

earthUI: una interfaz interactiva para el paquete earth (MARS)

Traducción al español, elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

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

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

Este artículo de acceso abierto se distribuye bajo los términos de la [licencia Creative Commons Atribución 4.0](#).

Resumen. earthUI es una interfaz gráfica de usuario para el paquete de R **earth**, que implementa Multivariate Adaptive Regression Splines (MARS). Ofrece tres modos de uso — modelado predictivo general, valuación inmobiliaria y análisis de zona de mercado — y guía al usuario a través de la importación de datos, la configuración del modelo, el ajuste y la interpretación de los diagnósticos y salidas gráficas de earth. Este artículo documenta los requisitos de formato de datos de earthUI, el flujo de trabajo de modelado, las pantallas de salida y la referencia completa de funciones.

Palabras clave: MARS, multivariate adaptive regression splines, earth, R, interfaz gráfica de usuario, valuación inmobiliaria

Nota del editor (julio de 2026): Este artículo describe earthUI 0.10.0. Poco antes de la publicación, earthUI 0.11.0 absorbió las aplicaciones complementarias glmnetUI y mgcvUI (descritas en los dos artículos siguientes), consolidando los flujos de trabajo de MARS, elastic-net y GAM en una sola aplicación. La interfaz y el flujo de trabajo descritos aquí no se ven afectados por la consolidación.

Índice

1. Introducción	4
1.1. ¿Qué es earthUI?	4
1.2. Un paquete, tres rutinas	4
1.3. Tres modos de propósito	4
1.4. Funciones específicas para bienes raíces	5
1.5. Instalación de earthUI y vProlog	5
1.6. Primeros pasos	6
2. Requisitos de los datos de entrada del MLS	7
2.1. Formato y estructura del archivo	7
2.2. Columnas requeridas para el modo Appraisal	7
2.3. Convenciones especiales de nomenclatura de columnas	8
2.4. Calidad e integridad de los datos	8
2.5. Ubicación de la fila del inmueble a valorar	9
3. Modo de Propósito General (General Purpose Mode)	10
3.1. Panorama general	10
3.2. El flujo de trabajo de la barra lateral	10
3.3. Pestañas del panel principal	11

3.4.	Persistencia de la configuración	11
3.5.	Modo oscuro (Dark Mode)	11
3.6.	Configuración regional y de idioma (Locale)	12
3.6.1.	Países admitidos	12
3.6.2.	Anulación del tamaño de papel	12
3.6.3.	Guardar valores predeterminados	13
4.	Modo de Valuación (Appraisal Mode)	14
4.1.	Manejo de la Fila del Inmueble a Valuar	14
4.2.	Effective Date y Sale Age	14
4.3.	Designaciones de Columnas Special	14
4.4.	Residual Contribution (Estimación de Características Poco Frecuentes)	15
4.5.	Panorama de los Ajustes RCA	16
5.	Modo de Análisis de Zona de Mercado (Market Area Analysis Mode)	17
5.1.	Diferencias con el Modo de Valuación	17
5.2.	Cuándo Usar el Modo de Mercado	17
6.	Selección de variables	18
6.1.	Variable(s) objetivo	18
6.2.	La tabla de predictores	18
6.3.	Detección de tipos de datos y anulaciones	19
6.4.	Indicadores Factor y Linear	19
6.5.	Referencia de tipos especiales de columna	19
6.6.	Objetivos múltiples	20
7.	Selección de Parámetros	21
7.1.	Parámetros Básicos	21
7.2.	Pasada Hacia Adelante (Forward Pass)	21
7.3.	Interacciones Permitidas (Allowed Interactions)	22
7.4.	Poda	22
7.5.	Validación Cruzada	23
7.6.	Filtrado por Subconjunto (Subset Filtering)	23
7.7.	Valores de Parámetros Recomendados	23
7.7.1.	Parámetros de la Pasada Hacia Adelante	23
7.7.2.	Parámetros de Poda	25
7.7.3.	Parámetros de Validación Cruzada	25
7.7.4.	Modelo de Varianza	26
7.7.5.	Archivo de Salida de Earth	26
7.8.	Valores Predeterminados de Configuración	26
8.	Ajuste del modelo Earth	27
8.1.	El botón Fit	27
8.2.	Ventana modal de ajuste y salida de traza	27
8.3.	Registro del ajuste	27
9.	Descarga de datos	28
9.1.	Columnas de salida	28
9.2.	Orden de columnas y formato de Excel	28
9.3.	Puntuaciones CQA	29
9.4.	Ordenamiento y fila del inmueble a valuar	29
9.5.	Exportación automática para mgcvUI	29

9.6. Importación en glmnetUI	30
9.6.1. Beneficios clave	30
9.6.2. Importante: la columna de pesos debe coincidir	30
9.6.3. Flujo de trabajo	30
10.Cálculos RCA y descarga	31
10.1. Apertura del diálogo RCA	31
10.2. Interpolación de la puntuación CQA	31
10.3. Columnas de salida	31
10.4. La pestaña RCA Adjustments	32
11.Sales Comparison Grid	33
11.1. Descripción general y propósito	33
11.2. Modal de selección de comparables	33
11.3. Estructura de la cuadrícula	33
11.4. Precio de venta, concesiones y Net SP	34
11.5. Filas agrupadas (Location, Site, Age)	34
11.6. Filas de variables del modelo	34
11.7. CQA Residual y filas de características residuales	35
11.8. Fórmula del Adjusted Sale Price	35
11.9. Protección de hojas y celdas editables	35
11.10Trabajo con la cuadrícula en Excel	35
12.Descarga de informes	37
12.1. Generate Quarto Report	37
12.2. Convert Quarto Report	37
12.3. Contenido del informe	37
13.Los módulos de posprocesamiento: Elastic Net y GAM	39
13.1. ¿Por qué ajustes de apoyo?	39
13.2. Ejecución y estado compartido	39
13.3. La rutina de Elastic Net	39
13.4. La rutina GAM	39
13.5. Un flujo de trabajo práctico	40
14.Conjunto de datos de demostración: Appraisal_1.csv	41
14.1. Descripción	41
14.2. Columnas	42
14.3. Inicio rápido sugerido	42
15.Experimental: extracción de texto con Prolog & reglas de derivación	44
15.1. Qué hace	44
15.2. Requisitos: vProlog (Trealla), no SWI-Prolog	45
15.3. Cómo habilitarla y ejecutarla	45
15.4. Reglas de derivación	45
15.5. Archivo de datos activo versionado	46
15.6. Interfaz programática & archivos incluidos	46
15.7. Experimento relacionado: priores por región climática	46
15.7.1. Más allá de California: paquetes jurisdiccionales y módulos de zona de mercado	46

1 Introducción

1.1 ¿Qué es earthUI?

earthUI es una interfaz gráfica de usuario para el paquete **earth** de R, que implementa Multivariate Adaptive Regression Splines (MARS) (Friedman 1991). El motor de modelado es el paquete **earth** de Stephen Milborrow (Milborrow 2024), documentado en detalle por Milborrow (s.f.[a]) y Milborrow (s.f.[b]). Se ejecuta como una aplicación Shiny local — no hay inicio de sesión, ni servidor, ni cuentas. Se inicia desde R, se importa un conjunto de datos (CSV o Excel), se configura el modelo y se ajusta de manera interactiva.

La aplicación proporciona un flujo de trabajo completo: importación de datos, configuración de variables, ajuste del modelo, gráficos de diagnóstico, importancia de las variables, ecuaciones del modelo e informes descargables en formato HTML, Word o PDF.

1.2 Un paquete, tres rutinas

A partir de la versión actual, el paquete earthUI también *contiene* sus dos Post Processing Modules (módulos de posprocesamiento), que en versiones anteriores eran los paquetes separados glmnetUI y mgcvUI. Instalar earthUI instala las tres rutinas; cada una se ejecuta como su propia aplicación Shiny local:

- `earthUI::launch()` — la aplicación earth (MARS) descrita en este artículo, en el puerto 7878. `earth()` es el motor central de valuación.
- `earthUI::launch_glmnet()` — el Post Processing Module de elastic net (puerto 7879), que importa las funciones base del modelo earth y las reestima bajo regularización lasso/elastic-net (véase “glmnetUI Import” más adelante).
- `earthUI::launch_mgcv()` — el Post Processing Module de GAM (puerto 7880), que utiliza las ubicaciones de los nodos del modelo earth como puntos de partida para los términos suavizados de `mgcv` (véase “mgcvUI Auto-Export” más adelante).

Las tres rutinas comparten un mismo árbol de proyectos, una misma base de datos de configuración y una misma copia de la infraestructura de soporte, de modo que un proyecto abierto en una aplicación queda inmediatamente disponible para las demás. Las rutinas glmnet y mgcv constituyen evidencia de apoyo para la conclusión de valor del modelo earth, no flujos de trabajo en competencia: el patrón previsto consiste en ajustar y refinar primero el modelo earth, y luego ejecutar los módulos sobre el mismo proyecto cuando se desee corroboración.

1.3 Tres modos de propósito

earthUI organiza el trabajo en **proyectos** — entidades con nombre que mantienen juntos los archivos de entrada, los resultados y la configuración en una ubicación fija. Cada proyecto lleva un **Purpose** (propósito), elegido en el asistente New Project cuando se crea el proyecto. Existen tres modos:

- **General** — Regresión Earth para cualquier tipo de población o conjunto de datos. Este es el modo predeterminado. Proporciona el flujo de trabajo completo de modelado MARS sin adiciones específicas de dominio.
- **For Appraisal** (para valuación) — Regresión Earth adaptada a la valuación inmobiliaria. Agrega funciones específicas para la valuación de un inmueble individual, incluidos el manejo del inmueble a valuar, las designaciones especiales de columnas, el Residual Constraint Approach (RCA) (es.: enfoque por restricciones residuales) y la generación del Sales Comparison Grid.

- **Market Area Analysis** (análisis de zona de mercado) — Regresión Earth adaptada a estudios de zonas de mercado. Agrega funciones para analizar grupos de inmuebles en un mercado definido, en lugar de valorar un único inmueble a valorar.

En los tres modos, el motor central de modelado es idéntico — siempre se ajusta un modelo Earth (MARS). El ajuste de propósito controla qué herramientas y elementos de interfaz adicionales están disponibles.

El propósito es fijo por proyecto. El propósito de un proyecto — junto con su país, estado, condado y ciudad — queda bloqueado cuando se crea el proyecto. Cambiar a un proyecto diferente restablece earthUI a un estado predeterminado limpio (los datos importados, los resultados del modelo, las pestañas, la configuración de variables y los parámetros de earth se borran en su totalidad) y luego restaura la configuración guardada de ese proyecto según su propósito.

1.4 Funciones específicas para bienes raíces

Cuando se selecciona **For Appraisal** o **Market Area Analysis**, earthUI activa varias funciones diseñadas para el análisis de bienes raíces:

- **Designaciones especiales de columnas** — Cada predictor puede etiquetarse con un rol especial como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` o `display_only`. Estas designaciones controlan cómo se maneja la columna durante el ajuste, la salida y la generación del Sales Grid (cuadro de comparación). Véase el Capítulo 6 para la referencia completa.
- **Redondeo de latitud y longitud** — Las columnas designadas como latitud o longitud se redondean automáticamente a 3 decimales para prevenir el sobreajuste.
- **Column Sale Age** — Cuando una columna se designa como `contract_date` y se proporciona una Effective Date (fecha efectiva), earthUI calcula una columna `sale_age` (días entre la fecha de venta y la fecha efectiva) y la sustituye como predictor.
- **Cálculos RCA (solo valuación)** — En el modo de valuación, después de ajustar el modelo, earthUI puede calcular la salida del Residual Constraint Approach (RCA), que produce ajustes por comparable, resúmenes de ajustes netos/brutos y un precio de venta ajustado para el inmueble a valorar.
- **Sales Comparison Grid (solo valuación)** — Genera un libro de Excel de varias hojas con fórmulas para los ajustes de los comparables, desgloses residuales por característica y precios de venta ajustados. Véase el Capítulo 11.

***Sugerencia:** El modo de propósito General funciona con cualquier conjunto de datos — financiero, científico, de ingeniería o inmobiliario. Los modos de valuación y de mercado simplemente agregan funciones de conveniencia para los profesionales de bienes raíces.*

1.5 Instalación de earthUI y vProlog

earthUI está publicado en CRAN, por lo que la instalación estándar es una línea en R:

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

Este artículo describe la versión actual. Si CRAN aún ofrece una versión anterior cuando usted lea esto, la versión de desarrollo se instala desde GitHub (nótese el `subdir` — el paquete reside en la carpeta `pkg/` del repositorio):

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

vProlog (opcional). La funcionalidad experimental de procesamiento de observaciones (véase “Prolog Text Extraction” cerca del final de este artículo) utiliza el paquete vProlog — un motor Prolog embebido y autocontenido. Deliberadamente *no* es una dependencia obligatoria: earthUI se instala y se ejecuta sin él, y las funciones de Prolog simplemente permanecen deshabilitadas hasta que esté presente. vProlog se distribuye a través del repositorio r-universe del autor, que ofrece binarios precompilados para Windows y macOS (no se necesita ninguna cadena de herramientas de compilación):

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

No se requiere ninguna instalación de Prolog en el sistema (SWI-Prolog u otra) — el motor Trealla está compilado dentro del propio paquete. Todo lo demás en este artículo funciona sin vProlog.

