

mgcvUI: una interfaz interactiva para el paquete mgcv (modelos aditivos generalizados)

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

© 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. mgcvUI es una interfaz gráfica de usuario para el paquete de R mgcv, que ajusta modelos aditivos generalizados (GAM) mediante splines de regresión penalizada con selección automática de suavidad. Ofrece tres modos de uso — modelado predictivo general, tasación inmobiliaria y análisis de zona de mercado — y guía al usuario a través de la importación de datos, la especificación de términos suaves, el ajuste del modelo, los gráficos de diagnóstico y de efectos, y los informes descargables. Este artículo documenta los requisitos de formato de datos de mgcvUI, el flujo de trabajo de modelado, las pantallas de salida y la referencia completa de funciones.

Palabras clave: modelos aditivos generalizados, GAM, splines penalizados, suavizado, mgcv, R, interfaz gráfica de usuario, valuación inmobiliaria

Nota del editor (julio de 2026): Poco antes de la publicación, mgcvUI se consolidó dentro de earthUI (versión 0.11.0 y posteriores), que ahora ofrece el flujo de trabajo de GAM aquí descrito junto con los flujos de trabajo de MARS y elastic net. mgcvUI sigue disponible en su forma final publicada (0.3.0) en <https://github.com/wcraytor/mgcvUI>; el desarrollo futuro continúa en earthUI. Este artículo se refiere a mgcvUI tal como fue publicado.

Índice

1. Introducción	4
1.1. Historia y antecedentes	4
1.1.1. Modelos aditivos generalizados	4
1.1.2. El paquete mgcv	4
1.1.3. Comparación con MARS (earth) y elastic net (glmnet)	5
1.1.4. La cadena earthUI–mgcvUI	5
1.1.5. Flujo de trabajo recomendado	6
1.2. ¿Qué es mgcvUI?	6
1.3. Tres modos de uso	7
1.4. Funciones específicas del sector inmobiliario	7
1.5. Primeros pasos	7
2. Requisitos de los datos de entrada del MLS	9
2.1. Formato y estructura del archivo	9
2.2. Columnas requeridas en el modo de tasación	9
2.3. Convenciones de nombres de columnas especiales	10
2.4. Calidad e integridad de los datos	10
2.5. Ubicación de la fila del sujeto	10

3. Modo de uso general	11
3.1. Panorama general	11
3.2. El flujo de trabajo de la barra lateral	11
3.3. Importar desde earthUI (opcional)	12
3.4. Pestañas del panel principal	12
3.5. Persistencia de los ajustes	13
3.6. Modo oscuro	13
3.7. Configuración regional y de idioma	13
4. Modo de tasación	15
4.1. Tratamiento de la fila del sujeto	15
4.2. Fecha efectiva y antigüedad de la venta	15
4.3. Designaciones especiales de columnas	15
4.4. Panorama de los ajustes RCA	15
5. Modo de análisis de zona de mercado	16
5.1. Diferencias respecto del modo de tasación	16
5.2. Cuándo usar el modo de mercado	16
6. Selección de variables	17
6.1. Variable objetivo	17
6.2. Transformación de la respuesta	17
6.3. La tabla de predictores	17
6.4. Detección de tipos de datos y anulaciones	18
6.5. Referencia de tipos especiales de columna	18
7. Selección de parámetros	19
7.1. Preajustes de parámetros	19
7.2. Family	19
7.3. Method	19
7.4. Gamma (multiplicador de suavizado)	20
7.5. Validación cruzada	20
7.6. Select (contracción)	20
7.7. Base predeterminada	20
7.8. k predeterminada (dimensión de la base)	20
7.9. Tipo de tensor	21
7.10. Interacciones permitidas	21
7.11. Interacciones factor por término suave	21
7.12. Parámetros avanzados	22
8. Ajuste del modelo	23
8.1. El botón de ajuste	23
8.2. Qué ocurre durante el ajuste	23
8.3. Ajuste en segundo plano y progreso	23
9. Pestañas de resultados	24
9.1. Data	24
9.2. Equation	24
9.3. Summary	24
9.4. Variable Importance	24
9.5. Gráficos de contribución	25
9.5.1. Términos suaves univariantes — $s(x)$	25

9.5.2. Términos factor por término suave — $s(x, by = factor)$	25
9.5.3. Términos de interacción — $ti(x_1, x_2)$ y $te(x_1, x_2)$	25
9.5.4. Términos paramétricos: lineales y de factor	25
9.6. Correlation	25
9.7. Diagnostics	26
9.8. RCA Adjustments	26
9.9. Anova	26
9.10. Mgev Output	26
9.11. Sign Check	26
9.12. Concurvity	26
9.13. Basis Check	27
10.Descarga de datos	28
10.1. Columnas de salida	28
10.2. Puntuaciones CQA	28
11.Cálculos RCA y descarga	29
11.1. Abrir el diálogo RCA	29
11.2. Interpolación de la puntuación CQA	29
11.3. Columnas de salida	29
12.Cuadrícula de comparación de ventas	30
12.1. Diálogo de selección de comparables	30
12.2. Disposición del libro	30
12.3. Protección de las hojas y celdas residuales	30
13.Descarga de informes	31
13.1. Paso 1: generar el informe Quarto	31
13.2. Paso 2: convertir el informe Quarto	31
13.3. Contenido del informe	31
14.Exportación de las funciones del modelo	33
14.1. Funciones de R	33
14.2. Funciones de Python	33
14.3. Cabecera de C++	33
14.4. Base de datos SQLite	33
15.Comparación con earthUI y glmnetUI	34
15.1. Diferencias principales	34
15.2. Funciones compartidas	34
15.3. Flujo de trabajo recomendado	35
16.Conjunto de datos de demostración: Appraisal_1.csv	36
16.1. Descripción	36
16.2. Columnas	36
16.3. Inicio rápido sugerido	37
17.Requisitos del sistema y resolución de problemas	38
17.1. Plataformas admitidas	38
17.2. Versión de R	38
17.3. Paquetes de R requeridos	38
17.4. Dependencias opcionales	39
17.5. Notas específicas de cada plataforma	39

17.5.1. macOS	39
17.5.2. Windows	39
17.5.3. Linux	39
17.6. Degradación elegante	40
17.7. Resolución de problemas	40
17.7.1. “PDF option is not available”	40
17.7.2. “Settings will not persist”	40
17.7.3. “Diagnostics plot unavailable”	40
17.7.4. “Report error: number of items to replace...”	41
17.7.5. El puerto 7880 ya está en uso	41
17.7.6. La instalación de paquetes falla en Linux	41

1 Introducción

1.1 Historia y antecedentes

1.1.1. Modelos aditivos generalizados

Los **modelos aditivos generalizados** (GAM) fueron formalizados por Trevor Hastie y Robert Tibshirani en su monografía de 1990. Los GAM extienden el modelo lineal sustituyendo cada término lineal $\beta_j x_j$ por una función suave $f_j(x_j)$, de modo que el modelo queda así:

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

Cada f_j se estima a partir de los datos mediante una spline de regresión penalizada. Las funciones suaves permiten que el modelo capte relaciones no lineales sin que el usuario tenga que especificar de antemano una forma funcional. La penalización controla la ondulación de cada término suave, lo que evita el sobreajuste sin dejar de permitir la flexibilidad necesaria para seguir los patrones reales.

1.1.2. El paquete mgcv

El paquete de R **mgcv** (Mixed GAM Computation Vehicle) fue desarrollado por Simon Wood en la Universidad de Bath (Wood 2017). Publicado por primera vez en 2001, es hoy la implementación estándar de los GAM en R y se distribuye con cada instalación de R. Sus métodos subyacentes se describen en una serie de artículos que abarcan la estimación estable de múltiples parámetros de suavizado (Wood 2004), las splines de regresión thin-plate (Wood 2003), la estimación REML rápida y estable (Wood 2011) y la selección general de modelos suaves (Wood, Pya y Säfken 2016). Entre sus características principales se incluyen:

- **Selección automática de suavidad:** el parámetro de suavizado de cada término se estima como parte del ajuste del modelo, mediante REML (máxima verosimilitud restringida), GCV (validación cruzada generalizada) o ML (máxima verosimilitud).
- **Múltiples tipos de base:** splines de regresión thin-plate (tp), splines de regresión cúbica (cr), P-splines (ps) y B-splines (bs), cada uno con propiedades distintas.
- **Términos suaves de producto tensorial:** `te()` y `ti()` permiten interacciones suaves entre variables sin incurrir en la maldición de la dimensionalidad.
- **Selección de variables:** la opción `select = TRUE` añade una penalización adicional que puede contraer términos suaves hasta cero, con lo que realiza de hecho una selección de variables.
- **Diagnósticos de concurvidad:** miden el análogo de la multicolinealidad para los términos suaves.

1.1.3. Comparación con MARS (earth) y elastic net (glmnet)

La aplicación earth y sus dos Post Processing Modules (módulos de posprocesamiento; glmnet y mgcv) utilizan motores de modelado distintos. La tabla siguiente resume las diferencias principales:

Característica	mgcv (GAM)	earth (MARS)	glmnet (Elastic Net)
Tipo de modelo	No lineal suave (splines penalizadas)	Lineal por tramos (splines adaptativas)	Lineal (con regularización)
No linealidad	Automática mediante funciones suaves (thin plate, cúbicas, etc.)	Automática mediante funciones bisagra con nudos guiados por los datos	Solo mediante interacciones o expansión de base
Selección de variables	Opcional mediante contracción de doble penalización (select=TRUE)	Incorporada mediante pasos hacia adelante/atrás con poda por GCV	Incorporada mediante penalización L_1 (Lasso)
Interacciones	Términos suaves de producto tensorial (te(), ti())	Productos basados en bisagras (interpretables, hasta grado 3)	Productos cruzados por pares (difíciles de interpretar)
Interpretabilidad	Curvas suaves de efecto parcial; muy intuitivas	Las funciones g muestran efectos parciales lineales por tramos para cada variable	Los coeficientes son pesos lineales; fáciles de explicar individualmente
Más indicado para	Relaciones no lineales suaves, modelado flexible	Detección automática de no linealidad e interacciones	Datos de alta dimensión, selección de variables, muestras pequeñas
Ajustes RCA	Ajustes suaves mediante efectos parciales	Ajustes lineales por tramos mediante funciones g	Ajustes lineales por variable
Control del sobreajuste	Estimación del parámetro de suavizado por REML/GCV	Poda basada en GCV	Lambda validada por validación cruzada

MARS (Multivariate Adaptive Regression Splines) fue introducido por Jerome Friedman en 1991. Construye modelos lineales por tramos seleccionando de forma adaptativa funciones bisagra y las posiciones de sus nudos a partir de los datos. La implementación en R es el paquete **earth** de Stephen Milborrow.

