Interactions matrix.
penalty	2	GCV penalty per knot. Higher values produce simpler models.

7.2 Forward Pass

Parameter	Default	Description
nk	auto	Maximum terms before pruning. Default: $\min(200, \max(20, 2 \times \text{predictors})) + 1$.
thresh	0.001	Forward-step threshold. Smaller values allow more terms.
minspan	0 (auto)	Minimum span between knots. Negative values set the maximum knots per predictor.
endspan	0 (auto)	End span — minimum distance from a knot to the edge of the data.
newvar.penalty	0	Penalty for introducing a new variable (encourages reuse of existing predictors).

Parameter	Default	Description
fast.k	20	Number of parent terms to consider in the fast MARS algorithm.
fast.beta	1	Controls the fast MARS aging factor.

7.3 Allowed Interactions

When **degree** is set to 2 or higher, an interaction matrix appears below the basic parameters. This is an upper-triangular grid of checkboxes, one for each predictor pair. A checked box means the two predictors are allowed to interact; unchecking it forbids that specific interaction.

Clicking a predictor name (row or column label) toggles all interactions for that variable. **Allow All** and **Clear All** checkboxes at the top provide bulk control. The matrix uses sticky headers so column and row labels remain visible when scrolling.

An info alert reminds you: “Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).”

7.4 Pruning

Parameter	Default	Description
pmethod	backward	Pruning method: backward , none , exhaustive , forward , seqrep , or cv .
nprune	NULL	Maximum terms after pruning. Leave empty for automatic selection.

7.5 Cross-Validation

Parameter	Default	Description
nfold	10	Number of cross-validation folds. Set to 0 to disable CV.
ncross	20	Number of cross-validation repetitions.
stratify	on	Stratify CV samples so each fold has a similar response distribution.

When $\text{degree} \geq 2$, cross-validation is automatically enabled (nfold set to 10) to help guard against overfitting from interactions.

7.6 Subset Filtering

The **subset** text input accepts an R expression that filters which rows are used for model fitting. You can type an expression directly (e.g., `sale_age < 365 & area_id == 460`) or use the **Build filter...** button to construct one visually.

Build Filter Dialog — Click “Build filter...” to open a guided dialog. Each condition row has a column dropdown, operator (<, >, <=, >=, ==, !=), and a value input that adapts to the column type: numeric input for numbers, date picker for dates, and dropdown of unique values

for character/factor columns. Conditions are joined with AND (&) or OR (|) connectors. A preview at the bottom shows the expression and how many rows match. Click **Apply** to insert the expression into the text input.

Date columns must use `as.Date("...")` or `as.POSIXct("...")` wrappers in manual expressions. The Build Filter dialog handles this automatically.

Subset filtering is non-destructive — excluded rows remain in the dataset and receive predictions in the Excel export.

***Tip:** Use parentheses to group OR conditions when combining with AND. Without parentheses, $A \ \& \ B \ | \ C$ is evaluated as $(A \ \& \ B) \ | \ C$, which may not be what you intend.*

7.7 Recommended Parameter Values

earthUI displays a recommended value below key parameters in Section 4. These recommendations update reactively based on the number of fitting rows (n) and selected predictors (p). The formulas are derived from Friedman’s MARS paper, earth’s internal algorithms, and empirical testing.

7.7.1 Forward Pass Parameters

nk (max terms before pruning):

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

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

Earth’s default, $\min(200, \max(20, 2p)) + 1$, does not account for dataset size and can be too small for large datasets, constraining the forward pass before it explores all predictors adequately. The $\lfloor n/10 \rfloor$ term allows approximately one term per 10 observations — a standard rule of thumb for avoiding overparameterization. The cap of 100 prevents diminishing returns.

minspan (minimum observations between knots):

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

endspan (minimum observations from data boundaries):

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

n	minspan	endspan
30	5	6
50	6	6
100	7	8

n	minspan	endspan
200	9	12
300	11	15
500	15	16
1500	16	16

Earth's auto-calculated minspan uses Friedman's equation 43: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, which scales as $\ln(np)$ and grows too slowly for large datasets. At $n=1500$ it yields only ≈ 7 , giving roughly 245 candidate knot locations per continuous predictor — too many, allowing the forward pass to fit noise. The recommended formula targets approximately 100 candidates per predictor for large n , while deferring to earth's auto-calculation for small n (where it is well-tested). Endspan is set slightly larger than minspan to provide additional boundary protection, which helps when data has thin tails (common in real estate).