1.6 Primeros pasos

Para usar earthUI:

1. **Instale el paquete** desde CRAN como se describió anteriormente.
2. **Inicie la aplicación:** ejecute `earthUI::launch()` en R, o desde la línea de comandos ejecute `Rscript -e 'earthUI::launch()'`. La aplicación se abre en su navegador web en el puerto 7878. También puede acceder a la aplicación directamente navegando a `http://localhost:7878` en su navegador. (Los Post Processing Modules se inician de la misma manera: `launch_glmnet()` en el puerto 7879, `launch_mgcv()` en el puerto 7880.)
3. **Cree o abra un proyecto** en la Sección 1 de la barra lateral. El asistente New Project (nuevo proyecto) captura el nombre del proyecto, el Purpose (General, For Appraisal o Market Area Analysis), la ubicación (país, estado, condado, ciudad) y, opcionalmente, un archivo de datos inicial, que se copia en la carpeta `in/` del proyecto.
4. **Importe sus datos** en la Sección 2, que lista los archivos CSV y Excel de la carpeta `in/` del proyecto activo. Los archivos que usted agregue manualmente mediante Finder o Explorer aparecen tras hacer clic en Refresh.
5. **Configure las variables** — elija su(s) variable(s) objetivo y los predictores, establezca los tipos de datos y asigne cualquier rol especial de columna.
6. **Establezca los parámetros de Earth** — degree, nprune, filtros de subset y otros argumentos de `earth()`.
7. **Ajuste el modelo** — haga clic en “Fit Earth Model” y revise los resultados en el panel principal.
8. **Exporte** — descargue las predicciones como Excel, genere paquetes de informes Quarto o (en el modo de valuación) calcule los ajustes RCA y los Sales Comparison Grids.

La configuración se guarda del lado del servidor en una base de datos del proyecto, indexada por proyecto y propósito, mediante los botones “Save current as default” de las Secciones 3 y 4 (véase el Capítulo 3, “Settings Persistence”).

2 Requisitos de los datos de entrada del MLS

Para los flujos de trabajo de valuación inmobiliaria y análisis de mercado, los datos de entrada provienen típicamente de una exportación de un Multiple Listing Service (MLS). Este capítulo describe la estructura de archivo esperada y las columnas que earthUI puede utilizar.

2.1 Formato y estructura del archivo

earthUI acepta archivos **CSV** y **Excel** (.xlsx, .xls). Durante la importación, los nombres de las columnas se convierten automáticamente a **snake_case** — por ejemplo, “Living SqFt” se convierte en **living_sqft**, “Contract Date” se convierte en **contract_date** y “Sale Price” se convierte en **sale_price**. Esta normalización garantiza referencias de columnas consistentes a lo largo de todo el flujo de trabajo. El separador del CSV y la marca decimal utilizados durante la importación se determinan según la configuración regional (véase el Capítulo 3, “Locale & Regional Settings”).

El archivo de datos debe ser una tabla plana con una fila por inmueble y una columna por atributo. La primera fila del archivo debe contener los encabezados de las columnas.

2.2 Columnas requeridas para el modo Appraisal

Si bien earthUI funciona con cualquier conjunto de columnas, el flujo de trabajo completo de valuación (ajustes RCA + Sales Comparison Grid, cuadro de comparación de ventas) se beneficia de contar con las siguientes columnas en la exportación del MLS:

Columna	Tipo especial	Propósito
Sale Price	(target)	Variable de respuesta del modelo
Contract Date	contract_date	Se utiliza para calcular sale_age (días desde la fecha efectiva)
Listing Date	listing_date	Se utiliza junto con la fecha de contrato para calcular el DOM si no existe una columna DOM
Days on Market	dom	Días en el mercado; se muestra en el Sales Grid
Concessions	concessions	Concesiones de la venta; Net SP = Sale Price – Concessions en el Sales Grid
Living Area (SF)	living_area	Habilita los residuos por pie cuadrado (residual_sf , cqa_sf)
Lot Size	lot_size	Se agrupa en la fila "Site Size Dimensions" del Sales Grid
Site Dimensions	site_dimensions	Se agrupa con el tamaño del lote en el Sales Grid
Latitude	latitude	Se redondea a 3 decimales; se utiliza para el cálculo de proximidad y el grupo Location
Longitude	longitude	Se redondea a 3 decimales; se utiliza para el cálculo de proximidad y el grupo Location
Area ID	area	Se agrupa en la fila "Loc: Long Lat Area" del Sales Grid
Actual Age	actual_age	Se agrupa en la fila "Actual Age Effective Age" del Sales Grid
Effective Age	effective_age	Se agrupa con la edad real en el Sales Grid
Parcel Number	(none)	Se muestra en la fila APN del Sales Grid
Listing ID	(none)	Se muestra en la fila APN del Sales Grid
Address	display_only	Se muestra en la fila Address del Sales Grid; se excluye del modelo

Los nombres de las columnas de la hoja de cálculo pueden estar en un idioma extranjero — los nombres “especiales” están en inglés para que el programa de R pueda darles un tratamiento especial. En los demás casos, los nombres de columna proporcionados aparecen en los modelos de regresión, en los gráficos y (si se realizan avalúos) en el Intermediate Sales Grid.

No todas las columnas son requeridas. earthUI se adapta — si falta una columna, la característica correspondiente simplemente se omite. Por ejemplo, si no se designa una columna **concessions**, la fila Net SP del Sales Grid muestra el Sale Price sin deducción de concesiones. Sin embargo, para los modelos de precios inmobiliarios, ciertas columnas son altamente recomendadas para lograr un ajuste aceptable:

1. **Sale Age** — el número de días entre la fecha de venta del contrato y la fecha efectiva del avalúo o análisis. Si se utiliza un historial de ventas de varios años, especialmente para períodos superiores a 5 años, **sale_age** suele desempeñar un papel central en la estimación del precio de venta. De hecho, con frecuencia es tan importante que, sin él, earthUI no logra proporcionar modelo alguno.
2. **Living Area** — también se conoce con nombres como “Living Sqft,” “GLA” (gross living area, superficie habitable bruta), entre otros. Es otro determinante principal del precio de venta.
3. **Total Bath Count** — el número total de baños completos, de un cuarto, medios y de 3/4. Por ejemplo, dos baños completos y un medio baño corresponderían a un valor de 2.5.
4. **Garage Bays** o **Garage Area** — el número de espacios de estacionamiento en el garaje o la superficie del garaje en pies cuadrados.
5. **Lot Size** — la superficie del terreno del inmueble, típicamente en pies cuadrados o acres.
6. **Longitude**, **Latitude**, y, si está disponible, **Area ID**. Los ajustes correspondientes se combinarán bajo un único ajuste de Location en el Sales Grid.

2.3 Convenciones especiales de nomenclatura de columnas

earthUI identifica las columnas por su **designación de tipo especial**, no por su nombre de columna. Puede asignar a sus columnas cualquier nombre en la exportación del MLS — lo que importa es que asigne el tipo especial correcto en la tabla Variable Configuration (configuración de variables) del Capítulo 6.

Por ejemplo, su MLS podría exportar la superficie habitable como “GLA”, “Living SqFt”, “liv_area” o “gross_living_area”. Después de la importación (donde se convierte a snake_case), simplemente la designa como **living_area** en el menú desplegable Special. earthUI la utilizará entonces para los cálculos de residuos por pie cuadrado y para la agrupación en el Sales Grid, independientemente de su nombre original.

2.4 Calidad e integridad de los datos

- **Valores faltantes (NA):** Las filas con valores NA en cualquier columna predictora o de destino se eliminan automáticamente antes del ajuste. El **na.action** siempre se establece internamente en **na.fail**, y la eliminación de NA se realiza antes de la llamada a **earth()**.
- **Columnas de fecha:** Las fechas deben estar en un formato que R pueda interpretar (p. ej., 2025-06-15, 06/15/2025, June 15, 2025). earthUI detecta automáticamente las columnas de fecha cuando al menos el 80 % de los valores se interpretan correctamente. Los años de dos dígitos (p. ej., 06/15/25) se leen como el año coincidente más cercano, de modo que las exportaciones recientes del MLS se interpretan correctamente; los años de cuatro dígitos siempre se toman de forma literal.
- **Columnas numéricas:** El precio de venta, la superficie habitable, el tamaño del lote, las concesiones y campos similares deben ser numéricos. Si su MLS exporta los precios con

símbolos de moneda o separadores de miles (p. ej., “\$350,000” o “350.000,00”), puede ser necesario limpiarlos antes de la importación.

- **Columnas factor:** Las variables categóricas como el identificador de zona, el estilo o el estado deben contener un número manejable de valores únicos. earthUI detecta automáticamente las variables factor, pero usted puede anular dicha detección.

***Sugerencia:** Revise la columna T/F/NA de la tabla de predictores después de la importación — muestra los conteos de valores true/false/NA de cada columna. Las columnas con más del 30 % de valores faltantes se marcan en rojo. Considere excluir las columnas con alto porcentaje de NA o limpiar los datos antes de la importación.*

2.5 Ubicación de la fila del inmueble a valuar

En el modo **Appraisal**, la fila 1 debe corresponder al inmueble a valuar. Todas las filas restantes son ventas comparables. El precio de venta del inmueble a valuar puede dejarse en blanco (NA) o establecerse en cualquier valor — earthUI lo trata como NA durante el ajuste en cualquier caso.

En el modo **Market Area Analysis**, colocar un inmueble a valuar en la fila 1 es opcional, y ninguna fila se excluye automáticamente. Para excluir del ajuste la fila del inmueble a valuar (o cualquier otra fila), asígnele un valor de cero en una columna de ponderaciones designada o utilice un filtro de subconjunto.

En el modo **General**, no existe un tratamiento especial de filas — todas las filas se tratan por igual.

3 Modo de Propósito General (General Purpose Mode)

3.1 Panorama general

El modo General Purpose (propósito general) es el modo predeterminado al iniciar earthUI. Proporciona el flujo de trabajo completo de modelado MARS para cualquier conjunto de datos — no solo para bienes raíces. Puede utilizarse earthUI para datos científicos, análisis financiero, estudios de ingeniería o cualquier problema de regresión en el que se desee explorar relaciones no lineales e interacciones entre variables.

En el modo General, la interfaz omite las funciones específicas de bienes raíces (columnas especiales, antigüedad de la venta, redondeo de coordenadas, RCA). La barra lateral se simplifica para centrarse en la selección de variables, la configuración de parámetros, el ajuste del modelo y la exportación.

3.2 El flujo de trabajo de la barra lateral

La barra lateral está organizada en secciones numeradas y plegables que guían al usuario desde la selección del proyecto hasta la exportación:

1. Project (proyecto) — Permite elegir un proyecto existente o crear uno nuevo mediante el asistente New Project, que registra el propósito, el país y la ubicación administrativa (todos quedan bloqueados una vez creado el proyecto). El proyecto activo determina de dónde se leen los archivos de entrada y dónde se escriben todos los resultados — no existe un campo separado de carpeta de salida.

2. Import Data (importar datos) — Permite elegir un archivo CSV o Excel de la carpeta in/ del proyecto activo; un enlace Refresh vuelve a explorar la carpeta en busca de archivos agregados manualmente. Para archivos de Excel con varias hojas, aparece un selector de hoja. Los nombres de columna se convierten automáticamente a snake_case.

3. Variable Configuration (configuración de variables) — Selector de la variable objetivo (admite múltiples objetivos), Effective Date (fecha efectiva) y la tabla de predictores con casillas de verificación para Include, Factor, Linear y Latent. Véase el Capítulo 6 para los detalles completos.

4. Earth Call Parameters (parámetros de la llamada a earth) — Todos los argumentos de la función `earth()`: degree, penalty, nk, método de poda, validación cruzada y más. Véase el Capítulo 7 para la referencia completa de parámetros.

Execute Prolog Processing (ejecutar procesamiento Prolog) — Una sección opcional, sin numerar, entre las Secciones 4 y 5, deshabilitada a menos que la función experimental de Prolog esté habilitada en Settings. Véase el Capítulo 14.

5. Fit Earth Model (ajustar el modelo earth) — Un campo de semilla aleatoria y el botón que ejecuta el modelo de forma asíncrona. Véase el Capítulo 8 para los detalles del ajuste.

6. Download Estimated Target Variable(s) & Residuals (descargar variables objetivo estimadas y residuos) — (titulada “Download Estimated Sale Prices & Residuals” en los modos de valuación y de mercado) Exporta las predicciones, los residuos, las puntuaciones CQA y las contribuciones por función g como un archivo de Excel. Véase el Capítulo 9.

7–8. RCA Adjustments y Sales Grid — Solo en modo de valuación. Véanse los Capítulos 10 y 11.

9. Generate Quarto Report (generar informe Quarto) — Escribe un paquete de informe Quarto autocontenido (fuente .qmd más recursos) en la carpeta de salida del proyecto. Numerada como 7 fuera del modo de valuación. Véase el Capítulo 12.

10. Convert Quarto Report (convertir informe Quarto) — Renderiza cualquier archivo .qmd a HTML, Word o PDF. Véase el Capítulo 12.

3.3 Pestañas del panel principal

Después del ajuste, el panel principal ofrece diez pestañas:

Pestaña	Contenido
Data	Vista previa de los datos importados. En modo de valuación, se divide en las tablas Subject Property (inmueble a valuar) y Comparable Sales (ventas comparables).
Equation	La ecuación del modelo ajustado mostrada en LaTeX/MathJax, con cada función base y su coeficiente.
Summary	Métricas clave (R^2 , GRSq, GCV, RSS, $CV R^2$) y una tabla de coeficientes.
Variable Importance	Gráfico de barras y tabla clasificada de las puntuaciones de importancia de los predictores.
Contribution	Gráficos de contribución de las funciones g. Seleccione una función g del menú desplegable; los predictores univariados se muestran como gráficos de líneas, y las interacciones bivariadas ($\text{degree} \geq 2$) como gráficos 3D de perspectiva y de contorno.
Correlation	Mapa de calor de las correlaciones entre predictores.
Diagnostics	Tres gráficos: Residuals vs Fitted, Normal Q-Q y Actual vs Predicted.
RCA Adjustments	(solo en modo de valuación) Histogramas de Residual Adj %, Net Adj % y Gross Adj % entre los comparables, que se completan una vez finalizado el Paso 7 (Calculate RCA Adjustments).
Residual Contribution	Estima las contribuciones de mercado de características raras o excluidas mediante la regresión conjunta de los residuos del modelo sobre columnas que el modelo nunca utilizó (véase subsection 4.4).
ANOVA	Tabla de descomposición ANOVA que muestra la contribución de cada función base al RSS.
Earth Output	Salida de texto sin procesar de 'earth::summary()', incluidos los detalles del paso de poda.