La regresión **elastic net**, implementada en el paquete **glmnet** de Friedman, Hastie y Tibshirani (2010), combina las penalizaciones Lasso (L_1) y ridge (L_2) para realizar simultáneamente la selección de variables y la contracción de los coeficientes.

1.1.4. La cadena earthUI–mgcvUI

Un flujo de trabajo potente combina el descubrimiento automático de nudos de earthUI con la estimación suave de mgcvUI. Cuando earthUI exporta un archivo de resultados **.rds** y mgcvUI lo importa:

1. **Earth descubre los nudos** — MARS encuentra posiciones de bisagra guiadas por los datos (puntos de cambio) para cada variable.
2. **mgcvUI las usa como anclas** — los nudos de earth se convierten en la base de las splines de regresión cúbica (base `cr`) del GAM, lo que da a los términos suaves una ventaja inicial con posiciones de nudo que reflejan la estructura real de los datos.
3. **mgcv refina el ajuste** — la estimación por splines penalizadas suaviza el ajuste lineal por tramos de earth y produce curvas suaves de efecto parcial, al tiempo que conserva la forma general descubierta por earth.

Esta cadena tiende un puente entre el modelado exploratorio (earth) y el modelado confirmatorio (GAM), y combina lo mejor de ambos enfoques.

1.1.5. Flujo de trabajo recomendado

Para la tasación inmobiliaria y aplicaciones similares que requieren modelos interpretables y defendibles:

1. **Empiece con earthUI** para descubrir las variables importantes, las relaciones no lineales y las interacciones presentes en sus datos. La selección automática de base de earth y sus funciones `g` ofrecen excelentes intuiciones iniciales.
2. **Refine con mgcvUI** si desea efectos parciales más suaves. Importe los nudos de earthUI en mgcvUI para pasar sin fricción de los modelos lineales por tramos a los modelos suaves.
3. **Use glmnetUI** cuando necesite una selección de variables agresiva (Lasso), cuando el número de predictores se aproxime o supere al número de observaciones, o cuando desee un modelo puramente lineal con regularización por motivos de defendibilidad.

1.2 ¿Qué es mgcvUI?

mgcvUI es una interfaz gráfica de usuario para el paquete `mgcv` de R. 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 un GAM de manera interactiva.

A partir de la versión actual, mgcvUI ya no es un paquete que se instale por separado: se distribuye *dentro* del paquete earthUI como su Post Processing Module de GAM, y se inicia con `earthUI::launch_mgcv()`. La aplicación en sí — y todo lo descrito en este artículo — no ha cambiado; solo difieren la instalación y el comando de inicio. Las tres rutinas (earth, glmnet, mgcv) comparten un mismo árbol de proyectos, una misma base de datos de configuración y una misma copia de la infraestructura de soporte.

La aplicación proporciona un flujo de trabajo completo: importación de datos, configuración de variables con especificación de términos suaves, ajuste del modelo con procesamiento en segundo plano, gráficos de diagnóstico, curvas suaves de efecto parcial e informes descargables en formato Word, PDF o HTML.

Los **modelos aditivos generalizados** sustituyen la relación lineal βx por una función suave $f(x)$ para cada predictor. Esto significa que:

- **No hace falta especificar formas funcionales** — son los datos los que determinan la forma de cada relación.
- **Efectos parciales suaves** — la contribución de cada predictor es una curva suave, no una recta, lo que facilita ver cómo una variable afecta a la respuesta en distintos valores.
- **Selección automática de suavidad** — mgcv determina cuán ondulada debe ser cada curva mediante REML o GCV, equilibrando el ajuste frente al sobreajuste.

1.3 Tres modos de uso

Todo proyecto tiene un **purpose** (propósito o modo de uso), que se elige al crearlo en la Sección 1 de la barra lateral (**Project**) y queda fijado a partir de entonces. El propósito del proyecto activo determina qué herramientas y elementos de la interfaz aparecen; para trabajar en otro modo, cree o abra un proyecto con ese propósito. Los tres modos son:

- **General** — regresión GAM para cualquier tipo de población o conjunto de datos. Este es el modo predeterminado. Ofrece el flujo de trabajo completo de GAM sin añadidos específicos de ningún dominio.
- **For Appraisal** (para tasación) — regresión GAM adaptada a la tasación inmobiliaria. Añade funciones específicas para la valuación de una sola propiedad, incluidos el tratamiento de la propiedad sujeto, las designaciones especiales de columnas y el Residual Constraint Approach (RCA).
- **Market Area Analysis** (análisis de zona de mercado) — regresión GAM adaptada a los estudios de zona de mercado. Añade funciones para analizar grupos de propiedades dentro de un mercado definido, con tratamiento opcional de la fila del sujeto.

En los tres modos el motor de modelado subyacente es idéntico — siempre se ajusta un GAM mediante `mgcv::gam()`. El ajuste del propósito controla qué herramientas y elementos de interfaz adicionales están disponibles.

1.4 Funciones específicas del sector inmobiliario

Cuando se selecciona **For Appraisal** o **Market Area Analysis**, mgcvUI activa varias funciones diseñadas para el análisis inmobiliario:

- **Designaciones especiales de columnas** — cada predictor puede etiquetarse con un papel 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 trata la columna durante el ajuste y en las salidas.
- **Column Sale Age** — cuando una columna se designa como `contract_date` y se proporciona una fecha efectiva, mgcvUI calcula una columna `sale_age` (días entre la fecha de venta y la fecha efectiva) y la sustituye como predictor.
- **Cálculos RCA (solo en tasación)** — en el modo de tasación, después de ajustar el modelo mgcvUI puede calcular la salida del Residual Constraint Approach (RCA), que produce ajustes por comparable, resúmenes de ajuste neto/bruto y un precio de venta ajustado para la propiedad sujeto.

***Consejo:** El modo General funciona con cualquier conjunto de datos — financiero, científico, de ingeniería o inmobiliario. Los modos de tasación y de mercado simplemente añaden funciones de conveniencia para los profesionales del sector inmobiliario.*

1.5 Primeros pasos

Para usar mgcvUI:

1. **Instale earthUI** en R: `install.packages("earthUI")` — la rutina de mgcv está incluida. (Véase “Instalación de earthUI y vProlog” en el artículo sobre earthUI de este número para la versión de desarrollo y los componentes opcionales).
2. **Inicie la aplicación:** ejecute `earthUI::launch_mgcv()` en R. La aplicación se abre en su navegador web en el puerto 7880.
3. **Cree o abra un proyecto** en la Sección 1 de la barra lateral. Un proyecto fija el propósito (General, For Appraisal o Market Area Analysis) y la ubicación de trabajo (país, estado, condado, ciudad), así como sus carpetas de entrada y salida. Los proyectos se almacenan

bajo una raíz **regProj** compartida y se comparten con las aplicaciones hermanas earthUI y glmnetUI. Abrir un proyecto con un propósito distinto es la forma de cambiar de modo.

4. **Importe sus datos** en la Sección 2 seleccionando un archivo CSV o Excel de la carpeta de entrada del proyecto activo. Añada archivos a esa carpeta y haga clic en **Refresh** (actualizar) para listarlos.
5. **Importe desde earthUI (opcional)** — la Sección 3 permite importar un archivo **.rds** de resultados de earthUI para sembrar el GAM con los nudos descubiertos por earth. Omita este paso si no dispone de un resultado de earthUI.
6. **Configure las variables** — elija su variable objetivo y sus predictores, fije los tipos de datos, asigne los papeles especiales de columna y marque las variables como factor o lineal.
7. **Fije los parámetros de mgcv** — family, method, gamma, tipo de base, k, tipo de tensor, interacciones y otras opciones.
8. **Ajuste el modelo** — haga clic en “Fit Mgcvm GAM Model” (ajustar el modelo GAM de mgcv) y revise los resultados en el panel principal.
9. **Exporte** — descargue las predicciones en Excel, genere y convierta un informe Quarto o (en modo de tasación) calcule los ajustes RCA y una cuadrícula de comparación de ventas.

Los ajustes se guardan automáticamente en una base de datos SQLite y se restauran cuando vuelve a cargar el mismo archivo de entrada.

2 Requisitos de los datos de entrada del MLS

Para los flujos de trabajo de tasación inmobiliaria y análisis de mercado, sus datos de entrada provienen normalmente de una exportación de un Multiple Listing Service (MLS, servicio de listado múltiple). Este capítulo describe la estructura de archivo esperada y las columnas que mgcvUI puede utilizar.

2.1 Formato y estructura del archivo

mgcvUI acepta archivos **CSV** y **Excel** (.xlsx, .xls). Al importarlos, los nombres de las columnas se convierten automáticamente a **snake_case** — por ejemplo, “Living SqFt” se convierte en **living_sqft**, “Contract Date” en **contract_date** y “Sale Price” en **sale_price**. Esta normalización garantiza referencias de columna coherentes a lo largo de todo el flujo de trabajo. El separador de CSV y la marca decimal utilizados durante la importación vienen determinados por la configuración regional (véase el capítulo 3, “Configuración regional y de idioma”).

Su archivo de datos debe ser una tabla plana con una fila por propiedad y una columna por atributo. La primera fila del archivo debe contener los encabezados de columna.

2.2 Columnas requeridas en el modo de tasación

Aunque mgcvUI funciona con cualquier conjunto de columnas, el flujo de trabajo completo de tasación se beneficia de disponer de las siguientes columnas:

Columna	Tipo especial	Finalidad
Sale Price	(objetivo)	Variable de respuesta del modelo
Contract Date	contract_date	Se usa para calcular sale_age (días desde la fecha efectiva)
Listing Date	listing_date	Se usa 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 las exportaciones
Concessions	concessions	Concesiones de venta; Net SP = Sale Price – Concessions
Living Area (SF)	living_area	Habilita los residuos por pie cuadrado (residual_sf , cqa_sf)
Lot Size	lot_size	Columna de superficie del terreno
Site Dimensions	site_dimensions	Agrupada con el tamaño del terreno
Latitude	latitude	Redondeada a 3 decimales
Longitude	longitude	Redondeada a 3 decimales
Area ID	area	Identificador de zona de mercado o vecindario
Actual Age	actual_age	Edad de la propiedad
Effective Age	effective_age	Edad efectiva de la propiedad
Address	display_only	Se muestra en las exportaciones; se excluye del modelo

Los nombres de las columnas de la hoja de cálculo pueden estar en otro idioma — los nombres “especiales” están en inglés para que el programa de R pueda darles un tratamiento especial. Por lo demás, los nombres de columna proporcionados aparecen en los modelos de regresión, en los gráficos y, si se realizan tasaciones, en los informes de salida.