penalty (GCV penalty per knot):

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

This is earth's own default — 2 for additive models, 3 for interaction models (the higher penalty compensates for the larger search space).

newvar.penalty (penalty for introducing a new predictor):

Recommended: **0.1** when predictors are correlated (e.g., living area with bedroom count, bathroom count, lot size). This biases the forward pass toward reusing predictors already in the model rather than introducing correlated alternatives. The mechanism: each new predictor's RSS improvement is multiplied by $1/(1 + \text{penalty})$. With $\text{newvar.penalty} = 0.1$, a new variable must improve RSS by at least 10% more than another knot on an existing variable. This produces simpler models with fewer predictors without affecting final coefficient estimates (the penalty is removed after selection). Leave at 0 if predictors are not correlated.

7.7.2 Pruning Parameters

pmethod: Recommended: **backward**. The default GCV-based backward pruning is deterministic — the same data always produces the same model, with no dependence on random seeds. This is important for reproducibility, especially in appraisal work. The **cv** method uses cross-validation for pruning which introduces seed dependence.

nprune: Recommended: **leave empty** (NULL). Let GCV select the optimal number of terms. Setting **nprune** imposes a hard cap that overrides GCV's judgment.

7.7.3 Cross-Validation Parameters

nfold (CV folds):

$$\text{nfold} = \min(15, \max(10, \lfloor n/100 \rfloor))$$

n	nfold
30	10
100	10
500	10
1000	10
1500	15

With `pmethod = "backward"`, cross-validation does not affect the model — it only computes the diagnostic CVR^2 and provides residuals for the variance model. The floor of 10 and cap of 15 reflect this: enough folds for a stable diagnostic without unnecessary computation.

`ncross` (CV repetitions):

$$ncross = \max(3, \lceil 100/n \rceil)$$

n	<code>ncross</code>	total residuals
30	4	120
50	3	150
100	3	300
1500	3	4500

The variance model (`varmod.method = "lm"`) fits on the cross-validation residuals. The formula targets at least 100 total residuals ($n \times ncross$) for a stable variance estimate, with a floor of 3 (required by earth when a variance model is enabled).

7.7.4 Variance Model

`varmod.method`: Recommended: **lm**. This fits prediction intervals using a linear regression of absolute CV residuals on the predicted response. It requires `nfold > 0` and `ncross >= 3`.

7.7.5 Earth Output File

After every successful fit, earthUI writes `<filename>_earth_output_<timestamp>.txt` to the output folder. This file contains the model terms, summary statistics (R^2 , $GRSq$, $CVRSq$), variance model details, and trace log. One file is created per fit, providing a cumulative record for comparing parameter configurations.

7.8 Settings Defaults

A radio button at the top of Section 4 controls which defaults are loaded:

- **Use last settings for input file** — restores the settings you used last time with this particular file
- **Use default settings** — applies your saved custom defaults
- **Earth defaults** — resets all parameters to factory values

Click **Save current as default** to store the current parameter configuration as your personal defaults.

8 Fitting the Earth Model

8.1 The Fit Button

Section 5 of the sidebar contains a single green button: **Fit Earth Model**. This button is always visible (not inside a collapsible section). Before fitting, earthUI validates your configuration — a target must be selected and at least one predictor must be included. In appraisal/market modes, latitude and longitude columns are rounded, and sale age is computed from the effective date and contract date (if designated).

8.2 Fitting Modal & Trace Output

When you click Fit, a dark modal overlay appears with:

- **Header** — “Fitting Earth Model” with an elapsed timer counting seconds
- **Trace log** — a scrollable, monospace text area showing real-time output from the `earth()` function: dataset info, forward pass progress, cross-validation folds, and completion time
- **Color coding** — default lines in green, cross-validation lines in yellow, errors in red

earthUI fits models asynchronously using `callr::r_bg()`, which runs the computation in a separate R process. This keeps the application responsive during long-running fits. A 300ms polling observer reads output from the background process and streams it to the trace log.

When fitting completes, a “Done in X.Xs” message appears and a close button (X) is added to the modal. **The modal does not auto-dismiss** — you close it when you are ready to review results.

On success, a green checkmark is appended to the Fit button. If an error occurs, the error message is displayed in the trace log and a notification appears.