3.4 Persistencia de la configuración

earthUI guarda su configuración del lado del servidor en una base de datos de proyectos (`projects.sqlite` en la raíz de `regProj`), indexada por proyecto y propósito. El guardado es **accionado por botones**: el botón “Save current as default” (guardar la configuración actual como predeterminada) de la Sección 3 guarda el objetivo, la fecha efectiva y la configuración de los predictores, mientras que el botón de la Sección 4 guarda los parámetros de la llamada a earth y la matriz Allowed Interactions. Cada botón actualiza únicamente su propia mitad, de modo que nunca se sobrescriben entre sí, y **el ajuste del modelo no guarda automáticamente** — puede experimentarse libremente sin sobrescribir los valores predeterminados guardados. Los diversos archivos de entrada de un proyecto (un pequeño extracto de prueba, el conjunto de datos completo) comparten una sola configuración y, dado que la configuración reside en el árbol del proyecto, viaja con él cuando se sincroniza, se respalda o se comparte la carpeta `regProj`.

3.5 Modo oscuro (Dark Mode)

Haga clic en el ícono de luna/sol en la esquina superior derecha para alternar entre los temas claro y oscuro. La preferencia de tema se guarda en el almacenamiento local y persiste entre sesiones.

3.6 Configuración regional y de idioma (Locale)

earthUI admite convenciones internacionales de formato de números, fechas y CSV mediante un sistema de configuración regional basado en países. El menú desplegable **Country** (país) del menú **Settings** (configuración; ícono de engranaje) de la barra de menú superior selecciona un preajuste para los 31 países admitidos. Cada preajuste configura:

- **Separador CSV** — coma (,) para EE. UU./Reino Unido/Japón o punto y coma (;) para la mayor parte de Europa, donde la coma se usa como marca decimal.
- **Marca decimal** — punto (.) o coma (,).
- **Separador de miles** — coma (EE. UU./Reino Unido/Japón), punto (Alemania/Italia/España), espacio (Finlandia/Francia/Polonia/países bálticos/Ucrania/Rusia) o apóstrofo (Suiza).
- **Formato de fecha** — MM/DD/YYYY (EE. UU.), DD/MM/YYYY (la mayor parte de Europa) o YYYY-MM-DD (Suecia/Lituania/Japón/Canadá).
- **Tamaño de papel** — Letter (EE. UU./Canadá/México) o A4 (el resto del mundo).

3.6.1. Países admitidos

País	Sep. CSV	Decimal	Miles	Fecha	Papel	
EE. UU., Canadá (csv)	,	.	,	MDY/YMD	Letter	
Reino Unido, Irlanda, Australia, Nueva Zelanda	,	.	,	DMY	A4	
Alemania, Austria, Bélgica, Países Bajos, Italia, España, Portugal, Dinamarca, Brasil	;	,	.	DMY	A4	
Francia, Polonia, Rep. Checa, Noruega, Letonia, Estonia	;	,	(espacio)	DMY	A4	
Finlandia, Suecia	;	,	(espacio)	DMY/YMD	A4	
Suiza	;	,	'	DMY	A4	
Ucrania, Rusia	;	,	(espacio)	DMY	A4	
Turquía	;	,	.	DMY	A4	
Japón, Corea del Sur	,	.	,	YMD	A4	
México	,	.	,	DMY	Letter	
Lituania	;	,	(espacio)	YMD	A4	

3.6.2. Anulación del tamaño de papel

Debajo del selector de país, un menú desplegable **Paper** (papel; Letter o A4) permite anular el tamaño de página del informe sin cambiar de país. El separador CSV, la marca decimal y el orden del formato de fecha siguen el preajuste del país seleccionado; al cambiar de país, la anulación del papel se restablece al valor predeterminado de ese país.

3.6.3. Guardar valores predeterminados

Haga clic en **Save** (guardar) en el menú Settings para almacenar sus preferencias regionales de forma global. Estos valores predeterminados se aplican a todas las sesiones futuras, independientemente del proyecto que se abra. La configuración por proyecto (objetivo, predictores, parámetros) se guarda por separado en la base de datos del proyecto; los valores regionales predeterminados persisten en todos los proyectos mediante una base de datos SQLite de nivel de usuario. Este enfoque de dos niveles está diseñado para organizaciones, como las firmas de auditoría, que trabajan con datos de múltiples países — establezca su país más frecuente como predeterminado y ajústelo cuando sea necesario.

***Sugerencia:** el formato de los números en los ejes de los gráficos y en las etiquetas de pendiente se adapta automáticamente a su configuración regional. La configuración alemana usa puntos para los miles (200.000), la finlandesa usa espacios (200 000) y la suiza usa apóstrofes (200'000). No se muestran símbolos de moneda — earthUI es agnóstico respecto de la moneda.*

4 Modo de Valuación (Appraisal Mode)

Cuando se selecciona **For Appraisal** (para avalúo) como Purpose (propósito), earthUI se configura para la valuación de un solo inmueble. Todas las funciones descritas en el Capítulo 3 permanecen disponibles; este capítulo cubre únicamente las adiciones específicas del modo de valuación.

4.1 Manejo de la Fila del Inmueble a Valuar

En el modo de valuación, **la fila 1 del conjunto de datos es el inmueble a valuar** y todas las filas restantes son ventas comparables. El archivo de entrada debe organizarse en consecuencia (véase el Capítulo 2). El precio de venta del inmueble a valuar puede dejarse en blanco o fijarse en cualquier valor — earthUI lo trata automáticamente como NA durante el ajuste.

Después de la importación, la pestaña Data se divide en dos secciones: **Subject Property** (inmueble a valuar; fila 1) y **Comparable Sales** (ventas comparables; filas 2+). La fila 1 se excluye siempre del ajuste del modelo — la notificación “Skipping row 1 (subject). Fitting on N rows.” así lo confirma. Tras el ajuste, el modelo genera de todos modos predicciones para la fila del inmueble a valuar, mostradas como `est_<target>` en la salida.

4.2 Effective Date y Sale Age

Un campo **Effective Date** (fecha efectiva) aparece en la sección Variable Configuration (configuración de variables) en todos los modos de propósito (con la fecha de hoy como valor predeterminado) — los análisis generales y de mercado también tienen una fecha efectiva, no solo los avalúos — pero solo se *utiliza* en los modos de valuación y de mercado. En ellos, si se designa una columna como `contract_date` en el menú desplegable Special, earthUI calcula una columna `sale_age` — el número de días enteros entre la fecha de contrato de cada venta y la fecha efectiva. Esta columna reemplaza a la columna de fecha original como predictor.

La primera vez que se hace clic en Fit (ajustar) después de designar una fecha de contrato, earthUI crea la columna `sale_age` y notifica que debe hacerse clic en Fit nuevamente para incluirla.

4.3 Designaciones de Columnas Special

En los modos de valuación y de mercado, aparece un menú desplegable **Special** (especial) para cada predictor en la tabla Variable Configuration. La lista completa de los tipos especiales y sus efectos:

Tipo Special	Efecto
no	Predeterminado — sin manejo especial
contract_date	Activa el cálculo automático de <code>sale_age</code> a partir del Effective Date
listing_date	Se usa junto con la fecha de contrato para calcular DOM cuando no existe una columna <code>dom</code>
dom	Identifica la columna de Days on Market (días en el mercado); se muestra en el Sales Grid (cuadro de comparación)
concessions	Identifica las concesiones de venta; se usa en la fórmula de Net SP (Sale Price – Concessions) en el Sales Grid
latitude	Valores redondeados a 3 decimales; se usan para la proximidad de Haversine y el grupo Location en el Sales Grid
longitude	Valores redondeados a 3 decimales; se usan para la proximidad de Haversine y el grupo Location en el Sales Grid
area	Se agrupa en la fila "Loc: Long Lat Area" del Sales Grid
living_area	Habilita los residuos por pie cuadrado (<code>residual_sf</code> , <code>cqa_sf</code>) en la salida de descarga
lot_size	Se agrupa en la fila "Site Size Dimensions" del Sales Grid
site_dimensions	Se agrupa con el tamaño del lote en el Sales Grid
actual_age	Se agrupa en la fila "Actual Age Effective Age" del Sales Grid
effective_age	Se agrupa con la edad real en el Sales Grid
display_only	Columna incluida en las exportaciones pero excluida del ajuste del modelo (se permiten varias)
remarks	(Experimental) Columna de comentarios de texto libre que el paso opcional de Prolog analiza y convierte en columnas de características (se permiten varias); véase el Capítulo 14
list	(Experimental) Columna delimitada por comas que el paso opcional de Prolog divide en columnas one-hot TRUE/FALSE (se permiten varias); véase el Capítulo 14

Solo se permite una columna por tipo especial (excepto `display_only`, `remarks` y `list`). Asignar un tipo especial a una segunda columna lo elimina automáticamente de la primera. Una pequeña insignia azul aparece junto al nombre de la variable, mostrando el tipo especial asignado.

4.4 Residual Contribution (Estimación de Características Poco Frecuentes)

Las características demasiado poco frecuentes para el canal de regresión — un taller presente en un puñado de ventas — pertenecen a la capa residual del RCA. La pestaña **Residual Contribution** (contribución residual) estima su contribución de mercado mediante la regresión de los residuos del modelo ajustado sobre columnas que el modelo nunca utilizó, por mínimos cuadrados ordinarios. Seleccionar varias columnas las estima de manera *conjunta*, de modo que las amenidades correlacionadas (p. ej., `outbuilding_sqft` junto con `pr_has_workshop`) se reparten el crédito en lugar de contabilizarse dos veces. Las banderas binarias producen estimaciones de suma global y listan como evidencia el residuo de cada venta marcada; las columnas numéricas producen contribuciones por unidad; el error estándar reportado es de un SE ($t \geq 2$ como criterio habitual de significancia). Los términos respaldados por menos de diez observaciones reciben una advertencia automática CAUTION en la nota generada para el expediente de trabajo, la cual documenta el método, las estimaciones y las salvedades en un solo párrafo para el informe. Solo los predictores que el modelo efectivamente *utilizó* se excluyen de la lista de candidatos — una bandera poco frecuente ofrecida al ajuste pero nunca seleccionada por la pasada hacia adelante sigue siendo elegible, dado que el modelo no absorbió nada de su valor.

4.5 Panorama de los Ajustes RCA

El botón **Calculate RCA Adjustments & Download** (calcular ajustes RCA y descargar; sección 7 de la barra lateral, visible solo en el modo de valuación después del ajuste) calcula ajustes derivados del mercado para cada comparable en relación con el inmueble a valorar. El flujo de trabajo completo del RCA se describe en el Capítulo 10.

Después de calcular los ajustes RCA, el botón **Generate Sales Grid & Download** (generar el Sales Grid y descargar; sección 8 de la barra lateral) queda disponible. El flujo de trabajo del Sales Grid se describe en el Capítulo 11.

***Sugerencia:** La puntuación CQA (estado, calidad y atractivo) que se asigna al inmueble a valorar controla qué proporción de la distribución de los residuos se atribuye al inmueble a valorar. Una puntuación de 5.00 sitúa al inmueble a valorar en la mediana de los comparables.*

5 Modo de Análisis de Zona de Mercado (Market Area Analysis Mode)

Cuando se selecciona **Market Area Analysis** (análisis de zona de mercado) como Purpose, earthUI ofrece las mismas funciones específicas de bienes raíces que el modo de valuación (columnas especiales, antigüedad de la venta, redondeo de coordenadas), pero está orientado al análisis de un grupo de inmuebles en lugar de la valuación de un solo inmueble.

5.1 Diferencias con el Modo de Valuación

- **Sin exclusión automática del inmueble a valorar** — a diferencia del modo de valuación, el modo de mercado no excluye automáticamente la fila 1 del ajuste. Se incluyen todas las filas; para dejar la fila del inmueble a valorar fuera del ajuste, asígnele un valor de cero en una columna de pesos designada (como en el conjunto de datos de demostración) o utilice un filtro de subconjunto.
- **Sin secciones de RCA ni de Sales Grid** — los pasos “Calculate RCA Adjustments & Download” y “Generate Sales Grid & Download” no están disponibles. El modo de mercado se centra en el ajuste del modelo y su salida, no en los ajustes por comparable.
- **Numeración de la barra lateral** — sin los pasos de RCA y Sales Grid, la sección Generate Quarto Report lleva el número 7 en lugar del 9 (la sección Convert Quarto Report conserva el 10).

5.2 Cuándo Usar el Modo de Mercado

El modo Market Area Analysis es apropiado cuando:

- Se construye un modelo de regresión para un vecindario o zona de mercado con el fin de comprender los determinantes del valor
- Se analiza cómo variables como la superficie en pies cuadrados, la edad, el tamaño del lote y la ubicación afectan los precios de venta en un grupo de inmuebles
- Se prepara el soporte para un análisis de condiciones de mercado o una delimitación de vecindario
- Se trabaja con un conjunto de datos que incluye un inmueble a valorar en la fila 1, pero se desea la opción de incluirlo en el ajuste o excluirlo de él

***Sugerencia:** El modo de mercado también es útil para la regresión inmobiliaria general cuando se desean las funciones de columnas especiales (antigüedad de la venta, redondeo de coordenadas) pero no se necesita el flujo de trabajo de ajustes RCA.*

6 Selección de variables

La Sección 3 de la barra lateral — **Variable Configuration** (configuración de variables) — es donde se eligen las columnas que participan en el modelo y cómo se tratan.