No todas las columnas son obligatorias. mgcvUI se adapta — si falta una columna, la función correspondiente simplemente se omite. Sin embargo, para los modelos de precios inmobiliarios se recomiendan especialmente ciertas columnas:

1. **Sale Age** — el número de días transcurridos entre la fecha de contrato de venta y la fecha efectiva. Si se utiliza un historial de ventas de varios años, **sale_age** suele desempeñar un papel central.
2. **Living Area** — también recibe nombres como “Living Sqft” o “GLA” (gross living area, superficie habitable bruta).
3. **Total Bath Count** — el número total de baños completos, de un cuarto, de medio baño y de tres cuartos.
4. **Garage Bays** o **Garage Area** — plazas de garaje o superficie en pies cuadrados.
5. **Lot Size** — la superficie de terreno de la propiedad.
6. **Longitude**, **Latitude** y, si está disponible, **Area ID** para la variación geográfica de precios.

2.3 Convenciones de nombres de columnas especiales

mgcvUI identifica las columnas por su **designación de tipo especial**, no por su nombre de columna. Puede dar a sus columnas el nombre que desee en la exportación del MLS — lo que importa es que asigne el tipo especial correcto en la tabla de configuración de variables (capítulo 6).

2.4 Calidad e integridad de los datos

- **Valores faltantes (NA)**: las filas con valores NA en cualquier columna de predictor o de objetivo se eliminan automáticamente antes del ajuste. mgcvUI reduce los datos únicamente a las columnas utilizadas en el modelo, de modo que los NA en columnas no utilizadas no provocan la pérdida de filas.
- **Columnas numéricas**: el precio de venta, la superficie habitable, el tamaño del terreno y campos similares deben ser numéricos.
- **Columnas de factor**: las columnas de texto, TRUE/FALSE y de tipo factor de R se tratan automáticamente como categóricas — solo pueden entrar en el modelo como factores. Una columna **numérica** (por ejemplo, un código numérico de zona o de estilo) se convierte en factor únicamente cuando se marca su casilla **Factor**; mgcvUI nunca infiere el carácter categórico a partir del rango de valores de una columna numérica, de modo que predictores numéricos discretos como **bath_count** siguen siendo continuos (suavizados). El papel especial es independiente de la decisión sobre el factor.

***Consejo:** Revise la columna NA de la tabla de predictores después de la importación. Las columnas con muchos valores faltantes pueden provocar la eliminación de filas. Las columnas con $\geq 50\%$ de NA se marcan en rojo y se excluyen automáticamente.*

2.5 Ubicación de la fila del sujeto

En el modo **Appraisal**, la fila 1 debe ser la **propiedad sujeto**. Todas las filas restantes son ventas comparables. La fila del sujeto se excluye del ajuste del modelo. Después del ajuste, el modelo genera de todos modos predicciones para la fila del sujeto.

En el modo **Market Area Analysis**, colocar el sujeto en la fila 1 es opcional.

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

3 Modo de uso general

3.1 Panorama general

El modo de uso general es el predeterminado al iniciar mgcvUI. Ofrece el flujo de trabajo completo de GAM para cualquier conjunto de datos — no solo inmobiliarios. Puede usar mgcvUI para datos científicos, análisis financiero, estudios de ingeniería o cualquier problema de regresión en el que quepa esperar relaciones no lineales suaves.

En el modo General, la interfaz omite las funciones específicas del sector inmobiliario (columnas especiales, antigüedad de la venta, 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 se organiza en secciones numeradas y plegables que le guían desde la selección del proyecto hasta la exportación del informe. La mayoría de las secciones aparecen solo una vez abierto un proyecto activo, y varias son específicas del propósito del proyecto. La carpeta de salida se deriva del proyecto activo — no hay un campo aparte para la carpeta de salida.

1. Project — crear o abrir un proyecto. Un proyecto fija el propósito y la ubicación de trabajo (país, estado, condado, ciudad) y sus carpetas de entrada (`<os>_in/`) y de salida. Los proyectos se almacenan bajo una raíz **regProj** compartida (por omisión `~/regProj`, modificable con la variable de entorno `REGPROJ_ROOT` o desde el desplegable Settings) y se comparten con las aplicaciones hermanas **earthUI** y **glmnetUI**, de modo que los modelos de las tres conviven lado a lado. Haga clic en **Close Project** (cerrar proyecto) para cambiar.

2. Import Data — seleccionar un archivo CSV o Excel de la carpeta de entrada del proyecto activo. Añada archivos a la carpeta mediante el Finder o el Explorador, y haga clic en **Refresh** para listarlos. Para los archivos de Excel con varias hojas aparece un selector de hoja. Los nombres de columna se convierten automáticamente a `snake_case`.

3. Import from earthUI (opcional) — importar un archivo `.rds` de resultados de **earthUI** para sembrar el GAM con las posiciones de nudo descubiertas por **earth** (y para sembrar las selecciones de variables). Este paso es opcional — omítalo si no dispone de un resultado de **earthUI** que importar.

4. Variable Configuration — selector de la variable objetivo, transformación de la respuesta (none, log, log10) y tabla de predictores con casillas para Include, Factor y Linear. La columna Special aparece solo en los modos Appraisal y Market. Véase el capítulo 6 para todos los detalles.

5. MgcV Call Parameters — toda la configuración del modelo: preajustes de parámetros, family, method, gamma, validación cruzada, select, tipo de base, k, tipo de tensor, matriz de interacciones y parámetros avanzados. Véase el capítulo 7 para la referencia completa de parámetros.

6. Fit MgcV GAM Model — el botón que ejecuta el modelo. Los resultados aparecen en las pestañas de la derecha.

7. Download Output — exporta las predicciones, los residuos, las puntuaciones CQA y las contribuciones por variable como archivo de Excel a la carpeta de salida del proyecto. Disponible en todos los modos de uso una vez ajustado el modelo.

8. Calculate RCA Adjustments & Download — calcula los ajustes RCA por variable para cada comparable e interpola el residuo del sujeto mediante CQA. Aparece solo en los modos Appraisal y Market, después del ajuste.

9. Generate Sales Grid & Download — genera la Sales Comparison Grid (cuadrícula de comparación de ventas), recomendando comparables según el porcentaje de ajuste bruto. Aparece solo en el modo Appraisal.

10. Generate Quarto Report — escribe un paquete de informe Quarto (`.qmd`) autocontenido — con todos los gráficos y tablas — en la carpeta de salida del proyecto. Disponible una vez ajustado el modelo.

11. Convert Quarto Report — renderiza el paquete `.qmd` generado a HTML, Word o PDF. El PDF requiere una instalación de LaTeX; si no se detecta ninguna, la opción de PDF queda oculta. Véase el capítulo 13.

Los pasos específicos de cada propósito se omiten (y los pasos restantes se reenumeran) en los demás modos: en el modo Market falta el paso de la Sales Grid, y en el modo General faltan tanto el paso de RCA como el de la Sales Grid.

3.3 Importar desde earthUI (opcional)

La Sección 3 de la barra lateral — **Import from earthUI** — permite importar un archivo `.rds` de resultados de earthUI. Esto habilita la cadena `earth-mgcv`: las posiciones de nudo de earth, guiadas por los datos, se convierten en puntos de anclaje para los términos suaves del GAM.

Cuando se importa un resultado de earthUI, `mgcvUI`:

- Muestra un resumen con la variable objetivo, el R cuadrado, el número de términos y las posiciones de nudo por variable.
- Selecciona automáticamente el preajuste de parámetros “Earth Pipeline” (base de spline de regresión cúbica, $\gamma = 1.4$).
- Fija la variable objetivo para que coincida con la de earth.
- Marca previamente Include para las variables predictoras de earth.
- Marca previamente Factor para las variables categóricas de earth y Linear para los predictores lineales de earth.
- Rellena previamente las interacciones detectadas por earth (resaltadas en amarillo en la matriz de interacciones y bloqueadas frente a cambios manuales).
- Ofrece un botón **Export Knots CSV** para descargar los nudos de earth como archivo CSV.

La tabla de configuración de variables muestra entonces una nota que explica que las selecciones de variables (Include, Linear, Factor) y los nudos de las splines se *sembraron* a partir del modelo de earth. A diferencia de `glmnetUI`, donde el ajuste de `glmnet` queda ligado a la base de earth, `mgcvUI` ajusta el GAM a partir de sus selecciones actuales — earth solo proporciona una ventaja inicial, y todos los controles siguen siendo editables.

Un botón **Clear** elimina los datos importados de earth y restablece el modo autónomo.

3.4 Pestañas del panel principal

Tras la importación de los datos, el panel principal ofrece las siguientes pestañas (las que dependen del modelo se llenan después del ajuste):

Pestaña	Contenido
Data	DataTable interactiva con los datos importados. En los modos de tasación y de mercado se divide en tablas de propiedad sujeto y de ventas comparables.
Equation	Fórmula completa del GAM representada con MathJax, más una tabla con las definiciones de las funciones suaves.
Summary	Métricas clave (R^2 , R^2 de CV, desviación explicada, AIC, n, número de términos suaves) y tablas de términos suaves y paramétricos.
Variable Importance	Gráfico de barras horizontales de los estadísticos F (términos suaves) y de los valores $ t $ (paramétricos), más una tabla ordenada.
Contribution	Curvas suaves de efecto parcial interactivas en plotly, con bandas de confianza y marcadores de los nudos de earth.
Correlation	Mapa de calor de las correlaciones entre predictores numéricos (disponible antes del ajuste).
Diagnostics	Residuos frente a ajustados, gráfico Q-Q, histograma, respuesta frente a ajustados (mediante gratia). Diagrama de dispersión de valores reales frente a predichos.
RCA Adjustments	Histogramas del análisis RCA (modo de tasación o de mercado, tras el cálculo).
Anova	Tabla ANOVA combinada de los términos paramétricos y suaves.
Mgcv Output	Texto sin formato: tiempos, fórmula, resumen del modelo, familia, concurvidad.
Sign Check	Comprobación de la coherencia de dirección entre earth y el GAM (cuando se ha importado earth).
Concurvity	Matrices de concurvidad global y por pares con codificación por colores.
Basis Check	Pruebas de <code>gam.check()</code> sobre la adecuación de la dimensión de la base.