8.3 Fit Log

Every fit (success or failure) writes a log file to your output folder:

- **Filename:** `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Contents:** timestamp, all trace output from the fitting process, and any error messages
- **Location:** the Project Output Folder specified in Section 2 (defaults to `~/Downloads`)

***Tip:** If `callr` is not installed, earthUI falls back to synchronous fitting with a simple progress bar. Install `callr` for the full trace-output experience.*

9 Downloading Data

After fitting, download an Excel file with predictions and diagnostics. This output is used in Step 7 (RCA) to assign a CQA (Condition/Quality/Appeal) rating to the subject property. The output is sorted by `residual_sf` and `cqa_sf` to help you assess where the subject falls in the ranking. If the model is good quality, then the properties should be ranked from lowest appealing to most appealing based on residual features that did not go into the regression. The middle value should be approximately 0, the lower half negative values and the upper half positive values. You should find the worst quality homes, or “fixers” near the bottom of the ranking and the nicest homes at the top. There will usually be exceptions for anomalies such as foreclosures, short sales, probate (inheritance related) sales, and quick sales needed for job change or other reasons. Investigation of anomalies usually turns up a pertinent reason for the price anomaly.

The button label adapts to the purpose mode:

- General/Market: **Download Output (Excel)**
- Appraisal: **Download Intermediate Output (Excel)**

The filename format is `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`.

9.1 Output Columns

For each target variable, the following columns are appended:

Column	Description
<code>est_<target></code>	Model prediction, e.g. <code>est_sale_price</code> (1 decimal place)
<code>residual</code>	Actual minus predicted (1 dp)
<code>cqa</code>	Comparative Quality Analysis score (2 dp, range 0–10)
<code>residual_sf</code>	Residual divided by living area (if designated, 1 dp)
<code>cqa_sf</code>	CQA calculated from ranking via <code>residual_sf</code> (2 dp)
<code><variable>_contribution</code>	Per-g-function contribution value (1 dp, one column per g-function)
<code>basis</code>	Intercept value contribution, same for all properties (1 dp)
<code>calc_residual</code>	Actual minus (basis + all contributions) — verification column (1 dp)

For multi-target models, a `<i>` suffix is added to distinguish columns for each target.

9.2 Column Ordering & Excel Formatting

The ranking columns are placed at the **leftmost position** in the output file, in this order: `residual_sf`, `cqa_sf`, `residual`, `cqa`. This makes it easy to scan the sorted list and evaluate where the subject falls in the CQA distribution.

These columns have Excel number formatting applied:

Column	Format	Example
<code>residual_sf</code>	Numeric, 2 decimal places	12.34
<code>cqa_sf</code>	Numeric, 2 decimal places	7.25
<code>residual</code>	Numeric, 0 decimal places	15,200
<code>cqa</code>	Numeric, 2 decimal places	6.80

9.3 CQA Scores

The CQA (Comparative Quality Analysis) score ranks each row's residual against all other residuals. For a given row, the CQA is the percentage of rows with a smaller signed residual, multiplied by 10. This produces a 0–10 scale where:

- **High CQA** (≈ 9 – 10) — the property sold for much more than the model predicted (large positive residual)
- **Low CQA** (≈ 0 – 1) — the property sold for much less than predicted (large negative residual)
- **CQA ≈ 5** — the residual is near the median of all residuals

When a `living_area` column is designated, `cqa_sf` provides the same ranking based on per-square-foot residuals.

9.4 Sorting & Subject Row

In appraisal and market modes, comparable rows are sorted by `residual_sf` descending (or `residual` if no living area is designated). The subject row (row 1) remains in position 1 and is not sorted. In general mode, rows are exported in their original order.

This sorting, combined with the leftmost ranking columns, allows the appraiser to quickly scan the comparables from most over-predicted to most under-predicted, and assess where the subject's assigned CQA score falls in that distribution.

Factor levels in the prediction data are aligned with the training data before calling `predict()`. Rows with unseen factor levels will produce NA predictions.

Tip: The `calc_residual` column should always equal `residual` (within rounding). If they differ significantly, it indicates a computation issue worth investigating.

9.5 mgcvUI Auto-Export

On every successful fit with `degree` ≤ 2 , `earthUI` automatically saves the full result object as an `.rds` file to the Project Output Folder. The filename follows the pattern `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

This file can be loaded by `mgcvUI` (a companion Shiny app for GAM modeling) using `readRDS()`. `mgcvUI` uses the `earth` model's knot locations and basis functions as starting points for GAM smooth terms, enabling a seamless transition from MARS to GAM modeling.