6.1 Variable(s) objetivo

El menú desplegable **Target (response) variable(s)** (variable(s) objetivo o de respuesta), situado en la parte superior de la Sección 3, enumera todas las columnas del conjunto de datos. Seleccione una columna para un modelo estándar de respuesta única, o varias columnas para un modelo de respuesta múltiple. Cuando se seleccionan varios objetivos, el modelo ajusta todas las respuestas simultáneamente mediante `earth(cbind(y1, y2, ...) ~ .)`.

Las columnas seleccionadas como objetivos se excluyen automáticamente de la lista de predictores.

6.2 La tabla de predictores

Debajo del selector de objetivos, una tabla enumera todas las columnas restantes con los siguientes campos:

Columna	Descripción
Variable	Nombre de la columna (el nombre completo se muestra en una descripción emergente si está truncado), precedido por un conmutador $\times/+$ (véase más abajo). En los modos de valuación/mercado, una insignia azul muestra el tipo especial asignado.
Type	Menú desplegable de tipo de datos: numeric , integer , character , logical , factor , Date , POSIXct
Include	Casilla de verificación — incluir esta columna como predictor en el modelo
Factor	Casilla de verificación — tratar esta columna como variable categórica
Linear	Casilla de verificación — forzar la entrada lineal únicamente (sin funciones bisagra (<i>hinge</i>))
Latent	Casilla de verificación — mantener esta columna completamente fuera del modelo (excluida de <i>earth</i> y de los datos exportados a <i>glmnetUI</i> / <i>mgcvUI</i>), reservándola para el trabajo con variables latentes
Special	Menú desplegable (solo valuación/mercado) — véase más abajo la referencia de tipos especiales de columna
T/F/NA	Tres conteos por columna: para las lógicas, TRUE/FALSE/NA; para las numéricas, $\neq 0/=0/NA$. Se muestran en rojo cuando falta más del 30 % de los valores

Una línea de ayuda encima de la tabla explica las abreviaturas: “Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`).”

Desactivación suave en lugar de eliminación. El conmutador $\times$ junto al nombre de cada variable atenúa la fila y mantiene la columna fuera del modelo sin eliminarla de los datos; cambia a un $+$ verde que vuelve a habilitar la columna. Las columnas nunca se destruyen.

6.3 Detección de tipos de datos y anulaciones

earthUI detecta automáticamente los tipos de datos al importar. Se reconocen las columnas numéricas, enteras, lógicas, de tipo factor y de fecha. Las columnas de caracteres que parecen fechas (al menos el 80 % de los valores se interpretan según formatos de fecha comunes) se clasifican como **Date**.

Puede anular cualquier detección cambiando el menú desplegable **Type**. Cuando se cambia una columna a **character** o **factor**, la casilla Factor se marca automáticamente. El cambio de tipos afecta la forma en que la columna se pasa a la función **earth()**.

6.4 Indicadores Factor y Linear

- **Factor** — Cuando está marcada, la columna se trata como variable categórica. Esto es apropiado para columnas como **style**, **area_id** o **grade**, que representan grupos discretos en lugar de mediciones continuas. Las columnas de tipo factor entran en el modelo como variables indicadoras (dummy).
- **Linear** — Cuando está marcada, la columna se fuerza a entrar en el modelo de forma lineal — no se crean funciones bisagra (hinge) (lineales por tramos) para ella. Esto es útil cuando se sabe que una variable tiene una relación estrictamente lineal con el objetivo.

6.5 Referencia de tipos especiales de columna

En los modos de valuación y mercado, el menú desplegable **Special** ofrece las siguientes opciones. Cada tipo puede asignarse como máximo a una columna (excepto **display_only**, **remarks** y **list**, que admiten varias):

Tipos de fecha y hora:

- **contract_date** — Activa el cálculo automático de **sale_age**. La columna de fecha original se reemplaza por una columna de enteros que mide los días entre la fecha de venta y la Effective Date (fecha efectiva de valuación).
- **listing_date** — Se utiliza como alternativa para calcular los Days on Market (días en el mercado; DOM = fecha de contrato – fecha de listing) cuando no se designa una columna **dom** explícita.
- **dom** — Identifica la columna de Days on Market. Se muestra en la fila APN y en la fila Date of Sale del Sales Grid (cuadro de comparación).

Tipos monetarios:

- **concessions** — Identifica las concesiones de venta (créditos del vendedor, incentivos al comprador, etc.). Se utiliza en el Sales Grid para calcular el precio de venta neto: Net SP = Sale Price – Concessions.

Tipos de tamaño y ubicación:

- **latitude** — Los valores se redondean automáticamente a 3 decimales para evitar el sobreajuste. Se utiliza para los cálculos de proximidad de Haversine (distancia del inmueble a valuar a cada comparable) y se agrupa en la fila Location del Sales Grid.
- **longitude** — Recibe el mismo tratamiento de redondeo que la latitud. Se utiliza para la proximidad y el grupo Location.
- **area** — Normalmente un identificador de vecindario o de zona. Se agrupa con la latitud y la longitud en la fila “Loc: Long | Lat | Area” del Sales Grid.
- **living_area** — Habilita los cálculos de residuos por pie cuadrado (**residual_sf** y **cqa_sf**) en el archivo de descarga.
- **lot_size** — Se agrupa en la fila “Site Size | Dimensions” del Sales Grid.
- **site_dimensions** — Se agrupa con el tamaño del lote en el Sales Grid (p. ej., “75x120”).

Tipos de antigüedad:

- **actual_age** — Se agrupa en la fila “Actual Age | Effective Age” del Sales Grid.
- **effective_age** — Se agrupa con la antigüedad real en el Sales Grid.

Tipos de visualización:

- **display_only** — La columna se incluye en las exportaciones a Excel pero se excluye por completo del ajuste del modelo. Utilícelo para campos de dirección, números de MLS, identificadores de parcela u otros datos de referencia que no deban ser predictores. Varias columnas pueden tener esta designación.

Tipos de extracción de texto (experimental):

- **remarks** — Una columna de texto libre (p. ej. **public_remarks**) que será analizada en columnas de características por el paso opcional de Prolog. No hace nada hasta que esa funcionalidad se habilita en Settings. Se admiten varias columnas. Véase el Capítulo 14.
- **list** — Una columna delimitada por comas (p. ej. **existing_structures**) que será dividida en columnas one-hot **<col>_<value>** de tipo TRUE/FALSE por el mismo paso. Se admiten varias columnas. Véase el Capítulo 14.

6.6 Objetivos múltiples

Al seleccionar más de una variable objetivo se ajusta un modelo Earth de respuesta múltiple. Cuando se seleccionan varios objetivos:

- El modelo predice todas las respuestas simultáneamente, compartiendo las funciones base entre los objetivos
- El botón **wp (response weights)** (pesos de respuesta) se activa, lo que permite asignar un peso numérico a cada objetivo
- Los modelos de varianza (**varmod.method**) se desactivan (no se admiten para respuesta múltiple)
- Las pestañas de resultados muestran métricas, ecuaciones y gráficos de diagnóstico por respuesta

***Consejo:** Las columnas designadas como **display_only** permanecen en el conjunto de datos y aparecen en la exportación a Excel, pero se excluyen por completo de la tabla de predictores. Utilice esta opción para columnas de identificadores, direcciones u otros campos de referencia.*

7 Selección de Parámetros

La Sección 4 de la barra lateral — **Earth Call Parameters** (parámetros de la llamada a `earth`) — brinda acceso a todos los argumentos aceptados por la función `earth()`. Cada parámetro cuenta con un icono azul de ayuda (?) con una explicación emergente (tooltip). Los parámetros están organizados en subsecciones plegables.

7.1 Parámetros Básicos

Parámetro	Valor predeterminado	Descripción
subset	(vacío)	Expresión de filtrado de filas. Véase “Filtrado por Subconjunto (Subset Filtering)” más abajo.
weights	NULL	Selector de columna para ponderaciones por caso (fila). Solo se listan las columnas numéricas.
wp	NULL	Ponderaciones de respuesta para modelos con múltiples variables objetivo. El botón abre un diálogo con una entrada numérica por objetivo (valor predeterminado 1.0 cada una).
keepxy	off	Conserva x, y, subset y weights en el objeto del modelo.
trace	0	Nivel de traza para la salida del ajuste (0 a 5).
glm	none	Familia GLM opcional: gaussian , binomial o poisson .
degree	1	Orden máximo de interacción. Establecer $\text{degree} \geq 2$ activa automáticamente la validación cruzada y muestra la matriz Allowed Interactions.
penalty	2	Penalización GCV por nodo. Valores más altos producen modelos más simples.

7.2 Pasada Hacia Adelante (Forward Pass)

Parámetro	Valor predeterminado	Descripción
nk	auto	Número máximo de términos antes de la poda. Se completa automáticamente con el valor recomendado para los datos cargados (véase “Valores de Parámetros Recomendados” más abajo).

Parámetro	Valor predeterminado	Descripción
thresh	0.001	Umbral del pasada hacia adelante. Valores más pequeños permiten más términos.
minspan	auto	Separación mínima entre nodos. Se completa automáticamente con el valor recomendado. Los valores negativos fijan el número máximo de nodos por predictor.
endspan	auto	Separación en los extremos — distancia mínima entre un nodo y el borde de los datos. Se completa automáticamente con el valor recomendado.
newvar.penalty	0	Penalización por introducir una nueva variable (favorece la reutilización de los predictores existentes).
fast.k	20	Número de términos padre a considerar en el algoritmo fast MARS.
fast.beta	1	Controla el factor de envejecimiento de fast MARS.

7.3 Interacciones Permitidas (Allowed Interactions)

Cuando **degree** se establece en 2 o más, aparece una matriz de interacciones debajo de los parámetros básicos. Se trata de una cuadrícula triangular superior de casillas de verificación, una por cada par de predictores. Una casilla marcada indica que se permite la interacción entre los dos predictores; desmarcarla prohíbe esa interacción específica.

Al hacer clic en el nombre de un predictor (etiqueta de fila o de columna) se alternan todas las interacciones de esa variable. Las casillas **Allow All** (permitir todas) y **Clear All** (borrar todas) en la parte superior ofrecen control en bloque. La matriz utiliza encabezados fijos, de modo que las etiquetas de columnas y filas permanecen visibles al desplazarse.

Una alerta informativa le recuerda: “Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).” (Los términos de interacción aumentan el riesgo de sobreajuste. Se ha activado la validación cruzada, de 10 pliegues.)

7.4 Poda

Parámetro	Valor predeterminado	Descripción
pmethod	backward	Método de poda: backward , none , exhaustive , forward , seqrep o cv .
nprune	NULL	Número máximo de términos después de la poda. Déjelo vacío para la selección automática.

7.5 Validación Cruzada

Parámetro	Valor predeterminado	Descripción
nfold	10	Número de pliegues de la validación cruzada. Establézcalo en 0 para desactivar la CV.
ncross	20	Número de repeticiones de la validación cruzada.
stratify	on	Estratifica las muestras de CV de modo que cada pliegue tenga una distribución similar de la respuesta.

Cuando $\text{degree} \geq 2$, la validación cruzada se activa automáticamente (nfold se establece en 10) para ayudar a protegerse del sobreajuste debido a las interacciones.

7.6 Filtrado por Subconjunto (Subset Filtering)

La entrada de texto **subset** acepta una expresión de R que filtra qué filas se utilizan para el ajuste del modelo. Puede escribir una expresión directamente (p. ej., `sale_age < 365 & area_id == 460`) o utilizar el botón **Build filter...** (construir filtro) para construirla visualmente.

Build Filter Dialog (diálogo de construcción de filtros) — Haga clic en “Build filter...” para abrir un diálogo guiado. Cada fila de condición tiene una lista desplegable de columnas, un operador (<, >, <=, >=, ==, !=) y una entrada de valor que se adapta al tipo de columna: entrada numérica para números, selector de fechas para fechas y lista desplegable de valores únicos para columnas de tipo carácter/factor. Las condiciones se unen con conectores AND (&) u OR (|). Una vista previa en la parte inferior muestra la expresión y cuántas filas coinciden. Haga clic en **Apply** (aplicar) para insertar la expresión en la entrada de texto.

Las columnas de fecha deben usar los envoltorios `as.Date("...")` o `as.POSIXct("...")` en las expresiones manuales. El diálogo Build Filter lo maneja automáticamente.

El filtrado por subconjunto no es destructivo — las filas excluidas permanecen en el conjunto de datos y reciben predicciones en la exportación a Excel.

***Consejo:** Utilice paréntesis para agrupar las condiciones OR cuando las combine con AND. Sin paréntesis, $A \ \& \ B \ | \ C$ se evalúa como $(A \ \& \ B) \ | \ C$, lo cual puede no ser lo que usted pretende.*

7.7 Valores de Parámetros Recomendados

earthUI muestra un valor recomendado debajo de los parámetros clave de la Sección 4. Estas recomendaciones se actualizan de forma reactiva en función del número de filas de ajuste (n) y de los predictores seleccionados (p). Las fórmulas se derivan del artículo de MARS de Friedman, de los algoritmos internos de earth y de pruebas empíricas. Los campos **nk**, **minspan** y **endspan** se completan previamente con sus valores recomendados al cargarse los datos; una edición manual se conserva.