3.5 Persistencia de los ajustes

mgcvUI guarda automáticamente su configuración en una base de datos SQLite, indexada por el nombre del archivo de entrada. Cuando vuelve a cargar el mismo archivo se restauran todos los ajustes: selección del objetivo, casillas de los predictores, tipos de datos, parámetros de mgcv y matriz de interacciones. El propósito es una propiedad del proyecto activo y no un ajuste global; el último archivo de entrada utilizado se recuerda por proyecto y se carga automáticamente cuando el proyecto se vuelve a abrir.

3.6 Modo oscuro

Haga clic en el icono de luna/sol de la esquina superior derecha para alternar entre los temas Nord Light y Nord Dark. La preferencia de tema se guarda en el almacenamiento local y persiste entre sesiones. Todos los elementos de la interfaz se adaptan al tema seleccionado.

3.7 Configuración regional y de idioma

mgcvUI admite convenciones internacionales de formato numérico y de CSV mediante un sistema de configuración regional basado en el país. El desplegable **Settings** de la barra de título ofrece selectores **Country** y **Paper** para más de 30 países admitidos. Cada preajuste configura:

- **Separador de CSV** — coma (,) para EE. UU., Reino Unido y Japón, o punto y coma (;) para gran parte de Europa.
- **Signo decimal** — punto (.) o coma (,).
- **Separador de millares** — coma (EE. UU./Reino Unido/Japón), punto (Alemania/Italia/España), espacio (Finlandia/Francia/Polonia/Bálticos) o apóstrofo (Suiza).

- **Tamaño de papel** — Letter (EE. UU./Canadá/México) o A4 (en el resto de los casos).

Haga clic en **Save** para almacenar globalmente sus preferencias regionales. La sección Import Data dispone además de un selector **Locale** por archivo para importaciones puntuales que usen convenciones distintas. El mismo desplegable Settings contiene el campo de la **carpeta raíz regProj**, que reubica el árbol de proyectos compartido que usan mgcvUI, earthUI y glmnetUI.

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

4 Modo de tasación

Cuando el propósito del proyecto activo es **For Appraisal**, mgcvUI se configura para la valuación de una sola propiedad. Todas las funciones descritas en el capítulo 3 siguen disponibles; este capítulo trata únicamente los añadidos específicos de la tasación.

4.1 Tratamiento de la fila del sujeto

En el modo de tasación, **la fila 1 de su conjunto de datos es la propiedad sujeto** y todas las filas restantes son ventas comparables. Su archivo de entrada debe organizarse en consecuencia (véase el capítulo 2). El precio de venta del sujeto puede dejarse en blanco o fijarse en cualquier valor — mgcvUI lo trata automáticamente como NA durante el ajuste.

Tras el ajuste, el modelo genera predicciones para la fila del sujeto, que constituyen la base del flujo de trabajo de ajustes RCA.

4.2 Fecha efectiva y antigüedad de la venta

En los modos de tasación y de mercado aparece un campo **Effective Date** en la sección de configuración de variables (con la fecha de hoy por omisión). Si designa una columna como **contract_date** en el desplegable Special, mgcvUI calcula una columna **sale_age** — el número de días enteros entre la fecha de contrato de cada venta y la fecha efectiva. Esta columna se añade como predictor.

Cuando cambia la fecha efectiva, **sale_age** se recalcula automáticamente.

4.3 Designaciones especiales de columnas

En los modos de tasación y de mercado aparece un desplegable **Special** para cada predictor de la tabla de configuración de variables. Véase el capítulo 6 para la referencia completa de los tipos especiales y sus efectos.

4.4 Panorama de los ajustes RCA

El botón **Calculate RCA Adjustments & Download** (visible en los modos de tasación y de mercado tras el ajuste) calcula ajustes derivados del mercado para cada comparable respecto del sujeto. El flujo de trabajo RCA completo se describe en el capítulo 11.

***Consejo:** La puntuación CQA que asigne al sujeto controla qué parte de la distribución de residuos se le atribuye. Una puntuación de 5.00 sitúa al sujeto en la mediana de los comparables.*

5 Modo de análisis de zona de mercado

Cuando el propósito del proyecto activo es **Market Area Analysis**, mgcvUI ofrece las mismas funciones específicas del sector inmobiliario que el modo de tasación (columnas especiales, antigüedad de la venta), pero está orientado a analizar un grupo de propiedades en lugar de valorar un único sujeto.

5.1 Diferencias respecto del modo de tasación

- **El tratamiento de la fila del sujeto es flexible** — en el modo de mercado, la fila 1 no se trata automáticamente como propiedad sujeto.
- **Sin Sales Grid** — el paso “Generate Sales Grid & Download” no está disponible.

5.2 Cuándo usar el modo de mercado

El modo de análisis de zona de mercado es apropiado cuando usted:

- Construye un modelo GAM para un vecindario o zona de mercado con el fin de comprender los factores no lineales que determinan el valor
- Analiza cómo variables como la superficie, la antigüedad, el tamaño del terreno y la ubicación afectan a los precios de venta en un grupo de propiedades
- Prepara el respaldo para un análisis de condiciones de mercado o una delimitación de vecindario
- Trabaja con un conjunto de datos en el que desea las funciones de columnas especiales (antigüedad de la venta) pero no necesita el flujo de trabajo de la cuadrícula de ventas

***Consejo:** El modo de mercado también resulta útil para la regresión inmobiliaria general cuando se desean las funciones de columnas especiales pero no se necesita el flujo de trabajo de valuación de una sola propiedad.*

6 Selección de variables

La Sección 4 de la barra lateral — **Variable Configuration** — es donde se elige qué columnas participan en el modelo y cómo se tratan.

6.1 Variable objetivo

El desplegable **Target (response) variable** situado en la parte superior de la Sección 4 lista todas las columnas numéricas de su conjunto de datos. Seleccione una columna como variable de respuesta (por ejemplo, el precio de venta). La columna objetivo se excluye automáticamente de la lista de predictores.

6.2 Transformación de la respuesta

Debajo del selector de objetivo, un desplegable **Response Transform** ofrece tres opciones:

- **None (raw)** — usar la respuesta tal cual. Apropiado cuando la relación entre los predictores y la respuesta es aproximadamente aditiva en la escala original.
- **Log (natural)** — aplicar la transformación logarítmica natural antes del ajuste. Al deshacer la transformación, esto hace que los ajustes sean proporcionales (%) en lugar de absolutos (\$). Resulta especialmente útil para los precios de venta, donde un ajuste de \$10,000 significa cosas muy distintas a distintos niveles de precio.
- **Log10** — aplicar la transformación logarítmica en base 10.

***Consejo:** La transformación logarítmica hace que los ajustes temporales sean proporcionales (%) en lugar de absolutos (\$). En los modelos inmobiliarios que abarcan un amplio rango de precios, las transformaciones logarítmicas suelen producir residuos mejor comportados.*

Cuando se selecciona una transformación logarítmica, los valores ≤ 0 de la respuesta se filtran automáticamente. Todas las salidas (predicciones, contribuciones, residuos) se retransforman automáticamente a la escala original.

6.3 La tabla de predictores

Debajo del selector de objetivo, una tabla desplazable lista todas las columnas restantes. Las columnas de casillas usan encabezados verticales rotados para mayor compacidad. El orden de las columnas es:

Columna	Descripción
Variable	Nombre de la columna (fuente monoespaciada, truncado con puntos suspensivos). Celda con borde.
Type	Desplegable de tipo de dato: numeric, integer, character, factor, logical, Date, POSIXct. Desplegable con borde.
Include	Casilla — incluir esta columna como predictor en el modelo. Encabezado vertical.
Factor	Casilla — tratar como factor categórico (crea un coeficiente por nivel). Encabezado vertical.
Linear	Casilla — forzar una relación lineal (sin término suave, solo βx). Encabezado vertical.
Special	Desplegable (solo en los modos de tasación y de mercado) — véase más abajo la referencia de tipos especiales de columna. Desplegable con borde.
NAs	Recuento y porcentaje de valores faltantes, con código de color: rojo ($\geq 50\%$), naranja ($\geq 20\%$), gris (>0). Celda con borde.

Factor frente a suave frente a lineal: por omisión, las variables numéricas incluidas reciben un término suave $f(x)$. Marcar **Factor** crea un término categórico (un coeficiente por nivel). Marcar **Linear** crea un término lineal simple βx en lugar de un término suave. Las variables marcadas a la vez como Factor y Linear se tratan como Factor.

6.4 Detección de tipos de datos y anulaciones

mgcvUI detecta automáticamente los tipos de datos en la importación. Se reconocen las columnas numéricas, enteras, lógicas, de factor y de fecha. Puede anular cualquier detección cambiando el desplegable **Type**. Cambiar los tipos afecta a cómo se trata la columna en el modelo.

6.5 Referencia de tipos especiales de columna

En los modos de tasación y de mercado, el desplegable **Special** ofrece las siguientes opciones:

Ponderación:

- **weight** — columna de peso de las observaciones (solo se permite una; las filas con peso = 0 se excluyen del ajuste)

Tipos de fecha y hora:

- **contract_date** — activa el cálculo automático de **sale_age** a partir de la fecha efectiva
- **sale_age** — identifica una columna de antigüedad de venta ya calculada (días desde la fecha efectiva), que se usa tal cual
- **listing_date** — se usa como alternativa para calcular los días en el mercado
- **dom** — identifica la columna de días en el mercado

Tipo de registro:

- **sale_type** — identifica la columna de tipo de registro que distingue el sujeto de las ventas comparables

Tipos monetarios:

- **concessions** — identifica las concesiones de venta (créditos del vendedor, incentivos al comprador, etc.)

Tipos de tamaño y ubicación:

- **latitude** — los valores se redondean automáticamente a 3 decimales
- **longitude** — mismo tratamiento de redondeo que la latitud
- **area** — identificador de zona de mercado o vecindario
- **living_area** — habilita los cálculos de residuos por pie cuadrado (**residual_sf** y **cqa_sf**)
- **lot_size** — columna de superficie del terreno
- **site_dimensions** — agrupada con el tamaño del terreno

Tipos de edad:

- **actual_age** — columna de edad de la propiedad
- **effective_age** — edad efectiva de la propiedad

Tipos de visualización:

- **display_only** — la columna se incluye en las exportaciones a Excel pero se excluye por completo del ajuste del modelo. Úselo para campos de dirección, números de MLS, identificadores de parcela u otros datos de referencia.

***Consejo:** Las columnas designadas como **display_only** permanecen en su conjunto de datos y aparecen en la exportación a Excel, pero quedan excluidas por completo de la tabla de predictores. Úselo para columnas de identificador, direcciones u otros campos de referencia.*

7 Selección de parámetros

La Sección 5 de la barra lateral — **Mgcv Call Parameters** — da acceso a todas las opciones de configuración del GAM. Cada parámetro tiene un icono azul de ayuda (?) con una explicación emergente.