Models with `degree` > 2 are skipped because `mgcvUI` only supports pairwise interactions. A manual **Export for mgcvUI** button is also available in the sidebar for on-demand export.

9.6 glmnetUI Import

The same `.rds` file saved for `mgcvUI` can also be imported into `glmnetUI` (a companion Shiny app for elastic net regression). In `glmnetUI`, navigate to **Section 2: Import from earthUI** and use the Browse button to select the `.rds` file.

`glmnetUI` uses the standard `earth::model.matrix()` approach: the `earth` model's basis functions (hinges, interactions, and linear terms) become the columns of `glmnet`'s design matrix. `glmnet` then performs regularized regression on these basis columns, selecting and shrinking them via lasso/elastic net. This combines `earth`'s adaptive basis construction with `glmnet`'s regularization.

9.6.1 Key Benefits

- **Interpretable interactions:** earth's hinge-based interactions (including 3-way terms) have clear geometric meaning, unlike glmnet's native cross-product interactions which are difficult to explain in court or audit settings.
- **Automatic nonlinearity:** earth's hinge functions capture nonlinear relationships that a standard glmnet model with raw predictors would miss.
- **Variable selection:** glmnet can further prune the earth basis, dropping hinge terms that don't contribute.

9.6.2 Important: Weight Column Must Match

Warning

If earthUI uses a weight column (e.g., to exclude the subject row with weight=0), the same weight column must be designated in glmnetUI's variable configuration. If the weight settings differ, the row counts will not align and the earth basis cannot be appended to glmnet's model matrix.

9.6.3 Workflow

1. Fit an earth model in earthUI (any degree).
2. The `.rds` result file is saved automatically to the Project Output Folder.
3. In glmnetUI, open **Section 2: Import from earthUI** and browse to the `.rds` file.
4. Verify that the **same data file** and **same weight column** are used in both apps.
5. Configure glmnet parameters in Section 5 (the Interaction Matrix is automatically disabled when an earthUI import is active).
6. Click **Fit Model**. The earth basis columns appear in the Equation, Coefficients, and Contributions tabs.

10 RCA Calculations & Downloading

The **Calculate RCA Adjustments & Download** button appears in sidebar section 7, visible only in appraisal mode after a model has been fitted. RCA (Reconciliation by Comparable Adjustment) produces market-derived, per-comparable adjustments relative to the subject property.

10.1 Opening the RCA Dialog

Clicking the button opens a small modal dialog with:

- **Score type** — radio buttons to choose **CQA** or **CQA per SF** (the SF option is available only when a `living_area` column is designated). If you choose “CQA per SF,” then based on the CQA score you assign the subject, its residual score will be its living area times the `residual_sf` that matches the given `CQA_SF` score.
- **CQA value** — numeric input for the subject’s CQA score (0.00–9.99, default 5.00). This score represents where you believe the subject falls in the quality distribution of the comparables.
- **Generate** button — computes the RCA output and downloads it as Excel

10.2 CQA Score Interpolation

When you click Generate, earthUI interpolates the subject’s residual from the comparable CQA/residual pairs:

1. The comparables’ CQA scores and residuals are sorted by CQA ascending
2. Linear interpolation (`stats::approx()`) maps your entered CQA value to a residual
3. If using CQA per SF, the per-SF residual is converted back to a total residual by multiplying by the subject’s living area
4. The subject’s estimated value is computed as: model prediction + interpolated residual

10.3 Output Columns

The RCA Excel file (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) includes all intermediate output columns plus:

Column	Description
<code>subject_value</code>	Model prediction + interpolated residual (row 1 only)
<code>subject_cqa</code>	The CQA score you entered (row 1 only)
<code><variable>_adjustment</code>	Subject contribution minus comp contribution (per g-function)
<code>residual_adjustment</code>	Subject residual minus comp residual
<code>net_adjustments</code>	Sum of all adjustments (contribution + residual)
<code>gross_adjustments</code>	Sum of absolute values of all adjustments
<code>adjusted_sale_price</code>	Comp sale price + net adjustments

The adjustment columns tell you, for each comparable, how much the model attributes the difference to each variable. `adjusted_sale_price` is the comparable’s sale price after applying all model-derived adjustments — a set of adjusted sale prices that should cluster around the subject’s estimated value.