7.7.1. Parámetros de la Pasada Hacia Adelante

nk (máximo de términos antes de la poda):

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

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

El valor predeterminado de `earth`, $\min(200, \max(20, 2p)) + 1$, no tiene en cuenta el tamaño del conjunto de datos y puede resultar demasiado pequeño para conjuntos grandes, restringiendo la pasada hacia adelante antes de que explore adecuadamente todos los predictores. El término $\lfloor n/10 \rfloor$ permite aproximadamente un término por cada 10 observaciones — una regla práctica estándar para evitar la sobreparametrización. El tope de 100 evita rendimientos decrecientes.

minspan (mínimo de observaciones entre nodos):

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

endspan (mínimo de observaciones desde los límites de los datos):

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

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

El `minspan` calculado automáticamente por `earth` utiliza la ecuación 43 de Friedman: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, que escala como $\ln(np)$ y crece demasiado lentamente para conjuntos de datos grandes. Con $n=1500$ produce solo ≈ 7 , lo que da aproximadamente 245 ubicaciones candidatas de nodo por predictor continuo — demasiadas, lo que permite que la pasada hacia adelante ajuste el ruido. La fórmula recomendada apunta a aproximadamente 100 candidatos por predictor para n grande, mientras que se remite al cálculo automático de `earth` para n pequeño (donde está bien probado). `Endspan` se fija ligeramente por encima de `minspan` para proporcionar protección adicional en los límites, lo cual ayuda cuando los datos tienen colas delgadas (algo común en bienes raíces).

penalty (penalización GCV por nodo):

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

Este es el propio valor predeterminado de `earth` — 2 para modelos aditivos, 3 para modelos con interacciones (la penalización más alta compensa el espacio de búsqueda más grande).

newvar.penalty (penalización por introducir un nuevo predictor):

Recomendado: **0.1** cuando los predictores están correlacionados (p. ej., la superficie habitable con el número de recámaras, el número de baños o el tamaño del lote). Esto sesga la pasada

hacia adelante hacia la reutilización de los predictores ya presentes en el modelo en lugar de introducir alternativas correlacionadas. El mecanismo: la mejora en RSS de cada nuevo predictor se multiplica por $1/(1 + \text{penalty})$. Con `newvar.penalty = 0.1`, una nueva variable debe mejorar el RSS al menos un 10 % más que otro nodo sobre una variable existente. Esto produce modelos más simples con menos predictores sin afectar las estimaciones finales de los coeficientes (la penalización se elimina después de la selección). Déjelo en 0 si los predictores no están correlacionados. Tenga en cuenta que `earth` no admite un `newvar.penalty` distinto de cero junto con ponderaciones por caso — `earthUI` desactiva la entrada (forzando 0) siempre que se designa una columna de ponderación.

7.7.2. Parámetros de Poda

pmethod: Recomendado: **backward**. La poda hacia atrás basada en GCV, la predeterminada, es determinista — los mismos datos producen siempre el mismo modelo, sin dependencia de semillas aleatorias. Esto es importante para la reproducibilidad, especialmente en el trabajo de valuación. El método `cv` utiliza validación cruzada para la poda, lo que introduce dependencia de la semilla.

nprune: Recomendado: **dejar vacío** (NULL). Deje que GCV seleccione el número óptimo de términos. Establecer `nprune` impone un tope rígido que anula el criterio de GCV.

7.7.3. Parámetros de Validación Cruzada

nfold (pliegues de CV): recomendado por banda de tamaño muestral:

n (filas de ajuste)	nfold
0–299	5
300–399	6
400–499	7
500–599	8
600–699	9
700+	10

Más pliegues entrenan cada pliegue con más datos (una estimación menos sesgada), pero reducen el pliegue de prueba reservado. La tabla escala gradualmente hasta el estándar de 10 pliegues a medida que crece la muestra, manteniendo cada pliegue de prueba lo suficientemente grande (aproximadamente 50+ filas) para una métrica estable por pliegue, y recurre a 5 pliegues para muestras pequeñas — donde la estimación de CVR^2 sigue siendo ruidosa, por lo que conviene aumentar `ncross` (p. ej. 15–20) para promediar la varianza. El tope de 10 refleja que cada pliegue adicional supone un ajuste completo de `earth` a cambio de una ganancia diagnóstica insignificante. Con `pmethod = "backward"`, la validación cruzada no afecta al modelo — solo calcula el CVR^2 diagnóstico y proporciona residuos para el modelo de varianza.

ncross (repeticiones de CV): recomendado por banda de tamaño muestral:

n (filas de ajuste)	ncross
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

Las repeticiones promedian la suerte en la partición de pliegues para estabilizar el CVR²; las muestras más grandes requieren menos repeticiones porque una sola validación cruzada ya es más estable. Los residuos de la CV repetida también alimentan el modelo de varianza.

7.7.4. Modelo de Varianza

varmod.method: Recomendado: **lm** — el valor predeterminado de Milborrow: un modelo de varianza lineal robusto para intervalos de predicción (la dispersión se estima a partir de residuos ruidosos, por lo que un modelo simple generaliza mejor). Si **lm** no converge con sus datos, pruebe **rlm** o **const** (homocedástico, siempre converge), o recurra a **earth** (que también maneja una dispersión no lineal). Requiere **nfold > 1** y **ncross > 1**.

7.7.5. Archivo de Salida de Earth

Después de cada ajuste exitoso, earthUI escribe `<filename>_earth_output_<timestamp>.txt` en la carpeta de salida. Este archivo contiene los términos del modelo, estadísticas de resumen (R^2 , GRSq, CVRSq), detalles del modelo de varianza y el registro de traza. Se crea un archivo por ajuste, lo que proporciona un registro acumulativo para comparar configuraciones de parámetros.

7.8 Valores Predeterminados de Configuración

Un botón de opción en la parte superior de la Sección 4 controla qué valores predeterminados se cargan:

- **Use last settings for input file** (usar la última configuración para el archivo de entrada) — restaura la configuración que usted utilizó la última vez con este archivo en particular
- **Use default settings** (usar la configuración predeterminada) — aplica sus valores predeterminados personalizados guardados
- **Earth defaults** (valores predeterminados de earth) — restablece todos los parámetros a los valores de fábrica

Haga clic en **Save current as default** (guardar la configuración actual como predeterminada) en la parte inferior de la Sección 4 para almacenar los parámetros actuales de earth y la matriz Allowed Interactions; el botón correspondiente en la Sección 3 almacena la configuración de la variable objetivo, la fecha efectiva y los predictores. Ambos se guardan por proyecto y propósito en la base de datos del proyecto.

8 Ajuste del modelo Earth

8.1 El botón Fit

La sección 5 de la barra lateral contiene el botón **Fit Earth Model** (ajustar el modelo earth) junto con un campo **Random seed** (semilla aleatoria). La semilla hace reproducible la asignación de los pliegues de validación cruzada — un menú desplegable recuerda las últimas 10 semillas utilizadas, y la semilla empleada queda registrada en la pestaña **Earth Output** después del ajuste. Antes de ajustar, earthUI valida su configuración — debe estar seleccionada una variable objetivo y debe incluirse al menos un predictor. En los modos de valuación y de mercado, las columnas de latitud y longitud se redondean, y la antigüedad de la venta se calcula a partir de la fecha efectiva y la fecha de contrato (si están designadas).

8.2 Ventana modal de ajuste y salida de traza

Al hacer clic en **Fit**, aparece una superposición modal oscura con:

- **Encabezado** — “Fitting Earth Model” con un cronómetro que cuenta los segundos transcurridos y un botón **Abort** (abortar) que cancela el ajuste en segundo plano
- **Trace log** (registro de traza) — un área de texto monoespaciada y desplazable que muestra en tiempo real la salida de la función `earth()`: información del conjunto de datos, progreso del pasado hacia adelante, pliegues de validación cruzada y tiempo de finalización
- **Codificación por colores** — líneas normales en verde, líneas de validación cruzada en amarillo, errores en rojo

earthUI ajusta los modelos de forma asíncrona mediante `callr::r_bg()`, que ejecuta el cálculo en un proceso de R separado. Esto mantiene la aplicación con capacidad de respuesta durante ajustes de larga duración. Un observador con sondeo cada 300ms lee la salida del proceso en segundo plano y la transmite al Trace log.

Cuando el ajuste concluye, aparece un mensaje “Done in X.Xs”, se agrega un botón de cierre (X) a la ventana modal y se retira el fondo oscurecido para que las pestañas de resultados puedan renderizarse bajo demanda. La ventana modal se cierra luego automáticamente una vez que las salidas visibles terminan de renderizarse — o bien usted puede cerrarla antes.

En caso de éxito, se añade una marca de verificación verde al botón **Fit**. Si ocurre un error, el mensaje de error se muestra en el Trace log y aparece una notificación.

8.3 Registro del ajuste

Cada ajuste (exitoso o fallido) escribe un archivo de registro en su carpeta de salida:

- **Nombre del archivo:** `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Contenido:** marca de tiempo, toda la salida de traza del proceso de ajuste y cualquier mensaje de error
- **Ubicación:** la carpeta de salida del proyecto activo

***Sugerencia:** Si callr no está instalado, earthUI recurre al ajuste síncrono con una barra de progreso simple. Instale callr para disfrutar de la experiencia completa de salida de traza.*

9 Descarga de datos

Después del ajuste, la sección 6 de la barra lateral — titulada “Download Estimated Target Variable(s) & Residuals” en el modo general y “Download Estimated Sale Prices & Residuals” en los modos de valuación y de mercado — descarga un archivo de Excel con predicciones y diagnósticos. Esta salida se utiliza en el Paso 7 (RCA) para asignar una calificación CQA (estado, calidad y atractivo; Condition-Quality-Appeal) al inmueble a valorar. La salida se ordena por **residual_sf** y **cqa_sf** para ayudarle a evaluar dónde se ubica el inmueble a valorar dentro del ranking. Si el modelo es de buena calidad, los inmuebles deberían quedar ordenados desde el menos atractivo hasta el más atractivo según las características residuales que no entraron en la regresión. El valor central debería ser aproximadamente 0, la mitad inferior valores negativos y la mitad superior valores positivos. Debería encontrar las viviendas de peor calidad, o “fixers” (para remodelar), cerca del final del ranking y las viviendas más atractivas en la parte superior. Habitualmente habrá excepciones por anomalías tales como ejecuciones hipotecarias, ventas cortas (short sales), ventas sucesorias (relacionadas con herencias) y ventas rápidas motivadas por un cambio de empleo u otras razones. La investigación de las anomalías suele revelar una razón pertinente para la anomalía en el precio.

La etiqueta del botón se adapta al modo de propósito:

- General/Mercado: **Download Output (Excel)**
- Valuación: **Download Intermediate Output (Excel)**

El formato del nombre de archivo es <datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx.

9.1 Columnas de salida

Para cada variable objetivo se anexan las siguientes columnas:

Columna	Descripción
est_<target>	Predicción del modelo, p. ej. est_sale_price (1 decimal)
residual	Valor real menos el predicho (1 decimal)
cqa	Puntuación Condition-Quality-Appeal (2 decimales, rango 0–10)
residual_sf	Residuo dividido entre la superficie habitable (si está designada, 1 decimal)
cqa_sf	CQA calculado a partir del ranking mediante residual_sf (2 decimales)
<variable>_contribution	Valor de la contribución por función g (1 decimal, una columna por función g)
basis	Contribución al valor del intercepto, igual para todos los inmuebles (1 decimal)
calc_residual	Valor real menos (basis + todas las contribuciones) — columna de verificación (1 decimal)

Para modelos con múltiples variables objetivo, se agrega un sufijo **_<i>** para distinguir las columnas de cada objetivo.

9.2 Orden de columnas y formato de Excel

Las columnas de ranking se colocan en la **posición más a la izquierda** del archivo de salida, en este orden: **residual_sf**, **cqa_sf**, **residual**, **cqa**. Esto facilita recorrer la lista ordenada y evaluar dónde se ubica el inmueble a valorar dentro de la distribución de CQA.

A estas columnas se les aplica formato numérico de Excel:

Columna	Formato	Ejemplo
<code>residual_sf</code>	Numérico, 2 decimales	12.34
<code>cqa_sf</code>	Numérico, 2 decimales	7.25
<code>residual</code>	Numérico, 0 decimales	15,200
<code>cqa</code>	Numérico, 2 decimales	6.80

9.3 Puntuaciones CQA

La puntuación CQA (Condition-Quality-Appeal) clasifica el residuo de cada fila frente a todos los demás residuos. Para una fila dada, el CQA es el porcentaje de filas con un residuo con signo menor, multiplicado por 10. Esto produce una escala de 0–10 en la que:

- **CQA alto** (≈ 9 –10) — el inmueble se vendió por mucho más de lo que el modelo predijo (residuo positivo grande)
- **CQA bajo** (≈ 0 –1) — el inmueble se vendió por mucho menos de lo predicho (residuo negativo grande)
- **CQA ≈ 5** — el residuo se encuentra cerca de la mediana de todos los residuos

Cuando se designa una columna `living_area`, `cqa_sf` proporciona el mismo ranking con base en los residuos por pie cuadrado.

9.4 Ordenamiento y fila del inmueble a valorar

En los modos de valuación y de mercado, las filas de comparables se ordenan por `residual_sf` de forma descendente (o por `residual` si no se designó superficie habitable). La fila del inmueble a valorar (fila 1) permanece en la posición 1 y no se ordena. En el modo general, las filas se exportan en su orden original.

Este ordenamiento, combinado con las columnas de ranking situadas más a la izquierda, permite al valuator recorrer rápidamente los comparables desde el más sobrepredicho hasta el más subpredicho, y evaluar dónde cae la puntuación CQA asignada al inmueble a valorar dentro de esa distribución.

Los niveles de las variables factor en los datos de predicción se alinean con los datos de entrenamiento antes de llamar a `predict()`. Las filas con niveles de factor no vistos producirán predicciones NA.