7.1 Preajustes de parámetros

Un desplegable situado en la parte superior ofrece dos preajustes que configuran valores predeterminados razonables para los flujos de trabajo habituales:

Preajuste	Base	Ajustes
Standalone (descubrimiento)	tp (thin plate)	k = auto (10), gamma = 1.2, select = TRUE
Earth Pipeline (refinamiento)	cr (regresión cúbica)	k = auto (10), gamma = 1.4, select = TRUE

El preajuste **Earth Pipeline** se selecciona automáticamente cuando se importan nudos de earthUI. La base de spline de regresión cúbica (cr) es necesaria para la integración de los nudos de earth porque permite especificar posiciones de nudo exactas.

7.2 Family

Elija la familia de distribución para su variable de respuesta:

Familia	Caso de uso
gaussian	Respuestas continuas (por ejemplo, el precio de venta). La más común.
Gamma	Respuestas continuas positivas con varianza proporcional a la media.
poisson	Datos de recuento (por ejemplo, el número de ventas).
binomial	Resultados binarios.
inverse.gaussian	Datos positivos con fuerte asimetría a la derecha.

7.3 Method

El método de estimación del parámetro de suavizado:

Método	Descripción
REML	Máxima verosimilitud restringida (predeterminado, recomendado). El más robusto frente al sobreajuste.
GCV.Cp	Validación cruzada generalizada. Tiende a seleccionar modelos algo más complejos.
ML	Máxima verosimilitud. Similar a REML, pero puede suavizar de menos.
P-REML	Variante REML de Pearson.
P-ML	Variante ML de Pearson.
GACV.Cp	Validación cruzada aproximada generalizada.

7.4 Gamma (multiplicador de suavizado)

El parámetro gamma multiplica los grados de libertad efectivos en el criterio de suavizado, lo que favorece ajustes más suaves (menos ondulados). El valor predeterminado es 1.2.

Valor	Efecto
1.0	Suavizado estándar
1.2–1.4	Algo más suave que el predeterminado (recomendado)
2.0+	Mucho más suave — adecuado para datos ruidosos o muestras pequeñas

Los valores más altos de gamma protegen frente al sobreajuste al penalizar la complejidad con mayor severidad. El preajuste Earth Pipeline usa 1.4 para suavizar más al refinar los nudos de earth.

7.5 Validación cruzada

Cuando está marcada (por omisión), mgcvUI ejecuta una validación cruzada de 10 pliegues tras el ajuste para calcular un R cuadrado de validación cruzada. Esto proporciona una estimación honesta de la capacidad predictiva fuera de la muestra.

7.6 Select (contracción)

Cuando está marcada (por omisión), añade una penalización adicional a cada término suave que puede contraerlo hasta cero. Esto permite la selección automática de variables — los términos suaves poco importantes quedan efectivamente eliminados del modelo. Se implementa mediante `select = TRUE` en `gam()`.

7.7 Base predeterminada

El tipo de función base de spline para los términos suaves:

Base	Nombre completo	Descripción
tp	Spline de regresión thin plate	Predeterminada. Isotrópica (no hace falta colocar nudos). Buena elección de propósito general.
cr	Spline de regresión cúbica	Permite colocar nudos de forma explícita. Necesaria para integrar los nudos de earth.
ps	P-spline	Base de B-splines con nudos equiespaciados y penalización de diferencias. Computacionalmente eficiente.
bs	B-spline	Base de B-splines flexible.

7.8 k predeterminada (dimensión de la base)

La dimensión de la base k controla la ondulación máxima de cada término suave. Un valor de 0 (el predeterminado) significa “automático” — mgcvUI usa $k = 10$ o el número de nudos de earth, según lo que corresponda.

- **k pequeña (3–5):** curvas muy suaves, casi lineales.
- **k media (8–15):** flexibilidad moderada (suele ser suficiente).

- **k grande (20+)**: muy flexible — riesgo de sobreajuste sin una regularización fuerte.

Consejo: Los grados de libertad efectivos (EDF) de la pestaña Summary indican cuántos grados de libertad usa realmente cada término suave. Si el EDF se acerca a $k - 1$, la base puede ser demasiado pequeña — aumente k . Si el EDF es mucho menor que k , el término suave está bien penalizado.

7.9 Tipo de tensor

Para las interacciones entre variables continuas hay disponibles dos tipos de producto tensorial:

Tipo	Función	Descripción
ti	ti(x1, x2)	Interacción tensorial — modela solo el efecto de interacción, y los efectos principales se tratan por separado mediante términos $s()$ univariantes. Preferible por su interpretabilidad.
te	te(x1, x2)	Producto tensorial — modela todo el efecto conjunto, incluidos los efectos principales. Más difícil de descomponer para los ajustes RCA.

7.10 Interacciones permitidas

Una sección plegable **Allowed Interactions** muestra una matriz triangular superior de casillas para todos los predictores suaves incluidos (no lineales y no factores). Cada casilla habilita un término suave de producto tensorial entre el par de variables correspondiente.

- **Allow All / Clear All** — botones de activación masiva.
- **Haga clic en el nombre de una variable** — alterna todas las interacciones de esa variable.
- **Haga clic con el botón derecho en el nombre de una variable** — “Block from main effect” — marca la variable para uso exclusivo en interacciones (sin término suave univariante).

Cuando se importan nudos de earthUI, las interacciones detectadas por earth quedan premarcadas y bloqueadas (resaltadas con fondo amarillo). Si el modelo de earth usó $\text{degree} = 1$ (sin interacciones), se muestra un mensaje informativo.

Advertencia

Habilitar interacciones tensoriales aumenta considerablemente la complejidad del modelo. Use interacciones cuando disponga de conocimiento del dominio o de evidencia detectada por earth que las respalde. Verifique que los ajustes resultantes sean razonables y explicables.

7.11 Interacciones factor por término suave

Aparece una matriz aparte de **Factor-by-Smooth Interactions** cuando se incluyen a la vez variables de factor y variables suaves. Cada casilla crea un término $s(x, \text{by} = \text{factor})$ — una curva suave distinta para cada nivel del factor.

7.12 Parámetros avanzados

Una sección plegable “Advanced” expone ajustes adicionales:

- **Optimizer** — algoritmo de estimación del parámetro de suavizado: `outer/newton` (predeterminado), `outer/bfgs` o `efs` (Fellner-Schall extendido).
- **Scale** — parámetro de escala. 0 (predeterminado) = estimar a partir de los datos. Fíjelo en un valor conocido para modelos de escala fija.
- **Discrete covariate method** — ajuste discretizado rápido para conjuntos de datos grandes (`discrete = TRUE`).
- **Threads** — número de hilos paralelos (1–32). Solo tiene efecto cuando `discrete = TRUE`.

8 Ajuste del modelo

8.1 El botón de ajuste

La Sección 6 de la barra lateral contiene el botón **Fit MgcV GAM Model**. Al hacer clic en él se ejecuta el modelo con su configuración actual.

8.2 Qué ocurre durante el ajuste

Cuando hace clic en Fit, mgcvUI:

1. **Valida la configuración** — comprueba que haya al menos un predictor y al menos 10 filas completas. Advierte si menos del 50 % de las filas están completas (indicando las columnas responsables de la mayoría de los NA).
2. **Prepara los datos** — reduce los datos únicamente a las columnas relevantes para el modelo, elimina las filas con NA, aplica los pesos (si se han designado), excluye la fila del sujeto (en el modo de tasación) y aplica las transformaciones de la respuesta.
3. **Construye la fórmula del GAM** — construye términos suaves $s(x, bs="cr", k=10)$ para los predictores numéricos, términos de factor para los categóricos, términos lineales para las variables marcadas como Linear y productos tensoriales para las interacciones seleccionadas. Los nudos de earth se incorporan como posiciones de nudo explícitas en los términos con base **cr**.
4. **Trata los casos límite** — descarta las variables constantes, convierte los predictores binarios en factores, limita k al número de valores únicos menos 1 y completa los nudos de earth con cuantiles de los datos cuando hacen falta más nudos.
5. **Ajusta el GAM** — llama a `mgcv::gam()` con la fórmula construida y los parámetros family, method, gamma, select y demás.
6. **Ejecuta la validación cruzada** (si está habilitada) — CV de 10 pliegues que calcula el R cuadrado fuera de la muestra.
7. **Actualiza todas las pestañas de resultados** — se llenan Summary, Equation, los gráficos de contribución, Diagnostics y todas las demás pestañas.

8.3 Ajuste en segundo plano y progreso

Cuando el paquete **callr** está disponible, el ajuste del modelo se ejecuta en un proceso en segundo plano para que la aplicación siga respondiendo. Una superposición de progreso de estilo terminal y tema oscuro muestra:

- Un **cronómetro** con los segundos y minutos transcurridos.
- Un **registro de traza con código de color**: azul para el estado, amarillo para el progreso de los pliegues de CV, rojo para los errores y verde para los resultados.
- Un botón **Abort** para detener el proceso de ajuste si es necesario.
- Un botón **Close** que aparece al finalizar.

Si **callr** no está disponible, el ajuste se ejecuta de forma síncrona (la aplicación se congela hasta que termina).

Tras un ajuste correcto aparece una marca de verificación blanca en el botón Fit y una línea de estado muestra: “R-sq = X, CV R-sq = X, Dev = X %, AIC = X, n = X”.

9 Pestañas de resultados

9.1 Data

Muestra los datos importados en una DataTable interactiva con desplazamiento horizontal y 15 filas por página. En los modos de tasación y de mercado, la pestaña se divide en una tabla de **Subject Property** (fila 1) y una tabla de **Comparable Sales** (las filas restantes); en el modo General se muestra una sola tabla.

Cada celda se mantiene en una única línea y se trunca al ancho de la columna, de modo que los campos de texto libre extensos (como las observaciones sobre la propiedad) no estiren una fila a lo largo de la página. Un clic simple selecciona una fila; **haga doble clic en cualquier celda** para abrir una ventana emergente con su contenido completo sin truncar.

9.2 Equation

Muestra el modelo GAM ajustado en dos partes:

Ecuación del modelo (representada con MathJax):

- El lado izquierdo muestra la respuesta con la transformación correspondiente (log, log10) y la función de enlace.
- El lado derecho muestra el valor del intercepto, los términos suaves como $f_i(\text{variable})$, los términos de factor agrupados por variable con el recuento de niveles y los términos lineales con los valores de sus coeficientes.
- La familia y el método se muestran encima de la ecuación.