Tip: A CQA score of 5.00 places the subject at the median of the comparables. Scores above 5 indicate the subject is above average quality; below 5 indicates below average.

11 Sales Comparison Grid

11.1 Overview & Purpose

The **Sales Comparison Grid** is an Excel workbook generated in appraisal mode (sidebar section 8) after computing RCA adjustments. It presents the subject property alongside selected comparables in a structured grid format, with Excel formulas that automatically compute adjustments and an adjusted sale price for each comparable.

The grid is designed for the appraiser’s workfile — it combines the regression-derived adjustments from the Earth model with editable cells where the appraiser can allocate the CQA residual to specific property features. The output filename is `SalesGrid_<YYYYMMDD_HHMMSS>.xlsx`.

11.2 Comp Selection Modal

Clicking “Generate Sales Grid & Download” opens a modal dialog where you select which comparables to include:

- **Recommended comps** are pre-checked. These are comparables with a gross adjustment percentage below 25% of sale price, sorted by gross adjustment percentage ascending (smallest adjustments first).
- **Additional comps** are listed below, unchecked by default. These have larger gross adjustments but are still available for selection.
- A maximum of **30 comps** can be selected, producing up to **10 sheets** (3 comps per sheet).

After confirming your selection, the selected comps are sorted by gross adjustment percentage ascending, so the most similar comparables appear on Sheet 1.

11.3 Grid Layout

Each sheet has a 20-column layout accommodating the subject property plus 3 comparables. The columns for each entity are:

Entity	Columns (5 per entity)
Subject	Label, Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC)
Each Comp	Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC), Adjustment

The rows from top to bottom are:

1. **Title** — sheet name (e.g., “Intermediate Sales Comparable Grid — Sheet 1 of 3”)
2. **Headers** — “Subject” and comp column headers with address
3. **Address** — full property address
4. **APN | MLS# | DOM | Subj.Prox** — parcel number, listing ID, days on market, and Haversine distance (miles) from subject
5. **Sales Price | Concess. | Net SP** — sale price, concessions, and Net Sale Price formula
6. **Regression Features** header row
7. **BASE VALUE** — the model intercept
8. **Date of Sale | OffMkt | OnMkt** — contract date, sale age, and DOM
9. **Grouped rows** (conditional) — Location, Site Size, and/or Age groups
10. **Model variable rows** — one row per predictor (excluding grouped variables)
11. **Blank separator row**

12. **Residual Features** header row
13. **CQA|Residual** — CQA score + remaining residual formula
14. **Residual feature rows** — named + blank rows for appraiser entry
15. **Total VC / Net Adjustment**
16. **Net Adjustment %**
17. **Gross Adjustment %**
18. **Adjusted Sale Price** — formula row
19. **Copyright** footer

11.4 Sale Price, Concessions & Net SP

Row 5 shows three values for each comparable:

- **Sale Price** — the comparable’s sale price from the data
- **Concessions** — the value from the column designated as **concessions** (0 if not designated)
- **Net SP** — an Excel formula: Sale Price – Concessions

The subject column shows “N/A” for sale price (since it is unknown) and any concessions value if available.

11.5 Grouped Rows (Location, Site, Age)

When certain special types are designated and those variables appear in the model, earthUI creates grouped rows that combine related variables:

Loc: Long | Lat | Area — Appears when any of **longitude**, **latitude**, or **area** are in the model. Shows factual values for each constituent variable, a combined Value Contribution (sum of the individual VCs), and for comps, an adjustment (subject combined VC – comp combined VC). Styled with light blue background on VC cells.

Site Size | Dimensions — Appears when **lot_size** or **site_dimensions** are in the model. Same structure as the Location group.

Actual Age | Effective Age — Appears when **actual_age** or **effective_age** are in the model. Same structure.

Variables consumed by grouped rows are **excluded from the individual model variable rows** below, preventing double-counting in the adjustment totals.