***Sugerencia:** La columna `calc_residual` siempre debería ser igual a `residual` (dentro del margen de redondeo). Si difieren de manera significativa, ello indica un problema de cálculo que conviene investigar.*

9.5 Exportación automática para mgcvUI

En cada ajuste exitoso con `degree` ≤ 2 , earthUI guarda automáticamente el objeto de resultado completo como un archivo `.rds` en la carpeta de salida del proyecto. El nombre del archivo sigue el patrón `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

Este archivo puede cargarse con `readRDS()` en el módulo de posprocesamiento mgcv (mgcv Post Processing Module) — incluido en el paquete earthUI y que se inicia con `earthUI::launch_mgcv()`. La rutina GAM utiliza las ubicaciones de los nodos y las funciones base del modelo earth como puntos de partida para los términos suaves del GAM, lo que permite una transición fluida del modelado MARS al modelado GAM.

Los modelos con `degree` > 2 se omiten porque mgcvUI solo admite interacciones por pares. Cuando earthUI se inicia en modo Trilogy (desde la Trilogy UI), una sección adicional **Lock**

Model Output (bloquear la salida del modelo) le permite bloquear el archivo `.rds` de un ajuste determinado — junto con su base de predictores y la configuración compartida de CQA — como el modelo earth canónico que glmnetUI y mgcvUI importan.

9.6 Importación en glmnetUI

El mismo archivo `.rds` guardado para la rutina GAM también puede importarse en el **elastic net Post Processing Module** (módulo de posprocesamiento de red elástica) — incluido igualmente en el paquete earthUI y que se inicia con `earthUI::launch_glmnet()`. En la aplicación glmnet, diríjase a **Section 2: Import from earthUI** y utilice el botón **Browse** para seleccionar el archivo `.rds`.

glmnetUI utiliza el enfoque estándar de `earth::model.matrix()`: las funciones base del modelo earth (funciones bisagra, interacciones y términos lineales) se convierten en las columnas de la matriz de diseño de glmnet. glmnet realiza entonces una regresión regularizada sobre estas columnas base, seleccionándolas y contrayéndolas mediante lasso/red elástica. Esto combina la construcción adaptativa de bases de earth con la regularización de glmnet.

9.6.1. Beneficios clave

- **Interacciones interpretables:** las interacciones basadas en funciones bisagra de earth (incluidos los términos de 3 vías) tienen un significado geométrico claro, a diferencia de las interacciones nativas de producto cruzado de glmnet, que resultan difíciles de explicar en contextos judiciales o de auditoría.
- **No linealidad automática:** las funciones bisagra de earth capturan relaciones no lineales que un modelo glmnet estándar con predictores sin transformar pasaría por alto.
- **Selección de variables:** glmnet puede podar aún más la base de earth, descartando los términos bisagra que no contribuyen.

9.6.2. Importante: la columna de pesos debe coincidir

Advertencia

Si earthUI utiliza una columna de pesos (p. ej., para excluir la fila del inmueble a valuar con `weight=0`), la misma columna de pesos debe designarse en la configuración de variables de glmnetUI. Si las configuraciones de pesos difieren, los conteos de filas no coincidirán y la base de earth no podrá anexarse a la matriz del modelo de glmnet.

9.6.3. Flujo de trabajo

1. Ajuste un modelo earth en earthUI (de cualquier grado).
2. El archivo de resultados `.rds` se guarda automáticamente en la carpeta de salida del proyecto.
3. En glmnetUI, abra **Section 2: Import from earthUI** y navegue hasta el archivo `.rds`.
4. Verifique que se utilicen el **mismo archivo de datos** y la **misma columna de pesos** en ambas aplicaciones.
5. Configure los parámetros de glmnet en la Sección 5 (la Interaction Matrix se deshabilita automáticamente cuando hay una importación de earthUI activa).
6. Haga clic en **Fit Model**. Las columnas de la base de earth aparecen en las pestañas Equation, Coefficients y Contributions.

10 Cálculos RCA y descarga

El botón **Calculate RCA Adjustments & Download** (calcular ajustes RCA y descargar) aparece en la sección 7 de la barra lateral, visible únicamente en el modo de avalúo después de que se ha ajustado un modelo. RCA (Residual Constraint Approach, es.: enfoque por restricciones residuales) produce ajustes por comparable derivados del mercado, relativos al inmueble a valorar.

10.1 Apertura del diálogo RCA

Al hacer clic en el botón se abre un pequeño diálogo modal con:

- **Score type** (tipo de puntuación) — botones de opción para elegir **CQA** o **CQA per SF** (la opción por SF está disponible solo cuando se ha designado una columna `living_area`). Si elige “CQA per SF,” entonces, con base en la puntuación CQA que asigne al inmueble a valorar, su puntuación residual será su superficie habitable multiplicada por el `residual_sf` que corresponda a la puntuación `CQA_SF` dada.
- **CQA value** (valor CQA) — campo de entrada para la puntuación CQA del inmueble a valorar (0.00–10.00). Esta puntuación representa el lugar que, a su juicio, ocupa el inmueble a valorar en la distribución de calidad de los comparables.
- Botón **Generate** (generar) — calcula la salida RCA y la descarga como archivo de Excel. El botón permanece deshabilitado hasta que se introduce una puntuación válida.

10.2 Interpolación de la puntuación CQA

Cuando hace clic en Generate, earthUI interpola el residuo del inmueble a valorar a partir de los pares CQA/residuo de los comparables:

1. Las puntuaciones CQA y los residuos de los comparables se ordenan por CQA de forma ascendente
2. La interpolación lineal (`stats::approx()`) transforma el valor CQA introducido en un residuo
3. Si se utiliza CQA per SF, el residuo por SF se convierte de nuevo en un residuo total multiplicándolo por la superficie habitable del inmueble a valorar
4. El valor estimado del inmueble a valorar se calcula como: predicción del modelo + residuo interpolado

10.3 Columnas de salida

El archivo de Excel del RCA (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) incluye todas las columnas de salida intermedias más:

Columna	Descripción
<code>subject_value</code>	Predicción del modelo + residuo interpolado (solo fila 1)
<code>subject_cqa</code>	La puntuación CQA que usted introdujo (solo fila 1)
<code><variable>_adjustment</code>	Contribución del inmueble a valorar menos contribución del comparable (por función g)
<code>residual_adjustment</code>	Residuo del inmueble a valorar menos residuo del comparable
<code>net_adjustments</code>	Suma de todos los ajustes (contribución + residuo)
<code>gross_adjustments</code>	Suma de los valores absolutos de todos los ajustes
<code>residual_pct</code>	Ajuste residual como fracción del precio de venta del comparable

Columna	Descripción
net_adj_pct	Ajustes netos como fracción del precio de venta del comparable
gross_adj_pct	Ajustes brutos como fracción del precio de venta del comparable
adjusted_sale_price	Precio de venta del comparable + ajustes netos

Las columnas de ajustes indican, para cada comparable, cuánto de la diferencia atribuye el modelo a cada variable. **adjusted_sale_price** es el precio de venta del comparable después de aplicar todos los ajustes derivados del modelo — un conjunto de precios de venta ajustados que debería agruparse en torno al valor estimado del inmueble a valorar.

10.4 La pestaña RCA Adjustments

Una vez calculada la salida del RCA, la pestaña **RCA Adjustments** (ajustes RCA) del panel principal muestra histogramas de las distribuciones de Residual Adj %, Net Adj % y Gross Adj % entre los comparables (un conjunto por variable objetivo en los modelos multiobjetivo), lo que ofrece una comprobación visual rápida de cuán estrechamente se agrupan los ajustes.

Sugerencia: Una puntuación CQA de 5.00 sitúa el inmueble a valorar en la mediana de los comparables. Las puntuaciones superiores a 5 indican que el inmueble a valorar es de calidad superior al promedio; las inferiores a 5 indican calidad inferior al promedio.

11 Sales Comparison Grid

11.1 Descripción general y propósito

El **Sales Comparison Grid** (cuadro de comparación de ventas) es un libro de Excel generado en el modo de avalúo (sección 8 de la barra lateral) después de calcular los ajustes del RCA. Presenta el inmueble a valorar junto con los comparables seleccionados en un formato de cuadrícula estructurado, con fórmulas de Excel que calculan automáticamente los ajustes y un precio de venta ajustado para cada comparable.

La cuadrícula está diseñada para el expediente de trabajo del valuador — combina los ajustes derivados de la regresión del modelo Earth con celdas editables en las que el valuador puede asignar el residuo CQA a características específicas del inmueble. El nombre del archivo de salida es **SalesGrid_<YYYYMMDD_HHMMSS>.xlsx**.

11.2 Modal de selección de comparables

Al hacer clic en “Generate Sales Grid & Download” (generar y descargar el Sales Grid) se abre un cuadro de diálogo modal donde se seleccionan los comparables que se desea incluir:

- Los **comparables recomendados** aparecen premarcados. Se trata de comparables con un porcentaje de ajuste bruto inferior al 25 % del precio de venta, ordenados por antigüedad de la venta en orden ascendente (las ventas más recientes primero) y limitados a 30.
- Los **comparables adicionales** se listan a continuación, sin marcar de forma predeterminada. Estos presentan ajustes brutos del 25 % o más, pero siguen disponibles para su selección (se muestran hasta 50, ordenados por porcentaje de ajuste bruto ascendente).
- Puede seleccionarse un máximo de **30 comparables**, lo que produce hasta **10 hojas** (3 comparables por hoja).

Tras confirmar la selección, los comparables seleccionados se ordenan por porcentaje de ajuste bruto ascendente, de modo que los comparables más similares aparecen en la Hoja 1.

11.3 Estructura de la cuadrícula

Cada hoja tiene una estructura de 20 columnas que acomoda el inmueble a valorar más 3 comparables. Las columnas para cada entidad son:

Entidad	Columnas (5 por entidad)
Subject (inmueble a valorar)	Label, Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC)
Cada comparable	Factual Value 1, Factual Value 2, Factual Value 3, Value Contribution (VC), Adjustment

Las filas, de arriba hacia abajo, son:

1. **Title** — nombre de la hoja (p. ej., “Intermediate Sales Comparable Grid — Sheet 1 of 3”)
2. **Headers** — encabezados de columna “Subject” y de los comparables, con la dirección
3. **Address** — dirección completa del inmueble
4. **APN | MLS# | DOM | Subj.Prox** — número de parcela, identificador del listing, días en el mercado y distancia Haversine (en millas) respecto del inmueble a valorar
5. **Sales Price | Concess. | Net SP** — precio de venta, concesiones y fórmula del precio de venta neto
6. Fila de encabezado **Regression Features** (características de regresión)

7. **BASE VALUE** — el intercepto del modelo (valor base)
8. **Date of Sale | OffMkt | OnMkt** — fecha de contrato, antigüedad de la venta y DOM
9. **Filas agrupadas** (condicionales) — grupos de Location, Site Size y/o Age
10. **Filas de variables del modelo** — una fila por predictor (excluidas las variables agrupadas)
11. **Fila separadora en blanco**
12. Fila de encabezado **Residual Features** (características residuales)
13. **CQA|Residual** — puntuación CQA + fórmula del residuo restante
14. **Filas de características residuales** — filas con nombre + filas en blanco para captura por parte del valuador
15. **Total VC / Net Adjustment** (ajuste neto)
16. **Net Adjustment %**
17. **Gross Adjustment %** (porcentaje de ajuste bruto)
18. **Adjusted Sale Price** (precio de venta ajustado) — fila de fórmula
19. Pie de página de **Copyright**

11.4 Precio de venta, concesiones y Net SP

La fila 5 muestra tres valores para cada comparable:

- **Sale Price** — el precio de venta del comparable según los datos
- **Concessions** — el valor de la columna designada como **concessions** (0 si no se ha designado)
- **Net SP** — una fórmula de Excel: precio de venta – concesiones

La columna del inmueble a valorar muestra “N/A” como precio de venta (dado que se desconoce) y el valor de las concesiones, si está disponible.

11.5 Filas agrupadas (Location, Site, Age)

Cuando se designan ciertos tipos especiales y esas variables aparecen en el modelo, earthUI crea filas agrupadas que combinan variables relacionadas:

Loc: Long | Lat | Area — Aparece cuando cualquiera de **longitude**, **latitude** o **area** está en el modelo. Muestra los valores factuales de cada variable constituyente, una contribución al valor combinada (suma de las VC individuales) y, para los comparables, un ajuste (VC combinada del inmueble a valorar – VC combinada del comparable). Las celdas de VC se presentan con fondo azul claro.

Site Size | Dimensions — Aparece cuando **lot_size** o **site_dimensions** están en el modelo. Misma estructura que el grupo de Location.

Actual Age | Effective Age — Aparece cuando **actual_age** o **effective_age** están en el modelo. Misma estructura.

Las variables consumidas por las filas agrupadas quedan **excluidas de las filas individuales de variables del modelo** que aparecen debajo, lo que evita la doble contabilización en los totales de ajustes.

11.6 Filas de variables del modelo

Debajo de las filas agrupadas (o directamente debajo de **BASE VALUE** si no hay filas agrupadas), una fila por cada predictor restante del modelo muestra:

- **Valores factuales** para el inmueble a valorar y cada comparable (los valores reales de los datos)
- **Value Contribution (VC)** — la contribución de la función g al valor predicho para esa fila
- **Adjustment** (solo comparables) — VC del inmueble a valorar menos VC del comparable

11.7 CQA|Residual y filas de características residuales

La fila **CQA|Residual** muestra la puntuación CQA de cada inmueble y contiene una fórmula para el **residuo restante** (Remaining Residual) — la porción del residuo aún no asignada a características específicas. La fórmula es:

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

Debajo se encuentran las **filas de características residuales** — 5 filas con nombre (Location, View/Appeal, Condition, Quality, Other) más 6 filas en blanco. Son celdas de entrada donde el valuador captura contribuciones al valor correspondientes a características no capturadas por el modelo. A medida que el valuador introduce valores, la fórmula del residuo restante disminuye automáticamente.