Tabla de **Smooth Function Definitions** (definiciones de las funciones suaves):

- Cadena con la fórmula del término (por ejemplo, `s(living_sqft, bs="cr", k=8)`)
- Nombre de la variable
- Tipo (s, te, ti o lineal)
- Tipo de función base
- Valor de k
- Posiciones de los nudos de earth (si se han importado)

9.3 Summary

Seis tarjetas de métricas mostradas en la parte superior:

- **R-squared** — proporción de desviación explicada
- **CV R-squared** — R cuadrado de validación cruzada (si la CV está habilitada)
- **Dev. Explained %** — porcentaje de desviación explicada
- **AIC** — criterio de información de Akaike
- **n** — número de observaciones utilizadas
- **Smooths** — número de términos suaves

Debajo de las tarjetas:

- **Tabla Smooth Terms** — término, EDF (grados de libertad efectivos), Ref.df, estadístico F, valor p
- **Tabla Parametric Terms** — término, estimación, error estándar, valor t, valor p

9.4 Variable Importance

Un gráfico de barras horizontales que muestra la importancia relativa de cada término del modelo:

- **Términos suaves:** ordenados por el estadístico F (barras azules)

- **Términos paramétricos:** ordenados por el valor $|t|$ (barras verdes)

Debajo del gráfico, una DataTable muestra el término, el tipo, el EDF, el estadístico y el valor p, ordenados por estadístico descendente.

9.5 Gráficos de contribución

Todo término del modelo que contribuye al valor predicho tiene un gráfico de contribución interactivo en **plotly**, mostrado en una tarjeta con borde. La pestaña muestra gráficos para todos los tipos de término:

9.5.1. Términos suaves univariantes — $s(x)$

- **Línea azul** con banda de confianza del 95 %
- **Puntos grises** que muestran los residuos parciales (en los modelos sin transformación logarítmica)
- **Líneas verticales rojas discontinuas** en las posiciones de los nudos de earth (si se ha importado earth)
- **Ventana emergente al pasar el cursor:** valor de la variable, contribución (\$), IC del 95 %, tasa (pendiente por unidad)

En los **modelos con transformación logarítmica**, el eje y se retransforma a contribuciones en dólares mediante la fórmula $\bar{y} \times (e^{f(x)} - 1)$, lo que hace que las curvas sean directamente interpretables en términos monetarios.

Para las variables de **latitud y longitud**, la pendiente se muestra por cada 0.001 grados en lugar de por grado, ya que ese es un incremento geográfico más significativo.

9.5.2. Términos factor por término suave — $s(x, \text{by} = \text{factor})$

Una línea de color por cada nivel del factor, superpuestas en un mismo gráfico. Útil para ver cómo varía el efecto de un predictor entre grupos (por ejemplo, en distintos vecindarios).

9.5.3. Términos de interacción — $ti(x_1, x_2)$ y $te(x_1, x_2)$

Mapa de calor que muestra la superficie de contribución en 2D. Escala de color: rojo (negativo) → blanco (cero) → azul (positivo). Al pasar el cursor se muestran los valores de ambas variables y el importe de la contribución.

9.5.4. Términos paramétricos: lineales y de factor

- **Términos lineales numéricos:** diagrama de dispersión de los datos con la recta ajustada superpuesta
- **Términos de factor:** gráfico de barras con la contribución media por nivel del factor

9.6 Correlation

Una matriz en mapa de calor que muestra las correlaciones de Pearson entre todos los predictores numéricos. Disponible inmediatamente después de importar los datos — no requiere ajustar el modelo.

- Gradiente de color: azul (-1) → blanco (0) → rojo $(+1)$
- Los valores de correlación se imprimen en cada celda
- Útil para identificar la concurvidad (el análogo suave de la multicolinealidad) antes del ajuste

9.7 Diagnostics

Dos conjuntos de visualizaciones de diagnóstico:

Panel de diagnóstico (mediante `gratia::appraise`): cuatro gráficos — residuos frente a ajustados, gráfico Q-Q, histograma de los residuos y respuesta frente a valores ajustados.

Diagrama de dispersión **Actual vs Predicted**: valores observados frente a predichos con una línea de referencia discontinua a 45 grados. En los modelos con transformación logarítmica, los valores se retransforman a la escala original.

9.8 RCA Adjustments

Tres histogramas que se muestran tras el cálculo del RCA (solo en los modos de tasación y de mercado):

- **Residual Adjustment %** — distribución de los ajustes por residuo como porcentaje del precio de venta
- **Net Adjustment %** — distribución de los ajustes netos
- **Gross Adjustment %** — distribución de los ajustes brutos

Cada histograma muestra la media, la mediana y la desviación típica en el subtítulo, con líneas de referencia discontinuas para la media y la mediana.

9.9 Anova

Una tabla ANOVA combinada que fusiona los términos paramétricos y suaves del resumen del GAM. Muestra el término, el tipo (paramétrico o suave) y todos los estadísticos asociados.

9.10 Mgcvc Output

Salida de texto sin formato que muestra:

- Los tiempos del ajuste
- La fórmula del GAM
- La salida de impresión del modelo
- La salida completa de `summary(model)`
- La familia y la función de enlace
- Las medidas globales de concurvidad

9.11 Sign Check

Cuando se importan nudos de earthUI, esta pestaña compara la dirección de cada variable en el modelo de earth con su dirección en el GAM:

- **earth_direction**: creciente, decreciente o mixta (según los signos de las funciones bisagra)
- **gam_direction**: creciente, decreciente o mixta (estimada a partir de la derivada del término suave sobre una malla de 100 puntos)
- **consistent**: TRUE/FALSE
- **warning**: descripción de cualquier discrepancia

Una línea de estado muestra “All directions consistent” (todas las direcciones son coherentes) o el número de variables con incoherencias.

9.12 Concurvity

La concurvidad es el análogo suave de la multicolinealidad. Esta pestaña muestra:

- **Tabla Overall Concurvity:** filas para las medidas worst, observed y estimate, y columnas para cada término suave. Con código de color: blanco (<0.5), amarillo ($0.5-0.8$), rojo (>0.8).
- **Tabla Pairwise Concurvity:** matriz por pares del peor caso, con el mismo código de color.

Los valores superiores a ~ 0.8 en la fila “worst” indican que un término suave puede aproximarse bien mediante los demás términos suaves, lo que sugiere redundancia.

9.13 Basis Check

Salida de `mgcv::gam.check()`: pruebas de adecuación de la dimensión de la base para cada término suave. Si la k de un término suave es demasiado pequeña, esta prueba lo señalará con un valor p bajo, lo que sugiere que debería aumentar k .

10 Descarga de datos

Tras el ajuste, el botón **Download Output (Excel)** (sección 7 de la barra lateral) exporta un archivo de Excel con las predicciones y los diagnósticos a la carpeta de salida de mgcv del proyecto activo.

10.1 Columnas de salida

Columna	Descripción
<code>basis</code>	Valor del intercepto retransformado (igual para todas las filas)
<code><var>_contribution</code>	Contribución de cada término suave a la predicción. En los modelos logarítmicos, se asigna proporcionalmente a valores monetarios.
<code>est_<target></code>	Predicción del modelo (por ejemplo, <code>est_sale_price</code>), retransformada desde la escala logarítmica si procede
<code>residual</code>	Valor real – predicho
<code>calc_residual</code>	Verificación: valor real – suma de las contribuciones (retransformada)
<code>cqa</code>	Puntuación de condición, calidad y atractivo (CQA), en escala 0–10
<code>residual_sf</code>	Residuo / superficie habitable (si se ha designado <code>living_area</code>)
<code>cqa_sf</code>	CQA calculada a partir de la ordenación por <code>residual_sf</code>

En el modo de tasación, la fila 1 (el sujeto) tiene `residual`, `cqa` y `cqa_sf` fijados en NA. Las filas se ordenan por `residual_sf` descendente cuando se ha designado una columna `living_area`.

Tras completarse correctamente aparece una marca de verificación blanca en el botón de descarga.

10.2 Puntuaciones CQA

La CQA ordena el residuo de cada fila frente a todos los demás en una escala de 0 a 10:

- **CQA alta (~9–10)**: se vendió por mucho más de lo predicho — probablemente condición, calidad o atractivo superiores
- **CQA baja (~0–1)**: se vendió por mucho menos de lo predicho — probablemente condición inferior o venta forzada
- **CQA ~5**: cerca de la mediana

11 Cálculos RCA y descarga

El flujo de trabajo RCA (Residual Constraint Approach) está disponible en los modos **Appraisal** y **Market Area Analysis**, después de ajustar el modelo. La propiedad sujeto debe estar en la fila 1.

11.1 Abrir el diálogo RCA

Haga clic en el botón **Calculate RCA Adjustments & Download** de la sección 8 de la barra lateral. Aparece un diálogo modal con:

- **Score type:** CQA o CQA per SF (si se ha designado `living_area`). La opción por pie cuadrado escala el residuo del sujeto según la superficie habitable.
- **Subject CQA Score:** admite valores de 0.00 a 10.00. El campo aparece inicialmente en blanco; cuando se ha importado un modelo de earthUI, el tipo de puntuación y el valor que usó en earthUI se trasladan desde los ajustes compartidos del proyecto.

Haga clic en **Generate** para calcular y descargar. El archivo de Excel se escribe en la carpeta de salida de mgcv del proyecto activo (`<os>_out_mgcgv`).

11.2 Interpolación de la puntuación CQA

1. Se ordenan las puntuaciones CQA y los residuos de los comparables
2. Una interpolación lineal (`approx()`) asigna a su valor de CQA un residuo
3. Valor del sujeto = predicción del modelo + residuo interpolado
4. Las filas con peso 0 se tratan como si tuvieran el mismo residuo que el sujeto

11.3 Columnas de salida

Columna	Descripción
<code>subject_value</code>	Predicción del modelo + residuo interpolado
<code>subject_cqa</code>	Puntuación CQA introducida por el usuario
<code><var>_adjustment</code>	Contribución del sujeto – contribución del comparable (en dólares)
<code>residual_adjustment</code>	Residuo del sujeto – residuo del comparable
<code>net_adjustments</code>	Suma de todos los ajustes
<code>gross_adjustments</code>	Suma de los valores absolutos de los ajustes
<code>residual_adj_pct</code>	Ajuste por residuo como % del precio de venta
<code>net_adj_pct</code>	Ajustes netos como % del precio de venta
<code>gross_adj_pct</code>	Ajustes brutos como % del precio de venta
<code>adjusted_sale_price</code>	Precio de venta real + ajustes netos

Tras un cálculo correcto aparece una marca de verificación blanca en el botón.