11.6 Model Variable Rows

Below the grouped rows (or directly below BASE VALUE if no grouped rows), one row per remaining model predictor shows:

- **Factual values** for subject and each comp (the actual data values)
- **Value Contribution (VC)** — the g-function’s contribution to the predicted value for that row
- **Adjustment** (comps only) — subject VC minus comp VC

11.7 CQA|Residual & Residual Feature Rows

The **CQA|Residual** row shows each property’s CQA score and contains a formula for the **remaining residual** — the portion of the residual not yet allocated to specific features. The formula is:

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

Below this are **residual feature rows** — 5 named rows (Location, View/Appeal, Condition, Quality, Other) plus 6 blank rows. These are input cells where the appraiser enters value contributions for features not captured by the model. As the appraiser fills in values, the Remaining Residual formula automatically decreases.

For each comp, the adjustment column contains a formula: subject feature VC – comp feature VC.

11.8 Adjusted Sale Price Formula

The **Adjusted Sale Price** row contains Excel formulas:

- **Subject:** Sum of all Value Contribution cells from BASE VALUE through the last residual feature row. This represents the model's total prediction for the subject plus any appraiser-allocated residual features.
- **Comps:** Net SP + sum of all Adjustment cells from the first grouped or variable row through the last residual feature row. This gives the comparable's sale price after all regression-derived and appraiser-entered adjustments.

11.9 Sheet Protection & Editable Cells

Each sheet is **protected** to prevent accidental modification of formulas and data. The protection locks:

- All formula cells (Net SP, remaining residual, adjustments, adjusted sale price, etc.)
- All data cells (addresses, sale prices, factual values, value contributions from the model)
- Labels and headers

The only **unlocked** (editable) cells are the residual feature Value Contribution inputs — the cells under the CQA|Residual row where the appraiser enters breakdowns. These are styled with a light yellow background to indicate they are editable.

11.10 Working with the Grid in Excel

1. **Open the file** in Excel or a compatible spreadsheet application.
2. **Review the regression adjustments** — the model-derived adjustments are pre-populated and locked.
3. **Allocate the residual** — in the residual feature rows (yellow cells), enter your assessment of how much of the remaining residual is attributable to each feature (Location, View, Condition, Quality, etc.). The Remaining Residual formula will decrease as you allocate.
4. **Check the Adjusted Sale Price** — this formula automatically updates as you enter residual feature values.
5. **Multiple sheets** — if you selected more than 3 comps, navigate between sheets. Each sheet has the same subject in the left column with 3 different comps.

***Tip:** The goal is to allocate the CQA residual across the feature rows until the Remaining Residual is close to zero. This demonstrates that you can account for the difference between each comparable's regression estimate and its actual sale price.*

12 Downloading Reports

After fitting a model, the Download Report section (sidebar section 7, 8, or 9 depending on the purpose mode) lets you generate a comprehensive formatted report.

12.1 Report Formats

Three formats are available via the format dropdown:

- **HTML** — an `.html` file with KaTeX math rendering, embedded images, and a table of contents. Uses the Flatly Bootstrap theme and Roboto Condensed font.
- **Word** — a `.docx` file suitable for editing and distribution. Includes a table of contents, page numbers, and uses a custom reference template for consistent styling.
- **PDF** — typeset with LuaLaTeX for professional-quality output. Paper size follows the locale setting (Letter or A4). Includes landscape pages for large interaction matrices and uses Roboto Condensed with Latin Modern Math for equations.

Reports are rendered via Quarto and saved directly to the Project Output Folder as `<datafile>_report_<YYYYMMDD_HHMMSS>.<format>`. Rendering runs in the background — a modal dialog shows elapsed time and Quarto progress while the app remains responsive. A notification confirms the file location on completion.

All operations (model fitting, data downloads, RCA calculations, sales grid generation, and report rendering) are logged with start/end timestamps and elapsed times to `<datafile>_earthui_log.txt` in the output folder, for troubleshooting and performance monitoring.

12.2 Report Contents

Every report includes the following sections:

1. **Dataset Description** — number of observations, target variable(s), predictors used, categorical predictors
2. **Model Specification** — degree, cross-validation status, number of terms, predictors in model
3. **Allowed Interactions** (degree ≥ 2 only) — the interaction matrix with checkmarks, formatted for the output type
4. **Results Summary** — R^2 , GRSq, GCV, RSS, and CV R^2 (if cross-validation was enabled)
5. **Model Equation** — the complete Earth model equation in LaTeX notation
6. **Coefficients & Basis Functions** — table of all terms, their coefficients, and the basis functions that define them
7. **Variable Importance** — bar chart and ranked table of predictor importance scores
8. **g-Function Contributions** — plots for each g-function: line plots for univariate terms, 3D perspective and contour plots for bivariate interactions
9. **Correlation Matrix** — heatmap of predictor correlations
10. **Diagnostics** — Residuals vs Fitted, Normal Q-Q, and Actual vs Predicted plots
11. **ANOVA Decomposition** — table showing each basis function's contribution to RSS
12. **Earth Output** — raw text output from `earth::summary()`, including pruning pass details

For multi-response models, sections 4–8 and 10 are repeated for each target variable.