Para cada comparable, la columna de ajuste contiene una fórmula: VC de la característica del inmueble a valorar — VC de la característica del comparable.

11.8 Fórmula del Adjusted Sale Price

La fila **Adjusted Sale Price** contiene fórmulas de Excel:

- **Inmueble a valorar:** suma de todas las celdas de contribución al valor desde BASE VALUE hasta la última fila de características residuales. Representa la predicción total del modelo para el inmueble a valorar más las características residuales asignadas por el valuador.
- **Comparables:** Net SP + suma de todas las celdas de ajuste desde la primera fila agrupada o de variable hasta la última fila de características residuales. Esto proporciona el precio de venta del comparable después de todos los ajustes derivados de la regresión y los introducidos por el valuador.

11.9 Protección de hojas y celdas editables

Cada hoja está **protegida** para evitar la modificación accidental de fórmulas y datos. La protección bloquea:

- Todas las celdas de fórmula (Net SP, residuo restante, ajustes, precio de venta ajustado, etc.)
- Todas las celdas de datos (direcciones, precios de venta, valores factuales, contribuciones al valor provenientes del modelo)
- Etiquetas y encabezados

Las únicas celdas **desbloqueadas** (editables) son las entradas de contribución al valor de las características residuales — las celdas bajo la fila **CQA|Residual** donde el valuador captura los desgloses. Estas celdas se presentan con fondo amarillo claro para indicar que son editables.

11.10 Trabajo con la cuadrícula en Excel

1. **Abra el archivo** en Excel o en una aplicación de hoja de cálculo compatible.
2. **Revise los ajustes de la regresión** — los ajustes derivados del modelo vienen prellenados y bloqueados.
3. **Asigne el residuo** — en las filas de características residuales (celdas amarillas), capture su estimación de qué parte del residuo restante es atribuible a cada característica (Location, View, Condition, Quality, etc.). La fórmula del residuo restante disminuirá a medida que asigne valores.
4. **Verifique el Adjusted Sale Price** — esta fórmula se actualiza automáticamente conforme se capturan los valores de las características residuales.

5. **Hojas múltiples** — si seleccionó más de 3 comparables, navegue entre las hojas. Cada hoja contiene el mismo inmueble a valorar en la columna izquierda con 3 comparables distintos.

Sugerencia: el objetivo es asignar el residuo CQA entre las filas de características hasta que el residuo restante se aproxime a cero. Esto demuestra que es posible explicar la diferencia entre la estimación de regresión de cada comparable y su precio de venta real.

12 Descarga de informes

Después de ajustar un modelo, la generación de informes es un flujo de trabajo de Quarto en dos pasos: **Generate Quarto Report** (generar informe Quarto; sección 9 de la barra lateral en el modo de avalúo, 7 en los demás casos) escribe un paquete de informe autocontenido, y **Convert Quarto Report** (convertir informe Quarto; sección 10) lo renderiza — al igual que cualquier otro archivo Quarto — a los formatos finales.

12.1 Generate Quarto Report

Al hacer clic en **Generate Quarto Report** se escribe un paquete Quarto autocontenido bajo la carpeta de salida del proyecto, en `<base>_qmd/`, que contiene el archivo fuente `.qmd` ya poblado, todos los recursos gráficos pregenerados (PNG y PDF), los datos del informe y una plantilla de referencia de Word. En este paso no se realiza ninguna renderización — el paquete puede editarse a mano, combinarse con otras fuentes Quarto mediante *includes* de Quarto, o previsualizarse de manera independiente con `quarto preview`.

12.2 Convert Quarto Report

La sección 10 ofrece un selector de archivos (que apunta por defecto al paquete generado más recientemente en el proyecto actual — aunque puede navegarse hasta cualquier archivo `.qmd`) y casillas de verificación de formato:

- **HTML** — un archivo `.html` con renderización matemática mediante KaTeX, imágenes incrustadas y una tabla de contenido. Utiliza el tema Flatly de Bootstrap y la fuente Roboto Condensed.
- **Word** — un archivo `.docx` apto para edición y distribución. Incluye una tabla de contenido y números de página, y utiliza una plantilla de referencia personalizada para un estilo uniforme.
- **PDF** — compuesto tipográficamente con LuaLaTeX para una salida de calidad profesional. El tamaño de papel sigue la configuración regional (Letter o A4). Incluye páginas apaisadas para matrices de interacción grandes y utiliza Roboto Condensed con Latin Modern Math para las ecuaciones. La opción PDF se ofrece únicamente cuando se detecta una distribución de LaTeX.

Haga clic en **Convert** para renderizar los formatos seleccionados. La renderización se ejecuta en segundo plano — un cuadro de diálogo modal muestra el tiempo transcurrido y el progreso de Quarto mientras la aplicación permanece receptiva. Una notificación confirma la ubicación del archivo al finalizar.

Todas las operaciones (ajuste de modelos, descargas de datos, cálculos RCA, generación de Sales Grids y renderización de informes) se registran con marcas de tiempo de inicio/fin y tiempos transcurridos en `<datafile>_earthui_log.txt` dentro de la carpeta de salida, para diagnóstico de problemas y supervisión del rendimiento.

12.3 Contenido del informe

Todo informe generado incluye las secciones siguientes:

1. **Descripción del conjunto de datos** — número de observaciones, variable(s) objetivo, predictores utilizados, predictores categóricos
2. **Especificación del modelo** — grado, estado de la validación cruzada, número de términos, predictores incluidos en el modelo
3. **Interacciones permitidas** (solo con grado ≥ 2) — la matriz de interacciones con marcas de verificación, formateada según el tipo de salida

4. **Resumen de resultados** — R^2 , GRSq, GCV, RSS y CV R^2 (si la validación cruzada estaba habilitada)
5. **Ecuación del modelo** — la ecuación completa del modelo Earth en notación LaTeX
6. **Coeficientes y funciones base** — tabla de todos los términos, sus coeficientes y las funciones base que los definen
7. **Importancia de las variables** — gráfico de barras y tabla ordenada de las puntuaciones de importancia de los predictores
8. **Contribuciones de las funciones g** — gráficos para cada función g: gráficos de líneas para los términos univariados, gráficos de perspectiva 3D y de contorno para las interacciones bivariadas
9. **Matriz de correlación** — mapa de calor de las correlaciones entre predictores
10. **Diagnósticos** — gráficos de Residuals vs Fitted (residuos frente a valores ajustados), Normal Q-Q y Actual vs Predicted (valores reales frente a predichos)
11. **Descomposición ANOVA** — tabla que muestra la contribución de cada función base al RSS
12. **Salida de Earth** — salida de texto sin procesar de `earth::summary()`, incluidos los detalles de la pasada de poda

En los modelos de respuesta múltiple, las secciones 4–8 y 10 se repiten para cada variable objetivo.

Sugerencia: Los informes requieren el paquete de R *quarto* y una instalación de Quarto. Para la salida en PDF debe instalarse además una distribución de LaTeX (p. ej., mediante `tinytex::install_tinytex()`) — de lo contrario, la casilla PDF permanece oculta.

13 Los módulos de posprocesamiento: Elastic Net y GAM

El paquete earthUI incluye dos módulos de posprocesamiento — una rutina de elastic net (red elástica) construida sobre `glmnet` y una rutina GAM construida sobre `mgcv`. Cada una se trata en profundidad en su propio artículo de este número; esta sección explica *para qué* sirven dentro del flujo de trabajo de earthUI y cómo cooperan las tres rutinas.

13.1 ¿Por qué ajustes de apoyo?

El modelo earth produce la conclusión de valor. Los módulos de posprocesamiento existen para **corroborarla**: dos estimaciones adicionales del mismo inmueble, producidas por motores metodológicamente distintos y ajustadas a los mismos datos del proyecto. Cuando una reestimación mediante elastic net de la base de earth y un suavizado GAM independiente llegan a un valor sustancialmente igual, la conclusión del valuador ya no descansa sobre los supuestos de un solo modelo — una ventaja significativa en contextos de revisión y litigio. Cuando *discrepan*, esa discrepancia es en sí misma diagnóstica: por lo general apunta a datos escasos, a un término excesivamente flexible o a un comparable atípico que conviene reexaminar. Ninguno de los módulos prevalece sobre el modelo earth; constituyen evidencia de apoyo, presentada deliberadamente sin ponderación.

13.2 Ejecución y estado compartido

Cada rutina es una aplicación Shiny local independiente:

- `earthUI::launch_glmnet()` — la aplicación de elastic net, en el puerto 7879.
- `earthUI::launch_mgcv()` — la aplicación GAM, en el puerto 7880.

Las tres aplicaciones leen el mismo árbol de proyecto y la misma base de datos de configuración, de modo que el proyecto abierto en earthUI queda inmediatamente disponible en los módulos — junto con los campos de arrastre que earthUI guarda para él: la Response (variable objetivo), la Effective Date (fecha efectiva), el tipo y valor de CQA del inmueble a valuar para el RCA, y la designación de superficie habitable. No es necesario reingresarlos; los módulos los leen para que los mismos valores fluyan a través de los tres ajustes.

13.3 La rutina de Elastic Net

La rutina de `glmnet` importa las funciones base del modelo earth — sus funciones bisagra (hinges), interacciones y términos lineales — como columnas de una matriz de diseño y las reestima bajo regularización lasso/elastic-net (véase “`glmnetUI Import`” más arriba para los detalles del mecanismo y la salvedad sobre la columna de pesos). El resultado conserva los términos interpretables y geométricamente significativos de earth, a la vez que permite que el regularizador contraiga o elimine aquellos que no justifican su permanencia. Véase el artículo sobre `glmnetUI` en este número para el tratamiento completo.

13.4 La rutina GAM

La rutina de `mgcv` va en la dirección opuesta: en lugar de reutilizar la base de earth, ajusta términos suaves penalizados nuevos, usando las ubicaciones de los nodos del modelo earth como puntos de partida (véase “`mgcvUI Auto-Export`” más arriba). Dado que el GAM vuelve a derivar las formas funcionales bajo un paradigma de suavizado distinto, su concordancia con el modelo earth constituye una corroboración genuinamente independiente y no un mero eco. Se admiten modelos earth con $\text{degree} \leq 2$. Véase el artículo sobre `mgcvUI` en este número para el tratamiento completo.

13.5 Un flujo de trabajo práctico

1. Ajuste y refine **primero el modelo earth**, en earthUI, hasta quedar satisfecho con él. Cada ajuste exitoso ($\text{grado} \leq 2$) exporta automáticamente el archivo **.rds** que consumen los módulos.
2. Abra las rutinas de **glmnet** y **mgcv** sobre el mismo proyecto, importe el resultado de earth y ajuste. Trabaje en una rutina a la vez — los ajustes son de larga duración y las rutinas se coordinan a través del proyecto compartido, por lo que el ritmo sensato es secuencial, no paralelo.
3. Itere con libertad. Cada aplicación conserva sus resultados mientras usted alterna entre ellas, y nada obliga a generar informes por el camino.
4. Solo cuando los tres resultados le satisfagan, regrese a cada rutina y genere su informe Cuarto. Cada método registra su ajuste y su conclusión de valor en el registro de trilogía del proyecto, manteniendo los tres informes trazables a una misma corrida comparativa.

***Sugerencia:** Los módulos son opcionales en todos los niveles: el flujo de trabajo propio de earthUI — ajuste, RCA, Sales Grids, informes — está completo sin ellos. Recorra a ellos cuando una conclusión amerite corroboración: un inmueble a valorar difícil, un segmento de mercado con pocos datos o un encargo con probabilidades de ser objeto de revisión.*

14 Conjunto de datos de demostración: Appraisal_1.csv

earthUI incluye un conjunto de datos MLS de demostración para explorar el flujo de trabajo de valuación. Carguelo programáticamente con:

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

O bien, en la aplicación Shiny, cópielo en la carpeta `in/` de un proyecto (o adjúntelo como archivo de datos inicial al crear el proyecto) y selecciónelo en la Sección 2 (Import Data, importación de datos).

14.1 Descripción

El archivo contiene 1,501 ventas residenciales más 1 inmueble a valuar en la fila 1 (1,502 filas en total), provenientes de una exportación MLS simulada. Los datos representan ventas de viviendas unifamiliares en un mercado con múltiples zonas, con una variedad de tamaños, antigüedades y ubicaciones de los inmuebles.

No se trata de datos reales, sino que se basan en un vecindario realista del norte de California. Toda la información de identificación ha sido alterada o eliminada.