12 Cuadrícula de comparación de ventas

La Sales Comparison Grid (cuadrícula de comparación de ventas) solo está disponible en el modo **Appraisal**, después de calcular los ajustes RCA. Genera un libro de Excel con formato apto para incluirlo en informes de tasación.

12.1 Diálogo de selección de comparables

Al hacer clic en el botón **Generate Sales Grid & Download** se abre un diálogo modal que lista todas las ventas comparables. Los comparables se dividen en dos grupos:

- **Recommended comps** (premarcados): ajuste bruto $< 25\%$ del precio de venta, ordenados por antigüedad de la venta (los más recientes primero). Máximo 30 comparables.
- **Other comps** (sin marcar por omisión): los comparables restantes, ordenados por porcentaje de ajuste bruto.

Seleccione sus comparables y haga clic en **Generate Sales Grid** para crear el libro de Excel.

Tipos especiales obligatorios: deben designarse `contract_date` y `living_area`. **Tipos especiales recomendados:** `latitude`, `longitude` y `lot_size` — la cuadrícula funcionará sin ellos, pero algunos campos quedarán en blanco.

12.2 Disposición del libro

El libro generado contiene varias hojas, cada una con 3 comparables uno al lado del otro:

- **Filas de encabezado:** direcciones del sujeto y de los comparables, precios de venta, fechas de venta y días en el mercado
- **Filas agrupadas:** ubicación (latitud, longitud, zona), tamaño del terreno (superficie del lote, dimensiones del sitio) y edad (edad real, edad efectiva), con encabezados de grupo y subelementos
- **Filas de variables del modelo:** una fila por término suave con la contribución del sujeto, las contribuciones de los comparables y los importes de ajuste
- **Filas de características residuales:** celdas vacías y desbloqueadas para que el tasador introduzca características no captadas por el modelo (por ejemplo, condición, calidad, vistas)
- **Filas de resumen:** % de ajuste neto, % de ajuste bruto y precio de venta ajustado, con fórmulas de Excel activas

12.3 Protección de las hojas y celdas residuales

Cada hoja está protegida para evitar ediciones accidentales, pero las celdas de valor de las características residuales están explícitamente desbloqueadas. Esto permite que los tasadores introduzcan valores para características ausentes del modelo sin alterar los cálculos de ajuste basados en fórmulas.

13 Descarga de informes

La generación de informes es un flujo de trabajo Quarto de dos pasos, que sustituye al informe de un solo paso de las versiones anteriores.

13.1 Paso 1: generar el informe Quarto

El botón **Generate Quarto Report** (disponible una vez que el modelo está ajustado) escribe un paquete `.qmd` autocontenido en la carpeta de salida de `mgcv` del proyecto activo. El paquete contiene el código fuente de Quarto, los archivos de gráficos prerrenderizados (PNG y PDF), una carga de datos `report_data.rds` con todas las tablas y los metadatos del modelo, y una plantilla de estilo `reference.docx` para la salida en Word. Como el paquete es autocontenido, puede editar el código fuente `.qmd` o combinarlo con otros documentos Quarto antes de renderizarlo.

Los gráficos se prerrenderizan al generar el paquete; si falla un gráfico concreto, se escribe una imagen de marcador de posición y el resto del paquete se genera con normalidad.

13.2 Paso 2: convertir el informe Quarto

La sección **Convert Quarto Report** renderiza un archivo `.qmd` a uno o varios formatos de salida mediante `quarto::quarto_render()`. Por omisión toma el último paquete generado, pero el botón **Browse** permite seleccionar cualquier archivo `.qmd`. Hay tres formatos disponibles:

- **HTML** — documento HTML autocontenido. Funciona en todas las plataformas sin software adicional. Opción predeterminada recomendada.
- **Word (.docx)** — renderizado con la plantilla de estilo `reference.docx` incluida. Adecuado para edición y distribución.
- **PDF** — **requiere una instalación de LaTeX** (véanse los requisitos del sistema más abajo). Si no se detecta LaTeX, la opción de PDF se oculta automáticamente de las casillas de formato.

Los renderizadores heredados de un solo paso, `render_gam_report()` y `export_gam_docx()`, siguen disponibles para uso programático.

13.3 Contenido del informe

El informe generado incluye:

- **Dataset & Model:** archivo de datos, tamaño de la muestra, familia, método e información sobre el ajuste
- **Results Summary:** R^2 , R^2 de CV, desviación explicada, AIC y otras métricas de ajuste relacionadas
- **Model Equation:** la ecuación del modelo con notación f_i para los términos suaves y coeficientes explícitos para los términos lineales, además de las definiciones de las funciones suaves. Las ecuaciones largas se dividen en varias líneas en el PDF.
- **Tablas de resumen:** términos suaves (EDF, F, valor p) y términos paramétricos (estimación, error estándar, t, valor p)
- **Interacciones habilitadas:** una matriz de los términos de interacción tensorial y de factor por término suave incluidos en el modelo (se muestra solo cuando hay interacciones)
- **Variable Importance:** gráfico de barras ordenado por el estadístico F (términos suaves) o por $|t|$ (paramétricos)
- **Smooth Effects** (todos los tipos de término):
 - Gráficos de términos suaves univariantes (versiones estáticas en `ggplot2` de los gráficos interactivos)

- Gráficos multilínea de factor por término suave
- Mapas de calor de interacción (términos ti y te)
- Gráficos de términos paramétricos (dispersión para los numéricos, barras para los de factor)
- Mapa de calor de **correlaciones entre predictores**
- **Diagnostics**: diagnósticos de residuos y diagrama de dispersión de valores reales frente a predichos
- **mgcv Summary**: la salida de texto completa del modelo ajustado
- **Apéndice de notación**: una sección final sin numerar que explica la notación de las funciones de contribución al valor empleada en el informe

Tras una generación correcta aparece una marca de verificación blanca en el botón Generate Quarto Report.

14 Exportación de las funciones del modelo

mgcvUI puede exportar las funciones suaves del GAM ajustado como código autónomo en varios lenguajes. Cada término suave se evalúa sobre una malla de 200 puntos y se exporta como tabla de consulta con interpolación lineal, lo que permite usar el modelo fuera de R sin ninguna dependencia de mgcv.

En la versión actual, la exportación de funciones es programática. Las funciones auxiliares exportadas toman un objeto de resultado ajustado (producido por `fit_gam()`) y escriben los archivos que se describen a continuación:

- `generate_r_code()`, `generate_python_code()` y `generate_cpp_code()` — archivos de código fuente por lenguaje
- `export_functions_sqlite()` — base de datos SQLite con tablas de consulta
- `export_functions_zip()` — agrupa cualquier combinación de los cuatro formatos en un archivo zip

14.1 Funciones de R

Archivo: `gam_functions.R`

- Comentarios de cabecera con la variable de respuesta, el R^2 , el R^2 de CV y la fórmula.
- Una constante `INTERCEPT`.
- Una `approxfun()` por cada variable suave, con vectores `x/y` precalculados.
- Los coeficientes de los términos lineales como constantes con nombre.
- Uso: sumar el intercepto, cada `g_var(value)` y cada `coef_var * value`.

14.2 Funciones de Python

Archivo: `gam_functions.py`

- Importación de `numpy`.
- Constante `INTERCEPT`.
- Por variable: arreglos `_x_var` y `_fx_var`, y función `g_var(x)` que usa `np.interp()`.
- Los coeficientes de los términos lineales como constantes `COEF_*`.

14.3 Cabecera de C++

Archivo: `gam_functions.hpp`

- Solo cabecera, con `#pragma once`.
- Espacio de nombres `gam_functions`.
- Función auxiliar `interp()` basada en `std::lower_bound`.
- Por variable: función inline con arreglos `static const`.
- Los coeficientes de los términos lineales como `constexpr double`.

14.4 Base de datos SQLite

Archivo: `gam_functions.sqlite`

- Tabla `smooth_grids`: columnas (`variable`, `x`, `fx`) — las tablas de consulta de 200 puntos.
- Tabla `model_info`: `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.
- Tabla `linear_terms`: `variable`, `coeficiente`.
- Tabla `intercept`: valor del intercepto.

15 Comparación con earthUI y glmnetUI

La aplicación earth y sus Post Processing Modules forman un único conjunto de herramientas para el modelado de regresión. Comparten el mismo formato de datos, los mismos tipos especiales de columna, el mismo flujo de trabajo RCA y los mismos conjuntos de datos de demostración, pero usan motores de modelado distintos.

15.1 Diferencias principales

Característica	mgcvUI	earthUI	glmnetUI
Método	GAM (mgcv)	MARS/Earth (earth)	Elastic net (glmnet)
Relaciones	No lineales suaves	Lineales por tramos	Lineales
Selección de variables	Contracción con <code>select=TRUE</code>	Poda por pasos	Penalización Lasso
Interacciones	Productos tensoriales (te/ti)	Productos de bisagras (grado 2-3)	Productos por pares
Efectos parciales	Curvas suaves con IC	Funciones g (por tramos)	Pendientes lineales
Cadena con earth	Importa los nudos de earth	Fuente de los nudos	Importa la base de bisagras
Exportación de código	R, Python, C++, SQLite	R, Python, C++, SQLite	—
Formatos de informe	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

Conviene señalar un elemento diferenciador exclusivo de earthUI: earthUI incluye una función experimental de Prolog/DCG que extrae atributos estructurados de las propiedades a partir de las observaciones de texto libre del MLS.

15.2 Funciones compartidas

Las tres herramientas ofrecen:

- Un árbol de proyectos **regProj** compartido (por omisión `~/regProj`) — los proyectos, los archivos de entrada y las salidas de cada método se comparten entre las aplicaciones hermanas
- Importación de CSV y Excel con conversión de los nombres de columna a `snake_case`
- Compatibilidad con la configuración regional de más de 30 países, con separador de CSV, signo decimal y tamaño de papel
- Designaciones especiales de columnas (`contract_date`, `living_area`, `area`, `latitude`, `longitude`, etc.)
- Cálculo automático de `sale_age` a partir de la fecha efectiva
- Alternancia entre modo oscuro y claro (tema Nord)
- Persistencia de los ajustes (SQLite y/o almacenamiento local)
- Mapa de calor de correlaciones, importancia de las variables y diagnósticos
- Puntuación CQA y flujo de trabajo de ajustes RCA
- Sales Comparison Grid con selección de comparables y salida a Excel con formato
- Compatibilidad con los mismos conjuntos de datos de demostración (`Appraisal_1.csv`)

15.3 Flujo de trabajo recomendado

- **Empiece con earthUI** para descubrir las variables importantes, las relaciones no lineales y las interacciones. La construcción automática de modelos de earth ofrece el mejor punto de partida.
- **Refine con mgcvUI** si desea efectos parciales más suaves. Importe los nudos de earthUI para una transición sin fricción.
- **Use glmnetUI** cuando necesite un modelo puramente lineal con selección agresiva de variables o restricciones de signo en los coeficientes.
- **Compare las tres** sobre los mismos datos para valorar si está justificada la complejidad adicional de los modelos no lineales.