Tip: Reports require the `quarto` R package. For PDF output, a LaTeX distribution (e.g., TinyTeX) must also be installed.

13 Demo Dataset: Appraisal_1.csv

earthUI includes a demo MLS dataset for exploring the appraisal workflow. Load it programmatically with:

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

Or import it directly through the Shiny app file upload.

13.1 Description

The file contains 1,502 residential sales (plus 1 subject property in row 1) from a simulated MLS export. The data represents single-family home sales in a multi-area market with a range of property sizes, ages, and locations.

This is not real data, but is based on a realistic neighborhood in Northern California. All identification information has been altered or removed.

13.2 Columns

Column	Type	Special Type	Description
weight	numeric	—	Observation weight (0 = exclude from fitting)
id	numeric	display_only	Internal record ID
property_id	numeric	display_only	MLS property identifier
listing_id	character	display_only	MLS listing number
parcel_number	character	display_only	County assessor parcel number (APN)
street_address	character	display_only	Property address
city_name	character	display_only	City
postal_code	character	display_only	ZIP code
county_name	character	display_only	County
contract_date	Date	contract_date	Sale contract date (computes sale_age)
sale_age	numeric	—	Days from contract date to effective date (pre-computed)
coe_date	Date	display_only	Close of escrow date
listing_status	character	display_only	Listing status (e.g., "Sold")
sale_price	numeric	(target)	Sale price — response variable
rent	numeric	—	Monthly rent (for multi-target models)
list_price	numeric	display_only	Listing price
original_list_price	numeric	display_only	Original listing price
living_sqft	numeric	living_area	Gross living area in square feet
beds_total	integer	—	Number of bedrooms
baths_total	numeric	—	Total bath count (e.g., 2.5 = 2 full + 1 half)
lot_size	numeric	lot_size	Lot size in square feet
area_id	integer	area	MLS area identifier
area_text	character	display_only	Area name
age	numeric	actual_age	Property age in years
year_built	integer	display_only	Year of construction
latitude	numeric	latitude	Latitude (rounded to 3 dp for model)
longitude	numeric	longitude	Longitude (rounded to 3 dp for model)
latitude6	numeric	display_only	Full-precision latitude
longitude6	numeric	display_only	Full-precision longitude
garage_spaces	integer	—	Number of garage bays
fp_count	integer	—	Number of fireplaces
no_of_stories	numeric	—	Number of stories
style	character	—	Architectural style
view	character	—	View type (e.g., "Neighborhood", "Hills")
days_on_market	integer	dom	Days on market
listing_date	Date	listing_date	Listing date
sale_concessions	numeric	concessions	Seller concessions

13.3 Suggested Quick Start

1. Launch earthUI: `earthUI::launch()`
2. Import `Appraisal_1.csv` via the file upload
3. Set Purpose to **For Appraisal**
4. Select `sale_price` as the target
5. Assign special types as shown in the table above
6. Include predictors: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (as factor), `age`, `latitude`, `longitude`, `garage_spaces`
7. Set degree to 1, click **Fit Earth Model**
8. Download intermediate output (Step 6), review the CQA ranking
9. Compute RCA adjustments (Step 7) with a CQA score of ~ 5.00
10. Generate the Sales Comparison Grid (Step 8)

Tip: The demo dataset is a good way to explore all of earthUI's features before importing your own MLS data. The suggested predictors above typically produce a model with R^2 above 0.85.

References

- Friedman, Jerome H. (1991). "Multivariate Adaptive Regression Splines". In: *The Annals of Statistics* 19.1, pp. 1–141. DOI: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Milborrow, Stephen (Oct. 1, 2024). *earth: Multivariate Adaptive Regression Splines*. R package. URL: <https://CRAN.R-project.org/package=earth>.
- (n.d.[a]). *Notes on the earth Package*. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (n.d.[b]). *Variance Models in earth*. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.