14.2 Columnas

Columna	Tipo	Tipo especial	Descripción
weight	numeric	—	Peso de la observación (0 = excluir del ajuste)
id	numeric	display_only	ID interno del registro
property_id	numeric	display_only	Identificador MLS del inmueble
listing_id	character	display_only	Número de listing MLS
parcel_number	character	display_only	Número de parcela (APN) del Assessor del condado (autoridad de valuación fiscal)
street_address	character	display_only	Dirección del inmueble
city_name	character	display_only	Ciudad
postal_code	character	display_only	Código postal (ZIP)
county_name	character	display_only	Condado
contract_date	Date	contract_date	Fecha del contrato de venta (calcula sale_age)
sale_age	numeric	—	Días desde la fecha del contrato hasta la fecha efectiva (precalculado)
coe_date	Date	display_only	Fecha de cierre de escrow
listing_status	character	display_only	Estado del listing (p. ej., "Sold")
sale_price	numeric	(target)	Precio de venta — variable de respuesta
rent	numeric	—	Renta mensual (para modelos con múltiples variables objetivo)
list_price	numeric	display_only	Precio del listing
original_list_price	numeric	display_only	Precio original del listing
living_sqft	numeric	living_area	Superficie habitable bruta en pies cuadrados
beds_total	integer	—	Número de dormitorios
baths_total	numeric	—	Número total de baños (p. ej., 2.5 = 2 completos + 1 medio baño)
lot_size	numeric	lot_size	Tamaño del lote en pies cuadrados
area_id	integer	area	Identificador de zona MLS
area_text	character	display_only	Nombre de la zona
age	numeric	actual_age	Antigüedad del inmueble en años
year_built	integer	display_only	Año de construcción
latitude	numeric	latitude	Latitud (redondeada a 3 decimales para el modelo)
longitude	numeric	longitude	Longitud (redondeada a 3 decimales para el modelo)
latitude6	numeric	display_only	Latitud con precisión completa
longitude6	numeric	display_only	Longitud con precisión completa
garage_spaces	integer	—	Número de plazas de garaje
fp_count	integer	—	Número de chimeneas
no_of_stories	numeric	—	Número de pisos
style	character	—	Estilo arquitectónico
view	character	—	Tipo de vista (p. ej., "Neighborhood", "Hills")
days_on_market	integer	dom	Días en el mercado
listing_date	Date	listing_date	Fecha del listing
sale_concessions	numeric	concessions	Concesiones del vendedor

14.3 Inicio rápido sugerido

1. Inicie earthUI: `earthUI::launch()`
2. Cree un proyecto con el Purpose (propósito) **For Appraisal** (para avalúo) en el asistente New Project (nuevo proyecto), adjuntando **Appraisal_1.csv** como archivo de datos inicial
3. Seleccione **Appraisal_1.csv** en la Sección 2 (Import Data)
4. Seleccione **sale_price** como variable objetivo
5. Asigne los tipos especiales como se muestra en la tabla anterior
6. Incluya los predictores: **sale_age**, **living_sqft**, **baths_total**, **lot_size**, **area_id** (como factor), **age**, **latitude**, **longitude**, **garage_spaces**
7. Establezca el grado en 1 y haga clic en **Fit Earth Model** (ajustar el modelo earth)
8. Descargue la salida intermedia (Paso 6) y revise la clasificación CQA
9. Calcule los ajustes RCA (Paso 7) con una puntuación CQA de ~5.00

10. Genere el Sales Comparison Grid (Paso 8)

Sugerencia: El conjunto de datos de demostración es una buena manera de explorar todas las funciones de earthUI antes de importar sus propios datos MLS. Los predictores sugeridos arriba producen típicamente un modelo con R^2 superior a 0.85.

15 Experimental: extracción de texto con Prolog & reglas de derivación

La versión 0.10.0 introduce una facilidad **opcional y genuinamente experimental** para extraer características listas para el modelo a partir de las columnas de texto descriptivo de una exportación MLS — ante todo la columna “Public Remarks”, pero también los comentarios del agente y las listas de características delimitadas por comas, como las características de las habitaciones o las estructuras existentes. El motor de extracción es Prolog y, en particular, las **gramáticas de cláusulas definidas (Definite Clause Grammars, DCGs)**: las reglas que reconocen frases como “move in ready,” “3 car garage,” u “owned solar” en el texto de los comentarios se escriben como reglas gramaticales de Prolog.

Esta funcionalidad aún se encuentra en desarrollo. Hoy puede utilizarse *parcialmente* para la extracción — la gramática incluida ya reconoce un conjunto útil de características —, pero adaptarla al vocabulario de su propio MLS exige leer y escribir Prolog y, en particular, DCGs. Si ese no es su caso, omita este capítulo por completo: la funcionalidad está desactivada por defecto y el resto de earthUI funciona exactamente igual sin ella.

Advertencia

Se trata de una funcionalidad experimental en desarrollo activo. Su gramática de extracción no está completa ni validada para trabajo de valuación en producción — las columnas extraídas deben cotejarse con el texto de origen antes de utilizarse con confianza en un modelo. Usar la funcionalidad con eficacia requiere conocimientos prácticos de Prolog y de DCGs. Todo lo que hace es aditivo y reversible; si algo funciona mal, desactive el interruptor de Settings (configuración) y continúe con normalidad.

15.1 Qué hace

Los listings de la MLS contienen una gran cantidad de información relevante para el valor en texto libre y en campos de listas delimitadas. Este paso opcional convierte ese texto en columnas ordinarias que su regresión puede utilizar, mediante dos roles especiales de columna (Capítulo 6):

- Las **columnas remarks** son analizadas por la gramática DCG de Prolog para producir una columna de característica por cada atributo extraído. El prefijo de columna se construye con las primeras letras del nombre de la columna de origen: **public_remarks** produce columnas **pr_***, **agent_remark** produce **ar_***, y así sucesivamente.
- Las **columnas list** (p. ej., **existing_structures = "Shed, Workshop"**) se dividen en columnas one-hot **<col>_<value>** de tipo TRUE/FALSE. Esta parte es puro R y no necesita Prolog en absoluto.

La gramática incluida (**inst/prolog/mls_remarks.pl**) extrae actualmente, entre otras cosas: un nivel graduado de estado con una puntuación **condition_score** de 1–5; indicadores de vista y tipos de vista; indicadores de remodelación de cocina y baño; pisos de madera y de otros materiales; una lista y un conteo de mejoras; plazas de garaje y superficie de garaje; piscina, spa y otras características exteriores; número de chimeneas; construcciones auxiliares (cobertizo, granero, taller, casita, casa de huéspedes, perrera, etc.) con tamaños obtenidos de dimensiones como “40x50”; estacionamiento y almacenamiento para vehículos recreativos (RV); propiedad versus arrendamiento de paneles solares, almacenamiento en baterías; tamaño del lote (incluida la conversión de acres a pies cuadrados); montos de precio ligados al contexto de la venta y asociados a un sujeto (reducción de precio, cuotas de HOA, arrendamiento o liquidación del sistema solar, Mello-Roos); indicios de ubicación; y una puntuación de sentimiento de aspectos destacados/preocupaciones en $[-1, 1]$.

15.2 Requisitos: vProlog (Trealla), no SWI-Prolog

El análisis de los comentarios (remarks) se ejecuta sobre el paquete de R **vProlog** — un motor Trealla Prolog autocontenido. No está en CRAN; instálelo desde el repositorio r-universe del autor, como se muestra en “Instalación de earthUI y vProlog” al comienzo de este artículo (binarios precompilados para Windows y macOS; código fuente en github.com/wcraytor/vProlog). No se requiere ni se utiliza ninguna instalación de Prolog en el sistema (SWI-Prolog u otra). Si vProlog no está presente, el análisis de comentarios se omite con una nota, mientras que la división de columnas `list` se ejecuta de todos modos. La función auxiliar `vProlog_available()` informa si tanto el motor como la gramática incluida están presentes.

15.3 Cómo habilitarla y ejecutarla

1. Abra **Settings** (engranaje) y marque **Enable Prolog / list column processing** (habilitar el procesamiento de columnas Prolog/list). El interruptor está desactivado por defecto, lleva la nota “advanced feature — for users with Prolog expertise” (funcionalidad avanzada, para usuarios con experiencia en Prolog), se recuerda por proyecto y se muestra en el diálogo Project Information (información del proyecto).
2. En la tabla de predictores, establezca el rol **Special** de las columnas pertinentes en `remarks` o `list` (varias columnas pueden llevar cada rol).
3. Utilice la sección **Execute Prolog Processing** (ejecutar el procesamiento Prolog), situada entre las Secciones 4 y 5 de la barra lateral. Ofrece dos pasos:

1. Expand lists & remarks (expandir listas y comentarios) — ejecuta la gramática DCG sobre cada columna `remarks` y divide en columnas one-hot cada columna `list`. La expansión es *incremental*: solo se procesan las columnas designadas desde la última ejecución, y el botón se desactiva una vez que todo lo designado ya ha sido expandido. Una barra de progreso da seguimiento al análisis.

2. Execute derivation rules (ejecutar reglas de derivación) — aplica sus propias reglas `derive/2` de Prolog sobre las columnas expandidas (véase más abajo). Las reglas se someten a una verificación de sintaxis y a una validación de ejecución en seco antes de ejecutarse.

15.4 Reglas de derivación

Para cada fila de datos, earthUI aserta las columnas de la fila como hechos `cell(Name, Value)` y recoge todas las soluciones de sus reglas `derive(Col, Val)` — cada `Col` distinto se convierte en una nueva columna, y gana el *primer* valor encontrado por columna, de modo que puede añadirse a continuación una cláusula de respaldo. Los predicados auxiliares (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) dan soporte a la agregación y al manejo de NA, y los hechos `drop/1` descartan columnas después de ejecutarse las reglas. Por ejemplo:

```
% 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).
```

Las reglas residen en un archivo `<project>_rules.pl` por proyecto, generado a partir de una plantilla completamente comentada, y se editan mediante **Edit derivation rules...** (editar reglas de derivación) en un editor de código flotante y arrastrable que también puede acoplarse a la barra lateral. Un botón **Check** (verificar) valida las reglas (exploración de sintaxis, verificación de carga y una ejecución en seco contra la fila 1 de sus datos) sin ejecutarlas.

15.5 Archivo de datos activo versionado

Después de cualquiera de los dos pasos, el data frame aumentado se escribe en la carpeta de entrada del proyecto como `<original_name>_<YYMMDD_HHMMSS>.xlsx` y se convierte en el archivo de datos **activo** del proyecto — al reabrir el proyecto se cargan los datos aumentados, de modo que las expansiones nunca se recalculan, y ese mismo archivo es el que consumen `glmnetUI` y `mgcvUI`. Las columnas de origen con los comentarios sin procesar se excluyen automáticamente del modelo; las columnas generadas *no* se incluyen automáticamente como predictores. El análisis de texto puede producir muchas columnas, así que marque como **Latent** (latente) las que (todavía) no desee en la regresión, o desactívelas con `×` (Capítulo 6).

15.6 Interfaz programática & archivos incluidos

Toda la cadena de procesamiento está exportada para uso por lotes: `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` orquesta las pasadas, y está construida sobre `remarks_to_columns()`, `split_list_column()`, `apply_user_rules()` y `validate_rules()`. Las fuentes Prolog incluidas bajo `inst/prolog/` son `mls_remarks.pl` (la gramática DCG y la agregación de características), `mls_analyze.pl` (un controlador de entrada/salida de archivos para el análisis por lotes), `mls_rules.pl` (los predicados auxiliares disponibles para las reglas de derivación) y `climate_regions_ca.pl` (más abajo).

15.7 Experimento relacionado: priores por región climática

Una facilidad experimental complementaria aborda las características *de datos escasos* — aquellas que la regresión no puede estimar porque aparecen solo en un puñado de comparables (un enfriador evaporativo, digamos). Tres funciones auxiliares de R — `climate_region_for()`, `market_area_profile()` y `climate_feature_priors()` — clasifican la ubicación de un proyecto (condado + ciudad) en una de siete regiones climáticas de California reconocidas por los compradores (una reducción práctica de las 16 zonas climáticas de edificación de la California Energy Commission) y devuelven priores ordinales de contribución al valor para características de climatización (HVAC) y enfriamiento (calefacción, aire acondicionado, enfriador evaporativo, ventilador de casa completa, filtración) por región. El módulo Prolog `climate_regions_ca.pl` refleja la misma clasificación para su uso dentro de las reglas de derivación; ambos lados se generan a partir de una única fuente canónica en R, de modo que no pueden divergir. Los priores están concebidos para *informar* la valuación de características y servir como verificación de plausibilidad de los coeficientes estimados — las asignaciones de región (en particular para condados que abarcan varias regiones) deben ser revisadas por el valuator.

15.7.1. Más allá de California: paquetes jurisdiccionales y módulos de zona de mercado

California es la primera jurisdicción, no la frontera. `earthUI` se utiliza internacionalmente, y el conocimiento de los priores climáticos se está reestructurando en forma de **paquetes**

jurisdiccionales (jurisdiction packs): tablas de hechos autocontenidas, con forma de datos (definiciones de zonas, asignaciones de unidades administrativas, priores de características), indexadas por los códigos de país y de nivel administrativo del sistema del proyecto, de modo que un paquete puede describir un estado de los EE. UU., un Land alemán o directamente un país pequeño. Una capa complementaria de **módulos de zona de mercado** captura el vocabulario local y priores refinados para una zona de mercado con nombre propio.

Ambos están diseñados para ser creados *en el terreno* por Valuation Engineers y valuadores, en su propio idioma — y, siendo realistas, con asistencia de LLM: el mundo no cuenta con muchos programadores de Prolog, por lo que los artefactos de campo son deliberadamente tablas y no programas. El formato de paquete es un esquema documentado con un ejemplo completo desarrollado (la propia California), expresamente para que un asistente como Claude pueda redactar un paquete válido a partir de la descripción de un mercado en lenguaje llano; earthUI genera el Prolog subyacente a partir del paquete, y un validador bloqueante (verificaciones de esquema, de cobertura, de integridad referencial y de ejecución en seco) debe aprobarse antes de que un paquete pueda habilitarse. Todo paquete creado en el terreno lleva una procedencia obligatoria — el autor responsable, la fecha y la autoridad citada —, de modo que quien respalda el conocimiento es el profesional responsable, no la herramienta de redacción, y los informes pueden citar el paquete y la versión exactos utilizados.

Sugerencia: *Un patrón práctico: expanda los comentarios, consolide los indicadores granulares `pr_*` en un puñado de conteos o booleanos derivados con reglas `derive/2`, elimine con `drop/1` los indicadores sin procesar que consolidó y conserve solo las columnas derivadas como predictores candidatos. Esto mantiene corta la lista de predictores y manejable la colinealidad.*

Referencias

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