16 Conjunto de datos de demostración: Appraisal_1.csv

16.1 Descripción

El conjunto de datos de demostración se comparte con earthUI y glmnetUI. Se encuentra en la carpeta `demo_mls/` del código fuente de mgcvUI o, si earthUI está instalado:

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

El archivo contiene aproximadamente 1.500 registros de ventas residenciales procedentes de una exportación simulada de un MLS, con la propiedad sujeto en la fila 1 y las ventas comparables en las filas restantes. Los datos representan ventas de viviendas unifamiliares en un mercado con varias zonas y con una diversidad de tamaños, edades y ubicaciones de las propiedades.

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

16.2 Columnas

Columna	Tipo	Tipo especial	Descripción
sale_type	character	sale_type	Código de tipo de registro que distingue el sujeto de las ventas comparables
weight	numeric	weight	Peso de la observación (0 = excluir del ajuste)
id	numeric	display_only	Identificador interno del registro
property_id	numeric	display_only	Identificador de propiedad del MLS
listing_id	character	display_only	Número de listado del MLS
parcel_number	character	display_only	Número de parcela del catastro del condado (APN)
street_address	character	display_only	Dirección de la propiedad
city_name	character	display_only	Ciudad
postal_code	character	display_only	Código postal
county_name	character	display_only	Condado
contract_date	Date	contract_date	Fecha de contrato de venta (calcula sale_age)
sale_age	numeric	—	Días desde la fecha de contrato hasta la fecha efectiva (precalculado)
coe_date	Date	display_only	Fecha de cierre del depósito en garantía
listing_status	character	display_only	Estado del listado (por ejemplo, "Sold")
sale_price	numeric	(objetivo)	Precio de venta — variable de respuesta
rent	numeric	—	Renta mensual (para modelos con varios objetivos)
list_price	numeric	display_only	Precio de listado
original_list_price	numeric	display_only	Precio de listado original
living_sqft	numeric	living_area	Superficie habitable bruta en pies cuadrados
ppsf	numeric	display_only	Precio por pie cuadrado
beds_total	integer	—	Número de dormitorios
baths_total	numeric	—	Número total de baños (por ejemplo, 2.5 = 2 completos + 1 medio)
lot_size	numeric	lot_size	Tamaño del terreno en pies cuadrados
area_id	integer	area	Identificador de zona del MLS
area_text	character	display_only	Nombre de la zona
age	numeric	actual_age	Edad de la propiedad 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

Columna	Tipo	Tipo especial	Descripción
no_of_stories	numeric	—	Número de plantas
style	character	—	Estilo arquitectónico
view	character	—	Tipo de vistas (por ejemplo, “Neighborhood”, “Hills”)
days_on_market	integer	dom	Días en el mercado
listing_date	Date	listing_date	Fecha de listado
sale_concessions	numeric	concessions	Concesiones del vendedor

16.3 Inicio rápido sugerido

1. Inicie la rutina de mgcv: `earthUI::launch_mgcv()`
2. Cree un proyecto nuevo (Sección 1) con el propósito **For Appraisal**
3. Coloque **Appraisal_1.csv** en la carpeta de entrada del proyecto y selecciónelo en la Sección 2 (haga clic en **Refresh** si es necesario)
4. Seleccione **sale_price** como objetivo
5. Opcionalmente, fije Response Transform en **Log (natural)** para obtener ajustes proporcionales
6. Asigne los tipos especiales tal como se indica en la tabla anterior
7. Incluya los predictores **sale_age**, **living_sqft**, **baths_total**, **lot_size**, **area_id** (como factor), **age**, **latitude**, **longitude** y **garage_spaces**
8. Mantenga los valores predeterminados: REML, gamma = 1.2, base tp, k = auto, select = TRUE
9. Haga clic en **Fit Mgcv GAM Model**
10. Revise las pestañas Summary, Contribution (gráficos de términos suaves) y Diagnostics
11. Descargue la salida (paso 7) y revise la ordenación por CQA
12. Calcule los ajustes RCA (paso 8) con una puntuación CQA de ~5.00
13. Genere una Sales Comparison Grid (paso 9) con los comparables recomendados

17 Requisitos del sistema y resolución de problemas

17.1 Plataformas admitidas

mgcvUI funciona en **macOS**, **Windows** y **Linux**. La aplicación se desarrolla y se prueba principalmente en macOS con RStudio. Más abajo se ofrecen notas específicas de cada plataforma.

17.2 Versión de R

Se requiere $R \geq 4.1.0$. Se recomienda encarecidamente RStudio Desktop (2023.06 o posterior) — incluye **Quarto** y **pandoc** (necesarios para renderizar los informes) y ofrece un entorno cómodo para iniciar la aplicación.

17.3 Paquetes de R requeridos

Los siguientes paquetes se instalan automáticamente como dependencias:

Paquete	Finalidad
shiny	Marco de aplicaciones web
mgcv	Ajuste de los modelos GAM
valengrCore	Núcleo compartido de tasación (papeles especiales, traspaso desde earthUI)
ggplot2, plotly	Gráficos de contribución estáticos e interactivos
gratia	Extracción de términos suaves y diagnósticos
DT	Tablas de datos interactivas
quarto, rmarkdown	Generación y renderizado de informes Quarto
readxl	Importación de archivos de Excel
shinyFiles	Exploradores de archivos y carpetas
RSQLite, DBI	Bases de datos de ajustes y de proyectos (SQLite)
jsonlite	Serialización del estado de las variables

17.4 Dependencias opcionales

Componente	Cuándo se necesita
LaTeX (TinyTeX, MiKTeX o MacTeX)	Solo para generar informes en PDF. Si no se detecta ninguna instalación de LaTeX, la opción de PDF se oculta automáticamente del desplegable de formatos. Instálelo con: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	Fuente Roboto Condensed para los gráficos de ggplot2. Si no está disponible o se trabaja sin conexión, la aplicación recurre automáticamente a la fuente de palo seco del sistema.
earth ($\geq 5.3.0$)	Solo se necesita para la cadena de importación desde earthUI. No es necesario para el uso autónomo.
callr	Ajuste del modelo en segundo plano (mantiene la aplicación receptiva). Sin él, el ajuste se ejecuta de forma síncrona.
writexl, openxlsx	Descargas a Excel (salida del modelo, RCA, cuadrícula de ventas). La aplicación lo indica si faltan.
bslib	Temas de Bootstrap 5 (Nord Light/Dark).
officer	Exportación programática heredada a Word (<code>export_gam_docx()</code>).

17.5 Notas específicas de cada plataforma

17.5.1. macOS

La plataforma con menos problemas. Se recomienda Homebrew para las bibliotecas del sistema. Si se necesitan informes en PDF:

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

17.5.2. Windows

Funciona bien con RStudio Desktop. Notas principales:

- **Equipos corporativos o restringidos:** si el directorio temporal o el directorio de datos del usuario están restringidos, la persistencia de los ajustes y la generación de informes pueden fallar. La aplicación mostrará una advertencia en la consola de R pero seguirá funcionando sin persistencia.
- **LaTeX:** instale TinyTeX desde R (el mismo comando que en macOS más arriba) o instale MiKTeX desde miktex.org.
- **Rutas de archivo largas:** Windows tiene un límite de 260 caracteres para las rutas. Mantenga corta la ruta de su directorio de trabajo.

17.5.3. Linux

La plataforma más variable entre distribuciones. Los usuarios de Ubuntu o Debian pueden necesitar bibliotecas del sistema antes de que los paquetes de R compilen:

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

Para los informes en PDF: `tinytex::install_tinytex()` o instale `texlive-full` desde el gestor de paquetes del sistema.

En servidores sin interfaz gráfica, los gráficos interactivos de plotly se renderizan como alternativas estáticas en los informes. La propia aplicación Shiny requiere una conexión desde un navegador web.

17.6 Degradación elegante

mgcvUI está diseñado para funcionar incluso cuando faltan componentes opcionales:

Componente ausente	Comportamiento
Sin LaTeX	La opción de PDF se oculta del desplegable de formatos de informe. HTML y Word siguen disponibles.
Sin internet o fallo de las fuentes	La fuente Roboto Condensed se sustituye por la de palo seco del sistema. Se registra un mensaje en la consola.
Sin RSQLite o sistema de archivos de solo lectura	Los ajustes no persisten entre sesiones. Se muestra una advertencia en la consola. Por lo demás, la aplicación funciona con normalidad.
Fallo de <code>gratia::appraise</code>	La pestaña Diagnostics y los informes muestran un mensaje de marcador de posición. Use <code>mgcv::gam.check()</code> en la consola como alternativa.
Directorio temporal sin permiso de escritura	La generación del informe falla con un mensaje de error claro. Se resuelve fijando la variable de entorno <code>TMPDIR</code> en una ubicación con permiso de escritura.

17.7 Resolución de problemas

17.7.1. “PDF option is not available”

No se ha detectado ninguna instalación de LaTeX. Ejecute en R:

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

Después reinicie la aplicación. La opción de PDF aparecerá.

17.7.2. “Settings will not persist”

No se puede crear el directorio de la base de datos SQLite ni escribir en él. Compruebe los permisos de:

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

17.7.3. “Diagnostics plot unavailable”

La función `gratia::appraise()` falla ocasionalmente con modelos complejos (muchas interacciones tensoriales, familias no estándar). Se trata de un problema conocido del paquete original. Solución alternativa: ejecute `mgcv::gam.check(model)` en la consola de R para ver directamente los gráficos de diagnóstico.

17.7.4. “Report error: number of items to replace...”

Este era un problema conocido de las versiones anteriores, relacionado con el paquete **gratia**. Asegúrese de tener instalada la última versión de **mgcvUI**. La versión actual captura este error y produce el informe con un marcador de posición en la sección de diagnósticos.

17.7.5. El puerto 7880 ya está en uso

Si la aplicación no arranca porque el puerto 7880 está ocupado (por una sesión anterior), ejecute:

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

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

17.7.6. La instalación de paquetes falla en Linux

Faltan bibliotecas del sistema. Instale las cabeceras de desarrollo enumeradas en la sección de Linux más arriba y vuelva a intentar `install.packages()`.

Referencias

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